UNIVERSITY OF SOUTHAMPTON

FACULTY OF LAW, ARTS AND SOCIAL SCIENCES

School of Social Sciences

Modelling Complex Longitudinal Survey Data

by

Marcel de Toledo Vieira

Thesis for the degree of Doctor of Philosophy

September 2005

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF LAW, ARTS AND SOCIAL SCIENCES

SCHOOL OF SOCIAL SCIENCES

Doctor of Philosophy

MODELLING COMPLEX LONGITUDINAL SURVEY DATA

by Marcel de Toledo Vieira

Modelling methods for longitudinal complex survey data are investigated in this thesis.

An empirical investigation using longitudinal survey data is conducted. Variance effects of clustering are identified and results indicate that clustering impacts may be stronger for longitudinal studies than for cross-sectional studies. Earlier empirical evidence that those impacts could be less the more complex the analysis, which may sometimes be used to justify ignoring the complex sampling scheme in longitudinal analysis, is thus contradicted. A theoretical discussion is provided in order to support the major empirical results. The considered longitudinal regression modelling methods are reviewed in the complex survey context.

The adoption of covariance structure models for longitudinal survey data is emphasised in this dissertation as this approach includes a wide range of modelling techniques and has application in the social sciences. A weighted estimation procedure (S_w), which considers covariates, is proposed for estimating the population covariance matrix Σ. Further developments on variance estimation methods for $\hat{\Sigma}$ considering the complex survey approach are accomplished by adopting a Taylor expansion technique in order to extend asymptotically distribution-free (ADF) methods. By adopting S_w, modifications to point estimation methods, such as unweighted (ULS) and generalised least squares (GLS), for a vector parameter are also proposed. A pseudo maximum likelihood (PML) for covariance structure models is also derived via maximisation of the pseudo log likelihood function. The behaviour of the proposed estimation procedures are assessed by simulation.

ADF variance estimation methodology for GLS point estimators is extended. A method for estimating the asymptotic covariance matrix of the PML point estimator is proposed under the complex survey data approach. Some extensions to model fitting statistics when working with longitudinal data in a complex survey design framework are developed. We propose modifying the Wald goodness of fit test in the context of models for covariance structures, which is shown to be equivalent to modifying the scaled test statistics. Furthermore, we also propose a modification for the Wald significance test for nested hypothesis. Goodness of fit indices are also modified in order be utilised in the complex survey data context. An additional simulation study is adopted for evaluating the proposed variance estimation methods.

Contents

List of Tables

List of Figures

Acknowledgements

I, first of all, thank God for His love and care, for all the strength, courage, patience and determination He has given me.

I am deeply indebted to my supervisor, Professor Chris Skinner, for his encouragement and for the high quality of his academic supervision through out this project. His advice has been absolutely essential for the progress of my research.

The friendly and outstanding academic environment in the Division of Social Statistics and S3RI has been also determining to the development of my activities. I am very gratefull to Professor Danny Pfefferman and Professor Ray Chambers for their comments. Acknowledgement is also due to Dr. Ann Berrington for her support in preparing the BHPS data set. A special thank to all Research Office staff and to Dr. Sandy Mackinnon and Darren Hampton for their support. I am indebted to my fellow colleagues for their friendship and support, especially Gail, Nikolaos, Ebrahim, Amos, Gabriele, Fátima, Leslie, Leonardo, and Solange.

I am indebted to Dr. Pedro Luis Nascimento Silva for introducing me to Survey Sampling. In particular, without his enormous support and encouragement I would not have come to Southampton. I am also grateful to Dr. Ronaldo Rocha Bastos, Jane Azevedo da Silva, and Dr. Tufi Machado Soares for introducing me to Statistics.

I want also to thank Fr. Michael Ryan and Sr. Mairead Noonan for their immense support, and all my friends from CathSoc and University Chaplaincy for their friendship, especially Robert and Clair. Life in England was considerably easier also because of Henrique and family, Geraldo and Ana Maria, and John and Margaret.

I owe the deepest gratitude to my adored wife Anneliese for being always at my side especially over the last four years. I thank her for the patience, care, dedication and unconditional love. Without her I would have never been able to finish this thesis. I thank my parents José Celso e Maria Fátima for their love, and for having always encouraged me for academic education.

Finally, I dedicate this thesis to my beloved son Daniel born in Southampton, and to the memories of my brother in law Klaus and my uncle José Luis.

This research was supported by grant number 200286/2001.3 from the National Council for Scientific and Technological Development (CNPq) from the Brazilian Ministry of Science and Technology. The Division of Social Statistics sponsored my participation in most of the conferences and extra courses attended while undertaking my studies.

Chapter 1

Introduction

1.1 Longitudinal surveys

Sample surveys have been a substantial source of human and institutional data for both descriptive and analytic use, since the XIX century (Skinner, Holt and Smith, 1989). The collection of longitudinal survey data has become more purposeful, and longitudinal research has become more valued by the governments and the scientific community, after the World War II (Menard, 1991)[1].

Survey data has never been as indispensable as it is for XXI century society, especially because this type of data is the main source of information when regarding demographic and social characteristics of the population, economic activity, lifestyle patterns, and public opinion (Barnett, 1991). Furthermore both everyday life decisions and scientific research in different subject areas are based to a great extent on samples (Cochran, 1977).

Longitudinal survey data allow, for example, the periodic measurement of individual's demographic and socio-economic changes in their conditions (Berrington, 2002). We may briefly describe two different classes of longitudinal studies: (*i*) repeated cross-sections studies, where data is collected repeatedly through time with the same variables being measured on different samples of cases (see Menard, 2002; Kalton and Citro, 2000; and Firebaugh, 1997); (*ii*) panel studies[2] (or repeated measures data studies), where the same and (or) different variables are measured on the same units at least at two time points. Methodology discussed in this thesis shall be appropriate mainly to analysing repeated measures data.

> *"The term panel data refers to the pooling of observations on a cross-section of households, countries, firms, etc. over several time periods. This*

---

[1] Periodic collection of census data has been conducted at the national level since the XVII (Menard, 2002).
[2] Terminology usually adopted by economists and sociologists.

can be achieved by surveying a number of households or individuals and following them over time" (Baltagi, 2001).

Panel data is particularly adequate for investigating changes at the individual level. Furthermore longitudinal studies also allow us to distinguish the degree of variation in the response variable across time for one person from the variation among subjects, and in principle also to make stronger causal interpretations (Diggle *et al.*, 2002) mainly regarding inferences about changes, by determining the direction and magnitude of causal relationships (Menard, 1991). Repeated measures data analysis techniques may be adopted for disentangling persistent from transient effects and for controlling for individuals' pasts when evaluating effects (Duncan, 2000). Furthermore, panel data is capable, for example, of providing measures before and after important social policy events. See Rose (2000b) for a comprehensive review on the characteristics of household panel studies.

In panel data studies, data is obtained *prospectively*, i.e. individuals are followed advanced in time. Nevertheless, longitudinal data on each individual may also be obtained *retrospectively*, i.e. by selecting a cross-sectional sample and obtaining data by inquiring retrospective questions. According to Skinner (2000), common inaccuracies of retrospective measurements are: (*i*) under reporting, i.e. failure to declare events; (*ii*) over reporting, i.e. reporting of episodes which did not happen; (*iii*) 'telescoping', i.e. susceptibility to over or under estimate time since event; and (*iv*) 'heaping', i.e. tendency to round time since episode to round number. Moreover, respondents may understate change within the reference time interval.

Examples of a combination of both prospective and retrospective measurements are the cohort studies, which are longitudinal studies that involve the selection of a sample of individuals born within a specific period of time and following them over time. It is frequent in this type of study to collect information on events since the previous wave.

Some examples of longitudinal studies carried out in Great Britain are the British Household Panel Survey (BHPS; see Taylor *et al.*, 2001), the Youth Cohort Study (YCS, see Courtenay, 1997), the Office for National Statistics (ONS) Longitudinal Study (LS, see Hattersley and Creeser, 1995), and the English Longitudinal Study of Ageing (ELSA, see Marmot *et al.*, 2003). See Rose (2000b) for various applications involving the analysis of social and economic longitudinal survey data.

A very brief overview of models appropriate for the analysis of longitudinal data shall be presented in Section 1.2. Section 1.3 shall provide some consideration on methodological issues related to sampling in longitudinal surveys. Some important aspects related to statistical inference, including *finite population* and *superpopulation* models, shall be discussed in Section 1.4. The motivation and aims for this research project are summarized in Section 1.5. Finally, Section 1.6 presents concisely the outline of this thesis.

1.2 Models for the analysis of longitudinal survey data

Let y denote the *study variable*, or *survey variable*, which is quantified in the survey and is relevant in the analysis. As in Chambers and Skinner (2003, Part D) and most of the longitudinal data literature, in this thesis we shall adopt i to denote an individual and t to denote time. In this context, we thus denote the *survey variable of interest* as y_{it} for individual i at time t.

In this thesis we shall be concerned mainly with the analytic use of longitudinal surveys[3]. Hence, several different modelling techniques could be adopted when analysing longitudinal survey data. We shall present in the current section a short overview of models appropriate for the analysis of longitudinal data.

The interest for fitting models to longitudinal complex survey data has been growing in the last few years (see Feder, Nathan and Pfeffermann, 2000). Methods for modelling longitudinal data have to consider the variation in y across the population as well as across time (Chambers and Skinner, 2003). Consequently specific statistical techniques are required when analyzing this type of data, as we should consider the inter-correlation among observations on one subject.

One factor that could guide one's decision on the adoption of a specific modelling technique is the nature of the data to be analysed, for example whether the response variable of interest is continuous or discrete. Furthermore, different modelling approaches would also be chosen depending on whether continuous (with t taking any value in a given interval) or discrete time data is available (note that in this thesis we shall only consider

[3] See Skinner, Holt and Smith (1989, Chapter 1, Section 1.1) for a discussion on the distinction between analytic and descriptive uses of sample surveys.

the discrete time case, as $t = 1,\ldots,T$, where T is the number of waves of the survey and t are equally spaced in time).

Perhaps the simplest approach to be adopted when modelling repeated measures data is the so-called two-stage or derived variable strategy (see for example Diggle *et al.*, 2002, Chapter 1). This technique is based on reducing the repeated values of the response variable into one or two summaries and analysing each summary variable as a function of the covariates. Nevertheless, this approach is limited to time-constant covariates.

Alternatively, we may choose to model response variable individual values in terms of time-varying (and/or time-constant) covariates. In this context, we could model the marginal mean similarly to the cross-sectional data modelling case. However, repeated observations on the same variable are likely to be correlated and such correlation has to be considered in this marginal analysis. Moreover, when adopting this modelling approach we would have the capability of separately modelling the mean and the covariance. This class on longitudinal regression model shall be discussed in more details in the Chapter 2 of this thesis. For further information see, for example, Diggle *et al.* (2002).

An approach alternative to the marginal analysis is random effects modelling, which presupposes the correlation among repeated observations is caused by the fact that model coefficients vary across individuals. This class of longitudinal regression models is especially appropriate when we aim to produce inferences about individuals (Diggle *et al.*, 2002). A conditional version of the random effects modelling approach, i.e. fixed effects models, is also an approach that could be adopted when analysing longitudinal data. Skinner (2003b) considers, however, this specification to be unnatural to some extent from a survey sampling viewpoint, as this considers the inference being conditional on the effects in the sample. For additional information on random effects models for longitudinal data see, for example, Goldstein (1995, Chapter 6), and Hand and Crowder (1996). Note that both marginal models and random effects models are suitable for discrete time data.

Transition models could also be adopted when analysing longitudinal survey data. This kind of models involves the regression of the response variable of interest on its values at previous time point (or points) in addition to the covariates. When working with discrete response variables one could adopt logistic or multinomial logistic regression models, for example, for the analysis of transitions. For supplementary material and application involving transition models see, for example, Mealli and Pudney (2003), Diggle *et al.* (2002), Skinner (2000), and Ermisch (2000).

Event history analysis (EHA) is also an alternative methodology for performing analysis of transitions, as (*i*) an event may be described as a transition between states and (*ii*) a transition may be interpreted as a type of event. Techniques of EHA deal with intervals of time which units spend in the different categories of the response variable (Skinner, 2003b). EHA can be carried out on either discrete or continuous time recorded longitudinal data. In most situations, discrete time methods give very similar results to continuous time methods, although in practical terms the latter are more often adopted. In discrete time event history, models may be fitted via standard logistic regression techniques. For further information on the analysis of event history data see, for example, Allison (1984) and Lawless (2003).

When analysing one-way transition data the term survival analysis is usually adopted. Basic survival analysis models are then special cases of EHA and consider the time of occurrence of a single terminal event as the outcome of interest.

Furthermore, the characteristics of longitudinal survey data adapts adequately with the adoption of structural equation modelling (or covariance structure analysis[4]). Wiley, Schmidt and Bramble (1973) define the covariance structure analysis as the group of models and statistical methods (including multivariate analysis of variance, regression and factor analysis), which are employed for the structural analysis of covariance matrices predominantly in the social sciences. For further information on structural equation modelling, see for example, Bollen (1989), Raykov and Marcoulides (2000), Loehlin (1987), Finkel (1995), Long (1983), Bentler and Weeks (1980), Browne (1982), Schoenberg (1989), Bentler and Dudgeon (1996), and MacCallum and Austin (2000). We shall discuss methods for covariance structure modelling from Chapter 4 and beyond.

---

[4] According to Long (1983), this statistical modelling approach is also referred to as linear structural relations modelling, moments structure modelling, latent variable equation systems in structured linear models, and LISREL modelling.

1.3 Sampling in longitudinal surveys

The *sampling scheme* used by the survey designer to select the sample is one of the crucial features that have to be considered in survey data analysis. In this section we give some consideration on methodological issues surrounding sampling in longitudinal surveys.

1.3.1 Sampling designs

We first consider the cross-sectional case.

The *target population* is the *finite population* about which we demand information. Moreover it shall be denoted by $\mathcal{U}$ with N distinct elements, which are labelled so that

$$\mathcal{U} = \{1,2,\ldots,N\},$$

without loss of generality.

Let $\mathcal{S} = \{1,\ldots,n\}$ represent a sample, where $\mathcal{S}$ is a subset of $\mathcal{U}$, n is the sample size, and $1 \leq n < N$. For each sampled unit, the values of a set of *study variable*s are observed. Let the function $p(\mathcal{S})$ represent a *sampling scheme*, defined for every $\mathcal{S}$.

The *sampling scheme* of a survey may be defined as the procedure adopted to select a sample $\mathcal{S}$ from $\mathcal{U}$ (see, for example, Nascimento Silva, 1996, Section 2.3). Let ζ represent the set of all possible samples (or subsets of $\mathcal{U}$).

Let s denote the actual selected sample and $\Pr(\mathcal{S}=s)=p(s)$, where the operator Pr(.) denotes the *probability of*. The *sampling scheme* $p(\mathcal{S})$ has the following attributes,

$$0 \leq p(\mathcal{S}) < 1 \text{ for all } \mathcal{S} \subset \zeta,$$

and

$$\sum_{s} p(\mathcal{S}) = 1.$$

Let w_i denote *sampling* (or *probability*) *weights*, which could be taken to be reciprocals of the inclusion probability, i.e. $w_i = 1/\pi_i$, where

$$\pi_i = \Pr(i \in s) = \sum_{s \ni i} p(s)$$

is the probability of selecting the ith individual in the sample s, which is assumed to be known. A sample scheme is a probability sample scheme if every element of $\mathcal{U}$ has a chance to be selected in s, i.e.

$\pi_i > 0$, all $i \in U$.

Let $E_p(.)$ denote the expectation with respect to the sampling distribution of statistics over repeated samples s generated by the sampling design $p(s)$. Moreover, let z_i be any variable, which does not depend upon the choice of the sample s. We may now consider the following general results, (Isaki and Fuller, 1982)

$$E_p\left(\sum_{i=1}^{n} w_i z_i\right) = E_p\left(\sum_{i=1}^{n} \pi_i^{-1} z_i\right) = \sum_{i=1}^{N} z_i ,$$

and (1.1)

$$p\lim_{n\to\infty}\left[\frac{\sum_{i=1}^{n} w_i z_i}{\sum_{i=1}^{N} z_i}\right] = 1,$$

which shows the design unbiasedness and consistency of the Horvitz-Thompson estimator (Horvitz and Thompson, 1952). In (1.1), $p\lim$ denotes *probability limit of.* This result shall be very useful and shall be referred to later in this thesis. Note that in this dissertation the model expectation[5] shall be denoted by $E(.)$.

Sampling schemes with a less variable set of weights are often desired because weights with high variability could cause the increase in weighted estimators' variability (Korn and Graubard, 1995).

In most social surveys populations, characteristics such as indicator variables, strata and clusters, which define subgroups and are normally used by the survey designer, may be referred as *design variables* (see Chambers and Skinner, 2003, Section 1.2).

Examples of *sampling schemes* include stratification, clustering and multistage sampling. Primary sampling units (PSUs), could be selected within each stratum using either equal or unequal probability sampling, with or without replacement. From each PSU, secondary sampling units (SSU) could be selected using either equal or unequal probability sampling with or without replacement, and so forth in the presence of subsequent sampling stages.

We may additionally classify *sampling designs* as informative when they depend directly upon the *study variable* (see, for example, Chambers, Dorfman and Sverchkov,

[5] The 'classic' model-based expectations do not take the randomisation due to the sample design into account.

2003, Section 11.2). Otherwise the *sampling scheme* could be categorized as non-informative.

1.3.2 Sampling designs for longitudinal surveys

There is usually a high cost associated[6] with repeated sampling of individuals over time when compared to a cross-sectional *sampling scheme*, although the analysis of longitudinal survey data has several methodological advantages (see Section 1.1). Sampling designs for prospective longitudinal surveys involve collecting initially a cross-sectional sample at wave one and then following its elements from wave two and beyond. The longitudinal features of the longitudinal survey have to be taken into account when designing the sample for the first wave (Kalton and Citro, 2000).

Let s_T denote longitudinal sample. In longitudinal studies, waves of data are collected usually equally spaced through time on the same individuals. Under these general circumstances, sample data would thus follow the form $\{y_{it}; i \in s_T, t = 1,\ldots,T\}$, when it is assumed no nonresponse (Skinner, 2003b). We additionally assume here that the population $\mathcal{U}$ and the sample are determined at the time of first wave of the survey, remaining fixed thereafter.

Sampling in longitudinal studies may become a more sophisticated issue than under a cross-sectional perspective, as it is fundamental that the longitudinal sample continues representative of the population (Rose, 2000a). Skinner (2003b) provides some further discussion on this issue, including the fact that (*i*) survey designers could face problems like attrition and other types of non-response[7]; moreover (*ii*) panel rotation may be adopted and new respondents may be allowed to join the survey[8]; furthermore (*iii*) tracing respondents[9] could become a potential cause for incomplete data, specially when we have long time windows among the waves; and (*iv*) additionally there are possible issues involving measurement error in longitudinal surveys. See also Skinner (2000) for further

[6] See Menard (2002) for further information on the costs of longitudinal research.
[7] Note that cumulative non-response as result of attrition is likely to be a more considerable issue in longitudinal surveys than in cross-sectional ones (Rose, 2000). The risk of panel attrition usually increases with the number of waves of data collection (Kalton and Citro, 2000). See Duncan (2000), Winkels and Withers (2000), and Menard (2002) for information on the analytical problems caused by attrition.
[8] Often 'births' that enter population may be sampled while co-habitants with sample members may also be included in the longitudinal study with the aim of updating the sample.
[9] See also Kalton and Citro (2000) for further information on tracking and tracing panel members.

information on measurement error in panel surveys, and Kalton and Citro (2000), and Duncan (2000), for additional material on other sources of non-sampling error.

In the context of longitudinal data, the *sampling weights* w_i may be adjusted, at each wave, for taking account of previous wave respondents' absence through refusal at the present wave, or through some other way of sample attrition. Thus, the essential procedure for calculating sampling weights for a panel survey is practically equivalent to that for a cross-sectional study although there are a numerous further complexities to be taken into account (Kalton and Brick, 2000). Hence the *longitudinal weight* at any wave generally account for losses between each immediate pair of waves up to that point and the initial design (Taylor *et al.*, 2001). In this thesis the adjusted *sampling weights* are called *longitudinal weights* and shall be denoted by w_{iT}^*, $i \in s$. See Kalton and Brick (2000), for further information on the many issues involved on the production of sampling weights in longitudinal surveys. Furthermore, for financial reasons, samples in most household longitudinal surveys are selected by a *complex design*, frequently involving multi-way stratification and multi-stage clustering and unequal probability sampling.

1.4 Inference in the presence of complex sampling

1.4.1 Finite populations and superpopulation models

It is usually assumed that $\mathcal{U}$ is fixed. Nevertheless, in the longitudinal survey framework, we could concede the population to change over time.

Recall that y denotes the *study variable*. In this thesis we shall assume that there is neither non-response nor measurement error, i.e. we suppose that y is accounted properly for all sample units.

We may define a *finite population* parameter as any descriptive function of the values $y_1, \ldots, y_N$, which are the population values of y.

Under the *superpopulation* model the values $y_1, \ldots, y_N$ are assumed to be a joint realisation of the random vectors $Y_1, \ldots, Y_N$ (Skinner, Holt and Smith, 1989). We may denote the joint distribution of $Y_1,\ldots,Y_N$ by ξ. *Superpopulation* parameters are characteristics of ξ, which may also be seen as a model for the N-dimensional distribution of $Y_1,\ldots,Y_N$ (Särndal, Swenson and Wretman, 1992).

Superpopulation modelling may be briefly described as the exercise one could adopt for specifying the characteristics of the mechanism considered to have generated $y_1,\ldots,y_N$ (Särndal, Swenson and Wretman, 1992). This shall be the approach adopted in this thesis.

We shall denote both *finite population* and *superpopulation* parameters by Greek letters. In order to make a distinction we shall add a subscript N to the *finite population* (or *census*) parameters.

Some further notation and essential terminology, which are necessary for later chapters, are also set out here. Estimators of the unknown *superpopulation* characteristics, for example a parameter vector $\underline{\theta}$, shall be denoted by $\hat{\underline{\theta}}$ when calculated from sample results. Therefore, the *hat* in $\hat{\underline{\theta}}$ expresses that this is a function of sampling observations.

1.4.2 Inference approaches

Two types of statistical inference could be adopted (Särndal, Swenson and Wretman, 1992): (*i*) inference about the *finite population* $\mathcal{U}$ itself (i.e. about the current condition of $\mathcal{U}$); and (*ii*) inference about a model or a *superpopulation* considered to have generated $\mathcal{U}$ (i.e. about $\mathcal{U}$'s underlying process). Inference type (*i*) shall not be relevant to this thesis.

Moreover, there are two possible approaches for performing inference: (*i*) the *design-based* or randomization approach; and (*ii*) the *model-based* or prediction approach.

According to approach (*i*), inference is conducted with respect to the sampling distribution of statistics over repeated samples $\mathcal{S}$ produced by the *sampling scheme* $p(\mathcal{S})$, assuming that the values of the *finite population* are seen as fixed constants, but unknown (Skinner, Holt and Smith, 1989; Cochran, 1977). In this case, the stochastic structure is produced by the *sampling design* because the sample is the random element (Särndal, Swenson and Wretman, 1992). The randomization approach assumes that the unique aspect that is random and controlled by the laws of probability is the process that controls which population elements are included in the sample (Kott, 1991).

On the contrary, when approach *(ii)* is adopted, inference is conducted with respect to the sampling distribution of statistics over repeated realizations $y_1,\ldots,y_N$ (of N random variables $Y_1,\ldots,Y_N$) generated by the model denoted by ξ, assuming that s is fixed (Skinner, Holt and Smith, 1989). The statistical literature is predominantly *model-based*. The purely prediction approach bases the inference uniquely on the model, and it pays no attention to $p(S)$ and its inclusion probabilities π_i (Särndal, Swenson and Wretman, 1992). In this case the inference is done to the specific realised s and not to any other samples.

Both *design-based* and *model-based* approaches utilize a frequentist approach to inference (see Chambers and Skinner, 2003), which we adopt in this dissertation. Information on the Bayesian approach for inference in the context of complex survey data may be also found in Chamber and Skinner (2003, Part A).

For both parameter and standard error estimation, the classical *design-based* survey methods are usually more robust than *model-based* methods (Korn and Graubard, 1995). But the second approach is frequently more efficient than the first one. When both a *superpopulation* model and a probability sampling distribution are obtainable, the measures of variance and bias of an estimator could be defined with regard to the combined model (see Royall and Cumberland, 1981, Section 2.3).

Additionally, at this stage it is important to mention two wide alternative sets of analytic methodologies (Skinner, Holt and Smith, 1989). When the *disaggregated approach* is adopted, modelling procedures are performed allowing for different patterns within and between the categories of the *design variables*. Otherwise, in an *aggregated analysis* (or *marginal modelling*) the target model parameter is defined irrespective of the *design variables*.

1.5 Motivation and aims of the thesis

Researchers and other users of longitudinal survey data often make use of *standard* (or *classic*) statistical techniques, which in most of the cases do not take account of the complex sample designs. When statistical procedures are based upon *standard* formulation assumptions, it is usually assumed that data are generated plainly from the concerned population model, without any consideration of the *sampling design* (Chambers and Skinner, 2003). Those techniques may assume that the data are realizations of independent and identically distributed (iid) random vectors (Scott and Holt, 1982; and Skinner, 1986), which is rare in practice (Chambers and Skinner, 2003).

The *standard* formulation of the inference methods is often *not valid* when analysing data derived using a complex *sampling scheme*, i.e. when data is selected according to the *complex survey design approach*. In some situations, it may be not recommended to ignore the *sampling design*, especially when the scheme has a direct effect on inference procedures (Skinner, 1986). Furthermore, even when the *sampling design* is considered ignorable, the stochastic hypothesis incorporated by the standard procedures could not satisfactorily reproduce the population complexities underlying the sampling (Chambers and Skinner, 2003).

Complex *sampling schemes* may be the cause of a correlation structure among observations, as elements in the same cluster are likely to be more similar than elements in different clusters. This phenomenon is called in the literature by positive *intra-cluster correlation*. Under the presence of positive *intra-cluster correlation*, the use of *standard* methods could lead, for example, to the under-estimation of the standard errors. Therefore, test statistics could be affected and confidence intervals based on those standard errors could fail to achieve the desired nominal coverage level (Scott and Holt, 1982).

Scott and Holt (1982) discuss the effects of *intra-cluster correlation* on ordinary least squares methods (OLS) for linear regression. Skinner (1986) clarifies the effects of two-stage sampling on statistical inference of functions of population moments under both *design-based* and *model-based* viewpoints.

After doing an extensive literature review we have diagnosed that further investigation is needed on methods for the analysis of longitudinal complex survey data. This is our main motivation. In this thesis we focus our attention mainly on the analytic use of sample surveys. It is our main target to investigate statistical methods for the analysis of longitudinal data collected under complex *sampling designs*. It shall be one of our aims to formulate statistical methods that take this issue into account appropriately, when estimating a target model vector parameter $\underline{\theta}$.

According to Skinner (2003b), we may regard longitudinal analysis as a form of multivariate analysis and the issue of how to take the sampling design into account may be seen as a generalisation of this problem for cross-section surveys. This is generally the strategy we shall follow in this thesis. In particular, we shall be mainly concerned with studying statistical techniques for analysis and modelling of longitudinal survey data, such as covariance structure models and random effects models. We shall initially investigate the variance effects of clustering for longitudinal studies in the context of complex survey data. We shall also evaluate empirically the effects of ignoring survey design characteristics when using standard methodologies.

1.6 Outline of the thesis

This dissertation is organized in eight chapters. The present chapter has provided an overview of models appropriate for the analysis of longitudinal data. It has also discussed issues related to sampling and inference, including *finite population* and *superpopulation* models, and *sampling designs* in the longitudinal context. We have also introduced some notation and reviewed some aspects of longitudinal data. In addition, our motivations and the aims of our research project have been summarized above.

The remainder of this thesis is organised as follows. In Chapter 2 we shall mainly review longitudinal regression models in the context of complex survey data. In addition, existing methods for model parameters variance estimation are reviewed. Moreover we shall discuss misspecification effects in that context.

Chapter 3 shall consider an empirical investigation using longitudinal survey data from the British Household Panel Survey (BHPS), applying methods discussed in Chapter 2. The main characteristics of the BHPS data set shall also be described. Moreover variance effects of clustering for longitudinal studies shall be identified and illustrated. Furthermore we shall include some conclusions and a theoretical discussion in order to provide an argumentation that supports our major empirical results. The main results of this chapter appear in Skinner and Vieira (2005).

In Chapter 4 we shall further discuss estimation procedures for the longitudinal regression model vector parameter $\underline{\beta}$. Moreover we shall describe methods on inference about the population covariance matrix Σ and for the variance estimation of $\hat{\Sigma}$. Furthermore we shall discuss and review estimation methods for the vector parameter of interest $\underline{\theta}$, upon which the covariance matrix depends, including unweighted least squares, generalised least squares under the classical approach, and maximum likelihood. We shall additionally propose some modified estimation methods, as unweighted least squares and generalised least squares under the complex survey approach. A pseudo maximum likelihood is also derived via maximisation of the pseudo log likelihood function, in the context of covariance structure models. In addition some discussion about the methods discussed and proposed in this chapter shall also be given.

Chapter 5 shall present the characteristics and results of a simulation study, which has the main object of evaluating the statistical properties of the point estimation procedures discussed in Chapter 4. We shall also compare the properties of the proposed methods with the traditional statistical techniques described in Chapter 4. Several observations shall be drawn from the simulation results.

In Chapter 6 we shall discuss methods for variance estimation of $\hat{\underline{\theta}}$ and diagnostics techniques for structural models for covariance matrices. We shall initially review variance estimation methods and model fitting statistics under the standard approach. Furthermore, we shall also provide some new developments on both variance estimation and model fitting statistics when working under the complex sampling approach.

Chapter 7 shall present the characteristics and results of a second simulation study, which has the main object of evaluating the statistical properties of the variance estimation procedures and model fitting statistics proposed in Chapter 6. We shall also compare the properties of the proposed methods with the classic ones also discussed in Chapter 6.

We shall include in Chapter 8 some conclusions about the main issues discussed in this thesis. Moreover we shall point out the main research achievements included in this dissertation. In addition we shall enumerate some possible further research topics and future outcomes of this doctoral research project.

Chapter 2

Regression models for longitudinal survey data

2.1 Introduction

We consider a finite population denoted by $\bar{U}$ (see Chapter 1, Sub-section 1.2.1), which is fixed on occasions $1,\ldots,T$. Let N represent the size of $\bar{U}$. Let $N_o = N \cdot T$.

Let $\underline{Y}_i = (Y_{i1},\ldots,Y_{iT})'$ be the random vector containing T repeated observations on the *study variable* for unit $i = 1,2,\ldots,N$ over the T waves of the survey. We mainly consider Y_{it} to be continuous. We assume a model in which

$$E(\underline{Y}_i) = \underline{\mu}_i(\underline{\beta}) \tag{2.1}$$

is the $T \times 1$ vector with their respective expected values, where

$$\underline{\mu}_i(\underline{\beta}) = [\mu(\underline{x}_{i1},\underline{\beta}),\ldots,\mu(\underline{x}_{iT},\underline{\beta})]'. \tag{2.2}$$

In (2.2) $\underline{x}_{it}$ is a $1 \times q$ vector with the q fixed covariates, $\underline{\beta}$ is a $q \times 1$ vector of unknown parameters and the function $\mu(.,.)$ is assumed known. Let $\underline{y}_i$ be the observed values of $\underline{Y}_i$ (data).

Additionally let

$$\Sigma = \mathrm{COV}(\underline{Y}_i) = E\{[\underline{Y}_i - \underline{\mu}_i][\underline{Y}_i - \underline{\mu}_i]'\}, \tag{2.3}$$

to be a $T \times T$ population variance-covariance matrix. It is assumed that Σ does not depend on i. In (2.3), and in the remaining of this dissertation, COV(.) denotes population covariance.

Longitudinal regression models are able to relate individual's behaviour at one time point, possibly to other behaviour at another point in time. The regression models mainly described in this chapter are usually referred in the literature as (*i*) cross-sectional time-series linear models, (*ii*) population-averaged, or (*iii*) marginal models. This approach, including information about its estimation procedures is described in Liang and Zeger (1986), Zeger, Liang and Albert (1988), and Diggle *et al.* (2002).

In the linear cross-sectional regression models context, the model parameters are estimated by performing a comparison of subjects with a specific value of x, to the remaining cases with other values. Distinctively, for longitudinal regression model parameters to be estimated, a comparison of responses on a subject over the time is additionally carried out, considering that x is allowed to modify with time. When fitting a longitudinal model, each subject could be seen as working as its particular control. Moreover, there is usually a large amount of variability across subjects caused by unobserved attributes. Nevertheless, those unmeasured characteristics are 'camouflaged' when cross-sectional model parameters are estimated. See Diggle *et al.* (2002, Chapter 1, Section 1.4), for further information on these issues.

The methodology discussed in this chapter is one of the possible approaches to longitudinal data analysis. In brief terms, we may say that these models allow for within-individual correlation. Furthermore the estimation procedures focus on the marginal distribution and the correlation structure is also estimated.

The longitudinal regression models discussed in this chapter may be included in the *aggregated approach*, discussed in Chapter 1, Sub-section 1.3.3, as their target model parameters are usually not defined with respect to the categories of the *design variables*. Marginal regression models are an alternative for analysing continuous response discrete time longitudinal data, avoiding thus the specification of random effects (Skinner, 2003b).

We review in Section 2.2 estimation procedures for longitudinal regression model parameters considering the classic case, which considers a general situation where the sampling scheme $p(s)$ is defined such that s is selected setting data to be independent and to obey the model introduced in expressions (2.1) and (2.2). Section 2.3 discusses parameters variance estimation in the classic case.

The estimation of longitudinal model parameters allowing for complex designs is reviewed in Section 2.4, whilst Section 2.5 includes variance estimation of estimators of parameters. Section 2.6 discusses misspecification effects in the present context. Section

2.7 gives information about Sudaan, which is one of the possible statistical software to be adopted for application of the techniques discussed in the current chapter. Moreover, Section 2.8 concludes this chapter.

It is illustrative at this point to discuss a class of linear regression models in the context of longitudinal data (see Example 2.1 below).

Example 2.1: uniform correlation model (UCM)

A special case of the model discussed above may be represented as

$$Y_{it} = \underline{x}_{it}\underline{\beta} + u_i + v_{it}, \text{ with } i = 1,\ldots,N \text{ and } t = 1,\ldots,T, \tag{2.4}$$

which defines a class of random effects models[10] (see Model A, in Skinner and Holmes, 2003; and Skinner, 2003b, Section 13.2). In (2.4), $\underline{x}_{it}$ is a $1\times q$ vector with the q fixed covariates, $\underline{\beta}$ is the $q\times 1$ vector of the unknown fixed coefficients for the x variables (which represent the effect of each covariate on the response variable), u_i are the permanent random effects (or unobservable individual specific factors), and v_{it} are transitory random effects. More generally we could re-write (2.4) as

$$Y_{it} = \underline{x}_{it}\underline{\beta} + \varepsilon_{it},$$

where ε_{it} is the error term and $E(\varepsilon_{it}) = 0$.

We assume that the random variables u_i and v_{it} are mutually independent with $E(u_i) = E(v_{it}) = 0$, $\text{VAR}(u_i) = \sigma_u^2$, $\text{VAR}(v_{it}) = \sigma_v^2$, $\text{COV}(v_{it}, v_{it'}) = 0$ (for all $t \neq t'$), and $\text{COV}(u_i, v_{it}) = 0$, where VAR(.) denotes population variance. The expectation of Y_{it}, may be represented as

$$E(Y_{it}) = \underline{x}_{it}\underline{\beta} = \underline{\mu}(\underline{x}_{it}, \underline{\beta}), \text{ with } t = 1,\ldots,T. \tag{2.5}$$

From (2.4) and (2.5), the variance of Y_{it} and the covariance between Y_{it} and $Y_{it'}$ are given respectively by

$$\begin{aligned}&\sigma^2 = \text{VAR}(Y_{it}) = E\{[\varepsilon_{it}]^2\} = E[u_i + v_{it}]^2 = E[u_i^2 + 2u_i v_{it} + v_{it}^2] = \\ &\sigma_u^2 + 2E(u_i \cdot v_{it}) + \sigma_v^2 = \sigma_u^2 + \sigma_v^2,\end{aligned} \tag{2.6}$$

and

[10] The model described in Example 2.1 could be also called a multilevel or mixed linear model.

$$\begin{aligned} \mathrm{COV}(Y_{it},Y_{it'}) &= E\{[u_i+v_{it}]\cdot[u_i+v_{it'}]\}= \\ &= E\{[u_i]^2\}+E\{u_i\cdot v_{it}\}+E\{u_i\cdot v_{it'}\}+E\{v_{it}\cdot v_{it'}\}=E\{[u_i]^2\}=\sigma_u^2, \end{aligned} \tag{2.7}$$

for all $t \neq t'$. The variance of any response is modelled as two separate components: (*i*) the variation on the same individual, σ_v^2; and (*ii*) variance across individuals (a covariance), σ_u^2 (see Lindsey, 1994). From the general definition of correlation and expressions (2.6) and (2.7),

$$\mathrm{CORR}(Y_{it},Y_{it'})=\rho=\frac{\sigma_u^2}{\sigma_u^2+\sigma_v^2},$$

where ρ denotes the population *intra-individual correlation*, and $0 \leq \rho < 1$. The *intra-individual correlation* is basically produced by the random individual effect. It is (see Jones, 1993, p. 14) (*i*) $\rho = 0$ when $\sigma_u^2 = 0$, meaning that there is no variance across units (what would reduce the case to simple regression); and (*ii*) ρ is 'close' to 1 when σ_v^2 is 'close' to 0, meaning that there is almost no variation on the same individual over time and no measurement error (the situation '$\rho = 1$' does not happen in practice). ■

We assume in this chapter that (*i*) the observations are equally spaced in time; and (*ii*) the number of individuals, *N*, is 'large' relative to the number of observations per individual, *T* (see Jones, 1993, Section 1.3).

We also suppose that (*iii*) the sample is selected on one occasion and then the same sample units are returned on to each of the $T-1$ subsequent waves of the survey, and for simplicity (*iv*) we assume no nonresponse (every Y_i has the same length *T*).

2.2 Estimation procedures for parameter $\underline{\beta}$ - Classical case

Suppose that $\underline{Y}_i = (Y_{i1},\ldots,Y_{iT})'$ is distributed as a *T*-dimensional multivariate normal, which is denoted by

$$\underline{Y}_i \sim \mathrm{N}_T[\underline{\mu}_i(\underline{\beta}),\Sigma], \tag{2.8}$$

where Σ is a $T \times T$ positive definite covariance matrix. Note the distinction in notation between *N*, which is adopted for the finite population size, and N which is adopted as usual for the Normal distribution.

Let $\underline{y}_i$ be $T \times 1$ vectors with the observed values for the response variable for each individual i in each wave. At this stage, we assume that the whole finite population is observed, and that $\underline{Y}_1,\ldots,\underline{Y}_N$ are mutually independent. So, we may define the census joint density function (or probability mass function) for all the observations as the product of the marginal normal densities (see Johnson and Wichern, 1998)

$$f(\underline{y}_1,\ldots,\underline{y}_N;\underline{\beta}) = \prod_{i=1}^{N}\left\{\frac{1}{(2\pi)^{T/2}|\Sigma|^{1/2}} e^{-[\underline{y}_i-\underline{\mu}_i(\underline{\beta})]'\Sigma^{-1}[\underline{y}_i-\underline{\mu}_i(\underline{\beta})]/2}\right\}$$
$$= (2\pi)^{-N_o/2}|\Sigma|^{-N/2} e^{-\sum_{i=1}^{N}[\underline{y}_i-\underline{\mu}_i(\underline{\beta})]'\Sigma^{-1}[\underline{y}_i-\underline{\mu}_i(\underline{\beta})]/2}. \quad (2.9)$$

Note the distinction in notation between Σ, which is adopted for the covariance matrix, and $\sum$, which is the standard symbol to represent the sum. The matrix Σ is treated here as known. In (2.9), the resulting expression may be called census likelihood function, which we denote by $\ell_N[\underline{\beta}]$.

The $q \times 1$ parameter vector $\underline{\beta}$ may then be estimated by maximising the logarithmic[11] census likelihood with respect to $\underline{\beta}$, which is defined as

$$L_N[\underline{\beta}] = \log \ell_N[\underline{\beta}] = \log\left[\frac{1}{(2\pi)^{N_o/2}}\frac{1}{|\Sigma|^{N/2}} e^{-\sum_{i=1}^{N}[\underline{y}_i-\underline{\mu}_i(\underline{\beta})]'\Sigma^{-1}[\underline{y}_i-\underline{\mu}_i(\underline{\beta})]/2}\right] =$$
$$= -\frac{1}{2}\left[N_o \log 2\pi + N\log|\Sigma| + \sum_{i=1}^{N}[\underline{y}_i - \underline{\mu}_i(\underline{\beta})]'\Sigma^{-1}[\underline{y}_i - \underline{\mu}_i(\underline{\beta})]\right]. \quad (2.10)$$

The maximum census likelihood estimator[12] $\hat{\underline{\beta}}_N$, for the parameter $\underline{\beta}$, may be obtained by minimising the exponent in the census likelihood function (2.10)

$$\sum_{i=1}^{N}[\underline{y}_i - \underline{\mu}_i(\underline{\beta})]'\Sigma^{-1}[\underline{y}_i - \underline{\mu}_i(\underline{\beta})], \quad (2.11)$$

with respect to $\underline{\beta}$. Expression (2.11) above is the sum of squares of the multivariate generalized distance from $\underline{Y}_i$ to $\mu_i(\underline{\beta})$.

---

[11] It is frequently easier to maximise the log of a function than the function itself. The adoption of the logarithm does not change the value of $\underline{\beta}$, as the log is a monotonic function (Bollen, 1989, Appendix 4A).

[12] The classic maximum likelihood estimation approach (ML) "*determines estimates for the model parameters that maximise the likelihood of observing the available data if one were to collect data from the same population again*" (Raykov and Marcoulides, 2000). Furthermore, ML estimators are consistent, asymptotically normal, and efficient, but they are "*derived under the assumption of a particular parametric form, generally the normal, for the distribution of the data vector* (Harville, 1977).

Now suppose that we only observe $\underline{y}_i$ for unit i in a sample s, denote $\{1,\ldots,n\}$. Thus expression (2.11) could be estimated by

$$\frac{N}{n}\cdot\sum_{i=1}^{n}\left[\underline{y}_i-\underline{\mu}_i(\underline{\beta})\right]'\Sigma^{-1}\left[\underline{y}_i-\underline{\mu}_i(\underline{\beta})\right]. \tag{2.12}$$

The maximum likelihood estimator of $\underline{\beta}$, is obtained by minimising expression (2.12). Alternatively, we may solve the following equation system,

$$\sum_{i=1}^{n}\left[\underline{y}_i-\underline{\mu}_i(\underline{\beta})\right]'\Sigma^{-1}\frac{\partial\underline{\mu}_i(\underline{\beta})}{\partial\underline{\beta}}=0, \tag{2.13}$$

which are usually known as the pseudo score equations for $\underline{\beta}$. Let X_i be the $T\times q$ matrix with covariates for individual i, so that $\mathrm{X}_i=(\underline{x}_{i1}',\ldots,\underline{x}_{iT}')'$. When $\underline{\mu}_i(\underline{\beta})=\mathrm{X}_i\underline{\beta}$, we have

$$\frac{\partial\underline{\mu}_i(\underline{\beta})}{\partial\underline{\beta}}=\mathrm{X}_i.$$

In this case, under the linear model presented in (2.1), the pseudo score equations have a closed-form solution for $\hat{\underline{\beta}}(\Sigma)$ given by (with standard matrix manipulations; see Jones, 1993)

$$\hat{\underline{\beta}}(\Sigma)=\left(\sum_{i=1}^{n}X_i'\Sigma^{-1}X_i\right)^{-1}\sum_{i=1}^{n}X_i'\Sigma^{-1}\underline{y}_i. \tag{2.14}$$

In general Σ is unknown. We shall consider estimation of Σ in Chapter 4. One alternative approach is to replace Σ by a $T\times T$ *working covariance matrix* V (Diggle *et al.*, 2002, p. 70). Let thus $\hat{\underline{\beta}}(\mathrm{V})$ denote the estimator when Σ is replaced by V. Thus, when $\underline{\mu}_i(\underline{\beta})=\mathrm{X}_i\underline{\beta}$,

$$\hat{\underline{\beta}}(\mathrm{V})=\left(\sum_{i=1}^{n}\mathrm{X}_i'\mathrm{V}^{-1}\mathrm{X}_i\right)^{-1}\sum_{i=1}^{n}\mathrm{X}_i'\mathrm{V}^{-1}\underline{y}_i. \tag{2.15}$$

Different estimators of $\underline{\beta}$ arise for different choices of V. We consider below some alternatives. Let R be a $T\times T$ *working correlation matrix* corresponding to V, so that (see for example, Press, 1972; and Rencher, 1998)

$$\mathrm{R}=\mathrm{D}_\mathrm{V}^{-1}\mathrm{V}\mathrm{D}_\mathrm{V}^{-1},$$
$$\mathrm{V}=\mathrm{D}_\mathrm{V}\mathrm{R}\mathrm{D}_\mathrm{V},$$

where

$$\mathrm{D}_\mathrm{V}=[\mathrm{diag}(\mathrm{V})]^{1/2}. \tag{2.16}$$

In (2.16), diag(V) may be obtained replacing off-diagonal elements of V by zeros. In practice, R could be chosen after calculating empirical estimates of the correlations, and could take several different commonly adopted patterns, as for example: (*i*) exchangeable, i.e. $R_{tt'} = 1$ for $t = t'$ and ρ otherwise, where ρ is the first lag correlation; (*ii*) stationary, e.g. $R_{tt'} = 1$ for $t = t'$, ρ if $0 < |t - t'| \leq g$, and 0 otherwise, where g is the maximum considered lag covariance; (*iii*) autoregressive, i.e. $R_{tt'} = 1$ for $t = t'$ and $\rho^{|t-t'|}$ otherwise; (*iv*) nonstationary, e.g. $R_{tt'} = 1$ for $t = t'$, $\rho_{tt'}$ if $0 < |t - t'| \leq g$, where $\rho_{tt'} = \rho_{t't}$; and (*v*) unstructured, i.e. $R_{tt'} = 1$ for $t = t'$, $\rho_{tt'}$ otherwise, where $\rho_{tt'} = \rho_{t't}$.

We refer to V and R as *working* covariance and correlation matrices respectively, because V (as in Zeger and Liang, 1986) is not expected to be the true value of Σ.

As a special case of $\hat{\underline{\beta}}(\mathrm{V})$, the simplest option for the *working* covariance matrix is $\mathrm{V} = I$, where I is an $T \times T$ identity matrix, i.e. we assume that repeated outcomes for a given individual are independent. It makes expression (2.15) equivalent to the ordinary least squares (OLS) estimator for $\underline{\beta}$, and equals to

$$\hat{\underline{\beta}}(I) = \left(\sum_{i=1}^{n} \mathrm{X}_i'\mathrm{X}_i \right)^{-1} \sum_{i=1}^{n} \mathrm{X}_i'\underline{y}_i . \tag{2.17}$$

According to Liang and Zeger (1986, Section 4), $\hat{\underline{\beta}}(\mathrm{V})$ is a consistent estimator for $\underline{\beta}$ whatever the choice[13] of a constant matrix V. This generally depends only upon the correct specification of the mean. We further discuss this statement for the complex survey data case.

Example 2.1: (*Continuation*)

Continuing the uniform correlation example, we have

$$\mathrm{VAR}(\underline{\varepsilon}_i) = \sigma^2 \begin{bmatrix} 1 & & & \\ \rho & 1 & & \\ \vdots & & \ddots & \\ \rho & \rho & \cdots & 1 \end{bmatrix}, \tag{2.18}$$

[13] However, specifying V closer to the true correlation increases the efficiency of the estimator.

where $\underline{\varepsilon}_i = [\varepsilon_{i1},\ldots,\varepsilon_{iT}]'$. Thus, it is assumed constant residual correlation between any pair of observations from the same individual, i.e. an exchangeable *working covariance matrix*. Note that σ^2 is a constant term which cancels out of (2.15) and hence does not need to be estimated for $\hat{\underline{\beta}}(\mathrm{V})$.

Let var(.) denote an estimator of VAR(.). The covariance matrix $\mathrm{VAR}(\underline{\varepsilon}_i)$, presented in expression (2.18) could be estimated by

$$\mathrm{var}(\underline{\varepsilon}_i) = \hat{\sigma}^2 \begin{bmatrix} 1 & & & \\ \hat{\rho} & 1 & & \\ \vdots & & \ddots & \\ \hat{\rho} & \hat{\rho} & \cdots & 1 \end{bmatrix} = \mathrm{V}, \qquad (2.19)$$

where (Jones, 1993)

$$\hat{\sigma}^2 = (1/n-q)\cdot\sum_{i=1}^{n}\hat{\underline{\varepsilon}}'_i \mathrm{V}^{-1}\hat{\underline{\varepsilon}}_i,$$

is an estimator of σ^2 (Shah, Barnwell and Bieler, 1997),

$$\hat{\rho} = \frac{\dfrac{\sum_{i=1}^{n}\left(\sum_{i=1}^{T}\hat{\varepsilon}^*_{it}\right)^2}{n\cdot T} - \dfrac{\sum_{i=1}^{n}\sum_{i=1}^{T}\left(\hat{\varepsilon}^*_{it}\right)^2}{n\cdot T}}{n\cdot(T-1)}$$

is the estimator of the intra-individual correlation, and by iterating between the estimation of $\underline{\beta}$ and the estimation of the intra-individual correlation (Liang and Zeger, 1986; and Shah, Barnwell and Bieler, 1997),

$$\hat{\varepsilon}^*_{it} = \frac{\hat{\varepsilon}_{it} - \overline{\hat{\varepsilon}_i}}{\hat{\sigma}_i},$$

$$\hat{\varepsilon}_{it} = y_{it} - \underline{x}_{it}\hat{\underline{\beta}}(\mathrm{V}), \qquad (2.20)$$

$$\hat{\sigma}_i^2 = \left[\left(\sum_{t=1}^{T}\hat{\varepsilon}^2_{it}\right)\Big/T\right] - \left(\overline{\hat{\varepsilon}_i}\right)^2,$$

and

$$\overline{\hat{\varepsilon}_i} = \left(\sum_{t=1}^{T}\hat{\varepsilon}_{it}\right)\Big/T. \blacksquare$$

2.3 Variance estimation in classical case

2.3.1 Weighted least squares

A variance estimator for $\underline{\hat{\beta}}(\mathrm{V})$ given in (2.18), assuming that the model is true, is (Diggle *et al.*, 2002, p. 60)

$$\mathrm{var_n}\left[\underline{\hat{\beta}}(\mathrm{V})\right] = \hat{\sigma}^2 \left[\sum_{i=1}^{n} \left(\mathrm{X}_i' \mathrm{V}^{-1} \mathrm{X}_i\right)\right]^{-1}, \qquad (2.21)$$

where (Jones, 1993; Diggle *et al.*, 2002, p. 63)

$$\hat{\sigma}^2 = (1/n - q) \cdot \sum_{i=1}^{n} \underline{\hat{\varepsilon}}_i' \mathrm{V}^{-1} \underline{\hat{\varepsilon}}_i ,$$

and X_i, V and $\underline{\hat{\varepsilon}}_i$ are defined earlier on. Expression (2.20), given above, may be adopted for calculating $\underline{\hat{\varepsilon}}_i$.

The estimator in (2.21) does not account for the complex survey design. It is also not robust for misspecifications of V. In (2.21) the subscript n is thus adopted for denoting 'naïve'. This estimator is usually referred in the literature as the weighted least squares estimator.

We shall adopt in Chapter 3 the term 'naïve' for this variance estimator in situations where we assume that $\mathrm{V} = I$, where I is an identity matrix as defined for (2.20). Then, (2.21) could be re-expressed as (Diggle *et al.*, 2002, p. 63)

$$\mathrm{var_n}\left[\underline{\hat{\beta}}(I)\right] = \hat{\sigma}^2 \left[\sum_{i=1}^{n} (X_i' X_i)\right]^{-1}. \qquad (2.22)$$

When (2.22) is adopted, we do not consider the correlation structure in the data. It may be somehow very risky to assume that $\mathrm{V} = I$ when it is not so, what could lead us to serious over- or underestimation of the variance of $\underline{\hat{\beta}}$ (Diggle *et al.*, 2002).

2.3.2 Robust variance estimator

The variance estimator discussed in the current section belongs to a class of estimation methods, which are referred in the statistical and econometric literature as Huber-White (White, 1980) or *sandwich* estimators. These methods produce consistent estimates without making any distributional assumption. The previous statement is valid even if the underlying model is incorrect (see Kauermann and Carroll, 2001).

We may produce estimates for the variance of $\hat{\underline{\beta}}(\mathrm{V})$ given in (2.15) which are robust for misspecifications of V, as the covariance matrix Σ is typically unknown. We shall discuss in this sub-section the robust variance estimator presented by Liang and Zeger (1986) as consistent when (2.1) holds, but where the working covariance matrix, V, may not reflect the true covariance structure. However, expressions presented below still regard situations where the sample is selected by a simple random sampling (srs) scheme. We may derive this estimator by writing

$$[\hat{\underline{\beta}}(\mathrm{V})-\underline{\beta}]=\left(\sum_{i=1}^{n}\mathrm{X}_i'\mathrm{V}^{-1}\mathrm{X}_i\right)^{-1}\cdot\sum_{i=1}^{n}\mathrm{X}_i'\mathrm{V}^{-1}\underline{\varepsilon}_i\,. \tag{2.23a}$$

In addition the variance of $[\hat{\underline{\beta}}(\mathrm{V})-\underline{\beta}]$ is given by (Liang and Zeger, 1986; see also Zeger, Liang, and Albert, 1988; Kauermann and Carroll, 2001; and Binder, 1983)

$$\begin{aligned}\mathrm{var}\left[\hat{\underline{\beta}}(\mathrm{V})-\underline{\beta}\right]&=\left(\sum_{i=1}^{n}\mathrm{X}_i'\mathrm{V}^{-1}\mathrm{X}_i\right)^{-1}\cdot\sum_{i=1}^{n}\mathrm{var}\left(\mathrm{X}_i'\mathrm{V}^{-1}\underline{\varepsilon}_i\right)\cdot\left(\sum_{i=1}^{n}\mathrm{X}_i'\mathrm{V}^{-1}\mathrm{X}_i\right)^{-1}\\&=\left(\sum_{i=1}^{n}\mathrm{X}_i'\mathrm{V}^{-1}\mathrm{X}_i\right)^{-1}\cdot\left[\sum_{i=1}^{n}\mathrm{X}_i'\mathrm{V}^{-1}\,\mathrm{var}(\underline{\varepsilon}_i)\mathrm{V}^{-1}\mathrm{X}_i\right]\cdot\left(\sum_{i=1}^{n}\mathrm{X}_i'\mathrm{V}^{-1}\mathrm{X}_i\right)^{-1},\end{aligned} \tag{2.23b}$$

where X_i is as defined for (2.14) whilst V is as defined for (2.15). The final expression in (2.23b) yields the robust variance estimator of $\hat{\underline{\beta}}(\mathrm{V})$, given by

$$\mathrm{var}_r\left[\hat{\underline{\beta}}(\mathrm{V})\right]=\left[\sum_{i=1}^{n}\mathrm{X}_i'\mathrm{V}^{-1}\mathrm{X}_i\right]^{-1}\left[\sum_{i=1}^{n}\left(\mathrm{X}_i'\mathrm{V}^{-1}\hat{\underline{\varepsilon}}_i\right)\left(\mathrm{X}_i'\mathrm{V}^{-1}\hat{\underline{\varepsilon}}_i\right)'\right]\left[\sum_{i=1}^{n}\mathrm{X}_i'\mathrm{V}^{-1}\mathrm{X}_i\right]^{-1},$$

where

$$\hat{\underline{\varepsilon}}_i=\underline{y}_i-\mathrm{X}_i\hat{\underline{\beta}}(\mathrm{V}).$$

The estimator given above does not account for the complex survey design, as it still assumes that observations for two different individuals are independent.

According to Diggle *et al.* (2002), an inadequate choice of V could affect the efficiency of inferences for $\underline{\beta}(\mathrm{V})$, even though that would not have any effect on its validity. Hence, confidence intervals and significance tests calculated based on $\mathrm{var}_r[\hat{\underline{\beta}}(\mathrm{V})]$ are asymptotically correct. However, the price of the consistency of

$\mathrm{var}_r[\hat{\underline{\beta}}(\mathrm{V})]$ is its inefficiency because it has in general larger variance than classical variance estimators. See Kauermann and Carroll (2001), and references contained therein, for further information regarding robust variance estimators.

2.4 Estimation of $\underline{\beta}$ allowing for complex design

The methods discussed in the previous two sections rely on the assumption that the data are obtained by srs from large populations. However, the sample structure of many social surveys is complex, involving stratification, clustering, and multiple stages of selection, yielding data that violate the srs assumption.

We assume here that $\underline{Y}_i$ is defined as in Section 2.2 and is distributed as stated in (2.08). Moreover we define the census joint density function and the census likelihood function as in (2.09). Let $\underline{y}_i$ and X_i be respectively a vector and a matrix as previously defined in Section 2.2.

Here we may still consider that the maximum census likelihood estimator $\hat{\underline{\beta}}_N$, for the parameter $\underline{\beta}$, could be obtained by minimising the exponent in the census likelihood function, presented in (2.11), with respect to $\underline{\beta}$.

As in Section 2.2, we may again suppose that we only observe $\underline{y}_i$ for units i in a sample s. As N may be unknown, in order to allow for complex sampling design, we could estimate expression (2.11) by

$$\sum_{i=1}^{n} w_i \left[\underline{y}_i - \underline{\mu}_i(\underline{\beta})\right]' \Sigma^{-1} \left[\underline{y}_i - \underline{\mu}_i(\underline{\beta})\right], \tag{2.24}$$

where w_i are *sampling weights* (see Chapter 1, Sub-section 1.3.1). The pseudo maximum likelihood estimator, PML (Skinner, 1989a), of $\underline{\beta}$, is obtained by minimising expression (2.24). Alternatively, we have to solve the following equation system,

$$\sum_{i=1}^{n} w_i \left[\underline{y}_i - \underline{\mu}_i(\underline{\beta})\right]' \Sigma^{-1} \frac{\partial \underline{\mu}_i(\underline{\beta})}{\partial \underline{\beta}} = 0 . \tag{2.25}$$

When $\underline{\mu}_i(\underline{\beta}) = \mathrm{X}_i \underline{\beta}$, we have

$$\frac{\partial \underline{\mu}_i(\underline{\beta})}{\partial \underline{\beta}} = \mathrm{X}_i .$$

The pseudo score equations would thus have a closed-form solution for $\underline{\hat{\beta}}(\Sigma)_{PML}$ given by

$$\underline{\hat{\beta}}(\Sigma)_{PML} = \left(\sum_{i=1}^{n} w_i \mathrm{X}_i' \Sigma^{-1} \mathrm{X}_i \right)^{-1} \sum_{i=1}^{n} w_i \mathrm{X}_i' \Sigma^{-1} \underline{y}_i , \tag{2.26}$$

where $\underline{\hat{\beta}}(\Sigma)_{PML}$ is a $q \times 1$ vector with the estimated coefficients, the *sampling weights* w_i are scalars, and n, X_i, Σ and $\underline{y}_i$ are as defined in Section 2.2.

We let $\underline{\hat{\beta}}(\mathrm{V})_{PML}$ denote the pseudo likelihood estimator when Σ is replaced by V, where V is a *working covariance matrix* as defined in Section 2.2. Thus, when $\underline{\mu}_i(\underline{\beta}) = \mathrm{X}_i \underline{\beta}$,

$$\underline{\hat{\beta}}(\mathrm{V})_{PML} = \left(\sum_{i=1}^{n} w_i \mathrm{X}_i' \mathrm{V}^{-1} \mathrm{X}_i \right)^{-1} \sum_{i=1}^{n} w_i \mathrm{X}_i' \mathrm{V}^{-1} \underline{y}_i . \tag{2.27}$$

Note that the estimator $\underline{\hat{\beta}}(\mathrm{V})$ defined in expression (2.15), in Section 2.2, is a special case of $\underline{\hat{\beta}}(\mathrm{V})_{PML}$ when the *sampling weights* are constant. Note that (2.27) is equivalent to

$$\underline{\hat{\beta}}(\mathrm{V})_{PML} = \left(\sum_{i=1}^{n} \mathrm{X}_i^{*\prime} \mathrm{V}^{-1} \mathrm{X}_i^{*} \right)^{-1} \sum_{i=1}^{n} \mathrm{X}_i^{*\prime} \mathrm{V}^{-1} \underline{y}_i^{*} , \tag{2.27a}$$

where

$$\mathrm{X}_i^{*} = \left(\underline{x}_{i1}^{*\prime}, \ldots, \underline{x}_{iT}^{*\prime} \right)', \quad \text{with} \quad \underline{x}_{it}^{*} = \sqrt{w_i} \cdot \underline{x}_{it} ,$$

and

$$\underline{y}_i^{*} = \left(y_{i1}^{*}, \ldots, y_{iT}^{*} \right)', \quad \text{with} \quad y_{it}^{*} = \sqrt{w_i} \cdot y_{it} .$$

Different estimators of $\underline{\beta}$ arise for different choices of V. See Section 2.2 for some alternative choices and further information on V.

We assume that repeated outcomes for a given individual are independent when we consider $\mathrm{V} = I$, where I is an $T \times T$ identity matrix, for $\underline{\hat{\beta}}(\mathrm{V})_{PML}$. It makes expression (2.27) equivalent to the ordinary least squares (OLS) estimator for $\underline{\beta}$ weighted by the sampling weighs w_i, and equal to (Kish and Frankel, 1974; and Fuller, 1975)

$$\underline{\hat{\beta}}(I)_{PML} = \left(\sum_{i=1}^{n} w_i \mathrm{X}_i'\mathrm{X}_i\right)^{-1} \sum_{i=1}^{n} w_i \mathrm{X}_i' \underline{y}_i , \tag{2.28}$$

where $\underline{\hat{\beta}}(I)_{PML}$ denotes the PML estimator for $\underline{\beta}$ when setting $\mathrm{V} = I$.

We now show, assuming V is constant, that $\underline{\hat{\beta}}_{PML}$ is approximately unbiased whichever choice of V we adopt. Although V could be sometimes chosen after calculating empirical estimates, we shall assume here that V is constant, i.e. that it does not depend upon the choice of the sample s. Thus, from (2.26), (2.27) and (2.28)

$$E_p\left[\underline{\hat{\beta}}(\mathrm{V})_{PML}\right] = E_p\left[\left(\sum_{i=1}^{n} w_i \mathrm{X}_i'\mathrm{V}^{-1}\mathrm{X}_i\right)^{-1} \sum_{i=1}^{n} w_i \mathrm{X}_i'\mathrm{V}^{-1}\underline{y}_i\right] \doteq$$

$$\doteq \left[E_p\left(\sum_{i=1}^{n} w_i \mathrm{X}_i'\mathrm{V}^{-1}\mathrm{X}_i\right)\right]^{-1} E_p\left(\sum_{i=1}^{n} w_i \mathrm{X}_i'\mathrm{V}^{-1}\underline{y}_i\right) = \left(\sum_{i=1}^{N} \mathrm{X}_i'\mathrm{V}^{-1}\mathrm{X}_i\right)^{-1} \sum_{i=1}^{N} \mathrm{X}_i'\mathrm{V}^{-1}\underline{y}_i .$$

The approximation above may be justified by expressions given by (1.1), in Chapter 1, Sub-section 1.3.1, taking $z_i = \mathrm{X}_i'\mathrm{V}^{-1}\mathrm{X}_i$ or $z_i = \mathrm{X}_i'\mathrm{V}^{-1}\underline{y}_i$. Moreover, if $E(\underline{y}_i) = \mathrm{X}_i'\underline{\beta}$ and if we assume that V is constant with respect to the model,

$$E\left\{E_p\left[\underline{\hat{\beta}}(\mathrm{V})_{PML}\right]\right\} \doteq \left(\sum_{i=1}^{N} \mathrm{X}_i'\mathrm{V}^{-1}\mathrm{X}_i\right)^{-1} \sum_{i=1}^{N} \mathrm{X}_i'\mathrm{V}^{-1}\mathrm{X}_i\underline{\beta} = \underline{\beta} . \tag{2.29}$$

Then, similarly to the Generalised Estimating Equation (GEE) approach discussed by Liang and Zeger (1986), $\underline{\hat{\beta}}(\mathrm{V})_{PML}$ will be approximately unbiased with respect to sampling, nonresponse and the model whatever the choice of a constant matrix V, if the model holds and the weights fully capture the sampling and nonresponse probabilities.

A common concern when working with a longitudinal sample s_T is that the underlying attrition mechanism may conduct to biased estimation of $\underline{\beta}$ (see Chapter 1, Section 1.3). One possible way of attempting to correct for this potential biasing effect is via the use of longitudinal survey weights, w_{iT}^*. We may assume here that (2.29) holds even when we substitute w_i by w_{iT}^*. This is a reasonable assumption as the longitudinal weights still capture the probabilities of selection π_i, where π_i is as described in Chapter 1, Sub-section 1.3.1.

2.5 Variance estimation for $\underline{\hat{\beta}}_{PML}$

When sample s is selected by a complex sampling scheme, a correlation structure among the observations, additional to the longitudinal correlation, may be caused. Then, in that situation if we assume that s is selected by a simple random sampling, we could generate erroneous estimates of the standard errors. An effect that could also interfere in the construction of confidence intervals and calculation of tests of significance.

In this section we present some information on two of the most commonly used variance estimation methods in survey sampling: the (*i*) Taylor linearization variance estimation approach (Sub-section 2.5.1); and the (*ii*) Jackknife replication variance estimation approach (Sub-section 2.5.2).

2.5.1 Linearization variance estimator

The robust variance estimation method discussed in Sub-section 2.3.2 involve Taylor expansions of the parameter point estimator $\underline{\hat{\beta}}(\mathrm{V})$, about the true model parameter $\underline{\beta}$ providing first order approximations. The robust variance estimator is thus a special case of the linearization variance estimator, which we present in more details in the current sub-section. We provide here further discussion about the well known linearization variance estimator, particularly in the context of complex survey schemes.

The linearization (commonly named Taylor series expansion or δ) method is based on the approximation of $\underline{\hat{\beta}}(\mathrm{V})_{PML}$, which is nonlinear in the observations, by a linear function. Variance estimation is thus performed making use of the first-order series approximation (see Binder, 1983). We have to assume that (*i*) $\underline{\hat{\beta}}(\mathrm{V})_{PML}$ is consistent for $\underline{\beta}$ (see Section 2.4), and that (*ii*) the expansion of the estimator (terms beyond the linear one) make negligible contribution to its variance[14] (Rust, 1985).

Let $\mathrm{var}_L[\underline{\hat{\beta}}(\mathrm{V})_{PML}]$ denote a linearization variance estimator, which is a $q \times q$ covariance matrix of $\underline{\hat{\beta}}(\mathrm{V})_{PML}$. We shall follow steps suggested by Binder (1995). Expression (2.23a) extends to

[14] This assumption may fail for 'small' sample sizes (Rust, 1985).

$$\left[\underline{\hat{\beta}}(\mathrm{V})_{PML}-\underline{\beta}\right]=\left(\sum_{i=1}^{n}w_i\mathrm{X}_i'\mathrm{V}^{-1}\mathrm{X}_i\right)^{-1}\cdot\sum_{i=1}^{n}w_i\mathrm{X}_i'\mathrm{V}^{-1}\underline{\varepsilon}_i\,, \tag{2.30}$$

where $\underline{\varepsilon}_i$ is the error term,

$$\underline{\varepsilon}_i=\underline{Y}_i-\mathrm{X}_i\underline{\beta}\,. \tag{2.31}$$

Thus (similarly to Skinner, 1989a, Sub-section 3.4.4), we have

$$\mathrm{var}_L\left[\underline{\hat{\beta}}(\mathrm{V})_{PML}-\underline{\beta}\right]=\left(\sum_{i=1}^{n}w_i\mathrm{X}_i'\mathrm{V}^{-1}\mathrm{X}_i\right)^{-1}\cdot\mathrm{var}_L\left[\sum_{i=1}^{n}w_i\mathrm{X}_i'\mathrm{V}^{-1}\underline{\varepsilon}_i\right]\cdot\left(\sum_{i=1}^{n}w_i\mathrm{X}_i'\mathrm{V}^{-1}\mathrm{X}_i\right)^{-1}, \tag{2.32}$$

where the subscript L denotes linearization.

For producing an estimate for $\mathrm{var}_L\left[\underline{\hat{\beta}}(\mathrm{V})_{PML}\right]$, we need to find an estimate for the middle term in the right hand side of (2.32),

$$\mathrm{var}_L\left[\sum_{i=1}^{n}w_i\mathrm{X}_i'\mathrm{V}^{-1}\underline{\varepsilon}_i\right]. \tag{2.33}$$

As we usually do not know the population parameter $\underline{\beta}$ in (2.31), we also can not observe $\underline{\varepsilon}_i$ given above. Expression (2.33) could then be estimated by

$$\mathrm{var}_L\left[\sum_{i=1}^{n}w_i\mathrm{X}_i'\mathrm{V}^{-1}\underline{\hat{\varepsilon}}_i\right], \tag{2.34}$$

where $\underline{\hat{\varepsilon}}_i=Y_i-\mathrm{X}_i\underline{\hat{\beta}}(\mathrm{V})_{PML}$, with $\underline{\hat{\beta}}(\mathrm{V})_{PML}$ given by (2.27).

We notice that (2.34) is basically an estimate for the variance of a weighted estimator (see Binder, 1983, Section 4.2), which is explicitly dependent upon the *sampling weight*s w_i. We could then rewrite (2.34) as

$$\mathrm{var}_L\left[\sum_{i=1}^{n}w_i\underline{Z}_i\right], \tag{2.35}$$

where $\underline{Z}_i=\mathrm{X}_i'\mathrm{V}^{-1}\underline{\hat{\varepsilon}}_i$ is a $q\times 1$ vector. In (2.35) we may notice that

$$\sum_{i=1}^{n}w_i\underline{Z}_i \tag{2.36}$$

has the form of an estimate of the population total of Z (see Binder 1995), which we denote by the $q\times 1$ vector of totals $\underline{\hat{Z}}$. Therefore,

$$\underline{\hat{Z}}=\sum_{i=1}^{n}w_i\underline{Z}_i\,. \tag{2.37}$$

We may thus adopt a standard technique for estimating the variance of (2.36) for the sampling scheme adopted for selecting the considered data set. We consider below an specific sampling design as an example on how to estimate the variance of $\hat{\underline{Z}}$.

Example 2.2: sampling of PSUs with replacement

We shall consider a multistage stratified sampling scheme that involves sampling with replacement at the first stage of PSUs from each of a total of H strata, and sampling with or without replacement at subsequent stages. We additionally consider equal or unequal selection probabilities at both the first and subsequent stages.

In order to explicitly consider stratification and clustering we may rewrite (2.37) as

$$\hat{\underline{Z}} = \sum_{h=1}^{H} \hat{\underline{Z}}_h = \sum_{h=1}^{H}\sum_{j=1}^{m_h} \hat{\underline{Z}}_{hj} = \sum_{h=1}^{H}\sum_{j=1}^{m_h}\sum_{i=1}^{n_{hj}} w_{hji}\underline{Z}_{hji},$$

where H is the number of strata in the sample, m_h is the sample number of PSUs in stratum h, n_{hj} is the sample number of individuals in PSU j in stratum h, and w_{hji} is the sampling weigh for individual i in PSU j in stratum h. From Shah *et al.* (1995, Sub-section 2.2.3), an estimator for the variance of $\hat{Z}$, considering the sampling scheme described above, is given by

$$\text{var}_L\left[\hat{\underline{Z}}\right] = \sum_{h=1}^{H} m_h \left\{ \left[\sum_{j=1}^{m_h} \left(\hat{\underline{Z}}_{hj} - \overline{\hat{\underline{Z}}}_h \right)^2 \right] \Big/ (m_h - 1) \right\}, \tag{2.38}$$

where $\hat{\underline{Z}}_{hj}$ is an estimator of a total in PSU j in stratum h, and $\overline{\hat{\underline{Z}}}_h$ is the mean of $\hat{\underline{Z}}_{hj}$ in stratum h. See also Cochran (1977, Section 11.9).

Therefore, the estimator

$$\text{var}_L\left[\hat{\underline{Z}}\right]$$

may be adopted for calculating

$$\text{var}_L\left[\sum_{i=1}^{n} w_i \underline{Z}_i\right],$$

when considering the sampling scheme described in this example. In the current context, expression (2.38) could thus be plugged in (2.32) and hence yielding an expression for $\text{var}_L[\hat{\underline{\beta}}(\text{V})_{PML}]$. ■

See Lavange, Koch and Schwartz (2001), for example, for applications of a similar procedure to allowing for complex sampling schemes in regression analyses of repeated measures data from various longitudinal studies.

Example 2.3: independent sampling of units

Under independent sampling of units, expression (2.23b) extends to

$$\begin{aligned}\operatorname{var}\left[\underline{\hat{\beta}}_{PML}(\mathrm{V})-\underline{\beta}\right]&=\left(\sum_{i=1}^{n} w_i \mathrm{X}_i' \mathrm{V}^{-1}\mathrm{X}_i\right)^{-1}\cdot\sum_{i=1}^{n}\operatorname{var}\left(w_i \mathrm{X}_i' \mathrm{V}^{-1}\underline{\varepsilon}_i\right)\cdot\left(\sum_{i=1}^{n} w_i \mathrm{X}_i' \mathrm{V}^{-1}\mathrm{X}_i\right)^{-1}\\ &=\left(\sum_{i=1}^{n} w_i \mathrm{X}_i' \mathrm{V}^{-1}\mathrm{X}_i\right)^{-1}\cdot\left[\sum_{i=1}^{n} w_i^2 \mathrm{X}_i' \mathrm{V}^{-1}\operatorname{var}(\underline{\varepsilon}_i)\mathrm{V}^{-1}\mathrm{X}_i\right]\cdot\left(\sum_{i=1}^{n} w_i \mathrm{X}_i' \mathrm{V}^{-1}\mathrm{X}_i\right)^{-1}.\end{aligned} \tag{2.39}$$

The last expression in (2.39) yields the robust variance estimator of $\underline{\hat{\beta}}_{PML}(\mathrm{V})$, given by

$$\operatorname{var}_r\left[\underline{\hat{\beta}}_{PML}(\mathrm{V})\right]=\left[\sum_{i=1}^{n} w_i \mathrm{X}_i' \mathrm{V}^{-1}\mathrm{X}_i\right]^{-1}\left[\sum_{i=1}^{n} w_i^2 \left(\mathrm{X}_i' \mathrm{V}^{-1}\underline{\hat{\varepsilon}}_i\right)\left(\mathrm{X}_i' \mathrm{V}^{-1}\underline{\hat{\varepsilon}}_i\right)'\right]\left[\sum_{i=1}^{n} w_i \mathrm{X}_i' \mathrm{V}^{-1}\mathrm{X}_i\right]^{-1} \tag{2.40}$$

where,

$$\underline{\hat{\varepsilon}}_i=\underline{y}_i-\mathrm{X}_i\underline{\hat{\beta}}_{PML}(\mathrm{V}).$$

The variance estimator given in (2.40) does not account for complex survey schemes. As in Sub-section 2.3.2 we also assume here that observations at two different individuals are independent. ■

Taylor series approximation of variances for complex statistics could be performed through different asymptotically equivalent procedures (Binder, 1995). In fact, Taylor series methods are responsible exclusively for the construction of a linear approximation of the statistic of interest, $\underline{\hat{\beta}}(\mathrm{V})_{PML}$ here. The variance expression, suitable for the actual survey sampling design, could then be employed for the linear approximation.

This method presents a very important property – it allows for the possibility that a clustered sampling design could generate data with a correlated error structure, and also for the likeliness of unequal unit variances both within and across clusters (Kott, 1991).

Nevertheless, the linearization method has also disadvantages: (*i*) the necessity for producing analytic expressions for the partial derivatives (see Skinner, Holt and Smith, 1989, Chapters 3 and 5); and (*ii*) it is biased because the number of clusters is finite, and the bias tends to be downward (Wolter, 1985, Chapter 8; and Kott, 1991).

Another difficulty with using Taylor series expansion to approximate the variance of $\hat{\underline{\beta}}$ may arise when the number of primary sampling units is 'small'. That could result in reduced degrees of freedom for the variance estimation and consequently the possibility for a 'weak' statistical inference (Korn and Graubard, 1995).

2.5.2 Jackknife variance estimator

The jackknife is a pseudo-replication method for variance estimation. The use of this class of variance estimation methods is nowadays considerably widespread among survey statisticians. The introduction of the jackknife estimation technique is attributed in the statistical literature to J. W. Tukey in the 1950's, who extended the bias reduction in parametric estimation methods previously developed by M. H. Quenouille in papers published in the 1940's and 50's (for full references see Miller, 1974).

In early stages, one of the jackknife's predominant applications had been to ratio estimation (see Durbin, 1959, among others). Tukey's original idea has then been described in details by Miller (1964), where we can also find further investigation mainly on the use of jackknife for the construction of approximate confidence intervals and significance tests, considering a simple non-linear parameter that is locally linear in the observations.

Additionally, Miller (1974) gives several other important references and presents a comprehensive literature review on early developments of the jackknife technique including its application to (*i*) inference on variances, (*ii*) regression problems, (*iii*) maximum likelihood estimation, (*iv*) transformation of statistics, and (*v*) multivariate analysis. Note that Wolter (1985; Chapter 4) also provides a comprehensive coverage of the topic.

Moreover the jackknife principle has been extended to handle stratified multistage sampling data in successive papers by (*i*) Lee (1973), for combined ratio estimators, where this method has been referred to as generalised repeated partial sample; (*ii*) Jones (1974), for the class of estimators that can be expressed as functions of sample means; (*iii*) Kish and Frankel (1974), where it is referred as jackknife repeated replication method; and (*iv*) Krewski and Rao (1981), where the consistency of jackknife variance estimators for both linear and non-linear statistics are theoretically established; among other papers.

A jackknife estimator may also perform variance estimation for $\hat{\underline{\beta}}_{PML}(\mathrm{V})$ by estimating numerous estimates of the population target parameter from distinct parts of the original sample. For the stratified multistage sampling case, we produce each estimate by removing a primary sampling unit (PSU) at a time (as in Jones, 1974), and its units from all subsequent sampling stages, in stratum h in order to make one replication. The variance estimation is thus obtained by calculating the variability among all the replicate estimates of $\underline{\beta}$ (see Drewski and Rao, 1981). Therefore (Shah, Barnwell, and Bieler, 1997),

$$\mathrm{var}_J\left[\hat{\underline{\beta}}_{PML}(\mathrm{V})\right]=\sum_{h=1}^{H}(m_h-1)/m_h\sum_{j=1}^{m_h}\left[\hat{\underline{\beta}}(\mathrm{V})_{PML(hj)}-\hat{\underline{\beta}}(\mathrm{V})_{PML(h.)}\right]^2, \tag{2.41}$$

where the subscript J denotes jackknife, H is the total number of strata in the sample, m_h is the number of PSUs in stratum h, $\hat{\underline{\beta}}_{PML(hj)}$ is the pseudo maximum likelihood estimator of β after omitting PSU j in stratum h,

$$\hat{\underline{\beta}}(\mathrm{V})_{PML(h.)}=\frac{1}{m_h}\sum_{j=1}^{m_h}\hat{\underline{\beta}}(\mathrm{V})_{PML(hj)}, \tag{2.42}$$

and $\hat{\underline{\beta}}(\mathrm{V})_{PML(hj)}$ is estimated using (2.27).

When applying the Jackknife variance estimation method, the sampling weights have to be inflated in the remaining PSUs, so that those become representative of all the elements (Lee, Forthofer and Lorimor, 1989). This may be obtained by dividing the sum of the sampling weights of the remaining PSUs by the factor

$$1-(sum1/sum2)$$

where $sum1$ is the sum of the sampling weights for individuals in the excluded PSU and $sum2$ is the sum of the weights for individuals in all PSUs in stratum h. We shall adopt this procedure for defining the replication weights when applying the Jackknife method in the next chapter.

Jackknife repeated replication methods usually work under the assumption that the PSUs are selected independently and with replacement. The adoption of jackknife for variance estimation for situations where PSUs are not selected independently could be a source of bias (Rust, 1985, Section 7, and references therein). That assumption could be substituted by an alternative presupposition that the *sampling fraction*[15] in every first stage stratum is 'small', say less than 10% (Shah, Barnwell, and Bieler, 1997).

In the next chapter of this thesis, we shall have a chance to perform an empirical comparison between jackknife and linearization variance estimation methods for a data set with a specific sampling design in both cross-sectional and longitudinal contexts.

2.6 Misspecification effects

Kish (1965) introduced the term design effect for a measure of the impact of a complex design, which is usually defined as the variance of a statistic under the given design divided by the variance of the corresponding statistic under a srs design.

Kish's design effect $deff_{Kish}(\hat{\theta})$ of $\hat{\theta}$, where here $\hat{\theta}$ is an estimator of any scalar parameter of interest θ, is thus given by

$$deff_{Kish}(\hat{\theta}) = \frac{\mathrm{VAR}_{true}(\hat{\theta})}{\mathrm{VAR}_{\mathrm{srs}}(\hat{\theta})},$$

where $\mathrm{VAR}_{true}\left(\hat{\theta}\right)$ is the *true* variance of $\hat{\theta}$ that considers the *true* sampling scheme used for the selection of the sample, and $\mathrm{VAR}_{\mathrm{srs}}\left(\hat{\theta}\right)$ is the hypothetical variance of $\hat{\theta}$ when the sample was selected by srs with replacement.

The original design effect measure present some disadvantages for analysis purposes. Kish's design effect has its applicability related mainly to the evaluation of precision gains or losses, when comparing alternative sampling schemes (see Lee, Forthofer and Lorimor, 1989). The measure $deff_{Kish}(\hat{\theta})$ would actually not be appropriate for analysis under one single design, according to Skinner, Holt and Smith (1989).

---

[15] The sampling fraction in a stratum is calculated as the number of PSUs selected in the sample divided by the population number of PSUs in the stratum (Shah, Barnwell, and Bieler, 1997).

Skinner (1989b) extends the concept of design effect by proposing the *misspecification effect* (*meff*), which is more suitable for analytic inference than the classic design effect. The *meff* is designed to measure the effects of incorrect specification of both the sampling scheme and the considered model.

The *meff* is defined in the same way as the *deff* with the same numerator, but with the denominator consisting of the expectation of a variance which ignores the complex design. Thus let (Skinner, Holt and Smith, 1989)

$$\mathrm{var}_0\left(\hat{\theta}\right) = \mathrm{var}_{\mathrm{iid}}\left(\hat{\theta}\right)$$

be a consistent estimator of the variance of $\hat{\theta}$, when we assume that the observations are independent and identically distributed (iid). The effect of the complex sampling scheme on $\mathrm{var}_0\left(\hat{\theta}\right)$ can be evaluated if we examine its distribution.

Skinner, Holt and Smith (1989) thus define the misspecification effect as

$$meff(\hat{\theta}, \mathrm{var}_0) = \frac{\mathrm{VAR}_{true}(\hat{\theta})}{E_{true}[\mathrm{var}_0(\hat{\theta})]},$$

which gives a measure of how much $\mathrm{var}_0(\hat{\theta})$ over- or underestimate $\mathrm{VAR}_{true}(\hat{\theta})$, and may be estimated, for example by

$$m\hat{e}ff(\hat{\theta}, \mathrm{var}_0) = \frac{\mathrm{var}_L(\hat{\theta})}{\mathrm{var}_0(\hat{\theta})}, \qquad \text{or} \qquad m\hat{e}ff(\hat{\theta}, \mathrm{var}_0) = \frac{\mathrm{var}_J(\hat{\theta})}{\mathrm{var}_0(\hat{\theta})}.$$

The *meff*, defined above, is thus a measure of relative bias of the variance estimator. Figure 2.1 below gives information on how to interpret misspecification effect results.

$meff\left(\hat{\theta}, \mathrm{var}_0\right)$	Bias [$\mathrm{var}_0\left(\hat{\theta}\right)$]	Interpretation
< 1	> 0	Overestimation of $\mathrm{VAR}(\hat{\theta})$
$= 1$	$= 0$	Correct estimation of $\mathrm{VAR}(\hat{\theta})$
> 1	< 0	Underestimation of $\mathrm{VAR}(\hat{\theta})$

Figure 2.1 – $meff\left(\hat{\theta}, \mathrm{var}_0\right)$ interpretation.

We may observe, for example, that: (*i*) $meff\left(\hat{\theta}, \text{var}_0\right) < 1$ when estimating $\text{var}_0(\hat{\theta})$ and ignoring stratification; and (*ii*) $meff\left(\hat{\theta}, \text{var}_0\right) > 1$ when estimating $\text{var}_0(\hat{\theta})$ and ignoring *sampling weights* or clustering. See Skinner, Holt and Smith (1989, Chapter 2), for some examples.

The misspecification effect proposed by Skinner, Holt and Smith (1989) may be either a design based or a model based measure. We may take both $\text{VAR}_{true}(\hat{\theta})$ and $E_{true}[\text{var}_0(\hat{\theta})]$ to refer to either the randomization distribution induced by the true sampling design or by the model.

The $meff(\hat{\theta}, \text{var}_0)$ is theoretically different from the $deff_{Kish}(\hat{\theta})$, because that depends upon two arguments $\hat{\theta}$ and var_0 (Skinner, Holt and Smith, 1989). However we may observe that their respective estimators, calculated from a unique sample coincide in practice.

In the context of the current chapter, we may define the misspecification effect as

$$meff = \frac{\text{VAR}[\underline{\hat{\beta}}(\text{V})_{PML,\kappa}]}{\text{var}_r[\underline{\hat{\beta}}(\text{V})_{PML,\kappa}]},$$

which may be estimated by replacing VAR by var_J or var_L , so that

$$m\hat{e}ff^{J} = \frac{\text{var}_J[\underline{\hat{\beta}}(\text{V})_{PML,\kappa}]}{\text{var}_r[\underline{\hat{\beta}}(\text{V})_{PML,\kappa}]},$$

or (2.43)

$$m\hat{e}ff^{L} = \frac{\text{var}_L[\underline{\hat{\beta}}(\text{V})_{PML,\kappa}]}{\text{var}_r[\underline{\hat{\beta}}(\text{V})_{PML,\kappa}]},$$

where $\underline{\hat{\beta}}(\text{V})_{PML}$ is as defined for (2.27) in Section 2.4, and the subscript κ denotes that $\underline{\hat{\beta}}(\text{V})_{PML,\kappa}$ is the κ^{th} element of $\underline{\hat{\beta}}(\text{V})_{PML}$, with $\kappa = 1, \ldots, q$.

Both $m\hat{e}ff^{J}$ and $m\hat{e}ff^{L}$ measure the influence of a sample design on inference procedures, although they do not account for the effects of the longitudinal structure of the data on inference procedures. In general, *meffs* will reflect the impact of weighting, clustering and stratification.

Note that we should always be cautious when interpreting estimated misspecification effect results, considering that (2.43) are estimators and therefore are subject sampling variability. For some discussion on this subject as well as empirical results, see Yu (2002).

2.7 Software

The statistical specialist sampling package SUDAAN (Shah, Barnwell, and Bieler, 1997) is an alternative to be adopted for the application of the techniques described in this chapter. SUDAAN was originally developed for analysis of complex survey data. This is one of the few software which supports for multiple nesting levels (see Horton and Lipsitz, 1999). This characteristic makes SUDAAN to be particularly appropriate for analysis of repeated measures and clustered data.

SUDAAN's *regress* procedure fits linear regression models to sample survey data, which could be either cross-sectional or longitudinal. Survey data longitudinal model parameters can be estimated using either expression (2.27) for $\hat{\underline{\beta}}(\mathrm{V})_{PML}$, with the working covariance matrix V assumed to be exchangeable (see Section 2.4), or (2.28) for $\hat{\underline{\beta}}(I)_{PML}$, which assumes that $\mathrm{V} = I$, depending on the procedure that is adopted in SUDAAN.

In addition, when assuming the classical case $\hat{\underline{\beta}}(\mathrm{V})$ may also be calculated by this software, which adopts either (2.15) considering the working covariance matrix V to be exchangeable (see Section 2.2) or (2.17), with $\hat{\underline{\beta}}(\mathrm{V} = I)$.

Furthermore SUDAAN is able to perform variance estimation for $\hat{\underline{\beta}}(\mathrm{V})_{PML}$, considering a linearization variance estimator (see Sub-section 2.5.1) and a jackknife variance estimator (see Sub-section 2.5.2). A weighted least square (see Sub-section 2.3.1) and robust variance estimator (see Sub-section 2.3.2) can also be calculated for $\hat{\underline{\beta}}(\mathrm{V})$ by this software.

When calculating $\hat{\underline{\beta}}_{PML}(\mathrm{V})$ by adopting (2.27), and estimating $\hat{\underline{\beta}}_{PML}(\mathrm{V})$'s variance via jackknife or linearization methods, this software should be able to consider both individual-level correlations over time and between cluster variability.

Therefore, we shall consider the utilization of SUDAAN for applying some of the techniques discussed in the current chapter in Chapter 3, including linearization variance estimation.

However, jackknife variance estimates in the longitudinal context shall not be produced by using the software described in the current section. We shall adopt the statistical computer software *R* (Ihaka and Gentleman, 1996) for programming this technique. We shall provide some description of this software later in this thesis, in Chapter 5. Moreover, see Zhang (2001) for some information on the limitations of using SUDAAN for variance estimation.

Other softwares that could handle complex survey data are, for example: STATA (svy commands; Stata Corp., 2003); SAS (PROC SURVEYREG; SAS Institute, 2004); and a 'Complex Sample add-on' more recently implemented for SPSS version 14.0 (SPSS Inc, 2005). Note that STATA (xt commands; Stata Corp., 2003), SAS (PROC GENMOD; SAS Institute, 2004), and S-PLUS (YAGS library or Oswald system; see Horton and Lipsitz, 1999), for example, could be adopted for fitting population-averaged or marginal longitudinal models, albeit without allowing for effects of sample clustering on variance estimation. Nevertheless, to our knowledge, none of the software listed above offer readily available tools that could be utilised for modelling longitudinal survey data by accounting for correlations over time as well as clustering, using the techniques discussed in the current chapter, which belong to a survey sampling approach.

Notice, that any standard software's longitudinal modelling procedure could be adopted for point estimation of $\underline{\beta}(\mathrm{V})$ by calculating $\hat{\underline{\beta}}_{PML}(\mathrm{V})$ given by (2.27a). This, however, does not eliminate the necessity of estimating $\hat{\underline{\beta}}_{PML}(\mathrm{V})$'s variance via methods that take clustering into account.

An alternative set of techniques for modelling longitudinal survey data allowing for clustering is the multilevel modelling approach (see, for example, Goldstein, 1995), which we shall not explicitly discuss in this Thesis. Thus, MLwiN (Rasbash *et al.*, 2002) or SAS (PROCs MIXED, GLIMMIX, and NLMIXED; SAS Institute, 2004), for example, could be adopted for handling both longitudinal correlations and clustering in a multilevel modelling set up.

2.8 Concluding remarks

In this chapter we have mainly presented a literature review on longitudinal regression models in the context of complex survey data. We have given information on the estimation of longitudinal regression model parameters and on how to estimate their variances considering initially the classical case.

We have additionally discussed methods for situations where we allow for complex survey schemes. We have considered the pseudo likelihood estimation method for model parameter estimation. A robust estimator for $\underline{\hat{\beta}}_{PML}$, which does not account for complex survey schemes, has then been discussed. In addition, we have given information on both jackknife replication variance estimation and Taylor linearization variance estimation approaches.

Moreover we have reviewed the main aspects of the misspecification effect, which we shall adopt in Chapter 3 as a measure of the influence of a sample design on longitudinal analytic inference procedures.

Furthermore we have pointed out in Section 2.7 the main characteristics of the software Sudaan, which is a statistical package that we shall mainly adopt in the following chapter for an empirical application, which shall involve the adoption of most of the techniques discussed in the current chapter.

Hence, we shall carry out in the following chapter an empirical investigation about variance effects of clustering in longitudinal studies, using longitudinal survey data from the British Household Panel Survey (BHPS) applying methods discussed in the current chapter.

Chapter 3

Variance effects of clustering in longitudinal studies – an empirical investigation

3.1 Introduction

In this chapter we shall study the misspecification effects (*meffs*), discussed in Chapter 2, Section 2.6. We expect *meffs* to behave empirically like Kish's design effects (*deffs*). There is some empirical evidence that design effects from clustering tend to decrease the more complex the analysis (Kish and Frankel, 1974). For example, design effects for regression coefficients are often found to be less than design effects for the mean of the dependent variable in the regression. An intuitive reason for such findings is that a regression model may 'explain away' certain differences between strata or multi-stage units thus diminishing the residual impact of the design. Evidence of design effects close to unity for such analyses may be used by some analysts of survey data to justify ignoring the sampling design in complex analyses.

In this chapter we present some evidence of an opposite tendency, for design effects to be higher for complex longitudinal analyses than for corresponding cross-sectional analyses. The design effects for some of the regression coefficients are found to increase the more waves are included in the analysis. A similar tendency is observed for estimates of the time-averaged mean of the dependent variable. The implication of these findings is that standard errors in analyses of longitudinal survey data could be misleading if the initial sample was clustered and if this clustering is ignored in the analysis. Moreover, it is our aim to provide some possible theoretical justification for that result.

Our empirical evidence is based upon data from the British Household Panel Study (BHPS). This survey follows longitudinally a sample of individuals selected in 1991 by two-stage sampling, with clustering by area. Data are collected in annual waves. Our analyses are based upon a subsample of women aged 16-39. The dependent variable is a gender role attitude score, derived from responses to six five-point questions, and treated as a continuous variable. Covariates include age group, economic activity and educational qualifications. Data are analysed for five waves of the survey when the gender role attitude questions were asked.

Methods discussed in Chapter 2 shall be applied in the empirical investigation presented in this chapter. We shall adopt longitudinal regression modelling, based upon a model considered by Berrington (2002), who studies persistence and changes in the gender role attitudes of women in Britain. Regression model fitting results allowing for both (*i*) longitudinal structure; (*ii*) clustering, (*iii*) stratification and (*iv*) sampling weights shall be produced in this chapter. Moreover, we shall also present values of design effects that capture only the effect of clustering, treating the weights as constant and ignoring stratification. Furthermore, jackknife (see Chapter 2, Sub-section 2.5.2) and Taylor linearization variance estimation (see Chapter 2, Sub-section 2.5.1) results shall be included here.

In this chapter we shall assume that the parameters of interest should not depend upon the sampling design, via the population structure underlying the sampling (Skinner, Holt and Smith, 1989). The reason for the adoption of this assumption is that the primary sampling units (PSUs) in the BHPS are postcode sectors, determined by the needs of the British postal system. Thus, we shall assume here that BHPS PSUs are not relevant to the definition of parameters of scientific interest.

The main characteristics of BHPS data set, including its *sampling design* features, shall be presented in details in Section 3.2, which also includes some further information on how the study variable considered in this chapter was originally constructed.

Section 3.3 mainly identifies and illustrates variance effects of clustering for longitudinal studies, but also includes some exploratory data analysis and cross-sectional model fitting results. It is in that section that we shall provide some evidence that the design effects for longitudinal analyses can be greater than for corresponding cross-sectional analyses, implying that more caution is required before ignoring the complex design in standard error estimation. Moreover, we present a succinct interpretation for some achieved model fitting results.

A theoretical discussion about the major results presented in this chapter is presented in Section 3.4. Finally, Section 3.5 includes some concluding remarks.

3.2 British Household Panel Survey

We have been able to notice great modifications in the role of men and women in the family in numerous countries over the last few decades. The relationship between changes in the gender role attitudes and behaviour modifications, as for example parenthood and labour force participation, has been an issue of great interest for social scientists (see, for example, Morgan and Waite, 1987; Fan and Marini, 2000; Berrington, 2002; and more recently Brooks and Bolzendahl, 2004).

We analyse here data from the BHPS, which is an on-going large nationally representative household panel survey of individuals in private domiciles[16], carried out by the Institute for Social and Economic Research, University of Essex (see Taylor *et al.*, 2001; and Berrington, 2002). The BHPS exemplifies the use of complex sampling as well as longitudinal schemes. This survey has the main objective of providing information about social and economic change at the individual and household level, "*to identify, model and forecast such changes, and also to investigate their causes and consequences in relation to a range of socio-economic variables*" (Taylor *et al.*, 2001).

3.2.1 *Sampling design*

The BHPS is a longitudinal survey and adopts a complex multistage *sampling scheme* for collecting data. It has additionally a multiple cohort *prospective* panel design. At wave one, in 1991, it involved an (*i*) approximately equal probability selection of households, and a (*ii*) multistage stratified clustered probability design with systematic sampling.

As PSUs, 250 postcode sectors were selected, with replacement and with probability of selection proportional to size using a systematic procedure. *Delivery points*[17] were selected as secondary sampling units, with the adoption of an analogous systematic procedure.

In addresses with up to 3 households present, all households were included, and in those with more than 3 households, a random selection procedure, using a Kish grid, was

[16] BHPS data is collected *prospectively* (see Chapter 1, Sub-section 1.3.2), which has usually better quality than data collected *retrospectivelly* (see Diggle *et al.*, 2002).
[17] The BHPS sampling design defines *delivery points* as addresses.

used for the selection of 3 households. Then, all resident household members aged 16 or over were selected. All adults selected at wave one, were followed from wave two and beyond. A result of this sampling design is that inclusion probabilities of adults have little variability.

A considerable number of PSUs are sampled for the BHPS. This protects against the occurrence of limited degrees of freedom when estimating variances (see Chapter 2, Sub-section 2.5.1). Korn and Graubard (1995), for example, recommend *sampling designs* with 'many' PSUs for more efficient variance estimation.

The BHPS data set includes *longitudinal weights* w_{iT}^{*} (see Chapter 1, Section 1.3), which are provided to individual cases who have responded at each wave up to and including the latest wave. Those weights allow for different selection probabilities, nonresponse at wave one and attrition. Cross-sectional weights were also produced for each wave and are also provided. For information regarding how weights are defined for the BHPS case, including some expressions, see Taylor *et al.* (2001), where further details about the BHPS's *sampling design* are also included.

3.2.2 BHPS subset

We consider in this study data from waves one, three, five, seven, and nine (collected respectively in 1991, 1993, 1995, 1997, and 1999). The subset of data adopted here was obtained and prepared with computational procedures adapted from Berrington (2002), which were provided by that author. We use in the analysis presented here the following BHPS subset of variables, (*i*) gender role attitude score, (*ii*) wave number – t, which is the time variable, (*iii*) age group, (*iv*) economic activity, (*v*) parenthood status, and (*vi*) educational level.

Covariates for the regression analysis that shall be performed in the current chapter were chosen on the basis of discussion in Berrington (2002) but cut down in number to facilitate a focus on the methodological questions of interest. We consider economic activity as the exploratory variable of highest scientific importance. That covariate discriminates between women who are at home looking after children and those following other kinds of activity in relation to the labour market. Two other covariates, age and education, are also included in our analysis because these have been considered to be related to gender role attitude as, for example, in Fan and Farini (2000).

Note that we shall thus include both age and time as candidate covariates in our model fitting procedures. This is only possible in longitudinal analysis. Hence, by doing that we try to control for both aging (developmental effect) and trend over time (historical effect) when studying variables like the women gender role attitudes. This strategy is so-called Palmore's method. Ideally we should also include the birth cohort factor as explanatory variable, but it is usually not done due to colinearity problems (Menard, 1991; and Firebaugh, 1997, Chapter 2).

The levels of each categorical variable adopted in the current chapter are described below in Table 3.1.

Variable	Values	Labels
Age group	1	16-21 years
	2	22-27 years
	3	28-33 years
	4	More than 34 years
Economic activity	1	Full time employed
	2	Part time employed
	3	Other inactive
	4	Full time student
	5	Family care
Parenthood status	0	Not a parent
	1	Parent
Educational level	1	First or higher degree
	2	Teaching qualification (QF) or other higher QF
	3	Nursing QF or GCE A levels
	4	GCE O levels or equivalent
	5	Other*

* Includes the following original categories: 'commercial QF', 'CSE Grade 2-5', 'apprenticeship', 'other QF', 'no QF', and 'still at school'.

Table 3.1 – Variable's categories and their respective labels.

The BHPS subset considered in this chapter includes additionally some *design variables*, as the cross-wave person identifier, strata and primary sampling unit.

Given our interest in whether women's primary labour market activity is 'caring for a family', we consider women aged 16-39, in 1991, in Britain, as our *study population*. Only women with full interview outcome (complete records) in all five waves are included in our data set, i.e. no case with panel attrition is considered so far. Then, we have also a balanced design because the number of observations in the sample is the same for all $n = 1340$ subjects. The individuals included in the adopted data subset are distributed fairly evenly across $m = 248$ postcode sectors (PSUs), where m is the number of clusters in the sample. As $T = 5$, where T is the number of waves, then $n_o = n \cdot T = 6700$ (see Chapter 2, Section 2.1), where n_o is thus the total number of rows in the data set.

The small average sample size of approximately five per PSU associated with the somewhat low intra cluster correlation for the variable of interest leads to relatively small impacts of the sampling design, as measured by misspecification effects (*meffs*). Since our aims in this chapter are mainly methodological ones, i.e. to compare *meffs* for different analyses, we have decided to group the postcode sectors into 47 geographically contiguous clusters, in order to create clearer comparisons, less confused by sampling errors which could be considerable when performing variance estimation.

3.2.3 Gender role attitudes

The women's role attitudes score is the *variable of interest* or the response variable in this chapter and that is a measure originally considered by Berrington (2002). In waves one, three, five, seven and nine, respondents were requested to complete a questionnaire with the same affirmations concerning the family, women's roles, and work out of the household (see Appendix A). The respondents were advised to express the magnitude to which they 'strongly agree', 'agree', 'neither agree nor disagree', 'disagree' or 'strongly disagree' with each affirmation. The statements were coded on a range from 1 to 5.

Factor analysis was adopted to establish which statements would be composite into a gender role attitude measure. The attitude score was thus constructed from the total score to the first six statements. Higher scores signify more egalitarian gender role attitudes[18]. Further information on how the variable attitude score was originally constructed may be found in Berrington (2002, Sub-section 10.3.4). For an example of an application involving gender role attitudes see Firebaugh (1997).

3.3 Clustering effects for longitudinal studies

It is our main objective in this chapter to identify variance effects of clustering for longitudinal studies. This section is organized as follows. Sub-section 3.3.1 presents some exploratory data analysis whilst Sub-sections 3.3.2 and 3.3.3 include some model fitting results with non-weighted parameter estimates in both cross-sectional and longitudinal contexts respectively.

To allow for the minor variations in the selection probabilities as well as wave one non response and attrition we shall use the *longitudinal sampling weights* provided at wave nine, denoted by w_{i9}^{*}, in all the weighted parameters calculated in the current chapter. Recall that the data set we adopt in this chapter includes only cases for individuals who have responded at each wave up to and including wave nine.

3.3.1 Exploratory data analysis

We present briefly some exploratory data analysis for the gender role attitude score in terms of three candidate covariates to be included in our model – age group, economic activity and educational level, which are all time-varying. See also Appendix B for an examination of the distribution of the study variable of interest y, and a graphical display of the repeated observations for the response variable for each individual.

Furthermore, for the following three tables, jackknife variance estimation for the estimated mean is performed taking the actual BHPS's complex survey design into account. See Chapter 2, Section 2.5, Sub-section 2.5.2 for information on jackknife variance estimation regarding regression model parameters estimators. For a jackknife

[18] The codes for affirmations 3, 4 and 5 (see Appendix A) were inverted.

variance of the estimator of the mean, substitute $\underline{\hat{\beta}}_{PML}$ with $\bar{y}_t$ in (2.41) and (2.42), where $\bar{y}_t$ is a weighted estimator for the mean, $\bar{Y}_t$, at wave t given by (Skinner, 2003b)

$$\bar{y}_t = \sum_{i=1}^{n} w_{i9}^* y_{it} \Bigg/ \left(\sum_{i=1}^{n} w_{i9}^* \right), \tag{3.1}$$

where $\bar{y}_t$ is a scalar, n is the sample size and y_{it} is the value of the variable of interest for individual i at wave t. The jackknife variance estimation for $\bar{y}_t$ shall be denoted here by $\text{var}_J[\bar{y}_t]$.

To assess the misspecification effect, we also calculate the unweighted estimator of the mean, denoted by $\bar{y}_t^*$, and assuming that the sample is selected by simple random sampling (srs), its well known standard variance estimator is

$$\text{var}[\bar{y}_t^*] = \frac{sd_t}{\sqrt{n}}, \tag{3.2}$$

where sd_t is the standard deviation of y considering wave t, and n is the sample size. In (3.2), both $\bar{y}_t^*$ and sd_t are scalars.

Let $\sqrt{\text{var}[.]}$ be an estimator of the standard error (s.e.) of the estimator of our target parameter. Tables from 3.2 to 3.4 show the attitude score weighted mean, $\bar{y}_t$, calculated using (3.1) and the sample size n for each considered wave by 'age group', 'economic activity', and 'educational level' respectively. Also shown are: (*i*) $\bar{y}_t$'s jackknife standard error (s.e.) estimates, given by $\sqrt{\text{var}_J[\bar{y}_t]}$; (*ii*) $\bar{y}_t^*$'s estimated s.e. given by $\sqrt{\text{var}[\bar{y}_t^*]}$, which is calculated from expression (3.2), considering that the sample is selected by srs; and (*iii*) misspecification effects - *meff* (see Chapter 2, Section 2.6). Note that $\bar{y}_t$'s linearization s.e. estimates are found to be very similar to $\sqrt{\text{var}_J[\bar{y}_t]}$, and therefore shall not be presented below.

For evaluating the influence of the BHPS *sampling design* on the particular cross-sectional inference procedures presented in Tables from 3.2 to 3.4, we shall adopt $meff^J$, which shall be calculated in the present situation as (see Chapter 2, Section 2.6)

$$\hat{meff}^J = \frac{\text{var}_J[\bar{y}_t]}{\text{var}[\bar{y}_t^*]}. \tag{3.3}$$

Note that $\bar{y}_t^*$ in (3.3) is not a weighted estimator, and that the variance estimator in the denominator of $\hat{meff}^J$ defined above is not a robust estimator, as considered in expression (2.43) in Chapter 2, Section 2.6.

Table 3.2 below gives results for mean attitude score by 'age group'.

Age Group	Statistics		Attitude Score: Wave 1	Wave 3	Wave 5	Wave 7	Wave 9
16-21	$\bar{y}_t$		21.67	21.55	21.6	22.59	-
	n		229	146	64	2	0
	s. e. mean	$\sqrt{\text{var}_J[\bar{y}_t]}$	0.28	0.30	0.49	3.63	-
		$\sqrt{\text{var}[\bar{y}_t^*]}$	0.25	0.29	0.46	3.00	-
	$\hat{meff}^J$		1.25	1.07	1.13	1.52	-
22-27	$\bar{y}_t$		20.12	20.43	20.47	20.72	20.54
	n		370	311	273	229	149
	s. e. mean	$\sqrt{\text{var}_J[\bar{y}_t]}$	0.22	0.19	0.24	0.27	0.31
		$\sqrt{\text{var}[\bar{y}_t^*]}$	0.20	0.20	0.23	0.24	0.29
	$\hat{meff}^J$		1.21	0.90	1.09	1.27	1.14
28-33	$\bar{y}_t$		19.75	19.72	19.77	19.79	20.12
	n		416	435	406	371	311
	s. e. mean	$\sqrt{\text{var}_J[\bar{y}_t]}$	0.20	0.20	0.21	0.21	0.23
		$\sqrt{\text{var}[\bar{y}_t^*]}$	0.21	0.19	0.18	0.18	0.21
	$\hat{meff}^J$		0.91	1.11	1.36	1.36	1.20
More than 34	$\bar{y}_t$		19.53	19.62	19.43	19.45	19.40
	n		325	448	597	738	880
	s. e. mean	$\sqrt{\text{var}_J[\bar{y}_t]}$	0.24	0.18	0.15	0.14	0.12
		$\sqrt{\text{var}[\bar{y}_t^*]}$	0.21	0.17	0.15	0.13	0.12
	$\hat{meff}^J$		1.31	1.12	1.00	1.16	1.00

Table 3.2 – Mean attitude score by age group.

The misspecification effects are reasonably close to one (see Figure 2.1, in Chapter 2, Section 2.6) for most of the age categories, in most of the waves. Two of the results for $\hat{meff}^J$ are slightly smaller than one and some of them are approximately equal to one. The $\hat{meff}^J$'s largest values are 1.52 for age '16-21' in wave 7, and 1.36 for age '28-33' in waves 5 and 7. Thus, we could say that in general the influence of the BHPS *sampling design* on this particular cross-sectional inference procedure is not very strong.

We may notice that older women are more traditional in their orientation than younger ones. Moreover, controlling by age group, the results show that the gender role attitude score does not vary substantially over time. Nevertheless, we may notice a slight positive trend over time in the attitude score mean for age groups '22-27' and '28-33'.

Table 3.3 gives results for mean attitude score by 'economic activity'.

Economic Activity	Statistics		Attitude Score				
			Wave 1	Wave 3	Wave 5	Wave 7	Wave 9
Full time employed	$\bar{y}_t$		21.59	21.35	20.03	20.99	20.98
	n		605	568	588	630	634
	s. e. mean	$\sqrt{\text{var}_J[\bar{y}_t]}$	0.15	0.14	0.17	0.16	0.14
		$\sqrt{\text{var}[\bar{y}_t^*]}$	0.15	0.14	0.15	0.14	0.14
	$m\hat{e}ff^J$		1.00	1.00	1.28	1.31	1.00
Part time employed	$\bar{y}_t$		19.41	19.28	19.57	19.26	19.06
	n		284	303	337	341	364
	s. e. mean	$\sqrt{\text{var}_J[\bar{y}_t]}$	0.22	0.19	0.19	0.18	0.18
		$\sqrt{\text{var}[\bar{y}_t^*]}$	0.21	0.18	0.18	0.17	0.17
	$m\hat{e}ff^J$		1.10	1.11	1.11	1.12	1.12
Other inactive	$\bar{y}_t$		19.20	20.24	20.32	19.66	19.12
	n		73	76	77	86	101
	s. e. mean	$\sqrt{\text{var}_J[\bar{y}_t]}$	0.51	0.47	0.42	0.40	0.49
		$\sqrt{\text{var}[\bar{y}_t^*]}$	0.47	0.44	0.41	0.34	0.41
	$m\hat{e}ff^J$		1.18	1.14	1.05	1.38	1.43
Full time student	$\bar{y}_t$		22.21	22.25	21.78	20.61	21.12
	n		85	71	39	21	16
	s. e. mean	$\sqrt{\text{var}_J[\bar{y}_t]}$	0.41	0.44	0.51	1.19	0.91
		$\sqrt{\text{var}[\bar{y}_t^*]}$	0.38	0.44	0.58	0.92	0.84
	$m\hat{e}ff^J$		1.16	1.00	0.77	1.67	1.17
Family care	$\bar{y}_t$		17.76	18.12	17.73	17.50	17.60
	n		293	322	299	262	225
	s. e. mean	$\sqrt{\text{var}_J[\bar{y}_t]}$	0.25	0.22	0.21	0.23	0.23
		$\sqrt{\text{var}[\bar{y}_t^*]}$	0.21	0.20	0.22	0.21	0.22
	$m\hat{e}ff^J$		1.42	1.21	0.91	1.20	1.09

Table 3.3 – Mean attitude score by economic activity.

In this case the misspecification effects are reasonably close to one for most of the economic activity groups, for almost all waves. Some of the $m\hat{e}ff^J$'s results are slightly smaller than or equal to one. The largest values are 1.67 for 'full time student' in wave 7 and 1.43 for 'other inactive' in wave 9. Hence, we could say that in general the influence of the BHPS *sampling design* on this particular cross-sectional inference procedure is also not very strong for most of the economic activity groups.

We may see that 'full time employed' and 'full time student' women have more egalitarian gender role attitudes than other categories. Furthermore, controlling by economic activity, we can not see any clear trend over time in the attitude score mean for most of the categories. Nevertheless, there is a possible slight negative trend over time in the attitude score mean for some economic activity groups. Table 3.4 below shows results for mean attitude score now controlled by 'educational level'.

Educational Level	Statistics		Attitude Score				
			Wave 1	Wave 3	Wave 5	Wave 7	Wave 9
First or higher degree	$\bar{y}_t$		21.30	21.21	21.13	20.73	20.59
	n		135	148	171	194	204
	s. e. mean	$\sqrt{\text{var}_J[\bar{y}_t]}$	0.33	0.29	0.36	0.34	0.26
		$\sqrt{\text{var}[\bar{y}_t^*]}$	0.34	0.29	0.29	0.27	0.26
	$m\hat{e}ff^J$		0.94	1.00	1.54	1.59	1.00
T. QF or other higher QF	$\bar{y}_t$		20.61	20.88	20.33	20.20	20.04
	n		159	222	268	309	382
	s. e. mean	$\sqrt{\text{var}_J[\bar{y}_t]}$	0.35	0.27	0.24	0.22	0.21
		$\sqrt{\text{var}[\bar{y}_t^*]}$	0.33	0.26	0.22	0.22	0.19
	$m\hat{e}ff^J$		1.12	1.08	1.19	1.00	1.22
Nursing QF or GCE A levels	$\bar{y}_t$		20.88	20.52	19.98	19.66	19.63
	n		247	247	235	207	177
	s. e. mean	$\sqrt{\text{var}_J[\bar{y}_t]}$	0.29	0.23	0.26	0.27	0.28
		$\sqrt{\text{var}[\bar{y}_t^*]}$	0.24	0.23	0.24	0.24	0.27
	$m\hat{e}ff^J$		1.46	1.00	1.17	1.27	1.08
GCE O levels or equivalent	$\bar{y}_t$		20.39	20.3	19.67	19.60	19.6
	n		434	388	356	336	306
	s. e. mean	$\sqrt{\text{var}_J[\bar{y}_t]}$	0.20	0.20	0.21	0.19	0.21
		$\sqrt{\text{var}[\bar{y}_t^*]}$	0.18	0.18	0.20	0.19	0.19
	$m\hat{e}ff^J$		1.23	1.23	1.10	1.00	1.22
Other	$\bar{y}_t$		19.18	19.12	19.19	19.24	18.96
	n		365	335	309	292	265
	s. e. mean	$\sqrt{\text{var}_J[\bar{y}_t]}$	0.22	0.22	0.20	0.24	0.25
		$\sqrt{\text{var}[\bar{y}_t^*]}$	0.21	0.20	0.22	0.22	0.23
	$m\hat{e}ff^J$		1.10	1.21	0.83	1.19	1.18

Table 3.4 – Mean attitude score by educational level.

For the table above, the misspecification effects are also sufficiently close to one for most of the economic activity groups in almost all the waves. Two of the $\hat{meff}^J$'s results are either slightly smaller than or approximately equal to one. The largest values are 1.59 and 1.54 for the category 'First or higher degree', in waves 7 and 5 respectively. Therefore, we could say that in general the influence of the BHPS *sampling design* on this particular cross-sectional inference procedure is also not very strong for most of the educational level groups.

From Table 3.4's results we may affirm that more educated women tend to have more egalitarian gender role attitudes than the less educated ones. Moreover, controlling by educational level, we can not see any strong trend over time in the attitude score mean for most of the categories. Nevertheless, there is a possible slight negative trend in the attitude score mean over time for most of the categories (except for category 'Other').

Table 3.5 shows some estimates for the mean attitude score, for each wave and for waves '1 to 9', fully considering the actual BHPS *sampling design* and longitudinal weights. The estimator $\bar{y}_t$ shall now be obtained by fitting the regression model described in Example 2.1, Chapter 2, considering a model only with the constant term. We shall thus calculate $\bar{y}_t$ by adopting (2.27) given in Chapter 2, Section 2.4. Cross-sectional results may be achieved by holding t fixed as $t=1$, $t=3$, $t=5$, $t=7$, and $t=9$ for waves one, three, five, seven, and nine respectively.

The following table thus give: (*i*) $\bar{y}_t$'s estimated robust s.e., represented by $\sqrt{\text{var}_r[\bar{y}_t]}$ (see Chapter 2, Sub-section 2.5.1, Example 2.3); (*ii*) $\bar{y}_t$'s estimated linearization s.e., represented by $\sqrt{\text{var}_L[\bar{y}_t]}$ (see Chapter 2, Sub-section 2.5.1); (*iii*) $\bar{y}_t$'s estimated jackknife s.e., represented by $\sqrt{\text{var}_J[\bar{y}_t]}$ (see Chapter 2, Sub-section 2.5.2); and (*iv*) $\hat{meff}^J$ and $\hat{meff}^L$, which shall be calculated in this situation respectively as

$$\hat{meff}^J = \frac{\text{var}_J[\bar{y}_t]}{\text{var}_r[\bar{y}_t]}, \quad \text{and} \quad \hat{meff}^L = \frac{\text{var}_L[\bar{y}_t]}{\text{var}_r[\bar{y}_t]}, \tag{3.4}$$

where the variance estimator in the denominator of both $\hat{meff}^J$ and $\hat{meff}^L$ is now a robust estimator, and $\bar{y}_t$ is a weighted estimator.

	One wave only					Waves
Wave	1	3	5	7	9	1 to 9
$\bar{y}_t$	20.24	20.15	19.93	19.84	19.76	19.98
$\sqrt{\text{var}_r[\bar{y}_t]}$	0.12	0.11	0.11	0.11	0.11	0.09
$\sqrt{\text{var}_J[\bar{y}_t]}$	0.12	0.10	0.10	0.11	0.10	0.09
$\sqrt{\text{var}_L[\bar{y}_t]}$	0.12	0.10	0.10	0.11	0.10	0.09
$m\hat{e}ff^J \doteq m\hat{e}ff^L$	1.00	0.83	0.82	1.09	0.93	1.00

Table 3.5 – Misspecification effects for mean attitude score considering the actual BHPS sampling design.

We notice a possible negative trend over time in the attitude score mean, which could be caused, for example, by the aging process of the survey respondents included in the sample subset considered in our investigation. Note that linearization s.e. estimates are approximately equal to jackknife ones. Further discussion shall be presented later in the current chapter regarding a comparison between these two variance estimation methods.

Moreover, misspecification effects are close or equal to one, for both when each wave is analysed separately and waves '1 to 9'. However, *meffs* are not considerably lower for waves '1 to 9' compared to each single wave, as one could expect. This issue shall be further investigated below. Note that *meffs* presented in Table 3.5 are in general smaller than those included in Tables 3.2 to 3.4. Recall that for Table 3.5, misspecification effects were given by (3.4) and calculated considering a robust variance estimator in the denominator, while that was not the case in Tables 3.2 to 3.4.

We shall provide in Table 3.6 estimates for the mean estimator for the attitude score for wave one, and then for waves '1 to 3', '1 to 5', '1 to 7', and '1 to 9'. We shall consider now the 'new clustering' BHPS *sampling design*, treat the weights as constant and ignore stratification in order to create clearer comparisons.

Wave	1	1 to 3	1 to 5	1 to 7	1 to 9
$\bar{y}_t$	20.07	20.03	19.95	19.88	19.83
$\sqrt{\text{var}_r[\bar{y}_t]}$	0.11	0.10	0.09	0.09	0.09
$\sqrt{\text{var}_L[\bar{y}_t]}$	0.12	0.12	0.12	0.12	0.12
$m\hat{e}ff^L$	1.51	1.50	1.68	1.81	1.84

Table 3.6 – Misspecification effects for mean attitude score considering the 'new clustering' BHPS sampling design.

Note that the *meffs* included in table above capture only the effect of clustering and are given by (3.4). Table 3.6 suggests a trend for the *meff* to augment with the number of waves included in the analysis. These misspecification effects are unquestionably subject to sampling error and there seems to be some authentic variation in meffs for cross-sectional estimates at separate waves but this variation does not appear to be enough to explain this tendency.

We shall expand the analysis by presenting in the next table weighted and unweighted estimates of difference of means, δ, for the attitude score among the categories of the variable age group. Let $\hat{\delta}$ represent an unweighted estimator of the contrast. We could obtain $\hat{\delta}$ by adopting (2.15) given in Chapter 2, Section 2.2. We would then need to fit a regression model with indicator variables for age group as covariates (see model in Example 2.1, Chapter 2). The resulting regression model has an intercept term, which is a domain mean, and three covariates representing the contrasts between women in age group '16-21' and women in other age groups. Note that Table 3.7 shall not include results for the intercept term.

Moreover let the variance of $\hat{\delta}$ be represented by $\mathrm{VAR}(\hat{\delta})$, which could be estimated by $\mathrm{var}(\hat{\delta})$ when not accounting for the complex survey design. Hence the estimator $\mathrm{var}(\hat{\delta})$ could be obtained by adopting (2.21) given in Chapter 2, Sub-section 2.3.1.

A weighted estimator, $\hat{\delta}_w$, for the measure of contrast may be calculated by adopting (2.27) given in Chapter 2, Section 2.4. We shall estimate $\mathrm{VAR}(\hat{\delta}_w)$ adopting the robust variance estimator described in Chapter 2, Sub-section 2.5.1, Example 2.3, for the contrast, denoted here by $\mathrm{var}_r[\hat{\delta}_w]$ to estimate the variance of $\hat{\delta}_w$.

Furthermore, we shall also estimate $\mathrm{VAR}(\hat{\delta}_w)$, considering the actual BHPS complex *sampling design*, adopting the linearization variance method (see Chapter 2, Sub-section 2.5.1) for the contrast, denoted here by $\mathrm{var}_L[\hat{\delta}_w]$ to estimate the variance of $\hat{\delta}_w$. Note that $\hat{\delta}_w$'s jackknife s.e. estimates are fond to be very similar to $\mathrm{var}_L[\hat{\delta}_w]$, and thus shall not be presented below.

Note that for producing results presented in Table 3.5 for each wave separately, i.e. in a cross-sectional analyses context, we shall consider that t is held fixed as $t=1$, $t=3$,

$t = 5$, and $t = 7$, for waves one, three, five and seven respectively. We shall not include results for wave 9 in Table 3.7, as there are no women in the age group '16-21' at wave 9.

For evaluating the influence of the BHPS sampling scheme on the results included in Table 3.7 we adopt $m\hat{e}ff^{L}$, which shall be calculated in the present situation as

$$m\hat{e}ff^{L} = \frac{\text{var}_L[\delta_w]}{\text{var}_r[\delta_w]}. \tag{3.5}$$

See also Chapter 2, Section 2.6. Note in (3.5) that $\hat{\delta}_w$ is a weighted estimator.

Results in Table 3.7 below are presented for each wave separately, where we may observe only complex sampling effects, and then for waves '1 to 9', where we may observe both longitudinal and complex sampling effects.

Wave		Contrast		s.e.			$m\hat{e}ff^{L}$
		δ	δ_w	$\sqrt{\text{var}(\delta)}$	$\sqrt{\text{var}_r[\delta_w]}$	$\sqrt{\text{var}_L[\delta_w]}$	
1	"22-27" *minus* "16-21"	-1.53	-1.57	0.32	0.33	0.36	1.07
	"28-33" *minus* "16-21"	-1.99	-1.94	0.32	0.33	0.34	1.18
	"34+" *minus* "16-21"	-2.10	-2.15	0.32	0.34	0.34	1.04
3	"22-27" *minus* "16-21"	-1.22	-1.12	0.35	0.36	0.35	0.98
	"28-33" *minus* "16-21"	-2.01	-1.84	0.34	0.35	0.35	0.97
	"34+" *minus* "16-21"	-2.00	-1.94	0.33	0.34	0.35	0.98
5	"22-27" *minus* "16-21"	-1.36	-1.17	0.51	0.54	0.52	1.06
	"28-33" *minus* "16-21"	-1.98	-1.86	0.49	0.52	0.52	0.92
	"34+" *minus* "16-21"	-2.25	-2.19	0.48	0.52	0.51	0.99
7	"22-27" *minus* "16-21"	-2.32	-1.77	2.14	2.10	2.09	0.96
	"28-33" *minus* "16-21"	-3.33	-2.76	2.13	2.09	2.08	0.99
	"34+" *minus* "16-21"	-3.62	-3.03	2.13	2.09	2.08	0.99
1 to 9	"22-27" *minus* "16-21"	-1.27	-0.86	0.20	0.18	0.21	1.25
	"28-33" *minus* "16-21"	-1.91	-1.25	0.19	0.21	0.22	1.16
	"34+ *minus* "16-21"	-2.20	-1.52	0.19	0.21	0.23	1.19

Table 3.7 – Mean difference in the attitude score by age group.

By looking across each row of Table 3.7 we can notice that weighted and unweighted means are very similar for wave one, although they differ more for waves three, five and seven, and waves '1 to 9'. Misspecification effects are sufficiently close to one in all situations but tend to be greater for waves '1 to 9' than for each wave separately. Misspecification effects results for waves '1 to 9' could be providing some evidence that variance effect of clustering may be stronger for longitudinal studies. We do not try to give any attitudinal interpretation for estimates given in the table above, as our interest in that case is mainly to evaluate the misspecification effects.

We shall additionally consider estimating δ for the attitude score among the categories of the economic activity. Results shall be produced now considering the 'new clustering', treating the weights as constant and ignoring stratification. In the table below $m\hat{e}ff^{L}$ is also given by (3.5).

Wave		s.e.			$\hat{meff}^{L}$
		δ	$\sqrt{\text{var}_r[\delta_w]}$	$\sqrt{\text{var}_L[\delta_w]}$	
1	Intercept	21.49	0.15	0.16	1.13
	"PT emp." *minus* "FT emp."	-2.31	0.26	0.25	0.93
	"O. inactive" *minus* "FT emp."	-2.48	0.49	0.38	0.60
	"FT stud." *minus* "FT emp."	0.72	0.41	0.43	1.10
	"Fam. Care" *minus* "FT emp."	-3.85	0.26	0.22	0.72
1 to 3	Intercept	21.08	0.12	0.12	1.01
	"PT emp." *minus* "FT emp."	-1.68	0.18	0.17	0.91
	"O. inactive" *minus* "FT emp."	-1.29	0.30	0.29	0.96
	"FT stud." *minus* "FT emp."	0.43	0.32	0.37	1.32
	"Fam. Care" *minus* "FT emp."	-2.76	0.19	0.13	0.49
1 to 5	Intercept	20.81	0.11	0.11	1.09
	"PT emp." *minus* "FT emp."	-1.20	0.14	0.14	0.93
	"O. inactive" *minus* "FT emp."	-0.83	0.22	0.18	0.68
	"FT stud." *minus* "FT emp."	0.30	0.25	0.27	1.14
	"Fam. Care" *minus* "FT emp."	-2.46	0.16	0.12	0.58
1 to 7	Intercept	20.70	0.10	0.11	1.38
	"PT emp." *minus* "FT emp."	-1.14	0.12	0.12	1.00
	"O. inactive" *minus* "FT emp."	-0.84	0.18	0.16	0.76
	"FT stud." *minus* "FT emp."	0.27	0.21	0.25	1.48
	"Fam. Care" *minus* "FT emp."	-2.34	0.14	0.11	0.66
1 to 9	Intercept	20.58	0.09	0.11	1.50
	"PT emp." *minus* "FT emp."	-1.03	0.10	0.10	0.89
	"O. inactive" *minus* "FT emp."	-0.80	0.17	0.15	0.81
	"FT stud." *minus* "FT emp."	0.41	0.20	0.24	1.44
	"Fam. Care" *minus* "FT emp."	-2.18	0.12	0.10	0.59

Table 3.8 – Mean difference in the attitude score by economic activity.

As before, there is some evidence in Table 3.8 of tendency for the *meff* to augment, from 1.13 with one wave to 1.50 with five waves. The *meffs* for the contrasts in the table above differ in size, some greater than and some less than one. This misspecification effects may be considered as a mixture of the classic variance inflating effect of clustering in surveys combined with the known variance reducing effect of blocking in an experiment. The fundamental characteristic of these results of interest here is that there is once more no tendency for the meffs to converge to one as the number of waves expands. If there is a trend, it is in the opposite direction.

Therefore, we could say that in general the influence of the BHPS *sampling design* on those contrast measures inference procedure should not be ignored. For the contrast of special scientific relevance, that between women who are 'full-time employed' and those who are 'at home caring for a family', the *meff* is invariably considerably inferior to one.

We may try to compare results included in Table 3.6 with those in Table 3.8. Such comparison validates Kish and Frankel (1974) statement that *meffs* for regression coefficients are likely not to be greater than those for the means of the response variable (see also Kish, 1980).

3.3.2 Cross-sectional models

This sub-section presents cross-sectional model fitting results with both non-weighted and weighted parameter estimates. Both jackknife and linearization variance estimation methods shall be adopted in this sub-section when taking the BHPS sampling design into account. We shall start by describing explicitly in Example 3.1 below a linear regression cross-sectional model (LM).

Example 3.1: linear cross-sectional model

In studies where $T=1$, i.e. in cross-sectional studies, we may consider that the response variable Y_i of individual i, is represented by the model

$$Y_{i1} = \underline{x}_{i1}\underline{\beta}^c + \varepsilon_{i1}, \text{ with } i = 1,\ldots,N, \qquad (3.6)$$

where the superscript c denotes 'cross-sectional', $\underline{x}_{i1}$ is the $1\times q$ vector with the q fixed covariates, which are also survey variables, $\underline{\beta}^c$ is the $q\times 1$ vector of the unknown fixed coefficients for the x variables, and the scalar ε_{i1} is the error term. In (3.6), if the explanatory variables are categorical, $\underline{\beta}^c$ may be interpreted as the average difference in Y when comparing each of the categories of the covariates with their reference levels.

Let $\sigma_1^2 = \mathrm{VAR}(Y_{i1})$, where VAR(.) denote population variance. Moreover we assume that $E(\underline{\varepsilon}_{i1}) = 0$ for all i. In (3.6), we assume that

$$Y_{i1} \sim N(\underline{x}_{i1}\underline{\beta}^c, \sigma_1^2). \qquad ■$$

Model fitting is undertaken in this sub-section for the model described above. We first include all the candidate covariates in the model (see Sub-section 3.2.2, variables from *iii* to *vi*). We thus exclude, one by one, variables which do not have their coefficients significantly from zero at the 5% level, starting with the least significant variables, i.e. we perform the well known backward elimination procedure. Only *final* model fitting results are presented in this sub-section, i.e. Tables 3.9 and 3.10 only include covariates found to be significant at 5% level.

Interactions shall not be considered in the model fitting results presented in this chapter, as our main interests are to identify and to illustrate variance effects of clustering for longitudinal studies, which do not require us to have very elaborate models.

The cross-sectional model parameters may be estimated by a cross-section pseudo maximum likelihood estimator, $\hat{\underline{\beta}}^c_{PML}$, calculated by (2.27) given in Chapter 2, Section 2.4, with t being held fixed as $t=1$ for wave one, for example. An unweighted $\hat{\underline{\beta}}^c_{PML}$ may be achieved when considering that w_i are constant (see Chapter 2, Section 2.2), and that would be equivalent to an ordinary least square estimator, which we shall denote here by $\hat{\underline{\beta}}^c_{OLS}$.

Both jackknife and linearization estimators shall be adopted here for the variance of $\hat{\underline{\beta}}^c_{PML}$ and shall be denoted by $\text{var}_J[\hat{\underline{\beta}}^c_{PML}]$ and $\text{var}_L[\hat{\underline{\beta}}^c_{PML}]$, respectively. See Chapter 2 for further information on these methods. Results for robust variance estimator shall also be produced and denoted here by $\text{var}_r[\hat{\underline{\beta}}^c_{PML}]$.

We shall also calculate a variance estimator for $\hat{\underline{\beta}}^c_{OLS}$ given by (Kmenta, 1971)

$$\text{var}_{\text{n}}\left[\hat{\underline{\beta}}^c_{OLS}\right] = \hat{\sigma}_1^2\left[\sum_{i=1}^{n}\left(\underline{x}_{i1}'\underline{x}_{i1}\right)\right]^{-1}, \tag{3.7}$$

where the subscript n denotes 'naïve', and

$$\hat{\sigma}_1^2 = (1/n-q)\cdot\sum_{i=1}^{n}\left(\hat{\varepsilon}_{i1}\right)^2$$

is the usual mean squared residual, and $\hat{\varepsilon}_i^c$ is a scalar, as defined in Example 3.1. We may consider (3.7) to be a cross-sectional version of (2.22) in Chapter 2, Subsection 2.3.1.

For evaluating the influence of the BHPS sampling scheme on the results included in Tables 3.9 and 3.10 we adopt $m\hat{e}ff^L$, which shall be calculated in the present situation as

$$m\hat{e}ff^L = \frac{\text{var}_L[\hat{\underline{\beta}}^c_{PML}]}{\text{var}_r[\hat{\underline{\beta}}^c_{PML}]}.$$

See also Chapter 2, Section 2.6. Note that Table 3.9 shall additionally include $m\hat{e}ff^L *$, which represents a misspecification effect calculated considering the 'new clustering' BHPS design, with constant sampling weights and ignoring stratification.

Table 3.9 presents the *final* LM for the women gender role attitude score as response variable, considering only wave one data.

Covariate		$\hat{\underline{\beta}}^c_{OLS}$	$\hat{\underline{\beta}}^c_{PML}$	s.e. $\sqrt{\text{var}_n[\hat{\beta}^c]}$	s.e. $\sqrt{\text{var}_r[\hat{\underline{\beta}}^c_{PML}]}$	s.e. $\sqrt{\text{var}_J[\hat{\underline{\beta}}^c_{PML}]}$	s.e. $\sqrt{\text{var}_L[\hat{\underline{\beta}}^c_{PML}]}$	$m\hat{e}ff^L$	$m\hat{e}ff^L *$
Intercept		23.15	23.23	0.44	0.44	0.46	0.45	1.05	0.95
Age group	16-21	-	-	-	-	-	-	-	-
	22-27	-0.99	-1.03	0.34	0.35	0.38	0.38	1.18	1.22
	28-33	-1.14	-1.11	0.34	0.36	0.37	0.36	1.00	1.38
	More than 34	-1.27	-1.36	0.35	0.37	0.37	0.36	0.95	0.94
Econ. activity	FT emp.	-	-	-	-	-	-	-	-
	PT emp.	-2.03	-1.88	0.27	0.29	0.26	0.26	0.80	0.97
	O. inactive	-2.57	-2.47	0.45	0.54	0.51	0.50	0.86	0.60
	FT stud.	-0.01	-0.10	0.47	0.49	0.50	0.50	1.04	0.93
	Fam. Care	-3.54	-3.47	0.27	0.29	0.30	0.29	1.00	0.77
Educ. level	Degree	-	-	-	-	-	-	-	-
	QF	-0.82	-0.80	0.42	0.44	0.44	0.44	1.00	0.77
	A-level	-0.63	-0.67	0.40	0.39	0.39	0.39	1.00	0.98
	O-level	-0.82	-0.93	0.36	0.36	0.37	0.36	1.00	0.62
	Other	-1.16	-1.25	0.38	0.38	0.39	0.39	1.05	0.83

Table 3.9 – LM for gender role attitude score – wave one.

Variables age group, economic activity and educational level are significant at the 5% level, considering both 'naïve', jackknife and linearization results. We notice one possible difference in the significance tests results (see category 'QF' of the educational level covariate), at the 5% level, when comparing results that considers data is selected by srs with those which account for complex survey design structure.

Misspecification effects are sufficiently close to one for most of the categories of all the covariates included in the model, when considering the actual BHPS sampling design. There are some differences between weighted and non-weighted parameter estimates, which are not very strong in this case. Therefore, we could say that the influence of the BHPS sampling scheme on the estimation of the cross-sectional model parameters is not very strong, when considering wave one data.

The $m\hat{e}ff^{L}*$ are also either very close to one or even smaller than one for some coefficients. Recall that these misspecification effects were calculated considering constant the sampling weights, which may have contributed to a reduction of $\text{var}_L[\underline{\hat{\beta}}^{c}_{PML}]$. These *meff* values shall be used in further comparisons that shall be performed later in this chapter.

Results in the table above suggest that the results for jackknife and linearization variance estimators are very similar for $\underline{\hat{\beta}}^{c}_{PML}$, when considering attitude score, from BHPS data, as response variable. There are a number of empirical and analytical investigations in the literature for comparing the performances of jackknife and linearization procedures. We can find in Rust (1985, Section 7) a comprehensive literature review on the comparison among variance estimation methods. That article gives also an extensive list of references on this issue.

Our result agrees with findings of Rust (1985) and most of the references included there, which say that both variance estimation methods present methodological advantages and similar accuracy. Shah, Barnwell, & Bieler (1997) also say that those two variance estimation methods usually yield very similar estimates. Additionally, our results could somehow be an indication that adopting the sampling scheme described in Example 2.2, in Sub-section 2.5.1 is a reasonable approximation for the BHPS sampling design when calculating linearization variance estimators in a cross-sectional context.

In terms of model interpretation, considering weighted parameters results, for the variable economic activity we can say, that women in ‘family care’ have on average an attitude score about 3.5 points less than women in ‘full time employment’ in 1991 (wave one data). Controlling for all the remaining covariates, it indicates that women in full time employment tend to be more egalitarian than women in family care. For the covariate educational level we can say that women with 'other' educational level, which is the lowest educational level, have an attitude score about 1.25 point less on average than those with first or higher degree. Controlling for all the remaining covariates, it indicates that women with higher educational levels tend to be more egalitarian in 1991.

Table 3.10 presents the *final* LM fitting results for the women gender role attitude score as response variable, considering only wave nine data.

Covariate†		$\hat{\underline{\beta}}^c_{OLS}$	$\hat{\underline{\beta}}^c_{PML}$	s.e. $\sqrt{\text{var}_n[\hat{\beta}^c]}$	$\sqrt{\text{var}_r[\hat{\underline{\beta}}^c_{PML}]}$	$\sqrt{\text{var}_J[\hat{\underline{\beta}}^c_{PML}]}$	$\sqrt{\text{var}_L[\hat{\underline{\beta}}^c_{PML}]}$	$m\hat{e}ff^L$
Intercept		21.65	21.66	0.34	0.35	0.38	0.37	1.12
Age group	16-21 (*)	-	-	-	-	-	-	-
	22-27 (**)	-	-	-	-	-	-	-
	28-33	-0.08	-0.07	0.34	0.35	0.39	0.38	1.18
	More than 34	-0.73	-0.70	0.31	0.31	0.31	0.31	1.00
Econ. activity	FT emp.	-	-	-	-	-	-	-
	PT emp.	-1.70	-1.76	0.23	0.23	0.22	0.21	0.83
	O. inactive	-1.71	-1.70	0.37	0.47	0.50	0.49	1.09
	FT stud.	-0.39	0.12	0.87	0.88	0.95	0.90	1.05
	Fam. Care	-3.39	-3.25	0.27	0.29	0.28	0.28	0.93
Educ. level	Degree	-	-	-	-	-	-	-
	QF	-0.40	-0.31	0.30	0.32	0.31	0.31	0.94
	A-level	-0.55	-0.51	0.35	0.37	0.36	0.36	0.95
	O-level	-0.18	-0.32	0.31	0.33	0.31	0.31	0.88
	Other	-0.55	-0.53	0.33	0.35	0.34	0.33	0.89

(*) There are no women aged 16-21 in the sample at wave 9.
(**) Reference category.
† Covariate parenthood status was originally significant for 1999 data. Nevertheless, for consistency with Table 3.9, that variable was not included in the final model presented in this table.

Table 3.10 – LM for gender role attitude score – wave nine.

The parameter estimates for wave nine data are similar to the wave one results. Note, however, that the base line category for the covariate age group is now '22-27', instead of '16-21'. In addition, we notice one difference in the significance tests results (see category 'A-level' of the educational level covariate), at the 5% level, when comparing 'naïve' and jackknife p-values.

Misspecification effects are sufficiently close to one for most of the categories of all the covariates included in the model. The largest *meff* result is 1.18 for the category '28-33' of covariate age group. Therefore, we could say that the influence of the BHPS sampling scheme on the estimation of the cross-sectional model parameters is not very strong, also when considering wave nine data. We shall not try to give any attitudinal interpretation for estimates given in Table 3.10.

3.3.3 Longitudinal model fit results

In this sub-section, we shall concentrate on identifying and illustrating variance effects of clustering for longitudinal studies. We shall also study patterns of changes in the gender role attitudes of women and establish the direction and magnitude of its relationship with a set of explanatory variables. We shall thus provide a description of patterns of change and a concise analysis of causal relations.

We shall adopt in this sub-section the model described in Chapter 2, Example 2.1. Note that in a more elaborate analysis we could have adopted a measurement error model for the attitude variable of interest, as for example considered in Fan and Marini (2000), with each of responses to the six statements considered as ordinal variables. However, we follow a more straightforward approach, treating the aggregate attitude score Y_{it} and the associated vector of coefficients $\underline{\beta}$ as scientifically interesting, with the measurement error included in the model error term ε_{it}.

The adoption of the uniform correlation model, which assumes an exchangeable structure for the *working correlation matrix* V, is supported by the fact that – a robust variance may be estimated for $\hat{\underline{\beta}}$ when n is large relative to T. According to Diggle *et al.* (2002), when the last statement is true, valid inferences may be obtained even when the correlation is misspecified. That is the case here as $n = 1340$ and $T = 5$. See also Chapter 2, Sub-sections 2.3.2 and 2.5.1 for information on robust variance estimators.

We again exclude covariates with coefficients which do not attain a 5% level of significance. As in Sub-section 3.3.2, we adopt here a backward elimination procedure for model selection. The candidate covariates in the model are variables from *ii* to *vi* listed in Sub-section 3.2.2. Only *final* longitudinal model fitting results are presented in this sub-section.

The model formulation that we work with in this sub-section assumes that a change in the dependent variable is a function of changes in the explanatory variables, even though the models do not explicitly involve change variables[19]. We compare score for women who belong to different categories of each explanatory variable, controlling for the remaining covariates.

Models parameters presented in this sub-section have interpretation for the population rather than for any subject, i.e. they describe how the population-averaged response depends on the explanatory variables. These models focus more on the difference in the population-averaged response among groups (treatment versus control) than on the change in an individual's response over time (see Zeger, Liang & Albert, 1988).

We thus present some longitudinal model fitting results including both $\underline{\hat{\beta}}(\mathrm{V})$, given by (2.15), in Chapter 2, Section 2.2, and $\hat{\beta}_{PML}(\mathrm{V})$, given by (2.27), in Section 2.4. We assume here that (2.29), included in Chapter 2, Section 2.4, holds even when we substitute w_i by w_{iT}^{*}. Some of the results presented in this sub-section shall be weighted, with longitudinal respondent weights w_{iT}^{*} (see Chapter 1, Section 1.3) being adopted. We thus substitute w_i by w_{it}^{*} for $\underline{\hat{\beta}}_{PML}(\mathrm{V})$ estimated here. Furthermore, we shall also produce some results considering the 'new clustering' described earlier in Sub-section 3.2.2, treating the weights as constant and ignoring stratification in order to capture only the effects of clustering.

We shall perform s.e. estimation for $\underline{\hat{\beta}}(\mathrm{V})$ considering a 'naïve' estimator, denoted by $\sqrt{\mathrm{var}_{\mathrm{n}}[\underline{\hat{\beta}}(\mathrm{V})]}$ (see Chapter 2, Sub-section 2.3.1). And for $\underline{\hat{\beta}}_{PML}(\mathrm{V})$, we shall adopt: (*i*) a robust estimator, denoted by $\sqrt{\mathrm{var}_r[\underline{\hat{\beta}}_{PML}(\mathrm{V})]}$; (*ii*) a jackknife estimator, denoted by $\sqrt{\mathrm{var}_J[\underline{\hat{\beta}}_{PML}(\mathrm{V})]}$; and (*iii*) a linearization estimator, denoted by $\sqrt{\mathrm{var}_L[\underline{\hat{\beta}}(\mathrm{V})_{PML}]}$. For further information of these methods see Chapter 2, Sub-sections 2.5.1, 2.5.2 and 2.5.3, respectively.

When estimating $\mathrm{VAR}[\underline{\hat{\beta}}(\mathrm{V})_{PML}]$ via linearization we assume that the *sampling scheme* described in Example 2.2, in Chapter 2, is a 'good' approximation for the BHPS

[19] According to Menard (1991), the use of any change measure is not a simple issue, because the reliability of the changes is frequently not the same as the reliability of the original variables.

sampling design, described in Sub-section 3.2.1. Therefore, method described in Chapter 2, Sub-section 2.5.1 is adopted for calculating $\text{var}_L[\underline{\hat{\beta}}(\text{V})_{PML}]$.

For evaluating the influence of the BHPS *sampling design* on the estimated longitudinal model parameters estimated in this sub-section, we adopt $m\hat{e}ff^{L}$ defined by (2.43) in Chapter 2, Section 2.6.

We include in Table 3.11 longitudinal model fitting results considering data from waves one, three, five, seven and nine.

				s.e.				
Covariate		$\underline{\hat{\beta}}(\text{V})$	$\underline{\hat{\beta}}_{PML}(\text{V})$	$\sqrt{\text{var}_{\text{n}}[\underline{\hat{\beta}}(\text{V})]}$	$\sqrt{\text{var}_{r}[\underline{\hat{\beta}}_{PML}(\text{V})]}$	$\sqrt{\text{var}_{J}[\underline{\hat{\beta}}_{PML}(\text{V})]}$	$\sqrt{\text{var}_{L}[\underline{\hat{\beta}}(\text{V})_{PML}]}$	$m\hat{e}ff^{L}$
Intercept		22.18	22.30	0.25	0.30	0.32	0.30	1.00
T		-0.04	-0.04	0.01	0.02	0.02	0.03	1.00
Age Group	16-21	0.0	0.0	-	-	-	-	-
	22-27	-0.70	-0.64	0.16	0.20	0.22	0.21	1.10
	28-33	-0.87	-0.80	0.19	0.22	0.24	0.24	1.19
	34+	-1.00	-1.00	0.22	0.25	0.28	0.28	1.25
Economic Activity	FT emp.	0.0	0.0	-	-	-	-	-
	PT emp.	-0.91	-0.86	0.10	0.11	0.12	0.11	1.00
	O. inactive	-0.73	-0.75	0.15	0.17	0.38	0.17	1.00
	FT stud.	0.17	0.20	0.20	0.21	0.22	0.24	1.31
	Fam. Care	-2.05	-2.01	0.11	0.14	0.16	0.14	1.00
Qualification	Degree	0.0	0.0	-	-	-	-	-
	QF	-0.51	-0.52	0.21	0.24	0.28	0.25	1.09
	A-level	-0.60	-0.65	0.21	0.26	0.29	0.26	1.08
	O-level	-0.43	-0.49	0.22	0.25	0.27	0.25	1.00
	Other	-1.17	-1.22	0.23	0.26	0.28	0.26	1.00

Table 3.11 – Five waves longitudinal model considering the actual BHPS sampling design.

We can see some differences when comparing the previous table's estimated parameter with those in Tables 3.7 and 3.9, specially regarding to coefficients of the explanatory variable 'economic activity'. It may be indicating that the longitudinal model is somehow picking up within-women changes in time, as well as cross-sectional differences between women.

We notice additionally that the results for $\sqrt{\text{var}_L[\underline{\hat{\beta}}(\text{V})_{PML}]}$ are again very similar to $\sqrt{\text{var}_J[\underline{\hat{\beta}}_{PML}(\text{V})]}$. The results of greatest interest in Table 3.11 are the misspecification effects, which are in general reasonably close to one. Additionally, we may also notice some differences between weighted and non-weighted parameter estimates, although they are not very strong in this case.

For the model fitting results presented above, the variables time, age group, economic activity and educational level may be considered significant at the 5% level. Overall, for the variable economic activity we can see, for example, that an average woman in 'family care' has an attitude score about two points less than an average woman in 'full time employment'. Controlling by the remaining covariates, it indicates that women in full time employment tend to be more egalitarian than women in family care. For the variable educational level we can see that an average woman with 'other' educational level has an attitude score about 1.2 less than an average woman with 'first or higher degree'. Controlling for the remaining covariates, it indicates that women with higher educational levels tend to be more egalitarian.

As we assume here an exchangeable *working correlation*, the estimated within individual correlation (see Chapter 2), ρ, is 0.59, which could also be interpreted as a fairly substantial between-woman component in the attitude scores unexplained by the chosen covariates.

She shall provide in Table 3.12 longitudinal model fitting results waves '1 to 9', considering now the 'new clustering' BHPS *sampling design*, treating the weights as constant and ignoring stratification in order to allow for clearer comparisons.

Covariate		$\hat{\underline{\beta}}(V)$	s.e. $\sqrt{var_r[\hat{\underline{\beta}}(V)]}$	$\sqrt{var_L[\hat{\underline{\beta}}(V)]}$	$m\hat{e}ff^L$ *
Intercept		22.20	0.29	0.30	1.06
t		-0.04	0.01	0.01	0.96
Age Group	16-21	0.0	-	-	-
	22-27	-0.70	0.19	0.25	1.65
	28-33	-0.87	0.21	0.27	1.59
	34+	-1.00	0.24	0.27	1.34
Economic Activity	FT emp.	0.0	-	-	-
	PT emp.	-0.91	0.10	0.10	0.91
	O. inactive	-0.73	0.17	0.15	0.81
	FT stud.	0.17	0.21	0.24	1.31
	Fam. Care	-2.05	0.12	0.10	0.67
Qualification	Degree	0.0	-	-	-
	QF	-0.51	0.23	0.21	0.84
	A-level	-0.60	0.24	0.24	1.01
	O-level	-0.43	0.23	0.20	0.73
	Other	-1.17	0.25	0.22	0.82

Table 3.12 – Five waves longitudinal model considering the 'new clustering' BHPS sampling design.

The misspecification effects presented in Table 3.12 for the coefficients of the intercept, and 'age group' are larger than the ones presented in Table 3.11, which considered the actual BHPS sampling design. For the covariates 'time', 'economic activity' and 'qualification' *meffs* are in general very similar when comparing Table 3.12 with 3.11. However, when comparing the misspecification effects presented in the table above with $m\hat{e}ff^L$ * included in Table 3.9 we may notice that *meffs* for waves '1 to 9' are in general greater than for wave 1. These results agree with Table 3.6's results and suggest that variance effects of clustering are stronger for longitudinal studies than for cross-sectional ones.

For further investigate the variance effects of clustering in longitudinal models, we also fit longitudinal models considering (*i*) waves one and three (Table 3.13), (*ii*) waves one, three and five (Table 3.14), and (*iii*) waves one, three, five and seven (3.15). Our aim is to verify how the *meffs* results behave when additional waves are included in the data set. Moreover, Tables 3.13 to 3.15 shall also include $\hat{meff}^{L}*$, i.e. a misspecification effect calculated considering the 'new clustering' BHPS design, with constant sampling weights and ignoring stratification. Thus, the following table presents information for the *final* longitudinal model, considering data from waves one and three.

Covariate		$\hat{\underline{\beta}}(V)$	$\hat{\underline{\beta}}_{PML}(V)$	s.e. $\sqrt{var_n[\hat{\underline{\beta}}(V)]}$	s.e. $\sqrt{var_r[\hat{\underline{\beta}}_{PML}(V)]}$	s.e. $\sqrt{var_L[\hat{\underline{\beta}}(V)_{PML}]}$	$\hat{meff}^{L}$	$\hat{meff}^{L}*$
Intercept		22.77	22.95	0.35	0.37	0.37	1.00	0.87
T		0.03	0.02	0.04	0.05	0.05	1.00	0.86
Age Group	16-21	-	-	-	-	-	-	-
	22-27	-0.97	-0.89	0.24	0.29	0.31	1.14	1.37
	28-33	-1.26	-1.21	0.26	0.30	0.30	1.00	1.40
	34+	-1.30	-1.28	0.27	0.31	0.32	1.07	1.10
Economic Activity	FT emp.	-	-	-	-	-	-	-
	PT emp.	-1.46	-1.37	0.18	0.19	0.18	0.90	0.95
	O. inactive	-1.25	-1.33	0.26	0.30	0.30	1.00	0.96
	FT stud.	-0.03	0.04	0.31	0.34	0.36	1.12	1.32
	Fam. Care	-2.45	-2.47	0.18	0.20	0.22	1.21	0.59
Qualification	Degree	-	-	-	-	-	-	-
	QF	-0.71	-0.80	0.33	0.34	0.34	1.00	0.64
	A-level	-0.71	-0.88	0.31	0.31	0.31	1.00	0.87
	O-level	-0.79	-0.94	0.30	0.30	0.30	1.00	0.62
	Other	-1.37	-1.49	0.30	0.32	0.33	1.06	0.83

Table 3.13 – Waves one and three longitudinal model.

We notice some differences in the model estimated coefficients when comparing previous table with Table 3.11, which could be due to the exclusion of data from waves five, seven and nine. For the covariate 'economic activity', the misspecification effects when considering the actual BHPS sampling scheme, are larger than the ones from the cross-sectional model for wave one presented in Table 3.9. For the remaining covariates, $\hat{meff}^{L}$ values are very similar when comparing Table 3.13 with Table 3.11. When considering the 'new clustering' BHPS sampling design, with constant sampling weights and ignoring stratification, *meffs* for the covariate 'age group' are larger in the table above than in Table 3.9. For the other covariates, $\hat{meff}^{L}*$ values are very similar when comparing Table 3.13 with Table 3.9. Furthermore, in the table above we can not see strong differences between weighted and non-weighted parameter estimates.

We do not give any attitudinal interpretation for estimated models parameters in Table 3.13, as our main interest in this situation is to evaluate the misspecification effects, when considering data from two waves.

Fitting results for the *final* longitudinal model, considering data from waves one, three and five, are included in Table 3.14 below.

Covariate		$\hat{\underline{\beta}}(\mathrm{V})$	$\hat{\underline{\beta}}_{PML}(\mathrm{V})$	s.e. $\sqrt{\mathrm{var}_n[\hat{\underline{\beta}}(\mathrm{V})]}$	s.e. $\sqrt{\mathrm{var}_r[\hat{\underline{\beta}}_{PML}(\mathrm{V})]}$	s.e. $\sqrt{\mathrm{var}_L[\hat{\underline{\beta}}(\mathrm{V})_{PML}]}$	$\hat{meff}^{L}$	$\hat{meff}^{L}*$
Intercept		22.64	22.74	0.30	0.35	0.34	0.94	0.87
T		-0.03	-0.03	0.02	0.03	0.03	1.00	0.69
Age Group	16-21	-	-	-	-	-	-	-
	22-27	-0.77	-0.67	0.19	0.23	0.24	1.09	1.44
	28-33	-1.09	-1.03	0.22	0.26	0.26	1.00	1.45
	34+	-1.24	-1.23	0.24	0.28	0.29	1.07	1.12
Economic Activity	FT emp.	-	-	-	-	-	-	-
	PT emp.	-0.99	-0.95	0.14	0.15	0.15	1.00	0.96
	O. inactive	-0.77	-0.81	0.20	0.22	0.21	0.91	0.68
	FT stud.	0.09	0.10	0.25	0.26	0.26	1.00	1.23
	Fam. Care	-2.20	-2.21	0.15	0.17	0.19	1.25	0.70
Qualification	Degree	-	-	-	-	-	-	-
	QF	-0.81	-0.81	0.28	0.30	0.32	1.14	0.75
	A-level	-0.82	-0.86	0.27	0.28	0.29	1.07	0.94
	O-level	-0.89	-0.96	0.26	0.29	0.27	0.87	0.59
	Other	-1.48	-1.50	0.27	0.30	0.29	0.93	0.78

Table 3.14 – Waves one, three and five longitudinal model.

We observe some differences in the model estimated coefficients when comparing the previous table with Tables 3.12 and 3.13, which could be due to the fact that those tables are based on different sets of waves. For most of the coefficients the misspecification effects when considering the actual BHPS sampling scheme, are similar to the values presented in Tables 3.9 and 3.13. When considering the 'new clustering' BHPS sampling design, with constant sampling weights and ignoring stratification, *meffs* for the covariate 'age group' are larger in the table above than in Tables 3.9 and 3.13. For the remaining covariates, $m\hat{e}ff^{L}*$ values are very similar when comparing Table 3.14 with Table 3.13 and Table 3.9. Similarly to the results from previous tables, there are some differences between weighted and non-weighted parameter estimates, which are not very strong. We also do not give any interpretation for estimated model parameters presented in Table 3.14.

Table 3.15 presents information for the *final* longitudinal model, considering data from waves one, three, five and seven.

Covariate		$\hat{\underline{\beta}}(V)$	$\hat{\underline{\beta}}_{PML}(V)$	s.e. $\sqrt{var_n[\hat{\underline{\beta}}(V)]}$	s.e. $\sqrt{var_r[\hat{\underline{\beta}}_{PML}(V)]}$	s.e. $\sqrt{var_L[\hat{\underline{\beta}}(V)_{PML}]}$	$m\hat{e}ff^{L}$	$m\hat{e}ff^{L}*$
Intercept		22.45	22.57	0.27	0.32	0.32	1.00	1.03
T		-0.04	-0.03	0.02	0.02	0.02	1.00	0.59
Age Group	16-21	-	-	-	-	-	-	-
	22-27	-0.74	-0.70	0.17	0.20	0.22	1.21	1.73
	28-33	-0.99	-0.96	0.20	0.23	0.25	1.18	1.68
	34+	-1.19	-1.24	0.23	0.26	0.27	1.08	1.26
Economic Activity	FT emp.	-	-	-	-	-	-	-
	PT emp.	-0.99	-0.91	0.11	0.13	0.13	1.00	1.06
	O. inactive	-0.80	-0.85	0.17	0.18	0.18	1.00	0.77
	FT stud.	0.06	0.10	0.22	0.22	0.25	1.29	1.39
	Fam. Care	-2.17	-2.16	0.13	0.15	0.17	1.28	0.78
Qualification	Degree	-	-	-	-	-	-	-
	QF	-0.70	-0.69	0.24	0.27	0.29	1.15	0.87
	A-level	-0.74	-0.80	0.23	0.27	0.29	1.15	0.94
	O-level	-0.65	-0.74	0.24	0.27	0.27	1.00	0.69
	Other	-1.25	-1.28	0.25	0.29	0.29	1.00	0.80

Table 3.15 – Waves one, three, five and seven longitudinal model.

For most of the model parameters, the misspecification effects presented in Table 3.15 calculated considering the actual BHPS sampling design are larger than the ones included in Table 3.14. If we consider the 'new clustering' BHPS sampling design, with constant sampling weights and ignoring stratification, $\hat{meff}^{L}*$ for most of the coefficients are also larger in the table above than in Table 3.14. We may still highlight the $\hat{meff}^{L}*$ values for the parameters for the categories of the covariate 'age group'', which are again larger than the cross-sectional models ones.

In Table 3.16 we try to summarise the misspecification effects results calculated accounting for the 'new clustering' BHPS sampling design, ignoring stratification and considering constant weights.

Covariate		Waves				
		1	1 to 3	1 to 5	1 to 7	1 to 9
Intercept		0.95	0.87	0.87	1.04	1.07
T		-	0.86	0.69	0.59	0.96
Age Group	22-27	1.22	1.37	1.44	1.73	1.64
	28-33	1.38	1.40	1.46	1.68	1.59
	34+	0.94	1.10	1.13	1.26	1.34
Economic Activity	PT Emp.	0.97	0.95	0.96	1.06	0.91
	O. inactive	0.60	0.96	0.68	0.77	0.81
	FT stud.	0.93	1.32	1.23	1.39	1.32
	Fam. Care	0.77	0.59	0.70	0.78	0.67
Qualification	QF	0.77	0.64	0.75	0.87	0.85
	A-level	0.98	0.87	0.94	0.94	1.01
	O-level	0.62	0.62	0.59	0.69	0.73
	Other	0.83	0.83	0.78	0.80	0.82

Table 3.16 – Misspecification effects for model parameters.

We can observe in Table 3.6 a general tendency in the *meff*s to increase when additional waves of the survey are included in the data set considered in the model fitting. The misspecification effects values for the parameters of the covariate 'age group' are good examples of that trend.

We could say that we have some empirical evidence that the clustering effects are larger for longitudinal models than for cross-sectional ones, when modelling the BHPS data. Indeed we could not find any tendency for the *meffs* to converge to one as the number of waves increases.

3.4 Discussion

This chapter compares empirically design effects for longitudinal estimators and cross-sectional ones. We have presented and examined empirical evidences that variance effects of clustering may be stronger for longitudinal studies than for cross-sectional ones. Design effects for certain kinds of longitudinal analysis may be greater than for corresponding cross-sectional analyses. Hence, it is important to allow for complex designs in longitudinal analysis. In this section, we consider a possible theoretical explanation for that postulation.

Furthermore, we have noticed some differences between weighted and non-weighted parameter estimates, although they are not very strong for most of the model parameters. Nevertheless, the results presented in this chapter have generally indicated that we should be aware of checking the importance of taking the *sampling design* into account when modelling longitudinal complex survey data.

Additionally, differences have been found between parameters' s.e. estimates when comparing non-weighted estimates, $\sqrt{\text{var}_{\text{n}}[\hat{\beta}(\text{V})]}$, and weighted estimates without taking the *sampling design* into account, $\sqrt{\text{var}_{r}[\hat{\underline{\beta}}_{PML}(\text{V})]}$, with weighted estimates considering the *sampling design*, $\sqrt{\text{var}_{J}[\hat{\underline{\beta}}_{PML}(\text{V})]}$ and $\sqrt{\text{var}_{L}[\hat{\underline{\beta}}_{PML}(\text{V})]}$.

We have observed only a few issues regarding significance tests results for cross-sectional model parameters, which are based upon the s.e. estimates. P-values usually seem to be more significant for non-weighted estimates, without taking the *sampling design* into account. That may, in some situations, cause possible false positive tests of covariate effects. This fact is also confirmed by the misspecification effects results that have been produced above.

Moreover, our results have somehow indicated that cluster effects arise when longitudinal regression models are fitted. We provide below some further discussion with the objective of giving some theoretical justification on that statement. We shall suppose that the principal component of the design effects arises from clustering. Consider converting the two-level model, described in Example 2.1, Chapter 2, into a simple three-level model (Goldstein, 1995)

$$Y_{ijt} = E(Y_{ijt}) + \eta_j + u_{ij} + v_{ijt}, \tag{3.8}$$

where Y_{ijt} is the attitude score for individual i, in cluster j, at wave t, $E(Y_{ijt})$ denotes the expected value of Y_{ijt}, an additional random term η_j with variance σ_η^2 represents the area effect, assumed independent of u_{ij} and v_{ijt}. Let σ_u^2 and σ_v^2 denote the variances of of u_{ij} and v_{ijt} respectively.

At first, we shall use the model in (3.8) to analyse the nature of misspecification effects in a cross-sectional analyses context, where t is held fixed as $t = 1$. In this situation, if we suppose that $E(Y_{ijt})$ is simply the mean of Y_{ijt} in (3.8) and that there is a common sample size m per PSU, then (Hansen, Hurwitz, and Madow, 1953; and Kish, 1965, p. 162)

$$meff \doteq 1 + (m-1)\rho_1^*, \tag{3.9a}$$

where (Skinner, 1989b)

$$\rho_1^* = \frac{\sigma_\eta^2}{(\sigma_\eta^2 + \sigma_u^2 + \sigma_v^2)}$$

is the intra-cluster correlation coefficient, a measure of cluster homogeneity, or

$$meff \doteq 1 + (\overline{m}-1)\rho_1^*, \tag{3.9b}$$

when the sample sizes per cluster are unequal, where $\overline{m}$ is the average sample size per cluster. Note that (3.9a) and (3.9b) were derived for the case of a clustered *sampling design* which selects m clusters, and all its elements, via srs, and for samples with a large m.

From (3.9), we have that the larger ρ_1^*, the larger $meff$. Note the distinction in notation between ρ_1^*, given above, and ρ which is the intra-individual correlation coefficient first defined in Chapter 2, Example 2.2.

We may now consider the longitudinal case, where now $E(Y_{ijt})$ is a longitudinal mean of Y_{ijt} for $t = 1,\ldots,T$. The same theory for *meff*s will apply, but with ρ_1^* being replaced by ρ^*, which shall represent the intra-cluster correlation for η_j and $u_{ij} + v_{ijt}$ averaged over the waves, i.e.

$$\rho^* = \frac{\sigma_\eta^2}{[\sigma_\eta^2 + \sigma_u^2 + (\sigma_v^2/T)]}.$$

Hence, under this model, the misspecification effect increases as T increases, if $\sigma_v^2 > 0$.

We may now extend the theoretical reasoning for this finding further, note that the model in (3.8) is seemingly an oversimplification as the area effects are likely to show some variation over time, suggesting that we write η_{jt} rather than η_j so that

$$Y_{ijt} = E(Y_{ijt}) + \eta_{jt} + e_{ijt},$$

where e_{ijt} are random individual effects, and $e_{ijt} = u_{ij} + v_{ijt}$. The intra-cluster correlation, ρ^*, would then become

$$\rho^* = \frac{\mathrm{VAR}(\overline{\eta}_j)}{\mathrm{VAR}(\overline{\eta}_j) + \mathrm{VAR}(\overline{e}_{ij})} = \frac{\sigma_\eta^2(1+(T-1)\rho_\eta^*)}{\sigma_\eta^2(1+(T-1)\rho_\eta^*) + \sigma_e^2(1+(T-1)\rho_e)}, \quad (3.10)$$

where

$$\mathrm{VAR}(\overline{e}_{ij}) = \mathrm{VAR}(\overline{u}_j + \overline{v}_{ij}),$$

$$\overline{\eta}_j = \sum_{i=1}^{T} \eta_{jt} \Big/ T,$$

$$\overline{e}_{ij} = \sum_{i=1}^{T} e_{ijt} \Big/ T,$$

and we have assumed that

$$\rho_\eta^* = \mathrm{CORR}(\eta_{js}, \eta_{jt}), \quad (3.11)$$

and

$$\rho_e = \mathrm{CORR}(e_{ijs}, e_{ijt}), \; s \neq t, \quad (3.12)$$

are both constant, free of s and t. In (3.11) and (3.12), $\mathrm{CORR}(.,.)$ denotes correlation and we have ρ_η^* and ρ_e representing the cluster correlation over time and the individual correlation over time respectively.

Our empirical finding that the misspecification effect for a longitudinal statistics is greater than that for the corresponding cross-section statistic then corresponds, from (3.9), to the inequality $\rho^* > \rho_1^*$ which itself corresponds, from (3.10) to the inequality $\rho_\eta^* > \rho_e$, i.e. that cluster effects are more stable than individual effects.

Remark 3.1: To understand the equalities in (3.10) and the subsequent theoretical discussion, let z_t be any variable. Therefore, we would have

$$\text{VAR}(z_1 + z_2 + \ldots + z_T) = \sum_{t=1}^{T} \text{VAR}(z_t) + \sum_{s \neq t} \text{COV}(z_s, z_t).$$

If we assume that $\text{VAR}(z_t) = \sigma_z^2$,

$$\text{var}(z_1 + z_2 + \ldots + z_T) = T\sigma_z^2 + \sum_{s \neq t} \sigma_z^2 \cdot \text{CORR}(z_s, z_t).$$

And if we assume that $\text{CORR}(z_s, z_t) = \rho_z$,

$$\text{VAR}(z_1 + z_2 + \ldots + z_T) = T\sigma_z^2 + T(T-1)\rho_z\sigma_z^2 = T\sigma_z^2(1 + (T-1)\rho_z).$$

Consequently, we would have as extreme cases, when there is no correlation over time, i.e. $\rho_z = 0$,

$$\text{VAR}(z_1 + z_2 + \ldots + z_T) = T\sigma_z^2,$$

and when there is perfect correlation over time, i.e. $\rho_z = 1$,

$$\text{VAR}(z_1 + z_2 + \ldots + z_T) = T^2\sigma_z^2. \blacksquare$$

Our main argument for the variance effects of clustering to be stronger for longitudinal studies than for cross-sectional ones is that the random cluster effects could be more correlated over the time than the random individual effects. See Example 3.2 below.

Example 3.2: If $\rho_\eta^* = 0.5$, $\rho_e = 0$ and $T = 5$,

$$\rho^* = \frac{3 \cdot \sigma_\eta^2}{3 \cdot \sigma_\eta^2 + \sigma_e^2} > \rho_1^*. \blacksquare$$

Our assertion is based on the presupposition that women's views regarding their roles individually are less stable (less correlated over time) than their views when considered as a group (or as a cluster), so that $\text{var}(\overline{\eta}_j)$ may be expected to decline more slowly with T than $\text{var}(\overline{e}_{ij})$. We may state that our results suggest that the cluster units could possibly be manifesting homogeneity over time. This seems plausible, in particular because individual responses include a measurement error component which tend to reduce ρ_e, whereas random individual measurement error effects may be expected to cancel out in the η_{jt} terms.

Our theoretical explanation may also be supported by the fact that the regression model parameters that we have included in this chapter could be alternatively interpreted as coefficient averages of cross-sectional models over time.

3.5 Concluding remarks

In this chapter we have presented some evidence that, at least for the specific analysis considered, variance-inflating impacts of complex sampling schemes tend to be greater for longitudinal analyses than for corresponding cross-sectional analyses. Our empirical evidence is based upon a regression analysis of longitudinal data on gender role attitudes from the BHPS. We have investigated reasons for this finding and suggest that it arises from a specific longitudinal feature of the analysis.

We have addressed longitudinal aspects of regression analyses of BHPS data on attitudes to gender roles and their relation to demographic and economic variables. In terms of attitudinal model interpretation, we have found from our longitudinal regression model fitting results that the factor that is mostly influencing the women gender role attitudes is the economic activity that the women are involved. Moreover, we confirm previous research results that younger women, employed women and those with higher educational levels tend to have more egalitarian attitudes concerning women's roles (see Berrington, 2002).

In the next chapter, we mainly introduce new approaches for making inference about random effects models. We shall consider the model as a covariance structure model for the $T \times T$ covariance matrix of y_{it}. We shall then consider the population variance-covariance matrix Σ to be a multivariate outcome for each time point and to be constrained to be functions of the $b \times 1$ parameter vector of interest $\underline{\theta}$, as discussed by Skinner (2003b, Section 13.2) and Skinner and Holmes (2003).

Chapter 4

Covariance structure models for complex survey data

4.1 Introduction

We return to the models considered in Chapter 2. We consider a finite population denoted by $\tilde{U}$ (see Chapter 1, Sub-section 1.3.1), which is fixed on occasions $1,\ldots,T$. Let N represent the size of $\tilde{U}$. Let $N_o = N \cdot T$.

Recall that $\underline{Y}_i = (Y_{i1},\ldots,Y_{iT})'$ is a random vector containing T repeated observations on the *study variable* for unit $i = 1,2,\ldots,N$ over the T waves of the survey. As in Chapter 2, let

$$E(\underline{Y}_i) = \underline{\mu}_i(\underline{\beta}) \tag{4.1}$$

be the $T \times 1$ vector with their respective expected values, where

$$\underline{\mu}_i(\underline{\beta}) = [\mu(\underline{x}_{i1},\underline{\beta}),\ldots,\mu(\underline{x}_{iT},\underline{\beta})]' \tag{4.2}$$

and $\underline{x}_{it}$ is a vector of values of the covariates. Again, note the distinction in notation between $\underline{Y}_i$, which represent random variables, and $\underline{y}_i$ which are observed values of $\underline{Y}_i$ (data).

As in Chapter 2, we also assume in this chapter that (*i*) the observations are equally spaced in time, and that (*ii*) the number of individuals is 'large' relative to the number of observations per individual. We also suppose here that (*iii*) the sample is selected on one occasion and then the same sample units are returned to on each of the $T-1$ subsequent waves of the survey; and again (*iv*) we assume that there is not any nonresponse.

A covariance structure model is a model for the $T \times T$ symmetric population variance-covariance matrix of $\underline{Y}_i$, which is

$$\Sigma = \mathrm{COV}(\underline{Y}_i) = E\{[\underline{Y}_i - \underline{\mu}_i][\underline{Y}_i - \underline{\mu}_i]'\} = \Sigma(\underline{\theta}), \tag{4.3}$$

where COV(.) denote population variance and population covariance, respectively. We assume in this chapter that this matrix is the same for each unit *i*, and that $k = T(T+1)/2$ distinct elements of the variance-covariance matrix $\Sigma(\underline{\theta})$ are constrained to be functions of the $b \times 1$ parameter vector of interest $\underline{\theta}$, with $b < k$, as for example in Fuller (1987, Sub-section 4.2.1), Skinner and Holmes (2003), and Hand and Crowder (1996).

If we consider Example 2.1 in Chapter 2, $\Sigma(\underline{\theta})$ would thus have diagonal values $\sigma_u^2 + \sigma_v^2$ and off-diagonal values σ_u^2. Such a covariance structure is called compound symmetric (see Crowder and Hand, 1990; and Jones, 1993).

Another example of this model is given by Diggle *et al.* (2002), who define Σ as

$$\Sigma(\underline{\theta}) = \sigma^2 V(\underline{\alpha}), \quad (4.4)$$

where $\sigma^2 = \text{VAR}(Y_{it})$ is an unknown scalar (assumed to be constant for *t* and *i*), $V(\underline{\alpha})$ is a $T \times T$ matrix with ones on the diagonal, $\underline{\alpha}$ is a parameter vector and we may write $\underline{\theta} = (\sigma^2, \underline{\alpha})$. Note the distinction between V in (4.4) and V, which is the *working covariance matrix* (see Chapter 2).

Jöreskog (1970), Anderson (1973) and Wiley, Schmidt and Bramble (1973) are possibly the earliest articles to suggest the adoption of structural analysis of covariance matrices for estimating variance components. More recently, Pourahmadi (1999), Pourahmadi (2000), and Pan and Mackenzie (2003) discuss modelling mean-covariance structures with applications to longitudinal data, in classical independent and identically distributed (iid) observations multivariate analysis. In this chapter, we shall pay particular attention to inference procedures for random effects models with longitudinal complex survey data. More specifically, we concentrate on new estimation methods for $\underline{\theta}$, allowing for complex survey data.

Estimation procedures for model parameters, previously discussed in Chapter 2, are extended in Section 4.2, whilst Section 4.3 describes methods on inference about the covariance matrix Σ and for the variance estimation of $\hat{\Sigma}$.

Section 4.4 discusses and reviews some estimation methods for the parameter $\underline{\theta}$, including unweighted least squares (Sub-section 4.4.1), generalised least squares under the classical approach (Sub-section 4.4.2), and maximum likelihood (Sub-section 4.4.4), and proposes some new methods, as unweighted least squares for complex surveys (also in Sub-section 4.4.1), generalised least squares under the complex survey approach (Sub-section 4.4.3), and pseudo maximum likelihood (Sub-section 4.4.5). Some concluding remarks are given in Section 4.5.

It is illustrative at this stage to discuss a class of random effects models in Example 4.1 below, additionally to Example 2.1 introduced in Chapter 2. Models discussed in both

Example 2.1 and Example 4.1 may be referred to as variance component models, error component models, mixed effects models, multilevel models and hierarchical models (see Skinner and Holmes, 2003).

Example 4.1: transitory random effects as a first-order autoregressive process

We consider the following class of random effects model (see Model B, in Skinner and Holmes, 2003), given by

$$Y_{it} = \underline{x}_{it}\underline{\beta} + u_i + v_{it}, \text{ with } t = 1,\ldots,T. \tag{4.5}$$

This model is so-called exponential correlation model and it is more elaborate than the one presented in Example 2.1, in Chapter 2. The transitory random effects v_{it} are generated by a stochastic first-order autoregressive process (AR1)[20]

$$v_{it} = \gamma v_{it-1} + \varepsilon_{it}, \text{ with } t = 1,\ldots,T. \tag{4.6}$$

If $\gamma = 0$, $v_{it} = \varepsilon_{it}$, and the AR1 model described here reduces to the uniform correlation model introduced in Chapter 2, Example 2.1. In (4.6), (see Jones, 1993)

$$-1 < \gamma < 1 ,$$

is considered as a regression parameter, ε_{it} are the residuals, with

$$E(\varepsilon_{it}) = 0,$$

and

$$\text{VAR}(\varepsilon_{it}) = \sigma_\varepsilon^2.$$

If we let $\text{VAR}(v_{it}) = \sigma_v^2$ and assume that v_{it} and ε_{it} are mutually independent and stationary, (see Skinner and Holmes, 2003; Crowder and Hand, 1990; and Jones, 1993)

$$\sigma_v^2 = \frac{\sigma_\varepsilon^2}{1-\gamma^2}. \tag{4.7}$$

Taking the expectation of Y_{it}, yields the model

$$E(Y_{it}) = \underline{x}_{it}\underline{\beta} = \underline{\mu}(\underline{x}_{it}, \underline{\beta}), \text{ with } t = 1,\ldots,T, \tag{4.8}$$

which equals (2.5) in Chapter 2. From (4.5), (4.7) and (4.8), the variance of Y_{it} and the covariance between Y_{it} and $Y_{it'}$ are respectively given by

$$\sigma^2 = \text{VAR}(Y_{it}) = \sigma_u^2 + \sigma_v^2 = \sigma_u^2 + \frac{\sigma_\varepsilon^2}{1-\gamma^2},$$

and

---

[20] Jones (1993, section 3.5), and Hand and Crowder (1996, Chapter 6) discuss some issues regarding the adoption of more complicated error structures than AR1.

$$\text{COV}(Y_{it},Y_{it'}) = E\{[u_i]^2\} + E\{v_{it} \cdot v_{it'}\} = \sigma_u^2 + \gamma^{|t-t'|}\sigma_v^2.$$

This model assumes that the covariance within individuals decreases when the distance in time (lag) increases. Note that the AR1 are the most popular models among other specific structures usually adopted in longitudinal analysis, according to Pourahmadi (1999).

For Example 2.1, in Chapter 2, we may define the parameter vector $\underline{\theta}$ as $\underline{\theta} = (\sigma_u^2, \sigma_v^2)'$, with $b = 2$. For the current example, the parameter vector $\underline{\theta}$ may be defined as $\underline{\theta} = (\sigma_u^2, \sigma_v^2, \gamma)'$[21]. ■

When working in the context of the models discussed in the previous example, modelling procedures be considered to belong to the *disaggregated approach* (see Chapter 1, Section 1.4, Sub-section 1.4.2). The target model parameters are defined with respect to some of the *design variables* (see Chapter 1, Section 1.3, Sub-section 1.3.1), including for example individual indicator variables and random effects.

In this chapter, the approach we adopt for making inference about random effects models is to consider the model as a covariance structure model for Σ, which is thus treated as a multivariate outcome for each time point.

4.2 Estimation procedures for parameter $\underline{\beta}$ given $\underline{\theta}$

Recall the estimation procedures for $\underline{\beta}$, first considered in Chapter 2, Sections 2.2 and 2.4. Suppose, following (2.8), that

$$\underline{Y}_i \sim \text{N}_T[\underline{\mu}_i(\underline{\beta}), \Sigma(\underline{\theta})], \qquad (4.9)$$

where $\Sigma(\underline{\theta})$ is a $T \times T$ positive definite covariance matrix.

The census joint density function in (2.9) may now depend on $\underline{\theta}$ as well as $\underline{\beta}$ and becomes

$$f(y_1,\ldots,y_N;\underline{\beta},\underline{\theta}) = (2\pi)^{-N_o/2}|\Sigma(\underline{\theta})|^{-N/2} e^{-\sum_{i=1}^{N}[\underline{y}_i - \underline{\mu}_i(\underline{\beta})]'\Sigma(\theta)^{-1}[\underline{y}_i - \underline{\mu}_i(\underline{\beta})]/2}. \qquad (4.10)$$

The expression in (4.10) is denoted by $\ell_N\lfloor\underline{\beta},\underline{\theta}\rfloor$.

If $\underline{\theta}$ is held fixed and the linear model holds then the pseudo maximum likelihood argument in Chapter 2, Section 2.4 leads to the following expression[22]

---

[21] And we may alternatively define the parameter vector as $\underline{\alpha} = (\alpha_1, \alpha_2)$, where $\alpha_1 = \sigma_u^2/\sigma^2$ and $\alpha_2 = \gamma$, when adopting for example Diggle *et al.* (2002)'s formulation.

[22] Note that Sutradhar and Kovacevic (2000, section 2.5) adopt a similar approach.

$$\underline{\hat{\beta}}_{PML}(\Sigma(\underline{\theta})) = \left(\sum_{i=1}^{n} w_i \underline{x}_i' \Sigma(\underline{\theta})^{-1} \underline{x}_i\right)^{-1} \sum_{i=1}^{n} w_i \underline{x}_i' \Sigma(\underline{\theta})^{-1} \underline{y}_i, \tag{4.11}$$

where V in (2.27) is replaced by $\Sigma(\underline{\theta})$. Note that for calculating $\underline{\hat{\beta}}_{PML}(\Sigma(\underline{\theta}))$, Σ needs to be estimated to be plugged in (4.11). We shall consider methods for calculating $\hat{\Sigma}$ in the following section while methods for estimating $\underline{\theta}$ shall be discussed in Section 4.4.

4.3 Inference about Σ

4.3.1 Estimation of Σ

In this section we consider the estimation of the $T \times T$ variance-covariance matrix of $\underline{Y}_i$, Σ, presented in (4.3). We first examine the finite population covariance matrix S_N, denoted by

$$S_N = [S_N]_{tt'} = \begin{pmatrix} \mathrm{var}_N(\underline{Y}_{i1}) & & & \\ \mathrm{cov}_N(\underline{Y}_{i2}, \underline{Y}_{i1}) & \mathrm{var}_N(\underline{Y}_{i2}) & & \\ \vdots & & \ddots & \\ \mathrm{cov}_N(\underline{Y}_{iT}, \underline{Y}_{i1}) & \mathrm{cov}_N(\underline{Y}_{iT}, \underline{Y}_{i2}) & \cdots & \mathrm{var}_N(\underline{Y}_{iT}) \end{pmatrix}, \tag{4.12}$$

where the subscript N indicates a finite population (or *census*) parameter.

Under the classical set up, it is natural to let $\mathrm{var}_N(.)$ and $\mathrm{cov}_N(.)$, in (4.12), be the population level estimator for the variance and covariance, respectively given by (Skinner, Holt and Smith, 1989, p. 13)

$$\mathrm{var}_N(\underline{Y}_{it}) = (N-1)^{-1} \sum_{i=1}^{N} (y_{it} - \hat{\mu}_{N\,it})^2,$$

and

$$\mathrm{cov}_N(\underline{Y}_t, \underline{Y}_{t'}) = (N-1)^{-1} \sum_{i=1}^{N} (y_{it} - \hat{\mu}_{N\,it})(y_{it'} - \hat{\mu}_{N\,it'}),$$

where $\hat{\mu}_{N\,it}$ is a population level estimator of μ_{it}. Thus, S_N depends upon the estimated mean vector $\underline{\hat{\mu}}_{N\,i}(\hat{\mu}_{N\,i1}, \ldots, \hat{\mu}_{N\,iT})'$.

In the simple case, as for example in Pourahmadi (1999, Section 2.2), when $\underline{\mu}_i = \underline{\mu}$ it is natural to set $\hat{\mu}_{N\,it} = \overline{Y}_t$, where

$$\overline{Y}_t = N^{-1} \sum_{i=1}^{N} y_{it},$$

is the finite population mean of the response variable at time t.

In general, the $T \times 1$ mean vector $\underline{\mu}_i = \underline{\mu}_i(\underline{\beta})$ depends upon the covariates, in which case we might set $\underline{\hat{\mu}}_{Ni} = \underline{\mu}_i(\underline{\hat{\beta}}_N)$ for some estimator $\underline{\hat{\beta}}_N$ of $\underline{\beta}$. For example, we might set $\underline{\hat{\mu}}_{Ni} = \underline{\mu}_i(\underline{\hat{\beta}}_{PML_N}(\mathrm{V} = I))$, which is equivalent to $\underline{\hat{\beta}}_{N,OLS}$ (see Chapter 2, Section 2.2, expression (2.17); Pourahmadi, 1999, Section 2.3; and also Pourahmadi, 2000), where $\underline{\hat{\beta}}_{N,OLS}$ is the census OLS estimator of $\underline{\beta}$, i.e. the value of $\underline{\beta}$ which minimises

$$\sum_{i=1}^{N} [\underline{y}_i - \underline{\mu}_i(\underline{\beta})]' [\underline{y}_i - \underline{\mu}_i(\underline{\beta})].$$

Thus, (Rencher, 1998)

$$\mathrm{var}_N(\underline{Y}_{it}) = (N-q)^{-1} \sum_{i=1}^{N} \left(y_{it} - \underline{x}_{it} \underline{\hat{\beta}}_{N,OLS}\right)^2, \tag{4.13}$$

where q is the number of fixed covariates, as defined in example 4.1, and

$$\mathrm{cov}_N(\underline{Y}_t, \underline{Y}_{t'}) = (N-q)^{-1} \sum_{i=1}^{N} \left(y_{it} - \underline{x}_{it} \underline{\hat{\beta}}_{N,OLS}\right)\left(y_{it'} - \underline{x}_{it'} \underline{\hat{\beta}}_{N,OLS}\right). \tag{4.14}$$

Note that the term $(-q)$ in the denominator of (4.13) and (4.14) is a degrees of freedom bias correction term.

Let $N^* = (N-q)$. Under the assumptions of the classical approach, the matrix S_N is a natural estimator of Σ, as (see, for example, Rencher, 1998, Sub-section 7.4.4; and Johnson and Wichern , 1998, Chapter 7, Result 7.2)

$$E(\mathrm{S}_N) = E\left[\begin{pmatrix} \mathrm{var}_N(\underline{Y}_{i1}) & & & \\ \mathrm{cov}_N(\underline{Y}_{i2}, \underline{Y}_{i1}) & \mathrm{var}_N(\underline{Y}_{i2}) & & \\ \vdots & & \ddots & \\ \mathrm{cov}_N(\underline{Y}_{iT}, \underline{Y}_{i1}) & \mathrm{cov}_N(\underline{Y}_{iT}, \underline{Y}_{i2}) & \cdots & \mathrm{var}_N(\underline{Y}_{iT}) \end{pmatrix}\right] =$$

$$= E\left[\begin{pmatrix} (N^*)^{-1}\sum_{i=1}^{N}(y_{i1} - \underline{x}_{i1}\underline{\hat{\beta}}_{N,OLS})^2 & & & \\ (N^*)^{-1}\sum_{i=1}^{N}(y_{i2} - \underline{x}_{i2}\underline{\hat{\beta}}_{N,OLS})(y_{i1} - \underline{x}_{i1}\underline{\hat{\beta}}_{N,OLS}) & (N^*)^{-1}\sum_{i=1}^{N}(y_{i2} - \underline{x}_{i2}\underline{\hat{\beta}}_{N,OLS})^2 & & \\ \vdots & & \ddots & \\ (N^*)^{-1}\sum_{i=1}^{N}(y_{iT} - \underline{x}_{iT}\underline{\hat{\beta}}_{N,OLS})(y_{i1} - \underline{x}_{i1}\underline{\hat{\beta}}_{N,OLS}) & (N^*)^{-1}\sum_{i=1}^{N}(y_{iT} - \underline{x}_{iT}\underline{\hat{\beta}}_{N,OLS})(y_{i2} - \underline{x}_{i2}\underline{\hat{\beta}}_{N,OLS}) & \cdots & (N^*)^{-1}\sum_{i=1}^{N}(y_{iT} - \underline{x}_{iT}\underline{\hat{\beta}}_{N,OLS})^2 \end{pmatrix}\right] =$$

$$= \begin{pmatrix} \mathrm{VAR}(\underline{Y}_{i1}) & & & \\ \mathrm{COV}(\underline{Y}_{i2}, \underline{Y}_{i1}) & \mathrm{VAR}(\underline{Y}_{i2}) & & \\ \vdots & & \ddots & \\ \mathrm{COV}(\underline{Y}_{iT}, \underline{Y}_{i1}) & \mathrm{COV}(\underline{Y}_{iT}, \underline{Y}_{i2}) & \cdots & \mathrm{VAR}(\underline{Y}_{iT}) \end{pmatrix} = \Sigma(\underline{\theta}) \cdot \tag{4.15}$$

Now suppose, that we only observe $\underline{y}_i$ for units i in a sample s and that s was selected from a complex sampling design. We shall propose S_w to be a weighted sample covariance matrix,

$$S_w = [S_w]_{tt'} = \begin{pmatrix} \text{var}_w(\underline{Y}_{i1}) & & & \\ \text{cov}_w(\underline{Y}_{i2}, \underline{Y}_{i1}) & \text{var}_w(\underline{Y}_{i2}) & & \\ \vdots & & \ddots & \\ \text{cov}_w(\underline{Y}_{iT}, \underline{Y}_{i1}) & \text{cov}_w(\underline{Y}_{iT}, \underline{Y}_{i2}) & \cdots & \text{var}_w(\underline{Y}_{iT}) \end{pmatrix},$$

$$= \begin{pmatrix} \hat{N}^{-1}\sum_{i=1}^{n} w_i (y_{i1} - \hat{\mu}_{i1})^2 & & & \\ \hat{N}^{-1}\sum_{i=1}^{n} w_i (y_{i2} - \hat{\mu}_{i2})(y_{i1} - \hat{\mu}_{i1}) & \hat{N}^{-1}\sum_{i=1}^{n} w_i (y_{i2} - \hat{\mu}_{i2})^2 & & \\ \vdots & & \ddots & \\ \hat{N}^{-1}\sum_{i=1}^{n} w_i (y_{iT} - \hat{\mu}_{iT})(y_{i1} - \hat{\mu}_{i1}) & \hat{N}^{-1}\sum_{i=1}^{n} w_i (y_{iT} - \hat{\mu}_{iT})(y_{i2} - \hat{\mu}_{i2}) & \cdots & \hat{N}^{-1}\sum_{i=1}^{n} w_i (y_{iT} - \hat{\mu}_{iT})^2 \end{pmatrix},$$

where w_i is as defined in Section 4.2,

$$\hat{N} = \sum_{i=1}^{n} w_i ,$$

and

$$\hat{\mu}_{it} = \mu(\underline{x}_{it}, \underline{\hat{\beta}}_{PML}(\text{V})), \tag{4.16}$$

with $\underline{x}_{it}$ and $\underline{\hat{\beta}}_{PML}(\text{V})$ defined respectively as a $1 \times q$ vector with the q fixed covariates and a $q \times 1$ vector with the estimated pseudo-maximum likelihood coefficients.

Following Isaki and Fuller (1982), (see also Chapter 1, Section 1.3, Sub-section 1.3.1)

$$E_p(\hat{N}) = N ,$$

and for large n, we may treat (4.17)

$$\frac{\hat{N}}{N} \doteq 1 .$$

Under the complex design approach,

$$E_p(S_w) = E_p \begin{pmatrix} \hat{N}^{-1}\sum_{i=1}^{n} w_i (y_{i1} - \hat{\mu}_{i1})^2 & & & \\ \hat{N}^{-1}\sum_{i=1}^{n} w_i (y_{i2} - \hat{\mu}_{i2})(y_{i1} - \hat{\mu}_{i1}) & \hat{N}^{-1}\sum_{i=1}^{n} w_i (y_{i2} - \hat{\mu}_{i2})^2 & & \\ \vdots & & \ddots & \\ \hat{N}^{-1}\sum_{i=1}^{n} w_i (y_{iT} - \hat{\mu}_{iT})(y_{i1} - \hat{\mu}_{i1}) & \hat{N}^{-1}\sum_{i=1}^{n} w_i (y_{iT} - \hat{\mu}_{iT})(y_{i2} - \hat{\mu}_{i2}) & \cdots & \hat{N}^{-1}\sum_{i=1}^{n} w_i (y_{iT} - \hat{\mu}_{iT})^2 \end{pmatrix},$$

which, assuming n large and (4.17), is approximately

$$
\doteq \begin{pmatrix} N^{-1}E_p\left[\sum_{i=1}^{n} w_i (y_{i1}-\hat{\mu}_{i1})^2\right] & & & \\ N^{-1}E_p\left[\sum_{i=1}^{n} w_i (y_{i2}-\hat{\mu}_{i2})(y_{i1}-\hat{\mu}_{i1})\right] & N^{-1}E_p\left[\sum_{i=1}^{n} w_i (y_{i2}-\hat{\mu}_{i2})^2\right] & & \\ \vdots & & \ddots & \\ N^{-1}E_p\left[\sum_{i=1}^{n} w_i (y_{iT}-\hat{\mu}_{iT})(y_{i1}-\hat{\mu}_{i1})\right] & N^{-1}E_p\left[\sum_{i=1}^{n} w_i (y_{iT}-\hat{\mu}_{iT})(y_{i2}-\hat{\mu}_{i2})\right] & \cdots & N^{-1}E_p\left[\sum_{i=1}^{n} w_i (y_{iT}-\hat{\mu}_{iT})^2\right] \end{pmatrix},
$$

which, following (1.1) in Chapter 1, Section 1.3, Sub-section 1.3.1, is approximately

$$
\doteq \begin{pmatrix} N^{-1}\sum_{i=1}^{N} (y_{i1}-\hat{\mu}_{N\,i1})^2 & & & \\ N^{-1}\sum_{i=1}^{N} (y_{i2}-\hat{\mu}_{N\,i2})(y_{i1}-\hat{\mu}_{N\,i1}) & N^{-1}\sum_{i=1}^{N} (y_{i2}-\hat{\mu}_{N\,i2})^2 & & \\ \vdots & & \ddots & \\ N^{-1}\sum_{i=1}^{N} (y_{iT}-\hat{\mu}_{N\,iT})(y_{i1}-\hat{\mu}_{N\,i1}) & N^{-1}\sum_{i=1}^{N} (y_{iT}-\hat{\mu}_{N\,iT})(y_{i2}-\hat{\mu}_{N\,i2}) & \cdots & N^{-1}\sum_{i=1}^{N} (y_{iT}-\hat{\mu}_{N\,iT})^2 \end{pmatrix}.
$$

Thus,

$$
E_p(\mathrm{S}_w) \doteq \frac{(N-q)}{N}\mathrm{S}_N \doteq \mathrm{S}_N
$$

assuming $\hat{\mu}_{it}$ is consistent for $\hat{\mu}_{N\,it}$, and from (4.15), we have

$$
E(E_p(\mathrm{S}_w)) \doteq E(\mathrm{S}_N) = \Sigma(\underline{\theta}),
$$

for any fixed choice of V, where V is as previously defined in Sub-section 4.2.1 and in Chapter 2, Section 2.2.

It is also convenient at this stage to define the unweighted sample covariance matrix,

$$
\mathrm{S} = [\mathrm{S}]_{tt'} = \begin{pmatrix} n^{-1}\sum_{i=1}^{n} (y_{i1}-\hat{\mu}_{i1})^2 & & & \\ n^{-1}\sum_{i=1}^{n} (y_{i2}-\hat{\mu}_{i1})(y_{i1}-\hat{\mu}_{i1}) & n^{-1}\sum_{i=1}^{n} (y_{i2}-\hat{\mu}_{i2})^2 & & \\ \vdots & & \ddots & \\ n^{-1}\sum_{i=1}^{n} (y_{iT}-\hat{\mu}_{iT})(y_{i1}-\hat{\mu}_{i1}) & n^{-1}\sum_{i=1}^{n} (y_{iT}-\hat{\mu}_{iT})(y_{i2}-\hat{\mu}_{i2}) & \cdots & n^{-1}\sum_{i=1}^{n} (y_{iT}-\hat{\mu}_{iT})^2 \end{pmatrix}, \quad (4.18)
$$

Note that, as S is defined here as a special case of S_w when the *sampling weight*s are constant, we shall not consider the degrees of freedom bias correction term in the denominator of each element of (4.18).

4.3.2 Variance estimation for $\hat{\Sigma}$

Under the classical set up and under normality (Browne, 1987; see also Jöreskog and Goldberger, 1972; Swain, 1975; Fuller, 1987, Sub-section 4.2.1, Theorem 4.2.2; Shapiro and Browne, 1987; Amemiya and Anderson, 1990; Ghosh, 1996; and Yuan and Bentler, 1997b),

$$
n \cdot \mathrm{S} \sim \mathrm{W}_T(\Sigma, n-1),
$$

where W_T is T-variate Wishart distribution with n degrees of freedom and parameter Σ, with $\Sigma = \Sigma(\underline{\theta})$ and n denoting the sample size. See Krzanowski (2000, Section 7.3), Johnson and Wichern (1998, Section 4.4), and Rencher (1998, Chapter 2), for example, for information on the properties of the Wishart distribution.

We shall initially present in this sub-section a brief review of variance estimation methods for $\hat{\Sigma}$ that assume multivariate normality for $\underline{Y}_i$ and no covariates. From the classical multivariate analysis literature, we may write the following expression, (Jöreskog and Goldberger, 1972; see also Swain, 1975; and Jöreskog and Sörbom, 1997)

$$a\mathrm{COV}(\mathrm{S}_{tt'}, \mathrm{S}_{t''t'''}) = n^{-1} \cdot \{\Sigma_{tt''} \cdot \Sigma_{t't'''} + \Sigma_{tt'''} \cdot \Sigma_{t't''}\}, \tag{4.19}$$

which is an asymptotic covariance $k \times k$ matrix of S.

Let $vech[\mathrm{S}]$ and $vech[\Sigma(\underline{\theta})]$ be $k \times 1$ vectors formed from the nonduplicated elements of S and $\Sigma(\underline{\theta})$, respectively (see Browne, 1982)

$$vech[\mathrm{S}] = [S_{11}, S_{21}, \ldots, S_{T1}, S_{22}, \ldots, S_{T2}, \ldots, S_{TT}],$$

and (4.20)

$$vech[\Sigma(\theta)] = [\Sigma_{11}, \Sigma_{21}, \ldots, \Sigma_{T1}, \Sigma_{22}, \ldots, \Sigma_{T2}, \ldots, \Sigma_{TT}].$$

Moreover, let the *residual covariance matrix* be denoted by E, so that

$$\mathrm{E} = [\mathrm{S} - \Sigma(\underline{\theta})]. \tag{4.21a}$$

Then, we may alternatively write

$$vech[\mathrm{E}] = vech[\mathrm{S}] - vech[\Sigma(\underline{\theta})],$$

where the operator *vech* is as defined for (4.20). Note so forth the distinction in notation between E, which is adopted for the *residual covariance matrix*, and E which is adopted as the usual symbol for the expectation operator (see Chapter 1, Section 1.3, Sub-section 1.3.1).

We consider that (Browne, 1977)

$$E\{vech[\mathrm{E}]\} = 0, \tag{4.21b}$$

as for example in Fuller (1987, Sub-section 4.2.1), which is equivalent to

$$E(\mathrm{S}) = \Sigma(\underline{\theta}). \tag{4.22}$$

Let

$$\mathrm{C} = n \cdot \mathrm{COV}\{vech[\mathrm{E}], vech[\mathrm{E}]'\}.$$

Let $\bar{\mathrm{c}}_{tt',t''t'''}$ represent a typical element of $\bar{\mathrm{C}}$, where

$$\bar{\mathrm{C}} = \lim_{n \to \infty} \mathrm{C}.$$

Under the assumptions of normality and iid observations with no weighting, the elements of the matrix $\overline{\text{C}}$ have the following form (Browne, 1982),

$$\overline{c}_{tt',t''t'''} = \Sigma_{tt''} \cdot \Sigma_{t't'''} + \Sigma_{tt'''} \cdot \Sigma_{t't''} . \qquad (4.23)$$

We treat $\overline{\text{C}}/n$ as the asymptotic covariance matrix of $vech[\text{S}]$. We seek a consistent estimator of this matrix and let U be a consistent estimator of $\overline{\text{C}}$, i.e.

$$p\lim_{n\to\infty}(\text{U}-\text{C}) = 0 .$$

See Bollen (1989, Appendix B), for a brief overview on asymptotic distribution theory.

Let $\text{u}_{tt',t''t'''}$ represent a typical element of the $k \times k$ matrix U. Additionally let the $T \times T$ matrix W be any consistent estimator of Σ (Bentler and Weeks, 1980; Swain, 1975), such as S (or S_w). Under the classical set up, we may set (Jöreskog and Goldberger, 1972; Swain, 1975)

$$\text{u}_{tt',t''t'''} = \omega_{tt''}\omega_{t't'''} + \omega_{tt'''}\omega_{t't''} , \qquad (4.24)$$

where $\omega_{tt'}$ represent a typical element of W. In this case U is consistent for C as required (Browne, 1982).

So far in this sub-section we have discussed procedures suitable for normally distributed variables under iid assumptions. We shall consider below techniques for non-normally distributed data or asymptotically distribution-free (ADF) methods. Thus, for data with arbitrary distribution, matrix $\overline{\text{C}}$ would have the following typical element

$$\overline{c}^{*}_{tt',t''t'''} = \Sigma_{tt't''t'''} - \Sigma_{tt'} \cdot \Sigma_{t''t'''} ,$$

according to Browne (1984; see also Bentler and Dudgeon, 1996, for example), where $\Sigma_{tt't''t'''}$ are forth-order moments about the mean.

Suppose again that there are no covariates and again that the complex design is ignored so that $\underline{\mu}_i = \underline{\mu}$ and $\hat{\mu}_{it} = \overline{y}_t$, where μ is as introduced in Section 4.1 and Section 4.3, Sub-section 4.3.1, for example. In addition, let (see, for example, Bentler and Dudgeon, 1996; and Olsson, Foss, and Troye, 2003)

$$m_{tt',t''t'''} = n^{-1}\sum_{i=1}^{n}(y_{it} - \overline{y}_t)(y_{it'} - \overline{y}_{t'})(y_{it''} - \overline{y}_{t''})(y_{it'''} - \overline{y}_{t'''}) , \qquad (4.25)$$

where $m_{tt',t''t'''}$ are fourth-order sample moments about the mean. Let U^* be a $k \times k$ matrix and let $\text{u}^*_{tt',t''t'''}$ represent a typical element of U^*. An assumption that all eighth-order moments are finite guarantees that $[\mathrm{S}]_{tt'}$ will be a consistent estimator of $[\Sigma]_{tt'}$ without specifying a particular distribution for $\underline{Y}_i$, so that (Browne, 1982; Browne, 1984)

$$\mathrm{u}^*_{tt',t''t'''} = m_{tt't''t'''} - \mathrm{S}_{tt'} \cdot \mathrm{S}_{t''t'''}, \tag{4.26}$$

will be a consistent estimator of $\overline{\mathrm{c}}^*_{tt',t''t'''}$.

Note that U could be a stochastic matrix so that (Browne, 1984)

$$\overline{\mathrm{U}}^* = \lim_{n\to\infty} \mathrm{U}^*, \tag{4.27}$$

where $\overline{\mathrm{U}}^*$ is a matrix of constant rank (see also, for example, Satorra, 1989, Remark 3.1).

Expression (4.26) is an estimator of the asymptotic covariance between the 'product variables' $(y_{it} - \overline{Y}_t)(y_{it'} - \overline{Y}_{t'})$ and $(y_{it''} - \overline{Y}_{t''})(y_{it'''} - \overline{Y}_{t'''})$. Each element of U^* may thus be calculated by (4.26). See also Bollen (1989, p. 425-427), and Muthén and Satorra (1995) for further information. Moreover, if n is larger than k, and $\overline{\mathrm{C}}$ is positive definite, matrix U^* described by (4.26) will certainly be positive definite, according to Browne (1984).

Recall that previous discussion assumes $\underline{\mu}_i = \underline{\mu}$. We propose initially to generalise $m_{tt't''t'''}$, given by (4.25), to

$$m_{tt't''t'''} = n^{-1}\sum_{i=1}^{n} \hat{\varepsilon}_{it}\hat{\varepsilon}_{it'}\hat{\varepsilon}_{it''}\hat{\varepsilon}_{it'''}, \tag{4.28}$$

where

$$\underline{\hat{\varepsilon}}_i = \underline{y}_i - \underline{\hat{\mu}}_i. \tag{4.29}$$

In (4.29), $\underline{\hat{\varepsilon}}_i$, $\underline{y}_i$ and $\underline{\hat{\mu}}_i$ are $T \times 1$ vectors, and $\underline{\hat{\varepsilon}}_i = [\hat{\varepsilon}_{i1},\ldots,\hat{\varepsilon}_{iT}]'$, with

$$\underline{\hat{\mu}}_i = \underline{x}_{it}\underline{\hat{\beta}}_{PML}(\mathrm{V}). \tag{4.30}$$

In (4.30), $\underline{\hat{\beta}}_{PML}(\mathrm{V})$ is given by (2.15) in Chapter 2, Section 2.2, which is a special case of $\underline{\hat{\beta}}_{PML}(\mathrm{V})$ when the *sampling weights* are constant.

Moreover, we may alternatively represent S, given in (4.18), as

$$[\mathrm{S}]_{tt'} = n^{-1}\sum_{i=1}^{n} \underline{\hat{\varepsilon}}_i \underline{\hat{\varepsilon}}_i', \tag{4.31}$$

where $\underline{\hat{\varepsilon}}_i$ is as defined above in expressions (4.29) and (4.30).

Recall that $vech[\mathrm{S}]$ is a $k \times 1$ vector formed from the non-redundant duplicated elements of S and write

$$\underline{\bar{c}} = vech[\mathrm{S}] = \begin{bmatrix} \bar{c}_1 \\ \vdots \\ \bar{c}_k \end{bmatrix} = \begin{bmatrix} n^{-1}\sum_{i=1}^{n}\hat{\varepsilon}_{i1}\hat{\varepsilon}_{i1} \\ n^{-1}\sum_{i=1}^{n}\hat{\varepsilon}_{i2}\hat{\varepsilon}_{i1} \\ \vdots \\ n^{-1}\sum_{i=1}^{n}\hat{\varepsilon}_{iT}\hat{\varepsilon}_{i(T-1)} \\ n^{-1}\sum_{i=1}^{n}\hat{\varepsilon}_{iT}\hat{\varepsilon}_{iT} \end{bmatrix} = n^{-1}\sum_{i=1}^{n}\underline{c}_i, \tag{4.32}$$

where

$$\underline{c}_i = \begin{bmatrix} \hat{\varepsilon}_{i1}\hat{\varepsilon}_{i1} \\ \hat{\varepsilon}_{i2}\hat{\varepsilon}_{i1} \\ \vdots \\ \hat{\varepsilon}_{iT}\hat{\varepsilon}_{i(T-1)} \\ \hat{\varepsilon}_{iT}\hat{\varepsilon}_{iT} \end{bmatrix},$$

with

$$E(\underline{\bar{c}}) = vech[\Sigma].$$

Note the distinction in notation between the italicised c ($\underline{\bar{c}}$, $\bar{c}_1$ and $\underline{c}_i$, for example) introduced above, and $\mathrm{c}_{tt',t''t'''}$ which was introduced earlier in this sub-section and represents a typical element of matrix C.

According to Browne (1984), $vech[\mathrm{S}]$ is generally asymptotically normally distributed under the classical set up. See Press (1972, Section 4.4) for information on multivariate central limit theorems. By adopting (4.26), (4.28), and (4.31), the $k \times k$ covariance matrix $\mathrm{VAR}\{vech[\mathrm{S}]\}$ could be estimated, assuming independence, by (Skinner, 1989a, Section 3.4; see also Fuller, 1987, Sub-section 4.2.1, Example 4.2.3)

$$\mathrm{var}(\underline{\bar{c}}) = n^{-1}(n-1)^{-1}\sum_{i=1}^{n}(\underline{c}_i - \underline{\bar{c}})(\underline{c}_i - \underline{\bar{c}})',$$

which may alternatively expressed as

$$\begin{aligned}
[\mathrm{var}(\underline{\bar{c}})]_{(tt'),(t''t''')} &= n^{-1}(n-1)^{-1}\sum_{i=1}^{n}\left[\left(\hat{\varepsilon}_{it}\hat{\varepsilon}_{it'} - n^{-1}\sum_{i}^{n}\hat{\varepsilon}_{it}\hat{\varepsilon}_{it'}\right)\cdot\left(\hat{\varepsilon}_{it''}\hat{\varepsilon}_{it'''} - n^{-1}\sum_{i}^{n}\hat{\varepsilon}_{it''}\hat{\varepsilon}_{it'''}\right)\right] = \\
&= n^{-1}(n-1)^{-1}\sum_{i=1}^{n}(\hat{\varepsilon}_{it}\hat{\varepsilon}_{it'} - S_{tt'})\cdot(\hat{\varepsilon}_{it''}\hat{\varepsilon}_{it'''} - S_{t''t'''}) \\
&= (n-1)^{-1}\left[n^{-1}\sum_{i=1}^{n}\hat{\varepsilon}_{it}\hat{\varepsilon}_{it'}\hat{\varepsilon}_{it''}\hat{\varepsilon}_{it'''} - S_{tt'}S_{t''t'''}\right] \\
&= (n-1)^{-1}\cdot \mathrm{u}^{*}_{tt',t''t'''},
\end{aligned} \tag{4.33}$$

where $\underline{\bar{c}}$ is as defined by (4.32) and the subscripts t denote time. Fuller (1987) considers this type of estimation method to be robust to deviations from distributional assumptions and that $\text{var}(\underline{\bar{c}})$ is a consistent estimator of $\text{VAR}\{vech[\text{S}]\}$. Note that by adopting (4.33) for estimating $\text{VAR}\{vech[\text{S}]\}$ we are dividing $u^{*}_{tt',t''t'''}$ by $(n-1)$, as in Skinner (1989a) instead of n.

Inference techniques appropriate to non-normal data could be generalised for the analysis of complex survey data, as suggested for example by Skinner (1989a), followed by Satorra (1992), and Muthén and Satorra (1995). We have considered above, the case where the sampling weights are constant.

Furthermore, we shall consider the case of unequal weights. We consider below a weighted version of the *residual covariance matrix*. Let E_c be a $T \times T$ defined as

$$\text{E}_c = \text{S}_w - \Sigma(\underline{\theta}),$$

where the subscript c denotes 'complex', S_w is the weighted sample covariance matrix, as defined in Sub-section 4.3.1.

We may alternatively define

$$vech[\text{E}_c] = vech[\text{S}_w] - vech[\Sigma(\underline{\theta})],$$

where $vech[\text{E}_c]$ is a $k \times 1$ vector formed from the nonduplicated elements of E_c.

Similarly to (4.21b) and (4.22) we assume that

$$E\{vech[\text{E}_c]\} = 0,$$

$$E_p\{vech[\text{E}_c]\} = 0,$$

and

$$n \cdot \text{COV}\{vech[\text{E}_c], vech[\text{E}_c]'\} = \text{C}_c,$$

where n is the sample size, $E(.)$ denotes the model expectation, $E_p(.)$ denotes the expectation with respect to the sampling distribution of statistics over repeated samples s generated by the sampling design, and C_c is a $k \times k$ nonnegative definite matrix defined, under complex survey design.

We assume that (similarly to Browne, 1984)

$$\bar{\text{C}}_c = \lim_{n \to \infty} \text{C}_c,$$

the limiting covariance matrix exists.

Let $vech[\text{S}_w]$, be a $k \times 1$ vector formed from the nonduplicated elements of S_w, defined in Sub-section 4.3.1, and w_i denote the *sampling weight* for individual i, so that

$$vech[\mathrm{S}_w]=\begin{bmatrix}\left(\sum_{i=1}^{n}w_i\right)^{-1}\sum_{i=1}^{n}w_i\hat{\varepsilon}_{i1}\hat{\varepsilon}_{i1}\\ \left(\sum_{i=1}^{n}w_i\right)^{-1}\sum_{i=1}^{n}w_i\hat{\varepsilon}_{i2}\hat{\varepsilon}_{i1}\\ \vdots\\ \left(\sum_{i=1}^{n}w_i\right)^{-1}\sum_{i=1}^{n}w_i\hat{\varepsilon}_{iT}\hat{\varepsilon}_{i(T-1)}\\ \left(\sum_{i=1}^{n}w_i\right)^{-1}\sum_{i=1}^{n}w_i\hat{\varepsilon}_{iT}\hat{\varepsilon}_{iT}\end{bmatrix}=\left(\sum_{i=1}^{n}w_i\right)^{-1}\sum_{i=1}^{n}w_i\underline{c}_i, \tag{4.34}$$

where $\underline{c}_i$ is a $k\times 1$ vector which is defined as for (4.32). In (4.34), $\underline{\hat{\varepsilon}}_i$ is as defined in expressions (4.29) and (4.30) now with $\underline{\hat{\beta}}_{PML}(\mathrm{V})$ instead given by (2.27) in Chapter 2, Section 2.4, which allows for unequal *sampling weight*s.

Note that

$$\underline{\bar{c}}_w = vech[\mathrm{S}_w]=\frac{\sum_{i=1}^{n}w_i\underline{c}_i}{\sum_{i=1}^{n}w_i}, \tag{4.35}$$

is being expressed as a ratio of two totals.

We shall adopt below the linearization variance estimation method for developing a derivation of the approach recommended by Browne (1984). See also Chapter 2, Section 2.5, Sub-section 2.5.1 of this thesis for some characteristics of the Taylor linearization method. We shall propose an estimator for $\mathrm{VAR}(vech[\mathrm{S}_w])$, which considers complex sampling schemes, following Skinner (1989a). Hence, we may explicitly rewrite (4.35) as

$$vech[\mathrm{S}_w]=\frac{\underline{\bar{z}}}{\bar{w}},$$

where

$$\bar{w}=n^{-1}\sum_{i=1}^{n}w_i,$$

and (4.36)

$$\underline{\bar{z}}=n^{-1}\sum_{i=1}^{n}\underline{z}_i,$$

with

$$\underline{z}_i = w_i\cdot\underline{c}_i.$$

In (4.36), n is the sample size, w_i is the sampling weight for individual i, $\underline{c}_i$ is defined as for (4.32), and $\underline{\bar{z}}$ and $\underline{z}_i$ are $k\times 1$ vectors. A first order Taylor expansion gives (see Binder, 1983; Woodruff, 1971; Cochran, 1977, p. 169-171; Skinner, 1989b; and Shah *et al.*, 1995)

$$vech[\mathrm{S}_w] \doteq \frac{\mu_z}{\mu_w} + \frac{1}{n}\sum_{i=1}^{n}\left(z_i - \frac{\mu_z}{\mu_w} w_i\right)\cdot\frac{1}{\mu_w}.$$

Thus, the variance of $vech[\mathrm{S}_w]$ may be approximated by

$$\mathrm{VAR}\{vech[\mathrm{S}_w]\} \doteq \mathrm{VAR}\left(\frac{1}{n}\cdot\sum_{i=1}^{n}\underline{u}_i\right) = \left(\frac{1}{n}\right)^2\cdot\mathrm{VAR}\left(\sum_{i=1}^{n}\underline{u}_i\right) = \left(\frac{1}{n}\right)^2\cdot\mathrm{VAR}(\underline{\mathrm{B}}), \tag{4.37}$$

where the $k\times 1$ vector $\underline{u}_i$ is given by

$$\underline{u}_i = \frac{1}{\underline{\mu}_w}\cdot\left(\underline{z}_i - \frac{\underline{\mu}_z}{\underline{\mu}_w} w_i\right), \tag{4.38}$$

and

$$\underline{\mathrm{B}} = \sum_{i=1}^{n}\underline{u}_i\,.$$

We could then estimate $\underline{u}_i$ and $\underline{\mathrm{B}}$ respectively by

$$\hat{\underline{u}}_i = \frac{1}{\overline{w}}\cdot\left(\underline{z}_i - \frac{\overline{\underline{z}}}{\overline{w}} w_i\right) \tag{4.38b}$$

and

$$\hat{\underline{\mathrm{B}}} = \sum_{i=1}^{n}\hat{\underline{u}}_i\,.$$

Note that $\hat{\underline{\mathrm{B}}}$, above, is a $k\times 1$ vector of totals and thus has the form of an estimate of a population total vector $\underline{\mathrm{B}}$.

We may then revisit Example 2.2 discussed in Chapter 2, where we considered a multistage stratified sampling scheme that involves sampling with replacement at the first stage of primary sampling units (PSUs) from each of a total of *H* strata, and sampling with or without replacement at subsequent stages. We also consider equal or unequal selection probabilities at both the first and subsequent stages. We may thus explicitly consider stratification and clustering by rewriting $\hat{\underline{\mathrm{B}}}$ as

$$\hat{\underline{\mathrm{B}}} = \sum_{h=1}^{H}\hat{\underline{\mathrm{B}}}_h = \sum_{h=1}^{H}\sum_{j=1}^{m_h}\hat{\underline{\mathrm{B}}}_{hj} = \sum_{h=1}^{H}\sum_{j=1}^{m_h}\sum_{i=1}^{n_{hj}}\hat{\underline{u}}_{hji}\,,$$

where *H* is the number of strata in the sample, m_h is the sample number of PSUs in stratum *h*, n_{hj} is the sample number of individual size in PSU *j* in stratum *h*, and $\hat{\underline{u}}_{hji}$ is the $k\times 1$ vector for individual *i* in PSU *j* in stratum *h*. From Shah *et al.* (1995, Sub-section 2.2.3), an estimator for the covariance matrix of $\hat{\underline{\mathrm{B}}}$, considering the sampling scheme described above, is given by

$$\mathrm{var}_L[\hat{\underline{\mathrm{B}}}]_{v,l} = \sum_{h=1}^{H} m_h\left\{\left[\sum_{j=1}^{m_h}\left(\hat{\underline{\mathrm{B}}}_{hj,v} - \overline{\hat{\underline{\mathrm{B}}}}_{h,v}\right)'\left(\hat{\underline{\mathrm{B}}}_{hj,l} - \overline{\hat{\underline{\mathrm{B}}}}_{h,l}\right)\right]\Big/(m_h - 1)\right\}, \tag{4.39}$$

with subscripts v and l denoting respectively $v=(t,t')$ and $l=(t'',t''')$. In (4.39) $\underline{\hat{\mathrm{B}}}_{hj}$ is an estimator of a total vector in PSU j in stratum h, and $\overline{\underline{\hat{\mathrm{B}}}}_h$ is the mean of $\underline{\hat{\mathrm{B}}}_{hj}$ in stratum h. See also Cochran (1977, Section 11.9). Note that (4.37) could thus be estimated by

$$\left(\frac{1}{n}\right)^2 \cdot \mathrm{var}_L[\underline{\hat{\mathrm{B}}}]_{v,l}. \tag{4.39b}$$

We shall also propose a special case of the estimator considered above when working under independence assumptions. We could consider that the population consists of only one stratum and that each individual i is a PSU. Under this circumstances, (4.39) could be rewritten as

$$\mathrm{var}[\underline{\hat{\mathrm{B}}}]_{v,l} = \frac{n}{n-1} \cdot \sum_{i=1}^{n} \left(\underline{\hat{u}}_{i,v} - \overline{\underline{\hat{u}}}_v\right)' \left(\underline{\hat{u}}_{i,l} - \overline{\underline{\hat{u}}}_l\right),$$

where $\underline{\hat{u}}_i$ is given by (4.38b), and

$$\overline{\underline{\hat{u}}} = \frac{1}{n} \cdot \left(\sum_{i=1}^{n} \underline{\hat{u}}_i\right). \tag{4.39c}$$

Note that (4.38b) could be rewritten as

$$\underline{\hat{u}}_i = \frac{n}{\sum_{i=1}^{n} w_i} \cdot \left[w_i \cdot \left(\underline{c}_i - \overline{\underline{c}}_w\right)\right],$$

with $\overline{\underline{c}}_w$ given by (4.35), and that (4.39c) is a null vector. Thus, $\mathrm{VAR}(vech[\mathrm{S}_w])$ could alternatively be estimated by

$$\begin{aligned}
\left[\mathrm{var}(\overline{\underline{c}}_w)\right]_{v,l} &= \frac{1}{n^2} \cdot \frac{n}{(n-1)} \cdot \left(n \Big/ \sum_{i=1}^{n} w_i\right)^2 \cdot \sum_{i=1}^{n} w_i^2 \left(\underline{c}_{i,v} - \overline{\underline{c}}_{w,v}\right) \cdot \left(\underline{c}_{i,l} - \overline{\underline{c}}_{w,l}\right) \\
&= \frac{1}{n^2} \cdot \frac{n}{(n-1)} \cdot \left(n \Big/ \sum_{i=1}^{n} w_i\right)^2 \cdot \sum_{i=1}^{n} w_i^2 \left(\underline{c}_{i,v} - \left(\sum_{i=1}^{n} w_i\right)^{-1} \sum_{i=1}^{n} w_i \underline{c}_{i,v}\right) \cdot \left(\underline{c}_{i,l} - \left(\sum_{i=1}^{n} w_i\right)^{-1} \sum_{i=1}^{n} w_i \underline{c}_{i,l}\right) \\
&= \frac{1}{n(n-1)} \cdot \left(n \Big/ \sum_{i=1}^{n} w_i\right)^2 \cdot \sum_{i=1}^{n} w_i^2 \left(\hat{\varepsilon}_{it}\hat{\varepsilon}_{it'} - S_{w\,tt'}\right) \cdot \left(\hat{\varepsilon}_{it''}\hat{\varepsilon}_{it'''} - S_{w\,t''t'''}\right).
\end{aligned} \tag{4.40}$$

Note that (4.40) may also be derived from the standard expression for a variance of a ratio estimate, in the simple random sampling context (see Cochran, 1977, Chapter 6, Section 6.3; see also Särndal, Swenson and Wretman, 1992, Chapter 5, Section 5.3), when ignoring the finite population correction. Thus, note that even though the estimator given in (4.40) allows for unequal sampling weights, it does not fully accounts for complex sampling designs. Note that (4.40) reduces to (4.33) when the sampling weights are constant.

4.4 Estimation methods for parameter $\underline{\theta}$

In this section we are interested in estimating $\underline{\theta}$. From Section 4.3 we suppose that S_w (or S) is a good estimator of $\Sigma(\underline{\theta})$. We should like to estimate $\underline{\theta}$ by an estimator $\hat{\underline{\theta}}$, such that S_w (or S) and $\hat{\Sigma} = \Sigma(\hat{\underline{\theta}})$ are 'close' (Jöreskog and Goldberger, 1972; Swain, 1975). We let $F(S_w, \Sigma)$ be the fitting function[23], which measures the distance between S_w and Σ, and following Browne (1984) let

$$F\left(S_w, \Sigma(\hat{\underline{\theta}})\right) = \min_{\underline{\theta} \in \Theta} F(S_w, \Sigma(\underline{\theta})), \qquad (4.41)$$

where Θ is a parameter space, which is a subset of a b-dimensional Euclidean space, with b as defined in Section 4.1. For further information, see also Browne (1987), Shapiro (1986), and Satorra (1989). The minimisation of $F(S_w, \Sigma(\underline{\theta}))$, with respect to $\underline{\theta}$ yields $\hat{\underline{\theta}}$, which is the minimum discrepancy function estimator of $\underline{\theta}$. We assume this is uniquely defined by (4.41). For simplification we shall denote the fitting function F by $F(\underline{\theta})$.

We now consider the choice of $F(\underline{\theta})$. The fitting function $F(\underline{\theta})$ may be formulated in different ways. We discuss in this section three important well known choices of $F(\underline{\theta})$, which lead to different estimation methods of $\underline{\theta}$, (*i*) unweighted least squares, (*ii*) generalised least squares, and (*iii*) maximum likelihood. We additionally propose fitting functions, and consequently estimation methods, which allow for data from complex surveys.

We assume the following properties (see Browne, 1982; and Browne, 1984): (*i*) $F(\underline{\theta})$ is a scalar; (*ii*) $F(\underline{\theta}) \geq 0$; (*iii*) $F(S_w, \Sigma(\underline{\theta})) = 0$ if and only if $S_w = \Sigma$; and (*iv*) $F(\underline{\theta})$ is a twice continuously differentiable function of S_w and Σ. These will be assumed throughout the current section, and in the remaining of this thesis even when no explicit discussion is made. See also Jöreskog and Goldberger (1972), Browne (1987), Shapiro (1986), Satorra and Bentler (1986), Satorra and Bentler (1988), Bollen (1989), Satorra (1989), Satorra (1992), and Bentler and Dudgeon (1996), for example, for some discussion on these assumptions.

If $\underline{\theta}$ is identified[24] and $\Sigma(\underline{\theta})$ is continuous in $\underline{\theta}$, the minimisation of any $F(S_w, \Sigma(\underline{\theta}))$ will provide consistent $\hat{\underline{\theta}}$ (see Browne, 1982; and Browne, 1984, *Proposition* 1) assuming S_w

[23] It is so called discrepancy function, and less frequently loss function.

[24] Let us consider an unknown parameter as a parameter whose identification status is unknown. According to Bollen (1989, p. 89): "*If an unknown parameter in* $\underline{\theta}$ *can be written as a function of one or more elements of* Σ,

(or S) is consistent for $\Sigma(\underline{\theta})$. Moreover, under appropriate regularity conditions (see Swain, 1975, with proof; see also Browne, 1982; Lee, 1985; Shapiro, 1986; Browne, 1987; Satorra and Bentler 1988; Chou, Bentler and Satorra, 1991; Satorra, 1992; Satorra and Bentler, 1994; Yuan and Bentler, 1997a; Yuan and Bentler, 1997b),

$$\sqrt{n} \cdot (\hat{\underline{\theta}} - \underline{\theta})$$

is asymptotically normal with zero mean and a nonnegative definite covariance matrix.

4.4.1 Unweighted least squares (ULS)

The ULS approach is analogous to the OLS regression. But the ULS addresses attention to the observed and predicted covariances, while the OLS works with the response variable observed and predicted values. So forth, let *tr* refer to the trace of a matrix. The ULS method has the following fitting function (see Jöreskog and Goldberger, 1972; Knight, 1978; Long, 1983; and Bollen, 1989),

$$F(\underline{\theta})_{ULS} = \frac{1}{2} \cdot tr\{[\mathrm{S} - \Sigma(\underline{\theta})]^2\}, \tag{4.42a}$$

which is equivalent to one half the sum of squares of each element in the $T \times T$ *residual covariance matrix* E, defined in Subsection 4.3.2. When minimising $F(\underline{\theta})_{ULS}$, we minimise the differences between the empirical variances and covariances and the corresponding ones predicted by the model that we consider. The square in (4.42a) denotes the square of a matrix, i.e. the matrix multiplied by itself.

Under a complex design, unweighted sample covariance matrix S may not be consistent for Σ and it is more natural to consider the following adaptation to the ULS fitting function,

$$F(\underline{\theta})_{ULSC} = \frac{1}{2} \cdot tr\{[\mathrm{S}_w - \Sigma(\underline{\theta})]^2\}, \tag{4.42b}$$

where *ULSC* indicates that complex survey data is considered, and S_w is the weighted sample covariance matrix defined in Sub-section 4.3.1. Note again that S is a special case of S_w when the *sampling weights* are constants.

The estimates for $\underline{\theta}$ have to be produced aiming the minimisation of $F(\underline{\theta})_{ULS}$ or $F(\underline{\theta})_{ULSC}$. The fitting function $F(\underline{\theta})_{ULSC}$, for example, is minimised when $\hat{\underline{\theta}}_{ULSC}$ is chosen by differentiating $F(\underline{\theta})_{ULSC}$,

---

that parameter is identified. If all unknown parameters in $\underline{\theta}$ are identified, then the model is identified. Moreover, if a parameter is not identifiable in a model it has no consistent estimator (Bentler and Weeks, 1980), although identifiability does not necessarily mean that there is a consistent estimator.

$$\frac{\partial F(\underline{\theta})_{ULSC}}{\partial \underline{\theta}} = 0,$$

with respect to each of the b unknown parameters included in the parameter vector $\underline{\theta}$.

Once $\underline{\hat{\theta}}_{ULSC}$ is estimated, a sufficient condition for it to minimise $F(\underline{\theta})_{ULSC}$ is that the following matrix (see Bollen, 1989),

$$\Omega_{ULS} = \begin{bmatrix} \frac{\partial^2 F(\underline{\theta})_{ULS}}{\partial(\hat{\gamma})^2} & 0 & 0 \\ 0 & \frac{\partial^2 F(\underline{\theta})_{ULS}}{\partial(\hat{\sigma}_v^2)^2} & 0 \\ 0 & 0 & \frac{\partial F(\underline{\theta})_{ULS}}{\partial(\hat{\sigma}_u^2)^2} \end{bmatrix},$$

is positive definite, when $\underline{\theta} = (\sigma_u^2, \sigma_v^2, \gamma)'$, for example. In addition to being positive definite, Ω should be regular in order to guarantee the identification of $\underline{\theta}$. Note that this condition could not hold in situations where any element of $\underline{\theta}$ approaches zero. Matrix Ω_{ULS} is a $b \times b$ diagonal matrix with the second partial derivatives of $F(\underline{\theta})_{ULS}$, with respect to $\underline{\theta}$, on the diagonal.

As long as $\underline{\theta}$ is identified and S_w (or S) is consistent for Σ, the ULSC (or ULS) method leads to consistent estimation of $\underline{\theta}$. Although the ULS type of methods has a very intuitive, computationally cheap, and an easy to understand fitting function, it has some disadvantages. It does not lead to the asymptotically most efficient estimator of $\underline{\theta}$, and it is neither scale invariant nor scale free (Jöreskog and Goldberger, 1972). More information on scale invariant and freeness is given in Bollen (1989), and briefly later in Sub-section 4.5.4, which regards maximum likelihood estimators.

We consider below a continuation for Example 4.1.

Example 4.1 (*Continuation I*)

Assuming the model discussed in Example 4.1, the $T \times T$ matrix $\Sigma(\underline{\theta})$ is represented as

$$\Sigma(\underline{\theta}) = \sigma^2 \cdot \begin{bmatrix} 1 & & & \\ \frac{\sigma_u^2 + \gamma\,\sigma_v^2}{\sigma^2} & 1 & & \\ \vdots & \vdots & \ddots & \\ \frac{\sigma_u^2 + \gamma^{(T-1)}\sigma_v^2}{\sigma^2} & \frac{\sigma_u^2 + \gamma^{(T-2)}\sigma_v^2}{\sigma^2} & \cdots & 1 \end{bmatrix} =$$

$$= \begin{pmatrix} \sigma_u^2 + \sigma_v^2 & & & \\ \sigma_u^2 + \gamma\sigma_v^2 & \sigma_u^2 + \sigma_v^2 & & \\ \vdots & \vdots & \ddots & \\ \sigma_u^2 + \gamma^{(T-1)}\sigma_v^2 & \sigma_u^2 + \gamma^{(T-2)}\sigma_v^2 & \cdots & \sigma_u^2 + \sigma_v^2 \end{pmatrix},$$

and

$$\hat{\Sigma} = \Sigma(\hat{\underline{\theta}}) = \begin{pmatrix} \hat{\sigma}_u^2 + \hat{\sigma}_v^2 & & & \\ \hat{\sigma}_u^2 + \hat{\gamma}\,\hat{\sigma}_v^2 & \hat{\sigma}_u^2 + \hat{\sigma}_v^2 & & \\ \vdots & \vdots & \ddots & \\ \hat{\sigma}_u^2 + \hat{\gamma}^{(T-1)}\hat{\sigma}_v^2 & \hat{\sigma}_u^2 + \hat{\gamma}^{(T-2)}\hat{\sigma}_v^2 & \cdots & \hat{\sigma}_u^2 + \hat{\sigma}_v^2 \end{pmatrix}.$$

We may expand (4.42a), for simplicity for the case where $T = 3$, as

$$F(\underline{\theta})_{ULSC} = \frac{1}{2}\left[S_{w,11} - \sigma_u^2 - \sigma_v^2\right]^2 + \left[S_{w,21} - \sigma_u^2 - \gamma\sigma_v^2\right]^2 + \left[S_{w,31} - \sigma_u^2 - \gamma^2\sigma_v^2\right]^2 + \frac{1}{2}\left[S_{w,22} - \sigma_u^2 - \sigma_v^2\right]^2$$
$$+ \left[S_{w,32} - \sigma_u^2 - \gamma\sigma_v^2\right]^2 + \frac{1}{2}\left[S_{w,33} - \sigma_u^2 - \sigma_v^2\right]^2,$$

where $S_{w,11}$, $S_{w,21}$, $S_{w,31}$, $S_{w,22}$, $S_{w,32}$ and $S_{w,33}$ are elements of the symmetric $\left[S_w\right]_{tt'}$ as defined in Sub-section 4.3.1.

The partial derivatives are

$$\frac{\partial F(\underline{\theta})_{ULSC}}{\partial \gamma} = -2\sigma_v^2 \cdot \left[S_{w,21} - \sigma_u^2 - \gamma\sigma_v^2\right] - 4\gamma\sigma_v^2 \cdot \left[S_{w,31} - \sigma_u^2 - \gamma^2\sigma_v^2\right] - 2\sigma_v^2 \cdot \left[S_{w,32} - \sigma_u^2 - \gamma\sigma_v^2\right], \tag{4.43}$$

$$\frac{\partial F(\underline{\theta})_{ULSC}}{\partial \sigma_u^2} = -S_{w,11} + 9\sigma_u^2 + 3\sigma_v^2 - 2S_{w,21} + 4\gamma\sigma_v^2 - 2S_{w,31} + 2\lambda^2\sigma_v^2 - S_{w,22} - 2S_{w,32} - S_{w,33}, \tag{4.44}$$

and

$$\frac{\partial F(\underline{\theta})_{ULSC}}{\partial \sigma_v^2} = -S_{w,11} + 3\cdot\left(\sigma_u^2 + \sigma_v^2\right) - 2\gamma \cdot \left[S_{w,21} - \sigma_u^2 - \gamma\sigma_v^2\right] - 2\gamma^2 \cdot \left[-2\gamma^2 \cdot S_{w,21} - \sigma_u^2 - \gamma^2\sigma_v^2\right]$$
$$- S_{w,22} - 2\gamma \cdot \left[S_{w,32} - \sigma_u^2 - \gamma\sigma_v^2\right] - S_{w,33}. \tag{4.45}$$

Setting (4.43), (4.44) and (4.45) to zero and solving this system of three equations and three parameters γ, σ_u^2 and σ_v^2, we find that the solution is respectively

$$\hat{\sigma}^2_{v,ULSC} = \frac{1}{12} \cdot \frac{\left[-3S_{w,21} + 2S_{w,11} + 2S_{w,22} - 3S_{w,32} + 2S_{w,33}\right]^2}{-3S_{w,21} + 3S_{w,31} - 3S_{w,32} + S_{w,11} + S_{w,22} + S_{w,33}},$$

$$\hat{\sigma}^2_{u,ULSC} = \frac{1}{4} \cdot \left\{ \frac{-3\left[S_{w,32}\right]^2 - 3\left[S_{w,21}\right]^2 - 6\left[S_{w,21}S_{w,32}\right] + 4S_{w,31} \cdot \left[S_{w,11} + S_{w,22} + S_{w,33}\right]}{-3S_{w,21} + 3S_{w,31} - 3S_{w,32} + S_{w,11} + S_{w,22} + S_{w,33}} \right\},$$

and

$$\hat{\gamma}_{ULSC} = 3 \cdot \frac{S_{w,21} - 2S_{w,31} + S_{w,32}}{-3S_{w,21} + 2S_{w,11} + 2S_{w,22} - 3S_{w,32} + 2S_{w,33}} .$$

Explicit solutions for σ_u^2 and σ_v^2 when assuming the model discussed in Chapter 2, Example 2.1 with $T = 5$ are presented in Appendix C. ■

When we can not determine a final solution analytically, iterative numerical techniques[25] would be required. Loehlin (1987; see also, for example, Jöreskog and Goldberger, 1972; Swain, 1975; McDonald, 1980; and Bentler and Weeks, 1980) suggests various procedures, as steepest descent, Fletcher-Powell, Newton-Raphson, and the EM algorithm as good alternatives. These techniques generally adopt search procedures across all possible solutions for the parameters values. See Jöreskog and Goldberger (1972), Pourahmadi (2000), and Pan and Mackenzie (2003), for example, for further information on Newton-Raphson type methods.

4.4.2 Generalised least squares (GLS) under the classical approach

In essence, the ULS method attaches equal weight to all elements of the resultant *residual covariance matrix* E. That does not make any allowance for different variances and covariances of different elements. The GLS estimation procedure[26] was proposed by Jöreskog and Goldberger (1972)[27] and weights observations to allow for unequal variances or nonzero covariances of the residuals. This method is thus a generalisation of the ULS estimation method (Bentler and Weeks, 1980).

Let us define the following quadratic form or objective function (see Browne, 1977),

$$F(\underline{\theta})_{GLS} = \{vech[\mathrm{S}] - vech[\Sigma(\underline{\theta})]\}' \mathrm{U}^{-1} \{vech[\mathrm{S}] - vech[\Sigma(\underline{\theta})]\} , \qquad (4.46)$$

where U is a $k \times k$ positive definite weight matrix defined in (4.24), with k as defined in Section 4.1, and $vech[\mathrm{S}]$ and $vech[\Sigma(\theta)]$ are as defined in Sub-section 4.3.2.

We assume that $\underline{\theta}$ is identified and that $\Sigma(\underline{\theta})$ is positive definite. Expression (4.46) represents a GLS distance function, under the classical set up. For further information, see Knight (1978), Dahm, Melton and Fuller (1983), Shapiro (1986), Shapiro and Browne (1987); Fuller (1987, Chapter 4), Skinner (1989a), and Yuan and Bentler (1997a).

We consider a GLS approach in which U is defined in terms of the covariance matrix of $vech[\mathrm{E}]$. See Sub-section 4.3.2, expression (4.24). According to Browne (1977), U could

[25] Most of the softwares that could be adopted for performing numerical procedures usually find a minimum rather than a maximum.
[26] This method is also known as minimum chi-square analysis (see Muthén and Satorra, 1995).
[27] Jöreskog and Goldberger (1972) adopt the well known Aitkens's generalised least squares principle.

alternatively be represented as (see also, for example, Satorra and Bentler, 1989; Kano, Berkane and Bentler, 1990, Section 2; Muthén and Satorra, 1995, Section 3.2; and Bentler and Dudgeon, 1996)

$$U = 2 \cdot K'(W \otimes W)K , \qquad (4.46b)$$

where K is a $k \times T^2$ transition (or elimination) matrix, and $\otimes$ is the operator for the right Kronecker product (see McDonald, 1980). Matrix K is denoted 'transition' because (see, for example, Le, 1990; and Chou, Bentler and Satorra, 1991)

$$K' \cdot vec(S) = vech(S),$$

where $vec(S)$ is a $T^2 \times 1$ column vector obtained by stacking the columns of S. In (4.46), we would then have (Shapiro, 1986; and Browne, 1987)

$$U^{-1} = \frac{1}{2} \cdot K^*(W^{-1} \otimes W^{-1})K^{*\prime},$$

with

$$K^* = (K'K)^{-1}K'.$$

By considering (4.46b), Browne (1977) shows that (4.46) then reduces to (see also Jöreskog and Goldberger, 1972; and Swain, 1975)

$$F(\underline{\theta})^{1}_{GLS} = \left(\frac{1}{2}\right) \cdot tr\left\{[(S - \Sigma(\underline{\theta}))W^{-1}]^2\right\}, \qquad (4.47)$$

where the $T \times T$ matrix W is any consistent estimator of Σ as defined in Sub-section 4.3.2, such as S when considering the classic set up. Note that expression (4.47) is computed more easily than (4.46). If W^{-1} is an identity matrix, then $F(\underline{\theta})_{GLS}$ reduces to $F(\underline{\theta})_{ULS}$.

The estimator $\hat{\underline{\theta}}_{GLS}$, which minimises $F(\underline{\theta})_{GLS}$ is referred to as a GLS estimator of $\underline{\theta}$. Note that $\hat{\underline{\theta}}_{GLS}$ has some important properties. It is: (*i*) a consistent estimator of $\underline{\theta}$, i.e.

$$p\lim_{n \to \infty} \hat{\underline{\theta}}_{GLS} = \underline{\theta};$$

(*ii*) asymptotically distributed as a multivariate normal with a known asymptotic covariance matrix[28]; and it is (*iii*) scale invariant and scale free[29] (Jöreskog and Goldberger, 1972). Browne (1982, Section 1.2) and Swain (1975), for example, provide some further information respectively on invariance under changes of scale and scale freeness. Fuller

[28] It allows for tests of statistical significance.

[29] The fitting function $F[S, \Sigma(\underline{\theta})]$ is scale invariant if $F[S, \Sigma(\underline{\theta})] = F[DSD, D\Sigma(\underline{\theta})D]$, where **D** is a diagonal, non-singular matrix with positive values on the diagonal. If the main diagonal of **D** contains the inverses of the standard deviations of the observed variables, the term **DSD** may be substituted for **S**, and $D\Sigma(\underline{\theta})D$ may be substituted for $\Sigma(\underline{\theta})$. An estimator is scale free when an equivalency between the structural parameters and estimates in a model with the original variables, and those in a model with linearly transformed variables, is maintained (Bollen, 1989, p. 109).

(1987, Sub-section 4.2.1) also provides an extensive discussion on the characteristics of the least squares approach methods.

Under the classical approach, the estimator $\hat{\underline{\theta}}_{GLS}$ obtained by the minimisation of (4.46) has minimum asymptotic variance if $\overline{\mathrm{U}} = \overline{\mathrm{C}}$ (see Browne, 1984, *Proposition* 3), where $\overline{\mathrm{U}}$ and $\overline{\mathrm{C}}$ are as defined in Sub-section 4.3.2. For further information, see also Bollen (1989, p. 426).

Moreover, the efficiency of $\hat{\underline{\theta}}_{GLS}$ depends on which matrix W has been chosen. Then, how to select W is an important issue on GLS estimation. Bollen (1989) discusses some assumptions, which lead to the choice of a suitable weighting matrix W. The key assumption is that the asymptotic distribution of the elements of matrix S is multivariate normal with means $\mathrm{COV}(Y_{it}, Y_{it'})$, and that the asymptotic covariances of $\mathrm{cov}(Y_{it}, Y_{it'})$ and $\mathrm{cov}(Y_{it''}, Y_{it'''})$ are equal to (4.19). See Sub-section 4.3.2 and Bollen (1989, p. 427). A requirement for this second assumption to hold is that the data follow the presuppositions of the classical set up, and that the fourth-order moments of Y_i exist.

When W is so adopted, $\hat{\underline{\theta}}_{GLS}$ has the properties (*i*) to (*iii*) listed above and it is also asymptotically efficient, under the assumptions of the classical set up (Jöreskog and Goldberger, 1972; Bentler and Weeks, 1980; Swain, 1975). But even though several choices of W^{-1} are consistent estimators of Σ^{-1}, a natural choice is (Jöreskog and Goldberger, 1972; Swain, 1975; Amemiya and Anderson, 1990; Yuan and Bentler, 1998)

$$\mathrm{W}^{-1} = \mathrm{S}^{-1},$$

which makes (4.47) equal to

$$F(\underline{\theta})^2_{GLS} = \left(\frac{1}{2}\right) \cdot tr\left\{[(\mathrm{S} - \Sigma(\underline{\theta}))\mathrm{S}^{-1}]^2\right\} = \left(\frac{1}{2}\right) \cdot tr\left\{[I - \Sigma(\underline{\theta})\mathrm{S}^{-1}]^2\right\}, \qquad (4.48a)$$

which is minimised when $\hat{\underline{\theta}}_{GLS}$ is chosen by differentiating $F(\underline{\theta})^2_{GLS}$ with respect to each of the unknown parameters (Jöreskog and Goldberger, 1972). Then, another alternative of representing the GLS estimator is as the solution of estimating equations. Let the scalar θ_{j} denotes any of the b elements included in the $b \times 1$ parameter vector $\underline{\theta}$, with $\mathrm{j} = 1, \ldots, b$. We may obtain $\hat{\underline{\theta}}^2_{GLS}$ by solving

$$\frac{F(\underline{\theta})^2_{GLS}}{\partial \theta_{\mathrm{j}}} = tr\left\{\mathrm{S}^{-1} \cdot [\Sigma(\underline{\theta}) - \mathrm{S}] \cdot \mathrm{S}^{-1} \cdot \frac{\partial \Sigma(\underline{\theta})}{\partial \theta_{\mathrm{j}}}\right\} = 0.$$

See Result D.1 in the Appendix D. Note the distinction in notation between j, introduced above, and the italicised *j* introduced in Chapter 2 for denoting the *j*th primary sampling unit.

In the GLS type fitting function, it appears to be advisable that elements of $[\mathrm{S} - \Sigma(\underline{\theta})]$ are weighted according to their variances and covariances with other elements. Another alternative would thus be (see Swain, 1975; Amemiya and Anderson, 1990)

$$\mathrm{W} = \Sigma(\underline{\theta}),$$

which makes (4.47) equal to

$$F(\underline{\theta})^{3}_{GLS} = \left(\frac{1}{2}\right) \cdot tr\left\{\left[(\mathrm{S} - \Sigma(\underline{\theta}))\Sigma(\underline{\theta})^{-1}\right]^{2}\right\} = \left(\frac{1}{2}\right) \cdot tr\left\{\left[\mathrm{S}\Sigma(\underline{\theta})^{-1} - I\right]^{2}\right\}. \qquad (4.48b)$$

Although $\hat{\underline{\theta}}_{GLS}$ has several appropriate characteristics, it has also some limitations (Bollen, 1989), as: (*i*) the general asymptotic standard errors and chi-square tests of significance could not be precise if the assumption described above are not fulfilled (or specially when the distribution of the marginal variables has very thin tails); (*ii*) the asymptotic covariance matrix for $\hat{\underline{\theta}}_{GLS}$ may be very complicated; and (*iii*) its properties are 'just' asymptotic. Moreover, according to Bollen (1989): "*very little is known about the small sample behaviour of* $\hat{\underline{\theta}}_{GLS}$ *but it appears that it has bias toward zero in small samples*".

4.4.3 Generalised least squares (GLS) under the complex survey approach

It is our aim in this sub-section to propose an adaptation for the GLS approach, allowing its application for complex survey data.

We suppose that S_w is a consistent estimator of the $T \times T$ population variance-covariance matrix $\Sigma(\underline{\theta})$. Recall that $vech[\mathrm{S}_w]$ and $vech[\Sigma(\underline{\theta})]$ are $k \times 1$ vectors formed from the nonduplicated elements of S_w and $\Sigma(\underline{\theta})$, respectively. We may consider a discrepancy function with the following quadratic form or objective function,

$$F(\theta)_{GLSC} = \{vech[\mathrm{S}_w] - vech[\Sigma(\theta)]\}' \mathrm{U}^{-1}\{vech[\mathrm{S}_w] - vech[\Sigma(\theta)]\}\ , \qquad (4.49)$$

where the subscript *GLSC* expresses that $F(\underline{\theta})$ could be considered as an alternative GLS fitting function for complex survey data, and U is a $k \times k$ positive definite weight matrix, as defined previously in the current chapter.

One choice of fitting function analogous to (4.47) would be

$$F(\underline{\theta})^{1}_{GLSC} = \left(\frac{1}{2}\right) \cdot tr\left\{\left[(\mathrm{S}_w - \Sigma(\underline{\theta}))\mathrm{W}^{-1}\right]^{2}\right\}, \qquad (4.50)$$

where W is any consistent estimator of Σ, as discussed previously in the current chapter, such as S_w when considering the complex survey data set up.

Thus, a natural strategy that could be adopted, if we consider that the data set was collected from a complex sampling scheme, is replace W in (4.50) by S_w,

$$F(\underline{\theta})^2_{GLSC} = \left(\frac{1}{2}\right)\cdot tr\left\{\left[(S_w - \Sigma(\underline{\theta}))S_w^{-1}\right]^2\right\} = \left(\frac{1}{2}\right)\cdot tr\left\{\left[I - \Sigma(\underline{\theta})S_w^{-1}\right]^2\right\}, \qquad (4.51)$$

which is minimised when $\hat{\underline{\theta}}_{GLSC}$ is chosen by differentiating $F(\underline{\theta})^2_{GLSC}$ with respect to each of the unknown parameters. Then, another alternative of representing the GLSC estimator is as the solution of estimating equations. We may then obtain $\hat{\underline{\theta}}^2_{GLSC}$ by solving

$$\frac{F(\underline{\theta})^2_{GLSC}}{\partial\theta_j} = tr\left\{S_w^{-1}\cdot\left[\Sigma(\underline{\theta}) - S_w\right]\cdot S_w^{-1}\cdot\frac{\partial\Sigma(\underline{\theta})}{\partial\theta_j}\right\} = 0, \qquad (4.51a)$$

where θ_j is as defined in Sub-section 4.4.2. See Result D.1 in Appendix D. Furthermore, note that (4.51) may be alternatively expressed as (see Fuller, 1987, p. 334)

$$F(\underline{\theta})^2_{GLSC} = \frac{1}{2}(n-1)\sum_{t=1}^{T}(\lambda_t - 1)^2, \qquad (4.51b)$$

were $\lambda_1,\ldots,\lambda_t$ are eigenvalues of $S_w^{-1/2}\Sigma(\underline{\theta})S_w^{-1/2}$.

Note that in order to reduce (4.49) to (4.50), and then to (4.51), we have to assume that $n\cdot S_w$ is distributed as a multivariate Wishart distribution (see Section 4.3, Sub-section 4.3.2), when defining matrix U. We shall suppose here that if the sampling weights are not excessively variable, this assumption might still provide a good approximation. We shall thus treat matrix U with W replaced by S_w only as a working estimator of C_c, where C_c is as defined in Section 4.3, Sub-section 4.3.2.

From Sub-section 4.4.2, an additional choice of fitting function that could lead to consistent estimates of $\underline{\theta}$ is

$$F(\underline{\theta})^3_{GLSC} = \left(\frac{1}{2}\right)\cdot tr\left\{\left[S_w\Sigma(\underline{\theta})^{-1} - I\right]^2\right\}. \qquad (4.52)$$

Note that, according to Fuller (1987, Sub-section 4.2.1, Example 4.2.1), deviations from normality in $\underline{Y}_i$ may have large effects on the distribution of the estimators for variance components, included in the parameter vector $\underline{\theta}$, that rely upon normality assumptions.

Moreover, matrix U in (4.49) could be misspecified if $\overline{U} \neq \overline{C}$, where $\overline{U}$ and $\overline{C}$ are as defined in Sub-section 4.3.2. Under this circumstance, the GLS estimator would not be necessarily asymptotic efficient. This may happen in practice if C depends upon the complex survey design but U does not. In that situation we could is substitute C by C_c, where C_c is also as defined in Sub-section 4.3.2.

Recall that we have proposed earlier in Sub-section 4.3.2 an adaptation of ADF methods, originally proposed by Browne (1982) and Browne (1984), for the analysis of complex survey data. Under the assumptions of iid observations with no weighting, for example, an ADF type estimator for $\underline{\theta}$ is acquired by minimizing (4.49), with respect to $\underline{\theta}$, and substituting U by U^*, with $u^*_{tt',t''t'''}$ given by (4.26).

As stated previously, an estimator $\hat{\underline{\theta}}_{GLS}$ with minimum asymptotic variance could be achieved by defining the matrix U as a consistent estimator of C_c, introduced earlier in Section 4.3, Subsection 4.3.2, say $\hat{C}_c$, so that

$$F(\underline{\theta})^4_{GLSC} = \{vech[S_w] - vech[\Sigma(\underline{\theta})]\}' \hat{C}_c^{-1} \{vech[S_w] - vech[\Sigma(\underline{\theta})]\}, \qquad (4.53)$$

where

$$\hat{C}_c = n \cdot \mathrm{var}\{vech[S_w]\}_{v,l}, \qquad (4.53b)$$

with the $k \times k$ positive definite matrix $\mathrm{var}\{vech[S_w]\}_{v,l}$ given by (4.39b). The estimator given in (4.53) could thus be considered an ADF method, and thus the use of $\mathrm{var}[vech(S_w)]$ as a weight matrix should yield asymptotic optimality (in the sense of leading to efficient estimators) for any distribution of $\underline{Y}_i$, where $\underline{Y}_i$ is as defined in Section 4.2.

Another alternative could be to calculate $\mathrm{var}\{vech[S_w]\}_{v,l}$ by adopting (4.40). In this case, the resultant ADF estimator would allow for unequal sampling weights, but it would not fully consider complex sampling designs.

However, ADF methods have to be used very carefully according, for example, to Bollen (1989, p. 432), Satorra (1992), Muthén and Satorra (1995), Satorra (2000), and Satorra and Bentler (2001), mainly because fourth order moments are large in number and could be unstable in small samples. The larger is T the more important last statement is. Note that the sample size necessary for convergence may be larger for this method than for the other methods considered in the present chapter. This type of method could consequently result in computational overload and lack of robustness in samples of 'small' and 'moderate' sizes.

We shall evaluate the ADF fitting function $F(\underline{\theta})^4_{GLSC}$, and compare its performance with other methods, through an extensive simulation study, which results are presented in the next chapter. We shall also assess the behaviour of the proposed methods when the sample size is reduced.

4.4.4 Maximum likelihood (ML)

We initially show in the present sub-section how to obtain the ML fitting function directly from the multivariate normal probability distribution function (as in Jöreskog, 1970; and Bollen, 1989, Appendix 4A). See Chapter 2, Section 2.2, and Section 4.2 earlier in the present chapter for additional information on ML estimation. Under the classical set up, if we take the distribution of Y_i to be of the multivariate normal form, the log likelihood function with respect to β for given $\underline{\theta}$ is

$$L[\underline{\beta},\underline{\theta}] = \log \ell[\underline{\beta},\underline{\theta}] = -\frac{1}{2}\left[n_o \log 2\pi + n\log|\Sigma(\underline{\theta})| + \sum_{i=1}^{n}[y_i - \mu_i(\underline{\beta})]'\Sigma(\underline{\theta})^{-1}[y_i - \mu_i(\underline{\beta})]\right]. \tag{4.54}$$

One alternative is to estimate the parameter vector $\underline{\theta}$ simultaneously with $\underline{\beta}$ via maximisation of (4.54), i.e. by minimising

$$\begin{aligned} & n\cdot\log|\Sigma(\underline{\theta})| + \sum_{i=1}^{n}[y_i - \mu_i(\underline{\beta})]'\Sigma(\underline{\theta})^{-1}[y_i - \mu_i(\underline{\beta})] = \\ & = tr\left[\Sigma(\underline{\theta})^{-1}\sum_{i=1}^{n}(y_i - \mu_i)(y_i - \mu_i)'\right] + n\cdot\log|\Sigma(\underline{\theta})| = \\ & = n\cdot tr\left[\Sigma(\underline{\theta})^{-1}n^{-1}\sum_{i=1}^{n}(y_i - \mu_i)(y_i - \mu_i)'\right] + n\cdot\log|\Sigma(\underline{\theta})|. \end{aligned} \tag{4.55}$$

We may substitute μ_i for its estimator $\hat{\mu}_i$, defined by (4.16) with constant weights, in (4.55) so that it equals

$$n\cdot\left\{tr\left[S\Sigma(\underline{\theta})^{-1}\right] + \log|\Sigma(\underline{\theta})|\right\} \tag{4.56}$$

where

$$S = n^{-1}\sum_{i=1}^{n}(y_i - \hat{\mu}_i)(y_i - \hat{\mu}_i)'. \tag{4.57}$$

Matrices S and $\Sigma(\underline{\theta})$ are assumed to be positive definite, and $|\Sigma(\underline{\theta})|$ denote the determinant of $\Sigma(\underline{\theta})$. We may define from (4.56) the ML fitting function, as (see Jöreskog, 1970; Jöreskog and Goldberger, 1972; Wiley, Schmidt and Bramble, 1973; and Jöreskog and Goldberger, 1975)

$$F(\underline{\theta})_{ML} = \log|\Sigma(\underline{\theta})| + tr\left[S\Sigma(\underline{\theta})^{-1}\right] - \log|S| - T,$$

or equivalently (4.58)

$$F(\underline{\theta})_{ML} = tr\left[S\Sigma(\underline{\theta})^{-1}\right] - \log\left|S\Sigma(\underline{\theta})^{-1}\right| - T,$$

which may be considered as a distance measure between S and $\Sigma(\underline{\theta})$ (Browne, 1984), and has to be minimised in order to estimate the parameter vector $\underline{\theta}$. Note that the term *n*, present in (4.56), is not present in (4.58) as it does not affect the estimation of $\underline{\theta}$.

The term $-\log|S| - T$ is included in (4.58) because it is desirable to make F_{ML} an appropriate fitting function (see, for example, Loehlin, 1987), by allowing it to equal zero when $\hat{\Sigma} = S$. This term does not depend upon $\underline{\beta}$ and $\underline{\theta}$, hence does not change the values of $\underline{\beta}$ and $\underline{\theta}$ minimising this function.

The included term in (4.58) guarantees that $F_{ML} = 0$ when $\Sigma(\hat{\underline{\theta}}) = S$, because (*i*) S and $\Sigma(\underline{\theta})$ would have the same determinant, and (*ii*) $S\Sigma(\underline{\theta})^{-1} = I$, with the sum of the diagonal elements equal to *T*, as the identity matrix *I* is a $T \times T$ matrix.

Bollen (1989, Appendix 4B) and Amemiya and Anderson (1990), for example, provide an alternative derivation for (4.58), which is based upon the fact that the sample covariance matrix S has a Wishart distribution, as it was stated earlier (see Sub-section 4.3.2, in the current chapter).

The fitting function F_{ML} is minimised when $\hat{\underline{\theta}}_{ML}$ is chosen by differentiating $F(\underline{\theta})_{ML}$ with respect to each of the unknown parameters (Wiley, Schmidt and Bramble, 1973). Then, another alternative of representing the ML estimator is as the solution of estimating equations. We may obtain $\hat{\underline{\theta}}_{ML}$ by solving

$$\frac{\partial F(\underline{\theta})_{ML}}{\partial \theta_j} = tr\left\{\Sigma(\underline{\theta})^{-1} \cdot [\Sigma(\underline{\theta}) - S] \cdot \Sigma(\underline{\theta})^{-1} \cdot \frac{\partial \Sigma(\underline{\theta})}{\partial \theta_j}\right\} = 0,$$

where θ_j is as defined in Sub-section 4.4.2. See matrix differential calculus results in Appendix D (Result D.2).

Once $\hat{\underline{\theta}}_{ML}$ is estimated, a sufficient condition for it to minimise $F(\underline{\theta})_{ML}$ is that the following matrix (Bollen, 1989),

$$\Omega_{ML} = \begin{bmatrix} \frac{\partial^2 F(\underline{\theta})_{ML}}{\partial(\hat{\gamma})^2} & 0 & 0 \\ 0 & \frac{\partial^2 F(\underline{\theta})_{ML}}{\partial(\hat{\sigma}_v^2)^2} & 0 \\ 0 & 0 & \frac{\partial F(\underline{\theta})_{ML}}{\partial(\hat{\sigma}_u^2)^2} \end{bmatrix},$$

is positive definite, again when $\underline{\theta} = (\sigma_u^2, \sigma_v^2, \gamma)'$, for example. Ω_{ML} is a matrix with the second partial derivatives of $F(\underline{\theta})_{ML}$, with respect to $\underline{\theta}$, on the diagonal. Recall from Sub-section 4.4.1 that Ω should also be regular in order to guarantee the identification of $\underline{\theta}$.

Two different strategies could be adopted when minimising $F(\underline{\theta})_{ML}$. A first option could be (*i*) to estimate $\underline{\beta}$ separately by adopting (2.15) from Chapter 2, Section 2.2, (*ii*) plug $\hat{\underline{\beta}}$ in (4.57) and (4.58), and then (*iii*) minimise (4.58) with respect only to $\underline{\theta}$. As a second alternative, we could minimise (4.58) simultaneously with respect to $\underline{\beta}$ and $\underline{\theta}$ as stated earlier in the current sub-section. We shall present, in the next chapter, simulation results produced by considering the first alternative estimation procedure.

ML estimators have many considerable and desirable asymptotic properties. They are (*i*) asymptotically unbiased, (*ii*) consistent, (*iii*) asymptotic efficient, and (*iv*) asymptotically normally distributed (see, for example, Bollen, 1989, for further information on $\hat{\underline{\theta}}_{ML}$ characteristics; Jöreskog and Goldberger, 1972; Swain, 1975; Knight, 1978; and Pourahmadi, 2000). Beyond that, tests of statistical significance are possible for $\hat{\underline{\theta}}_{ML}$, given the characteristics of its asymptotic covariance matrix.

Note that, according to Anderson (1973), under normality and under the classical set up, $\hat{\underline{\theta}}_{GLS}$ given by (4.48a) converges to the maximum likelihood estimator, and consequently it has the same asymptotic properties for linear covariance structures. Furthermore, Fuller (1987, Sub-section 4.2.2) performs an impeccable study on the relationships between least squares and maximum likelihood methods in the context of covariance structure models. See also Jöreskog and Goldberger (1972), Swain (1973), Browne (1977), Bentler and Weeks (1980), Satorra (1992), Bentler and Dudgeon (1996), and Ogasawara (2005) for further information.

According to Browne (1987), $F(\underline{\theta})_{ML}$ is the most often adopted fitting function. Other important characteristics are: (*v*) $F(\underline{\theta})_{ML}$ is generally scale invariant and scale free (Swain, 1975), and (*vi*) the ML estimator yields a χ^2 test for the model goodness of fit (see, for example, Wiley, Schmidt, and Bramble, 1973; Swain, 1975; Bentler and Weeks, 1980; Browne, 1984; Fuller, 1987, Theorem 4.2.1; and Shapiro, 1986).

See Pourahmadi (1999), Pourahmadi (2000), Pan and Mackenzie (2003) for more recent studies on ML estimation, in the context of modelling mean-covariance structures with balanced longitudinal data.

4.4.5 Pseudo maximum likelihood (PML)

The estimation methods for $\underline{\theta}$, discussed in the previous sub-section, were not originally designed to handle complex survey data. In this sub-section we shall adapt the ML fitting function, allowing for complex sampling design, using the PML approach earlier discussed for $\underline{\beta}$ in Section 4.2 (where we assumed that $\underline{\theta}$ was known) and also in Chapter 2.

The log census likelihood with respect to $\underline{\beta}$ for given $\underline{\theta}$ is given in Chapter 2, Section 2.4. The parameter vector $\underline{\theta}$ could be estimated simultaneously with $\underline{\beta}$ via maximisation of the pseudo log likelihood, i.e. by minimising

$$N \cdot \log|\Sigma(\underline{\theta})| + \sum_{i=1}^{n} w_i [y_i - \mu_i(\underline{\beta})]' \Sigma(\underline{\theta})^{-1} [y_i - \mu_i(\underline{\beta})],$$

which is the weighted estimate of the $\log \ell_N[\underline{\beta},\underline{\theta}]$, ignoring constants. Alternatively, since N may be unknown, we may minimise

$$\begin{aligned}
&\sum_{i=1}^{n} w_i \cdot \log|\Sigma(\underline{\theta})| + tr\left[\sum_{i=1}^{n} w_i [y_i - \mu_i(\underline{\beta})][y_i - \mu_i(\underline{\beta})]' \Sigma(\underline{\theta})^{-1}\right] = \\
&\sum_{i=1}^{n} w_i \cdot \log|\Sigma(\underline{\theta})| + \sum_{i=1}^{n} w_i \cdot tr\left[\left(\sum_{i=1}^{n} w_i\right)^{-1} \cdot \sum_{i=1}^{n} w_i [y_i - \mu_i(\underline{\beta})][y_i - \mu_i(\underline{\beta})]' \Sigma(\underline{\theta})^{-1}\right] = \qquad (4.59) \\
&\sum_{i=1}^{n} w_i \cdot \left\{\log|\Sigma(\underline{\theta})| + tr\left[\left(\sum_{i=1}^{n} w_i\right)^{-1} \cdot \sum_{i=1}^{n} w_i [y_i - \mu_i(\underline{\beta})][y_i - \mu_i(\underline{\beta})]' \Sigma(\underline{\theta})^{-1}\right]\right\}.
\end{aligned}$$

We may substitute μ_i by its estimator $\hat{\mu}_i$, as defined in (4.16), in (4.59) so that it equals

$$\sum_{i=1}^{n} w_i \cdot \left\{tr\left[S_w \Sigma(\underline{\theta})^{-1}\right] + \log|\Sigma(\underline{\theta})|\right\} \qquad (4.60)$$

where

$$S_w = \left(\sum_{i=1}^{n} w_i\right)^{-1} \sum_{i=1}^{n} w_i (y_i - \hat{\mu}_i)(y_i - \hat{\mu}_i)'. \qquad (4.61)$$

Note that the minimisation of the $\log \ell_N[\underline{\beta}, \underline{\theta}]$ is unaffected by scale multiplication of w_i. In (4.61), S_w is the weighted sample covariance matrix (see Sub-section 4.3.1).

In a similar way to the F_{ML} case in (4.58), we add the following term,

$$-\log|S_w| - T,$$

into the braces in (4.60).

Then the pseudo maximum likelihood fitting function, $F(\underline{\theta})_{PML}$ could be defined as

$$F(\underline{\theta})_{PML} = tr\left[S_w \Sigma(\underline{\theta})^{-1}\right] + \log|\Sigma(\underline{\theta})| - \log|S_w| - T,$$

or equivalently (4.62)

$$F(\underline{\theta})_{PML} = tr\left[S_w \Sigma(\underline{\theta})^{-1}\right] - \log\left|S_w \Sigma(\underline{\theta})^{-1}\right| - T.$$

The fitting function $F(\underline{\theta})_{PML}$ is minimised when $\hat{\underline{\theta}}_{PML}$ is chosen by differentiating $F(\underline{\theta})_{PML}$ with respect to each of the unknown parameters. We may obtain $\hat{\underline{\theta}}_{PML}$ by solving

$$\frac{\partial F(\underline{\theta})_{PML}}{\partial \theta_j} = tr\left\{\Sigma(\underline{\theta})^{-1} \cdot \left[\Sigma(\underline{\theta}) - S_w\right] \cdot \Sigma(\underline{\theta})^{-1} \cdot \frac{\partial \Sigma(\underline{\theta})}{\partial \theta_j}\right\} = 0, \quad (4.62a)$$

where θ_j is as defined in Sub-section 4.4.2. See matrix differential calculus results in Appendix D (Result D.2). Note that $F(\underline{\theta})_{PML}$ reduces to $F(\underline{\theta})_{ML}$ when the weights w_i are constant.

By comparison with (4.51a), the PML estimator given by (4.62a) is found to be asymptotically equivalent to the GLSC2 estimator, if the model holds, i.e. if S_w converges to $\Sigma(\underline{\theta})$. Moreover, following Fuller (1987, p. 334), (4.62) may alternatively be expressed as

$$F(\underline{\theta})_{PML} = \sum_{t=1}^{T} (\log \lambda_t + \lambda_t^{-1}), \quad (4.62b)$$

were $\lambda_1, \ldots, \lambda_t$ is as defined earlier for (4.51b), in Section 4.4, Sub-section 4.4.3. Thus if the model holds, both GLSC2 and PML estimators are obtained by minimizing (see Fuller, 1987, p. 335)

$$\sum_{t=1}^{T} (\lambda_t - 1)^2,$$

therefore confirming that the GLSC2 and PML estimators may be considered asymptotic equivalent.

As in the ML case, the included terms in (4.62) guarantee that $F(\underline{\theta})_{PML} = 0$ when $\Sigma(\hat{\underline{\theta}}) = S_w$, because (*i*) S_w and $\Sigma(\underline{\theta})$ would have the same determinant, and (*ii*) $S_w \Sigma(\underline{\theta})^{-1} = I$, with the sum of the diagonal elements equal to *T*, as the identity matrix *I* is a $T \times T$ matrix.

Note that, similarly to the case of the ML method, we shall (*i*) to estimate $\underline{\beta}$ separately by adopting (2.27) from Chapter 2, Section 2.4, then (*ii*) plug $\hat{\underline{\beta}}$ in (4.61) and (4.62), and then (*iii*) minimise (4.62) with respect only to $\underline{\theta}$, instead of minimising (4.62) simultaneously with respect to $\underline{\beta}$ and $\underline{\theta}$.

In cases where the solution for (4.62a) is not found analytically, iterative numerical techniques would be required. Iterative numerical routines, as for example Nelder and Mead (1965) simplex algorithm or Newton-Raphson iteration, could be adopted in this situation.

4.5 Discussion

In this chapter we have proposed some inference procedures for covariance structure models, in the context of longitudinal complex survey data. We have paid particular attention to estimation methods for the parameter vector of interest $\underline{\theta}$, allowing for complex surveys. We have initially extended estimation methods for model parameters, previously discussed in Chapter 2, considering in the present chapter that the variance-covariance matrices $\Sigma(\underline{\theta})$ are constrained to be functions of $\underline{\theta}$.

We have also discussed methods for making inference about the covariance matrix Σ and proposed a new method for calculating $\hat{\Sigma}$. Further developments when considering the complex survey approach have also accomplished here, mainly by adopting Taylor expansion techniques in order to extend asymptotically distribution-free (ADF) methods for estimating the variance of $\hat{\Sigma}$.

We have produced in this chapter a review on classic estimation methods for the parameter vector $\underline{\theta}$, including unweighted least squares, generalised least squares under the classical approach, and maximum likelihood. Furthermore we may consider another achievement presented in this chapter the proposition of new estimation methods, including generalised least squares type of estimators, under the complex survey approach, and also pseudo maximum likelihood.

Throughout the present chapter and previous chapters we have assumed that the full random sample of individuals is available for the observed variables at each wave of the survey. Bollen (1989, Chapter 8, p. 369-376) gives an introductory overview of the problem of missing values in the covariance structure modelling context, including alternative estimators of covariance matrix, explicit estimation methods, and systematic missing values.

We shall present in Chapter 5 the characteristics and results of a simulation study, which shall have as main objective to evaluate the statistical properties of the point estimation procedures discussed and proposed in the current chapter.

Chapter 5

Simulation study I

5.1 Introduction

The use of simulation techniques in statistics has its origins in the beginning of the 20th century (see Morgan, 1984). Lewis and Orav (1989) define simulation as a controlled statistical procedure (experiment) based on repeated sampling carried out on a computer. We present in this chapter the characteristics and results of a simulation study, which has the main objective of evaluating the statistical properties of the point estimation procedures proposed in Chapter 4.

Hence, we shall consider two unweighted least squares (ULS) type of fitting functions, the classic $F(\underline{\theta})_{ULS}$, and the modified $F(\underline{\theta})_{ULSC}$, where $\underline{\theta}$ is our parameter vector of interest (as defined in Chapter 4, Section 4.1). For additional information on ULS and unweighted least squares for complex survey data (ULSC) methods, see Chapter 4, Sub-section 4.4.1.

We shall also evaluate three proposed generalised least squares fitting functions for complex survey data (GLSC), $F(\underline{\theta})^2_{GLSC}$, $F(\underline{\theta})^3_{GLSC}$, and $F(\underline{\theta})^4_{GLSC}$. See Chapter 4, Sub-section 4.4.3, for further information on GLSC methods. Moreover, three different choices of fitting functions shall be evaluated when considering the generalised least squares method (GLS), $F(\underline{\theta})^2_{GLS}$, $F(\underline{\theta})^3_{GLS}$, and $F(\underline{\theta})^4_{GLS}$, where $F(\underline{\theta})^4_{GLS}$ is a special case of $F(\underline{\theta})^4_{GLSC}$ when considering that the sampling weights are constant. For additional information on the GLS approach see Chapter 4, Sub-section 4.4.2.

We shall additionally consider maximum likelihood type methods (ML) for estimating $\underline{\theta}$. See Chapter 4, Sub-section 4.4.4 for additional information on the ML estimation method. Furthermore, we shall evaluate the proposed pseudo maximum likelihood (PML) fitting function F_{PML}, which could be adopted when analysing complex survey data. See Chapter 4, Sub-section 4.4.5 for additional information on the PML estimation method.

Thus, in the current chapter we also aim to compare the properties of the proposed methods with the traditional statistical techniques also described in Chapter 4. We shall present here detailed information on how the simulation study is implemented, which is given in Section 5.2. Moreover, results and some further discussion are presented in sections 5.3 and 5.4, respectively. Concluding remarks shall be included in Section 5.5.

5.2 Simulation procedures

A simulation experiment is by definition a method that one could use for obtaining approximate answers for probabilistic problems. After programming the point estimation procedures developed and described in Chapter 4, we shall apply them and evaluate their statistical properties in this simulation study in terms of both bias and variance.

This simulation study shall involve simulating $d = 1,\dots,D$ replicate samples. Note the distinction in notation between **D**, which is a diagonal, non-singular matrix with positive values down its diagonal introduced earlier in Chapter 4, Sub-section 4.4.2, and D which is the number of simulated repeated datasets.

Each replicate sample considered in this study is based upon a BHPS data subset, which was adopted earlier in Chapter 3, with size 1340 subjects. Initially we fix the sample size of each replicate sample as $n^{sim} = 1340$ (results included in Sub-section 5.3.2), the same as is the BHPS subset. The values of the x variable are fixed and the values of Y_{it} are simulated from different models, independently for each replicate. The superscript *sim* is added to denote *simulation*.

To assess the effect of sample size we also repeat this exercise but with reduced sample sizes, $n^{sim} = 500$ (results presented in Sub-section 5.3.3), $n^{sim} = 200$ (results in Sub-section 5.3.4), and $n^{sim} = 100$ (results in Sub-section 5.4.5). These reduced samples are obtained by a simple random sampling without replacement scheme. We undertook two approaches to sampling: (*i*) select a sample and hold the same values for x in every replicate but varying Y; and (*ii*) select a different sample, with different values for x, for every replicate. We found little difference in the results and we shall report here for results for (*ii*).

We shall initially adopt a uniform correlation model (UCM),

$$Y_{it} = \underline{x}_{it}\underline{\beta} + u_i + v_{it}, \tag{5.1}$$

which was introduced on Example 2.1, Chapter 2, for generating sample values of Y_{it} that we are going to use in this simulation study, with

$$u_i \sim N\left(0, \sigma_u^2\right),$$

and

$$v_{it} \sim N\left(0, \sigma_v^2\right).$$

In (5.1), Y_{it} is the value (scalar) for the response variable for unit $i = 1,2,\dots,n^{sim}$ at wave t of the survey, $\underline{x}_{it}$ is a $1 \times q$ vector with the q fixed covariates, $\underline{\beta}$ is the $q \times 1$ vector of the unknown fixed coefficients for the x variables, u_i are the permanent random effects, and v_{it} are transitory random effects.

In order to represent the possible effects of complex sampling in our simulation, we shall also consider the following alternative UCM type of model, which allows the impact of clustering,

$$Y_{ijt} = \underline{x}_{ijt}\underline{\beta} + \eta_j + u_{ij} + v_{ijt}, \tag{5.2}$$

shall also be considered in this simulation study, with

$$\eta_j \sim N(0, \sigma_\eta^2),$$

$$u_{ij} \sim N(0, \sigma_u^2),$$

and

$$v_{ijt} \sim N(0, \sigma_v^2),$$

where Y_{ijt} is the value (scalar) for the study variable for unit $i = 1, \ldots, n^{sim}$, in cluster $j = 1, \ldots, m^{sim}$, at wave t of the survey, η_j are random cluster effects, with m^{sim} denoting the number of clusters in each simulation sample, and $\underline{\beta}$, u_i and v_{it} are as defined for (5.1).

We shall refer to the model described in (5.1) as UCM in the remaining of this chapter, and to the model described in (5.2) as UCM-C, where C denotes *cluster*.

For simplicity, we shall not attempt to allow for the impact of either stratification or unequal probability sampling in this simulation study, although we shall consider the properties of some weighted estimators where the weights are those taken for the BHPS for the 1340 women. This will enable us to study the impact of weighting in circumstances when it is not needed for bias correction.

In order to obtain the parameters that are necessary for producing the simulating replicates we shall initially fit the models described above in expressions (5.1) and (5.2). The simulation parameter $\underline{\beta}$ is a $q \times 1$ vector, where q is the number of fixed covariates, and is estimated by (see expression (2.15), in Chapter 2, Section 2.2)

$$\underline{\beta}^{sim}(\mathrm{V}) = \left(\sum_{i=1}^{1340} \mathrm{X}_i' \mathrm{V}^{-1} \mathrm{X}_i \right)^{-1} \sum_{i=1}^{1340} \mathrm{X}_i' \mathrm{V}^{-1} \underline{y}_i, \tag{5.3}$$

where $\underline{y}_i$ are $T \times 1$ vectors with the observed values for the response variable for each individual i included in the considered BHPS subset, X_i are $T \times q$ matrices with the values for the q explanatory variables for each individual i in each wave, and V is a $T \times T$ exchangeable covariance matrix.

In order to consider realistic values for simulation parameters, we shall adopt the gender role attitude score (see Chapter 3, Section 3.2, Sub-section 3.2.3) as the dependent variable and (*i*) wave number (time covariate), (*ii*) age group (with four categories), (*iii*) economic activity (with five categories), and (*iv*) educational level (with five categories), as explanatory variables, we obtain for the UCM model (see Tables 3.1 and 3.11, in Chapter 3)

$$\underline{\beta} = \underline{\hat{\beta}}^{sim} = \begin{bmatrix} 22.18 \\ -0.04 \\ -0.70 \\ -0.87 \\ -1.00 \\ -0.91 \\ -0.73 \\ 0.17 \\ -2.05 \\ -0.51 \\ -0.60 \\ -0.43 \\ -1.17 \end{bmatrix}. \qquad (5.4)$$

Then, for simulation purposes we shall also assume that

$$\begin{aligned} \sigma_u^2 &= \sigma_u^{2\ sim} \cong 6.744, \\ \sigma_v^2 &= \sigma_v^{2\ sim} \cong 4.965, \\ \rho &= \rho^{sim} \cong 0.576. \end{aligned} \qquad (5.5)$$

Note that $\sigma_u^{2\ sim}$, $\sigma_v^{2\ sim}$ and ρ^{sim} are calculated via Swamy-Arora method, which is described by Stata Corp. (2003), for estimating variance components. Note that, according to Boomsma (1985), a choice of values for $\sigma_u^{2\ sim}$ and $\sigma_v^{2\ sim}$ close to zero could lead for an increase in the occurrence of improper solutions and inaccuracies of the point estimates. This issue shall not be further investigated in this study.

When fitting a UCM-C model, we obtain

$$\underline{\beta} = \underline{\beta}^{sim,C} = \begin{bmatrix} 22.16 \\ -0.04 \\ -0.70 \\ -0.87 \\ -1.00 \\ -0.91 \\ -0.73 \\ 0.17 \\ -2.05 \\ -0.51 \\ -0.60 \\ -0.42 \\ -1.17 \end{bmatrix}, \tag{5.6}$$

and for simulation purposes we shall also assume that

$$\begin{aligned} \sigma_\eta^2 &= \sigma_\eta^{2\ sim,C} \cong 0.088 \\ \sigma_u^2 &= \sigma_u^{2\ sim,C} \cong 7.135, \\ \sigma_v^2 &= \sigma_v^{2\ sim,C} \cong 4.981, \\ \rho &= \rho^{sim,C} \cong 0.589. \end{aligned} \tag{5.7}$$

The variance components $\sigma_a^{2\ sim,C}$, $\sigma_u^{2\ sim,C}$ and $\sigma_v^{2\ sim,C}$, and $\rho^{sim,C}$ are calculated by applying a GLS type method, which is described by Goldstein (1995) and Rasbash *et al.* (2002).

5.2.1 Simulation under multivariate normality

We initially simulate each replicate d from

$$\underline{y}_i^{(d)} = \mathrm{N}_T\left[\mathrm{X}_i\underline{\beta}, \begin{pmatrix} \sigma^2 & \rho\sigma^2 & \cdots & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \cdots & \rho\sigma^2 \\ \vdots & \rho\sigma^2 & \ddots & \vdots \\ \rho\sigma^2 & \rho\sigma^2 & \cdots & \rho\sigma^2 \end{pmatrix}\right], \text{ with } i = 1,\ldots,n^{sim},\ d = 1,\ldots,D, \tag{5.8}$$

and fix X_i for each d as the values from the BHPS data, where $\underline{y}_i^{(d)}$ are $T \times 1$ vectors with the T-dimensional normally distributed simulated values for the response variable for each individual i in each wave for replication d, and X_i are as defined for (5.3). We shall adopt for simulation purposes the values for $\underline{\beta}$, σ_u^2, and σ_v^2, which are presented respectively in (5.4) and (5.5), when considering a UCM model.

Thus, the simulation shall involve generating

$$z_i^{(d)} \sim N(0,1) \text{ setting } u_i^{(d)} = \sigma_u \cdot z_i^{(d)}, \tag{5.9}$$

with σ_u given in (5.5), generating

$$r_{it}^{(d)} \sim N(0,1) \text{ setting } v_{it}^{(d)} = \sigma_v \cdot r_{it}^{(d)}, \tag{5.10}$$

with σ_v given in (5.5), and finally generating

$$y_{it}^{(d)} = \underline{x}_{it}\underline{\beta} + u_i^{(d)} + v_{it}^{(d)}, \tag{5.11}$$

with $\underline{\beta}$ given by (5.4).

Additionally, when considering a UCM-C model, we shall adopt estimated values included in (5.6) and (5.7) respectively for $\underline{\beta}$, and σ_η^2, σ_u^2 and σ_v^2. Thus,

$$y_{ijt}^{(d)} = \underline{x}_{ijt}\underline{\beta} + \eta_j^{(d)} + u_{ij}^{(d)} + v_{ijt}^{(d)},$$

where

$$\eta_j^{(d)} = \sigma_\eta \cdot c_j^{(d)},$$

$$u_{ij}^{(d)} = \sigma_u \cdot z_{ij}^{(d)},$$

and

$$v_{ijt}^{(d)} = \sigma_v \cdot r_{ijt}^{(d)} \quad ,$$

with

$$c_j^{(d)} \sim N(0,1),$$

$$z_i^{(d)} \sim N(0,1),$$

and

$$r_{ijt}^{(d)} \sim N(0,1).$$

5.2.2 Simulation under multivariate Student's t-distribution

Statistical procedures usually depend somehow upon the data distribution. We thus also aim in this simulation study to evaluate how the methods described in Chapter 4, including the ones we propose, behave under departures from the normality distributional assumptions. Hence we shall alternatively consider,

$$\underline{y}_i^{t(d)} = \mathrm{t}_{T,\upsilon=5}\left[\mathrm{X}_i\beta, \begin{pmatrix} \sigma^2 & \rho\sigma^2 & \cdots & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \cdots & \rho\sigma^2 \\ \vdots & \rho\sigma^2 & \ddots & \vdots \\ \rho\sigma^2 & \rho\sigma^2 & \cdots & \rho\sigma^2 \end{pmatrix}\right], \text{with } i = 1,\ldots,n^{sim},\ d = 1,\ldots,D,$$

where $\underline{y}_i^{t(d)}$ are $T \times 1$ vectors with the T-dimensional t distributed simulated values for the response variable for each individual i in each wave for replication d, and X_i, σ^2, and ρ are as defined for (5.8). Note the distinction in notation between t, which is so far adopted in this thesis for denoting time, and t_υ which is adopted as usual for the t distribution with υ degrees of freedom.

We also consider here for simulation purposes the values for σ_u, σ_v and $\underline{\beta}$, which are presented respectively in (5.4) and (5.5), when considering a UCM model. Thus, the simulation involve initially generating

$$z_i^{t^*(d)} \sim t_{\upsilon=5}$$

and

$$r_{it}^{t^*(d)} \sim t_{\upsilon=5},$$

where the superscript t* denote that z_i and r_{it} follow a t distribution, with non-standardised variance. Both $z_i^{t^*(d)}$ and $r_{it}^{t^*(d)}$ may be standardised respectively by

$$z_i^{t(d)} = z_i^{t^*(d)} \Big/ \sqrt{\mathrm{var}(t_{\upsilon=5})} \qquad (5.12)$$

and

$$r_{it}^{t(d)} = r_{it}^{t^*(d)} \Big/ \sqrt{\mathrm{var}(t_{\upsilon=5})}. \qquad (5.13)$$

In (5.12) and (5.13), $\mathrm{var}(t_{\upsilon=5})$ may be obtained from (Johnson, Kotz and Balakrishnan, 1995)

$$m_p(t_\upsilon) = \upsilon^{p/2} \cdot \frac{\Gamma\left[\frac{1}{2}(p+1)\right]\Gamma\left[\frac{1}{2}(\upsilon-p)\right]}{\Gamma\left[\frac{1}{2}\right]\Gamma\left[\frac{1}{2}\upsilon\right]},$$

if p is even, where $m_p(t_\upsilon)$ is the pth central moment of a t_υ distribution. For $p = 2$,

$$m_2(t_\upsilon) = \mathrm{var}(t_\upsilon) = \frac{\upsilon}{\upsilon - 2},$$

and with $\upsilon = 5$,

$$\mathrm{var}(t_{\upsilon=5}) \cong 1.666667.$$

Thus,

$$z_i^{t(d)} \sim t_{\upsilon=5}(0,1),$$

and

$$r_{it}^{t(d)} \sim t_{\upsilon=5}(0,1).$$

We may then set

$$u_i^{t(d)} = \sigma_u \cdot z_i^{t(d)},$$

and

$$v_{it}^{t(d)} = \sigma_v \cdot r_{it}^{t(d)},$$

and finally generate

$$y_{it}^{t(d)} = \underline{x}_{it}\underline{\beta} + u_i^{t(d)} + v_{it}^{t(d)},$$

with $\underline{\beta}$ given by (5.4) and (5.6).

Furthermore, when considering a UCM-C model, we consider values included in (5.6) and (5.7) for σ_η^2, σ_u, σ_v and $\underline{\beta}$. And thus,

$$y_{ijt}^{t(d)} = \underline{x}_{ijt}\underline{\beta} + \eta_j^{t(d)} + u_{ij}^{t(d)} + v_{ijt}^{t(d)},$$

where

$$\eta_j^{t(d)} = \sigma_\eta \cdot c_j^{t(d)},$$

$$u_{ij}^{t(d)} = \sigma_u \cdot z_{ij}^{t(d)},$$

and

$$v_{ijt}^{t(d)} = \sigma_v \cdot r_{ijt}^{t(d)},$$

with

$$c_j^{t(d)} \sim t_{\upsilon=5}(0,1)$$

$$z_{ij}^{t(d)} \sim t_{\upsilon=5}(0,1),$$

and

$$r_{ijt}^{t(d)} \sim t_{\upsilon=5}(0,1).$$

5.3 Results

Let $\underline{\theta} = (\sigma_u^2, \sigma_v^2, \gamma)'$ be our $b \times 1$ parameter vector of interest, with $b = 3$. Recall that the estimators of $\underline{\theta}$ for which simulation results are produced in the current chapter, were defined in Chapter 4:

(*i*) see expression (4.42a) for $\underline{\hat{\theta}}_{ULS}$;

(*ii*) see expression (4.48a) for $\underline{\hat{\theta}}_{GLS}^2$;

(*iii*) see expression (4.48b) for $\underline{\hat{\theta}}_{GLS}^3$;

(*iv*) see expression (4.46), with matrix U being substituted by U^* given by (4.26), (4.28), and (4.31), for $\underline{\hat{\theta}}_{GLS}^4$;

(*v*) see expression (4.58) for $\underline{\hat{\theta}}_{ML}$;

(*vi*) see expression (4.42b) for $\underline{\hat{\theta}}_{ULSC}$;

(*vii*) see expression (4.51) for $\underline{\hat{\theta}}_{GLSC}^2$;

(*viii*) see expression (4.52) for $\underline{\hat{\theta}}_{GLSC}^3$;

(*ix*) see expressions (4.53) and (4.53b), with $\mathrm{var}\{vech[\mathrm{S}_w]\}_{v,l}$ calculated by adopting (4.40), for $\hat{\underline{\theta}}^{4}_{GLSC}$; and

(*x*) see expression (4.62) for $\hat{\underline{\theta}}_{PML}$.

Note that $\hat{\underline{\theta}}_{ULSC}$, $\hat{\underline{\theta}}^{2}_{GLSC}$, $\hat{\underline{\theta}}^{3}_{GLSC}$, $\hat{\underline{\theta}}^{4}_{GLSC}$, and $\hat{\underline{\theta}}_{PML}$ are weighted point estimators, which were proposed in Chapter 4 and allow for complex survey data.

We shall adopt the mean square error as a criterion for evaluating an estimator. Moreover (Cochran, 1977),

$$\mathrm{MSE}(\hat{\theta}_{\mathrm{j}}) = \mathrm{VAR}(\hat{\theta}_{\mathrm{j}}) + [\mathrm{BIAS}(\hat{\theta}_{\mathrm{j}})]^2, \qquad (5.14)$$

where θ_{j} denotes any of the b elements included in the parameter vector $\underline{\theta}$, $\mathrm{VAR}(\hat{\theta}_{\mathrm{j}})$ is the true variance of $\hat{\theta}_{\mathrm{j}}$, $\mathrm{BIAS}(\hat{\theta}_{\mathrm{j}})$ denote the true bias of an estimator and may be estimated by

$$\mathrm{bias}(\hat{\theta}_{\mathrm{j}}) = \hat{E}(\hat{\theta}_{\mathrm{j}}) - \theta_{\mathrm{j}}. \qquad (5.15)$$

In (5.15), $\hat{E}(\hat{\theta}_{\mathrm{j}})$ is the estimated expected value of $\hat{\theta}_{\mathrm{j}}$. Information on how $\mathrm{VAR}(\hat{\theta}_{\mathrm{j}})$ is estimated and how $\hat{E}(\hat{\theta}_{\mathrm{j}})$ is calculated shall be provided later on in this section. Let $\mathrm{rel\ bias}(\hat{\theta}_{\mathrm{j}})$ denote the estimated relative bias of $\hat{\theta}_{\mathrm{j}}$, so that

$$\mathrm{rel\ bias}(\hat{\theta}_{\mathrm{j}}) = \frac{\mathrm{bias}(\hat{\theta}_{\mathrm{j}})}{\theta_{\mathrm{j}}} \cdot 100.$$

For each replicate we shall estimate $\hat{\theta}^{(d)}_{\mathrm{j}}$, via each of the estimation methods listed above. We then calculate

$$\hat{E}(\hat{\theta}_{\mathrm{j}}) = \frac{1}{D}\sum_{d=1}^{D}\hat{\theta}^{(d)}_{\mathrm{j}},$$

where $\hat{E}(\hat{\theta}_{\mathrm{j}})$ is as introduced above, and is the mean of our parameter of interest estimated over repeated simulation of the datasets.

Moreover,

$$\mathrm{var}(\hat{\theta}_{\mathrm{j}}) = \frac{1}{D-1}\sum_{d=1}^{D}\left[\hat{\theta}^{(d)}_{\mathrm{j}} - \hat{E}(\hat{\theta}_{\mathrm{j}})\right]^2,$$

where $\mathrm{var}(\hat{\theta}_{\mathrm{j}})$ is a simulation estimator of $\mathrm{VAR}(\hat{\theta}_{\mathrm{j}})$ in (5.1).

The $MSE(\hat{\theta}_{\mathrm{j}})$ may thus be estimated by

$$\mathrm{mse}(\hat{\theta}_{\mathrm{j}}) = \mathrm{var}(\hat{\theta}_{\mathrm{j}}) + \left[\mathrm{bias}(\hat{\theta}_{\mathrm{j}})\right]^2,$$

the simulation standard error of $\hat{E}(\hat{\theta}_{\mathrm{j}})$ by

$$se[\hat{E}(\hat{\theta}_j)] = \sqrt{\mathrm{var}(\hat{\theta}_j)} \Big/ \sqrt{D},$$

and an approximately 95% simulation confidence interval for θ_j may be defined as

$$\hat{E}(\hat{\theta}_j) \pm 1.96 \cdot se[\hat{E}(\hat{\theta}_j)]. \tag{5.16}$$

Furthermore, the coefficient of variation of $\hat{\theta}_j$ shall be calculated by

$$cv[\hat{\theta}_j] = \frac{\sqrt{\mathrm{var}(\hat{\theta}_j)}}{\hat{E}(\hat{\theta}_j)} \cdot 100.$$

We shall initially fit an UCM model and assume that the $T \times T$ population covariance matrix $\Sigma(\underline{\theta})$ has the following structure, (see Sub-section 4.1, Chapter 4)

$$\Sigma(\underline{\theta}) = \begin{pmatrix} \sigma_u^2 + \sigma_v^2 & & & \\ \sigma_u^2 & \sigma_u^2 + \sigma_v^2 & & \\ \vdots & \vdots & \ddots & \\ \sigma_u^2 & \sigma_u^2 & \cdots & \sigma_u^2 + \sigma_v^2 \end{pmatrix}.$$

Moreover, in order to evaluate how the methods behave in the context of a more complex model, we shall also fit, for each *d*, a transitory random effects as a first-order autoregressive process model (AR1), which is described in Chapter 4, Example 4.1. In this case, it is assumed that $\Sigma(\underline{\theta})$ is structured as

$$\Sigma(\underline{\theta}) = \begin{pmatrix} \sigma_u^2 + \sigma_v^2 & & & \\ \sigma_u^2 + \gamma^{|s-t|}\sigma_v^2 & \sigma_u^2 + \sigma_v^2 & & \\ \vdots & \vdots & \ddots & \\ \sigma_u^2 + \gamma^{|s-t|}\sigma_v^2 & \sigma_u^2 + \gamma^{|s-t|}\sigma_v^2 & \cdots & \sigma_u^2 + \sigma_v^2 \end{pmatrix}.$$

It should be observed that each evaluated method shall be analysing exactly the same data for each situation considered in this simulation study.

Note that Sub-sections from 5.3.2 to 5.3.5 shall only present summarised simulation results. More complete tables are confined to Appendix E, for simulation circumstances where $n^{sim} = 1340$ is considered. Detailed results for the remaining considered sample sizes have also been produced, although they shall not be included in the thesis.

5.3.1 Software and minimisation procedures

In this simulation study, we shall adopt the statistical computer software *R* (Ihaka and Gentleman, 1996) both for generating the replicated data for the simulation study and for programming techniques that have been studied and developed in Chapter 4. *R* is available through the Internet under the General Public License.

This software has the following desired characteristics (see Dalgaard, 2002): (*i*) it is a complete programming language (what makes it very flexible), which is very similar to the S language; (*ii*) it allows for statistical analysis and graphics; and (*iii*) it allows for further computations on the results of a statistical procedure.

The estimation methods that are being evaluated in this chapter are all based on the minimisation of fitting functions *F* (see Chapter 4). See Appendix C for explicit solutions for parameters when adopting ULSC (and ULS) and PML (and ML) estimation methods and fitting a UCM model. Note that according to analytic results reported in Appendix C, ULSC and PML are found to be equivalent when considering a UCM model. This may represent an interesting finding as ULS type methods are considered to be computationally cheap when compared to other estimation approaches, while ML methods have many considerable and desirable asymptotic properties, which were comprehensively described in Chapter 4, Section 4.4, Sub-section 4.4.4.

For the remaining estimation methods and for all situations involving AR1 model fitting, we shall perform the necessary minimisations by adopting an iterative numerical method. We shall utilise a Newton type algorithm for carrying out numerical derivatives of *F*, as similarly suggested for example by Pourahmadi (1999). See also Dennis and Schnabel (1983, chapters 5 and 6), Schnabel, Koontz and Weiss (1985), and Bollen (1989, Appendix 4C). That method is readily available in the function **nlm** included in the software *R* (R Development Core Team, 2003). Note that results presented in this chapter shall be produced considering only replicates for which the numerical method achieved convergence for every considered point estimation method. Information on non-convergence shall be reported for each method and considered simulation situation (as in Boomsma, 1985), whenever it occurs.

One of the reasons for the adoption of a numerical method for minimising the fitting functions *F* is that, according to Bollen (1989) and Long (1993), ULS, GLS and ML fitting functions usually come out with equations that are typically nonlinear in the parameters. Therefore, in general explicit solutions for the parameters are often not achievable.

The selection of initial values may affect the numerical minimisation procedures in various ways. That could influence the number of iterations necessary, for example. Furthermore a choice of starting values that are far from the final ones may augment the probability of finding a local minimum rather than the global minima, or even of not finding a convergent solution. For further discussion on this issues see Boomsma (1985). See also Bollen (1989, Appendix 4C), where a very brief consideration on this subject is provided. For the reasons explained above, we have adopted as starting values for $\underline{\hat{\theta}}$, the values given in expression (5.5), which are referred to by Boomsma (1985) as ideal starting values.

5.3.2 Simulation results for samples of size $n^{sim} = 1340$

The results presented in the current sub-section are produced by using replications of the whole simulation population, with $n^{sim} = 1340$ (see Sub-section 5.2.1), $D = 1000$ (as, for example, adopted in Satorra, 1992; and Muthén and Satorra, 1995), and generated by making use of both UCM and UCM-C models.

Table 5.1 includes results that were produced when fitting both UCM and AR1 models and considering normality conditions, with simulated Y_{it} values generated by a UCM model.

Note that in all tables included in the current chapter, and in those presented in Appendix E, no results for the component γ shall be presented when a UCM model is fitted as we have $\underline{\theta} = (\sigma_u^2, \sigma_v^2)'$ in this situation. Moreover, in situations where a AR1 model is fitted, relative bias and coefficient of variation shall not be reported (NR) for the component γ as simulated Y_{it} values, as stated above, are generated by a UCM model, and UCM are special cases of the AR1 models when $\gamma = 0$.

Additionally, all bias results that were found not to be significantly different from zero at the 95% level shall be indicated with "NS", when the interval given by (5.16) includes the true value of the parameter of interest. That shall indicate situations where the simulations do not provide enough evidence that the estimator is biased.

parameter		UCM model rel bias	UCM model $cv(\hat{\theta})$	UCM model mse $(\times 1000)$	AR1 model rel bias	AR1 model $cv(\hat{\theta})$	AR1 model mse $(\times 1000)$
$\underline{\hat{\theta}}_{ULS}$	$\hat{\sigma}_u^2$	-0.88%	4.26%	84.707	-0.89%	4.31%	86.497
	$\hat{\sigma}_v^2$	0.06%[NS]	1.87%	8.655	0.08%[NS]	2.05%	10.432
	$\hat{\gamma}$	-	-	-	NR [NS]	NR	0.537
$\underline{\hat{\theta}}^2_{GLS}$	$\hat{\sigma}_u^2$	-1.42%	4.29%	90.683	-1.45%	4.32%	92.071
	$\hat{\sigma}_v^2$	-0.74%	1.90%	10.162	-0.64%	2.02%	10.942
	$\hat{\gamma}$	-	-	-	NR [NS]	NR	0.355
$\underline{\hat{\theta}}^3_{GLS}$	$\hat{\sigma}_u^2$	-0.60%	4.27%	83.481	-0.61%	4.29%	84.500
	$\hat{\sigma}_v^2$	0.47%	1.88%	9.297	0.45%	2.00%	10.415
	$\hat{\gamma}$	-	-	-	NR [NS]	NR	0.348
$\underline{\hat{\theta}}^4_{GLS}$	$\hat{\sigma}_u^2$	-1.46%	4.33%	92.427	-1.48%	4.35%	93.569
	$\hat{\sigma}_v^2$	-0.74%	1.93%	10.365	-0.65%	2.04%	11.162
	$\hat{\gamma}$	-	-	-	NR [NS]	NR	0.358
$\underline{\hat{\theta}}_{ML}$	$\hat{\sigma}_u^2$	-0.88%	4.26%	84.707	-0.89%	4.29%	85.822
	$\hat{\sigma}_v^2$	0.06%[NS]	1.87%	8.655	0.08%[NS]	1.99%	9.820
	$\hat{\gamma}$	-	-	-	NR [NS]	NR	0.348
$\underline{\hat{\theta}}_{ULSC}$	$\hat{\sigma}_u^2$	-0.34%	4.57%	95.062	-0.33%	4.61%	96.495
	$\hat{\sigma}_v^2$	-0.27%	2.01%	10.131	-0.29%	2.20%	12.048
	$\hat{\gamma}$	-	-	-	NR [NS]	NR	0.593
$\underline{\hat{\theta}}^2_{GLSC}$	$\hat{\sigma}_u^2$	-0.96%	4.61%	98.968	-0.95%	4.64%	99.971
	$\hat{\sigma}_v^2$	-1.18%	2.06%	13.616	-1.13%	2.18%	14.590
	$\hat{\gamma}$	-	-	-	NR [NS]	NR	0.394
$\underline{\hat{\theta}}^3_{GLSC}$	$\hat{\sigma}_u^2$	-0.03%[NS]	4.58%	95.402	0.003%[NS]	4.61%	96.600
	$\hat{\sigma}_v^2$	0.18%	2.02%	10.170	0.10%[NS]	2.14%	11.379
	$\hat{\gamma}$	-	-	-	NR [NS]	NR	0.390
$\underline{\hat{\theta}}^4_{GLSC}$	$\hat{\sigma}_u^2$	-1.06%	4.65%	101.470	-1.05%	4.67%	102.199
	$\hat{\sigma}_v^2$	-1.29%	2.09%	14.572	-1.22%	2.21%	15.466
	$\hat{\gamma}$	-	-	-	NR [NS]	NR	0.402
$\underline{\hat{\theta}}_{PML}$	$\hat{\sigma}_u^2$	-0.34%	4.57%	95.062	-0.35%	4.77%	103.246
	$\hat{\sigma}_v^2$	-0.27%	2.01%	10.131	-0.25%	2.92%	21.001
	$\hat{\gamma}$	-	-	-	NR [NS]	NR	1.391

[NS] – denotes 'absolute bias not significantly different from zero at 95% level'.
NR – denotes 'no reported'.

Table 5.1 – Evaluation of $\underline{\hat{\theta}}$ with normally distributed errors (population – replications generated by UCM model).

By looking across each row of Table 5.1 we can notice that, in general, there is not much difference between mse when comparing different point estimation methods. Those differences are noticeable, but not enormous. Moreover, we may observe that the impact of the relative bias is relatively small when compared to the coefficient of variation, i.e. the c.v. is the main component contributing to the mean square error of for all the estimators in the simulation situation considered above, under normality conditions and a sample of size 1340. This shall be true in most cases for the remaining situations considered in this sub-section.

When fitting a UCM model, ULS, ML and GLS3 are the methods with the lowest relative bias, although the differences to the remaining unweighted methods are not gigantic. In terms of variance, methods are broadly very similar, with ULS and ML presenting a marginally smaller c.v. than the other methods. When considering the weighted methods, GLSC3 is the one with the somewhat smallest relative bias, while all weighted methods behave almost identically in terms of variance.

Generally, very similar remarks may be made when considering the AR1 model fitting. However, we may notice that the PML method has a reasonably larger mse for estimating γ comparatively to the other weighed methods and also to the unweighted ones.

The estimator GLS4, which is the distribution free method (ADF), proposed in Chapter 4, had somewhat the largest bias among the unweighted methods, although in terms of coefficient of variation results that method was very similar to the other ones, when fitting both UCM and AR1 models. Very similar conclusions may be drawn for GLSC4 when comparing than to the remaining weighted methods.

Results above, and in all the following tables included in the current chapter, illustrate our finding that for situations where a UCM is considered, ULS and ML point estimators are equivalent (see Subsection 5.3.1 and Appendix C). Under these circumstances the ULS estimator could be considered as an alternative to GLS type methods which are computationally more expensive, as ML methods have several desirable asymptotic properties. Furthermore, we may observe that even for situations where an AR1 model is fitted ULS results are very similar to those calculated for the ML estimator. Note that our finding agrees with Bollen (1989, p. 112), which says that ML and ULS estimates are usually very close.

In terms of the asymptotic equivalence between GLS2 and ML methods (Anderson, 1973; and Fuller, 1987, Sub-section 4.2.2, for example), discussed also in Chapter 4, Section 4.4, we may observe that there is not a large difference between the mse results when comparing these two methods and also for GLSC2 and PML. That of course will become less clear in later sub-sections when simulations with smaller samples sizes shall be considered.

We may also observe that weighted methods presented larger variance than the unweighted methods especially because, considering the framework adopted in this simulation study, there would be no need for utilising weighted estimators for bias correction. This result is expected from the survey sampling literature (see, for example, Pfeffermann and LaVange, 1989; and Skinner, 2003a, and references therein). Recall that we do not allow for the impact of either stratification or unequal probability sampling when generating values of Y_{it}.

Furthermore, according to our results, weighted methods appear to have slightly lower bias when compared to the unweighted ones. This shall become clearer in later sub-sections when the sample size is reduced. Nevertheless, we shall not claim here that this is always true, as there is not any clear theoretical reason for this result. It is important also to notice that the impact of bias is relatively less important of that caused by the coefficient of variation, at least at this stage when we are considering samples of size 1340.

Table 5.2 includes results for both UCM and AR1 models considering normality conditions, with simulated Y_{it} values generated by an UCM-C model. Recall that the UCM-C, introduced earlier in Section 5.2, allows for the impact of clustering.

parameter		UCM model			AR1 model		
		rel bias	$cv(\hat{\theta})$	mse $(\times 1000)$	rel bias	$cv(\hat{\theta})$	mse $(\times 1000)$
$\underline{\hat{\theta}}_{ULS}$	$\hat{\sigma}_u^2$	-0.87%	4.37%	101.771	-0.88%	4.38%	102.345
	$\hat{\sigma}_v^2$	0.06%^{NS}	1.90%	8.935	0.09%^{NS}	2.11%	11.037
	$\hat{\gamma}$	-	-	-	NR ^{NS}	NR	0.522
$\underline{\hat{\theta}}^2_{GLS}$	$\hat{\sigma}_u^2$	-1.44%	4.40%	108.740	-1.47%	4.42%	110.245
	$\hat{\sigma}_v^2$	-0.78%	1.92%	10.478	-0.68%	2.09%	11.796
	$\hat{\gamma}$	-	-	-	NR ^{NS}	NR	0.347
$\underline{\hat{\theta}}^3_{GLS}$	$\hat{\sigma}_u^2$	-0.58%	4.38%	100.470	-0.59%	4.40%	101.475
	$\hat{\sigma}_v^2$	0.48%	1.91%	9.687	0.47%	2.09%	11.462
	$\hat{\gamma}$	-	-	-	NR ^{NS}	NR	0.344
$\underline{\hat{\theta}}^4_{GLS}$	$\hat{\sigma}_u^2$	-1.44%	4.39%	108.463	-1.47%	4.42%	110.053
	$\hat{\sigma}_v^2$	-0.78%	1.93%	10.647	-0.69%	2.10%	11.992
	$\hat{\gamma}$	-	-	-	NR ^{NS}	NR	0.352
$\underline{\hat{\theta}}_{ML}$	$\hat{\sigma}_u^2$	-0.87%	4.37%	101.771	-0.88%	4.39%	102.905
	$\hat{\sigma}_v^2$	0.06%^{NS}	1.90%	8.935	0.09%^{NS}	2.07%	10.711
	$\hat{\gamma}$	-	-	-	NR ^{NS}	NR	0.343
$\underline{\hat{\theta}}_{ULSC}$	$\hat{\sigma}_u^2$	-0.34%	4.69%	114.744	-0.33%	4.69%	114.705
	$\hat{\sigma}_v^2$	-0.22%	2.01%	10.136	-0.23%	2.23%	12.378
	$\hat{\gamma}$	-	-	-	NR ^{NS}	NR	0.603
$\underline{\hat{\theta}}^2_{GLSC}$	$\hat{\sigma}_u^2$	-1.00%	4.73%	119.557	-0.99%	4.76%	120.744
	$\hat{\sigma}_v^2$	-1.17%	2.04%	13.453	-1.11%	2.19%	14.741
	$\hat{\gamma}$	-	-	-	NR ^{NS}	NR	0.403
$\underline{\hat{\theta}}^3_{GLSC}$	$\hat{\sigma}_u^2$	-0.01%^{NS}	4.70%	115.135	0.01%^{NS}	4.72%	116.240
	$\hat{\sigma}_v^2$	0.25%	2.03%	10.413	0.18%	2.20%	12.127
	$\hat{\gamma}$	-	-	-	NR ^{NS}	NR	0.400
$\underline{\hat{\theta}}^4_{GLSC}$	$\hat{\sigma}_u^2$	-1.07%	4.73%	120.163	-1.06%	4.76%	121.693
	$\hat{\sigma}_v^2$	-1.25%	2.07%	14.255	-1.19%	2.22%	15.439
	$\hat{\gamma}$	-	-	-	NR ^{NS}	NR	0.411
$\underline{\hat{\theta}}_{PML}$	$\hat{\sigma}_u^2$	-0.34%	4.69%	114.744	-0.32%	4.72%	115.825
	$\hat{\sigma}_v^2$	-0.22%	2.01%	10.136	-0.25%	2.18%	11.909
	$\hat{\gamma}$	-	-	-	NR ^{NS}	NR	0.398

^{NS} – denotes 'absolute bias not significantly different from zero at 95% level'.
NR – denotes 'no reported'.

Table 5.2 – Evaluation of $\underline{\hat{\theta}}$ with normally distributed errors (population – replications generated by UCM-C model).

Overall similar conclusions may be drawn when comparing Table 5.2 (data generated by a UCM-C model) with Table 5.1 (data generated by a UCM model). Methods ULS, ML and GLS3 are the unweighted methods with the lowest relative bias. We may notice that the effects of clustering did not lead to an increase in the bias of the evaluated point estimators. However, as expected from the survey sampling literature (see, for example, Deming, 1950; Kish, 1957; Kish and Frankel, 1974; and Kish, 1980) we may notice that most methods have had an increase in their mse for both fitted models (UCM and AR1), mainly because of a modest inflation of variance.

Table 5.3 includes results for both UCM and AR1 models considering non-normality conditions (t distribution, as described in Section 5.2, Sub-section 5.2.2), with simulated Y_{it} values generated by an UCM model.

parameter		UCM model rel bias	UCM model $cv(\hat{\theta})$	UCM model mse $(\times 1000)$	AR1 model rel bias	AR1 model $cv(\hat{\theta})$	AR1 model mse $(\times 1000)$
$\hat{\underline{\theta}}_{ULS}$	$\hat{\sigma}_u^2$	-1.15%	7.10%	230.181	-1.14%	7.14%	232.672
	$\hat{\sigma}_v^2$	-0.10%^{NS}	3.33%	27.306	-0.11%^{NS}	3.45%	29.276
	$\hat{\gamma}$	-	-	-	NR^{NS}	NR	0.510
$\hat{\underline{\theta}}^{2}_{GLS}$	$\hat{\sigma}_u^2$	-1.71%	7.12%	235.758	-1.76%	7.14%	237.751
	$\hat{\sigma}_v^2$	-1.37%	3.18%	28.838	-1.25%	3.26%	29.443
	$\hat{\gamma}$	-	-	-	NR^{NS}	NR	0.360
$\hat{\underline{\theta}}^{3}_{GLS}$	$\hat{\sigma}_u^2$	-0.87%	7.11%	229.277	-0.90%	7.12%	230.385
	$\hat{\sigma}_v^2$	0.55%	3.54%	31.976	0.56%	3.66%	34.145
	$\hat{\gamma}$	-	-	-	NR^{NS}	NR	0.367
$\hat{\underline{\theta}}^{4}_{GLS}$ †	$\hat{\sigma}_u^2$	-2.43%	7.50%	270.097	-2.47%	7.50%	271.024
	$\hat{\sigma}_v^2$	-3.06%	3.16%	46.184	-2.86%	3.24%	44.592
	$\hat{\gamma}$	-	-	-	NR^{NS}	NR	0.367
$\hat{\underline{\theta}}_{ML}$	$\hat{\sigma}_u^2$	-1.15%	7.10%	230.181	-1.18%	7.12%	231.553
	$\hat{\sigma}_v^2$	-0.10%^{NS}	3.33%	27.306	-0.06%^{NS}	3.43%	28.984
	$\hat{\gamma}$	-	-	-	NR^{NS}	NR	0.358
$\hat{\underline{\theta}}_{ULSC}$	$\hat{\sigma}_u^2$	-0.67%	7.98%	288.059	-0.64%	8.02%	290.805
	$\hat{\sigma}_v^2$	-0.49%	3.52%	30.869	-0.53%	3.67%	33.537
	$\hat{\gamma}$	-	-	-	NR	NR	0.584
$\hat{\underline{\theta}}^{2}_{GLSC}$	$\hat{\sigma}_u^2$	-1.31%	7.99%	290.443	-1.32%	8.01%	292.149
	$\hat{\sigma}_v^2$	-1.91%	3.32%	35.139	-1.82%	3.43%	36.137
	$\hat{\gamma}$	-	-	-	NR^{NS}	NR	0.412
$\hat{\underline{\theta}}^{3}_{GLSC}$	$\hat{\sigma}_u^2$	-0.35%	8.00%	289.460	-0.34%	8.01%	289.996
	$\hat{\sigma}_v^2$	0.25%	3.90%	37.802	0.21%^{NS}	4.13%	42.373
	$\hat{\gamma}$	-	-	-	NR^{NS}	NR	0.430
$\hat{\underline{\theta}}^{4}_{GLSC}$ †	$\hat{\sigma}_u^2$	-2.25%	8.28%	320.966	-2.28%	8.22%	317.109
	$\hat{\sigma}_v^2$	-3.80%	3.35%	61.251	-3.63%	3.44%	59.502
	$\hat{\gamma}$	-	-	-	NR^{NS}	NR	0.407
$\hat{\underline{\theta}}_{PML}$	$\hat{\sigma}_u^2$	-0.67%	7.98%	288.059	-0.75%	8.54%	328.958
	$\hat{\sigma}_v^2$	-0.49%	3.52%	30.869	-0.46%	3.79%	35.675
	$\hat{\gamma}$	-	-	-	NR^{NS}	NR	1.411

^{NS} – denotes 'absolute bias not significantly different from zero at 95% level'.
NR – denotes 'no reported'.
† Numerical minimisations did not achieve convergence for 0.1% of the replicates.

Table 5.3 – Evaluation of $\hat{\underline{\theta}}$ with $t_{\upsilon=5}(0,1)$ distributed errors (population – replications generated by UCM model).

When comparing Table 5.3 (data generated by a UCM model with t distributed errors) with Table 5.2 (data generated by UCM model with normally distributed errors), we may notice that all methods have had a large increase in their mse for both fitted models (UCM and AR1), caused by a modest increase in bias and variance. Perhaps contrary to what we expected the ADF point estimators, GLS4 and GLSC4, have not performed better than the other methods, although those had generally the smallest coefficient of variation for estimating σ_v^2 and also small values of mean square error for estimating γ.

Table 5.4 includes results for both UCM and AR1 models considering non-normality conditions (t distribution), with simulated Y_{it} values generated now by an UCM-C model.

parameter		UCM model			AR1 model		
		rel bias	$cv(\hat{\theta})$	mse $(\times 1000)$	rel bias	$cv(\hat{\theta})$	mse $(\times 1000)$
$\underline{\hat{\theta}}_{ULS}$	$\hat{\sigma}_u^2$	-1.05%	7.53%	294.868	-1.10%	7.59%	300.327
	$\hat{\sigma}_v^2$	0.26%	3.47%	30.183	0.33%	3.58%	32.251
	$\hat{\gamma}$	-	-	-	NR ^{NS}	NR	0.544
$\underline{\hat{\theta}}^2_{GLS}$	$\hat{\sigma}_u^2$	-1.63%	7.54%	300.991	-1.67%	7.58%	304.406
	$\hat{\sigma}_v^2$	-1.02%	3.28%	28.792	-0.90%	3.34%	29.207
	$\hat{\gamma}$	-	-	-	NR ^{NS}	NR	0.385
$\underline{\hat{\theta}}^3_{GLS}$	$\hat{\sigma}_u^2$	-0.76%	7.53%	294.094	-0.78%	7.58%	297.780
	$\hat{\sigma}_v^2$	0.92%	3.66%	36.009	0.92%	3.72%	37.081
	$\hat{\gamma}$	-	-	-	NR ^{NS}	NR	0.378
$\underline{\hat{\theta}}^4_{GLS}$ †	$\hat{\sigma}_u^2$	-2.56%	7.44%	307.947	-2.61%	7.47%	311.410
	$\hat{\sigma}_v^2$	-2.75%	3.21%	42.941	-2.54%	3.28%	41.374
	$\hat{\gamma}$	-	-	-	NR ^{NS}	NR	0.367
$\underline{\hat{\theta}}_{ML}$	$\hat{\sigma}_u^2$	-1.05%	7.53%	294.868	-1.08%	7.57%	298.410
	$\hat{\sigma}_v^2$	0.26%	3.47%	30.183	0.30%	3.53%	31.332
	$\hat{\gamma}$	-	-	-	NR ^{NS}	NR	0.375
$\underline{\hat{\theta}}_{ULSC}$	$\hat{\sigma}_u^2$	-0.50%	8.03%	333.788	-0.50% ^{NS}	8.09%	338.859
	$\hat{\sigma}_v^2$	-0.02% ^{NS}	3.73%	34.592	-0.03% ^{NS}	3.82%	36.242
	$\hat{\gamma}$	-	-	-	NR ^{NS}	NR	0.594
$\underline{\hat{\theta}}^2_{GLSC}$	$\hat{\sigma}_u^2$	-1.17%	8.04%	336.526	-1.17%	8.08%	339.501
	$\hat{\sigma}_v^2$	-1.49%	3.44%	33.999	-1.42%	3.50%	34.589
	$\hat{\gamma}$	-	-	-	NR ^{NS}	NR	0.411
$\underline{\hat{\theta}}^3_{GLSC}$	$\hat{\sigma}_u^2$	-0.16% ^{NS}	8.03%	335.589	-0.13% ^{NS}	8.08%	339.568
	$\hat{\sigma}_v^2$	0.74%	4.07%	43.179	0.67%	4.14%	44.139
	$\hat{\gamma}$	-	-	-	NR ^{NS}	NR	0.414
$\underline{\hat{\theta}}^4_{GLSC}$ †	$\hat{\sigma}_u^2$	-2.27%	7.92%	339.247	-2.26%	7.95%	341.682
	$\hat{\sigma}_v^2$	-3.47%	3.38%	56.301	-3.32%	3.44%	54.795
	$\hat{\gamma}$	-	-	-	NR ^{NS}	NR	0.396
$\underline{\hat{\theta}}_{PML}$	$\hat{\sigma}_u^2$	-0.50%	8.03%	333.788	-0.48% ^{NS}	8.07%	337.358
	$\hat{\sigma}_v^2$	-0.02% ^{NS}	3.73%	34.592	-0.05% ^{NS}	3.80%	35.804
	$\hat{\gamma}$	-	-	-	NR ^{NS}	NR	0.406

^{NS} – denotes ‘absolute bias not significantly different from zero at 95% level’.
NR – denotes ‘no reported’.
† Numerical minimisations did not achieve convergence for 0.2% of the replicates.

Table 5.4 – Evaluation of $\underline{\hat{\theta}}$ with $t_{\upsilon=5}(0,1)$ distributed errors (population – replications generated by UCM-C model).

Results presented in Table 5.4 (data generated by a UCM-C model with t distributed errors) above lead us to make similar conclusions to those made previously for Table 5.2 (data generated by a UCM-C model with normally distributed errors). Nevertheless, we may observe that most methods have had a large increase in their mse for both fitted models (UCM and AR1), mainly because of a modest increase in bias and also inflation of the coefficient of variation. Estimators GLS4 and GLSC4, i.e. ADF methods, have again have generally not performed better than the other methods, although those were the methods with the lowest levels of variance in all situations considered in the table above for estimating all the components of $\underline{\theta}$.

Moreover, when comparing Table 5.4 with Table 5.3 (data generated by UCM model with normally distributed errors), we may notice that all methods have had an increase in their mse for both fitted models (UCM and AR1), caused by a modest inflation in the coefficient of variation.

5.3.3 Simulation results for samples of size $n^{sim} = 500$

This subsection presents results that are produced making use of samples of size $n^{sim} = 500$ selected by srs from replications of the whole simulation population, as described in Section 5.2, with $D = 1000$, and generated by UCM and UCM-C models.

Table 5.5 includes results for both UCM and AR1 models considering normality conditions, as described in Section 5.2, Subsection 5.2.1. The values of Y_{it} are generated here by an UCM model.

		UCM model			AR1 model		
parameter		rel bias	$cv(\hat{\theta})$	mse $(\times 1000)$	rel bias	$cv(\hat{\theta})$	mse $(\times 1000)$
$\hat{\underline{\theta}}_{ULS}$	$\hat{\sigma}_u^2$	-1.92%	7.08%	236.411	-2.06%	7.21%	246.391
	$\hat{\sigma}_v^2$	0.47%	3.25%	26.827	0.66%	3.58%	33.167
	$\hat{\gamma}$	-	-	-	NR	NR	1.482
$\hat{\underline{\theta}}^2_{GLS}$	$\hat{\sigma}_u^2$	-3.40%	7.16%	270.239	-3.51%	7.25%	278.756
	$\hat{\sigma}_v^2$	-1.74%	3.39%	34.812	-1.41%	3.62%	36.252
	$\hat{\gamma}$	-	-	-	NR [NS]	NR	1.002
$\hat{\underline{\theta}}^3_{GLS}$	$\hat{\sigma}_u^2$	-1.18%	7.12%	231.661	-1.25%	7.19%	236.511
	$\hat{\sigma}_v^2$	1.59%	3.27%	33.447	1.60%	3.48%	37.112
	$\hat{\gamma}$	-	-	-	NR [NS]	NR	0.970
$\hat{\underline{\theta}}^4_{GLS}$	$\hat{\sigma}_u^2$	-3.41%	7.19%	272.544	-3.52%	7.33%	283.505
	$\hat{\sigma}_v^2$	-1.72%	3.44%	35.533	-1.40%	3.68%	37.363
	$\hat{\gamma}$	-	-	-	NR [NS]	NR	1.038
$\hat{\underline{\theta}}_{ML}$	$\hat{\sigma}_u^2$	-1.92%	7.08%	236.411	-2.01%	7.16%	242.442
	$\hat{\sigma}_v^2$	0.47%	3.25%	26.827	0.59%	3.47%	30.888
	$\hat{\gamma}$	-	-	-	NR [NS]	NR	0.968
$\hat{\underline{\theta}}_{ULSC}$	$\hat{\sigma}_u^2$	-0.66%	7.54%	257.025	-0.73%	7.65%	265.017
	$\hat{\sigma}_v^2$	-0.33%	3.43%	29.139	-0.23% [NS]	3.76%	34.875
	$\hat{\gamma}$	-	-	-	NR [NS]	NR	1.659
$\hat{\underline{\theta}}^2_{GLSC}$	$\hat{\sigma}_u^2$	-2.36%	7.65%	279.108	-2.39%	7.72%	284.412
	$\hat{\sigma}_v^2$	-2.78%	3.55%	48.376	-2.55%	3.77%	49.353
	$\hat{\gamma}$	-	-	-	NR [NS]	NR	1.104
$\hat{\underline{\theta}}^3_{GLSC}$	$\hat{\sigma}_u^2$	0.21% [NS]	7.57%	262.204	0.23% [NS]	7.64%	266.679
	$\hat{\sigma}_v^2$	0.90%	3.48%	32.485	0.79%	3.67%	35.328
	$\hat{\gamma}$	-	-	-	NR [NS]	NR	1.053
$\hat{\underline{\theta}}^4_{GLSC}$	$\hat{\sigma}_u^2$	-2.48%	7.77%	289.052	-2.50%	7.87%	296.105
	$\hat{\sigma}_v^2$	-2.96%	3.63%	52.260	-2.72%	3.90%	53.785
	$\hat{\gamma}$	-	-	-	NR [NS]	NR	1.178
$\hat{\underline{\theta}}_{PML}$	$\hat{\sigma}_u^2$	-0.66%	7.54%	257.025	-0.66%	7.61%	261.675
	$\hat{\sigma}_v^2$	-0.33%	3.43%	29.139	-0.34%	3.64%	32.779
	$\hat{\gamma}$	-	-	-	NR [NS]	NR	1.054

[NS] – denotes 'absolute bias not significantly different from zero at 95% level'.
NR – denotes 'no reported'.

Table 5.5 – Evaluation of $\hat{\underline{\theta}}$ with normally distributed errors ($n^{sim} = 500$, replications generated by UCM model).

From results included in Table 5.5 we can make very similar remarks to those made for results presented in Table 5.1, in terms of comparing the methods. Again, the estimator GLSC4 has had the largest bias among the weighted methods, although in terms of efficiency this was very similar to the other methods, when fitting both models. Results in Table 5.5 indicate increases of both bias and variance when compared to those in Table 5.1, caused by a reduction in the number of cases from 1340 to 500.

Table 5.6 includes results for both UCM and AR1 models considering normality conditions, with simulated Y_{it} values generated now by an UCM-C model.

parameter		UCM model			AR1 model		
		Rel bias	$cv(\hat{\theta})$	mse $(\times 1000)$	rel bias	$cv(\hat{\theta})$	mse $(\times 1000)$
$\hat{\underline{\theta}}_{ULS}$	$\hat{\sigma}_u^2$	-1.88%	7.49%	299.989	-1.97%	7.51%	302.818
	$\hat{\sigma}_v^2$	0.35%	3.12%	24.603	0.48%	3.48%	30.963
	$\hat{\gamma}$	-	-	-	NR [NS]	NR	1.579
$\hat{\underline{\theta}}^2_{GLS}$	$\hat{\sigma}_u^2$	-3.33%	7.53%	333.952	-3.39%	7.55%	337.235
	$\hat{\sigma}_v^2$	-1.83%	3.27%	33.972	-1.56%	3.50%	35.533
	$\hat{\gamma}$	-	-	-	NR [NS]	NR	1.011
$\hat{\underline{\theta}}^3_{GLS}$	$\hat{\sigma}_u^2$	-1.15%	7.54%	296.629	-1.19%	7.57%	299.035
	$\hat{\sigma}_v^2$	1.45%	3.14%	30.319	1.41%	3.36%	33.820
	$\hat{\gamma}$	-	-	-	NR [NS]	NR	0.964
$\hat{\underline{\theta}}^4_{GLS}$	$\hat{\sigma}_u^2$	-3.25%	7.63%	339.025	-3.34%	7.64%	342.691
	$\hat{\sigma}_v^2$	-1.83%	3.34%	35.057	-1.55%	3.57%	36.561
	$\hat{\gamma}$	-	-	-	NR [NS]	NR	1.052
$\hat{\underline{\theta}}_{ML}$	$\hat{\sigma}_u^2$	-1.88%	7.49%	299.989	-1.93%	7.52%	302.626
	$\hat{\sigma}_v^2$	0.35%	3.12%	24.603	0.42%	3.35%	28.582
	$\hat{\gamma}$	-	-	-	NR [NS]	NR	0.964
$\hat{\underline{\theta}}_{ULSC}$	$\hat{\sigma}_u^2$	-0.62%	7.90%	323.596	-0.61%	7.94%	327.029
	$\hat{\sigma}_v^2$	-0.49%	3.29%	27.269	-0.52%	3.71%	34.468
	$\hat{\gamma}$	-	-	-	NR	NR	1.756
$\hat{\underline{\theta}}^2_{GLSC}$	$\hat{\sigma}_u^2$	-2.25%	8.01%	345.944	-2.21%	8.04%	347.401
	$\hat{\sigma}_v^2$	-2.94%	3.44%	49.103	-2.80%	3.72%	51.887
	$\hat{\gamma}$	-	-	-	NR	NR	1.141
$\hat{\underline{\theta}}^3_{GLSC}$	$\hat{\sigma}_u^2$	0.20% [NS]	7.93%	329.596	0.27% [NS]	7.97%	333.385
	$\hat{\sigma}_v^2$	0.74%	3.34%	29.390	0.53%	3.62%	33.573
	$\hat{\gamma}$	-	-	-	NR	NR	1.099
$\hat{\underline{\theta}}^4_{GLSC}$	$\hat{\sigma}_u^2$	-2.31%	8.21%	363.150	-2.30%	8.23%	364.697
	$\hat{\sigma}_v^2$	-3.11%	3.54%	53.292	-2.96%	3.81%	55.654
	$\hat{\gamma}$	-	-	-	NR	NR	1.208
$\hat{\underline{\theta}}_{PML}$	$\hat{\sigma}_u^2$	-0.62%	7.90%	323.596	-0.56%	7.94%	326.520
	$\hat{\sigma}_v^2$	-0.49%	3.29%	27.269	-0.59%	3.59%	32.435
	$\hat{\gamma}$	-	-	-	NR	NR	1.094

[NS] – denotes 'absolute bias not significantly different from zero at 95% level'.
NR – denotes 'no reported'.

Table 5.6 – Evaluation of $\underline{\hat{\theta}}$ with normally distributed errors ($n^{sim} = 500$, replications generated by UCM-C model).

Similar conclusions may be drawn when comparing Tables 5.6 and 5.5. Nevertheless, we may observe that most methods have had an increase in their mse for both fitted models (UCM and AR1), mainly because of a modest increase in the variance especially when estimating the variance component σ_u^2 .

Table 5.7 includes results for both UCM and AR1 models considering non-normality conditions (t distribution), with simulated Y_{it} values generated by an UCM model.

parameter		UCM model			AR1 model		
		rel bias	$cv(\hat{\theta})$	mse $(\times 1000)$	rel bias	$cv(\hat{\theta})$	mse $(\times 1000)$
$\underline{\hat{\theta}}_{ULS}$	$\hat{\sigma}_u^2$	-2.61%	12.37%	690.612	-2.65%	12.37%	691.055
	$\hat{\sigma}_v^2$	0.27%NS	5.59%	77.544	0.32%NS	5.78%	83.130
	$\hat{\gamma}$	-	-	-	NR NS	NR	1.381
$\underline{\hat{\theta}}^2_{GLS}$	$\hat{\sigma}_u^2$	-4.07%	12.41%	719.647	-4.15%	12.41%	721.556
	$\hat{\sigma}_v^2$	-3.03%	5.20%	85.411	-2.76%	5.30%	84.220
	$\hat{\gamma}$	-	-	-	NR NS	NR	0.975
$\underline{\hat{\theta}}^3_{GLS}$	$\hat{\sigma}_u^2$	-1.87%	12.40%	689.077	-1.91%	12.40%	689.311
	$\hat{\sigma}_v^2$	2.02%	6.53%	119.593	2.00%	6.57%	120.415
	$\hat{\gamma}$	-	-	-	NR NS	NR	0.975
$\underline{\hat{\theta}}^4_{GLS}$ †	$\hat{\sigma}_u^2$	-6.06%	11.53%	700.446	-6.14%	11.53%	704.369
	$\hat{\sigma}_v^2$	-6.01%	5.31%	150.514	-5.55%	5.37%	139.265
	$\hat{\gamma}$	-	-	-	NR	NR	0.956
$\underline{\hat{\theta}}_{ML}$	$\hat{\sigma}_u^2$	-2.61%	12.37%	690.612	-2.66%	12.37%	691.220
	$\hat{\sigma}_v^2$	0.27%NS	5.59%	77.544	0.35%NS	5.67%	80.097
	$\hat{\gamma}$	-	-	-	NR NS	NR	0.943
$\underline{\hat{\theta}}_{ULSC}$	$\hat{\sigma}_u^2$	-1.50%	13.55%	820.430	-1.45%	13.55%	820.790
	$\hat{\sigma}_v^2$	-0.65%	5.89%	85.472	-0.71%	6.15%	93.284
	$\hat{\gamma}$	-	-	-	NR	NR	1.542
$\underline{\hat{\theta}}^2_{GLSC}$	$\hat{\sigma}_u^2$	-3.13%	13.63%	837.429	-3.12%	13.63%	836.813
	$\hat{\sigma}_v^2$	-4.26%	5.43%	111.429	-4.10%	5.59%	112.306
	$\hat{\gamma}$	-	-	-	NR	NR	1.120
$\underline{\hat{\theta}}^3_{GLSC}$	$\hat{\sigma}_u^2$	-0.67%NS	13.57%	828.043	-0.61%NS	13.56%	828.145
	$\hat{\sigma}_v^2$	1.29%	7.19%	135.003	1.12%	7.24%	135.141
	$\hat{\gamma}$	-	-	-	NR	NR	1.114
$\underline{\hat{\theta}}^4_{GLSC}$ †	$\hat{\sigma}_u^2$	-5.37%	12.39%	756.275	-5.41%	12.37%	755.286
	$\hat{\sigma}_v^2$	-7.57%	5.55%	206.132	-7.16%	5.70%	195.401
	$\hat{\gamma}$	-	-	-	NR NS	NR	1.091
$\underline{\hat{\theta}}_{PML}$	$\hat{\sigma}_u^2$	-1.50%	13.55%	820.430	-1.45%	13.54%	819.841
	$\hat{\sigma}_v^2$	-0.65%	5.89%	85.472	-0.70%	6.02%	89.435
	$\hat{\gamma}$	-	-	-	NR	NR	1.079

NS – denotes 'absolute bias not significantly different from zero at 95% level'.
NR – denotes 'no reported'.
† Numerical minimisations did not achieve convergence for 0.1% of the replicates.

Table 5.7 – Evaluation of $\underline{\hat{\theta}}$ with $t_{\upsilon=5}(0,1)$ distributed errors ($n^{sim} = 500$, replications generated by UCM model).

When comparing Table 5.7 with Table 5.6, we may notice that all methods have had a large increase in their mse for both fitted models (UCM and AR1), caused by an inflation of both bias and variance. The ADF method GLSC4 has shown the best performance, measured by the mean square error, among the weighted methods for estimating σ_u^2 when fitting a UCM model and for estimating both σ_u^2 and γ when fitting a AR1 model, although that method is the one with the largest levels of bias. The reason, thus, for its better performance are the c.v. results. Generally similar commentaries could be made if we consider the unweighted ADF method GLS4.

Table 5.8 includes results for both UCM and AR1 models considering non-normality conditions (t distribution), with simulated Y_{it} values generated in the current situation by an UCM-C model.

parameter		UCM model rel bias	UCM model $cv(\hat{\theta})$	UCM model mse $(\times 1000)$	AR1 model rel bias	AR1 model $cv(\hat{\theta})$	AR1 model mse $(\times 1000)$
$\underline{\hat{\theta}}_{ULS}$	$\hat{\sigma}_u^2$	-2.54%	12.10%	759.045	-2.59%	12.14%	764.228
	$\hat{\sigma}_v^2$	0.51%	5.55%	77.756	0.58%	5.76%	84.189
	$\hat{\gamma}$	-	-	-	NR [NS]	NR	1.360
$\underline{\hat{\theta}}_{GLS}^2$	$\hat{\sigma}_u^2$	-4.05%	12.10%	788.304	-4.11%	12.14%	794.531
	$\hat{\sigma}_v^2$	-2.65%	5.23%	81.645	-2.39%	5.36%	82.071
	$\hat{\gamma}$	-	-	-	NR [NS]	NR	0.969
$\underline{\hat{\theta}}_{GLS}^3$	$\hat{\sigma}_u^2$	-1.78%	12.17%	761.232	-1.82%	12.20%	765.733
	$\hat{\sigma}_v^2$	2.18%	6.08%	107.520	2.17%	6.20%	111.122
	$\hat{\gamma}$	-	-	-	NR [NS]	NR	0.953
$\underline{\hat{\theta}}_{GLS}^4$ †	$\hat{\sigma}_u^2$	-6.16%	11.72%	828.391	-6.21%	11.73%	832.693
	$\hat{\sigma}_v^2$	-5.58%	5.16%	136.241	-5.17%	5.28%	128.515
	$\hat{\gamma}$	-	-	-	NR [NS]	NR	0.952
$\underline{\hat{\theta}}_{ML}$	$\hat{\sigma}_u^2$	-2.54%	12.10%	759.045	-2.59%	12.14%	764.005
	$\hat{\sigma}_v^2$	0.51%	5.55%	77.756	0.59%	5.67%	81.555
	$\hat{\gamma}$	-	-	-	NR [NS]	NR	0.929
$\underline{\hat{\theta}}_{ULSC}$	$\hat{\sigma}_u^2$	-1.18%	13.11%	882.082	-1.16%	13.12%	883.958
	$\hat{\sigma}_v^2$	-0.43%	5.82%	83.823	-0.47%	6.04%	90.227
	$\hat{\gamma}$	-	-	-	NR	NR	1.546
$\underline{\hat{\theta}}_{GLSC}^2$	$\hat{\sigma}_u^2$	-2.88%	13.06%	881.941	-2.86%	13.09%	886.110
	$\hat{\sigma}_v^2$	-3.93%	5.52%	107.947	-3.78%	5.69%	109.668
	$\hat{\gamma}$	-	-	-	NR	NR	1.114
$\underline{\hat{\theta}}_{GLSC}^3$	$\hat{\sigma}_u^2$	-0.33% [NS]	13.21%	904.824	-0.26% [NS]	13.25%	910.477
	$\hat{\sigma}_v^2$	1.42%	6.30%	106.310	1.24%	6.44%	109.230
	$\hat{\gamma}$	-	-	-	NR	NR	1.084
$\underline{\hat{\theta}}_{GLSC}^4$ †	$\hat{\sigma}_u^2$	-5.44%	12.45%	877.106	-5.41%	12.49%	880.550
	$\hat{\sigma}_v^2$	-7.16%	5.44%	190.393	-6.84%	5.55%	182.239
	$\hat{\gamma}$	-	-	-	NR	NR	1.083
$\underline{\hat{\theta}}_{PML}$	$\hat{\sigma}_u^2$	-1.18%	13.11%	882.082	-1.11%	13.13%	885.635
	$\hat{\sigma}_v^2$	-0.43%	5.82%	83.823	-0.49%	5.97%	88.146
	$\hat{\gamma}$	-	-	-	NR	NR	1.056

[NS] – denotes 'absolute bias not significantly different from zero at 95% level'.
NR – denotes 'no reported'.
† Numerical minimisations did not achieve convergence for 0.1% of the replicates.

Table 5.8 – Evaluation of $\underline{\hat{\theta}}$ with $t_{\upsilon=5}(0,1)$ distributed errors ($n^{sim} = 500$, replications generated by UCM-C model).

By looking across each row of Table 5.8 we can notice that, when fitting both UCM and AR1 models ML and ULS are the methods with the lowest mse for estimating σ_u^2 and σ_v^2 among the unweighted methods while GLSC4 is the method with the lowest mse among the weighted methods for estimating the variance component σ_u^2 and PML is the method with the lowest mse for estimating σ_v^2 (and γ for the AR1 model). Note that ADF methods have performed better than the other methods in some situations, although that method is the one with the worst bias results.

When comparing results presented in Table 5.8 with those presented in Tables 5.5 and 5.6, we may observe that all methods have had a large increase in their mse for both fitted models (UCM and AR1), mainly because of a modest bias inflation and also an increase in the variance. Moreover, when comparing Table 5.8 with Table 5.7, we may notice most methods have also had an increase in their mse for both fitted models (UCM and AR1), caused in general by modest variance inflation.

Tables included in this sub-section have all indicated an increase of both bias and coefficient of variation when compared to those presented in the previous sub-section, caused presumably by a reduction in the number of cases to 500 units.

5.3.4 Simulation results for samples of size $n^{sim} = 200$

In the current sub-section results are produced making use of samples of size $n^{sim} = 200$ selected by srs from replications of the whole simulation population, and generated by UCM and UCM-C models. The number of simulated repeated datasets had to be reduced to $D = 700$ in this sub-section, mainly as result of computation problems with the function **nlm** from the software *R* (see Sub-section 5.3.1 earlier), which had more difficulties in yielding converged solutions with the sample size decrease to 200.

Table 5.9 includes results for both UCM and AR1 models considering normality conditions. The values of the simulated Y_{it} are generated here by an UCM model.

parameter		UCM model rel bias	UCM model $cv(\hat{\theta})$	UCM model mse $(\times 1000)$	AR1 model rel bias	AR1 model $cv(\hat{\theta})$	AR1 model mse $(\times 1000)$
$\underline{\hat{\theta}}_{ULS}$	$\hat{\sigma}_u^2$	-5.61%	11.68%	695.721	-5.84%	11.86%	722.191
	$\hat{\sigma}_v^2$	0.74%	5.09%	66.111	1.05%	5.65%	83.011
	$\hat{\gamma}$	-	-	-	NR ^{NS}	NR	3.733
$\underline{\hat{\theta}}^2_{GLS}$	$\hat{\sigma}_u^2$	-9.21%	12.14%	937.882	-9.43%	12.29%	968.127
	$\hat{\sigma}_v^2$	-4.68%	5.56%	123.275	-3.93%	5.89%	117.028
	$\hat{\gamma}$	-	-	-	NR ^{NS}	NR	2.626
$\underline{\hat{\theta}}^3_{GLS}$	$\hat{\sigma}_u^2$	-3.77%	11.77%	647.897	-3.91%	11.87%	661.138
	$\hat{\sigma}_v^2$	3.51%	5.20%	101.633	3.46%	5.55%	110.864
	$\hat{\gamma}$	-	-	-	NR ^{NS}	NR	2.394
$\underline{\hat{\theta}}^4_{GLS}$ †	$\hat{\sigma}_u^2$	-9.23%	12.76%	997.576	-9.43%	12.90%	1,024.906
	$\hat{\sigma}_v^2$	-4.60%	5.83%	128.387	-3.84%	6.22%	124.593
	$\hat{\gamma}$	-	-	-	NR ^{NS}	NR	2.834
$\underline{\hat{\theta}}_{ML}$	$\hat{\sigma}_u^2$	-5.61%	11.68%	695.721	-5.77%	11.80%	713.529
	$\hat{\sigma}_v^2$	0.74%	5.09%	66.111	0.95%	5.45%	76.810
	$\hat{\gamma}$	-	-	-	NR ^{NS}	NR	2.369
$\underline{\hat{\theta}}_{ULSC}$	$\hat{\sigma}_u^2$	-2.44%	12.56%	709.745	-2.47%	12.70%	725.534
	$\hat{\sigma}_v^2$	-1.23%	5.39%	73.560	-1.18%	6.02%	90.670
	$\hat{\gamma}$	-	-	-	NR ^{NS}	NR	4.357
$\underline{\hat{\theta}}^2_{GLSC}$	$\hat{\sigma}_u^2$	-6.62%	12.94%	863.855	-6.57%	13.07%	875.016
	$\hat{\sigma}_v^2$	-7.16%	5.90%	200.225	-6.72%	6.26%	195.212
	$\hat{\gamma}$	-	-	-	NR	NR	2.913
$\underline{\hat{\theta}}^3_{GLSC}$	$\hat{\sigma}_u^2$	-0.27%^{NS}	12.79%	740.548	-0.10%^{NS}	12.89%	753.650
	$\hat{\sigma}_v^2$	1.83%	5.53%	86.426	1.34%	5.85%	91.071
	$\hat{\gamma}$	-	-	-	NR	NR	2.624
$\underline{\hat{\theta}}^4_{GLSC}$ †	$\hat{\sigma}_u^2$	-6.96%	13.81%	970.834	-6.95%	13.86%	975.971
	$\hat{\sigma}_v^2$	-7.35%	6.19%	214.237	-6.87%	6.63%	210.403
	$\hat{\gamma}$	-	-	-	NR	NR	3.147
$\underline{\hat{\theta}}_{PML}$	$\hat{\sigma}_u^2$	-2.44%	12.56%	709.745	-2.31%	12.66%	720.447
	$\hat{\sigma}_v^2$	-1.23%	5.39%	73.560	-1.41%	5.74%	83.859
	$\hat{\gamma}$	-	-	-	NR	NR	2.603

^{NS} – denotes 'absolute bias not significantly different from zero at 95% level'.
NR – denotes 'no reported'.
† Numerical minimisations did not achieve convergence for 0.3% of the replicates.

Table 5.9 – Evaluation of $\underline{\hat{\theta}}$ with normally distributed errors ($n^{sim} = 200$, replications generated by UCM model).

In terms of methods comparison, results included in Table 5.9 are very similar to those in Tables 5.5 and 5.1. We may however highlight here that ULS, ML and GLS3 are the methods with the lowest relative bias, now with larger differences to the remaining unweighted methods. In terms of variance, methods are generally very similar with ULS and ML with a smaller coefficient of variation than the other methods. Once more, the estimator GLS4 has had the largest relative bias among the unweighted methods, although in terms of efficiency this was reasonably similar to the other methods, when fitting both models. Results in Table 5.9 indicate inflation in both bias and variance when compared to those in Tables 5.5 and 5.1, caused by a reduction in the number of cases to 200 units now.

Table 5.10 includes results for both UCM and AR1 models under normality conditions, with the simulated values of Y_{it} generated by an UCM-C model.

parameter		UCM model			AR1 model		
		Rel bias	$cv(\hat{\theta})$	mse $(\times 1000)$	rel bias	$cv(\hat{\theta})$	mse $(\times 1000)$
$\hat{\underline{\theta}}_{ULS}$	$\hat{\sigma}_u^2$	-5.53%	11.87%	814.974	-5.71%	12.04%	842.151
	$\hat{\sigma}_v^2$	0.50%	4.87%	60.111	0.76%	5.37%	73.996
	$\hat{\gamma}$	-	-	-	NR [NS]	NR	3.487
$\hat{\underline{\theta}}^2_{GLS}$	$\hat{\sigma}_u^2$	-9.25%	12.30%	1,095.938	-9.48%	12.44%	1,130.456
	$\hat{\sigma}_v^2$	-4.94%	5.29%	123.229	-4.21%	5.47%	112.102
	$\hat{\gamma}$	-	-	-	NR	NR	2.275
$\hat{\underline{\theta}}^3_{GLS}$	$\hat{\sigma}_u^2$	-3.61%	12.03%	769.231	-3.76%	12.12%	783.029
	$\hat{\sigma}_v^2$	3.30%	5.03%	93.928	3.34%	5.36%	103.672
	$\hat{\gamma}$	-	-	-	NR	NR	2.128
$\hat{\underline{\theta}}^4_{GLS}$ †	$\hat{\sigma}_u^2$	-9.07%	12.82%	1,138.103	-9.32%	12.94%	1,170.130
	$\hat{\sigma}_v^2$	-4.93%	5.40%	125.634	-4.18%	5.62%	115.401
	$\hat{\gamma}$	-	-	-	NR	NR	2.510
$\hat{\underline{\theta}}_{ML}$	$\hat{\sigma}_u^2$	-5.53%	11.87%	814.974	-5.71%	11.98%	835.011
	$\hat{\sigma}_v^2$	0.50%	4.87%	60.111	0.77%	5.17%	68.844
	$\hat{\gamma}$	-	-	-	NR	NR	2.071
$\hat{\underline{\theta}}_{ULSC}$	$\hat{\sigma}_u^2$	-2.48%	12.73%	835.511	-2.47%	12.87%	853.695
	$\hat{\sigma}_v^2$	-1.69%	5.15%	70.607	-1.70%	5.74%	86.032
	$\hat{\gamma}$	-	-	-	NR	NR	4.221
$\hat{\underline{\theta}}^2_{GLSC}$	$\hat{\sigma}_u^2$	-6.69%	13.20%	1,023.774	-6.68%	13.29%	1,034.695
	$\hat{\sigma}_v^2$	-7.68%	5.64%	213.783	-7.24%	5.84%	202.739
	$\hat{\gamma}$	-	-	-	NR	NR	2.702
$\hat{\underline{\theta}}^3_{GLSC}$	$\hat{\sigma}_u^2$	-0.29% [NS]	12.92%	865.788	-0.19% [NS]	12.97%	874.047
	$\hat{\sigma}_v^2$	1.40%	5.34%	77.608	1.03%	5.66%	83.737
	$\hat{\gamma}$	-	-	-	NR	NR	2.506
$\hat{\underline{\theta}}^4_{GLSC}$ †	$\hat{\sigma}_u^2$	-6.51%	13.81%	1,090.377	-6.49%	13.86%	1,095.724
	$\hat{\sigma}_v^2$	-8.00%	5.86%	230.711	-7.48%	6.12%	218.364
	$\hat{\gamma}$	-	-	-	NR	NR	2.925
$\hat{\underline{\theta}}_{PML}$	$\hat{\sigma}_u^2$	-2.48%	12.73%	835.511	-2.41%	12.80%	844.362
	$\hat{\sigma}_v^2$	-1.69%	5.15%	70.607	-1.79%	5.46%	79.171
	$\hat{\gamma}$	-	-	-	NR	NR	2.442

[NS] – denotes 'absolute bias not significantly different from zero at 95% level'.
NR – denotes 'no reported'.
† Numerical minimisations did not achieve convergence for 0.4% of the replicates.

Table 5.10 – Evaluation of $\hat{\underline{\theta}}$ with normally distributed errors ($n^{sim} = 200$, replications generated by UCM-C model).

Similar conclusions may be drawn when comparing Tables 5.10 and 5.9. Nevertheless, we may observe that most methods have had an increase in their mse for both fitted models (UCM and AR1), mainly because of a modest increase in variance especially when estimating the variance component σ_u^2 .

Table 5.11 includes results for both UCM and AR1 models under non-normality conditions (t distribution), with the simulated values of Y_{it} generated by an UCM model.

parameter		UCM model			AR1 model		
		rel bias	$cv(\hat{\theta})$	mse $(\times 1000)$	rel bias	$cv(\hat{\theta})$	mse $(\times 1000)$
$\hat{\underline{\theta}}_{ULS}$	$\hat{\sigma}_u^2$	-5.46%	19.82%	1,733.078	-5.61%	19.98%	1,760.609
	$\hat{\sigma}_v^2$	0.98%	8.27%	174.310	1.18%	8.65%	192.143
	$\hat{\gamma}$	-	-	-	NR ^{NS}	NR	3.690
$\hat{\underline{\theta}}^2_{GLS}$	$\hat{\sigma}_u^2$	-8.98%	19.90%	1,859.575	-9.16%	19.99%	1,882.111
	$\hat{\sigma}_v^2$	-6.54%	7.77%	235.527	-5.87%	8.02%	225.340
	$\hat{\gamma}$	-	-	-	NR ^{NS}	NR	2.742
$\hat{\underline{\theta}}^3_{GLS}$ ‡	$\hat{\sigma}_u^2$	-3.69%	20.01%	1,751.308	-3.76%	20.07%	1,761.696
	$\hat{\sigma}_v^2$	5.04%	9.31%	298.359	4.92%	9.49%	304.357
	$\hat{\gamma}$	-	-	-	NR ^{NS}	NR	2.635
$\hat{\underline{\theta}}^4_{GLS}$ †	$\hat{\sigma}_u^2$	-12.22%	20.02%	2,082.938	-12.47%	20.07%	2,110.137
	$\hat{\sigma}_v^2$	-10.89%	7.99%	417.203	-9.97%	8.24%	380.816
	$\hat{\gamma}$	-	-	-	NR ^{NS}	NR	2.826
$\hat{\underline{\theta}}_{ML}$	$\hat{\sigma}_u^2$	-5.46%	19.82%	1,733.078	-5.56%	19.89%	1,746.055
	$\hat{\sigma}_v^2$	0.98%	8.27%	174.310	1.02%	9.30%	220.255
	$\hat{\gamma}$	-	-	-	NR ^{NS}	NR	5.719
$\hat{\underline{\theta}}_{ULSC}$	$\hat{\sigma}_u^2$	-2.77%	20.20%	1,789.713	-2.73%	20.32%	1,810.858
	$\hat{\sigma}_v^2$	-1.03%	8.93%	195.121	-1.09%	9.25%	209.201
	$\hat{\gamma}$	-	-	-	NR	NR	4.030
$\hat{\underline{\theta}}^2_{GLSC}$	$\hat{\sigma}_u^2$	-6.72%	20.33%	1,840.788	-6.67%	20.38%	1,847.667
	$\hat{\sigma}_v^2$	-9.20%	8.32%	349.445	-8.79%	8.56%	340.997
	$\hat{\gamma}$	-	-	-	NR	NR	3.104
$\hat{\underline{\theta}}^3_{GLSC}$ ‡	$\hat{\sigma}_u^2$	-0.80% ^{NS}	20.37%	1,859.864	-0.58% ^{NS}	20.42%	1,876.328
	$\hat{\sigma}_v^2$	3.45%	10.29%	308.483	2.92%	10.33%	299.905
	$\hat{\gamma}$	-	-	-	NR	NR	2.909
$\hat{\underline{\theta}}^4_{GLSC}$ †	$\hat{\sigma}_u^2$	-10.46%	20.68%	2,057.537	-10.58%	20.64%	2,058.255
	$\hat{\sigma}_v^2$	-13.73%	8.73%	604.661	-13.01%	8.89%	564.782
	$\hat{\gamma}$	-	-	-	NR	NR	3.122
$\hat{\underline{\theta}}_{PML}$	$\hat{\sigma}_u^2$	-2.77%	20.20%	1,789.713	-2.60%	20.25%	1,800.507
	$\hat{\sigma}_v^2$	-1.03%	8.93%	195.121	-1.24%	9.10%	202.969
	$\hat{\gamma}$	-	-	-	NR	NR	2.777

^{NS} – denotes 'absolute bias not significantly different from zero at 95% level'.
NR – denotes 'no reported'.
† Numerical minimisations did not achieve convergence for 0.4% of the replicates.
‡ Numerical minimisations did not achieve convergence for 0.1% of the replicates.

Table 5.11 – Evaluation of $\hat{\underline{\theta}}$ with $t_{\upsilon=5}(0,1)$ distributed errors ($n^{sim}=200$, replications generated by UCM model).

When comparing Table 5.11 with Table 5.10, we may observe that all methods have had a large increase in their mse for both fitted models (UCM and AR1), caused in most cases by a moderate increase in bias and inflation in variance results. Dissimilar to results in Tables 5.7 and 5.8, results in Table 5.10 suggest that the ADF methods, i.e. GLS4 and GLSC4, have not performed better than the remaining methods (those were not even more efficient in any situation). We may notice that, when fitting both UCM and AR1 models, ML and ULS are the methods with the lowest mse for estimating both variance components σ_u^2 and σ_v^2 among the unweighted methods, while PML and ULSC are the methods with the lowest mse among the weighted methods.

Table 5.12 includes results for both UCM and AR1 models produced when considering non-normality conditions (t distribution), with simulated values of Y_{it} generated now by an UCM-C model.

parameter		UCM model			AR1 model		
		rel bias	$cv(\hat{\theta})$	mse $(\times 1000)$	rel bias	$cv(\hat{\theta})$	mse $(\times 1000)$
$\underline{\hat{\theta}}_{ULS}$	$\hat{\sigma}_u^2$	-5.16%	22.00%	2,408.666	-5.42%	22.04%	2,418.613
	$\hat{\sigma}_v^2$	0.62%^{NS}	8.62%	187.740	0.99%	8.98%	206.525
	$\hat{\gamma}$	-	-	-	NR	NR	3.606
$\underline{\hat{\theta}}^2_{GLS}$	$\hat{\sigma}_u^2$	-8.75%	22.07%	2,513.558	-9.00%	22.13%	2,536.356
	$\hat{\sigma}_v^2$	-7.13%	8.08%	265.693	-6.35%	8.37%	252.291
	$\hat{\gamma}$	-	-	-	NR	NR	2.700
$\underline{\hat{\theta}}^3_{GLS}$ ‡*	$\hat{\sigma}_u^2$	-3.32%	22.28%	2,476.950	-3.52%	22.32%	2,482.437
	$\hat{\sigma}_v^2$	4.83%	9.92%	326.058	4.96%	10.21%	345.776
	$\hat{\gamma}$	-	-	-	NR	NR	2.585
$\underline{\hat{\theta}}^4_{GLS}$ †	$\hat{\sigma}_u^2$	-13.00%	18.55%	2,239.679	-13.14%	18.79%	2,289.138
	$\hat{\sigma}_v^2$	-11.54%	8.23%	461.896	-10.56%	8.45%	418.473
	$\hat{\gamma}$	-	-	-	NR^{NS}	NR	2.846
$\underline{\hat{\theta}}_{ML}$	$\hat{\sigma}_u^2$	-5.16%	22.00%	2,408.666	-5.36%	22.06%	2,423.559
	$\hat{\sigma}_v^2$	0.62%^{NS}	8.62%	187.740	0.97%	8.89%	201.996
	$\hat{\gamma}$	-	-	-	NR	NR	2.401
$\underline{\hat{\theta}}_{ULSC}$	$\hat{\sigma}_u^2$	-1.26%^{NS}	24.94%	3,171.113	-1.29%^{NS}	24.94%	3,169.397
	$\hat{\sigma}_v^2$	-1.62%	9.05%	203.206	-1.59%	9.42%	219.305
	$\hat{\gamma}$	-	-	-	NR^{NS}	NR	4.043
$\underline{\hat{\theta}}^2_{GLSC}$	$\hat{\sigma}_u^2$	-5.59%	24.98%	3,063.605	-5.60%	25.00%	3,068.562
	$\hat{\sigma}_v^2$	-10.02%	8.52%	394.959	-9.54%	8.83%	384.248
	$\hat{\gamma}$	-	-	-	NR	NR	3.073
$\underline{\hat{\theta}}^3_{GLSC}$	$\hat{\sigma}_u^2$	0.96%^{NS}	25.27%	3,399.987	1.07%^{NS}	25.30%	3,415.449
	$\hat{\sigma}_v^2$	2.96%	10.35%	303.464	2.61%	10.66%	313.668
	$\hat{\gamma}$	-	-	-	NR	NR	2.908
$\underline{\hat{\theta}}^4_{GLSC}$ †	$\hat{\sigma}_u^2$	-11.15%	19.80%	2,261.811	-11.06%	19.80%	2,254.076
	$\hat{\sigma}_v^2$	-14.73%	8.62%	672.754	-14.05%	8.85%	633.530
	$\hat{\gamma}$	-	-	-	NR	NR	3.247
$\underline{\hat{\theta}}_{PML}$	$\hat{\sigma}_u^2$	-1.26%^{NS}	24.94%	3,171.113	-1.50%^{NS}	23.26%	2,748.618
	$\hat{\sigma}_v^2$	-1.62%	9.05%	203.206	-1.64%	11.28%	312.244
	$\hat{\gamma}$	-	-	-	NR^{NS}	NR	6.755

^{NS} – denotes 'absolute bias not significantly different from zero at 95% level'.
NR – denotes 'no reported'.
† Numerical minimisations did not achieve convergence for 0.4% of the replicates.
‡ Numerical minimisations did not achieve convergence for 0.1% of the replicates.

* $D = 699$, as 1 (0.14%) of the estimates was improper (see Boomsma, 1985) or inadmissible. By inadmissible, similarly to Hox and Maas (2001), we shall mean in this chapter either negative variance estimates or estimates 'much' larger than the simulated population values.

Table 5.12 – Evaluation of $\underline{\hat{\theta}}$ with $t_{\upsilon=5}(0,1)$ distributed errors ($n^{sim} = 200$, replications generated by UCM-C model).

By looking across each row of Table 5.12 we can notice that, when fitting UCM models, GLS4 is the method with lower mse for estimating σ_u^2, and ML and ULS are the methods with the lowest mse for estimating σ_v^2 among the unweighted methods, while GLSC4 is the method with lower mse for estimating σ_u^2, and PML and ULSC are the methods with the lowest mse for estimating σ_v^2 among the weighted methods. When fitting AR1 models, GLS4 is the method with lower mse for estimating σ_u^2, and ML is the method with lower mse for estimating σ_v^2 and γ among the unweighted methods, while GLSC4 is the method with the lowest mse for estimating σ_u^2, ULSC for estimating σ_v^2 and GLSC3 for estimating γ.

Note that ADF methods, i.e. GLS4 and GLSC4, have performed better than the other methods in some situations, although that method is generally the method with larger bias results. The reason, again, for the good performance of those methods are the low variance results.

When comparing results presented in Table 5.12 with those presented in Tables 5.9 and 5.10, we may observe that all methods have had a large increase in their mse for both fitted models (UCM and AR1), generally because of a modest increase in bias and also an inflation in variance results. Moreover, when comparing Table 5.12 with Table 5.11, we may observe that most methods have also had an increase in their mse for both fitted models (UCM and AR1), caused in general by a modest increase in variance.

Tables included in this sub-section have all indicated inflation in both bias and variance when compared to those presented in the two previous sub-sections, related to a reduction in the number of cases to 200 units.

5.3.5 Simulation results for samples of size $n^{sim} = 100$

This subsection presents results that are produced making use of samples of size $n^{sim} = 100$ selected by srs from replications of the whole simulation population, with $D = 300$ (unless otherwise stated), and generated by UCM and UCM-C models. The number of simulated repeated datasets had to be further reduced to $D = 300$ (unless otherwise stated) in this sub-section, for the same reasons stated in the previous sub-section. Note that, according to Boomsma (1985) and Bollen (1989, p. 255), there is some indications that non-convergent solutions could be more common for samples with size less than 150.

Table 5.13 includes results for both UCM and AR1 models considering normality conditions, as described earlier in Section 5.2, Sub-section 5.2.1. Note that the simulated values of Y_{it} are generated by an UCM model.

parameter		UCM model			AR1 model		
		rel bias	$cv(\hat{\theta})$	mse $(\times 1000)$	rel bias	$cv(\hat{\theta})$	mse $(\times 1000)$
$\hat{\underline{\theta}}_{ULS}$	$\hat{\sigma}_u^2$	-9.94%	17.18%	1,538.694	-10.35%	17.58%	1,617.425
	$\hat{\sigma}_v^2$	0.89%	6.84%	119.462	1.44%	7.63%	152.690
	$\hat{\gamma}$	-	-	-	NR ^{NS}	NR	6.872
$\hat{\underline{\theta}}^2_{GLS}$	$\hat{\sigma}_u^2$	-16.76%	17.77%	2,272.044	-17.21%	18.04%	2,361.748
	$\hat{\sigma}_v^2$	-9.70%	8.41%	374.188	-8.15%	8.93%	329.595
	$\hat{\gamma}$	-	-	-	NR ^{NS}	NR	6.312
$\hat{\underline{\theta}}^3_{GLS}$	$\hat{\sigma}_u^2$	-6.43%	17.69%	1,434.730	-6.56%	17.98%	1,478.911
	$\hat{\sigma}_v^2$	6.41%	7.19%	245.501	6.23%	7.70%	260.823
	$\hat{\gamma}$	-	-	-	NR ^{NS}	NR	4.747
$\hat{\underline{\theta}}^4_{GLS}$ †	$\hat{\sigma}_u^2$	-15.79%	19.44%	2,353.018	-16.16%	19.51%	2,404.134
	$\hat{\sigma}_v^2$	-9.89%	9.04%	404.644	-8.46%	9.62%	367.664
	$\hat{\gamma}$	-	-	-	NR ^{NS}	NR	6.530
$\hat{\underline{\theta}}_{ML}$	$\hat{\sigma}_u^2$	-9.94%	17.18%	1,538.694	-10.19%	17.4%	1,582.899
	$\hat{\sigma}_v^2$	0.89%	6.84%	119.462	1.25%	7.51%	146.308
	$\hat{\gamma}$	-	-	-	NR ^{NS}	NR	4.745
$\hat{\underline{\theta}}_{ULSC}$	$\hat{\sigma}_u^2$	-4.58%	18.28%	1,478.769	-4.66%	18.61%	1,530.982
	$\hat{\sigma}_v^2$	-2.89%	7.40%	148.031	-2.80%	8.28%	178.880
	$\hat{\gamma}$	-	-	-	NR ^{NS}	NR	8.149
$\hat{\underline{\theta}}^2_{GLSC}$	$\hat{\sigma}_u^2$	-12.79%	18.68%	1,951.320	-12.80%	18.89%	1,979.546
	$\hat{\sigma}_v^2$	-14.06%	9.14%	639.369	-13.07%	9.69%	596.107
	$\hat{\gamma}$	-	-	-	NR	NR	6.598
$\hat{\underline{\theta}}^3_{GLSC}$	$\hat{\sigma}_u^2$	-0.24% ^{NS}	18.96%	1,627.318	0.16% ^{NS}	19.10%	1,665.179
	$\hat{\sigma}_v^2$	3.04%	7.74%	179.725	2.06%	8.21%	183.488
	$\hat{\gamma}$	-	-	-	NR	NR	5.524
$\hat{\underline{\theta}}^4_{GLSC}$ †	$\hat{\sigma}_u^2$	-12.30%	20.80%	2,202.142	-12.28%	20.93%	2,218.344
	$\hat{\sigma}_v^2$	-14.45%	9.92%	692.289	-13.45%	10.47%	648.446
	$\hat{\gamma}$	-	-	-	NR	NR	7.627
$\hat{\underline{\theta}}_{PML}$	$\hat{\sigma}_u^2$	-4.58%	18.28%	1,478.769	-4.32%	18.41%	1,495.924
	$\hat{\sigma}_v^2$	-2.89%	7.40%	148.031	-3.23%	8.05%	175.395
	$\hat{\gamma}$	-	-	-	NR	NR	5.336

^{NS} – denotes 'absolute bias not significantly different from zero at 95% level'.
NR – denotes 'no reported'.
† Numerical minimisations did not achieve convergence for 0.7% of the replicates.

Table 5.13 – Evaluation of $\hat{\underline{\theta}}$ with normally distributed errors ($n^{sim} = 100$, replications generated by UCM model).

When comparing the various evaluated methods, results included in Table 5.13 are very similar to those in Tables 5.9, 5.5 and 5.1. Once more, the estimator GLSC4 has had the largest relative bias among the weighted methods, although in terms of variance results these were very similar to the other methods, when fitting both models.

Table 5.14 includes results for both UCM and AR1 models under normality conditions. The simulated values of Y_{it} are now generated by an UCM-C model.

parameter		UCM model			AR1 model		
		rel bias	$cv(\hat{\theta})$	mse $(\times 1000)$	rel bias	$cv(\hat{\theta})$	mse $(\times 1000)$
$\hat{\underline{\theta}}_{ULS}$	$\hat{\sigma}_u^2$	-11.08%	16.40%	1,748.798	-11.65%	16.49%	1,814.117
	$\hat{\sigma}_v^2$	1.37%	6.62%	116.355	2.16%	8.42%	195.327
	$\hat{\gamma}$	-	-	-	NR ^{NS}	NR	8.638
$\hat{\underline{\theta}}^2_{GLS}$	$\hat{\sigma}_u^2$	-17.58%	17.03%	2,639.260	-17.94%	17.12%	2,706.602
	$\hat{\sigma}_v^2$	-9.61%	8.33%	369.611	-8.17%	8.87%	330.493
	$\hat{\gamma}$	-	-	-	NR ^{NS}	NR	5.740
$\hat{\underline{\theta}}^3_{GLS}$	$\hat{\sigma}_u^2$	-7.75%	17.32%	1,644.813	-8.05%	17.42%	1,675.613
	$\hat{\sigma}_v^2$	7.07%	6.90%	259.431	7.01%	7.48%	280.855
	$\hat{\gamma}$	-	-	-	NR ^{NS}	NR	5.259
$\hat{\underline{\theta}}^4_{GLS}$ †	$\hat{\sigma}_u^2$	-17.38%	18.33%	2,770.673	-17.91%	18.03%	2,814.489
	$\hat{\sigma}_v^2$	-9.80%	8.91%	398.392	-8.16%	9.27%	345.297
	$\hat{\gamma}$	-	-	-	NR ^{NS}	NR	7.176
$\hat{\underline{\theta}}_{ML}$	$\hat{\sigma}_u^2$	-11.08%	16.40%	1,748.798	-11.37%	16.52%	1,792.104
	$\hat{\sigma}_v^2$	1.37%	6.62%	116.355	1.79%	7.20%	141.255
	$\hat{\gamma}$	-	-	-	NR ^{NS}	NR	4.984
$\hat{\underline{\theta}}_{ULSC}$	$\hat{\sigma}_u^2$	-5.46%	17.16%	1,528.424	-5.60%	16.92%	1,494.091
	$\hat{\sigma}_v^2$	-2.93%	7.28%	145.113	-2.76%	8.93%	206.083
	$\hat{\gamma}$	-	-	-	NR ^{NS}	NR	9.785
$\hat{\underline{\theta}}^2_{GLSC}$	$\hat{\sigma}_u^2$	-13.05%	17.77%	2,132.430	-12.92%	17.79%	2,120.730
	$\hat{\sigma}_v^2$	-14.69%	9.18%	687.444	-13.95%	9.51%	648.607
	$\hat{\gamma}$	-	-	-	NR	NR	6.157
$\hat{\underline{\theta}}^3_{GLSC}$	$\hat{\sigma}_u^2$	-1.43%^{NS}	18.36%	1,718.589	-1.19%^{NS}	18.27%	1,706.759
	$\hat{\sigma}_v^2$	3.31%	7.74%	185.940	2.38%	8.19%	188.448
	$\hat{\gamma}$	-	-	-	NR	NR	5.812
$\hat{\underline{\theta}}^4_{GLSC}$ †	$\hat{\sigma}_u^2$	-13.10%	19.67%	2,419.104	-12.96%	19.46%	2,372.202
	$\hat{\sigma}_v^2$	-14.99%	9.83%	730.998	-14.16%	10.06%	682.340
	$\hat{\gamma}$	-	-	-	NR	NR	7.261
$\hat{\underline{\theta}}_{PML}$	$\hat{\sigma}_u^2$	-5.46%	17.16%	1,528.424	-5.21%	17.16%	1,521.837
	$\hat{\sigma}_v^2$	-2.93%	7.28%	145.113	-3.32%	7.73%	165.733
	$\hat{\gamma}$	-	-	-	NR	NR	5.383

^{NS} – denotes 'absolute bias not significantly different from zero at 95% level'.
NR – denotes 'no reported'.
† Numerical minimisations did not achieve convergence for 0.7% of the replicates.

Table 5.14 – Evaluation of $\hat{\underline{\theta}}$ with normally distributed errors ($n^{sim} = 100$, replications generated by UCM-C model).

Reasonably similar conclusions may be drawn when comparing results from Table 5.14 to those in Table 5.13. We may highlight here the performance of the ULS and ML methods when estimating the variance component σ_v^2, with much lower bias than the remaining methods. Furthermore, we may notice that most methods have had an increase in their mse for both fitted models (UCM and AR1), mainly because of a modest inflation in bias. Although we did not expect effects of clustering to lead to increase in bias, we suspect that most estimation methods were more sensitive to data generated by using UCM-C models than to those generated by UCM models, in the current situation with samples of size 100 comparatively to previous sub-sections which have considered larger sample sizes.

Table 5.15 includes results for the unweighted methods when fitting a UCM model considering now a stronger departure from the normality condition (with non-standardised t distribution; see Section 5.2, Sub-section 5.2.2). The values of Y_{it} are generated here by an UCM model, with $D = 200$.

parameter		rel bias	$cv(\hat{\theta})$	mse $(\times 1000)$
$\underline{\hat{\theta}}_{ULS}$	$\hat{\sigma}_u^2$	62.51%	26.20%	26,016.160
	$\hat{\sigma}_v^2$	71.42%	10.14%	13,320.640
$\underline{\hat{\theta}}_{GLS}^2$	$\hat{\sigma}_u^2$	50.36%	26.99%	19,027.960
	$\hat{\sigma}_v^2$	47.67%	10.57%	6,202.951
$\underline{\hat{\theta}}_{GLS}^3$	$\hat{\sigma}_u^2$	67.81%	27.01%	30,260.690
	$\hat{\sigma}_v^2$	87.75%	19.11%	22,156.120
$\underline{\hat{\theta}}_{GLS}^4$	$\hat{\sigma}_u^2$	42.55%	27.42%	15,183.655
	$\hat{\sigma}_v^2$	32.82%	11.97%	3,227.955
$\underline{\hat{\theta}}_{ML}$	$\hat{\sigma}_u^2$	62.51%	26.20%	26,016.160
	$\hat{\sigma}_v^2$	71.42%	10.14%	13,320.630

Table 5.15 – Evaluation of $\underline{\hat{\theta}}$ for UCM model, with $t_{\upsilon=5}$ distributed[29] errors ($n^{sim} = 100$).

In general, all the methods performed not very well for this non-standardised t distribution situation. However, we may still make some brief comparisons. The GLS4, i.e. the ADF method, is the method with the best performance for estimating both components of $\underline{\theta}$. This method has generally lower bias and variance in this situation with a more severe departure from the normality condition.

Table 5.16 includes results for both UCM and AR1 models considering non-normality conditions (t distribution), with simulated values of Y_{it} generated by an UCM model.

[29] With non-standardised variance.

parameter		UCM model rel bias	UCM model $cv(\hat{\theta})$	UCM model mse $(\times 1000)$	AR1 model rel bias	AR1 model $cv(\hat{\theta})$	AR1 model mse $(\times 1000)$
$\underline{\hat{\theta}}_{ULS}$	$\hat{\sigma}_u^2$	-12.66%	24.52%	2,814.730	-12.85%	24.59%	2,839.475
	$\hat{\sigma}_v^2$	1.21%[NS]	11.62%	344.702	1.46%	12.05%	373.731
	$\hat{\gamma}$	-	-	-	NR[NS]	NR	6.873
$\underline{\hat{\theta}}^2_{GLS}$	$\hat{\sigma}_u^2$	-19.26%	25.26%	3,578.367	-19.72%	25.42%	3,664.026
	$\hat{\sigma}_v^2$	-13.06%	11.18%	653.536	-11.53%	11.42%	579.363
	$\hat{\gamma}$	-	-	-	NR[NS]	NR	6.245
$\underline{\hat{\theta}}^3_{GLS}$ ‡*	$\hat{\sigma}_u^2$	-9.25%	24.89%	2,709.491	-9.28%	25.04%	2,738.277
	$\hat{\sigma}_v^2$	9.36%	14.62%	846.390	9.02%	14.77%	840.298
	$\hat{\gamma}$	-	-	-	NR[NS]	NR	5.547
$\underline{\hat{\theta}}^4_{GLS}$ †	$\hat{\sigma}_u^2$	-22.37%	26.72%	4,233.014	-23.06%	26.27%	4,277.594
	$\hat{\sigma}_v^2$	-18.70%	12.62%	1,121.805	-16.76%	12.50%	959.767
	$\hat{\gamma}$	-	-	-	NR[NS]	NR	8.025
$\underline{\hat{\theta}}_{ML}$	$\hat{\sigma}_u^2$	-12.66%	24.52%	2.814.730	-12.78%	24.59%	2,835.910
	$\hat{\sigma}_v^2$	1.21%[NS]	11.62%	344.702	1.61%	12.20%	385.172
	$\hat{\gamma}$	-	-	-	NR[NS]	NR	8.381
$\underline{\hat{\theta}}_{ULSC}$	$\hat{\sigma}_u^2$	-7.43%	24.77%	2,642.143	-7.21%	24.69%	2,623.082
	$\hat{\sigma}_v^2$	-2.86%	12.74%	397.800	-3.16%	13.27%	431.967
	$\hat{\gamma}$	-	-	-	NR	NR	8.132
$\underline{\hat{\theta}}^2_{GLSC}$	$\hat{\sigma}_u^2$	-15.14%	25.89%	3,238.309	-15.13%	25.97%	3,250.858
	$\hat{\sigma}_v^2$	-17.92%	11.63%	1,016.030	-16.96%	11.84%	947.552
	$\hat{\gamma}$	-	-	-	NR	NR	6.844
$\underline{\hat{\theta}}^3_{GLSC}$ ‡	$\hat{\sigma}_u^2$	-3.36%	25.10%	2,726.981	-2.99%	25.13%	2,744.175
	$\hat{\sigma}_v^2$	5.88%	16.15%	806.333	4.88%	16.65%	810.591
	$\hat{\gamma}$	-	-	-	NR	NR	6.468
$\underline{\hat{\theta}}^4_{GLSC}$	$\hat{\sigma}_u^2$	-18.99%	27.06%	3,826.830	-19.39%	27.24%	3,902.797
	$\hat{\sigma}_v^2$	-23.52%	13.02%	1,608.108	-22.14%	13.25%	1,470.509
	$\hat{\gamma}$	-	-	-	NR	NR	8.779
$\underline{\hat{\theta}}_{PML}$	$\hat{\sigma}_u^2$	-7.43%	24.77%	2,642.143	-7.05%	24.72%	2,628.038
	$\hat{\sigma}_v^2$	-2.86%	12.74%	397.800	-2.91%	14.59%	515.340
	$\hat{\gamma}$	-	-	-	NR	NR	9.056

NS – denotes 'absolute bias not significantly different from zero at 95% level'.
NR – denotes 'no reported'.

† Numerical minimisations did not achieve convergence for 1.0% of the replicates.
‡ Numerical minimisations did not achieve convergence for 0.3% of the replicates.

* $D = 299$, as 1 (0.3%) of the estimates was inadmissible.

Table 5.16 – Evaluation of $\underline{\hat{\theta}}$ with $t_{\upsilon=5}(0,1)$ distributed errors ($n^{sim} = 100$, replications generated by UCM model).

When comparing Table 5.16 with Table 5.14, we may notice that all methods have had a large increase in their mse for both fitted models (UCM and AR1), caused in most cases by an inflation in bias and a reasonably large increase in variance. Dissimilarly to results in Tables 5.7 and 5.8, but similarly to results included in Table 5.11, results in Table 5.16 suggest that the ADF methods have generally not performed better than the other methods. Nevertheless, we may highlight here that the GLS4 method had in the considered situation much lower c.v. results than the GLS3 method for estimating the variance component σ_v^2.

Table 5.17 includes results for both UCM and AR1 models considering non-normality conditions (t distribution), with simulated values of Y_{it} generated by an UCM-C model.

parameter		UCM model			AR1 model		
		rel bias	$cv(\hat{\theta})$	mse $(\times 1000)$	rel bias	$cv(\hat{\theta})$	mse $(\times 1000)$
$\underline{\hat{\theta}}_{ULS}$	$\hat{\sigma}_u^2$	-10.33%	28.91%	4,060.477	-10.76%	29.15%	4,133.001
	$\hat{\sigma}_v^2$	1.56%	10.84%	307.027	2.18%	11.41%	349.009
	$\hat{\gamma}$	-	-	-	NR ^{NS}	NR	6.956
$\underline{\hat{\theta}}_{GLS}^2$	$\hat{\sigma}_u^2$	-16.73%	29.27%	4,558.109	-17.28%	29.45%	4,651.307
	$\hat{\sigma}_v^2$	-12.30%	10.98%	605.528	-10.64%	11.30%	533.901
	$\hat{\gamma}$	-	-	-	NR	NR	6.297
$\underline{\hat{\theta}}_{GLS}^3$ ‡	$\hat{\sigma}_u^2$	-7.11%	29.26%	4,115.177	-7.60%	29.54%	4,186.547
	$\hat{\sigma}_v^2$	9.45%	14.00%	804.026	9.80%	14.76%	889.899
	$\hat{\gamma}$	-	-	-	NR	NR	6.084
$\underline{\hat{\theta}}_{GLS}^4$ †	$\hat{\sigma}_u^2$	-21.82%	29.11%	5,181.865	-22.30%	29.07%	5,253.813
	$\hat{\sigma}_v^2$	-17.18%	11.74%	966.778	-15.43%	12.10%	850.643
	$\hat{\gamma}$	-	-	-	NR ^{NS}	NR	6.595
$\underline{\hat{\theta}}_{ML}$	$\hat{\sigma}_u^2$	-10.33%	28.91%	4,060.477	-10.80%	29.12%	4,125.330
	$\hat{\sigma}_v^2$	1.56%	10.84%	307.027	2.58%	11.84%	382.492
	$\hat{\gamma}$	-	-	-	NR	NR	8.711
$\underline{\hat{\theta}}_{ULSC}$	$\hat{\sigma}_u^2$	-5.12%	31.21%	4,709.861	-5.23%	31.44%	4,771.979
	$\hat{\sigma}_v^2$	-2.41%	12.27%	370.143	-2.26%	13.05%	416.320
	$\hat{\gamma}$	-	-	-	NR ^{NS}	NR	7.573
$\underline{\hat{\theta}}_{GLSC}^2$	$\hat{\sigma}_u^2$	-12.92%	31.62%	4,824.704	-12.97%	31.84%	4,879.829
	$\hat{\sigma}_v^2$	-17.39%	11.94%	991.670	-16.45%	12.28%	932.715
	$\hat{\gamma}$	-	-	-	NR ^{NS}	NR	6.283
$\underline{\hat{\theta}}_{GLSC}^3$ ‡	$\hat{\sigma}_u^2$	-1.22%^{NS}	31.52%	5,062.768	-1.36%^{NS}	31.83%	5,148.707
	$\hat{\sigma}_v^2$	6.46%	18.04%	1,018.290	6.32%	20.10%	1,231.931
	$\hat{\gamma}$	-	-	-	NR ^{NS}	NR	6.357
$\underline{\hat{\theta}}_{GLSC}^4$ †	$\hat{\sigma}_u^2$	-18.80%	30.92%	5,129.531	-18.69%	30.76%	5,083.114
	$\hat{\sigma}_v^2$	-22.40%	12.81%	1,489.880	-21.30%	12.90%	1,381.160
	$\hat{\gamma}$	-	-	-	NR	NR	6.923
$\underline{\hat{\theta}}_{PML}$	$\hat{\sigma}_u^2$	-5.12%	31.21%	4,709.861	-5.17%	31.35%	4,747.471
	$\hat{\sigma}_v^2$	-2.41%	12.27%	370.143	-2.31%	13.07%	417.936
	$\hat{\gamma}$	-	-	-	NR ^{NS}	NR	5.348

^{NS} – denotes 'absolute bias not significantly different from zero at 95% level'.
NR – denotes 'no reported'.

† Numerical minimisations did not achieve convergence for 1.0% of the replicates.
‡ Numerical minimisations did not achieve convergence for 0.3% of the replicates.

Table 5.17 – Evaluation of $\underline{\hat{\theta}}$ with $t_{\upsilon=5}(0,1)$ distributed errors ($n^{sim} = 100$, replications generated by UCM-C model).

By looking across each row of Table 5.17 we can notice that, when fitting UCM models ML and ULS are the methods with lower mse for estimating both variance components σ_u^2 and σ_v^2 among the unweighted methods, while PML and ULSC are the methods with lower mse for estimating also both σ_u^2 and σ_v^2 among the weighted methods. When fitting AR1 models, ML is the method with lower mse for estimating σ_u^2 and γ, and ULS is the method with lower mse for estimating σ_v^2 among the unweighted methods, while PML is the method with lower mse for estimating σ_u^2 and γ, and ULSC is the method with lower mse for estimating σ_v^2 among the unweighted methods. Note that now with a sample of size 100, ADF methods have not performed better than the other methods in most situations, although that method is the one with the lowest variance some circumstances. We may highlight again that the GLS4 method had much lower c.v. results than the GLS3 method for estimating the variance component σ_v^2.

When comparing results presented in Table 5.17 with those presented in Tables 5.13 and 5.14, we may notice that all methods have had a large increase in their mse for both fitted models (UCM and AR1), generally because of a modest increase in bias and also a reasonably large inflation of variance. Moreover, when comparing Table 5.17 with Table 5.16, we may notice most methods have also had reasonably large increase in their mse for both fitted models (UCM and AR1), caused in general by a large inflation in variance.

Moreover, we may observe that tables presented in the current sub-section have all presented an increase in both bias and variance when compared to those included in the three previous sub-sections, probably caused by a reduction in the number of cases to 100 units.

5.4 Further discussion

The current section summarizes in an alternative manner some of the results already shown in the previous section. It is our aim here to further evaluate the general behaviour of each of estimation methods summarised in Section 5.3 for different (*i*) sample sizes (1340, 500, 200, and 100), (*ii*) whether or not under normality condition, (*iii*) model fitting complexities (UCM and AR1), and (*iv*) model adopted for generating data (whether or not allowing for clustering).

In Figure 5.1, for simplicity, we shall focus our attention on unweighted estimators relative bias results for σ_u^2 and σ_v^2 (denoted by "Sigma2u" and "Sigma2v" respectively in the figure) estimation, when fitting AR1 models. Note that in the figure's legend, "Normal" shall mean normal simulated values of Y_{it} generated via UCM model, while "t" shall mean t distributed simulated values of Y_{it} generated via UCM model.

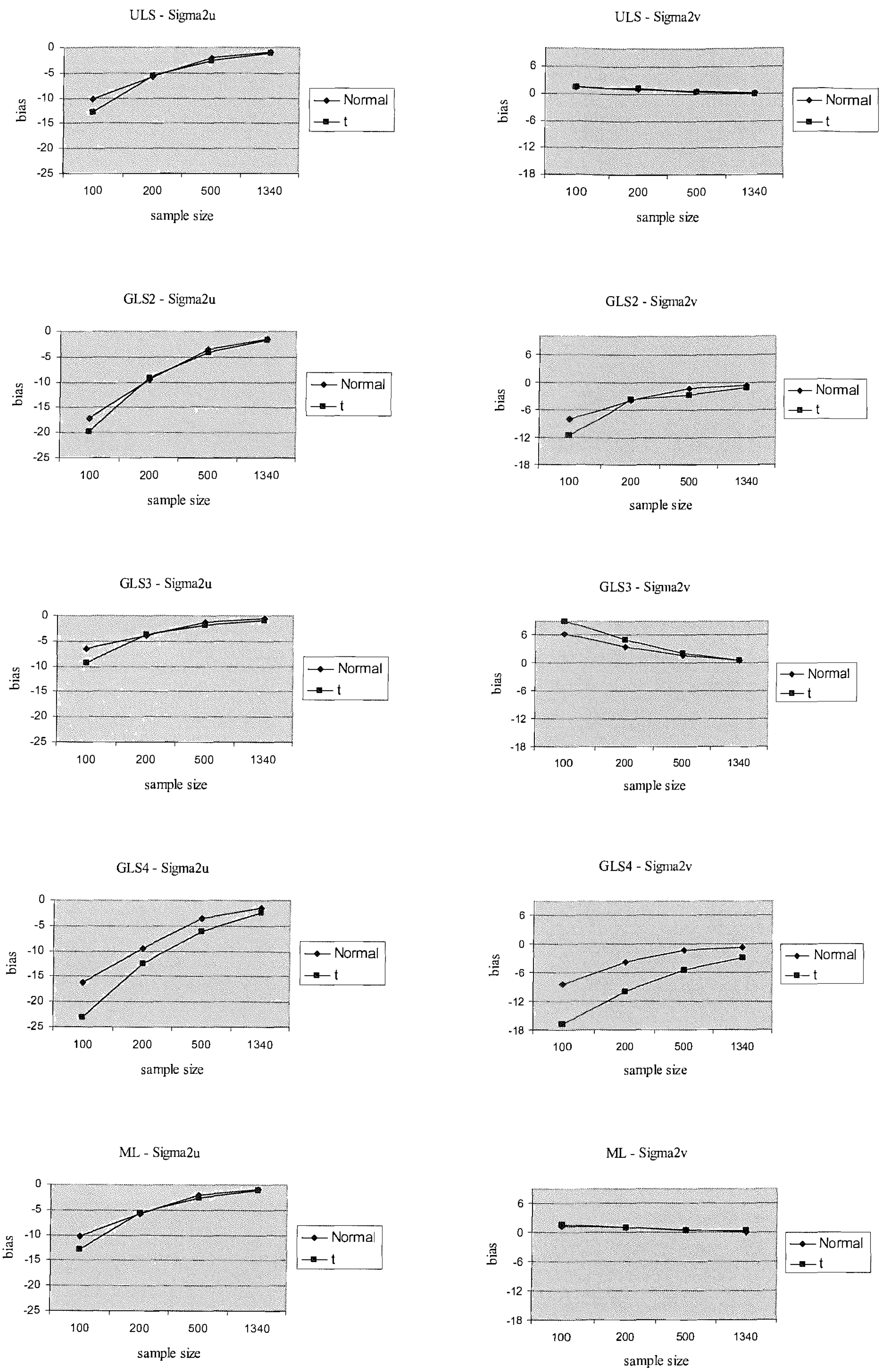

Figure 5.1 – Relative bias for σ_u^2 and σ_v^2 estimation when fitting an AR1 model.

By looking at Figure 5.1 above we may make several observations. In general all methods perform better for normal than for t distributed simulated values of Y_{it}, when the sample size is fixed. Both ULS and ML, and somewhat GLS3, when considering either distributions, have a positive bias which is very close to zero even for the smallest considered sample size when estimating σ_v^2 , while these methods have a negative relative bias, which gets near zero for samples sizes of size at least 200, when estimating σ_u^2 . Methods GLS2 and GLS4, which is the ADF unweighted method, have negative bias for estimating both variance components σ_u^2 and σ_v^2 that gets close to zero for samples sizes larger or equal to 500 when considering either distributions. All but the GLS4 method seem not to be much affected by the considered departure from the normality assumption, especially when considering samples of size 200 or more. Curiously the ADF GLS4 method needed the largest considered sample size, i.e. $n = 1340$, for the results produced under both considered distributions to become similar, even though we might expect that method to behave better under departures from normality conditions when compared to the remaining ones.

Table 5.18 includes information on how sensitive each method is for departures of the normality assumptions and sample size reductions, now with data generated by a UCM-C model, which allows for effects of clustering.

parameter		$\text{mse}_{t_{\upsilon=5}(0,1)}/\text{mse}_{N(0,1)}$ with $n^{sim}=100$		$\text{mse}_{n^{sim}=100}/\text{mse}_{n^{sim}=500}$ with $N(0,1)$	
		UCM model	AR1 model	UCM model	AR1 model
$\underline{\hat{\theta}}_{ULS}$	$\hat{\sigma}_u^2$	2.32	2.28	5.83	5.99
	$\hat{\sigma}_v^2$	2.64	1.79	4.73	6.31
	$\hat{\gamma}$	-	0.81	-	5.47
$\underline{\hat{\theta}}^2_{GLS}$	$\hat{\sigma}_u^2$	1.73	1.72	7.90	8.03
	$\hat{\sigma}_v^2$	1.64	1.62	10.88	9.30
	$\hat{\gamma}$	-	1.10	-	5.68
$\underline{\hat{\theta}}^3_{GLS}$	$\hat{\sigma}_u^2$	2.50	2.50	5.55	5.60
	$\hat{\sigma}_v^2$	3.10	3.17	8.56	8.30
	$\hat{\gamma}$	-	1.16	-	5.46
$\underline{\hat{\theta}}^4_{GLS}$	$\hat{\sigma}_u^2$	1.87	1.87	8.17	8.21
	$\hat{\sigma}_v^2$	2.43	2.46	11.36	9.44
	$\hat{\gamma}$	-	0.92	-	6.82
$\underline{\hat{\theta}}_{ML}$	$\hat{\sigma}_u^2$	2.32	2.30	5.83	5.92
	$\hat{\sigma}_v^2$	2.64	2.71	4.73	4.94
	$\hat{\gamma}$	-	1.75	-	5.17
$\underline{\hat{\theta}}_{ULSC}$	$\hat{\sigma}_u^2$	3.08	3.19	4.72	4.57
	$\hat{\sigma}_v^2$	2.55	2.02	5.32	5.98
	$\hat{\gamma}$	-	0.77	-	5.57
$\underline{\hat{\theta}}^2_{GLSC}$	$\hat{\sigma}_u^2$	2.26	2.30	6.16	6.10
	$\hat{\sigma}_v^2$	1.44	1.44	14.00	12.50
	$\hat{\gamma}$	-	1.02	-	5.40
$\underline{\hat{\theta}}^3_{GLSC}$	$\hat{\sigma}_u^2$	2.95	3.02	5.21	5.12
	$\hat{\sigma}_v^2$	5.48	6.54	6.33	5.61
	$\hat{\gamma}$	-	1.09	-	5.29
$\underline{\hat{\theta}}^4_{GLSC}$	$\hat{\sigma}_u^2$	2.12	2.14	6.66	6.50
	$\hat{\sigma}_v^2$	2.04	2.02	13.72	12.26
	$\hat{\gamma}$	-	0.95	-	6.01
$\underline{\hat{\theta}}_{PML}$	$\hat{\sigma}_u^2$	3.08	3.12	4.72	4.66
	$\hat{\sigma}_v^2$	2.55	2.52	5.32	5.11
	$\hat{\gamma}$	-	0.99	-	4.92

Table 5.18 – Comparative evaluation of the considered estimation methods.

Interestingly, results included in Table 5.18 above indicate effects of non-normality on all the evaluated methods, as a result of the increases in the variance of the estimators. Note that results presented above are samples of size 100. Results for samples of larger sizes were somewhat similar. Moreover: (*i*) among the unweighted methods, GLS2 and GLS4 are the methods less sensitive to departures from the normal distribution assumption when fitting a UCM model; (*ii*) among the weighted methods, GLSC2 and GLSC4 are the methods less sensitive to departures from the normal distribution assumption when fitting a UCM model; (*iii*) among the unweighted methods, ULS and GLS2 are the methods less sensitive to departures from the normal distribution assumption when fitting a AR1 model; (*iv*) among the weighted methods, GLSC4, GLSC2 and ULSC are the methods less sensitive to departures from the normal distribution assumption when fitting a AR1 model.

Regarding sample size impacts, we may notice from Table 5.18 that when fitting both UCM and AR1 models, GLS2 and GLS4 are in general the methods more sensitive to sample size reductions among the unweighted methods, while GLS2C and GLS4C are the more sensitive ones among the weighed methods.

Table 5.19 provides information on the performance of each of the evaluated methods when applied for clustered generated data compared to non-clustered data, when considering $n^{sim}=100$ and t distributed errors.

unweighted method		UCM model	AR1 model	weighted method		UCM model	AR1 model
$\hat{\underline{\theta}}_{ULS}$	$\hat{\sigma}_u^2$	1.44	1.46	$\hat{\underline{\theta}}_{ULSC}$	$\hat{\sigma}_u^2$	1.78	1.82
	$\hat{\sigma}_v^2$	0.89	0.93		$\hat{\sigma}_v^2$	0.93	0.96
	$\hat{\gamma}$	-	1.01		$\hat{\gamma}$	-	0.93
$\hat{\underline{\theta}}^2_{GLS}$	$\hat{\sigma}_u^2$	1.27	1.27	$\hat{\underline{\theta}}^2_{GLSC}$	$\hat{\sigma}_u^2$	1.49	1.50
	$\hat{\sigma}_v^2$	0.93	0.92		$\hat{\sigma}_v^2$	0.98	0.98
	$\hat{\gamma}$	-	1.01		$\hat{\gamma}$	-	0.92
$\hat{\underline{\theta}}^3_{GLS}$	$\hat{\sigma}_u^2$	1.52	1.53	$\hat{\underline{\theta}}^3_{GLSC}$	$\hat{\sigma}_u^2$	1.86	1.88
	$\hat{\sigma}_v^2$	0.95	1.06		$\hat{\sigma}_v^2$	1.26	1.52
	$\hat{\gamma}$	-	1.10		$\hat{\gamma}$	-	0.98
$\hat{\underline{\theta}}^4_{GLS}$	$\hat{\sigma}_u^2$	1.22	1.23	$\hat{\underline{\theta}}^4_{GLSC}$	$\hat{\sigma}_u^2$	1.34	1.30
	$\hat{\sigma}_v^2$	0.86	0.89		$\hat{\sigma}_v^2$	0.93	0.94
	$\hat{\gamma}$	-	0.82		$\hat{\gamma}$	-	0.79
$\hat{\underline{\theta}}_{ML}$	$\hat{\sigma}_u^2$	1.44	1.45	$\hat{\underline{\theta}}_{PML}$	$\hat{\sigma}_u^2$	1.78	1.81
	$\hat{\sigma}_v^2$	0.89	0.99		$\hat{\sigma}_v^2$	0.93	0.81
	$\hat{\gamma}$	-	1.04		$\hat{\gamma}$	-	0.59

Table 5.19 – Further comparative evaluation of the considered estimation methods, $\mathrm{mse}_{\text{UCM-C Data}}/\mathrm{mse}_{\text{UCM Data}}$.

By looking across each row of Table 5.19 we can notice that in most cases there is overall a modest increase in the mean square error when comparing clustered data results (generated by UCM-C model) with non-clustered data ones (generated by UCM model). In particular, we may observe that the ADF methods, i.e. GLS4 and GLSC4, are among the considered methods the ones which suffer the lesser impact from clustering.

5.5 Concluding remarks

We have presented in this chapter the characteristics and the main results of an extensive simulation study which has mainly the aim of evaluating the estimation methods we propose in Chapter 4. In Section 5.3, methods were compared mainly in terms of their bias and variance, for different types of data (clustered and non-clustered), different distributions (normal and t), considering various sample sizes (1340, 500, 200, and 100). Some additional observations were drawn in Section 5.4, in order to further investigate the impacts of departures from normality conditions, of clustering and of sample sizes.

Additionally to all the interpretation already included in this chapter, the main conclusion we may draw from the results is that overall most of the proposed methods have a reasonably good performance in terms of bias and variance when compared to the classical methods. As expected, weighted estimation methods had larger variance than the unweighted methods especially because, considering the current simulation setup, there would be no need for utilising weighted estimators for bias correction. We may recapitulate here that we do not allow for the impact of either stratification or unequal probability sampling in this simulation study. Moreover, our results indicated that weighted methods appear to have slightly lower bias when compared to the weighted ones, although there is not any clear theoretical reason for this result, as we have stated earlier.

Regarding the ADF method GLS4, this has overall not always performed as we expected when dealing with t distributed data, although very often these were at least the most efficient methods in some situations and generally less sensitive to clustered data. These methods have behaved particularly better for samples size 500 and 200 relatively to the remaining estimation methods when considering non-normality conditions particularly with simulated Y_{it} values generated by an UCM-C model, which allows for clustering. However, in general our results agree with Bollen (1989, p. 432), Satorra (1992), and Yuan and Bentler (1997b), for example, and we would therefore also recommend that ADF methods should be used carefully in situations where only samples of small size are available. Yuan and Bentler (1997b) considers that in some situations $n = 1000$, for example, may not be considered large enough. Olsson, Foss, and Troye (2003) have also reported that ADF methods should be considered reliable for samples larger than 1000, when working with simple models, and larger than 5000, when working with more complex models.

Nevertheless we may highlight here the performance of the GLS4 estimation method when considering now non-standardised t distributed data, generated by an UCM model (see

Table 5.15). That method was in that situation with a stronger departure from the normality conditions is the method with the best performance for estimating $\underline{\theta}$, with the smallest values for relative bias and coefficient of variation.

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. For fully understanding the reasons for that method not to perform similarly to the other proposed methods we may need to undertake further investigations in the future.

The ML (and PML) point estimators have in general produced relatively 'good' performance in terms of bias and variance, even in situations where the normality assumption was violated, as reported for example by Satorra and Bentler (1986). Furthermore, we would also recommend that if there is enough evidence of normality, and specially if the sample size is 'too small', that normal theory methods should be adopted.

A Shapiro-Wilk normality test (Shapiro and Wilk, 1965) has been adopted for evaluating the univariate distribution of each component of the estimators considered in the present simulation study. That test did not reject the null hypothesis of normality at the 95% level for most of the estimators in most of the situations considered, even for samples of size 100, especially in circumstances where simulated values of Y_{it} are generated under normality conditions.

We shall consider in the following chapter, methods for $\hat{\underline{\theta}}$'s variance estimation. We shall discuss existing variance estimation methods and propose alternative methods which are able to handle complex survey data. We shall also consider testing techniques for structural models for covariance matrices.

Chapter 6

Variance estimation for $\hat{\underline{\theta}}$ and covariance structure model evaluation

6.1 Introduction

Recall that $\mho$ denotes a finite population, which is fixed on occasions $1,\ldots,T$, with size N. Let $\underline{Y}_i = (\underline{Y}_{i1},\ldots,\underline{Y}_{iT})'$ be a random vector containing T repeated observations on a variable for unit $i = 1,2,\ldots,N$, over the T waves of the longitudinal survey.

We shall make in the present chapter the same assumptions described in Chapter 4, Section 4.1. Furthermore, we shall consider that the fitting function $F(\underline{\theta})$ satisfies the conditions enumerated in Chapter 4, Section 4.4, where $\underline{\theta}$ is a $b \times 1$ parameter vector, introduced in Chapter 4, Section 4.1.

It is our main aim in this chapter to discuss methods for variance estimation for generalised least squares (GLS; see Chapter 4, sub-sections 4.4.2 and 4.4.3) and (pseudo) maximum likelihood (ML and PML; see Chapter 4, subsections 4.4.4 and 4.4.5) point estimators of $\underline{\theta}$. We shall consider here the covariance structure model given by (4.3), also in Chapter 4, Section 4.1.

Additionally, testing techniques for structural models for covariance matrices shall also be reviewed in the current chapter in the classical independent and identically distributed (iid) observations context, while we shall discuss model fitting statistics for those models in a complex survey design framework.

As in previous chapters of this thesis, we shall be aiming to study methods under the survey sampling approach for covariance structure modelling of complex survey data (aggregate modelling). We shall not be considering here methods for disaggregated covariance structure modelling, i.e. a multilevel modelling approach for covariance models (for current developments considering this approach see, for example, Lee, 1990; Muthén and Satorra, 1995, Section 5; Muthén, 1997; Lee and Poon, 1998; Hox and Maas, 2001; and more recently Yuan and Hayashi, 2005). We may also acknowledge here that Yuan and Bentler (1998) have developed a structural equation modelling estimation method that is robust to the presence of outliers, although those shall not be considered here.

The purpose of the present chapter is then two-fold. We shall review variance estimation methods (Section 6.2) and model fitting statistics (Section 6.4, Sub-section 6.4.1), under the classical approach. Furthermore, we shall also provide some new developments on both variance estimation (Section 6.3) and model fitting statistics when working under the complex sampling approach (Section 6.4, Sub-section 6.4.2). Some very brief concluding remarks shall be presented in Section 6.5.

6.2 Variance estimation for $\hat{\underline{\theta}}$ - Classical case

We shall review in this section variance estimation methods which rely on the assumption that the data are obtained by simple random sampling (srs) from large populations. We shall consider variance estimation only for the point estimation methods that are built upon the same presupposition, i.e. generalised least squares (GLS; see Chapter 4, Sub-section 4.4.2) and maximum likelihood (ML; see Chapter 4, Sub-section 4.4.4). Moreover, we shall initially assume that $\underline{Y}_i$ is multivariate normally distributed.

Consider first the $\hat{\underline{\theta}}_{GLS}$, a generalised least squares estimator for the parameter vector $\underline{\theta}$, that may be obtained by minimizing the GLS fitting function $F(\underline{\theta})_{GLS}$ given by (4.46) and (4.47) in Chapter 4, Section 4.4, Subsection 4.4.2. Let S be the unweighted sample covariance matrix, given by expression (4.18) in Chapter 4, Section 4.3, Sub-section 4.3.1. According to Bentler and Weeks (1980), Swain (1975), and Bollen (1989, p. 114), for example, an asymptotic covariance matrix of $\hat{\underline{\theta}}_{GLS}$ is given by

$$a\text{COV}\left(\hat{\underline{\theta}}_{GLS}\right) = \left(\frac{2}{n}\right) \cdot \left[E\left(\frac{\partial^2 F(\underline{\theta})_{GLS}}{\partial \underline{\theta}\, \partial \underline{\theta}'} \right) \right]^{-1}, \qquad (6.01)$$

when U in (4.46), from Chapter 4, Section 4.4, Subsection 4.4.2, is given by (4.24), from Chapter 4, Section 4.3, Sub-section 4.3.2, with

$$p \lim_{n \to \infty} \mathrm{W}^{-1} = \tau\, \Sigma^{-1}, \qquad (6.02)$$

where n is the sample size, W is any consistent estimator of Σ (such as S), and τ is any constant (typically $\tau = 1$), as defined in Chapter 4. Note that in (6.01), the matrix in brackets is the information matrix.

We may substitute $\hat{\underline{\theta}}_{GLS}$ for $\underline{\theta}$ in (6.01) in order to calculate $a\text{cov}(\hat{\underline{\theta}}_{GLS})$, which is an estimated asymptotic covariance matrix for $\hat{\underline{\theta}}_{GLS}$. Note that the matrix $a\text{cov}(\hat{\underline{\theta}}_{GLS})$ has the estimated asymptotic variances of $\hat{\underline{\theta}}_{GLS}$ down its main diagonal and the estimated covariances in its off-diagonal elements.

Similarly, Bollen (1989, Appendix 4B) shows that the asymptotic covariance matrix of $\hat{\underline{\theta}}_{ML}$ is also given by (6.01), substituting $\hat{\underline{\theta}}_{ML}$ for $\hat{\underline{\theta}}_{GLS}$, and $F(\underline{\theta})_{ML}$ for $F(\underline{\theta})_{GLS}$, where $F(\underline{\theta})_{ML}$ is a ML fitting function. We may also substitute $\hat{\underline{\theta}}_{ML}$ for $\underline{\theta}$ in (6.01) in order to calculate $a\text{cov}(\hat{\underline{\theta}}_{ML})$, which is an estimator for the asymptotic covariance matrix of $\hat{\underline{\theta}}_{ML}$. See also Bollen (1989, p. 107-109), Jöreskog (1970), Swain (1975), Bentler and Weeks (1980), and Matsueda and Bielby (1986), for additional information.

Formulas for the asymptotic covariance matrix under non-normality have also been developed. According to Bollen (1989), when the multivariate normality of $\underline{Y}_i$ is violated the consistency of the ML, GLS and ULS estimators is not affected. However, it may lead to the estimated asymptotic covariance matrix discussed above being inconsistent.

Let (Browne, 1982; Browne, 1984)

$$\Delta = \frac{\partial\{vech[\Sigma(\underline{\theta})]\}}{\partial\underline{\theta}}, \tag{6.03}$$

be a $k \times b$ matrix of partial derivatives of elements of $vech[\Sigma(\underline{\theta})]$ with respect to the elements of $\underline{\theta}$, i.e. the Jacobian matrix of $vech[\Sigma(\underline{\theta})]$, with columns

$$\Delta_{\mathrm{j}} = \frac{\partial\{vech[\Sigma(\underline{\theta})]\}}{\partial\theta_{\mathrm{j}}}$$

with $\mathrm{j} = 1,\ldots,b$, where b is the dimension of population parameter vector $\underline{\theta}$, and the $k \times 1$ vector $vech[\Sigma(\underline{\theta})]$ and $k = T(T+1)/2$ are as introduced in Chapter 4, Section 4.1, with T being as defined in Section 6.1. See also Muthén and Satorra (1995).

Under fairly general conditions, from the central limit theorem, Jöreskog and Goldberger (1972) prove that

$$\sqrt{n} \cdot \{vech[\mathrm{S}] - vech[\Sigma(\underline{\theta})]\} \to_L \mathrm{N}_k(0,\mathrm{C}) \tag{6.04}$$

as $n \to \infty$, where $\to_L$ denotes convergence in distribution, and $\mathrm{N}_k(0,\mathrm{C})$ denotes a k-dimensional multivariate normal with zero mean vector, and a $k \times k$ nonnegative definite variance-covariance matrix $\mathrm{C} = \mathrm{VAR}\{\sqrt{n} \cdot vech[\mathrm{S}]\}$, introduced in Chapter 4, Section 4.3, Sub-section 4.3.2, given by (4.19) when working under normality assumptions. Fuller (1987,

Appendix 4.B, Theorem 4.B.4) also proves result (6.04). For further discussion on this subject, see Jöreskog and Goldberger (1972), Swain (1975), Browne (1984), Satorra and Bentler (1986), Shapiro (1986), Browne (1987), Satorra and Bentler (1988), Le (1990), Satorra (1989), Chou, Bentler and Satorra (1991), Satorra and Bentler (1994), Bentler and Dudgeon (1996), Satorra (1992), and Yuan and Bentler (1998). For some material on convergence in distributions see, for example, Bollen (1989, Appendix B).

From (6.04), if we adopt $\hat{\underline{\theta}}_{GLS}$, obtained by minimizing (4.46) from Chapter 4, Section 4.4, Sub-section 4.4.4, for estimating $\underline{\theta}$, we would have the following asymptotic $b \times b$ covariance matrix of $\hat{\underline{\theta}}$ (Browne, 1984; see also Skinner, 1989a; Satorra and Bentler, 1988; Satorra, 1989; Chou, Bentler and Satorra, 1991; Satorra, 1992; Satorra and Bentler, 1994; Muthén and Satorra, 1995; and Bentler and Dudgeon, 1996)

$$a\mathrm{COV}\left(\hat{\underline{\theta}}\right) = n^{-1}\left(\Delta'\mathrm{U}^{-1}\Delta\right)^{-1}\Delta'\mathrm{U}^{-1}\mathrm{C}\,\mathrm{U}^{-1}\Delta\left(\Delta'\mathrm{U}^{-1}\Delta\right)^{-1}, \tag{6.05}$$

for a finite sample size n, where U is the $k \times k$ weight matrix first defined in Chapter 4, Section 4.3, Sub-section 4.3.2, assuming that (Browne, 1984): (*i*) $\hat{\underline{\theta}}_{GLS}$ is a consistent estimator of $\underline{\theta}$; (*ii*) Δ and $\Sigma(\underline{\theta})$ are continuous functions of $\underline{\theta}$; (*iii*) Δ is of full rank b; and (*iv*) systematic errors caused by lack of fit of the model to the population variance covariance matrix are not large relative to random sampling errors in S.

When $\Delta'\mathrm{U}^{-1}\mathrm{C}\,\mathrm{U}^{-1}\Delta$ is asymptotically equal to $\Delta'\mathrm{U}^{-1}\Delta$, i.e. when U is consistent for C, (6.05) simplifies to (Browne, 1984; see also Satorra, 1989; Satorra, 1992; Muthén and Satorra, 1995; and Bentler and Dudgeon, 1996)

$$a\mathrm{COV}\left(\hat{\underline{\theta}}\right) = n^{-1}\left(\Delta'\mathrm{U}^{-1}\Delta\right)^{-1}, \tag{6.06}$$

which holds for both $\hat{\underline{\theta}}_{GLS}$ and for asymptotically distribution-free (ADF) methods estimates (see Muthén and Satorra, 1995), discussed earlier in Chapter 4. See also Fuller (1987, Sub-section 4.2.1, Theorem 4.2.1).

For estimating $a\mathrm{COV}\left(\hat{\underline{\theta}}\right)$ we need to evaluate $\Delta = \Delta(\underline{\theta})$ at $\hat{\underline{\theta}}$. Furthermore, we have to estimate the non-singular matrix C, which could be estimated considering a variety of ways. If the sample is selected from a normally distributed population, we may have

$$\hat{\mathrm{C}} = \mathrm{U} = 2 \cdot \mathrm{K}'(\mathrm{W} \otimes \mathrm{W})\mathrm{K}, \tag{6.07}$$

or equivalently (4.24) from Chapter 4, Section 4.3, Sub-section 4.3.2, as suggested by Browne (1987), Shapiro (1986) and Muthén and Satorra (1995), for example. In (6.07), as defined in Chapter 4, Section 4.4, Sub-section 4.4.2, K is a $k \times T^2$ transition matrix, $\otimes$ is the operator for the right Kronecker product, and W is any consistent estimator of Σ, such as S. See Chapter 4, Sub-section 4.4.2.

We may observe from (6.05) that we do not necessarily need to have $\hat{C} = U$. Therefore, when estimating normal theory parameter point estimates we could estimate their variances by choosing a $\hat{C}$ appropriate for situations where non-normality occurs (Satorra, 1989; Muthén and Satorra, 1995). Recall that under the classical set up, $\mathrm{VAR}\{vech[\mathrm{S}]\}$ may be estimated by (4.33) from Chapter 4, Section 4.3, Sub-section 4.3.2 (see also Skinner, 1989a).

According to Muthén and Satorra (1995), the adoption of (4.33) for estimating C leads to asymptotic optimal (again in the sense of leading to efficient estimators, as in Chapter 4, Sub-section 4.4.3) estimates for $a\mathrm{COV}(\hat{\underline{\theta}})$, i.e. robust standard errors, for any distribution of $\underline{Y}_i$, where $\underline{Y}_i$ is as defined in Section 6.1, when plugging (4.33) into (6.05).

According to Browne (1987), violation of multivariate normality of $\underline{Y}_i$ can severely invalidate the use of (6.06) for estimating the variance of the estimators. The greater is the kurtosis of the distribution of $\underline{Y}_i$, the stronger the previous statement is. Notice, however, that if $\underline{Y}_i$ has a distribution with null fourth order cumulant, or no kurtosis, then we would not face the problems described above. Note, however, that Satorra (2002) has shown that for a broad class of linear-latent variable models, under certain model and independence assumptions, normal theory point estimators and normal theory associated variance estimators are valid, even with non-normally distributed data.

We also acknowledge that Yuan and Bentler (1997b) have proposed a correction to the estimator of the asymptotic covariance of the ADF point estimator (especially for situations with small sample sizes), although that shall not be considered here.

6.3 Variance estimation for $\hat{\underline{\theta}}$ under the complex survey approach

In the current sub-section, we discuss methods for variance estimation of $\hat{\underline{\theta}}$ assuming that the sample is selected under the complex survey design approach. Under this circumstance we may not assume that the observations are iid.

Let us assume that

$$\sqrt{n} \cdot \{vech[\mathrm{S}_w] - vech[\Sigma(\underline{\theta})]\} \rightarrow_L \mathrm{N}_k(0, \mathrm{C}_c) \qquad (6.08)$$

also holds, with $C_c = \mathrm{VAR}\{\sqrt{n} \cdot vech[S_w]\}$, introduced in Chapter 4, Section 4.3, Sub-section 4.3.2. In (6.08), $vech[S_w]$ denotes a $k \times 1$ vector formed from the nonduplicated elements of S_w, which is the weighted sample covariance matrix, as defined in Chapter 4, Sub-section 4.3.1, recalling that $E(E_p(S_w)) \doteq E(S_N) = \Sigma$, where S_N is the finite population covariance matrix given by expression (4.12).

Observe that in situations where the sample has been selected with unequal selection probabilities and clustering, (6.07) and (4.33) are not appropriate for estimating C_c. Skinner and Holmes (2003) suggest the adoption of

$$a\mathrm{COV}(\hat{\underline{\theta}}) = n^{-1}(\Delta' U^{-1} \Delta)^{-1} \Delta' U^{-1} C_c\, U^{-1} \Delta (\Delta' U^{-1} \Delta)^{-1} \quad (6.09a)$$

for calculating $a\,\mathrm{cov}(\hat{\underline{\theta}})$, with a choice of estimator for C_c that takes the complex sampling design into account. As previously adopted in this thesis for a stratified multistage scheme, let H be the number of stratum in the sample, m_h the sample number of PSUs in strata h, and n_{hj} the sample number of individual size in PSU j in stratum h. When considering a multistage stratified sampling design (see Example 2.2, Chapter 2), we propose that C_c could be estimated by adopting (see Chapter 4, Section 4.3, Sub-section 4.3.2)

$$\{\hat{C}_c\}_{v,l} = n \cdot \mathrm{var}\{vech[S_w]\}_{v,l}, \quad (6.09b)$$

which allows for complex surveys, with subscripts v and l denoting respectively $v = (t, t')$ and $l = (t'', t''')$, and $\mathrm{var}\{vech[S_w]\}_{v,l}$ given by (4.39b), from Chapter 4, Section 4.3, Sub-section 4.3.2.

We could then calculate $a\,\mathrm{cov}(\hat{\underline{\theta}})$ by plugging (6.09b) into (6.09a). Note that the estimator we proposed above has as special cases the estimators proposed by Skinner (1989a), followed by Satorra (1992; when considering no sampling weighting and no covariates), and Skinner and Holmes (2003; also when considering no covariates). A variance estimator of this type would then be, according to Skinner (1989a; see also Satorra, 1992; and Muthén and Satorra, 1995), appropriate to non-normal complex survey data, although a 'large' number of PSUs units would be required to guarantee the asymptotic approximation to be correct, especially when working with large models. Note that this approach may also be adopted for estimating distribution-free variances of point estimators obtained under the normality assumption (Muthén and Satorra, 1995), and thus could be called robust normal theory analysis (Muthén and Satorra, 1995). Moreover, as (6.09b) belongs to a class of ADF methods, that estimator could suffer from problems described in Chapter 4 when working with small samples in situations where T is large, as discussed earlier in this thesis.

A special case of the estimator given by (6.09b) when working under independence assumptions is (see again Chapter 4, Section 4.3, Sub-section 4.3.2)

$$\frac{1}{n-1}\cdot\left(n\Big/\sum_{i=1}^{n} w_i\right)^2\cdot\sum_{i=1}^{n} w_i^2\left(\hat{\varepsilon}_{it}\hat{\varepsilon}_{it'}-S_{w\,tt'}\right)\cdot\left(\hat{\varepsilon}_{it''}\hat{\varepsilon}_{it'''}-S_{w\,t''t'''}\right). \tag{6.10}$$

By plugging (6.10) into (6.09a) for estimating $a\operatorname{cov}(\underline{\hat{\theta}})$, we would be adopting an estimator that allows for unequal sampling weights but does not fully accounts for complex sampling designs. Moreover, the estimator given in (6.10) is also suitable for non-normal $\underline{Y}_i$.

Alternatively we may follow the approach of Binder (1983) for estimation of $a\mathrm{COV}(\underline{\hat{\theta}}_{PML})$, the asymptotic covariance matrix of the PML point estimator $\underline{\hat{\theta}}_{PML}$. Recall that $F(\underline{\theta})_{PML}$ is given by (4.62) in Chapter 4, Sub-section 4.4.5. Let $\underline{\phi}(\underline{\theta})$ denote the $b\times 1$ pseudo-score function with jth element given by (see Chapter 4, Sub-section 4.4.5 and Appendix D, Result D2)

$$\phi_{\mathrm{j}}(\underline{\theta})=\frac{\partial F(\underline{\theta})_{PML}}{\partial\theta_{\mathrm{j}}}=tr\left\{\Sigma(\underline{\theta})^{-1}\cdot\left[\Sigma(\underline{\theta})-\mathrm{S}_w\right]\cdot\Sigma(\underline{\theta})^{-1}\cdot\frac{\partial\Sigma(\underline{\theta})}{\partial\theta_{\mathrm{j}}}\right\}, \tag{6.11}$$

with $\mathrm{j}=1,\ldots,b$. To obtain an expression for the estimator of the variance-covariance matrix of the pseudo maximum likelihood estimator $\underline{\hat{\theta}}_{PML}$, we may take a Taylor expansion of $\underline{\phi}(\underline{\hat{\theta}}_{PML})$ at $\underline{\hat{\theta}}_{PML}=\underline{\theta}$. Thus (Binder, 1983)

$$0=\underline{\phi}(\underline{\hat{\theta}}_{PML})\cong\underline{\phi}(\underline{\theta})+\frac{\partial\underline{\phi}(\underline{\theta})}{\partial\underline{\theta}}\left(\underline{\hat{\theta}}_{PML}-\underline{\theta}\right),$$

so that

$$\underline{\phi}(\underline{\theta})\cong-\frac{\partial\underline{\phi}(\underline{\theta})}{\partial\underline{\theta}}\left(\underline{\hat{\theta}}_{PML}-\underline{\theta}\right). \tag{6.12}$$

Let

$$\mathrm{I}(\underline{\theta})=-\frac{\partial\underline{\phi}(\underline{\theta})}{\partial\underline{\theta}}$$

be the $b\times b$ pseudo information matrix. Note here the distinction in notation between I(.) introduced above, and I which denotes an identity matrix in previous chapters. Moreover, it follows that (Binder, 1983)

$$\operatorname{cov}\left[\underline{\phi}(\underline{\hat{\theta}}_{PML})\right]=\left[\mathrm{I}(\underline{\hat{\theta}}_{PML})\right]\cdot\operatorname{cov}(\underline{\hat{\theta}}_{PML})\cdot\left[\mathrm{I}(\underline{\hat{\theta}}_{PML})\right]', \tag{6.13}$$

after taking variances of both sides of (6.12). Furthermore, expression (6.13) is equivalent to (Binder, 1983)

$$\operatorname{cov}\left(\underline{\hat{\theta}}_{PML}\right)=\left[\mathrm{I}\left(\underline{\hat{\theta}}_{PML}\right)\right]^{-1}\cdot\operatorname{cov}\left[\underline{\phi}\left(\underline{\hat{\theta}}_{PML}\right)\right]\cdot\left[\mathrm{I}\left(\underline{\hat{\theta}}_{PML}\right)\right]^{-1}, \tag{6.13b}$$

where $\operatorname{cov}\left(\underline{\hat{\theta}}_{PML}\right)$ is a $b\times b$ matrix, the $b\times b$ matrix $\operatorname{cov}[\underline{\phi}(\underline{\theta})]$ is a consistent estimator of the variance-covariance matrix of $\underline{\phi}(\underline{\theta})$ for fixed $\underline{\theta}$, and $\operatorname{cov}[\underline{\phi}(\underline{\hat{\theta}}_{PML})]$ is the value of $\operatorname{cov}[\underline{\phi}(\underline{\theta})]$ when evaluated at $\underline{\theta}=\underline{\hat{\theta}}_{PML}$.

Observe that we may express $\phi_{\mathrm{j}}(\underline{\theta})$ as constant plus a ratio of two totals, (see Appendix D, Result D3)

$$\phi_{\mathrm{j}}(\underline{\theta})=tr\left[\Sigma(\underline{\theta})^{-1}\cdot\frac{\partial\Sigma(\underline{\theta})}{\partial\theta_{\mathrm{j}}}\right]+\frac{\sum_{i=1}^{n}w_i z_{\mathrm{j}i}}{\sum_{i=1}^{n}w_i} \tag{6.14}$$

where

$$z_{\mathrm{j}i}=-\left(\underline{y}_i-\underline{\hat{\mu}}_i\right)'\Sigma(\underline{\theta})^{-1}\frac{\partial\Sigma(\underline{\theta})}{\partial\theta_{\mathrm{j}}}\Sigma(\underline{\theta})^{-1}\left(\underline{y}_i-\underline{\hat{\mu}}_i\right),$$

where $\underline{\hat{\mu}}_i$ is defined in Section 6.2, expression (4.30), with $\underline{\hat{\beta}}_{PML}(\mathrm{V})$ given by (2.27) from Chapter 2. By re-writing (6.11) as (6.14), and ignoring the constant term, we allow $\operatorname{cov}[\underline{\phi}(\underline{\hat{\theta}}_{PML})]$ to be formulated as the covariance matrix of means, considering $\underline{\hat{\mu}}_i$ as fixed.

Let $a_i=w_i z_i$, $i=1,\ldots,n$. Then,

$$R_{\mathrm{j}}=\frac{\mu_a}{\mu_w}=\frac{E(\overline{a}_{\mathrm{j}})}{E(\overline{w})}, \tag{6.15}$$

where j is as defined for (6.11),

$$\overline{a}_{\mathrm{j}}=n^{-1}\cdot\sum_{i=1}^{n}a_{\mathrm{j}i},$$

and

$$\overline{w}=n^{-1}\cdot\sum_{i=1}^{n}w_i.$$

We can notice that both μ_a and μ_w are linear statistics and we could then apply the Taylor series linearization method for estimating R_{j}'s variance. See Chapter 2, Section 2.5, Sub-section 2.5.1 for some characteristics of the linearization method. It follows that (Woodruff, 1971; see also Chapter 4, Section 4.3, Sub-section 4.3.2)

$$R_j \doteq \frac{\mu_a}{\mu_w} + \frac{1}{n}\sum_{i=1}^{n}\left(a_{ji} - \frac{\mu_a}{\mu_w} w_i\right)\cdot\frac{1}{\mu_w}.$$

The variance of R_j may thus be approximated by

$$\mathrm{VAR}\{R_j\} \doteq \mathrm{VAR}\left(\frac{1}{n}\cdot\sum_{i=1}^{n} u_{ji}\right) = \left(\frac{1}{n}\right)^2\cdot \mathrm{VAR}\left(\sum_{i=1}^{n} u_{ji}\right) = \left(\frac{1}{n}\right)^2\cdot \mathrm{VAR}(\mathrm{B}_j),$$

where

$$u_{ji} = \frac{1}{\mu_w}\cdot\left(a_{ji} - \frac{\mu_a}{\mu_w} w_i\right), \qquad (6.16a)$$

and

$$\mathrm{B}_j = \sum_{i=1}^{n} u_{ji}. \qquad (6.16b)$$

We could then adopt

$$\hat{u}_{ji} = \frac{1}{\overline{w}}\cdot\left(a_{ji} - \frac{\overline{a}_j}{\overline{w}} w_i\right)$$

and

$$\hat{\mathrm{B}}_j = \sum_{i=1}^{n} \hat{u}_{ji},$$

for estimating (6.16a) and (6.16b), respectively. Let,

$$\underline{\hat{\mathrm{B}}} = [\hat{\mathrm{B}}_1, \ldots, \hat{\mathrm{B}}_b]',$$

be a $b \times 1$ vector of totals. Taylor series approximation for R_j imply that (see Shah *et al.*, 1995, Sub-section 5.1)

$$\mathrm{cov}[\underline{\phi}(\underline{\theta})] = \left(\frac{1}{n}\right)^2\cdot \mathrm{cov}[\underline{\hat{\mathrm{B}}}],$$

for any complex sampling design.

We may then revisit again Example 2.2 (from Chapter 2), where we considered a multistage stratified sampling scheme. In that situation, we could calculate $\mathrm{cov}[\underline{\hat{\mathrm{B}}}]$ by (see Shah *et al.*, 1995, Sub-section 2.2.3)

$$\mathrm{cov}_L\left[\underline{\hat{\mathrm{B}}}\right]_{j,k} = \sum_{h=1}^{H} m_h \left\{\left[\sum_{j=1}^{m_h}\left(\underline{\hat{\mathrm{B}}}_{hj,j} - \overline{\underline{\hat{\mathrm{B}}}}_{h,j}\right)'\left(\underline{\hat{\mathrm{B}}}_{hj,k} - \overline{\underline{\hat{\mathrm{B}}}}_{h,k}\right)\right] \Big/ (m_h - 1)\right\}. \qquad (6.17)$$

Notice that for evaluating the matrix $\mathrm{I}(\underline{\theta})$, it is required to differentiate

$$\frac{\partial \Sigma(\underline{\theta})}{\partial \underline{\theta}}$$

with respect to $\underline{\theta}$, i.e. it is required to differentiate $\Sigma(\underline{\theta})$ with respect to $\underline{\theta}$ twice. However, we could avoid this complication by assuming that the model is correct (see Result D.4 in Appendix D), i.e. that $E[\mathrm{S}_w]=\Sigma(\underline{\theta})$. We could then alternatively define the information matrix as

$$\mathrm{I}(\underline{\theta})=E\left[-\frac{\partial\phi(\underline{\theta})}{\partial\underline{\theta}}\right].$$

In this situation, the jkth element of $\mathrm{I}(\underline{\theta})$ is

$$\mathrm{I}(\underline{\theta})_{\mathrm{jk}}=tr\left[\Sigma(\underline{\theta})^{-1}\frac{\partial\Sigma(\underline{\theta})}{\partial\theta_{\mathrm{j}}}\Sigma(\underline{\theta})^{-1}\frac{\partial\Sigma(\underline{\theta})}{\partial\theta_{\mathrm{k}}}\right],$$

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

Replication variance estimation methods, such as Jackknife (Chapter 2, Section 2.5, Sub-section 2.5.2), could also be adopted for variance estimation of $\hat{\underline{\theta}}$ assuming that the sample is selected under complex survey sampling, although we shall not consider this approach here.

6.4 Model fitting tests

Recall from Chapter 4, Section 4.1, expression (4.3), that

$$\Sigma=\Sigma(\underline{\theta}), \tag{6.18}$$

which shall be our covariance structure hypothesis. Model fit measures are used to assist in the evaluation of whether (6.18) is valid or not, and if not, such measures could help to calculate the deviation of Σ from $\Sigma(\underline{\theta})$, according to Bollen (1989).

We shall review model fitting statistics under the classical approach in Sub-section 6.4.1, while we shall propose some new developments on fitting measures when working under the complex sampling approach in Sub-section 6.4.2. Note that we shall not discuss incremental fit indices (or model modification indices) and model components fit measures (or model parameter tests) here. For a comprehensive review on these subjects, considering the classical approach, see Bollen (1989). For further information, see also Jöreskog and Sörbom (1989), Jöreskog and Sörbom (1997), Matsueda and Bielby (1986), Satorra (1989), Bentler and Dudgeon (1996), and Yuan and Bentler (1997b).

In this section, let $\hat{\underline{\theta}}$ denote an estimator of $\underline{\theta}$ which minimizes either $F(\underline{\theta})_{ML}$ or $F(\underline{\theta})_{GLS}$ in Sub-section 6.4.1 (unless otherwise stated), and $F(\underline{\theta})_{PML}$ or $F(\underline{\theta})_{GLSC}$, when working under the complex sampling context, which shall be considered in Sub-section 6.4.2 (unless otherwise stated).

6.4.1 Model testing in the classical context

In this section we shall be working under the assumptions of iid observations.

Both Σ and $\Sigma(\underline{\theta})$ are unknown population parameters. Thus for calculating model fit measures we would in fact need to consider their estimators S and $\Sigma(\hat{\underline{\theta}})$, where $\Sigma(\hat{\underline{\theta}})$ is the covariance matrix evaluated at $\hat{\underline{\theta}}$ (or fitted covariance matrix, as suggested by Jöreskog and Sörbom, 1989; and Jöreskog and Sörbom, 1997).

Most of the model fit measures proposed in the literature consider functions of S and Σ. Bollen (1989) discusses some limitations of overall model fit measures, as for example: (*i*) these measures are inapplicable to exactly identified modes (as S always equals $\Sigma(\hat{\underline{\theta}})$ in this circumstance); and (*ii*) these may contradict the fit of parts of the model (for example, the global fit may be favourable but parameter estimates could not be statistically significant). See Jöreskog and Sörbom (1997) for further discussion on limitations of overall model fit statistics.

In order to perform a goodness of fit test, we may initially define a null and a generic alternative hypothesis as respectively (see, for example, Bentler and Weeks, 1980; Browne, 1982; Satorra and Bentler, 1986; Satorra and Bentler, 1988; Satorra, 1989; and Satorra and Bentler, 1994)

$$H_0 : \Sigma = \Sigma(\underline{\theta})$$

against

$H_1 : \Sigma$ is an unrestricted covariance matrix (any $T \times T$ positive definite matrix). Let the population *residual covariance matrix* be denoted by E_p, so that

$$E_p = [\Sigma - \Sigma(\underline{\theta})]. \tag{6.19}$$

When H_0 is true, E_p is a zero matrix. The sample *residual covariance matrix* $\hat{E}$, defined as

$$\hat{E} = [S - \Sigma(\hat{\underline{\theta}})],$$

could help us to identify components of the variance-covariance matrix that are not well fit. See Chapter 4, Section 4.3, Sub-section 4.3.2.

Bollen (1989) considers $\hat{E}$ as the simplest model fit measure. In fact, let

$$[S_{tt'} - \Sigma(\hat{\underline{\theta}})_{tt'}]$$

be individual sample residual covariances, where $S_{tt'}$ and $\Sigma(\hat{\underline{\theta}})_{tt'}$ are the tt'th elements in S and $\Sigma(\underline{\theta})$ respectively. Note that the mean or the median of absolute individual residuals may be utilized as summary measures for evaluating model fit.

Furthermore, Jöreskog and Sörbom (1989) proposed the following statistic for summarising the residuals,

$$\text{RMR} = \sqrt{2 \cdot \sum_{t=1}^{T} \sum_{t'=1}^{t} \frac{[S_{tt'} - \Sigma(\hat{\underline{\theta}})_{tt'}]^2}{T(T+1)}},$$

where RMR stands for root mean-square residual. This measure may be adopted to compare the fit of two different models for the same data. Moreover, according to Jöreskog and Sörbom (1989) and Jöreskog and Sörbom (1997), the RMR would work better if $\underline{Y}_i$ are standardized, and according to Bollen (1989), this measure usually leads to very similar conclusions when compared to the mean absolute value of the unstandardized residuals.

Sample residuals are not only affected by differences between Σ and $\Sigma(\underline{\theta})$, but also by the scales of $\underline{Y}_i$ and by sampling fluctuations (errors). A direct solution for the scales issue may be to calculate correlation residuals as (Bollen, 1989)

$$r_{tt'} - \hat{r}_{tt'},$$

where $r_{tt'}$ is the sample correlation between $\underline{Y}_{it}$ and $\underline{Y}_{it'}$, and $\hat{r}_{tt'}$ denotes the model predicted correlation, so that

$$\hat{r}_{tt'} = \frac{\Sigma(\hat{\underline{\theta}})_{tt'}}{\sqrt{\Sigma(\hat{\underline{\theta}})_{tt} \cdot \Sigma(\hat{\underline{\theta}})_{t't'}}}.$$

Although this correlation residual is allowed to range from -2 to $+2$, we should expect values rather close to zero for models with a reasonably good fit.

Regarding sampling errors, we know that even when H_0 is true the expected amplitude of the individual sample residual covariances depends on n, and (Bollen, 1989)

$$\lim_{n \to \infty} \hat{\text{E}} = \text{E}_{\text{p}}.$$

Therefore, we shall introduce below a simultaneous significance test, based on sample residuals. Notice that the H_0 discussed above could be tested via a chi-square test (or a minimum discrepancy test statistic).

Under multivariate normality of $\underline{Y}_i$ and no covariates, if H_0 holds, if $\underline{\theta}$ is identified, if (6.03) is of full rank b, it holds that (Jöreskog and Goldberger, 1972; see also, for example, Wiley, Schmidt, and Bramble, 1973; Swain, 1975; Bentler and Weeks, 1980; Browne, 1982; Browne, 1984; Shapiro and Browne, 1987; Fuller, 1987, Theorem 4.2.1; Shapiro, 1986; Matsueda and Bielby, 1986; Satorra and Bentler, 1986; Browne, 1987; Satorra and Bentler, 1988; Satorra, 1989; Amemiya and Anderson, 1990; Muthén and Satorra, 1995; Bentler and Dudgeon, 1996; Yuan and Bentler, 1997b; and Bollen, 1989)

$$n \cdot F(\hat{\underline{\theta}})_{ML} \tag{6.20}$$

has an asymptotic chi-square distribution with $k-b$ degrees of freedom (df), i.e.

$$n \cdot F(\hat{\underline{\theta}})_{ML} \sim \chi^2_{k-b} \tag{6.21}$$

under H_0, when $F(\underline{\theta})_{ML}$ is evaluated at the final estimates and the model is true, where b is the number of free parameters.

Remark 6.1: We shall examine initially the $n \cdot F(\hat{\underline{\theta}})_{ML}$ asymptotic distribution by using the likelihood ratio principle, as in Bollen (1989).

We may consider $\Sigma(\hat{\underline{\theta}})$, which is the fitted covariance matrix at H_0. Let $\log \ell_0$ denote the maximised value of the log of the likelihood function (see Chapter 4, Section 4.4, Sub-section 4.4.4) under H_0, given by

$$\log \ell_0 = -\frac{n}{2} \cdot \{tr[S\Sigma(\hat{\underline{\theta}})^{-1}] + \log|\Sigma(\hat{\underline{\theta}})|\}, \tag{6.22}$$

ignoring the unnecessary constants. Assuming again normality and having a sufficiently large n (Jöreskog and Sörbom, 1997), a model fit test could be performed working with a likelihood ratio statistic, with (6.22) as the log of the numerator. Moreover, we may consider defining H_1 by specifying $\Sigma(\hat{\underline{\theta}})$ for example to the unweighted sample covariance matrix S in (6.22). We could then define the log likelihood function for H_1 as

$$\log \ell_1 = -\frac{n}{2} \cdot \{T + \log|S|\}, \tag{6.23}$$

also when ignoring the unnecessary constants. In this case, the expression given by (6.23) could then be adopted as the log of the denominator of the likelihood ratio statistic. Note that defining H_1 as above makes us to test H_0 against a perfect fit situation, with $\Sigma(\hat{\underline{\theta}}) = S$.

It is well known that (see, for example, Specht and Warren, 1976)

$$-2\log\left[\frac{\ell_0}{\ell_1}\right] \sim \chi^2_{k-b},$$

when n is large. Notice that, from (6.22) and (6.23),

$$-2\log\left[\frac{\ell_0}{\ell_1}\right]=n\cdot[\log|\Sigma(\hat{\underline{\theta}})|+tr(\Sigma(\hat{\underline{\theta}})^{-1}\mathrm{S})-\log|\mathrm{S}|-T],$$

which is n times the ML fitting function (see Chapter 4, Section 4.4, Sub-section 4.4.4), when substituting $\Sigma(\underline{\theta})$ for the ML fitted covariance matrix. We have therefore shown that (6.21) is valid. ■

It is also true, when $\underline{Y}_i$ is multivariate normally distributed, that (see proof in Jöreskog and Goldberger, 1972; and Fuller, 1987, Sub-section 4.2.1, Theorem 4.2.2; see also Browne, 1982; Browne, 1984; Shapiro, 1986; Satorra and Bentler, 1988; Matsueda and Bielby, 1986; Amemiya and Anderson, 1990; and Muthén and Satorra, 1995)

$$n\cdot F(\hat{\underline{\theta}})_{GLS} \tag{6.24}$$

has the same property as $n\cdot F(\hat{\underline{\theta}})_{ML}$, i.e.

$$n\cdot F(\hat{\underline{\theta}})_{GLS}\sim\chi^2_{k-b} \tag{6.25}$$

when (6.02) holds, and in large samples. Note that (6.25) is a Wald goodness of fit test statistic, which may be obtained as the minimum value of $F(\underline{\theta})_{GLS}$ given by expression (4.46), introduced in Chapter 4, Section 4.4, Subsection 4.4.2, i.e. when (4.46) is evaluated at $\hat{\underline{\theta}}_{GLS}$. See Buse (1982), Satorra (1989), and Satorra (1992) for some further discussion.

We shall not consider here the case of the ULS fitting function as some adjustments would be necessary in order to make the chi-square measure also asymptotically valid for that method (see Browne, 1984; Jöreskog and Sörbom, 1989; and Jöreskog and Sörbom, 1997).

Although the adoption of chi-square statistics to test covariance structure model fit is frequent in the literature (Teachman *et al.*, 2001), we have to be careful when working with that measure. The reason is that the chi-square approximation presupposes that (Bollen, 1989): (*i*) $\underline{Y}_i$ has no kurtosis; (*ii*) n is sufficiently large[30]; and (*iii*) H_0 holds exactly. In practice, often at least one of these assumptions is infringed, as for example with the incidence of non-normal data with excessive kurtosis (Browne, 1982; Browne, 1984; and Shapiro and Browne, 1987), which can gravely invalidate the chi-squared asymptotic distribution of the test statistic (Browne, 1987). Moreover, as models are only a tentative representation of the reality, the prospect that the model does not hold exactly has to be taken into account (Browne, 1984). Note that, under certain assumptions and conditions discussed

[30] For situations where the sample size is not large enough, there have been developed adjusted multiplying factors that could be substituted for n in (6.27) and (6.28). For further information see Browne (1982).

by Satorra (2002), in the context of linear-latent variable models, normal theory chi-square statistics could still be valid, when the model holds, despite non-normality.

Another disadvantage of adopting the chi-square test is that this may become too powerful with increases in the sample size or with 'too little' power to detect noticeable deviations in rather small samples (Browne, 1982; Satorra and Saris, 1985; and Teachman *et al.*, 2001; for further information on statistical power in covariance structure models, see also Matsueda and Bielby, 1986). Therefore, Jöreskog and Sörbom (1997) recommend that the chi-square statistic should be adopted as a goodness-of-fit measure instead of a test statistic, using the df as a guide to evaluate whether χ^2 is large or small.

Browne (1984), followed by Satorra and Bentler (1988), propose more generic goodness-of-fit tests of structural models for covariance matrices, which does not assume normality[31] and leads to precise test statistics.

Remark 6.2: In principle, we could perform a model fit test by adopting (6.25), with $F(\underline{\theta})_{GLS}$ given by expression (4.46) from Chapter 4, Section 4.4, Sub-section 4.4.2, with matrix U given by (4.33), i.e. an ADF GLS fitting function. Following Browne (1982) and Browne (1984), we shall assume that $\hat{\underline{\theta}}$ is an estimator (of $\underline{\theta}$) that, as $n \to \infty$, $\sqrt{n} \cdot (\hat{\underline{\theta}} - \underline{\theta})$ has an asymptotic normal distribution with a null mean vector and a finite covariance matrix. Note that according to Browne (1984), both GLS and ML types of point estimators, when considering regularity conditions described in that article, have the above property.

Let Δ^* be a $k \times (k-b)$ matrix valued function of $\underline{\theta}$, with rank $(k-b)$, so that $\Delta^{*}{}'\Delta = 0$ when Δ is given by (6.03). Moreover, let $\hat{\Delta}^*$ denote Δ^* evaluated at $\underline{\theta} = \hat{\underline{\theta}}$ (either $\hat{\underline{\theta}}_{ML}$ or $\hat{\underline{\theta}}_{GLS}$, according to Satorra, 1989; see also Muthén and Satorra, 1995). If $\hat{\mathrm{C}}$ is any consistent estimator of C, as for example $\hat{\mathrm{C}}$ given by (4.33), then (see proof in Browne, 1984)

$$X = n \cdot \{vech[\hat{\mathrm{E}}]' \hat{\Delta}^* (\hat{\Delta}^{*}{}' \hat{\mathrm{C}} \hat{\Delta}^*)^{-1} \hat{\Delta}^{*}{}' vech[\hat{\mathrm{E}}]\} \sim \chi^2_{k-b},$$

or equivalently (Browne, 1982; Browne, 1984; see also, for example, Satorra and Bentler, 1988; and Bentler and Dudgeon, 1996)

$$X = n \cdot vech[\hat{\mathrm{E}}]' \cdot \{\hat{\mathrm{C}}^{-1} - \hat{\mathrm{C}}^{-1} \cdot \hat{\Delta}(\hat{\Delta}' \hat{\mathrm{C}}^{-1} \hat{\Delta})^{-1} \hat{\Delta}' \hat{\mathrm{C}}^{-1}\} \cdot vech[\hat{\mathrm{E}}] \sim \chi^2_{k-b}. \qquad (6.26)$$

[31] See Amemiya and Anderson (1990), for information on conditions under which some normal theory variance estimation procedures and fit test would be asymptotically correct even when data is not normally distributed, considering the case of a class of factor analysis models.

Observe that, according to Browne (1984, Corollary 4.1), if $\hat{\mathrm{C}} = \mathrm{U}$, (6.26) reduces to (6.25). Moreover, according to Satorra and Bentler (1988), (6.26) is chi-squared distributed regardless of the distribution of $\underline{Y}_i$, and the matrix U adopted in the point estimation procedure, if the model holds, n is large enough, and C is properly estimated.

Nevertheless, according to Satorra and Bentler (1994), Bentler and Dudgeon (1996), Satorra (2000), and Satorra (2002), for example, X given by (6.26) could possibly be affected by the same problem which act on ADF methods in general when working with small samples, especially in situations with large T. Furthermore, according to Bentler and Dudgeon (1996), there is some empirical evidence that n would be required to be unrealistic large for the theory to hold (at least $n = 5000$).

Satorra and Bentler (1988) have proposed two types of corrections for standard test statistics, suitable for a more general distribution than the elliptical class[32] (see also Bentler and Dudgeon, 1996): a (*i*) Bartlett-type scaled statistic (X_S), and a (*ii*) Satterthwaite-type adjusted statistic (X_A). The scaled statistic (*i*) is a mean-correct chi-square and is given by

$$X_S = (\hat{\eta}^*)^{-1} \cdot n \cdot F(\hat{\underline{\theta}}), \tag{6.27}$$

where (Satorra and Bentler, 1988; see also Satorra, 1989; Chou, Bentler and Satorra, 1991; Satorra, 1992; Satorra and Bentler, 1994; Muthén and Satorra, 1995; and Satorra and Bentler, 2001)

$$\hat{\eta}^* = \frac{tr\{[\mathrm{U}^{-1} - \mathrm{U}^{-1} \cdot \hat{\Delta}(\hat{\Delta}'\mathrm{U}^{-1}\hat{\Delta})^{-1}\hat{\Delta}'\mathrm{U}^{-1}] \cdot \hat{\mathrm{C}}\}}{(k-b)}, \tag{6.28}$$

The statistic (6.27) is asymptotically chi-squared distributed with $(k-b)$ degrees of freedom, under H_0. In (6.28), U is a consistent estimator of $n \cdot a\mathrm{COV}(\mathrm{S}_{tt'}, \mathrm{S}_{t''t'''})$ adopted for performing the point estimation (for example, when minimizing (4.46) in Chapter 4, Section 4.4, Sub-section 4.4.2). However, as (6.27) does not assume that U is correctly specified (Bentler and Dudgeon, 1996), we shall assume that U is given by (4.24), from Chapter 4, Section 4.3, Sub-section 4.3.2, which assumes normality. Hence, one possible strategy could be to utilise the conventional normal theory point estimators and adopt the corrected variance estimators and test statistics (as suggested by Browne, 1984; Satorra, 1992; and Muthén and Satorra, 1995).

Note that Chou, Bentler and Satorra (1991) have found in an extensive simulation study, that under certain circumstances scaled test statistic given by (6.27) performed better than the

[32] Elliptical distributions are symetric with no skewness and have the normal as a special case, but also include platykurtic and leptokurtic distributions. For further information on elliptical distributions see Browne (1984), and references therein, and Satorra and Bentler (1988).

ADF test statistic proposed by Browne (1982) and Browne (1984), also discussed above, given by (6.26). Moreover, Satorra and Neudecker (1997) have developed asymptotic chi-squared scaled statistics in the context of testing equality of moment vectors, considering in principle a disaggregated analysis approach.

The statistic *(ii)* is adjusted for mean and variance and is given by

$$X_A = (\hat{\eta}^+)^{-1} \cdot n \cdot F(\hat{\underline{\theta}}), \tag{6.29}$$

where (Satorra and Bentler, 1988; see also Satorra, 1992; and Satorra and Bentler, 1994)

$$\hat{\eta}^+ = \frac{d}{tr\{[\mathrm{U}^{-1} - \mathrm{U}^{-1} \cdot \hat{\Delta}(\hat{\Delta}'\mathrm{U}^{-1}\hat{\Delta})^{-1}\hat{\Delta}'\mathrm{U}^{-1}] \cdot \hat{\mathrm{C}}\}}, \tag{6.30}$$

with d denoting the nearest integer to

$$d^* = \frac{\{tr[(\mathrm{U}^{-1} - \mathrm{U}^{-1} \cdot \hat{\Delta}(\hat{\Delta}'\mathrm{U}^{-1}\hat{\Delta})^{-1}\hat{\Delta}'\mathrm{U}^{-1}) \cdot \hat{\mathrm{C}}]\}^2}{tr[(\mathrm{U}^{-1} - \mathrm{U}^{-1} \cdot \hat{\Delta}(\hat{\Delta}'\mathrm{U}^{-1}\hat{\Delta})^{-1}\hat{\Delta}'\mathrm{U}^{-1}) \cdot \hat{\mathrm{C}}]^2}. \tag{6.31}$$

The statistic (6.29) is asymptotically chi-squared distributed with d degrees of freedom, under H_0. We shall consider in (6.30) and (6.31) that U is as defined for (6.28). Note that (6.29) is based on both scaling and a degrees of freedom adjustment (Satorra and Bentler, 1988). Moreover, according to Bentler and Dudgeon (1996), (6.29) is more nearly approximately chi-squared distributed than (6.27). See Satorra and Bentler (1988) for details on how the expressions for X_S and X_A are obtained.

We shall not consider here the direct approach proposed by Browne (1982), Browne (1984, with proof), and extended by Shapiro and Browne (1987), which involves applying simple corrections for kurtosis in order to make standard results applicable to a class of elliptical distributions. According to Satorra and Bentler (1994), these corrections may be very sensitive to violations of the elliptical assumption so that the corrected test statistic could be even more distinct from the reference distribution than the uncorrected statistic.

Olsson, Foss and Troye (2003) have recently shown that Browne (1984) ADF test statistic, and also those proposed by Satorra and Bentler (1988), have their expected values decreased with increases in the kurtosis.

Note that we shall not consider here the testing approach developed by Lee (1985), which also does assume normality. Moreover, Yuan and Chan (2002) have proposed another rescaled test statistic which has (6.29) as a special case, although that shall not be considered here. ■

The covariance structure model testing statistics given above are suitable in situations where it is acknowledged that the distribution of $\underline{Y}_i$ is non-normal. In that situation we still need to work under the assumptions that n is sufficiently large and that H_0 holds exactly. Furthermore, we still need to be cautious regarding the power of the tests with increases in the sample size. Note that Satorra and Saris (1985) have developed a method for calculating the power of the likelihood ratio test in the covariance structure analysis context, although we shall not consider that procedure here.

We shall now discuss some alternative asymptotically equivalent (Satorra, 1989; Buse, 1982) significance tests for the difference in model fit chi-square statistics for nested models: (*i*) likelihood ratio test (LRT) or chi-square difference test (Wilks, 1938); (*ii*) Lagrangian multiplier test (LMT; Silvey, 1959) or efficient score test; and (*iii*) Wald test (WT; Wald, 1943); which all assume that $F(\underline{\theta})$ is asymptotic optimal, i.e. leads to efficient estimators and chi-square statistics[33]. Moreover, these tests have been developed under the assumption that the asymptotic covariance matrix of $vech[\mathrm{S}]$ and the information matrix of the model are non-singular (Satorra, 1989; see Shapiro, 1986, for situations where this assumption is violated).

In general, these types of test aim to compare an 'initial' model with a restricted model, which has a sub-vector of parameters that is set to be equal to zero. Suppose that the $b \times 1$ parameter vector $\underline{\theta}$ is rewritten as

$$\underline{\theta} = \begin{pmatrix} \underline{\theta}^R \\ \underline{\theta}^U \end{pmatrix},$$

where $\underline{\theta}^R$ is a $b^* \times 1$ vector that corresponds to the restraints imposed to the unrestricted model, with $b^* < b$, and $\underline{\theta}^U$ is a $(b - b^*) \times 1$ vector. We may then define the null and alternative nested hypotheses as respectively

$$H_0 : \underline{\theta}^R = \underline{0}$$

against

$$H_1 : \Sigma = \Sigma(\underline{\theta}).$$

Let

$$\underline{\hat{\theta}}_{ML,r} = \begin{pmatrix} \underline{0} \\ \underline{\hat{\theta}}_r^U \end{pmatrix}$$

be the ML estimator when considering the restrictive (nested) model, and let

[33] See Satorra (1989), for generalised test statistics that do not require such assumption.

$$\hat{\underline{\theta}}_{ML,u} = \begin{pmatrix} \hat{\underline{\theta}}_u^R \\ \hat{\underline{\theta}}_u^U \end{pmatrix}$$

be the ML estimator when considering the covariance structure model without constraints.

We shall initially discuss the LRT, which has (Jöreskog, 1970; see also Swain, 1975; Buse, 1982; Satorra and Saris, 1985; Lee, 1985; Shapiro and Browne, 1987; Browne, 1987; Matsueda and Bielby, 1986; Satorra, 1989; Bollen, 1989; and Yanagihara, Tonda and Matsumoto, 2005)

$$-2[\log \ell(\hat{\underline{\theta}}_{ML,r}) - \log \ell(\hat{\underline{\theta}}_{ML,u})], \tag{6.32}$$

or alternatively

$$n \cdot [F(\hat{\underline{\theta}}_{ML,r}) - F(\hat{\underline{\theta}}_{ML,u})] \tag{6.33}$$

as its test statistic. The chi-square difference test statistic is the most often adopted test adopted to compare nested models (Satorra, 1989). Under mild regularity conditions, it has χ^2 limiting distribution with df equal to the difference between the df for the chi-square test statistics for the restricted and unrestricted models, i.e. the number of restrictions constrained by the null hypothesis. Notice that (6.32) may be obtained from Remark 6.1 presented above.

The LMT aims to confront the fit of a restrictive model to a less restrictive model. Its test statistic is based on

$$\frac{\partial \log \ell[\underline{\theta}]}{\partial \underline{\theta}}, \tag{6.34}$$

where $\log \ell[\underline{\theta}]$ is the unrestricted log likelihood function, as given by expression (4.54) in Chapter 4, Section 4.4, Sub-section 4.4.4. Note that the unrestricted components of $\hat{\underline{\theta}}_{ML,r}$ are expected to have a zero partial derivative, considering that the solution for the ML estimator $\hat{\underline{\theta}}_{ML,r}$ is obtained by setting these partial derivatives to zero. Nevertheless, the restrictions have to hold exactly so that the elements of (6.34), equivalent to the restricted parameters, are zero. We may thus check whether the restrictions are valid or not by substituting $\hat{\underline{\theta}}_{ML,r}$ for $\underline{\theta}$ in (6.34). Moreover, observe that sample errors must be considered as these could cause non-zero values even in cases where the restrictive model is valid. Therefore, let (Bollen, 1989)

$$\left[\frac{\partial \log \ell[\underline{\theta}]}{\partial \underline{\theta}}\right]' \cdot \left\{-E\left[\frac{\partial^2 \log \ell[\underline{\theta}]}{\partial \underline{\theta} \partial \underline{\theta}'}\right]\right\}^{-1} \cdot \left[\frac{\partial \log \ell[\underline{\theta}]}{\partial \underline{\theta}}\right]\Bigg|_{\underline{\theta}=\hat{\underline{\theta}}_{ML,r}}, \tag{6.35}$$

be the statistic for the LMT, which may be alternatively expressed as (see, for example, Buse, 1982; and Bollen, 1989)

$$\frac{n}{2}\cdot\left[\frac{\partial F(\underline{\theta})}{\partial\underline{\theta}}\right]'\cdot\left\{E\left[\frac{\partial^2 F(\underline{\theta})}{\partial\underline{\theta}\partial\underline{\theta}'}\right]\right\}^{-1}\cdot\left[\frac{\partial F(\underline{\theta})}{\partial\underline{\theta}}\right]\Bigg|_{\underline{\theta}=\hat{\underline{\theta}}_{ML,r}}.$$

The LMT statistic has also a χ^2 limiting distribution with df equal to the difference between the df for the chi-square test statistics for the restricted and unrestricted models.

We may now discuss the WT. Observe that $\underline{\theta}_r^R$ is zero by conception for the restrictive model. Moreover, if the restrictive model is true, $\hat{\underline{\theta}}_u^R$ is also expected to be zero within sampling error. Let (see, for example, Buse, 1982; and Bollen, 1989)

$$(\hat{\underline{\theta}}_u^R)'\cdot\left\{\left[\frac{\partial\hat{\underline{\theta}}_u^R}{\partial\hat{\underline{\theta}}_{ML,u}}\right]'\cdot\left[a\operatorname{cov}(\hat{\underline{\theta}}_{ML,u})\right]\cdot\left[\frac{\partial\hat{\underline{\theta}}_u^R}{\partial\hat{\underline{\theta}}_{ML,u}}\right]'\right\}^{-1}\cdot(\hat{\underline{\theta}}_u^R) \qquad (6.36)$$

be the WT statistic, which establishes how much $\hat{\underline{\theta}}_{ML,u}$ deviates from the restrictions enforced by the nested model. In (6.36), $a\operatorname{cov}(\hat{\underline{\theta}}_{ML,u})$ is an estimator of the asymptotic covariance matrix of $\hat{\underline{\theta}}_{ML,u}$ while the term in braces { . } is an estimator of the asymptotic covariance matrix of $\hat{\underline{\theta}}_u^R$. The WT statistic given by (6.36) has asymptotically a $\chi^2_{b^*}$ distribution (Buse, 1982; and Bollen, 1989), where b^* is as defined above.

Notice that the three significance tests of difference in model fit statistics for nested models, reviewed above, present the power limitations that we discussed earlier for the case of the chi-square tests (Bollen, 1989). Moreover, we acknowledge here that Satorra (2000) and Satorra and Bentler (2001) have extended the Satorra and Bentler (1988) corrections, discussed in the Remark 6.2, to LRT, LMT and WT considering a disaggregated analysis approach. Furthermore Yanagihara, Tonda and Matsumoto (2005) have recently provided some further theoretical investigation on the effects of non-normality on asymptotic distributions of some likelihood ratio criteria for testing covariance structures under the normal assumption.

We shall additionally consider in this section some alternative approaches for performing model selection. We shall then consider in this sub-section some further descriptive ad hoc measures of overall model fit, which were developed following the fact that LRT statistic could present different values for different sample sizes, for a given difference between the 'initial' model and the restricted model (Matsueda and Bielby, 1986).

We may thus discuss the Jöreskog and Sörbom (1989) goodness of fit indices (GFI), given by

$$\text{GFI}_{ML} = 1 - \left\{ \frac{tr[(\Sigma(\hat{\underline{\theta}})^{-1}\text{S} - I)^2]}{tr[(\Sigma(\hat{\underline{\theta}})^{-1}\text{S})^2]} \right\}, \tag{6.37}$$

when considering $F(\underline{\theta})_{ML}$, where I is a $T \times T$ identity matrix. This index computes relatively the magnitude of the variances and covariances in S, predicted by $\Sigma(\hat{\underline{\theta}})$ when adopting ML estimators (see also Bollen, 1989). Jöreskog and Sörbom (1989) also proposed versions of the GFI for $F(\underline{\theta})_{ULS}$, which is given by

$$\text{GFI}_{ULS} = 1 - \left\{ \frac{tr[(\text{S} - \Sigma(\hat{\underline{\theta}}))^2]}{tr[(\text{S})^2]} \right\}. \tag{6.38}$$

A GLS version of the goodness of fit indices was proposed by Tanaka and Huba (1985), and is given by (see also Bollen, 1989)

$$\text{GFI}_{GLS} = 1 - \left\{ \frac{tr[(I - \Sigma(\hat{\underline{\theta}})\text{S}^{-1})^2]}{T} \right\}, \tag{6.39}$$

or equivalently (see also Jöreskog and Sörbom, 1989; and Jöreskog and Sörbom, 1997)

$$\text{GFI}_{GLS} = 1 - \left\{ \frac{\{vech[\text{S}] - vech[\Sigma(\hat{\underline{\theta}})]\}' \text{U}^{-1} \{vech[\text{S}] - vech[\Sigma(\hat{\underline{\theta}})]\}}{vech[\text{S}]' \text{U}^{-1} vech[\text{S}]} \right\}, \tag{6.40}$$

which has as numerator the GLS fitting function given by expression (4.46), introduced in Chapter 4, Section 4.4, Subsection 4.4.2, when evaluated at $\hat{\underline{\theta}}_{GLS}$.

Jöreskog and Sörbom (1989) proposed additionally an adjusted fit index (AGFI), which penalises the models with more parameters, and is given by (see also Jöreskog and Sörbom, 1997; Matsueda and Bielby, 1986; and Bollen, 1989)

$$\text{AGFI} = 1 - \left(\frac{k}{\text{df}} \right) \cdot (1 - \text{GFI}), \tag{6.41}$$

where k is as defined in Chapter 4, Section 4.1. Observe that both GFI and AGFI achieve their maximum of one when S is perfectly predicted by $\Sigma(\hat{\underline{\theta}})$[34], and that they do not explicitly depend upon the sample size. Nevertheless, according to Jöreskog and Sörbom (1997) and Bollen (1989), the sampling distribution of these measures may be influenced by the size of n.

[34] Authough these measures are expected to be between zero and one, according to Jöreskog and Sörbom (1989) and Jöreskog and Sörbom (1997), it is in theory possible, in situations where the models fits worse than 'no model at all', for them to become negative.

We do not expect to have entirely reviewed the vast literature on covariance structure model fitting evaluation procedures in this sub-section. See, for example: (*i*) Browne (1982, Section 1.8) for information on a cross validation approach for model selection; and (*ii*) Matsueda and Bielby (1986), and Bollen (1989) for material on the Critical N statistic proposed by Hoelter (1983); which shall both not be discussed here.

6.4.2 Model testing under complex sampling

In this sub-section, we shall consider some further developments on covariance structure model fitting statistics when assuming that the sample is selected under the complex survey design approach, i.e. we shall not assume that the observations are iid. In covariance structure modelling, accurate computation of fitting statistics is as essential as estimation of standard errors of point estimators (Muthén and Satorra, 1995). Furthermore, according to Skinner, Holt and Smith (1989), ignoring the characteristics of the complex samples can lead us to calculate invalid statistical tests.

For calculating model fit measures in the present context we shall adopt S_w, the weighted sample covariance matrix, as defined in Chapter 4, Sub-section 4.3.1, as an estimator of Σ, recalling again that $E(E_p(S_w)) \doteq E(S_N) = \Sigma$ (see Chapter 4, Section 4.3, Sub-section 4.3.1). We shall then consider model fit measures which are functions of S_w and Σ. Note that in the current context, overall model fit measures also have the same limitations discussed by Bollen (1989), and Jöreskog and Sörbom (1997), and reviewed in the previous sub-section. Moreover, we shall work here with the same H_0, and a generic alternative hypothesis H_1, as defined in Sub-section 6.4.1.

For examining (6.19), i.e. for identifying components of the variance-covariance matrix that are not well fit, we shall now adopt the sample weighted *residual covariance matrix* $\hat{E}_c$, defined as

$$\hat{E}_c = S_w - \Sigma(\underline{\hat{\theta}}),$$

where the subscript c denotes 'complex'. See Chapter 4, Section 4.3, Sub-section 4.3.2. Moreover, let

$$[S_{w,tt'} - \Sigma(\underline{\hat{\theta}})_{tt'}]$$

be the individual weighted sample residual covariances, where $S_{w,tt'}$ is the tt'th element in S_w and $\Sigma(\hat{\underline{\theta}})_{tt'}$ is as defined in Sub-section 6.4.1. The RMR measure proposed by Jöreskog and Sörbom (1989) may be adapted to

$$\text{RMR}_c = \sqrt{2 \cdot \sum_{t=1}^{T} \sum_{t'=1}^{t} \frac{[S_{w,tt'} - \Sigma(\hat{\underline{\theta}})_{tt'}]^2}{T(T+1)}},$$

which has as special case the RMR measure when the sampling weights are constant, and could also be utilised for comparing the fit of two different models for the same data.

Theory developed in the categorical complex survey data analysis and modelling literature (see Rao and Scott, 1979; followed, for example, by Rao and Scott, 1981; Rao and Scott, 1984; Rao and Scott, 1987; Rao and Thomas, 1988; Rao and Thomas, 1989; Skinner, 1989a; and more recently, Rao and Thomas, 2003) suggests that, under complex sampling, $n \cdot F(\underline{\theta})_{PML}$, where $F(\underline{\theta})_{PML}$ is the PML fitting function given by (4.62) from Chapter 4, Section 4.4, Sub-section 4.4.5, would not be asymptotically chi-squared distributed. In that context, simple corrections have been proposed, which may be adapted so that they could be applied to $n \cdot F(\underline{\theta})_{PML}$ in order to make it approximately chi-squared distributed. Note that we shall be considering below the complex survey data set up, and following an approach proposed by Skinner (1989a, Section 3.4).

Remark 6.3: We shall initially consider the case of a GLSC type estimator, obtained by minimizing the $F(\underline{\theta})_{GLSC}$ fitting function given by expression (4.49), introduced in Chapter 4, Section 4.4, Subsection 4.4.3, with matrix U given by

$$\text{U} = 2 \cdot \text{K}'(\text{W} \otimes \text{W})\text{K}, \qquad (6.42)$$

where W is any consistent estimator of Σ. Under this situation, a Wald goodness of fit test statistic is given by (Skinner, 1989a)

$$X^2_{\text{W},srs} = n \cdot \{vech[\text{S}_w] - vech[\Sigma(\theta)]\}' \text{U}^{-1} \{vech[\text{S}_w] - vech[\Sigma(\theta)]\},$$

which implies, (see Chapter 4, Section 4.4, Sub-section 4.4.3)

$$X^2_{\text{W},srs} = n \cdot \left(\frac{1}{2}\right) \cdot tr\{[I - \Sigma(\underline{\theta})\text{S}_w^{-1}]^2\},$$

when S_w is considered, for example, as a choice for W in (6.42). Note that the subscript srs is added above to emphasize that U is defined under simple random sampling assumptions.

Note, nevertheless, that according to Skinner (1989a), the test statistic $X^2_{\text{W},srs}$ is no longer asymptotically chi-squared distributed, but in fact asymptotically (see also Rao and Scott, 1981; Rao and Scott, 1984; Rao and Scott, 1987; Rao and Thomas, 1988; Rao and Thomas, 1989; Rao and Thomas, 2003)

$$X^2_{\mathrm{W},srs} \sim \sum_{d=1}^{k-b} \lambda_d \chi^2_1$$

under H_0, where χ^2_1 are independent chi-squared distributed random variables, and λ_d are non-zero eigenvalues of

$$\mathrm{H} = \mathrm{U}^{-1}\mathrm{C}_c - \mathrm{U}^{-1}\mathrm{C}_c\mathrm{U}^{-1} \cdot \Delta(\Delta'\mathrm{U}^{-1}\Delta)^{-1}\Delta',$$

where C_c is the asymptotic covariance matrix of $vech[\mathrm{S}_w]$. As a one moment approximation, Skinner (1989a)

$$\frac{(k-b)\cdot X^2_{\mathrm{W},srs}}{tr(\mathrm{H})}, \tag{6.43}$$

is asymptotically distributed as a χ^2_{k-b}, and may be adopted for testing the goodness of fit of a covariance structure model when assuming that the sample is selected under the complex survey design approach.

Note that as $F(\underline{\theta})_{PML}$ is asymptotic equivalent to $F(\underline{\theta})^2_{GLSC}$ (see Chapter 4, Section 4.4, Sub-section 4.4.5; see also Fuller, 1987, p. 334-335), given by (4.51) in Chapter 4, Section 4.4, Sub-section 4.4.3, the approach proposed above is also valid for situation where $F(\underline{\theta})_{PML}$ is adopted for estimating $\underline{\theta}$.

We could also consider substituting matrix U by $\hat{\mathrm{C}}_c$, i.e.

$$X^2_{\mathrm{W}} = n \cdot \{vech[\mathrm{S}_w] - vech[\Sigma(\theta)]\}' \hat{\mathrm{C}}_c \{vech[\mathrm{S}_w] - vech[\Sigma(\theta)]\},$$

where $\hat{\mathrm{C}}_c$, given by (6.09b), is a consistent estimator for C_c (Skinner and Holmes, 2003), which was defined in (6.08). In this context, X^2_{W} is approximately distributed as χ^2_{k-b} under H_0 (Skinner, 1989a; and Skinner and Holmes, 2003). Note that when U is consistent for C_c,

$$X^2_{\mathrm{W}} = X^2_{\mathrm{W},srs} = \frac{(k-b)\cdot X^2_{\mathrm{W},srs}}{tr(\mathrm{H})},$$

as $tr(\mathrm{H}) = k-b$ in that situation. ■

Furthermore, in our context we shall consider methods for non-normal data which, according to Skinner (1989a), followed by Satorra (1992) and Muthén and Satorra (1995), are special cases of techniques for complex survey data (see also Chapter 4).

Remark 6.4: Let us assume here that (*i*) the model is correct, and that (*ii*) $\hat{C}_c$, given by (6.09b), is consistent for C_c. It is also assumed here that $\hat{\underline{\theta}}$ is an estimator (of $\underline{\theta}$) which, as $n \to \infty$, $\sqrt{n} \cdot (\hat{\underline{\theta}} - \underline{\theta})$ has an asymptotic normal distribution with a null mean vector and a finite covariance matrix. We shall assume that both GLSC and PML estimators, proposed in Chapter 4, have the above property, when considering regularity conditions described in Browne (1984). Hence, we could propose modifying (6.26) by substituting $\hat{C}$ by $\hat{C}_c$, given by (6.09b), and $\hat{E}$ by $\hat{E}_c$, so that the test statistic becomes

$$X_c = n \cdot vech[\hat{E}_c]' \cdot \{\hat{C}_c^{-1} - \hat{C}_c^{-1} \cdot \hat{\Delta}(\hat{\Delta}'\hat{C}_c^{-1}\hat{\Delta})^{-1}\hat{\Delta}'\hat{C}_c^{-1}\} \cdot vech[\hat{E}_c], \qquad (6.44)$$

which, following earlier argument from Remark 6.2 has χ^2_{k-b} asymptotic distribution under H_0, when assuming (*i*) and (*ii*) included above, where $\hat{\Delta}$ is given by (6.03), with $\hat{\Delta}$ evaluated at $\underline{\theta} = \hat{\underline{\theta}}$ (either $\hat{\underline{\theta}}_{PML}$ or $\hat{\underline{\theta}}_{GLSC}$). Observe that (6.44), as well as (6.26), may also be affected by the fact that the ADF estimator $\hat{C}_c$ could produce unreliable results when the sample size is not large enough and T is large (see Section 6.4.1; and Remark 6.2), as discussed earlier in this thesis.

Furthermore, we also propose modifying the corrections developed by Satorra and Bentler (1988) for standard test statistics in the classical context, so that those could be used when working with complex survey data. We shall propose substituting $\hat{C}_c$ for $\hat{C}$ in (6.28), with $\hat{C}_c$ given by (6.09b), so that (see Remark 6.2)

$$X_{S,c} = (\hat{\eta}_c^*)^{-1} \cdot n \cdot F(\hat{\underline{\theta}}),$$

with (6.45)

$$\hat{\eta}_c^* = tr\{[U^{-1} - U^{-1} \cdot \hat{\Delta}(\hat{\Delta}'U^{-1}\hat{\Delta})^{-1}\hat{\Delta}'U^{-1}] \cdot \hat{C}_c\}/(k-b),$$

where we shall still considerer that matrix U is as defined in (6.42). Note that the Satorra and Bentler modified test statistic proposed in (6.45) is equivalent to test statistic we proposed in (6.43), following Skinner (1989a) approach. Therefore, we may confirm that $X_{S,c}$ is also asymptotically chi-squared distributed with $(k-b)$ degrees of freedom.

Moreover, we could follow a similar strategy and substitute $\hat{C}_c$ for $\hat{C}$ also in (6.30) and (6.31). It follows that

$$X_{A,c} = (\hat{\eta}_c^+)^{-1} \cdot n \cdot F(\hat{\underline{\theta}}),$$

with

$$\hat{\eta}_c^{+} = \frac{d_c}{tr\{[\mathrm{U}^{-1} - \mathrm{U}^{-1}\cdot\hat{\Delta}(\hat{\Delta}'\mathrm{U}^{-1}\hat{\Delta})^{-1}\hat{\Delta}'\mathrm{U}^{-1}]\cdot\hat{\mathrm{C}}_c\}},$$

where d_c denotes the nearest integer to

$$d_c^{*} = \frac{\{tr[(\mathrm{U}^{-1} - \mathrm{U}^{-1}\cdot\hat{\Delta}(\hat{\Delta}'\mathrm{U}^{-1}\hat{\Delta})^{-1}\hat{\Delta}'\mathrm{U}^{-1})\cdot\hat{\mathrm{C}}_c]\}^2}{tr[(\mathrm{U}^{-1} - \mathrm{U}^{-1}\cdot\hat{\Delta}(\hat{\Delta}'\mathrm{U}^{-1}\hat{\Delta})^{-1}\hat{\Delta}'\mathrm{U}^{-1})\cdot\hat{\mathrm{C}}_c]^2}.$$

Note that the modifications we propose above have as special case those modifications developed by Satorra (1992), where a similar approach is suggested in a context without covariates. From argumentation in Satorra (1992), and Muthén and Satorra (1995), $X_{A,c}$ should be distributed as χ_d. ■

We may now consider a Wald significance test for nested hypothesis, i.e. difference in model fit chi-square statistics for nested models (discussed in the previous sub-section in the classical case), for situations where the PML fitting function is adopted. Note that we shall follow an approach proposed by Skinner (1989a, Section 3.4).

Remark 6.5: We shall assume that the asymptotic covariance matrix of $vech[\mathrm{S}_w]$ and the information matrix of the model are non-singular. Thus, let $\underline{\hat{\theta}}_{PML,r}$ be the PML estimator for the restrictive (nested) model, and let $\underline{\hat{\theta}}_{PML,u}$ be the PML estimator for the covariance structure model without constraints. Moreover, recall that $\underline{\theta}^R$ is a $b^* \times 1$ vector when $\underline{\theta}^R = \underline{0}$ corresponds to the restraints imposed to the unrestricted model, where $b^* < b$. We propose a modification to the Wald test, given by (6.36), so that

$$\mathrm{WT}_c = \underline{\hat{\theta}}_{PML,u}^{R}{}'\cdot\left\{\left[\frac{\partial\underline{\hat{\theta}}_{PML,u}^{R}}{\partial\underline{\hat{\theta}}_{PML,u}}\right]'\cdot\left[a\,\mathrm{cov}(\underline{\hat{\theta}}_{PML,u})\right]\cdot\left[\frac{\partial\underline{\hat{\theta}}_{PML,u}^{R}}{\partial\underline{\hat{\theta}}_{PML,u}}\right]'\right\}^{-1}\cdot\underline{\hat{\theta}}_{PML,u}^{R}, \qquad (6.46)$$

where we could adopt the approach of Binder (1983), adapted to the covariance structure models context proposed earlier in Section 6.3, for calculating $a\,\mathrm{cov}(\underline{\hat{\theta}}_{PML,u})$. If the null hypothesis is true, and if $\sqrt{n^{-1}}\cdot(\underline{\hat{\theta}}_{PML} - \underline{\theta})$ is asymptotically normal distributed (see Chapter 4, Section 4.4), then the modified WT_c statistic introduced above should be asymptotically $\chi^2_{b^*}$ distributed (Skinner, 1989a, p. 84) under H_0. ■

We may still propose in the current sub-section modifying the overall model fit descriptive measures, such as the Jöreskog and Sörbom (1989) goodness of fit indices (GFI) discussed in the previous sub-section. Let

$$\mathrm{GFI}_{c,PML} = 1 - \left\{ \frac{tr[(\Sigma(\hat{\underline{\theta}})^{-1}\mathrm{S}_w - I)^2]}{tr[(\Sigma(\hat{\underline{\theta}})^{-1}\mathrm{S}_w)^2]} \right\},$$

be a modified version GFI for complex survey data, when considering $F(\underline{\theta})_{PML}$. Observe that the GFI_c calculates the relative magnitude between S_w and $\Sigma(\hat{\underline{\theta}})$ when adopting the PML estimator (see also Bollen, 1989). In a similar way we may also propose a GFI version for the $F(\underline{\theta})_{ULSC}$, given by

$$\mathrm{GFI}_{c,ULS} = 1 - \left\{ \frac{tr[(\mathrm{S}_w - \Sigma(\hat{\underline{\theta}}))^2]}{tr[(\mathrm{S}_w)^2]} \right\}.$$

Furthermore, we could also modify Tanaka and Huba (1985) GLS version of the goodness of fit indices, so that

$$\mathrm{GFI}^1_{c,GLSC} = 1 - \left\{ \frac{tr[(I - \Sigma(\hat{\underline{\theta}})\mathrm{S}_w^{-1})^2]}{T} \right\},$$

or

$$\mathrm{GFI}^2_{c,GLSC} = 1 - \left\{ \frac{\{vech[\mathrm{S}_w] - vech[\Sigma(\hat{\underline{\theta}})]\}' \mathrm{U}^{-1} \{vech[\mathrm{S}_w] - vech[\Sigma(\hat{\underline{\theta}})]\}}{vech[\mathrm{S}_w]' \mathrm{U}^{-1} vech[\mathrm{S}_w]} \right\}.$$

Modified adjusted fit indices (AGFI_c), could thus be calculated as

$$\mathrm{AGFI}_c = 1 - \left(\frac{k}{\mathrm{df}} \right) \cdot (1 - \mathrm{GFI}_c).$$

6.5 Discussion

In this chapter we have proposed variance estimation procedures for covariance structure models point estimators, in the context of longitudinal complex survey data. We have discussed methods for variance estimation for GLS and PML point estimators of $\underline{\theta}$ that take the sampling scheme into account.

Moreover, goodness of fit testing techniques for structural models for covariance matrices have been reviewed for the classical iid observations case. We have then proposed modifications to some test statistics to be used in situations where the complex sampling approach is considered.

We shall present in Chapter 7 the characteristics and results of a second simulation study, which shall have as main objective to evaluate the statistical properties of the variance estimation procedures discussed and proposed in the current chapter. Note that the testing procedures discussed in Section 6.4 shall not be considered in the following chapter as result of time constraints.

Chapter 7

Simulation study II

7.1 Introduction

We presented in Chapter 5 the results of a simulation study, which had the main purpose of evaluating the statistical properties of the point estimation procedures proposed in Chapter 4, and also to compare their properties with those of the traditional methods that were reviewed also in Chapter 4.

The current chapter is an extension of Chapter 5. The variance estimation procedures proposed in Chapter 6 shall have their statistical properties examined in the current chapter. We shall thus present here the characteristics and main results of a second simulation study, where we shall contrast the behaviour of the proposed methods with that of the classic techniques, also discussed in Chapter 6.

In this chapter we shall initially present detailed information on how this simulation study is implemented, which is given in Section 7.2. Moreover, results and some further discussion are presented respectively in sections 7.3 and 7.4, while Section 7.5 includes some concluding remarks.

7.2 Simulation procedures

After programming the variance estimation procedures described in Chapter 6, we shall apply those and assess their statistical properties through the current simulation study.

As in Chapter 5, this simulation study shall involve simulating $d = 1,\ldots,D$ replicate samples. We shall consider again the BHPS subset, which has $n = 1340$, which we considered in Chapter 5, and examined earlier in Chapter 3, for generating simulation samples in the current chapter. Note that we shall hold the values of the x variable as fixed and that the values of Y_{it} shall be simulated from a model, which we shall describe below, independently for each replicate. The superscript *sim* is added again to denote *simulation*. Once more, we shall work with $T = 5$.

As in Chapter 5, for simplicity we shall not pursue to allow for the impact of either stratification or unequal probability sampling in this second simulation study. Furthermore, we shall only evaluate here the properties of variance estimators for unweighted point estimators, described in Chapter 4.

We shall assess the variance estimation methods considering different sample sizes. Moreover, we shall consider a two-stage cluster sampling scheme that involves simple random sampling with replacement of primary sampling units (PSUs). We shall initially select $m^{sim} = 47$, where m^{sim} denotes the number of clusters in the sample. We shall then consider $m^{sim} = 20$ and $m^{sim} = 15$. Let n_j^{sim*} denote the size of cluster j in the original BHPS subset adopted here. In each selected PSU, $j = 1,\ldots,m^{sim}$, n_j^{sim} secondary sampling units (SSUs) shall be selected also by simple random sampling with replacement. We shall consider several scenarios: (*i*) $n_j^{sim} = n_j^{sim*}$, (*ii*) $n_j^{sim} = 15$, (*iii*) $n_j^{sim} = 10$, and (*iv*) $n_j^{sim} = 5$.

Moreover, we shall suppose that the BHPS subset, adopted for generating simulation samples, is grouped in $m = 47$ primary sampling units, instead of the original 248 clusters. See Chapter 3 for further information on (*i*) how the original clusters were combined for creating the new clustering, and (*ii*) the reasons why we have chosen to work with a new aggregated clustering.

Let Y_{ijt} be the value (scalar) for the study variable for unit $i = 1,2,\ldots,n^{sim}$, in cluster $j = 1,\ldots,m^{sim}$, at wave t of the survey. As in Chapter 5, we shall use the following uniform correlation model, which allows for the impact of clustering,

$$Y_{ijt} = \underline{x}_{ijt}\underline{\beta} + \eta_j + u_{ij} + v_{ijt}, \tag{7.1}$$

with

$$\eta_j \sim N\left(0,\sigma_\eta^2\right),\ u_{ij} \sim N\left(0,\sigma_u^2\right), \text{and } v_{ijt} \sim N\left(0,\sigma_v^2\right),$$

for generating the values of Y_{ijt} that we are going to use in the simulation study. As in Chapter 5, we shall refer to the model described above as UCM-C, where C denotes *cluster*. We shall thus generate

$$y_{ijt}^{(d)} = \underline{x}_{ijt}\underline{\beta} + \eta_j^{(d)} + u_{ij}^{(d)} + v_{ijt}^{(d)},$$

where $\eta_j^{(d)}$, $u_{ij}^{(d)}$, and $v_{ijt}^{(d)}$ are as given in Sub-section 5.2.1, Chapter 5.

Note that we shall consider again a gender role attitude score (see Chapter 3, Sub-section 3.2.3) as the dependent variable, and the same covariates adopted and listed in Section 5.2, Chapter 5. We shall adopt the values for $\underline{\beta}$, σ_u^2, σ_v^2, and ρ as those given by (5.6) and (5.7), in Chapter 5. In particular, we shall consider here different choices for σ_η^2,

$$\sigma_\eta^{2\ sim,C} \cong 0.15,$$

$$\sigma_\eta^{2\ sim,C} \cong 0.45,$$

and

$$\sigma_\eta^{2\ sim,C} \cong 0.75,$$

to enable the evaluation of effects of different impacts of clustering on the considered variance estimation procedures.

7.3 Simulation results

Let $\text{var}(\hat{\underline{\theta}})$ be an estimator of our $b \times b$ covariance matrix of interest, $\text{VAR}(\hat{\underline{\theta}})$. Let alternatively $\text{VAR}(\hat{\theta}_j)$ represent the diagonal element of $\text{VAR}(\hat{\underline{\theta}})$ that includes the variance of the estimator for the component θ_j of $\underline{\theta}$. We shall estimate $\text{VAR}(\hat{\theta}_j)$ in this simulation study by

$$\text{VAR}(\hat{\theta}_j) = \frac{1}{D-1}\sum_{d=1}^{D}[\hat{\theta}_j^{(d)} - \hat{E}(\hat{\theta}_j)]^2, \tag{7.2}$$

where

$$\hat{E}(\hat{\theta}_j) = \frac{1}{D}\sum_{d=1}^{D}\hat{\theta}_j^{(d)}.$$

An approximately 95% simulation confidence interval for the true covariance matrix is given by

$$\text{VAR}(\hat{\theta}_j) \pm 1.96 \cdot \sqrt{\frac{\text{var}(z^{(d)})}{D}},$$

where

$$z^{(d)} = [\hat{\theta}_j^{(d)} - \hat{E}(\hat{\theta}_j)]^2$$

and

$$\text{var}(z^{(d)}) = \frac{1}{D-1}\sum_{d=1}^{D}[z^{(d)} - \bar{z}],$$

with

$$\bar{z} = \frac{1}{D}\sum_{d=1}^{D} z^{(d)}.$$

Moreover, let

$$\text{MSE}[\text{var}(\hat{\theta}_j)] = \text{VAR}[\text{var}(\hat{\theta}_j)] + \{\text{BIAS}[\text{var}(\hat{\theta}_j)]\}^2, \tag{7.3}$$

where $\mathrm{VAR}[\mathrm{var}(\hat{\theta}_j)]$ is simulation variance of $\mathrm{var}(\hat{\theta}_j)$, $\mathrm{BIAS}[\mathrm{var}(\hat{\theta}_j)]$ denotes the simulation estimate of the bias of the variance estimator, and may be calculated by

$$\mathrm{bias}[\mathrm{var}(\hat{\theta}_j)] = \hat{E}[\mathrm{var}(\hat{\theta}_j)] - \mathrm{VAR}(\hat{\theta}_j), \tag{7.4}$$

where $\hat{E}[\mathrm{var}(\hat{\theta}_j)]$ is the simulation estimated expected value of $\mathrm{var}(\hat{\theta}_j)$. We shall provide information on how $\mathrm{VAR}[\mathrm{var}(\hat{\theta}_j)]$ and $\hat{E}[\mathrm{var}(\hat{\theta}_j)]$ are calculated later in the current section. Recall that $\mathrm{var}(\hat{\theta}_j)$ represents a diagonal element of $b \times b$ covariance matrix estimator $\mathrm{var}(\underline{\hat{\theta}})$.

For evaluating the impacts of simulation errors, an approximately 95% simulation confidence interval for the bias of the variance estimator is given by

$$\mathrm{bias}[\mathrm{var}(\hat{\theta}_j)] \pm 1.96 \cdot \sqrt{\frac{\mathrm{var}\left(z'^{(d)}\right)}{D}}, \tag{7.5}$$

where

$$z'^{(d)} = \mathrm{var}(\hat{\theta}_j)^{(d)} - z^{(d)}$$

and

$$\mathrm{var}\left(z'^{(d)}\right) = \frac{1}{D-1} \sum_{d=1}^{D} \left[z'^{(d)} - \bar{z}'\right],$$

with

$$\bar{z}' = \frac{1}{D} \sum_{d=1}^{D} z'^{(d)}.$$

Furthermore, let $\mathrm{rel\ bias}[\mathrm{var}(\hat{\theta}_j)]$ denote the estimated relative bias of $\mathrm{var}(\hat{\theta}_j)$, so that

$$\mathrm{rel\ bias}[\mathrm{var}(\hat{\theta}_j)] = \frac{\mathrm{bias}[\mathrm{var}(\hat{\theta}_j)]}{\mathrm{VAR}(\hat{\theta}_j)} \cdot 100.$$

Note that, as in Muthén and Satorra (1995), we shall not consider in this study a relative bias of less than 10% to be 'practically significant'.

For each replicate we shall estimate $\mathrm{var}(\hat{\theta}_j)^{(d)}$, via each of the variance estimation methods included in Chapter 6. We may then calculate

$$\hat{E}[\mathrm{var}(\hat{\theta}_j)] = \frac{1}{D} \sum_{d=1}^{D} \mathrm{var}(\hat{\theta}_j)^{(d)}. \tag{7.6}$$

Moreover,

$$\mathrm{var}[\mathrm{var}(\hat{\theta}_j)] = \frac{1}{D-1} \sum_{d=1}^{D} \left[\mathrm{var}(\hat{\theta}_j)^{(d)} - \hat{E}[\mathrm{var}(\hat{\theta}_j)]\right]^2,$$

where $\mathrm{var}[\mathrm{var}(\hat{\theta}_j)]$ is a simulation estimator of $\mathrm{VAR}[\mathrm{var}(\hat{\theta}_j)]$ in (7.3).

The $MSE[\text{var}(\hat{\theta}_j)]$ may thus be estimated by

$$\text{mse}[\text{var}(\hat{\theta}_j)] = \text{var}[\text{var}(\hat{\theta}_j)] + \{\text{bias}[\text{var}(\hat{\theta}_j)]\}^2,$$

the standard error of $\hat{E}[\text{var}(\underline{\hat{\theta}})]$ by

$$se\left[\hat{E}[\text{var}(\hat{\theta}_j)]\right] = \sqrt{\text{var}[\text{var}(\hat{\theta}_j)]} \Big/ \sqrt{D},$$

and an approximately 95% simulation confidence interval for $\text{VAR}(\hat{\theta}_j)$ may be defined as

$$\hat{E}[\text{var}(\hat{\theta}_j)] \pm 1.96 \cdot se\left[\hat{E}[\text{var}(\hat{\theta}_j)]\right].$$

Furthermore, the coefficient of variation of $\text{var}(\underline{\hat{\theta}})$ is calculated by

$$cv\left[\text{var}(\hat{\theta}_j)\right] = \frac{\sqrt{\text{var}[\text{var}(\hat{\theta}_j)]}}{\hat{E}[\text{var}(\hat{\theta}_j)]} \cdot 100.$$

We aim to present in this chapter results based on $D = 10{,}000$ replicates. After some investigation our results strongly suggested that a smaller number of replicates, as for example $D = 1000$ or $D = 5000$ was not large enough for providing reliable and stable simulation estimates, given by (7.2), of the true covariance matrix in this simulation study. Recall that in Chapter 5 we have adopted an iterative numerical method (Newton type algorithm; see Section 5.3, Sub-section 5.3.1) for carrying out numerical derivatives and perform the necessary minimisations of fitting functions *F*.

The use of several other alternatives for performing the necessary numerical minimizations was also considered: (*i*) a Nelder and Mead (1965) method; (*ii*) a quasi-Newton method or variable metric algorithm, proposed simultaneously by Broyden, Fletcher, Goldfarb and Shanno in 1970 (see Nocedal and Wright, 1999); (*iii*) a conjugate gradients method (Fletcher and Reeves, 1964); (*iv*) Byrd *et al.* (1995) method, which is a modification of method (*ii*); and (*v*) a stochastic global optimization method proposed by Belisle (1992). All these methods are also implemented in the software *R*, in the function **optim**. In essence, we may report here that methods (*i*) to (*iii*) provided virtually the same result as those given by the adopted Newton type algorithm, while methods (*iv*) and (*v*) had difficulties in yielding converged solutions even for the largest sample size considered.

Recall that analytic solutions for PML (and ML) estimation methods have been produced when fitting a UCM model, and are included in Appendix C. Thus, as in Chapter 5, we shall perform here necessary minimisations for the GLS method by adopting an iterative numerical method. Note that we have evaluated the accuracy of the adopted numerical minimisation procedure for the evaluation of variance estimators in this simulation study by calculating

(7.2) and (7.6) for the PML (and ML) method, considering both analytic and numerical methods. Difference was found only on the seventh decimal place for $D = 10{,}000$, which suggests that the adopted Newton type algorithm provides enough precision.

Note that, as in Chapter 5, each evaluated method shall be analysing exactly the same data for each situation considered in this second simulation study. We shall again be fitting for each d replicate a UCM (see Chapter 2, Example 2.1; and Chapter 5, Section 5.3) model. We shall then evaluate the variance estimation methods considering Y_{ijt} generated by a UCM-C model, as described in Section 7.2.

Let $\text{var}_{\text{n}}(\hat{\underline{\theta}}_{ML})$, where the subscript n denotes naïve, be a linearization variance estimator for the $\hat{\underline{\theta}}_{ML}$ point estimator under the assumption of simple random sampling, based upon expressions (6.13b) and (6.17) in Chapter 6, Section 6.3, when considering that the population consists of only one stratum, and that each individual i is a PSU. For information on how $\hat{\underline{\theta}}_{ML}$ is estimated, see Chapter 4, Sub-section 4.4.4.

Let $\text{var}_{c}(\hat{\underline{\theta}}_{ML})$, where c denotes complex, be a linearization variance estimator for $\hat{\underline{\theta}}_{ML}$, which is based upon expressions (6.13b) and (6.17) when considering a two-stage sample design, with m primary sampling units (PSUs) being selected with replacement, and $H = 1$, where H is the number of stratum in the sample, and constant sampling weights.

Let $\text{var}_{n}(\hat{\underline{\theta}}^{2}_{GLS})$ be a variance estimator for $\hat{\underline{\theta}}^{2}_{GLS}$, when adopting (6.06) from Chapter 6, Section 6.2, with matrix U given by (4.24) in Chapter 4, Section 4.3, Sub-section 4.3.2, when considering the unweighted covariance estimator S to be a consistent estimator of Σ. Note that the estimator $\hat{\underline{\theta}}^{2}_{GLS}$ is given by (4.48a), in Chapter 4, Section 4.4, Sub-section 4.4.2.

Let $\text{var}_{adf}(\hat{\underline{\theta}}^{2}_{GLS})$ be a distribution free (ADF) variance estimator for $\hat{\underline{\theta}}^{2}_{GLS}$, when adopting expression (6.05) from Chapter 6, Section 6.2, with matrix U given by (4.24), in Chapter 4, Section 4.3, Sub-section 4.3.2, when considering S to be a consistent estimator of Σ, and matrix C being estimated by (4.33). Note that $\text{var}_{adf}(\hat{\underline{\theta}}^{2}_{GLS})$ does not consider the complex sampling design when estimating $\text{VAR}(\hat{\underline{\theta}}^{2}_{GLS})$.

Let $\text{var}_{c}(\hat{\underline{\theta}}^{2}_{GLS})$ be an ADF variance estimator for $\hat{\underline{\theta}}^{2}_{GLS}$ that takes the complex sampling scheme into account, when adopting expression (6.09a) from Chapter 6, Section 6.2, with matrix U given by (4.24), in Chapter 4, Section 4.3, Sub-section 4.3.2, when considering S to be a consistent estimator of Σ, with C being estimated by (6.09b), considering a two-stage sample design as described above for $\text{var}_c(\hat{\underline{\theta}}_{ML})$. As suggested by Muthén and Satorra (1995), we shall evaluate here distribution-free variance estimators for normal theory point estimators.

Recall that we are not allowing for unequal probability sampling in this simulation study. Therefore, we shall only gauge here the statistical properties of variance estimators for unweighted point estimators.

Moreover, note that this chapter only presents summarised simulation results. Additionally, all bias results that were found not to be significantly different from zero at the 95% level are flagged with "NS", when the interval given by (7.5) includes the value zero. As in Chapter 5, that shall illustrate circumstances where the simulations do not provide sufficient evidence that the variance estimator is biased. Note that more detailed tables with results are confined to Appendix F, when considering $\sigma_\eta^{2\ sim,C} \cong 0.75$. Complete results for the remaining considered values of σ_η^2 have also been calculated, although they shall not be presented in the thesis.

We shall include initially in Table 7.1 results which were produced when considering $m^{sim} = 47$ and $n_j^{sim} = n_j^{sim*}$, and different choices for σ_η^2.

Estimator		rel bias			$cv(\text{var}(\hat{\underline{\theta}}))$		
		$\sigma_\eta^2 = 0.15$	$\sigma_\eta^2 = 0.45$	$\sigma_\eta^2 = 0.75$	$\sigma_\eta^2 = 0.15$	$\sigma_\eta^2 = 0.45$	$\sigma_\eta^2 = 0.75$
$\text{var}_n\left(\hat{\underline{\theta}}_{ML}\right)$	$\text{var}(\hat{\sigma}_u^2)$	-1.79%NS	-9.52%	-19.89%	11.04%	11.33%	11.65%
	$\text{var}(\hat{\sigma}_v^2)$	-1.68%NS	-3.97%	-4.56%	7.65%	7.57%	7.61%
$\text{var}_n\left(\hat{\underline{\theta}}^2_{GLS}\right)$	$\text{var}(\hat{\sigma}_u^2)$	-2.51%NS	-10.15%	-20.64%	8.92%	9.18%	9.65%
	$\text{var}(\hat{\sigma}_v^2)$	-6.25%	-9.37%	-9.26%	5.84%	5.85%	5.83%
$\text{var}_{adj}\left(\hat{\underline{\theta}}^2_{GLS}\right)$	$\text{var}(\hat{\sigma}_u^2)$	-2.40%NS	-10.15%	-20.63%	11.07%	11.36%	11.69%
	$\text{var}(\hat{\sigma}_v^2)$	-4.69%	-7.84%	-7.79%	7.67%	7.60%	7.64%
$\text{var}_c\left(\hat{\underline{\theta}}_{ML}\right)$	$\text{var}(\hat{\sigma}_u^2)$	-1.14%NS	-2.81%	-4.62%	24.14%	25.90%	29.04%
	$\text{var}(\hat{\sigma}_v^2)$	-0.70%NS	-3.10%	-3.84%	22.43%	22.98%	23.04%
$\text{var}_c\left(\hat{\underline{\theta}}^2_{GLS}\right)$	$\text{var}(\hat{\sigma}_u^2)$	-1.75%NS	-3.51%	-5.56%	24.33%	26.02%	29.09%
	$\text{var}(\hat{\sigma}_v^2)$	-3.77%	-6.99%	-7.17%	22.82%	23.37%	23.35%

NS – denotes 'absolute bias not significantly different from zero at 95% level'.

Table 7.1 – Evaluation of $\text{var}(\hat{\underline{\theta}})$ considering $m^{sim} = 47$ and $n_j^{sim} = n_j^{sim*}$.

Simulation results presented above allow us to make several observations. Methods that do not take the sampling scheme into account, i.e. the first three evaluated methods, clearly underestimate the variance. This result agrees with the survey sampling literature (see, for example, Kish and Frankel, 1974; and Skinner, Holt and Smith, 1989) and thus confirms the relevance of further investigating complex survey sampling issues when performing covariance structure modelling (see also Muthén and Satorra, 1995). Furthermore, we may notice here that the underestimation tends to increase rapidly with increases in σ_{η}^{2}, i.e. for larger the impact of clustering there is an inflation in the downward relative bias.

Yet regarding methods that do not account for clustering, the naïve variance estimation method for $\underline{\hat{\theta}}_{ML}$ seems to be not very sensitive to small impacts of clustering. For $\sigma_{\eta}^{2} = 0.15$, this method performed satisfactorily and did not seriously underestimate the variance. Nevertheless, for larger values of σ_{η}^{2} this method presents very similar levels of relative bias as the other methods that do not consider the sampling design, especially for $\mathrm{var}(\hat{\sigma}_{u}^{2})$. The ADF variance estimators for the GLS point estimator $\underline{\hat{\theta}}_{GLS}^{2}$ seems to be performing slightly better in terms of relative bias at lower levels of clustering than the naïve estimator $\mathrm{var}_{n}(\underline{\hat{\theta}}_{GLS}^{2})$, although that is not clearly true when higher impacts of clustering are permitted.

Both methods that allow for clustering and take the sampling design into account have achieved noticeable improvements in terms of relative bias, when compared to methods that ignore the sampling scheme characteristics. Methods that perform variance estimation under the complex survey approach seem, nevertheless, to be still biased downwards as expected (Wolter, 1985, Chapter 8; and Kott, 1991; see also Chapter 2, Section 2.5, Sub-section 2.5.1). Those methods also appear to be sensitive in terms of bias to increases in the impact of clustering, although this inflation in bias is perceptibly not as accelerated as it is for the case of methods that do not take the sampling scheme into account.

Furthermore, we may observe that methods that consider the sampling design presented larger variance than the remaining classic methods as expected (see for example, Kott, 1991; and Korn and Graubard, 1995), as result of reduced degrees of freedom for the variance estimation. See also Raj (1968, Chapter 9) for some discussion on the relationship between the stability of a variance estimator and the number of degrees of freedom involved. Moreover, the *cv*s for both $\mathrm{var}_{c}(\underline{\hat{\theta}}_{ML})$ and $\mathrm{var}_{c}(\underline{\hat{\theta}}_{GLS}^{2})$ appear to have a slight tendency of increasing with larger impacts of clustering. This pattern however cannot be observed for the coefficient of variation of methods that do not account for clustering, which tend to have their variance reasonably stable with inflations in σ_{η}^{2}.

Table 7.2 includes results that were produced when considering $m^{sim} = 47$ and $n_j^{sim} = 15$, and different choices for σ_{η}^{2}.

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

^{NS} – denotes 'absolute bias not significantly different from zero at 95% level'.

Table 7.2 – Evaluation of $\mathrm{var}(\hat{\underline{\theta}})$ considering $m^{sim} = 47$ and $n_j^{sim} = 15$.

Under the current considered simulation situation, with 15 secondary sampling units selected from each primary sampling unit, we may again note that methods that do not take the sampling scheme into account generally underestimate the variance, especially for larger values of σ_{η}^{2}, i.e. for larger impacts of clustering. Once more, the naïve variance estimation method for $\hat{\underline{\theta}}_{ML}$ appear not to be very sensitive to small impacts of clustering, even though for larger values of σ_{η}^{2} this method behaves very similarly to other methods that do not consider the sampling design, at least for $\mathrm{var}(\hat{\sigma}_u^2)$.

Variance estimation methods that take the sampling scheme into account tend to lead to noticeable improvements in terms of relative bias when compared to methods that ignore the sampling scheme characteristics, especially for $\sigma_{\eta}^{2} = 0.45$ and $\sigma_{\eta}^{2} = 0.75$. These methods under the current situation, distinctively from the previous considered scenario, do not appear to be clearly sensitive in terms of bias to increases in σ_η^2. Moreover, we may observe again here that methods that consider the sampling design presented larger variance than the remaining methods and that those variances seem to have a slightly tendency of increasing with larger impacts of clustering, at least for $\text{var}(\hat{\sigma}_u^2)$.

Table 7.3 includes results that were produced when considering $m^{sim} = 47$ and reducing the number of selected SSUs per cluster to $n_j^{sim} = 10$, and different choices for σ_η^2. From this table onwards we shall neither report nor make any further detailed comments on simulation results for variance estimation methods that do not account for the sampling design as results follow a similar pattern of those presented in Tables 7.1 and 7.2.

Estimator		rel bias			$cv(\text{var}(\hat{\underline{\theta}}))$		
		$\sigma_\eta^2 = 0.15$	$\sigma_\eta^2 = 0.45$	$\sigma_\eta^2 = 0.75$	$\sigma_\eta^2 = 0.15$	$\sigma_\eta^2 = 0.45$	$\sigma_\eta^2 = 0.75$
$\text{var}_r\left(\hat{\underline{\theta}}_{ML}\right)$	$\text{var}(\hat{\sigma}_u^2)$	-2.22% NS	-2.09% NS	-5.10%	26.25%	27.21%	28.13%
	$\text{var}(\hat{\sigma}_v^2)$	-0.56% NS	-0.51% NS	1.01% NS	22.37%	22.44%	22.45%
$\text{var}_c\left(\hat{\underline{\theta}}^2_{GLS}\right)$	$\text{var}(\hat{\sigma}_u^2)$	-4.00%	-3.64%	-5.98%	26.50%	27.35%	28.24%
	$\text{var}(\hat{\sigma}_v^2)$	-8.32%	-9.09%	-8.17%	22.76%	22.79%	22.81%

NS – denotes 'absolute bias not significantly different from zero at 95% level'.

Table 7.3 – Evaluation of $\text{var}(\hat{\underline{\theta}})$ considering $m^{sim} = 47$ and $n_j^{sim} = 10$.

Under the present scenario, considering simulation with 470 cases, variance estimation methods that account for the sampling scheme generally led to appreciable improvements in terms of relative bias when compared to methods that ignore the sampling scheme characteristics. In fact, we may also observe here that the variance estimator for $\hat{\underline{\theta}}_{ML}$ had an evidently better performance in terms of bias than those for $\hat{\underline{\theta}}^2_{GLS}$, especially when estimating $\text{VAR}(\sigma_v^2)$.

These variance estimation methods under the complex survey approach have again not shown a clear tendency of a relative bias inflation with increases in σ_η^2, although in terms of variance we may notice here that these methods appear to have a slightly trend of increasing their *cv*s with larger impacts of clustering.

Furthermore, these methods had a larger variance than the methods that do not take the sampling design into account, as under all the scenarios considered later in this chapter. For this reason we shall not make any further comments in this section on this specific aspect of the simulation results.

Table 7.4 includes results that were produced when considering $m^{sim}=47$ and $n_j^{sim}=5$, and again different choices for σ_η^2.

Estimator		rel bias			$cv(\text{var}(\hat{\underline{\theta}}))$		
		$\sigma_\eta^2=0.15$	$\sigma_\eta^2=0.45$	$\sigma_\eta^2=0.75$	$\sigma_\eta^2=0.15$	$\sigma_\eta^2=0.45$	$\sigma_\eta^2=0.75$
$\text{var}_c\left(\hat{\underline{\theta}}_{ML}\right)$	$\text{var}(\hat{\sigma}_u^2)$	-2.54%NS	-3.92%	-6.55%	30.56%	31.46%	32.25%
	$\text{var}(\hat{\sigma}_v^2)$	-2.93%	-2.82%	-1.19%NS	24.08%	23.96%	24.09%
$\text{var}_c\left(\hat{\underline{\theta}}^2_{GLS}\right)$	$\text{var}(\hat{\sigma}_u^2)$	-4.69%	-6.24%	-9.07%	30.77%	31.67%	32.40%
	$\text{var}(\hat{\sigma}_v^2)$	-17.42%	-16.77%	-16.32%	24.55%	24.47%	24.40%

NS – denotes 'absolute bias not significantly different from zero at 95% level'.

Table 7.4 – Evaluation of $\text{var}(\hat{\underline{\theta}})$ considering $m^{sim}=47$ and $n_j^{sim}=5$.

Note that we are still evaluating situations with a sampling design with a 'large' number of PSUs, but now with a limited number of SSUs selected per cluster and therefore a sample size reduced to 235. Under this scenario, only the variance estimator that accounts for the sampling scheme for $\hat{\underline{\theta}}_{ML}$ has led to noticeable improvements in terms of relative bias when compared to methods that ignore the sampling scheme characteristics. Variance estimators for $\hat{\underline{\theta}}^2_{GLS}$, specifically for $\text{var}(\hat{\sigma}_v^2)$, presented very poor bias results. Therefore, again we may note that the variance estimator for $\hat{\underline{\theta}}_{ML}$ had a better performance in terms of bias than those for $\hat{\underline{\theta}}^2_{GLS}$.

Once more the two considered variance estimation methods have not shown a very obvious trend of a relative bias increase for larger σ_η^2, and we found that these methods may have a slightly tendency of inflating their *cv*s with larger impacts of clustering.

Table 7.5 includes results that were produced when considering $m^{sim}=20$ and $n_j^{sim}=n_j^{sim*}$, and different choices for σ_η^2.

Estimator		rel bias			$cv(\text{var}(\hat{\underline{\theta}}))$		
		$\sigma_\eta^2 = 0.15$	$\sigma_\eta^2 = 0.45$	$\sigma_\eta^2 = 0.75$	$\sigma_\eta^2 = 0.15$	$\sigma_\eta^2 = 0.45$	$\sigma_\eta^2 = 0.75$
$\text{var}_c\left(\hat{\underline{\theta}}_{ML}\right)$	$\text{var}(\hat{\sigma}_u^2)$	-3.39%	-3.91%	-1.78%NS	37.71%	40.01%	42.83%
	$\text{var}(\hat{\sigma}_v^2)$	-0.49%NS	-1.68%NS	-0.40%NS	35.23%	35.62%	35.49%
$\text{var}_c\left(\hat{\underline{\theta}}^2_{GLS}\right)$	$\text{var}(\hat{\sigma}_u^2)$	-4.80%	-5.05%	-3.01%	37.97%	40.07%	42.90%
	$\text{var}(\hat{\sigma}_v^2)$	-8.06%	-10.28%	-8.75%	35.79%	36.19%	36.09%

NS – denotes 'absolute bias not significantly different from zero at 95% level'.

Table 7.5 – Evaluation of $\text{var}(\hat{\underline{\theta}})$ considering $m^{sim} = 20$ and $n_j^{sim} = n_j^{sim*}$.

Notice that the number of selected PSUs is now reduced to 20, but the number of SSUs selected per cluster is equal to the cluster size in the original BHPS subset. Under this situation, the variance estimators that take the sampling scheme into account for both $\hat{\underline{\theta}}_{ML}$ and $\hat{\underline{\theta}}^2_{GLS}$ have led to marked betterments in terms of relative bias when compared to the classic methods and ADF methods that ignore the sampling scheme characteristics.

Again the considered variance estimation methods have not shown a very evident trend in the relative bias with increases in σ_η^2, although differently from previous results there appears to be a trend in the opposite direction, i.e. of the bias to decrease with an inflation in the impact of clustering. Furthermore, we may observe more clearly now that these methods have a some tendency of having their *cv*s increasing with larger σ_η^2.

Table 7.6 includes results that were produced when considering $m^{sim} = 20$ and $n_j^{sim} = 15$, i.e. 300 cases, and again different choices for σ_η^2.

Estimator		rel bias			$cv(\text{var}(\hat{\underline{\theta}}))$		
		$\sigma_\eta^2 = 0.15$	$\sigma_\eta^2 = 0.45$	$\sigma_\eta^2 = 0.75$	$\sigma_\eta^2 = 0.15$	$\sigma_\eta^2 = 0.45$	$\sigma_\eta^2 = 0.75$
$\text{var}_c\left(\hat{\underline{\theta}}_{ML}\right)$	$\text{var}(\hat{\sigma}_u^2)$	-5.17%	-5.25%	-4.69%	38.07%	39.03%	40.75%
	$\text{var}(\hat{\sigma}_v^2)$	-1.54%NS	-0.69%NS	-0.49%NS	33.55%	33.79%	34.44%
$\text{var}_c\left(\hat{\underline{\theta}}^2_{GLS}\right)$	$\text{var}(\hat{\sigma}_u^2)$	-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%

NS – denotes 'absolute bias not significantly different from zero at 95% level'.

Table 7.6 – Evaluation of $\text{var}(\hat{\underline{\theta}})$ considering $m^{sim} = 20$ and $n_j^{sim} = 15$.

Variance estimators that take the sampling scheme into account for both $\underline{\hat{\theta}}_{ML}$ and $\underline{\hat{\theta}}^2_{GLS}$ have generally led to noticeable improvements in terms of relative bias when compared to methods that ignore the sampling scheme characteristics. Variance estimators for $\underline{\hat{\theta}}^2_{GLS}$, specifically for $\text{var}(\hat{\sigma}^2_v)$, presented inadequate relative bias results. In consequence, we may observe that the variance estimator for $\underline{\hat{\theta}}_{ML}$ had a better performance in terms of bias than those for $\underline{\hat{\theta}}^2_{GLS}$.

Again the two considered variance estimation methods have generally not shown a very obvious trend of a relative bias increase for larger σ^2_η, and we found that these methods may have a slight tendency of inflating their *cv*s with larger impacts of clustering. However, we may note that, for $\text{var}(\hat{\sigma}^2_v)$, there may be a trend of $\text{var}_c(\underline{\hat{\theta}}_{ML})$ to decline with rises in the clustering impacts.

Table 7.7 includes results that were produced when considering $m^{sim} = 20$ and $n_j^{sim} = 10$, and various choices for σ^2_η.

Estimator		rel bias			$cv(\text{var}(\hat{\theta}))$		
		$\sigma^2_\eta = 0.15$	$\sigma^2_\eta = 0.45$	$\sigma^2_\eta = 0.75$	$\sigma^2_\eta = 0.15$	$\sigma^2_\eta = 0.45$	$\sigma^2_\eta = 0.75$
$\text{var}_c\left(\underline{\hat{\theta}}_{ML}\right)$	$\text{var}(\hat{\sigma}^2_u)$	-7.90%	-8.37%	-4.70%	40.65%	41.57%	42.69%
	$\text{var}(\hat{\sigma}^2_v)$	-0.79%NS	-3.30%	-0.87%NS	34.49%	34.91%	34.69%
$\text{var}_c\left(\underline{\hat{\theta}}^2_{GLS}\right)$	$\text{var}(\hat{\sigma}^2_u)$	-11.12%	-10.80%	-7.68%	41.07%	42.00%	43.04%
	$\text{var}(\hat{\sigma}^2_v)$	-17.08%	-18.17%	-17.58%	35.22%	35.55%	35.29%

NS – denotes 'absolute bias not significantly different from zero at 95% level'.

Table 7.7 – Evaluation of $\text{var}(\underline{\hat{\theta}})$ considering $m^{sim} = 20$ and $n_j^{sim} = 10$.

Note that the number of SSUs selected per cluster is now reduced to 10 and consequently the sample size decreases to 200. Under this situation, only the variance estimator that accounts for the sampling scheme for $\underline{\hat{\theta}}_{ML}$ has generally led to noticeable improvements in terms of bias when compared to the classic variance estimation methods. Variance estimators for $\underline{\hat{\theta}}^2_{GLS}$ presented very poor bias results.

Once more the two considered variance estimation methods have not shown a very obvious trend of a relative bias increase for larger σ_η^2. If there is a tendency, it is in the opposite direction. Again, it appears that these methods may have a marginal trend of increasing their *cv*s for larger impacts of clustering.

Table 7.8 includes results that were produced when considering $m^{sim} = 20$ and $n_j^{sim} = 5$, and different choices for σ_η^2.

Estimator		rel bias			$cv(\text{var}(\hat{\underline{\theta}}))$		
		$\sigma_\eta^2 = 0.15$	$\sigma_\eta^2 = 0.45$	$\sigma_\eta^2 = 0.75$	$\sigma_\eta^2 = 0.15$	$\sigma_\eta^2 = 0.45$	$\sigma_\eta^2 = 0.75$
$\text{var}_c\left(\hat{\underline{\theta}}_{ML}\right)$	$\text{var}(\hat{\sigma}_u^2)$	-8.75%	-8.05%	-9.14%	48.44%	48.26%	50.12%
	$\text{var}(\hat{\sigma}_v^2)$	-2.45%NS	-4.51%	-4.41%	37.13%	36.45%	36.85%
$\text{var}_c\left(\hat{\underline{\theta}}^2_{GLS}\right)$	$\text{var}(\hat{\sigma}_u^2)$	-13.69%	-12.01%	-13.70%	48.82%	48.79%	50.75%
	$\text{var}(\hat{\sigma}_v^2)$	-30.50%	-31.03%	-30.08%	38.28%	38.01%	38.30%

NS – denotes 'absolute bias not significantly different from zero at 95% level'.

Table 7.8 – Evaluation of $\text{var}(\hat{\underline{\theta}})$ considering $m^{sim} = 20$ and $n_j^{sim} = 5$.

The number of SSUs selected per cluster is now further reduced to 5 and consequently the sample size decreases to 100 cases, which is the smallest *n* considered so far in this simulation study. Under the current scenario, again only the variance estimator that accounts for the sampling scheme for $\hat{\underline{\theta}}_{ML}$ has generally led to some improvements in terms of bias when compared to the classic variance estimation methods. The achieved gain in bias reduction is however reduced here if compared to previous considered situations. Note that variance estimators for both $\hat{\underline{\theta}}^2_{GLS}$ presented inferior bias results.

The evaluated variance estimation methods have not shown now any evidence of a tendency of a relative bias increase for larger σ_η^2. Distinctively from previous results we may not observe any obvious trend of an increase in the *cv*s for larger impacts of clustering.

Table 7.9 includes results that were produced when considering $m^{sim} = 15$ and $n_j^{sim} = n_j^{sim*}$, and distinct choices for σ_η^2.

Estimator		rel bias			$cv(\text{var}(\hat{\underline{\theta}}))$		
		$\sigma_\eta^2 = 0.15$	$\sigma_\eta^2 = 0.45$	$\sigma_\eta^2 = 0.75$	$\sigma_\eta^2 = 0.15$	$\sigma_\eta^2 = 0.45$	$\sigma_\eta^2 = 0.75$
$\text{var}_c\left(\hat{\underline{\theta}}_{ML}\right)$	$\text{var}(\hat{\sigma}_u^2)$	-5.07%	-4.03%	-7.84%	44.25%	46.94%	49.91%
	$\text{var}(\hat{\sigma}_v^2)$	-1.95%NS	-1.75%NS	-2.54%NS	41.64%	42.33%	41.95%
$\text{var}_c\left(\hat{\underline{\theta}}^2_{GLS}\right)$	$\text{var}(\hat{\sigma}_u^2)$	-6.60%	-5.67%	-9.23%	44.34%	47.03%	49.86%
	$\text{var}(\hat{\sigma}_v^2)$	-12.29%	-11.50%	-12.43%	42.60%	43.04%	42.40%

NS – denotes 'absolute bias not significantly different from zero at 95% level'.

Table 7.9 – Evaluation of $\text{var}(\hat{\underline{\theta}})$ considering $m^{sim} = 15$ and $n_j^{sim} = n_j^{sim*}$.

Notice that the number of selected PSUs is now reduced to 15, but the number of SSUs selected per cluster is equal to the cluster size as for results presented earlier in Table 7.5. Under this situation, the variance estimators that take the sampling scheme into account for both $\hat{\underline{\theta}}_{ML}$ and $\hat{\underline{\theta}}^2_{GLS}$ have led to some marked gains in terms of bias when compared to the classic methods and ADF methods that ignore the sampling scheme characteristics. Variance estimators for $\hat{\underline{\theta}}^2_{GLS}$ again presented poor bias results.

Again the considered variance estimation methods have not shown a very evident trend in the relative bias with increases in σ_η^2. Moreover, we may notice that these methods have some tendency of having their variances to be inflated with larger σ_η^2.

Table 7.10 includes results that were produced when considering $m^{sim} = 15$ and $n_j^{sim} = 15$, and different choices for σ_η^2.

Estimator		rel bias			$cv(\text{var}(\hat{\underline{\theta}}))$		
		$\sigma_\eta^2 = 0.15$	$\sigma_\eta^2 = 0.45$	$\sigma_\eta^2 = 0.75$	$\sigma_\eta^2 = 0.15$	$\sigma_\eta^2 = 0.45$	$\sigma_\eta^2 = 0.75$
$\text{var}_c\left(\hat{\underline{\theta}}_{ML}\right)$	$\text{var}(\hat{\sigma}_u^2)$	-4.90%	-3.68%	-3.88%	44.79%	45.31%	48.16%
	$\text{var}(\hat{\sigma}_v^2)$	-2.07%NS	-2.96%	-2.47%NS	39.99%	39.53%	39.00%
$\text{var}_c\left(\hat{\underline{\theta}}^2_{GLS}\right)$	$\text{var}(\hat{\sigma}_u^2)$	-7.89%	-5.04%	-6.48%	45.12%	45.80%	48.38%
	$\text{var}(\hat{\sigma}_v^2)$	-17.09%	-18.63%	-18.28%	41.12%	40.23%	40.08%

NS – denotes 'absolute bias not significantly different from zero at 95% level'.

Table 7.10 – Evaluation of $\text{var}(\hat{\underline{\theta}})$ considering $m^{sim} = 15$ and $n_j^{sim} = 15$.

Under this scenario, variance estimators that account for the sampling design for $\hat{\underline{\theta}}_{ML}$ and $\hat{\underline{\theta}}^2_{GLS}$ have led to some noticeable improvements in terms of bias when compared to methods that ignore features of the sampling scheme. Variance estimators for $\hat{\underline{\theta}}^2_{GLS}$, when considering $\text{var}(\hat{\sigma}_v^2)$, presented very poor bias results. Therefore, again we may note that the variance estimator for $\hat{\underline{\theta}}_{ML}$ led to better results than those for $\hat{\underline{\theta}}^2_{GLS}$.

Once more the evaluated variance estimation methods have not shown a very obvious trend of a relative bias increase for larger σ_η^2, and we found that these methods may have some tendency of inflating their *cv*s with larger impacts of clustering.

Table 7.11 includes results that were produced when considering $m^{sim} = 15$ and $n_j^{sim} = 10$, and different choices for σ_η^2.

Estimator		rel bias			$cv(\text{var}(\hat{\underline{\theta}}))$		
		$\sigma_\eta^2 = 0.15$	$\sigma_\eta^2 = 0.45$	$\sigma_\eta^2 = 0.75$	$\sigma_\eta^2 = 0.15$	$\sigma_\eta^2 = 0.45$	$\sigma_\eta^2 = 0.75$
$\text{var}_c\left(\hat{\underline{\theta}}_{ML}\right)$	$\text{var}(\hat{\sigma}_u^2)$	-5.48%	-6.11%	-4.87%	47.86%	47.80%	50.19%
	$\text{var}(\hat{\sigma}_v^2)$	-3.41%	-2.68%NS	-1.38%NS	41.05%	40.43%	40.87%
$\text{var}_c\left(\hat{\underline{\theta}}^2_{GLS}\right)$	$\text{var}(\hat{\sigma}_u^2)$	-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%

NS – denotes 'absolute bias not significantly different from zero at 95% level'.

Table 7.11 – Evaluation of $\text{var}(\hat{\underline{\theta}})$ considering $m^{sim} = 15$ and $n_j^{sim} = 10$.

Under the current considered situation, the number of SSUs selected per cluster is now reduced to 10 and therefore the sample size is diminished to 150. We may observe that variance estimators for $\hat{\underline{\theta}}_{ML}$ and $\hat{\underline{\theta}}^2_{GLS}$ have led to some gains in bias when compared to classic methods. Variance estimators for $\hat{\underline{\theta}}^2_{GLS}$, when considering $\text{var}(\hat{\sigma}_v^2)$, produced nevertheless very large relative bias results.

We cannot observe a very clear tendency of an increase in bias for larger impacts of clustering under this scenario. If there is a trend, it is in the other direction. Furthermore, we may not notice here any evident trend of an increase in the *cv*s for larger impacts of clustering.

Table 7.12 includes results that were produced when considering $m^{sim}=15$ and $n_j^{sim}=5$, and once more different choices for σ_η^2.

Estimator		rel bias			$cv(\text{var}(\hat{\underline{\theta}}))$		
		$\sigma_\eta^2=0.15$	$\sigma_\eta^2=0.45$	$\sigma_\eta^2=0.75$	$\sigma_\eta^2=0.15$	$\sigma_\eta^2=0.45$	$\sigma_\eta^2=0.75$
$\text{var}_c(\hat{\underline{\theta}}_{ML})$	$\text{var}(\hat{\sigma}_u^2)$	-12.02%	-12.78%	-14.03%	55.57%	57.32%	58.49%
	$\text{var}(\hat{\sigma}_v^2)$	-4.64%	-4.94%	-5.26%	43.06%	42.49%	43.26%
$\text{var}_c(\hat{\underline{\theta}}^2_{GLS})$	$\text{var}(\hat{\sigma}_u^2)$	-16.84%	-17.45%	-18.82%	56.55%	58.42%	59.75%
	$\text{var}(\hat{\sigma}_v^2)$	-35.69%	-37.07%	-37.11%	44.73%	44.35%	45.37%

Table 7.12 – Evaluation of $\text{var}(\hat{\underline{\theta}})$ considering $m^{sim}=15$ and $n_j^{sim}=5$.

The number of SSUs selected per cluster is now further contracted to 5 and therefore the sample size is reduced to 75 cases, which shall be the smallest n considered in this second simulation study. Under this scenario, only the variance estimator that accounts for the sampling scheme for $\hat{\underline{\theta}}_{ML}$ has generally led to some improvements in terms of bias when compared to the classic variance estimation methods. This gain in bias terms is, as for results of Table 7.8, albeit reduced when compared to the majority of the previous considered situations. Variance estimators for $\hat{\underline{\theta}}^2_{GLS}$ presented considerably poor relative bias results.

The evaluated variance estimation methods have shown now some evidence of a tendency of a relative bias increase for larger σ_η^2, and also some trend of increases in variance for larger impacts of clustering.

7.4 Further discussion

The present section illustrates some of the results presented in the previous section. It is our aim here to highlight the general behaviour of the proposed variance estimation methods that account for the complex sampling scheme. We shall again consider different (*i*) sampling design characteristics, (*ii*) sample sizes, and (*iii*) clustering impacts.

In Figure 7.1, we shall initially consider simulation results for $\text{var}_c(\hat{\underline{\theta}}_{ML})$ and $\text{var}_c(\hat{\underline{\theta}}^2_{GLS})$, which shall be denoted by "var for ML" and "var for GLS2", respectively. More specifically we shall only consider relative bias results for $\text{var}(\hat{\sigma}_u^2)$. Furthermore, note that we shall adopt "+" to denote situations where the sample size is variable, i.e. in scenarios where the number SSUs selected per cluster is equal to the cluster size in the original BHPS subset.

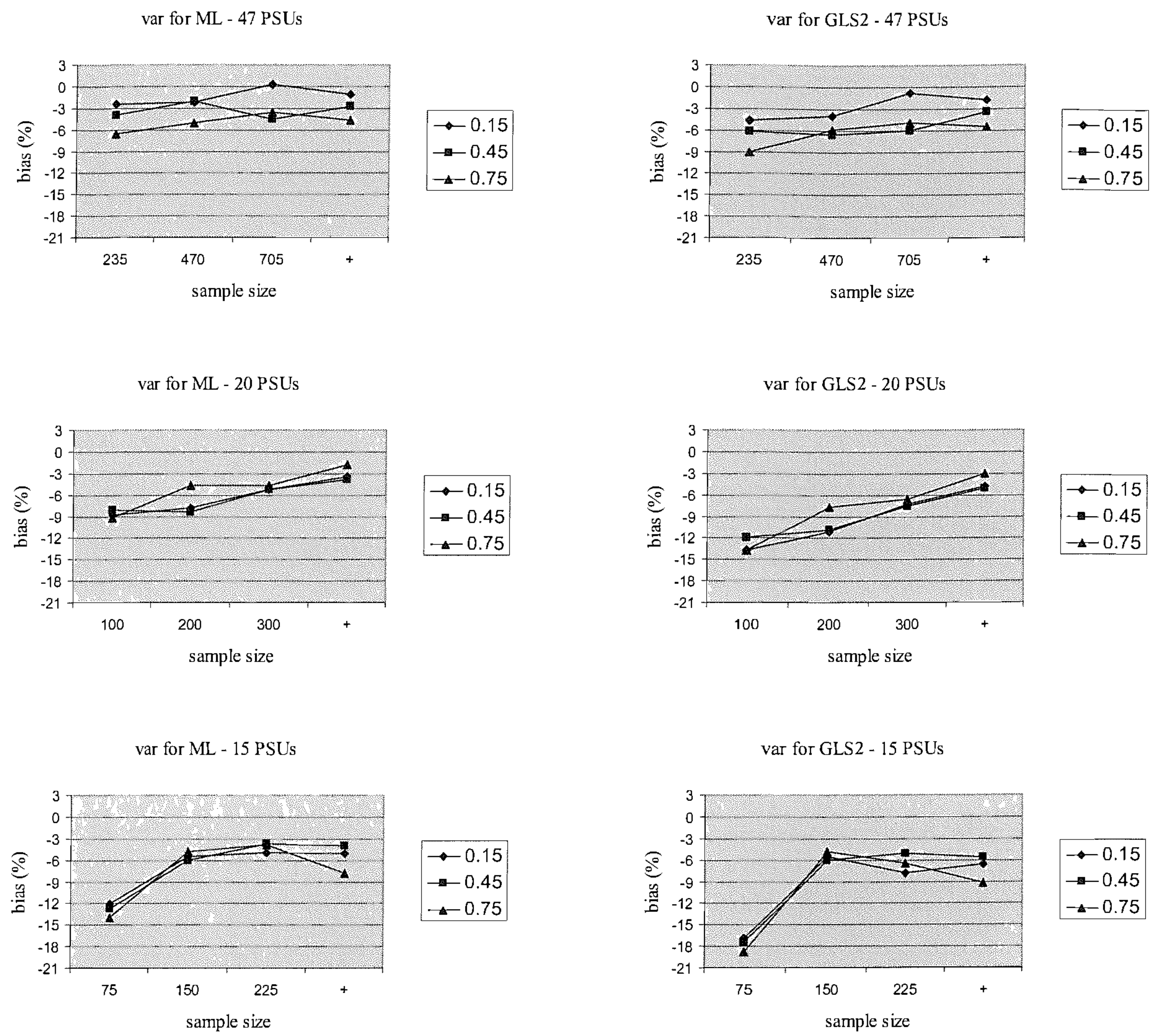

Figure 7.1 – Relative bias for $\text{var}_c(\hat{\underline{\theta}}_{ML})$ and $\text{var}_c(\hat{\underline{\theta}}^2_{GLS})$.

By looking at Figure 7.1 above we may draw various observations. An immediate conclusion is that, interestingly, both methods above evaluated follow, for every pair of graphs, a very similar pattern but with distinct bias levels, i.e. with $\text{var}_c(\hat{\underline{\theta}}_{ML})$ presenting lower bias in all situations. Furthermore, $\text{var}_c(\hat{\underline{\theta}}^2_{GLS})$ appear to be more sensitive to sample size reductions than $\text{var}_c(\hat{\underline{\theta}}_{ML})$. As stated in the previous section, under most scenarios with various sampling designs and samples sizes, it is not very clear whether there is a tendency of an increase in bias for larger impacts of clustering, although this may be true for some situations. The method $\text{var}_c(\hat{\underline{\theta}}_{ML})$ produces reasonably 'good' bias results in all situations with sample sizes of 100 or more, while $\text{var}_c(\hat{\underline{\theta}}^2_{GLS})$ generally required n to be at least 150.

In Figure 7.2, we shall present simulation results for approximately 95% simulation confidence intervals for the absolute bias, given by (7.5), for the variance estimators $\text{var}_c(\hat{\underline{\theta}}_{ML})$ and $\text{var}_c(\hat{\underline{\theta}}^2_{GLS})$. We shall only contemplate results for $\text{var}(\hat{\sigma}^2_u)$ and $\sigma^2_\eta = 0.15$.

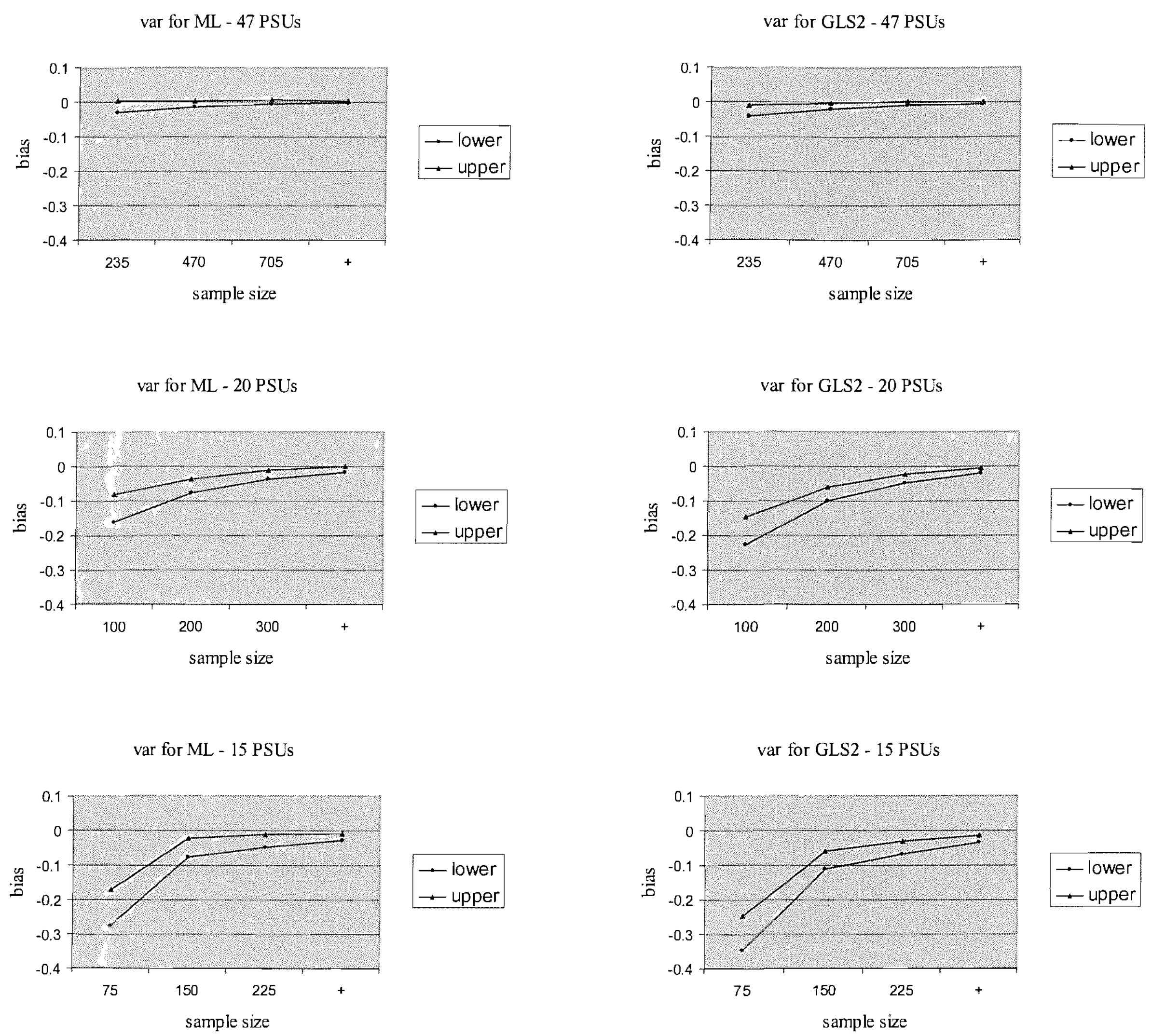

Figure 7.2 – 95% simulation confidence intervals for bias for $\text{var}_c(\hat{\underline{\theta}}_{ML})$ and $\text{var}_c(\hat{\underline{\theta}}^2_{GLS})$.

The graphs included in the figure presented above agree with those in Figure 7.1, as these indicate that $\text{var}_c(\hat{\underline{\theta}}_{ML})$ and $\text{var}_c(\hat{\underline{\theta}}^2_{GLS})$ appear to have very similar absolute bias patterns for different sampling designs and sample sizes, although $\text{var}_c(\hat{\underline{\theta}}_{ML})$ presents lower bias in all situations. Moreover, we may observe similar levels of bias for both estimators in scenarios with samples of size of approximately 200 or more.

In Figure 7.3, we shall consider coefficient of variation simulation results for $\text{var}_c(\hat{\underline{\theta}}_{ML})$ and $\text{var}_c(\hat{\underline{\theta}}^2_{GLS})$. Again we shall only consider results for $\text{var}(\hat{\sigma}^2_u)$.

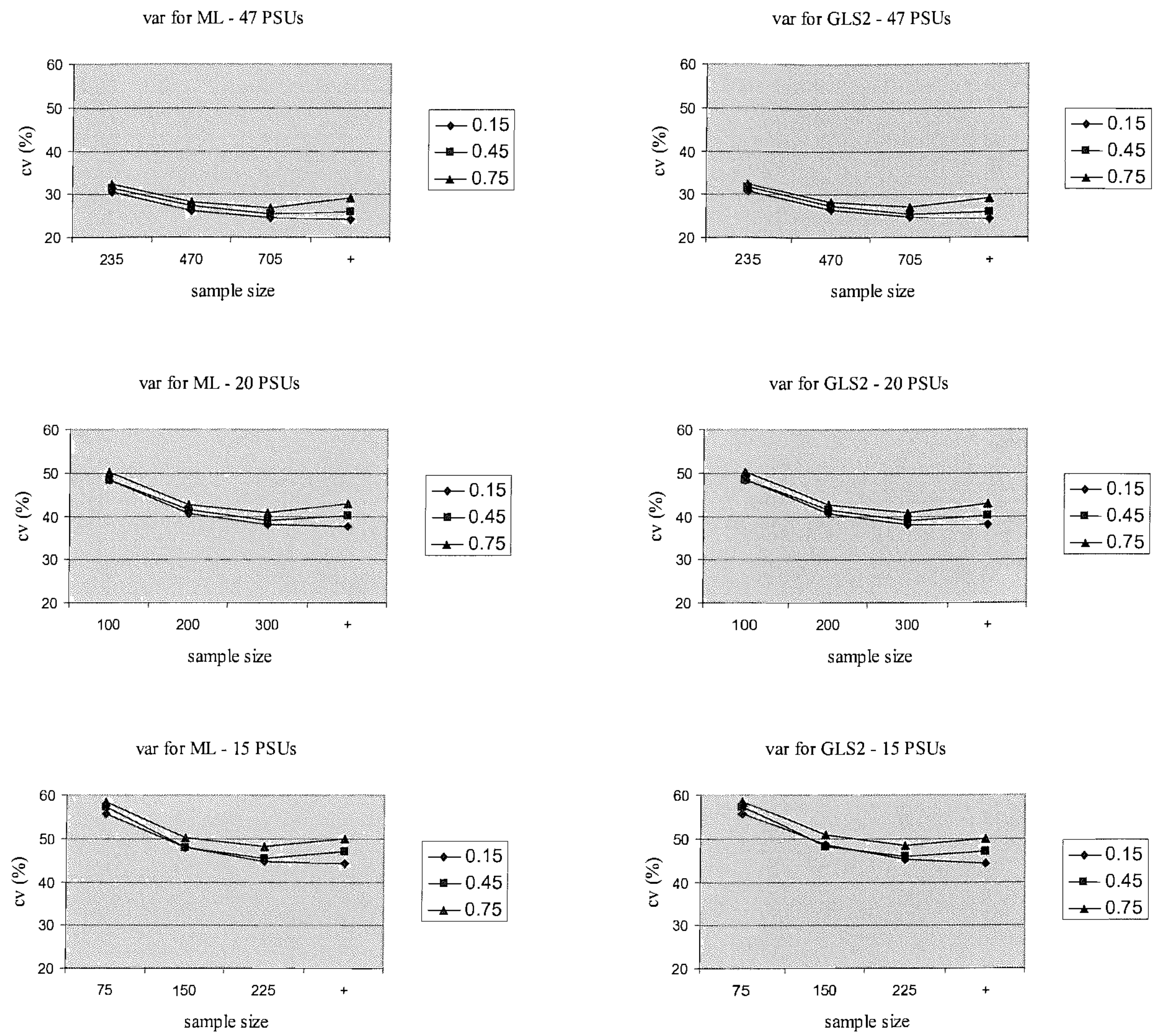

Figure 7.3 – Coefficient of variation for $\text{var}_c(\hat{\underline{\theta}}_{ML})$ and $\text{var}_c(\hat{\underline{\theta}}^2_{GLS})$.

By examining Figure 7.3 we may observe that both evaluated methods have virtually the same coefficient of variation in most of the situations considered. If there is any difference between them, it occurs in situations with sample sizes of around 200 or less, with $\text{var}_c(\hat{\underline{\theta}}_{ML})$ having slightly smaller *cv*s. In general, as noticed in the previous section, it appears that in most situations there is some tendency of increases in variance for larger values of σ^2_η.

7.5 Concluding remarks

This chapter included the features and the essential results of a second simulation study which had primarily the objective of assessing variance estimation methods we propose in Chapter 6. In Section 7.3 the methods were compared mainly in terms of their bias and variance, for different designs and considering several different sample sizes. Note that the proposed methods were also contrasted with classic variance estimation methods and with distribution free methods that do not account for the sampling scheme. Different choices for σ_η^2 were considered in order to allow us to assess the effects of different impacts of clustering on the considered variance estimation procedures. Some illustrations were provided in Section 7.4, where further considerations were made.

From all the observations already presented in this chapter, we may draw some conclusions. In summary, simulation results included here suggest that methods that do not take the sampling scheme into account underestimate the variance, in some situations seriously. Additionally, the underestimation has a tendency to increase rapidly for larger impacts of clustering. Nevertheless, the naïve variance estimation method for $\hat{\underline{\theta}}_{ML}$ appears to be not very sensitive to small impacts of clustering, at least in situations with sample sizes larger than around 200 cases. The ADF variance estimators for the GLS type point estimators, that do not consider the sampling design, seem to present some advantages over naïve variance estimators in terms of bias, at least for smaller values of σ_η^2.

In essence, both methods that allow for clustering and take the sampling design into account tend to lead to noticeable improvements in terms of relative bias when compared to methods that ignore the sampling scheme characteristics, at least for situations where the sample size is over around 200 cases. These methods presented larger values of coefficient of variation when compared to classic variance estimation methods. Furthermore, for some scenarios the variances appear to have somewhat a trend of increasing with larger impacts of clustering.

Certainly the most interesting result reported in the current chapter is the difference in performance when comparing results for $\text{var}_c(\hat{\underline{\theta}}^2_{GLS})$, extended in Chapter 6, with those for $\text{var}_c(\hat{\underline{\theta}}_{ML})$, proposed in the previous chapter following Binder (1983)'s approach. The variance estimator proposed under the complex survey approach for $\hat{\underline{\theta}}^2_{GLS}$ presented poor bias results in situations in samples sizes less than around 200 cases. The variance estimator for

$\hat{\underline{\theta}}_{ML}$ had an evidently better performance in terms of bias than those for $\hat{\underline{\theta}}^2_{GLS}$ in all considered situations, and led to reasonably 'good' results in all situations with sample sizes larger than 100.

In terms of variance, both estimators that take the sampling design into account performed very similarly, with virtually the same coefficient of variation in most of the simulation scenarios considered.

Some simulation results have also been produced for $\text{var}_n(\hat{\underline{\theta}}^3_{GLS})$, $\text{var}_{adf}(\hat{\underline{\theta}}^3_{GLS})$, and $\text{var}_c(\hat{\underline{\theta}}^3_{GLS})$, although those shall not be reported in this thesis. Those results were in general similar to the ones calculated and presented in the current chapter for $\text{var}_n(\hat{\underline{\theta}}^2_{GLS})$, $\text{var}_{adf}(\hat{\underline{\theta}}^2_{GLS})$, and $\text{var}_c(\hat{\underline{\theta}}^2_{GLS})$, respectively. Moreover, by comparing the results for the three considered variance estimation methods for $\hat{\underline{\theta}}^3_{GLS}$ we could draw generally very similar conclusions to those considerations presented in this chapter when performing comparisons among variance estimation methods for $\hat{\underline{\theta}}^2_{GLS}$. In addition, $\text{var}_c(\hat{\underline{\theta}}^3_{GLS})$ have also tended to present poor bias results in situations with 'small' samples sizes, similarly to $\text{var}_c(\hat{\underline{\theta}}^2_{GLS})$. The estimator $\hat{\underline{\theta}}^3_{GLS}$ is given by (4.48b), in Chapter 4, Section 4.4, Sub-section 4.4.2.

We acknoledge here that we have assessed in this simulation study variance estimators of estimators of variance components in a context of random effects models. We have not checked whether or not results would have been different when evaluating variance estimators of factor loadings and regression coefficients.

Note that an alternative variance estimation method for GLS type point estimator in situations where $\text{var}_c(\hat{\underline{\theta}}^2_{GLS})$ and $\text{var}_c(\hat{\underline{\theta}}^3_{GLS})$ did not perform very well could be a replication method, as for example the Jackknife method that was described earlier in Chapter 2, Section 2.5, Sub-section 2.5.2. An evaluation of this approach could be included in a potential extention of the current simulation study.

Moreover, an interesting exercise could be to compare the results presented in the current chapter, and those included earlier in Chapter 5, with results produced by adopting estimation procedures that are currently implemented in comercial softwares, for example LISREL (Jöreskog and Sörbom, 1997) and Mplus (Muthén and Muthén, 2005). This has not been done.

Chapter 8

Conclusions

We shall attempt to summarise in this final chapter the principal results and conclusions included in this doctoral dissertation. We shall emphasise here the main achievements of the research project. Potential further research topics shall also be addressed.

This thesis consists of investigating methods for the analysis of longitudinal complex survey data, with the analytic use of sample surveys as the main focus of interest. Specialised statistical techniques are required when analyzing longitudinal data, as the inter-correlation among observations on one subject should be accounted for. Beyond that, complex sample schemes may additionally cause correlation structure among observations.

We may subdivide this dissertation in three major areas of study: (*a*) impacts of complex sampling schemes on the analysis and modelling of longitudinal survey data (Chapters 2 and 3); (*b*) covariance structure models point estimation (Chapters 4 and 5); and (*c*) variance estimation and covariance structure model fitting assessment (Chapters 6 and 7).

Statistical practitioners and longitudinal survey data users generally utilise *standard* statistical tools, which in most situations do not consider the characteristics of the complex sample schemes. In particular, part (*a*) of the thesis has answered the following general question of interest, to our knowledge for the first time: what are the possible impacts of complex sampling designs on longitudinal analysis?

Therefore, this thesis initially reviews estimation procedures for longitudinal regression model parameters considering both the classical (Liang and Zeger, 1986; Zeger and Liang, 1986; Zeger, Liang and Albert, 1988; Jones, 1993; and Diggle *et al.*, 2002) and the complex survey data contexts (Kish and Frankel, 1974; Fuller, 1975; and Skinner, 1989a). The pseudo maximum likelihood method (PML) for estimation of longitudinal model parameters allowing for complex designs is discussed. Furthermore several methods for variance estimation are examined, including linearization (Binder, 1983; Rust, 1985; Skinner, 1989a; and Shah *et al.*, 1995) and jackknife (Krewski and Rao, 1981; Rust, 1985; and Shah, Barnwell, and Bieler, 1997) approaches. Issues regarding design effects (Kish, 1965) and misspecification effects (*meffs*; Skinner, 1989b) are also addressed considering the longitudinal context.

An empirical investigation using longitudinal survey data from the British Household Panel Survey (BHPS) is executed applying methods mentioned above. The main characteristics of the BHPS data set (Berrington, 2002) and of the BHPS sampling design (Taylor *et al.*, 2001) are detailed. Model fitting results are produced considering two different setups, allowing for: (*i*) longitudinal structure and survey weights, original clustering, and stratification; and (*ii*) longitudinal structure and a new aggregated clustering, treating the weights as constant and ignoring stratification.

Variance effects of clustering for longitudinal studies are then identified. The empirical results indicate that variance effects of clustering could be stronger for longitudinal studies than for cross-sectional studies, at least for certain kinds of longitudinal analysis. The main argument for that result is that the random cluster effects could be more correlated over the time than the random individual effects, suggesting that the cluster units could possibly be manifesting homogeneity over time. Our theoretical argumentation is also supported by the fact that attitude scores of individual women are affected both by measurement error and genuine changes in attitudes.

The results of our empirical study provide one specific achievement of this dissertation, which could be valuable for longitudinal data analysts, such as demographers, sociologists and epidemiologists. In general the evidence presented in Chapter 3 indicates that we should recognise the importance of taking the survey *sampling design* into account when analysing longitudinal survey data.

Nevertheless, we should stress that the patterns discovered for *meffs* in these kinds of analyses may well not be extended to other kinds of longitudinal analyses. To conjecture about the class of models and estimators for which the patterns observed in part (*a*) of this thesis might apply, we suggest that inflated misspecification effects for longitudinal analyses will occur when the longitudinal design allows temporal ‘random’ variation in individual responses to be extracted from between-person differences and thus to decrease the component of standard errors due to these differences, but provides less ‘explanation’ of between cluster differences, so that the relative influence of this component of standard errors increases.

Further research could also be performed for better understanding of this phenomenon. As a possible topic for further research, we could evaluate the variance effects of clustering considering the use of data from other longitudinal studies. One candidate survey could be the Brazilian Monthly Labour Force Survey (BMLFS). This is a large scale survey carried out by the Brazilian National Bureau of Statistics (IBGE). The BMLFS has as main objective to estimate level and change in employment, unemployment and other labour force characteristics. It covers 6 Brazilian metropolitan areas (approximately 25% of Brazilian population). Note that this is not actually a panel survey. In fact, that survey adopts a 4-8-4-rotation scheme with a sample overlap of 75% each moth. Brazilian users of that data have been pointing out some current problems, as for example the inexistence of any treatment for non-response, and the lack of use of the survey repetition for improving estimates, i.e. each month has been treated as a cross-section survey.

Following the findings presented in part (*a*), it has also been a purpose of this thesis to propose statistical methods for the analysis of longitudinal data collected under complex sampling designs. Part (*b*) of this dissertation discusses mainly point estimation procedures for covariance structure models parameters. A weighted estimation procedure (S_w) is proposed for estimating the population covariance matrix Σ, considering covariates. Additionally, a review of established variance estimation methods for $\hat{\Sigma}$ is carried out, while further developments on this issue when considering the complex survey approach are also accomplished mainly by adopting Taylor expansion techniques in order to modify or to extend asymptotically distribution-free (ADF) methods.

Note that the use of covariance structure models is emphasised in the research project because that approach includes a wide range of modelling techniques, which have a potential use, especially in the social sciences, for the structural analysis of covariance matrices.

Classical estimation methods such as unweighted least squares, generalised least squares, and maximum likelihood (ML) are discussed. These methods are then modified and unweighted least squares and generalised least squares under the complex survey approach methods are proposed by considering the weighted sample covariance matrix S_w. In addition, a pseudo maximum likelihood for covariance structure models is also derived via maximisation of the pseudo log likelihood function.

A possible area for further research, when considering part (*b*) of the thesis, is the extension of the restricted (or modified) maximum likelihood (REML) procedure. Patterson and Thompson (1971) introduced this estimation method originally for estimation of variance components in generalised linear models. Note that the loss in degrees of freedom, which is result of the estimation of the fixed effects, is taken into account by the REML (Harville, 1977; and Smyth and Verbyla, 1996). This issue is not considered under the traditional ML approach, which makes the estimators for the variance components, according to Robinson (1987) biased downwards, mainly when the number of fixed parameters is 'large' relative to the sampling size. The REML generally produces less biased estimates of the variance components than the ML method (Jones, 1993; Hocking, 1985; and Diggle *at al.*, 2002). Following the pseudo maximum likelihood method discussed in this thesis a refinement for the REML approach could be proposed for situations where complex survey data is considered. Note that Robinson (1987) provides a brief discussion about the use of variance components estimation techniques, including REML, to the estimation of standard errors in complex surveys.

An initial simulation study is carried out in this research project with the main objective of examining the statistical properties of the point estimation procedures proposed in this thesis. Moreover, the properties of the proposed methods are compared with those of the classic methods also discussed in this dissertation. Situations with clustered and non-clustered data are considered, alongside two different distributions (normal and t) and various alternative sample sizes.

The principal conclusions we may extract from the simulation results are: (*i*) overall most of the proposed methods have satisfactory performances in terms of bias and variance when compared to the classical methods; (*ii*) ADF methods do not always perform as we expected when dealing with departures from normality assumptions, although frequently these were the most efficient methods and generally less sensitive to clustering; (*iii*) ADF methods should be adopted carefully in situations where only samples of small size are available (agreeing with Bollen, 1989; Satorra, 1992; Yuan and Bentler, 1997b, and Olsson, Foss, and Troye, 2003, for example); (*iv*) ADF methods are the ones with the best performance in a situation with stronger departures from the normality conditions; and (*v*) maximum likelihood and pseudo maximum likelihood estimators have in general produced satisfactory performance in terms of bias and variance, even in situations where the normality assumption was violated (agreeing with Satorra and Bentler, 1986).

Part (*c*) of this thesis considers methods for variance estimation for generalised least squares and (pseudo) maximum likelihood point estimators, initially in the classical independent and identically distribution approach, both under normality assumptions and under departures from normality conditions.

Moreover, under the complex survey data approach, we extend in this research project ADF variance estimation methodology developed by Skinner (1989a), followed by Satorra

(1992), Muthén and Satorra (1995), and Skinner and Holmes (2003). In addition, we propose a method for estimating the asymptotic covariance matrix of the PML point estimator in the context of covariance structure models, by following the approach of Binder (1983).

Furthermore, testing techniques for structural models for covariance matrices are also reviewed in the classical context, including goodness of fit tests, test statistics for comparing nested models and goodness of fit indices.

We also consider methods that work under normality assumptions and methods that allow for departures from normality conditions, including distribution free methods, and scaled and adjusted test statistics.

Model testing is an important step in any model fit procedure. According to Menard (1991), in longitudinal models there is an increase to problems comparably to cross-sectional models in this regard. Eltinge (1999) acknowledged the need to improve and to develop new techniques for model assessment and diagnostic in the complex survey data context. Classic measures that are frequently used in model testing are appropriate for situations where data is obtained from a simple random sampling design.

Therefore, we propose some new developments on model fitting statistics when working with longitudinal data in a complex survey design framework. We initially modify the root mean-square residual measure proposed by Jöreskog and Sörbom (1989). Moreover, following Rao and Scott's conceptions (Rao and Scott, 1979; followed by, for example, Rao and Scott, 1981; Skinner, 1989a; and more recently, Rao and Thomas, 2003), we propose modifying the Wald goodness of fit test in the context of models for covariance structures. Note that our proposition is equivalent to modifying the scaled test statistics developed by Satorra and Bentler (1986) and Satorra and Bentler (1988). Furthermore, we also propose a modification for the Wald significance test for nested hypothesis, following an approach suggested by Skinner (1989a, Section 3.4). Goodness of fit indices proposed by Jöreskog and Sörbom (1989) are also modified in order be utilised in the complex survey data context.

A second simulation study is performed with the principal objective of evaluating the statistical properties of the variance estimation procedures discussed and proposed in this dissertation. Proposed methods are compared with classic variance estimation methods and with distribution free methods that do not account for the sampling scheme. Different impacts of clustering on the variance estimation procedures are considered, alongside different specifications for the sampling design and consequently several different sample sizes.

Simulation results suggest essentially that: (*i*) methods that do not take the sampling scheme into account underestimate the variance, in some situations very gravely; (*ii*) underestimation tend to increase rapidly with inflations in the impacts of clustering; (*iii*) ADF methods that allow for clustering and take the sampling design into account tend to lead to noticeable improvements in terms of relative bias when compared to methods that ignore the sampling scheme characteristics, in situations where the sample size is over around 200 cases; and (*iv*) the variance estimator we propose for estimating the variance of the maximum likelihood point estimator has an evidently better performance in terms of bias than those proposed for estimating the variance for the generalised least squares (GLS) estimators.

Another suggestion for some further investigation could be evaluating the behaviour of alternative variance estimation methods, such as the Jackknife replication method, for GLS type point estimators in situations where the proposed methods do not perform very well.

Moreover, further research could still involve evaluating the consequences of using standard model fit test techniques without consideration to the complexity of the sample, by extending for example the second simulation study performed in this dissertation.

Notice that a substantial portion of the research performed in both parts (*b*) and (*c*) of this dissertation involves further investigation and further evaluation of ADF methods when adapted to the complex survey data context. Another potential area for further research is thus to investigate the reasons why those methods did not perform relatively well in some situations considered in both simulation studies. Certainly this is a topic that goes beyond the purposes of the present thesis.

Appendix A

Gender role attitude statements

Table A.1 presents the set of statements concerning the family, women's roles, and work out of the household, included in waves one, three, five, seven and nine.

1	A pre-school child is likely to suffer if his or her mother works
2	All in all, family life suffers when the woman has a full-time job
3	A woman and her family would all be happier if she goes out to work
4	Both the husband and wife should contribute to the household income
5	Having a full-time job is the best way for a woman to be an independent person
6	A husband's job is to earn money; a wife's job is to look after the home and family
7	Children need a father to be as closely involved in their upbringing as the mother
8	Employers should make special arrangements to help mothers combine jobs and childcare
9	A single parent can bring up children as well as a couple

Source: Berrington (2002).

Table A.1 – Gender role attitude statements.

Note that the answers to these affirmations may possibly present some measurement error as some of the statements are very imprecise and abstract to have a standard interpretation (Berrington, 2002).

Appendix B

Evaluation of the distribution of *y* and individual profiles over time

We provide below a brief evaluation of the distribution of the dependent variable gender role attitude score. Figure B.1 presents histograms for that variables for waves one, three, five, seven and nine, respectively.

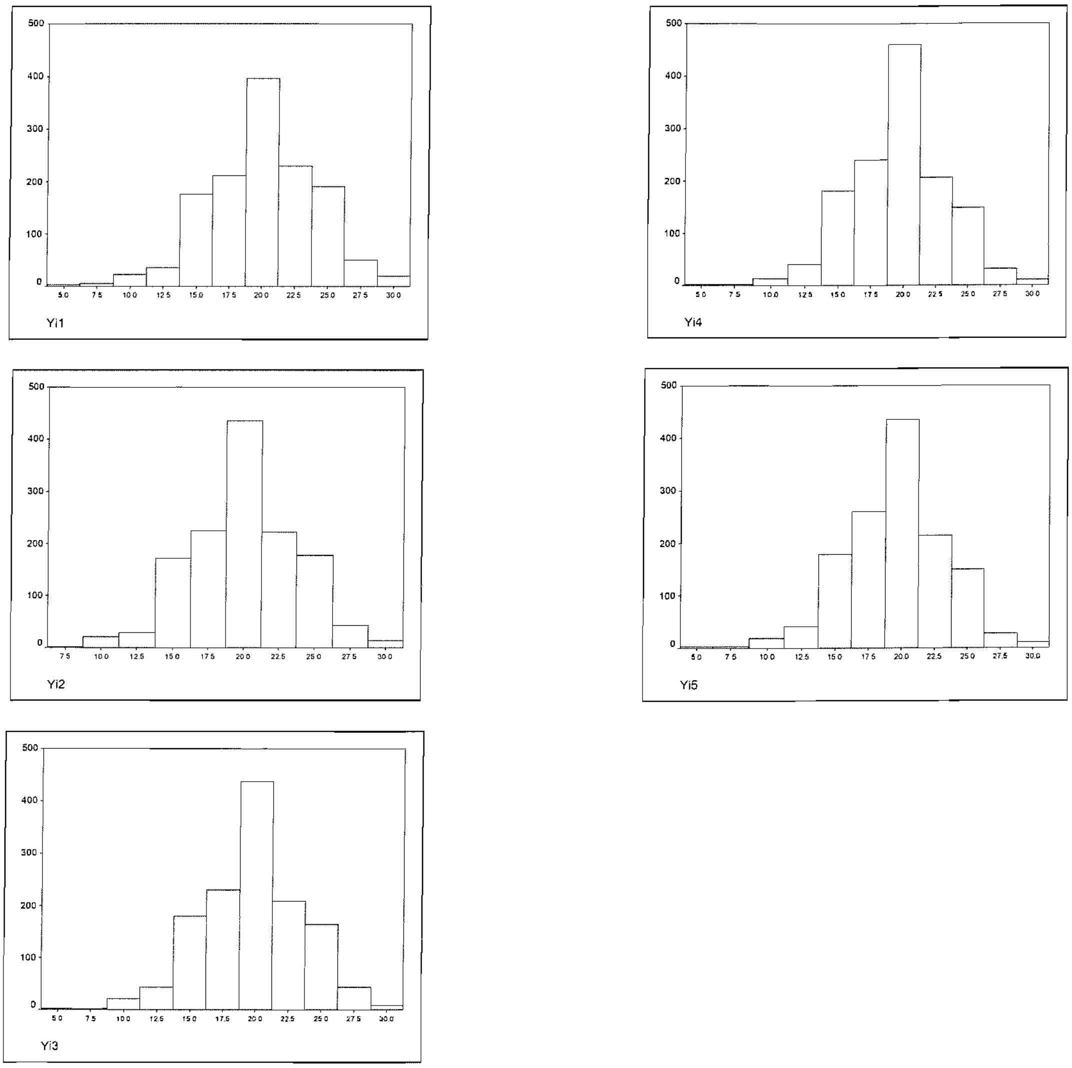

Figure B.1 – Histograms for *y* for waves 1, 3, 5, 7 and 9.

By observing the graphs presented above we do not have enough evidence that the univariate marginal distributions of y are not normal appearance. However it is possible for variables to have normal marginal distributions but not have multivariate normal distribution. We may also evaluate the multivariate distribution by presenting below in Figure B.2 scatter plots of pairs of observations of y on different time t (Johnson and Wichern, 1998).

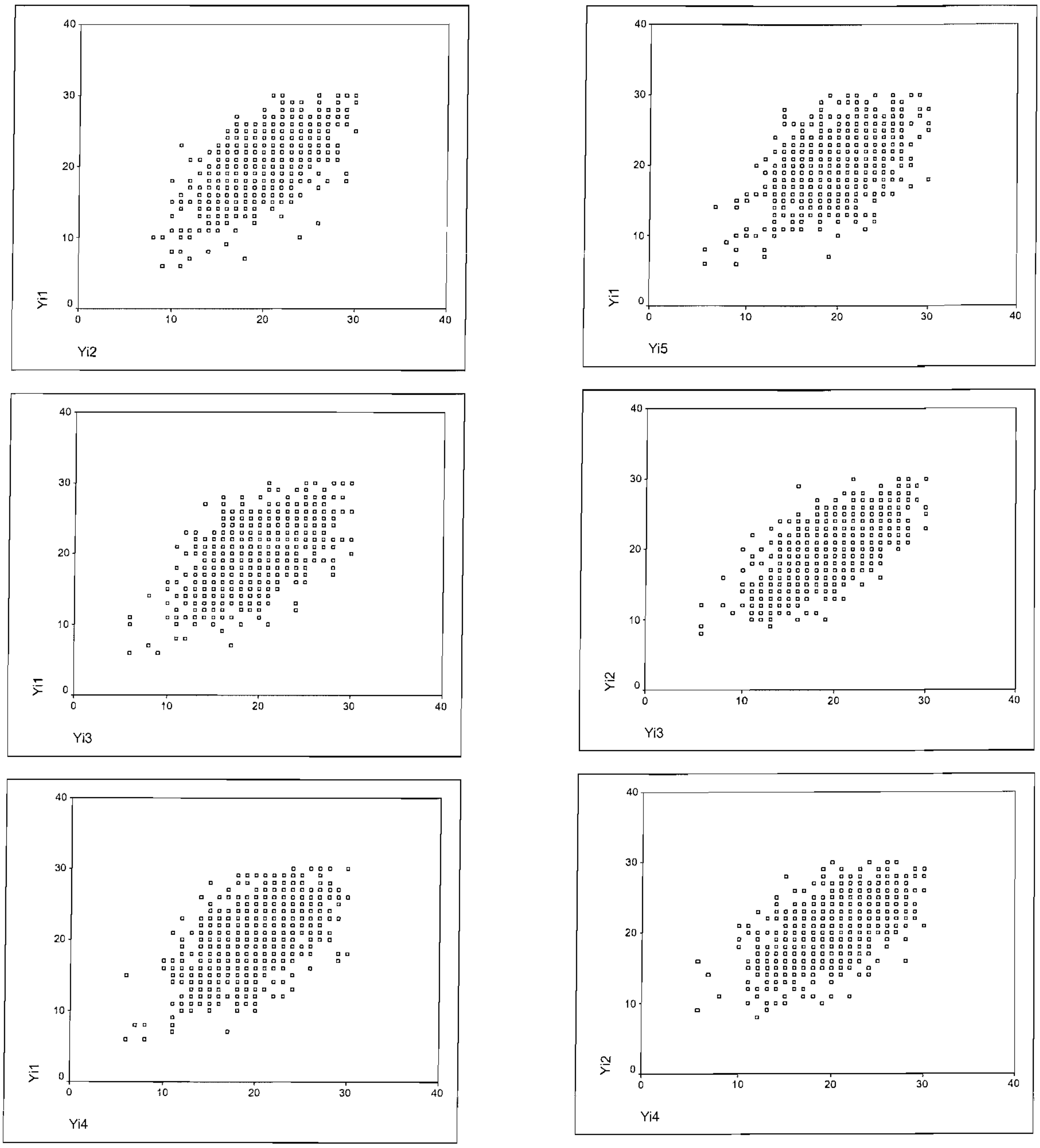

Figure B.2 – Scatter Plots for $\underline{y}_{it} \times \underline{y}_{it'}$.

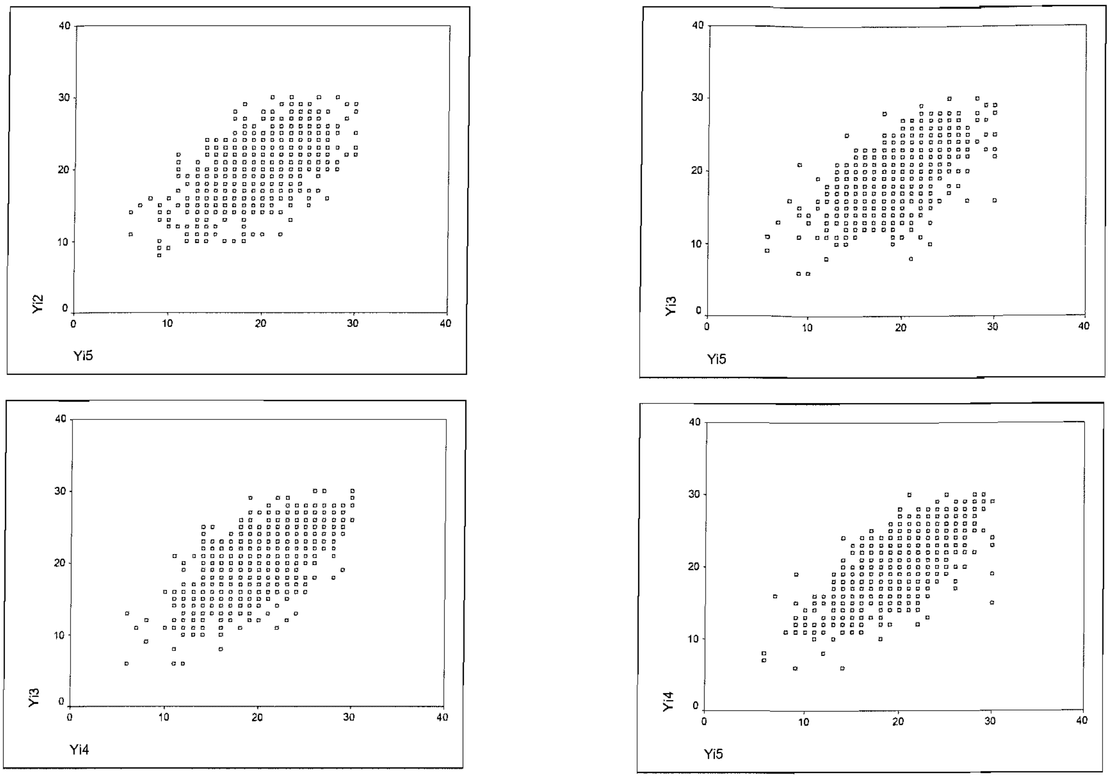

Figure B.2 – Scatter Plots for $\underline{y}_{it} \times \underline{y}_{it'}$ (continued).

If y is multivariate normally distributed we would expect to have to scatter plots above having elliptical appearance. By analysing Figure B.2, we cannot observe any characteristic that could give us enough evidence to conclude that y is not multivariate normally distributed.

Moreover, we may present below a graphical display of the repeated observations for the response variable for each individual.

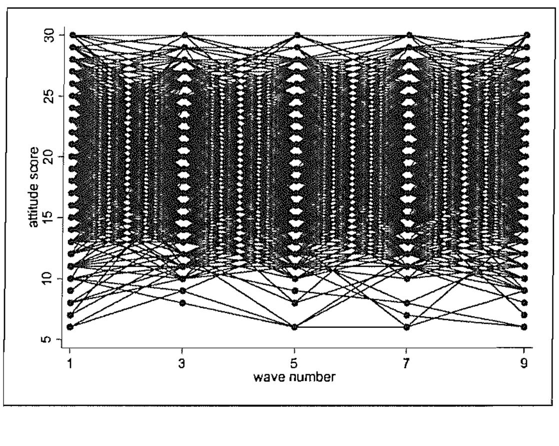

Figure B.3 – Data on the attitude scores of 1340 women over a nine-year period.

On Figure B.3, lines connect the repeated observation for each woman. Note however, as all the observations from the adopted BHPS subset (see Chapter 3, Section 3.2, Sub-section 3.2.2) were used, it is very hard to observe individual response profiles. We shall represent below only a sample of 25 women selected from the original data subset, mainly in order to exemplify the type of pattern we may expect to observe.

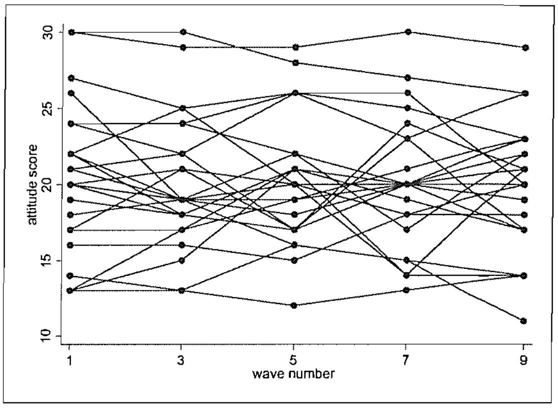

Figure B.4 – Data on the attitude scores of 25 women over a nine-year period.

Appendix C

Explicit solution for parameters when adopting ULSC and PML estimation methods and fitting a UCM model

Assuming the model discussed in Chapter 2, Example 2.1, the $T \times T$ matrix $\Sigma(\underline{\theta})$ may be represented as

$$\Sigma(\underline{\theta}) = \begin{pmatrix} \sigma_u^2 + \sigma_v^2 & & & \\ \sigma_u^2 & \sigma_u^2 + \sigma_v^2 & & \\ \vdots & \vdots & \ddots & \\ \sigma_u^2 & \sigma_u^2 & \cdots & \sigma_u^2 + \sigma_v^2 \end{pmatrix}, \qquad \text{(C.1)}$$

and

$$\hat{\Sigma} = \Sigma(\hat{\underline{\theta}}) = \begin{pmatrix} \hat{\sigma}_u^2 + \hat{\sigma}_v^2 & & & \\ \hat{\sigma}_u^2 & \hat{\sigma}_u^2 + \hat{\sigma}_v^2 & & \\ \vdots & \vdots & \ddots & \\ \hat{\sigma}_u^2 & \hat{\sigma}_u^2 & \cdots & \hat{\sigma}_u^2 + \hat{\sigma}_v^2 \end{pmatrix}.$$

We may expand the unweighted least squares for complex survey data fitting function given by (4.43b) in Chapter 4, Sub-section 4.4.1, as

$$\begin{aligned} F(\underline{\theta})_{ULSC} = \frac{1}{2}\Big\{ & \left[S_{w,11} - \sigma_u^2 - \sigma_v^2\right]^2 + \left[S_{w,22} - \sigma_u^2 - \sigma_v^2\right]^2 + \left[S_{w,33} - \sigma_u^2 - \sigma_v^2\right]^2 + \left[S_{w,44} - \sigma_u^2 - \sigma_v^2\right]^2 \\ & + \left[S_{w,44} - \sigma_u^2 - \sigma_v^2\right]^2 \Big\} + \left[S_{w,12} - \sigma_u^2\right]^2 + \left[S_{w,13} - \sigma_u^2\right]^2 + \left[S_{w,14} - \sigma_u^2\right]^2 + \left[S_{w,15} - \sigma_u^2\right]^2 \\ & + \left[S_{w,23} - \sigma_u^2\right]^2 + \left[S_{w,24} - \sigma_u^2\right]^2 + \left[S_{w,25} - \sigma_u^2\right]^2 + \left[S_{w,34} - \sigma_u^2\right]^2 + \left[S_{w,35} - \sigma_u^2\right]^2 + \left[S_{w,45} - \sigma_u^2\right]^2 \end{aligned}$$

for $T = 5$, where $S_{w,tt'}$ are elements of the symmetric matrix $\left[\mathrm{S}_w\right]_{tt'}$ as defined in Sub-section 4.3.1.

The partial derivatives are

$$\frac{\partial F(\underline{\theta})_{ULSC}}{\partial \sigma_u^2} = -S_{w,11} - S_{w,22} - S_{w,33} - S_{w,44} - S_{w,55} - 2S_{w,12} - 2S_{w,13} - 2S_{w,14} - 2S_{w,15} - 2S_{w,23} - 2S_{w,24} - 2S_{w,25} - 2S_{w,34} - 2S_{w,35} - 2S_{w,45} + 25\sigma_u^2 + 5\sigma_v^2, \quad \text{(C.2)}$$

and

$$\frac{\partial F(\underline{\theta})_{ULSC}}{\partial \sigma_v^2} = -S_{w,11} - S_{w,22} - S_{w,33} - S_{w,44} - S_{w,55} + 5\sigma_u^2 + 5\sigma_v^2 \quad \text{(C.3)}$$

Setting (C.2) and (C.3) to zero and solving this system of two equations and two parameters σ_u^2 and σ_v^2, we find that the solution is respectively

$$\hat{\sigma}_{u,ULSC}^2 = \frac{1}{10}\left[S_{w,12} + S_{w,13} + S_{w,14} + S_{w,15} + S_{w,23} + S_{w,24} + S_{w,25} + S_{w,34} + S_{w,35} + S_{w,45}\right], \quad \text{(C.4)}$$

and

$$\hat{\sigma}_{v,ULSC}^2 = \frac{1}{5}\left[S_{w,11} + S_{w,22} + S_{w,33} + S_{w,44} + S_{w,55}\right] - \frac{1}{10}\left[S_{w,12} + S_{w,13} + S_{w,14} + S_{w,15} + S_{w,23} + S_{w,24} + S_{w,25} + S_{w,34} + S_{w,35} + S_{w,45}\right]. \quad \text{(C.5)}$$

The estimators $\hat{\sigma}_{u,ULS}^2$ and $\hat{\sigma}_{v,ULS}^2$ may be obtained by substituting S_w for S in (C.4) and (C.5).

Furthermore, analytic solutions for the pseudo maximum likelihood estimators $\hat{\sigma}_{u,PML}^2$ and $\hat{\sigma}_{v,PML}^2$ may also be found when assuming the model discussed in Chapter 2, Example 2.1, with $\Sigma(\underline{\theta})$ given by (C.1) and $T = 5$. We may expand the PML fitting function (4.63). We have performed this expansion with the aid of the computer software Maple version 9.51 and shall not present detailed results here because of the long length. By taking the partial derivatives of $F(\underline{\theta})_{PML}$ with respect to σ_u^2 and σ_v^2, and then setting the two equations to zero solving that system. For this specific situation we have found that $\hat{\sigma}_{u,PML}^2$ and $\hat{\sigma}_{v,PML}^2$ would be also given by (C.4) and (C.5) respectively.

Appendix D

Differential calculus results relevant to Chapters 4 and 6

Let (see Chapter 4, Sub-section 4.4.2)

$$F(\underline{\theta})^2_{GLSC} = \left(\frac{1}{2}\right) \cdot tr\left\{\left[(S_w - \Sigma(\underline{\theta}))S_w^{-1}\right]^2\right\}. \tag{D.1}$$

Result D.1 We may obtain an expression for $\hat{\underline{\theta}}^2_{GLSC}$ by solving

$$\frac{F(\underline{\theta})^2_{GLSC}}{\partial\theta_j} = tr\left\{S_w^{-1} \cdot \left[\Sigma(\underline{\theta}) - S_w\right] \cdot S_w^{-1} \cdot \frac{\partial\Sigma(\underline{\theta})}{\partial\theta_j}\right\} = 0. \tag{D.2}$$

Proof.

As both S_w and $\Sigma(\underline{\theta})$ are symmetric matrices, we may re-write (D.1) so that

$$F(\underline{\theta})^2_{GLSC} = \left(\frac{1}{2}\right) \cdot tr\left\{\left[(S_w - \Sigma(\underline{\theta}))S_w^{-1}\right]\left[(S_w - \Sigma(\underline{\theta}))S_w^{-1}\right]\right\} = \left(\frac{1}{2}\right) \cdot tr\left\{S_w^{-1}(S_w - \Sigma(\underline{\theta}))^2 S_w^{-1}\right\} =$$
$$= \left(\frac{1}{2}\right) \cdot tr\left\{I - 2\Sigma(\underline{\theta})S_w^{-1} + S_w^{-1}\Sigma(\underline{\theta})^2 S_w^{-1}\right\}, \tag{D.3}$$

where I denotes an identity matrix.

We may then differentiate (D.3) with respect to $\underline{\theta}$. It follows that

$$\frac{\partial F(\underline{\theta})^2_{GLSC}}{\partial\theta_j} = \left(\frac{1}{2}\right) \cdot tr\left\{\frac{\partial I}{\partial\theta_j} - \frac{\partial 2\Sigma(\underline{\theta})S_w^{-1}}{\partial\theta_j} + \frac{\partial S_w^{-1}\Sigma(\underline{\theta})^2 S_w^{-1}}{\partial\theta_j}\right\} =$$
$$= \left(\frac{1}{2}\right) \cdot tr\left\{-2 \cdot \frac{\partial S_w^{-1}\Sigma(\underline{\theta})}{\partial\theta_j} + \frac{\partial S_w^{-2}\Sigma(\underline{\theta})^2}{\partial\theta_j}\right\} ==$$
$$= \left(\frac{1}{2}\right) \cdot tr\left\{-2S_w^{-1} \cdot \frac{\partial\Sigma(\underline{\theta})}{\partial\theta_j} + 2S_w^{-2}\Sigma(\underline{\theta}) \cdot \frac{\partial\Sigma(\underline{\theta})}{\partial\theta_j}\right\} = tr\left\{S_w^{-2}\Sigma(\underline{\theta}) \cdot \frac{\partial\Sigma(\underline{\theta})}{\partial\theta_j} - S_w^{-1} \cdot \frac{\partial\Sigma(\underline{\theta})}{\partial\theta_j}\right\} =$$
$$= tr\left\{\left[S_w^{-2}\Sigma(\underline{\theta}) - S_w^{-1}\right] \cdot \frac{\partial\Sigma(\underline{\theta})}{\partial\theta_j}\right\} = tr\left\{\left[S_w^{-1}\Sigma(\underline{\theta}) - I\right] \cdot S_w^{-1} \cdot \frac{\partial\Sigma(\underline{\theta})}{\partial\theta_j}\right\}, \tag{D.4}$$

which equals (D.2). ■

Moreover, let (see Chapter 4, Sub-section 4.4.2)

$$F(\underline{\theta})^2_{GLS} = \left(\frac{1}{2}\right) \cdot tr\left\{\left[(\mathrm{S} - \Sigma(\underline{\theta}))\mathrm{S}^{-1}\right]^2\right\}.$$

We may show that $\hat{\underline{\theta}}^2_{GLS}$ may be obtained by solving

$$\frac{F(\underline{\theta})^2_{GLS}}{\partial\theta_j} = tr\left\{\mathrm{S}^{-1} \cdot \left[\Sigma(\underline{\theta}) - \mathrm{S}\ \right] \cdot \mathrm{S}^{-1} \cdot \frac{\partial\Sigma(\underline{\theta})}{\partial\theta_j}\right\} = 0,$$

by substituting S_w by S in (D.4). ■

Furthermore, let matrices S_w and $\Sigma(\underline{\theta})$ be positive definite, and $|\Sigma(\underline{\theta})|$ denote the determinant of $\Sigma(\underline{\theta})$, and let (see Chapter 4, Sub-section 4.4.4)

$$F(\underline{\theta})_{PML} = tr\left[\mathrm{S}_w\Sigma(\underline{\theta})^{-1}\right] - \log\left|\mathrm{S}_w\Sigma(\underline{\theta})^{-1}\right|,$$

be the pseudo maximum likelihood fitting function, when ignoring the term $-T$ in (4.62).

Result D.2 We may obtain an expression for the $\hat{\underline{\theta}}_{PML}$ by solving

$$\frac{\partial F(\underline{\theta})_{PML}}{\partial\theta_j} = tr\left\{\Sigma(\underline{\theta})^{-1} \cdot \left[\Sigma(\underline{\theta}) - \mathrm{S}_w\right] \cdot \Sigma(\underline{\theta})^{-1} \cdot \frac{\partial\Sigma(\underline{\theta})}{\partial\theta_j}\right\} = 0. \qquad \text{(D.5)}$$

Proof.

Let

$$\frac{\partial F(\underline{\theta})_{PML}}{\partial\theta_j} = \frac{\partial tr\left[\mathrm{S}_w\Sigma(\underline{\theta})^{-1}\right]}{\partial\theta_j} - \frac{\partial \log\left|\mathrm{S}_w\Sigma(\underline{\theta})^{-1}\right|}{\partial\theta_j}. \qquad \text{(D.6)}$$

We may split (D.6) in two terms,

$$\frac{\partial tr\left[\mathrm{S}_w\Sigma(\underline{\theta})^{-1}\right]}{\partial\theta_j}$$

and

$$-\frac{\partial \log\left|\mathrm{S}_w\Sigma(\underline{\theta})^{-1}\right|}{\partial\theta_j}.$$

Let,

$$\frac{\partial tr\left[S_w \Sigma(\underline{\theta})^{-1}\right]}{\partial \theta_j} = tr\frac{\partial\left[S_w \Sigma(\underline{\theta})^{-1}\right]}{\partial \theta_j} = tr\left[\frac{\partial S_w}{\partial \theta_j}\cdot \Sigma(\underline{\theta})^{-1} + S_w \frac{\partial \Sigma(\underline{\theta})^{-1}}{\partial \theta_j}\right],$$

and as

$$\frac{\partial S_w}{\partial \theta_j} = 0,$$

then

$$\frac{\partial tr\left[S_w \Sigma(\underline{\theta})^{-1}\right]}{\partial \theta_j} = tr\left\{S_w \cdot \left[\left[-\Sigma(\underline{\theta})^{-1}\right]\frac{\partial \Sigma(\underline{\theta})}{\partial \theta_j}\left[\Sigma(\underline{\theta})^{-1}\right]\right]\right\}. \tag{D.6}$$

Let,

$$-\frac{\partial \log\left|S_w \Sigma(\underline{\theta})^{-1}\right|}{\partial \theta_j} = -tr\left\{\left[S_w \Sigma(\underline{\theta})^{-1}\right]^{-1}\cdot \frac{\partial\left[S_w \Sigma(\underline{\theta})^{-1}\right]}{\partial \theta_j}\right\} =$$
$$= -tr\left\{S_w^{-1}\Sigma(\underline{\theta})\cdot\left[\frac{\partial S_w}{\partial \theta_j}\Sigma(\underline{\theta})^{-1} + S_w \frac{\partial\left[\Sigma(\underline{\theta})^{-1}\right]}{\partial \theta_j}\right]\right\} =$$
$$= -tr\left\{S_w^{-1}\Sigma(\underline{\theta})S_w \cdot \frac{\partial\left[\Sigma(\underline{\theta})^{-1}\right]}{\partial \theta_j}\right\} = -tr\left\{\Sigma(\underline{\theta})\cdot\left[-\Sigma(\underline{\theta})^{-1}\right]\frac{\partial \Sigma(\underline{\theta})}{\partial \theta_j}\left[\Sigma(\underline{\theta})^{-1}\right]\right\} =$$
$$= -tr\left\{\left[-\Sigma(\underline{\theta})^{-1}\right]\frac{\partial \Sigma(\underline{\theta})}{\partial \theta_j}\right\} = tr\left\{\Sigma(\underline{\theta})^{-1}\frac{\partial \Sigma(\underline{\theta})}{\partial \theta_j}\right\}.$$

(D.7)

We may now plug (D.6) and (D.7) back in (D.6). It follows that

$$\frac{\partial F(\underline{\theta})_{PML}}{\partial \theta_j} = tr\left\{S_w \cdot \left[\left[-\Sigma(\underline{\theta})^{-1}\right]\frac{\partial \Sigma(\underline{\theta})}{\partial \theta_j}\left[\Sigma(\underline{\theta})^{-1}\right]\right]\right\} + tr\left\{\Sigma(\underline{\theta})^{-1}\frac{\partial \Sigma(\underline{\theta})}{\partial \theta_j}\right\} =$$
$$= tr\left\{\Sigma(\underline{\theta})^{-1}\frac{\partial \Sigma(\underline{\theta})}{\partial \theta_j} + \Sigma(\underline{\theta})^{-1}\cdot(-S_w)\cdot\Sigma(\underline{\theta})^{-1}\cdot\frac{\partial \Sigma(\underline{\theta})}{\partial \theta_j}\right\} =$$
$$= tr\left\{\Sigma(\underline{\theta})^{-1}\cdot\left[\frac{\partial \Sigma(\underline{\theta})}{\partial \theta_j} + (-S_w)\cdot\Sigma(\underline{\theta})^{-1}\cdot\frac{\partial \Sigma(\underline{\theta})}{\partial \theta_j}\right]\right\} = \tag{D.8}$$
$$= tr\left\{\Sigma(\underline{\theta})^{-1}\cdot\left[I + (-S_w)\cdot\Sigma(\underline{\theta})^{-1}\right]\cdot\frac{\partial \Sigma(\underline{\theta})}{\partial \theta_j}\right\},$$

which equals (D.5). In (D.8), I denotes an identity matrix. ■

Let the maximum likelihood fitting function be given by (see Chapter 4, Sub-section 4.4.4)

$$F(\underline{\theta})_{ML} = tr\left[S\Sigma(\underline{\theta})^{-1}\right] - \log\left|S\Sigma(\underline{\theta})^{-1}\right|,$$

when ignoring the term $-T$ in (4.55).

We may show that an expression for the $\hat{\underline{\theta}}_{ML}$ may be obtained by solving

$$\frac{\partial F(\underline{\theta})_{ML}}{\partial \theta_{\mathrm{j}}} = tr\left\{\Sigma(\underline{\theta})^{-1}\cdot\left[\Sigma(\underline{\theta}) - \mathrm{S}\right]\cdot\Sigma(\underline{\theta})^{-1}\cdot\frac{\partial \Sigma(\underline{\theta})}{\partial \theta_{\mathrm{j}}}\right\} = 0,$$

by substituting S_w by S in (D.6) and (D.7). ■

Result D.3 Note that (D.5) may be expressed as constant plus a ratio of two totals, (see Chapter 6, Section 6.3)

$$tr\left[\Sigma(\underline{\theta})^{-1}\cdot\frac{\partial \Sigma(\underline{\theta})}{\partial \theta_{\mathrm{j}}}\right] + \frac{\sum_{i=1}^{n} w_i z_{\mathrm{j}i}}{\sum_{i=1}^{n} w_i}$$

where

$$z_{\mathrm{j}i} = -\left(\underline{y}_i - \hat{\underline{\mu}}_i\right)'\Sigma(\underline{\theta})^{-1}\frac{\partial \Sigma(\underline{\theta})}{\partial \theta_{\mathrm{j}}}\Sigma(\underline{\theta})^{-1}\left(\underline{y}_i - \hat{\underline{\mu}}_i\right).$$

Proof.

$$\frac{\partial F(\underline{\theta})_{PML}}{\partial \theta_{\mathrm{j}}} = tr\left\{\Sigma(\underline{\theta})^{-1}\cdot\left[\Sigma(\underline{\theta}) - \mathrm{S}_w\right]\cdot\Sigma(\underline{\theta})^{-1}\cdot\frac{\partial \Sigma(\underline{\theta})}{\partial \theta_{\mathrm{j}}}\right\} =$$

$$= tr\left\{\left[I - \Sigma(\underline{\theta})^{-1}\mathrm{S}_w\right]\Sigma(\underline{\theta})^{-1}\cdot\frac{\partial \Sigma(\underline{\theta})}{\partial \theta_{\mathrm{j}}}\right\} =$$

$$= tr\left\{\left[\Sigma(\underline{\theta})^{-1} - \Sigma(\underline{\theta})^{-1}\mathrm{S}_w\Sigma(\underline{\theta})^{-1}\right]\cdot\frac{\partial \Sigma(\underline{\theta})}{\partial \theta_{\mathrm{j}}}\right\} =$$

$$= tr\left[\Sigma(\underline{\theta})^{-1}\cdot\frac{\partial \Sigma(\underline{\theta})}{\partial \theta_{\mathrm{j}}}\right] - tr\left[\Sigma(\underline{\theta})^{-1}\mathrm{S}_w\Sigma(\underline{\theta})^{-1}\cdot\frac{\partial \Sigma(\underline{\theta})}{\partial \theta_{\mathrm{j}}}\right].$$

Note that

$$-tr\left[\Sigma(\underline{\theta})^{-1}\mathrm{S}_w\Sigma(\underline{\theta})^{-1}\cdot\frac{\partial \Sigma(\underline{\theta})}{\partial \theta_{\mathrm{j}}}\right] = \frac{\sum_{i=1}^{n} w_i\left[-\left(\underline{y}_i - \hat{\underline{\mu}}_i\right)'\Sigma(\underline{\theta})^{-1}\frac{\partial \Sigma(\underline{\theta})}{\partial \theta_{\mathrm{j}}}\Sigma(\underline{\theta})^{-1}\left(\underline{y}_i - \hat{\underline{\mu}}_i\right)\right]}{\sum_{i=1}^{n} w_i}.$$ ■

Result D.4 When estimating of $a\mathrm{COV}(\hat{\underline{\theta}}_{PML})$, by assuming that the model is correct, i.e. that

$$E[\mathrm{S}_w] = \Sigma(\underline{\theta}), \tag{D.9}$$

we could define the information matrix alternatively as (see Chapter 6, Section 6.3)

$$\mathrm{I}(\underline{\theta}) = E\left[-\frac{\partial \phi(\underline{\theta})}{\partial \underline{\theta}}\right].$$

In this situation, the jk^{th} element of $\mathrm{I}(\underline{\theta})$ is

$$\mathrm{I}(\underline{\theta})_{\mathrm{jk}} = tr\left[\Sigma(\underline{\theta})^{-1}\frac{\partial \Sigma(\underline{\theta})}{\partial \theta_{\mathrm{j}}}\Sigma(\underline{\theta})^{-1}\frac{\partial \Sigma(\underline{\theta})}{\partial \theta_{\mathrm{k}}}\right],$$

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

Proof.

Recall that

$$\phi_{\mathrm{j}}(\underline{\theta}) = tr\left\{\Sigma(\underline{\theta})^{-1}\cdot[\Sigma(\underline{\theta}) - \mathrm{S}_w]\cdot\Sigma(\underline{\theta})^{-1}\cdot\frac{\partial \Sigma(\underline{\theta})}{\partial \theta_{\mathrm{j}}}\right\},$$

and thus

$$E\left[-\frac{\phi_{\mathrm{j}}(\underline{\theta})}{\partial \theta_{\mathrm{j}}}\right] = tr\left\{E\left[\frac{\partial\left(\Sigma(\underline{\theta})^{-1}(\Sigma(\underline{\theta}) - \mathrm{S}_w)\Sigma(\underline{\theta})^{-1}\right)}{\partial \theta_{\mathrm{j}}}\cdot\frac{\partial \Sigma(\underline{\theta})}{\partial \theta_{\mathrm{j}}} + \Sigma(\underline{\theta})^{-1}(\Sigma(\underline{\theta}) - \mathrm{S}_w)\Sigma(\underline{\theta})^{-1}\cdot\frac{\partial^2 \Sigma(\underline{\theta})}{\partial \theta_{\mathrm{j}}\partial \theta_{\mathrm{j}}}\right]\right\},$$

and as we assume (D.9),

$$E\left[-\frac{\phi_{\mathrm{j}}(\underline{\theta})}{\partial \theta_{\mathrm{j}}}\right] = tr\left\{E\left[\frac{\partial\left(\Sigma(\underline{\theta})^{-1}(\Sigma(\underline{\theta}) - \mathrm{S}_w)\Sigma(\underline{\theta})^{-1}\right)}{\partial \theta_{\mathrm{j}}}\cdot\frac{\partial \Sigma(\underline{\theta})}{\partial \theta_{\mathrm{j}}}\right]\right\} =$$

$$= tr\left\{E\left[\left(\frac{\partial \Sigma(\underline{\theta})^{-1}}{\partial \theta_{\mathrm{j}}}\cdot(\Sigma(\underline{\theta}) - \mathrm{S}_w)\Sigma(\underline{\theta})^{-1} + \Sigma(\underline{\theta})^{-1}\frac{\partial\left((\Sigma(\underline{\theta}) - \mathrm{S}_w)\Sigma(\underline{\theta})^{-1}\right)}{\partial \theta_{\mathrm{j}}}\right)\cdot\frac{\partial \Sigma(\underline{\theta})}{\partial \theta_{\mathrm{j}}}\right]\right\},$$

$$E\left[-\frac{\phi_{\mathrm{j}}(\underline{\theta})}{\partial \theta_{\mathrm{j}}}\right] = tr\left\{E\left[\left(\Sigma(\underline{\theta})^{-1}\frac{\partial\left((\Sigma(\underline{\theta}) - \mathrm{S}_w)\Sigma(\underline{\theta})^{-1}\right)}{\partial \theta_{\mathrm{j}}}\right)\cdot\frac{\partial \Sigma(\underline{\theta})}{\partial \theta_{\mathrm{j}}}\right]\right\} =$$

$$= tr\left\{E\left[\left(\Sigma(\underline{\theta})^{-1}\cdot\left(\frac{\partial \Sigma(\underline{\theta})}{\partial \theta_{\mathrm{j}}}\Sigma(\underline{\theta})^{-1} + (\Sigma(\underline{\theta}) - \mathrm{S}_w)\frac{\partial \Sigma(\underline{\theta})^{-1}}{\partial \theta_{\mathrm{j}}}\right)\right)\cdot\frac{\partial \Sigma(\underline{\theta})}{\partial \theta_{\mathrm{j}}}\right]\right\}$$

It follows that,

$$E\left[-\frac{\phi_a(\underline{\theta})}{\partial\theta_{\mathrm{j}}}\right]=tr\left\{\Sigma(\underline{\theta})^{-1}\frac{\partial\Sigma(\underline{\theta})}{\partial\theta_{\mathrm{j}}}\Sigma(\underline{\theta})^{-1}\frac{\partial\Sigma(\underline{\theta})}{\partial\theta_{\mathrm{k}}}\right\}.$$

Thus,

$$\mathrm{I}(\underline{\theta})_{\mathrm{jk}}=tr\left[\Sigma(\underline{\theta})^{-1}\frac{\partial\Sigma(\underline{\theta})}{\partial\theta_{\mathrm{j}}}\Sigma(\underline{\theta})^{-1}\frac{\partial\Sigma(\underline{\theta})}{\partial\theta_{\mathrm{k}}}\right]. \qquad \blacksquare$$

Appendix E

More detailed results obtained in simulation study I

Tables E.1 to E.8 below provide more detailed information about the results obtained in simulation study I, when considering $n^{sim} = 1340$.

parameter		pop value	$\hat{E}(\hat{\theta})$	rel bias	min	max	$\text{var}(\hat{\theta})$	$cv(\hat{\theta})$	s.e $[\hat{E}(\hat{\theta})]$	conf. interval for $\hat{E}(\hat{\theta})$		mse
										lower bound	upper bound	
$\hat{\underline{\theta}}_{ULS}$	$\hat{\sigma}_u^2$	6.744011	6.684898	-0.88%	5.728593	7.709990	0.081212	4.26%	0.009012	6.666875	6.702922	0.084707
	$\hat{\sigma}_v^2$	4.965116	4.968341	0.06%	4.677198	5.229038	0.008644	1.87%	0.002940	4.962461	4.974221	0.008655
$\hat{\underline{\theta}}^2_{GLS}$	$\hat{\sigma}_u^2$	6.744011	6.648070	-1.42%	5.699861	7.681686	0.081479	4.29%	0.009027	6.630017	6.666124	0.090683
	$\hat{\sigma}_v^2$	4.965116	4.928314	-0.74%	4.642093	5.201775	0.008808	1.90%	0.002968	4.922378	4.934250	0.010162
$\hat{\underline{\theta}}^3_{GLS}$	$\hat{\sigma}_u^2$	6.744011	6.703272	-0.60%	5.742010	7.724002	0.081821	4.27%	0.009045	6.685181	6.721363	0.083481
	$\hat{\sigma}_v^2$	4.965116	4.988362	0.47%	4.691102	5.243107	0.008757	1.88%	0.002959	4.982444	4.994280	0.009297
$\hat{\underline{\theta}}^4_{GLS}$	$\hat{\sigma}_u^2$	6.744011	6.645528	-1.46%	5.641908	7.686811	0.082728	4.33%	0.009096	6.627337	6.663719	0.092427
	$\hat{\sigma}_v^2$	4.965116	4.928468	-0.74%	4.649028	5.197126	0.009022	1.93%	0.003004	4.922460	4.934475	0.010365
$\hat{\underline{\theta}}_{ML}$	$\hat{\sigma}_u^2$	6.744011	6.684898	-0.88%	5.728593	7.709990	0.081212	4.26%	0.009012	6.666875	6.702922	0.084707
	$\hat{\sigma}_v^2$	4.965116	4.968341	0.06%	4.677198	5.229038	0.008644	1.87%	0.002940	4.962461	4.974221	0.008655
$\hat{\underline{\theta}}_{ULSC}$	$\hat{\sigma}_u^2$	6.744011	6.721061	-0.34%	5.759628	7.798238	0.094536	4.57%	0.009723	6.701615	6.740507	0.095062
	$\hat{\sigma}_v^2$	4.965116	4.951611	-0.27%	4.659896	5.242011	0.009949	2.01%	0.003154	4.945302	4.957919	0.010131
$\hat{\underline{\theta}}^2_{GLSC}$	$\hat{\sigma}_u^2$	6.744011	6.679199	-0.96%	5.714510	7.773242	0.094767	4.61%	0.009735	6.659729	6.698668	0.098968
	$\hat{\sigma}_v^2$	4.965116	4.906411	-1.18%	4.614110	5.187970	0.010170	2.06%	0.003189	4.900033	4.912789	0.013616
$\hat{\underline{\theta}}^3_{GLSC}$	$\hat{\sigma}_u^2$	6.744011	6.741997	-0.03%	5.782132	7.830875	0.095398	4.58%	0.009767	6.722462	6.761531	0.095402
	$\hat{\sigma}_v^2$	4.965116	4.974265	0.18%	4.680629	5.269621	0.010086	2.02%	0.003176	4.967913	4.980617	0.010170
$\hat{\underline{\theta}}^4_{GLSC}$	$\hat{\sigma}_u^2$	6.744011	6.672204	-1.06%	5.686740	7.807193	0.096314	4.65%	0.009814	6.652576	6.691832	0.101470
	$\hat{\sigma}_v^2$	4.965116	4.901303	-1.29%	4.610579	5.185511	0.01050	2.09%	0.003240	4.894822	4.907784	0.014572
$\hat{\underline{\theta}}_{PML}$	$\hat{\sigma}_u^2$	6.744011	6.721061	-0.34%	5.759628	7.798238	0.094536	4.57%	0.009723	6.701615	6.740507	0.095062
	$\hat{\sigma}_v^2$	4.965116	4.951611	-0.27%	4.659896	5.242011	0.009949	2.01%	0.003154	4.945302	4.957919	0.010131

Table E.1 – Evaluation of $\hat{\underline{\theta}}$ for UCM model, normally distributed errors (population, replications generated by UCM model) - further results.

parameter		pop value	$\hat{E}(\hat{\theta})$	rel bias	min	max	$\mathrm{var}(\hat{\theta})$	$cv(\hat{\theta})$	s.e [$\hat{E}(\hat{\theta})$]	conf. interval for $\hat{E}(\hat{\theta})$		mse
										lower bound	upper bound	
$\underline{\hat{\theta}}_{ULS}$	$\hat{\sigma}_u^2$	6.744011	6.684080	-0.89%	5.637698	7.722812	0.082905	4.31%	0.009105	6.665870	6.702290	0.086497
	$\hat{\sigma}_v^2$	4.965116	4.969155	0.08%	4.641652	5.264862	0.010416	2.05%	0.003227	4.962701	4.975610	0.010432
	$\hat{\gamma}$	0.000000	-0.000190	-	-0.064315	0.072885	0.000537	-	0.000733	-0.001656	0.001276	0.000537
$\underline{\hat{\theta}}^2_{GLS}$	$\hat{\sigma}_u^2$	6.744011	6.646476	-1.45%	5.599646	7.688807	0.082558	4.32%	0.009086	6.628304	6.664649	0.092071
	$\hat{\sigma}_v^2$	4.965116	4.933144	-0.64%	4.611927	5.207543	0.009920	2.02%	0.003150	4.926844	4.939443	0.010942
	$\hat{\gamma}$	0.000000	0.000140	-	-0.060618	0.058398	0.000355	-	0.000596	-0.001052	0.001331	0.000355
$\underline{\hat{\theta}}^3_{GLS}$	$\hat{\sigma}_u^2$	6.744011	6.702699	-0.61%	5.650292	7.756588	0.082793	4.29%	0.009099	6.684501	6.720897	0.084500
	$\hat{\sigma}_v^2$	4.965116	4.987288	0.45%	4.644202	5.256862	0.009923	2.00%	0.003150	4.980988	4.993588	0.010415
	$\hat{\gamma}$	0.000000	0.000006	-	-0.060909	0.06066	0.000348	-	0.000590	-0.001174	0.001186	0.000348
$\underline{\hat{\theta}}^4_{GLS}$	$\hat{\sigma}_u^2$	6.744011	6.644307	-1.48%	5.540856	7.692414	0.083628	4.35%	0.009145	6.626017	6.662597	0.093569
	$\hat{\sigma}_v^2$	4.965116	4.933038	-0.65%	4.615804	5.213083	0.010133	2.04%	0.003183	4.926671	4.939404	0.011162
	$\hat{\gamma}$	0.000000	0.000129	-	-0.063480	0.060272	0.000358	-	0.000598	-0.001068	0.001325	0.000358
$\underline{\hat{\theta}}_{ML}$	$\hat{\sigma}_u^2$	6.744011	6.683989	-0.89%	5.632939	7.722781	0.082220	4.29%	0.009068	6.665854	6.702124	0.085822
	$\hat{\sigma}_v^2$	4.965116	4.969213	0.08%	4.633311	5.236255	0.009803	1.99%	0.003131	4.962951	4.975475	0.009820
	$\hat{\gamma}$	0.000000	0.000048	-	-0.060818	0.059997	0.000348	-	0.000590	-0.001133	0.001228	0.000348

Table E.2a – Evaluation of $\underline{\hat{\theta}}$ for AR1 model, normally distributed errors (population, replications generated by UCM model) - further results, part a.

parameter		pop value	$\hat{E}(\hat{\theta})$	rel bias	min	max	$\mathrm{var}(\hat{\theta})$	$cv(\hat{\theta})$	s.e [$\hat{E}(\hat{\theta})$]	conf. interval for $\hat{E}(\hat{\theta})$ lower bound	conf. interval for $\hat{E}(\hat{\theta})$ upper bound	mse
$\underline{\hat{\theta}}_{ULSC}$	$\hat{\sigma}_u^2$	6.744011	6.721986	-0.33%	5.678600	7.795814	0.096010	4.61%	0.009798	6.702389	6.741583	0.096495
	$\hat{\sigma}_v^2$	4.965116	4.950683	-0.29%	4.628642	5.255275	0.011840	2.20%	0.003441	4.943801	4.957564	0.012048
	$\hat{\gamma}$	0.000000	-0.001128	-	-0.081778	0.086721	0.000592	-	0.000769	-0.002667	0.000411	0.000593
$\underline{\hat{\theta}}^2_{GLSC}$	$\hat{\sigma}_u^2$	6.744011	6.680122	-0.95%	5.619967	7.793276	0.095890	4.64%	0.009792	6.660538	6.699707	0.099971
	$\hat{\sigma}_v^2$	4.965116	4.909048	-1.13%	4.575270	5.209533	0.011446	2.18%	0.003383	4.902281	4.915814	0.014590
	$\hat{\gamma}$	0.000000	-0.001236	-	-0.073310	0.059544	0.000393	-	0.000627	-0.002490	0.000017	0.000394
$\underline{\hat{\theta}}^3_{GLSC}$	$\hat{\sigma}_u^2$	6.744011	6.744201	0.003%	5.696572	7.827110	0.096600	4.61%	0.009829	6.724544	6.763858	0.096600
	$\hat{\sigma}_v^2$	4.965116	4.970292	0.10%	4.624750	5.272025	0.011353	2.14%	0.003369	4.963553	4.977030	0.011379
	$\hat{\gamma}$	0.000000	-0.001380	-	-0.074022	0.061348	0.000388	-	0.000623	-0.002625	-0.000135	0.000390
$\underline{\hat{\theta}}^4_{GLSC}$	$\hat{\sigma}_u^2$	6.744011	6.672996	-1.05%	5.562603	7.824269	0.097156	4.67%	0.009857	6.653282	6.692709	0.102199
	$\hat{\sigma}_v^2$	4.965116	4.904301	-1.22%	4.567990	5.213887	0.011767	2.21%	0.003430	4.897440	4.911161	0.015466
	$\hat{\gamma}$	0.000000	-0.001211	-	-0.076389	0.062811	0.000400	-	0.000633	-0.002477	0.000054	0.000402
$\underline{\hat{\theta}}_{PML}$	$\hat{\sigma}_u^2$	6.744011	6.720216	-0.35%	4.081795	7.815855	0.102680	4.77%	0.010133	6.699950	6.740482	0.103246
	$\hat{\sigma}_v^2$	4.965116	4.952879	-0.25%	4.608172	8.050978	0.020852	2.92%	0.004566	4.943747	4.962012	0.021001
	$\hat{\gamma}$	0.000000	-0.000337	-	-0.073796	1.000229	0.001390	-	0.001179	-0.002695	0.002021	0.001391

Table E.2b – Evaluation of $\underline{\hat{\theta}}$ for AR1 model, normally distributed errors (population, replications generated by UCM model) - further results, part b.

parameter		pop value	$\hat{E}(\hat{\theta})$	rel bias	min	max	$var(\hat{\theta})$	$cv(\hat{\theta})$	s.e $[\hat{E}(\hat{\theta})]$	conf. interval for $\hat{E}(\hat{\theta})$		mse
										lower bound	upper bound	
$\hat{\underline{\theta}}_{ULS}$	$\hat{\sigma}_u^2$	7.221	7.158534	-0.87%	5.985205	8.340737	0.097869	4.37%	0.009893	7.138749	7.178320	0.101771
	$\hat{\sigma}_v^2$	4.981	4.984107	0.06%	4.693027	5.292875	0.008925	1.90%	0.002988	4.978132	4.990082	0.008935
$\hat{\underline{\theta}}^2_{GLS}$	$\hat{\sigma}_u^2$	7.221	7.117084	-1.44%	5.977545	8.266750	0.097942	4.40%	0.009897	7.097291	7.136877	0.108740
	$\hat{\sigma}_v^2$	4.981	4.942235	-0.78%	4.652814	5.269032	0.008975	1.92%	0.002996	4.936243	4.948227	0.010478
$\hat{\underline{\theta}}^3_{GLS}$	$\hat{\sigma}_u^2$	7.221	7.179365	-0.58%	5.988974	8.376041	0.098736	4.38%	0.009937	7.159492	7.199238	0.100470
	$\hat{\sigma}_v^2$	4.981	5.005060	0.48%	4.713303	5.304449	0.009108	1.91%	0.003018	4.999025	5.011096	0.009687
$\hat{\underline{\theta}}^4_{GLS}$	$\hat{\sigma}_u^2$	7.221	7.116674	-1.44%	5.963821	8.251120	0.097579	4.39%	0.009878	7.096918	7.136431	0.108463
	$\hat{\sigma}_v^2$	4.981	4.942078	-0.78%	4.654301	5.289679	0.009132	1.93%	0.003022	4.936035	4.948122	0.010647
$\hat{\underline{\theta}}_{ML}$	$\hat{\sigma}_u^2$	7.221	7.158534	-0.87%	5.985205	8.340737	0.097869	4.37%	0.009893	7.138749	7.178320	0.101771
	$\hat{\sigma}_v^2$	4.981	4.984107	0.06%	4.693027	5.292875	0.008925	1.90%	0.002988	4.978132	4.990082	0.008935
$\hat{\underline{\theta}}_{ULSC}$	$\hat{\sigma}_u^2$	7.221	7.196278	-0.34%	5.919973	8.383219	0.114133	4.69%	0.010683	7.174912	7.217645	0.114744
	$\hat{\sigma}_v^2$	4.981	4.969890	-0.22%	4.666173	5.316020	0.010012	2.01%	0.003164	4.963561	4.976218	0.010136
$\hat{\underline{\theta}}^2_{GLSC}$	$\hat{\sigma}_u^2$	7.221	7.148928	-1.00%	5.920690	8.348793	0.114363	4.73%	0.010694	7.127540	7.170316	0.119557
	$\hat{\sigma}_v^2$	4.981	4.922792	-1.17%	4.616839	5.292325	0.010065	2.04%	0.003173	4.916447	4.929137	0.013453
$\hat{\underline{\theta}}^3_{GLSC}$	$\hat{\sigma}_u^2$	7.221	7.220128	-0.01%	5.919321	8.421725	0.115134	4.70%	0.010730	7.198668	7.241589	0.115135
	$\hat{\sigma}_v^2$	4.981	4.993504	0.25%	4.689925	5.327507	0.010257	2.03%	0.003203	4.987099	4.999909	0.010413
$\hat{\underline{\theta}}^4_{GLSC}$	$\hat{\sigma}_u^2$	7.221	7.143911	-1.07%	5.864213	8.282758	0.114220	4.73%	0.010687	7.122537	7.165286	0.120163
	$\hat{\sigma}_v^2$	4.981	4.918713	-1.25%	4.612373	5.311029	0.010375	2.07%	0.003221	4.912271	4.925155	0.014255
$\hat{\underline{\theta}}_{PML}$	$\hat{\sigma}_u^2$	7.221	7.196278	-0.34%	5.919973	8.383219	0.114133	4.69%	0.010683	7.174912	7.217645	0.114744
	$\hat{\sigma}_v^2$	4.981	4.969890	-0.22%	4.666173	5.316020	0.010012	2.01%	0.003164	4.963561	4.976218	0.010136

Table E.3 - Evaluation of $\hat{\underline{\theta}}$ for UCM model, normally distributed errors (population, replications generated by UCM-C model) - further results.

parameter		pop value	$\hat{E}(\hat{\theta})$	rel bias	min	max	$\mathrm{var}(\hat{\theta})$	$cv(\hat{\theta})$	s.e $[\hat{E}(\hat{\theta})]$	conf. interval for $\hat{E}(\hat{\theta})$		mse
										lower bound	upper bound	
$\underline{\hat{\theta}}_{ULS}$	$\hat{\sigma}_u^2$	7.221	7.157192	-0.88%	5.950597	8.286306	0.098273	4.38%	0.009913	7.137366	7.177019	0.102345
	$\hat{\sigma}_v^2$	4.981	4.985438	0.09%	4.689615	5.356894	0.011017	2.11%	0.003319	4.978799	4.992076	0.011037
	$\hat{\gamma}$	0.000	0.000070	-	-0.060675	0.096566	0.000522	-	0.000722	-0.001375	0.001515	0.000522
$\underline{\hat{\theta}}^2_{GLS}$	$\hat{\sigma}_u^2$	7.221	7.115183	-1.47%	5.928219	8.252438	0.099048	4.42%	0.009952	7.095279	7.135088	0.110245
	$\hat{\sigma}_v^2$	4.981	4.947157	-0.68%	4.591754	5.315491	0.010651	2.09%	0.003264	4.940630	4.953684	0.011796
	$\hat{\gamma}$	0.000	0.000192	-	-0.054302	0.064516	0.000347	-	0.000589	-0.000986	0.001371	0.000347
$\underline{\hat{\theta}}^3_{GLS}$	$\hat{\sigma}_u^2$	7.221	7.178638	-0.59%	5.942271	8.357754	0.099681	4.40%	0.009984	7.158670	7.198606	0.101475
	$\hat{\sigma}_v^2$	4.981	5.004386	0.47%	4.680100	5.344372	0.010915	2.09%	0.003304	4.997778	5.010994	0.011462
	$\hat{\gamma}$	0.000	0.000184	-	-0.050321	0.065750	0.000344	-	0.000587	-0.000990	0.001357	0.000344
$\underline{\hat{\theta}}^4_{GLS}$	$\hat{\sigma}_u^2$	7.221	7.114976	-1.47%	5.945531	8.245595	0.098812	4.42%	0.009940	7.095095	7.134857	0.110053
	$\hat{\sigma}_v^2$	4.981	4.946783	-0.69%	4.591640	5.324690	0.010821	2.10%	0.003290	4.940204	4.953362	0.011992
	$\hat{\gamma}$	0.000	0.000136	-	-0.052156	0.066484	0.000352	-	0.000593	-0.001051	0.001322	0.000352
$\underline{\hat{\theta}}_{ML}$	$\hat{\sigma}_u^2$	7.221	7.157445	-0.88%	5.937630	8.323543	0.098866	4.39%	0.009943	7.137559	7.177331	0.102905
	$\hat{\sigma}_v^2$	4.981	4.985282	0.09%	4.662017	5.331072	0.010693	2.07%	0.003270	4.978742	4.991822	0.010711
	$\hat{\gamma}$	0.000	0.000183	-	-0.050583	0.065333	0.000343	-	0.000586	-0.000988	0.001354	0.000343

Table E.4a – Evaluation of $\underline{\hat{\theta}}$ for AR1 model, normally distributed errors (population, replications generated by UCM-C model) - further results, part a.

parameter		pop value	$\hat{E}(\hat{\theta})$	rel bias	min	max	$\text{var}(\hat{\theta})$	$cv(\hat{\theta})$	s.e $[\hat{E}(\hat{\theta})]$	conf. interval for $\hat{E}(\hat{\theta})$ lower bound	upper bound	mse
$\underline{\hat{\theta}}_{ULSC}$	$\hat{\sigma}_u^2$	7.221	7.196827	-0.33%	5.884590	8.333739	0.114121	4.69%	0.010683	7.175462	7.218193	0.114705
	$\hat{\sigma}_v^2$	4.981	4.969334	-0.23%	4.637085	5.381083	0.012242	2.23%	0.003499	4.962336	4.976332	0.012378
	$\hat{\gamma}$	0.000	-0.000963	-	-0.072566	0.103611	0.000602	-	0.000776	-0.002515	0.000589	0.000603
$\underline{\hat{\theta}}^2_{GLSC}$	$\hat{\sigma}_u^2$	7.221	7.149489	-0.99%	5.881007	8.343415	0.115630	4.76%	0.010753	7.127983	7.170996	0.120744
	$\hat{\sigma}_v^2$	4.981	4.925740	-1.11%	4.603064	5.318797	0.011688	2.19%	0.003419	4.918903	4.932578	0.014741
	$\hat{\gamma}$	0.000	-0.001159	-	-0.055659	0.078758	0.000402	-	0.000634	-0.002427	0.000109	0.000403
$\underline{\hat{\theta}}^3_{GLSC}$	$\hat{\sigma}_u^2$	7.221	7.221972	0.01%	5.876044	8.430054	0.116239	4.72%	0.010781	7.200410	7.243535	0.116240
	$\hat{\sigma}_v^2$	4.981	4.990018	0.18%	4.671968	5.363621	0.012046	2.20%	0.003471	4.983077	4.996960	0.012127
	$\hat{\gamma}$	0.000	-0.001126	-	-0.056961	0.080755	0.000398	-	0.000631	-0.002389	0.000136	0.000400
$\underline{\hat{\theta}}^4_{GLSC}$	$\hat{\sigma}_u^2$	7.221	7.144790	-1.06%	5.845829	8.285207	0.115885	4.76%	0.010765	7.123260	7.166320	0.121693
	$\hat{\sigma}_v^2$	4.981	4.921895	-1.19%	4.579014	5.325428	0.011945	2.22%	0.003456	4.914983	4.928808	0.015439
	$\hat{\gamma}$	0.000	-0.001222	-	-0.059107	0.078794	0.000410	-	0.000640	-0.002503	0.000058	0.000411
$\underline{\hat{\theta}}_{PML}$	$\hat{\sigma}_u^2$	7.221	7.197732	-0.32%	5.877930	8.380295	0.115283	4.72%	0.010737	7.176258	7.219206	0.115825
	$\hat{\sigma}_v^2$	4.981	4.968527	-0.25%	4.649510	5.347761	0.011753	2.18%	0.003428	4.961670	4.975383	0.011909
	$\hat{\gamma}$	0.000	-0.001142	-	-0.056550	0.080050	0.000397	-	0.000630	-0.002402	0.000118	0.000398

Table E.4b – Evaluation of $\underline{\hat{\theta}}$ for AR1 model, normally distributed errors (population, replications generated by UCM-C model) - further results, part b.

parameter		pop value	$\hat{E}(\hat{\theta})$	rel bias	min	max	$var(\hat{\theta})$	$cv(\hat{\theta})$	s.e [$\hat{E}(\hat{\theta})$]	conf. interval for $\hat{E}(\hat{\theta})$ lower bound	conf. interval for $\hat{E}(\hat{\theta})$ upper bound	Mse
$\underline{\hat{\theta}}_{ULS}$	$\hat{\sigma}_u^2$	6.744011	6.666327	-1.15%	5.370200	8.947343	0.224146	7.10%	0.014971	6.636384	6.696270	0.230181
	$\hat{\sigma}_v^2$	4.965116	4.960034	-0.10%	4.539989	5.919019	0.027281	3.33%	0.005223	4.949588	4.970480	0.027306
$\underline{\hat{\theta}}^2_{GLS}$	$\hat{\sigma}_u^2$	6.744011	6.628686	-1.71%	5.360541	8.878588	0.222458	7.12%	0.014915	6.598856	6.658517	0.235758
	$\hat{\sigma}_v^2$	4.965116	4.896950	-1.37%	4.514558	5.382400	0.024192	3.18%	0.004918	4.887113	4.906787	0.028838
$\underline{\hat{\theta}}^3_{GLS}$	$\hat{\sigma}_u^2$	6.744011	6.685308	-0.87%	5.374801	8.979263	0.225831	7.11%	0.015028	6.655253	6.715363	0.229277
	$\hat{\sigma}_v^2$	4.965116	4.992641	0.55%	4.547729	6.484145	0.031219	3.54%	0.005587	4.981466	5.003816	0.031976
$\underline{\hat{\theta}}^4_{GLS}$	$\hat{\sigma}_u^2$	6.744011	6.580430	-2.43%	5.387002	10.665942	0.243338	7.50%	0.015599	6.549232	6.611629	0.270097
	$\hat{\sigma}_v^2$	4.965116	4.813360	-3.06%	4.312328	5.306838	0.023154	3.16%	0.004812	4.803736	4.822983	0.046184
$\underline{\hat{\theta}}_{ML}$	$\hat{\sigma}_u^2$	6.744011	6.666327	-1.15%	5.370200	8.947343	0.224146	7.10%	0.014971	6.636384	6.696270	0.230181
	$\hat{\sigma}_v^2$	4.965116	4.960034	-0.10%	4.539989	5.919019	0.027281	3.33%	0.005223	4.949588	4.970480	0.027306
$\underline{\hat{\theta}}_{ULSC}$	$\hat{\sigma}_u^2$	6.744011	6.698981	-0.67%	5.186295	11.520522	0.286031	7.98%	0.016912	6.665156	6.732806	0.288059
	$\hat{\sigma}_v^2$	4.965116	4.940688	-0.49%	4.447631	6.317477	0.030272	3.52%	0.005502	4.929684	4.951692	0.030869
$\underline{\hat{\theta}}^2_{GLSC}$	$\hat{\sigma}_u^2$	6.744011	6.655871	-1.31%	5.162969	11.419777	0.282674	7.99%	0.016813	6.622245	6.689497	0.290443
	$\hat{\sigma}_v^2$	4.965116	4.870154	-1.91%	4.421098	5.460875	0.026121	3.32%	0.005111	4.859932	4.880376	0.035139
$\underline{\hat{\theta}}^3_{GLSC}$	$\hat{\sigma}_u^2$	6.744011	6.720677	-0.35%	5.196864	11.566581	0.288915	8.00%	0.016998	6.686682	6.754672	0.289460
	$\hat{\sigma}_v^2$	4.965116	4.977467	0.25%	4.461252	7.565965	0.037650	3.90%	0.006136	4.965195	4.989738	0.037802
$\underline{\hat{\theta}}^4_{GLSC}$	$\hat{\sigma}_u^2$	6.744011	6.592567	-2.25%	5.182134	13.148404	0.298031	8.28%	0.017264	6.558040	6.627095	0.320966
	$\hat{\sigma}_v^2$	4.965116	4.776220	-3.80%	4.244927	5.284544	0.025569	3.35%	0.005057	4.766107	4.786334	0.061251
$\underline{\hat{\theta}}_{PML}$	$\hat{\sigma}_u^2$	6.744011	6.698981	-0.67%	5.186295	11.520522	0.286031	7.98%	0.016912	6.665156	6.732806	0.288059
	$\hat{\sigma}_v^2$	4.965116	4.940688	-0.49%	4.447631	6.317477	0.030272	3.52%	0.005502	4.929684	4.951692	0.030869

Table E.5 – Evaluation of $\underline{\hat{\theta}}$ for UCM model, $t_{\upsilon=5}(0,1)$ distributed errors (population, replications generated by UCM model) - further results.

parameter		pop value	$\hat{E}(\hat{\theta})$	rel bias	min	max	$\mathrm{var}(\hat{\theta})$	$cv(\hat{\theta})$	s.e $[\hat{E}(\hat{\theta})]$	conf. interval for $\hat{E}(\hat{\theta})$ lower bound	upper bound	mse
$\underline{\hat{\theta}}_{ULS}$	$\hat{\sigma}_u^2$	6.744011	6.666937	-1.14%	5.340912	8.970265	0.226732	7.14%	0.015058	6.636821	6.697052	0.232672
	$\hat{\sigma}_v^2$	4.965116	4.959424	-0.11%	4.486214	5.917083	0.029244	3.45%	0.005408	4.948608	4.970239	0.029276
	$\hat{\gamma}$	0.000000	-0.000891	-	-0.068276	0.080259	0.000509	-	0.000714	-0.002319	0.000536	0.000510
$\underline{\hat{\theta}}^2_{GLS}$	$\hat{\sigma}_u^2$	6.744011	6.625544	-1.76%	5.334642	8.871264	0.223716	7.14%	0.014957	6.595630	6.655458	0.237751
	$\hat{\sigma}_v^2$	4.965116	4.903214	-1.25%	4.475744	5.425369	0.025611	3.26%	0.005061	4.893093	4.913336	0.029443
	$\hat{\gamma}$	0.000000	0.000838	-	-0.062923	0.069544	0.000359	-	0.000599	-0.000361	0.002037	0.000360
$\underline{\hat{\theta}}^3_{GLS}$	$\hat{\sigma}_u^2$	6.744011	6.683550	-0.90%	5.342382	8.964003	0.226730	7.12%	0.015058	6.653435	6.713665	0.230385
	$\hat{\sigma}_v^2$	4.965116	4.993010	0.56%	4.514113	6.669555	0.033367	3.66%	0.005776	4.981457	5.004562	0.034145
	$\hat{\gamma}$	0.000000	0.000720	-	-0.061434	0.073544	0.000367	-	0.000606	-0.000491	0.001931	0.000367
$\underline{\hat{\theta}}^4_{GLS}$	$\hat{\sigma}_u^2$	6.744011	6.577232	-2.47%	5.380775	10.456871	0.243209	7.50%	0.015595	6.546041	6.608422	0.271024
	$\hat{\sigma}_v^2$	4.965116	4.823245	-2.86%	4.272679	5.296753	0.024465	3.24%	0.004946	4.813353	4.833137	0.044592
	$\hat{\gamma}$	0.000000	0.000635	-	-0.062524	0.058065	0.000367	-	0.000606	-0.000576	0.001847	0.000367
$\underline{\hat{\theta}}_{ML}$	$\hat{\sigma}_u^2$	6.744011	6.664119	-1.18%	5.339409	8.934012	0.225171	7.12%	0.015006	6.634108	6.694131	0.231553
	$\hat{\sigma}_v^2$	4.965116	4.962346	-0.06%	4.507670	6.018046	0.028976	3.43%	0.005383	4.951580	4.973112	0.028984
	$\hat{\gamma}$	0.000000	0.000751	-	-0.061883	0.072220	0.000358	-	0.000598	-0.000445	0.001947	0.000358

Table E.6a – Evaluation of $\underline{\hat{\theta}}$ for AR1 model, $t_{\upsilon=5}(0,1)$ distributed errors (population, replications generated by UCM model) - further results, part a.

parameter		pop value	$\hat{E}(\hat{\theta})$	rel bias	min	max	$\text{var}(\hat{\theta})$	$cv(\hat{\theta})$	s.e $[\hat{E}(\hat{\theta})]$	conf. interval for $\hat{E}(\hat{\theta})$ lower bound	upper bound	mse
$\underline{\hat{\theta}}_{ULSC}$	$\hat{\sigma}_u^2$	6.744011	6.700974	-0.64%	5.146865	11.530258	0.288953	8.02%	0.016999	6.666977	6.734971	0.290805
	$\hat{\sigma}_v^2$	4.965116	4.938693	-0.53%	4.375454	6.319095	0.032839	3.67%	0.005731	4.927232	4.950155	0.033537
	$\hat{\gamma}$	0.000000	-0.001696	-	-0.068527	0.097586	0.000581	-	0.000762	-0.003220	-0.000171	0.000584
$\underline{\hat{\theta}}^2_{GLSC}$	$\hat{\sigma}_u^2$	6.744011	6.654936	-1.32%	5.137076	11.410044	0.284215	8.01%	0.016859	6.621219	6.688653	0.292149
	$\hat{\sigma}_v^2$	4.965116	4.874646	-1.82%	4.380560	5.445977	0.027952	3.43%	0.005287	4.864072	4.885220	0.036137
	$\hat{\gamma}$	0.000000	-0.000405	-	-0.067243	0.087184	0.000412	-	0.000642	-0.001689	0.000879	0.000412
$\underline{\hat{\theta}}^3_{GLSC}$	$\hat{\sigma}_u^2$	6.744011	6.720960	-0.34%	5.159738	11.55016	0.289465	8.01%	0.017014	6.686932	6.754987	0.289996
	$\hat{\sigma}_v^2$	4.965116	4.975695	0.21%	4.410660	8.047166	0.042261	4.13%	0.006501	4.962694	4.988697	0.042373
	$\hat{\gamma}$	0.000000	-0.000364	-	-0.066796	0.147210	0.000429	-	0.000655	-0.001675	0.000946	0.000430
$\underline{\hat{\theta}}^4_{GLSC}$	$\hat{\sigma}_u^2$	6.744011	6.590565	-2.28%	5.176667	12.682419	0.293563	8.22%	0.017134	6.556298	6.624833	0.317109
	$\hat{\sigma}_v^2$	4.965116	4.785079	-3.63%	4.220377	5.317978	0.027088	3.44%	0.005205	4.774669	4.795488	0.059502
	$\hat{\gamma}$	0.000000	-0.000537	-	-0.065458	0.070371	0.000406	-	0.000637	-0.001811	0.000738	0.000407
$\underline{\hat{\theta}}_{PML}$	$\hat{\sigma}_u^2$	6.744011	6.693142	-0.75%	0.404484	11.505795	0.326371	8.54%	0.018066	6.657011	6.729274	0.328958
	$\hat{\sigma}_v^2$	4.965116	4.942368	-0.46%	4.400935	6.530737	0.035157	3.79%	0.005929	4.930509	4.954226	0.035675
	$\hat{\gamma}$	0.000000	0.000594	-	-0.066930	1.000175	0.001411	-	0.001188	-0.001782	0.002969	0.001411

Table E.6b – Evaluation of $\underline{\hat{\theta}}$ for AR1 model, $t_{\upsilon=5}(0,1)$ distributed errors (population, replications generated by UCM model) - further results, part b.

parameter		pop value	$\hat{E}(\hat{\theta})$	rel bias	min	max	$var(\hat{\theta})$	$cv(\hat{\theta})$	s.e [$\hat{E}(\hat{\theta})$]	conf. interval for $\hat{E}(\hat{\theta})$		mse
										lower bound	upper bound	
$\underline{\hat{\theta}}_{ULS}$	$\hat{\sigma}_u^2$	7.221	7.144893	-1.05%	5.557217	10.178167	0.289076	7.53%	0.017002	7.110888	7.178897	0.294868
	$\hat{\sigma}_v^2$	4.981	4.994016	0.26%	4.384034	5.784704	0.030014	3.47%	0.005478	4.983059	5.004973	0.030183
$\underline{\hat{\theta}}^2_{GLS}$	$\hat{\sigma}_u^2$	7.221	7.103215	-1.63%	5.515122	10.061680	0.287118	7.54%	0.016945	7.069326	7.137104	0.300991
	$\hat{\sigma}_v^2$	4.981	4.929945	-1.02%	4.337172	5.462889	0.026186	3.28%	0.005117	4.919711	4.940180	0.028792
$\underline{\hat{\theta}}^3_{GLS}$	$\hat{\sigma}_u^2$	7.221	7.165865	-0.76%	5.581923	10.244999	0.291055	7.53%	0.017060	7.131744	7.199986	0.294094
	$\hat{\sigma}_v^2$	4.981	5.026982	0.92%	4.406344	6.077280	0.033894	3.66%	0.005822	5.015338	5.038626	0.036009
$\underline{\hat{\theta}}^4_{GLS}$	$\hat{\sigma}_u^2$	7.221	7.035859	-2.56%	5.471824	9.302476	0.273670	7.44%	0.016543	7.002773	7.068945	0.307947
	$\hat{\sigma}_v^2$	4.981	4.844238	-2.75%	4.315569	5.378118	0.024238	3.21%	0.004923	4.834392	4.854085	0.042941
$\underline{\hat{\theta}}_{ML}$	$\hat{\sigma}_u^2$	7.221	7.144893	-1.05%	5.557217	10.178167	0.289076	7.53%	0.017002	7.110888	7.178897	0.294868
	$\hat{\sigma}_v^2$	4.981	4.994016	0.26%	4.384034	5.784704	0.030014	3.47%	0.005478	4.983059	5.004973	0.030183
$\underline{\hat{\theta}}_{ULSC}$	$\hat{\sigma}_u^2$	7.221	7.185028	-0.50%	5.742142	10.604283	0.332494	8.03%	0.018234	7.148559	7.221497	0.333788
	$\hat{\sigma}_v^2$	4.981	4.979864	-0.02%	4.398174	6.019979	0.034591	3.73%	0.005881	4.968101	4.991626	0.034592
$\underline{\hat{\theta}}^2_{GLSC}$	$\hat{\sigma}_u^2$	7.221	7.136706	-1.17%	5.684590	10.483784	0.329421	8.04%	0.018150	7.100406	7.173006	0.336526
	$\hat{\sigma}_v^2$	4.981	4.906872	-1.49%	4.353484	5.568116	0.028504	3.44%	0.005339	4.896194	4.917550	0.033999
$\underline{\hat{\theta}}^3_{GLSC}$	$\hat{\sigma}_u^2$	7.221	7.209400	-0.16%	5.774508	10.676616	0.335455	8.03%	0.018315	7.172769	7.246031	0.335589
	$\hat{\sigma}_v^2$	4.981	5.017985	0.74%	4.419076	6.498920	0.041811	4.07%	0.006466	5.005053	5.030917	0.043179
$\underline{\hat{\theta}}^4_{GLSC}$	$\hat{\sigma}_u^2$	7.221	7.057246	-2.27%	5.630109	9.887106	0.312431	7.92%	0.017676	7.021894	7.092597	0.339247
	$\hat{\sigma}_v^2$	4.981	4.808065	-3.47%	4.317640	5.366489	0.026395	3.38%	0.005138	4.797790	4.818340	0.056301
$\underline{\hat{\theta}}_{PML}$	$\hat{\sigma}_u^2$	7.221	7.185028	-0.50%	5.742142	10.604283	0.332494	8.03%	0.018234	7.148559	7.221497	0.333788
	$\hat{\sigma}_v^2$	4.981	4.979864	-0.02%	4.398174	6.019979	0.034591	3.73%	0.005881	4.968101	4.991626	0.034592

Table E.7 – Evaluation of $\underline{\hat{\theta}}$ for UCM model, $t_{\upsilon=5}(0,1)$ distributed errors (population, replications generated by UCM-C model) - further results.

parameter		pop value	$\hat{E}(\hat{\theta})$	rel bias	min	max	$var(\hat{\theta})$	$cv(\hat{\theta})$	s.e $[\hat{E}(\hat{\theta})]$	conf. interval for $\hat{E}(\hat{\theta})$ lower bound	conf. interval for $\hat{E}(\hat{\theta})$ upper bound	mse
$\underline{\hat{\theta}}_{ULS}$	$\hat{\sigma}_u^2$	7.221	7.141699	-1.10%	5.621463	10.199508	0.294038	7.59%	0.017148	7.107404	7.175994	0.300327
	$\hat{\sigma}_v^2$	4.981	4.997204	0.33%	4.372269	5.798344	0.031988	3.58%	0.005656	4.985892	5.008515	0.032251
	$\hat{\gamma}$	0.000	0.000981	-	-0.073462	0.075796	0.000543	-	0.000737	-0.000493	0.002455	0.000544
$\underline{\hat{\theta}}^2_{GLS}$	$\hat{\sigma}_u^2$	7.221	7.100190	-1.67%	5.537889	10.021532	0.289811	7.58%	0.017024	7.066142	7.134237	0.304406
	$\hat{\sigma}_v^2$	4.981	4.936322	-0.90%	4.359079	5.457860	0.027211	3.34%	0.005216	4.925889	4.946754	0.029207
	$\hat{\gamma}$	0.000	0.000738	-	-0.065002	0.063380	0.000384	-	0.000620	-0.000502	0.001978	0.000385
$\underline{\hat{\theta}}^3_{GLS}$	$\hat{\sigma}_u^2$	7.221	7.164536	-0.78%	5.603583	10.217272	0.294592	7.58%	0.017164	7.130209	7.198863	0.297780
	$\hat{\sigma}_v^2$	4.981	5.026929	0.92%	4.427374	6.034731	0.034971	3.72%	0.005914	5.015101	5.038756	0.037081
	$\hat{\gamma}$	0.000	0.000601	-	-0.065850	0.065446	0.000378	-	0.000615	-0.000628	0.001831	0.000378
$\underline{\hat{\theta}}^4_{GLS}$	$\hat{\sigma}_u^2$	7.221	7.032869	-2.61%	5.496093	9.297604	0.276017	7.47%	0.016614	6.999642	7.066097	0.311410
	$\hat{\sigma}_v^2$	4.981	4.854272	-2.54%	4.310482	5.396218	0.025314	3.28%	0.005031	4.844209	4.864334	0.041374
	$\hat{\gamma}$	0.000	0.000851	-	-0.064921	0.059076	0.000366	-	0.000605	-0.000358	0.002061	0.000367
$\underline{\hat{\theta}}_{ML}$	$\hat{\sigma}_u^2$	7.221	7.142958	-1.08%	5.579357	10.146751	0.292319	7.57%	0.017097	7.108764	7.177153	0.298410
	$\hat{\sigma}_v^2$	4.981	4.996088	0.30%	4.405379	5.760269	0.031105	3.53%	0.005577	4.984933	5.007242	0.031332
	$\hat{\gamma}$	0.000	0.000641	-	-0.065593	0.064833	0.000375	-	0.000612	-0.000584	0.001865	0.000375

Table E.8a – Evaluation of $\underline{\hat{\theta}}$ for AR1 model, $t_{\upsilon=5}(0,1)$ distributed errors (population, replications generated by UCM-C model) - further results, part a.

parameter		pop value	$\hat{E}(\hat{\theta})$	rel bias	min	max	$\mathrm{var}(\hat{\theta})$	$cv(\hat{\theta})$	s.e [$\hat{E}(\hat{\theta})$]	conf. interval for $\hat{E}(\hat{\theta})$		mse
										lower bound	upper bound	
$\hat{\underline{\theta}}_{ULSC}$	$\hat{\sigma}_u^2$	7.221	7.185248	-0.50%	5.796745	10.639245	0.337581	8.09%	0.018373	7.148501	7.221994	0.338859
	$\hat{\sigma}_v^2$	4.981	4.979642	-0.03%	4.383212	6.027588	0.036241	3.82%	0.006020	4.967602	4.991682	0.036242
	$\hat{\gamma}$	0.000	-0.00076	-	-0.085915	0.072488	0.000593	-	0.000770	-0.002301	0.00078	0.000594
$\hat{\underline{\theta}}^2_{GLSC}$	$\hat{\sigma}_u^2$	7.221	7.136812	-1.17%	5.698179	10.446093	0.332413	8.08%	0.018232	7.100348	7.173277	0.339501
	$\hat{\sigma}_v^2$	4.981	4.910334	-1.42%	4.374390	5.544706	0.029595	3.50%	0.005440	4.899453	4.921214	0.034589
	$\hat{\gamma}$	0.000	-0.000913	-	-0.071793	0.065284	0.000411	-	0.000641	-0.002194	0.000369	0.000411
$\hat{\underline{\theta}}^3_{GLSC}$	$\hat{\sigma}_u^2$	7.221	7.211366	-0.13%	5.787531	10.653409	0.339475	8.08%	0.018425	7.174516	7.248216	0.339568
	$\hat{\sigma}_v^2$	4.981	5.014580	0.67%	4.437965	6.436482	0.043011	4.14%	0.006558	5.001464	5.027697	0.044139
	$\hat{\gamma}$	0.000	-0.001014	-	-0.072722	0.072522	0.000413	-	0.000642	-0.002299	0.000271	0.000414
$\hat{\underline{\theta}}^4_{GLSC}$	$\hat{\sigma}_u^2$	7.221	7.057676	-2.26%	5.646515	9.905772	0.315008	7.95%	0.017748	7.022179	7.093173	0.341682
	$\hat{\sigma}_v^2$	4.981	4.815447	-3.32%	4.309541	5.373717	0.027387	3.44%	0.005233	4.804981	4.825914	0.054795
	$\hat{\gamma}$	0.000	-0.000804	-	-0.070577	0.064559	0.000395	-	0.000629	-0.002061	0.000454	0.000396
$\hat{\underline{\theta}}_{PML}$	$\hat{\sigma}_u^2$	7.221	7.186326	-0.48%	5.755407	10.576838	0.336155	8.07%	0.018335	7.149657	7.222995	0.337358
	$\hat{\sigma}_v^2$	4.981	4.978736	-0.05%	4.417755	5.980887	0.035799	3.80%	0.005983	4.96677	4.990703	0.035804
	$\hat{\gamma}$	0.000	-0.000994	-	-0.072398	0.070146	0.000405	-	0.000636	-0.002266	0.000279	0.000406

Table E.8b – Evaluation of $\hat{\underline{\theta}}$ for AR1 model, $t_{\upsilon=5}(0,1)$ distributed errors (population, replications generated by UCM-C model) - further results, part b.

Appendix F

More detailed results obtained in simulation study II

Tables F.1 to F.12 below provide more detailed information about the results obtained in simulation study II, when considering $\sigma_\eta^{2\ sim,C} \cong 0.75$.

		pop	$\hat{E}(\text{var}(\hat{\underline{\theta}}))$	rel bias	$\text{var}(\text{var}(\hat{\underline{\theta}}))$	$cv(\text{var}(\hat{\underline{\theta}}))$	$s.e.[\hat{E}(\text{var}(\hat{\underline{\theta}}))]$	mse
$\text{var}_n(\hat{\underline{\theta}}_{ML})$	$\text{var}(\hat{\sigma}_u^2)$	0.148558	0.119016	-19.89%	1.92×10^{-4}	11.65%	1.39×10^{-4}	1.07×10^{-3}
	$\text{var}(\hat{\sigma}_v^2)$	0.010553	0.010072	-4.56%	5.88×10^{-7}	7.61%	7.67×10^{-6}	8.19×10^{-7}
$\text{var}_n(\hat{\underline{\theta}}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	0.148536	0.117877	-20.64%	1.29×10^{-4}	9.65%	1.14×10^{-4}	1.07×10^{-3}
	$\text{var}(\hat{\sigma}_v^2)$	0.010606	0.009624	-9.26%	3.15×10^{-7}	5.83%	5.61×10^{-6}	1.28×10^{-6}
$\text{var}_{adf}(\hat{\underline{\theta}}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	0.148536	0.117892	-20.63%	1.90×10^{-4}	11.69%	1.38×10^{-4}	1.13×10^{-3}
	$\text{var}(\hat{\sigma}_v^2)$	0.010606	0.009780	-7.79%	5.59×10^{-7}	7.64%	7.48×10^{-6}	1.24×10^{-6}
$\text{var}_c(\hat{\underline{\theta}}_{ML})$	$\text{var}(\hat{\sigma}_u^2)$	0.148558	0.141689	-4.62%	1.69×10^{-3}	29.04%	4.11×10^{-4}	1.74×10^{-3}
	$\text{var}(\hat{\sigma}_v^2)$	0.010553	0.010148	-3.84%	5.47×10^{-6}	23.04%	2.34×10^{-5}	5.63×10^{-6}
$\text{var}_c(\hat{\underline{\theta}}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	0.148536	0.140274	-5.56%	1.67×10^{-3}	29.09%	4.08×10^{-4}	1.73×10^{-3}
	$\text{var}(\hat{\sigma}_v^2)$	0.010606	0.009845	-7.17%	5.28×10^{-6}	23.35%	2.30×10^{-5}	5.86×10^{-6}

Table F.1a – Evaluation of $\text{var}(\hat{\underline{\theta}})$ considering $m^{sim} = 47$ and $n_j^{sim} = n_j^{sim*}$ (part a).

		min	max	95% ci for pop		95% ci for $\hat{E}(\text{var}(\hat{\underline{\theta}}))$		95% ci for bias	
				lower bound	upper bound	lower bound	upper bound	lower bound	upper bound
$\text{var}_n(\hat{\underline{\theta}}_{ML})$	$\text{var}(\hat{\sigma}_u^2)$	0.0764	0.1968	0.144337	0.152779	0.118744	0.119288	-0.033726	-0.025329
	$\text{var}(\hat{\sigma}_v^2)$	0.0074	0.0138	0.010262	0.010844	0.010057	0.010087	-0.000770	-0.000189
$\text{var}_n(\hat{\underline{\theta}}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	0.0828	0.1696	0.144327	0.152745	0.117654	0.118100	-0.034832	-0.026457
	$\text{var}(\hat{\sigma}_v^2)$	0.0075	0.0120	0.010314	0.010898	0.009613	0.009635	-0.001272	-0.000689
$\text{var}_{adf}(\hat{\underline{\theta}}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	0.0761	0.1891	0.144327	0.152745	0.117622	0.118162	-0.034817	-0.026442
	$\text{var}(\hat{\sigma}_v^2)$	0.0073	0.0135	0.010314	0.010898	0.009765	0.009794	-0.001116	-0.000533
$\text{var}_c(\hat{\underline{\theta}}_{ML})$	$\text{var}(\hat{\sigma}_u^2)$	0.0442	0.4654	0.144337	0.152779	0.140883	0.142496	-0.011067	-0.002642
	$\text{var}(\hat{\sigma}_v^2)$	0.0037	0.0210	0.010262	0.010844	0.010102	0.010194	-0.000697	-0.000110
$\text{var}_c(\hat{\underline{\theta}}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	0.0465	0.4760	0.144327	0.152745	0.139474	0.141074	-0.012452	-0.004043
	$\text{var}(\hat{\sigma}_v^2)$	0.0039	0.0198	0.010314	0.010898	0.009800	0.009890	-0.001054	-0.000465

Table F.1b – Evaluation of $\text{var}(\hat{\underline{\theta}})$ considering $m^{sim} = 47$ and $n_j^{sim} = n_j^{sim*}$ (part b).

		pop	$\hat{E}(\text{var}(\hat{\underline{\theta}}))$	rel bias	$\text{var}(\text{var}(\hat{\underline{\theta}}))$	$cv(\text{var}(\hat{\underline{\theta}}))$	$s.e.[\hat{E}(\text{var}(\hat{\underline{\theta}}))]$	mse
$\text{var}_n(\hat{\underline{\theta}}_{ML})$	$\text{var}(\hat{\sigma}_u^2)$	0.253266	0.224323	-11.43%	1.06×10^{-3}	14.54%	3.26×10^{-4}	1.90×10^{-3}
	$\text{var}(\hat{\sigma}_v^2)$	0.019143	0.019085	-0.30%	2.69×10^{-6}	8.59%	1.64×10^{-5}	2.69×10^{-6}
$\text{var}_n(\hat{\underline{\theta}}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	0.253008	0.221465	-12.47%	6.34×10^{-4}	11.37%	2.52×10^{-4}	1.63×10^{-3}
	$\text{var}(\hat{\sigma}_v^2)$	0.019455	0.018066	-7.14%	9.77×10^{-7}	5.47%	9.88×10^{-6}	2.91×10^{-6}
$\text{var}_{adj}(\hat{\underline{\theta}}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	0.253008	0.221131	-12.60%	1.04×10^{-3}	14.61%	3.23×10^{-4}	2.06×10^{-3}
	$\text{var}(\hat{\sigma}_v^2)$	0.019455	0.018285	-6.01%	2.53×10^{-6}	8.69%	1.59×10^{-5}	3.90×10^{-6}
$\text{var}_c(\hat{\underline{\theta}}_{ML})$	$\text{var}(\hat{\sigma}_u^2)$	0.253266	0.244278	-3.55%	4.30×10^{-3}	26.85%	6.56×10^{-4}	4.38×10^{-3}
	$\text{var}(\hat{\sigma}_v^2)$	0.019143	0.019333	0.99%	1.80×10^{-5}	21.98%	4.25×10^{-5}	1.81×10^{-5}
$\text{var}_c(\hat{\underline{\theta}}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	0.253008	0.240589	-4.91%	4.22×10^{-3}	27.00%	6.50×10^{-4}	4.37×10^{-3}
	$\text{var}(\hat{\sigma}_v^2)$	0.019455	0.018521	-4.80%	1.73×10^{-5}	22.43%	4.15×10^{-5}	1.81×10^{-5}

Table F.2a – Evaluation of $\text{var}(\hat{\underline{\theta}})$ considering $m^{sim} = 47$ and $n_j^{sim} = 15$ (part a).

		min	max	95% ci for pop lower bound	95% ci for pop upper bound	95% ci for $\hat{E}(\text{var}(\hat{\underline{\theta}}))$ lower bound	95% ci for $\hat{E}(\text{var}(\hat{\underline{\theta}}))$ upper bound	95% ci for bias lower bound	95% ci for bias upper bound
$\text{var}_n(\hat{\underline{\theta}}_{ML})$	$\text{var}(\hat{\sigma}_u^2)$	0.1282	0.3820	0.246268	0.260264	0.223683	0.224962	-0.035884	-0.021952
	$\text{var}(\hat{\sigma}_v^2)$	0.0140	0.0264	0.018610	0.019676	0.019053	0.019117	-0.000588	0.000476
$\text{var}_n(\hat{\underline{\theta}}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	0.1424	0.3326	0.245983	0.260033	0.220971	0.221958	-0.038501	-0.024535
	$\text{var}(\hat{\sigma}_v^2)$	0.0145	0.0217	0.018917	0.019993	0.018047	0.018085	-0.001923	-0.000850
$\text{var}_{adj}(\hat{\underline{\theta}}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	0.1216	0.3868	0.245983	0.260033	0.220498	0.221765	-0.038844	-0.024859
	$\text{var}(\hat{\sigma}_v^2)$	0.0133	0.0254	0.018917	0.019993	0.018254	0.018316	-0.001705	-0.000631
$\text{var}_c(\hat{\underline{\theta}}_{ML})$	$\text{var}(\hat{\sigma}_u^2)$	0.0834	0.6438	0.246268	0.260264	0.242993	0.245564	-0.015992	-0.001933
	$\text{var}(\hat{\sigma}_v^2)$	0.0075	0.0389	0.018610	0.019676	0.019250	0.019416	-0.000346	0.000729
$\text{var}_c(\hat{\underline{\theta}}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	0.0831	0.6454	0.245983	0.260033	0.239316	0.241862	-0.019452	-0.005336
	$\text{var}(\hat{\sigma}_v^2)$	0.0068	0.0393	0.018917	0.019993	0.018439	0.018602	-0.001474	-0.000390

Table F.2b – Evaluation of $\text{var}(\hat{\underline{\theta}})$ considering $m^{sim} = 47$ and $n_j^{sim} = 15$ (part b).

		pop	$\hat{E}(\text{var}(\hat{\underline{\theta}}))$	rel bias	$\text{var}(\text{var}(\hat{\underline{\theta}}))$	$cv(\text{var}(\hat{\underline{\theta}}))$	$s.e.[\hat{E}(\text{var}(\hat{\underline{\theta}}))]$	mse
$\text{var}_n(\hat{\underline{\theta}}_{ML})$	$\text{var}(\hat{\sigma}_u^2)$	0.372950	0.335870	-9.94%	3.60×10^{-3}	17.86%	6.00×10^{-4}	4.98×10^{-3}
	$\text{var}(\hat{\sigma}_v^2)$	0.028285	0.028519	0.83%	8.90×10^{-6}	10.46%	2.98×10^{-5}	8.96×10^{-6}
$\text{var}_n(\hat{\underline{\theta}}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	0.368915	0.330089	-10.52%	2.08×10^{-3}	13.82%	4.56×10^{-4}	3.59×10^{-3}
	$\text{var}(\hat{\sigma}_v^2)$	0.029370	0.026728	-9.00%	3.23×10^{-6}	6.73%	1.80×10^{-5}	1.02×10^{-5}
$\text{var}_{adj}(\hat{\underline{\theta}}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	0.368915	0.329514	-10.68%	3.50×10^{-3}	17.97%	5.92×10^{-4}	5.06×10^{-3}
	$\text{var}(\hat{\sigma}_v^2)$	0.029370	0.026951	-8.24%	8.08×10^{-6}	10.55%	2.84×10^{-5}	1.39×10^{-5}
$\text{var}_c(\hat{\underline{\theta}}_{ML})$	$\text{var}(\hat{\sigma}_u^2)$	0.372950	0.353928	-5.10%	9.91×10^{-3}	28.13%	9.96×10^{-4}	1.03×10^{-2}
	$\text{var}(\hat{\sigma}_v^2)$	0.028285	0.028569	1.01%	4.11×10^{-5}	22.45%	6.41×10^{-5}	4.12×10^{-5}
$\text{var}_c(\hat{\underline{\theta}}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	0.368915	0.346859	-5.98%	9.60×10^{-3}	28.24%	9.80×10^{-4}	1.01×10^{-2}
	$\text{var}(\hat{\sigma}_v^2)$	0.029370	0.026971	-8.17%	3.78×10^{-5}	22.81%	6.15×10^{-5}	4.36×10^{-5}

Table F.3a – Evaluation of $\text{var}(\hat{\underline{\theta}})$ considering $m^{sim} = 47$ and $n_j^{sim} = 10$ (part a).

		min	max	95% ci for pop lower bound	95% ci for pop upper bound	95% ci for $\hat{E}(\text{var}(\hat{\underline{\theta}}))$ lower bound	95% ci for $\hat{E}(\text{var}(\hat{\underline{\theta}}))$ upper bound	95% ci for bias lower bound	95% ci for bias upper bound
$\text{var}_n(\hat{\underline{\theta}}_{ML})$	$\text{var}(\hat{\sigma}_u^2)$	0.1511	0.6885	0.362386	0.383514	0.334694	0.337046	-0.047513	-0.026571
	$\text{var}(\hat{\sigma}_v^2)$	0.0198	0.0427	0.027492	0.029078	0.028460	0.028577	-0.000555	0.001029
$\text{var}_n(\hat{\underline{\theta}}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	0.1986	0.5451	0.358512	0.379318	0.329194	0.330983	-0.049084	-0.028496
	$\text{var}(\hat{\sigma}_v^2)$	0.0212	0.0347	0.028540	0.030200	0.026692	0.026763	-0.003468	-0.001811
$\text{var}_{adf}(\hat{\underline{\theta}}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	0.1482	0.7059	0.358512	0.379318	0.328354	0.330674	-0.049684	-0.029045
	$\text{var}(\hat{\sigma}_v^2)$	0.0178	0.0401	0.028540	0.030200	0.026895	0.027007	-0.003246	-0.001586
$\text{var}_c(\hat{\underline{\theta}}_{ML})$	$\text{var}(\hat{\sigma}_u^2)$	0.1170	0.9991	0.362386	0.383514	0.351976	0.355879	-0.029511	-0.008458
	$\text{var}(\hat{\sigma}_v^2)$	0.0108	0.0673	0.027492	0.029078	0.028444	0.028695	-0.000514	0.001088
$\text{var}_c(\hat{\underline{\theta}}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	0.1164	1.0119	0.358512	0.379318	0.344939	0.348779	-0.032397	-0.011643
	$\text{var}(\hat{\sigma}_v^2)$	0.0088	0.0590	0.028540	0.030200	0.026851	0.027092	-0.003233	-0.001558

Table F.3b – Evaluation of $\text{var}(\hat{\underline{\theta}})$ considering $m^{sim} = 47$ and $n_j^{sim} = 10$ (part b).

		pop	$\hat{E}(\text{var}(\hat{\underline{\theta}}))$	rel bias	$\text{var}(\text{var}(\hat{\underline{\theta}}))$	$cv(\text{var}(\hat{\underline{\theta}}))$	$s.e.[\hat{E}(\text{var}(\hat{\underline{\theta}}))]$	mse
$\text{var}_n(\hat{\underline{\theta}}_{ML})$	$\text{var}(\hat{\sigma}_u^2)$	0.722939	0.659269	-8.81%	2.75×10^{-2}	25.13%	1.66×10^{-3}	3.15×10^{-2}
	$\text{var}(\hat{\sigma}_v^2)$	0.057222	0.056528	-1.21%	7.17×10^{-5}	14.98%	8.47×10^{-5}	7.22×10^{-5}
$\text{var}_n(\hat{\underline{\theta}}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	0.715733	0.640339	-10.53%	1.57×10^{-2}	19.54%	1.25×10^{-3}	2.13×10^{-2}
	$\text{var}(\hat{\sigma}_v^2)$	0.061321	0.051435	-16.12%	2.55×10^{-5}	9.82%	5.05×10^{-5}	1.23×10^{-4}
$\text{var}_{adf}(\hat{\underline{\theta}}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	0.715733	0.635932	-11.15%	2.59×10^{-2}	25.30%	1.61×10^{-3}	3.22×10^{-2}
	$\text{var}(\hat{\sigma}_v^2)$	0.061321	0.051285	-16.37%	6.18×10^{-5}	15.32%	7.86×10^{-5}	1.62×10^{-4}
$\text{var}_c(\hat{\underline{\theta}}_{ML})$	$\text{var}(\hat{\sigma}_u^2)$	0.722939	0.675565	-6.55%	4.75×10^{-2}	32.25%	2.18×10^{-3}	4.97×10^{-2}
	$\text{var}(\hat{\sigma}_v^2)$	0.057222	0.056543	-1.19%	1.86×10^{-4}	24.09%	1.36×10^{-4}	1.86×10^{-4}
$\text{var}_c(\hat{\underline{\theta}}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	0.715733	0.650831	-9.07%	4.45×10^{-2}	32.40%	2.11×10^{-3}	4.87×10^{-2}
	$\text{var}(\hat{\sigma}_v^2)$	0.061321	0.051313	-16.32%	1.57×10^{-4}	24.40%	1.25×10^{-4}	2.57×10^{-4}

Table F.4a - Evaluation of $\text{var}(\hat{\underline{\theta}})$ considering $m^{sim} = 47$ and $n_j^{sim} = 5$ (part a).

		min	max	95% ci for pop lower bound	95% ci for pop upper bound	95% ci for $\hat{E}(\text{var}(\hat{\underline{\theta}}))$ lower bound	95% ci for $\hat{E}(\text{var}(\hat{\underline{\theta}}))$ upper bound	95% ci for bias lower bound	95% ci for bias upper bound
$\text{var}_n(\hat{\underline{\theta}}_{ML})$	$\text{var}(\hat{\sigma}_u^2)$	0.2620	1.8880	0.702144	0.743734	0.656021	0.662517	-0.084126	-0.043070
	$\text{var}(\hat{\sigma}_v^2)$	0.0325	0.1022	0.055606	0.058838	0.056362	0.056694	-0.002298	0.000920
$\text{var}_n(\hat{\underline{\theta}}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	0.2648	1.3164	0.695157	0.736309	0.637887	0.642792	-0.095502	-0.055141
	$\text{var}(\hat{\sigma}_v^2)$	0.0353	0.0762	0.059614	0.063028	0.051336	0.051534	-0.011578	-0.008183
$\text{var}_{adf}(\hat{\underline{\theta}}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	0.2457	1.7180	0.695157	0.736309	0.632779	0.639085	-0.100014	-0.059445
	$\text{var}(\hat{\sigma}_v^2)$	0.0293	0.1019	0.059614	0.063028	0.051131	0.051439	-0.011733	-0.008329
$\text{var}_c(\hat{\underline{\theta}}_{ML})$	$\text{var}(\hat{\sigma}_u^2)$	0.2019	2.5506	0.702144	0.743734	0.671295	0.679836	-0.067949	-0.026654
	$\text{var}(\hat{\sigma}_v^2)$	0.0185	0.1340	0.055606	0.058838	0.056276	0.056810	-0.002294	0.000947
$\text{var}_c(\hat{\underline{\theta}}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	0.1731	2.1944	0.695157	0.736309	0.646698	0.654963	-0.085223	-0.044437
	$\text{var}(\hat{\sigma}_v^2)$	0.0183	0.1257	0.059614	0.063028	0.051068	0.051559	-0.011712	-0.008292

Table F.4b - Evaluation of $\text{var}(\hat{\underline{\theta}})$ considering $m^{sim} = 47$ and $n_j^{sim} = 5$ (part b).

		pop	$\hat{E}(\text{var}(\hat{\theta}))$	rel bias	$\text{var}(\text{var}(\hat{\theta}))$	$cv(\text{var}(\hat{\theta}))$	$s.e.[\hat{E}(\text{var}(\hat{\theta}))]$	mse
$\text{var}_n(\hat{\theta}_{ML})$	$\text{var}(\hat{\sigma}_u^2)$	0.333080	0.281794	-15.40%	2.57×10^{-3}	17.99%	5.07×10^{-4}	5.20×10^{-3}
	$\text{var}(\hat{\sigma}_v^2)$	0.024174	0.023940	-0.97%	8.06×10^{-6}	11.86%	2.84×10^{-5}	8.12×10^{-6}
$\text{var}_n(\hat{\theta}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	0.331252	0.277689	-16.17%	1.72×10^{-3}	14.93%	4.15×10^{-4}	4.59×10^{-3}
	$\text{var}(\hat{\sigma}_v^2)$	0.025080	0.022561	-10.04%	4.31×10^{-6}	9.20%	2.07×10^{-5}	1.06×10^{-5}
$\text{var}_{adf}(\hat{\theta}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	0.331252	0.276975	-16.39%	2.49×10^{-3}	18.02%	4.99×10^{-4}	5.44×10^{-3}
	$\text{var}(\hat{\sigma}_v^2)$	0.025080	0.022773	-9.20%	7.28×10^{-6}	11.85%	2.70×10^{-5}	1.26×10^{-5}
$\text{var}_c(\hat{\theta}_{ML})$	$\text{var}(\hat{\sigma}_u^2)$	0.333080	0.327163	-1.78%	1.96×10^{-2}	42.83%	1.40×10^{-3}	1.97×10^{-2}
	$\text{var}(\hat{\sigma}_v^2)$	0.024174	0.024078	-0.40%	7.30×10^{-5}	35.49%	8.54×10^{-5}	7.30×10^{-5}
$\text{var}_c(\hat{\theta}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	0.331252	0.321287	-3.01%	1.90×10^{-2}	42.90%	1.38×10^{-3}	1.91×10^{-2}
	$\text{var}(\hat{\sigma}_v^2)$	0.025080	0.022885	-8.75%	6.82×10^{-5}	36.09%	8.26×10^{-5}	7.30×10^{-5}

Table F.5a – Evaluation of $\text{var}(\hat{\theta})$ considering $m^{sim} = 20$ and $n_j^{sim} = n_j^{sim*}$ (part a).

		min	max	95% ci for pop lower bound	95% ci for pop upper bound	95% ci for $\hat{E}(\text{var}(\hat{\theta}))$ lower bound	95% ci for $\hat{E}(\text{var}(\hat{\theta}))$ upper bound	95% ci for bias lower bound	95% ci for bias upper bound
$\text{var}_n(\hat{\theta}_{ML})$	$\text{var}(\hat{\sigma}_u^2)$	0.1514	0.5558	0.323479	0.342681	0.280800	0.282788	-0.060754	-0.041752
	$\text{var}(\hat{\sigma}_v^2)$	0.0149	0.0380	0.023513	0.024835	0.023884	0.023996	-0.000891	0.000428
$\text{var}_n(\hat{\theta}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	0.1608	0.5291	0.321716	0.340788	0.276876	0.278501	-0.062958	-0.044103
	$\text{var}(\hat{\sigma}_v^2)$	0.0161	0.0325	0.024395	0.025765	0.022520	0.022602	-0.003200	-0.001833
$\text{var}_{adf}(\hat{\theta}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	0.1516	0.5523	0.321716	0.340788	0.275996	0.277953	-0.063688	-0.044801
	$\text{var}(\hat{\sigma}_v^2)$	0.0138	0.0365	0.024395	0.025765	0.022720	0.022826	-0.002989	-0.001621
$\text{var}_c(\hat{\theta}_{ML})$	$\text{var}(\hat{\sigma}_u^2)$	0.0521	1.5054	0.323479	0.342681	0.324417	0.329909	-0.015554	0.003786
	$\text{var}(\hat{\sigma}_v^2)$	0.0050	0.0860	0.023513	0.024835	0.023911	0.024246	-0.000773	0.000587
$\text{var}_c(\hat{\theta}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	0.0526	1.4812	0.321716	0.340788	0.318586	0.323989	-0.019532	-0.000332
	$\text{var}(\hat{\sigma}_v^2)$	0.0044	0.0806	0.024395	0.025765	0.022723	0.023046	-0.002896	-0.001490

Table F.5b – Evaluation of $\text{var}(\hat{\theta})$ considering $m^{sim} = 20$ and $n_j^{sim} = n_j^{sim*}$ (part b).

		pop	$\hat{E}(\text{var}(\hat{\theta}))$	rel bias	$\text{var}(\text{var}(\hat{\theta}))$	$cv(\text{var}(\hat{\theta}))$	$s.e.[\hat{E}(\text{var}(\hat{\theta}))]$	mse
$\text{var}_n(\hat{\theta}_{ML})$	$\text{var}(\hat{\sigma}_u^2)$	0.583875	0.517089	-11.44%	1.33×10^{-2}	22.27%	1.15×10^{-3}	1.77×10^{-2}
	$\text{var}(\hat{\sigma}_v^2)$	0.044425	0.043912	-1.15%	3.32×10^{-5}	13.13%	5.76×10^{-5}	3.35×10^{-5}
$\text{var}_n(\hat{\theta}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	0.577675	0.505002	-12.58%	7.83×10^{-3}	17.53%	8.85×10^{-4}	1.31×10^{-2}
	$\text{var}(\hat{\sigma}_v^2)$	0.046465	0.040540	-12.75%	1.21×10^{-5}	8.57%	3.47×10^{-5}	4.72×10^{-5}
$\text{var}_{adf}(\hat{\theta}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	0.577675	0.502581	-13.00%	1.26×10^{-2}	22.33%	1.12×10^{-3}	1.82×10^{-2}
	$\text{var}(\hat{\sigma}_v^2)$	0.046465	0.040557	-12.71%	2.94×10^{-5}	13.37%	5.42×10^{-5}	6.43×10^{-5}
$\text{var}_c(\hat{\theta}_{ML})$	$\text{var}(\hat{\sigma}_u^2)$	0.583875	0.556483	-4.69%	5.14×10^{-2}	40.75%	2.27×10^{-3}	5.22×10^{-2}
	$\text{var}(\hat{\sigma}_v^2)$	0.044425	0.044208	-0.49%	2.32×10^{-4}	34.44%	1.52×10^{-4}	2.32×10^{-4}
$\text{var}_c(\hat{\theta}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	0.577675	0.539852	-6.55%	4.86×10^{-2}	40.83%	2.20×10^{-3}	5.00×10^{-2}
	$\text{var}(\hat{\sigma}_v^2)$	0.046465	0.040783	-12.23%	2.04×10^{-4}	35.00%	1.43×10^{-4}	2.36×10^{-4}

Table F.6a – Evaluation of $\text{var}(\hat{\theta})$ considering $m^{sim} = 20$ and $n_j^{sim} = 15$ (part a).

		min	max	95% ci for pop lower bound	95% ci for pop upper bound	95% ci for $\hat{E}(\mathrm{var}(\hat{\underline{\theta}}))$ lower bound	95% ci for $\hat{E}(\mathrm{var}(\hat{\underline{\theta}}))$ upper bound	95% ci for bias lower bound	95% ci for bias upper bound
$\mathrm{var}_n(\hat{\underline{\theta}}_{ML})$	$\mathrm{var}(\hat{\sigma}_u^2)$	0.2299	1.1772	0.567364	0.600386	0.514832	0.519346	-0.083017	-0.050437
	$\mathrm{var}(\hat{\sigma}_v^2)$	0.0273	0.0746	0.043183	0.045667	0.043799	0.044025	-0.001749	0.000731
$\mathrm{var}_n(\hat{\underline{\theta}}_{GLS2})$	$\mathrm{var}(\hat{\sigma}_u^2)$	0.2678	0.9397	0.561332	0.594018	0.503268	0.506737	-0.088670	-0.056561
	$\mathrm{var}(\hat{\sigma}_v^2)$	0.0270	0.0566	0.045157	0.047773	0.040472	0.040608	-0.007222	-0.004617
$\mathrm{var}_{adj}(\hat{\underline{\theta}}_{GLS2})$	$\mathrm{var}(\hat{\sigma}_u^2)$	0.2171	1.0978	0.561332	0.594018	0.500381	0.504781	-0.091163	-0.058911
	$\mathrm{var}(\hat{\sigma}_v^2)$	0.0248	0.0740	0.045157	0.047773	0.040451	0.040663	-0.007209	-0.004597
$\mathrm{var}_c(\hat{\underline{\theta}}_{ML})$	$\mathrm{var}(\hat{\sigma}_u^2)$	0.0838	2.3166	0.567364	0.600386	0.552039	0.560928	-0.043954	-0.010713
	$\mathrm{var}(\hat{\sigma}_v^2)$	0.0059	0.1338	0.043183	0.045667	0.043909	0.044506	-0.001481	0.001055
$\mathrm{var}_c(\hat{\underline{\theta}}_{GLS2})$	$\mathrm{var}(\hat{\sigma}_u^2)$	0.0718	2.1821	0.561332	0.594018	0.535532	0.544172	-0.054183	-0.021348
	$\mathrm{var}(\hat{\sigma}_v^2)$	0.0076	0.1302	0.045157	0.047773	0.040503	0.041063	-0.007006	-0.004348

Table F.6b – Evaluation of $\mathrm{var}(\hat{\underline{\theta}})$ considering $m^{sim} = 20$ and $n_j^{sim} = 15$ (part b).

		pop	$\hat{E}(\mathrm{var}(\hat{\underline{\theta}}))$	rel bias	$\mathrm{var}(\mathrm{var}(\hat{\underline{\theta}}))$	$cv(\mathrm{var}(\hat{\underline{\theta}}))$	$s.e.[\hat{E}(\mathrm{var}(\hat{\underline{\theta}}))]$	mse
$\mathrm{var}_n(\hat{\underline{\theta}}_{ML})$	$\mathrm{var}(\hat{\sigma}_u^2)$	0.839370	0.763898	-8.99%	4.29×10^{-2}	27.12%	2.07×10^{-3}	0.048610
	$\mathrm{var}(\hat{\sigma}_v^2)$	0.066161	0.065438	-1.09%	1.13×10^{-4}	16.22%	1.06×10^{-4}	0.000113
$\mathrm{var}_n(\hat{\underline{\theta}}_{GLS2})$	$\mathrm{var}(\hat{\sigma}_u^2)$	0.828626	0.738846	-10.83%	2.45×10^{-2}	21.18%	1.57×10^{-3}	0.032557
	$\mathrm{var}(\hat{\sigma}_v^2)$	0.070983	0.059060	-16.80%	3.97×10^{-5}	10.67%	6.30×10^{-5}	0.000182
$\mathrm{var}_{adj}(\hat{\underline{\theta}}_{GLS2})$	$\mathrm{var}(\hat{\sigma}_u^2)$	0.828626	0.732514	-11.60%	4.01×10^{-2}	27.33%	2.00×10^{-3}	0.049310
	$\mathrm{var}(\hat{\sigma}_v^2)$	0.070983	0.058499	-17.59%	9.39×10^{-5}	16.56%	9.69×10^{-5}	0.000250
$\mathrm{var}_c(\hat{\underline{\theta}}_{ML})$	$\mathrm{var}(\hat{\sigma}_u^2)$	0.839370	0.799886	-4.70%	1.17×10^{-1}	42.69%	3.41×10^{-3}	0.118144
	$\mathrm{var}(\hat{\sigma}_v^2)$	0.066161	0.065589	-0.87%	5.18×10^{-4}	34.69%	2.28×10^{-4}	0.000518
$\mathrm{var}_c(\hat{\underline{\theta}}_{GLS2})$	$\mathrm{var}(\hat{\sigma}_u^2)$	0.828626	0.765029	-7.68%	1.08×10^{-1}	43.04%	3.29×10^{-3}	0.112463
	$\mathrm{var}(\hat{\sigma}_v^2)$	0.070983	0.058505	-17.58%	4.26×10^{-4}	35.29%	2.06×10^{-4}	0.000582

Table F.7a – Evaluation of $\mathrm{var}(\hat{\underline{\theta}})$ considering $m^{sim} = 20$ and $n_j^{sim} = 10$ (part a).

		min	max	95% ci for pop lower bound	95% ci for pop upper bound	95% ci for $\hat{E}(\mathrm{var}(\hat{\underline{\theta}}))$ lower bound	95% ci for $\hat{E}(\mathrm{var}(\hat{\underline{\theta}}))$ upper bound	95% ci for bias lower bound	95% ci for bias upper bound
$\mathrm{var}_n(\hat{\underline{\theta}}_{ML})$	$\mathrm{var}(\hat{\sigma}_u^2)$	0.2734	2.2053	0.816075	0.862665	0.759838	0.767958	-0.098422	-0.052355
	$\mathrm{var}(\hat{\sigma}_v^2)$	0.0338	0.1478	0.064348	0.067974	0.065230	0.065646	-0.002523	0.001090
$\mathrm{var}_n(\hat{\underline{\theta}}_{GLS2})$	$\mathrm{var}(\hat{\sigma}_u^2)$	0.3117	1.5081	0.805662	0.851590	0.735779	0.741914	-0.112221	-0.067174
	$\mathrm{var}(\hat{\sigma}_v^2)$	0.0383	0.0839	0.069046	0.072920	0.058936	0.059183	-0.013846	-0.009987
$\mathrm{var}_{adj}(\hat{\underline{\theta}}_{GLS2})$	$\mathrm{var}(\hat{\sigma}_u^2)$	0.2711	1.9905	0.805662	0.851590	0.728590	0.736437	-0.118717	-0.073344
	$\mathrm{var}(\hat{\sigma}_v^2)$	0.0279	0.1053	0.069046	0.072920	0.058309	0.058689	-0.014412	-0.010543
$\mathrm{var}_c(\hat{\underline{\theta}}_{ML})$	$\mathrm{var}(\hat{\sigma}_u^2)$	0.1630	3.2148	0.816075	0.862665	0.793193	0.806578	-0.062897	-0.015905
	$\mathrm{var}(\hat{\sigma}_v^2)$	0.0155	0.2518	0.064348	0.067974	0.065143	0.066034	-0.002413	0.001281
$\mathrm{var}_c(\hat{\underline{\theta}}_{GLS2})$	$\mathrm{var}(\hat{\sigma}_u^2)$	0.1272	2.9557	0.805662	0.851590	0.758575	0.771482	-0.086641	-0.040389
	$\mathrm{var}(\hat{\sigma}_v^2)$	0.0095	0.1861	0.069046	0.072920	0.058101	0.058910	-0.014435	-0.010507

Table F.7a – Evaluation of $\mathrm{var}(\hat{\underline{\theta}})$ considering $m^{sim} = 20$ and $n_j^{sim} = 10$ (part a).

		pop	$\hat{E}(\text{var}(\hat{\underline{\theta}}))$	rel bias	$\text{var}(\text{var}(\hat{\underline{\theta}}))$	$cv(\text{var}(\hat{\underline{\theta}}))$	$s.e.[\hat{E}(\text{var}(\hat{\underline{\theta}}))]$	mse
$\text{var}_n(\hat{\underline{\theta}}_{ML})$	$\text{var}(\hat{\sigma}_u^2)$	1.631626	1.461839	-10.41%	0.317611	38.55%	0.005636	0.346439
	$\text{var}(\hat{\sigma}_v^2)$	0.133216	0.127079	-4.61%	0.000835	22.74%	0.000289	0.000873
$\text{var}_n(\hat{\underline{\theta}}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	1.584676	1.374454	-13.27%	0.176290	30.55%	0.004199	0.220483
	$\text{var}(\hat{\sigma}_v^2)$	0.148327	0.107566	-27.48%	0.000293	15.92%	0.000171	0.001955
$\text{var}_{adf}(\hat{\underline{\theta}}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	1.584577	1.349494	-14.84%	0.277523	39.04%	0.005269	0.332788
	$\text{var}(\hat{\sigma}_v^2)$	0.148316	0.103244	-30.39%	0.000598	23.68%	0.000244	0.002629
$\text{var}_c(\hat{\underline{\theta}}_{ML})$	$\text{var}(\hat{\sigma}_u^2)$	1.631626	1.482539	-9.14%	0.552066	50.12%	0.007430	0.574293
	$\text{var}(\hat{\sigma}_v^2)$	0.133216	0.127345	-4.41%	0.002202	36.85%	0.000469	0.002236
$\text{var}_c(\hat{\underline{\theta}}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	1.584479	1.367448	-13.70%	0.481593	50.75%	0.006940	0.528695
	$\text{var}(\hat{\sigma}_v^2)$	0.148273	0.103666	-30.08%	0.001577	38.30%	0.000397	0.003566

Table F.8a – Evaluation of $\text{var}(\hat{\underline{\theta}})$ considering $m^{sim} = 20$ and $n_j^{sim} = 5$ (part a).

		min	max	95% ci for pop		95% ci for $\hat{E}(\text{var}(\hat{\underline{\theta}}))$		95% ci for bias	
				lower bound	upper bound	lower bound	upper bound	lower bound	upper bound
$\text{var}_n(\hat{\underline{\theta}}_{ML})$	$\text{var}(\hat{\sigma}_u^2)$	0.3572	5.7767	1.583848	1.679404	1.450793	1.472885	-0.216107	-0.123140
	$\text{var}(\hat{\sigma}_v^2)$	0.0523	0.3040	0.129458	0.136974	0.126513	0.127646	-0.009848	-0.002398
$\text{var}_n(\hat{\underline{\theta}}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	0.3329	4.0834	1.538211	1.631141	1.366225	1.382684	-0.254576	-0.165550
	$\text{var}(\hat{\sigma}_v^2)$	0.0544	0.2013	0.144066	0.152588	0.107231	0.107902	-0.044975	-0.036518
$\text{var}_{adf}(\hat{\underline{\theta}}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	0.2918	4.8223	1.538109	1.631045	1.339167	1.359820	-0.280019	-0.189831
	$\text{var}(\hat{\sigma}_v^2)$	0.0384	0.2435	0.144054	0.152578	0.102765	0.103723	-0.049292	-0.040822
$\text{var}_c(\hat{\underline{\theta}}_{ML})$	$\text{var}(\hat{\sigma}_u^2)$	0.2079	7.9503	1.583848	1.679404	1.467976	1.497102	-0.195964	-0.101884
	$\text{var}(\hat{\sigma}_v^2)$	0.0277	0.4288	0.129458	0.136974	0.126426	0.128265	-0.009659	-0.002055
$\text{var}_c(\hat{\underline{\theta}}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	0.1340	8.1074	1.538011	1.630947	1.353845	1.381051	-0.262512	-0.171233
	$\text{var}(\hat{\sigma}_v^2)$	0.0210	0.3605	0.144012	0.152534	0.102888	0.104444	-0.048866	-0.040318

Table F.8b – Evaluation of $\text{var}(\hat{\underline{\theta}})$ considering $m^{sim} = 20$ and $n_j^{sim} = 5$ (part b).

		pop	$\hat{E}(\text{var}(\hat{\underline{\theta}}))$	rel bias	$\text{var}(\text{var}(\hat{\underline{\theta}}))$	$cv(\text{var}(\hat{\underline{\theta}}))$	$s.e.[\hat{E}(\text{var}(\hat{\underline{\theta}}))]$	mse
$\text{var}_n(\hat{\underline{\theta}}_{ML})$	$\text{var}(\hat{\sigma}_u^2)$	0.477175	0.386391	-19.03%	7.22×10^{-3}	21.99%	8.50×10^{-4}	1.55×10^{-2}
	$\text{var}(\hat{\sigma}_v^2)$	0.033756	0.033059	-2.06%	2.23×10^{-5}	14.30%	4.73×10^{-5}	2.28×10^{-5}
$\text{var}_n(\hat{\underline{\theta}}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	0.472661	0.378737	-19.87%	4.72×10^{-3}	18.14%	6.87×10^{-4}	1.35×10^{-2}
	$\text{var}(\hat{\sigma}_v^2)$	0.035210	0.030817	-12.48%	1.19×10^{-5}	11.22%	3.46×10^{-5}	3.12×10^{-5}
$\text{var}_{adf}(\hat{\underline{\theta}}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	0.472661	0.377801	-20.07%	6.87×10^{-3}	21.93%	8.29×10^{-4}	1.59×10^{-2}
	$\text{var}(\hat{\sigma}_v^2)$	0.035210	0.031017	-11.91%	1.93×10^{-5}	14.17%	4.39×10^{-5}	3.69×10^{-5}
$\text{var}_c(\hat{\underline{\theta}}_{ML})$	$\text{var}(\hat{\sigma}_u^2)$	0.477175	0.439764	-7.84%	4.82×10^{-2}	49.91%	2.20×10^{-3}	4.96×10^{-2}
	$\text{var}(\hat{\sigma}_v^2)$	0.033756	0.032900	-2.54%	1.91×10^{-4}	41.95%	1.38×10^{-4}	1.91×10^{-4}
$\text{var}_c(\hat{\underline{\theta}}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	0.472661	0.429026	-9.23%	4.57×10^{-2}	49.86%	2.14×10^{-3}	4.77×10^{-2}
	$\text{var}(\hat{\sigma}_v^2)$	0.035210	0.030832	-12.43%	1.71×10^{-4}	42.40%	1.31×10^{-4}	1.90×10^{-4}

Table F.9a – Evaluation of $\text{var}(\hat{\underline{\theta}})$ considering $m^{sim} = 15$ and $n_j^{sim} = n_j^{sim*}$ (part a).

		min	max	95% ci for pop lower bound	95% ci for pop upper bound	95% ci for $\hat{E}\left(\text{var}(\hat{\underline{\theta}})\right)$ lower bound	95% ci for $\hat{E}\left(\text{var}(\hat{\underline{\theta}})\right)$ upper bound	95% ci for bias lower bound	95% ci for bias upper bound
$\text{var}_n\left(\hat{\underline{\theta}}_{ML}\right)$	$\text{var}(\hat{\sigma}_u^2)$	0.1751	0.8897	0.463731	0.490619	0.384726	0.388056	-0.104039	-0.077433
	$\text{var}(\hat{\sigma}_v^2)$	0.0192	0.0539	0.032823	0.034689	0.032967	0.033152	-0.001624	0.000237
$\text{var}_n\left(\hat{\underline{\theta}}_{GLS2}\right)$	$\text{var}(\hat{\sigma}_u^2)$	0.1885	0.7027	0.459261	0.486061	0.377390	0.380083	-0.107103	-0.080651
	$\text{var}(\hat{\sigma}_v^2)$	0.0191	0.0451	0.034227	0.036193	0.030749	0.030884	-0.005371	-0.003409
$\text{var}_{adj}\left(\hat{\underline{\theta}}_{GLS2}\right)$	$\text{var}(\hat{\sigma}_u^2)$	0.1716	0.8724	0.459261	0.486061	0.376177	0.379425	-0.108092	-0.081533
	$\text{var}(\hat{\sigma}_v^2)$	0.0179	0.0507	0.034227	0.036193	0.030931	0.031104	-0.005171	-0.003207
$\text{var}_c\left(\hat{\underline{\theta}}_{ML}\right)$	$\text{var}(\hat{\sigma}_u^2)$	0.0505	2.0874	0.463731	0.490619	0.435462	0.444067	-0.050993	-0.023733
	$\text{var}(\hat{\sigma}_v^2)$	0.0034	0.1221	0.032823	0.034689	0.032630	0.033171	-0.001813	0.000108
$\text{var}_c\left(\hat{\underline{\theta}}_{GLS2}\right)$	$\text{var}(\hat{\sigma}_u^2)$	0.0401	2.1364	0.459261	0.486061	0.424834	0.433219	-0.057163	-0.030012
	$\text{var}(\hat{\sigma}_v^2)$	0.0033	0.1145	0.034227	0.036193	0.030576	0.031088	-0.005382	-0.003367

Table F.9b – Evaluation of $\text{var}(\hat{\underline{\theta}})$ considering $m^{sim} = 15$ and $n_j^{sim} = n_j^{sim*}$ (part b).

		pop	$\hat{E}\left(\text{var}(\hat{\underline{\theta}})\right)$	rel bias	$\text{var}\left(\text{var}(\hat{\underline{\theta}})\right)$	$cv\left(\text{var}(\hat{\underline{\theta}})\right)$	$s.e.\left[\hat{E}\left(\text{var}(\hat{\underline{\theta}})\right)\right]$	mse
$\text{var}_n\left(\hat{\underline{\theta}}_{ML}\right)$	$\text{var}(\hat{\sigma}_u^2)$	0.765576	0.679436	-11.25%	3.10×10^{-2}	25.90%	1.76×10^{-3}	3.84×10^{-2}
	$\text{var}(\hat{\sigma}_v^2)$	0.059867	0.058165	-2.84%	7.83×10^{-5}	15.21%	8.85×10^{-5}	8.12×10^{-5}
$\text{var}_n\left(\hat{\underline{\theta}}_{GLS2}\right)$	$\text{var}(\hat{\sigma}_u^2)$	0.757259	0.660279	-12.81%	1.79×10^{-2}	20.27%	1.34×10^{-3}	2.73×10^{-2}
	$\text{var}(\hat{\sigma}_v^2)$	0.064683	0.053032	-18.01%	2.90×10^{-5}	10.15%	5.39×10^{-5}	1.65×10^{-4}
$\text{var}_{adj}\left(\hat{\underline{\theta}}_{GLS2}\right)$	$\text{var}(\hat{\sigma}_u^2)$	0.757259	0.654696	-13.54%	2.92×10^{-2}	26.12%	1.71×10^{-3}	3.98×10^{-2}
	$\text{var}(\hat{\sigma}_v^2)$	0.064683	0.052598	-18.68%	6.81×10^{-5}	15.69%	8.25×10^{-5}	2.14×10^{-4}
$\text{var}_c\left(\hat{\underline{\theta}}_{ML}\right)$	$\text{var}(\hat{\sigma}_u^2)$	0.765576	0.735903	-3.88%	1.26×10^{-1}	48.16%	3.54×10^{-3}	1.26×10^{-1}
	$\text{var}(\hat{\sigma}_v^2)$	0.059867	0.058387	-2.47%	5.19×10^{-4}	39.00%	2.28×10^{-4}	5.21×10^{-4}
$\text{var}_c\left(\hat{\underline{\theta}}_{GLS2}\right)$	$\text{var}(\hat{\sigma}_u^2)$	0.757259	0.708222	-6.48%	1.17×10^{-1}	48.38%	3.43×10^{-3}	1.20×10^{-1}
	$\text{var}(\hat{\sigma}_v^2)$	0.064683	0.052861	-18.28%	4.49×10^{-4}	40.08%	2.12×10^{-4}	5.89×10^{-4}

Table F.10a – Evaluation of $\text{var}(\hat{\underline{\theta}})$ considering $m^{sim} = 15$ and $n_j^{sim} = 15$ (part a).

		min	max	95% ci for pop lower bound	95% ci for pop upper bound	95% ci for $\hat{E}\left(\text{var}(\hat{\underline{\theta}})\right)$ lower bound	95% ci for $\hat{E}\left(\text{var}(\hat{\underline{\theta}})\right)$ upper bound	95% ci for bias lower bound	95% ci for bias upper bound
$\text{var}_n\left(\hat{\underline{\theta}}_{ML}\right)$	$\text{var}(\hat{\sigma}_u^2)$	0.2466	1.8227	0.744097	0.787055	0.675987	0.682885	-0.107148	-0.064978
	$\text{var}(\hat{\sigma}_v^2)$	0.0349	0.1008	0.058193	0.061541	0.057992	0.058339	-0.003363	-0.000028
$\text{var}_n\left(\hat{\underline{\theta}}_{GLS2}\right)$	$\text{var}(\hat{\sigma}_u^2)$	0.3140	1.3703	0.735879	0.778639	0.657656	0.662902	-0.117775	-0.076035
	$\text{var}(\hat{\sigma}_v^2)$	0.0350	0.0776	0.062885	0.066481	0.052927	0.053138	-0.013434	-0.009855
$\text{var}_{adj}\left(\hat{\underline{\theta}}_{GLS2}\right)$	$\text{var}(\hat{\sigma}_u^2)$	0.2359	1.7913	0.735879	0.778639	0.651345	0.658048	-0.123461	-0.081514
	$\text{var}(\hat{\sigma}_v^2)$	0.0303	0.0935	0.062885	0.066481	0.052437	0.052760	-0.013872	-0.010284
$\text{var}_c\left(\hat{\underline{\theta}}_{ML}\right)$	$\text{var}(\hat{\sigma}_u^2)$	0.0885	3.8642	0.744097	0.787055	0.728957	0.742850	-0.051284	-0.007907
	$\text{var}(\hat{\sigma}_v^2)$	0.0048	0.2103	0.058193	0.061541	0.057941	0.058833	-0.003199	0.000251
$\text{var}_c\left(\hat{\underline{\theta}}_{GLS2}\right)$	$\text{var}(\hat{\sigma}_u^2)$	0.0805	3.8247	0.735879	0.778639	0.701507	0.714937	-0.070537	-0.027387
	$\text{var}(\hat{\sigma}_v^2)$	0.0057	0.1794	0.062885	0.066481	0.052446	0.053276	-0.013657	-0.009975

Table F.10b – Evaluation of $\text{var}(\hat{\underline{\theta}})$ considering $m^{sim} = 15$ and $n_j^{sim} = 15$ (part b).

		pop	$\hat{E}(\text{var}(\hat{\theta}))$	rel bias	$\text{var}(\text{var}(\hat{\theta}))$	$cv(\text{var}(\hat{\theta}))$	$s.e.[\hat{E}(\text{var}(\hat{\theta}))]$	mse
$\text{var}_n(\hat{\theta}_{ML})$	$\text{var}(\hat{\sigma}_u^2)$	1.096834	0.997739	-9.03%	9.65×10^{-2}	31.14%	3.11×10^{-3}	0.106361
	$\text{var}(\hat{\sigma}_v^2)$	0.088289	0.086617	-1.89%	2.54×10^{-4}	18.39%	1.59×10^{-4}	0.000257
$\text{var}_n(\hat{\theta}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	1.078658	0.956722	-11.30%	5.52×10^{-2}	24.55%	2.35×10^{-3}	0.070020
	$\text{var}(\hat{\sigma}_v^2)$	0.096592	0.076531	-20.77%	9.20×10^{-5}	12.53%	9.59×10^{-5}	0.000494
$\text{var}_{adj}(\hat{\theta}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	1.078658	0.945363	-12.36%	8.87×10^{-2}	31.50%	2.98×10^{-3}	0.106421
	$\text{var}(\hat{\sigma}_v^2)$	0.096592	0.075123	-22.23%	2.06×10^{-4}	19.11%	1.44×10^{-4}	0.000667
$\text{var}_c(\hat{\theta}_{ML})$	$\text{var}(\hat{\sigma}_u^2)$	1.096834	1.043433	-4.87%	2.74×10^{-1}	50.19%	5.24×10^{-3}	0.277060
	$\text{var}(\hat{\sigma}_v^2)$	0.088289	0.087068	-1.38%	1.27×10^{-3}	40.87%	3.56×10^{-4}	0.001268
$\text{var}_c(\hat{\theta}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	1.078658	0.985440	-8.64%	2.51×10^{-1}	50.85%	5.01×10^{-3}	0.259816
	$\text{var}(\hat{\sigma}_v^2)$	0.096592	0.075423	-21.92%	9.97×10^{-4}	41.86%	3.16×10^{-4}	0.001445

Table F.11a – Evaluation of $\text{var}(\hat{\theta})$ considering $m^{sim}=15$ and $n_j^{sim}=10$ (part a).

		min	max	95% ci for pop		95% ci for $\hat{E}(\text{var}(\hat{\theta}))$		95% ci for bias	
				lower bound	upper bound	lower bound	upper bound	lower bound	upper bound
$\text{var}_n(\hat{\theta}_{ML})$	$\text{var}(\hat{\sigma}_u^2)$	0.2649	3.3397	1.065946	1.127722	0.991649	1.003829	-0.129368	-0.068603
	$\text{var}(\hat{\sigma}_v^2)$	0.0416	0.1656	0.085871	0.090707	0.086305	0.086929	-0.004081	0.000755
$\text{var}_n(\hat{\theta}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	0.3224	2.2463	1.047954	1.109362	0.952119	0.961324	-0.151731	-0.091927
	$\text{var}(\hat{\sigma}_v^2)$	0.0459	0.1154	0.093947	0.099237	0.076343	0.076719	-0.022688	-0.017415
$\text{var}_{adj}(\hat{\theta}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	0.2584	3.5117	1.047954	1.109362	0.939527	0.951199	-0.163348	-0.103027
	$\text{var}(\hat{\sigma}_v^2)$	0.0379	0.1610	0.093947	0.099237	0.074842	0.075405	-0.024103	-0.018816
$\text{var}_c(\hat{\theta}_{ML})$	$\text{var}(\hat{\sigma}_u^2)$	0.1066	7.3566	1.065946	1.127722	1.033170	1.053697	-0.084396	-0.022187
	$\text{var}(\hat{\sigma}_v^2)$	0.0095	0.3221	0.085871	0.090707	0.086371	0.087766	-0.003707	0.001283
$\text{var}_c(\hat{\theta}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	0.1068	6.9005	1.047954	1.109362	0.975618	0.995262	-0.123959	-0.062262
	$\text{var}(\hat{\sigma}_v^2)$	0.0086	0.3226	0.093947	0.099237	0.074804	0.076042	-0.023860	-0.018459

Table F.11b – Evaluation of $\text{var}(\hat{\theta})$ considering $m^{sim}=15$ and $n_j^{sim}=10$ (part b).

		pop	$\hat{E}(\text{var}(\hat{\theta}))$	rel bias	$\text{var}(\text{var}(\hat{\theta}))$	$cv(\text{var}(\hat{\theta}))$	$s.e.[\hat{E}(\text{var}(\hat{\theta}))]$	mse
$\text{var}_n(\hat{\theta}_{ML})$	$\text{var}(\hat{\sigma}_u^2)$	2.222778	1.882306	-15.32%	0.740876	45.73%	0.008607	0.856797
	$\text{var}(\hat{\sigma}_v^2)$	0.178319	0.168233	-5.66%	0.001933	26.14%	0.000440	0.002035
$\text{var}_n(\hat{\theta}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	2.127350	1.738628	-18.27%	0.415753	37.09%	0.006448	0.566858
	$\text{var}(\hat{\sigma}_v^2)$	0.204034	0.135696	-33.49%	0.000668	19.05%	0.000259	0.005338
$\text{var}_{adj}(\hat{\theta}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	2.127993	1.700549	-20.09%	0.627852	46.60%	0.007927	0.810560
	$\text{var}(\hat{\sigma}_v^2)$	0.203879	0.128130	-37.15%	0.001287	28.00%	0.000359	0.007025
$\text{var}_c(\hat{\theta}_{ML})$	$\text{var}(\hat{\sigma}_u^2)$	2.222778	1.911003	-14.03%	1.249229	58.49%	0.011177	1.346432
	$\text{var}(\hat{\sigma}_v^2)$	0.178319	0.168942	-5.26%	0.005340	43.26%	0.000731	0.005428
$\text{var}_c(\hat{\theta}_{GLS2})$	$\text{var}(\hat{\sigma}_u^2)$	2.128328	1.727843	-18.82%	1.065758	59.75%	0.010329	1.226146
	$\text{var}(\hat{\sigma}_v^2)$	0.203844	0.128204	-37.11%	0.003383	45.37%	0.000582	0.009104

Table F.12a – Evaluation of $\text{var}(\hat{\theta})$ considering $m^{sim}=15$ and $n_j^{sim}=5$ (part a).

		min	max	95% ci for pop lower bound	95% ci for pop upper bound	95% ci for $\hat{E}(\text{var}(\hat{\underline{\theta}}))$ lower bound	95% ci for $\hat{E}(\text{var}(\hat{\underline{\theta}}))$ upper bound	95% ci for bias lower bound	95% ci for bias upper bound
$\text{var}_n\left(\hat{\underline{\theta}}_{ML}\right)$	$\text{var}(\hat{\sigma}_u^2)$	0.3221	10.8445	2.158452	2.287104	1.865436	1.899177	-0.402035	-0.278463
	$\text{var}(\hat{\sigma}_v^2)$	0.0622	0.4623	0.173285	0.183353	0.167371	0.169095	-0.015091	-0.005044
$\text{var}_n\left(\hat{\underline{\theta}}_{GLS2}\right)$	$\text{var}(\hat{\sigma}_u^2)$	0.3602	10.9200	2.064886	2.189814	1.725989	1.751266	-0.447438	-0.329581
	$\text{var}(\hat{\sigma}_v^2)$	0.0134	0.2540	0.198347	0.209721	0.135190	0.136203	-0.073957	-0.062677
$\text{var}_{adf}\left(\hat{\underline{\theta}}_{GLS2}\right)$	$\text{var}(\hat{\sigma}_u^2)$	0.1205	10.6746	2.065488	2.190498	1.685012	1.716087	-0.487046	-0.367415
	$\text{var}(\hat{\sigma}_v^2)$	0.0371	0.4986	0.198198	0.209560	0.127427	0.128834	-0.081387	-0.070069
$\text{var}_c\left(\hat{\underline{\theta}}_{ML}\right)$	$\text{var}(\hat{\sigma}_u^2)$	0.1481	15.2083	2.158452	2.287104	1.889096	1.932909	-0.374646	-0.248460
	$\text{var}(\hat{\sigma}_v^2)$	0.0247	0.6780	0.173285	0.183353	0.167510	0.170375	-0.014497	-0.004221
$\text{var}_c\left(\hat{\underline{\theta}}_{GLS2}\right)$	$\text{var}(\hat{\sigma}_u^2)$	0.0522	12.8831	2.065815	2.190841	1.707598	1.748089	-0.461204	-0.339339
	$\text{var}(\hat{\sigma}_v^2)$	0.0131	0.8297	0.198163	0.209525	0.127063	0.129344	-0.081334	-0.069906

Table F.12b – Evaluation of $\text{var}(\hat{\underline{\theta}})$ considering $m^{sim} = 15$ and $n_j^{sim} = 5$ (part b).

Appendix G

R code used in both simulation studies for the pseudo maximum likelihood estimator

We do not intend to present in the current appendix the complete code we have written for implementing the techniques we have studied and developed in this Thesis, as that would require a reasonably large number of pages. With mainly illustration purposes, we shall thus only present an extract with some attention to procedures related to pseudo maximum likelihood estimation. Figure G.1 below provide the R code for weighted estimator of the covariance matrix included in Chapter 4, Section 4.3, Sub-section 4.3.1.

```
#Calculating yi-MU
EE<-Ys-tMU
EE<-as.matrix(EE)

#using wave LONGITUDINAL WEIGHTS (WAVE 9)
sigmaw001<-matrix(0,5,5)
sigmaw001[1,1]<-sum(w*(EE[,1])^2)/sum(w)
sigmaw001[1,2]<-sum(w*(EE[,1])*((EE[,2])))/sum(w)
sigmaw001[1,3]<-sum(w*(EE[,1])*((EE[,3])))/sum(w)
sigmaw001[1,4]<-sum(w*(EE[,1])*((EE[,4])))/sum(w)
sigmaw001[1,5]<-sum(w*(EE[,1])*((EE[,5])))/sum(w)
sigmaw001[2,1]<-sigmaw001[1,2]
sigmaw001[2,2]<-sum(w*(EE[,2])^2)/sum(w)
sigmaw001[2,3]<-sum(w*(EE[,2])*((EE[,3])))/sum(w)
sigmaw001[2,4]<-sum(w*(EE[,2])*((EE[,4])))/sum(w)
sigmaw001[2,5]<-sum(w*(EE[,2])*((EE[,5])))/sum(w)
sigmaw001[3,1]<-sigmaw001[1,3]
sigmaw001[3,2]<-sigmaw001[2,3]
sigmaw001[3,3]<-sum(w*(EE[,3])^2)/sum(w)
sigmaw001[3,4]<-sum(w*(EE[,3])*((EE[,4])))/sum(w)
sigmaw001[3,5]<-sum(w*(EE[,3])*((EE[,5])))/sum(w)
sigmaw001[4,1]<-sigmaw001[1,4]
sigmaw001[4,2]<-sigmaw001[2,4]
sigmaw001[4,3]<-sigmaw001[3,4]
sigmaw001[4,4]<-sum(w*(EE[,4])^2)/sum(w)
sigmaw001[4,5]<-sum(w*(EE[,4])*((EE[,5])))/sum(w)
sigmaw001[5,1]<-sigmaw001[1,5]
sigmaw001[5,2]<-sigmaw001[2,5]
sigmaw001[5,3]<-sigmaw001[3,5]
sigmaw001[5,4]<-sigmaw001[4,5]
sigmaw001[5,5]<-sum(w*(EE[,5])^2)/sum(w)

comment(sigmaw001)<-"Sw - Weighted S"
print(comment(sigmaw001))
print(sigmaw001)
```

Figure G.1 – R code for weighted estimator of the covariance matrix.

Figure G.2 provide the R code for the pseudo maximum likelihood estimators when fitting an UCM model (with analytic solution), and an AR1 model (with numerical minimisation) respectively.

```
# pml UCM

mlX1<-matrix(0,1,3)

mlXX1<-matrix(0,1,2)

mlXX1[1]<-(1/10)*s21w+(1/10)*s31w+(1/10)*s41w+(1/10)*s51w+(1/10)*s32w+(1/10)*s42w+
(1/10)*s52w+(1/10)*s43w+(1/10)*s53w+(1/10)*s54w

mlXX1[2]<-(1/5)*s11w-(1/10)*s21w-(1/10)*s31w-(1/10)*s41w-(1/10)*s51w+(1/5)*s22w-
(1/10)*s32w-(1/10)*s42w-(1/10)*s52w+(1/5)*s33w-(1/10)*s43w-(1/10)*s53w+(1/5)*s44w-
(1/10)*s54w+(1/5)*s55w

mlX1[1]<-mlXX1[1,1]
mlX1[2]<-mlXX1[1,2]
mlX1[3]<-mlXX1[1,1]/(mlXX1[1,1]+mlXX1[1,2])

ml<-mlX1

theta100apmltre[f,]<-ml

# pml ar1
source("/home/marcel/Marcel/point1/Fpml.txt")
min9<-nlm(Fpml,c(7.135,4.981,0))
comment(min9$estimate)<-"Theta estimates - Fpml ar1"
print(comment(min9$estimate))
print(min9$estimate)
ml<-t(matrix(min9$estimate))

theta100apml[f,]<-ml
```

Figure G.2 – R code for the pseudo maximum likelihood estimator.

Figure G.3 provide the R code for estimating the variance of the pseudo maximum likelihood point estimator, following the approach of Binder (1983), proposed in Chapter 6, Section 6.3.

```
#Variance estimation for THETA hat pml

#dSigma(THETA)/dsigmau - d1Sigma
d1Sigma<-matrix(1,5,5)
#dSigma(THETA)/dsigmav - d2Sigma
d2Sigma<-diag(5)

#Sigma(THETA)
SigmaVector<-
c(ml[1]+ml[2],ml[1],ml[1],ml[1],ml[1],ml[1],ml[1]+ml[2],ml[1],ml[1],ml[1],ml[1],ml[1],
ml[1]+ml[2],ml[1],ml[1],ml[1],ml[1],ml[1],ml[1]+ml[2],ml[1],ml[1],ml[1],ml[1],ml[1],ml[1]+
ml[2])
Sigma<-matrix(SigmaVector,5,5)
```

Figure G.3 – R code for estimating the variance of the pseudo maximum likelihood point estimator.

```
#Inverse of Sigma
SigmaI<-solve(Sigma)

#Calculate terms A1 and A2 (see your personal notes!)
A1<-SigmaI%*%d1Sigma%*%SigmaI
A2<-SigmaI%*%d2Sigma%*%SigmaI

#Calculating MUi
#calculated earlier

#Calculating yi-MU
#calculated earlier

#Calculating vector z1i
z1i<-matrix(0,dim,1)
for (k in 1:dim)
{
z1i[k,]<-t(as.matrix(EE[k,]))%*%A1%*%as.matrix(EE[k,])
}

#Calculating vector z2i
z2i<-matrix(0,dim,1)
for (k in 1:dim)
{
z2i[k,]<-t(as.matrix(EE[k,]))%*%A2%*%as.matrix(EE[k,])
}

#Create matrix z
z<-cbind(z1i,z2i)
z<-as.matrix(z)

#Calculate u(theta)1
utheta1i<-sum(w*z1i)/sum(w)

#Calculate u(theta)2
utheta2i<-sum(w*z2i)/sum(w)

#Create vector uthetai
uthetai<-cbind(utheta1i,utheta2i)

#Calculate vector (capital) ZZ1
ZZ1<-matrix(0,dim,1)
for (k in 1:dim)
{
ZZ1[k,]<-((w[k,]*z1i[k,])-(utheta1i*w[k,]))/(sum(w))
}

#Calculate vector (capital) ZZ2
ZZ2<-matrix(0,dim,1)
for (k in 1:dim)
{
ZZ2[k,]<-((w[k,]*z2i[k,])-(utheta2i*w[k,]))/(sum(w))
}

# Create ZZ matrix
ZZ<-cbind(ZZ1,ZZ2)

#Calculate ZZhat vector
ZZhat<-cbind(sum(ZZ1),sum(ZZ2))

#Reminder: number of clusters => M

#Add cluster id to ZZ1 and ZZ2
ZZ1<-cbind(ZZ1,sz1[,24])
ZZ2<-cbind(ZZ2,sz1[,24])
```

Figure G.3 – R code for estimating the variance of the pseudo maximum likelihood point estimator (continued).

```
#create vector for allocating Sum ZZ1 in each cluster
ZZ1hatm<-matrix(0,M,1)

for (m in 1:M)
{
ZZ1temp<-ZZ1[ZZ1[,2]==m,]
ZZ1hatm[m,]<-sum(ZZ1temp[,1])
}

#create vector for allocating Sum ZZ2 in each cluster
ZZ2hatm<-matrix(0,M,1)

for (m in 1:M)
{
ZZ2temp<-ZZ2[ZZ2[,2]==m,]
ZZ2hatm[m,]<-sum(ZZ2temp[,1])
}

#create matrix ZZhatm
ZZhatm<-cbind(ZZ1hatm,ZZ2hatm)

#calculate vector ZZhatmbar
ZZhatmbar<-cbind(sum(ZZ1hatm)/M,sum(ZZ2hatm)/M)

#Caculate CovLZZhat
matrix<-matrix(0,2,2)
for (m in 1:M)
{
matrix0<-t(ZZhatm[m,]-ZZhatmbar)%*%(ZZhatm[m,]-ZZhatmbar)
matrix<-matrix0+matrix
}

matrix<-(M/(M-1))*matrix

#calculate Information Matrix for Theta (p)ml
InfthetaML<-matrix(0,2,2)
InfthetaML[1,1]<-sum(diag(SigmaI%*%d1Sigma%*%SigmaI%*%d1Sigma))
InfthetaML[1,2]<-sum(diag(SigmaI%*%d1Sigma%*%SigmaI%*%d2Sigma))
InfthetaML[2,1]<-sum(diag(SigmaI%*%d2Sigma%*%SigmaI%*%d1Sigma))
InfthetaML[2,2]<-sum(diag(SigmaI%*%d2Sigma%*%SigmaI%*%d2Sigma))

#calculate covariance matrix for theta
covthetaML<-(solve(InfthetaML))%*%matrix%*%(solve(InfthetaML))

covML<-c(covthetaML[1,1],covthetaML[2,2])

varPLM[f,]<-covML
```

Figure G.3 – R code for estimating the variance of the pseudo maximum likelihood point estimator (continued).

Glossary

y	study variable or survey variable
$Y_1,\ldots,Y_N$	random vectors
$y_1,\ldots,y_N$	population values of y or a joint realisation of $Y_1,\ldots,Y_N$
$\underline{Y}_i$	random vector containing T repeated observations on Y for individual i
$\underline{y}_i$	$T \times 1$ vectors with the observed values for the response variable for each individual i
$\underline{x}_{it}$	$1 \times q$ vector with q fixed covariates
X_i	$T \times q$ matrices with covariates for individual i
q	number of fixed covariates
z_i	any variable
$\underline{x}_i^c$	$1 \times q$ vector with the q fixed covariates (cross-sectional context)
$\mathcal{U}$	finite population
N	number of elements in the finite population or population size
T	number of waves of the survey
ξ	joint distribution of $Y_1,\ldots,Y_N$ under model
S	sample or a subset of $\mathcal{U}$
n	number of elements in the sample or sample size
$p(S)$	sampling scheme
ζ	set of all possible samples
s	actual selected sample
s_T	longitudinal sample
w_i	sampling weight for individual i
π_i	inclusion probabilities for individual i
w_{iT}^*	longitudinal weight for individual i at time T
PSU	primary sampling unit
H	number of strata
m	number of PSUs in the sample

srs	simple random sample (or sampling)
ε_{it}	error term
$\underline{\varepsilon}_i$	$T \times 1$ vector with errors
u_i	permanent random effects or unobservable individual specific factors
v_{it}	transitory random effects
η_j	random area effects
σ^2	model variance of Y_{it}
$\hat{\sigma}^2$	estimator of σ^2
σ_u^2	model variance of u_i
σ_v^2	model variance of v_{it}
σ_ε^2	model variance of ε_{it}
σ_η^2	model variance of η_j
ρ	model intra-individual correlation
ρ^*	model intra-cluster correlation
$\underline{\beta}$	$q \times 1$ vector of the unknown fixed coefficients for the x variables
$\hat{\underline{\beta}}_N$	maximum census likelihood estimator for the parameter $\underline{\beta}$
$\hat{\underline{\beta}}$	estimator for the parameter $\underline{\beta}$
$\underline{\beta}^c$	$q \times 1$ vector of the unknown fixed coefficients for the x variables (cross-sectional context)
$\hat{\underline{\beta}}^c$	cross-section estimator for the regression coefficient
$\overline{Y}_t$	finite population mean for the study variable at wave t
$\overline{y}_t$	weighted estimator for $\overline{Y}_t$
$\overline{y}_t^*$	unweighted estimator for $\overline{Y}_t$
sd_t	standard deviation of y considering wave t
$\sqrt{\text{var}[.]}$	an estimator the standard error, also denoted by s.e.
δ	contrast

$\hat{\delta}$	an unweighted estimator of the contrast
$\hat{\delta}_w$	a weighted estimator of the contrast
V	$T \times T$ working covariance matrix
g	Maximum considered lag covariance
R	$T \times T$ working correlation matrix
I	$T \times T$ identity matrix
Σ	$T \times T$ model variance-covariance matrix
$\Sigma(\underline{\theta})$	$T \times T$ model variance-covariance matrix constrained to be function of $\underline{\theta}$
$\underline{\theta}$	generic target model parameter or a $1 \times b$ parameter vector of interest
$\hat{\underline{\theta}}$	an estimator of the parameter of interest $\underline{\theta}$
k	number of distinct elements of Σ
S_N	$T \times T$ finite population covariance matrix
S_w	$T \times T$ weighted sample covariance matrix
S	$T \times T$ unweighted sample covariance matrix
W	any consistent estimator of Σ
$\omega_{tt'}$	typical element of W
U	positive definite weight matrix
$u_{tt',t''t'''}$	typical element of U
$m_{tt',t''t'''}$	forth-order moments about the mean
K	transition matrix
F	fitting function
Θ	parameter space
E_p	population residual covariance matrix
E	residual covariance matrix
E_c	weighted residual covariance matrix
VAR(.)	model variance
COV(.)	model covariance
aCOV(.)	asymptotic covariance

CORR(.)	model correlation
$\mathrm{var}_N(.)$	population level estimator of the variance
$\mathrm{cov}_N(.)$	population level estimator of the covariance
var(.)	estimator of the variance
$\mathrm{var}_n(.)$	'naïve' variance estimator
$\mathrm{var}_r(.)$	robust variance estimator
$\mathrm{var}_J(.)$	jackknife variance estimator
$\mathrm{var}_L(.)$	linearization variance estimator
$\mathrm{var}_{df}(.)$	distribution free variance estimator
$\mathrm{var}_c(.)$	distribution free variance estimator that accounts for the sampling design
GEE	generalised estimating equation
LM	linear regression cross-sectional model
AR1	stochastic first-order autoregressive process
UCM	uniform correlation model
ADF	asymptotically distribution-free
iid	independent and identically distributed
N_T	T-dimensional multivariate normal distribution
N_k	k-dimensional multivariate normal distribution
W_T	T-dimensional multivariate Wishart distribution
t	t distribution
t*	t distribution with non-standardised variance
$\ell_N[.]$	census likelihood function
$L_N[.]$	logarithmic census likelihood
ML	maximum likelihood
PML	pseudo maximum likelihood
OLS	ordinary least squares
ULS	unweighted least squares
ULSC	unweighted least squares for complex survey data
GLS	generalised least squares

GLSC	generalised least squares for complex survey data
$E(.)$	model expectation
$E_p(.)$	expectation with respect to the sampling distribution of statistics over repeated samples s generated by the sampling design $p(s)$
$p\lim$	probability limit of
$vech[.]$	a vector formed from the nonduplicated elements of a matrix
$vec[.]$	a vector obtained by stacking the columns of a matrix
$tr(.)$	trace of a matrix
$\lvert . \rvert$	determinant of a matrix
$\otimes$	right Kronecker product
$Deff_{Kish}(.)$	Kish's design effect
$\mathrm{VAR}_{true}(.)$	the *true* variance that considers the *true* sampling scheme used for the selection of the sample
$\mathrm{VAR}_{srs}(.)$	hypothetical variance when considering that the sample was selected by srs with replacement
$\mathrm{var}_0(.)$	a consistent estimator of the variance, when we assume that the observations are iid
$meff(.)$	misspecification effect
RMR	Root mean-square residual
LRT	likelihood ratio test
LMT	lagrangian multiplier test
WT	Wald test
GFI	goodness of fit index
AGFI	adjusted goodness of fit index
MSE(.)	mean square error of an estimator
cv	coefficient of variation
D	number of simulated replicates

References

Allison, P. D. (1984) *Event History Analysis – Regression for Longitudinal Event Data.* Newbury Park: Sage.

Amemiya, Y. and Anderson, T. W. (1990) Asymptotic Chi-square Tests for a Large Class of Factor Analysis Models. *The Annals of Statistics*, Vol. 18, N. 3, 1453-1463.

Anderson, T. W. (1973) Asymptotically Efficient Estimation of Covariance Matrices with Linear Structure. *Annals of Statistics*, Vol. 1, N. 1, 135-141.

Baltagi, B. H. (2001) *Econometric Analysis of Panel Data.* 2 ed., Chichester, John Wiley & Sons.

Barnett, V. (1991) *Sample Survey Principles & Methods.* London, Edward Arnold.

Belisle, C. J. P. (1992) Convergence theorems for a class of simulated annealing algorithms on Rd. *Journal of Applied Probability*, Vol. 29, 885-895.

Bentler, P. M. and Dudgeon, P. (1996) Covariance Structure Analysis: Statistical Practice, Theory, and Directions. *Annual Review of Psychology*, N. 47, 563-592.

Bentler, P. M. and Weeks, D. G. (1980) Linear Structural Equations with Latent Variables. *Psychometrika*, Vol. 45, N. 3, 289-308.

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.

Binder, D. A. (1995) Taylor Linearization for Single Phase and Two Phase Samples: A Cookbook Approach. *Proceedings of the Survey Research Methods Section*, American Statistical Association.

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

Boomsma, A. (1985) Nonconvergence, Improper Solutions, and Starting Values in LISREL Maximum Likelihood Estimation. *Psychometrika*, Vol. 50, N. 2, 229-242.

Brooks, C. and Bolzendahl, C. (2004) The transformation of US gender role attitudes: cohort replacement, social-structural change, and ideological learning. *Social Science Research*, Vol 33, 106-133.

Browne, M. W. (1977) Generalized Least-Squares Estimators in the Analysis of Covariance Structures. In Aigner, D. J. and Goldberger, A. S. eds. *Latent Variables in Socio-Economic Models*. Amsterdam, North-Holland.

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.

Browne, M. W. (1987) Robustness of Statistical Inference in Factor Analysis and Related Models. *Biometrika*, Vol. 74, N. 2, 375-384.

Bronson, R. (1991) *Matrix Methods: An Introduction*. 2nd. Ed. San Diego, Academic Press.

Buse, A. (1982) The Likelihood Ratio, Wald, and Lagrange Multiplier Tests: An Expository Note. *The American Statistician*, Vol. 36, N. 3, 153-157.

Byrd, R. H., Lu, P., Nocedal, J. and Zhu, C. (1995) A limited memory algorithm for bound constrained optimization. Journal of Scientific Computing, Vol. 16, 1190-1208.

Chambers, R. L., Dorfman, A. H. and Sverchkov, M. Yu. (2003) Nonparametric Regression with Complex Survey Data. In Chambers, R. L. and Skinner, C. J. eds. *Analysis of Survey Data*. Chichester, John Wiley & Sons.

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

Chou, C-P, Bentler, P. M. and Satorra, A. (1991) Scaled Test Statistic and Robust Standard Errors for Non-normal Data in Covariance Structure Analysis: A Monte Carlo Study. *British Journal of Mathematical and Statistical Psychology*, Vol. 44, 347-357.

Cochran, W. G. (1977) *Sampling Techniques*. 3rd. Ed. New York, John Wiley & Sons.

Courtenay, G. (1997) Youth Cohort Study of England and Wales. 2nd edition. Colchester, Essex: UK Data Archive, SN: 3093.

Crowder, M. J. and Hand, D. J. (1990) *Analysis of Repeated Measures*. London, Chapman and Hall.

Dahm, P. F., Melton, B. E. and Fuller, W. A. (1983) Generalized Least Squares Estimation of a genotypic Covariance Matrix. *Biometrics*, Vol. 39, N. 3, 587-597.

Dalgaard, P. (2002) *Introductory Statistics with R*. New York, Springer.

Dancan, G. (2000) Using Panel Studies to Understand Household Behaviour and Well-being. In Rose, D. ed. *Researching Social and Economic Change: the Uses of Household Panel Studies*. London, Routledge.

David, F. N. (1949) The Moments of the z and F Distributions. *Biometrika*, Vol. 36, N. 3/4, 394-403.

Deming, W. E. (1950) *Some Theory of Sampling*. New York, Dover Publications.

Dennis, J. E. and Schnabel, R. B. (1983) *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Englewood Cliffs, NJ, Prentice-Hall.

Diggle, P. J., Liang, K. & Zeger, S. L. (1996) *Analysis of Longitudinal Data*. Oxford: Oxford University Press.

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

Durbin, J. (1959) A Note on the Application of Quenouille's Method of Bias Reduction to the Estimation of Ratios. *Biometrika*, Vol. 46, N. 3/4, 477-480.

Eltinge, J. L. (1999) *Assessment of Information Capacity and Sensitivity in the Analysis of Complex Surveys*. Bulletin of the International Statistical Institute. 52nd Session Proceedings.

Ermisch, J. (2000) Using Panel Data to Analyse Household and Family Dynamics. In Rose, D. ed. *Researching Social and Economic Change: the Uses of Household Panel Studies*. London, Routledge.

Fan, P. -L. and Marini, M. M. (2000) Influences on gender-role attitudes during the transition to adulthood. Social Science Research, Vol. 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*, Vol. 26, N. 1, 53-65.

Finkel, S. E. (1995) *Causal Analysis with Panel Data*. Newbury Park: Sage.

Firebaugh, G. (1997) *Analysing Repeated Surveys*. Newbury Park: Sage.

Fletcher, R. and Reeves, C. M. (1964) Function minimization by conjugate gradients. *Computer Journal*. Vol. 7, 148-154.

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

Fuller, W. A. (1987) *Measurement Error Models*. New York, John Wiley & Sons.

Ghosh, M. (1996) Wishart Distribution via Induction. *The American Statistician*, Vol. 50, N. 3, 243- 246.

Goldstein, H. (1995) *Multilevel Statistical Models*. 2nd ed., London, Arnold.

Hand, D., and Crowder, M. (1996) *Practical Longitudinal Data Analysis*. London, Chapman & Hall.

Hansen, M. H., Hurwitz, W. N. and Madow, W. G. (1953) *Sample Survey Methods and Theory*. Vol. 2, New York, Wiley.

Hardin, J. and Hilbe, J. (2001) *Generalized Linear Models and Extensions*. College Station, Stata.

Harville, D. A. (1977) Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems. *Journal of the American Statistical Association*, Vol. 72, n. 358, 320-338.

Harville, D. A. (1997) *Matrix Algebra From a Statistician's Perspective*. New York, Springer.

Hattersley, L. and Creeser, R. (1995) *LS Series 7: Longitudinal Study 1971-1991. History, organisation and quality of data*. London, HMSO.

Hocking, R. R. (1985) *The Analysis of Linear Models*. Monterey, Brooks/Cole.

Hoelter, J. W. (1983) The Analysis of Covariance Structures: Goodness of fit indices. *Sociological Methods and Research*. Vol. 11, 325-344.

Horton, N. J. & Lipsitz, S. R. (1999) Review of Software to Fit Generalized Estimating Equation Regression Models. *The American Statistician*. Vol. 53, 160-169.

Horvitz, D. G. and Thompson, D. J. (1952) A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, Vol. 47, n. 260, 663-685.

Hox, J. J. and Maas, C. J. M. (2001) The Accuracy of Multilevel Structural Equation Modelling with Pseudobalanced Groups and Small Samples. *Structural Equation Modelling*, Vol. 8, n. 2, 157-174.

Ihaka, R. and Gentleman, R. (1996) R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics*, Vol. 5, n. 3, 299-314.

Isaki, C. T. and Fuller, W. A. (1982) Survey Design Under the Regression Superpopulation Model. *Journal of the American Statistical Association*. Vol. 77, n. 377, 89-96.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1995) *Continuous Univariate Distributions*. Vol. 2, 2nd ed., New York, John Wiley & Sons.

Johnson, R. A. and Wichern, D. W. (1998) *Applied Multivariate Statistical Analysis*. 4th ed., Upper Saddle River, Prentice-Hall.

Jones, H. L. (1974) Jackknife Estimation of Function of Stratum Means. *Biometrika*, Vol. 61, N. 2, 343-348.

Jones, R. H. (1993) *Longitudinal Data with Serial Correlation: A State-space Approach*. London, Chapman and Hall.

Jöreskog, K. G. (1970) A General Method for Analysis of Covariance Structures. *Biometrika*, Vol. 57, N. 2, 239-251.

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

Jöreskog, K. G. and Goldberger, A. S. (1975) Estimation of a Model with Multiple Indicators and Multiple Causes of a Single Latent Variable. *Journal of the American Statistical Association*, 70, 631-639.

Jöreskog, K. G. and Sörbom, D. (1989) *Lisrel 7: A Guide to the Program and Applications.* Chicago, SPSS publications.

Jöreskog, K. G. and Sörbom, D. (1997) *Lisrel 8: User's Reference Guide.* Lincolnwood, Scientific Software International.

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.

Kalton, G. and Citro, C. F. (2000) Panel Surveys – Adding the Fourth Dimention. In Rose, D. ed. *Researching Social and Economic Change: the Uses of Household Panel Studies.* London, Routledge.

Kano, Y., Berkane, M. and Bentler, P. M. (1990) Covariance structure analysis with heterogeneous kurtosis parameters. *Biometrika*, Vol. 77, n. 3, 575-585.

Kauermann, G. and Carroll, R. J. (2001) A Note on the Efficiency of Sandwich Covariance Matrix Estimation. *Journal of the American Statistical Association*, Vol. 96, N. 456, 1387-1396.

Kish, L. (1957) Confidence Intervals for Clustered Samples. *American Sociological Review*, Vol. 22, N. 2, 154-165.

Kish, L. (1965) *Survey Sampling*. New York, Wiley.

Kish, L. (1980) Design and Estimation for Domains. *The Statistician*, Vol. 29, N. 4, 209-222.

Kish, L. and Frankel, M. R. (1974) Inference from Complex Samples. *Journal of the Royal Statistical Society – Series B*, Vol. 36, N. 1, 1-37.

Kmenta, J. (1971) *Elements of Econometrics*. New York, Macmillan.

Knight, A. (1978) Common Factor Analysis: Some Recent Developments in Theory and Practice. *The Statistician*, Vol. 27, N. 1, 27-42.

Koch, G. G., Freeman, Jr. D. H. and Freeman, J. L. (1975) Strategies in the Multivariate Analysis of Data from Complex Surveys. *International Statistical Review*, Vol. 43, N. 1, 59-78.

Korn, E. L. and Graubard, B. I. (1995) Analysis of Large Health Surveys: Accounting for the Sampling Design. *Journal of the Royal Statistical Society – Series A*, Vol. 158, N. 2, 263-295.

Kott, P. S. (1991) A Model-Based Look at Linear Regression with Survey Data. *The American Statistician*, Vol. 45, N. 2, 107-112.

Kott, P. S. (1994) Regression Analysis of Repeated Survey Data (with Available Software). *Proceedings of the Survey Research Methods Section*, American Statistical Association.

Krewski and Rao (1981) Inference from Stratified Samples: Properties of the Linearization, Jackknife and Balanced Repeated Replication Methods. *The Annals of Statistics*, Vol. 9, N. 5, 1010-1019.

Krzanowski, W. J. (2000) *Principles of Multivariate Analysis – A User's Perspective*. Oxford, Oxford University Press.

Lavange, L. M., Koch, G. G. and Schwartz, T. A. (2001) Applying sample survey methods to clinical trials data. *Statistics in Medicine*, Vol. 20, 2609-2623.

Lawless, J. F. (2003) Event History Analysis and Longitudinal Surveys. In Chambers, R. L. and Skinner, C. J. eds. *Analysis of Survey Data*. Chichester, John Wiley & Sons.

Lee, K-H. (1973) Variance Estimation in Stratified Sampling. *Journal of the American Statistical Association*, Vol. 68, N. 342, 336-342.

Lee, E. S., Forthofer, R. N. and Lorimor, R. J. (1989) *Analyzing Complex Survey Data*. London: Sage Publications.

Lee, S-Y (1985) On Testing Functional Constrains in Structural Equation Models. *Biometrika*, Vol. 8, 749-766.

Lee, S-Y. (1990) Multilevel Analysis of Structural Equation Models. *Biometrika*, Vol. 77, N. 4, 763-772.

Lee, S-Y and Poon, W-Y (1998) Analysis of two-level structural equation models via EM type algorithms. *Statistica Sinica*, Vol. 8, 749-766.

Lewis, P. A. W. (1989) *Simulation Methodology for Statisticians, Operations Analysts and Engineers, Volume 1*. Belmont, Wadsworth.

Liang, K. and Zeger, S. L. (1986) Longitudinal Data Analysis Using Generalised Linear Models. *Biometrika*, Vol. 73, n. 1, 13-22.

Lindsey, J. K. (1994) *Models for Repeated Measurements*. Oxford, Oxford University Press.

Loehlin, J. C. (1987) *Latent Variable Models – An Introduction to Factor, Path, and Structural Analysis*. Hillsdale, Lawrence Erlbaum Associates.

Long, J. S. (1983) *Covariance Structure Models – An Introduction to LISREL*. Newbury Park: Sage.

MacCallum, R. C. and Austin, J. T. (2000) Applications of Structural Equation Modeling in Psychological Research. *Annual Review of Psychology*, Vol. 51, 201-226.

Marmot, M., Banks, J., Blundell, R., Lessof, C. and Nazroo, J. (2003) *Health, Wealth and Lifesyles of the Older Population in England - The 2002 English Longitudinal Study of Ageing*. London, Institute for Fiscal Studies.

Matsueda, R. L. and Bielby, W. T. (1986) Statistical Power in Covariance Structure Models. *Sociological Methodology*, Vol. 16, 120-158.

McDonald, R. P. (1980) A Simple Comprehensive Model for the Analysis of Covariance Structures: Some Remarks on Applications. *British Journal of Mathematical and Statistical Psychology*, Vol. 33, 161-183.

Mealli, F. and Pudney, S. (2003) Applying Heterogeneous Transition Models in Labour Economics: the Role of Youth Training in Labour Market Transitions. In Chambers, R. L. and Skinner, C. J. eds. *Analysis of Survey Data*. Chichester, John Wiley & Sons.

Menard, S. (1991) *Longitudinal Research*. Newbury Park: Sage.

Menard, S. (2002) *Longitudinal Research*. 2nd ed. Newbury Park: Sage.

Miller, Jr. R. G. (1964) A Trustworthy Jackknife. *The Annals of Mathematical Statistics*, Vol. 35, n. 4, 1594-1605.

Miller, Jr. R. G. (1974) The Jackknife – A Review. *Biometrika*, Vol. 61, n. 1, 1-15.

Morgan, B. J. T. (1984) *Elements of Simulation*. London, Chapman and Hall.

Morgan, S. P. and Waite, L. J. (1987) Parenthood and the attitudes of young adults. *American Sociological Review*, Vol. 52, 541-547.

Muthén, B. O. (1997) Latent Variable Modelling of Longitudinal and Multilevel Data. *Sociological Methodology*, Vol. 27, 453-480.

Muthén, L. K. and Muthén, B. O. (2005) *Mplus User's Guide Excerps*. Los Angeles, CA, Muthén & Muthén.

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

Nascimento Silva, P. L. (1996) *Utilizing Auxiliary Information in Sample Survey Estimation and Analysis*. PhD Thesis, Southampton, University of Southampton, Department of Social Statistics.

Nelder, J. A. and Mead, R. (1965) A simplex algorithm for function minimization. *Computer Journal*, Vol. 7, 308-313.

Nocedal, J. and Wright, S. J. (1999) *Numerical Optimization*. New York, Springer.

Ogasawara, H. (2005) Asymptotic Robustness of the Asymptotic Biases in Structural Equation Modeling. *Computational Statistics and Data Analysis*, Vol. 49, 771-783.

Olsson, U. H., Foss, T. and Troye, S. V. (2003) Does the ADF Fit Function Decrease When the Kurtosis Increases? *British Journal of Mathematical and Statistical Psychology*, Vol. 56, 289-303.

Pan, J. and Mackenzie, G. (2003) On Modelling Mean-covariance Structures in Longitudinal Studies. *Biometrika*, Vol. 90, n. 1, 239-244.

Patterson, H. D. and Thompson, R. (1971) Recovery of Inter-Block Information when Block Sizes are Unequal. *Biometrika*, Vol. 58, n. 3, 545-554.

Pfeffermann, D. and LaVange, L. (1989) Regression Models for Stratified Multi-Stage Cluster Samples. In Skinner, C. J., Holt, D. and Smith, T. M. F. eds. *Analysis of Complex Surveys*. Chichester, John Wiley & Sons.

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

Pourahmadi, M. (2000) Maximum Likelihood Estimation of Generalised Linear Models for Multivariate Normal Covariance Matrix. *Biometrika*, Vol. 87, n. 2, 425-435.

Press, S. J. (1972) *Applied Multivariate Analysis*. New York, Holt, Rinehart and Winston, Inc.

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

Raj, D. (1968) *Sampling Theory*. New York, McGraw Hill.

Rao, J. N. K. and Scott (1979) Chi-squared Tests for Analysis of Categorical Data from Complex Surveys. *Proceedings of the Survey Research Methods Section*, American Statistical Association.

Rao, J. N. K. and Scott (1981) The Analysis of Categorical Data From Complex Sample Surveys Maximum Likelihood Estimation of Generalised Linear Models for Multivariate Normal Covariance Matrix. *Biometrika*, Vol. 87, n. 2, 425-435.

Rao, J. N. K. and Scott (1984) On Chi-squared Tests for Multiway Contingency Tables with Cell Proportions Estimated from Survey Data. *The Annals of Statistics*, Vol. 12, n. 1, 46-60.

Rao, J. N. K. and Scott (1987) On Simple Adjustments to Chi-squared Tests with Sample Survey Data. *The Annals of Statistics*, Vol. 15, n. 1, 385-397.

Rao, J. N. K. and Thomas, D. R. (1988) The Analysis of Cross-Classified Categorical Data from Complex Sample Surveys. *Sociological Methodology*, Vol. 18, 213-269.

Rao, J. N. K. and Thomas, D. R. (1989) Chi-Squared Tests for Contingency Tables. In Skinner, C. J., Holt, D. and Smith, T. M. F. eds. *Analysis of Complex Surveys*. Chichester, John Wiley & Sons.

Rao, J. N. K. and Thomas, D. R. (2003) Analysis of Categorical Response Data from Complex Surveys: An Appraisal and Upgrade. In Chambers, R. L. and Skinner, C. J. eds. *Analysis of Survey Data*. Chichester, John Wiley & Sons.

Rasbash, J., Browne, W., Goldstein, H., Yang, M., Plewis, I., Healy, M., Woodhouse, G., Draper, D., Langford, I., and Lewis, T. (2002) *A user's guide to MLwiN – Version 2.1d for use with MLwiN 1.10*. London, Centre for Multilevel Modelling, Institute of Education, University of London.

Raykov, T. and Marcoulides, G. A. (2000) *A First Course in Structural Equation Modelling*. Mahwah, Lawrence Erlbaum Associates.

Rencher, A. C. (1998) *Multivariate Statistical Inference and Applications*. New York, John Wiley & Sons.

Robinson, D. L. (1987) Estimation and Use of Variance Components. *Statistician*, Vol. 36, n. 1, 3-14.

Robinson, G. K. (1991) That BLUP is a Good Thing: The estimation of Random Effects. *Statistical Science*, Vol. 6, n. 1, 15-32.

Rose, D. (2000a) Household panel studies – An overview. In Rose, D. ed. *Researching Social and Economic Change: the Uses of Household Panel Studies*. London, Routledge.

Rose, D. ed. (2000b) *Researching Social and Economic Change: the Uses of Household Panel Studies*. London, Routledge.

Royall, R. M. and Cumberland, W. G. (1981) An empirical study of the ratio estimator and estimators of its variance. *Journal of the American Statistical Association*, Vol. 76, 66-88.

Rust, K. (1985) Variance Estimation for Complex Estimators in Sample Surveys. *Journal of Official Statistics*, Vol. 1, n. 4, p. 381-397, Statistics Sweden.

Särndal, C.-E., Swenson, B. and Wretman, J. (1992) *Model Assisted Survey Sampling*. New York, Springer.

SAS Institute (2004) *SAS/STAT User Guide*. Cary, NC, SAS Publishing.

Satorra, A. (1989) Alternative Test Criteria in Covariance Structure Analysis: A Unified Approach. *Psychometrika*, Vol. 54, N. 1, p. 131-151.

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

Satorra, A. (2000) Scaled and Adjusted Restricted Tests in Multi-sample Analysis of Moment Structures. In Heijmans, D. D. H., Pollock, D. S. G. and Satorra, A. eds. *Innovations in Multivariate Statistical Analysis: A Festschrift for Henz Neudecker*. Dordrecht, Kluwer Academic Publishers.

Satorra, A. (2002) Asymptotic Robustness in Multiple Group Linear-latent Variable Models. *Econometric Theory*, 18, 297-312.

Satorra, A. and Bentler, P. M. (1986) Some Robustness Properties of Goodness of Fit Statistics in Covariance Structure Analysis. *Proceedings of the Business and Economics Statistics Section*, p. 549-554, Alexandria, VA: American Statistical Association.

Satorra, A. and Bentler, P. M. (1988) Scaling Corrections for Chi-Square Statistics in Covariance Structure Analysis. *Proceedings of the Business and Economics Statistics Section*, p. 308-313, Alexandria, VA: American Statistical Association.

Satorra, A. and Bentler, P. M. (1994) Corrections to Test Statistics and Standard Errors in Covariance Structure Analysis. In Eye, A. von and Clogg, C. C. eds. *Latent Variables Analysis: Applications for Developmental Research*. Thousand Oaks, CA, Sage.

Satorra, A. and Bentler, P. M. (2001) A Scaled Difference Chi-square Test Statistic for Moment Structure Analysis. *Psychometrika*, Vol. 66, N. 4, 507-514.

Satorra, A. and Neudecker, H. (1997) Compact Matrix Expressions for Generalized Wald Tests of Equality of Moment Vectors. *Journal of Multivariate Analysis*, Vol. 63, 259-276.

Satorra, A. and Saris, W. E. (1985) Power of the Likelihood Ratio Test in Covariance Structure Analysis. *Psychometrika*, Vol. 50, N. 1, 83-90.

Schnabel, R. B., Koontz, J. E. and Weiss, B. E. (1985) A modular system of algorithms for unconstrained minimization. *ACM Trans. Math. Software*, 11, 419-440.

Schoenberg, R. (1989) Covariance Structure Models. *Annual Review of Sociology*, Vol. 15, 425-440.

Scott, A. J. and Holt, D. (1982) The effect of two-stage sampling on ordinary lest-squares methods. *Journal of the American Statistical Association*, Vol. 77, 848-854.

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.

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.

Shapiro, A. (1986) Asymptotic Theory of Overparameterized Structural Models. *Journal of the American Statistical Association*, Vol. 81, N. 393, 142-149.

Shapiro, A., and Browne, M. (1987) Analysis of Covariance Structures under Elliptical Distributions. *Journal of the American Statistical Association*, Vol. 82, N. 400, 1092-97.

Shapiro, S. S., and Wilk, M. B. (1965) An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, Vol. 52, N. 3/4, 591-611.

Silvey, S. D. (1959) The Lagrangian Multiplier Test. *The Annals of Mathematical Statistics*, Vol. 30, N. 2, 389-407.

Skinner, C. J. (1986) Design Effects of Two-Stage Sampling. *Journal of the Royal Statistical Society – Series B*, Vol. 48, N. 1, 89-99.

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

Skinner, C. J. (1989b) Introduction to Part A. In Skinner, C. J., Holt, D. and Smith, T. M. F. eds. *Analysis of Complex Surveys*. Chichester, John Wiley & Sons.

Skinner, C. J. (2000) Dealing with measurement error in panel analysis. In Rose, D. ed. *Researching Social and Economic Change: the Uses of Household Panel Studies*. London, Routledge.

Skinner, C. J. (2003a) Introduction to Part B – Categorical Response Data. In Chambers, R. L. and Skinner, C. J. eds. *Analysis of Survey Data*. Chichester, John Wiley & Sons.

Skinner, C. J. (2003b) Introduction to Part D – Longitudinal Data. In Chambers, R. L. and Skinner, C. J. eds. *Analysis of Survey Data*. Chichester, John Wiley & Sons.

Skinner, C. J. & Coker, O. (1996) Regression Analysis for Complex Survey Data with Missing Values of a Covariance. *Journal of the Royal Statistical Society – Series A*, Vol. 159, Part 2, 265-274.

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, John Wiley & Sons.

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

Skinner, C. J. and Vieira, M. D. T. (2005) Design Effects in the Analysis of Longitudinal Survey Data. *S3RI Methodology Working Papers*, M05/13, 23pp, Southampton, Southampton Statistical Sciences Research Institute.

Smyth, G. K., and Verbyla, A. P. (1996). A conditional approach to residual maximum likelihood estimation in generalized linear models. *Journal of the Royal Statistical Society* B, 58, 565-572.

Specht, D. A. and Warren, R. D. (1976) Comparing Causal Models. *Sociological Methodology*, Vol. 7, 46-82.

SPSS Inc (2005) *SPSS Complex Samples 14.0 – Specifications*. Chicago, SPSS Inc.

Stata Corp. (2003) *Stata Statistical Software: Release 8.0 – Cross-sectional time-series*. College Station, TX: Stata Corporation, 207-210.

Sutradhar, B. C. and Kovacevic, M. (2000) Analysing ordinal longitudinal survey data: Generalised estimating equations approach. *Biometrika*, Vol. 87, N. 4, 837-848.

Swain, A. J. (1975) A Class of Factor Analysis Estimation Procedures with Common Asymptotic Sampling Properties. *Psychometrika*, Vol. 40, N. 3, 315-335.

Tanaka, J. S., and Huba, G. J. (1985) A fit index for covariance structure models under arbitrary GLS estimation. *British Journal of Mathematical and Statistical Psychology*, Vol. 38, 197-201.

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.

Teachman, J., Duncan, G. J., Yeung, W. J., and Levy, D. (2001) Covariance Structure Models for Fixed and Random Effects. *Sociological Methods and Research*, Vol. 30, N. 2, 271-288.

Wald, A. (1943) Tests of Statistical Hypothesis Concerning Several Parameters when the Number of Observations is Large. *Transactions of the American Mathematical Society*, Vol. 54, N. 3, 426-482.

Welch, B. L. (1937) The significance of the difference between two means when the population variances are unequal. *Biometrika*, 29, 350-362.

White, H. (1980) A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, Vol. 48, N. 4, 817-838.

Wiley, D. E., Schmidt, W. H. and Bramble, W. J. (1973) Studies of a Class of Covariance Structure Models. *Journal of the American Statistical Association*, Vol. 68, N. 342, 317-323.

Wilks, S. S. (1938) The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypothesis. *The Annals of Mathematical Statistics*, Vol. 9, N. 1, 60-62.

Winkels, J. W. and Withers, S. D. (2000) Panel Attrition. In Rose, D. ed. *Researching Social and Economic Change: the Uses of Household Panel Studies*. London, Routledge.

Wolter, K. M. (1985) *Introduction to Variance Estimation*. New York, Springer.

Woodruff, R. S. (1971) A Simple Method for Approximating the Variance of a Complicated Estimate. *Journal of the American Statistical Association*, Vol. 66, N. 334, 411-414.

Yanagihara, H., Tonda, T. and Matsumoto, C. (2005) The Effects of Non-normality on Asymptotic Distributions of some Likelihood Ratio Test Criteria for Testing Covariance Structures Under Normal Assumption. *Journal of Multivariate Analysis*, too appear.

Yuan, K-H. and Bentler, P. M. (1997a) Mean and Covariance Structure Analysis: Theoretical and Practical Improvements. *Journal of the American Statistical Association*, Vol. 92, N. 438, 767-774.

Yu, W. (2002) Design Effect Variation for Estimates Derived from the MEPS (1996-1998), *2002 Proceedings of the American Statistical Association*, Alexandria, VA: American Statistical Association: 3893-3898.

Yuan, K-H. and Bentler, P. M. (1997b) Improving Parameter Tests in Covariance Structure Analysis. *Computational Statistics and Data Analysis*, Vol. 26, 177-198.

Yuan, K-H. and Bentler, P. M. (1998) Structural Equation Modeling with Robust Covariances. *Sociological Methodology*, Vol. 28, 363-396.

Yuan, K-H. and Chan, W. (2002) Fitting Structural Equation Models Using Estimating Equations: A Model Segregation Approach. *British Journal of Mathematical and Statistical Psychology*, Vol. 55, 41-62.

Yuan, K-H. and Hayashi, K. (2005) On Muthén's Maximum Likelihood for Two-level Covariance Structure Models. *Psychometrika*, Vol. 70, N. 1, 1-21.

Zeger, S. L. and Liang, K. (1986) Longitudinal Data Analysis for Discrete and Continuous Outcomes. *Biometrics*, Vol. 42, n. 1, 121-130.

Zeger, S. L., Liang, K. and Albert, P. S. (1988) Models for Longitudinal Data: A Generalized Estimating Equation Approach. *Biometrika*, Volume 44, Issue 4, 1049-1060.

Zhang, F. (2001) An Empirical Study of the Limitation of Using SUDAAN for Variance Estimation. In Zhang, F, Weng, S, Salvucci, S., and Hu, M-X. eds. *A Study of Variance Estimation Methods*. Washington, National Center for Education Statistics, Working Paper N. 2001-18.