CORRECTED SCORE ESTIMATORS IN MULTIVARIATE REGRESSION MODELS WITH HETEROSCEDASTIC MEASUREMENT ERRORS

Wannaporn Junthopas

A Dissertation Submitted in Partial
Fulfillment of the Requirements for the Degree of
Doctor of Philosophy (Statistics)
School of Applied Statistics
National Institute of Development Administration
2016

CORRECTED SCORE ESTIMATORS IN MULTIVARIATE REGRESSION MODELS WITH HETEROSCEDASTIC MEASUREMENT ERRORS

Wannaporn Junthopas

School of Applied Statistics

Associate Professor .... Major Advisor
(Jirawan Jitthavech, Ph.D.)

Associate Professor .... Co-Advisor
(Vichit Lorchirachoonkul, Ph.D.)

The Examining Committee Approved This Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor Philosophy (Statistics).

Associate Professor .... Committee Chairperson
(Montip Tiensuwan, Ph.D.)

Professor .... Committee
(Samruam Chongcharoen, Ph.D.)

Associate Professor .... Committee
(Jirawan Jitthavech, Ph.D.)

Associate Professor .... Committee
(Vichit Lorchirachoonkul, Ph.D.)

Assistant Professor .... Dean
(Sutep Tongngam, Ph.D.)

March 2017

ABSTRACT

Title of Dissertation Corrected Score Estimators in Multivariate Regression Models with Heteroscedastic Measurement Errors

Author Miss Wannaporn Junthopas

Degree Doctor of Philosophy (Statistics)

Year 2016

In this study, the knowledge of parameter estimation theory based on the corrected score (CS) approach is extended in a linear multivariate multiple regression model with heteroscedastic measurement errors (HME) and an unknown HME variance. The heteroscedasticity of the HME variance is assumed to be capable of being grouped into similar patterns where the sample of observations are assembled into several sub-samples with the property that the variances of the measurement error (ME) are homoscedastic within a group but heteroscedastic between groups. In each group, the variance of the ME of the surrogate variable is estimated by the pooled variance of the variable with HMEs observed in repeated measurements.

The statistical properties of the proposed CS estimator are analytically investigated based on the specific model in which there are two independent variables of which one is measured with HME. To evaluate the performance of the proposed CS estimator via a simulation study, datasets are generated based on two forms of heteroscedasticity: the step-up function form and the step-down function form. From the simulation results, the ordinary least squares (OLS) estimation of the parameters of the precisely observed variable is unaffected by HME, but the parameter estimators of the variable measured with HME are underestimated. The CS method outperforms the OLS method since the absolute bias and mean square error of the CS estimator are less than those of the OLS estimator when either the number of repeated measurements or the sample size increases, and the bias of the CS estimator approaches zero when the sample size increases. The results of the simulation study show conformance to the theoretical proof.

ACKNOWLEDGEMENTS

This dissertation would not have been completed without the help and support of several people and my organization. I am sincerely grateful to my advisors, Associate Professor Dr. Jirawan Jitthavech and Associate Professor Dr. Vichit Lorchirachoonkul, who gave their valuable time to advise and train me as well as moderate this research and, along those lines, improved this manuscript significantly. Besides my advisors, I would also like to thank the rest of my dissertation committee, Professor Dr. Samruam Chongcharoen and Associate Professor Dr. Montip Tiensuwan, for helpful comments and suggestions.

I am also immensely grateful to Khon Kaen University and the National Science and Technology Development Agency for supporting and funding this research. My most sincere thanks are extended to Assistant Professor Pramote Krongyuth who provided me with the opportunity to join this Ph.D study, and Instructor Yupaporn Tongprasit who helped by providing the documents and support for joining the study. I would also like to thank Mr. Somchai Injhorhor who controlled and looked after the fund.

Finally, I must express a big thanks to my family and friends for their encouragement, help, and great love throughout the period of this study.

Wannaporn Junthopas

March 2017

TABLE OF CONTENTS

LIST OF TABLES

Tables **Page**

LIST OF FIGURES

Figures **Page**

CHAPTER 1

INTRODUCTION

1.1 Background

In statistical analysis, the process of gathering and measuring data on variables is very important and researchers require that variables are measured without errors. However, in practice, it is almost always impossible for a variable to be observed precisely. Imprecise measurements violate the ordinary least squares (OLS) assumption that the independent variables and random errors are independent, with the OLS assumption no longer being true. Consequently, this problem causes estimators to no longer be of value and leads to the wrong conclusions in statistical inference. The study of affecting and correcting of measurement error must be importance. The ME problem can occur in either dependent variables (Y) or independent variables (Z); the coefficients and variances of OLS estimators based on the ME in Y are still unbiased whereas those based on the ME in Z are biased and are also inconsistent. Consequently, the ME Problems in Z are more complicated than in Y (Gujarati, 2006: 346).

The heteroscedastic measurement error (HME) model is a statistical model whose variables have been measured with errors, and the variances of the MEs change across observations. The focus of a model with HME is on the process of gathering and measuring data in variables because, In some environments, the precise measurement of a specific variable is impracticable or very expensive in terms of time and effort and, furthermore, the error variances of this variable across observations may not be static. HME models have been widely applied in epidemiology, analytical chemistry, and botany, as can be seen in several studies (Kulathinal, Kuulasmaa and Gasbarra, 2002; Cheng and Riu, 2006; De Castro, Augustin, Döring, and Rummel, 2008; Galea and Bolfarine, 2008; Veenendaal, Mantlana, Pammenter, Weber, Huntsman-Mapila and Lloyd, 2008; Patriota, Bolfarine and De Castro, 2009).

As mentioned above, the ME in Z is a more serious problem than the ME in Y. The presence of HMEs also causes biased and inconsistent parameter estimates and leads to wrong conclusions in statistical inference. For this study, the HME model in Z is of interest. In the case of either measurement error (ME) or HME models, the methods used to correct the bias of the estimators can be grouped into either functional modeling or structural modeling. Several methods based on functional modeling can be applied, such as regression calibration, simulation-extrapolation, conditional score, corrected score (CS), and instrumental variables. In linear functional modelling, estimators from these methods have been shown to be asymptotically consistent (Buzas, Stefanski, and Tosteson, 2005).

One of the methods providing an efficient estimator is the corrected score (CS) approach, which was introduced by Nakamura (1990). This method deals with parameter estimation in the presence of ME in an independent variable based on estimation equations by looking for the biased correction term to correct the biased score estimation function, i.e. by finding $U^*(\theta, X, Y)$ such that $E\left(U^*(\theta, X, Y)\right) = U(\theta, Z, Y)$, where $U^*(\theta, X, Y)$ is the observable score function of the independent variables and $U(\theta, Z, Y)$ is the unobservable score function of the independent variables. By using the CS approach, Nakamura (1990) proposed a CS function for four models: the generalized linear model, the normal regression model, the Poisson regression model, and the gamma regression model. After that, the CS approach became the focus of attention in the literature. The proof of an asymptotic distribution of the CS estimators was presented and its application described by Giménez and Bolfarine (1997) in a simple linear regression and a comparative calibration model. In a comparison of the four approaches for consistent estimators, a number of methods have been used: sufficiency and conditional scores, maximum likelihood estimation, CS functions, and moment estimators. The results showed that, for small to moderate sample sizes, there is no one estimator more efficient than the others (Giménez and Bolfarine, 2000). Additionally, a CS estimator in a comparative calibration model with unknown variance of ME (Giménez and Patat, 2005), a CS estimator in one of several simple linear regression models with HME (De Castro, Bolfarine and Castilho, 2006), and a CS estimator in a heteroscedastic comparative calibration model (Giménez and Galea, 2013; Giménez and Patat, 2014) have been reported. However, most of the literature assumes that the ME variance and/or HME variance is known, except in the paper by Giménez and Patat (2005), in which a simple method for estimating the unknown ME variance was presented. The model is assumed that the providing of repeated measurements can be used to adapt in parameter estimation. Subsequently, the unknown ME variance could be estimated by using pooled variance. The assumption of known ME variance or HME variance is commonly applied in studies of parameter estimation in a model with only one independent variable (de Castro, Bolfarine and Castilho, 2006; Giménez and Galea, 2013; Chen, Hanfelt and Huang, 2015). However, Giménez and Patat (2005, 2014) proposed a method for estimating the parameter in a comparative calibration model under unknown ME variance condition.

HME could occur in several types of model, such as a linear regression model or a comparative calibration model. Likewise, a linear multivariate regression model could be in danger from HME. Therefore, this study is aimed at measuring HME errors in the latter type of model by using the CS approach.

There are three types of data susceptible to MEs: validation data, replication data, and instrumental data (Carroll, Ruppert, Stefanski and Crainiceanu, 2006: 33). The first type refers to data if variable Z is observable directly but subject to MEs and is referred to as an imprecise measurements variable. The second type, based on the assumption that the mean of replicated measurements can be reduced the variation from single measurement of Z. The last type refers to data of the second measurement variable being observed instead of data of the imprecise measurements variable.

In this study, an estimation approach based on the CS in a linear multivariate regression model with HME is proposed. An imprecise independent variable in the HME model and using replicated data is of interest. The assumption is that the HME variance is unknown and can be estimated based on grouped heteroscedasticity when replicated data are provided. This process is not complicated and can be obtained from the necessary information used to estimate the variance of the HMEs when it is assumed to be unknown.

1.2 Objectives of the Study

1) To derive the CS estimators of parameters in linear multivariate multiple regression models with HMEs.

2) To investigate the properties of the proposed estimators.

1.3 Scope of the Study

The proposed estimators are derived based on a linear multivariate multiple regression model with HMEs under the following scope:

1) The data are assumed to be multivariate normally distributed.

2) There are s independent variables in the model of which $s - s_1$ have been measured with errors in the form of additive HME.

3) The variance-covariance matrix of the random error is assumed to be known and the variance of the heterogeneous random MEs is assumed to be unknown.

4) The random errors and the heterogeneous random MEs in 3) are assumed to be mutually independent.

1.4 Definitions

1.4.1 HME Models

An HME model is a statistical model whose variables have been measured with errors and the variances of the MEs change across observations (the covariance matrix of the MEs has non-constant variances). An example of an HME model using simple linear regression is shown as follows:

$$y = \beta_0 + \beta_1 z + \varepsilon , \tag{1.1}$$

$$x = z + u , \tag{1.2}$$

Under the assumptions $\varepsilon \sim N\left(0,\sigma_{\varepsilon}^{2}\right)$,$u \sim N\left(0,\sigma_{uj}^{2}\right)$, and that Z , X , and U are uncorrelated, the independent variable Z is imprecisely observed. Assume that the observable variable X is measured by Z with error U . When U is normally distributed with mean zero and the error structure of U has a non-constant variance σ_{uj}^{2} , this denotes an HME model, and U is called a random HME. The covariance matrix of U can be expressed as:

$$E\left(\mathbf{uu}'\right)=\begin{pmatrix} \sigma_{u1}^{2} & 0 & \cdots & 0 \\ 0 & \sigma_{u2}^{2} & \cdots & 0 \\ & & \vdots & \\ 0 & 0 & \cdots & \sigma_{un}^{2} \end{pmatrix}. \tag{1.3}$$

1.4.2 Additive Structure of HME

Random HME, when added to an imprecise independent variable, is referred to as an additive structure of HME. For example, equation (1.2) has a surrogate observable variable X instead of $Z+U$, which gives an additive structure of HME.

1.4.3 Grouped Heteroscedasticity

The variance of MEs for observations is the same within groups of the data points but differs across the groups. Assume that there are g groups of data points where $g=1,2,...h$. The first group $(g=1)$ consists of n_1 observations, the second group $(g=2)$ consists of n_2 observations, until the last group $(g=h)$ consists of n_h observations. The variance of group $g=1,2,...h$ are $\sigma_{u1}^{2},\sigma_{u2}^{2},...,\sigma_{uh}^{2}$, respectively. For example, when the covariance matrix of U is grouped heteroscedastic, equation (1.3) becomes

$$E\left(\mathbf{uu}'\right)=\begin{pmatrix} \sigma_{u1}^2 & & & & & & & & \\ & \ddots & & & & & & & \\ & & \sigma_{u1}^2 & & & & \underline{0} & & \\ & & & \sigma_{u2}^2 & & & & & \\ & & & & \ddots & & & & \\ & & & & & \sigma_{u2}^2 & & & \\ & & \underline{0} & & & & \ddots & & \\ & & & & & & & \sigma_{uh}^2 & & \\ & & & & & & & & \ddots & \\ & & & & & & & & & \sigma_{uh}^2 \end{pmatrix} \begin{matrix} \}n_1 \\ \\ \}n_2 \\ \vdots \\ \}n_g \end{matrix} . \tag{1.4}$$

1.4.1 The CS Approach

The CS approach is a method for estimating the parameters in an ME or HME model. This technique deals with the parameter estimation in the absence of ME or HME in an independent variable based on estimation equations. The basic idea is based on extracting the bias correction term to correct the biased score estimation function to evaluate $U_c\left(\theta, X, Y\right)$ in

$$E\left(U_c\left(\theta, X, Y\right)\right)=U\left(\theta, Z, Y\right) , \tag{1.5}$$

where $U_c\left(\theta, X, Y\right)$ is the observable corrected score function of independent variables instead of the imprecisely observable Z and $U\left(\theta, Z, Y\right)$ is the imprecisely observable score function of independent variables.

1.4.1 The Bias of the Estimator

The bias of an estimator is the difference between the expectation of the estimator and the true value of the parameter being estimated. In an ideal scenario, an estimator with a small bias is more appropriate than one with a large bias. If the bias is zero, the estimator is referred to as unbiased.

Let $\hat{\boldsymbol{\beta}}$ be the vector of estimators for the vector of parameters $\boldsymbol{\beta}$ in the model, The bias of the vector of the estimators $\hat{\boldsymbol{\beta}}$ is given by

$$Bias\,of\ \hat{\boldsymbol{\beta}} = E\left(\hat{\boldsymbol{\beta}}\right) - \boldsymbol{\beta}\ . \tag{1.6}$$

1.4.1 The Mean Square Error of the Estimator

The mean square error (MSE) of an estimator is equal to the sum of the variance and the squared bias. It is a well-known performance measure for an estimator where a small MSE is more appropriate than a large one. If the bias of an estimator is zero, the MSE of the estimator is equal to its variance.

The MSE of the vector of the estimators $\hat{\boldsymbol{\beta}}$ is given by

$$MSE\left(\hat{\boldsymbol{\beta}}\right) = E\left[\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)'\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)\right] ,$$

$$= tr\left(\operatorname{var}\left(\hat{\boldsymbol{\beta}}\right)\right) + \left[E\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)\right]'\left[E\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)\right]. \tag{1.7}$$

CHAPTER 2

LITERATURE REVIEW

This chapter is organized as follows. ME and HME models are reviewed in Section 2.1, and the effects of and methods to correct ME and HME problems are outlined in Section 2.2. Finally, details of linear multivariate multiple regression models are introduced in Section 2.3.

2.1 MEs and HMEs

The problems of ME have been concentrated upon for a long time and MEs occur whenever a variable in the model of the study cannot be accurately observed. Incidences of this problem happen for many reasons and can cause a specific variable to become impracticable or very expensive in terms of time and effort to elucidate, such as recorder, instrument, or sampling error (Buonaccorsi, 2010: 1).

Researchers require that the variables of the data are measured without error, but, in practice, every measurement is usually carried out with errors. Furthermore, in many situations, the variance of the error may not be static, such as when data of a variable are observed in different areas, temperature, or environment, or data are recorded during a distinct period. These situations bring about the HME problem. If the model of a study is fitted with a variable with ME or HME without adjustment, then it leads to bias in parameter estimation. Statistical models and methods for analyzing and correcting the problem are called ME or HME models. In linear models, both ME and HME could occur if the dependent variables, independent variables, or both are measured with errors.

Effects of ME or HME on a dependent variable are shown as follows:

1) The OLS estimators are unbiased.
2) The variances of OLS estimators are also unbiased.

3) The estimated variances of the estimators in cases with ME or HME are larger than those without.

Effects of ME or HME on the independent variable are shown as follows:

1) The OLS estimators are biased.
2) The OLS estimators are also inconsistent.

In the case of ME or HME in a dependent variable, the error in the variable is included in the common error term of the model. This makes the estimated variances of the estimators larger than usual. However, the OLS estimators and their variances are still unbiased. Meanwhile, the OLS estimators are bias and inconsistent in case of ME or HME in an independent variable and constitute a serious problem (Gujarati, 2006: 346).

When considering the area of ME or HME models, there are important terms that need to be distinguished between: non-differential versus differential ME, and functional versus structural modeling.

When differentiating between non-differential and differential ME, the measurement error U, which is the error in X, acting as a surrogate of unobserved variable Z, is non-differential if X is associated with the dependent Y variable whenever Z is available. It can be expressed in technical terms as the distribution of Y given (Z,X) dependent only on Z, $f_{y|z,x} = f_{y|z}$, i.e. X has no information about Y whenever information on Z is available. Otherwise, the differential is expressed as $f_{y|z,x} \neq f_{y|z}$.

Non-differential ME is assumed in many studies because it is superior to differential ME in cases where the parameters in models for responses given true covariates can be estimated even when the true covariates are not observable. Examples of non-differential ME are explained in the Framingham study: long-term systolic blood pressure (Z) is the objective in the study but single day blood pressure (X) can be observed instead of the true long-term blood pressure. It can be seen that X is a surrogate of the true Z, i.e. X has no information about Z (Carroll, 2005).

To differentiate between functional and structural modeling, methods to correct the bias of the estimators can be grouped into either functional or structural models. Functional modeling is superior to structural modeling in cases where there is no need to make assumptions on Z as opposed to those requiring the specification of the distribution of Z (Carroll, Ruppert, Stefanski and Crainiceanu, 2006: 25; Fuller, 1987: 2). Several methods based on functional modeling: regression calibration, simulation-extrapolation (SIMEX), conditional score, CS, and instrumental variables need to be applied. Moreover, these methods obtain asymptotically consistent estimators in a linear functional model (Buzas, Stefanski and Tosteson, 2005). On the other hand, the method based on structural modeling is the likelihood method.

As mentioned above, the problems of ME or HME in Z are more complicated than in Y. The next topics review the concepts of ME and HME models. The major effects caused by ME and HME are described and illustrated based on linear models with ME and HME in their independent variables.

2.1.1 ME Models

ME models are characterized by circumstances where independent variable Z cannot be measured precisely, and so surrogate variable X is observed instead. The true model of the study involves the corresponding relationship between Z and dependent variable Y but the data that can be observed consist of observations of the variables X and Y. In this case, the statistical model infers that surrogate variable X can be expressed as the independent variable Z measured with errors U : $X = Z + U$. The additive structure of ME is assumed where U refers to random ME with zero mean and homogeneous variance. An example of a classical ME model in a simple linear regression is as follows:

$$y_j = \beta_0 + \beta_1 z_j + \varepsilon_j , \tag{2.1}$$

$$x_j = z_j + u_j , \tag{2.2}$$

Denote y_j as the response at the j^{th} observation,

z_j as the imprecisely observed value of independent variable the j^{th} observation,

x_j as the value which could be observed in independent variable z with the measurement error u at the j^{th} observation,

ε_j as the random error at the j^{th} observation,

u_j as the random measurement error at the j^{th} observation,

$j = 1, 2, \ldots, n$, and

n as the sample size.

Equations (2.1) and (2.2) are called "the equation-error-model" where the variables ε and U are independent. The random error is distributed as normal with mean zero and constant variance σ_ε^2, and the random ME is distributed as normal with mean zero and constant variance σ_u^2, which can be expressed as

$$\varepsilon_j \sim N\left(0, \sigma_\varepsilon^2\right),$$

$$u_j \sim N\left(0, \sigma_u^2\right).$$

Real-life Examples of the Occurrence of ME are described as follow:

In epidemiology, the result of the diagnosis of some diseases such as AIDS, cancer etc. is accessed by an indirect procedure. For example, blood pressure, blood test or x-ray diagram is used to access for disease may get the false conclusion.

In analytical chemistry, the measurements of chemical substances density or chemical level can be expressed a mistake from some sources such as the instruments, laboratories setting and/or self-reporting by researcher. Additionally, when measuring dietary intake, it is measured through the questionnaire; food frequency, exercise frequency, physical body.

MEs can be occurred in many design experiments, such as measuring water or fertilizer in agricultural industry, temperature or light setting in laboratories, measuring nutrient or calories level in dietary intake.

2.1.2 HME Models

When discussing the ME model above, the model is assumed that the variance of measurement random error is homogeneity. However, the variances of the errors can change across observations, which lead to the occurrence of HMEs.

Following the model from equations (2.1) and (2.2), the assumptions are that variables ε and U are independent, the random error is distributed as normal with mean zero and constant variance σ_{ε}^{2}, and the measurement random error is distributed as normal with mean zero. However, if it is assumed that the variance of the error u_j changes across the j observations, u_j has an HME variance σ_{uj}^{2}, i.e.

$$\varepsilon_j \sim N\left(0, \sigma_{\varepsilon}^{2}\right),$$

$$u_j \sim N\left(0, \sigma_{uj}^{2}\right).$$

HME models have been widely applied in epidemiology, analytical chemistry and botany to avoid the violations of bias in parameter estimation.

Consider one of the OLS assumptions where the independent variables and errors are independent. If the HMEs are taken into consideration, this assumption is no longer true. OLS estimators based on HME have the following effects:

1) The estimation leads to inconsistent estimates.
2) The parameter estimate is a biased estimate of the true coefficient.

Consider a simple linear regression model with HMEs:

$$y_j = \beta_0 + \beta_1 z_j + \varepsilon_j, \tag{2.3}$$

$$x_j = z_j + u_j, \tag{2.4}$$

where x_j is the observed value of z_j, $j = 1, 2, ..., n$ and assumed that

$$\begin{pmatrix} \varepsilon_j \\ u_j \end{pmatrix} \sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_e^2 & 0 \\ 0 & \sigma_{uj}^2 \end{pmatrix}\right),$$

with z_j, ε_j and u_j are independent.

Substituting $z_j = x_j - u_j$ into equation (2.3) gives

$$\begin{aligned} y_j &= \beta_0 + \beta_1\left(x_j - u_j\right) + \varepsilon_j \\ &= \beta_0 + \beta_1 x_j + \left(\varepsilon_j - \beta_1 u_j\right) \end{aligned}$$

Let $v_j = \varepsilon_j - \beta_1 u_j$, then

$$y_j = \beta_0 + \beta_1 x_j + v_j, \tag{2.5}$$

Here, $v_j = \varepsilon_j - \beta_1 u_j$ is a compound equation and measurement error, and

$$E\left(v_j\right)=E\left(\varepsilon_j-\beta_1 u_j\right)=0\,, \tag{2.6}$$

$$\begin{aligned}\operatorname{cov}\left(x_j,v_j\right)&=E\left[\left(x_j-E\left(x_j\right)\right)\left(v_j-E\left(v_j\right)\right)\right]\\&=E\left[\left(z_j+u_j-E\left(z_j+u_j\right)\right)\left(\varepsilon_j-\beta_1 u_j\right)\right]\\&=E\left[u_j\left(\varepsilon_j-\beta_1 u_j\right)\right]\\&=-\beta_1\sigma_{uj}^2\,.\end{aligned} \tag{2.7}$$

It can be seen that $\operatorname{cov}\left(x_j,v_j\right)\neq 0$, which violates the OLS assumption that the independent variables and errors are uncorrelated.

Moreover, considering equation (2.5) and by OLS estimation yields

$$\begin{aligned}\hat{\beta}_1&=\frac{\sum\limits_{j=1}^{n}\left(x_j-\bar{x}\right)y_j}{\sum\limits_{j=1}^{n}\left(x_j-\bar{x}\right)^2}\\&=\beta_1+\frac{\sum\limits_{j=1}^{n}\left(x_j-\bar{x}\right)v_j}{\sum\limits_{j=1}^{n}\left(x_j-\bar{x}\right)^2}\end{aligned}$$

$$E\left(\hat{\beta}_1\right)=\beta_1+\frac{\sum\limits_{j=1}^{n}\left(x_j-\bar{x}\right)v_j}{\sum\limits_{j=1}^{n}\left(x_j-\bar{x}\right)^2}\,. \tag{2.8}$$

It can be seen that the OLS estimator $\hat{\beta}_1$ is a biased estimator of β_1.

The probability limit of $\hat{\beta}_1$ is

$$p\lim\hat{\beta}_1=p\lim\beta_1+p\lim\left(\frac{\sum\limits_{j=1}^{n}\left(x_j-\bar{x}\right)v_j}{\sum\limits_{j=1}^{n}\left(x_j-\bar{x}\right)^2}\right)$$

$$=\beta_1+\frac{p\lim\left(\frac{1}{n}\sum\limits_{j=1}^{n}\left(x_j-\bar{x}\right)v_j\right)}{p\lim\left(\frac{1}{n}\sum\limits_{j=1}^{n}\left(x_j-\bar{x}\right)^2\right)}$$

$$p\lim\hat{\beta}_1=\beta_1+\frac{\operatorname{cov}\left(x_j,v_j\right)}{\operatorname{var}\left(x_j\right)}$$

$$=\beta_1+\frac{\operatorname{cov}\left(x_j,\varepsilon_j-\beta_1u_j\right)}{\operatorname{var}\left(z_j+u_j\right)}.$$

Substituting $\operatorname{cov}\left(x_j,v_j\right)=-\beta_1\sigma_{uj}^2$ and using the independence property of z_j and u_j yields

$$p\lim\hat{\beta}_1=\beta_1+\frac{-\beta_1\sigma_{uj}^2}{\sigma_z^2+\sigma_{u_j}^2}=\beta_1\left[\frac{1}{1+\left(\sigma_{u_j}^2/\sigma_z^2\right)}\right]. \tag{2.9}$$

It can be seen that the term $\frac{1}{1+\left(\sigma_{u_j}^2/\sigma_z^2\right)}$ in (2.9) is always less than one. Consequently, the OLS estimator $\hat{\beta}_1$ is an inconsistent estimator of β_1.

2.2 Methods to Correct the ME/HME Problem

Several methods to correct the ME or HME problem have been put forward. Methods based on functional modeling can be divided into approximately consistent methods (remove most of the bias) and fully consistent methods (remove the bias to achieve asymptotic consistency) (Buzas, Stefanski and Tosteson, 2005). The first group of methods uses either regression calibration or the SIMEX method. The second group consists of conditional score method, corrected score (CS) approach, and method of using some instrumental variables. The concept for each method is described below.

2.2.1 Regression Calibration

The concept of this method is based on regression analysis by predicting the unobservable Z from a surrogate X, then using the prediction of X as the independent variable of the model and regress Y onto the prediction of X. Pierce and Kellerer (2004: 863) noted that, “It is very convenient that essentially the same methods for ongoing analyses can be employed as if the variable Z were observed". Regression calibration was first developed by Prentice (1982) in his studies on the proportional hazard model. After that, this approach was modified for use in epidemiology by Clayton (1992) and extended to logistic regression by Rosner, Willett and Spiegelman (1989) and Carroll and Stefanski (1990). The details of the concept are explained by Carroll, Ruppert, Stefanski and Crainiceanu (2006: 65-95).

For the step of modeling a surrogate X on the unobservable Z, additional data such as using instrumental variables or replicated observations are required. The steps for parameter estimation by using regression calibration are:

i) Regress X on Z, then the prediction of X is obtained.
ii) Estimate the parameters in the model by regressing the prediction of X onto Y.

Regression calibration is fully consistent in linear models, which also applies to a generalized linear model. However, it is approximately consistent (ineffective in reducing bias) in non-linear models.

2.2.2 SIMEX

SIMEX is a technique first developed by Cook and Stefanski (1995) consisting of a combination of extrapolation and simulation. The extrapolation method is a mathematical procedure designed to enable one to estimate the unknown values of a parameter from known values. The parameter estimation is obtained by using a simulation to obtain values for the surrogate variable X and compute the corresponding regression. Details and examples of this method are available in Carroll, Ruppert, Stefanski and Crainiceanu (2006: 97-126).

SIMEX is fully consistent in linear models but approximately consistent in non-linear ones.

2.2.3 Conditional Score

The basic theory behind conditional scoring is based on solving the parameter estimation by using estimating equations under the condition that an estimating score is unbiased if it has an expectation of zero. This method was first proposed by Gleser (1981) based on the derivation of the estimators in linear regression by maximizing the joint density of the observed data with respect to all of the unknown parameters, including the unobservable Z. Next, the conditional score was developed in the parameter estimation in logistic regression (Carroll and Stefanski, 1990). It can be seen that this method is difficult to compute in complex models such as logistic regression and it requires the distribution of the measurement error U has normal distributed.

2.2.4 CS

The CS method was developed by Stefanski and Carroll (1987) and Nakamura (1990). The basic concept is the same as the conditional score method, and is based on the using of an estimation equation to solve the parameters in the model. This method can correct inconsistent estimators and also does well in models with no assumptions concerning the distribution of the unobserved variable (functional modeling). The details of this method are as follows.

To consider the estimation of parameter θ, let *z* be the column vector of independent variables and *y* be a random variable whose distribution depends on *z*; the *z*'s are termed covariates and the *y*' s dependent variables. Let Z and Y denote the set of independent variables *z*'s and dependent variables *y*'s, respectively, and denote

$L(\theta, Z, Y)$ as the likelihood function of θ given Z and Y,

$l(\theta, Z, Y)$ as the log-likelihood function of θ given Z and Y, and

$U(\theta, Z, Y)$ as the score function of θ given Z and Y.

If Z can be observed without errors, then by maximum likelihood estimator (MLE), the estimator of θ is obtained from the satisfying of $E\left(U(\theta, Z, Y)\right) = 0$. Otherwise, if Z is imprecisely observed and is measured with errors which can be expressed as X, the log-likelihood function $l(\theta, Z, Y)$ is replaced with $l(\theta, X, Y)$, then $E\left(U(\theta, X, Y)\right)$ is bounded away from zero although the sample sizes approaches infinity. Consequently, the MLE which satisfies $U(\theta, X, Y) = \frac{\partial l(\theta, X, Y)}{\partial \theta} = 0$ is no longer consistency (Augustin, 2004).

Because of the above mentioned, the model which is observed the variables Y and X, instead of Y and Z provides the in efficiency estimators. The ME is corrected by constructing unbiased estimating functions based on the observable Y and X. The concept is to correct the bias term by finding the observable CS function $U_c(\theta, X, Y)$ such that

$$E\left(U_c(\theta, X, Y)\right) = U(\theta, Z, Y). \tag{2.10}$$

For simplicity, the log-likelihood is used for finding a function $l_c(\theta, X, Y)$, then the function $l_c(\theta, X, Y)$ is satisfied

$$E\left(l_c(\theta, X, Y)\right) = l(\theta, Z, Y). \tag{2.11}$$

Then, by the regularity condition,

$$U_c(\theta, X, Y) = \frac{\partial l_c(\theta, Z, Y)}{\partial \theta}. \tag{2.12}$$

$U_c(\theta, X, Y)$ is called the CS function and $l_c(\theta, X, Y)$ denotes a corrected log-likelihood function.

It follows from the property for unbiasedness, $E\left(U(\theta, Z, Y)\right) = 0$ and $E\left(U_c(\theta, X, Y)\right) = U(\theta, Z, Y)$, that $U_c(\theta, X, Y)$ is also conditionally unbiased provided $E\left(U_c(\theta, X, Y)\right) = 0$. Thus, by applying the general theory of M-estimation, the estimating equation written as

$$\sum_{j=1}^{n} U_c\left(\theta, X_j, Y_j\right) = 0, \tag{2.13}$$

possesses a consistent, asymptotically normally sequence of solutions (Nakamura, 1990).

Nakamura (1990) proposed the CS function for correcting the inconsistent estimators in MEs in models with independent variables. The idea is indicated that "the conditional distribution of the corrected estimate given the true independent variables and the dependent variables is centered around the maximum likelihood estimate, which in turn is centered around the true parameter value". This is to find the function of corrected log-likelihood of the x' s and y' s, $l_c(\theta, X, Y)$, such that

$E\left(l_c\left(\theta, X, Y\right)\right) = l\left(\theta, Z, Y\right)$, where θ denotes the parameters of the model, and Z , X , and Y represents the set of imprecisely observed independent variables z' s , observed independent variables z' s, and dependent variables y' s, respectively. Moreover, Nakamura derived a CS function for four different models as shown here.

1) Generalized Linear Model
2) Normal Regression Model
3) Poisson Regression Model
4) Gamma Regression Model

After that, Huang and Wang (2001) proposed the CS in logistic regression and then, Chen, Hanfelt and Huang (2015) extened the studied in this model.

Giménez and Bolfarine (2000) considered comparisons between four approaches of consistent estimators in functional comparative calibration models (a special case of linear multivariate ME models). The model is assumed that there are $p+1, \left(p \geq 1\right)$ disposal measuring instruments of n subjects. Let y_{ji} be the observed value at the i^{th} disposal measuring instruments at the j^{th} subject, $j = 1, 2, ..., n$ by $i = 1, 2, ..., p$, and x_j be the observed measures of the unobserved z_j .To consider the linear model

$$\mathbf{y}_{j.} = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 z_j + \boldsymbol{\varepsilon}_{j.}, \tag{2.14}$$

$$x_j = z_j + u_j, \tag{2.15}$$

where $\mathbf{y}'_{j.} = \left(y_{j1}, y_{j2}, ..., y_{jp}\right)'$, $\boldsymbol{\beta}_0 = \left(\beta_{01}, \beta_{02}, ..., \beta_{0p}\right)'$ is the $p \times 1$ vector associated with the additive bias of the p measuring devices, $\boldsymbol{\beta}_1 = \left(\beta_{11}, \beta_{12}, ..., \beta_{1p}\right)'$ is the $p \times 1$ vector associated with the multiplicative bias of the p measuring devices, $\boldsymbol{\varepsilon}_{j.} = \left(\varepsilon_{j1}, \varepsilon_{j2}, ..., \varepsilon_{jp}\right)'$ is the $p \times 1$ vector of errors with the errors within the subject j that are not independent, $\boldsymbol{\varepsilon}_{j.} \sim N\left(0, \Sigma\right)$, $\Sigma = diag\left(\sigma_1^2, \sigma_2^2, ..., \sigma_p^2\right)$, $u_j \sim N\left(0, \sigma_u^2\right)$ and $\boldsymbol{\varepsilon}_{j.}$ and u_j are mutually independent.

The approaches for the four consistent estimators are: 1) sufficiency and conditional scores, 2) maximum likelihood estimation 3), the CS function, and 4) moment estimators. The asymptotic distributions associated with the different estimators were studied and they showed that the conditional score function is equivalent to the maximum likelihood estimator. Note that the asymptotic relative efficiency was a criterion for this study. The results showed that, for small and moderate sample sizes, there is no one estimator more efficient than the others. However, the sufficiency score estimator given the poor performance when comparing with the naive estimator (the OLS estimator without ME adjustment).

Giménez and Patat (2005) focused on CS estimation in a comparative calibration model (a special case of linear multivariate models) with balanced repeated measurements and constant variances of MEs. Assumes the model in (2.14) and (2.15), for the instrument $i,\, i=1,2,...,p$, the k repeated measurement at the j^{th} observation, for $k=1,2,...,r$ are observed in Y and $k=1,2,...,m$, are observed in X , then the model can be expressed as

$$y_{ijk}=\beta_{0i}+\beta_{1i}z_j+\varepsilon_{ijk}, \tag{2.16}$$

$$x_{jk}=z_j+u_{jk},, \tag{2.17}$$

where $j=1,2,...,n$, ε_{ijk} and u_{jk} are independent and identically distributed with $N\left(0,\sigma_i^2\right)$ and $N\left(0,\sigma_u^2\right)$, respectively.

The CS function of this model is given by

$$U^*_{j(\beta_q)}=\begin{cases}\mathbf{\Sigma}^{-1}\sum\limits_{k=1}^{r}\left(\mathbf{y}_{jk}-\boldsymbol{\alpha}-\boldsymbol{\beta}\bar{x}_{j\cdot}\right), q=0\\ \mathbf{\Sigma}^{-1}\sum\limits_{k=1}^{r}\left(\mathbf{y}_{jk}-\boldsymbol{\alpha}-\boldsymbol{\beta}\bar{x}_{j\cdot}\right)\bar{x}_{j\cdot}+\boldsymbol{\beta}\dfrac{\boldsymbol{\sigma}_u^2}{m}, q=1\end{cases},$$

repeated measurement of the imprecisely observed variable Z obtained the necessary information for estimating the variances of MEs which can be evaluated by

$$\hat{\sigma}_u^2=\frac{\sum\limits_{j=1}^{n}\sum\limits_{k=1}^{m}\left(x_{jk}-\bar{x}_{j\cdot}\right)^2}{n(m-1)},$$

The proposed consistent and asymptotically normal distributed estimator can be expressed as

$$\hat{\beta}_{0i} = \bar{y}_{i..} - \hat{\beta}_{1i}\bar{x}_{i..}\,,\ \hat{\beta}_{1i} = \frac{S_{\bar{x}\,\bar{y}_i}}{S_{\bar{x}\,\bar{x}} - \hat{\sigma}_u^2 \big/ m}\,,\ i = 1, 2, ..., p\,,$$

with $S_{\bar{x}\,\bar{x}} = \frac{1}{n}\sum_{j=1}^{n}\left(\bar{x}_{j.} - \bar{x}_{..}\right)^2,\ \bar{x}_{..} = \frac{1}{n}\sum_{j=1}^{n}\bar{x}_{j.}\,,$

$$S_{\bar{x}\,\bar{y}_i} = \frac{1}{n}\sum_{j=1}^{n}\left(\bar{x}_{j.} - \bar{x}_{..}\right)\left(\bar{y}_{ij.} - \bar{y}_{i..}\right),$$

$$\bar{y}_{i..} = \frac{1}{nr}\sum_{j=1}^{n}\sum_{k=1}^{r} y_{ijk} = \frac{1}{n}\sum_{j=1}^{n}\bar{y}_{ij.}\ .$$

The performance of the estimators in simulation study showed that for m is fixed, the bias and the standard deviation (SD) decreased when r increased.

Additionally, the CS estimator in a comparative calibration model with unbalanced repeated measurements is derived by Giménez and Patat (2014). Based on the model in (2.16) and (2.17), setting $i = 1$, the unbalanced repeated measurements model becomes

$$y_{jk} = \beta_0 + \beta_1 z_j + \varepsilon_{jk}\,,\ k = 1, 2, ..., r_j\,, \tag{2.18}$$

$$x_{jh} = z_j + u_{jh}\,,\ h = 1, 2, ..., m_j\,,\ j = 1, 2, ..., n\ , \tag{2.19}$$

The estimators in the model can be expressed as

$$\hat{\beta}_0 = \tilde{y}_{..} - \hat{\beta}_1\tilde{x}_{..}\,,\ \hat{\beta}_1 = \frac{\tilde{S}_{xy}}{\tilde{S}_{xx} - b\hat{\sigma}_u^2}\,,\ \hat{\sigma}_u^2 = \frac{\sum_{j=1}^{n}\sum_{k=1}^{m_j}\left(x_{jh} - \bar{x}_{j.}\right)^2}{M - n}\,,$$

with $S_{xy} = \sum_{j=1}^{n}\lambda_j\left(\bar{x}_{j.} - \bar{x}_{..}\right)\left(\bar{y}_{j.} - \bar{y}_{..}\right),\ \tilde{S}_{xx} = \sum_{j=1}^{n}\lambda_j\left(\bar{x}_{j.} - \bar{x}_{..}\right)^2,$

$\tilde{y}_{..} = \sum_{j=1}^{n}\lambda_j y_{j.}\,,\ \tilde{x}_{..} = \sum_{j=1}^{n}\lambda_j\bar{x}_{j.}\,,$ and

$b = \sum_{j=1}^{n}\frac{\lambda_j}{m_j}$ where $\lambda_j = \frac{r_j}{R},\ j = 1, 2, ..., n,$

$R = \sum_{j=1}^{n} r_j\ ,\ m = \sum_{j=1}^{n} m_j\ .$

DeCastro, Bolfarine and Castilho (2006) considered consistent estimation based on the CS approach in simple linear regression under MEs with known different variances in various situations. For one regression line, assumed variable z_j was measured with bias. Consider the simple linear regression model in (2.3) and (2.4), assume that ε_j and u_j are independent with σ_{ej}^2 and σ_{uj}^2 known. The corrected likelihood estimator is derived as

$$l^*\left(\beta, Z, Y\right) = constant - \frac{1}{2}\sum_{j=1}^{n}\frac{\left(y_j - \beta_0 - \beta_1 z_j\right)^2 - \beta_1^2\sigma_{uj}^2}{\sigma_{ej}^2},$$

and the CS estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ are given by

$$\hat{\beta}_0 = \frac{\sum_{j=1}^{n}\left(y_j - \hat{\beta}_1 x_j\right)^2 \Big/ \sigma_{ej}^2}{\sum_{j=1}^{n} 1\big/\sigma_{uj}^2},$$

$$\hat{\beta}_1 = \frac{\sum_{j=1}^{n}\left(x_j y_j \big/ \sigma_{ej}^2\right) - \sum_{j=1}^{n}\left(y_j \big/ \sigma_{ej}^2\right)\sum_{j=1}^{n}\left(x_j \big/ \sigma_{ej}^2\right)\left[\sum_{j=1}^{n}\left(1\big/\sigma_{ej}^2\right)\right]^{-1}}{\sum_{j=1}^{n}\left(x_j^2 \big/ \sigma_{ej}^2\right) - \left[\sum_{j=1}^{n}\left(x_j^2 \big/ \sigma_{ej}^2\right)\right]^2\left[\sum_{j=1}^{n}\left(1\big/\sigma_{ej}^2\right)\right]^{-1} - \sum_{j=1}^{n}\left(\sigma_{uj}^2 \big/ \sigma_{ej}^2\right)}.$$

Patriota, Lemonte and Bolfarine (2010) derived a bias-adjustment scheme to eliminate the second-order biases of the maximum-likelihood estimates in a heteroskedastic multivariate MEs regression model using the general matrix formulae for the second-order bias derived by Patriota and Lemonte (2009). Via a simulation study, the bias correction derived in this paper was very effective, even when the sample size was large. The bias correction yields the HME become nearly unbiased estimator.

Giménez and Galae (2013) derived a CS estimator for a multivariate model containing HMEs with known variances. By assuming that the model for their study followed the model of Giménez and Bolfarine (2000), they focused on one independent variable and assumed that the errors within unit j were independent and that $u_j \sim N\left(0, \sigma_j^2\right)$. In addition, this CS estimator was considered for assessing the local influence of the effects of minor perturbations. A perturbation vector was introduced to the CS function (which is independent of the incidental parameters) and a methodology to find the density of the corresponding perturbed model (including structural and incidental parameters) was proposed.

A comparison of conditional score and CS showed that both methods are fully consistent in the linear model. In Poisson regression, the parameter estimation by conditional score has more efficient than CS in some cases, but, for other models, if the CS function exists, CS estimators are more easy to derive than conditional score estimators.

2.2.5 Instrumental Variables

In this method, parameter estimation is derived from additional data (information about the unobservable Z) when repeated measurements or validation data cannot be used. The additional data comes from instrumental variables for which the following conditions must hold true:

1) They must be non-differential ME.
2) They must be correlated with the unobservable Z.
3) They must be independent of $X - Z$.

The disadvantage of this method is that it uses a second parameter, i.e. the parameter estimation comes from the instrumental variables which are not directly analyzed on the surrogate X and, based on the requirements mentioned above, the use of instrumental variable makes the assumption weaker than the initial model.

2.3 The Multivariate Multiple Regression Model

A multivariate multiple regression model is an extension of multiple regression models where the effects on a set of dependent variables are modeled simultaneously, and the focus is on the problem of modeling the relationship between p dependent variables $\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_p$ and a single set of independent variables $z_1, z_2, ..., z_s$:

$$\mathbf{y}_1 = \beta_{01} + \beta_{11} z_1 + \ldots + \beta_{s1} z_s + \boldsymbol{\varepsilon}_1$$
$$\mathbf{y}_2 = \beta_{02} + \beta_{12} z_1 + \ldots + \beta_{s2} z_s + \boldsymbol{\varepsilon}_2$$
$$\vdots$$
$$\mathbf{y}_p = \beta_{0p} + \beta_{1p} z_1 + \ldots + \beta_{sp} z_s + \boldsymbol{\varepsilon}_p \,,$$

The matrix form of this model can be written as

$$\begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1p} \\ y_{21} & y_{22} & \cdots & y_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{np} \end{pmatrix} = \begin{pmatrix} 1 & z_{11} & \cdots & z_{1,s-1} & z_{1s} \\ 1 & z_{21} & \cdots & z_{2,s-1} & z_{2s} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & z_{n1} & \cdots & z_{n,s-1} & z_{ns} \end{pmatrix} \begin{pmatrix} \beta_{01} & \beta_{02} & \cdots & \beta_{0p} \\ \beta_{11} & \beta_{12} & \cdots & \beta_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{s1} & \beta_{s2} & \cdots & \beta_{sp} \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} & \varepsilon_{12} & \cdots & \varepsilon_{1p} \\ \varepsilon_{21} & \varepsilon_{22} & \cdots & \varepsilon_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon_{n1} & \varepsilon_{n2} & \cdots & \varepsilon_{np} \end{pmatrix}$$

or

$$\mathbf{Y} = \mathbf{ZB} + \mathbf{E} \,, \qquad (2.20)$$

where $\mathbf{Y}$ is the $n \times p$ matrix of dependent variable ,

$\mathbf{Z}$ is the $n \times (s+1)$ matrix of independent variables including a constant unit vector,

$\mathbf{B}$ is the $(s+1) \times p$ matrix of parameters including a constant term,

$\mathbf{E}$ is the $n \times p$ matrix of random errors,

n is the number of observations in the model,

p is the number of dependent variables in the model, and

r is the number of independent variables in the model.

Denote $j = 1, 2, \ldots, n$ by $i = 1, 2, \ldots, p$, then

$\mathbf{Y} = (\mathbf{y}_{1\cdot}, \mathbf{y}_{2\cdot}, \ldots, \mathbf{y}_{n\cdot})'$ represents the j^{th} row of $\mathbf{Y}$ where $\mathbf{y}_{j\cdot} = (y_{j1}, y_{j2}, \ldots, y_{jp})'$,

$\mathbf{Z} = (\mathbf{z}_{1\cdot}, \mathbf{z}_{2\cdot}, \ldots, \mathbf{z}_{n\cdot})'$ represents the j^{th} row of $\mathbf{Z}$ where $\mathbf{z}_{j\cdot} = (1, z_{j1}, \ldots, z_{js})'$,

$\mathbf{B} = (\boldsymbol{\beta}_{\cdot 1}, \boldsymbol{\beta}_{\cdot 2}, \ldots, \boldsymbol{\beta}_{\cdot p})'$ represents the i^{th} row of $\mathbf{B}$ where $\boldsymbol{\beta}_{\cdot i} = (\beta_{oi}, \beta_{1i}, \ldots, \beta_{si})'$, and

$\mathbf{E} = (\boldsymbol{\varepsilon}_{\cdot 1}, \boldsymbol{\varepsilon}_{\cdot 2}, \ldots, \boldsymbol{\varepsilon}_{\cdot p})'$ represents the i^{th} row of $\mathbf{E}$ where $\boldsymbol{\varepsilon}_{\cdot i} = (\varepsilon_{1i}, \varepsilon_{2i}, \ldots, \varepsilon_{si})'$.

Assumptions:

1) $E(\boldsymbol{\varepsilon}_{\cdot i}) = 0$,

2) $\operatorname{cov}(\boldsymbol{\varepsilon}_{\cdot i}, \boldsymbol{\varepsilon}_{\cdot l}) = \sigma_{il}\mathbf{I}, i, l = 1, 2, ..., p$.

The p observations on the j^{th} sample unit have covariance matrix $\Sigma = \{\sigma_{il}\}$, but observations from different trials are uncorrelated.

The representation in the i^{th} component has the form

$$\left(\mathbf{y}_{\cdot 1}, \mathbf{y}_{\cdot 2}, ..., \mathbf{y}_{\cdot p}\right) = \mathbf{Z}\left(\boldsymbol{\beta}_{\cdot 1}, \boldsymbol{\beta}_{\cdot 2}, ..., \boldsymbol{\beta}_{\cdot p}\right) + \left(\boldsymbol{\varepsilon}_{\cdot 1}, \boldsymbol{\varepsilon}_{\cdot 2}, ..., \boldsymbol{\varepsilon}_{\cdot p}\right).$$

The implicit multiple regression model for the i^{th} component can be written as

$$\mathbf{y}_{\cdot i} = \mathbf{Z}\boldsymbol{\beta}_{\cdot i} + \boldsymbol{\varepsilon}_{\cdot i} \ . \tag{2.21}$$

For maximum likelihood estimation, let the multivariate multiple regression model in equation (2.14) hold with full rank $(\mathbf{Z}) = s+1, n \geq (s+1) + p$, and let the errors $\mathbf{E}$ be normally distributed, $\mathbf{y}_{j\cdot} = \left(y_{j1}, y_{j2}, ..., y_{jp}\right)'$ be the p-dimension of dependent variables, and $\mathbf{z}_{j\cdot} = \left(1, z_{j1}, ..., z_{js}\right)'$ be the vector of independent variables. Then, the likelihood function can be written as

$$L(\mathbf{B}, \Sigma) = \prod_{j=1}^{n} \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2}\left(\mathbf{y}_j - \mathbf{B}'\mathbf{z}_j\right)' \Sigma^{-1}\left(\mathbf{y}_j - \mathbf{B}'\mathbf{z}_j\right)\right). \tag{2.22}$$

Following this, the log-likelihood function becomes

$$l(\mathbf{B}, \Sigma) = \sum_{j=1}^{n} -\frac{p}{2}\log(2\pi) - \frac{1}{2}\left(\mathbf{y}_j - \mathbf{B}'\mathbf{z}_j\right)' \Sigma^{-1}\left(\mathbf{y}_j - \mathbf{B}'\mathbf{z}_j\right), \tag{2.23}$$

and the derivative of $l(\mathbf{B}, \Sigma)$ yields the score function

$$U(\mathbf{B}, \Sigma) = \begin{pmatrix} \dfrac{\partial l(\mathbf{B}, \Sigma)}{\partial \beta} \\ \dfrac{\partial l(\mathbf{B}, \Sigma)}{\partial \Sigma^{-1}} \end{pmatrix}$$

The maximum likelihood estimator of $\boldsymbol{B}$ can be expressed as

$$\hat{\mathbf{B}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y},$$

and the implicit form with the i^{th} component can be written as

$$\hat{\boldsymbol{\beta}}_{\cdot i} = \left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{y}_{\cdot i}\,. \qquad (2.24)$$

2.3.1 HME in a Multivariate Multiple Regression Model

When considering a linear multivariate regression model with HME, assumed that, in the model in equation (2.14), there are p components of dependent variables $Y_1, Y_2, ..., Y_p$ corresponding to s independent variables, where $Z_1, ..., Z_{s-1}$ refers to the precise measurement of independent variables, and $X_s = Z_s + U_s$ refers to an independent variable observed instead of Z_s with additive HME. The HME model is specified as follows:

$$\mathbf{Y} = \mathbf{ZB} + \mathbf{E},$$
$$\mathbf{x}_s = \mathbf{z}_s + \mathbf{u}_s, \qquad (2.25)$$

The additive HME equation in implicit form with the j^{th} observation $\left(j = 1, 2, ..., n\right)$ can be specified as

$$x_{js} = z_{js} + u_{js} \qquad (2.26)$$

For the random errors matrix $\mathbf{E} = \left(\bar{\boldsymbol{\varepsilon}}_1\ \bar{\boldsymbol{\varepsilon}}_2\ ...\ \bar{\boldsymbol{\varepsilon}}_n\right)'$, the rows are independent with $\boldsymbol{\varepsilon}_j$ distributed as $N_p\left(0, \boldsymbol{\Sigma}_j\right)$, where

$$\boldsymbol{\Sigma}_j = \begin{pmatrix} \sigma_{11}^2 & \cdots & \sigma_{1p} \\ \vdots & \ddots & \vdots \\ \sigma_{p1} & \cdots & \sigma_{pp}^2 \end{pmatrix}_j, \qquad (2.27)$$

u_{js} are distributed as $N\left(0, \sigma_{uj}^2\right)$, $z_1, ..., z_{s-1}$, ε_{ji} and u_{js} are mutually independent, $\sigma_{ii'}^2 > 0$, and $\sigma_{uj}^2 > 0$, $i = 1, 2, ..., p,\ j = 1, 2, ..., n$.

Substituting the additive HME equation (2.26) : $z_{js} = x_{js} - u_{js}$ into equation (2.20), the matrix form of equation (2.20) becomes

$$\begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1p} \\ y_{21} & y_{22} & \cdots & y_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{np} \end{pmatrix} = \begin{pmatrix} 1 & z_{11} & \cdots & z_{1,s-1} & x_{1s}-u_{1s} \\ 1 & z_{21} & \cdots & z_{2,s-1} & x_{2s}-u_{2s} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & z_{n1} & \cdots & z_{n,s-1} & x_{ns}-u_{ns} \end{pmatrix} \begin{pmatrix} \beta_{01} & \beta_{02} & \cdots & \beta_{0p} \\ \beta_{11} & \beta_{12} & \cdots & \beta_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{s1} & \beta_{s2} & \cdots & \beta_{sp} \end{pmatrix}$$

$$+\begin{pmatrix} \varepsilon_{11} & \varepsilon_{12} & \cdots & \varepsilon_{1p} \\ \varepsilon_{21} & \varepsilon_{22} & \cdots & \varepsilon_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon_{n1} & \varepsilon_{n2} & \cdots & \varepsilon_{np} \end{pmatrix} \tag{2.28}$$

The implicit form with the j^{th} observation $\left(j=1,2,...,n\right)$ can be written as

$$i=1: \quad y_{j1}=\beta_{01}+\beta_{11}z_{j1}+...+\beta_{s-1,1}z_{j,s-1}+\beta_{s1}\left(x_{js}-u_{js}\right)+\varepsilon_{11},$$

$$i=2: \quad y_{j2}=\beta_{02}+\beta_{12}z_{j1}+...+\beta_{s-1,2}z_{j,s-1}+\beta_{s2}\left(x_{js}-u_{js}\right)+\varepsilon_{12},$$

$$\vdots$$

$$i=p: \quad y_{jp}=\beta_{0p}+\beta_{1p}z_{j1}+...+\beta_{s-1,p}z_{j,s-1}+\beta_{sp}\left(x_{js}-u_{js}\right)+\varepsilon_{1p},$$

and we can rewrite the implicit form with the j^{th} observation $\left(j=1,2,...,n\right)$ as:

$$i=1: \quad y_{j1}=\beta_{01}+\beta_{11}z_{j1}+...+\beta_{s-1,1}z_{j,s-1}+\beta_{s1}x_{js}+\upsilon_{j1}$$

$$i=2: \quad y_{j2}=\beta_{02}+\beta_{12}z_{j1}+...+\beta_{s-1,2}z_{j,s-1}+\beta_{s2}x_{js}+\upsilon_{j2}$$

$$\vdots$$

$$i=p: \quad y_{jp}=\beta_{0p}+\beta_{1p}z_{j1}+...+\beta_{s-1,p}z_{j,s-1}+\beta_{sp}x_{js}+\upsilon_{jp}$$

where $\upsilon_{ji}=\varepsilon_{ji}-\beta_{si}u_{js}$, $i=1,2,...,p,\quad j=1,2,...,n.$

Consequently, the matrix from of a linear multivariate multiple regression model with HME can be specified as

$$\begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1p} \\ y_{21} & y_{22} & \cdots & y_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{np} \end{pmatrix} = \begin{pmatrix} 1 & z_{11} & \cdots & z_{1,s-1} & x_{1s} \\ 1 & z_{21} & \cdots & z_{2,s-1} & x_{2s} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & z_{n1} & \cdots & z_{n,s-1} & x_{ns} \end{pmatrix} \begin{pmatrix} \beta_{01} & \beta_{02} & \cdots & \beta_{0p} \\ \beta_{11} & \beta_{12} & \cdots & \beta_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{s1} & \beta_{s2} & \cdots & \beta_{sp} \end{pmatrix} + \begin{pmatrix} \upsilon_{11} & \upsilon_{12} & \cdots & \upsilon_{1p} \\ \upsilon_{21} & \upsilon_{22} & \cdots & \upsilon_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \upsilon_{n1} & \upsilon_{n2} & \cdots & \upsilon_{np} \end{pmatrix}$$

or

$$\mathbf{Y}=\mathbf{XB}+\mathbf{V}\ . \tag{2.29}$$

The OLS estimator of $\mathbf{B}$ is

$$\hat{\mathbf{B}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y}\,. \tag{2.30}$$

The implicit form with the i^{th} component can be written as

$$\tilde{\boldsymbol{\beta}}_{\cdot i} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}_{\cdot i}\;. \tag{2.31}$$

In this case, $\tilde{\boldsymbol{\beta}}_{\cdot i}$ is a biased estimator of $\boldsymbol{\beta}_{\cdot i}$ where the bias term is $E\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\upsilon}_{\cdot i}\right]$,

and its estimated variance is $\left(\sigma_i^2 + \beta_{si}^2\sigma_{uj}^2\right)\left(\mathbf{X}'\mathbf{X}\right)^{-1}$.

CHAPTER 3

METHODOLOGY

In this study, the aim is to propose an estimation approach based on the CS in a linear multivariate multiple regression model with HME. The situation of an HME model with imprecisely observed independent variables and repeated measurements is of interest. The CS approach consists of four stages: first, construct a corrected log-likelihood function; second, evaluate the CS functions of the parameters; third, estimate the variances of the HMEs; and fourth, evaluate the CS estimators of the parameters. This section also covers the derivation of the bias and MSE of the CS estimator.

3.1 The Study Model

Consider a linear multivariate measurement error regression model in which the p correlated dependent variables $Y_1, Y_2, ..., Y_p$ explained by s independent variables, where the first s_1 independent variables $Z_1, Z_2, ..., Z_{s_1}$ are precisely observed and the last $\left(s - s_1\right)$ independent variables $Z_{s_1+1}, Z_{s_1+2}, ..., Z_s$ are imprecisely observed via their corresponding variables $X_{s_1+1}, X_{s_1+2}, ..., X_s$ with additive HME. Let r_j be the number of repeated measurements of Y_j, $j = 1, 2, ..., n$, where n is the number of observations, and X_q, $q = s_1 + 1, s_1 + 2, \ldots, s$, y_{jik} and x_{jqk} be the k^{th} repeated measurement of the j^{th} observation of Y_i and X_q, $i = 1, 2, ..., p$, $j = 1, 2, ..., n$, $q = s_1 + 1,$, $s_1 + 2, ..., s$, $k = 1, 2, ..., r_j$. Subsequently, the linear multivariate measurement error regression model can be expressed in matrix form as:

$$\mathbf{Y} = \mathbf{ZB} + \mathbf{E}, \tag{3.1}$$

$$\bar{\mathbf{x}}_q = \mathbf{z}_q + \bar{\mathbf{u}}_q \ , \ q = s_1 + 1, s_1 + 2, ..., s \ . \tag{3.2}$$

Denote $\mathbf{Y}$ as the $n \times p$ matrix of the average measurements of dependent variables with the j^{th} observation and the i^{th} component,

$\mathbf{Z}$ as the $n \times (s+1)$ matrix of precise measurements of independent variables including a constant unit vector,

$\mathbf{B}$ as the $(s+1) \times p$ matrix of parameters including a constant term,

$\mathbf{E}$ as the $n \times p$ matrix of random errors with the j^{th} observation and and the i^{th} component,

$\bar{\mathbf{x}}_q$ as the $n \times 1$ vector of the averages of measurements of X_q with the j^{th} observation,

$\mathbf{z}_q$ as the $n \times 1$ vector of imprecise measurements of the q^{th} independent variables, and

$\bar{\mathbf{u}}_q$ as the $n \times 1$ vector of the averages of the heterogeneous random ME with the j^{th} observation.

Let $\bar{y}_{ji\cdot}$, $i = 1, 2, ..., p$, $j = 1, 2, ..., n$ be the average of the observed values of the dependent variable for the j^{th} observation and the i^{th} component, corresponding to the independent variable z_{jq}, $q = 1, 2, ..., s$ where $z_{j1}, ..., z_{js_1}$ can be observed precisely whereas $z_{j(s_1+1)}, ..., z_{js}$ cannot. Then,

$$\bar{y}_{ji\cdot} = \frac{\sum_{k=1}^{r_j} y_{jik}}{r_j}, \ j = 1, 2, ..., n, \ i = 1, 2, ..., p \ .$$

Assuming that $\bar{x}_{jq\cdot} = z_{jq} + \bar{u}_{jq\cdot}$ is observed instead of z_{jq}, $q = s_1 + 1, s_1 + 2, ..., s$ in the k^{th} repeated measurement of dependent variable Y_i at the j^{th} observation. Then,

$$\bar{x}_{jq\cdot} = \frac{\sum_{k=1}^{r_j} x_{jqk}}{r_j}, \ j = 1, 2, ..., n, \ q = s_1 + 1, s_1 + 2, ..., s,$$

$$\bar{u}_{jq\cdot} = \frac{\sum_{k=1}^{r_j} u_{jqk}}{r_j}, \ j = 1, 2, ..., n, \ q = s_1 + 1, s_1 + 2, ..., s, \text{ and}$$

$$\bar{\varepsilon}_{ji\cdot} = \frac{\sum_{k=1}^{r_j} \varepsilon_{jik}}{r_j}, \ j = 1, 2, ..., n, \ i = 1, 2, ..., p,$$

where ε_{jik} is the mutually independent random error in the k^{th} repeated measurement of dependent variable Y_i at the j^{th} observation.

The random measurement error of the k^{th} measurements of X_q at the j^{th} observation, u_{jqk} are independent variables across the measurements of the observation, and distributed as $N\left(0, \sigma_{ujq}^2\right)$, the variance σ_{ujq}^2 is assumed to be associated with HME but unknown.

The matrix notation can be written in the form

$$\mathbf{Y} = \begin{bmatrix} \bar{y}_{11\cdot} & \bar{y}_{12\cdot} & \cdots & \bar{y}_{1p\cdot} \\ \bar{y}_{21\cdot} & \bar{y}_{22\cdot} & \cdots & \bar{y}_{2p\cdot} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{y}_{n1\cdot} & \bar{y}_{n2\cdot} & \cdots & \bar{y}_{np\cdot} \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{y}}_1' \\ \bar{\mathbf{y}}_2' \\ \vdots \\ \bar{\mathbf{y}}_n' \end{bmatrix}.$$

Denote the j^{th} row of $\mathbf{Y}$ where

$$\bar{\mathbf{y}}_j = \begin{bmatrix} \bar{y}_{j1\cdot} & \bar{y}_{j2\cdot} & \ldots & \bar{y}_{jp\cdot} \end{bmatrix}', \ j = 1, 2, ..., n, \tag{3.3}$$

$$\mathbf{Z} = \begin{bmatrix} z_{10} & z_{11} & z_{12} & \cdots & z_{1s_1} & z_{1(s_1+1)} & \cdots & z_{1s} \\ z_{20} & z_{21} & z_{22} & \cdots & z_{2s_1} & z_{2(s_1+1)} & \cdots & z_{2s} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ z_{n0} & z_{n1} & z_{n2} & \cdots & z_{ns_1} & z_{n(s_1+1)} & \cdots & z_{ns} \end{bmatrix} = \begin{bmatrix} \mathbf{z}_1' \\ \mathbf{z}_2' \\ \vdots \\ \mathbf{z}_n' \end{bmatrix}.$$

Denote the j^{th} row of $\mathbf{Z}$ where

$$\mathbf{z}_j = \begin{bmatrix} z_{j0} & z_{j1} & \dots & z_{js_1} & z_{j(s_1+1)} & \cdots & z_{js} \end{bmatrix}', \; j = 1, 2, ..., n, \tag{3.4}$$

$$\mathbf{B} = \begin{bmatrix} \beta_{01} & \beta_{02} & \cdots & \beta_{0p} \\ \beta_{11} & \beta_{12} & \cdots & \beta_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{s1} & \beta_{s2} & \cdots & \beta_{sp} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\beta}_0 & \boldsymbol{\beta}_1 & \dots & \boldsymbol{\beta}_s \end{bmatrix}'.$$

Denote the i^{th} column of $\mathbf{B}$ where

$$\boldsymbol{\beta}_q = \left(\beta_{q1} \quad \beta_{q2} \quad \dots \quad \beta_{qp} \right)', \; q = 0, 1, 2, ..., s, \tag{3.5}$$

$$\mathbf{E} = \begin{bmatrix} \bar{\varepsilon}_{11\cdot} & \bar{\varepsilon}_{12\cdot} & \cdots & \bar{\varepsilon}_{1p\cdot} \\ \bar{\varepsilon}_{21\cdot} & \bar{\varepsilon}_{22\cdot} & \cdots & \bar{\varepsilon}_{2p\cdot} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{\varepsilon}_{n1\cdot} & \bar{\varepsilon}_{n2\cdot} & \cdots & \bar{\varepsilon}_{np\cdot} \end{bmatrix} = \begin{bmatrix} \bar{\boldsymbol{\varepsilon}}_1' \\ \bar{\boldsymbol{\varepsilon}}_2' \\ \vdots \\ \bar{\boldsymbol{\varepsilon}}_n' \end{bmatrix}.$$

Denote the j^{th} row of $\mathbf{E}$ where

$$\bar{\boldsymbol{\varepsilon}}_j = \begin{bmatrix} \bar{\varepsilon}_{j1\cdot} & \bar{\varepsilon}_{j2\cdot} & \dots & \bar{\varepsilon}_{jp\cdot} \end{bmatrix}', \; j = 1, 2, ..., n,$$

Elements of the random error vector $\bar{\boldsymbol{\varepsilon}}_j$ are independent and identically distributed as $N_p\left(0, \boldsymbol{\Sigma}_j\right)$. The $p \times p$ variance-covariance matrix of $\bar{\boldsymbol{\varepsilon}}_j$ is assumed to be known and represented as

$$\boldsymbol{\Sigma}_j = \begin{bmatrix} \bar{\sigma}_{11j} & \cdots & \bar{\sigma}_{1pj} \\ \vdots & \ddots & \vdots \\ \bar{\sigma}_{p1j} & \cdots & \bar{\sigma}_{ppj} \end{bmatrix}, \; j = 1, 2, ..., n \; , \tag{3.6}$$

where $\bar{\sigma}_{ii'j} = \sigma_{ii'jk} / r_j$, $i, i' = 1, 2, \ldots, p$, $k = 1, 2, \ldots, r_j$.

Let $\mathbf{x}$ be a surrogate for independent variables $\mathbf{z}$ with the last $(s - s_1)$ variables imprecisely observed in model (3.1), and be expressed in vector notation as

$$\mathbf{X} = \left[\mathbf{x}_1 \quad \mathbf{x}_2 \quad \ldots \quad \mathbf{x}_n\right]', \tag{3.7}$$

$$\mathbf{x}_j = \left[z_{j0} \quad z_{j1} \quad \ldots \quad z_{js_1} \quad \bar{x}_{j(s_1+1)} \quad \bar{x}_{j(s_1+2)} \quad \ldots \quad \bar{x}_{js}\right]', \tag{3.8}$$

Table 3.1 Data Layout of the Study

j	k	y_{j1k}	y_{j2k}	$\cdots$	y_{jpk}	z_{j1}	z_{j2}	$\cdots$	z_{js_1}	$x_{j(s_1+1)k}$	$\cdots$	x_{jsk}
1	1	y_{111}	y_{121}		y_{1p1}	z_{11}	z_{12}		z_{1s_1}	$x_{1(s_1+1)1}$		x_{1s1}
	2	y_{112}	y_{122}		y_{1p2}					$x_{1(s_1+1)2}$		x_{1s2}
	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
	r_1	y_{11r_1}	y_{12r_1}		y_{1pr_1}					$x_{1(s_1+1)r_1}$		x_{1sr_1}
2	1	y_{211}	y_{221}		y_{2p1}	z_{21}	z_{22}		z_{2s_1}	$x_{2(s_1+1)1}$		x_{2s1}
	2	y_{212}	y_{222}		y_{2p2}					$x_{2(s_1+1)2}$		x_{2s2}
	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
	r_2	y_{21r_1}	y_{22r_2}		y_{2pr_2}					$x_{2(s_1+1)r_2}$		x_{2sr_2}
	$\vdots$						$\vdots$					
n	1	y_{n11}	y_{n21}		y_{np1}	z_{n1}	z_{n2}		z_{ns_1}	$x_{n(s_1+1)1}$		x_{ns1}
	2	y_{n12}	y_{n22}		y_{np2}					$x_{n(s_1+1)2}$		x_{ns2}
	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
	r_n	y_{n1r_n}	y_{n2r_n}		y_{npr_n}					$x_{n(s_1+1)r_n}$		x_{nsr_n}

3.2 Estimating Parameters by the CS Approach

Denote $\bullet(\mathbf{B},\mathbf{Z},\mathbf{Y})$ to be a function of $\mathbf{B}$ given $\mathbf{Z}$ and $\mathbf{Y}$ with independent variables Z observed precisely, and $\bullet(\mathbf{B},\mathbf{X},\mathbf{Y})$ to be a function of $\mathbf{B}$ given $\mathbf{Z}$ and $\mathbf{Y}$ with X as a surrogate for independent variables Z, which cannot be observed precisely, with repeated measurements. Then, the notations are specified as

$L(\mathbf{B},\mathbf{Z},\mathbf{Y})$ as the likelihood function of $\mathbf{B}$ given $\mathbf{Z}$ and $\mathbf{Y}$,

$l(\mathbf{B},\mathbf{Z},\mathbf{Y})$ as the log-likelihood function of $\mathbf{B}$ given $\mathbf{Z}$ and $\mathbf{Y}$, and

$U(\mathbf{B},\mathbf{Z},\mathbf{Y})$ as the score function of $\mathbf{B}$ given $\mathbf{Z}$ and $\mathbf{Y}$.

Because Z cannot be observed precisely, X is the observable value of Z with errors, and $U(\mathbf{B},\mathbf{X},\mathbf{Y})$ becomes the score function instead of $U(\mathbf{B},\mathbf{Z},\mathbf{Y})$. The main idea of the CS approach is to construct unbiased estimating functions on the observable variables Y and X by looking for the biased correction term to correct the biased score estimation function. The procedure of the CS approach consists of four stages:

Stage 1: Construct a Corrected Log-Likelihood Function

An unbiased estimating function in the absence of imprecisely measured variables is constructed by looking for a function $U_c(\mathbf{B},\mathbf{X},\mathbf{Y})$ which satisfies the property that

$$E\left(U_c(\mathbf{B},\mathbf{X},\mathbf{Y})\right)=U(\mathbf{B},\mathbf{Z},\mathbf{Y}). \tag{3.9}$$

In an indirect way, we use the log-likelihood and look for function $l_c(\mathbf{B},\mathbf{X},\mathbf{Y})$,

$$E\left(l_c(\mathbf{B},\mathbf{X},\mathbf{Y})\right)=l(\mathbf{B},\mathbf{Z},\mathbf{Y}). \tag{3.10}$$

A function $l_c(\mathbf{B},\mathbf{X},\mathbf{Y})$ is called a corrected log-likelihood function.

Stage 2: Evaluate the Corrected Score Functions of $\mathbf{B}$

Under regularity conditions, a corrected score function is given by

$$U_c(\mathbf{B},\mathbf{X},\mathbf{Y}) =: \frac{\partial l_c(\mathbf{B},\mathbf{X},\mathbf{Y})}{\partial \mathbf{B}}. \tag{3.11}$$

Stage 3: Estimate the variances of HME based on grouped heteroscedasticity

Under the assumption of grouped heteroscedasticity, σ_{ujq}^2 is homogeneous within g subsets of observations but heterogeneous across the subsets (Judge, Griffiths, Hill, Lütkepohl and Lee, 1985: 428). The observations can be separated into g groups of size n_h, $h = 1, 2, ..., g$, with homogeneous variance within the groups and heteroscedastic variances among them.

Stage 4: Evaluate the Corrected Score Estimators of $\mathbf{B}$

By following the unbiased property $E\left(U\left(\mathbf{B}, \mathbf{Z}, \mathbf{Y}\right)\right) = 0$ and $E\left(U_c\left(\mathbf{B}, \mathbf{X}, \mathbf{Y}\right)\right) = U\left(\mathbf{B}, \mathbf{Z}, \mathbf{Y}\right)$, so that $U_c\left(\mathbf{B}, \mathbf{X}, \mathbf{Y}\right)$ is also conditionally unbiased, and $E\left(U_c\left(\mathbf{B}, \mathbf{X}, \mathbf{Y}\right)\right) = 0$. Thus, by applying the general theory of M-estimation, the estimating equations $\sum_{j=1}^{n} U_c\left(\mathbf{B}, X_j, Y_j\right) = 0$ possess a consistent, asymptotically normal sequence of solutions (Nakamura, 1990).

Based on the model of the study described in (3.1) and (3.2), and following the 4 stages for estimating parameters by CS approach is given by

Stage 1: The corrected log-likelihood function is satisfied the following condition,

$$E\left[l_c\left(\mathbf{B}, \mathbf{x}_j, \bar{\mathbf{y}}_j\right)\right] = l_c\left(\mathbf{B}, \mathbf{z}_j, \bar{\mathbf{y}}_j\right). \tag{3.12}$$

The likelihood function $L\left(\mathbf{B}, \mathbf{Z}, \mathbf{Y}\right)$ of the study model is defined as

$$L\left(\mathbf{B}, \mathbf{Z}, \mathbf{Y}\right) = \prod_{j=1}^{n} \frac{1}{\left(2\pi\right)^{p/2}} \cdot \frac{1}{\left|\boldsymbol{\Sigma}_j\right|^{1/2}} \cdot \exp\left\{-\frac{1}{2}\left(\bar{\mathbf{y}}_j - \mathbf{B}'\mathbf{z}_j\right)' \boldsymbol{\Sigma}_j^{-1}\left(\bar{\mathbf{y}}_j - \mathbf{B}'\mathbf{z}_j\right)\right\}. \tag{3.13}$$

By substituting $\mathbf{Z}$ and $\mathbf{B}$ using the vector notations in (3.3) and (3.4) into the likelihood function in (3.13), the log likelihood function $l\left(\mathbf{B}, \mathbf{z}_j, \bar{\mathbf{y}}_j\right)$ becomes

$$l\left(\mathbf{B}, \mathbf{z}_j, \bar{\mathbf{y}}_j\right) = c_1 + c_2 - \frac{1}{2}\left(\bar{\mathbf{y}}_j - \mathbf{B}'\mathbf{z}_j\right)' \boldsymbol{\Sigma}_j^{-1}\left(\bar{\mathbf{y}}_j - \mathbf{B}'\mathbf{z}_j\right). \tag{3.14}$$

where $c_1 = -\frac{p}{2}\log\left(2\pi\right)$ and $c_2 = -\frac{1}{2}\log\left|\boldsymbol{\Sigma}_j\right|$.

Substituting $\mathbf{z}_j$ in (3.12) into (3.14) and taking the expectation by using the relationship $E\left(\bar{x}_{jq\cdot}\right)=z_{jq}$, $q=s_1+1,s_1+2,...,s$ yields

$$E\left[l\left(\mathbf{B},\mathbf{x}_j,\bar{\mathbf{y}}_j\right)\right]=c_1+c_2-\frac{1}{2}\left\{\left(\bar{\mathbf{y}}_j-\mathbf{B}'\mathbf{z}_j\right)'\boldsymbol{\Sigma}_j^{-1}\left(\bar{\mathbf{y}}_j-\mathbf{B}'\mathbf{z}_j\right)+\sum_{q=s-f+1}^{s}\boldsymbol{\beta}_q'\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\beta}_q\left(\sigma_{ujq}^2/r_j\right)\right\}.$$

(3.15)

Stage 2: The CS functions of $\mathbf{B}$ under regularity conditions, which yields a CS function given by

$$U_{c\,j\boldsymbol{\beta}_q}\left(\mathbf{B},\mathbf{x}_j,\bar{\mathbf{y}}_j\right)=\frac{\partial l_c\left(\mathbf{B},\mathbf{x}_j,\bar{\mathbf{y}}_j\right)}{\partial\boldsymbol{\beta}_q},\ q=0,1,...,s. \tag{3.16}$$

Under equation (3.16), the corrected log-likelihood function of the study model can be written as

$$l_c\left(\mathbf{B},\mathbf{x}_j,\bar{\mathbf{y}}_j\right)=c_1+c_2-\frac{1}{2}\left\{\left(\bar{\mathbf{y}}_j-\mathbf{B}'\mathbf{x}_j\right)'\boldsymbol{\Sigma}_j^{-1}\left(\bar{\mathbf{y}}_j-\mathbf{B}'\mathbf{x}_j\right)-\sum_{q=s_1+1}^{s}\boldsymbol{\beta}_q'\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\beta}_q\left(\sigma_{ujq}^2/r_j\right)\right\}. \tag{3.17}$$

Stage 3: Let the number of groups be g; the size of the h^{th} group be n_h, $h=1,2,...,g$; σ_{uqh}^2 be the h^{th} heteroscedastic variance of the ME of X_q in the h^{th} group; and r_h be the number of repeated measurements of each observation in the h^{th} group, then the sample variance of X_q in the h^{th} group is given by

$$S_{uqh}^2=\frac{\sum_{j_h=1}^{n_h}\sum_{k=1}^{r_h}\left(x_{qj_hk}-\bar{x}_{qj_h\cdot}\right)^2}{n_hr_h} \tag{3.18}$$

Stage 4: The CS estimators of $\mathbf{B}$ are determined by

$$\sum_{j=1}^{n}U_{c\,j\hat{\boldsymbol{\beta}}_q}\left(\hat{\mathbf{B}},\mathbf{x}_j,\bar{\mathbf{y}}_j\right)=0,\ q=0,1,...,s, \tag{3.19}$$

where $\hat{\mathbf{B}} = \left[\hat{\boldsymbol{\beta}}_0 \quad \hat{\boldsymbol{\beta}}_1 \quad \ldots \quad \hat{\boldsymbol{\beta}}_s \right]'$ is a matrix of the CS estimators of $\mathbf{B}$ and $\hat{\boldsymbol{\beta}}_q = \left[\hat{\beta}_{q1} \quad \hat{\beta}_{q2} \quad \cdots \quad \hat{\beta}_{qp} \right]'$, $q = 0,1,2,\ldots,s$, with a consistent, asymptotically normal sequence of solutions.

It can be deduced from (3.11) and (3.19) that the CS score function can be expressed as

$$U_{c_{j\hat{\boldsymbol{\beta}}_q}}\left(\hat{\mathbf{B}}, \mathbf{x}_j, \bar{\mathbf{y}}_j\right) = \begin{cases} \boldsymbol{\Sigma}_j^{-1}\left[\left(\bar{\mathbf{y}}_j - \mathbf{B}'\mathbf{x}_j\right) z_{jq}\right], q = 0,1,\ldots,(s-f), \\ \boldsymbol{\Sigma}_j^{-1}\left[\left(\bar{\mathbf{y}}_j - \mathbf{B}'\mathbf{x}_j\right)\bar{x}_{jq\cdot} + \boldsymbol{\beta}_q\left(\sigma_{ujq}^2 / r_j\right)\right], q = s_1+1, s_1+2,\ldots,s. \end{cases} \tag{3.20}$$

From (3.12), the $p(s+1)$ estimating equations can be written as

$$\begin{aligned}
\sum_{j=1}^{n} \boldsymbol{\Sigma}_j^{-1}\left(\hat{\mathbf{B}}'\mathbf{x}_j\right) &= \sum_{j=1}^{n} \boldsymbol{\Sigma}_j^{-1}\bar{\mathbf{y}}_j \\
\sum_{j=1}^{n} z_{j1}\boldsymbol{\Sigma}_j^{-1}\left(\hat{\mathbf{B}}'\mathbf{x}_j\right) &= \sum_{j=1}^{n} \boldsymbol{\Sigma}_j^{-1} z_{j1}\bar{\mathbf{y}}_j \\
&\vdots \\
\sum_{j=1}^{n} z_{js_1}\boldsymbol{\Sigma}_j^{-1}\left(\hat{\mathbf{B}}'\mathbf{x}_j\right) &= \sum_{j=1}^{n} \boldsymbol{\Sigma}_j^{-1} z_{js_1}\bar{\mathbf{y}}_j \\
\sum_{j=1}^{n}\left(\bar{x}_{j(s_1+1)\cdot}\boldsymbol{\Sigma}_j^{-1}\left(\hat{\mathbf{B}}'\mathbf{x}_j\right) + \left(\sigma_{uj(s_1+1)}^2 \big/ r_j\right)\hat{\boldsymbol{\beta}}_{(s_1+1)}\right) &= \sum_{j=1}^{n} \boldsymbol{\Sigma}_j^{-1}\bar{x}_{j(s_1+1)\cdot}\bar{\mathbf{y}}_j \\
&\vdots \\
\sum_{j=1}^{n}\left(\bar{x}_{js\cdot}\boldsymbol{\Sigma}_j^{-1}\left(\hat{\mathbf{B}}'\mathbf{x}_j\right) + \left(\sigma_{ujs}^2 \big/ r_j\right)\hat{\boldsymbol{\beta}}_s\right) &= \sum_{j=1}^{n} \boldsymbol{\Sigma}_j^{-1}\bar{x}_{js\cdot}\bar{\mathbf{y}}_j\ .
\end{aligned} \tag{3.21}$$

Solving the $p(s+1)$ linear equations in (3.21) yields

$$vec\left(\hat{\mathbf{B}}_{cs}\right) = \left[\left\{\left(\mathbf{X}\otimes\mathbf{I}_p\right)'\mathbf{V}^{-1}\left(\mathbf{X}\otimes\mathbf{I}_p\right)\right\} + \mathbf{C}\right]^{-1}\left(\mathbf{X}\otimes\mathbf{I}_p\right)'\mathbf{V}^{-1} vec\left(\mathbf{Y}'\right), \tag{3.22}$$

which can be expressed as

$$vec\left(\hat{\mathbf{B}}_{cs}\right) = \left[\mathbf{I}_{p(s+1)} - \boldsymbol{\Psi}\right] vec\left(\hat{\mathbf{B}}_{ols/hme}\right), \tag{3.23}$$

where $\boldsymbol{\Psi} = \left(\left(\left(\mathbf{X} \otimes \mathbf{I}_p \right)' \mathbf{V}^{-1} \left(\mathbf{X} \otimes \mathbf{I}_p \right) \right)^{-1} \mathbf{C} \left(\mathbf{I}_{p(s+1)} + \left(\left(\mathbf{X} \otimes \mathbf{I}_p \right)' \mathbf{V}^{-1} \left(\mathbf{X} \otimes \mathbf{I}_p \right) \right)^{-1} \mathbf{C} \right)^{-1} \right.$,

$\mathbf{V}^{-1}$ is a block-diagonal matrix of size np, $diag\left(\boldsymbol{\Sigma}_1^{-1} \quad \boldsymbol{\Sigma}_2^{-1} \quad \ldots \quad \boldsymbol{\Sigma}_n^{-1} \right)$, $\mathbf{C}$ is a block-diagonal matrix of size $p\left(s+1\right)$ where the first $\left(s_1+1\right)$ diagonal square submatrices of size p are zero and the last $\left(s-s_1\right)$ diagonal square submatrices of size p are the estimates of $-\sum_{j=1}^{n} \boldsymbol{\Sigma}_j^{-1} \left(\sigma^2_{uj(s_1+1)} / r_j \right), \ldots, -\sum_{j=1}^{n} \boldsymbol{\Sigma}_j^{-1} \left(\sigma^2_{ujs} / r_j \right)$, respectively, and

$$vec\left(\hat{\mathbf{B}}_{ols/hme} \right) = \left(\left(\mathbf{X} \otimes \mathbf{I}_p \right)' \mathbf{V}^{-1} \left(\mathbf{X} \otimes \mathbf{I}_p \right) \right)^{-1} \left(\mathbf{X} \otimes \mathbf{I}_p \right)' \mathbf{V}^{-1} vec\left(\mathbf{Y}' \right). \qquad (3.24)$$

Consider the special case where the covariance matrix of the random error is invariant, i.e. $\boldsymbol{\Sigma}_j = \boldsymbol{\Sigma}, \forall j = 1, 2, \ldots, n$. In this case, the term $\left(\mathbf{X} \otimes \mathbf{I}_p \right)' \mathbf{V}^{-1}$ in (3.22) can be reduced to

$$\left(\mathbf{X} \otimes \mathbf{I}_p \right)' \mathbf{V}^{-1} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ z_{11} & z_{21} & \cdots & z_{n1} \\ \vdots & \vdots & \ddots & \vdots \\ z_{1s_1} & z_{2s_1} & \cdots & z_{ns_1} \\ \bar{x}_{1(s_1+1)\cdot} & \bar{x}_{2(s_1+1)\cdot} & \cdots & \bar{x}_{n(s_1+1)\cdot} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{x}_{1s\cdot} & \bar{x}_{2s\cdot} & \cdots & \bar{x}_{ns\cdot} \end{bmatrix} \otimes \boldsymbol{\Sigma}^{-1}. \qquad (3.25)$$

which leads to expressing the term $\left(\mathbf{X} \otimes \mathbf{I}_p \right)' \mathbf{V}^{-1} \left(\mathbf{X} \otimes \mathbf{I}_p \right) + \mathbf{C}$ as

$$\left(\mathbf{X}\otimes\mathbf{I}_p\right)'\mathbf{V}^{-1}\left(\mathbf{X}\otimes\mathbf{I}_p\right)+\mathbf{C}$$

$$=\begin{bmatrix} n & \sum_{j=1}^{n} z_{j1} & \cdots & \sum_{j=1}^{n} z_{js_1} & \sum_{j=1}^{n}\bar{x}_{j(s_1+1)\cdot} & \cdots & \sum_{j=1}^{n}\bar{x}_{js\cdot} \\ \sum_{j=1}^{n} z_{j1} & \sum_{j=1}^{n} z_{j1}^2 & \cdots & \sum_{j=1}^{n} z_{j1}z_{js_1} & \sum_{j=1}^{n} z_{j1}\bar{x}_{j(s_1+1)\cdot} & \cdots & \sum_{j=1}^{n} z_{j1}\bar{x}_{js\cdot} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \sum_{j=1}^{n} z_{js_1} & \sum_{j=1}^{n} z_{js_1}z_{j1} & \cdots & \sum_{j=1}^{n} z_{js_1}^2 & \sum_{j=1}^{n} z_{js_1}\bar{x}_{j(s_1+1)\cdot} & \cdots & \sum_{j=1}^{n} z_{js_1}\bar{x}_{js\cdot} \\ \sum_{j=1}^{n}\bar{x}_{j(s_1+1)\cdot} & \sum_{j=1}^{n}\bar{x}_{j(s_1+1)\cdot}z_{j1} & \cdots & \sum_{j=1}^{n}\bar{x}_{j(s_1+1)\cdot}z_{js_1} & \sum_{j=1}^{n}\bar{x}_{j(s_1+1)\cdot}^2-\sum_{j=1}^{n}\left(S_{uj(s_1+1)}^2/r_j\right) & \cdots & \sum_{j=1}^{n}\bar{x}_{j(s_1+1)\cdot}\bar{x}_{js\cdot} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \sum_{j=1}^{n}\bar{x}_{js\cdot} & \sum_{j=1}^{n}\bar{x}_{js\cdot}z_{j1} & \cdots & \sum_{j=1}^{n}\bar{x}_{js\cdot}z_{js_1} & \sum_{j=1}^{n}\bar{x}_{js\cdot}\bar{x}_{j(s_1+1)\cdot} & \cdots & \sum_{j=1}^{n}\bar{x}_{js\cdot}^2-\sum_{j=1}^{n}\left(S_{ujs}^2/r_j\right) \end{bmatrix}\otimes\mathbf{\Sigma}^{-1}.$$

$$=\left(\mathbf{X}'\mathbf{X}+\mathbf{C}_u\right)\otimes\mathbf{\Sigma}^{-1}, \tag{3.26}$$

where $\mathbf{C}_u$ is a diagonal matrix of size $(s+1)$ where the first (s_1+1) diagonal elements are zero and the last $(s-s_1)$ elements are the estimates of $-\sum_{j=1}^{n}\left(\sigma_{uj(s_1+1)}^2/r_j\right),-\sum_{j=1}^{n}\left(\sigma_{uj(s_1+2)}^2/r_j\right),\ldots,-\sum_{j=1}^{n}\left(\sigma_{ujs}^2/r_j\right)$.

From (3.26), the inverse of $\left(\mathbf{X}\otimes\mathbf{I}_p\right)'\mathbf{V}^{-1}\left(\mathbf{X}\otimes\mathbf{I}_p\right)+\mathbf{C}$ can be expressed as

$$\left(\left(\mathbf{X}\otimes\mathbf{I}_p\right)'\mathbf{V}^{-1}\left(\mathbf{X}\otimes\mathbf{I}_p\right)+\mathbf{C}\right)^{-1}=\left(\mathbf{X}'\mathbf{X}+\mathbf{C}_u\right)^{-1}\otimes\mathbf{\Sigma}. \tag{3.27}$$

For simplicity of notation, let $vec(\mathbf{F})=\left(\mathbf{X}\otimes\mathbf{I}_p\right)'\mathbf{V}^{-1}vec(\mathbf{Y}')$. Substituting in (3.25) into the right hand side of the definition of $vec(\mathbf{F})$ yields

$$vec(\mathbf{F})=vec\left[\mathbf{\Sigma}^{-1}\begin{pmatrix}\sum_{j=1}^{n} z_{j0}\bar{\mathbf{y}}_j & \sum_{j=1}^{n} z_{j1}\bar{\mathbf{y}}_j & \cdots & \sum_{j=1}^{n} z_{js_1}\bar{\mathbf{y}}_j & \sum_{j=1}^{n}\bar{x}_{j(s_1+1)}\bar{\mathbf{y}}_j & \cdots & \sum_{j=1}^{n}\bar{x}_{js}\bar{\mathbf{y}}_j\end{pmatrix}\right]. \tag{3.28}$$

Substituting (3.27) and (3.28) into (3.22) gives

$$vec\left(\hat{\mathbf{B}}_{cs}\right)=vec\left(\mathbf{\Sigma F}\left(\mathbf{X'X}+\mathbf{C}_u\right)^{-1}\right)$$

$$=vec\left(\left(\sum_{j=1}^{n} z_{j0}\bar{\mathbf{y}}_j \quad \sum_{j=1}^{n} z_{j1}\bar{\mathbf{y}}_j \quad \cdots \quad \sum_{j=1}^{n} z_{js_1}\bar{\mathbf{y}}_j \quad \sum_{j=1}^{n} \bar{x}_{j(s_1+1)}\bar{\mathbf{y}}_j \quad \cdots \quad \sum_{j=1}^{n} \bar{x}_{js}\bar{\mathbf{y}}_j\right)\left(\mathbf{X'X}+\mathbf{C}_u\right)^{-1}\right),$$

(3.29)

where the i^{th} CS estimator, $\hat{\boldsymbol{\beta}}_{i\_cs}$, $i=1,2,\ldots,p$ can be expressed as

$$\hat{\boldsymbol{\beta}}_{i\_cs}=\left(\mathbf{X'X}+\mathbf{C}_u\right)^{-1}\mathbf{X'}\bar{\mathbf{y}}_i\,. \tag{3.30}$$

which can be reduced the same results given by Giménez and Patat (2005) when the MEs are homogeneous. In this study, the heterogeneous variance of the MEs is unknown and estimated by the pooled sample variance. In the case of grouped heteroscedasticity, the observations are grouped into several subsets such that the variance of MEs is homogeneous within a group but heterogeneous across the groups (Judge, Griffiths, Hill, Lütkepohl and Lee, 1985). Therefore, the estimates of the q^{th} diagonal element of $\mathbf{C}_u$, $-\sum_{j=1}^{n}\left(\sigma_{ujq}^2/r_j\right)$, can be written as $-\sum_{h=1}^{g} n_h S_{uqh}^2/r_h$, $q=s_1+1,\ s_1+2,\ \ldots,\ s$ where S_{uqh}^2 is evaluated from (3.18).

The estimator directly obtained by the OLS method without score correcting is the case in (3.30) where $\mathbf{C}_u$ is a zero matrix. Following this, the $\hat{\boldsymbol{\beta}}_{i\_ols/hme}$ estimator can be expressed as

$$\hat{\boldsymbol{\beta}}_{i\_ols/hme}=\left(\mathbf{X'X}\right)^{-1}\mathbf{X'}\bar{\mathbf{y}}_i\,, \tag{3.31}$$

which is the OLS estimator obtained by substituting X_q for unobservable Z_q, $q=s_1+1,\ s_1+2,\ \ldots,\ s$ without correcting the HMEs.

Consider the specific case where $s_1=1$, $s=2$, i.e. the independent variable Z_2 is imprecisely measured by X_2 with HME and the covariance matrix of the random error is invariant, $\Sigma_j=\Sigma, \forall j=1,2,...,n$. From (3.21), the CS estimators of $\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2$ can be expressed as

$$\sum_{j=1}^{n} \boldsymbol{\Sigma}^{-1}\mathbf{b}_0 + \sum_{j=1}^{n} \boldsymbol{\Sigma}^{-1} z_{j1}\mathbf{b}_1 + \sum_{j=1}^{n} \boldsymbol{\Sigma}^{-1}\bar{x}_{j2\cdot}\mathbf{b}_2 = \sum_{j=1}^{n} \boldsymbol{\Sigma}^{-1}\bar{\mathbf{y}}_j$$

$$\sum_{j=1}^{n} \boldsymbol{\Sigma}^{-1} z_{j1}\mathbf{b}_0 + \sum_{j=1}^{n} \boldsymbol{\Sigma}^{-1} z_{j1}^2\mathbf{b}_1 + \sum_{j=1}^{n} \boldsymbol{\Sigma}^{-1} z_{j1}\bar{x}_{j2\cdot}\mathbf{b}_2 = \sum_{j=1}^{n} \boldsymbol{\Sigma}^{-1} z_{j1}\bar{\mathbf{y}}_j$$

$$\sum_{j=1}^{n} \boldsymbol{\Sigma}^{-1}\bar{x}_{j2\cdot}\mathbf{b}_0 + \sum_{j=1}^{n} \boldsymbol{\Sigma}^{-1} z_{j1}\bar{x}_{j2\cdot}\mathbf{b}_1 + \sum_{j=1}^{n} \boldsymbol{\Sigma}^{-1}\left(\bar{x}_{j2\cdot}^2 - \frac{\hat{\sigma}_{uj2}^2}{r_j}\right)\mathbf{b}_2 = \sum_{j=1}^{n} \boldsymbol{\Sigma}^{-1}\bar{x}_{j2\cdot}\bar{\mathbf{y}}_j \;. \qquad (3.32)$$

Solving the $3p$ linear equations in (3.32) yields

$$vec\left(\hat{\mathbf{B}}_{cs}\right) = vec\left(\left(\sum_{j=1}^{n} z_{j0}\bar{\mathbf{y}}_j \quad \sum_{j=1}^{n} z_{j1}\bar{\mathbf{y}}_j \quad \sum_{j=1}^{n} \bar{x}_{j2\cdot}\bar{\mathbf{y}}_j\right)\left(\mathbf{X}'\mathbf{X} + \hat{\boldsymbol{\Sigma}}_u\right)^{-1}\right). \qquad (3.33)$$

The i^{th} vector in $vec(\hat{\mathbf{B}}_{cs})$ in (2.34) can be written as

$$\hat{\boldsymbol{\beta}}_{i\_cs} = \left(\mathbf{X}'\mathbf{X} + \mathbf{C}_u\right)^{-1}\mathbf{X}'\bar{\mathbf{y}}_i\,, \qquad (3.34)$$

where $\hat{\boldsymbol{\beta}}_{i\_cs} = \left[\hat{\beta}_{0i} \quad \hat{\beta}_{1i} \quad \hat{\beta}_{2i}\right]'$, $\bar{\mathbf{y}}_i = \left[\bar{y}_{1i\cdot} \quad \bar{y}_{2i\cdot} \quad \cdots \quad \bar{y}_{ni\cdot}\right]'$, $i = 1, 2, ..., p$,

$$\mathbf{X} = \begin{bmatrix} 1 & z_{11} & \bar{x}_{12\cdot} \\ 1 & z_{21} & \bar{x}_{22\cdot} \\ \vdots & \vdots & \vdots \\ 1 & z_{n1} & \bar{x}_{n2\cdot} \end{bmatrix}, \quad \mathbf{C}_u = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -A \end{bmatrix}, \text{ and } A = \sum_{h=1}^{g} \frac{n_h S_{u2h}^2}{r_h}\,.$$

3.3 The Bias of the CS Estimator

The bias and variance of the CS estimator with p correlations of dependent variables $Y_1, Y_2, ..., Y_p$ corresponding to s independent variables, where the first s_1 independent variables $Z_1, Z_2, ..., Z_{s_1}$ are precisely observed and the last $\left(s - s_1\right)$ independent variables, $Z_{s_1+1}, Z_{s_1+2}, ..., Z_s$ are imprecisely observed, are derived here.

Equation (3.22) can be expressed as

$$vec\left(\hat{\mathbf{B}}_{cs}\right)$$

$$=\left[\left(\left(\mathbf{X}\otimes\mathbf{I}_p\right)'\mathbf{V}^{-1}\left(\mathbf{X}\otimes\mathbf{I}_p\right)\right)^{-1}-\mathbf{\Psi}\left(\left(\mathbf{X}\otimes\mathbf{I}_p\right)'\mathbf{V}^{-1}\left(\mathbf{X}\otimes\mathbf{I}_p\right)\right)^{-1}\right]\left(\mathbf{X}\otimes\mathbf{I}_p\right)'\mathbf{V}^{-1}vec\left(\mathbf{Y}'\right)$$

$$vec\left(\hat{\mathbf{B}}_{cs}\right)=vec\left(\hat{\mathbf{B}}_{ols}\right)-\mathbf{\Psi}vec\left(\hat{\mathbf{B}}_{ols}\right). \tag{3.35}$$

The bias and variance of the CS estimators are given by

$$Bias\left[vec\left(\hat{\mathbf{B}}_{cs}\right)\right]=Bias\left[vec\left(\hat{\mathbf{B}}_{ols/hme}\right)\right]-E\left[\left(\mathbf{\Psi}\right)vec\left(\hat{\mathbf{B}}_{ols/hme}\right)\right], \tag{3.36}$$

$$Var\left[vec\left(\hat{\mathbf{B}}_{cs}\right)\right]=Var\left[\left\{\mathbf{I}_{p(s+1)}-\mathbf{\Psi}\right\}vec\left(\hat{\mathbf{B}}_{ols/hme}\right)\right], \tag{3.37}$$

Lemma 1 In a grouped heteroscedasticity, the n observations can be grouped into h groups such that the variance of the measurement errors, σ^2_{u2h}, is homogeneous within the h^{th} group but heterogeneous across the groups. Let r_j be the number of repeated measurements of the j^{th} observation and u_{j2k} be the random measurement error of the j^{th} observation of x_{j2} in the k^{th} repeated measurement independently distributed as $N(0,\sigma^2_{uj2})$. Then, as $n\to\infty$,

$$\sum_{j=1}^{n}\bar{u}^2_{j2\cdot}=n\left(S^2_{\bar{u}_2}+\bar{u}^2_2\right)\to\sum_{h=1}^{g}\frac{n_h\sigma^2_{u2h}}{r_h}.$$

Proof. The proof of Lemma 1 is shown in Appendix A.

Assumptions: (A1) $z_1, z_2, \varepsilon_{jik}$, and u_{j2k} are independent,

(A2) $\sigma^2_{\bar{u}_2}\ll\sigma^2_{z_2}$, and $\bar{u}^2_2\ll\sigma^2_{z_2}$.

Theorem 1 In the linear multivariate measurement error regression model described in (3.1) and (3.2) where $s_1 = 1$ and $s = 2$, $\sigma^2_{\bar{u}_2} \ll \sigma^2_{z_2}$ and $\bar{u}^2_2 \ll \sigma^2_{z_2}$, $\hat{\beta}_{1i\_cs}$ is an unbiased estimator but $\hat{\beta}_{0i\_cs}$ and $\hat{\beta}_{2i\_cs}$ are asymptotically unbiased estimators, $i = 1, 2, \ldots, p$.

Proof.

The bias of the $\hat{\boldsymbol{\beta}}_{i\_cs}$ can be written from (3.34) as

$$\begin{aligned} \textit{Bias of } \hat{\boldsymbol{\beta}}_{i\_cs} &= E\left[\left(\mathbf{X}'\mathbf{X} + \hat{\boldsymbol{\Sigma}}_{\mathrm{u}}\right)^{-1} \mathbf{X}'\bar{\mathbf{y}}_{i\cdot}\right] - \boldsymbol{\beta}_i \\ &= E\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}'\bar{\mathbf{y}}_{i\cdot} - \left(\mathbf{X}'\mathbf{X}\right)^{-1} \hat{\boldsymbol{\Sigma}}_{\mathrm{u}} \left[\mathbf{I} + \left(\mathbf{X}'\mathbf{X}\right)^{-1} \hat{\boldsymbol{\Sigma}}_{\mathrm{u}}\right]^{-1} \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}'\bar{\mathbf{y}}_{i\cdot}\right] - \boldsymbol{\beta}_i \end{aligned} \tag{3.38}$$

The bias of the $\hat{\boldsymbol{\beta}}_{i\_ols/hme}$ estimator from (3.31) can be written as

$$\begin{aligned} \textit{Bias of } \hat{\boldsymbol{\beta}}_{i\_ols/hme} &= E\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}'\bar{\mathbf{y}}_{i}\right] - \boldsymbol{\beta}_i \\ &= E\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}'\left(\mathbf{X}\boldsymbol{\beta}_i + \mathbf{v}_i\right)\right] - \boldsymbol{\beta}_i \\ &= E\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}'\mathbf{v}_i\right], \end{aligned} \tag{3.39}$$

where $\mathbf{v}_i = \bar{\boldsymbol{\varepsilon}}_i - \beta_{2i}\bar{\mathbf{u}}_2$. Then the bias of the $\hat{\boldsymbol{\beta}}_{i\_cs}$ estimator can be expressed in terms of $\hat{\boldsymbol{\beta}}_{i\_ols/hme}$ and its associated bias as

$$\textit{Bias of } \hat{\boldsymbol{\beta}}_{i\_cs} = \textit{Bias of } \hat{\boldsymbol{\beta}}_{i\_ols/hme} - E\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{C}_{\mathrm{u}} \left[\mathbf{I} + \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{C}_{\mathrm{u}}\right]^{-1} \hat{\boldsymbol{\beta}}_{i\_ols/hme}\right]. \tag{3.40}$$

From the definition of $\mathbf{X}$ in (3.30) and by using the independent property of $\mathbf{z}_1$ and $\mathbf{x}_2$, the inverse of $\mathbf{X}'\mathbf{X}$ can be expressed in terms of the statistics of the observations as

$$\left(\mathbf{X}'\mathbf{X}\right)^{-1}=\frac{1}{n}\begin{bmatrix}\frac{S_{z_1}^2+\bar{z}_1^2}{S_{z_1}^2}+\frac{\left(\bar{z}_2+\bar{u}_2\right)^2}{S_{z_2}^2+S_{\bar{u}_2}^2} & -\frac{\bar{z}_1}{S_{z_1}^2} & -\frac{\left(\bar{z}_2+\bar{u}_2\right)}{S_{z_2}^2+S_{\bar{u}_2}^2}\\ -\frac{\bar{z}_1}{S_{z_1}^2} & \frac{1}{S_{z_1}^2} & 0\\ -\frac{\left(\bar{z}_2+\bar{u}_2\right)}{S_{z_2}^2+S_{\bar{u}_2}^2} & 0 & \frac{1}{S_{z_2}^2+S_{\bar{u}_2}^2}\end{bmatrix},$$

which is denoted by

$$\left(\mathbf{X}'\mathbf{X}\right)^{-1}=\begin{bmatrix}q^{11} & q^{12} & q^{13}\\ q^{21} & q^{22} & q^{23}\\ q^{31} & q^{32} & q^{33}\end{bmatrix}. \qquad (3.41)$$

From (3.39) and the definition of $\mathbf{v}_i$, the term $\mathbf{X}'\mathbf{v}_i$ can be obviously expressed as

$$\mathbf{X}'\mathbf{v}_i=\begin{bmatrix}\sum_{j=1}^{n}\left(\bar{\varepsilon}_{ji\cdot}-\beta_{2i}\bar{u}_{j2\cdot}\right)\\ \sum_{j=1}^{n}z_{j1}\left(\bar{\varepsilon}_{ji\cdot}-\beta_{2i}\bar{u}_{j2\cdot}\right)\\ \sum_{j=1}^{n}\bar{x}_{j2\cdot}\left(\bar{\varepsilon}_{ji\cdot}-\beta_{2i}\bar{u}_{j2\cdot}\right)\end{bmatrix},$$

which is denoted by

$$\mathbf{X}'\mathbf{v}_i=\begin{bmatrix}d_1\\ d_2\\ d_3\end{bmatrix}. \qquad (3.42)$$

By substituting (3.41) and (3.42) into (3.39), the bias of the $\hat{\boldsymbol{\beta}}_{i\_ols/hme}$ estimator becomes

$$\begin{bmatrix}Bias\ of\ \hat{\beta}_{0i\_ols/hme}\\ Bias\ of\ \hat{\beta}_{1i\_ols/hme}\\ Bias\ of\ \hat{\beta}_{2i\_ols/hme}\end{bmatrix}=E\begin{bmatrix}q^{11}d_1+q^{12}d_2+q^{13}d_3\\ q^{21}d_1+q^{22}d_2+q^{23}d_3\\ q^{31}d_1+q^{32}d_2+q^{33}d_3\end{bmatrix}. \qquad (3.43)$$

From (3.41), it can be easily seen that the term $\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{C}_u = \begin{bmatrix} 0 & 0 & -Aq^{13} \\ 0 & 0 & -Aq^{23} \\ 0 & 0 & -Aq^{33} \end{bmatrix}$ which

leads to the expression of the last term $E\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{C}_u\left[\mathbf{I}+\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{C}_u\right]^{-1}\right]$ as

$$E\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{C}_u\left(\mathbf{I}+\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{C}_u\right)^{-1}\right] = E\left(\frac{1}{1-Aq^{33}}\right)\begin{bmatrix} Aq^{13}\hat{\beta}_{2i\_ols/hme} \\ Aq^{23}b_{2i\_ols/hme} \\ Aq^{33}b_{2i\_ols/hme} \end{bmatrix}. \qquad (3.44)$$

Substituting the bias of $\hat{\boldsymbol{\beta}}_{i\_ols/hme}$ in (3.43) and the RHS in (3.44) into (3.40) yields

$$\begin{bmatrix} Bias\ of\ \hat{\beta}_{0i\_cs} \\ Bias\ of\ \hat{\beta}_{1i\_cs} \\ Bias\ of\ \hat{\beta}_{2i\_cs} \end{bmatrix} = E\begin{bmatrix} q^{11}d_1+q^{12}d_2+q^{13}d_3 \\ q^{21}d_1+q^{22}d_2+q^{23}d_3 \\ q^{31}d_1+q^{32}d_2+q^{33}d_3 \end{bmatrix} + E\left[\left(\frac{1}{1-Aq^{33}}\right)\begin{bmatrix} Aq^{13}\hat{\beta}_{2i\_ols/hme} \\ Aq^{23}\hat{\beta}_{2i\_ols/hme} \\ Aq^{33}\hat{\beta}_{2i\_ols/hme} \end{bmatrix}\right].$$

(3.45)

From the definitions of q^{ij} and d_i in (3.41) and (3.42) respectively, the bias of $\hat{\boldsymbol{\beta}}_{i\_ols/hme}$ in (3.43) can be expressed in terms of statistical properties of the variables as

$$Bias\ of\ \hat{\beta}_{0i\_ols/hme} = -\beta_{2i}E\left[\frac{\left(\bar{z}_2+\bar{u}_2\right)^2\bar{u}_2}{S_{z_2}^2\left(1+\frac{S_{\bar{u}_2}^2}{S_{z_2}^2}\right)}\right] + \beta_{2i}E\left[\frac{\bar{z}_2+\bar{u}_2}{S_{z_2}^2\left(1+\frac{S_{\bar{u}_2}^2}{S_{z_2}^2}\right)}\sum_{j=1}^{n}\frac{\bar{u}_{j2\cdot}^2}{n}\right].$$

(3.46)

By Lemma 1, the bias of $\hat{\beta}_{0i\_ols/hme}$ in (3.33) can be written after the first order approximation under the assumption: $S_{\bar{u}_2}^2 \ll S_{z_2}^2$ as

$$Bias\ of\ \hat{\beta}_{0i\_ols/hme} \simeq \frac{\beta_{2i}}{S_{z_2}^2}\left[\bar{z}_2\sigma_{\bar{u}_2}^2\left(1-\frac{2}{n}\right)+\frac{3\bar{z}_2}{nS_{z_2}^2}\left(\frac{2}{n}-1\right)+\frac{2\bar{z}_2}{n^2S_{z_2}^2}E\left(\sum_{j=1}^{n}\sum_{k=j+1}^{n}\bar{u}_{j2\cdot}^2\bar{u}_{k2\cdot}^2\right)\left(\frac{2}{n}-1\right)\right].$$

(3.47)

As $n \to \infty$, the bias of $\hat{\beta}_{0i\_ols/hme}$ in (3.47) approaches

$$Bias\ of\ \hat{\beta}_{0i\_ols/hme} \to \frac{\beta_{2i}\bar{z}_2\sigma^2_{\bar{u}_2}}{S^2_{z_2}}. \tag{3.48}$$

Now consider the bias of $\hat{\beta}_{0i\_cs}$ which is given in (3.45) as

$$Bias\ of\ \hat{\beta}_{0i\_cs} = Bias\ of\ \hat{\beta}_{0i\_ols/hme} + E\left(\frac{Aq^{13}\hat{\beta}_{2i\_ols/hme}}{1-Aq^{33}}\right). \tag{3.49}$$

Substituting the bias of $\hat{\beta}_{2i\_ols/hme}$ from (3.43) into the last term of (3.45) yields

$$E\left(\frac{Aq^{13}\hat{\beta}_{2i\_ols/hme}}{1-Aq^{33}}\right) = E\left[\frac{Aq^{13}\left(\beta_{2i}+q^{31}d_1+q^{32}d_2+q^{33}d_3\right)}{1-Aq^{33}}\right]. \tag{3.50}$$

From the definitions of q^{ij} and d_i in (3.41) and (3.42) respectively, the last term of (3.45) can be expressed in terms of statistical properties of the variables as

$$E\left(\frac{Aq^{13}\hat{\beta}_{2i\_ols/hme}}{1-Aq^{33}}\right)$$

$$\simeq \frac{1}{S^2_{z_2}}E\left[\begin{array}{l} -\bar{z}_2S^2_{\bar{u}_2}\beta_{2i} - \frac{(\bar{z}_2+\bar{u}_2)^2 S^2_{\bar{u}_2}\beta_{2i}\bar{u}_2}{S^2_{z_2}} + \frac{(\bar{z}_2+\bar{u}_2)^2\left(S^2_{\bar{u}_2}\right)^2\beta_{2i}\bar{u}_2}{\left(S^2_{z_2}\right)^2} + \frac{(\bar{z}_2+\bar{u}_2)S^2_{\bar{u}_2}\beta_{2i}}{S^2_{z_2}}\left(\sum\limits_{j=1}^{n}\frac{z_{j2}\bar{u}_{j2\cdot}}{n}\right) \\ -\frac{(\bar{z}_2+\bar{u}_2)\left(S^2_{\bar{u}_2}\right)^2}{\left(S^2_{z_2}\right)^2}\left(\sum\limits_{j=1}^{n}\frac{z_{j2}\bar{u}_{j2\cdot}}{n}\right) + \frac{(\bar{z}_2+\bar{u}_2)S^2_{\bar{u}_2}\beta_{2i}}{S^2_{z_2}} - \frac{(\bar{z}_2+\bar{u}_2)\left(S^2_{\bar{u}_2}\right)^3\beta_{2i}}{\left(S^2_{z_2}\right)^2} \end{array}\right]$$

(3.51)

It can be shown that under the condition $u_{j2k} \sim N(0,\sigma^2_{uj2})$

$E(\bar{u}_2S^2_{\bar{u}_2}) = 0;\ E(\bar{u}_2^2S^2_{\bar{u}_2}) \sim O(1/n^2);\ E(\bar{u}_2^3S^2_{\bar{u}_2}) \sim O(1/n^2);\ E(\bar{u}_2^2S^4_{\bar{u}_2}) \sim O(1/n^2);$

$E(S^6_{\bar{u}_2}) \sim O(1/n^3);\ E(S^4_{\bar{u}_2}) = 3/n + O(1/n^2);\ E(\bar{u}_2^2\, S^4_{\bar{u}_2}) = 0;\ E(S^2_{\bar{u}_2}\sum\limits_{j=1}^{n}\bar{u}_{j2.}) = 0;$

$E(S^2_{\bar{u}_2}\bar{u}_2\sum\limits_{j=1}^{n}\bar{u}_{j2.}) = 0;\ E(S^2_{\bar{u}_2}\bar{u}_2^2\sum\limits_{j=1}^{n}\bar{u}_{j2.}) \sim O(1/n^2).$

As $n \to \infty$, the last term of (3.49) approaches

$$E\left(\frac{Aq^{13}\hat{\beta}_{2i\_ols/hme}}{1-Aq^{33}}\right) \to -\frac{\beta_{2i}\bar{z}_2\sigma^2_{\bar{u}_2}}{S^2_{z_2}}. \tag{3.52}$$

Therefore, from (3.48) and (3.52), it can be concluded that the bias of $\hat{\beta}_{0i\_cs}$ approaches zero as $n \to \infty$.

Similarly, by following the same approach, and the bias of $\hat{\beta}_{1i\_cs}$ in (3.45) can be written as

$$Bias\ of\ \hat{\beta}_{1i\_cs} = E\left(q^{21}d_1 + q^{22}d_2 + q^{23}d_3\right) + E\left(\frac{Aq^{23}\hat{\beta}_{2i\_ols/hme}}{1-Aq^{33}}\right). \tag{3.53}$$

Under the assumption that z_{j1} and x_{j2k} are independent, $q^{23} = 0$ and $z_1, z_2, \varepsilon_{jik}$, and u_{j2k} are independent. Then, equation (3.53) becomes

$$Bias\ of\ \hat{\beta}_{1i\_cs} = E\left(q^{21}Q_1 + q^{22}Q_2\right)$$

$$= E\left[-\frac{\bar{z}_1}{nS^2_{z_1}}\sum_{j=1}^{n}\left(\bar{\varepsilon}_{ji\cdot} - \beta_{2i}\bar{u}_{j2\cdot}\right) + \frac{1}{nS^2_{z_1}}\sum_{j=1}^{n} z_{j1}\left(\bar{\varepsilon}_{ji\cdot} - \beta_{2i}\bar{u}_{j2\cdot}\right)\right]$$

$$= E\left[-\frac{\bar{z}_1}{S^2_{z_1}}\left(\bar{\varepsilon}_{i\cdot} - \beta_{2i}\bar{u}_2\right) + \frac{1}{S^2_{z_1}}\left(\sum_{j=1}^{n}\frac{z_{j1}\bar{\varepsilon}_{ji\cdot}}{n} - \beta_{2i}\sum_{j=1}^{n}\frac{z_{j1}\bar{u}_{j2\cdot}}{n}\right)\right] = 0.$$

Thus, $\hat{\beta}_{1i\_cs}$ is an unbiased estimator.

Next, Consider the bias of $\hat{\beta}_{2i\_cs}$ in (3.45) given by

$$Bias\ of\ \hat{\beta}_{2i\_cs} = E\left(q^{31}d_1 + q^{32}d_2 + q^{33}d_3\right) + E\left(\frac{Aq^{33}\hat{\beta}_{2i\_ols}}{1-Aq^{33}}\right). \tag{3.54}$$

The term $E\left(q^{31}d_1 + q^{32}d_2 + q^{33}d_3\right)$ in (3.54) can be expressed as $Bias\ of\ \hat{\beta}_{2i\_ols/hme}$ as follows

$Bias\ of\ \hat{\beta}_{2i\_ols/hme}$

$$= E\left[\begin{array}{l} -\dfrac{\bar{z}_2 + \bar{u}_2}{S_{z_2}^2 + S_{\bar{u}_2}^2}\left(\bar{\varepsilon}_{i\cdot} - \beta_{2i}\bar{u}_2\right) \\ +\dfrac{1}{S_{z_2}^2 + S_{\bar{u}_2}^2}\left(\sum_{j=1}^{n}\dfrac{z_{j2}\bar{\varepsilon}_{ji\cdot}}{n} + \sum_{j=1}^{n}\dfrac{\bar{u}_{j2\cdot}\bar{\varepsilon}_{ji\cdot}}{n} - \beta_{2i}\sum_{j=1}^{n}\dfrac{z_{j2}\bar{u}_{j2\cdot}}{n} - \beta_{2i}\sum_{j=1}^{n}\dfrac{\bar{u}_{j2\cdot}^2}{n}\right) \end{array}\right]. \tag{3.55}$$

Under the assumptions that $z_1, z_2, \varepsilon_{jik}$,and u_{j2k} are independent, $S_{\bar{u}_2}^2 \ll S_{z_2}^2, \bar{u}_2^2 \ll S_{z_2}^2$, $E\left(u_{jqk}\right) = 0$, $E\left(\boldsymbol{\varepsilon}_j\right) = 0$, and using Lemma 1, equation (3.55) yields

$Bias\ of\ \hat{\beta}_{2i\_ols/hme}$

$$\simeq -\frac{\beta_{2i}\sigma_{\bar{u}_2}^2}{S_{z_2}^2}\left(1 - \frac{1}{n}\right) + \frac{\beta_{2i}}{\left(S_{z_2}^2\right)^2}\left(1 - \frac{1}{n}\right)\left\{\frac{3}{n} + \frac{2}{n^2}E\left(\sum_{j=1}^{n}\sum_{k=j+1}^{n}\bar{u}_{j2\cdot}^2\bar{u}_{k2\cdot}^2\right)\right\}. \tag{3.56}$$

As $n \to \infty$, equation (3.56) approaches

$$Bias\ of\ \hat{\beta}_{2i\_ols/hme} \to -\frac{\beta_{2i}\sigma_{\bar{u}_2}^2}{S_{z_2}^2}. \tag{3.57}$$

Consider that the term $E\left(\dfrac{Aq^{33}\hat{\beta}_{2i\_ols/hme}}{1 - Aq^{33}}\right)$ in (3.54), can be expressed as

$$E\left(\frac{Aq^{33}\hat{\beta}_{2i\_ols/hme}}{1 - Aq^{33}}\right) = E\left[\frac{Aq^{33}\left(\beta_{2i} + q^{31}d_1 + q^{32}d_2 + q^{33}d_3\right)}{1 - Aq^{33}}\right]$$

$$\simeq E\left[\frac{S_{\bar{u}_2}^2}{S_{z_2}^2}\left\{\begin{array}{l} \beta_{2i} - \dfrac{\bar{z}_2 + \bar{u}_2}{S_{z_2}^2}\left(1 - \dfrac{S_{\bar{u}_2}^2}{S_{z_2}^2}\right)\left(\bar{\varepsilon}_{i\cdot} - \beta_{2i}\bar{u}_2\right) \\ +\dfrac{1}{S_{z_2}^2}\left(1 - \dfrac{S_{\bar{u}_2}^2}{S_{z_2}^2}\right)\left(\sum_{j=1}^{n}\dfrac{z_{j2}\bar{\varepsilon}_{ji\cdot}}{n} + \sum_{j=1}^{n}\dfrac{\bar{u}_{j2\cdot}\bar{\varepsilon}_{ji\cdot}}{n} - \beta_{2i}\sum_{j=1}^{n}\dfrac{z_{j2}\bar{u}_{j2\cdot}}{n} - \beta_{2i}S_{\bar{u}_2}^2\right) \end{array}\right\}\right]$$

$$E\left(\frac{Aq^{33}\hat{\beta}_{2i\_ols/hme}}{1-Aq^{33}}\right)$$

$$\simeq \frac{\beta_{2i}}{S_{z_2}^2}\left[\begin{array}{l} E\left(S_{\bar{u}_2}^2\right)+\frac{1}{S_{z_2}^2}E\left(\left(\bar{z}_2+\bar{u}_2\right)\bar{u}_2S_{\bar{u}_2}^2\right)-\frac{1}{\left(S_{z_2}^2\right)^2}E\left(\left(\bar{z}_2+\bar{u}_2\right)\bar{u}_2\left(S_{\bar{u}_2}^2\right)^2\right) \\ -\frac{1}{nS_{z_2}^2}E\left(S_{\bar{u}_2}^2\sum_{j=1}^{n}z_{j2}\bar{u}_{j2\cdot}\right)+\frac{1}{n\left(S_{z_2}^2\right)^2}E\left(\left(S_{\bar{u}_2}^2\right)^2\sum_{j=1}^{n}z_{j2}\bar{u}_{j2\cdot}\right) \\ -\frac{1}{\left(S_{z_2}^2\right)^2}E\left(\left(S_{\bar{u}_2}^2\right)^2\right)+\frac{1}{\left(S_{z_2}^2\right)^2}E\left(\left(S_{\bar{u}_2}^2\right)^3\right) \end{array}\right]. \tag{3.58}$$

As $n \to \infty$, equation (3.58) approaches

$$E\left(\frac{Aq^{33}\hat{\beta}_{2i\_ols/hme}}{1-Aq^{33}}\right) \to \frac{\beta_{2i}\sigma_{\bar{u}_2}^2}{S_{z_2}^2}. \tag{3.59}$$

Thus, from (3.57) and (3.59), it can be concluded that the bias of $\hat{\beta}_{2i\_cs}$ approaches zero as $n \to \infty$.

□

In summary, the biases of the CS and OLS estimators of the slope parameter of the precisely observed variable, $\hat{\beta}_{1i}$, are both zero. The biases of $\hat{\beta}_{0i\_ols/hme}$ and $\hat{\beta}_{2i\_ols/hme}$ asymptotically approach to $\frac{\beta_{2i}\bar{z}_2\sigma_{\bar{u}_2}^2}{S_{z_2}^2}$, and $-\frac{\beta_{2i}\sigma_{\bar{u}_2}^2}{S_{z_2}^2}$ respectively. In the case that $\beta_{2i} > 0$, $\hat{\beta}_{0i\_ols/hme}$ may be either an overestimated or underestimated parameter depending on the signs of β_{0i} and $\bar{z}_2$ whereas $\hat{\beta}_{2i\_ols/hme}$ is definitely an underestimated parameter. On the other hand, $\hat{\beta}_{0i\_cs}$ and $\hat{\beta}_{2i\_cs}$ are asymptotically unbiased estimators.

The algorithm of the CS approach in a linear multivariate regression model with HME is shown in Figure 1.

Define model and conditions

↓

Stage 1: Construct the corrected log-likelihood function i.e. finding $l_c(\mathbf{B},\mathbf{X},\mathbf{Y})$ which satisfies

$$E\left(l_c(\mathbf{B},\mathbf{X},\mathbf{Y})\right) = l(\mathbf{B},\mathbf{Z},\mathbf{Y})$$

↓

Stage 2: Evaluate the CS functions of $\mathbf{B}$;

$$U_{c\, j\boldsymbol{\beta}_q}(\mathbf{B},\mathbf{X},\mathbf{Y}) = \frac{\partial l_c(\mathbf{B},\mathbf{X},\mathbf{Y})}{\partial \boldsymbol{\beta}_q}$$

↓

Stage 3: Estimate the variances of HME ; S^2_{uqh} ;

$$S^2_{uqh} = \frac{\sum_{j_h=1}^{n_h} \sum_{k=1}^{r_h} \left(x_{qj_hk} - \bar{x}_{qj_h\cdot}\right)^2}{n_h r_h}$$

↓

Stage 4: Evaluate the CS estimators of $\mathbf{B}$;

$$\sum_{j=1}^{n} U_c\left(\mathbf{B}, X_j, Y_j\right) = 0 ,$$

Figure 3.1 Algorithm of the CS Approach for the Study Model

CHAPTER 4

SIMULATION STUDY

In this chapter, an evaluation of the performance of the proposed CS estimator compared to the OLS estimator is described. The reduced form of the model in (3.1)-(3.2) is considered when there are two independent variables of which one is measured with HME based on grouped heteroscedasticity whereas the other is not, and the random errors have a constant variance-covariance matrix defined by $s_1 = 1$, $s = 2$, and $\Sigma_j = \Sigma, \forall j = 1, 2, ..., n$. The CS estimator used follows equation (3.34). In section 4.1, details of the simulation settings are provided. Next, section 4.2 contains the simulation procedure, and the results of simulation study are shown in Section 4.3.

4.1 Simulation Settings

The objective of the simulation study is to empirically analyze the parameter estimations by the OLS and CS methods when varying the sample size n and the number of repeated measurements at the j^{th} observation, r_j. The proposed CS estimator is compared to the OLS estimator by considering bias and mean square error (MSE). Data sets are generated from the model defined in (3.1) and (3.2) with two dependent variables $(p = 2)$ and two independent variables $(s = 2)$. One of the independent variables, Z_1, is precisely observable and is generated with the uniform distribution $U[-1,1]$ whereas the other, Z_2, cannot be precisely observed and is generated with the standard normal distribution $N(0,1)$. The parameters in the model are set as follows: $\beta_{0i} = 0$, $\beta_{1i} = \beta_{2i} = 1$, $i = 1, 2$, and the variance-covariance matrix was set as $\mathbf{\Sigma}_j = \begin{pmatrix} 0.8 & 0.5 \\ 0.5 & 1.0 \end{pmatrix}$, $j = 1, 2, ..., n$. The surrogate variable X_2 instead of Z_2 is observed with the same r repeated measurements at each observation. The observations for each sample size are grouped into five sub-samples such that the variance of the random measurement error is homogeneous within a group but heterogeneous between groups.

The HME variance in the simulation is set in two forms: the step-up function form $(F1)$ and the step-down function form $(F2)$, as specified in Table 4.2, and are referred to as HME forms from now on. Each HME form is grouped into five sub-samples of equal size $(h=1,2,3,4,5)$. The random measurement error u_2 in the h^{th} group is distributed as $N\left(0,\sigma_{u2h}^2\right)$. In the simulation, three sample sizes, n, are specified: $50, 100$, and 500, and the number of repeated measurements, r: 5, 10, 20, and 40. One hundred replications are simulated for a particular combination of n and r. The data layout of the simulation study for one case is shown in Table 4.1.

Table 4.1 Data Layout of the Study When $s_1=1$, $s=2$, and $p=2$

j	k	y_{j1k}	y_{j2k}	z_{j1}	z_{j2}	$\Rightarrow x_{j2k}$	
1	1	y_{111}	y_{121}	z_{11}	z_{12}	x_{121}	
	2	y_{112}	y_{122}			x_{122}	$\bar{x}_{12\cdot}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	
	r	y_{11r}	y_{12r}			x_{12r}	
2	1	y_{211}	y_{221}	z_{21}	z_{22}	x_{221}	
	2	y_{212}	y_{222}			x_{222}	$\bar{x}_{22\cdot}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	
	r	y_{21r}	y_{22r}			x_{22r}	
	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
n	1	y_{n11}	y_{n21}	z_{n1}	z_{n2}	x_{n21}	
	2	y_{n12}	y_{n22}			x_{n22}	$\bar{x}_{n2\cdot}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	
	r	y_{n1r}	y_{n2r}			x_{n2r}	

4.2 Simulation Procedure

Step 1. Generate normal independent random variables with zero mean and unit variance, $a \sim N(0,1)$, consisting of 4,000,000 observations, using the CALL STREAMINIT with RAND('NORMAL') routine in SAS version 9.3. The seed number to generate a was 5837259.

Step 2. Separate the datasets in Step 1 into four sets:

Set 1: Dataset for preparing the random errors of the 1st component consisting of 1,000,000 observations, called a_e1.

Set 2: Dataset for preparing the random errors of the 2nd component consisting of 1,000,000 observations, called a_e2.

Set 3: Dataset of the independent variable Z_2 consisting of 1,000,000 observations, called z_2.

Set 4: Dataset for preparing the HME random error consisting of 1,000,000 observations, called a_u.

Step 3. Check the independence of all four sets in Step 2 using the SAS statement PROC CORR. If they are independent, then go to Step 4. If not, go to Step 1.

Step 4. Transform the normal independent random variables in Set 1 and Set 2 to be multivariate random errors with correlation coefficient $\rho = 0.5$ and the variances of ε_{j1} and ε_{j2} to be 0.8 and 1.0, respectively. The variance-covariance matrix was set as

$$\boldsymbol{\Sigma}_j = \begin{pmatrix} 0.8 & 0.5 \\ 0.5 & 1.0 \end{pmatrix}, \forall j,\ j = 1, 2, \ldots, n.$$

Transform the normal independent random variable in Set 4 to be heterogeneous random measurement error u_2 distributed as $N\left(0, \sigma_{u2h}^2\right)$. The HME variance was set in two forms: the step-up function form $(F1)$ and the step-down function form $(F2)$, as specified in Table 4.2.

Table 4.2 HME Variances for HME Forms $F1$ and $F2$

HME Form	HME Variance				
	σ^2_{u21}	σ^2_{u22}	σ^2_{u23}	σ^2_{u24}	σ^2_{u25}
$F1$	0.1	0.2	0.4	0.6	0.8
$F2$	0.8	0.6	0.4	0.2	0.1

Step 5. Generate independent random variables Z_1 distributed as $U[-1,1]$, $z_1 \sim U[-1,1]$, consisting of 1,000,000 observations, using the SAS routine CALL STREAMINIT with RAND('UNIFORM').

Step 6. Construct the multivariate model where $\beta_{oi} = 0$ and $\beta_{1i} = \beta_{2i} = 1$, $i = 1, 2$, and uses the random variable z_2 in Step 2, random errors ε_{j1} and ε_{j2}, and the random heteroscedastic measurement error u_2 in Step 4, then the model in (3.1) to (3.2) can be expressed as

$$y_i = \beta_{0i} + \beta_{1i} z_1 + \beta_{2i} z_2 + \varepsilon_i,$$

$$= z_1 + z_2 + \varepsilon_i, \text{ for } i = 1, 2, \text{ and}$$

$$x_2 = z_2 + u_2.$$

Next, check that the assumptions and properties of the model, i.e. z_1, z_2, ε_i for $i = 1, 2$, and u_2 are independent using the SAS statement PROC CORR. If the assumptions are appropriate, use the dataset in this step for the population and go to Step 7. If not, go to Step 1 and change the seed number before regenerating the normal independent random variables.

Step 7. Test the sample for the model without repeated measurements from the population in Step 6 using SAS statement PROC SURVEY. For each HME form, there were combinations of datasets for n = 50, 100, and 150, and r = 5, 10, 20, and 40, with 100 simulated replications. Next, for each case, test the rest of the repeated measurement samples and construct the repeated measurements model.

Step 8. Check the assumptions and properties of the model, i.e. z_1, z_2, ε_i for $i = 1, 2$, and u_2 are independent for each case using the PROC CORR statement. If the assumptions are appropriate, the dataset in this step is used as the population and go to Step 9. If not, go to Step 7 and the change seed number in PROC SURVEY statement.

Step 9. For each case, evaluate the study model corresponding to model (3.1) to (3.2) where $p = 2,\ s = 2,\ s_1 = 1$:

$$\bar{y}_i = \beta_{0i} + \beta_{1i} z_1 + \beta_{2i} z_2 + \bar{\varepsilon}_i, \quad i = 1, 2,$$

$$\bar{x}_2 = z_2 + \bar{u}_2.$$

Step 10. Estimate the parameters of the model in Step 9 using the OLS method; regress $\bar{y}_i$ on $z_1, \bar{x}_2$ using SAS statement PROC GLM.

Step 11. Estimate the parameters of the model in Step 9 by the CS method where the CS estimator satisfies (3.34).

Step 12. Evaluate the simulated sample mean; the SE of the sample mean; the p-value of the t-test for the sample means: $H_0 : \beta_{qi} = 1,\ q = 1, 2,\ i = 1, 2$, at significance level 0.05; the bias, and the MSE of the CS estimator compared to the OLS estimator:

The simulated sample mean of $\hat{\beta}_{qi}\left(\bar{\hat{\beta}}_{qi}\right) = \dfrac{\sum_{s=1}^{100} \hat{\beta}_{qi(s)}}{100}$, $q = 1, 2,\ i = 1, 2$.

The Hypothesis for testing the parameters is given by

$$H_0 : \beta_{qi} = 1 \quad \text{vs.} \quad H_1 : \beta_{qi} \neq 1 \quad \text{for } q = 1, 2,\ i = 1, 2,$$

The t-statistic is $t = \dfrac{\bar{\hat{\beta}}_{qi} - 1}{SE\left(\bar{\hat{\beta}}_{qi}\right)}$, $q = 1, 2,\ i = 1, 2$. The p-value = $P\left(t \geq t_{0.025,1}\right)$.

$$\text{Bias of } \hat{\beta}_{qi} = \frac{\sum_{s=1}^{100} \left(\hat{\beta}_{qi(s)} - \beta_{qi}\right)}{100}, q = 1, 2,\ i = 1, 2.$$

$$\text{MSE of } \hat{\beta}_{qi} = \frac{\sum_{s=1}^{100} \left(\hat{\beta}_{qi(s)} - \beta_{qi}\right)^2}{100}, q = 1, 2,\ i = 1, 2.$$

The details of population data are shown in Appendix C.

4.3 Results of the Simulation Study

4.3.1 HME Variance Estimation

The simulation results in Tables 4.3 and 4.4 provide the simulated sample means of the HME sample variance, S_{u2h}^2, and the standard error (SE) under the model based on grouped heteroscedasticity for HME forms $F1$ and $F2$, respectively. The results reveal that, for all situations, the number of repeated measurements does not seem to affect the magnitude of the bias of S_{u2h}^2, whereas S_{u2h}^2 is close to the population parameter σ_{u2h}^2 when the sample size increases, and the SE decreases when either the number of repeated measurements or the sample size increases. On the other hand, for the same number of repeated measurements, the SE increases when the variance of grouped heteroscedasticity is large. The SE is smallest when $n=500, r=40, h=1$ $(SE=0.0024)$ and $n=500, r=40, h=5$ $(SE=0.0024)$ for HME forms $F1$ and $F2$, respectively.

4.3.2 CS Estimation

So as to gauge the performance of the estimation parameters, Tables 4.5 to 4.8 provide the simulated sample means, and SE of the CS estimator is compared to the OLS estimator. Tables 4.3 and 4.4 show data on the performance of the estimator b_{1i}, $i=1,2$ (the observations z_1 were precisely observed) for HME forms $F1$ and $F2$, respectively. The results show that all simulation conditions corresponding p-values > 0.05, and so the tests fails to reject the null hypothesis $H_0:\beta_{1i}=1,\ i=1,2$ for both the OLS and CS methods. The SE decreases when either the number of repeated measurements or sample size increases for both methods, whereas the SE for the CS method is slightly higher than that of the OLS method for all situations.

Tables 4.5 and 4.6 provide the results of the performance of estimator b_{2i}, $i=1,2$ (the observations x_2 are observed with HME). The result for the OLS method reveal that all situations has corresponding p-values > 0.05, which lead to rejection of the null hypothesis $H_0 : \beta_{2i} = 1,\ i = 1, 2$, and the bias is negative; thus, the sample means of $b_{2i},\ i = 1, 2$, are underestimated. Meanwhile, the t-test for the sample means of $b_{2i},\ i = 1, 2$ for the CS method is no different from the true value $\beta_{2i},\ i = 1, 2$, $(p > 0.05)$. The SE for the CS method is slightly higher than that for the OLS method when the number of repeated measurements is small $(r = 5, 10, 20)$ and is close to the same value when the number is large $(r = 40)$.

In Table 4.7, the bias and MSE of the CS estimator compared with the OLS estimator for estimator $b_{1i},\ i = 1, 2$ show that the magnitude of bias of the two estimators seem no different from each other and is close to zero when the repeated measurements and the sample size are large $(r = 40,\ n = 500)$ for both HME forms. In addition, Table 4.8 provides the bias and MSE of estimator $b_{2i},\ i = 1, 2$; the magnitude of bias and the MSE of the CS estimator are lower than those of the OLS estimator for all cases, and the magnitude of bias is close to zero when the number of repeated measurements and the sample size is large.

Table 4.3 The Simulated Sample Mean and SE of s_{u2h}^2 for HME Form $F1$

HME Form	n	r	Sample Mean and SE * of HME Variance									
			h=1		h=2		h=3		h=4		h=5	
$F1$	50	5	0.1061	(0.0260)	0.2124	(0.0475)	0.4263	(0.1016)	0.5895	(0.1296)	0.8100	(0.1869)
		10	0.1045	(0.0163)	0.2032	(0.0333)	0.4046	(0.0601)	0.5876	(0.0883)	0.7782	(0.1199)
		20	0.1008	(0.0113)	0.2073	(0.0246)	0.3990	(0.0449)	0.5960	(0.0655)	0.7931	(0.0834)
		40	0.1009	(0.0080)	0.2026	(0.0177)	0.4093	(0.0318)	0.5960	(0.0432)	0.7964	(0.0522)
	100	5	0.1021	(0.0175)	0.2070	(0.0346)	0.3875	(0.0664)	0.6170	(0.0990)	0.8086	(0.1268)
		10	0.1026	(0.0111)	0.2065	(0.0252)	0.3996	(0.0464)	0.5958	(0.0725)	0.7789	(0.0829)
		20	0.0999	(0.0073)	0.2004	(0.0172)	0.3986	(0.0328)	0.5932	(0.0458)	0.7840	(0.0569)
		40	0.1006	(0.0052)	0.2006	(0.0099)	0.4018	(0.0216)	0.6038	(0.0309)	0.7972	(0.0366)
	500	5	0.1015	(0.0072)	0.2019	(0.0168)	0.4004	(0.0292)	0.5961	(0.0424)	0.8000	(0.0597)
		10	0.1004	(0.0047)	0.2001	(0.0113)	0.3966	(0.0201)	0.5965	(0.0330)	0.7916	(0.0393)
		20	0.1009	(0.0038)	0.2010	(0.0078)	0.3998	(0.0144)	0.5957	(0.0231)	0.7973	(0.0263)
		40	0.1000	(0.0024)	0.2004	(0.0053)	0.3991	(0.0104)	0.5990	(0.0126)	0.7930	(0.0172)

* the SE is presented in parentheses

Table 4.4 The Simulated Sample Mean and SD of S^2_{u2h} for HME Form $F2$

HME Form	n	r	Sample Mean and SE* of HME Variance									
			h=1		h=2		h=3		h=4		h=5	
$F2$	50	5	0.7913	(0.1791)	0.6003	(0.1303)	0.4101	(0.1090)	0.2037	(0.0453)	0.1068	(0.0224)
		10	0.7995	(0.1179)	0.6031	(0.1035)	0.4042	(0.0669)	0.2085	(0.0377)	0.1042	(0.0165)
		20	0.7944	(0.0832)	0.6016	(0.0626)	0.4020	(0.0394)	0.2056	(0.0244)	0.1019	(0.0102)
		40	0.7979	(0.0669)	0.6008	(0.0446)	0.4013	(0.0282)	0.1999	(0.0160)	0.1013	(0.0080)
	100	5	0.7928	(0.1268)	0.6019	(0.0949)	0.4072	(0.0689)	0.2057	(0.0367)	0.1071	(0.0187)
		10	0.7896	(0.0817)	0.6098	(0.0717)	0.3970	(0.0438)	0.2013	(0.0230)	0.1023	(0.0112)
		20	0.7955	(0.0559)	0.6073	(0.0456)	0.4054	(0.0340)	0.2031	(0.0168)	0.1028	(0.0081)
		40	0.7872	(0.0361)	0.5997	(0.0307)	0.3970	(0.0241)	0.1994	(0.0098)	0.1022	(0.0059)
	500	5	0.7971	(0.0531)	0.5949	(0.0414)	0.4010	(0.0323)	0.2000	(0.0182)	0.1020	(0.0077)
		10	0.7921	(0.0360)	0.6029	(0.0287)	0.4025	(0.0191)	0.2005	(0.0116)	0.1022	(0.0051)
		20	0.7961	(0.0278)	0.6054	(0.0209)	0.4003	(0.0141)	0.2001	(0.0070)	0.1012	(0.0031)
		40	0.7964	(0.0170)	0.6025	(0.0136)	0.4019	(0.0093)	0.2005	(0.0047)	0.1005	(0.0024)

* the SE is presented in parentheses

Table 4.5 Statistics and MSEs for $\hat{\beta}_{11}$, $\hat{\beta}_{12}$ under HME Form $F1$

n	r	Parameter	Sample Mean		SE		p-value	
			OLS	**CS**	**OLS**	**CS**	**OLS**	**CS**
50	5	β_{11}	0.9984	0.9986	0.01022	0.01025	0.8764	0.8905
		β_{12}	0.9945	0.9946	0.01153	0.01166	0.6323	0.6462
	10	β_{11}	1.0067	1.0066	0.00699	0.00701	0.3401	0.3487
		β_{12}	0.9997	0.9996	0.00769	0.00774	0.9648	0.9538
	20	β_{11}	1.0023	1.0023	0.00515	0.00517	0.6611	0.6635
		β_{12}	1.0007	1.0007	0.00537	0.00539	0.1690	0.1710
	40	β_{11}	0.9979	0.9978	0.00322	0.00322	0.5081	0.5048
		β_{12}	0.9970	0.9970	0.00394	0.00394	0.4507	0.4487
100	5	β_{11}	0.9959	0.9962	0.00668	0.00666	0.5452	0.5726
		β_{12}	0.9988	0.9990	0.00768	0.0078	0.8731	0.9017
	10	β_{11}	1.0014	1.0013	0.00523	0.00524	0.7849	0.8037
		β_{12}	1.0024	1.0022	0.00589	0.00591	0.6900	0.7066
	20	β_{11}	0.9984	0.9983	0.00373	0.00374	0.6771	0.6619
		β_{12}	0.9972	0.9972	0.00417	0.00418	0.5111	0.5051
	40	β_{11}	0.9963	0.9962	0.00265	0.00265	0.1612	0.1591
		β_{12}	1.0006	1.0006	0.00290	0.00291	0.8348	0.8402
500	5	β_{11}	1.0005	1.0006	0.00298	0.00300	0.8766	0.8416
		β_{12}	1.0013	1.0015	0.00337	0.00342	0.6946	0.6695
	10	β_{11}	1.0020	1.0020	0.00240	0.00241	0.4094	0.4149
		β_{12}	1.0012	1.0012	0.00275	0.00276	0.6706	0.6754
	20	β_{11}	1.0006	1.0006	0.00161	0.00161	0.7290	0.7292
		β_{12}	0.9986	0.9986	0.00175	0.00175	0.4185	0.4192
	40	β_{11}	1.0005	1.0006	0.00123	0.00123	0.6551	0.6538
		β_{12}	1.0000	1.0000	0.00123	0.00123	0.9776	0.9757

Table 4.6 Statistics and MSEs for $\hat{\beta}_{11}$, $\hat{\beta}_{12}$ under HME Form $F2$

n	r	Parameter	Sample Mean		SE		p-value	
			OLS	**CS**	**OLS**	**CS**	**OLS**	**CS**
50	5	β_{11}	0.9926	0.9924	0.00998	0.01006	0.4615	0.4490
		β_{12}	0.9947	0.9945	0.01148	0.01158	0.6465	0.6329
	10	β_{11}	1.0018	1.0010	0.00761	0.00766	0.1572	0.1537
		β_{12}	1.0004	1.0006	0.00757	0.00763	0.9593	0.9426
	20	β_{11}	1.0035	1.0035	0.00622	0.00623	0.5804	0.5735
		β_{12}	1.0003	1.0004	0.00600	0.00601	0.9583	0.9493
	40	β_{11}	0.9975	0.9975	0.00371	0.00371	0.5002	0.5023
		β_{12}	0.9981	0.9981	0.00438	0.00438	0.6640	0.6659
100	5	β_{11}	1.0048	1.0048	0.00715	0.00719	0.5041	0.5019
		β_{12}	1.0005	1.0006	0.00735	0.00743	0.9413	0.9367
	10	β_{11}	1.0001	0.9998	0.00596	0.00599	0.9904	0.9774
		β_{12}	1.0042	1.0040	0.00569	0.00570	0.4620	0.4888
	20	β_{11}	1.0014	1.0013	0.00403	0.00403	0.7381	0.7494
		β_{12}	1.0029	1.0028	0.00407	0.00408	0.4855	0.4949
	40	β_{11}	0.9985	0.9984	0.00281	0.00281	0.5831	0.5804
		β_{12}	1.0006	1.0006	0.00291	0.00291	0.8436	0.8471
500	5	β_{11}	0.9985	0.9985	0.00354	0.00358	0.6674	0.6712
		β_{12}	0.9977	0.9977	0.00373	0.00378	0.5454	0.5510
	10	β_{11}	0.9980	0.9979	0.00218	0.00219	0.3676	0.3504
		β_{12}	1.0012	1.0011	0.00233	0.00234	0.5993	0.6252
	20	β_{11}	1.0004	1.0004	0.00168	0.00169	0.7960	0.7990
		β_{12}	0.9987	0.9987	0.00178	0.00178	0.4547	0.4528
	40	β_{11}	0.9991	0.9991	0.00110	0.00110	0.4328	0.4332
		β_{12}	0.9988	0.9988	0.00125	0.00125	0.3530	0.3537

Table 4.7 Statistics and MSEs for $\hat{\beta}_{21}$, $\hat{\beta}_{22}$ under HME Form $F1$

n	r	Parameter	Sample Mean		SE		p-value	
			OLS	**CS**	**OLS**	**CS**	**OLS**	**CS**
50	5	β_{21}	0.9782	1.0016	0.00261	0.00265	<.0001	0.5470
		β_{22}	0.9800	1.0034	0.00301	0.00307	<.0001	0.2652
	10	β_{21}	0.9908	1.0019	0.00196	0.00195	<.0001	0.3200
		β_{22}	0.9904	1.0015	0.00207	0.00207	<.0001	0.4601
	20	β_{21}	0.9962	1.0018	0.00143	0.00143	.0095	0.2065
		β_{22}	0.9956	1.0012	0.00166	0.00162	.0090	0.4776
	40	β_{21}	0.9949	0.9977	0.00112	0.00112	<.0001	0.0565
		β_{22}	0.9955	0.9983	0.00125	0.00124	.0005	0.1750
100	5	β_{21}	0.9781	0.9993	0.00223	0.00227	<.0001	0.7563
		β_{22}	0.9762	0.9975	0.00216	0.00218	<.0001	0.2448
	10	β_{21}	0.9882	0.9988	0.00143	0.00148	<.0001	0.4112
		β_{22}	0.9882	0.9988	0.00145	0.00147	<.0001	0.4000
	20	β_{21}	0.9964	1.0017	0.00107	0.00109	.0009	0.1196
		β_{22}	0.9960	1.0013	0.00125	0.00127	.0017	0.3060
	40	β_{21}	0.9979	1.0006	0.00069	0.00069	.0031	0.4149
		β_{22}	0.9977	1.0004	0.00078	0.00078	.0050	0.6000
500	5	β_{21}	0.9797	1.0003	0.00099	0.00102	<.0001	0.7962
		β_{22}	0.9797	1.0003	0.00115	0.00116	<.0001	0.6695
	10	β_{21}	0.9892	0.9996	0.00063	0.00064	<.0001	0.5398
		β_{22}	0.9893	0.9997	0.00076	0.00077	<.0001	0.6737
	20	β_{21}	0.9950	1.0003	0.00047	0.00048	<.0001	0.4867
		β_{22}	0.9948	1.0001	0.00053	0.00054	<.0001	0.8814
	40	β_{21}	0.9977	1.0003	0.00035	0.00036	<.0001	0.3959
		β_{22}	0.9972	0.9998	0.00040	0.00040	<.0001	0.6853

Table 4.8 Statistics and MSEs for $\hat{\beta}_{21}$, $\hat{\beta}_{22}$ under HME Form $F2$

n	r	Parameter	Sample Mean		SE		p-value	
			OLS	**CS**	**OLS**	**CS**	**OLS**	**CS**
50	5	β_{21}	0.9815	1.0037	0.00303	0.00315	<.0001	0.2467
		β_{22}	0.9828	1.0050	0.00318	0.00330	<.0001	0.1337
	10	β_{21}	0.9910	1.0022	0.00199	0.00203	<.0001	0.2817
		β_{22}	0.9914	1.0026	0.00217	0.00224	.0001	0.2416
	20	β_{21}	0.9965	1.0023	0.00149	0.00150	.0224	0.1307
		β_{22}	0.9961	1.0018	0.00159	0.00159	.0156	0.2483
	40	β_{21}	0.9959	0.9987	0.00106	0.00107	.0002	0.2281
		β_{22}	0.9950	0.9978	0.00124	0.00125	<.0001	0.0786
100	5	β_{21}	0.9784	0.9990	0.00203	0.00217	<.0001	0.9573
		β_{22}	0.9809	1.0024	0.00188	0.00191	<.0001	0.2151
	10	β_{21}	0.9892	0.9998	0.00161	0.00165	<.0001	0.8952
		β_{22}	0.9901	1.0006	0.00164	0.00170	<.0001	0.7070
	20	β_{21}	0.9967	1.0020	0.00115	0.00114	.0046	0.0754
		β_{22}	0.9962	1.0016	0.00109	0.00109	.0008	0.1485
	40	β_{21}	0.9979	1.0005	0.00077	0.00077	.0069	0.4927
		β_{22}	0.9979	1.0005	0.00074	0.00074	.0047	0.4666
500	5	β_{21}	0.9792	1.0001	0.00086	0.00083	<.0001	0.8813
		β_{22}	0.9792	1.0002	0.00092	0.00093	<.0001	0.8546
	10	β_{21}	0.9893	0.9998	0.00070	0.00070	<.0001	0.7986
		β_{22}	0.9898	1.0004	0.00072	0.00073	<.0001	0.5999
	20	β_{21}	0.9954	1.0007	0.00046	0.00046	<.0001	0.1590
		β_{22}	0.9952	1.0004	0.00054	0.00053	<.0001	0.4561
	40	β_{21}	0.9977	1.0003	0.00040	0.00040	<.0001	0.4563
		β_{22}	0.9978	1.0005	0.00039	0.00039	<.0001	0.2228

Table 4.9 The Bias and MSE for $\hat{\beta}_{11}$, $\hat{\beta}_{12}$ with HME Forms $F1$ and $F2$

n	r	Parameter	F1				F2			
			Bias		MSE		Bias		MSE	
			OLS	**CS**	**OLS**	**CS**	**OLS**	**CS**	**OLS**	**CS**
50	5	β_{11}	-0.00160	-0.00140	0.01034	0.01041	-0.00740	-0.00760	0.00991	0.01008
		β_{12}	-0.00550	-0.00540	0.01320	0.01349	-0.00530	-0.00550	0.01308	0.01331
	10	β_{11}	0.00670	0.00660	0.00489	0.00491	0.00180	0.00100	0.00585	0.00593
		β_{12}	-0.00030	-0.00040	0.00585	0.00593	0.00040	0.00060	0.00568	0.00576
	20	β_{11}	0.00230	0.00230	0.00263	0.00265	0.00350	0.00350	0.00384	0.00386
		β_{12}	0.00070	0.00070	0.00292	0.00294	0.00030	0.00040	0.00356	0.00358
	40	β_{11}	-0.00210	-0.00220	0.00103	0.00103	-0.00250	-0.00250	0.00137	0.00137
		β_{12}	-0.00300	-0.00300	0.00155	0.00155	-0.00190	-0.00190	0.00190	0.00190
100	5	β_{11}	-0.00410	-0.00380	0.00443	0.00441	0.00480	0.00480	0.00508	0.00514
		β_{12}	-0.00120	-0.00100	0.00584	0.00587	0.00050	0.00060	0.00535	0.00547
	10	β_{11}	0.00140	0.00130	0.00271	0.00272	0.00010	-0.00020	0.00352	0.00355
		β_{12}	0.00240	0.00220	0.00344	0.00346	0.00420	0.00400	0.00322	0.00323
	20	β_{11}	-0.00160	-0.00170	0.00138	0.00139	0.00140	0.00130	0.00160	0.00160
		β_{12}	-0.00280	-0.00280	0.00173	0.00174	0.00290	0.00280	0.00170	0.00170
	40	β_{11}	-0.00370	-0.00380	0.00071	0.00071	-0.00150	-0.00160	0.00078	0.00078
		β_{12}	0.00061	0.00060	0.00083	0.00084	0.00060	0.00060	0.00084	0.00084

Table 4.9 (Continued)

n	r	Para-meter	F1 Bias OLS	F1 Bias CS	F1 MSE OLS	F1 MSE CS	F2 Bias OLS	F2 Bias CS	F2 MSE OLS	F2 MSE CS
500	5	β_{11}	0.00050	0.00060	0.00088	0.00089	-0.00150	-0.00150	0.00125	0.00127
		β_{12}	0.00130	0.00150	0.00011	0.00116	-0.00230	-0.00230	0.00138	0.00142
	10	β_{11}	0.00200	0.00200	0.00057	0.00058	-0.00200	-0.00210	0.00047	0.00048
		β_{12}	0.00120	0.00120	0.00075	0.00075	0.00120	0.00110	0.00054	0.00055
	20	β_{11}	0.00060	0.00060	0.00026	0.00026	0.00040	0.00040	0.00028	0.00028
		β_{12}	-0.00140	-0.00140	0.00030	0.00030	-0.00130	-0.00130	0.00032	0.00032
	40	β_{11}	0.00050	0.00060	0.00015	0.00015	-0.00090	-0.00090	0.00020	0.00019
		β_{12}	0.00000	0.00000	0.00015	0.00015	-0.00120	-0.00120	0.00016	0.00016

Table 4.10 The Bias and MSE for $\hat{\beta}_{21}$, $\hat{\beta}_{22}$ with HME Forms $F1$ and $F2$

n	r	Para-meter	F1				F2			
			Bias		MSE		Bias		MSE	
			OLS	**CS**	**OLS**	**CS**	**OLS**	**CS**	**OLS**	**CS**
50	5	β_{21}	-0.02180	0.00160	0.001148	0.000770	-0.01850	0.00370	0.001250	0.000997
		β_{22}	-0.02000	0.00340	0.001298	0.000945	-0.01720	0.00500	0.001290	0.001104
	10	β_{21}	-0.00920	0.00190	0.000464	0.000380	-0.00900	0.00220	0.000475	0.000413
		β_{22}	-0.00960	0.00150	0.000514	0.000426	-0.00860	0.00260	0.000539	0.000503
	20	β_{21}	-0.00380	0.00180	0.000216	0.000206	-0.00350	0.00230	0.000233	0.000228
		β_{22}	-0.00440	0.00120	0.000291	0.000275	-0.00390	0.00180	0.000265	0.000255
	40	β_{21}	-0.00510	-0.00230	0.000149	0.000130	-0.00410	-0.00130	0.000130	0.000120
		β_{22}	-0.00450	-0.00170	0.000174	0.000155	-0.00500	-0.00220	0.000180	0.000160
100	5	β_{21}	-0.02190	-0.00070	0.000974	0.000509	-0.02160	0.00100	0.000877	0.000424
		β_{22}	-0.02380	-0.00250	0.001027	0.000478	-0.02910	0.00240	0.000718	0.000368
	10	β_{21}	-0.01180	-0.00120	0.000341	0.000217	-0.01080	-0.00020	0.000374	0.000270
		β_{22}	-0.01180	-0.00120	0.000347	0.000215	-0.00990	0.00060	0.000364	0.000287
	20	β_{21}	-0.00360	0.00170	0.000126	0.000120	-0.00330	0.00200	0.000141	0.000132
		β_{22}	-0.00400	0.00130	0.000171	0.000162	-0.00380	0.00160	0.000131	0.000121
	40	β_{21}	-0.00210	0.00060	0.000052	0.000047	-0.00210	0.00050	0.000064	0.000060
		β_{22}	-0.00230	0.00040	0.000066	0.000061	-0.00210	0.00050	0.000058	0.000054

Table 4.10 (Continued)

n	r	Parameter	F1 Bias OLS	F1 Bias CS	F1 MSE OLS	F1 MSE CS	F2 Bias OLS	F2 Bias CS	F2 MSE OLS	F2 MSE CS
500	5	β_{21}	-0.02030	0.00030	0.000510	0.000103	-0.02080	0.00010	0.000503	0.000072
		β_{22}	-0.02030	0.00030	0.000543	0.000134	-0.02080	0.00020	0.000518	0.000087
	10	β_{21}	-0.01080	-0.00060	0.000156	0.000040	-0.02070	-0.00080	0.000163	0.000049
		β_{22}	-0.01070	-0.00070	0.000171	0.000058	-0.02020	0.00040	0.000154	0.000052
	20	β_{21}	-0.00500	0.00030	0.000046	0.000023	-0.00460	0.00070	0.000042	0.000022
		β_{22}	-0.00520	0.00010	0.000055	0.000029	-0.00480	0.00040	0.000052	0.000028
	40	β_{21}	-0.00230	0.00030	0.000018	0.000013	-0.00230	0.00030	0.000021	0.000016
		β_{22}	-0.00280	-0.00020	0.000024	0.000016	-0.00220	0.00050	0.000020	0.000015

4.4 The Proposed CS Estimator with Real-life Data

Data from an oximetry study is used to demonstrate the performance of parameter estimation by the proposed CS estimator and comes as a part of the MethComp package for the R program (Carstensen, Gurrin, and Ekstrom, 2015), the data is shown in Appendix D. The study examines the percentage of oxygen in the blood of sick children at the Royal Children's hospital in Melbourne. The CO oximetry method (CO) is compared to the pulse oximetry method (PULSE). 61 children are measured where 53 had 3 replicates in each method, 4 children had 2 replicates and another one had 1. Based on a functional comparative calibration model, the CS estimators of the intercept parameter ($\hat{\beta}_0$) and the slope parameter ($\hat{\beta}_1$) are $\hat{\beta}_0 = 0.0126$ and $\hat{\beta}_1 = 0.8622$, respectively (Giménez and Patat, 2014).

When regressing the PULSE variable onto the CO variable using the OLS method without adjustment for HME, the results show that $\hat{\beta}_{0\_ols/hme} = 0.0343$ and $\hat{\beta}_{1\_ols/hme} = 0.8092$, respectively. It can be seen that $\hat{\beta}_{1\_ols/hme}$ is underestimated when compares with $\hat{\beta}_1 = 0.8622$ (as a reference estimator) and the scatterplot of residual against CO as Figure 4.1 shows that heteroscedasticity has occurred. Thus, the proposed CS estimator is used to correct the HME for this data.

Assuming the HMEs are in the form of grouped heteroscedasticity, the data is separated into 2 groups. The CO data are put into descending order and the PULSE data order in tandem with the CO data. Following this, the data are separated into 2 groups: the first group has 30 observations and the second group has 31 observations. The proposed CS estimators are evaluated where $\hat{\beta}_{0\_cs} = 0.0031$ and $\hat{\beta}_{1\_cs} = 0.8680$. The results show that the bias of the proposed CS estimator is smaller than the bias of the OLS estimator. Moreover, in Figure 4.2, the CS method shows that the residuals are not heteroscedastic. Thus, it can be concluded that the CS estimator outperforms the OLS estimator and the parameter estimation by the proposed CS estimator shows conformance to the parameter estimation by Giménez and Patat (2014).

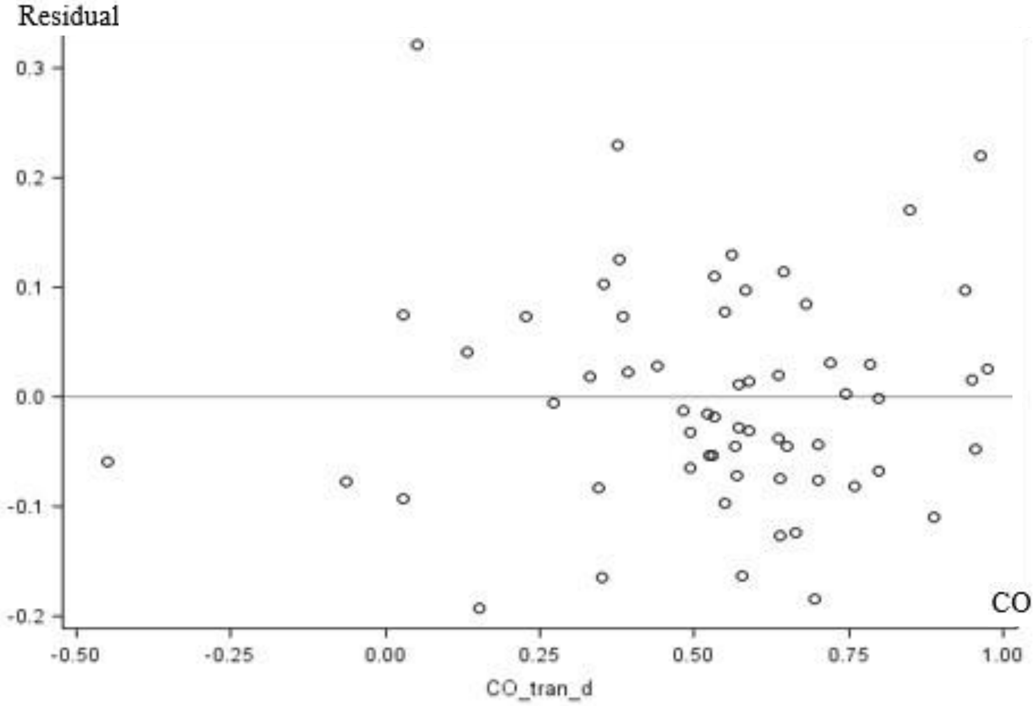

Figure 4.1 Scatter Plot of Residuals Against CO for the OLS Method

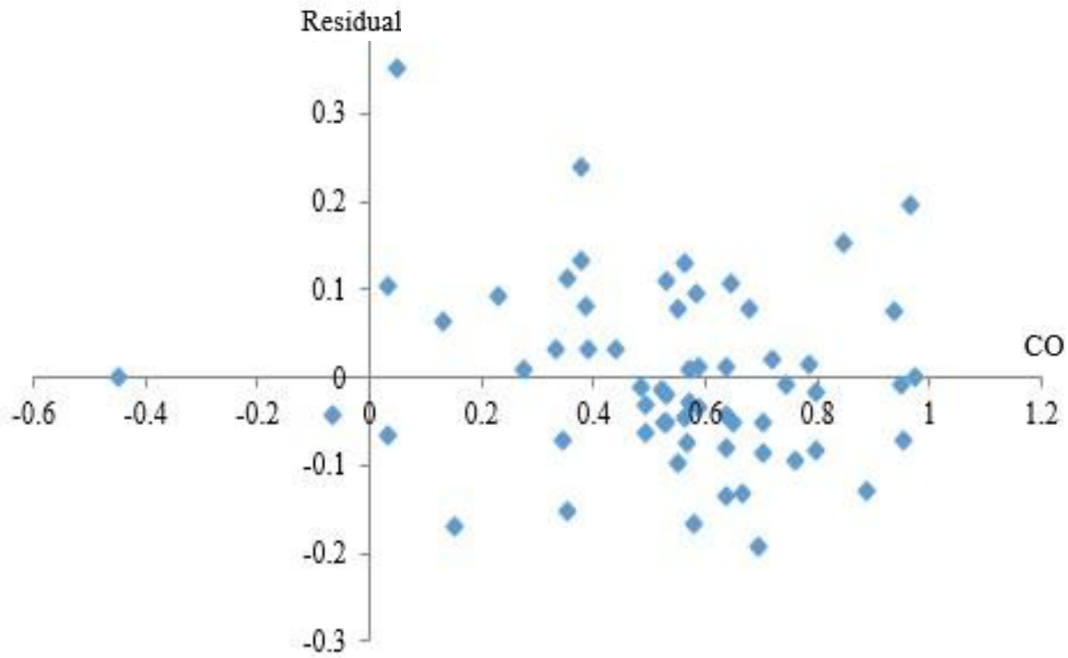

Figure 4.2 Scatter Plot of Residuals Against CO for the CS Method

CHAPTER 5

CONCLUSIONS

5.1 Conclusions

This study extends the estimation theory based on the CS to cover a linear multivariate multiple regression model consisting of s_1 precisely observed independent variables and $\left(s-s_1\right)$ independent variables with HMEs. The random error at the j^{th} observation is distributed independently across observations as $N\left(0,\boldsymbol{\Sigma}_j\right)$. The assumption in this model is that the HME variance is unknown and is estimated based on grouped heteroscedasticity; each group can be evaluated from pooled variances by a variable with HME observed in repeated measurements. HME violates the OLS assumption of dependency in the independent variables and random errors; the OLS estimator based on HMEs in X have attenuated bias and are also inconsistent. The CS approach in this study is identified as a method for correcting attenuated bias because it is one of the methods based on functional modeling to correct the HME and provides a fully consistent (asymptotically removes all bias) method for a linear model.

The proposed estimating parameter procedure is developed in four stages:

Stage 1: Construct a corrected log-likelihood function.

Stage 2: Evaluate the CS functions of $\mathbf{B}$.

Stage 3: Estimate the variances of HME based on grouped heteroscedasticity.

Stage 4: Evaluate the CS estimators of $\mathbf{B}$.

Based on the model in equations (3.1) to (3.2), the estimation of HME variance is evaluated from the pooled variance by a variable with HME observed in repeated measurements is given by

$$S_{uqh}^{2}=\frac{\sum_{j_h=1}^{n_h}\sum_{k=1}^{r_h}\left(x_{qj_hk}-\bar{x}_{qj_h\cdot}\right)^2}{n_hr_h}. \quad (5.1)$$

The proposed CS estimator is given by

$$vec\left(\hat{\mathbf{B}}_{cs}\right)=\left[\left\{\left(\mathbf{X}\otimes\mathbf{I}_p\right)'\mathbf{V}^{-1}\left(\mathbf{X}\otimes\mathbf{I}_p\right)\right\}+\mathbf{C}\right]^{-1}\left(\mathbf{X}\otimes\mathbf{I}_p\right)'\mathbf{V}^{-1}vec\left(\mathbf{Y}'\right), \quad (5.2)$$

where $\mathbf{\Psi}=\left(\left(\mathbf{X}\otimes\mathbf{I}_p\right)'\mathbf{V}^{-1}\left(\mathbf{X}\otimes\mathbf{I}_p\right)\right)^{-1}\mathbf{C}\left(\mathbf{I}_{p(s+1)}+\left(\left(\mathbf{X}\otimes\mathbf{I}_p\right)'\mathbf{V}^{-1}\left(\mathbf{X}\otimes\mathbf{I}_p\right)\right)^{-1}\mathbf{C}\right)^{-1}$,

$\mathbf{V}^{-1}$ is a block-diagonal matrix of size np, $diag\left(\mathbf{\Sigma}_1^{-1} \quad \mathbf{\Sigma}_2^{-1} \quad \cdots \quad \mathbf{\Sigma}_n^{-1}\right)$, $\mathbf{C}$ is a block-diagonal matrix of size $p(s+1)$ where the first (s_1+1) diagonal squar submatrices of size *p* are zero and the last $(s-s_1)$ diagonal square submatrices of size *p* are the estimates of $-\sum_{j=1}^{n}\mathbf{\Sigma}_j^{-1}\left(\sigma_{uj(s_1+1)}^2/r_j\right),\ldots,-\sum_{j=1}^{n}\mathbf{\Sigma}_j^{-1}\left(\sigma_{ujs}^2/r_j\right)$, respectively, and

$$vec\left(\hat{\mathbf{B}}_{ols/hme}\right)=\left(\left(\mathbf{X}\otimes\mathbf{I}_p\right)'\mathbf{V}^{-1}\left(\mathbf{X}\otimes\mathbf{I}_p\right)\right)^{-1}\left(\mathbf{X}\otimes\mathbf{I}_p\right)'\mathbf{V}^{-1}vec\left(\mathbf{Y}'\right).$$

By setting the covariance matrix of the random error as invariant, $\Sigma_j=\Sigma, \forall j=1,2,...,n$, the proposed CS estimator yields

$$vec\left(\hat{\mathbf{B}}_{cs}\right)$$
$$=vec\left(\left(\sum_{j=1}^{n}z_{j0}\bar{\mathbf{y}}_j \quad \sum_{j=1}^{n}z_{j1}\bar{\mathbf{y}}_j \quad \cdots \quad \sum_{j=1}^{n}z_{js_1}\bar{\mathbf{y}}_j \quad \sum_{j=1}^{n}\bar{x}_{j(s_1+1)}\bar{\mathbf{y}}_j \quad \cdots \quad \sum_{j=1}^{n}\bar{x}_{js}\bar{\mathbf{y}}_j\right)\left(\mathbf{X}'\mathbf{X}+\mathbf{C}_u\right)^{-1}\right) \quad (5.3)$$

The i^{th} vector in $vec(\hat{\mathbf{B}}_{cs})$ can be written as

$$\hat{\boldsymbol{\beta}}_{i\_cs}=\left(\mathbf{X}'\mathbf{X}+\mathbf{C}_u\right)^{-1}\mathbf{X}'\bar{\mathbf{y}}_i, \quad (5.4)$$

where the estimates of the q^{th} diagonal element of $\mathbf{C}_u$, $-\sum_{j=1}^{n}\left(\sigma_{ujq}^2 / r_j\right)$, can be written as $-\sum_{h=1}^{g} n_h S_{uqh}^2 / r_h$, $q = s_1 + 1, s_1 + 2, \ldots, s$.

In the case of independently and identically distributed random errors and homogeneous measurement error, the analytical results agree with the findings of Gimenez and Patat (2005). The theoretical proof for the bias of CS estimators is shown based on the specific case, under assumptions (A1) and (A2). In the specific case where the multivariate regression model consists of p dependent variables, one precisely observed independent variable and one independent variable with HME, it is shown that the estimates of $\beta_{0i\_ols/hme}$ and $\beta_{2i\_ols/hme}$ are biased but the estimates of β_{0i_cs} and β_{2i_cs} are asymptotically unbiased, and that the estimates of $\beta_{1i\_ols/hme}$ and β_{1i_cs} are both unbiased for $i = 1, 2, \ldots, p$.

The results of the simulation study show that the OLS estimation of the parameter of the variable not measured with HME is not affected by HME. Meanwhile the estimation of the parameter of the variable measured with HME reveals an underestimated estimator. The CS method outperforms the OLS method since the magnitude of bias and the MSE of the CS estimator are far lower than the OLS estimator when either the number of repeated measurements or sample size increases, and are close to zero when the sample size is large. These results assert that the simulation study conforms to the theoretical proof.

5.2 Discussion

The MEs of the surrogate variables are assumed to be independently distributed with heteroscedastic variances. However, the heteroscedasticity in this study is restricted to the form where the sample of observations can be grouped into several sub-samples with the property that the variance of the measurement error is homogeneous within a group but heterogeneous between groups. In each group, the variance of the measurement error of the surrogate variable is estimated by the pooled variance of the variable with HMEs observed in the repeated measurements.

5.3 Future Research

In future research, the effect of the number of groups and varying the number of repeated measurements in each group of grouped heteroscedasticity could be considered. Some other approaches to solving the problem of HME variance estimation should be investigated intensively to support other types of HME.

BIBLIOGRAPHY

Augustin, T. 2004. An Exact Corrected Log-Likelihood Function for Cox's Proportional Hazards Model under Measurement Error and Some Extensions. **Scandinavian Journal of Statistics**. 31 (March): 43-50.

Augustin, T., Döring, A. and Rummel, D. 2008. Regression Calibration for Cox Regression Under Heteroscedastic Measurement Error - Determining Risk Factors of Cardiovascular Diseases from Error-prone Nutritional Replication Data. In **Recent Advances in Linear Models and Related Areas**. Shalabh, Christian Heumann. Heidelberg: Springer. Pp. 253–278.

Buonaccorsi, J. P. 2010. **Measurement Error Models, Methods and Applications**. New York: Chapman and Hall.

Buzas, J. S.; Stefanski, L. A. and Tosteson, T. D. 2003. Measurement Error. In **Handbook of Epidemiology**. W. Ahrens and I. Pigeot, eds. Berlin, Heidelberg: Springer Berlin Heidelberg. Pp. 729–765.

Carroll, R. J. 2005. Measurement Error in Epidemiologic Studies. In **Encyclopedia of Biostatistics**. Retrieved May 24, 2016 from http://doi.org/10.1002/ 0470011815.b2a03082

Carroll, R. J., and Stefanski, L. A. 1990. Approximate Quasi-Likelihood Estimation in Models with Surrogate Predictors. **Journal of the American Statistical Association**. 85 (411): 652–663.

Carroll, R. J.; Ruppert, D.; Stefanski, L.A. and Crainiceanu, C.M. 2006. **Measurement Error in Nonlinear Models: A Modern Perspective**. 2nd ed. Northwestern, IL: Chapman and Hall.

Carstensen B.; Gurrin, L. and Ekstrom, C. 2012. Functions for Analysis of Method Comparison Studies. In **Meth Comp.** R Package (Version 1.15). Retrieved September 10, 2015 from https://cran.r project.org/ web/ packages/MethComp/

Chen, J., Hanfelt, J. J. and Huang, Y. 2015. A Simple Corrected Score for Logistic Regression with Errors-in-Covariates. **Communications in Statistics - Theory and Methods.** 44 (10): 2024-2036.

Cheng, C. L. and Riu, J. 2006. On Estimating Linear Relationships when both Variables are Subject to Heteroscedastic Measurement Errors. **Technometrics**. 48 (4): 511-519.

Clayton, D. G. 1991. Models for the Analysis of Cohort and Case-Control Studies with Inaccurately Measured Exposures. In **Statistical Models for Longitudinal Studies of Health**. J.H. Dwyer, M. Feinleib, P. Lipsert et al., eds. New York: Oxford University Press. Pp. 301-331.

Cook, J. and Stefanski, L. A. 1994. Simulation-Extrapolation Estimation in Parametric Measurement Error Models. **Journal of the American Statistical Association.** 89 (428): 1314-1328.

de Castro, M.; Bolfarine, H. and Castilho, M.V. 2006. Consistent Estimation and Testing in Comparing Analytical Bias Models. **Environmetrics**. 17 (March): 167-182.

de Castro, M.; Bolfarine, H. and Galea, M. 2008. Hypothesis Testing in an Errors-in Variables Model with Heteroscedastic Measurement Errors. **Statistics in Medicine**. 27 (November): 5217-5234.

Fuller, W. A. 1997. **Measurement error models**. New York: John Wiley & Sons.

Giménez, P. and Bolfarine, H. 1997. Corrected Score Functions in Classical Error-in Variables and Incidental Parameters Models. **The Australian Journal of Statistics**. 39 (September): 325-344.

Giménez, P. and Bolfarine, H. 2000. Comparing Consistent Estimators in Comparative Calibration Models. **Journal of Statistical Planning and Inference**. 86 (April): 143-155.

Giménez, P. and Galae, M. 2013. Influence Measures on Corrected Score Estimators in Functional Heteroscedastic Measurement Error Models. **Journal of Multivariate Analysis**. 114 (February): 1-15.

Giménez, P. and Patat, M. L. 2005. Estimation in Comparative Calibration Models with Replicate Measurement. **Statistics & Probability**. 71 (February): 155-164.

Giménez, P. and Patat, M. L. 2014. Local Influence for Functional Comparative Calibration Models with Replicated Data. **Statistical Papers**. 55 (2): 431-454.

Gleser, L. J. 1981. Estimation in Multivariate Errors in Variables Regression Model: Large Sample Results. **The Annals of Statistics.** 9 (1): 24-44.

Gujarati, D.N. 2006. **Essentials of Econometrics**. 3rd ed. New York: McGraw-Hill.

Huang, Y. and Wang, C. Y. 2001. Consistent Functional Methods for Logistic Regression with Errors in Covariates. **Journal of the American Statistical Association**. 96 (456): 1469–1482.

Johnson, R. A. and Wichern, D.W. 2002. **Applied Multivariate Statistical Analysis**. 5th ed. Upper Saddle River, NJ: Prentice-Hall.

Judge, G.G. et. al. 1985. **The Theory and Practice of Econometrics.** 2nd ed. New York: Wiley.

Kulathinal, S. B.; Kuulasmaa, K. and Gasbarra, D. 2002. Estimation of an Errors-in-Variables Regression Model when the Variances of the Measurement Errors Vary between the Observations. **Statistics in Medicine**. 21 (April): 1089-1101.

Nakamura, T. 1990. Corrected Score Function for Errors-in-Variables Models: Methodology and Application to Generalized Linear Models. **Biometrika**. 77 (1): 127-137.

Patriota, A. G.; Bolfarine, H. and de Castro, M. 2009. A Heteroscedastic Structural Errors-in-Variables Model with Equation Error. **Statistical Methodology**. 6 (July): 408-423.

Patriota, A. G. and Lemonte, A.J. 2009. Bias Correction in a Multivariate Normal Regression Model with General Parameterization. **Statistics and Probability Letters.** 79 (August): 1655-1662.

Patriota, A. G.; Lemonte, A.J. and Bolfarine, H. 2011. Improved Maximum Likelihood Estimators in a Heteroskedastic Errors-in-Variables Model. **Statistical Papers**. 52 (May): 455-467.

Pierce, D. A. and Kellerer, A. M. 2004. Adjusting for Covariate Errors with Nonparametric Assessment of the True Covariate Distribution. **Biometrika**. 91 (4): 863-876.

Prentice R. L. 2016. Covariate Measurement Errors and Parameter Estimation in a Failure Time Regression Model. **Biometrika Trust**. 69 (2): 331–342.

Rosner B.; Willett, W. C. and Spiegelman, D. 1990. Correction of Logistic Regression Relative Risk Estimates and Confidence Intervals for Systematic Within-Person Measurement Error. **Statistics in Medicine**. 8 (4): 1051–1069.

Veenendaal, E.M. et al. 2008. Growth Form and Seasonal Variation in Leaf Gas Exchange of Colophospermum Mopane Savanna Trees in Northwest Botswana. **Tree Physiology**. 28 (3): 417-424.

APPENDICES

Appendix A

Proof of Lemma 1

Lemma 1. In a grouped heteroscedasticity, the *n* observations can be grouped into *h* groups such that the variance of the measurement errors, σ^2_{u2h}, is homogeneous within the h^{th} group but heterogeneous across the groups. Let r_j be the number of repeated measurements of the j^{th} observation and u_{j2k} be the random measurement error of the j^{th} observation of x_{j2} in the k^{th} repeated measurement independently distributed as $N(0,\sigma^2_{uj2})$. Then, as $n \to \infty$,

$$\sum_{j=1}^{n} \bar{u}^2_{j2\cdot} = n\left(S^2_{\bar{u}_2} + \bar{u}^2_2\right) \to \sum_{h=1}^{g} \frac{n_h \sigma^2_{u2h}}{r_h}.$$

Proof.

The sample variance of the average of the random measurement errors can be written as

$$S^2_{\bar{u}_2} = \frac{1}{n}\sum_{j=1}^{n}\left(\bar{u}_{j2.} - \bar{u}_2\right)^2, \tag{A.1}$$

which can be estimated by the pooled variance as

$$S^2_{\bar{u}_2} = \frac{1}{n}\sum_{h=1}^{g}\frac{n_h S^2_{u2h}}{r_h}, \tag{A.2}$$

where S^2_{u2h} is the sample variance of the measurement error in the h^{th} group of n_h observations, r_h is the number of repeated measurements of the observation in the h^{th} group.

From (A.1), the sum of squares of the average of random measurement errors of the j^{th} observation can be expressed as

$$\sum_{j=1}^{n}\bar{u}^2_{j2.} = n\left(S^2_{\bar{u}_2} + \bar{u}^2_2\right). \tag{A.3}$$

Substituting (A.2) into (A.3) yields

$$\sum_{j=1}^{n} \bar{u}_{j2.}^{2} = \sum_{h=1}^{g} \frac{n_h S_{u2h}^{2}}{r_h} + n\bar{u}_{2}^{2} . \tag{A.4}$$

As $n \to \infty$, (A.4) becomes

$$\sum_{j=1}^{n} \bar{u}_{j2.}^{2} \to \sum_{h=1}^{g} \frac{n_h \sigma_{u2h}^{2}}{r_h} .$$

□

Appendix B

Population Data

Table B.1 Pearson Correlation for Generator Variables of HME Form $F1$ and $F2$

Variable		a_e1	a_e2	a_z_2	a_u
a_e1	Pearson Correlation	1.0000	0.0015	0.0003	0.0001
	Sig. (2-tailed)	-	0.1290	0.7850	0.3409
a_e2	Pearson Correlation	0.0015	1.0000	-0.0005	-0.0018
	Sig. (2-tailed)	0.1290	-	0.6291	0.0754
a_z_2	Pearson Correlation	0.0003	-0.0005	1.0000	0.0016
	Sig. (2-tailed)	0.7850	0.6291	-	0.1216
a_u	Pearson Correlation	0.0010	-0.0018	0.0016	1.0000
	Sig. (2-tailed)	0.3409	0.0754	0.1216	-

Table B.2 Covariance Estimated for Population Data of HME Form $F1$ and $F2$

Variable	e1	e2
e1	0.79996	0.4485
e2	0.4485	1.0018

Table B.3 HME Variances Estimated for Grouped Heteroscedasticity of $F1$

$\hat{\sigma}^2_{u21}$	$\hat{\sigma}^2_{u22}$	$\hat{\sigma}^2_{u23}$	$\hat{\sigma}^2_{u24}$	$\hat{\sigma}^2_{u25}$
0.099616	0.19966	0.39945	0.59882	0.79522

Table B.4 Pearson Correlation for Variables in the Model of HME Form $F1$

Variable		y1_true	y2_true	e1	e2	z1	z2	x2	u
y1_true	Pearson Correlation	1	0.9138	0.3951	0.1974	0.2561	0.88278	0.8401	0.0015
	Sig. (2-tailed)	-	< .0001*	< .0001*	< .0001*	< .0001*	< .0001*	< .0001*	0.1250
y2_true	Pearson Correlation	0.9138	1	0.2175	0.4330	0.2509	0.8659	0.8239	0.0008
	Sig. (2-tailed)	< .0001*	-	< .0001*	< .0001*	< .0001*	< .0001*	< .0001*	0.4102
e1	Pearson Correlation	0.3951	0.2175	1	0.5009	0.0014	0.0003	0.0005	0.0008
	Sig. (2-tailed)	< .0001*	< .0001*	-	< .0001*	0.1768	0.7850	0.6043	0.4003
e2	Pearson Correlation	0.1974	0.4330	0.5009	1	0.0004	−0.0003	−0.0005	−0.0008
	Sig. (2-tailed)	< .0001*	< .0001*	< .0001	-	0.6835	0.6835	0.6023	0.4107
z1	Pearson Correlation	0.2561	0.2509	0.0014	0.0004	1	0.0010	0.0011	0.0004
	Sig. (2-tailed)	< .0001*	< .0001*	0.1768	0.6835	-	0.3013	0.2725	0.7106
z2	Pearson Correlation	0.88277	0.8659	0.0003	−0.0003	0.0010	1	0.9515	0.0013
	Sig. (2-tailed)	< .0001*	< .0001*	0.7850	0.6835	0.3013	-	< .0001*	0.2092
x2	Pearson Correlation	0.8401	0.8239	0.00052	−0.0005	0.0011	0.9515	1	0.3088
	Sig. (2-tailed)	< .0001*	< .0001*	0.6043	0.6023	0.2725	< .0001*	-	< .0001*
u	Pearson Correlation	0.0015	0.0008	0.0008	−0.0008	0.0004	0.0013	0.3088	1
	Sig. (2-tailed)	0.1250	0.4102	0.4003	0.4107	0.7106	0.2092	< .0001*	-

* significance at level 0.0001

Table B.5 Characteristics of the OLS Estimators for Population HME Form $F1$

Parameter	Estimate	SE	t	p-value
β_{01}	-0.0002	0.0009	-0.24	0.8075
β_{11}	1.0021	0.0015	646.73	<.0001*
β_{21}	1.0001	0.0004	2237.04	<.0001*
β_{02}	0.0000	0.0010	0.00	0.9999
β_{12}	1.0007	0.0017	577.12	<.0001*
β_{22}	0.9999	0.0005	1998.48	<.0001*

* significance at level 0.0001

Table B.6 Characteristics of the OLS Estimators without adjustment for Population HME Form $F1$

Parameter	Estimate	SE	t	p-value
$\beta_{01\_ols/hme}$	-0.0015	0.0011	-1.38	0.1668
$\beta_{11\_ols/hme}$	1.0021	0.0019	532.97	<.0001*
$\beta_{21\_ols/hme}$	0.9052	0.0005	1754.44	<.0001*
$\beta_{02\_ols/hme}$	-0.0013	0.0012	-1.09	0.2753
$\beta_{12\_ols/hme}$	1.0007	0.0020	491.47	<.0001*
$\beta_{22\_ols/hme}$	0.9048	0.0006	1619.22	<.0001*

* significance at level 0.0001

Table B.7 Pearson Correlation for Variables in the Model of HME Form $F2$

Variable		y1_true	y2_true	e1	e2	z1	z2	x2	u
y1_true	Pearson Correlation	1	0.9135	0.3962	0.1970	0.2558	0.8824	0.8395	0.0019
	Sig. (2-tailed)	-	< .0001*	< .0001*	< .0001*	< .0001*	< .0001*	< .0001*	0.0589
y2_true	Pearson Correlation	0.9135	1	0.2185	0.4331	0.2507	0.8654	0.8233	0.0016
	Sig. (2-tailed)	< .0001*	-	< .0001*	< .0001*	< .0001*	< .0001*	< .0001*	0.1181
e1	Pearson Correlation	0.3962	0.2185	1	0.5009	0.0014	0.0010	0.0010	0.0002
	Sig. (2-tailed)	< .0001*	< .0001*	-	< .0001*	0.1768	0.3409	0.3322	0.8346
e2	Pearson Correlation	0.1970	0.4331	0.5009	1	0.0004	–0.0011	–0.0012	–0.0005
	Sig. (2-tailed)	< .0001*	< .0001*	< .0001	-	0.6835	0.2876	0.2458	0.6276
z1	Pearson Correlation	0.2558	0.2507	0.0014	0.0004	1	0.0003	0.0007	0.0012
	Sig. (2-tailed)	< .0001*	< .0001*	0.1768	0.6835	-	0.7407	0.5034	0.2506
z2	Pearson Correlation	0.8824	0.8654	0.0010	–0.0011	0.0003	1	0.9515	0.0013
	Sig. (2-tailed)	< .0001*	< .0001*	0.3409	0.2876	0.7407	-	< .0001*	0.0861
x2	Pearson Correlation	0.8395	0.8233	0.0010	–0.0012	0.0007	0.9512	1	0.3101
	Sig. (2-tailed)	< .0001*	< .0001*	0.3322	0.2458	0.5034	< .0001*	-	< .0001*
u	Pearson Correlation	0.0019	0.0016	0.0002	–0.0005	0.0012	0.0017	0.3101	1
	Sig. (2-tailed)	0.0589	0.1181	0.8346	0.6276	0.2506	0.0861	< .0001*	-

* significance at level 0.0001

Table B.8 HME Variances Estimated for Grouped Heteroscedasticity of $F2$

$\hat{\sigma}^2_{u21}$	$\hat{\sigma}^2_{u22}$	$\hat{\sigma}^2_{u23}$	$\hat{\sigma}^2_{u24}$	$\hat{\sigma}^2_{u25}$
0.7967	0.6017	0.4018	0.1999	0.1005

Table B.9 Characteristics of the OLS Estimators for Population HME Form $F2$

Parameter	Estimate	SE	t	p-value
β_{01}	-0.0002	0.0009	-0.25	0.8059
β_{11}	1.0021	0.0015	646.74	<.0001*
β_{21}	1.0004	0.0004	2233.77	<.0001*
β_{02}	0.0000	0.0010	0	0.9981
β_{12}	1.0007	0.0017	577.12	<.0001*
β_{22}	0.9995	0.0005	1994.18	<.0001*

* significance at level 0.0001

Table B.10 Characteristics of the OLS Estimators without adjustment for Population HME Form $F2$

Parameter	Estimate	SE	t	p-value
$\beta_{01\_ols/hme}$	0.0008	0.0011	0.70	0.4816
$\beta_{11\_ols/hme}$	1.0010	0.0019	532.03	<.0001*
$\beta_{21\_ols/hme}$	0.9047	0.0005	1749.86	<.0001*
$\beta_{02\_ols/hme}$	0.0010	0.0012	0.84	0.4016
$\beta_{12\_ols/hme}$	0.9997	0.0020	490.94	<.0001*
$\beta_{22\_ols/hme}$	0.9038	0.0006	1615.21	<.0001*

* significance at level 0.0001

Appendix C

Oximetry Data

The Oximetry Data for examining the percentage oxygen of the blood of sick children at the Royal Children's hospital in Melbourne is shown as follows:

No.	obs	rep	CO	PULSE
1	1	1	0.5497	0.3889
2	1	2	0.5102	0.4102
3	1	3	0.5297	0.4320
4	2	1	0.3414	0.3274
5	2	2	0.3194	0.3076
6	2	3	0.3334	0.3274
7	3	1	0.6856	0.6585
8	3	2	0.6048	0.5248
9	3	3	0.6213	0.5248
10	4	1	0.2181	-0.1224
11	4	2	0.2842	0.3475
12	4	3	0.3174	0.5248
13	5	1	0.4959	0.5006
14	5	2	0.4475	0.4102
15	5	3	0.5078	0.3274
16	6	1	0.5497	0.5754
17	6	2	0.5702	0.5497
18	6	3	0.5322	0.5497
19	7	1	0.7993	0.7202
20	7	2	0.7299	0.6585
21	7	3	0.8653	0.6585
22	8	1	0.5807	0.4543
23	8	2	0.5754	0.4543
24	8	3	0.7636	0.5248
25	9	1	0.3889	0.3680
26	9	2	0.3763	0.3680
27	9	3	0.4123	0.3889
28	10	1	0.0872	0.2311
29	10	2	0.0539	0.4102

No.	obs	rep	CO	PULSE
32	11	2	0.9542	0.7533
33	11	3	1.0911	0.8653
34	12	1	-0.5446	-0.5006
35	12	2	-0.4748	-0.4102
36	12	3	-0.3254	-0.2499
37	13	1	0.5886	0.3680
38	13	2	0.6764	0.3889
39	13	3	0.8179	0.4771
40	14	1	0.3973	0.6021
41	14	2	0.4320	0.6021
42	14	3	0.2997	0.5006
43	15	1	0.7137	0.6021
44	15	2	0.7234	0.5497
45	15	3	0.6674	0.5248
46	16	1	0.5054	0.2499
47	16	2	0.5497	0.2881
48	16	3	0.6795	0.4771
49	17	1	0.8179	0.7884
50	17	2	1.1301	0.9080
52	18	1	0.5650	0.4102
53	18	2	0.5573	0.4543
54	18	3	0.5754	0.4771
55	19	1	0.4059	0.3889
56	19	2	0.4588	0.4771
57	19	3	0.4543	0.3889
58	20	1	0.4498	0.3274
59	20	2	0.5396	0.4771
61	21	1	0.4320	0.4771
62	21	2	0.5754	0.6021
63	21	3	1.1508	0.8653
64	22	1	0.5396	0.2688
65	22	2	0.6795	0.7884
66	22	3	0.3475	0.2688
67	23	1	0.5624	0.5754
68	23	2	0.5967	0.6585
69	23	3	0.5886	0.5754
70	24	1	0.6645	0.6021
71	24	2	0.7848	0.7202
72	24	3	0.7884	0.6021
73	25	1	0.7776	0.5497
74	25	2	0.3995	0.4102
76	26	1	0.5650	0.5006

No.	obs	rep	CO	PULSE
77	26	2	0.5676	0.5248
78	26	3	0.5807	0.5006
79	27	1	0.2574	0.0696
80	27	2	0.3700	0.0872
81	27	3	0.4276	0.3076
82	28	1	0.1761	0.1224
83	28	2	0.2126	0.3680
84	28	3	0.2978	0.3889
85	29	1	0.5522	0.4771
86	29	2	0.5297	0.4771
87	29	3	0.5174	0.3889
88	30	1	0.5860	0.4102
89	30	2	0.6382	0.4102
90	30	3	0.6948	0.4543
91	31	1	-0.0087	-0.4102
92	31	2	-0.0959	-0.0696
93	31	3	-0.0889	0.1943
94	32	1	0.7602	0.5006
95	32	2	0.6917	0.5754
96	32	3	0.6498	0.5006
97	33	1	0.6382	0.7533
98	33	2	0.6645	0.6021
99	33	3	0.6326	0.6585
100	34	1	0.8217	0.6585
101	34	2	0.8030	0.5754
102	34	3	0.7671	0.6021
103	35	1	0.6440	0.5754
104	35	2	0.4342	0.3076
105	35	3	0.6326	0.3889
106	36	1	0.7234	0.5006
107	36	2	0.3995	0.4771
108	36	3	0.8736	0.3680
109	37	1	0.8372	0.6021
110	37	2	0.5913	0.6021
111	37	3	0.4771	0.3274
112	38	1	0.7365	0.6297
113	38	2	0.6825	0.7202
114	38	3	0.6213	0.6585

No.	obs	rep	CO	PULSE
115	39	1	0.5322	0.5754
118	40	1	0.3095	0.2126
119	40	2	0.3354	0.2311
120	40	3	0.3910	0.2499
121	41	1	0.5676	0.3889
122	41	2	0.5272	0.3680
123	41	3	0.5548	0.3889
124	42	1	0.2688	0.4320
125	42	2	0.4771	0.4543
126	42	3	0.3174	0.3889
127	43	1	0.0400	0.1581
128	43	2	0.0139	0.1224
129	43	3	0.0365	0.1224
130	44	1	0.1224	0.1761
131	44	2	0.1206	0.1761
132	44	3	0.1509	0.1943
133	45	1	0.4959	0.4543
134	45	2	0.5886	0.4771
135	45	3	0.6297	0.4771
136	46	1	0.7499	0.5248
137	46	2	0.5650	0.5248
138	46	3	0.6382	0.5006
139	47	1	0.6048	0.5006
140	47	2	0.5728	0.5497
141	47	3	0.5860	0.5248
142	48	1	0.3475	0.3076
143	48	2	0.3597	0.4320
144	48	3	0.4276	0.6585
145	49	1	0.9640	0.7202
146	49	2	0.8612	0.6886
147	49	3	0.8411	0.5248
148	50	1	0.0452	-0.0872
149	50	2	0.0139	0.0174
151	51	1	0.1153	-0.1047
152	51	2	0.1743	0.0348
153	51	3	0.1671	-0.0348
154	52	1	0.4145	0.3076
155	52	2	0.5754	0.4320
156	52	3	0.5833	0.4771
157	53	1	0.3414	0.3889
158	53	2	0.4959	0.5006

No.	**obs**	**rep**	**CO**	**PULSE**
159	53	3	0.3174	0.3680
160	54	1	1.0320	1.0607
161	54	2	0.1384	0.2688
162	54	3	0.5102	0.5248
163	55	1	0.9640	1.0607
164	55	2	0.8067	0.7533
165	55	3	0.7776	0.8653
166	56	1	0.7398	0.7202
167	56	2	0.7741	0.6585
168	56	3	0.8411	0.7202
169	57	1	1.1579	1.0607
170	57	2	1.0320	0.9542
171	57	3	0.6241	0.6585
172	58	1	0.8778	0.7202
173	58	2	0.9494	0.8256
174	58	3	1.0210	0.9080
175	59	1	0.9260	1.1950
176	59	2	0.9306	0.9080
177	59	3	1.0376	1.0048
178	60	1	0.8333	0.5754
179	60	2	0.7042	0.5248
180	60	3	0.7432	0.6021
181	61	1	0.4982	0.4102
182	61	2	0.5174	0.3680
183	61	3	0.4611	0.3274

Dependent Variable: PULSE

Number of Observations Read	61
Number of Observations Used	61

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	1	2.90633	2.90633	279.96	<.0001
Error	59	0.61249	0.01038		
Corrected Total	60	3.51881			

Root MSE	0.10189	**R-Square**	0.8259
Dependent Mean	0.46595	**Adj R-Sq**	0.8230
Coeff Var	21.86648		

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	**1**	0.03427	0.02891	1.19	0.2407
CO	**1**	0.80922	0.04836	16.73	<.0001

FIGURE C.1 SAS Output of the Oximetry Data

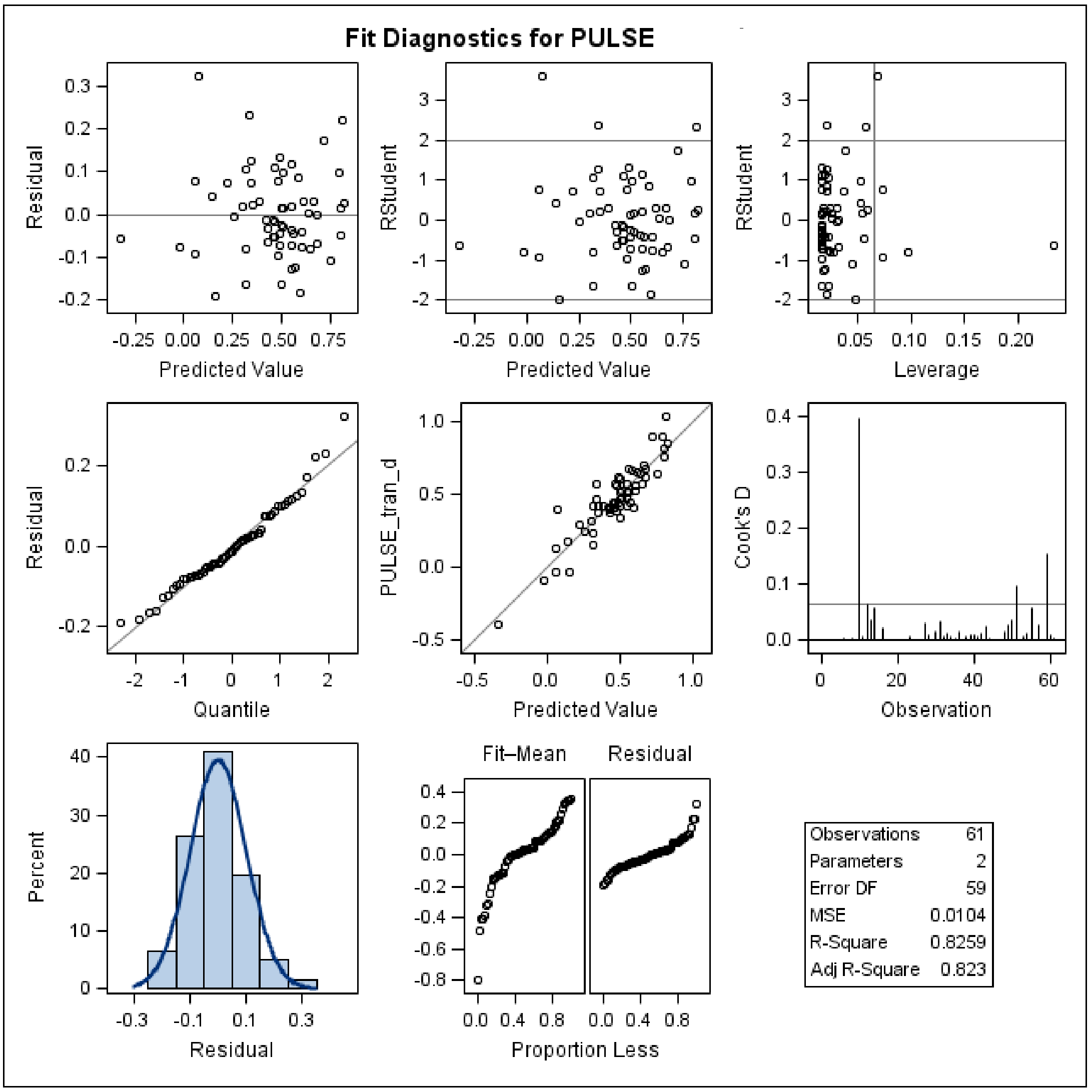

FIGURE C.2 Residual Plot of the Oximetry Data

BIOGRAPHY

NAME	Wannaporn Junthopas
ACADEMIC BACKGROUND	Bachelor Second Class Honour Degree with a major in Statistics from Khon Kaen University, Khon Kaen Province, Thailand in 2005. Master's Honour Degree Applied Statistics and Information Technology from National Institute of Development Administration, Bangkok, Thailand in 2009.
PRESENT POSITION	Lecturer, Statistics Program Faculty of Science, Khon Kaen University, Khon Kaen Province, Thailand.
PUBLICATION	Wannaporn Junthopas and Jirawan Jitthavech. 2012. The Comparison of Tests and Corrections for Heteroscedasticity Problem in Simple Linear Regression. **Burapha Science Journal**. 17 (1): 97-107. Wannaporn Junthopas. 2015. Corrected Score Estimator in a Linear Multivariate Regression Model with Measurements Errors of Heteroscedasticity Groups. In

Proceedings of the International Conference on Applied Statistics 2015 (ICAS2015), 15 - 17 May 2015.
Chonburi, Thailand.