UNIVERSITY FOR DEVELOPMENT STUDIES

UNIVERSITY FOR DEVELOPMENT STUDIES

ESTIMATION MAXIMIZATION CONCEPT OF ASSESSING RESIDUALS: A STRUCTURAL EQUATION MODELLING APPROACH

BY

ABDUL-RAHAMAN ABDUL-AZIZ (MPhil., BSc.)

(UDS/DAS/0033/15)

A Thesis submitted to the Department of STATISTICS, University FOR DEVELOPMENT STUDIES, in partial fulfillment of the requirement for the degree of

DOCTOR of Philosophy in APPLIED STATISTICS.

OCTOBER, 2019

DECLARATION

I hereby declare that this submission is my own work towards the award of PhD. (Applied Statistics) degree and that, to the best of my knowledge, it contains no material previously published by another person nor material which had been accepted for the award of any other degree of the University, except where due acknowledgement had been made in the text.

Student Name and ID:

ABDUL-RAHAMAN ABDUL-AZIZ……………….…… …………………

(UDS/DAS/0033/15) Signature Date

Supervisors'

I hereby declare that the preparation and presentation of the dissertation/thesis was supervised in accordance with the guidelines on supervision of dissertation/thesis laid down by the University for Development Studies

PROF. ALBERT LUGUTERAH ……………….…… …………………

(Main supervisor) Signature Date

PROF. BASHIRU I.I. SAEED ……………….…… …………………

(Co-supervisor) Signature Date

DEDICATION

This work is dedicated to my lovely and caring wife and children for their continuous

support and inspiration in my life.

Sherifa Sani,

Sameeha Abdul-Aziz

And

Ramlah abdul-Aziz

UNIVERSITY FOR DEVELOPMENT STUDIES

ii

ACKNOWLEDGEMENT

My sincere thanks go to the Almighty God for His protection and guidance throughout this programme.

My deepest and profound gratitude to Prof. Albert Luguterah and Prof. Bashiru I. I. Saeed for spending a lot of their time reading this thesis, criticizing where necessary and offering very good suggestions and pieces of advice.

Again, I wish to thank my lovely and caring parents Alhaji Abdul-Rahaman Ahmed and Hajia Afishetu Abdulai for their continuous support and prayers in my life.

Further, I wish to express my appreciation to Mr. Ahmed Hafiz A. Rahman, Baba Ahmed A. Rahman, Mustapha A. Rahman and all my brothers for being there for me.

Finally, I wish to thank all the staff of the Statistics Department of Kumasi Technical University (KsTU), especially Mr. Kwame Annin for their spirit of encouragement.

To all other Tom, Dick and Harry who have in diverse ways contributed to this work but their names have not been mentioned, I deeply appreciate you all.

God bless you all.

UNIVERSITY FOR DEVELOPMENT STUDIES

ABSTRACT

The main objective of this study was to apply estimation maximization concept to assess residuals through structural equation modelling. This was achieved through simulation setup. Data was obtained from test scores of some selected course lecturers of Kumasi Technical University to validate the results from the simulation. The results showed that the estimated residuals of the measurement errors using all three estimators correlate negatively with the estimated residuals associated with measurement errors of items that load on the same factor. These correlations are strongest when using the Bartlett's method-based estimator and weakest when using the regression method-based estimator. Thus, the Bartlett's method-based residual estimators are among the three estimators that achieved very close values. Also, it can be deduced from the results on the various simulation of quantile-quantile plots that all these methods demonstrate the ability to detect outliers and potential influential observation in a SEM framework. It is worth noting that the Anderson-Rubin method provided a quantile-quantile plot which was more efficient in terms of visual display for detecting outliers and potential influential observations as compared to the other class of residual estimators.

Finally, it was therefore found from the comparative model fits information, by comparing among the three existing residual estimators, that the Bartlett's based method gave better residual parameter estimates over the regression based method and the Anderson Rubin based method. However, the estimation maximization method gave better residual parameter estimates than the other three existing methods. It is therefore worth noting that this present study contribution to knowledge is demonstration of the fact that estimation maximization method could be a better residual estimator within the SEM framework compared to other existing methods.

iv

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

UNIVERSITY FOR DEVELOPMENT STUDIES

LIST OF SYMBOLS AND THEIR DIMENSION

Symbol	Dimension Definition
Variable	
x	$q \times 1$ observed indicators of ξ
y	$p \times 1$ observed indicators of η
δ	$q \times 1$ measurement errors of x
ε	$p \times 1$ measurement errors of y
Coefficient	
Λ_x	$q \times n$ coefficients relating x to ξ
Λ_y	$p \times m$ coefficients relating y to η
Covariance matrices	
θ_δ	$q \times q$ $E(\delta\delta')$ (i.e. covariance matrix of δ)
θ_ε	$p \times p$ $E(\varepsilon\varepsilon')$ (i.e. covariance matrix of ε)

Notation for the SEM and Residual Estimators

Variables

η_i - factor scores of the endogenous latent variables

ξ_i - factor scores of the exogenous latent variables

L_i - stacked vector of factor scores of the endogenous and exogenous latent variables variables such that $L_i = \begin{bmatrix} \eta_i \\ \xi_i \end{bmatrix}$

ζ_i - latent errors in equations

y_i - indicators of η_i

x_i - indicators of ξ_i

z_i - indicators of L_i such that $z_i = \begin{bmatrix} y_i \\ x_i \end{bmatrix}$

ε_i - measurement errors of y_i

δ_i - measurement errors of x_i

ν_i - measurement errors of z_i such that $\nu_i = \begin{bmatrix} \varepsilon_i \\ \delta_i \end{bmatrix}$

Coefficients

B - coefficient matrix of the endogenous latent variables

Γ - coefficient matrix of the exogenous latent variables

M - re-expressed coefficient matrix such that $M = [(I - B)^{-1} - \Gamma]$

Λ_y - coefficient matrix relating y_i to η_i

Λ_x - coefficient matrix relating x_i to ξ_i

Λ - coefficient matrix relating z_i to L_i such that $\Lambda = \begin{bmatrix} \Lambda_y & 0 \\ 0 & \Lambda_x \end{bmatrix}$

Weight Matrices

W_r - regression method-based weight matrix such that $W_r = \Sigma_{LL} \Lambda' \Sigma_{zz}^{-1}$

W_b - Bartlett's method-based weight matrix such that $W_b = (\Lambda' \Sigma_{vv}^{-1} \Lambda)^{-1} \Lambda' \Sigma_{vv}^{-1}$

W_{ar} - Anderson-Rubin method-based weight matrix such that $W_{ar} = F^{-1} \Lambda' \Sigma_{vv}^{-1}$

Where $F^2 = \left(\Lambda' \Sigma_{vv}^{-1} \Sigma_{zz} \Sigma_{vv}^{-1} \Lambda \right)$

Covariance Matrices

Φ - covariance matrix of ξ_i

$\Sigma_{\eta\eta}$ - covariance matrix of η_i

Σ_{LL} - covariance matrix of L_i such that $\Sigma_{LL} = \begin{bmatrix} \Sigma_{\eta\eta} & (I - B)^{-1} \Gamma \Phi \\ \Phi \Gamma' (I - B)^{-T} & \Phi \end{bmatrix}$

Ψ - covariance matrix of ζ_i

Θ_ε - covariance matrix of ε_i

Θ_δ - covariance matrix of δ_i

Σ_{vv} - covariance matrix of ν_i such that $\Sigma_{vv} = \begin{bmatrix} \theta_\varepsilon & 0 \\ 0 & \theta_\delta \end{bmatrix}$

Σ_{zz} - covariance matrix of z_i such that

xii

CHAPTER ONE

INTRODUCTION

1.1 Background to the Study

Structural equation models (SEM) are extensions of the usual linear regression models potentially involving unobservable random variables that are not error terms. Also, a random variable may appear as an independent variable in one equation and a dependent variable in another. A random variable in a structural equation model can be classified in two ways. It may be either latent or manifest, and either exogenous or endogenous (Blalock 1971, Goldberger 1972, Goldberger & Duncan 1973, Aigner & Goldberger 1977, Bielby & Hauser 1977, Bentler & Weeks 1980, Aigner et al, 1984, Joreskog & Wold 1982, and Bollen 1989). Latent variable is a random variable that is not observable. Manifest variable is a random variable that is observable. It is part of the data set. Exogenous Variable is a random variable that is not written as a function of any other variable in the model. Endogenous variable is random variable that is written as a function of at least one other variable in the model.

According to Ringle et al (2009) the manifest model links every construct variable to the measured variables for which it is related to and therefore indicating the synthesis of many variables into a combine (and at times latent) variables. On the other hand, the structural model (Ringle et al, 2009) links the combine (latent) variables in a model to every other. Therefore, a procedure for computing purposes, usually known as estimation method, becomes imperative in order to estimate the parameter values that describes these associations. Under the SEM concept, both the explanatory and the response variables could be construct or manifest (Lee & Xia, 2008). Therefore, structural equation modeling is said to be a method for assessing a number of associations as well

as examining a quantitative value to every case base on the covariances among the variables. The quantitative values are known as parameter estimates, which are numeric approximations of the degree as well as the direction of without-variable associations that might be considered under a population (Bollen, 1989; Kline, 2011). SEM is also the commonest method which is applied across varied fields including education, psychology, sociology, economics, marketing research, just to list a few (Monecke & Leisch, 2012). SEM is basically calculates the coefficients all together for multiple regression regarding a system in which explanatory as well as the response variables are supposed to be interrelated in possibly complicated ways since some particular variables can be both response and explanatory variables while some response variables have many explanatory variables, among others (Bollen, 1989; Haenlein & Kaplan, 2004; Kline, 2011). Moreover, SEM seeks, mainly, to identify a particular set of parameter estimates (i.e., path coefficients, error terms, etc.) which reduces the total difference between the implied covariances through the model and those measured under population. Thus, generally, SEM is made up of manifest model(s) as well as the structural model (Bollen, 1989; Kline, 2011).

SEM is a general concept which adapts to different aspects of the life of people, both socially and scientifically. Contrary to many other statistical methodologies that lay emphasis on modelling single and/or many variables, SEM often lay place prominence in modelling measured observations which in terms of covarience matrices so that the parameters attained could minimize the disparity among the measured and the predicted convariance during the modelling process. A residual in SEM mainly refers to the disparity among measured and predicted covariance. As a result of the aforementioned reason the term residual in SEM is significantly different form that which apply in

traditional statistics. Thus SEM differ hugely as it emphasizes on the measured convariance observations.

Moreover, SEM is very amenable technique and has the ability to model intricate problems with pictorial models which the other statistical methodologies would not be able to generate. Again it has the capability of modeling every facet of human life by adopting and utilizing manifest observations and observed variables along with their error terms. As a result of its convenience it is utilized in societal issues and scientific related disciplines.

1.2 Statement of Problem

There is substantial disparity between structural equation modelling and other statistical methodologies. This can be attributed to the former's ability in modelling the covariances associated with measured or indicator variables contrary to the latter which can model only the individual elements. Again, the process of analyzing residuals clearly demonstrates the how SEM differ from other statistical methodologies. For quite a number of statistical methodologies, analyzing residuals basically involve displaying graphics regarding residuals. Diagnostics of a model using residuals, testing hypothesis about residuals to examine assumptions of a model as well as spotting possible outliers and controlling elements are among the commonest methods utilized regarding the application of residuals in many statistical methodologies other than SEM. Meanwhile, many established model diagnostics, have been utilized in assessing fitness of any hypothesized structural equation model as well as a number simulation researches that have examined the influence of misspecifying models, departure from assumptions as well as outliers.

However, it is worth noting that the use of residual analysis and by extension residual estimators akin to what is often utilized under other statistical methodologies to estimate residuals for a structural equation model has been largely neglected. Moreover, the estimation maximization method has rarely been utilized in SEM. To this end, the study applied the estimation maximization (EM) concept, through simulation, to analyze residuals under the SEM framework.

1.3 Objectives of the Study

The study is generally aimed at modelling residuals, via simulation, in structural equation models using the EM method.

Also, this study would specifically seek to:

i.   assess some of the finite sample properties of a group of residual estimators which are functions, in terms of weight, of the measured variables;

ii.   determine the ability of residual estimators in detecting outliers and influential observations;

iii.   compare the EM method against other methods of residual estimations.

1.4 Significance of the Study

The research would provide sufficient basis for assessing, estimation maximization, residuals in SEM concept. It will also bring to fore how key properties, which are asymptotically in nature and linearly functions of the measured observations, can be obtained via suggested simulations. Derivations of various estimators of residuals would provide basis in theory for application to real life situations across all disciplines or field of study.

The estimators utilized here would make it easier for adoption and application as its weaknesses and strengths would be highlighted. The selection and application of a weight in order to make the estimators robust would make them suitable for its application different data sets.

This study is very general, chiefly linear, chiefly cross-sectional statistical modeling technique in terms of assessing errors or residuals. That is, researchers are more likely to use this study to determine whether or not which residual estimator to use in estimating residuals.

Also, the study has shed light on how outliers and influential observation within the SEM framework, which often cause model fit indices to be exaggerated, could be detected through an efficient graphical display residuals.

Compared to other dimensions of SEM, this study in particular is a relatively young field, having its roots in papers that appeared in recent years. As such, the methodology is still developing, and even fundamental concepts are subject to challenge and revision. This rapid change is a source of excitement for some researchers and a source of frustration for others for furthers studies.

1.5 Organization of the Study

Chapter one entails the background, the statement of problem, the study objectives, significance of the study and the organization of the study. Chapter two reviewed literature based on SEM, the concept of spotting outliers and comparing various estimation methods in SEM. Chapter three looks at the adopted SEM model, how it can be identified and subsequently evaluating the model under certain specific tests. Chapter five looked at the properties, statistically, of the various residual estimations and the

derivation of conditionally unbiased, univocal, among others. Again, the Chapter looked at the weighing concept of the various existing residual estimators and their derivations. The chapter also entails the derivations of the various residual estimations utilized in the study in examining how to spot outliers and influential observations and which estimator estimates residual better in SEM. The Chapter six comprised the summary of key findings, the conclusions and the recommendation. The recommendation were divided into two sections; general and further studies for research. Lastly, the chapter also entails the gains made in terms of contribution to knowledge.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

A number of researches have been carried out on residuals in the structural equation modeling framework and its related issues, in this Chapter, a review of literature on these previous studies would be discussed.

2.2 Asymptotic Properties of SEM

Maximum likelihood (ML) estimation of univicality in SEMs with categorical outcomes are numerical integration with many dimensions (Muthén, 2010), making the estimation process computationally tedious. Moreover, estimation methods like ML, relying on finite sample properties, tend to produce either reliable or unreliable outcomes depending of the sample involved (Asparouhov & Muthén, 2010b; Hox & Maas, 2001; Meuleman & Billiet, 2009). In univocal analysis, estimation of the parameters on the between various techniques is based on as many observed variables as there are latent variables in the data, needing a sufficient number of latent variables in order for the asymptotic properties of estimators to hold (Asparouhov & Muthén, 2010b).

As volition to classical estimation methods, Bayesian estimation could help to overcome some of the weaknesses when estimating in univocal correlation in SEMs with finite samples. The estimation maximization approach has recently been applied successfully in many complicated SEMs, such as two-level nonlinear SEMs (Song & Lee, 2004), multivariate latent curve models (Song et al, 2009), or semiparametric SEMs (Song et al, 2013; Yang & Dunson, 2010). Bayesian analysis can also deal with very convoluted models in cases where classical approaches such as ML estimation often unsuccessful,

as it is the case for the estimation of complicated multilevel SEMs with categorical indicators (Asparouhov & Muthén, 2010b; Muthén & Asparouhov, 2012). Also, sampling based estimation maximization methods do not rely on asymptotic theory and exhibit better finite sample performances for factor analyses (Lee & Song, 2004) or when there are only few clusters in multilevel models (Muthén & Asparouhov, 2012; Hox et al, 2012), and could overcome convergence difficulties in finite samples (Depaoli & Clifton, 2015). Moreover, variations in estimation accuracy for a model with continuous and categorical indicator variables, as well as the influence of different prior input in Bayesian analysis were examined. Inaccurate informative prior conditions were included in the simulation to examine the influence of possible prior misspecifications. Again, they compared Mplus (Muthén & Muthén, 1998, 2012) and Stan (Stan Development Team, 2014b) in their performance with regard to the estimation technique. With increasing computational power and the availability of Markov chain Monte Carlo (MCMC) technique, the number of statistical software packages for Bayesian data analysis has evolved hugely, with widely different implementations of MCMC techniques and algorithms.

In the conventional SEM architecture, ML estimation emphasis on covariances rather than observed variables and/or individual observations, with the goal of minimizing the variation between selected covariances and the covariances predicted by the model at stake (Bollen, 1989). If the model is rightly specified, ML produces consistent and asymptotically unbiased, asymptotically efficient, and asymptotically normally distributed parameter estimates (Bollen, 1989). Challenges with ML estimation technique arise when numerical integration is required with multiplex measurements of integration, as same is observed in confirmatory factor analysis (CFA) for categorical

outcomes (Muthén, 2010; Wirth & Edwards, 2007). Computational demands of numerical integration technique increases as the quantity of factors, currently hitting its limits at a maximum of three to four dimensions of integration, corresponding to models with at most three to four latent variables (Asparouhov & Muthén, 2012a). An alternative estimation technique is weighted least squares (WLS) estimation, which relies on a polychoric correlation matrix using pairwise information and can be utilised even with a comparably more number of latent variables (Asparouhov & Muthén, 2012a). The robust WLSMV estimator as implemented in Mplus uses a diagonal weight matrix for the fitting function, with standard errors and a mean- and variance-adjusted chi-square test statistic that utilises the full weight matrix (Muthén & Muthén, 1998–2012; Asparouhov & Muthén, 2007).

A more extensive study was done on the WLSMV estimation technique and the two-level WLS estimator employed in Mplus by Muthén (1983, 1984) and Asparouhov and Muthén (2007). It is worth noting that the categorical weighted least squares technique is not peculiar to Mplus, but similar, if not identical, application of WLS estimation in other software should produce similar or identical results. However, the WLSMV estimator has been demonstrated to perform well for SEMs with ordinal indicator variables (Beauducel & Herzberg, 2006; Nussbeck et al, 2006) or multilevel SEMs with continuous and categorical indicator variables (Asparouhov & Muthén, 2007; Hox et al, 2010). Previous Monte Carlo researches also showed a deterioration of performance in complex models (Nussbeck et al, 2006) and convergence problems for multilevel SEMs (Depaoli & Clifton, 2015). Furthermore, ML estimation is based on finite sample theory, Bayesian technique does not assume asymptotic arguments and can give better reliable outcome for finite samples (Lee & Song, 2004; Song & Lee, 2012). For instance, in

multilevel SEM with few number of independent observations, Bayes estimation might perform better than ML (Asparouhov & Muthén, 2010b; Baldwin & Fellingham, 2013; Hox et al., 2012). Nonetheless, ML and Bayesian approaches are asymptotically equivalent, with Bayesian estimates partaking in the optimal properties of ML estimates (Song & Lee, 2012). The likelihood often relies on the estimation method, while the prior does not, and the posterior approximates the likelihood (Lynch, 2007). Statistical evolution in SEM basically depend on asymptotics, or properties of estimators in finite samples. In small samples, SEM parameter estimates can be biased, but they become unbiased in large samples (Bentler, 1993). SEM parameters cannot easily be ranked in terms of their relative sampling variability (their relative efficiency), and their relative performance is context-dependent. However, in large samples, some SEM estimators become clear "winners" in that they achieve the smallest sampling variability compared to other estimators. This property is called *asymptotic efficiency*. To put it simply, asymptotic efficiency means the same thing as efficiency, but a large sample is required before the property always holds.

Furthermore, just a few studies have compared these estimation techniques for computing univocal correlations in SEMs (Hox et al, 2012; Depaoli & Clifton, 2015). However, Hox et al (2012) solely focused on between-level finite samples in the context of international comparative surveys, assuming asymptotic properties. In other applied researches, involving the assessment of interpersonal relationships with family members, friends, or colleagues, within-level finite sample often do not comprise less than two and/or more than ten observed variables. Moreover, one current study compared Bayesian with frequentist approaches for a multilevel SEM with dichotomous indicators (Depaoli & Clifton, 2015). The outcome of their study showed the merits of Bayesian

estimation as compared to WLSMV (weighted least squares, mean and variance adjusted) when utilized in combination with informative priors. According to previous researches (Asparouhov & Muthén, 2010b; Lee et al, 2010), their results indicated that for models with categorical items in the observed variable there was a substantial effect of the choice of priors (Depaoli & Clifton, 2015), a phenomenon termed prior assumption dependence (Asparouhov & Muthén, 2010b). As there is less empirical information available in data with categorical indicators, inaccurate prior information may have particularly detrimental effects in these situations. Nonetheless, simulation researches investigating the effect of inaccurate priors on the estimation of multilevel SEMs are scarce. For this section, the current study fills the gap, to the best of our knowledge, by first assessing the finite sample properties of residual estimators which are functions, in terms of weight, of the measured variables in SEMs.

2.3 Detecting Outliers and Influential Observation in SEM

Issues associated with outliers are often looked at in in textbooks, whilst in practical sense academics tend to have divergent views on its meaning and how it can rightfully be determined and managed, if possible (Mark and Jiaqi, 2017). Managing outliers of various kind require different techniques. According to Aguinis et al (2013) there are 14 varied perspectives about outliers including, but not restricted to, issues of high leverage and the ability to overwhelm parameter estimation and model fitness in SEM. Also, outliers of different nature require different treatments, and Aguinis et al (2013) summarized the definitions of outliers in three categories: (a) those due to correctable errors such as input error, (b) those exhibiting idiosyncratic characteristics and of interest themselves (c) those exerting disproportionately large influence on the substantive

conclusion regarding a model of interest. Again, Muthén (2015) defined outliers as extreme observations that have the ability to exaggerate model coefficients.

Outliers are different from controlling observations as was established by Pek and MacCallum (2011). Moreover, Mark and Jiaqi (2017) noted that outliers often cause dissimilar stir on model adequacy as well as parameter estimation. Some techniques for spotting outliers and possible controlling observations in SEM were the likelihood, Mahalanobis and Cook's distances (Aguinis et al, 2013; Pek & MacCallum, 2011; and Yuan & Zhang, 2012). Other studies Yuan and Zhong (2008, 2013) and Asparouhov and Muthén (2015) identified a linear notation for modelling outlier residuals in SEM. However, contemporary methods for identifying and controlling outliers and possible controlling observation in SEM require scientist to utilize special programs which creates more burden for researchers (Sterba & Pek, 2012; Yuan & Zhang, 2012).

In a normal SEM model, very little portion of outliers and potential controlling observations can have a huge impact on model fit and parameter estimates. For instance, Yuan and Bentler (2001) and Yuan and Zhang (2015) demonstrated mathematically that existence of outliers can hugely inflate the Type I error rates of likelihood ratio test (LRT) and associated test statistics balancing for non-normality when using maximum likelihood (ML). The LRT statistic could be exaggerated by, at least, five times in figures as opined by Yuan and Zhong (2008). It was also showed that in confirmatory factor analysis (CFA), by Yuan and Zhong (2008) and Yuan and Hayashi (2010), that about 3% of outliers could necessarily bias the estimates of factor loading by not less than 50% and increase the covariance estimates and the latent factor variance about 3–10 times, as oppose to about 3% of bad controlling observations which could yield even higher biases

12

on all parameter estimates. Yuan and Zhong (2013) again demonstrated mathematically with improved actual data sets that outliers produce worse fit indices; including but not limited to RMSEA and CFI, whereas potential controlling observations can lead to poor RMSEA value but better CFI for some cases.

According to Yuan and Bentler (2001) and Yuan and Zhong (2013), SEM based on normal-theory is not sturdy to outliers to the extent that a little presence outliers and potential controlling observations could bias both the model fits and parameter. However, robust modelling by substituting the normality assumption with an error term that follow a heavier-tailed t distribution which has since been developed in regression models and multilevel models (Pinheiro et al, 2001; Gelman & Hill, 2006). Moreover, using the t-based SEM and other robust SEM methods are preferred, as opposed to deleting outliers and controlling observations directly, since the complex nature of SEM makes it very challenging to apply common methods including Mahalanobis and Cook's distance to identify outliers and potential controlling observations (Flora et al, 2012; Sterba & Pek, 2012). Outliers and controlling observations are not mutually exclusive, notwithstanding the conceptual disparity, as some outliers can also exert strong influence on research results (Pek & MacCallum, 2011; Yuan & Zhang, 2012; Aguinis et al, 2013; O'Connell et al, 2015). They again assessed how the commonest fitness statistics perform base on simulations that were conducted regarding ML-Norm, ML-t, as well as using the Huber-type weights for a contaminated data. Further they examined how effective the criteria of information was in terms of the choice of either ML-Normal or ML-t for various conditions of misspecifying a model, a given sample sizes as well as the proportions of outliers and controlling observations.

Also, Pek and MacCallum (2011) opined in their research that there was an emphatic distinction between outliers and controlling observations under SEM which they further indicated that both terms have dissimilar effects estimated parameters as well as model fitness statistics. They affirmed that outliers are items or observation that lie afar from majority of the observations under consideration. For regression analysis which has just one independent variable then the outliers refer to the observations that depart largely from the predicted value considering the line of regression. However, for multivariate analysis including SEM, an observation length considered from the center of majority of the data points is often quantifiable using Mahalanobis distance. However, controlling items or observation is said to be those that have huge effect on the model fitness statistics as well as the estimated parameters. That is to say that the absence of the influential observation shows remarkable changes in terms of the estimated parameters and model fitness. Though there are conceptual variations between outliers and controlling observations, they sometimes may overlap in terms of their effect as some outliers can equally have huge effect on the results. Meanwhile, some of the techniques noted for the identification outliers as well as controlling observations under SEM consist of Cook's, Mahalanobis, and likelihood distances. Past researches (Aguinis et al, 2013; O'Connell et al, 2015; Pek & MacCallum, 2011; Yuan & Zhang, 2012) have succinctly discussed the tools as well as the procedures required for the identification of outliers and controlling observations.

Aguinis et al (2013) reviewed 232 varied methods on organizational science matters about outliers and controlling observations, and just five of them were related to SEM, in spite of the popularity of SEM in the recent times. The main possible reason is that practical guidelines on managing outliers and potential controlling observations were

evolved just recently (Pek & MacCallum, 2011; Aguinis et al, 2013). Notwithstanding the reported essence of outliers and controlling observations, detection and diagnostics of such observations were rarely performed and practice in real research, and in particular the use of SEM methods that are robust has been very scarce though SEM is notably made up of a measurement model(s) and a structural model (Bollen, 1989; Kline, 2011). Again, a plausible backdrop for current techniques regarding how outliers could be spotted and managed and by extension controlling variables demand for sophisticated software in SEM (Sterba & Pek, 2012; Yuan & Zhang, 2012), hence making the reseachers work more difficult, particularly when they are not in tune with the software. Asparouhov and Muthén (2015) modeled SEM concept by proposing mathematical equations, with linearity, and also defining what outliers and controlling data sets stands for based on earlier works done by Yuan and Zhong (2008, 2013). Meanwhile, t-distribution technique in SEM is preferred alongside like robust methodologies in SEM instead of basically taking off outliers and controlling observation. This often arises in SEM which normally present complex situations which does not allow it to utilize universal methods including Mahalanobis as well as Cook's distances in identifying outliers and controlling observations (Flora et al, 2012; Sterba & Pek, 2012).

Furthermore, SEM usually is broadly embodied by measurement models and manifest models (Bollen, 1989; Kline, 2011). The process of identifying outliers and controlling data is a normal routine in regression analysis but not much work has been done about the situation under SEM (Pek & MacCallum, 2011). Yuan and Bentler (2001) demonstrated in a mathematical sense the presence of outliers and how it can hugely impact the likelihood ratio test statistic to the extent that it can raise its value, at least, six times. Yuan and Zhong (2008) and Yuan and Hayashi (2010) have further indicated that

15

during CFA process, the factor scores loaded can be biased due to outliers which could constitute at least 3% of the estimates. Also, 232 varied methods reviewed from a chunk of published papers on issues bordering on outliers and controlling observation, just five of them were utilized in SEM though it has been widely used in many other areas in the last 20 years (Aguinis et al, 2013).

2.4 Residual Estimators in SEM

Varied methods utilized in SEM to estimate could be viewed based on covariance (such as ML) as well as component based (such as PLS, GSCA), or the frequentist approach (such as ML, PLS, GSCA) as well as the Bayesian method (such as MCMC). Methods such as the covariance based were developed for modelling, evaluating as well as validating. On the other hand, the component based methods were meant to achieve how to compute and predict (Tenenhaus, 2008). In simple sense, the main difference is that covariance based was designed for to test models whilst the component based methods were meant to provide succinct meaning to variances as well as predict (Hulland et al, 2010; Tenenhaus, 2008). Meanwhile the frequentist technique usually identifies values of parameters which are due to measured data whereas the Bayesian methods looks at estimate obtained from a parameter which are theoretical depictions of relations that rely on measured data. Again, adding to the varied reasons and dimensions of ML, PLS, GSCA, as well as MCMC usually varies in terms of how robust they appear due to different data scenarios. This is attributable, but not limited, to size of the sample, variables considered, misspecifying the model as well as the kind of measurement-manifest observation link.

16

Also, inference and deductions made from outcomes of modeling generally rely on the methods adopted and implemented in SEM. It remains though to point out whether hypothetical model normally presents correct information based on an application of a study or simulated study that has the capacity to shine light on the effect of misspecified parameter among methodologies of estimations (Asparouhov & Muthén, 2010; Hwang, et al, 2010). Moreover, the degree upon which parameters could be affected as a results of misspecifying a given model relies on the architectural makeup of the sample utilized (Henseler, 2010; Tanaka, 1987) and overall complexity of the model (Tanaka, 1987).

Whether or not the link amid measured and manifest observations are developing or contemplative in its form, is essential to the method of research since it is conceptually motivated in applied study. For SEM, manifest variables could be utilized as the basis for measured observations in modeling (Bollen & Lennox, 1991), and also for representing the unified figures of the measured values (Curtis & Jackson, 1962). It is a necessity in specifying SEM to mirror the right conceptual links, however the estimation methodologies most often differ in terms of how they perform based on the kind of association described. Developing indicator models were often deemed unsuitable for classical maximum likelihood method but for recently (Chin, 1998; Ringle et al, 2009). In contemporary times, it has been noted that ML was much possible to over-estimate a parameter under contemplative model, particularly where the sample size is small. Ringle et al (2009) opined contrary to the aforeassigned notion that PLS was could possibly under-estimate parameters under contemplative models. Meanwhile, owing to the amenable nature of GSCA to contain either developing or contemplative items is effectively on record, though the assertion widely relies on conceptually motivated

anticipations of the methodology without evidence from experimental studies (Hwang & Takane, 2004).

Many estimation methodologies as well as modifications of these methods have been researched upon and utilize in SEM, bordering on ML, plus ML which are robust standard errors, GLS and WLS (Muthén & Muthén, 1998-2010). Meanwhile, it is notable fact that these methodologies are not efficient when subjected to certain assumptions. For instance, ML as well as WLS basically fail to give definite parameters where the sample is not large (Hoogland & Boomsma, 1998; Hu et al, 1992; Olsson et al, 2000); the higher the degree of precision to produce an estimate under MLR the more generally it is restricted to estimates of standard errors rather than coefficient of the structural or measurement pathway. GLS is to a large extent unaffected by model misspecifications that may lead to overwhelming fitness (Olsson et al, 1999). By reacting to these hindrances as well as related estimation methodologies, more estimation methods have been utilized in estimating under SEM, such as PLS (Wold, 1975), standard structured component modeling (Hwang & Takane, 2004; Kline, 2011), as well as MCMC (Hastings, 1970).

According to Hoyle (2000) the commonest method of estimating parameters in SEM is maximum-likelihood. Studies on ML is across wide range of fields as well as data conditions and its challenges are on record. One of the conditions under which ML performs abysmally is when the sample is not large (Kline, 2011). Owing to this challenge, it is very essential that studies pay attention to the performance of other methods in terms of improving parameter estimates; partial least squares (PLS), generalized structural components analysis (GSCA) as well as Markov Chain Monte

Carlo (MCMC). should how weak or strong any different methodology in the presence of small sample study was well understood, researchers could have been adequately resourced to undertake decisions from an informed position regarding the selection of an exact method of estimating and making sense of results

Over the period advances has been made in the methods used in SEM. Even more pronounce are the different methodologies that have been established such as LS, WLS, PLS, GSCA as well as MCMC approaches. However, it is imperative to underscore the fact that these different methods are yet to be comprehended as their performance in terms of using real life data is normally challenging to predict (Henseler, 2012; Hwang et al, 2010 & Malhotra et al, 2010). Some estimation methods, besides what has been described earlier in this study, were developed for specific use in SEM whenever assumptions underpinning ML were violated, particularly robust ML and WLS (Henseler, 2012; Hwang et al, 2010; Malhotra et al, 2010). It is worth noting that it is almost impossible to compare and examine the performance of these different estimation methods in one study. Therefore, the current study will mainly focus on differential performance of the regression, Bartlett's, Anderson-Rubin and the EM methods to estimate residuals emanating from both manifest and construct variables in SEM.

CHAPTER THREE

STRUCTURAL EQUATION MODELING

3.1 Introduction

This chapter provides the basic concepts of SEM. The key main steps in SEM comprising model specification, identification, estimation and evaluation were also highlighted. This chapter also looks into details the concept of residual estimators under SEM. It also comprised the derivation of residuals and its diagnostics based on varied residual estimators. It again looks at how to estimate latent variables or factor scores and residuals which are not direct in terms of latent variables are unobserved. Further, the focus will be on utilizing individual cases in constructing the matrices as oppose to relying on covariance matrix of the measured variables.

3.2 Meaning of Structural Equation Modeling

The structural equation model (SEM) is a statistical methodology that adopts a confirmation approach (for example, hypothesis testing) for the analysis of a structural theory related to some phenomena. Typically, this theory represents 'causal' processes that generate observations in several variables (Bentler, 1988). The term structural equation model conveys two important aspects of the procedure: (a) the causal processes under consideration are represented by a series of structural equations (i.e. regression), and (b) that these structural relationships can be graphically modeled to allow a conceptualization clearer than the theory in question. The hypothesized model can be statistically tested in a simultaneous analysis of the entire system of variables to determine the extent to which it is consistent with the data. If the goodness of adaptation is adequate, the model supports the plausibility of the postulated relations between the variables; if it is inadequate, the tenacity of these reports is rejected.

20

Different aspects of SEM differentiate it from the previous generation of multivariate procedures. First of all, as noted above, it requires a confirmatory approach rather than an exploratory approach to data analysis (although some aspects of the latter can be addressed). Moreover, requiring that the scheme of inter-variable relations be specified a priori, SEM lends itself to the analysis of data for inferential purposes. In contrast, most other multivariate procedures are essentially descriptive (for example, exploratory factor analysis), so hypothesis testing is difficult, if not impossible. Second, while traditional multivariable procedures are not able to assess or correct the measurement error, the SEM provides explicit estimates of these error variance parameters. Indeed, alternative methods (for example those rooted in regression or the general linear model) assume that errors in explanatory (i.e. independent) variables disappear. Therefore, applying these methods when there is an error in the explanatory variables is equivalent to ignoring the error, which can lead, in the final analysis, to serious inaccuracies, especially when the errors are considerable. Such errors are avoided when the corresponding SEM analysis is used (in general terms). Third, although data analyzes using the above methods are based only on observed measurements, those using SEM procedures may incorporate observed (non-latent) and unobserved variables. Finally, there are no alternative methods of wide and simple application for the modeling of multivariate relationships or for the estimation of indirect effects and/or intervals; these important features are available using the SEM methodology (Byrne, 2003).

For these heavily desirable characteristics, SEM has become a famous method in terms of non-experimental studies, in which methodology for purposes of testing theoretical related issues are not properly developed and considerations of ethics make experimental design not feasible (Bentler, 1980). The modeling of structural equations can be used

21

very effectively to address numerous research problems involving non-experimental research (Arbuckle, 2007).

3.2.1 Basic Concepts

Latent versus observed variables

In behavioral science, researchers are often interested in studying theoretical constructions that cannot be directly observed. These abstract phenomena are called latent variables or factors. Examples of latent variables in psychology are self-concept and motivation; in sociology impotence and anomie; in education, verbal ability and teacher expectancy; and in economics, capitalism and the social class. Consequently, as the latent variables are not directly observed, they cannot be measured directly. Therefore, the researcher must be defined operationally to represent the latent variables of interest in terms of behavior. As such, the unobserved variable is linked to an observable variable that allows it to be measured. Performance assessment, then the direct measurement of an observed variable, albeit indirect measurement of an unobserved variable (i.e. the underlying construct) (Byrne, 2010).

It is important to note that the term behavior is used here in the broadest sense, in particular to record notes on a meter. Therefore, the observation may include, for example, answers to self-report a scale of attitude scores on a performance test, observation scores in vivo representing a specific task or physical activity, encoded interview questions and answers, similar (Byrne, 2010). These measured values (i.e. measurements) are referred to as observed or open variables. In the context of the SEM methodology, they serve as indicators of the underlying construct they are intended to represent. Given the need to bridge observed and unobserved latent variables,

methodologists should now explain why researchers recommend that researchers be careful when choosing assessment measures. Although the selection of psychometric tools is important for the credibility of all the study results, this selection becomes even more critical when it is assumed that the observed measure represents an underlying construct (Byrne, 2010).

Exogenous versus endogenous latent variables

When working with SEM models, it is helpful to distinguish between exogenous latent variables and endogenous variables. Exogenous latent variables are synonyms for independent variables. They 'cause' variations in the values of other model latent variables. The variations in the values of the exogenous variables are not explained by the model. Rather, it is believed that they are influenced by other factors outside the model. The underlying variables such as gender, age and socioeconomic status are examples of such external factors. The endogenous latent variables are synonyms of dependent variables and, as such, are directly or indirectly influenced by the exogenous variables of the model. The variability of the values of endogenous variables is explained by the model, since all the latent variables that affect them are included in the specifications of the model (Byrne, 2010).

The factor analytic model

The oldest and best known statistical method for studying the relationships between the quantities of observed and latent variables is that of factor analysis. Using this data analysis approach, the researcher investigates the covariation between a set of observed variables to gather information about their underlying latent constructs (ie factors). There are two types of factor analysis: exploratory factor analysis (EFA) and confirmatory

factor analysis (CFA). We now turn to a brief description of each. Exploratory factor analysis (EFA) has been developed for situations where the links between observed and latent variables are unknown or uncertain (Byrne, 2010).

The analysis therefore proceeds as exploratory to determine how and to what extent the observed variables are related to their underlying factors. In general, the researcher wants to identify the minimum number of factors that underlie (or explain) covariation among the observed variables. Suppose a researcher develops a new tool to measure five facets of the physical self-concept (e.g, health, fair play, physical appearance, coordination, and physical strength).

After formulating the questionnaire elements designed to measure these five latent constructs, it would perform an EFA to determine to what extent the element's measurements (the observed variables) were related to the five latent constructs. In factor analysis, these relationships are represented by factor loads. The researcher hopes that the elements designed to measure health, for example, have high loads for this factor and low or negligible loads for the other four factors. This factorial analytical approach is considered exploratory as the researcher has no prior knowledge that the elements actually measure the factors sought (Comrey, 1992; Gorsuch, 1983; McDonald, 1985; Mulaik, 1972; Byrne, 2005a; Fabrigar et al, 1999; MacCallum et al, 1999; Preacher & MacCallum, 2003; Wood, Tataryn & Gorsuch, 1996).

Unlike EFA, confirmatory factor analysis (CFA) is used appropriately when the researcher knows something about the underlying structure of the latent variables. On the basis of theoretical knowledge, empirical research or both, it postulates a priori the

relationship between the observed measures and the underlying factors, and thus statistically verifies this hypothetical structure. For example, using the example above, the researcher would support the loading of articles designed to measure self-understanding of sports skills based on this specific factor and not on health, physical appearance, coordination or the size of the concept, even the physical force. As a result, the prior specification of the CFA model would allow all elements of sportsmanship who understand each other to self-understand to play freely on this factor, but limiting themselves to factors remaining at zero. The model would then be evaluated by statistical means to determine the appropriateness of its fit to the sample data (Bollen, 1989a; Byrne, 2003, 2005b; Long, 1983a.)

In summary, the factor analysis model (EFA or CFA) focuses only on how the observed variables are reported underlying latent factors and to what extent. More specifically, it is determined to what extent the observed variables are generated by the underlying latent constructions. The strength of the regression pathways of the factors with respect to the observed variables (factor loadings) is therefore of primary interest. Although inter-factorial relations are also of interest, no regression structure between them is taken into account in the factor analysis model. Since the CFA model focuses exclusively on the connection between factors and measured variables, it is in the context of SEM, which has been called a measurement model (Byrne, 2010).

The full latent variable model

Unlike the analytical factorial model, latent complete variable model (LV) to specify the regression structure between latent variables. In other words, the researcher can make a hypothesis on the impact of latent construction on the modeling of causal direction. This

model is called complete (or total) because it includes both a measurement model and a structural model: the measurement model that describes the relationship between latent variables and measured measurements (CFA model), and the structural model describes the connections among the latent variables themselves.

A complete LV model that specifies the direction of the cause of a single direction is called a recursive model; what allows reciprocal effects or feedback is called a non-recursive model. Only the applications of the recursive model are considered in this study (Byrne, 2010).

3.3 Fitting the Structural Equation Model

There are often four principal steps in terms of fitting SEM. These include specifying the model, identifying the model, estimating the parameters and subsequently evaluating the model. The section here also looked at diagnostics that available to examine a given model under SEM. In the first step, under the model specification, the notations utilized under SEM would be introduced. For the second step, the model is identified by a theoretical examination to ascertain model identification and the possibility of estimation. In the third step, estimating the model concerns the application of estimation methods such maximum likelihood and the assumptions, which are normally used under the procedure, which underpins the distribution. The final step, which borders on evaluating the model, concerns the assessment of the model fitness and perhaps utilizing model diagnostics to examine the departure from assumptions, particularly in the midst of outliers and controlling observations (Byrne, 2010).

In the SEM concept, models are arrived at such that it specifies the associations between measured and manifest variables which often results in a system of linear equations so that the relationships among remain linear or can be transformed into linear. The variables in these linear equations are related by structural parameters which can be denoted as θ. As a result, the covariance matrix of the population measured variables denoted as Σ symbolized the function of the parameters, θ, contained in the model.

The basic hypothesis of SEM was defined by Bollen (1989) as

$$\Sigma = \Sigma(\theta)$$

where Σ connotes the covariance matrix of the populations observed variables, θ depicts vector of parameters of the model, and $\Sigma(\theta)$ represents the covariance matrix as a function of θ.

3.3.1 Model Specification

As in indicated section 3.3, the starting point of every SEM is the specification of the model the researcher is interested in using as a result of the relationship emanating from theory which needs examination. The hypothesized link created among the variables such as linear equations comprising factor (or latent), measured and the error terms are achieved through structural parameters. The structural parameters then summarizes the associations between the variables and thus defines the causal relationships among factor variables, measured variables and the factor and measured variables. These system of structural equations comprise only two main sub-systems; the factor variable model, which links factor variables whereas the measurement model is the one that links the indicator variables. The notation, which was developed first by Joreskog (1973, 1977), Wiley (1973), and Keesling (1972) and cited later by Bollen (1989), of our SEM is introduced under this section. The notation is usually known as LISREL, which stands

for Linear Structural RELationships, model notation because of the software that made it popular.

The path diagram shown in Figure 3.1 was adopted, from Hildreth (2013), throughout this study. A recursive model containing a mediation component was utilized to facilitate the application of various residual estimators to enable to estimation of both the observed and the manifest variables. As often is the case for all path diagrams, the squares as well as the rectangles represents the indicator or measured variables whiles the circular or ovals represents the factor variables and the error terms. Again, the single arrow head represents a causal link between variables where the variable at the tail is deemed to be responsible for the cause of the variable on the arrow head. Also, a double arrow head represents a reciprocated link among any two variables and a curved double arrow head denotes relationships among any two variables which are not analyzed.
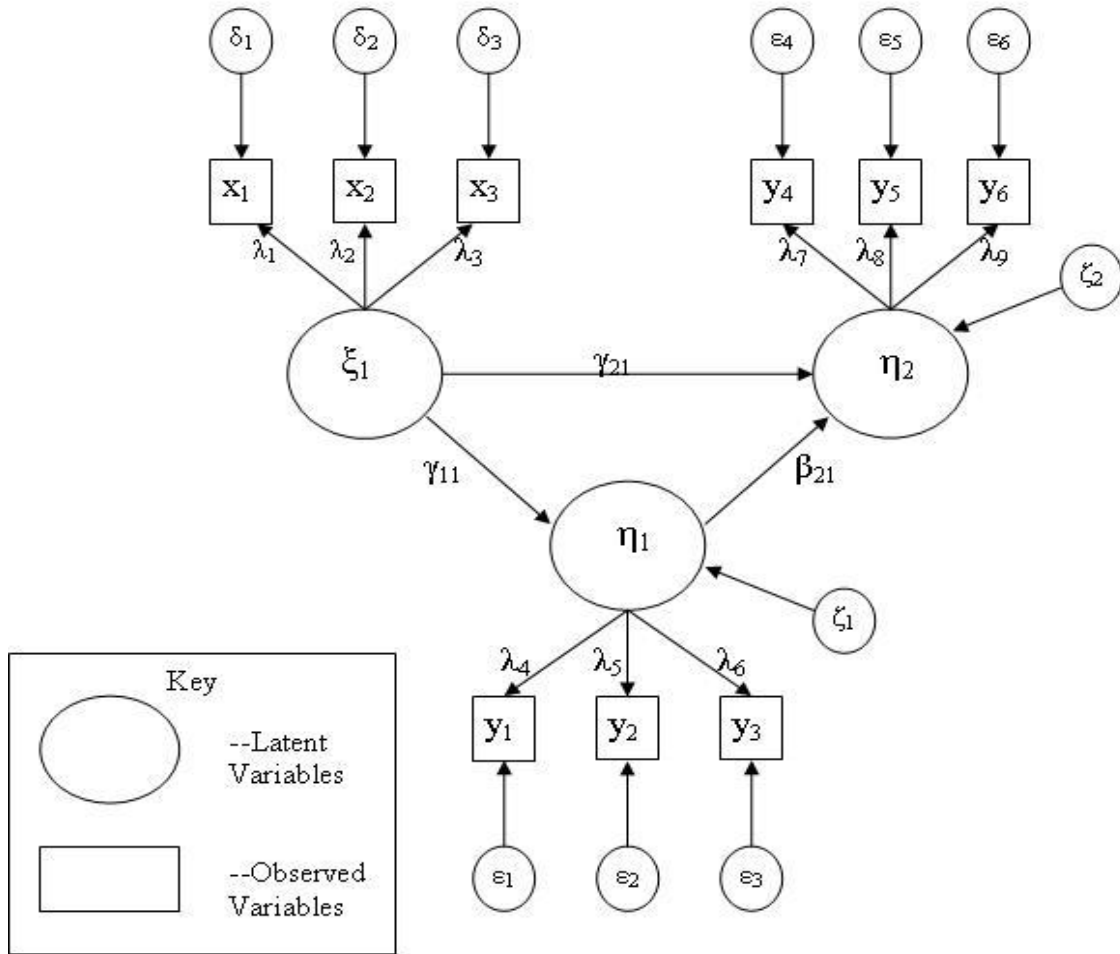
Figure 3.1: Adopted Hypothesized Recursive Structural Equation Model

    (Hildreth, 2013)

The factor variable model, which is also usually known as the structural or causal model, is made up of a system of linear equations that defines the relationships among factor variables. From Figure 3.1 above, consider a SEM framework. The factor (or latent) variables are taken to be either be exogenous, such as ξ_1, since their causes is derived from outside the model, or endogenous, such as η_1 and η_2, since their causes is from within the model. In Figure 3.1, it was hypothesized that ξ_1 is the cause of both η_1 and η_2. Also η_1 is the cause of η_2. The factor (or latent) variable of the hypothesized model in Figure 3.1 can be shown as follows:

$$\eta_1 = \gamma_{11}\xi_1 + \zeta_1 \qquad\qquad (3.1)$$

$$\eta_2 = \beta_{21}\eta_1 + \gamma_{21}\xi_2 + \zeta_2 \qquad\qquad (3.2)$$

29

where Equation (3.1) as well as Equation (3.2) are both linear in terms of the variables and parameters. Regarding the equations expressed in this chapter, such as Equation (3.1) and Equation (3.2), it is worth noting that these at the individual or observational stage. Despite the fact this is generally demonstrated by using the subscript i for every random variable, it is left out under this chapter for purposes of convenience. Therefore, the sizes regarding all matrices as well as vectors presented under the chapter here are equally representing the ith observation or individual.

Meanwhile, none of the constant terms are part since it is supposed (or assumed) that all variables have departed from their means indicating that $E(\eta_1) = 0$, $E(\eta_2) = 0$ and $E(\xi_3) = 0$. This assumption mainly helps in simplifying the algebraic maneuvering and does not influence either the general analysis or generalization. The structural parameter γ_{11} represents the expected change in η_1 associated with a one unit increase in ξ_1. The structural parameters γ_{21} and β_{21} have analogous interpretations such that γ_{21} represents the expected change in η_2 associated with a one unit change in ξ_1 holding η_1 constant and β_{21} represents the expected change in η_2 associated with a one unit increase in η_1 holding ξ_1 constant. In this example, γ_{21} represents the direct effect of ξ_1 on η_2; similarly, the structural parameter β_{21} represents the direct effect of η_1 on η_2. In this model, it is hypothesized that ξ_1 not only has a direct effect on η_2 but also an indirect effect that is mediated by η_1.

This indirect effect is then equal to $\gamma_{11}\beta_{21}$ leading to ξ_1 having a total effect on η_2 equal

to the sum of the direct and indirect effects, $\gamma_{21} + \gamma_{11}\beta_{21}$. The random errors ζ_1 and ζ_2

are assumed to have an expected value of zero and homoscedastic variances as well as

be independent (they are not autocorrelated), and uncorrelated with ξ_1.

Equation (3.1) and (3.2) can be rewritten in matrix notation as

$$\begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ \beta_{21} & 0 \end{bmatrix}\begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} + \begin{bmatrix} \gamma_{11} \\ \gamma_{21} \end{bmatrix}[\xi_1] + \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix}$$

Which can be written more compactly as

$$\eta = B\eta + \Gamma\xi + \zeta \qquad (3.3)$$

where η represents an $m \times 1$ vector of endogenous construct variables, ξ is $n \times 1$ vector

representing the exogenous latent variables. It is supposes that when every $k = 1, ..., m$:

then $E(\zeta_k) = 0$; ζ_k are homoscedastic; ζ_k are independent whiles η and ζ are

uncorrelated.

The parameters of the structural part summarizes the associations amidst the construct

variables are contained in the $m \times m$ matrix B as well as the $n \times n$ matrix Γ. For B, we

supposed that: $(I - B)$ for I represents the identity matrix and it is nonsingular; and the

main diagonal of B comprise zeros. The Γ matrix contains the structural parameters that

link the exogenous latent variables to the endogenous latent variables. This matrix

consists of elements γ_{kl} where k denotes the row position and l denotes the column

position. The element γ_{kl} represents the expected direct change in η_k associated with a

one unit increase in ξ_l where $l = 1, ..., n$; ξ_l may also cause a change in η_k indirectly

31

via other latent variables in η which are then calculated using elements in Γ and B (Bollen, 1989).

For the covariance matrices, we supposed Ψ and Φ are symmetric while generally assuming that these are invertible. Often there is no particular matrix which is provided for covariance of the endogenous construct variables, represented by $\Sigma_{\eta\eta}$, since the matrix is defined as a function of B, Γ, Ψ, and Φ. Algebraic simplification will indicate that $\Sigma_{\eta\eta} = (I - B)^{-1}(\Gamma\Phi\Gamma' + \Psi)(I - B)^{-T}$.

3.3.2 Measurement Model

Despite the fact that the construct variables model reduces the associations theoretically amidst the construct variables that a study has hypothesized, these associations are only examined when measure of the construct variables are gathered so that the measured variables become proxies of the construct variables. Thus the manifest model relates the construct variables to the measured variables.

For instance, Figure 3.1 indicates that every construct variable has three observed variables, which is linked to just a factor. The indicators for η_1 are y_1, y_2 and y_3, the indicators for η_2 are y_4, y_5 and y_6, and the indicators for ξ_1 are x_1, x_2 and x_3.

The manifest model related to Figure 3.1 is given by

$$
\begin{aligned}
x_1 &= \lambda_1\xi_1 + \delta_1 & y_1 &= \lambda_1\eta_1 + \in_1 & y_4 &= \lambda_4\eta_4 + \in_4 \\
x_2 &= \lambda_2\xi_2 + \delta_2 & y_2 &= \lambda_2\eta_2 + \in_2 & y_5 &= \lambda_5\eta_5 + \in_5 \\
x_3 &= \lambda_3\xi_3 + \delta_3 & y_3 &= \lambda_3\eta_3 + \in_3 & y_6 &= \lambda_6\eta_6 + \in_6
\end{aligned}
\tag{3.4}
$$

The Equations contained under (3.4) can be written more compactly in matrix notation as:

$$x = \Lambda_x \xi + \delta \tag{3.5}$$

$$y = \Lambda_y \eta + \in \tag{3.6}$$

for $\quad x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad \Lambda_x = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix}, \quad \xi = \xi_1, \quad \delta = \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{bmatrix}, \quad \Lambda_y = \begin{bmatrix} \lambda_4 & 0 \\ \lambda_5 & 0 \\ \lambda_6 & 0 \\ 0 & \lambda_7 \\ 0 & \lambda_8 \\ 0 & \lambda_9 \end{bmatrix}, \quad \eta = \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} \quad$ and

$\in = \begin{bmatrix} \in_1 \\ \in_2 \\ \in_3 \\ \in_4 \\ \in_5 \\ \in_6 \end{bmatrix}$

diagonals of the matrices θ_δ as well as θ_ε comprise error variances of the manifests x_h and y_j, respectively. The off-diagonal items are the covariances of the mamifest errors of indicators x_h and x_j in θ_δ and y_h and y_j in θ_ε. Usually the off-diagonal items are set to zero, but this should not be the case. Manifest errors related to two indicators could be associated for varied reasons such as being collected at the same point in time or by using the same question at different points in time such that there is a common source of variability for these items (Bollen, 1989).

The manifest model in this study utilized Bollen (1989) notations as given below.

$$x = \Lambda_x \xi + \delta$$

$$y = \Lambda_y \eta + \varepsilon$$

Assumptions made regarding the model are as follows.

1. $E(\eta) = 0, E(\xi) = 0, E(\delta) = 0, E(\varepsilon) = 0$

2. ε not associated with η, ξ, and δ

3. δ not associated with η, ξ, and ε

3.4 Identification

After the specification of a model, the next phase is to find out whether or not it is identifiable. Thus, if there is the theoretical possibility of computing derivations of unique parameter estimates. Identifiability can be demonstrated when every unknown parameter under a given SEM are functions of known parameters while theses function eventually leads to unique solutions (Bollen, 1989). Whenever this achieved then the unknown parameters are identified else they are deemed to be unidentified. For practical purposes, there exist known necessary as well as sufficient condition for identifiability which is heighted in the subsequent sections (Byrne, 2010).

3.5 Necessary Conditions

The initial two rules known conditions that are necessary regarding identifiability but not sufficient (required for the possibility of identifiability but on their own are enough to ensure model identifiability). The two rules must hold in order to determine whether a model satisfies the sufficient conditions of model identification. Bollen (1989) and Kline (2011) has a lot more explanation regarding these rules.

3.5.1 Scaling the Latent Variable

The key rule under SEM for necessary condition is referred to as scaling the construct variable. The main problem that emanates from construct variables is the fact that there

is no establish metric related to the construct variables. Since these kind of variables are not observed their metric is equally not observed and thus not known (Kline, 2011). One of the ways to solve the problem is to set one of the paths, λ_g, regarding every construct variable under η and ξ to a unique element under y as well as x to 1. This process is referred to as scaling the construct variable and it is necessary for every construct variable in a SEM. By this, the scaling of construct variables under η and ξ is then the measurement scale of the unique items in y as well as x_i. For our model in Figure 3.1, the metric is set through $\lambda_1 = \lambda_4 = \lambda_7 = 1$ though it could be achieved through different indicators as measurement scale. Metric of the construct variable could also be achieved by standardizing the variance of every construct variable under η and ξ to be equal to 1. By this approach, standardizing the variance for our model in Figure 3.1 is achieved by letting $\phi_{11} = \psi_{11} = \psi_{22} = 1$.

3.5.2 t Rule

Under this necessary condition, the quantity of specific items under Σ, represented as t, must either be less than or equals the number of unknown parameters contained in θ, $\frac{p(p+1)}{2}$. For our model in Figure 3.1, $t = 21$ (i.e. the parameters which are not known comprise; 1 element under B, 2 elements in Γ, 1 element under Φ, 2 elements in Ψ, 2 elements under Λ_x, 4 elements in Λ_y, 3 elements under θ_δ, and 6 elements in θ_ε) while $\frac{p(p+1)}{2} = \frac{9(10)}{2} = 45$. As $t = 21 < \frac{p(p+1)}{2} = 45$, the condition is satisfied (Kline, 2011).

3.6 Sufficient Conditions

Despite the fact that fixing the metric and meeting the t-rule are necessary conditions for any SEM, they do not ensure identifiability of a model. Thus this brought about more

identification rules. For SEM, the sufficient condition is known as two-step rule. This rule comprises numerous varied rules which depends on the subsystem as well as the structure of the said model (Byrne, 2010).

3.6.1 Two-Step Rule

Under the two-step rule, it is indicated that when the manifest model as well as the construct variable model are each identified then the entire model is deemed to be identified. The intial step assesses the construct model and handles it as if it were a path analysis. That is to assume that the construct variables measured variables. Many rules, under this step, need to be utilized to confirm the identifiability as opined by Bollen (1989) and Kline (2011). For the second step, the manifest model is assessed as if it were a CFA. According to Bollen (1989), Kline (2011), as well as Brown (2006) there are many rules, under this step, that can be utilized to confirm whether the model is identifiable.

Meanwhile the identification rule regarding the construct variable model identifiability arises if very unknown parameter under B, Γ, Φ, and Ψ is expressed as a function of either one or more items in the covariance matrix Σ of the population. Similarly, to identify a manifest model then every parameter under Λ_x, Λ_y, θ_δ, and θ_ε should be expressed as a function of the items under Σ. The determination of this could be tedious which therefore give rise to the introduction of many rules, all of which assume that the construct variables have measurement scale as well as other constraints (for instance, constraining particular items under θ_δ and θ_ε to be equal to zero) has been formalized (Byrne, 2010). Moreover, the CFA models are either standard (where every indicator can only load on a factor while θ_δ as well as θ_ε are diagonal matrices) or non-standard (for

36

every indicator loads on many factors while either $\theta_\delta \ or \ \theta_\varepsilon$ or even both are not diagonal matrices).

3.7 Empirical Identification

The rules mentioned in the preceding sections are key for ascertaining whether a given SEM is identifiable on theory basis. However, these rules do not ensure that a SEM is identifiable empirically. Many issues which are considered crucial can result in empirical under-identification. This situation, at the data stage, may arise whenever two variables exhibit strong collinearity which can, to a very large extent, lessen the quantity of elements below $p(p+1)/2$. Again, problems may emanate, at the model stage, if the coefficients of the various paths basically zero which effectively removes such paths from the model. This situation can cause a factor to have very few indicators to be identified. For instance, a CFA which contains 3 indicators with 1 factor where one path equals zero then the said model cannot be identified since would have been left basically with 2 indicators. Meanwhile, other problems may cause empirical under-identification particularly if the assumptions underpinning the model are either violated or the said model is misspecified (Kline, 2011).

3.8 Estimation Method and Distributional Assumptions

After the specification and theoretical identification of a structural equation model, the next stage is the estimation of the parameters of the model. The foundation underpinning the estimation methods under SEM is introduced under this section which focuses on maximum likelihood estimation procedure. The basic assumptions regarding the distribution of normality for maximum likelihood are together with the properties of maximum likelihood estimators are looked at here (Kline, 2011).

Under SEM, every method of estimation are based on the association between the implied covariance matrix of the measure variables, (θ), as well as the sample covariance matrix of the measured variables S. The main aim for every method of estimation is utilizing the sample covariance matrix, S, to achieve estimates for the structural parameters in θ for the given model so that $\Sigma(\theta)$ is close to S. In order to assess what "close" stands for, then a fitting function must be selected which is minimized. Several possibilities are available in terms of choosing for this fitting function, which is represented as $(S, \Sigma(\theta))$, is a function of the sample covariance matrix as well as the implied covariance matrix. For practical purposes, there exist four key properties that a fitting function must possess (Bollen, 1989 and Browne, 1984). These are

1. $F(S, \Sigma(\theta))$ is scalar

2. $F(S, \Sigma(\theta)) \geq 0$

3. $F(S, \Sigma(\theta)) = 0$ whenever $S = \Sigma(\theta)$

4. $F(S, \Sigma(\theta))$ is a twice continuously differentiable function in terms of S as well as $\Sigma(\theta)$.

The commonest method of estimation is ML which demand that assumptions should be made concerning the construct (latent) variables as well as the error terms. The commonest assumption is that of normality which indicates that $\sim \mathcal{N}(0, \Phi)$, $\eta \sim \mathcal{N}(0, \Sigma_{\eta\eta})$, $\delta \sim \mathcal{N}(0, \theta_\delta)$, $\varepsilon \sim \mathcal{N}(0, \theta_\varepsilon)$, as well as $\zeta \sim \mathcal{N}(0, \Psi)$. Therefore, this means that $x = \Lambda_x \eta + \delta$ as well as $y = \Lambda_y \eta + \varepsilon$ are deemed to be normally distributed so that $x \sim \mathcal{N}(0, \Sigma_{xx})$ for $\Sigma_{xx} = \Lambda_x \Phi \Lambda'_x + \theta_\delta$ and $y \sim \mathcal{N}(0, \Sigma_{yy})$

for $\Sigma_{yy} = \Lambda_y (I - B)^{-1} (\Gamma \Phi \Gamma' + \Psi)(I - B)^{-T} \Lambda'_y + \theta_\varepsilon$. Based on the assumption of normality the MLfor the fitting function is given by Bollen, (1989):

$$F_{ML} = \log |\Sigma(\theta)| + Tr(S\Sigma^{-1}(\theta)) - \log |S| - (p + q), \qquad (3.7)$$

while log connotes the natural log and Tr is the trace of a matrix. Thus Equation (3.7) satisfies the properties for fitting function base on the assumption that $\Sigma(\theta)$ as well as S are matrices which are positive definite. The fitting function is then minimized by the ML in respect of the structural parameters often by an iterative numerical procedure, usually the Newton-Raphson algorithm (Bollen, 1989).

According to Casella and Berger (2002), suppose $X_1, X_2, \ldots$, be iid $f(x|\theta)$, when $\hat{\theta}$ is the maximum likelihood estimator of θ and let $\tau(\theta)$ be the continuous function of θ. According to Cox and Hinkley (1974), every member regarding the exponential dispersion family (normal distribution included) satisfies the regularity conditions and therefore below holds

$$\sqrt{n}[\tau(\hat{\theta}) - \tau(\theta)] \to \mathcal{N}(0, v(\theta))$$

for $v(\theta)$ is a Cramér-Rao lower bound. By this theory, the properties of the MLEs, $\hat{\theta}$, are given as (Casella and Berger, 2002 and Greene, 2008).

1. ML estimators are asymptotically unbiased so that

   $E[\tau(\hat{\theta}) - \tau(\theta)] = 0$ as $n \to \infty$.

2. ML estimators are said to be consistent. This means that when $\varepsilon > 0$,

   $P(|\tau(\hat{\theta}) - \tau(\theta)| \geq \varepsilon) = 0$ as $n \to \infty$.

3. MLEs are asymptotically efficient. MLEs are asymptotically normally distributed with mean θ and variance $(\frac{2}{(N-1)})(E[\frac{\partial^2 F_{ML}}{\partial\theta\partial\theta}])^{-1}$.

3.9 Model Evaluation

The next stage after estimating a SEM is the evaluation of the model to ascertain whether it fits the data set. A model can be evaluated at two levels. For the first part, the coefficients must be assessed whereas the second part deals with the general model fit. Here, the section looks at the tools required at every part (Hildreth, 2013).

3.9.1 Coefficient Evaluation

Irrespective of the statistical method under consideration, every model evaluation demands an examination of the estimated coefficients regarding the sign as well as the magnitude in comparison to the theoretical underpinning of the model. This particular phase, regrettably, is occasionally ignored since researchers may rely on only the general fit indices to arrive at decision regarding the fitness of the model while ignoring the theoretical interest the model embodies. Meanwhile every SEM is formulated to check a particular theory which demands a proper consideration of the values associated with estimated coefficients. It is very important to assess the magnitude as well as the sign of the coefficient in line with past study and the significance of coefficients both substantively and statistically (Bollen, 1989).

3.9.2 Overall Model Fit Measures

For evaluation of SEM, quite a number of statistical measures have so far been formulated. Therefore there are measures to assess the validity of the hypothesis that $\Sigma = \Sigma(\theta)$ through the measurement of the length or distance between Σ and (θ). For practical purposes, however, Σ and $\Sigma(\theta)$ are substituted with their sample counterparts S as well as $\Sigma(\hat{\theta})$ where S represents the covariance matrix of the sample and $\Sigma(\hat{\theta})$ represents the implied covariance matrix which is evaluated at the estimate $\hat{\theta}$ that reduces

the fitting function. Measures regarding fitness thus becomes functions of both S and $\Sigma(\hat{\theta})$ which are formulated ascertain the extent of "closeness" of S to $(\hat{\theta})$. These indices have the advantage of examining the whole model and may be able to reveal model inadequacies not revealed by coefficient estimates (Bollen, 1989; Kline, 2011).

Residual Matrix: The direct way of examining the hypothesis that $\Sigma = \Sigma(\theta)$ is by calculating $\Sigma - \Sigma(\theta)$. Whenever the hypothesis holds then $\Sigma - \Sigma(\theta)$ represents a null matrix so that nonzero items show the model specification error. For practical purposes, S and $\Sigma(\hat{\theta})$ are replaced with both unknown population matrices Σ as well as $\Sigma(\theta)$ respectively in order to establish the sample residual matrix $S - \Sigma(\hat{\theta})$. The sample residual matrix $S - \Sigma(\hat{\theta})$ comprise of individual items, $s_{hj} - \sigma_{hj}$, where s_{hj} represents the sample covariance between the hth and jth variables as well as σ_{hj} which represents the predicted covariance of the model between the hth and jth observations. In order to correct for metric variations as well as sampling error, it was opined by Jöreskog and Sörbom (1986) that the utilization of a normalized residual matrix for individual items equals

$$\frac{s_{hj} - \hat{\sigma}}{[(\hat{\sigma}\hat{\sigma} + \hat{\sigma})/N]^{1/2}}$$

where s_{hj} represents the covariance of the sample between variables h and j, while $\hat{\sigma}$ represents the predicted covariance of the model between variables h and j, while $\hat{\sigma}$ represents the model predicted variance of variable h and $\hat{\sigma}$ being the model predicted variance of variable j.

41

There have been two more fit statistics which are not either absolute statistics or increment statistics comprise the Akaike information criterion (AIC-Akaike, 1974) as well as the Bayesian information criterion or Schwartz-Bayes criterion (BIC or SBC-Schwartz, 1978) that are utilized under SEM just as applied under regression modelling.

3.9.3 Diagnostics in SEM

As pertains in other statistical methodologies, many studies have developed model diagnostic checks for model fitness statistics as well as assessing the departures from certain underlying assumptions in SEM context. Some of these assumption have been discussed extensively previous under Section (2.3). Thus, this section summarizes the works done regarding the influence of and methodologies to detect: outliers as well as controlling elements; departures from the normality assumption; departure from the assumption of linearity; departures from the assumption of homoskedasticity; and the violations of the assumption of independence. Against this backdrop many SEM diagnostics were established to assist in terms of identifying outliers and controlling elements such as: the distance measures of Bollen's A (Bollen, 1987, 1989; Mullen, Milne, 2007); the measurements that rely on the effect of an item on the likelihood (Coffman & Millsap, 2006); the measures that rely on the effect of an item regarding the measured covariance matrix; the case deletion process (Lee & Tang, 2004) as well as the forward search algorithm to spot many items would be identified by deletion diagnostics (Mavridis & Moustaki, 2008).

3.10 Methods Used in Developing Residual Estimators

Studies regarding residuals in SEM usually refer to residual matrix which is defined as the difference between the observed and the predicted covariance matrices, given by

42

$(S - \Sigma(\theta))$. Unlike the definition that considers the residuals linked to the individual cases that form the matrices. More importantly, the utilization of residuals in SEM, being the case for other statistical methodologies, would enable us to identify outliers and possible influential observations. The SEM framework normally considers three different cases of residuals; the residuals related to the observed error (measurement error) for those associated with the exogenous and endogenous factor scores (latent variables) δ and ϵ respectively as well as the residuals related to the error under the latent variable equations, ζ. Therefore, many earlier studies have established different methodologies for residual estimators. All the methods often utilize factor score estimators at different levels based on the methodology adopted. Generalization of the methods have been adopted for every case in terms of diagnosis (DiStefano, Zhu & Mindrila, 2009).

3.10.1 Weighted Function of the Observed Variable

The weighted function of the observed variable is the commonest factor score is often used in the derivation process Lawley and Maxwell (1971). The method consist of obtaining factor scores by taking the product of the observed variable matrix and the weight matrix denoted by W. This process then leads to the transformation of the observed variable into factor scores. The commonest choice of W is underpinned by the principle least squares which is known as the regression method estimator. This approach was utilized by Bollen and Arminger (1991) and Sánchez et al (2009). Bollen and Arminger (1991) developed another option for W, not common though, which was subsequently formalized by Raykov and Penev (2001) through the application of principle of weighted least squares which is known as the Bartlett's method estimator. However, the third choice of W, which is known as the Anderson Rubin method estimator (Anderson & Rubin, 1956) was an extension of Bartlett's method based on

assumption of orthogonality of the factor model. The derivations of the weight matrices are discussed in subsequent sections under this chapter.

3.10.2 Bayesian Expected a Posteriori Scores

Bartholomew (1980 and 1981) developed the Bayesian expected a posteriori scores (EAP) method which is also known as posterior means. This was subsequently adopted in obtaining values for residuals Mavridis and Moustaki (2008). Standard Bayesian principles in terms of statistics, as the main aim of EAP, is often used to obtain the posterior distribution of the factor scores. More importantly, the observed variables under the vectors x and y, which is generally known as z, is fixed. The values obtained are subsequently utilized as the prior distribution for the latent variables as Baye's theorem is used to calculate the posterior distribution. A measure of location from the distribution, in this case the posterior mean $E(L|z)$, would then be selected and adopted for the factor scores. Thurstone (1935) developed the ultimate formula for EAP scores based on the assumptions that the measured variables were normally distributed which happens to be a unique case for the regression method. However, the derivations of their theoretical framework differ significantly though the EAP scores are a unique case of the regression method scores.

3.10.3 Empirical Bayes Estimates

This method is a look alike of the EAP scores which is also known as posterior modes. Bartholomew (1984) initially proposed this method and Meredith and Tisak (1990) later developed it using latent growth curve technique, a unique case of SEM. Subsequently, Coffman and Millsap (2006) further developed separate fit statistics. The difference between empirical Bayes estimation and EAP scores comes from the selection of the

location measure of the posterior mode as against the posterior mean of the posterior distribution which utilizes factor scores. However, whenever there exists symmetry in the posterior distribution, that is satisfying the normal assumption, then both empirical Bayes estimates as well as EAP scores will be identical since the mean and mode under such condition are the same (Mavridis & Moustaki, 2008).

3.10.4 EM Algorithm

This method, which is the last, provides for the calculation of the residuals by utilizing EM-type algorithm under nonlinear SEM which was proposed by Lee and Zhu (2002). According to Lee and Zhu (2002) the latent variables are used as missing observation and estimation of the latent variables are then obtained by utilizing the EM-type. Lee and Lu (2003) utilized the EM method to propose a general form of Cook's distance measure for nonlinear SEM by laying emphasis on checking the normality assumption of residuals in SEM. The study here utilizes the uniform horizontal QQ plots which are constructed to spot outliers and controlling observations using a simulation setup.

3.11 Residual Estimators and Residuals for Weighted Function of the Observed Variables

Residual estimators detailed further in this chapter were constructed by using the weight functions observed previously in section 3.2.1. reasons that informed the selection of this particular method include the fact that; it was already the commonest utilized in factor analysis to develop factor scores, it is relatively easier to work with in practice, both EAP scores and empirical Bayes estimates produces the same factor scores for the measured variables and the weight matrix which are obtained through the principle of least squares when the error term is assumed to be normal (Hildreth, 2013).

This section therefore develops the residual estimators through the varied possible weight matrices discussed. It is imperative to provide a bit of information regarding the notation utilized here. Variables that has the subscript i represents the values for the ith individual or observation. Those that do not have subscript i were either structural parameters (such as Γ) or estimators. The study carefully considered a clear distinction between estimators and the resulting residuals in this chapter so as to avoid any confusion.

To construct residuals and define residual estimators in the SEM context, recall the equations associated with a SEM for the ith individual:

$$\eta_i = B\eta_i + \Gamma\xi_i + \zeta_i \tag{3.8}$$

$$y_i = \Lambda_y\eta_i + \epsilon_i \tag{3.9}$$

$$x_i = \Lambda_x\xi_i + \delta_i \tag{3.10}$$

Where, x_i represents a vector of the measured values for the exogenous latent variables for the ith individual; δ_i represents the vector of the measurement error for items linked to the exogenous latent variables for the ith observation,

At this point, it follows from the method of Bollen and Arminger (1991) that for reasons of convenience, we apply the definitions

$$z_i = \begin{bmatrix} y_i \\ x_i \end{bmatrix}, \qquad L_i = \begin{bmatrix} \eta_i \\ \xi_i \end{bmatrix}, \qquad v_i = \begin{bmatrix} \varepsilon_i \\ \delta_i \end{bmatrix}, \qquad \Lambda = \begin{bmatrix} \Lambda_y & 0 \\ 0 & \Lambda_x \end{bmatrix}$$

So that Eqns 4.2 and 4.3 would become

$$z_i = \Lambda L_i + v_i \tag{3.11}$$

46

Where at the ith of N independent values, z_i is a vector of $(p+q) \times 1$ measured variables, L_i is a vector of $(m+n) \times 1$ measurement errors for z_i, and $\Lambda \in \mathfrak{R}^{(p+q)\times(m\times n)}$ is a matrix of $(p+q) \times (m+n)$ of coefficients associating z_i to L_i so that $\Lambda \neq 0$.

Also, $\Sigma_{LL} \in \mathfrak{R}^{(m+n)\times(m+n)}$ and $\Sigma_{vv} \in \mathfrak{R}^{(p+q)\times(p+q)}$ are covariance matrices linked to (3.11)

$$\Sigma_{vv} = \begin{bmatrix} \theta_\varepsilon & 0 \\ 0 & \theta_\delta \end{bmatrix}$$

and

$$\Sigma_{LL} = \begin{bmatrix} \Sigma_{\eta\eta} & (I-B)^{-1} A\Phi \\ \Phi A'(I-B)^{-T} & \Phi \end{bmatrix}$$

$$\Sigma_{zz} = \begin{bmatrix} \Lambda_y (I-B)^{-1}(A\Phi A' + \Psi)(I-B)^{-T}\Lambda_y' + \theta_\varepsilon & \Lambda_y (I-B)^{-1} A\Phi\Lambda_x' \\ \Lambda_x \Phi A'(I-B)^{-T}\Lambda_y & \Lambda_x \Phi\Lambda_x' + \theta_\delta \end{bmatrix}$$

Thus, Σ_{vv}, Σ_{LL} and Σ_{zz} are the covariance matrices which are assumed to be positive definite (non-singular) and symmetric (Bollen & Arminger, 1991).

By rearrangement, Equations (3.8) and (3.11) then the residuals for the individual i would be defined by

$$v_i = z_i - \Lambda L_i \tag{3.12}$$

$$\zeta_i(\theta) = ML_i \tag{3.13}$$

For $M \in \mathfrak{R}^{m\times(m+n)} = \left[(I-B)^{-1} - A\right]$.

In real sense, the observations of L_i in Equations (3.12) and (3.13) are not known and therefore are replaced with their corresponding factor scores defined by

$$\hat{L}_i = Wz_i \tag{3.14}$$

47

Thus, the estimators of the residual in Equation (3.12) and (3.13) are expressed as

$$\hat{v}(\theta) = (I - \Lambda W)z \qquad (3.15)$$

$$\hat{\zeta}(\theta) = MWz \qquad (3.16)$$

Where $\hat{v}(\theta)$ and $\hat{\zeta}(\theta)$ are the notations that are used to represent the residual estimators which are functions of the population parameters in the vector θ. This contains the peculiar elements in B, A, Λ, Σ_{LL}, Σ_{vv} and Ψ (Bollen & Arminger, 1991). Meanwhile, the observations of θ are replaced with their sample counterpart $\hat{\theta}$ so that the residual estimators utilized in real sense are as follows

$$\hat{v}(\hat{\theta}) = (I - \hat{\Lambda}\hat{W})z \qquad (3.17)$$

$$\hat{\zeta}(\hat{\theta}) = \hat{M}\hat{W}z \qquad (3.18)$$

Thus the three most frequent selection of W that can be utilized in Equations (3.17) and (3.18) are given by

$$W_R = \Sigma_{LL}\Lambda'\Sigma_{zz}^{-1} \qquad (3.19)$$

$$W_B = \left(\Lambda'\Sigma_{vv}^{-1}\Lambda\right)^{-1}\Lambda'\Sigma_{vv}^{-1} \qquad (3.20)$$

$$W_{AR} = F^{-1}\Lambda'\Sigma_{vv}^{-1} \qquad (3.21)$$

where $F^2 = \left(\Lambda'\Sigma_{vv}^{-1}\Sigma_{zz}\Sigma_{vv}^{-1}\Lambda\right)$.

However, in practice, their sample counterparts $\hat{W}_R$, $\hat{W}_B$ and $\hat{W}_{AR}$ are utilized in equations (3.17) and (3.18) (Bollen & Arminger, 1991). Therefore, the three residual estimators which are considered for the measurement errors in this study are

$$\hat{v}_R(\hat{\theta}) = \left(I - \hat{\Lambda}\hat{\Sigma}_{LL}\hat{\Lambda}'\hat{\Sigma}_{zz}^{-1}\right)z \qquad (3.22)$$

$$\hat{v}_B(\hat{\theta}) = \hat{M}\left(\hat{\Lambda}'\hat{\Sigma}_{vv}^{-1}\hat{\Lambda}\right)^{-1}\hat{\Lambda}'\hat{\Sigma}_{vv}^{-1}z \qquad (3.23)$$

48

$$\hat{v}_{AR}\left(\hat{\theta}\right) = \hat{M}\hat{F}'\hat{\Lambda}'\hat{\Sigma}_{vv}^{-1}z \qquad (3.24)$$

where $\hat{F}^2 = \left(\hat{\Lambda}'\hat{\Sigma}_{vv}^{-1}\hat{\Sigma}_{zz}\hat{\Sigma}_{vv}^{-1}\hat{\Lambda}\right)$.

It is worthy of note that the estimators in Equations (3.19) and (3.21) are known as the regression method of measurement error residuals estimator and the regression method for latent errors of the residual estimators respectively. Again, the estimators in Equations (3.22) and (3.24) are known as the Bartlett's method of the measurement error residual estimator and the Barlett's method of the latent errors of the residual estimators respectively. Also, the estimators contained is Equations (3.21) and (3.24) are called the Anderson-Rubin method of measurement error residual estimator and the Anderson-Rubin method latent errors of the residual estimators respectively.

$$\Sigma_{zz} = \begin{bmatrix} \Lambda_y(I-B)^{-1}(\Gamma\Phi\Gamma' + \Psi)(I-B)^{-T}\Lambda_y' + \theta_\varepsilon & \Lambda_y(I-B)^{-1}\Gamma\Phi\Lambda_x' \\ \Lambda_x\Phi\Gamma, (I-B)^{-T}\Lambda_y' & \Lambda_x\Phi\Lambda_x + \theta_\delta \end{bmatrix}$$

3.12 Finite Sample Properties of $\hat{v}(\theta)$ and $\hat{\zeta}(\theta)$

According to Bollen and Arminger (1991) four main properties are basically desired for estimators in SEM concept. The selection process of the right estimators $v(\theta)$ and $\zeta(\theta)$ is often backed by the properties which are as follows

1. Conditional unbiasedness: This concerns the accuracy of an estimator. The property here concerns the conditional expectation of an estimator adopted for observations whose true factor scores are L and therefore are makes it desirable for the conditional bias to be 0.

2. Mean squared error (MSE): This property assesses the precision of an estimator. The MSE under this assesses the average squared errors which are defined to be the difference between the residuals attained through a particular estimator and

49

the true observations of the residual. Choosing an estimator that reduces this value must always be desired.

3. Structure preservation: Here, the selection or choice of an estimator must not change the associations among the error terms. This means that it is preferred that both covariance structures of the estimated residual as well as the true residual are the same.

4. Univocality: This refers to a situation where the residuals estimated associate or correlate mainly with their counterpart true residuals (i.e validity) but does not correlate with the noncorresponding residuals (i.e. invalidity). Any estimator that exhibit this property is considered as univocal.

The four properties stated above would be properly defined, by mathematical expressions, here subsequently. Equation (3.15) and Equation (3.16) are deemed to be linear function of factor score estimators of the category in Equation (3.14). Thus it is worth highlighting the properties of the factor score estimators since they mirror the residual estimators. These properties of the factor score estimators utilize the weighted matrices contained in Equations (3.19), (3.20), and (3.21) given that θ is known for all the elements. Based on previous studies by McDonald and Burr (1967), Lawley and Maxwell (1971) and Saris et al (1978) the factor score estimator is conditionally not biased as well as possessing smaller mean square error, when utilizing the Bartlett's method, for every conditionally not biased estimators. However, for the regression method, the factor score estimator only possess smaller mean square error for every estimator. Moreover, according Saris et al (1978) the Anderson-Rubin method estimator has structure preservation in that $E(\hat{L}\hat{L}) = E(LL') = \Sigma_{LL}$ (the original structure of the latent variables are preserved by W_{ar}). This holds based on the assumption of

50

orthogonality of the model (where $E(LL') = I$). Univocality was ascertained by McDonald and Burr (1967) for the Bartlett's method factor score.

Last but not least the residual estimators' distribution are addressed. Thus this property together with the four properties mentioned above would provide basis for developing the right residual diagnostics.

Theoretically the derivation of the finite sample properties rely on the assumption that there exist a parameter space Θ for the vector parameters θ (McDonald & Burr, 1967) which corresponds to;

1. All $\Lambda = \begin{bmatrix} \Lambda_y & 0 \\ 0 & \Lambda_x \end{bmatrix}$, $\Lambda_y \in \Re^{p \times m}$, $\Lambda_x \in \Re^{q \times n}$ where Λ has rank $(m + n)$.

2. All $\Gamma \in \Re^{m \times n}$ where Γ has rank n.

3. All $B \in \Re^{m \times m}$ such that $(I - B)^{-1}$ exists and diag $(B) = 0$.

4. All $\Sigma_{vv} = \begin{bmatrix} \theta_\varepsilon & 0 \\ 0 & \theta_\delta \end{bmatrix}$ where $\theta_\varepsilon \in \Re^{p \times p}$ and $\theta_\delta \in \Re^{q \times q}$ are symmetric as well as positive definite matrices.

5. All $\Phi \in \Re^{n \times n}$ are symmetric as well as positive definite matrices..

6. All $\Psi \in \Re^{m \times n}$ are symmetric as well as positive definite matrices.

7. All associated matrices $\Sigma_{\eta\eta} \in \Re^{m \times m}$, $\Sigma_{LL} \in \Re(m + n) \times (m + n)$, and $\Sigma_{zz} \in \Re(p + q) \times (p + q)$ are symmetric as well as positive definite matrices.

3.12.1 Conditional Unbiasedness

Under SEM, conditional unbiasedness concerns the accuracy of the residual estimator which is conditioned on every element with factor scores L. It is important for all estimators to be accurate in order for the estimated residuals to estimate the true residuals.

Therefore the following definitions are provided are utilized to secure results for the conditional unbiasedness of the estimators $\hat{v}(\theta)$ as well as $\hat{\zeta}(\theta)$. Subsequently, these results would be utilized to arrive at heuristic results in terms of the estimators $\hat{v}(\theta)$ and $\hat{\zeta}(\theta)$.

Definition 3.1 McDonald and Burr, (1967) indicated that the factor score estimator $\hat{L}$ was conditionally unbiased estimator of the true factor scores L which holds for all $L \in \Re(m + n) \times N$

$$E[\hat{L}|L] = L.$$

Definition 3.2 The models stipulated in Equations (3.8) and (3.11) as well as the residuals stated in Equation (3.12) means the estimator $\hat{v}(\theta)$ is conditionally unbiased which holds for all $\theta \in \Theta$ as well as for all $L \in \Re(m + n) \times N$ (McDonald & Burr, 1967)

$$E[(\hat{v}(\theta) - v(\theta))|L] = 0.$$

Definition 3.3 The models defined in Equations (3.8) and (3.11) as well as the residuals stated in Equation (3.13) indicate the estimator $\hat{\zeta}(\theta)$ is conditionally unbiased, which holds for all $\theta \in \Theta$ together for all $L \in \Re(m + n) \times N$ (McDonald & Burr, 1967)

$$E[(\hat{\zeta}(\theta) - \zeta(\theta)) |L] = 0.$$

Definition 3.4 Models defined in Equations (3.8) and (3.11) and the residuals stated in Equation (3.12) implies that the estimator $\hat{v}(\hat{\theta})$ is conditionally unbiased, which holds for all $\theta \in \Theta$ together for all $L \in \Re(m + n) \times N$ (McDonald & Burr, 1967)

$$E[(\hat{v}(\hat{\theta}) - v(\theta)) |L] = 0.$$

Definition 3.5 The models defined in Equations (3.8) and (3.11) and the residuals stated in Equation (3.13) shows that the estimator $\hat{\zeta}(\hat{\theta})$ is conditionally unbiased, which holds for all $\theta \in \Theta$ together for all $L \in \Re(m + n) \times N$ (McDonald & Burr, 1967)

$$E[(\hat{\zeta}(\hat{\theta}) - \zeta(\theta)) \,|L] \ = 0.$$

Thus Results 1 as well as 2 are formulated based the definitions from 3.8 to 3.10.

Result 1: For the models defined in Equation (3.8) and Equation (3.11),

a) The estimator $\hat{v}(\theta)$ defined in Equation (3.15) is a conditionally not biased estimator of the true residuals stated in Equation (3.12) which holds for only factor score estimator $\hat{L} = Wz$ is a conditionally not biased estimator of the true factor scores L.

b) For estimator $\hat{v}(\theta)$ defined in Equation (3.15) is said to be conditionally not biased estimator of the true residuals $v(\theta)$ stated in Equation (3.12) holds for only $W\Lambda = I$ (or alternatively only when $\Lambda W = I$).

c) Estimators under consideration in Equations $(3.22) - (3.24)$, the Bartlett's method estimator in Equation (3.23) holds for $W\Lambda = I$ in that the Bartlett's method residual estimator remains the only conditionally unbiased estimator.

Proof. Recall that

$$\hat{v}(\theta) \ = \ (I - \Lambda W)z$$

$$v(\theta) \ = \ z - \Lambda L$$

$$\hat{L} = Wz.$$

We now demonstrate that $\Lambda W \ = I$ is equivalent to $W\Lambda = I$ given all $\theta \in \Theta$. Subsequently it would be shown that every part of the three proofs holds. Particularly that this equivalence is helpful in proving part (b) as well as part (c) of the Result. Suppose $\Lambda W = I$, then

53

$$\Lambda W = I$$

$$I - \Lambda W = 0$$

$$\Lambda'(I - \Lambda W)\Lambda = 0$$

$$\Lambda'\Lambda - \Lambda'\Lambda W\Lambda = 0$$

$$\Lambda'\Lambda = \Lambda'\Lambda W\Lambda$$

$$(\Lambda'\Lambda)^{-1}\Lambda'\Lambda = (\Lambda'\Lambda)^{-1}\Lambda'\Lambda W\Lambda$$

$$I = W\Lambda.$$

Therefore it is proved that $\Lambda W = I$ equivalent to $W\Lambda = I$ for all $\theta \in \Theta$.

Now, we demonstrate or show that part (a) of Result 1 holds. Suppose that $\hat{v}(\theta)$ is conditionally unbiased estimator for (θ). Using Definition 3.2, it holds that for all $\theta \in \Theta$ together with all $L \in \Re(m + n) \times N$

$$E[(\hat{v}(\theta) - v(\theta))|L] = 0.$$

Then,
$$0 = E[(\hat{v}(\theta) - v(\theta))|L] = E[((I - \Lambda W)z - (z - \Lambda L))|L]$$

$$= E[(z - \Lambda W z - z + \Lambda L)|L]$$

$$= \Lambda[E(L|L) - E(\hat{L}|L)]$$

$$= \Lambda L - \Lambda E(\hat{L}|L)$$

$$(\Lambda'\Lambda)^{-1}\Lambda'\Lambda L = (\Lambda'\Lambda)^{-1}\Lambda'\Lambda E(\hat{L}|L)$$

$$L = E(\hat{L}|L) .$$

Therefore, given that $\hat{v}(\theta)$ is conditionally unbiased estimator of $v(\theta)$, then from Definition 3.1, $\hat{L}$ is equally conditionally unbiased estimator for L (McDonald & Burr, 1967).

Now, suppose that $\hat{L}$ is conditionally unbiased estimator of L so that Definition 3.1 is satisfied. By implication when all $L \in \Re(m + n) \times N,$

$$E[(\hat{v}(\theta) - v(\theta))|L] = E[((I - \Lambda W)z - (z - \Lambda L))|L]$$

$$= E[(z - \Lambda W z - z + \Lambda L)|L]$$

$$= \Lambda[E(L|L) - E(\hat{L}|L)]$$

$$= \Lambda[L - E(\hat{L}|L)]. \tag{3.25}$$

Rearrangement of Definition 3.1 holds given all $L \in \Re(m + n) \times N$ so that

$$L - E(\hat{L}|L) = 0$$

By substituting the above into Equation (3.25), then

$$E[(\hat{v}(\theta) - v(\theta))|L] = \Lambda 0$$

$$= 0$$

when all $\theta \in \Theta$ and all $L \in \Re(m + n) \times N$. Therefore as the factor score estimator $\hat{L}$ is conditionally unbiased estimator of L then by extension $\hat{v}(\theta)$ equally is conditionally unbiased estimator of $\hat{v}(\theta)$ using Definition 3.2. Thus $\hat{v}(\theta)$ is conditionally unbiased estimator of $v(\theta)$ whenever $\hat{L}$ is conditionally unbiased estimator for L (McDonald & Burr, 1967). Hence the prove of part (a) of Result 1 is confirmed.

Now part (b) of Result 1 is confirmed so that $\hat{v}(\theta)$ is conditionally unbiased estimator of $\hat{v}(\theta)$ whenever $W\Lambda = I$. To begin with, suppose $W\Lambda = I$.

We can easily show that generally the conditional bias of $\hat{v}(\theta)$ given $\theta \in \Theta$ and $L \in \Re(m + n) \times N$ is given by

$$E[\hat{v}(\theta) - v(\theta)|L] = (I - \Lambda W)\Lambda L \tag{3.26}$$

$$= 0.$$

55

Now, suppose $\hat{v}(\theta)$ is conditionally unbiased estimator of $v(\theta)$ to the extent that Definition 3.2 is satisfied. Thus by Equation (3.26), $(I - \Lambda W)\Lambda L = 0$ must be satisfied for all $L \in \Re(m + n) \times N$ given by

$$(I - \Lambda W)\Lambda \ = \ 0$$

$$\Lambda'(I - \Lambda W)\Lambda \ = \ 0$$

$$\Lambda'\Lambda - \Lambda'\Lambda W\Lambda \ = \ 0$$

$$\Lambda'\Lambda \ = \ \Lambda'\Lambda W\Lambda$$

$$I \ = \ W\Lambda$$

Therefore part (b) of Result 1 is proved as $\hat{v}(\theta)$ is conditionally unbiased estimator of the true residuals $v(\theta)$ whenever the factor score estimator $W\Lambda = I$.

For part (c) of Result 1 to be shown that it holds. Then it must be shown that $W\Lambda = I$ for $W = W_b$ (McDonald & Burr, 1967). Thus,

$$W_b\Lambda \ = \ (\Lambda'\Sigma_{vv}^{-1}\Lambda)^{-1}\Lambda'\Sigma_{vv}^{-1}\Lambda$$

$$= I$$

Therefore holds for $W = W_b$, $\Lambda W = I$.

Now, as the proofs of parts (a), (b), and (c) are combined then Result 1 is satisfied.

Result 2: The models defined in Equation (3.8) and Equation (3.11) given that $m \geq n$

a) For estimator $\hat{\zeta}(\theta)$ defined in Equation (3.16) is conditionally not biased estimator of the true residuals $v(\theta)$ stated in Equation (3.13) whenever the factor score estimator $\hat{L} = Wz$ is conditionally not biased estimator of the true factor scores L.

b) For estimator $\hat{\zeta}(\theta)$ defined in Equation (3.16) is conditionally not biased estimator of the true residuals $\zeta(\theta)$ stated in Equation (3.13) whenever $W\Lambda = I$.

56

c) The estimators considered in Equations $(3.125) - (3.27)$, then it means that the Bartlett's method estimator in Equation (3.26) holds for $W\Lambda = I$ in that the Bartlett's method residual estimator is the one and only one conditionally not biased estimator for those considered.

Proof. Remember that

$$\hat{\zeta}(\theta) \ = \ MWz$$

$$\zeta(\theta) \ = \ ML$$

$$\hat{L} = Wz.$$

To begin with it shall be demonstrated that part (a) of Result 2 is satisfied whenever $m \geq n$, the estimator $\hat{\zeta}(\theta)$ is conditionally not biased estimator of $\zeta(\theta)$ whenever $\hat{L}$ is conditionally not biased estimator for L (McDonald & Burr, 1967). Suppose $\hat{\zeta}(\theta)$ is conditionally not biased estimator of $\zeta(\theta)$. Using Definition 3.3 satisfies for all $\theta \in \Theta$ and $L \in \Re(m + n) \times N$

$$E[(\hat{\zeta}(\theta) - \zeta(\theta)) \, |L] = 0.$$

Thus,

$$0 = E[(\hat{\zeta}(\theta) - \zeta(\theta)) \, |L] \ = \ E[(M\hat{L} - ML) \, |L]$$

$$= \ M[E(\hat{L}|L) - E(L|L)]$$

$$= M\text{E} \, (\hat{L}|L) - ME(L|L)$$

satisfies for all $M \in \Re^{m \times (m+n)}$. Next we prove that $M\text{E} \, (\hat{L}|L) - ME(L|L) = 0$ is satisfied whenever $E(\hat{L}|L) = L$. Thus it is given by

$$D \ = \ E(\hat{L}|L) - L$$

$$= \ [D_1 \, D_2]'$$

as well as

$$M = [C \quad -\Gamma]$$

for $C = (I - B)^{-1}$. then,

$$0 = MD$$

$$= CD_1 - \Gamma D_2 \qquad (3.27)$$

is satisfied for all $B \in \mathfrak{R}^{m \times m}$ given that C is invertible as well as for all $\Gamma \in \mathfrak{R}^{m \times n}$. It must therefore be shown that $D = 0$ has the equivalence by showing that $E[\hat{L}|L] = L$. Given any $\Gamma \in \mathfrak{R}^{m \times n}$ which implies for $\Gamma \neq 0$, pick any value of κ where $\kappa = \frac{1}{k} \to 0$ for $k \to \infty$. Then define $\Gamma^* = \kappa \Gamma$ so that $\Gamma^* \to 0$ as $k \to \infty$ (Bollen & Aminger, 1991). Equation (3.27) then becomes

$$CD_1 - \Gamma^* D_2.$$

Therefore, suppose $k \to \infty$, $D_1 = 0$ for C which is invertible so that $C \neq 0$ for all $\theta \in \Theta$. Subsequently, assume $D_1 = 0$ which implies $\Gamma^* D_2 = 0$. Further suppose $m \geq n$ so that Γ^* can be picked to contain n linear independent columns of length m. Thus, the matrix $\Gamma' \Gamma^*$ of size $n \times n$ is invertible which means that

$$0 = (\Gamma^{*\prime} \Gamma^*)^{-1} \Gamma^{*\prime} \Gamma^* D_2$$

$$= D_2.$$

Thus $D = 0$ or $= E(\hat{L}|L)$. Hence, if $m \geq n$ for $\hat{\zeta}(\theta)$ is conditionally not biased estimator of $\zeta(\theta)$ it implies that from Definition 3.1 $\hat{L}$ is equally conditionally not biased estimator for L (Bollen & Aminger, 1991).

Now, suppose $\hat{L}$ is a conditionally not biased estimator of L to extent that Definition 3.1 is satisfied (McDonald & Burr, 1967). Thus $\theta \in \Theta$ and $L \in \mathfrak{R}(m + n) \times N,$

$$E[(\hat{\zeta}(\theta) - \zeta(\theta)) | L] = E[(M\hat{L} - ML) | L]$$

$$= M[E(\hat{L}|L) - E(L|L)]$$

$$= M[E(\hat{L}|L) - L]. \qquad (3.28)$$

through rearrangement of Definition 3.1 it satisfies for $L \in \Re(m + n) \times N$ so that

$$L - E(\hat{L}|L) = 0.$$

which when substituted into Equation (3.28) yields

$$E[(\hat{\zeta}(\theta) - \zeta(\theta)) | L] = M = 0$$

given that $\theta \in \Theta$ and for all $L \in \Re(m + n) \times N$. Therefore part (a) of Result 2 is proved to the extent that $\hat{\zeta}(\theta)$ is conditionally not biased estimator of $\zeta(\theta)$ whenever $\hat{L}$ is conditionally not biased estimator for L (McDonald & Burr, 1967).

Also, part (b) of Result 2 is satisfied if $\hat{\zeta}(\theta)$ is conditionally not biased estimator of $\zeta(\theta)$ whenever $W\Lambda = I$. In order to prove this, let us suppose $W\Lambda = I$. It therefore becomes easier to show generally that the conditional bias of $\hat{\zeta}(\theta)$ is given as

$$E[(\hat{\zeta}(\theta) - \zeta(\theta))] = M(W\Lambda - I)L. \qquad (3.29)$$

Thus for every $\theta \in \Theta$ and $L \in \Re(m + n) \times N$, then

$$M(W\Lambda - I)L = M(I - I)L$$

$$= 0.$$

We suppose next that $\hat{\zeta}(\theta)$ is conditionally not biased estimator of $\zeta(\theta)$ in that Definition 4.3 is satisfied. Thus, utilizing Equation (3.29), $M(W\Lambda - I)L = 0$ necessarily satisfies when $L \in \Re(m + n) \times N$ and when $M \in \Re^{m \times (m+n)}$ which means (considering $L = [I \; 0]$ given $(m + n) \times (m + n)$ matrix I) (Bollen & Aminger, 1991)

$$M(W\Lambda - I) = 0$$

satisfies when $M \in \Re^{m \times (m+n)}$. Consequently, we show that $M(W\Lambda - I) = 0$ exist whenever $M(W\Lambda + I)$ which can be defined as

$$D = (W\Lambda - I) = 0$$

$$= [D_1 \ D_2]'$$

and

$$M = [C \ -\Gamma]$$

given $C = (I - B)^{-1}$, it impies

$$0 = MD$$

$$= CD_1 - \Gamma D_2 \qquad (3.30)$$

exist when $B \in \Re^{m \times m}$ for C which is invertible as well as when $\Gamma \in \Re^{m \times n}$. Therefore it can be proved that $D = 0$ is equivalent to proving that $W\Lambda = I$. Given that $\Gamma \in \Re^{m \times n}$ for $\Gamma \neq 0$, pick any value of κ for $\kappa = \frac{1}{k} \to 0$ as $k \to \infty$. We define $\Gamma^* = \kappa\Gamma$ when $\Gamma^* \to 0$ as $k \to \infty$. Rewrite Equation (3.30) as

$$CD_1 - \Gamma^* D_2.$$

Thus suppose $k \to \infty$, $D_1 = 0$ when C is invertible means $C \neq 0$ when $\theta \in \Theta$. Consequently assume that $D_1 = 0$ which means that $\Gamma^* D_2 = 0$. Further suppose $m \geq n$ so that Γ^* could be picked to comprise n linear independent columns of length m (McDonald & Burr, 1967). Thus, the matrix $\Gamma'\Gamma^*$ with dimension $n \times n$ is deemed invertible such that

$$0 = (\Gamma^{*'}\Gamma^*)^{-1}\Gamma^{*'}\Gamma^* D_2$$

$$= D_2.$$

Therefore, part (b) of Result 2 is proved to the extent that $\hat{\zeta}(\theta)$ is conditionally not biased estimator of the true residuals $\zeta(\theta)$ whenever $W\Lambda = I$.

Finally, in order to prove part (c) of Result 2 we will show that $W\Lambda = I$ as $W = W_b$.

Thus,

$$W_b\Lambda = (\Lambda'\Sigma_{vv}^{-1}\Lambda)^{-1}\Lambda'\Sigma_{vv}^{-1}\Lambda$$

$$= I$$

Hence, as $W = W_b$ then $W\Lambda = I$ which proves part (c) of Result 2.

Now, the combination of proofs of parts (a), (b), and (c) then Result 2 is said to be proved accordingly (McDonald & Burr, 1967). Consequently, the following facts are worth noting:

1. The conditional biases of the regression and Anderson-Rubin method estimators equal 0 whenever either (a) $\theta = 0$ or (b) $L = 0$. However, it is not possible for $\theta = 0$ in real sense nor is it possible for $L = 0$. Therefore, the regression as well as the Anderson-Rubin method estimators are generally conditionally biased.

2. Generally, there exist no upper limit regarding the absolute value in terms of of the residuals gotten from the estimators contained in Equations (3.17) and (3.18) since the only constraint regarding the value of either θ or L are said to be real-valued.

3. Both the regression and Anderson-Rubin method estimators derive their values from θ and true factor scores L which defines the sign of conditional bias.

The Heuristic Result 1 demonstrates that both $\hat{v}_b(\hat{\theta}_{MLE})$ as well as $\hat{\zeta}_b(\hat{\theta}_{MLE})$ indicate close attributes in terms of $\hat{v}_b(\theta)$ and $\hat{\zeta}_b(\theta)$ so that $\hat{v}_b(\hat{\theta}_{MLE})$ as well as $\hat{\zeta}_b(\hat{\theta}_{MLE})$ were conditionally not biased estimators of $v(\theta)$ and $\zeta(\theta)$, respectively.

61

3.12.2 Mean Squared Error

Though the conditionally unbiased criteria is helpful in choosing from estimators, it basically examines accurateness but fails to examine the precision level of an estimator. Generally, it is imperative to have a precise estimator such that there is higher degree of closeness between the estimates residuals and their true value residual counterparts. The commonest criteria which is often used ascertain same is the Mean Squared Error (MSE) (Byrne, 2010). The following definition spanning from 3.15 to 3.18 defines the MSE under the SEM concept in this study which then establishes Result 3. Consequently, Heuristic Result 2 is established based on the earlier stated Results 4 as well as 5.

Definition 3.6 Consider the models defined in Equations (3.8) as well as (3.11) and the residuals stated Equation (3.12) (Bollen & Aminger, 1991), then

Mean square error of the estimator $\hat{v}(\theta)$ and $(\hat{\zeta}(\theta)$ stated in Equation (3.15) is as follows:

$$Tr\left[E(\hat{v}(\theta) - v(\theta))(\hat{v}(\theta) - v(\theta))'\right]$$

$$Tr\left[E(\hat{\zeta}(\theta) - \zeta(\theta))\,(\hat{\zeta}(\theta) - \zeta(\theta))'\right]$$

$$Tr\left[E(\hat{v}(\hat{\theta}) - v(\theta))\,(\hat{v}(\hat{\theta}) - v(\theta))'\right]$$

$$Tr\left[E(\hat{\zeta}(\hat{\theta}) - \zeta(\theta))\,(\hat{\zeta}(\hat{\theta}) - \zeta(\theta))'\right]$$

Result 3: Consider the models defined under Equations (3.8) as well as (3.11), when $\theta \in \Theta$ then the estimator $\hat{v}(\theta)$ as stated under Equation (3.22) realizes the smallest mean squared error from all the estimators of the class under Equation (3.15).

Proof. Recollect that

$$\hat{v}(\theta) = (I - \Lambda W)z$$

$$v(\theta) = z - \Lambda L.$$

First, it is imperative to identify the weight matrix $\boldsymbol{W}$ that will minimize the MSE of the estimator $\hat{v}(\theta)$ as stated by Definition 3.6. Therefore, it becomes appropriate to determining what weight matrix $\boldsymbol{W}$ minimizes

$$\mathcal{L} = Tr[E(\hat{v}(\theta) - v(\theta))(\hat{v}(\theta) - v(\theta))'].$$

Take the first derivative in respect of $\boldsymbol{W}$

$$\frac{\partial}{\partial W}\mathcal{L} = \frac{\partial}{\partial W}\,\text{Tr}\,[E(((I - \Lambda W)z - v)((I - \Lambda W)z - v$$

$$= \frac{\partial}{\partial W}\,\text{Tr}\,[(I - \Lambda W)\Sigma_{zz}(I - \Lambda W)' - 2(I - \Lambda W)\Sigma_{vv} + \Sigma_{vv}]$$

$$= \frac{\partial}{\partial W}\,\text{Tr}\,[\Sigma_{zz} - 2\Lambda'W\Sigma_{zz} + \Lambda W\Sigma_{zz}W'\Lambda' - \Sigma_{vv} + 2\Lambda'W\Sigma_{vv}]$$

$$= -2\Lambda'\Sigma_{zz} + 2\Lambda'\Lambda W\Sigma_{zz} + 2\Lambda'\Sigma_{vv}, \quad\quad\quad (3.31)$$

by equating Equation (3.31) to 0 and working for $\boldsymbol{W}$ then

$$-2\Lambda'\Sigma_{zz} + 2\Lambda'\Lambda W\Sigma_{zz} + 2\Lambda'\Sigma_{vv} = 0$$

$$-\Lambda'\Sigma_{zz} + \Lambda'\Lambda W\Sigma_{zz} + \Lambda'\Sigma_{vv} = 0$$

$$\Lambda'\Lambda W\Sigma_{zz} = \Lambda'\Sigma_{zz} - \Lambda'\Sigma_{vv}$$

$$(\Lambda'\Lambda)^{-1}\Lambda'\Lambda W\Sigma_{zz}\Sigma_{zz}^{-1} = (\Lambda'\Lambda)^{-1}\Lambda'\Sigma_{zz}\Sigma_{zz}^{-1} - (\Lambda'\Lambda)^{-1}\Lambda'\Sigma_{vv}\Sigma_{zz}^{-1}$$

$$IWI = (\Lambda'\Lambda)^{-1}\Lambda'I - (\Lambda'\Lambda)^{-1}\Lambda'\Sigma_{vv}\Sigma_{zz}^{-1}$$

$$W = (\Lambda'\Lambda)^{-1}\Lambda' - (\Lambda'\Lambda)^{-1}\Lambda'\Sigma_{vv}\Sigma_{zz}^{-1}$$

$$= (\Lambda'\Lambda)^{-1}\Lambda' - (\Lambda'\Lambda)^{-1}\Lambda'(\Sigma_{zz} - \Lambda\Sigma_{LL}\Lambda')\Sigma_{zz}^{-1}$$

$$= (\Lambda'\Lambda)^{-1}\Lambda' - (\Lambda'\Lambda)^{-1}\Lambda'\Lambda\Sigma_{zz}\Sigma_{zz}^{-1} + (\Lambda'\Lambda)^{-1}\Lambda'\Lambda\Sigma_{LL}\Lambda'\Sigma_{zz}^{-1}$$

$$= (\Lambda'\Lambda)^{-1}\Lambda' - (\Lambda'\Lambda)^{-1}\Lambda'I + I\Sigma_{LL}\Lambda'\Sigma_{zz}^{-1}$$

$$= (\Lambda'\Lambda)^{-1}\Lambda' - (\Lambda'\Lambda)^{-1}\Lambda' + \Sigma_{LL}\Lambda'\Sigma_{zz}^{-1}$$

$$= \Sigma_{LL}\Lambda'\Sigma_{zz}^{-1}$$

$$= W_r$$

that is from Equation (3.19).

Result 4: Consider the models defined under Equations (3.8) as well as (3.11), when $\theta \in \Theta$ for $m \geq n$ then the estimator $\hat{\zeta}(\theta)$ as stated under Equation (3.27) succeeds a minimum mean squared error from all the estimators contained in Equation (3.16).

Proof. Recollect that

$$\hat{\zeta}(\theta) \ = \ MWz$$

$$\zeta(\theta) \ = \ ML.$$

Now it is imperative to identify the weight matrix **W** that will minimize the MSE of the estimator $\hat{\zeta}(\theta)$ as stated by Definition 3.7. Hence, we find what weight matrix **W** will minimize (Bollen & Aminger, 1991)

$$\mathcal{L} = Tr[E(\hat{\zeta}(\theta) - \zeta(\theta))(\hat{\zeta}(\theta) - \zeta(\theta))'].$$

To begin with take its derivative in respect of **W**

$$\frac{\partial}{\partial W}\mathcal{L} = \frac{\partial}{\partial W}\text{Tr}\left[E((MWz—ML)\,((MWz - ML\right.$$

$$= \frac{\partial}{\partial W}\text{Tr}\left[MW\Sigma_{zz}W'M' - 2M'M\Sigma_{LL}\Lambda'W' - M\Sigma_{LL}M'\right]$$

$$= \ 2M'MW\Sigma_{zz} - 2M'M\Sigma_{LL}\Lambda'. \qquad\qquad (3.32)$$

while equating Equation (3.32) to 0 as well as working for **W**

$$2M'MW\Sigma_{zz} - 2M'M\Sigma_{LL}\Lambda' = 0$$

$$M'MW\Sigma_{zz} - M'M\Sigma_{LL}\Lambda' = 0$$

$$(MM')^{-1}MM'MW\Sigma_{zz} - (MM')^{-1}MM'M\Sigma_{LL}\Lambda' \ = \ 0$$

$$MW\Sigma_{zz} - M\Sigma_{LL}\Lambda' \ = \ 0$$

$$MW\Sigma_{zz}\Sigma_{zz}^{-1} - M\Sigma_{LL}\Lambda'\Sigma_{zz}^{-1} \ = \ 0$$

$$MW - M\Sigma_{LL}\Lambda'\Sigma_{zz}^{-1} \ = \ 0$$

$$M[W - \Sigma_{LL}\Lambda'\Sigma_{zz}^{-1}] \ = \ 0$$

satisfied when $M \in \Re^{m \times (m+n)}$. Now, it would be demonstrated that $M[W - \Sigma_{LL}\Lambda'\Sigma_{zz}^{-1}]$

$= 0$ is satisfied whenever $W = \Sigma_{LL}\Lambda'\Sigma_{zz}^{-1}$ (Bollen & Aminger, 1991). Thus it is defined

as follows

$$D = W - \Sigma_{LL}\Lambda'\Sigma_{zz}^{-1}$$

$$= [D_1 \, D_2]'$$

and

$$M = [C \ - \Gamma]$$

when $C = (I - B)^{-1}$, Thus

$$0 = MD$$

$$= CD_1 - \Gamma D_2$$

$$(3.33)$$

is established hen $B \in \Re^{m \times m}$ for C is deemed invertible as well as when $\Gamma \in \Re^{m \times n}$. Thus

it is imperative to prove that $D = 0$ which is the same as proving that $W = \Sigma_{LL}\Lambda'\Sigma_{zz}^{-1}$.

Consider any $\Gamma \in \Re^{m \times n}$ so that $\Gamma \neq 0$, pick any value of κ for $\kappa = \frac{1}{k} \to 0$ as $k \to \infty$. The

define $\Gamma^* = \kappa\Gamma$ so that $\Gamma^* \to 0$ as $k \to \infty$. Thus we rewrite Equation (3.33) as

$$CD_1 - \Gamma^* D_2.$$

$$0 = (\Gamma^{*'}\Gamma^*)^{-1}\Gamma^{*'}\Gamma^* D_2$$

$$= D_2$$

Heuristic Result 2: Consider the models defined under Equations (3.15) as well as

(3.19), then both estimators, $\hat{v}(\hat{\theta})$ and $\hat{\zeta}(\hat{\theta})$, attains a smaller MSE for all the estimators

contained in Equation (3.24) and Equation (3.25) respectively.

Proof. Recollect that

$$\hat{v}(\theta) = (I - \Lambda W)z$$
$$v(\theta) = z - \Lambda L.$$

Thus, the Lagrange function of concern is expressed as

$$\mathcal{L} = \text{Tr}\,[E(\hat{v}(\theta) - v(\theta))(\hat{v}(\theta) - v(\theta))' + 2C(\Lambda W - I)]$$

$$= \text{Tr}\,[\Sigma_{zz} - 2\Lambda W \Sigma_{zz} + \Lambda W \Sigma_{zz} W' \Lambda' - \Sigma_{vv} + 2\Lambda W \Sigma_{vv} + 2C(\Lambda W - I)] \quad (3.35)$$

for C is a $(p + q) \times (p + q)$ matrix of indeterminate multipliers (Lagrangian multipliers). Thus the partial derivatives of Equation (3.35) regarding W are considered so that

$$\frac{\partial}{\partial W}\mathcal{L} = -2\Lambda' \Sigma_{zz} + 2\Lambda' \Lambda W \Sigma_{zz} + 2\Lambda' \Sigma_{vv} + 2\Lambda' C'. \quad (3.36)$$

Now, equate Equation (3.30) to 0 while computing for W

$$2\Lambda' \Lambda W \Sigma_{zz} = 2\Lambda' \Sigma_{zz} - 2\Lambda' \Sigma_{vv} - 2\Lambda' C'$$

$$\Lambda' \Lambda W \Sigma_{zz} = \Lambda' \Sigma_{zz} - \Lambda' \Sigma_{vv} - 2\Lambda' C'$$

$$\Lambda' \Lambda W \Sigma_{zz} = \Lambda'(\Sigma_{zz} - \Sigma_{TJTJ} - C')$$

$$(\Lambda' \Lambda)^{-1} \Lambda' \Lambda W \Sigma_{zz} \Sigma_{zz}^{-1} = (\Lambda' \Lambda)^{-1} \Lambda'(\Sigma_{zz} - \Sigma_{TJTJ} - C')\Sigma_{zz}^{-1}$$

$$IWI = (\Lambda' \Lambda)^{-1} \Lambda'(\Sigma_{zz} - \Sigma_{vv} - C')\Sigma_{zz}^{-1}$$

$$W = (\Lambda' \Lambda)^{-1} \Lambda'(\Sigma_{zz} - \Sigma_{vv} - C')\Sigma_{zz}^{-1}$$

$$= Q \Sigma_{zz}^{-1}$$

$$= Q(\Lambda \Sigma_{LL} \Lambda' + \Sigma_{vv})^{-1}$$

$$= Q(\Lambda \Sigma_{LL} \Lambda' \Sigma_{vv}^{-1} \Sigma_{vv} + \Sigma_{vv} \Sigma_{vv}^{-1} \Sigma_{vv})^{-1}$$

$$= Q(\Lambda \Sigma_{LL} \Lambda' \Sigma_{vv}^{-1} + I)^{-1} \Sigma_{vv}^{-1}$$

$$= Q(\Lambda \Sigma_{LL} \Lambda' \Sigma_{vv}^{-1} \Lambda \Sigma_{LL} \Lambda'(\Lambda \Sigma_{LL} \Lambda')^{-1} + \Lambda \Sigma_{LL} \Lambda'(\Lambda \Sigma_{LL} \Lambda')^{-1})^{-1} \Sigma_{vv}^{-1}$$

$$= Q(\Lambda \Sigma_{LL} \Lambda' \Sigma_{vv}^{-1} \Lambda \Sigma_{LL} \Lambda' + \Lambda \Sigma_{LL} \Lambda')^{-1} \Lambda \Sigma_{LL} \Lambda' \Sigma_{vv}^{-1}$$

$$= Q^* \Lambda' \Sigma_{vv}^{-1}.$$

Now, it becomes imperative to find Q^* by assumption that $W\Lambda = I$ while solving for Q^*

$$W\Lambda = I$$

$$= Q^*\Lambda'\Sigma_{vv}^{-1}\Lambda$$

$$Q^* = (\Lambda'\Sigma_{vv}^{-1}\Lambda)^{-1}$$

thus,

$$W = Q^*\Lambda'\Sigma_{vv}$$

$$= \left(\Lambda'\Sigma_{vv}\Lambda\right)^{-1}\Lambda'\Sigma_{vv}$$

$$= W_b$$

using (3.20).

Hence $\hat{v}(\theta)$ attains the smallest MSE from all the conditional unbiased estimators.

Result 4: Consider the model defined under Equation (3.8) as well as (3.11), when $\theta \in \Theta$ for $m \geq n$ then the estimator, $\hat{\zeta}(\theta)$, stated under Equation (3.26) attains the smallest MSE from all the conditional unbiased estimators contained in Equation (3.16).

Proof. Recollect that

$$\hat{\zeta}(\theta) = MWz$$

$$\zeta(\theta) = ML$$

Thus the Lagrangian function of concern will become

$$\mathcal{L} = \text{Tr}\left[E(\hat{\zeta}(\theta) - \zeta(\theta))(\hat{\zeta}(\theta) - \zeta(\theta))' + 2C(W\Lambda - I)\right]$$

$$= \text{Tr}\left[MW\Sigma_{zz}W'M' - 2M'M\Sigma_{LL}\Lambda'W' - M\Sigma_{LL}M' + 2C(W\Lambda - I)\right] \quad (3.37)$$

for matrix C of size $a(m + n) \times (m + n)$ which is indeterminate (Lagrangian) multipliers (Bollen & Aminger, 1991). Thus the partial derivative for Equation (3.37) regarding $\boldsymbol{W}$ are taken in order that

67

$$\frac{\partial}{\partial W}\mathcal{L} = 2M'MW\Sigma_{zz} - 2M'M\Sigma_{LL}\Lambda' + 2C'\Lambda' \tag{3.38}$$

Therefore it becomes imperative to equate Equation (3.38) to 0 while solving for $\boldsymbol{W}$

$$2M'MW\Sigma_{zz} - 2M'M\Sigma_{LL}\Lambda' + 2C'\Lambda' \; = \; 0$$

$$M'MW\Sigma_{zz} - M'M\Sigma_{LL}\Lambda' + C'\Lambda' \; = \; 0$$

$$(MM')^{-1}MM'MW\Sigma_{zz} - (MM')^{-1}MM'M\Sigma_{LL}\Lambda' + (MM')^{-1}MC'\Lambda' = 0$$

$$MW\Sigma_{zz} - M\Sigma_{LL}\Lambda' + (MM')^{-1}MC'\Lambda' \; = \; 0$$

$$MW\Sigma_{zz} - M\Sigma_{LL}\Lambda' + MM'(MM')^{-1}(MM')^{-1}MC'\Lambda' \; = \; 0$$

$$MW\Sigma_{zz}\Sigma_{zz}^{-1} - M\Sigma_{LL}\Lambda'\Sigma_{zz}^{-1} + MM'(MM')^{-1}(MM')^{-1}MC'\Lambda'\Sigma_{zz}^{-1} \; = \; 0$$

$$M[W - \Sigma_{LL}\Lambda'\Sigma_{zz}^{-1} + M'(MM')^{-1}(MM')^{-1}MC'\Lambda'\Sigma_{zz}^{-1}] \; = \; 0$$

$$M[W - (\Sigma_{LL}\Lambda' + M'(MM')^{-1}(MM')^{-1}MC'\Lambda')\Sigma_{zz}^{-1}] \; = \; 0$$

$$M[W - Q\Sigma_{zz}^{-1}] \; = \; 0$$

$$M[W - Q^*\Lambda'\Sigma_{vv}^{-1}] \; = \; 0$$

that holds $M \in \Re^{m \times (m+n)}$ (see the proof of Result 6 for details on $Q\Sigma_{zz}^{-1} = Q^*\Lambda'\Sigma_{vv}^{-1}$).

Now, it can be demonstrated that $M[W - Q^*\Lambda'\Sigma_{zz}^{-1}] = 0$ holds whenever $W = Q^*\Lambda'\Sigma_{vv}^{-1}$

(Bollen & Aminger, 1991). Thus

$$D \; = \; W - Q\Lambda'\Sigma_{vv}^{-1}$$

$$= \; [D_1 \; D_2]'$$

and

$$M = \; [C \; - \Gamma]$$

where $C = (I - B)^{-1}$, Thus

$$0 \; = \; MD$$

$$= \; CD_1 - \Gamma D_2 \tag{3.39}$$

68

when $B \in \Re^{m \times m}$ for C which is invertible when $\Gamma \in \Re^{m \times n}$. It can then be shown that

$D = 0$ which is the same as proving that $W = Q^* \Lambda' \Sigma_{vv}^{-1}$. When $\Gamma \in \Re^{m \times n}$ so that $\Gamma \neq 0$,

select any value about κ for $\kappa = \frac{1}{k} \to 0$ as $k \to \infty$. Then define $\Gamma^* = \kappa \Gamma$ which implies

$\Gamma^* \to 0$ as $k \to \infty$. Then rewrite Equation (3.39) as

$$CD_1 - \Gamma^* D_2.$$

Suppose $k \to \infty$, $D_1 = 0$ as C is invertible meaning that $C \neq 0$ when $\theta \in \Theta$.

Now, assume $D_1 = 0$ meaning $\Gamma^* D_2 = 0$. Further suppose $m \geq n$ so Γ^* can be selected

to comprise n linear independently columns of magnitude m (Bollen & Aminger, 1991).

Thus, the matrix $\Gamma' \Gamma^*$ of size $n \times n$ is deemed invertible which means that

$$0 = (\Gamma^{*'} \Gamma^*)^{-1} \Gamma^{*'} \Gamma^* D_2$$

$$= D_2.$$

then $D = 0$ or $W = Q^* \Lambda' \Sigma_{vv}^{-1}$.

It is required, finally, to find Q^* through the assumption $W\Lambda = I$ while working for Q^*

$$W\Lambda = I$$

$$Q^* \Lambda' \Sigma_{vv}^{-1} \Lambda = I$$

$$Q^* = (\Lambda' \Sigma_{vv} \Lambda)^{-1}$$

thus,

$$W = Q^* \Lambda' \Sigma_{vv}$$

$$= \left(\Lambda' \Sigma_{vv} \Lambda\right)^{-1} \Lambda' \Sigma_{vv}$$

$$= W_b.$$

69

3.12.3 Structure Preservation

The structure preservation basically concerns the effect of the selected estimator would have on the structure of the covariance of the residual. It is often very crucial that the associations amidst the unobserved errors are not changed because of the selection of the residual estimator regarding some applications. Whenever, for instance, the residuals estimated are hugely associated or related then the residuals estimated together with other residual analysis noting departures from certain conditions and/or possible outliers regarding specific equation of the model under Equation (3.8) and Equation (3.11) can be attributed to departure from certain conditions and/or possible outliers in different equation of the same model. As a result, it is preferable that a residual estimator to be structure preserving. It is worth noting that Definitions 3.17 and 3.18 defined the structure preservation under this study which are therefore utilized to affirm Result 8 (Saris et al, 1978).

Definition 3.7 Given the model defined under Equations (3.8) as well as (3.11) and the residuals stated in Equation (3.12), then the residual estimators, $\hat{v}(\theta)$ and $\hat{\zeta}(\theta)$, defined by Equation (3.15) as well as Equation (3.16) are structure preserving whenever $\theta \in \Theta$ then the ensuing hold (Saris et al, 1978):

$$E[\hat{v}(\theta)\hat{v}(\theta)'] = \Sigma_{TJTJ} \tag{3.40}$$

$$E[\hat{\zeta}(\theta)\hat{\zeta}(\theta)'] = \Psi \tag{3.41}$$

$$E[\hat{v}(\theta)\hat{\zeta}(\theta)'] = 0. \tag{3.42}$$

Definition 3.8 Given the models defined under Equations (3.8) as well as (3.11) together with the residuals stated under Equation (3.12), then the residual estimators, $\hat{v}(\hat{\theta})$ and $\hat{\zeta}(\hat{\theta})$, are structure preserving whenever $\theta \in \Theta$ for the ensuing three conditions hold (Saris et al, 1978):

$$E[\hat{v}(\hat{\theta})\hat{v}(\hat{\theta})'] = \Sigma_{vv} \tag{3.43}$$

$$E[\hat{\zeta}(\hat{\theta})\hat{\zeta}(\hat{\theta})'] = \Psi \tag{3.44}$$

$$E[\hat{v}(\hat{\theta})\hat{\zeta}(\hat{\theta})'] = 0. \tag{3.45}$$

Result 5: Given the model defined under Equations (3.8) as well as (3.11) together with the stated residuals under Equations (3.12) and (3.13) when every $\theta \in \Theta$ then no residual estimators contained under Equations (3.15) and (3.17) holds under the conditions from Equations (3.37) to (3.39) regarding structure preservation but for the estimators considered under Equations (3.15) to (3.20) then the condition contained in Equation (3.34) holds if $\hat{v}(\theta) = \hat{v}(\theta)$ as well as $\hat{\zeta}(\theta) = \hat{\zeta}(\theta)$ (Saris et al, 1978).

Proof. The proof here comprises three phases. The initial two phases shows that no estimator under the class of Equations (3.12) and (3.13) holds for Equation (3.37) as well as Equation (3.39) respectively. Last but not least, is to demonstrate that the estimators considered under the condition in Equation (3.39) holds for $\hat{v}(\theta) = \hat{v}(\theta)$ as well as $\hat{\zeta}(\theta) = \hat{\zeta}(\theta)$.

Recollect that

$$\hat{v}(\theta) = (I - \Lambda W)z$$

$$v(\theta) = z - \Lambda L$$

$$\hat{\zeta}(\theta) = MWz$$

$$\zeta(\theta) = ML.$$

Thus, generally, the covariance matrices $[\hat{v}(\theta)\hat{v}(\theta)']$, $E[\hat{\zeta}(\theta)\hat{\zeta}(\theta)']$ as well as $E[\hat{v}(\theta)\hat{\zeta}(\theta)']$ would be written as:

$$E[\hat{v}(\theta)\hat{v}(\theta)'] = (I - \Lambda W)\Sigma_{zz}(I - \Lambda W)' \tag{3.46}$$

$$E[\hat{\zeta}(\theta)\hat{\zeta}(\theta)'] = MW\Sigma_{zz}W'M' \tag{3.47}$$

$$E[\hat{v}(\theta)\hat{\zeta}(\theta)'] = (I - \Lambda W)\Sigma_{zz}W'M'. \qquad (3.48)$$

To begin with, it is demonstrated below that no estimator under the class of Equation (3.8) holds for Equation (3.45). Thus, from Equation (3.46),

$$E[\hat{v}(\theta)\hat{v}(\theta)'] = (I - \Lambda W)\Sigma_{zz}(I - \Lambda W)'$$

$$= (I - \Lambda W)(\Lambda\Sigma_{LL}\Lambda' + \Sigma_{vv})(I - \Lambda W)'$$

$$= (I - \Lambda W)\Lambda\Sigma_{LL}\Lambda'(I - \Lambda W)' + (I - \Lambda W)\Sigma_{vv}(I - \Lambda W)'$$

when $\hat{v}(\theta)$ is structure preserving then it must satisfy every $\theta \in \Theta$ so that

$$\Sigma_{vv} = (I - \Lambda W)\Lambda\Sigma_{LL}\Lambda'(I - \Lambda W)' + (I - \Lambda W)\Sigma_{vv}(I - \Lambda W)' \quad (3.49)$$

Equation (3.49) is deemed satisfied only whenever $\Lambda W = I$ and $\Lambda W = 0$. Based on the presumptions of the parameter space Θ then both $\Lambda W = I$ and $\Lambda W = 0$ does not satisfy. Therefore, when every $\theta \in \Theta$ then there are no estimators $\hat{v}(\theta)$ under the class of Equation (3.15) holds for Equation (3.45) (Saris et al, 1978).

Secondly, it is demonstrated below that there are no estimators under the group in Equation (3.23) holds for Equation (3.46). From Equation (3.49),

$$E[\hat{\zeta}(\theta)\hat{\zeta}(\theta)'] = MW\Sigma_{zz}W'M'$$

$$= MW\Sigma_{zz}W'M'$$

$$= MW\Lambda\Sigma_{LL}\Lambda W'M' + MW\Sigma_{vv}M'W'..$$

When $\hat{\zeta}(\theta)$ shows structure preservation, then it has to satisfy whenever $\theta \in \Theta$ that

$$\Sigma_{\zeta\zeta} = MW\Lambda\Sigma_{LL}\Lambda W'M' + MW\Sigma_{vv}W'M'. \qquad (3.50)$$

Equation (3.50) satisfies only whenever $W\Lambda = I$ while $W = 0$. Based on the conditions of the parameter space Θ then $\Lambda W = I$ and $W = 0$ does not satisfy. Therefore, whenever

$\theta \in \Theta$ no residual estimators $\hat{\zeta}(\theta)$ within the group in Equation (3.16) holds for Equation (3.46).

Last but not least, we now show that regarding the estimators being considered when every $\theta \in \Theta$, Equation (3.47) holds if $\hat{v}(\theta) = \hat{v}(\theta)$ as well as $\hat{\zeta}(\theta) = \hat{\zeta}(\theta)$. Suppose $\hat{v}(\theta) = \hat{v}(\theta)$ and $\hat{\zeta}(\theta) = \hat{\zeta}(\theta)$ (Saris et al, 1978). Then base on Result 1 it is shown that $\Lambda W = I$. From Equation (3.50),

$$E[\hat{v}(\theta)\hat{\zeta}(\theta)'] = E[(I - \Lambda W)z([(I - B)^{-1} - \Gamma] \, Wz$$

$$= E[(I - I)zz'W'[(I - B)^{-1} - \Gamma$$

$$= 0.$$

Therefore, $\hat{v}(\theta)$ and $\hat{\zeta}(\theta)$ are structure preserving of the actual covariance structure between the observed errors and construct errors.

3.12.4 Univocality

Univocality criterion basically concerns the property of validity (i.e. how best residuals constructed associate with their counterpart true residuals they intend to measure) as well as invalidity (Heise & Bohrnstedt, 1970; McDonald & Burr, 1967). Generally it is preferable that both estimators, $\hat{v}(\theta)$ and $\hat{\zeta}(\theta)$, are univocal in order that they exhibit huge validity as well as smaller invalidity. Meanwhile, univocality is defined by Definition 3.12 as well as Definition 3.13 which are consequently utilized to arrive at Result 9 which is subsequently utilized to institute Heuristic Result 5.

Definition 3.9 Given the defined model under Equations (3.8) as well as (3.11) together with the stated residuals under Equation (3.12), then the estimators, $\hat{v}(\theta)$ and $\hat{\zeta}(\theta)$, are deemed univocal whenever every $\theta \in \Theta$ the conditions in Equations (3.51) – (3.54) are true:

$$E[v(\theta)\hat{v}(\theta)'] = \Delta_{v\hat{v}} \tag{3.51}$$

$$E[\zeta(\theta)\hat{\zeta}(\theta)'] = \Delta_{\zeta\hat{\zeta}} \tag{3.52}$$

$$E[v(\theta)\hat{\zeta}(\theta)'] = 0 \tag{3.53}$$

$$E[\zeta(\theta)\hat{v}(\theta)'] = 0 \tag{3.54}$$

for $\Delta_{v\hat{v}}$ as well as $\Delta_{\zeta\hat{\zeta}}$ are diagonal matrices (Heise & Bohrnstedt, 1970).

Definition 3.10 According to Heise and Bohrnstedt (1970) given the defined model under Equations (3.8) as well as (3.11) together with the stated residuals in Equation (3.12), then the estimators, $\hat{v}(\hat{\theta})$ as well as $\hat{\zeta}(\hat{\theta})$, are deemed univocal whenever $\theta \in \Theta$ the conditions in Equations (3.55) – (3.58) are true:

$$E[v(\theta)\hat{v}(\hat{\theta})'] = \Delta_{v\hat{v}}^* \tag{3.55}$$

$$E[\zeta(\theta)\hat{\zeta}(\hat{\theta})'] = \Delta_{\zeta\hat{\zeta}}^* \tag{3.56}$$

$$E[v(\theta)\hat{\zeta}(\hat{\theta})'] = 0 \tag{3.57}$$

$$E[\zeta(\theta)\hat{v}(\hat{\theta})'] = 0 \tag{3.58}$$

for $\Delta_{v\hat{v}}^*$ and $\Delta_{\zeta\hat{\zeta}}^*$ are diagonal matrices. As observed earlier, challenges arise during the derivation of the expected value regarding $\hat{\theta}$ as, generally, it does not have a close form solution (Schott, 2005). Therefore, as seen previously, derivation about $\hat{v}(\theta)$ and $\hat{\zeta}(\theta)$ would be utilized regarding their properties in order to specify Heuristic Result for $\hat{v}(\hat{\theta})$ and $\hat{\zeta}(\hat{\theta})$ (Heise & Bohrnstedt, 1970).

Result 6: Given the defined model under Equations (3.8) as well as (3.11) together with the stated residuals under Equations (3.12) and (3.13) whenever $\theta \in \Theta$ there exist no residual estimator within the group under Equations (3.15) and (3.16) satisfies the conditions from Equation (3.49) to Equation (3.52) to be univocal even when the estimators being considered in Equations (3.22) – (3.24) then the condition under Equation (3.58) hold whenever $\hat{v}(\theta) = \hat{v}_b(\theta)$ as well as $\hat{\zeta}(\theta) = \hat{\zeta}_b(\hat{\theta})$ (Heise & Bohrnstedt, 1970).

Proof. Here, the proof comprises of four phases. The initial three phases shows that whenever $\theta \in \Theta$ that there are no estimators from the group in Equations (3.22) – (3.27) holds for Equations (3.56) – (3.58). Last but not least, it would be proved that whenever $\theta \in \Theta$ then the estimators considered in Equations (3.22) – (3.24), the condition under Equation (3.59) holds for $\hat{v}(\theta) = \hat{v}(\theta)$.

Recollect

$$\hat{v}(\theta) = (I - \Lambda W)z$$

$$v(\theta) = z - \Lambda L$$

$$\hat{\zeta}(\theta) = MWz$$

$$\zeta(\theta) = ML$$

Thus, the covariance matrices $E[v(\theta)\hat{v}(\theta)']$, $E[\zeta(\theta)\hat{\zeta}(\theta)']$, $E[v(\theta)\hat{\zeta}(\theta)']$, as well as $E[\zeta(\theta)\hat{v}(\theta)']$ generally are given by:

$$E[v(\theta)\hat{v}(\theta)'] = \Sigma_{vv}(I - \Lambda W)' \tag{3.59}$$

$$E[\zeta(\theta)\hat{\zeta}(\theta)'] = M\Sigma_{LL}\Lambda'W'M' \tag{3.60}$$

$$E[v(\theta)\hat{\zeta}(\theta)'] = \Sigma_{vv}W'M' \tag{3.61}$$

$$E[\zeta(\theta)\hat{v}(\theta)'] = M\Sigma_{LL}\Lambda'(I - \Lambda W)' \tag{3.62}$$

75

Heuristic Result 5: Given the defined model under equations (3.8) as well as (3.11) together with the stated residuals under Equations (3.12) and (3.13), neither of these residual estimators in Equations $(3.22) - (3.27)$ hold for the conditions under Equations (3.64) to (3.66) regarding univocality but for the condition in Equation (3.66) which holds if $\hat{v}(\hat{\theta}) = \hat{v}(\hat{\theta})$ (Heise & Bohrnstedt, 1970).

Proof. As a result of MLE being consistent, both estimators, $\hat{v}(\hat{\theta})$ and $\hat{\zeta}(\hat{\theta})$, have same behavior as $\hat{v}(\theta)$ and $\hat{\zeta}(\theta)$ respectively as was noted regarding the proof in heuristic results earlier. From Result 9, it means that neither of these estimators are univocal but the condition under Equation (3.66) holds when utilizing the estimator $\hat{v}(\hat{\theta})$. On the contrary, it was indicated that the Bartlett's residual estimator is optimal under the univocal criteria based on the fact that the assumption underpinning orthogonal factor model was proved by McDonald and Burr (1967).

3.12.5 Distribution

There have not been any distribution assumptions, so far, made about either the construct variables or the error terms as well as the derivation of Results 1-9 or Heuristic Results 1-5 despite the fact that in real sense assumptions regarding distributions are imperative to calculate ML estimates. In making inference, however, about the true error through the estimates residual it becomes desirable to find the sampling or distribution of the residual. Due to the fact that both estimators, $\hat{v}(\theta)$ and $\hat{\zeta}(\theta)$, are function of z which makes it desirable to find the distribution of z. Recollect that $z = \Lambda L + v$ so that the distribution of z relies on the distribution of , v, as well as ζ. In theory, despite the fact that there is the possibility of choosing any distribution for, ζ, and v in practice the commonest way is by assuming that $L \sim \mathcal{N}(0, \Sigma_{LL})$, $v \sim \mathcal{N}(0, \Sigma_{vv})$, as well as $\zeta \sim$

76

$\mathcal{N}(0, \Psi)$ so that $z \sim \mathcal{N}(0, \Sigma_{zz})$. Based on normal assumption of Result 10 (Green, 2008).

Result 7: According to Green (2008), given the defined model under Equations (3.8) as well as (3.11) together with the stated residuals in Equations (3.12) and (3.13), when $z \sim \mathcal{N}(0, \Sigma_{zz})$ then whenever $\theta \in \Theta$ $\hat{v}(\theta) \sim \mathcal{N}(0, (I - \Lambda W)\Sigma_{zz}(I - \Lambda W$ and $\hat{\zeta}(\theta) \sim \mathcal{N}(0, MW\Sigma_{zz}W'M')$.

Proof. Greene (2008) established by a theory that when $x \sim \mathcal{N}(\mu, \Sigma)$ then $Ax + b \sim \mathcal{N}(A\mu + b, A\Sigma A')$. For this application suppose $\mu = 0$ and $\Sigma = \Sigma_{zz}$. Then for $\hat{v}(\theta)$ suppose $A = (I - \Lambda W)$ as well as $b = 0$. Consequently by Greene (2008) theory it implies that $\hat{v}(\theta) \sim \mathcal{N}(0, (I - \Lambda W)\Sigma_{zz}(I - \Lambda W))$. Now, for $\hat{\zeta}(\theta)$ suppose $A = MW$ and $b = 0$. Then by Greene (2008) theory $\hat{\zeta}(\theta) \sim \mathcal{N}(0, MW\Sigma_{zz}W'M')$.

Heuristic Result 4: Consider the defined model under Equations (3.8) as well as (3.11) together with the stated residuals in Equations (3.12) and (3.13), when $z \sim \mathcal{N}(0, \Sigma_{zz})$ then $\hat{v}(\hat{\theta}) \sim \mathcal{N}(0, (I - \Lambda W)\Sigma_{zz}(I - \Lambda W))$ as well as $\hat{\zeta}(\hat{\theta}) \sim \mathcal{N}(0, MW\Sigma_{zz}W'M')$ (Green, 2008).

Proof. Owing to the fact that MLE is applied to obtain items $\hat{\theta}$ it's anticipated that $\hat{v}(\hat{\theta})$ and $\hat{\zeta}(\hat{\theta})$ exhibit same behavior as $\hat{v}(\theta)$ and $\hat{\zeta}(\hat{\theta})$. By Result 10, it is established that $\hat{v}(\theta) \sim \mathcal{N}(0, (I - \Lambda W)\Sigma_{zz}(I - \Lambda W))$ as well as $\hat{\zeta}(\theta) \sim \mathcal{N}(0, MW\Sigma_{zz}W'M')$. Thus, $\hat{v}(\theta) \sim \mathcal{N}(0, (I - \Lambda W)\Sigma_{zz}(I - \Lambda W$ and $\hat{\zeta}(\hat{\theta}) \sim \mathcal{N}(0, MW\Sigma_{zz}W'M')$ for the notation $\sim$ connotes the approximate distribution.

The outcomes shows that based on the criteria used the optimal estimator can either be the regression estimator or the Bartlett's estimator. However, for no criteria was the Anderson-Rubin estimator optimal. This shows, at least, it is unreasonable option for

77

SEM. Thus, it is very possible to attribute it to the derivation of $\boldsymbol{W_{ar}}$ takes orthogonal factor model which cannot occur under the framework of SEM.

3.13 Asymptotic Properties of $\hat{\boldsymbol{v}}(\hat{\boldsymbol{\theta}})$ and $\hat{\boldsymbol{\zeta}}(\hat{\boldsymbol{\theta}})$

So far the derivations of the properties of $\hat{v}(\hat{\theta})$ and $\hat{\zeta}(\hat{\theta})$ are deemed to have finite sample properties. Also, since SEM generally considers huge samples, it is worth deriving the asymptotic properties. The following properties are worth considering when derivation are to be made about the asymptotic properties regarding these estimators (Grice, 2001; Wackwitz & Horn, 1971).

1. Consistency: This concerns whether an estimator converges to what it is estimating when the size of the sample becomes infinite. On the other hand, it concerns the asymptotic precision of the estimator. It is preferable when an estimator is consistent.

2. Efficiency: This concerns the asymptotic accuracy of the estimator. Whenever an estimator achieves Cramér-Rao lower bound then it is deemed efficient.

3. Asymptotic distribution: This concerns the examination of the distribution of the estimator regarding its convergence to when the sample size approaches infinite. It is relevant to derive given that it is the foundation for testing hypothesis regarding residuals as well as diagnostics.

4. Asymptotic structure preservation: This concept concerns the examination of the impact the selection of an estimator has regarding the associations among the error terms when the size of sample approaches infinite. Generally, the asymptotic covariance structure preferably should be the same as their counterparts of true residuals.

5. Asymptotic univocality: This refers to the degree of association between estimated residuals and their counterpart true residuals (i.e. asymptotic validity) as well as their noncorresponding residuals when the size of the sample becomes infinite. An estimator, preferable, should lead to estimates that associate only with their counterpart true residuals as well as not associate with noncorresponding residuals asymptotically.

The asymptotic properties stated above would be defined much broader in the rest of the section. Contrary to the finite sample properties of $\hat{v}(\hat{\theta})$ and $\hat{\zeta}(\hat{\theta})$, the possibility of deriving the asymptotic properties of $\hat{v}(\hat{\theta})$ and $\hat{\zeta}(\hat{\theta})$ without haven to make heuristic arguments by relying on $\hat{v}(\theta)$ and $\hat{\zeta}(\theta)$ properties (Grice, 2001; Wackwitz & Horn, 1971).

Despite the fact that finite sample properties about factor scores estimator were derived in earlier studies (Lawley and Maxwell, 1971; McDonald and Burr, 1967; Saris et al., 1978; Tucker, 1971) as well as examined via simulation setups (Grice, 2001; Wackwitz & Horn, 1971) and practical applications (Horn, 1965; Horn & Miller, 1966; Moseley & Klett, 1964), a more thorough literature survey shows that no similar work has been conducted concerning factor score estimators in terms of their asymptotic properties. Therefore, the optimal residual estimator would not be compared to the optimal factor score estimator whenever an assessment of the estimators been done through the same property.

3.13.1 Consistency

As done in regression analysis, residuals by themselves can be employed as an avenue of assessing assumptions underlying models. In such a case, it then becomes imperative

to consider the consistency of the estimators, $\hat{v}(\hat{\theta})$ and $\hat{\zeta}(\hat{\theta})$. Consistency, informally, of these istimators means that when the sample size increases and gets closer to infinity, $\hat{v}(\hat{\theta})$ and $\hat{\zeta}(\hat{\theta})$ get arbitrary closer to or converges to $v(\theta)$ and $\zeta(\theta)$ their true residuals. The following provide the underpinning definitions that establishes consistency (Lehman & Casella, 1998; Casella & Berger, 2003).

Definition 3.11 Given a defined model under Equations (3.8) as well as (3.11) together with the stated residuals contained in Equation (3.12), then $\hat{v}(\hat{\theta})$ is deemed to be consistent estimator of $v(\theta)$ whenever $\varepsilon > 0$ while for all $\theta \in \Theta$, (Lehman & Casella, 1998; Casella & Berger, 2003)

$$\lim_{n\to\infty} P_\theta \ (|\hat{v}(\hat{\theta}) - v(\theta)| \geq \varepsilon) = 0.$$

Definition 3.12 Given the defined model under Equations (3.8) as well as (3.11) together with the stated residuals contained in Equation (3.13), then $\hat{\zeta}(\hat{\theta})$ is said to be consistent estimator of $\zeta(\theta)$ whenever $\varepsilon > 0$ and for all $\theta \in \Theta$, (Lehman & Casella, 1998; Casella & Berger, 2003)

$$\lim_{n\to\infty} P_\theta \ (|\hat{\zeta}(\hat{\theta}) - \zeta(\theta)| \geq \varepsilon) = 0.$$

Theorem 3.1 Suppose $X_1, X_2, \ldots, X_n$ be iid $f(x|\theta)$, and suppose $L(\theta|x) = \Pi_{i=1}^n f(x_i|\theta)$ represent the likelihood function. Suppose $\hat{\theta}$ connotes the MLE of θ. Further suppose $\tau(\theta)$ be a continuous function of θ. Base on the regularity conditions in Miscellanea 10.6.2 on $f(x|\theta)$ and, thus, $(\theta|x)$, whenever $\theta \in \Theta$, (Casella & Berger, 2002).

$$\lim_{n\to\infty} P_\theta \ (|\tau(\hat{\theta}) - \tau(\theta)| \geq 0) = 0.$$

for $\tau(\hat{\theta})$ is consistent estimator of (θ).

Theorem 3.2 When W_n is a sequence of estimators for a parameter θ which holds for

    i. $\lim\limits_{n\to\infty} Var_\theta W_n = 0,$

    ii. $\lim\limits_{n\to\infty} Bias_\theta W_n = 0$

whenever $\theta \in \Theta$, thus W_n is a consistent sequence of estimators of θ.

By Definitions 3.14 and 3.15 as well as Theorem 3.1 and Theorem 3.2, then it establishes the consistency of these estimators $\hat{v}(\hat{\theta})$ and $\hat{\zeta}(\hat{\theta})$.

Result 8: Given a defined model under the Equations (3.8) as well as (3.11) together with the stated residuals in Equation (3.12), whenever $\in \Theta$, of the estimators being considered under Equations $(3.22) - (3.24)$, then the Bartlett's method residual estimator, $\hat{v}(\hat{\theta})$, defined under Equation (3.23) is said to be consistent estimator of the true observed errors $v(\theta)$ (Casella and Berger, 2002).

Proof. The proof is made up of two phases. For the first phase;

Recollect

$$\hat{v}(\hat{\theta}) = (I - \hat{\Lambda}\hat{W})z$$

$$\hat{v}(\theta) = (I - \Lambda W)z$$

$$v(\theta) = z - \Lambda L.$$

Firstly, it would be proved that the estimator $\hat{v}(\hat{\theta})$ converges to the estimator $\hat{v}(\theta)$ irrespective of the selection of $\boldsymbol{W}$. As $\hat{\theta} = \hat{\theta}$, $\hat{v}(\theta)$ is a function of θ and from important regularity conditions is satisfied, from Theorem 4.1 it can be expressed as

$$\lim\limits_{n\to\infty} P_\theta \left(|\hat{v}(\hat{\theta}) - \hat{v}(\theta)| \geq \varepsilon\right) = 0$$

where whenever $\theta \in \Theta$. Therefore the estimator $\hat{v}(\hat{\theta})$ converges to the estimator $\hat{v}(\theta)$.

81

Secondly, it would be proved that $\hat{v}(\theta)$ is a consistent estimator of $v(\theta)$ if $\boldsymbol{W} = \boldsymbol{W_b}$.

Hence becomes relevant to prove that

$$\lim_{n\to\infty} P_\theta \ (|\hat{v}(\theta) - v(\theta)| \geq \varepsilon) = 0 \qquad (3.61)$$

is satisfies if $\boldsymbol{W} = \boldsymbol{W_b}$. From the Chebyshev's Inequality,

$$\lim_{n\to\infty} P_\theta \ (|\hat{v}(\theta) - v(\theta)| \geq \varepsilon) \leq \lim_{n\to\infty} \frac{E(\hat{v}(\theta) - v(\theta))^2}{\varepsilon^2}$$

indicating that whenever $\theta \in \Theta$

$$\lim_{n\to\infty} E_\theta \ [(\hat{v}(\theta) - v(\theta))^2] = 0$$

thus $\hat{v}(\theta)$ is a consistent estimator of $v(\theta)$. Therefore an examination of the behavior of $\lim_{n\to\infty} E \ [(\hat{v}(\theta) - v(\theta))^2]$, the MSE of $\hat{v}(\theta)$, to find out the what conditions $\hat{v}(\theta)$ is a consistent estimator of $v(\theta)$. Moreover, by breaking down the MSE of an estimator and Theorem 3.2 (Casella and Berger, 2002),

$$\lim_{n\to\infty} E \ (\hat{v}(\theta) - v(\theta))^2 = \lim_{n\to\infty} \text{Var} \ (\hat{v}(\theta)) + \lim_{n\to\infty} \ [\text{Bias} \ (\hat{v}(\theta))]^2 \qquad (3.67)$$

so that $\hat{v}(\theta)$ is a consistent estimator of $v(\theta)$ whenever $\theta \in \Theta$

1. $\lim_{n\to\infty} V \, ar(\hat{v}(\theta)) = 0$

2. $\lim_{n\to\infty} \text{Bias} \ (\hat{v}(\theta)) = 0.$

To begin with, the limiting variance is considered:

$$\lim_{n\to\infty} \text{Var} \ (\hat{v}(\theta)) = \lim_{n\to\infty} \ \text{Var} \ ((I - \Lambda W)z)$$

$$= \lim_{n\to\infty} (I - \Lambda W)\Sigma_{zz}(I - \Lambda W)'$$

$$= (I - \Lambda W)\Sigma_{zz}(I - \Lambda W)' \qquad (3.68)$$

Therefore, based on the parameter space Θ assumptions, $\lim_{n \to \infty} Var(\hat{v}(\theta)) = 0$ if $\Lambda W = I$ or $W = (\Lambda'\Lambda)^{-1}\Lambda$ (Lehman & Casella, 1998; Casella & Berger, 2003). Neither of the weight matrices under Equations $(3.19) - (3.21)$ holds for the condition $W = (\Lambda'\Lambda)^{-1}\Lambda$. On the contrary, by Result 1 it is established that $\Lambda W = I$ if $\boldsymbol{W = W_b}$. Therefore, $\lim_{n \to \infty} Var(\hat{v}(\theta)) = 0$ regarding the Bartlett's method estimator (Casella and Berger, 2002).

Now the limiting bias is considered:

$$\lim_{n \to \infty} Bias(\hat{v}(\theta)) = \lim_{n \to \infty} E\left[\hat{v}(\theta) - v(\theta)\right]$$

$$= \lim_{n \to \infty} E\left[(I - \Lambda W)z - (z - \Lambda L)\right]$$

$$= \lim_{n \to \infty} (E[(z - \Lambda W z - z + \Lambda L)]$$

$$= \lim_{n \to \infty} (\Lambda W E(z) + \Lambda E(L)]$$

$$= \lim_{n \to \infty} (0)$$

$$= 0$$

Which satisfies every $\theta \in \Theta$ since it is supposed that $E(v) = E(L) = E(z) = z$. Therefore $\lim_{n \to \infty} Bias(\hat{v}(\theta)) = 0$ for every estimator considered. The outcome together with that of the limiting variance indicates that $\hat{v}(\theta)$ is a consistent estimator of (θ). Merging the two phases of the proof clearly shows that $\hat{v}(\hat{\theta})$ is a consistent estimator of $v(\theta)$ when only the Bartlett's method residual estimator is used (Casella and Berger, 2002).

3.13.2 Efficiency, Asymptotic and Limiting Variance, and Asymptotic Distribution

Despite the fact that consistency is key since it examines the asymptotic accuracy of estimators, it disregards the asymptotic precision of estimators. Therefore, this section

concerns the asymptotic precision as assessed by the asymptotic variance of the estimators. As previous assumed that $\hat{\theta}$ is attained via MLE, it becomes possible to adopt its properties to narrate that of the properties of efficiency as well as asymptotic distribution at the same time as asymptotic variance (Lehman & Casella, 1998; Casella & Berger, 2003).

Theorem 3.3 Suppose $X_1, X_2, \ldots,$ is iid $f(x|\theta)$, given that $\hat{\theta}$ connotes the MLE of θ, while $\tau(\theta)$ is a continuous function of θ (Casella and Berger, 2002; Lehman and Casella, 1998). Base on the regularity conditions stated earlier assumptions $F1 - F8$,

$$\sqrt{n}[\tau(\hat{\theta}) - \tau(\theta)] \to \mathcal{N}(0, \kappa(\theta))$$

for $\kappa(\theta)$ is a Cramér-Rao lower bound which yields

$$\kappa(\theta) = \frac{\partial \tau(\theta)}{\partial \theta} [I(\theta)]^{-1} \frac{\partial \tau(\theta)}{\partial \theta} T$$

for

$$I_{j,k} = E[\frac{d}{d\theta_j} \log f(x; \theta) \frac{d}{d\theta_k} \log f(x; \theta)]$$

while $\partial \tau(\theta)$ is the Jacobian matrix. Where $\tau(\hat{\theta})$ is consistent and asymptotically efficient estimator of $\tau(\theta)$ (Casella and Berger, 2002; Lehman and Casella, 1998).

Result 8 can easily established by Theorem 3.3

Result 9: Given defined model under Equations (3.8) as well as (3.11) together with the stated residuals in Equation (3.12), when $\theta \in \Theta$, the estimator $\hat{v}(\hat{\theta})$ converges to the estimator $\hat{v}(\theta)$ so that the estimator $\hat{v}(\hat{\theta})$ is asymptotically distributed normally (Lehman & Casella, 1998; Casella & Berger, 2003).

Proof. Recollect that

$$\hat{v}(\hat{\theta}) = (I - \hat{\Lambda}\hat{W})z$$

$$\hat{v}(\theta) = (I - \Lambda W)z$$

Directly following from Theorem 3.3 since $\hat{\theta} = \hat{\theta}$, $\hat{v}(\hat{\theta})$ is a continuous function of θ, while the key regularity conditions are satisfied. Thus, from Theorem 3.3 it is established that

$$\sqrt{n}[\hat{v}(\hat{\theta}) - \hat{v}(\theta)] \to \mathcal{N}(0, \kappa_v(\theta))$$

for $\kappa_v(\theta)$ is the Cramér-Rao lower bound. Based on these (i.e. the theorems) it implies that $\hat{v}(\hat{\theta})$ converges to the estimator $\hat{v}(\theta)$, and it is efficient, while asymptotically normally distributed (Casella and Berger, 2002; Lehman and Casella, 1998).

Definition 3.13 According to Casella and Berger, (2002), given two estimators W_n and V_n will hold for

$$\sqrt{n}[W_n - \tau(\theta)] \to \mathcal{N}(0, \sigma_W^2) \; \sqrt{n}[V_n - \tau(\theta)] \to \mathcal{N}(0, \sigma_V^2)$$

the distribution, then the asymptotic relative efficiency (ARE) of V_n for W_n is expressed as

$$ARE(V_n, \; W_n) = \frac{\sigma_W^2}{\sigma_V^2}$$

For purposes of applications, W_n and V_n implies a unique item in the vector of estimators under $\hat{v}(\hat{\theta})$ or $\hat{\zeta}(\hat{\theta})$ whereas $\tau(\theta)$ under Definition 3.13 implies the analogous unique item is either in $\hat{v}(\theta)$ or $\hat{\zeta}(\theta)$.

3.13.3 Asymptotic Structure Preservation

This concept concerns the effect of a given estimator on the asymptotic covariance structure regarding the residuals. As seen under structure preservations for a finite property, for some applications it is important that asymptotic relationships amidst error terms need not be changed as a result of the selection of an estimator.

Definition 3.14 Given a model under Equations (3.8) and (3.11), these residual estimators, $\hat{v}(\hat{\theta})$ and $\hat{v}(\hat{\theta})$, as stated under Equation (3.17) as well as Equation (3.18) are both asymptotically structure preserving when every $\theta \in \Theta$ then the conditions below are true (Casella and Berger, 2002):

$$\lim_{n\to\infty} E\left[\hat{v}(\hat{\theta})\hat{v}(\hat{\theta})'\right] = \Sigma_{vv} \qquad (3.69)$$

$$\lim_{n\to\infty} E\left[\hat{\zeta}(\hat{\theta})\hat{\zeta}(\hat{\theta})'\right] = \Psi \qquad (3.70)$$

$$\lim_{n\to\infty} E\left[\hat{v}(\hat{\theta})\hat{\zeta}(\hat{\theta})'\right] = 0. \qquad (3.71)$$

Thus none of these estimators is optimal regarding structure preservation for finite sample property.

3.13.4 Asymptotic Univocality

This concept concerns the properties the properties of asymptotic validity as well as asymptotic invalidity. This property determines how the estimated residuals associates with their analogous true residual counterparts (asymptotic validity) as well as their non-analogous true residuals (invalidity) while the sample size approaches infinity.

Definition 3.15 Given a model under Equation (3.8) as well as Equation (3.11) these residual estimators $\hat{v}(\hat{\theta})$ and $\hat{\zeta}(\hat{\theta})$ as stated in Equations (3.17) and (3.18) are both asymptotically univocal when every $\theta \in \Theta$ then the four conditions below are true (Casella and Berger, 2002):

$$\lim_{n\to\infty} E[v(\theta)\hat{v}(\hat{\theta})'] = \Delta_{v\hat{v}} \qquad (3.72)$$

$$\lim_{n\to\infty} E[\zeta(\theta)\hat{\zeta}(\hat{\theta})'] = \Delta_{\zeta\hat{\zeta}} \qquad (3.73)$$

$$\lim_{n\to\infty} E[v(\theta)\hat{\zeta}(\hat{\theta})'] = 0 \qquad (3.74)$$

$$\lim_{n\to\infty} E[\zeta(\theta)\hat{v}(\hat{\theta})'] = 0 \qquad (3.75)$$

for $\Delta_{v\hat{v}}$ as well as $\Delta_{\zeta\hat{\zeta}}$ are matrices in diagonal form.

UNIVERSITY FOR DEVELOPMENT STUDIES

87

CHAPTER FOUR

ESTIMATION MAXIMIZATION

4.1 Introduction

This chapter looks into the concept of estimation maximization and its application to structural equation modelling. The split estimation procedure, the E-step and the M-step, are examined here alongside relevant derivations and demonstrations.

4.2 A Simplified Model

But in order to avoid heavy formulas in the development of the algorithm, we shall use in the sequel with no loss of generality, a simplified model involving $p = 2$ explanatory blocks X^1 and X^2. The corresponding equation set for a given unit i, reads:

$$\begin{cases} y_i' & = t_i'D + g_ib' + \varepsilon_{iy'} \\ x_i^{1'} & = t_i^{1'}D^1 + f_i^1a^{1'} + \varepsilon_{i^{1'}} \\ x_i^{2'} & = t_i^{2'}D^2 + f_i^2a^{2'} + \varepsilon_{i^{2'}} \\ g_i & = f_i^1c^1 + f_i^2c^2 + \varepsilon_{ig} \end{cases} \qquad (4.1)$$

Such as $\theta = \{D, D^1, D^2, b, a^1, a^2, c^1, c^2, \sigma_Y^2, \sigma_1^2, \sigma_2^2\}$. Thus, in this case (cf. (2)), the dimension of θ is:

$$K = 5 + q_Y(r_T + 1) + \sum_{m=1}^{2} q_m (r_m + 1)$$

4.3 Estimation Using the EM Algorithm

In this work, likelihood maximization is carried out via an iterative EM algorithm (Dempster et al, 1977). Each iteration of the algorithm involves an Expectation (E) - step followed by a Maximization (M) - step. Dempster et al (1977) prove that the EM algorithm yields maximum likelihood estimates. Moreover, they proved that even if the

starting point is one where the likelihood is not convex, if an instance of the algorithm converges, it will converge to a (local) maximum of the likelihood. Another major advantage of the EM algorithm is that it can be used to "estimate" missing data. Thus if we consider LV s as missing data the EM algorithm will prove a general technique to maximize the likelihood of statistical models with LV's but also to estimate these LV's. In our SEM framework LV s correspond to factors. Thus, we will be able to estimate the factors at unit-level. We shall present the algorithm on the simplified model with no loss of generality (Dempster et al, 1977).

4.4 The EM algorithm

Let $z = (y, x^1, x^2)$ be the OV's and $h = (g, f^1, f^2)$ the LV's. The EM algorithm is based on the log-likelihood associated with the complete data (z, h).

4.4.1 The complete log-likelihood function

Let $p(z, h; \theta)$ denote the probability density of the complete data. The corresponding log-likelihood function is:

$$\mathcal{L}(\theta; z, h) = -\frac{1}{2} \sum_{i=1}^{n} \{ ln|\psi_Y| + ln|\psi_1| + ln|\psi_2|$$

$$+(y_i - D't_i - g_i b)'\psi_Y^{-1}(y_i - D't_i - g_i b)$$

$$+(x_i^1 - D^{1'}t_i^1 - f_i^1 a^1)'\psi_1^{-1}(x_i^1 - D^{1'}t_i^1 - f_i^1 a^1)$$

$$+(x_i^2 - D^{2'}t_i^2 - f_i^2 a^2)'\psi_2^{-1}(x_i^2 - D^{2'}t_i^2 - f_i^2 a^2)$$

$$+(g_i - c^1 f_i^1 - c^2 f_i^2)^2 + (f_i^1)^2 + (f_i^2)^2\} + \lambda$$

Where θ is the K-dimensional set of model parameters and λ, a constant. However, because of the simplification made in the section 2.4, in our case $=$ $\{D, D^1, D^2, b, a^1, a^2, c^1, c^2, \sigma_Y^2, \sigma_1^2, \sigma_2^2\}$. Indeed, $\psi_Y = \sigma_Y^2 Id_{q_{Y7}} \psi_1 = \sigma_1^2 Id_{q_1}$ and $\psi_2 = \sigma_2^2 Id_{q_2}$ (Foulley, 2002).

4.4.2 Estimation in SEM

To maximize this function, in the framework of EM algorithm, we have to solve (Foulley, 2002):

$$E_z^h[\frac{\partial}{\partial\theta}\mathcal{L}(\theta; z, h)] = 0 \qquad (4.2)$$

This demands that we know the derivatives of the log-likelihood function and the distribution $p_{z_i}^{h_i}$ of h_i conditional on z_i for each observation $i \in [1, n] \cdot$ Let us introduce the following notation (Foulley, 2002):

$$p_{z_i}^{h_i} = \mathcal{N}(M_i = \begin{pmatrix} m_{1i} \\ m_{2i} \\ m_{3i} \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix})$$

$$\overline{g_i} = E_{z_i}^{h_i}[g_i] = m_{1i}; \overline{\gamma_i} = E_{z_i}^{h_i}[g_i^2] = (E_{z_i}^{h_i}[g_i])^2 + V_{z_i}^{h_i}[g_i] = m_{1_i^2} + \sigma_{11}$$

$$f_i^{\overline{1}} = E_{z_i}^{h_i}[f_i^1] = m_{2i)} \varphi_i^{\overline{1}} = E_{z_i}^{h_i}[(f_i^1)^2] = (E_{z_i}^{h_i}[f_i^1])^2 + V_{z_i}^{h_i}[f_i^1] = m_{2_i^2} + \sigma_{22}$$

$$f_i^{\overline{2}} = E_{z_i}^{h_i}[f_i^2] = m_{3i)} \varphi_i^{\overline{2}} = E_{z_i^i}^{h}[(f_i^2)^2] = (E_{z_i}^{h_i}[f_i^2])^2 + V_{z_i}^{h_i}[f_i^2] = m_{3_i^2} + \sigma_{33}$$

For all $\overline{\overline{\xi}} \in \{\overline{g}, f^{\overline{1}}, f^{\overline{2}}, \overline{\gamma}, \varphi^{\overline{1}}, \varphi^{\overline{2}}\}7$ we denote $\overline{\overline{\xi}} = (\overline{\overline{\xi}}_i)_{i=1,...,n} \in \mathbb{R}^n$.

The parameters of the Gaussian distribution $p_{z_i}^{h_i}$ are explicit and have the following form (Foulley, 2002):

$$M_i = \Sigma_2^* \Sigma_3^{*-1} \mu^* \text{ and } \Sigma = \Sigma_1^* - \Sigma_2^* \Sigma_3^{*-1} \Sigma_2^{*\prime} \text{ where:}$$

$$\Sigma_1^* = \begin{pmatrix} (c^1)^2 + (c^2)^2 + 1 & c^1 & c^2 \\ c^1 & 1 & 0 \\ c^2 & 0 & 1 \end{pmatrix}$$

90

$$\Sigma_2^* = \begin{pmatrix} ((c^1)^2 + (c^2)^2 + 1)b' & c^1 a^{1'} & c^2 a^{2'} \\ c^1 b' & a^{1'} & 0_{(1,q_2)} \\ c^2 b' & 0_{(1,q_1)} & a^{2'} \end{pmatrix}$$

$$\Sigma_3^* = \begin{pmatrix} ((c^1)^2 + (c^2)^2 + 1)bb' + \Psi_Y & c^1 ba^{1'} & c^2 ba^{2'} \\ c^1 a^1 b' & a^1 a^{1'} + \Psi_1 & 0_{(q_1,q_2)} \\ c^2 a^2 b' & 0_{(q_2,q_1)} & a^2 a^{2'} + \Psi_2 \end{pmatrix}$$

$$\mu_i^* = \begin{pmatrix} D' y_i - t_i \\ -x_i^1 D^{1'} t_i^1 \\ -x_i^2 D^{2'} t_i^2 \end{pmatrix}$$

These results are demonstrated in section 4.6.3. Expressions of the first-order derivatives of $\mathcal{L}$ with respect to θ are also established in section 4.6.4 and written in the following forms with $m \in \{1,2\}$ (Foulley, 2002):

$$\begin{cases} \dfrac{\partial}{\partial D} \mathcal{L}(z, h) & = \sum_{i=1}^{n} \psi_Y^{-1} (y_i - D' t_i - g_i b) t_i' \\[2mm] \dfrac{\partial}{\partial D^{m'}} \mathcal{L}(z, h) & = \sum_{i=1}^{n} \psi_m^{-1} (x_i^m - D^{m'} t_i^m - f_i^m a^m) t_i^{m'} \\[2mm] \dfrac{\partial}{\partial b} \mathcal{L}(z, h) & = \sum_{i=1}^{n} g_i \psi_Y^{-1} (y_i - D' t_i - g_i b) \\[2mm] \dfrac{\partial}{\partial a^m} \mathcal{L}(z, h) & = \sum_{i=1}^{n} f_i^m \psi_m^{-1} (x_i^m - D^{m'} t_i^m - f_i^m a^m) \\[2mm] \dfrac{\partial}{\partial c^m} \mathcal{L}(z, h) & = \sum_{i=1}^{n} f_i^m (g_i - c^2 f_i^2 - c^1 f_i^1) \\[2mm] \dfrac{\partial}{\partial \sigma_Y^2} \mathcal{L}(z, h) & = n q_Y \sigma_Y^{-2} - \sigma_Y^{-4} \sum_{i=1}^{n} ||y_i - D' t_i - g_i b||^2 \\[2mm] \dfrac{\partial}{\partial \sigma_m^2} \mathcal{L}(z, h) & = n q_m \sigma_m^{-2} - \sigma_m^{-4} \sum_{i=1}^{n} ||x_i^m - D^{m'} t_i^m - f_i^m a^m||^2 \end{cases}$$

$$(4.3)$$

So, here formula (4.3) develops into:

$$
\begin{cases}
\sum_{i=1}^{n}(y_i - D't_i - \overline{g_i}b)t_i' & = 0 \\[2mm]
\sum_{i=1}^{n}(x_i^m - D^{m'}t_i^m - \overline{f_i^m}a^m)t_i^{m'} & = 0 \\[2mm]
\sum_{i=1}^{n}\overline{g_i}\,y_i - \overline{g_i}D't_i - \overline{\gamma_i}b & = 0 \\[2mm]
\sum_{i=1}^{n}\overline{f_i^m}\,x_i^m - \overline{f_i^m}D^{m'}t_i^m - \overline{\varphi_i^m}a^m & = 0 \\[2mm]
\sum_{i=1}^{n}\sigma_{12} + f_i^{\overline{1}}\overline{g_i} - c^2\sigma_{23} - c^2 f_i^{\overline{1}}f_i^{\overline{2}} - \varphi_i^{\overline{1}}c^1 & = 0 \\[2mm]
\sum_{i=1}^{n}\sigma_{31} + f_i^{\overline{2}}\overline{g_i} - c^2\varphi_i^{\overline{2}} - c^1\sigma_{32} - c^1 f_i^{\overline{1}}f_i^{\overline{2}} & = 0 \\[2mm]
nq_Y\sigma_Y^{-2} - \sigma_Y^{-4}\sum_{i=1}^{n}||y_i - D't_i||^2 + ||b||^2\overline{\gamma_i} - 2(y_i - D't_i)'\overline{g_i}b & = 0 \\[2mm]
nq_m\sigma_m^{-2} - \sigma_m^{-4}\sum_{i=1}^{n}||x_i^m - D^{m'}t_i^m||^2 + ||a^m||^2\overline{\varphi_i^m} - 2(x_i^m - D^{m'}t_i^m)'\overline{f_i^m}a^m & = 0
\end{cases}
$$

$$(4.4)$$

System of equations (4.4) is easy to solve and the obtained solutions will be given in the next section (Foulley, 2002).

4.4.3 Results

The explicit solution of the system (4.3) and also of (4.4) is the following (Myriam, 2015):

$$
\hat{b} = \frac{(\overline{\overline{g}y} - \overline{yt'})(\overline{tt'})^{-1}\overline{\overline{g}t}}{=,\gamma - \overline{\overline{g}}t'(\overline{tt'})^{-1}\overline{\overline{g}t}}
$$

$$
\widehat{a^m} = \frac{f^m x^m - \overline{x^m t^{m'}}(\overline{t^m t^{m'}})^{-1}f^m t^m}{\overline{\overline{\varphi^m}} - \overline{f^m}t^{m/}(\overline{t^m t^{m/}})^{-1}\overline{\overline{f^m}}t^m}
$$

92

$$\widehat{c^1} = \underline{(\sigma_{12} + f^{\overline{1}}\overline{g}} - - - -$$

$$\varphi^1\varphi^2 - (\sigma_{23} + f^1f^2)^2$$

$$\widehat{c^2} = - - - -$$

$$\varphi^1\varphi^2 - (\sigma_{23} + f^1f^2)^2$$

$$\widehat{D'} = (\overline{yt'} - \hat{b}\overline{\overline{g}t'})(\overline{tt'})^{-1}$$

$$D^{m'} = (\overline{x^m t^{m'}} - \widehat{a^m}f^m t^{m'})(\overline{t^m t^{m'}})^{-1}$$

$$\widehat{\sigma_Y^2} = \frac{1}{nq_Y}\sum_{i=1}^{n}\{||y_i - \widehat{D'}t_i||^2 + ||\hat{b}||^2\overline{\gamma_i} - 2(y_i - \widehat{D'}t_i)'\hat{b}\overline{g_i}\}$$

$$\widehat{\sigma_m^2} = \frac{1}{nq_m}\sum_{i=1}^{n}\{||x_i^m - \overline{D^{m'}}t_i^m||^2 + ||\widehat{a^m}||^2\overline{\varphi_i^m} - 2(x_i^m - \overline{D^{m'}}t_i^m)'\widehat{a^m}\overline{f_i^m}\}$$

4.4.4 The Algorithm

To estimate parameters in θ, we propose the following EM-algorithm. We denote $[t]$ the t^{ieth}-iteration of the algorithm.

1. Initialization = choice of the initial parameter values $\theta^{[0]}$.

In the initialization step, $\forall m \in \{1, p\}$ we propose to obtain $D^{m[0]}$ by multiple linear regression between X^m and T^m. Then, to initialize the others, we compute each approximated factor $\overline{f^m}[0]$ and $\overline{g}$ as first principal component of a PCA of $X^m - T^m D^{m[0]}$ and $Y - TD^{[0]}$. Thus, we initialize a^m, σ_m^2 (resp. , σ_y^2) by multiple linear regression between $X^m - T^m D^{m[0]}$ and $\overline{f^m}[0]$ (resp. between $Y - TD^{[0]}$ and $\overline{g}$). Finally, each $c^{m[0]}$ can be obtained by multiple linear regression between $\overline{g}$ and $\sum_{m=1}^{p}\overline{f^m}[0]$. In practice we use the functions $lm()$ and $PCA\ ()$ derived from the package *FactoMineR* Husson et al, (2008).

2. Current iteration $t \geq 1_7$ until stopping condition is met:

(a) E-step: with $\theta^{[t-1]}$,

i. Calculate explicitly distribution $p_{z_i}^{h_i}$ for each $i \in [1, nI \cdot$

ii. Estimate the factor-values $\overline{g}, \overline{f^m}[t], m \in \{1,2\}$.

iii. Calculate $\overline{\gamma}$ and $\overline{\varphi^m}, m[t] \in \{1,2\}$.

(b) M-step:

i. Update θ to $\theta^{[t]}$ by injecting $\overline{g}, \overline{\gamma}$ and $\overline{f^m}[t], \overline{\varphi^m}[t], m \in$

$\{1,2\}$ into the formulas in (7).

3. We used the following stopping condition with the smallest ε possible:

$$\sum_{k=1}^{K} \frac{|\theta^{*[t+1]}[k] - \theta^{*[t]}[k]|}{\theta^{*[t+1]}[k]} < \varepsilon \tag{4.5}$$

where θ^* is the K-dimensional vector containing the scalar values in all parameters in θ.

4.5 Numerical Results on Simulated Data

4.5.1 Data Generation

We consider $n = 400$ units and $q_Y = q_1 = q_2 = 40$. Therefore7 the 120 OV sY, X^1, X^2 are simulated so as to be structured respectively around three factors g, f^1, f^2. ‡ $factors\ f^1$ and f^2 are explanatory of g. Besides, we consider $r_T = r_1 = r_2 = 2$ i.e 2 covariates are simulated for each covariate matrix T, T^1 and T^2. The data is simulated as follows (Myriam, 2015).

1. Choice of θ:

(a) $D = D^1 = D^2 a)$ matrices filled in row-wise with the ordered

integer sequence ranging from 1 to 80 (indeed: $r_T * q_Y = r_1 * q_1 =$

$r_2 * q_2 = 2 * 40)$.

(b) $b = a^1 = a^2 =$ ordered integer sequence ranging from 1 to 40.

(c) $c^1 = c^2 = 1$

(d) $\sigma_Y^2 = \sigma_1^2 = \sigma_2^2 = 1$

2. Simulation of factors g, f^1, f^2

(a) Simulate vectors f^1 and f^2 of $n = 400$ normally distributed random numbers with mean 0 and variance 1 (abbreviated $\forall m, \in$

$$\{1,2\} f^m \sim \mathcal{N}(0, Id_{400}).$$

(b) We simulate ε^g according to distribution $\varepsilon^g \sim \mathcal{N}(0, Id_{400})$.

(c) We then calculate g as $g = f^1 c^1 + f^2 c^2 + \varepsilon^g$

3. Simulation of noises $\varepsilon^Y, \varepsilon^1, \varepsilon^2$ Each element of matrix ε^Y, (respectively $\varepsilon^1, \varepsilon^2$) is simulated independently from distribution $\mathcal{N}(0, \sigma_Y^2 = 1)$ (respectively, $\sigma_1^2 = 1$, $\sigma_2^2 = 1$).

4. Simulation of covariate matrices $T, \ T^1, T^2$

Each element of matrices $T, \ T^1, T^2$ is simulated according to the standard normal distribution.

5. Construction of $Y, X^1, X^2 Y, X^1, X^2$ are eventually calculated through formulas in the model (1).

This simulation scheme was performed 100 times, each time yielding a set of simulated data matrices (Y, X^1, X^2). Then for each simulated data7 we ran an estimation routine with a threshold value $\varepsilon = 10^{-2}$, yielding the average results presented in section 4.2. Thus from $400 * 120 = 48000$ scalar elements of data, we will estimate $3 * n = 1200$ scalar elements of factors plus $K = 5 + 3 * 40(2 + 1) = 365$ scalar parameters, i.e 1565 scalars (Myriam, 2015).

4.5.2 Results

Convergence was observed in almost all cases in less than five iterations. We assess the quality of the estimations as follows (Myriam, 2015).

- On the one hand, we calculate the absolute relative deviation between each simulated scalar parameter in θ^* and its estimation, and then average these deviations over the 100 simulations. We then produce a box-plot of the average absolute relative deviations (cf. fig. 4.1). This makes the interpretation easier since we only need to look at the box- plot's values and check that they are positive (because of the absolute value) and close to 0.

- On the other hand, to assess the quality of the factor estimations, we compute the 300 values of square correlations between the simulated concatenated factors (g, f^1, f^2) (respectively) and the corresponding estimations $((\overline{g}, f^{\overline{1}}, f^2))$.

The median of square correlations is 0.998 the first quartile is 0.997 and the third quartile is 0.999. So, factor g (respectively f^1 and f^2) turn out to be drawn towards the principal direction underlying the bundles made up by observed variables Y (respectively X^1 and X^2). Now we may legitimately wonder how the quality of estimations could be affected by the number of observations and the number of OV s in each block. In the following section we give a sensitivity analysis performed to investigate this (Myriam, 2015).

Subsequently, a performance of sensibility analysis on the simulated data presented in section 4.5.1 can be done. The purpose was to study the influence of the block-sizes (n, q_Y, q_1, q_2). On the quality of estimation, both of the parameters and the factors. To simplify the analysis, we imposed $q_Y = q_1 = q_2$ and varied n and q separately, i.e.

studied the cases n = 50; 100; 200; 400 with $q = 40$ and $q = 5$; 10; 20; 40 with n = 400. Each case was simulated 100 times. Therefore, we simulated 800 data-sets (Myriam, 2015).

To sum things up, the sample size n proved to have more impact on the quality of parameter estimation and factor reconstruction than the number of OV s. Now, the quality of factor reconstruction remains high for rather small values of n or q. We advise to use a minimal sample size of $n = 100$ to obtain really stable structural coefficients. Above this threshold n has but little impact on the biases and standard deviations of estimations.

4.5.3 An application to Environmental Data

4.5.3.1 Data Presentation

Mortier et al, (2014) applied their model to the data-set *genus* provided in the R-package SCGLR. Data-set *genus* was built from the CoForChange database. It gives the abundances of 27 common tree genera present in the tropical moist forest of the Congo-Basin, and the measurements of 40 geo-referenced environmental variables, for $n = 1000$ inventory plots (observations). Some of the geo-referenced environmental variables describe 16 physical factors pertaining to topography geology and rainfall description. The remaining variables characterize vegetation through the enhanced vegetation index (EVI) measured on 16 dates.

In this section, Mortier et al, (2014) modeled the tree abundances from the other variables while reducing the dimension of data. The dependent block of variables Y therefore consists of the $q_Y = 27$ tree species counts divided by the plot-surface. A PCA of the geo-referenced environmental variables and the photosynthetic activity variables

97

confirms that EVI measures are clearly separated from the other variables. Indeed, it shows two variable-bundles with almost orthogonal central directions. This justifies using our model (section 4.2) with $p = 2$ explanatory groups one of them (X^1) gathering $q_1 = 16$ rainfall measures and location variables (longitude, latitude and altitude), and the second one (X^2), the $q_2 = 23$ EVI measures. Besides, in view of the importance of the geological substrate on the spatial distribution of tree species in the Congo Basin showed by Fayolle et al, (2012). They chose to put nominal variable *geology* in a block T. This block therefore contains constant 1 plus all the indicator variables of geology but one, which will be the reference value. Geology having 5 levels T has thus 5 columns.

4.6 Model with Geologic Covariates

4.6.1 Model Specification

Here is the model used with the variable-blocks designed in section 4.2:

$$\begin{cases} Y & = TD + gb' + \varepsilon^Y \\ X^1 & = L_n d^{1'} + f^1 a^{1'} + \varepsilon^1 \\ X^2 & = L_n d^{2'} + f^2 a^{2'} + \varepsilon^2 \\ g & = f^1 c^1 + f^2 c^2 + \varepsilon^g \end{cases}$$

Where $n = 1000$, $q_Y = 27$, $q_1 = 16$, $q_2 = 23$ and $r_T = 5$. The first row of D is a parameter vector that contains the means of the Y's noted $D[1,]$ in Table 4.1, and the other rows, the overall effects of the geological substrates with respect to the reference one. Indeed, the next section presents the model's parameter-estimations where in Table 4.1 each row r of D is noted $D[r,]$ (Fayolle et al, 2012).

4.6.2 Results

With a threshold value $\varepsilon = 10^{-3}$, convergence was reached after 58 iterations. Some

parameter-estimations are presented in Tables 4.1 and 4.2.

Table 4.1: Application to the *genus* data with geologic covariate: estimations of

parameters D' and b'

Variables	Parameter estimates					
	D[1,]	D[2,]	D[3,]	D[4,]	D[5,]	b'
gen1	0.76	0.16	0.06	0.68	-0.12	-0.13
gen2	0.54	-0.28	-0.03	-0.03	-0.28	0.47
gen3	0.41	-0.23	-0.02	0.25	-0.37	0.29
gen4	0.12	-0.14	0.03	0.52	0.30	0.09
gen5	0.31	-0.15	0.19	-020	0.84	0.09
gen6	0.55	-0.12	-0.26	0.06	-0.02	0.14
gen7	0.46	0.06	-0.04	-0.37	0.43	0.14
gen8	0.55	0.04	-0.09	-0.16	0.04	0.42
⋮	⋮	⋮	⋮	⋮	⋮	⋮
gen27	0.27	0.41	0.69	-0.24	0.56	0.08

Table 4.2: Application to the *genus* data with geologic covariate: estimations of

parameters $d^{1'}$ and $a^{1'}$

Variables	Parameter estimates	
	$d^{1'}$	$a^{1'}$
altitude	4.43	0.62
pluvio_1	44.45	0.16
pluvio_2	2.48	-0.91
pluvio_3	4.32	-0.88
pluvio_4	9.65	-0.47
pluvio_5	8.56	-0.28
pluvio_6	6.68	0.26
pluvio_7	5.98	0.83
pluvio_8	4.78	0.81
⋮	⋮	⋮
lat	2.49	0.92

It can be seen in Tables 4.1 and 4.2 that for certain species, the geologic substrate seems to be of great importance (e.g. for genl, gen5, gen7, gen9, gen12, gen16, gen21, gen25, gen26, gen27), whereas for others, it only has a small impact on the abundances (e.g. for $gen2, gen6, gen8$, genl 10, genl 8, $gen20$, gen23). Moreover, Table 4.1 shows that only these are well accounted for by our model. Although we have carried out the analysis with variables *gen2*, *gen3*, *gen8*, *genlO*, *genll*, *genl5*, *genl 7*, *gen23*, *gen24* and *gen25*, the results are practically the same when we take all variables.

4.6.3 Calculation of the complete data log-likelihood function $\mathcal{L}$

Proof

Consider the case $p = 2$, $\psi_Y = \sigma_Y^2 Id_{q_Y}$, $\psi_1 = \sigma_1^2 Id_{q_1}$ and $\psi_2 = \sigma_2^2 Id_{q_{27}}$ and for observation i, the model is formulated as follows (Tami et al, 2015):

$$\begin{cases} y_i' & = t_i'D + g_i b' + \varepsilon_{i y'} \\ x_i^{1'} & = t_i^{1'} D^1 + f_i^1 a^{1'} + \varepsilon_{i^{1'}} \\ x_i^{2'} & = t_i^{2'} D^2 + f_i^2 a^{2'} + \varepsilon_{i^{2'}} \\ g_i & = f_i^1 c^1 + f_i^2 c^2 + \varepsilon_{i g} \end{cases}$$

We have,

$$p(z_i, h_i; \theta) = p(y_i, x_i^1, x_i^2, g_i, f_i^1, f_i^2; \theta)$$

$$= p(y_i, x_i^1, x_i^2 | g_i, f_i^1, f_i^2; \theta) p(g_i, f_i^1, f_i^2; \theta)$$

$$= p(y_i, x_i^1, x_i^2 | g_i, f_i^1, f_i^2; \theta) p(g_i | f_i^1, f_i^2; \theta) p(f_i^1, f_i^2; \theta)$$

$$= p(y_i, x_i^1, x_i^2 | g_i, f_i^1, f_i^2; \theta) p(g_i | f_i^1, f_i^2; \theta) p(f_i^1; \theta) p(f_i^2; \theta)$$

$$= p(y_i, x_i^1, x_i^2 | g_i, f_i^1, f_i^2; \theta) p(g_i | f_i^1, f_i^2; \theta) p(f_i^1) p(f_i^2)$$

$$= p(x_i^1, x_i^2 | y_i, g_i, f_i^1, f_i^2; \theta) p(y_i | g_i, f_i^1, f_i^2; \theta) p(g_i | f_i^1, f_i^2; \theta) p(f_i^1) p(f_i^2)$$

$$= p(x_i^1, x_i^2 | f_i^1, f_i^2; \theta) p(y_i | g_i; \theta) p(g_i | f_i^1, f_i^2; \theta) p(f_i^1) p(f_i^2)$$

$$= p(x_i^1 | x_i^2, f_i^1, f_i^2; \theta) p(x_i^2 | f_i^1, f_i^2; \theta) p(y_i | g_i; \theta) p(g_i | f_i^1, f_i^2; \theta) p(f_i^1) p(f_i^2)$$

$$= p(x_i^1|f_i^1;\theta)p(x_i^2|f_i^2;\theta)p(y_i|g_i;\theta)p(g_i|f_i^1,f_i^2;\theta)p(f_i^1)p(f_i^2)$$

Where $\theta = \{D, D^1, D^2, b, a^1, a^2, c^1, c^2, \psi_Y, \psi_1, \psi_2\}$ is the set of model parameters (Myriam, 2015). Therefore,

$$\mathcal{L}(\theta; z_i, h_i) = \mathcal{L}(\theta; x_i^1|f_i^1) + \mathcal{L}(\theta; x_i^2|f_i^2) + \mathcal{L}(\theta; y_i|g_i) + \mathcal{L}(\theta; g_i|f_i^1, f_i^2) + \mathcal{L}(f_i^1)$$
$$+ \mathcal{L}(f_i^2)$$

Because of the model and the normal distribution properties we obtain: $x_i^m|f_i^m \sim$ $\mathcal{N}(t_i^{m\prime}D^m + f_i^m a^{m\prime}, \psi_{X^m})$

$$y_i|g_i \sim \mathcal{N}(t_i'D + g_i b', \psi_Y)$$

$$g_i|f_i^1, f_i^2 \sim \mathcal{N}(f_i^1 c^1 + f_i^2 c^2, 1)$$

$$f_i^m \sim \mathcal{N}(0,1)$$

Then, we obtain the complete data log-likelihood function (Tami et al, 2015):

$$\mathcal{L}(\theta; z, h) = -\frac{1}{2}\sum_{i=1}^{n}\{ln|\psi_Y| + ln|\psi_1| + ln|\psi_2|$$

$$+(y_i - D't_i - g_i b)'\psi_Y^{-1}(y_i - D't_i - g_i b)$$

$$+(x_i^1 - D^{1\prime}t_i^1 - f_i^1 a^1)'\psi_1^{-1}(x_i^1 - D^{1\prime}t_i^1 - f_i^1 a^1)$$

$$+(x_i^2 - D^{2\prime}t_i^2 - f_i^2 a^2)'\psi_2^{-1}(x_i^2 - D^{2\prime}t_i^2 - f_i^2 a^2)$$

$$+(g_i - c^1 f_i^1 - c^2 f_i^2)^2 + (f_i^1)^2 + (f_i^2)^2\} + \lambda$$

Where λ a constant. Also, the set of model parameters

$\theta = \{D, D^1, D^2, b, a^1, a^2, c^1, c^2, \psi_Y, \psi_1, \psi_2\}$ in this case corresponds to

$\theta = \{D, D^1, D^2, b, a^1, a^2, c^1, c^2, \sigma_Y^2, \sigma_1^2, \sigma_2^2\}$ because of the simplification made earlier.

Indeed, $\psi_Y = \sigma_Y^2 Id_{q_Y}, \psi_1 = \sigma_1^2 Id_{q_1}$ and $\psi_2 = \sigma_2^2 Id_{q_2}$.

Therefore, we can also write the complete data log-likelihood function (Tami et al, 2015):

$$\mathcal{L}(\theta; z, h) = -\frac{1}{2}\sum_{i=1}^{n}\{q_Y ln(\sigma_Y^2) + q_1 ln(\sigma_1^2) + q_2 ln(\sigma_2^2)$$

$$+\sigma^{-2_Y}(y_i - D't_i - g_i b)'(y_i - D't_i - g_i b)$$

$$+\sigma^{-2_1}(x_i^1 - D^{1'}t_i^1 - f_i^1 a^1)'(x_i^1 - D^{1'}t_i^1 - f_i^1 a^1)$$

$$+\sigma^{-2_2}(x_i^2 - D^{2'}t_i^2 - f_i^2 a^2)'(x_i^2 - D^{2'}t_i^2 - f_i^2 a^2)$$

$$+(g_i - c^1 f_i^1 - c^2 f_i^2)^2 + (f_i^1)^2 + (f_i^2)^2\} + \lambda$$

4.6.4 Demonstration of the normality of the distribution of $h_i|z_i$

Proof

Consider the case $p = 2$, $\psi_Y = \sigma_Y^2 Id_{q_Y}$, $\psi_1 = \sigma_1^2 Id_{q_1}$ and $\psi_2 = \sigma_2^2 Id_{q_{27}}$ and for

observation i_7 the model is formulated as follows:

$$\begin{cases} y_i' &= t_i'D + g_i b' + \varepsilon_{iy'} \\ x_i^{1'} &= t_i^{1'}D^1 + f_i^1 a^{1'} + \varepsilon_{i1'} \\ x_i^{2'} &= t_i^{2'}D^2 + f_i^2 a^{2'} + \varepsilon_{i2'} \\ g_i &= f_i^1 c^1 + f_i^2 c^2 + \varepsilon_{ig} \end{cases}$$

To prove the normality of the distribution of $h_i|z_i$, we use the classical result about the

conditioning of normally distributed variables (Tami et al, 2015). Before using this result,

we calculate the joint distribution of $(g_i, f_i^1, f_i^2, y_i, x_i^1, x_i^2)$.

We know that for observation i,

$$y_i \sim \mathcal{N}(D't_i, b((c^1)^2 + (c^2)^2 + 1)b' + \Psi_Y)$$

$$x_i^m \sim \mathcal{N}(D^{m'}t_i^m, a^m a^{m'} + \Psi_m)$$

$$g_i \sim \mathcal{N}(0, (c^1)^2 + (c^2)^2 + 1)$$

$$f_i^m \sim \mathcal{N}(0,1)$$

102

Then, after computing the required covariances, the following was obtain,

$$(g_i, f_i^1, f_i^2) \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} (c^1)^2 + (c^2)^2 + 1 & c^1 & c^2 \\ c^1 & 1 & 0 \\ c^2 & 0 & 1 \end{pmatrix}\right)$$

And,

$$(y_i, x_i^1, x_i^2)$$

$$\sim \mathcal{N}\left(\begin{pmatrix} D't_i \\ D^{1'}t_i^1 \\ D^{2'}t_i^2 \end{pmatrix}, \begin{pmatrix} ((c^1)^2 + (c^2)^2 + 1)bb' + \Psi_Y & c^1 ba^{1'} & c^2 ba^{2'} \\ c^1 a^1 b' & a^1 a^{1'} + \Psi_1 & 0_{(q_1, q_2)} \\ c^2 a^2 b' & 0_{(q_2, q_1)} & a^2 a^{2'} + \Psi_2 \end{pmatrix}\right)$$

If two variables X_1 and X_2 are normally distributed such that (Myriam, 2015),

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}\left(\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right)$$

where, $\mu_1(r \times 1), \mu_2(s \times 1), \Sigma_{11}(r \times r), \Sigma_{12}(r \times s), \Sigma_{21}(s \times r)$ and $\Sigma_{22}(s \times s)$; then,

$$(X_1 | X_2 = x_2) \sim \mathcal{N}(M = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \varphi = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}) \quad (9)$$

Then, after compute the required covariances we obtain the joint distribution (Tami et al, 2015),

$$(g_i, f_i^1, f_i^2, y_i, x_i^1, x_i^2) \sim \mathcal{N}(M_i^*, \Sigma^*) \text{ such as,}$$

$$M_i^* = \begin{pmatrix} 0_{(3,1)} \\ D't_i \\ D^{1'}t_i^1 \\ D^{2'}t_i^2 \end{pmatrix} \text{ and } \Sigma^* = (^{\Sigma}\Sigma_2^* 1 *, \Sigma_3^{*)}. \text{ Where,}$$

$$\Sigma_1^* = \begin{pmatrix} (c^1)^2 + (c^2)^2 + 1 & c^1 & c^2 \\ c^1 & 1 & 0 \\ c^2 & 0 & 1 \end{pmatrix}$$

$$\Sigma_2^* = \begin{pmatrix} ((c^1)^2 + (c^2)^2 + 1)b' & c^1 a^{1'} & c^2 a^{2'} \\ c^1 b' & a^{1'} & 0_{(1,q_2)} \\ c^2 b' & 0_{(1,q_1)} & a^{2'} \end{pmatrix}$$

$$\Sigma_3^* = \begin{pmatrix} ((c^1)^2 + (c^2)^2 + 1)bb' + \Psi_Y & c^1 ba^{1'} & c^2 ba^{2'} \\ c^1 a^1 b' & a^1 a^{1'} + \Psi_1 & 0_{(q_1, q_2)} \\ c^2 a^2 b' & 0_{(q_2, q_1)} & a^2 a^{2'} + \Psi_2 \end{pmatrix}$$

Finally, we obtain the distribution, $h_i | z_i \sim \mathcal{N}(M_i, \Sigma)$ where, $M_i = \Sigma_2^* \Sigma_3^{*-1} \mu_i^*$

and $\Sigma = \Sigma_1^* - \Sigma_2^* \Sigma_3^{*-1} \Sigma_{27}^*{}'$ such that $\mu_i^* = \begin{pmatrix} D' y_i - t_i \\ -x_i^1 D^{1'} t_i^1 \\ -x_i^2 D^{2'} t_i^2 \end{pmatrix}$

4.6.5 Calculation of the first-order derivatives of $\mathcal{L}$

Proof.

We search the first-order derivatives of the complete data log-likelihood function (Tami

et al, 2015): $\qquad \mathcal{L}(\theta; z, h) = -\frac{1}{2} \sum_{i=1}^{n} \{ ln|\psi_Y| + ln|\psi_1| + ln|\psi_2|$

$$+ (y_i - D' t_i - g_i b)' \psi_Y^{-1} (y_i - D' t_i - g_i b)$$

$$+ (x_i^1 - D^{1'} t_i^1 - f_i^1 a^1)' \psi_1^{-1} (x_i^1 - D^{1'} t_i^1 - f_i^1 a^1)$$

$$+ (x_i^2 - D^{2'} t_i^2 - f_i^2 a^2)' \psi_2^{-1} (x_i^2 - D^{2'} t_i^2 - f_i^2 a^2)$$

$$+ (g_i - c^1 f_i^1 - c^2 f_i^2)^2 + (f_i^1)^2 + (f_i^2)^2 \} + \lambda$$

Where λ constant7 $\theta = \{D, D^1, D^2, b, a^1, a^2, c^1, c^2, \psi_Y, \psi_1, \psi_2\}$7 $\psi_Y = \sigma_Y^2 Id_{q_{Y7}}$ $\psi_1 = \sigma_1^2 Id_{q_1}$ and $\psi_2 = \sigma_2^2 Id_{q_2}$.

Therefore, there are matrix-parameters (D, D^1, D^2) vector-parameters (b, a^1, a^2) and scalar parameters $(c^1, c^2, \sigma_Y^2, \sigma_1^2, \sigma_2^2)$. Then, $\mathcal{L}$ is a sum of three types of functions: the logarithm, the square function and a quadratic form function $(w - X\beta)' \Gamma (w - X\beta)$7 where Γ is symmetric and $w(q \times 1)$, $X(q \times m)$, β $(m \times 1)$ and $\Gamma(q \times q)$. The first-order derivatives of the logarithm function and the square function are in our case trivial. The first-order derivative of $(w - X\beta)' \Gamma (w - X\beta)$ by X is less trivial but necessary. Let

us start by making explicit the first-order derivative of $(w - X\beta)'\Gamma(w - X\beta)$ with respect to X (Tami et al, 2015).

$$d_X[(w - X\beta)'\Gamma(w - X\beta)] = (w - X\beta)'\Gamma(-dX\beta) + (-dX\beta)'\Gamma(w - X\beta)$$

$$= -2(w - X\beta)'\Gamma(dX\beta)$$

$$= tr[-2(w - X\beta)'\Gamma(dX\beta)]$$

$$= tr[-2\beta(w - X\beta)'\Gamma dX]$$

$$= < -2\beta(w - X\beta)'\Gamma | dX >$$

Therefore,

$$\frac{d}{dX}[(w - X\beta)'\Gamma(w - X\beta)] = (-2\beta(w - X\beta)'\Gamma)'$$

$$= -2(\beta(w - X\beta)'\Gamma)'$$

$$= -2\Gamma(w - X\beta)\beta'$$

Likewise, we establish that:

$$\frac{\partial}{\partial D'}\mathcal{L}(z, h) = \sum_{i=1}^{n} \psi_Y^{-1}(y_i - D't_i - g_ib)t_i'$$

Similar reasoning can be applied to D^m and allows to obtain the second row of (5). Concerning the third and the fourth row of (5), we use the classical result:

$$\frac{\partial}{\partial \beta}[(w - X\beta)'\Gamma(w - X\beta)] = -2X'\Gamma(w - X\beta)$$

Eventually, the fifth, the sixth and the eighth rows of (5) are obtained in a trivial way (Tami et al, 2015).

Suppose we have a model for the complete data Y, with associated density $f(Y/\theta)$, where $\theta = (\theta_1, ..., \theta_d)$ is the unknown parameter. We write $Y = (Y_{obs}, Y_{mis})$, where Y_{obs} represents the observed part of Y and Y_{mis} denotes the missing values. The EM algorithm

finds the value of θ, θ^* that maximizes $f\left(Y_{obs}\,/\,\theta\right)$, that is, the MLE for θ based on the

observed data Y_{obs}.

The EM algorithm starts with an initial value $\theta^{(0)}$. Letting $\theta^{(t)}$ be the estimate θ at the

i^{th} iteration, iteration (t +1) of EM is as follows

E step: Find the expected complete-data log-likelihood if θ were $\theta^{(t)}$:

$$Q\left(\theta\,/\,\theta^{(t)}\right)=\int L\left(\theta\,/\,Y\right)f\left(Y_{mis}\,/\,Y_{obs},\theta=\theta^{(t)}\right)dY_{mis} \tag{4.6}$$

Where $L\left(\theta\,/\,Y\right)=\log f\left(Y\,/\,\theta\right)$

M step: Determine $\theta^{(t+1)}$ by maximizing this expected log-likelihood:

$$Q\left(\theta^{(t+1)}\,/\,\theta^{(t)}\right)\geq Q\left(\theta\,/\,\theta^{(t)}\right) \text{ for all } \theta \tag{4.7}$$

The M step of EM algorithm is easy to implement in broad classes of problems, such as

in exponential families, since it uses the identical computational method as ML

estimation from $L\left(\theta\,/\,Y\right)$. The E step of EM algorithm is also very easy to implement in

many problems, including many exponential family models, since it follows from

standard complete-data theory for means of conditional distributions.

4.7 Simulation Setup

In order to demonstrate the earlier derivations regarding some of the finite sample

properties (theoretical), a simulation setup was utilized to examine these properties

empirically. The setup, in this study, examined two main properties comprising

univocality as well as conditional bias. For starters, the model utilized under the

simulation setup was described and consequently an outline of how the study was

106

conducted. Subsequently, the outcome the simulation setup was explained and interrogated under chapter five.

4.7.1 Model

The construct (latent) variable model comprised two variables, ξ_1 as well as η_1, which yields

$$\eta_1 = 0.6\xi_1 + \zeta_1. \tag{4.8}$$

Related to this model were the covariance matrices Φ as well as Ψ. Suppose $\Phi = [1]$ as well as $\Psi = [1]$.

Now, the manifest model indicates that every construct (latent) variable contains three indicators while every indicator associated with just a factor. Based on the manifest model, the equations are expressed as

$$x_1 = 1.0\xi_1 + \delta_1, y_1 = 1.0\eta_1 + \varepsilon_1$$
$$x_2 = 0.8\xi_1 + \delta_2, y_2 = 0.8\eta_1 + \varepsilon_2$$
$$x_3 = 0.8\xi_1 + \delta_3, y_3 = 0.8\eta_1 + \varepsilon_3 \tag{4.9}$$

Related to the manifest model is the covariance matrix of manifest errors, Σ_{vv}. Suppose the manifest errors were not associated with every other so that $\Sigma_{vv} = I$.

Thus these models in Equation (4.8) as well as Equation (4.9) mirrors a basic model that can be utilized applied research.

4.7.2 Design of the Study

Simulation of this study was performed through SAS 9.2/JMP 13 by the steps outlined below.

1. By PROC IML,

   i. N values are simulated regarding the random variable $v_i(\theta)$ so that v_i (which is iid) is $\mathcal{N}(0,1)$.

   ii. N values are simulated regarding the random variable $\zeta_i(\theta)$ so that ζ_i (which is iid) is $\mathcal{N}(0,1)$.

   iii. N values are simulated regarding the random variable $\xi_i(\theta)$ so that ξ_i (which is iid) is $\mathcal{N}(0,1)$.

   iv. N values are subsequently generated regarding η_{1i} base on Equation (4.8).

   v. N values are subsequently generated regarding the random variables y_i as well as N values for the random variables x_i by utilizing the Equations under (4.9).

2. Relying on the data set generated in the first step (i.e. step 1), the model is estimated via PROC CALIS to obtain $\hat{\theta}$. When the model estimated in step 2 converges (i.e. the sample from step 1 results in a solution that converges in 100 or fewer iterations) but does not result in any Heywood cases, then calculate $\hat{v}(\theta)$ using PROC IML through Equation (4.8), $\hat{\zeta}(\theta)$ through Equation (3.16) (the estimated residuals assuming the elements of θ are known), $\hat{v}(\hat{\theta})$ through Equation (3.17), as well as $\hat{\zeta}(\hat{\theta})$ by Equation (3.19) (the estimated residuals assuming the elements of θ are unknown). However, when the model estimated under step 2 does not converge or yield Heywood case, then the data set would be disregarded while a fresh data set is simulated through step 1. Again, the following statistics are computed through the calculated residuals in step 2;

108

i. The average conditional bias if θ is established (known): Given observation i, the conditional bias is computed through Equation (3.29) as well as Equation (3.32) for each weighted matrix being considered by utilizing the known values of θ. Given every residual the average of the conditional bias is then computed.

ii. The average conditional bias if θ is not established (unknown): Given observation i, the conditional bias is computed by substituting θ with $\hat{\theta}$ under Equation (3.29) as well as Equation (3.32). The conditional bias is then computed for every weighted matrix being considered. Given every residual, the average of the conditional bias is computed.

iii. The correlation between the true residuals, the estimated residuals if θ is established (known), as well as the residuals estimated if θ is not known: Given every residual and residual estimator then the correlations between: (1) the true residuals obtained in Step 1 as well as the estimated residuals under Equation (3.15) as well as Equation (3.16) are computed; (2) the true residuals obtained under step 1 as well as the estimated residuals under Equation (3.17) and Equation (3.18) are computed; and (3) the estimated residuals through Equation (3.15) and Equation (3.16) as well as the estimated residuals by Equation (3.17) and Equation (3.18) are computed.

It worth noting that the process above utilized three dissimilar sample sizes, consisting of $250, 500$, as well as 1000. The samples utilized were meant to mirror samples that are often encountered in real sense. A sample 250 mirrors a usual sample encountered in many empirical studies and again what is considered a reasonable sample size by utilizing the q rule propounded by Jackson (2003). Using this rule, it implies that the ratio of sample size N to the quantity of parameters q must preferably be at least 10 with at least

20 being optimal. For this study, there were $q = 13$ parameters showing that the smallest

sample must be at least 130. Thus a sample size of $N = 200$ is considerably above the

smallest figure while being close to the preferable sample value of 260. Moreover a

sample of $N = 500$ is deemed to be reasonable for purposes of practical sense regarding

average sample type of researches that utilize SEM whereas a sample of $N = 1000$

mirrors a typical threshold often considered for huge sized researches.

It is worth noting that for every sample considered a total of 500 data points were

generated which resulted in convergence without Heywood scenarios. Again, in order to

CHAPTER FIVE

RESULTS AND DISCUSSIONS

5.1 Introduction

This section looks at the results obtained from the simulations and analysis of data for this study. The results obtained are based on the objectives of the study and this includes a tabulation of correlations bothering on univocal asymptotic property, the graphical display of detection of outliers and influential observations and finally the comparative analysis between the EM concept of residual estimation and other known methods of residual estimators. Subsequently, the findings arrived based on the results were juxtapose against the findings made in the literature under the discussion here.

5.2 Preliminary Analysis

Before the results from the simulation is presented, it is worth noting that data set created in order to comprehend the descriptive statistics of the data. Estimated error terms were obtained from Eqn 3.8 and Eqn 3.9 through known values of θ. Similarly, the estimated error terms from Eqn 3.10 and Eqn 3.11 were obtained through $\hat{\theta}$. The descriptive statistics (the mean, standard deviation, middle value, minimum and maximum values) of the actual residuals computed for both the unknown and known parameters respectively.

Table 5.1: Descriptive Statistics

Parameter	Mean	Standard deviation	Minimum	Median	Maximum
δ_1	-0.0577	0.9981	-4.0559	-0.0632	3.5487
δ_2	0.0580	0.9551	-2.7719	0.0219	3.2744
δ_3	0.0588	0.9587	-3.6771	0.0362	3.0963
$\hat{\delta}_{1(R)}$	-0.0464	0.8732	-2.7285	-0.0289	2.2709
$\hat{\delta}_{2(R)}$	0.0667	0.9377	-2.5370	0.0777	2.8124
$\hat{\delta}_{3(R)}$	0.0674	0.9596	-2.8723	0.0608	2.9448
$\hat{\delta}_{1(B)}$	-0.0678	0.7453	-2.6444	-0.0495	1.6537
$\hat{\delta}_{2(B)}$	0.0664	0.9357	-2.5360	0.0731	2.8014
$\hat{\delta}_{3(B)}$	0.0515	0.8537	-2.3207	-0.0259	2.9458
$\hat{\delta}_{1(AR)}$	-0.0255	0.7145	-2.0476	0.0212	1.7308
$\hat{\delta}_{2(AR)}$	0.0825	0.9919	-2.8877	0.0375	2.7625
$\hat{\delta}_{3(AR)}$	0.0833	0.9746	-2.8833	0.0524	2.9632
ε_1	0.0556	1.0066	-2.7458	0.0520	2.8788
ε_2	0.0280	1.0326	-2.7819	0.0519	3.2644
ε_3	0.0487	0.9387	-3.5871	0.0252	3.0853
$\hat{\varepsilon}_{1(R)}$	0.0245	0.6973	-2.1678	0.0281	2.7914
$\hat{\varepsilon}_{2(R)}$	0.0056	0.8367	-2.4270	0.0357	3.7114
$\hat{\varepsilon}_{3(R)}$	0.0016	0.8696	-2.7922	0.0201	3.8348
$\hat{\varepsilon}_{1(B)}$	0.0120	0.6652	-2.0943	0.0194	2.1947
$\hat{\varepsilon}_{2(B)}$	-0.0076	0.9164	-2.280	0.0511	3.1140
$\hat{\varepsilon}_{3(B)}$	-0.0116	0.8287	-2.6807	-0.0149	3.6357
$\hat{\varepsilon}_{1(AR)}$	0.0397	0.6755	-2.5382	0.1114	2.4505
$\hat{\varepsilon}_{2(AR)}$	0.0168	0.9573	-2.5397	0.05764	3.4815
$\hat{\varepsilon}_{3(AR)}$	0.0136	0.8824	-2.7455	0.0406	3.2994
ζ	0.0174	1.0171	-2.8458	-0.0158	3.1320
$\zeta_{(R)}$	0.0540	0.9498	-2.6644	0.0199	2.7844
$\zeta_{(B)}$	0.0687	1.4324	-4.2364	0.0522	4.8513

$\zeta_{(AR)}$	0.0454	1.1663	-3.5385	0.0239	4.1907
$\hat{\zeta}_{(R)}$	0.0665	0.9367	-2.8276	0.0477	2.9110
$\hat{\zeta}_{(B)}$	0.0832	1.4237	-4.3107	0.0190	4.6052
$\hat{\zeta}_{(AR)}$	0.0568	1.0996	-3.3556	0.0485	3.9430

It can be seen from Table 5.1 above that the mean of the actual error, say for δ_1, was -0.0577. This was not significantly dissimilar to mean of the estimated errors obtained from the Regression, Bartllet's and Anderson-Rubin methods which were -0.0464, -0.0678 and -0.0255 respectively. It, thus, can be noted that the measurement error (δ_1) associated with x yields negative means which were generally very close. Again, the mean (0.0580) of the actual error (δ_2) did not vary remarkably from the estimated error obtained from the Regression, Bartllet's and Andrson-Rubin which were 0.0667, 0.0664 and 0.0825 respectively. Thus, the measurement error (δ_2) for x provided positive means which were generally close. The same observation could be made for the actual error δ_3.

It was revealing to note that the standard deviation for the actual error (δ_1) was 0.9981 which was higher than its counterparts δ_2 and δ_3 which recorded 0.9551 and 0.9587 respectively. However, the standard deviation for the estimated error δ_1 obtained from the Regression estimator was lower than its counterparts obtained from both Bartllet's and Anderson-Rubin methods. Generally, it can be noted the actual error δ_3 and its equivalent estimated errors obtained from the Regression, Bartllet's and Anderson-Rubin produced higher values across all the descriptive statistics.

Now, the mean of the actual error (ε_1) did not differ overwhelmingly from the estimated errors obtained from the Regression, Bartllet's and Anderson-Rubin methods which yielded 0.0245, 0.0120 and 0.0397 respectively. It thus can be observed that both the actual and estimated errors (associated with y) yielded positive mean values, except for

two of the Bartllet's estimated errors. Also, the mean of the actual error (ε_2) did not vary so much from the estimated errors obtained from the Regression and Anderson-Rubin methods which were 0.0056 and 0.0168 respectively, but differ from the Bartllet's method which was -0.0076. It is worth noting, however, that the three methods yielded mean values which were not dissimilar to zero.

Meanwhile, the standard deviations for the actual error (ε_2) and its equivalent estimated errors recorded higher values whilst the others recorded lower values. Generally, it can be observed that both the actual and the estimated errors (ε_2) yielded higher values for all the descriptive statistics as compared to its counterparts obtained from all the three estimators.

Moreover, the means for both the actual and estimated errors of the latent observation were all positive, unlike what was observed under the measurement errors. It can be seen that the estimated errors yielded means which did not differ so much from one another for the Regression and Anderson-Rubin methods, but produced slightly higher means under the Bartllet's estimator. Again, the Bartllet's estimator recorded higher values in all the descriptive statistics than its counterpart estimators. Comparatively, the descriptive statistics obtained under the latent component for all the three methods of estimators were higher, particularly the standard deviations, than the measurement errors.

5.3 Assessing the finite sample properties of a class of residual estimators

Table 5.2: Average Conditional Bias

Method of Estimation	Residuals						
	δ_1	δ_2	δ_3	ε_1	ε_2	ε_3	ζ
Regression	0.000588	0.000411	0.000411	0.000124	0.000868	0.000868	-0.00094
	(0.000568)	(0.000371)	(0.000448)	(0.000132)	(0.000961)	(0.000101)	(-0.0009)
Bartlett's	0.000002	0.000010	0.000101	0.000011	0.000001	0.000002	0.00000
	(0.000000)	(0.000000)	(0.000010)	(0.000001)	(0.000000)	(0.000001)	(0.0000)
Anderson-Rubin	0.001083	0.000701	0.000712	0.000144	0.000111	0.000010	-0.00072
	(0.000101)	(0.000699)	(0.000679)	(0.000138)	(0.000101)	(0.000001)	(-0.0007)

From Table 5.2 above, it can be observed that the mean conditional biases of the measurement estimated errors were all positive while their counterpart, the latent estimated error, yielded negative values, except for the Bartllet's estimator. It is worth noting that both the Regression and Anderson-Rubin estimators yielded mean values which did not differ remarkably from the parameter, whether known or unknown. It can therefore be reasoned that both estimators reflects a conditional bias for this study but that the strength of biasness was so little. Moreover, it was revealing, as observed in the methodology (Chapter Four), that the mean values of conditional bias, obtained under the Bartllet's estimator were almost zero when the estimating parameter was both known and unknown. Thus, it can be indicated that conditional bias was achieved, to a great extent, for this study under the Bartlett's estimator.

Table 5.3: Correlation Structure of the True Residuals and the Estimated

Residuals when θ is known

	Regression method-based estimator						
	$\delta_{1(R)}$	$\delta_{2(R)}$	$\delta_{3(R)}$	$\varepsilon_{1(R)}$	$\varepsilon_{2(R)}$	$\varepsilon_{3(R)}$	$\zeta_{(R)}$
δ_1	0.8274	-0.2460	-0.2405	-0.0664	-0.0410	-0.0411	-0.1334
	(0.0143)	(0.0421)	(0.0421)	(0.0418)	(0.0475)	(0.0469)	(0.0440)
δ_2	-0.2707	0.9188	-0.1679	-0.0447	-0.0281	-0.0266	-0.0895
	(0.0419)	(0.0073)	0.0424	(0.0437)	(0.0450)	(0.0451)	(0.0474)
δ_3	-0.2668	-0.1693	0.9189	-0.0480	-0.0263	-0.0305	-0.0945
	(0.0400)	(0.0421)	(0.0071)	(0.0458)	(0.0437)	(0.0452)	(0.0459)
ε_1	-0.0641	-0.0379	-0.0420	0.8087	-0.2656	-0.2591	0.3995
	(0.0417)	(0.0430)	(0.0450)	(0.0152)	(0.0410)	(0.0401)	(0.0374)
ε_2	-0.0474	-0.0284	-0.0277	-0.2977	0.9112	-0.1818	0.2785
	(0.0453)	(0.0455)	(0.0467)	(0.0406)	(0.0081)	(0.0463)	(0.0637)
ε_3	-0.0463	-0.0263	-0.0312	-0.2941	-0.1841	0.9120	0.2844
	(0.0460)	(0.0462)	(0.0441)	(0.0391)	(0.0454)	(0.0076)	(0.0676)
ζ	-0.1269	-0.0790	-0.0808	0.3892	0.2467	0.2466	0.7925
	(0.0472)	(0.0448)	(0.0444)	(0.0390)	(0.0437)	(0.0437)	(0.0172)
	Bartlett's method-based estimator						
	$\delta_{1(B)}$	$\delta_{2(B)}$	$\delta_{3(B)}$	$\varepsilon_{1(B)}$	$\varepsilon_{2(B)}$	$\varepsilon_{3(B)}$	$\zeta_{(B)}$
δ_1	0.7052	-0.4118	-0.4065	-0.0001	0.0001	0.0000	-0.1982
	(0.0231)	(0.0373)	(0.0375)	(0.0438)	(0.0471)	(0.0462)	(0.0430)
δ_2	-0.5037	0.8671	-0.2834	-0.0002	-0.0006	0.0010	-0.1349
	(0.0335)	(0.0113)	(0.0407)	(0.0424)	(0.0449)	(0.0450)	(0.0467)
δ_3	-0.4999	-0.2851	0.8671	-0.0012	0.0030	-0.0016	-0.1398
	(0.0329)	(0.0392)	(0.0112)	(0.0452)	(0.0443)	(0.0448)	(0.0457)
ε_1	-0.0004	0.0025	-0.0019	0.7007	-0.4086	-0.4023	0.3942
	(0.0437)	(0.0431)	(0.0440)	(0.0227)	(0.0368)	(0.0353)	(0.0374)

116

ε_2	-0.0016	0.0006	0.0014	-0.5033	0.8662	-0.2829	0.2752
	(0.0458)	(0.0463)	(0.0472)	(0.0341)	(0.0117)	(0.0436)	(0.0644)
ε_3	-0.0004	0.0028	-0.0023	-0.5016	-0.2864	0.8674	0.2814
	(0.0466)	(0.0468)	(0.0436)	(0.0317)	(0.0426)	(0.0110)	(0.0659)
ζ	-0.0002	0.0010	-0.0008	-0.0052	0.0028	0.0032	0.7820
	(0.0466)	(0.0456)	(0.0454)	(0.0469)	(0.0463)	(0.0462)	(0.0182)
Anderson Rubin method-based estimator							
	$\delta_{1(AR)}$	$\delta_{2(AR)}$	δ_{3AR}	$\varepsilon_{1(AR)}$	$\varepsilon_{2(AR)}$	$\varepsilon_{3(AR)}$	$\zeta_{(AR)}$
δ_1	0.7540	-0.3400	-0.3350	0.0820	0.0504	0.0502	-0.2499
	(0.0195)	(0.0404)	(0.0396)	(0.0427)	(0.0473)	(0.0463)	(0.0420)
δ_2	-0.3938	0.8871	-0.2323	0.0585	0.0355	0.0368	-0.1711
	(0.0394)	(0.0098)	(0.0410)	(0.0442)	(0.0448)	(0.0446)	(0.0459)
δ_3	-0.3915	-0.2351	0.8871	0.0558	0.0377	0.0334	-0.1758
	(0.0369)	(0.0404)	(0.0095)	(0.0450)	(0.0436)	(0.0452)	(0.0453)
ε_1	0.0848	0.0533	0.0492	0.7599	-0.3067	-0.3003	0.3813
	(0.0418)	(0.0436)	(0.0446)	(0.0185)	(0.0399)	(0.0393)	(0.0376)
ε_2	0.0581	0.0362	0.0371	-0.3489	0.8902	-0.2112	0.2663
	(0.0459)	(0.0456)	(0.0472)	(0.0398)	(0.0095)	(0.0456)	(0.0630)
ε_3	0.0605	0.0391	0.0343	-0.3455	-0.2137	0.8911	0.2724
	(0.0463)	(0.0473)	(0.0434)	(0.0385)	(0.0449)	(0.0091)	(0.0637)
ζ	0.1683	0.1018	0.1003	0.2697	0.1707	0.1710	0.7565
	(0.0455)	(0.0449)	(0.0445)	(0.0421)	(0.0449)	(0.0445)	(0.0201)

Table 5.4: Correlation Structure of the True Residuals and the Estimated

Residuals when θ is Unknown

Regression method-based estimator							
	$\hat{\delta}_{1(R)}$	$\hat{\delta}_{2(R)}$	$\hat{\delta}_{3(R)}$	$\hat{\varepsilon}_{1(R)}$	$\hat{\varepsilon}_{2(R)}$	$\hat{\varepsilon}_{3(R)}$	$\hat{\zeta}_{(R)}$
δ_1	0.8243 (0.0148)	-0.2442 (0.0301)	-0.2388 (0.0307)	-0.0664 (0.0547)	-0.0408 (0.0524)	-0.0414 (0.0506)	-0.1319 (0.0255)
δ_2	-0.2695 (0.0264)	0.9165 (0.0076)	-0.1692 (0.0297)	-0.0421 (0.0662)	-0.0262 (0.0538)	-0.0249 (0.0551)	-0.0927 (0.0182)
δ_3	-0.2666 (0.0263)	-0.1712 (0.0316)	0.9167 (0.0074)	-0.0504 (0.0667)	-0.0277 (0.0534)	-0.0324 (0.0550)	-0.0927 (0.0181)
ε_1	-0.0630 (0.0543)	-0.0375 (0.0463)	-0.0415 (0.0490)	0.8055 (0.0156)	-0.2658 (0.291)	-0.2621 (0.0302)	0.3995 (0.0636)
ε_2	-0.0476 (0.0631)	-0.0287 (0.0511)	-0.0279 (0.0537)	-0.2961 (0.0246)	0.9087 (0.0084)	-0.1825 (0.0281)	0.2785 (0.0637)
ε_3	-0.0474 (0.0625)	-0.0243 (0.0528)	-0.0325 (0.0525)	-0.2945 (0.0249)	-0.1842 (0.0290)	0.9095 (0.0079)	0.2844 (0.0676)
ζ	-0.1268 (0.0395)	-0.0797 (0.0409)	-0.0815 (0.0410)	0.3880 (0.0324)	0.2458 (0.0312)	0.2422 (0.0310)	0.7872 (0.0176)
Bartlett's method-based estimator							
	$\hat{\delta}_{1(B)}$	$\hat{\delta}_{2(B)}$	$\hat{\delta}_{3(B)}$	$\hat{\varepsilon}_{1(B)}$	$\hat{\varepsilon}_{2(B)}$	$\hat{\varepsilon}_{3(B)}$	$\hat{\zeta}_{(B)}$
δ_1	0.7023 (0.0456)	-0.4058 (0.0339)	-0.4008 (0.0342)	-0.0005 (0.0438)	0.0005 (0.0473)	0.0004 (0.0457)	-0.1956 (0.0379)
δ_2	-0.4980 (0.0407)	0.8629 (0.0274)	-0.2856 (0.0434)	-0.0007 (0.0426)	-0.0010 (0.0450)	0.0006 (0.0449)	-0.1350 (0.0427)
δ_3	-0.4961 (0.0429)	-0.2879 (0.0441)	0.8630 (0.0284)	-0.0014 (0.0455)	0.0028 (0.0442)	-0.0019 (0.0448)	-0.1404 (0.0423)
ε_1	-0.0010 (0.0442)	0.0016 (0.0436)	-0.0024 (0.0443)	0.6980 (0.0409)	-0.4042 (0.0210)	-0.4021 (0.0293)	0.3935 (0.0638)
ε_2	-0.0020 (0.0463)	0.0002 (0.0461)	0.0010 (0.0471)	-0.4980 (0.0354)	0.8638 (0.0230)	-0.2824 (0.0347)	0.2752 (0.0644)
ε_3	-0.0010 (0.0464)	0.0025 (0.0469)	-0.0026 (0.0434)	-0.4994 (0.0345)	-0.2847 (0.0369)	0.8624 (0.0251)	0.2814 (0.0659)
	-0.0014	0.0010	-0.0016	-0.0022	0.0076	0.0007	0.7771

	$\hat{\delta}_{1(AR)}$	$\hat{\delta}_{2(AR)}$	$\hat{\delta}_{3AR}$	$\hat{\varepsilon}_{1(AR)}$	$\hat{\varepsilon}_{2(AR)}$	$\hat{\varepsilon}_{3(AR)}$	$\hat{\zeta}_{(AR)}$
ζ	(0.0468)	(0.0457)	(0.0453)	(0.0624)	(0.0480)	(0.0540)	(0.0214)
Anderson Rubin method-based estimator							
δ_1	0.7520	-0.3362	-0.3314	0.0814	0.0504	0.0502	-0.2499
	(0.0218)	(0.0441)	(0.0431)	(0.0490)	(0.0473)	(0.0463)	(0.0420)
δ_2	-0.3877	0.8799	-0.2392	0.0590	0.0355	0.0368	-0.1711
	(0.0460)	(0.0235)	(0.0454)	(0.0503)	(0.0448)	(0.0446)	(0.0459)
δ_3	-0.3870	-0.2421	0.8796	0.0564	0.0377	0.0334	-0.1758
	(0.0466)	(0.0469)	(0.0284)	(0.0508)	(0.0436)	(0.0452)	(0.0453)
ε_1	0.0850	0.0546	0.0506	0.7578	-0.3067	-0.3003	0.3813
	(0.0488)	(0.0486)	(0.0492)	(0.0200)	(0.0399)	(0.0393)	(0.0376)
ε_2	0.0578	0.0368	0.0378	-0.3421	0.8902	-0.2112	0.2663
	(0.0506)	(0.0483)	(0.0494)	(0.0472	(0.0095)	(0.0456)	(0.0630)
ε_3	0.0609	0.0404	0.0357	-0.3417	-0.2137	0.8911	0.2724
	(0.0527)	(0.0498)	(0.0461)	(0.0484)	(0.0449)	(0.0091)	(0.0637)
ζ	0.1684	0.1042	0.1032	0.2752	0.1707	0.1710	0.7565
	(0.0638)	(0.0569)	(0.0568)	(0.0603)	(0.0449)	(0.0445)	(0.0201)

It can be noted from Tables 5.3 and 5.4 above, that generally there were mean association among the actual estimated errors and the estimated errors off the main diagonal. Univocality, and for that matter validity, ideally would be achieved whenever the mean correlations computed for known parameter was as much close to its unknown parameter. The outcome, as contained in both tables 5.3 and 5.4, mirror the results obtained in the methodology (Chapter four), which indicated that none of the estimators, for all the three applied in this study, was univocally sound or satisfied. More so, all the actual estimated errors associated with so many non-equivalent estimated errors.

Also, the measurement error components which loaded the same factors for estimated errors associated negatively for all the estimators. Meanwhile, the Bartlett's estimator yielded an average of a better degree of association of values of about -0.49 compared to the Regression estimator which recorded a rather low mean association values of about -0.25. Again, for all the actual and estimated residuals, there were no associations among the components loaded on ono-equivalent factors for the measurement errors as they were not dissimilar to zero. Thus it can be indicated that the matter of non-validity was attributable to components on equivalent factors other than the non-equivalent factors.

Moreover, the Bartlett's estimator produced mean associations where the estimated errors linked to the measurement errors were not correlated with actual construct error. It was however revealing to observe that the actual construct errors were associated with the estimated parameters of the measurement errors for both Anderson-Rubin and the Regression methods, which affirmed the case observed in the methodology, but deemed to have recorded rather very low associations for the measurement errors that loaded on the exogenous manifest component and the construct errors. Thus, there exist somewhat moderate association between the actual manifest errors and the estimated errors linked to the measurement error components which loaded on the endogenous construct observations.

Again, very low negative associations were observed between the measurement errors which were linked to the components connecting to the exogenous construct observations and its estimated residuals of the construct errors. Moreover, there were somewhat moderate association between the estimated errors (residuals) for the construct errors and the measurement errors of components connecting to endogenous construct observations. Hence, the outcome on the univocality concept reflects what was obtained in the

120

theoretical computation or methodology. This indicates that there were, to a very little extent, issues of non-validity.

5.4 Using Residual Estimators to Detect Outliers and Influential Observations

Generally, the observations plotted lie within a uniform horizontal scale in each quartile. Slight deviations from the quantile horizontal scale shows evidence of an outlier. However, observations that depart farther away from the uniform horizontal scale indicates evidence of potential influential observation. The QQ plots proposed in this study differs from other methods based on the estimated residuals of the measurement errors in detecting potential outliers and influential observations at the overall model level.

100.0%	maximum	10
99.5%		10
97.5%		10
90.0%		8.9
75.0%	quartile	7.5
50.0%	median	5.4
25.0%	quartile	2.9
10.0%		1.25
2.5%		1.25
0.5%		1.25
0.0%	minimum	1.25
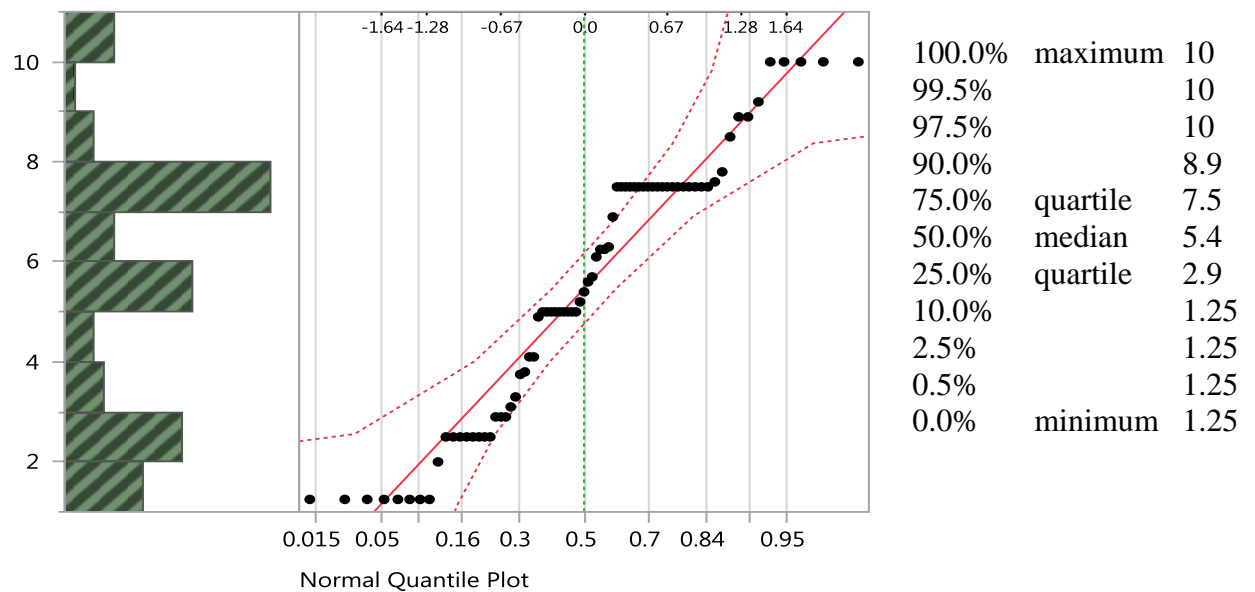
Figure 5.1: QQ plot for Anderson-Rubin based method

It can be noticed from Figure 5.1 above, based on the Anderson-Rubin method, that there was evidence of an influential observation within the first quartile (25th). Again, in the second quartile (50th) there were evidence of outliers which are observations deemed to lie close, about 0.5cm, to the horizontal plane whereas the observation which lie farther

121

away from the horizontal plane were identified as influential observation within the median. Also, there were evidence of both outliers and influential observations in the third quartile (75th). In the last quartile, observations can be seen lying almost on the horizontal plane and others lying within the 0.5cm distance which were all deemed to be outliers with some few observations found to lie outside the reference distance or father away from the horizontal plane and a such were deemed to be influential observations.
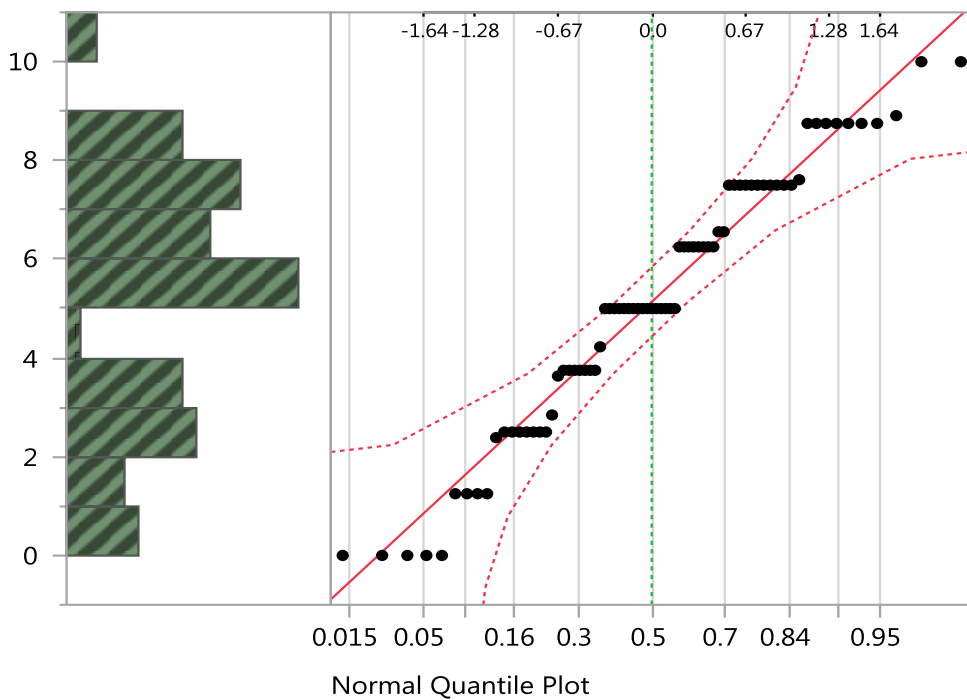
Figure 5.2: QQ plot for data in Anderson-Rubin based method

From the Anderson-Rubin method, as can be seen in Figure 5.2 above when real data was utilized, that there was evidence of an influential observation within the first quartile (25th). However, no outlier observation was seen in the 25th quartile. Moreover, in the 50th quartile (median), there were evidence of one outlier which was observation deemed to lie close, the horizontal quantile plane whereas the many observations were observed as influential observation within the median. Also, there were no evidence of outliers but there were some influential observations in the 75th quartile. In the last quartile,

observations can be seen lying almost on the horizontal quantile plane and others lying

outside the 1.5cm distance which were all deemed to be influential observations of the

real data applied.

100.0%	maximum	10
99.5%		10
97.5%		9.9999982
90.0%		9.999998
75.0%	quartile	9.999998
50.0%	median	6.666666
25.0%	quartile	3.333333
10.0%		0
2.5%		0
0.5%		0
0.0%	minimum	0

Normal Quantile Plot
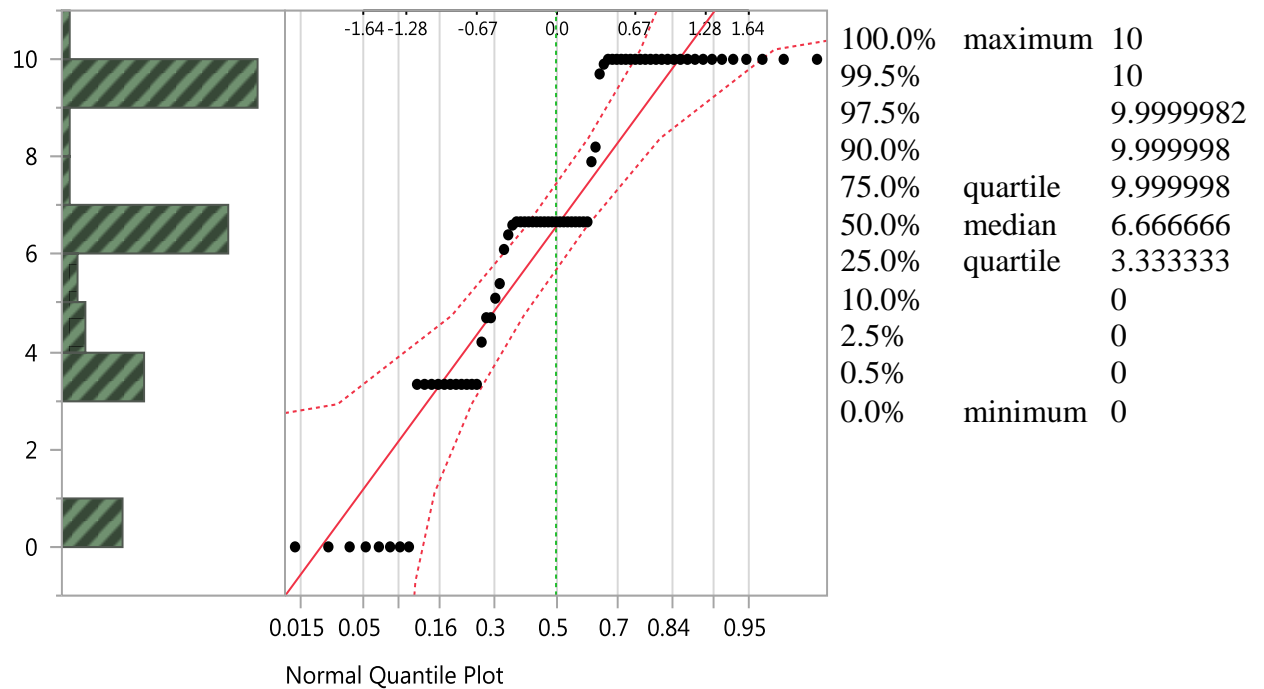
Figure 5.3: QQ plot for Bartlett's based method

From Figure 5.3 above, based on the Bartlett's method, that there was no evidence of

both outliers and influential observations within the first quartile (25th). Meanwhile, the

second quartile (50th) showed evidence of outliers which are observations deemed to lie

close, about 0.5cm, to the quantile horizontal plane whereas the observation which lie

farther away from the quantile horizontal plane were represents the influential

observations within the median. Also, there were evidence of both outliers and influential

observations in the third quartile (75th). In the last quartile, observations can be seen lying

almost on the horizontal plane and others lying within the 0.5cm distance which were all

deemed to be outliers with some few observations found to lie outside the reference

123

distance or father away from the horizontal plane and a such were deemed to be influential observations.
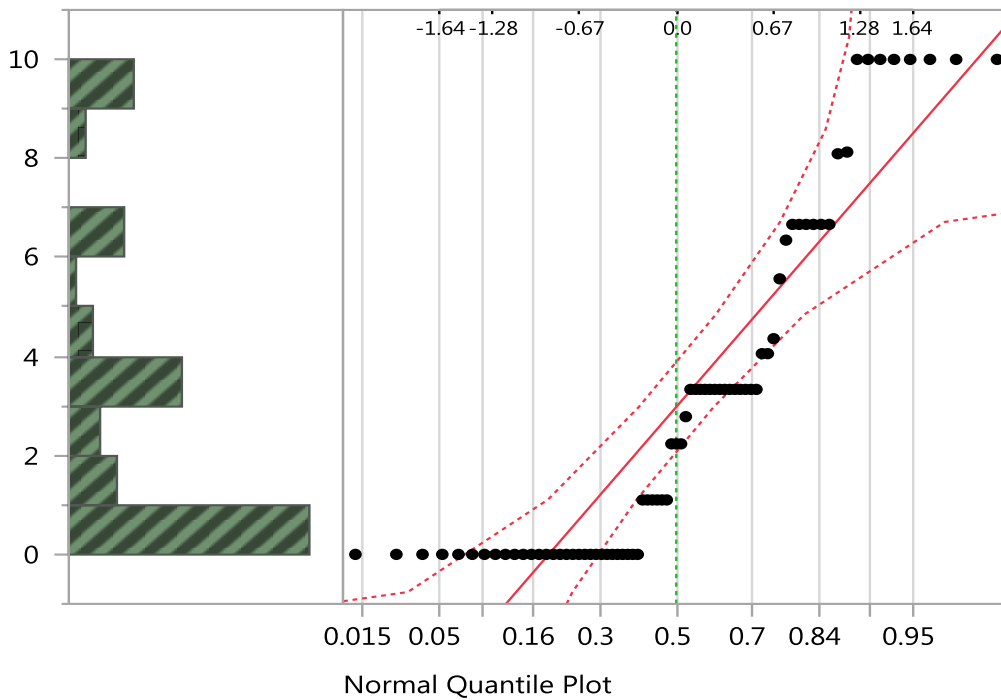
Figure 5.4: QQ plot for data in Bartlett's based method

From the Bartlett's based method, as can be seen in Figure 5.4 above with real data, that there was evidence of few outliers and three influential observations within the 25th quartile. Again, the 50th quartile showed few outliers and about two influential observations. Moreover, in the 75th quartile there were evidence of two outliers which were observation deemed to lie close, the quantile horizontal quantile plane with no evidence of influential observations within the third quartile.

100.0%	maximum	10
99.5%		10
97.5%		10
90.0%		10
75.0%	quartile	6.666666
50.0%	median	3.333333
25.0%	quartile	1.496333
10.0%		0
2.5%		0
0.5%		0
0.0%	minimum	0

Normal Quantile Plot
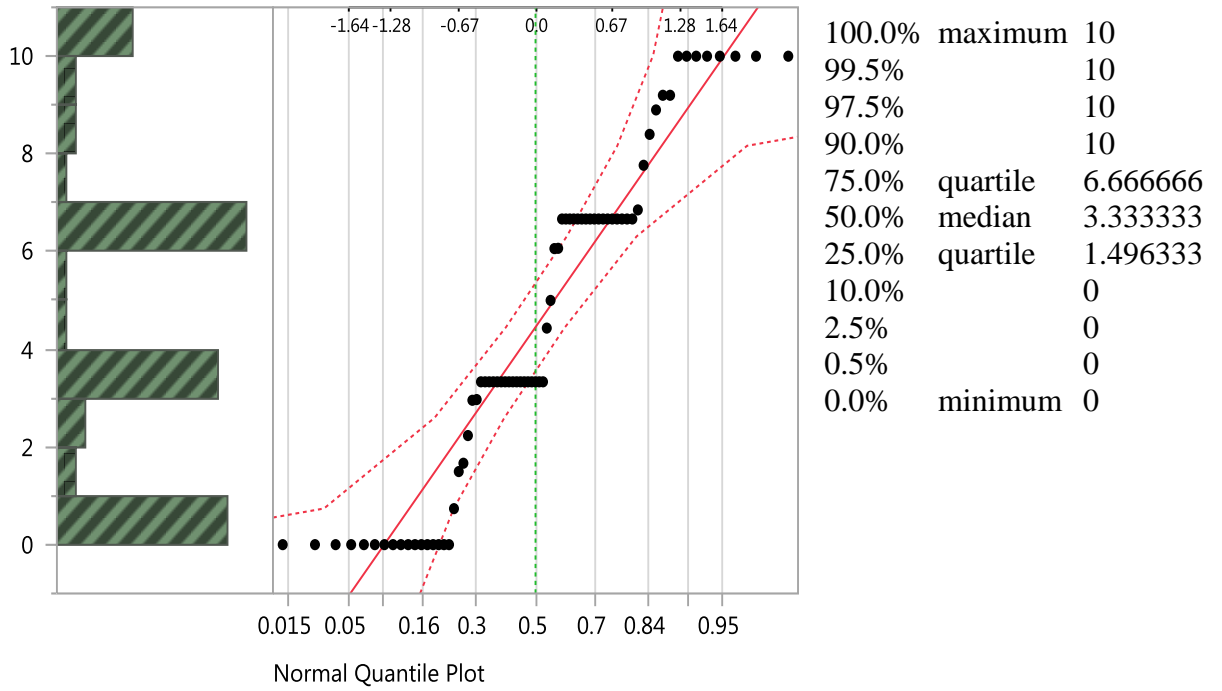
Figure 5.5: QQ plot for Regression based method

It can be noticed from Figure 5.5 above, based on the regression method, that there was evidence of about three outliers and two influential observations within the first quartile (25th). Also, the second quartile (50th) showed evidence of two outliers which are observations deemed to lie close, about 0.5cm, to the quantile horizontal plane whereas the two observations which lie farther away from the quantile horizontal plane were identified as influential observations within the median. Also, there were evidence of about four outliers and two influential observations in the third quartile (75th).

125
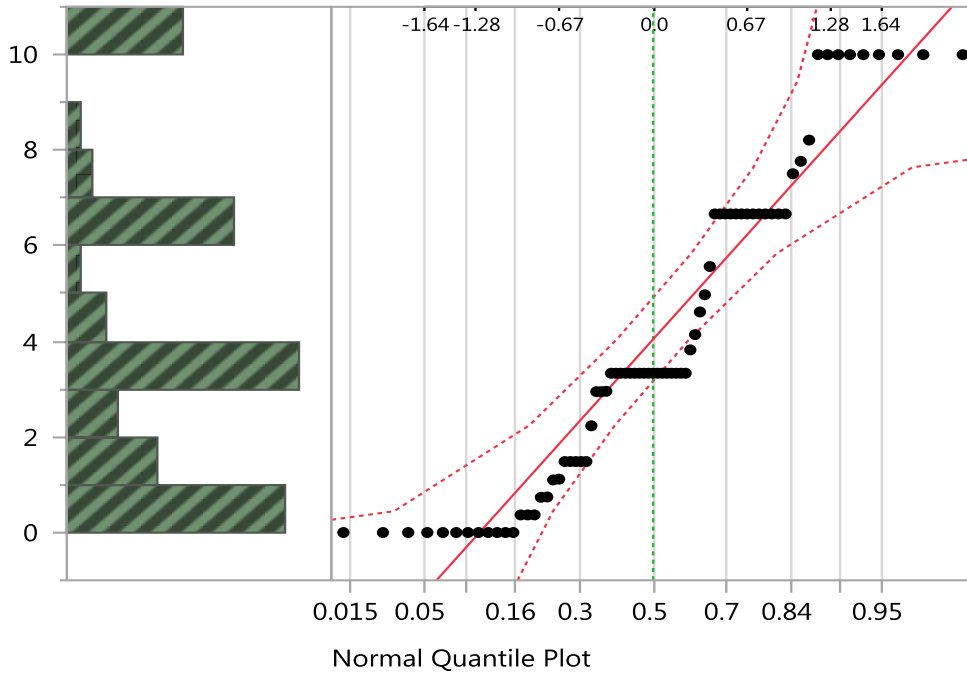
Figure 5.6: QQ plot for data in Regression based method

From the figure 5.6 above, using real data for the regression based method, it can be seen that at the 25th quartile, there were about five observations identified as outliers and two observations deemed to be influential variables. In the 50th quartile, only a single observation was seen as an outlier and about two observations identified as influential within the median. However, it was revealing to note that in the 75th quartile there were evidence of five outliers which lie within the close distance to the quantile horizontal plane with zero or no influential observation recorded. Again, the fourth quartile showed evidence of three observation deemed to be identified as outliers without any influential observation.

126
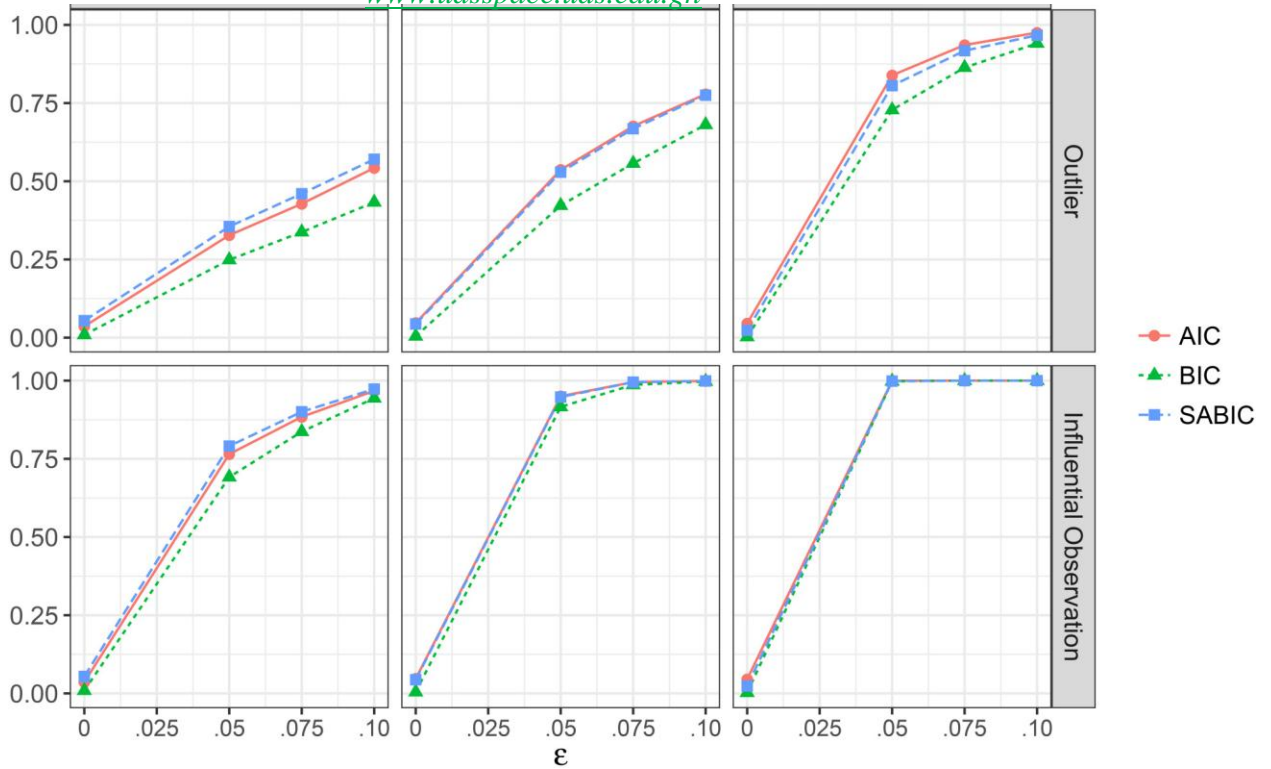
Figure 5.7: Information criteria for QQ plots

The fitting indices, as indicated in Figure 5.7, for the QQ plot of the various estimation methods clearly support the earlier view, based on the QQ plots, that Anderson-Rubin based method provides a better visual display for the detection of outliers and influential observations. As seen from Figure 5.6, the first plot from left, which represents the Anderson-Rubin based method plot shows a smaller AIC, BIC and SABIC as compared to Bartlett's and the Regression based methods which are second and third from left.

5.5 Comparing the EM Method against Other Methods of Residual Estimators

Determined factor scores were obtained at the preliminary stage in order comprehend the impact of the number of components for each manifest observation, type of manifest observation indicator association and the estimation technique for the parameters in a recursive SEM, mean absolute deviation of the standard error and the overall fitness of the equation.

127

The measurement observations of curious errors, manifest component and estimation methods for reasons of breaking down and arranging the outcome of the parameters considered here was laid out and explained under the circumstances for all four categories of estimators being compared.

Table 5.5: Parameter Estimates and Standard Errors of Residual Estimators

Parameter	Regression method	Bartlett's method	Anderson Rubin method	EM method
δ_1	0.598 (0.12)	0.599 (0.10)	0.600 (0.10)	0.600 (0.09)
δ_2	0.648 (0.14)	0.648 (0.13)	0.649 (0.14)	0.650 (0.14)
δ_3	0.699 (0.15)	0.701 (0.13)	0.703 (0.12)	0.700 (0.15)
ϵ_1	0.636 (0.11)	0.637 (0.10)	0.639 (0.10)	0.641 (0.09)
ϵ_2	0.572 (0.09)	0.573 (0.09)	0.574 (0.09)	0.578 (0.09)
ϵ_3	0.504 (0.09)	0.505 (0.09)	0.507 (0.08)	0.510 (0.08)
ζ_1	0.407 (0.24)	0.418 (0.19)	0.429 (0.16)	0.432 (0.23)
Test				
χ^2	28.29	25.29	26.82	57.80
RMSEA	0.040	0.026	0.034	0.023
p	0.205	0.335	0.264	0.001
SRMR	0.041	0.037	0.038	0.020
CFI	0.989	0.994	0.992	0.984
AIC	268.466	259.516	264.692	246.317

From Table 5.5 above, it can be seen hat the estimates residual (δ_3) linked to the measurement component recorded the highest estimates with higher standard errors across the four estimators whilst computed errors (ζ_1) associated with the manifest component recorded the lowest parameter estimates with rather higher values of standard errors. Also, on one hand, the estimates obtained under the Bartlett's method was higher than the other two known methods (regression and Anderson-Rubin) whiles the EM, on

the other hand, yielded the highest parameter estimates than all the other three methods of estimators (Regression, Bartlett's and Anderson-Rubin).

Again, the Regression method yielded fitness (χ^2 =28.29; p=0.205, RMSEA=0.040, CFI=0.989, SRMR = 0.041) for the study in terms of the estimates obtained for the parameters. For the Bartlett's method, it was observed that the fitness (χ^2= 25.29, p=0.335, RMSEA=0.026, CFI=0.994, SRMR=0.037) were somewhat dissimilar to the fitness indices obtained under the regression method. The SRMR stands better for purposes of comparison, particularly for a Chi-square distribution.

Also, the Anderson-Rubin method yielded goodness-of-fitness (χ^2= 26.82, p=0.264, RMSEA=0.034, CFI=0.992, SRMR=0.038). It is worth noting that the AIC preferred the Bartlett's estimator over and above the Regression and Anderson Rubin estimators with differential values 259.516, 268.466 and 264.692 respectively. This therefore indicates somewhat slight heavy tail in the distribution without considering EM method yet. Again, to a very large extent, is worth noting that most of the fitness figures and the estimates alike (contained in Table 5.5) under all the methods applied here were closer and therefore makes choice, a bit trivial, among the three existing estimators utilized in this study.

It was however observed that when the EM method was eventually applied to change the estimation technique utilized in the other three estimators it yielded higher estimates. The EM method produced a fitness indices (χ^2= 57.80, p< 0.001, RMSEA=0.023, CFI=0.984, SRMR=0.020). Close examination of both χ^2 and RMSEA indicates a kind

of not good fit, though the effect of CFI was pronounced but the SRMR really provided a better (supported in literature) fitness as compared to the other three existing methods.

Thus, it means that these methods applied here provided goodness of fitness indices which were close to show an obvious choice. Against this backdrop the standard errors mirroring the amount of error in estimating the parameters and its equivalent goodness of fitness could be utilized to further comprehend the specific residual method estimation that produced a better parameter estimate. This therefore supports the choice of Bartlett's and EM as they both recorded minimal standard errors. Also, the comparative fitness of the EM method was compared to the other three existing methods. Together, the AIC, BIC, and CAIC, strongly preferred the EM method against the other three methods with some amount of differentials though.

More so, the estimates shown in Table 5.5 demonstrates that much as the parameters were very close for the various estimators, there was an element of robustness in Bartlett's and the EM method in particular.

5.6 Discussions

It can be noted from Tables 5.3 and 5.4 above, that generally there were mean association among the actual estimated errors and the estimated errors off the main diagonal. Univocality, and for that matter validity, ideally would be achieved whenever the mean correlations computed for known parameter was as much close to its unknown parameter. The outcome mirror the results obtained in (Heise & Bohrnstedt 1970; Beauducel & Herzberg, 2006; Nussbeck et al, 2006) which indicated that none of the estimators, for

all the three applied in this study, was univocally sound or satisfied. More so, all the actual estimated errors associated with so many non-equivalent estimated errors.

Also, the measurement error components which loaded the same factors for estimated errors associated negatively for all the estimators. Meanwhile, the Bartlett's estimator yielded an average of a better degree of association compared to the Regression estimator which recorded a rather low mean association values corroborates Asparouhov & Muthén, (2010b). Again, for all the actual and estimated residuals, there were no associations among the components loaded on ono-equivalent factors for the measurement errors as they were not dissimilar to zero. Thus it can be indicated that the matter of non-validity was attributable to components on equivalent factors other than the non-equivalent factors which contradicts some previous studies (Asparouhov & Muthén, 2010b; Baldwin & Fellingham, 2013; Hox et al, 2012) as they employed Bayes estimation.

Moreover, the Bartlett's estimator produced mean associations where the estimated errors linked to the measurement errors were not correlated with actual manifest error. It was however revealing to observe that the actual manifest errors were associated with the estimated parameters of the measurement errors for both Anderson-Rubin and the Regression methods which supports Asparouhov & Muthén (2010b) but deemed to have recorded rather very low associations for the measurement errors that loaded on the exogenous manifest component and the manifest errors. Thus, there exist somewhat moderate association between the actual manifest errors and the estimated errors linked to the measurement error components which loaded on the endogenous manifest

131

observations. These findings mirror the views opined in earlier study by Nussbeck et al (2006) which adopted a Monte Carlo approach.

Again, very low negative associations were observed between the measurement errors which were linked to the components connecting to the exogenous manifest observations and its estimated residuals of the manifest errors affirms the position of Beauducel & Herzberg (2006). Moreover, there were somewhat moderate association between the estimated errors (residuals) for the manifest errors and the measurement errors of components connecting to endogenous manifest observations. Hence, the outcome on the univocality concept reflects what was obtained in previous study by Nussbeck et al, (2006).

Also, the study applied the group of residual estimators to spot outliers and possible controlling observations via uniform horizontal QQ plot. The study implemented the QQ plots in SEM using JMP software. The implantation experience supports Sterba and Pek (2012) and Yuan and Zhang (2012) who opined that identifying outliers in SEM was rarely accessible due to the complexity of modelling, unlike traditional modelling in statistics including linear regression models.

Our results showed that the presence of outliers and possible controlling observation which affirms the assertion made by Yuan and Zhong (2013) who spotted outliers and controlling observations through boxplots. Moreover, it was revealing to note that the ease with which outliers and possible controlling observations could be spotted in this study, particularly with the Anderson-Rubin technique which contradicts Aguinis et al,

(2013) who indicated that only outliers could be identified easily but noting the challenge in spotting controlling observations using boxplot under SEM framework.

Also, the present study found Aderson-Rubin technique the most efficient method of identifying outliers and possible controlling observations under SEM which corroborates the previous studies that utilized general techniques such as Mahalanobis and Cook's distances (Flora et al, 2012; Sterba & Pek, 2012). These residual estimators provided parameters and fitness that are insensitive to the influence of outliers and possible controlling observations as they were achieved through a robust procedure of estimation by assigning weights. Further, the paper affirms the views noted in other studies that outliers and possible controlling observations need be of interest in their own right and therefore can lead to crucial scientific findings (Aguinis et al, 2013; O'Connell et al, 2015; Mark & Jiaqi, 2017).

Again, the present study provides a different perspective to spotting outliers and possible controlling observations through a uniform horizontal QQ plots approach as was opined in earlier methodological works which provided accessible tools to identify outliers and possible controlling observations in SEM (Pek & MacCallum, 2011; Sterba & Pek, 2012; Mark & Jiaqi, 2017).

It is worth noting that despite the significant contributions of the study, there were some limitations that call further studies. To begin with, it should be emphasized that, in the current study, we only focused on situations where a small proportion of data is partitioned in each quantile, based on the moderate sample sized used. Also, the data used in the QQ plot was found to be normal and for that matter further studies could

ascertain new way(s) of detecting outliers and controlling observations for a non-normal data with the same or similar concept of residual estimators. Corrections for non-normality such as the Satorra-Bentler procedure which relies on sandwich estimator and higher-order moments of the sample data could be adopted as data used under the SEM concept often had skewness and kurtosis deviated from those of a normal distribution (Mark & Jiaqi, 2017).

CHAPTER SIX

SUMMARY OF FINDINGS, CONCLUSIONS AND RECOMMENDATIONS

6.1 Introduction

The main aim of the study was to apply, through simulation, the EM method to examine residuals in SEM. This chapter, therefore, entails the summary of the main findings. It also draws conclusions based on the objectives of the study and suggested recommendations to researchers and for further studies in the area.

6.2 Summary of Main Findings

To begin with, the mean values of conditional bias, obtained under the Bartllet's estimator were almost zero when the estimating parameter was both known and unknown. Thus, it can be indicated that conditional bias was achieved, to a great extent, for this study under the Bartlett's estimator.

Also, the Bartlett's estimator yielded an average of a better degree of association of values compared to the Regression estimator which recorded a rather low mean association values. Hence, the outcome on the univocality concept reflects what was obtained in the theoretical computation. This indicates that there were, to a very little extent, issues of non-validity.

Moreover, it is worth noting that the Anderson Rubin method of QQ plot provided a more efficient or visual display of spotting outliers and possible controlling observations as compared to the other group of estimators.

135

Further, it was revealing to note that the comparative fitness of the Bartlett's method was referred to the other two existing methods (i.e the Regression and Anderson Rubin). Together, the AIC, BIC, and CAIC, strongly preferred the EM method against the other three methods with some amount of differentials. More so, the estimates demonstrated that much as the parameters were very close for the various estimators, there was an element of robustness in Bartlett's and especially the EM method.

6.3 Conclusions

The main aim of the study was to apply, through simulation, the EM method to examine residuals and therefore provide the basis in SEM. To begin with, a group of residual estimators were examined in terms of its ability to meet some finite properties of an observed data set. Again, the study was specifically aimed at using QQ plots to spot outliers and possible controlling observations in a given data. Last but not least, the EM method was to be compared against the other three existing estimators in SEM in order to ascertain which method estimates better.

Objective 1

In conclusion, the results indicate that the mean values of conditional bias, obtained under the Bartllet's estimator were almost zero when the estimating parameter was both known and unknown. Thus, it can be indicated that conditional bias was achieved, to a great extent, for this study under the Bartlett's estimator. Also, the outcome mirror the results obtained in the methodology, which indicated that none of the estimators, for all the three applied in this study, was univocally sound or satisfied. More so, all the actual estimated errors associated with so many non-equivalent estimated errors. Again, very low negative associations were observed between the measurement errors which were linked to the components connecting to the exogenous manifest observations and its estimated residuals

of the manifest errors. Moreover, there were somewhat moderate association between the estimated residuals for the manifest errors and the measurement errors of components connecting to endogenous manifest observations. Hence, the outcome on the univocality concept reflects what was obtained in the theoretical computation. This indicates that there were, to a very little extent, issues of non-validity.

Objective 2

It can be deduced from the results on the various simulation of QQ plots that all these methods demonstrated the ability to spot outliers and possible influential observation under the SEM concept. It is worth noting that the AR method of QQ plot provided a more efficient or visual display of spotting outliers and possible influential observations as compared to the other group of estimators. This, therefore, can be deemed as an efficient way of expanding the Cook's method of spotting outliers and influential observation with the uniform horizontal QQ plot under the SEM concept.

Objective 3

It was unclear and difficult in arriving at a definite decision in terms of which residual estimator yielded better residual parameter estimates based on the model fit indices since the strength of one residual estimator may be the weakness of the other. It was worth noting that the Bartlett's estimator was preferred over and above the Regression and Anderson Rubin estimators with differential values. This therefore indicates somewhat slight heavy tail in the distribution without considering EM method yet. Again, to a very large extent, is worth noting that most of the fitness figures and the estimates under all the methods applied here were closer and therefore makes choice, a bit trivial, among the three existing estimators utilized in this study. The comparative fitness of the Bartlett's method was referred to the other two existing methods (i.e the Regression and Anderson

UNIVERSITY FOR DEVELOPMENT STUDIES

Rubin). Together, the EM method was strongly preferred against the other three methods with some amount of differentials. It is therefore worth noting that this present study contribution to knowledge is demonstration of the fact that EM method could be a better residual estimator within the SEM concept compared to other existing methods.

6.4 Recommendations

To begin with, attention must be given when developing residual diagnostics and tests to account for the lack of structure and univocality to ensure the validity of finite properties diagnostics and tests.

Again, the Anderson-Rubin method of estimator are conditionally biased and should be utilized whenever this property is to be ascertained under the SEM concept.

Moreover, the uniform scale QQ plots, in particular, the Anderson-Rubin method, in detecting outliers and potential influential observations in SEM concept is comparatively better.

Further, the Bartlett's method of estimator could be utilized whenever the structure for model is to be ascertained in SEM as it yielded better association among the other group of estimators.

Finally, the EM method could be employed for estimating residuals in SEM since it was deemed to provide better residual estimates as opposed to the other methods of estimators.

138

Future Works

Based on this current research, further expansion of the residual plots and Cook's distance to examine the behavior of the residual estimators under data contamination would necessary.

As the impact of outliers and controlling observations on fitness could be dissimilar based on the kind of SEM, complex models, and estimation, later research could improve upon the simulation circumstances in order to get other sides of the study covered.

Also, it should be determined if there are alternative ways, other than orthogonality, as was required for Anderson-Rubin method in SEM concept.

Again, later research should look at the plausibility for a hypothesized model to have more exogenous manifest components than endogenous manifest components.

More so, subsequent studies could utilized derivation of properties of the estimators theoretically should there be issues of violations in any assumption.

Further, later research could utilized much more small samples to observe the behavior of estimates and model fit indices as it turns to impact negatively on these indices given that it is a Chi-square distribution.

Finally, similar residual SEM studies should be developed and implemented in R since SAS and JMP are commercial software and often expensive to procure or obtain for such analysis.

Contribution to Knowledge

The simulation of uniform scale QQ plots, through the various residual estimators, to detect outliers and influential observations extends Cook's distance within the SEM framework.

Again, the simulation of EM method which was achieved by using a split or stepwise estimation procedure to obtain better residual estimates against other residual estimators (Regression, Bartlett's and Anderson Rubin methods) is, to the best of my knowledge, unprecedented and therefore provides a different method of estimating residual parameters within the SEM framework.

REFERENCES

Aguinis, H., Gottfredson, R. K., and Joo, H. (2013). Best-practice recommendations for defining, identifying, and handling outliers. Org. Res. Methods 16, 270–301. doi: 10.1177/1094428112470848.

Anderson, T. W. and Rubin, H. (1956). Statistical inference in factor analysis. Proceedings of the third Berkeley Symposium on mathematical statistics and probability, 5, 111-150.

Asparouhov, T., & Muthén, B. (2007). Computationally efficient estimation of multilevel high-dimensional latent variable models. In *Proceedings of the 2007 JSM meeting in Salt Lake City, Utah, Section on Statistics in Epidemiology* (pp. 2531–2535). Salt Lake City, Utah.

Asparouhov, T., & Muthén, B. (2010). *Bayesian analysis using Mplus: Technical implementation*. Retrieved December 18, 2012 from http://www.statmodel.com

Asparouhov, T., & Muthén, B. (2010a). *Bayesian analysis in Mplus: Technical implementation* (Mplus technical report, Version 3). Retrieved from http://www.statmodel.com/download/Bayes3.pdf

Asparouhov, T., & Muthén, B. (2010b). *Bayesian analysis of latent variable models using Mplus* (Technical report, Version 4). Retrieved from http://www.statmodel.com

Asparouhov, T., & Muthén, B. (2012a). *Comparison of computational methods for high dimensional item factor analysis* (Mplus technical report, Version 7). Retrieved from http:// www.statmodel.com/download/HighDimension11.pdf

Asparouhov, T., and Muthén, B. (2015). Structural equation models and mixture models with continuous non-normal skewed distributions. Struc. Eq. Model. A Multidiscip. J. 23, 1–19. doi: 10.1080/10705511.2014.947375.

Baldwin, S. A., & Fellingham, G. W. (2013). Bayesian methods for the analysis of small sample multilevel data with a complex variance structure. *Psychological Methods*, *18*(2), 151–164. doi: 10.1037/a0030642

Bartlett, M. S. (1937). The statistical conception of mental factors. British Journal of Psychology, 28, 97-104.

Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling: A Multidisciplinary Journal*, *13*(2), 186–203. doi: 10.1207/s15328007sem1302_2

141

Bentler, P. M. (1993). Structural equation models as nonlinear regression models. In K.

Binnewies, C., Sonnentag, S., & Mojza, E. J. (2010). Recovery during the weekend and fluctuations in weekly job performance: A week-level study examining intra-individual relationships. *Journal of Occupational and Organizational Psychology*, *83*(2), 419–441. doi: 10.1348/096317909X418049

Bollen, K. A. & Arminger, G. (1991). Observational residuals in factor analysis and structural equation models. Sociological Methodology, 21, 235-262.

Bollen, K. A. (1989). Structural equations with latent variables. New York, NY: JohnWiley & Sons.

Bollen, K. A., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin, 110*(2), 305-314.

Byrne, B. M., (2010). Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming, 2nd edition, Taylor & Francis

Carretero-Dios, H., Eid, M., & Ruch, W. (2011). Analyzing multitrait-mulitmethod data with multilevel confirmatory factor analysis: An application to the validation of the state trait cheerfulness inventory. *Journal of Research in Personality*, *45*(2), 153–164. doi: 10.1016/j.jrp.2010.12.007 cut-eGroup. (2011). *Client-specific 360-degree feedback questionnaire*. Unpublished questionnaire.

Chin, W. W. (1998). Issues and opinion on structural equation modeling. *Management Information Systems Quarterly, 22*(1).

Curtis, R. F., & Jackson, E. F. (1962). Multiple indicators in survey research. *American Journal of Sociology, 68*, 195-204.

Depaoli, S., & Clifton, J. P. (2015). A Bayesian approach to multilevel structural equation modeling with continuous and dichotomous outcomes. *Structural Equation Modeling: A Multidisciplinary Journal*, *22*(3), 327–351. doi: 10.1080/10705511.2014.937849

Dempster, A. P., Laird, N. M., Rubin, D. B., (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. Journal of the Royal Statistical Society. Series B (Methodological) 39 (1), 1–38.

Fayolle, A., Engelbrecht, B., Freycon, V., Mortier, F., Swaine, M., Réjou-Méchain, M., Doucet, J.-L., Fauvet, N., Cornu, G., Gourlet-Fleury, S., (2012). Geological Substrates Shape Tree Species and Trait Distributions in African Moist Forests. PLoS ONE 7 (8), e42381

Flora, D. B., LaBrish, C., and Chalmers, R. P. (2012). Old and new ideas for data screening and assumption testing for exploratory and confirmatory factor analysis. Front. Psychol. 3:55. doi: 10.3389/fpsyg.2012.00055

Foulley, J. L., (2002). Algorithme EM: Théorie et application au modèle mixte. Journal de la Société française de statistique 143 (3-4), 57–109.

Geiser, C., Eid, M., Nussbeck, F. W., Courvoisier, D. S., & Cole, D. A. (2010). Analyzing true change in longitudinal multitrait-multimethod studies: Application of a multimethod change model to depression and anxiety in children. *Developmental Psychology*, *46*(1), 29–45. doi: 10.1037/a0017888

Gelman, A., and Hill, J. (2006). Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge: Cambridge University.

Haagen, D. J. Bartholomew, & M. Deistler (Eds.), *Statistical modelling and latent variables*. Amsterdam, Netherlands: Elsevier Science. pp. 51–64.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika, 57*, 97-109.

Heise and bohrnstedt Heise, D. R. and Bohrnsted, G. W. (1970). Validity, invalidity, and reliability. Sociological Methodology, 2, 104-129.

Henseler, J. (2010). On the convergence of the partial least squares path modeling algorithm. *Computational Statistics, 25*, 107-120.

Henseler, J. (2012). Why generalized structured component analysis is not universally preferable to structural equation modeling. *Journal of the Academy of Marketing Science, 40*, 402-413.

Hoogland, J. J., & Boomsma, A. (1998). Robustness studies in covariance structure modeling: An overview and a meta-analysis. *Sociological Methods & Research, 26*(3), 329-367.

Hox, J. J., & Maas, C. J. (2001). The accuracy of multilevel structural equation modeling with pseudobalanced groups and small samples. *Structural Equation Modeling: A Multidisciplinary Journal*, *8*(2), 157–174. doi: 10.1207/S15328007SEM 0802_1

Hox, J. J., Maas, C. J., & Brinkhuis, M. J. (2010). The effect of estimation method and sample size in multilevel structural equation modeling. *Statistica Neerlandica*, *64*(2), 157–170. doi: 10.1111/j.1467-9574.2009. 00445.x

Hox, J. J., van de Schoot, R., Matthijsse, S. (2012). How few countries will do? Comparative survey analysis from a Bayesian perspective. *Survey Research Methods*, *6*(2), 87–93. doi: 10.18148/srm/2012.v6i2.5033.

Hoyle, R. H. (2000). Confirmatory factor analysis. In H. E. A. Tinsley & S. D. Brown (Eds.), *Handbook of applied multivariate statistics and mathematical modeling*. New York: Academic Press.

Hu, L., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin, 112*(2), 351-362.

Husson, F., Josse, J., Lê, S., (2008). FactoMineR: An R Package for Multivariate Analysis. Journal of Statistical Software, 18.

Hulland, J., Ryan, M. J., & Rayner, R. K. (2010). Modeling customer satisfaction: A comparative performance evaluation of covariance structure analysis versus partial least squares. In V. E. Vinzi, W. W. Chin, J. Henseler, & H. Wang (Eds.), *Handbook of Partial Least Squares*. Berlin: Springer-Verlag.

Hwang, H., & Takane, Y. (2004). Generalized structured component analysis. *Psychometrika, 69*(1), 81-99.

Hwang, H., Ho, M. H. R., & Lee, J. (2010). Generalized structured component analysis with latent interactions. *Psychometrika, 75*, 228-242.

Hwang, H., Malhotra, N. K., Kim, Y., Tomiuk, & Hong (2010). A comparative study on parameter recovery of three approaches to structural equation modeling. *Journal of Marketing Research, 47*, 699-712.

Kiersch, C. E., & Byrne, Z. S. (2015). Is being authentic being fair? Multilevel examination of authentic leadership, justice, and employee outcomes. *Journal of Leadership & Organizational Studies*, *22*(3), 292–303. doi: 10.1177/1548051 815570035

Kline, R. B. (2011). Principles and practice of structural equation modeling (1st Edition). New York, NY: Guilford.

Kline, R. B. (2011). Principles and Practice of Structural Equation Modeling (3rd Edition). New York, NY: The Guilford Press.

Kutner, M. H., Nachtsheim, C. J., Neter, J., and Li, W. (2004). Applied Linear Statistical Models (Fifth edition). New York, New York: McGraw-Hill/Irwin.

Lee, S.-Y., & Song, X.-Y. (2004). Evaluation of the Bayesian and maximum likelihood approaches in analysing structural equation models with small sample sizes. *Multivariate Behavioral Research*, *39*(4), 653–686. doi: 10.1207/s15327906 mbr3904 4

Lee, S.-Y., Song, X.-Y., & Cai, J.-H. (2010). A Bayesian approach for nonlinear structural equationmodels with dichotomous variables using logit and probit links. *Structural Equation Modeling: A Multidisciplinary Journal*, *17*(2), 280–302. doi: 10.1080/10705511003659425

Little, J. (2013). Multilevel confirmatory ordinal factor analysis of the life skills profile–16. *Psychological Assessment*, *25*(3), 810–825. doi: 10.1037/a0032574

Löfholm, C. A., Eichas, K., & Sundell, K. (2014). The Swedish implementation of multisystemic therapy for adolescents: Does treatment experience predict treatment adherence? *Journal of Clinical Child & Adolescent Psychology*, *43*(4), 643–655. doi: 10.1080/15374416.2014.883926

Lynch, S. M. (2007). *Introduction to applied Bayesian statistics and estimation for social scientists*. New York, NY: Springer.

Mahlke, J., Schultze, M., Koch, T., Eid, M., Eckert, R., & Brodbeck, F. (2016). A multilevel CFA-MTMM approach for multisource feedback instruments: Presentation and application of a new statistical model. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*, 91–110. doi: 10.1080/10705511. 2014.9901533

McDonald, R. P. & Burr, E. J. (1967). A comparison of four methods of constructing factor scores. Psychometrika, 32(4), 381-401.

Meuleman, B., & Billiet, J. (2009). A Monte Carlo sample size study: How many countries are needed for accurate multilevel SEM? *Survey Research Methods*, *3*, 45–58. doi: 10.18148/srm/2009.v3i1.666

Miranda, H., & Russell, M. (2011). Predictors of teacher directed student use of technology in elementary classrooms: A multilevel SEM approach using data from the USEIT study. *Journal of Research on Technology in Education*, *43*(4), 301–323. doi: 10.1080/15391523.2011.10782574

Mortier, F., Trottier, C., Cornu, G., Bry, X., (2014). SCGLR-An R Package for Supervised Component Generalized Linear Regression. Journal of Statistical Software.

Muthén, B. (1983). Latent variable structural equation modeling with categorical data. *Journal of Econometrics*, *22*(1), 43–65. doi: 10.1016/0304-4076(83)90093-3

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, *49*(1), 115–132. doi: 10.1007/BF02294210

Muthén, B. (2010). *Bayesian analysis in Mplus: A brief introduction* (Incomplete draft, Version 3). Retrieved from http://www.statmodel.com/download/IntroBayes Version% 203.pdf

Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods*, *17*(3), 313–335. doi: 10.1037/a0026802

Muthén, L. K., & Muthén, B. O. (1998-2010). M*plus* User's Guide. Sixth Edition. Los Angeles, CA: Muthén & Muthén.

Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide*. (7th ed.) [Computer software manual]. Los Angeles, CA: Muthén & Muthén.

Nussbeck, F. W., Eid, M., & Lischetzke, T. (2006). Analysing multitrait–multimethod data with structural equation models for ordinal variables applying the WLSMV estimator: What sample size is needed for valid results? *British Journal of Mathematical and Statistical Psychology*, *59*(1), 195–213. doi: 10.1348/0007 11005X67490

O'Connell, A. A., Yeomans-Moldanado, G., and McCoach, D. B. (2015). "Residual diagnostics and model assessment in a multilevel framework: recommendations toward best practice," in Advances in Multilevel Modeling for Educational Research: Addressing Practical Issues Found in Real-World Applications, eds J. R. Harring, L. M. Stapleton, and S. N. Beretvas (Charlotte, NC: Information Age), 97–135.

Olsson, U. H., Foss, T., Troye, S. V., & Howell, R. D. (2000) The performance of ML, GLS, and WLS estimation in structural equation modeling under conditions of misspecification and nonnormality. *Structural Equation Modeling, 7*(4), 557-595.

Olsson, U. H., Troye, S. V., & Howell, R. D. (1999). Theoretical fit and empirical fit: The performance of maximum likelihood versus generalized least squares estimation in structural equation models. *Multivariate Behavioral Research, 34*(1), 31-58.

Pek, J., and MacCallum, R. C. (2011). Sensitivity analysis in structural equation models: cases and their influence. Multiv. Behav. Res. 46, 202–228. doi: 10.1080/00273171.2011.561068.

UNIVERSITY FOR DEVELOPMENT STUDIES

Pinheiro, J. C., Liu, C., and Wu, Y. N. (2001). Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t distribution. J. Comput. Graph. Stat. 10, 249–276. doi: 10.1198/10618600152628059

Ringle, C. M., Götz, O., Wetzels, M., & Wilson, B. (2009). On the use of formative measurement specifications in structural equation modeling: A Monte Carlo simulation study to compare covariance-based and partial least squares model estimation methodologies. In *METEOR Research Memoranda (RM/09/014)*: Maastricht University.

Saçkes, M. (2014). How often do early childhood teachers teach science concepts? determinants of the frequency of science teaching in kindergarten. *European Early Childhood Education Research Journal*, *22*(2), 169–184. doi: 10.1080/1350293X.2012.704305

Sanchez, B. N., Houseman, E. A. & Ryan, L. M. (2009). Residual-based diagnostics for structural equation models. Biometrics, 65, 104-115. *Jo*

Song, X.-Y., & Lee, S.-Y. (2004). Bayesian analysis of two level nonlinear structural equation models with continuous and polytomous data. *British Journal of Mathematical and Statistical Psychology*, *57*(1), 29–52. doi: 10.1348/0007110 04849259

Song, X.-Y., & Lee, S.-Y. (2012). A tutorial on the Bayesian approach for analyzing structural equation models. *Journal of Mathematical Psychology*, *56*(3), 135–148. doi: 10.1016/j.jmp.2012.02.001

Song, X.-Y., Chen, F., & Lu, Z.-H. (2013). A Bayesian semiparametric dynamic two-level structural equation model for analyzing non-normal longitudinal data. *Journal of Multivariate Analysis*, *121*, 87–108. doi: 10.1016/j.jmva.2013.06.001

Song, X.-Y., Lee, S.-Y., & Hser, Y.-I. (2009). Bayesian analysis of multivariate latent curve models with nonlinear longitudinal latent effects. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(2), 245–266. doi: 10.1080/107055 10902751275

Stan Development Team. (2014b). *Stan: A C++library for probability and sampling*, version 2.2. Retrieved from http://mcstan.org/

Sterba, S. K., and Pek, J. (2012). Individual influence on model selection. Psychol. Methods 17:582. doi: 10.1037/a0029253

Tanaka, J. S. (1987). "How big is big enough?": Sample size and goodness of fit in structural equation models with latent variables. *Child Development, 58*(1), 134-146.

Tami, M., Bry, X., Lavergne, C., (2015). Estimation of structural equation models with factors by EM algorithm. In: JDS 2014 Rennes. Rennes, France.

Tenenhaus, M. (2008). Component-based structural equation modelling. *Total Quality Management, 19*, 871-886.

Wirth, R., & Edwards, M. C. (2007). Item factor analysis: Current approaches and future directions. *Psychological Methods*, *12*(1), 58–79. doi: 10.1037/1082-989X.12.1.58

Wold, H. (1975). Path models with latent variables: The NIPALS approach. In H. M. Blalock, A. Aganbegian, F. M. Borodkin, R. Boudon, & V. Capecchi (Eds.), *Quantitative sociology: International perspectives on mathematical and statistical modeling* (pp. 307-357). New York: Academic.

Yang, M., & Dunson, D. B. (2010). Bayesian semiparametric structural equation models with latent variables. *Psychometrika*, *75*(4), 675–693. doi: 10.1007/s11336-010-9174-4

Yuan, K.-H., and Bentler, P. M. (2001). Effect of outliers on estimators and tests in covariance structure analysis. Br. J. Math. Stat. Psychol. 54, 161–175. doi: 10.1348/000711001159366

Yuan, K.-H., and Hayashi, K. (2010). Fitting data to model: structural equation modeling diagnosis using two scatter plots. Psychol. Methods 15, 335–351. doi: 10.1037/a0020140

Yuan, K.-H., and Zhang, Z. (2012). Structural equation modeling diagnostics using R package semdiag and EQS. Struc. Eq. Modeli. A Multidisc. J. 19, 683–702. doi: 10.1080/10705511.2012.713282

Yuan, K.-H., and Zhang, Z. (2012). Structural equation modeling diagnostics using R package semdiag and EQS. Struc. Eq. Modeli. A Multidisc. J. 19, 683–702. doi: 10.1080/10705511.2012.713282

Yuan, K.-H., and Zhang, Z. (2015). rsem: Robust Structural EquationModeling with Missing Data and Auxiliary Variables. R package version 0.4.6. Available online at: https://CRAN.R-project.org/package=rsem

Yuan, K.-H., and Zhong, X. (2008). Outliers, leverage observations, and influential cases in factor analysis: Using robust procedures to minimize their effect. Sociol. Methodol. 38, 329–368. doi: 10.1111/j.1467-9531.2008.00198.x

Yuan, K.-H., and Zhong, X. (2013). Robustness of fit indices to outliers and leverage observations in structural equation modeling. Psychol. Methods 18, 121–136. doi: 10.1037/a0031604.

UNIVERSITY FOR DEVELOPMENT STUDIES

APPENDICES

APPENDIX A

```
/*conditional bias*/
proc means MEAN data = errors noprint;
var CBeps1r CBeps2r CBeps3r CBdelta1r CBdelta2r CBdelta3r          CBzetar
    CBeps1b CBeps2b CBeps3b CBdelta1b CBdelta2b CBdelta3b CBzetab
CBeps1ar CBeps2ar CBeps3ar CBdelta1ar CBdelta2ar CBdelta3ar CBzetaar
CBeps1hatr CBeps2hatr CBeps3hatr CBdelta1hatr CBdelta2hatr CBdelta3hatr  CBzetahatr
    CBeps1hatb CBeps2hatb CBeps3hatb CBdelta1hatb CBdelta2hatb CBdelta3hatb CBzetahatb
CBeps1hatar CBeps2hatar CBeps3hatar CBdelta1hatar CBdelta2hatar CBdelta3hatar CBzetahatar; output
out = condbias mean = CBeps1r CBeps2r CBeps3r CBdelta1r CBdelta2r CBdelta3r CBzetar
    CBeps1b CBeps2b CBeps3b CBdelta1b CBdelta2b CBdelta3b CBzetab
CBeps1ar CBeps2ar CBeps3ar CBdelta1ar CBdelta2ar CBdelta3ar CBzetaar
CBeps1hatr CBeps2hatr CBeps3hatr CBdelta1hatr CBdelta2hatr CBdelta3hatr CBzetahatr
    CBeps1hatb CBeps2hatb CBeps3hatb CBdelta1hatb CBdelta2hatb CBdelta3hatb CBzetahatb
CBeps1hatar CBeps2hatar CBeps3hatar CBdelta1hatar CBdelta2hatar CBdelta3hatar CBzetahatar;
run;


proc means SUM data = errors noprint;
var bias2eps1r bias2eps2r bias2eps3r bias2delta1r bias2delta2r bias2delta3r bias2zetar
bias2eps1b bias2eps2b bias2eps3b bias2delta1b bias2delta2b bias2delta3b bias2zetab
bias2eps1ar bias2eps2ar bias2eps3ar bias2delta1ar bias2delta2ar bias2delta3ar bias2zetaar
bias2eps1hatr bias2eps2hatr bias2eps3hatr bias2delta1hatr bias2delta2hatr bias2delta3hatr
bias2zetahatr
bias2eps1hatb bias2eps2hatb bias2eps3hatb bias2delta1hatb bias2delta2hatb bias2delta3hatb
bias2zetahatb
bias2eps1hatar bias2eps2hatar bias2eps3hatar bias2delta1hatar
bias2delta2hatar bias2delta3hatar bias2zetahatar; output out = mse sum =
MSEeps1r MSEeps2r MSEeps3r MSEdelta1r MSEdelta2r MSEdelta3r MSEzetar
MSEeps1b MSEeps2b MSEeps3b MSEdelta1b MSEdelta2b MSEdelta3b MSEzetab
MSEeps1ar MSEeps2ar MSEeps3ar MSEdelta1ar MSEdelta2ar MSEdelta3ar MSEzetaar
MSEeps1hatr MSEeps2hatr MSEeps3hatr MSEdelta1hatr MSEdelta2hatr MSEdelta3hatr MSEzetahat
MSEeps1hatb MSEeps2hatb MSEeps3hatb MSEdelta1hatb MSEdelta2hatb MSEdelta3hatb MSEzetaha
MSEeps1hatar MSEeps2hatar MSEeps3hatar MSEdelta1hatar MSEdelta2hatar MSEdelta3hatar
MSEzetahatar;
run;


/*univocality*/
proc corr data = errors COV outp = unir noprint;
var eps1R eps2R eps3R delta1R delta2R delta3R zetaR;
with eps1 eps2 eps3 delta1 delta2 delta3 zeta;
run;
proc corr data = errors COV outp = unib noprint;
var eps1B eps2B eps3B delta1B delta2B delta3B zetaB;
with eps1 eps2 eps3 delta1 delta2 delta3 zeta;
run;
proc corr data = errors COV outp = uniar noprint;
var eps1AR eps2AR eps3AR delta1AR delta2AR delta3AR zetaAR;
with eps1 eps2 eps3 delta1 delta2 delta3 zeta;
run;
```

```
proc corr data = errors COV outp = unirhat noprint;
var eps1hatR eps2hatR eps3hatR delta1hatR delta2hatR delta3hatR zetahatR;
with eps1 eps2 eps3 delta1 delta2 delta3 zeta;
run;
proc corr data = errors COV outp = unibhat noprint;
var eps1hatB eps2hatB eps3hatB delta1hatB delta2hatB delta3hatB zetahatB; with eps1 eps2 eps3 delta1 delta2
delta3 zeta;
run;
proc corr data = errors COV outp = uniarhat noprint;

var eps1hatAR eps2hatAR eps3hatAR delta1hatAR delta2hatAR delta3hatAR zetahatAR; with eps1 eps2 eps3
delta1 delta2 delta3 zeta;
run;
```

```
data kstu;
input id mec vec alg ana
sta;
/*mec = mechanics, closed
book
  vec = vectors, closed book
  alg = algebra,open book
  ana = analysis, open book
  sta=statistics,     open
book*/


proc means data =
mardia;
run;
```

```
/*create dummies for high leverages points and potential
outliers as taken from Yuan and Hayashi (2010)*/
data kstu;
set kstu;
if id = 81 or id = 87 then
highlev = 1; else highlev = 0;
if id = 1 then poutlier = 1;
else if id = 2 then poutlier = 2;
else if id = 3 then poutlier = 3;
else if id = 6 then poutlier = 4;
else if id = 23 then poutlier = 5;
else if id = 28 then poutlier = 6;
else if id = 29 then poutlier = 7;
else if id = 54 then poutlier = 8;
else if id = 56 then poutlier = 9;
else if id = 61 then poutlier = 10;
else if id = 81 then poutlier = 11;
else if id = 87 then poutlier = 12;
else if id = 88 then poutlier = 13;
else poutlier = 0;
run;
proc standard data = kstu out=kstustan
mean                =                 0;
run;

/*use PROC CALIS to fit the model*/
PROC CALIS data = kstustan outest = ests  outstat =           stat;
factor
closed -> mec = 1.0,
closed -> vec =
lambda2,
open -> alg = 1.0,
open -> ana =
lambda4,
open -> sta =
lambda5;
cov
closed  open  =
phi21;
pvar
mec = delta1,
vec = delta2,
alg = delta3,
ana = delta4,
sta = delta5,
closed = phi11,
open = phi22;
run;
```

152

```
/*read PROC CALIS estimates into PROC IML and create needed matrices*/

PROC IML;

/*id*/

use kstu;

read all var {id highlev poutlier} into idstuff;

/*original data*/

use kstu;

read all var {mec vec alg ana sta} into data;

/*center the data*/

cdata = data - data[:,];

means = data[:,];

/*factor   loadings*/

use ests;

read all var {lambda2} into lambda2hat where (_TYPE_="PARMS");

read all var {lambda4} into lambda4hat where (_TYPE_="PARMS");

read all var {lambda5} into lambda5hat where (_TYPE_="PARMS");

lambdahat = (1.0 || 0) //

(lambda2hat  ||  0)  //

(0 || 1.0) //

(0 || lambda4hat)//

(0 || lambda5hat);

/* covariance matrices*/

/*covariance matrix of the factors*/

use ests;
```

```
read all var {phi11} into phi11hat where (_TYPE_="PARMS");

read all var {phi22} into phi22hat where (_TYPE_="PARMS");

read all var {phi21} into phi21hat where (_TYPE_="PARMS");

Sigma_LLhat = (phi11hat || phi21hat)//

   (phi21hat || phi22hat);

/*covariance matrix of the measurement errors*/

use ests;

read all var {delta1} into delta1hat where (_TYPE_="PARMS");

read all var {delta2} into delta2hat where (_TYPE_="PARMS");

read all var {delta3} into delta3hat where (_TYPE_="PARMS");

read all var {delta4} into delta4hat where (_TYPE_="PARMS");

read all var {delta5} into delta5hat where (_TYPE_="PARMS");

Sigma_nunuhat = (delta1hat || J(1,4,0))//

(0 || delta2hat || J(1,3,0))//

(J(1,2,0) || delta3hat || J(1,2,0))//

(J(1,3,0) || delta4hat || 0)//

(J(1,4,0) || delta5hat);

/*covariance matrix of the observed variables*/

Sigma_zzhat = Lambdahat*Sigma_LLhat*Lambdahat'+Sigma_nunuhat;


/*construct weight matrices*/

/*regression method*/

Wrhat   =   Sigma_LLhat*Lambdahat'*inv(Sigma_zzhat);

/*Bartlett's method*/

Wbhat   =   (inv(Lambdahat'*inv(Sigma_nunuhat)*Lambdahat))*Lambdahat'*inv(Sigma_nunuhat);

/*Anderson-Rubin method*/

Asq   =   (Lambdahat'*inv(Sigma_nunuhat)*Sigma_zzhat*inv(Sigma_nunuhat)*Lambdahat);

call eigen(evals, evects, Asq);

A   =   evects*diag(sqrt(evals))*evects';

Ainv = inv(A);
Warhat = Ainv*Lambdahat'*inv(Sigma_nunuhat);
```

```
/*residuals*/
/*regression method based residuals*/
nuhatR = ((I(5) - Lambdahat*Wrhat)*cdata')';

/*Bartlett's method based residuals*/
nuhatB = ((I(5) - Lambdahat*Wbhat)*cdata')';
/*Anderson-Rubin method based residuals*/
nuhatAR = ((I(5) - Lambdahat*Warhat)*cdata')';

/*predicted values*/
/*regression method based predicted values*/
zhatR = (means'+Lambdahat*Wrhat*cdata')';
/*Bartlett's method based predicted values*/
zhatB = (means'+Lambdahat*Wbhat*cdata')';
/*Anderson-Rubin  method  based  predicted  values*/
zhatAR = (means'+Lambdahat*Warhat*cdata')';


 /*standardize the data*/

 /*calculate covariance matrices*/
 CovzhatnuhatR = Lambdahat*Wrhat*Sigma_zzhat*(I(5)-Lambdahat*Wrhat)';
 CovzhatnuhatB = Lambdahat*Wbhat*Sigma_zzhat*(I(5)-Lambdahat*Wbhat)';
 CovzhatnuhatAR = Lambdahat*Warhat*Sigma_zzhat*(I(5)-Lambdahat*Warhat)';
 VarzhatR = Lambdahat*Wrhat*Sigma_zzhat*Wrhat'*Lambdahat';
 VarzhatB = Lambdahat*Wbhat*Sigma_zzhat*Wbhat'*Lambdahat';
 VarzhatAR = Lambdahat*Warhat*Sigma_zzhat*Warhat'*Lambdahat';
 VarnuhatR = (I(5)-Lambdahat*Wrhat)*Sigma_zzhat*(I(5)-Lambdahat*Wrhat)';
 VarnuhatB = (I(5)-Lambdahat*Wbhat)*Sigma_zzhat*(I(5)-Lambdahat*Wbhat)';
 VarnuhatAR = (I(5)-Lambdahat*Warhat)*Sigma_zzhat*(I(5)-Lambdahat*Warhat)';
 CovLhatR = Wrhat*Sigma_zzhat*Wrhat';
 CovLhatB = Wbhat*Sigma_zzhat*Wbhat';
 CovLhatAR = Warhat*Sigma_zzhat*Warhat';


 delta1hatRst = inv(sqrt(VarnuhatR[1,1]))*nuhatR[,1];
 delta2hatRst = inv(sqrt(VarnuhatR[2,2]))*nuhatR[,2];
 delta3hatRst = inv(sqrt(VarnuhatR[3,3]))*nuhatR[,3];
 delta4hatRst = inv(sqrt(VarnuhatR[4,4]))*nuhatR[,4];
 delta5hatRst = inv(sqrt(VarnuhatR[5,5]))*nuhatR[,5];
 delta1hatBst = inv(sqrt(VarnuhatB[1,1]))*nuhatB[,1];
 delta2hatBst = inv(sqrt(VarnuhatB[2,2]))*nuhatB[,2];
 delta3hatBst = inv(sqrt(VarnuhatB[3,3]))*nuhatB[,3];
 delta4hatBst = inv(sqrt(VarnuhatB[4,4]))*nuhatB[,4];
 delta5hatBst = inv(sqrt(VarnuhatB[5,5]))*nuhatB[,5];
 delta1hatARst  =   inv(sqrt(VarnuhatAR[1,1]))*nuhatAR[,1];
 delta2hatARst  =   inv(sqrt(VarnuhatAR[2,2]))*nuhatAR[,2];
 delta3hatARst  =   inv(sqrt(VarnuhatAR[3,3]))*nuhatAR[,3];
 delta4hatARst  =   inv(sqrt(VarnuhatAR[4,4]))*nuhatAR[,4];
 delta5hatARst = inv(sqrt(VarnuhatAR[5,5]))*nuhatAR[,5];
```

155

```
CovzhatnuhatRst = inv(sqrt(diag(VarnuhatR)))*CovzhatnuhatR;

CovzhatnuhatBst = inv(sqrt(diag(VarnuhatB)))*CovzhatnuhatB;

CovzhatnuhatARst = inv(sqrt(diag(VarnuhatAR)))*CovzhatnuhatAR;

/*rotate the standardized data*/

d1mecRst = (zhatR[,1] || delta1hatRst[,1]);

if CovzhatnuhatRst[1,1]/VarzhatR[1,1] >=0 then thetaRd1Rst = -ATAN(abs(CovzhatnuhatRst[1,1]/VarzhatR[1,1]));

else thetaRd1Rst = 2*constant('PI')-ATAN(abs(CovzhatnuhatRst[1,1]/VarzhatR[1,1]));

Rd1Rst = (cos(thetaRd1Rst) || -sin(thetaRd1Rst)) //

(sin(thetaRd1Rst)  || cos(thetaRd1Rst));

d1mecRrst = d1mecRst*Rd1Rst;

 Sigma_mecd1Rst = (VarzhatR[1,1]  || CovzhatnuhatRst[1,1]) //
   (CovzhatnuhatRst[1,1] || 1);

 Ud1Rst = root(Sigma_mecd1Rst);

 Ld1Rst=Ud1Rst';

 d1mecrCst = (trisolv(4,Ld1Rst,d1mecRst'))';

 d2vecRst = (zhatR[,2] || delta2hatRst[,1]);

 if CovzhatnuhatRst[2,2]/VarzhatR[2,2] >=0 then thetaRd2Rst = -ATAN(abs(CovzhatnuhatRst[2,2]/VarzhatR[2,2]));

 else thetaRd2Rst = 2*constant('PI') - ATAN(abs(CovzhatnuhatRst[2,2]/VarzhatR[2,2]));

 Rd2Rst = (cos(thetaRd2Rst) || -sin(thetaRd2Rst)) //

(sin(thetaRd2Rst)  || cos(thetaRd2Rst));

d2vecRrst = d2vecRst*Rd2Rst;


Sigma_vecd2Rst = (VarzhatR[2,2]  || CovzhatnuhatRst[2,2]) //
   (CovzhatnuhatRst[2,2] || 1);

Ud2Rst = root(Sigma_vecd2Rst);

Ld2Rst=Ud2Rst';

d2vecrCst = (trisolv(4,Ld2Rst,d2vecRst'))';


d3algRst = (zhatR[,3] || delta2hatRst[,1]);

if CovzhatnuhatRst[3,3]/VarzhatR[3,3] >=0 then thetaRd3Rst = -ATAN(abs(CovzhatnuhatRst[3,3]/VarzhatR[3,3]));

else thetaRd3Rst = 2*constant('PI') - ATAN(abs(CovzhatnuhatRst[3,3]/VarzhatR[3,3]));

Rd3Rst = (cos(thetaRd3Rst) || -sin(thetaRd3Rst)) //

(sin(thetaRd3Rst)  || cos(thetaRd3Rst));

d3algRrst = d3algRst*Rd3Rst;
```

```
Sigma_algd3Rst  =  (VarzhatR[3,3]  ||  CovzhatnuhatRst[3,3])  //

  (CovzhatnuhatRst[3,3] || 1);

Ud3Rst  =  root(Sigma_algd3Rst);

Ld3Rst=Ud3Rst';

d3algrCst = (trisolv(4,Ld3Rst,d3algRst'))';


d4anaRst = (zhatR[,4] || delta4hatRst[,1]);

if CovzhatnuhatRst[4,4]/VarzhatR[4,4] >=0 then thetaRd4Rst = -ATAN(abs(CovzhatnuhatRst[4,4]/VarzhatR[4,4]));

else thetaRd4Rst = 2*constant('PI') - ATAN(abs(CovzhatnuhatRst[4,4]/VarzhatR[4,4]));

Rd4Rst = (cos(thetaRd4Rst) || -sin(thetaRd4Rst)) //

(sin(thetaRd4Rst)   ||   cos(thetaRd4Rst));

d4anaRrst = d4anaRst*Rd4Rst;


Sigma_anad4Rst  =  (VarzhatR[4,4]  ||  CovzhatnuhatRst[4,4])  //

  (CovzhatnuhatRst[4,4] || 1);

Ud4Rst   =   root(Sigma_anad4Rst);

Ld4Rst=Ud4Rst';

d4anarCst = (trisolv(4,Ld4Rst,d4anaRst'))';


d5staRst = (zhatR[,5] || delta5hatRst[,1]);

if CovzhatnuhatRst[5,5]/VarzhatR[5,5] >=0 then thetaRd5Rst = -ATAN(abs(CovzhatnuhatRst[5,5]/VarzhatR[5,5]));

else thetaRd5Rst = 2*constant('PI') - ATAN(abs(CovzhatnuhatRst[5,5]/VarzhatR[5,5]));

Rd5Rst = (cos(thetaRd5Rst) || -sin(thetaRd5Rst)) //

(sin(thetaRd5Rst)   ||   cos(thetaRd5Rst));

d5staRrst = d5staRst*Rd5Rst;


Sigma_stad5Rst  =  (VarzhatR[5,5]  ||  CovzhatnuhatRst[5,5])  //

  (CovzhatnuhatRst[5,5] || 1);
```

157

```
Ud5Rst  =  root(Sigma_stad5Rst);

Ld5Rst=Ud5Rst';

d5starCst = (trisolv(4,Ld5Rst,d5staRst'))';


d1mecBst = (zhatB[,1] || delta1hatBst[,1]);

if CovzhatnuhatBst[1,1]/VarzhatB[1,1] >=0 then thetaRd1Bst = -ATAN(abs(CovzhatnuhatBst[1,1]/VarzhatB[1,1]));

else thetaRd1Bst = 2*constant('PI') - ATAN(abs(CovzhatnuhatBst[1,1]/VarzhatB[1,1]));

Rd1Bst = (cos(thetaRd1Bst) || -sin(thetaRd1Bst)) //

(sin(thetaRd1Bst)  ||  cos(thetaRd1Bst));

d1mecBrst = d1mecBst*Rd1Bst;


Sigma_mecd1Bst  =  (VarzhatB[1,1]  ||  CovzhatnuhatBst[1,1])  //

  (CovzhatnuhatBst[1,1] || 1);

Ud1Bst  =  root(Sigma_mecd1Bst);

Ld1Bst=Ud1Bst';

d1mecbCst = (trisolv(4,Ld1Bst,d1mecBst'))';


d2vecBst = (zhatB[,2] || delta2hatBst[,1]);

if CovzhatnuhatBst[2,2]/VarzhatB[2,2] >=0 then thetaRd2Bst = -ATAN(abs(CovzhatnuhatBst[2,2]/VarzhatB[2,2]));

else thetaRd2Bst = 2*constant('PI') - ATAN(abs(CovzhatnuhatBst[2,2]/VarzhatB[2,2]));

Rd2Bst = (cos(thetaRd2Bst) || -sin(thetaRd2Bst)) //

(sin(thetaRd2Bst)  ||  cos(thetaRd2Bst));

d2vecBrst = d2vecBst*Rd2Bst;


Sigma_vecd2Bst  =  (VarzhatB[2,2]  ||  CovzhatnuhatBst[2,2])  //

  (CovzhatnuhatBst[2,2] || 1);

Ud2Bst  =  root(Sigma_vecd2Bst);

Ld2Bst=Ud2Bst';

d2vecBcst = (trisolv(4,Ld2Bst,d2vecBst'))';


d3algBst = (zhatB[,3] || delta3hatBst[,1]);

if CovzhatnuhatBst[3,3]/VarzhatB[3,3] >=0 then thetaRd3Bst = -ATAN(abs(CovzhatnuhatBst[3,3]/VarzhatB[3,3]));

else thetaRd3Bst = 2*constant('PI') - ATAN(abs(CovzhatnuhatBst[3,3]/VarzhatB[3,3]));

Rd3Bst = (cos(thetaRd3Bst) || -sin(thetaRd3Bst)) //

(sin(thetaRd3Bst)  ||  cos(thetaRd3Bst));

d3algBrst = d3algBst*Rd3Bst;
```

```
Sigma_algd3Bst  =  (VarzhatB[3,3]  ||  CovzhatnuhatBst[3,3])  //
   (CovzhatnuhatBst[3,3] || 1);

Ud3Bst   =   root(Sigma_algd3Bst);

Ld3Bst=Ud3Bst';

d3algbCst = (trisolv(4,Ld3Bst,d3algBst'))';


d4anaBst = (zhatB[,4] || delta4hatBst[,1]);

if CovzhatnuhatBst[4,4]/VarzhatB[4,4] >=0 then thetaRd4Bst = -ATAN(abs(CovzhatnuhatBst[4,4]/VarzhatB[4,4]));

else thetaRd4Bst = 2*constant('PI') - ATAN(abs(CovzhatnuhatBst[4,4]/VarzhatB[4,4]));


Rd4Bst = (cos(thetaRd4Bst) || -sin(thetaRd4Bst)) //

(sin(thetaRd4Bst) || cos(thetaRd4Bst));

d4anaBrst = d4anaBst*Rd4Bst;


Sigma_anad4Bst  =  (VarzhatB[4,4]  ||  CovzhatnuhatBst[4,4])  //
   (CovzhatnuhatBst[4,4] || 1);

Ud4Bst   =   root(Sigma_anad4Bst);

Ld4Bst=Ud4Bst';

d4anabCst = (trisolv(4,Ld4Bst,d4anaBst'))';
```

```
d5staBst = (zhatB[,5] || delta5hatBst[,1]);
if CovzhatnuhatBst[5,5]/VarzhatB[5,5] >=0 then thetaRd5Bst = -ATAN(abs(CovzhatnuhatBst[5,5]/VarzhatB[5,5]));
else thetaRd5Bst = 2*constant('PI') - ATAN(abs(CovzhatnuhatBst[5,5]/VarzhatB[5,5]));
Rd5Bst = (cos(thetaRd5Bst) || -sin(thetaRd5Bst)) //
(sin(thetaRd5Bst)   ||   cos(thetaRd5Bst));
d5staBrst = d5staBst*Rd5Bst;


Sigma_stad5Bst  =  (VarzhatB[5,5]  ||  CovzhatnuhatBst[5,5])  //
   (CovzhatnuhatBst[5,5] || 1);
Ud5Bst   =   root(Sigma_stad5Bst);
Ld5Bst=Ud5Bst';
d5stabCst = (trisolv(4,Ld5Bst,d5staBst'))';


d1mecARst = (zhatAR[,1] || delta1hatARst[,1]);
if CovzhatnuhatARst[1,1]/VarzhatAR[1,1] >=0 then thetaRd1ARst = -ATAN(abs(CovzhatnuhatARst[1,1]/VarzhatAR[1,1]));
else thetaRd1ARst = 2*constant('PI') - ATAN(abs(CovzhatnuhatARst[1,1]/VarzhatAR[1,1]));
Rd1ARst = (cos(thetaRd1ARst) || -sin(thetaRd1ARst)) //
(sin(thetaRd1ARst)   ||   cos(thetaRd1ARst));
d1mecARrst = d1mecARst*Rd1ARst;


Sigma_mecd1ARst  =  (VarzhatAR[1,1]  ||  CovzhatnuhatARst[1,1])  //
   (CovzhatnuhatARst[1,1] || 1);
Ud1ARst   =   root(Sigma_mecd1ARst);
Ld1ARst=Ud1ARst';
d1mecarCst = (trisolv(4,Ld1ARst,d1mecARst'))';


d2vecARst = (zhatAR[,2] || delta2hatARst[,1]);
if CovzhatnuhatARst[2,2]/VarzhatAR[2,2] >=0 then thetaRd2ARst = -ATAN(abs(CovzhatnuhatARst[2,2]/VarzhatAR[2,2]));
else thetaRd2ARst = 2*constant('PI') - ATAN(abs(CovzhatnuhatARst[2,2]/VarzhatAR[2,2]));
Rd2ARst = (cos(thetaRd2ARst) || -sin(thetaRd2ARst)) //
(sin(thetaRd2ARst)   ||   cos(thetaRd2ARst));
d2vecARrst = d2vecARst*Rd2ARst;
```

160

```
Sigma_vecd2ARst = (VarzhatAR[2,2] || CovzhatnuhatARst[2,2]) //
   (CovzhatnuhatARst[2,2] || 1);
Ud2ARst   =   root(Sigma_vecd2ARst);
Ld2ARst=Ud2ARst';
d2vecarCst = (trisolv(4,Ld2ARst,d2vecARst'))';


d3algARst = (zhatAR[,3] || delta3hatARst[,1]);
if CovzhatnuhatARst[3,3]/VarzhatAR[3,3] >=0 then thetaRd3ARst = -ATAN(abs(CovzhatnuhatARst[3,3]/VarzhatAR[3,3]));
else thetaRd3ARst = 2*constant('PI') - ATAN(abs(CovzhatnuhatARst[3,3]/VarzhatAR[3,3]));
Rd3ARst = (cos(thetaRd3ARst) || -sin(thetaRd3ARst)) //
(sin(thetaRd3ARst) || cos(thetaRd3ARst));

d3algARrst = d3algARst*Rd3ARst;


Sigma_algd3ARst  =  (VarzhatAR[3,3]  ||  CovzhatnuhatARst[3,3])  //
   (CovzhatnuhatARst[3,3] || 1);
Ud3ARst   =   root(Sigma_algd3ARst);
Ld3ARst=Ud3ARst';
d3algarCst = (trisolv(4,Ld3ARst,d3algARst'))';


d4anaARst = (zhatAR[,4] || delta4hatARst[,1]);
if CovzhatnuhatARst[4,4]/VarzhatAR[4,4] >=0 then thetaRd4ARst = -ATAN(abs(CovzhatnuhatARst[4,4]/VarzhatAR[4,4]));


else thetaRd4ARst = 2*constant('PI') - ATAN(abs(CovzhatnuhatARst[4,4]/VarzhatAR[4,4]));
Rd4ARst = (cos(thetaRd4ARst) || -sin(thetaRd4ARst)) //
(sin(thetaRd4ARst)   ||   cos(thetaRd4ARst));
d4anaARrst = d4anaARst*Rd4ARst;


Sigma_anad4ARst  =  (VarzhatAR[4,4]  ||  CovzhatnuhatARst[4,4])  //
   (CovzhatnuhatARst[4,4] || 1);
Ud4ARst   =   root(Sigma_anad4ARst);
Ld4ARst=Ud4ARst';
d4anaarCst = (trisolv(4,Ld4ARst,d4anaARst'))';


d5staARst = (zhatAR[,5] || delta5hatARst[,1]);
if CovzhatnuhatARst[5,5]/VarzhatAR[5,5] >=0 then thetaRd5ARst = -ATAN(abs(CovzhatnuhatARst[5,5]/VarzhatAR[5,5]));
else thetaRd5ARst = 2*constant('PI') - ATAN(abs(CovzhatnuhatARst[5,5]/VarzhatAR[5,5]));
Rd5ARst = (cos(thetaRd5ARst) || -sin(thetaRd5ARst)) //
(sin(thetaRd5ARst)   ||   cos(thetaRd5ARst));
d5staARrst = d5staARst*Rd5ARst;


Sigma_stad5ARst  =  (VarzhatAR[5,5]  ||  CovzhatnuhatARst[5,5])  //
   (CovzhatnuhatARst[5,5] || 1);
Ud5ARst   =   root(Sigma_stad5ARst);
Ld5ARst=Ud5ARst';
d5staarCst = (trisolv(4,Ld5ARst,d5staARst'))';
```

161

```
/*create data sets*/

mardiaunrst = data || delta1hatRst || delta2hatRst || delta3hatRst || delta4hatRst    || delta5hatRst   ||

    delta1hatBst || delta2hatBst || delta3hatBst || delta4hatBst || delta5hatBst ||

    delta1hatARst || delta2hatARst || delta3hatARst || delta4hatARst || delta5hatARst ||

    zhatR || zhatB || zhatAR || idstuff;

cname4 = {"mec" "vec" "alg" "ana" "sta"

  "delta1hatRst" "delta2hatRst" "delta3hatRst" "delta4hatRst" "delta5hatRst"

  "delta1hatBst" "delta2hatBst" "delta3hatBst" "delta4hatBst" "delta5hatBst"

  "delta1hatARst" "delta2hatARst" "delta3hatARst"  "delta4hatARst" "delta5hatARst"

  "mechatR" "vechatR" "alghatR" "anahatR" "stahatR"

  "mechatB" "vechatB" "alghatB" "anahatB" "stahatB"

  "mechatAR" "vechatAR"  "alghatAR"  "anahatAR"  "stahatAR"

 "id" "highlev" "poutlier"};

create  mardiaunrst  from  mardiaunrst[colname=cname4];

append from mardiaunrst;


mardiarst = data || d1mecRrst || d2vecRrst || d3algRrst || d4anaRrst || d5staRrst ||

    d1mecBrst || d2vecBrst || d3algBrst || d4anaBrst || d5staBrst ||

    d1mecARrst || d2vecARrst || d3algARrst || d4anaARrst || d5staARrst || idstuff;

cname2 = {"mec" "vec" "alg" "ana" "sta"

  "mechatRr" "delta1hatRr" "vechatRr" "delta2hatRr" "alghatRr" "delta3hatRr" "anahatRr" "delta4hatRr" "stahatRr" "delta5hatRr"

  "mechatBr" "delta1hatBr" "vechatBr" "delta2hatBr" "alghatBr" "delta3hatBr" "anahatBr" "delta4hatBr" "stahatBr" "delta5hatBr"

  "mechatARr" "delta1hatARr" "vechatARr" "delta2hatARr" "alghatARr" "delta3hatARr" "anahatARr" "delta4hatARr" "stahatARr"

  "delta5hatARr" "id" "highlev" "poutlier"};

create  mardiarst  from  mardiarst[colname=cname2];

append from mardiarst;


mardiarCst = data || d1mecrCst || d2vecrCst || d3algrCst || d4anarCst || d5starCst ||

    d1mecBCst || d2vecBCst || d3algBCst || d4anaBCst || d5staBCst ||

    d1mecARCst || d2vecARCst || d3algARCst || d4anaARCst || d5staARCst || idstuff;

cname3 = {"mec" "vec" "alg" "ana" "sta"
```

```
"mechatRr" "delta1hatRr" "vechatRr" "delta2hatRr" "alghatRr" "delta3hatRr" "anahatRr" "delta4hatRr" "stahatRr" "delta5hatRr"

"mechatBr" "delta1hatBr" "vechatBr" "delta2hatBr" "alghatBr" "delta3hatBr" "anahatBr" "delta4hatBr" "stahatBr" "delta5hatBr"

"mechatARr" "delta1hatARr" "vechatARr" "delta2hatARr" "alghatARr" "delta3hatARr" "anahatARr" "delta4hatARr" "stahatARr" "delta5hatARr" "id"

"highlev" "poutlier"};

create  mardiarCst  from  mardiarCst[colname=cname3];

append from mardiarCst;


mardiaCD = data || idstuff || CDR1 || CDB1 || CDAR1 || CDR2 || CDB2 || CDAR2 || CDR3 || CDB3 || CDAR3 ||

          CDR4 || CDB4 || CDAR4 || CDR5 || CDB5 || CDAR5;

cnameCD = {"mec" "vec" "alg" "ana" "sta" "id" "highlev" "poutlier" "CDR1" "CDB1" "CDAR1"

    "CDR2" "CDB2" "CDAR2" "CDR3" "CDB3" "CDAR3" "CDR4" "CDB4" "CDAR4" "CDR5" "CDB5" "CDAR5"};

create mardiaCD from mardiaCD[colname=cnameCD];

append from mardiaCD;

run;quit;
```

APPENDIX B

The SAS System

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
id	100	41.6800000	37.6309102	1.0000000	304.0000000
y1	100	5.7993333	3.6795223	1.2500000	28.0000000
y2	100	4.2506651	3.9488139	0	9.9999980
y3	100	6.5186658	3.3019791	0	10.0000000
y4	100	4.4969775	3.3565237	0	10.0000000
y5	100	5.1488889	2.6013855	0	10.0000000
y6	100	3.0098815	3.3554722	0	9.9999980
y7	100	6.1538963	3.3456306	0	9.9999980
y8	100	4.0708684	3.2291615	0	10.0000000
x1	100	5.0136142	0.8570500	1.1166660	6.7369670
x2	100	4.7986681	1.5069082	1.3862940	7.8720740
x3	100	3.5662491	1.4045131	1.0016740	6.4245910
poutlier	100	1.2133333	3.0944298	0	13.0000000

The SAS System

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
id	100	1.913728E-14	37.6309102	-40.6800000	262.3200000
y1	100	-2.36848E-17	3.6795223	-4.5493333	22.2006667
y2	100	-1.40332E-15	3.9488139	-4.2506651	5.7493329
y3	100	-3.76588E-15	3.3019791	-6.5186658	3.4813342
y4	100	-6.75016E-16	3.3565237	-4.4969775	5.5030225
y5	100	9.473903E-17	2.6013855	-5.1488889	4.8511111
y6	100	-1.58096E-15	3.3554722	-3.0098815	6.9901165
y7	100	-3.75403E-15	3.3456306	-6.1538963	3.8461017
y8	100	2.250052E-16	3.2291615	-4.0708684	5.9291316
x1	100	3.067176E-15	0.8570500	-3.8969482	1.7233528
x2	100	-1.3974E-15	1.5069082	-3.4123741	3.0734059
x3	100	-6.36528E-16	1.4045131	-2.5645751	2.8583419
poutlier	100	-2.60532E-16	3.0944298	-1.2133333	11.7866667

The SAS System

The CALIS Procedure
Covariance Structure Analysis: Model and Initial Values

Modeling Information

Maximum Likelihood Estimation

Data Set	WORK.STANPOLDEM
N Records Read	100
N Records Used	100
N Obs	100
Model Type	PATH
Analysis	Covariances

Variables in the Model

Endogenous	**Manifest**	x1 x2 x3 y1 y2 y3 y4 y5 y6 y7 y8
	Latent	poldem60 poldem65
Exogenous	**Manifest**	
	Latent	ind60

Number of Endogenous Variables = 13
Number of Exogenous Variables = 1

Initial Estimates for PATH List

Path			Parameter	Estimate
ind60	===>	x1		1.00000
ind60	===>	x2	lambda2	.
ind60	===>	x3	lambda3	.
poldem60	===>	y1		1.00000
poldem60	===>	y2	lambda5	.
poldem60	===>	y3	lambda6	.
poldem60	===>	y4	lambda7	.
poldem65	===>	y5		1.00000
poldem65	===>	y6	lambda9	.
poldem65	===>	y7	lambda10	.
poldem65	===>	y8	lambda11	.
ind60	===>	poldem60	gamma11	.
ind60	===>	poldem65	gamma21	.
poldem60	===>	poldem65	beta21	.

165

Initial Estimates for Variance Parameters

Variance Type	Variable	Parameter	Estimate
Exogenous	ind60	phi11	.
Error	poldem60	psi11	.
	poldem65	psi22	.
	x1	thetadelta11	.
	x2	thetadelta22	.
	x3	thetadelta33	.
	y1	thetaeps11	.
	y2	thetaeps22	.
	y3	thetaeps33	.
	y4	thetaeps44	.
	y5	thetaeps55	.
	y6	thetaeps66	.
	y7	thetaeps77	.
	y8	thetaeps88	.

Initial Estimates for Covariances Among Errors

Error of	Error of	Parameter	Estimate
y1	y5	thetaeps15	.
y2	y6	thetaeps26	.
y3	y7	thetaeps37	.
y4	y8	thetaeps48	.
y2	y4	thetaeps24	.
y6	y8	thetaeps68	.

The SAS System

The CALIS Procedure
Covariance Structure Analysis: Descriptive Statistics

Simple Statistics

Variable	Mean	Std Dev
y1	0	3.67952
y2	0	3.94881
y3	0	3.30198
y4	0	3.35652
y5	0	2.60139
y6	0	3.35547
y7	0	3.34563
y8	0	3.22916
x1	0	0.85705
x2	0	1.50691
x3	0	1.40451

The SAS System

The CALIS Procedure
Covariance Structure Analysis: Optimization

Initial Estimation Methods

1	Instrumental Variables Method
2	McDonald Method
3	Two-Stage Least Squares

**Optimization Start
Parameter Estimates**

N	Parameter	Estimate	Gradient
1	**lambda2**	2.01288	-0.04100
2	**lambda3**	1.61197	-0.08009
3	**lambda5**	1.39168	-0.02555
4	**lambda6**	1.01635	-0.06782
5	**lambda7**	1.45850	-0.03016
6	**lambda9**	1.25560	0.05569
7	**lambda10**	1.32677	0.10023

Optimization Start
Parameter Estimates

N	Parameter	Estimate	Gradient
8	lambda11	1.35445	-0.03200
9	gamma11	1.16533	-0.03268
10	gamma21	0.50643	0.05010
11	beta21	0.85790	0.26609
12	phi11	0.53396	0.00887
13	psi11	3.75732	0.01550
14	psi22	0.25538	-0.13875
15	thetadelta11	0.20058	-1.22852
16	thetadelta22	0.10735	-0.12016
17	thetadelta33	0.58520	0.20143
18	thetaeps11	9.05646	-0.0002096
19	thetaeps22	6.91171	0.00463
20	thetaeps33	6.27284	0.00322
21	thetaeps44	1.73107	-0.03792
22	thetaeps55	2.53518	-0.01754
23	thetaeps66	4.58725	-0.00584
24	thetaeps77	3.74358	-0.03508
25	thetaeps88	2.66364	-0.00261
26	thetaeps15	0.25456	-0.03118
27	thetaeps26	2.09001	-0.01324
28	thetaeps37	1.66880	0.00832
29	thetaeps48	-0.27790	-0.05450
30	thetaeps24	0.35598	-0.01342
31	thetaeps68	0.93788	-0.03419

Value of Objective Function = 1.174826315

The SAS System

The CALIS Procedure
Covariance Structure Analysis: Optimization
Levenberg-Marquardt Optimization
Scaling Update of More (1978)

Parameter Estimates 31

Functions (Observations) 66

Optimization Start

Active Constraints	0	**Objective Function** 1.174826315
Max Abs Gradient Element	1.2285182449	**Radius** 5.5804208865

eration	Restarts	Function Calls	Active Constraints	Objective Function	Objective Function Change	Max Abs Gradient Element	Lambda	Ratio Between Actual and Predicted Change
1	0	4	0	1.06421	0.1106	0.0495	0	0.881
2	0	6	0	1.06190	0.00231	0.00911	0	0.835
3	0	8	0	1.06174	0.000166	0.00276	0	0.720
4	0	10	0	1.06172	0.000018	0.00103	0	0.664
5	0	12	0	1.06172	2.324E-6	0.000365	0	0.627
6	0	14	0	1.06172	3.411E-7	0.000149	0	0.608
7	0	16	0	1.06172	5.38E-8	0.000058	0	0.599
8	0	18	0	1.06172	8.794E-9	0.000024	0	0.595
9	0	20	0	1.06172	1.463E-9	9.835E-6	0	0.592

Optimization Results

Iterations	9	**Function Calls**	23
Jacobian Calls	11	**Active Constraints**	0
Objective Function	1.0617167208	**Max Abs Gradient Element**	9.834553E-6
Lambda	0	**Actual Over Pred Change**	0.5924978695
Radius	0.0002413858		

Convergence criterion (GCONV=1E-8) satisfied.

169

The SAS System

The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation
Fit Summary

Modeling Info	Number of Observations	100
	Number of Variables	11
	Number of Moments	66
	Number of Parameters	31
	Number of Active Constraints	0
	Baseline Model Function Value	8.7109
	Baseline Model Chi-Square	644.6058
	Baseline Model Chi-Square DF	55
	Pr > Baseline Model Chi-Square	<.0001
Absolute Index	Fit Function	1.0617
	Chi-Square	78.5670
	Chi-Square DF	35
	Pr > Chi-Square	<.0001
	Elliptic Corrected Chi-Square	0.8351
	Pr > Elliptic Corr. Chi-Square	1.0000
	Z-Test of Wilson & Hilferty	3.9621
	Hoelter Critical N	47
	Root Mean Square Residual (RMR)	0.3393
	Standardized RMR (SRMR)	0.0624
	Goodness of Fit Index (GFI)	0.8645
Parsimony Index	Adjusted GFI (AGFI)	0.7445
	Parsimonious GFI	0.5502
	RMSEA Estimate	0.1297
	RMSEA Lower 90% Confidence Limit	0.0914
	RMSEA Upper 90% Confidence Limit	0.1681
	Probability of Close Fit	0.0009
	ECVI Estimate	2.0617
	ECVI Lower 90% Confidence Limit	1.7543
	ECVI Upper 90% Confidence Limit	2.4951
	Akaike Information Criterion	140.5670
	Bozdogan CAIC	243.4092

Fit Summary

Schwarz Bayesian Criterion	212.4092
McDonald Centrality	0.7479
Incremental Index Bentler Comparative Fit Index	0.9261
Bentler-Bonett NFI	0.8781
Bentler-Bonett Non-normed Index	0.8839
Bollen Normed Index Rho1	0.8085
Bollen Non-normed Index Delta2	0.9285
James et al. Parsimonious NFI	0.5588

The SAS System

The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation

PATH List

Path			Parameter	Estimate	Standard Error	t Value	Pr > |t|
ind60	===>	x1		1.00000			
ind60	===>	x2	lambda2	2.17437	0.22259	9.7685	<.0001
ind60	===>	x3	lambda3	1.77491	0.20174	8.7980	<.0001
poldem60	===>	y1		1.00000			
poldem60	===>	y2	lambda5	1.38908	0.30432	4.5645	<.0001
poldem60	===>	y3	lambda6	1.06516	0.23876	4.4611	<.0001
poldem60	===>	y4	lambda7	1.44005	0.28460	5.0599	<.0001
poldem65	===>	y5		1.00000			
poldem65	===>	y6	lambda9	1.20005	0.17861	6.7187	<.0001
poldem65	===>	y7	lambda10	1.29162	0.17395	7.4253	<.0001
poldem65	===>	y8	lambda11	1.31569	0.16995	7.7414	<.0001
ind60	===>	poldem60	gamma11	1.41058	0.45097	3.1279	0.0018
ind60	===>	poldem65	gamma21	0.53328	0.25888	2.0600	0.0394
poldem60	===>	poldem65	beta21	0.83824	0.17383	4.8221	<.0001

Variance Parameters

Variance Type	Variable	Parameter	Estimate	Standard Error	t Value	Pr > |t|
Exogenous	ind60	phi11	0.46503	0.11475	4.0527	<.0001
Error	poldem60	psi11	3.48278	1.38415	2.5162	0.0119
	poldem65	psi22	0.45677	0.29808	1.5323	0.1254
	x1	thetadelta11	0.26950	0.05010	5.3788	<.0001
	x2	thetadelta22	0.07216	0.10799	0.6682	0.5040
	x3	thetadelta33	0.50766	0.11042	4.5974	<.0001
	y1	thetaeps11	9.15236	1.60718	5.6947	<.0001
	y2	thetaeps22	6.90213	1.53367	4.5004	<.0001
	y3	thetaeps33	5.94858	1.10361	5.3901	<.0001
	y4	thetaeps44	2.12191	0.91001	2.3317	0.0197
	y5	thetaeps55	2.52335	0.51396	4.9096	<.0001
	y6	thetaeps66	5.00644	0.95592	5.2373	<.0001
	y7	thetaeps77	4.01480	0.82650	4.8576	<.0001
	y8	thetaeps88	3.02432	0.73435	4.1184	<.0001

Covariances Among Errors

Error of	Error of	Parameter	Estimate	Standard Error	t Value	Pr > |t|
y1	y5	thetaeps15	0.62545	0.65457	0.9555	0.3393
y2	y6	thetaeps26	2.25128	0.77074	2.9209	0.0035
y3	y7	thetaeps37	1.68137	0.72929	2.3055	0.0211
y4	y8	thetaeps48	0.14767	0.47015	0.3141	0.7535
y2	y4	thetaeps24	0.46071	0.85643	0.5379	0.5906
y6	y8	thetaeps68	1.28783	0.58977	2.1836	0.0290

Squared Multiple Correlations

Variable	Error Variance	Total Variance	R-Square
x1	0.26950	0.73453	0.6331
x2	0.07216	2.27077	0.9682
x3	0.50766	1.97266	0.7427
y1	9.15236	13.56042	0.3251
y2	6.90213	15.40767	0.5520
y3	5.94858	10.94984	0.4567

Squared Multiple Correlations

Variable	Error Variance	Total Variance	R-Square
y4	2.12191	11.26309	0.8116
y5	2.52335	6.79614	0.6287
y6	5.00644	11.15973	0.5514
y7	4.01480	11.14302	0.6397
y8	3.02432	10.42068	0.7098
poldem60	3.48278	4.40807	0.2099
poldem65	0.45677	4.27279	0.8931

The SAS System

The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation

Standardized Results for PATH List

Path			Parameter	Estimate	Standard Error	t Value	Pr > |t|
ind60	===>	x1		0.79567	0.04711	16.8879	<.0001
ind60	===>	x2	lambda2	0.98398	0.02430	40.4919	<.0001
ind60	===>	x3	lambda3	0.86177	0.03670	23.4784	<.0001
poldem60	===>	y1		0.57015	0.08650	6.5915	<.0001
poldem60	===>	y2	lambda5	0.74299	0.06903	10.7628	<.0001
poldem60	===>	y3	lambda6	0.67583	0.07186	9.4043	<.0001
poldem60	===>	y4	lambda7	0.90089	0.04684	19.2327	<.0001
poldem65	===>	y5		0.79291	0.05152	15.3893	<.0001
poldem65	===>	y6	lambda9	0.74255	0.06032	12.3094	<.0001
poldem65	===>	y7	lambda10	0.79981	0.05029	15.9052	<.0001
poldem65	===>	y8	lambda11	0.84248	0.04569	18.4394	<.0001
ind60	===>	poldem60	gamma11	0.45816	0.10354	4.4248	<.0001
ind60	===>	poldem65	gamma21	0.17593	0.08389	2.0971	0.0360
poldem60	===>	poldem65	beta21	0.85141	0.06707	12.6945	<.0001

Standardized Results for Variance Parameters

Variance Type	Variable	Parameter	Estimate	Standard Error	t Value	Pr > |t|
Exogenous	ind60	phi11	1.00000			
Error	poldem60	psi11	0.79009	0.09488	8.3274	<.0001
	poldem65	psi22	0.10690	0.06793	1.5737	0.1155
	x1	thetadelta11	0.36690	0.07498	4.8936	<.0001
	x2	thetadelta22	0.03178	0.04782	0.6644	0.5064
	x3	thetadelta33	0.25735	0.06326	4.0680	<.0001
	y1	thetaeps11	0.67493	0.09863	6.8429	<.0001
	y2	thetaeps22	0.44797	0.10258	4.3669	<.0001
	y3	thetaeps33	0.54326	0.09713	5.5928	<.0001
	y4	thetaeps44	0.18840	0.08440	2.2322	0.0256
	y5	thetaeps55	0.37129	0.08171	4.5442	<.0001
	y6	thetaeps66	0.44862	0.08959	5.0076	<.0001
	y7	thetaeps77	0.36030	0.08044	4.4791	<.0001
	y8	thetaeps88	0.29022	0.07698	3.7699	0.0002

Standardized Results for Covariances Among Errors

Error of	Error of	Parameter	Estimate	Standard Error	t Value	Pr > |t|
y1	y5	thetaeps15	0.06515	0.06780	0.9610	0.3366
y2	y6	thetaeps26	0.17169	0.05796	2.9622	0.0031
y3	y7	thetaeps37	0.15221	0.06500	2.3418	0.0192
y4	y8	thetaeps48	0.01363	0.04341	0.3140	0.7535
y2	y4	thetaeps24	0.03497	0.06508	0.5374	0.5910
y6	y8	thetaeps68	0.11942	0.05511	2.1671	0.0302