**Title**
PLSe: Efficient Estimators and Tests for Partial Least Square



**Author**
Huang, Wenjing

**Publication Date**
2013

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

# PLSe: Efficient Estimators and Tests for Partial Least Squares

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Psychology

by

Wenjing Huang

2013

ABSTRACT OF THE DISSERTATION

PLSe: Efficient Estimators and Tests for Partial Least Squares

By

Wenjing Huang

Doctor of Philosophy in Psychology

University of California, Los Angeles, 2013

Professor Peter Bentler, Chair

This dissertation extends Dijkstra's (2011) consistent partial least squares estimator for structural equation models by deriving new estimators that are efficient. The new methods allow formal testing of models via chi-square statistics and evaluation of parameter estimates by deriving standard error estimates, which are previously not directly available. Two approaches are developed: (1) PLSe1: a one-step improvement methodology based on PLSc-estimated factor loadings and TSLS-estimated structural parameters; (2) PLSe2: an optimal generalized least squares methodology using PLSc-

implied covariance matrix. The performances of the proposed methods are evaluated by Monte Carlo simulations. We generated data under a non-recursive structural equation model. We investigated the performances of the proposed estimators relative to the classical Maximum Likelihood estimator under a variety of sample sizes for both normal data (with PLSe1 and PLSe2) and non-normal data (with PLSe1 only). The results indicate that the proposed estimators provide estimates that are almost as good as the theoretically optimal ML estimator under normality. We also demonstrate that the standard error estimates closely correspond to the empirical Monte Carlo variation. Under non-normality, PLSe1 performs favorably with non-normal (i.e. robust) adjustments to model fit statistics and standard errors. The standard error estimates are consistent with corresponding sampling variance and the robust model fit chi-squares statistics are well calibrated. In particular, Satorra-Bentler's (1994) scaled chi-square statistic stands out clearly.

The dissertation of Wenjing Huang is approved.

Hongjing Lu

Steve Reise

Ying Nian Wu

Peter Bentler, Committee Chair

University of California, Los Angeles

2013

*To my parents and husband Li*

TABLE OF CONTENTS

LIST OF TABLES

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Dr. Peter Bentler for his original ideas and constant guidance, without which this dissertation would not be possible. I am grateful for his kind support and encouragement throughout my graduate study at UCLA. I truly appreciate the precious opportunity to work with him and to learn from him. I also would like to thank members of my committee, Drs. Steve Reise, Hongjing Lu and Ying Nian Wu, for the stimulating discussions and suggestions during my proposal meeting that led to a much improved research project.

I would like to use this opportunity to express my sincere gratitude towards Dr. Theo Dijkstra for kindly sharing his most recent developments and his MATLAB programs for implementing the consistent PLS method. This research project would not be possible without his earlier contributions.

I thank friends, faculty and office staff from the Psychology Department for the fine memorable years at UCLA. I deeply appreciate the support from computer scientists David Sookne and Eric Wu who helped create the EQS program and are now trying to implement the proposed methods in the next generation of EQS.

Finally, I would like to thank my parents and husband Li for their support during the long years of my graduate study.

VITA

| | |
|---|---|
| 1998-2002 | Bachelor of Arts in English, Business Administration Minor<br>Soochow University, Suzhou, China. |
| 2005-2008 | Research Assistant<br>Department of Psychology<br>University of California, Los Angeles |
| 2005-2008 | Master of Arts in Psychology<br>University of North Caroline, Chapel Hill<br>Chapel Hill, North Carolina |
| 2009 | Teaching Assistant<br>Department of Psychology<br>University of California, Los Angeles |
| 2010 & 2012 | Teaching Associate<br>Department of Psychology<br>University of California, Los Angeles |
| 2010-present | Graduate Student Researcher<br>Center for Advancing Longitudinal Drug Abuse Research<br>University of California, Los Angeles |

# CHAPTER 1

# INTRODUCTION

## 1.1. Confirmatory Factor Analysis and Structural Equation Modeling

Structural Equation Model (SEM) is widely used in many areas of social and behavioral sciences as a tool for multivariate data analysis (MacCallum & Austin, 2000), and is especially appropriate for theory testing (e.g., Bagozzi, 1980). Unlike single-equation regression models, SEM usually refers to a simultaneous equation system that involves multiple independent and dependent variables, some of which may be latent variables that are not directly observed but represented by observed/manifest variables. Together with appropriate estimation methods, SEM can account for measurement error in manifest variables so that the latent variable regression models are free from the influence of measurement error. In addition, SEM provides the advantage of an overall test of fit so the models become falsifiable.

There are different approaches to SEM, for example, the traditional Jöreskog-Keesling-Wiley model (Jöreskog, 1973, 1977; Jöreskog & Sörbom, 1994; Keesling, 1972; Wiley, 1973) developed in the LISREL program (Jöreskog & Sörbom, 1994), the Bentler-Weeks model (Bentler & Weeks, 1980) implemented in the EQS program (Bentler, 2006) and a number of other related models (e.g. Brown & Arminger, 1995; McArdle & McDonald, 1984) implemented in several tools, such as RAMONA, AMOS, SePath,

COSAN, and Mplus. These different approaches can be understood as covariance structure models (e.g., models for variances and covariances), expressed as a function of a relatively small set of parameters. In some models, not discussed further in this dissertation, means also may be structured.

The Jöreskog-Keesling-Wiley model (here we call it the LISREL model) makes a clear distinction between two main components of SEMs: the *structural model* that describes the potential causal relations among the endogenous and exogenous latent variables, and the *measurement model* that deals with the relations between latent variables (*common factors*) and their manifest indicators. The regression coefficients of manifest variables on latent variables are often referred to as *factor loadings*. The error terms associated with the observed indicators are called *unique factors*. While it is useful for learning and explanatory purposes, this approach requires the use of 8 equations and 12 parameter matrices to specify a model. As an alternative, the Bentler-Weeks model can handle all the linear structure of the LISREL model. As path analysis, it provides an equation for every dependent variable and covariances for independent variables, which is easy to understand and implement.

In spite of the different modeling frameworks, standard models such as factor analysis, path analysis and multivariate regression can be viewed as special types of SEMs. For example, Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) models can be viewed as SEMs that contain only the measurement part

(see Jöreskog, 1969), while path analysis (or simultaneous equations) can also be viewed as SEM that only has the structural part and involves only observed variables.

The classical SEM approaches usually employ estimation methods that require minimizing some suitably chosen fit function. The fit functions are derived under different distributional assumptions. For example, the Maximum Likelihood (ML) estimator requires the assumption that the joint distribution of the observed indicators is multivariate normal. If such an assumption is made, the ML estimates maximize the likelihood of observed data given the model specified. ML enjoys the properties of consistency, asymptotic normality, and asymptotic efficiency.

Despite the restrictive normality assumption, the ML parameter estimates are robust against violations of this assumption; although one needs to adjust the standard errors as well as the model fit chi-square when the data are non-normal. There are other estimation methods that require fewer assumptions, such as the "asymptotically distribution free" (ADF, see Browne, 1984) method. ADF is also known as AGLS or "arbitrary distribution Generalized Least Square" in the literature. ADF only requires the assumption of the existence of 8th-order moments for the observed variable distributions. The estimation procedure involves finding a sample-based weight matrix that depends on the sample third- and fourth-order moments. Asymptotically ADF is efficient among the class of generalized least squares estimators, but due to the size of the weight matrix (polynomial in the number of manifest variables) and the inherent

instability of sample higher-order moments, ADF often requires excessively large sample size, sometimes thousands or more, to find stable parameter estimates. Thus this method almost never works well unless the sample size is very large or the model is very small (Browne, 1984; Savalei & Bentler, 2006).

### 1.2. Partial Least Squares

While most of the development of SEM started in the 1960's and 1970's, another line of work called Partial Least Squares (PLS) was developed by Herman Wold during the same period (Wold, 1966, 1975, 1982) for "soft models and soft data" in social sciences. PLS, as another approach to traditional SEM, is also known as PLS Path Modeling and its purpose is for the analysis of high-dimensional data when specificity of the structure of the model is low, with strong inclination toward the models' use for prediction purposes.

The general applicability of LISREL models were questioned by Wold (1966), because in empirical studies, the distributions are often unknown or it would be difficult to assume them to be normal. In empirical studies, a sufficiently large sample size would also be difficult to obtain so that one could not use any of the distributional free approaches, such as ADF. In order to find a method that is free of distribution assumptions, that does not require a large sample size, and is relatively easy to apply in

practice, Wold and associates developed the Partial Least Squares (PLS) method with inspirations from principal component analysis and canonical correlation analysis.

PLS is considered a soft modeling approach because no strong assumptions are made with respect to the distributions of the observed variables or the sample sizes. Compared to the covariance based approaches, such as the LISREL model, PLS can deal with a much larger number of indicators per latent variable. In addition, PLS makes no assumptions about the population or scale of measurement (Fornell & Brookstein, 1982). PLS became more widely used in areas of marketing, political sciences and chemometrics where nominal and ordinal variables are common.

As with all statistical methods, PLS does require certain assumptions to be fulfilled. The most important assumption as noted by Chin and Newsted (1999) is predictor specification. That is, PLS requires that the systematic part of the regression models must be equal to the conditional expectation of the dependent variables (see Wold, 1975). This is no different from the specification assumptions made in any regression analysis.

Perhaps the most important distinction between the traditional SEM methods (e.g. estimation with ML or ADF) and PLS is that those traditional SEM estimation procedures aim to reproduce the sample covariance matrices, and all the parameters are estimated simultaneously. PLS, on the other hand, aims to maximize the amount of variance in the dependent variables explained by the independent variables. As

illustrated by Wold's (1975) NIPALS (nonlinear iterative partial least square) algorithm, the iterative procedure separately estimates the parameters in the measurement model in the first step and then estimates the regression coefficients in the structural model based on those parameter estimates obtained in the first step.

Apart from its obvious advantages in many applied fields, PLS has a few disadvantages that cannot be ignored. First, it lacks a classical parametric inferential framework (Vinzi, Trinchera & Amato, 2010). One has to resort to empirical confidence intervals and hypothesis testing procedures based on resampling methods such as jackknife and bootstrap (Chin, 1998). It also suffers from undesirable statistical properties of the estimates, e.g. coefficients are known to be biased (Cassel, Hackl & Westlund, 1999, 2000).

PLS researchers, e.g., Wold, Apel, Hui, and Lohmoller, have almost exclusively focused on the development of estimation algorithms for PLS. They have not addressed questions pertaining to the stochastic properties of the estimators. Exceptions do exist. For instance, Dijkstra (1983) summarized in broad terms what appears to be known about the asymptotic properties of ML and PLS estimators.

The problem known under the term *consistency at large* is the major drawback of PLS. A consistent estimator can be described as "one that converges in probability to the value of the parameter being estimated as the sample size increases" (McDonald, 1996, p. 248). However, due to the fact that PLS does not account for measurement error

during estimation procedures, it does not yield consistent estimates of what are called ''latent variables'' in formal structural equation modeling (Dijkstra, 1983, 1985). As a result, "the path coefficients estimated through PLS converge on the parameters of the latent-variable model [only] as both the sample size and the number of indicators of each latent variable become infinite" (McDonald, 1996). In most empirical studies, such *consistency at large* situations usually cannot be achieved. With finite sample size and relatively few indicators per latent variable, PLS tends to underestimate the correlations between the latent variables and overestimate the loadings (Dijkstra, 1983).

### 1.3. PLSc: A Consistent Correction Procedure for Partial Least Squares

During the early years of PLS' developments, Theo Dijkstra joined Herman Wold's research team and was suggested to write a PhD thesis on LISREL versus PLS in the context of Wold's basic design (Wold, 1982). Recently, Dijkstra (2011) proposed a new method which involves corrections to the PLS estimates to account for measurement error. He called it PLS-consistent (PLSc) to distinguish it from the traditional PLS estimators. This procedure, as demonstrated through his simulation studies, is effective in reducing the bias in parameter estimates.

While there are many later developments of PLS with different algorithms either for exploratory or confirmatory purposes, (see e.g., Hastie, Tibshirani & Friedman, 2001; Stone & Brooks, 1990; Frank & Friedman, 1993; Tenenhaus et al., 2005), this dissertation

study will follow the basic design due to Wold (1982) as implemented by Dijkstra (2011), where his new PLSc estimators are used for Confirmatory Factor Analysis (CFA) type of models.

In the class of factor models that Dijkstra (2011) analyzed, each indicator loads on one latent variable only. In addition, each latent variable is measured by blocks of at least two indicators. The indicators may correlate with one another across the blocks, but only through correlations among the latent variables. When the structural model coefficients (i.e. causal relationships between the latent variables) are of interest, they can be estimated in a second step using the second order moment matrix of the latent variables already provided by PLSc.

It is worth noting that SEM generally estimates all the parameters simultaneously, including factor loadings, factor covariance matrix, unique factor covariance matrix in the measurement model, the regression coefficients, and equation disturbance covariance matrix in the structural model. The use of PLSc in this study, however, focuses on the measurement model only. PLSc is used to obtain the correlations among the latent variables in the first step. In the second step, regression coefficients among the latent variables are estimated by Two-Stage Least Squares based on the PLSc estimates of the correlations among the latent variables.

### 1.4. Two-Stage Least Squares for SEM

One way to think of PLSc is that it is a data condensation procedure that provides consistent estimates of latent variable correlations. Once these correlations are known, one could essentially ignore the manifest indicators and treat the estimation of path coefficients in the structural regression models independently - using solely the PLSc estimates of factor correlations. According to the Continuous Mapping Theorem, if $T_n$ is consistent for $\theta$ and $g(\cdot)$ is a real-valued function continuous at point $\theta$, then $g(T_n)$ will be consistent for $g(\theta)$ (Mann & Wald, 1943). Thus parameter estimates of the regression coefficients obtained from the PLSc consistent estimates of the factor correlations will be consistent. One of the simplest and most widely used estimation methods available for consistently estimating simultaneous equations models is Two-Stage Least Squares (TSLS).

We could use Ordinary Least Squares (OLS) estimation for single-equation multivariate and/or multiple regression models. However, when we have a system of simultaneous equations, it is difficult to assume that the endogenous variables appearing on the right-hand side of the equations are uncorrelated with the equation disturbances. Such correlations lead to biased and inconsistent estimates. If instrumental variables are available, then TSLS estimation can correct the bias, which is a well-known result in econometrics (see, e.g., Mardia, Kent & Bibby, 1979). When applied to simultaneous equations estimation, TSLS essentially involves a sequence of single-equation instrumental variable estimation, using model-implied instruments.

### 1.5. Evaluating the Quality of the Estimates and the Fit of the Model

Combining PLSc developed by Dijkstra (2011) and the classic TSLS for simultaneous equations leads to consistent parameter estimates for both the measurement and structural models. However, we still do not know how good these parameter estimates are or how well-fitting the overall model is.

Dijkstra (2012) recently investigated these two issues with Monte Carlo experiments. According to his results, simulation-based methods such as the bootstrap are feasible considering the fact that PLSc usually converges quickly. An alternative approach would be the delta-method. With the Jacobian matrix calculated numerically, one could obtain standard error estimates based on normal or distribution-free asymptotic theory.

Regarding model fit, as described in Savalei and Bentler (2006), the goal of SEM is to see how well our proposed model, which is a set of specified causal and non-causal relationships among latent and observed variables, accounts for the observed relationships. The observed relationships are usually in the form of covariances, summarized in the sample covariance matrix of the observed variables, which we will call $\mathbf{S}$. The SEM specifies a certain hypothesized covariance structure model in the population, i.e., $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$, where $\boldsymbol{\Sigma}$ is the population covariance matrix, and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is a matrix-valued function of SEM parameters in $\boldsymbol{\theta}$. Under broad conditions $\mathbf{S}$ is a

consistent estimator of $\boldsymbol{\Sigma}$. Fitting the model $\boldsymbol{\Sigma}(\boldsymbol{\theta})$to $\mathbf{S}$ leads to a model-implied covariance matrix $\widehat{\boldsymbol{\Sigma}}$. The discrepancy between $\mathbf{S}$ and $\widehat{\boldsymbol{\Sigma}}$ provides a basis for model fit evaluation (Browne, 1974; Jöreskog, 1978; Bentler, 1983). With regards to model fit, the difference between PLS and SEM resembles the distinction between principal component analysis (PCA) and factor analysis. As a data condensation technique, PCA or PLS holds unique advantages over factor analysis or SEM. However, PCA or PLS does not specify a formally testable covariance structure model. Adding the capability to evaluate the quality of the model to PLS would be a major desirable feature.

In this dissertation, we propose to take PLSc-based consistent estimator $\widehat{\boldsymbol{\theta}}_c$ of the unknown parameters $\boldsymbol{\theta}$ of some covariance structure $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ model and improve the estimator to obtain efficient estimator $\widehat{\boldsymbol{\theta}}$ with minimum asymptotic variances. At the same time, we obtain standard errors for the efficient estimator and provide a statistical test of the null hypothesis $H_0$: $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$. Specifically, we extend Dijkstra's (2011) consistent partial least squares estimator for structural equation models to derive estimators that are efficient and allow evaluation of models via chi-square statistics and evaluation of parameter estimates via standard errors. Two approaches are developed: (1) a one-step improvement methodology and (2) an optimal generalized least squares methodology. The proposed methods are evaluated by Monte Carlo simulations with an SEM that involves confirmatory factor analysis measurement model and non-recursive latent structural model.

# CHAPTER 2

THEORY

## 2.1. An Overview of SEM

For the purpose of simplicity in demonstration, let us use the terminology of LISREL models to introduce SEM and let us assume the structural part of SEM deals only with causal relations between and among the latent variables, while the measurement part deals only with how the latent variables are defined by the observed variables.

The measurement part of SEM describes how the latent variables, including the endogenous latent variables $\boldsymbol{\eta}$ and exogenous latent variables $\boldsymbol{\xi}$, are defined by their indicators $\boldsymbol{y}$ and $\boldsymbol{x}$, respectively. The vector of observed variables $\boldsymbol{y}$ is presumed to following a factor analytic measurement model:

$$\boldsymbol{y} = \boldsymbol{\Lambda}_y \boldsymbol{\eta} + \boldsymbol{\varepsilon}, \tag{1}$$

where $\boldsymbol{\Lambda}_y$ is a factor loading matrix that shows which indicators of $\boldsymbol{y}$ load on which endogenous latent variables $\boldsymbol{\eta}$. Similarly, for the vector of observed variables $\boldsymbol{x}$, the measurement model is:

$$\boldsymbol{x} = \boldsymbol{\Lambda}_x \boldsymbol{\xi} + \boldsymbol{\delta}, \tag{2}$$

where $\boldsymbol{\Lambda}_x$ is a factor loading matrix that shows which indicators of $\boldsymbol{x}$ load on which exogenous latent variables $\boldsymbol{\xi}$. In LISREL models' terminology, $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ are vectors of

latent common factors, whereas $\boldsymbol{\varepsilon}$ and $\boldsymbol{\delta}$ are vectors of unique factors. The covariance matrices of the unique factors are denoted as $V(\boldsymbol{\varepsilon}) = \boldsymbol{\Omega}_{\varepsilon}$ and $V(\boldsymbol{\delta}) = \boldsymbol{\Omega}_{\delta}$. Let us assume $V(\boldsymbol{\varepsilon}, \boldsymbol{\delta}) = \mathbf{0}$ and let us use $\boldsymbol{\Omega}$ to represent the covariance matrix of the unique factors:

$$V\begin{pmatrix}\boldsymbol{\varepsilon}\\ \boldsymbol{\delta}\end{pmatrix} = \boldsymbol{\Omega} = \begin{pmatrix}\boldsymbol{\Omega}_{\varepsilon} & \\ & \boldsymbol{\Omega}_{\delta}\end{pmatrix}.$$

The structural component of SEM describes the relationship between the vectors of latent variables $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ in a set of simultaneous equations:

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta} = (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Gamma}\boldsymbol{\xi} + (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\zeta}, \tag{3}$$

where $(\mathbf{I} - \mathbf{B})$ is assumed to be nonsingular. The last equation is said to be in a *reduced form*. The covariance matrices of the common factors are denoted as $V(\boldsymbol{\xi}) = \boldsymbol{\Phi}$ and $V(\boldsymbol{\zeta}) = \boldsymbol{\Psi}$. Let us assume $V(\boldsymbol{\xi}, \boldsymbol{\zeta}) = \mathbf{0}$ and let us also assume the disturbance vector $\boldsymbol{\zeta}$ has the property that $E(\boldsymbol{\zeta}|\boldsymbol{\xi}) = 0$.

Let $(\mathbf{I} - \mathbf{B})^{-1} = \mathbf{A}$. The covariance matrix of the latent variables is:

$$V\begin{pmatrix}\boldsymbol{\eta}\\ \boldsymbol{\xi}\end{pmatrix} = \begin{pmatrix}\mathbf{A}(\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Gamma}' + \boldsymbol{\Psi})\mathbf{A}' & \mathbf{A}\boldsymbol{\Gamma}\boldsymbol{\Phi}\\ \boldsymbol{\Phi}\boldsymbol{\Gamma}'\mathbf{A}' & \boldsymbol{\Phi}\end{pmatrix} = \boldsymbol{\Upsilon}. \tag{4}$$

Let $\boldsymbol{\Lambda} = \begin{pmatrix}\boldsymbol{\Lambda}_y\\ \boldsymbol{\Lambda}_x\end{pmatrix}$, the covariance matrix of the observed variables is therefore:

$$V\begin{pmatrix}\boldsymbol{y}\\ \boldsymbol{x}\end{pmatrix} = \boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Upsilon}\boldsymbol{\Lambda}' + \boldsymbol{\Omega}. \tag{5}$$

Without considering the mean structure, the LISREL model specification contains eight parameter matrices, five of which ($\boldsymbol{\Lambda}_x$, $\boldsymbol{\Lambda}_y$, $\boldsymbol{\Omega}_{\varepsilon}$, $\boldsymbol{\Omega}_{\delta}$ and$\boldsymbol{\Phi}$) are related to the measurement model and three of which ($\boldsymbol{\Psi}$, $\mathbf{B}$ and $\boldsymbol{\Gamma}$) are related to the structural model. Let every one of these parameter matrices depend on a parameter vector $\boldsymbol{\theta}$, then the

covariance structure of observed variable can be written as $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^q$ and $q$ is number of free parameters.

### 2.1.1 Estimation

Based on the sample variance and covariances of the observed variables (collectively in the sample covariance matrix $\mathbf{S}$), the parameter vector $\boldsymbol{\theta}$ containing the unknown parameters can then be estimated. In the estimation stage, a fitting function is chosen according to the distribution assumption of the observed variables. The fitting function is minimized to obtain parameter estimates through an iterative process. To start the iteration, the initial values for all the parameters are plugged into the function so that the function can be evaluated. The parameter estimates are then modified to make the function smaller. The function is reevaluated in the same manner from one iteration to the next iteration until the value of the function changes by a very small designated number (this is called convergence).

Different assumptions about the distributions of the observed variables lead to different estimators. For example, the ordinary least squares (OLS) estimator is defined by minimizing the discrepancy function

$$F_{OLS}(\boldsymbol{\Sigma},\ \mathbf{S}) = \frac{1}{2} tr[(\boldsymbol{\Sigma} - \mathbf{S})^2],$$

where $tr(\cdot)$ represents the trace operator (sum of the diagonal elements). The OLS discrepancy function measures the sum of squared differences between the model implied covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ and the observed sample covariance matrix $\mathbf{S}$. The OLS estimator is consistent and it does not require assuming a specific distributional form of the observed variables. However, it is not efficient because the estimator has larger sampling variability than a more efficient alternative, such as the maximum likelihood (ML) estimator. The ML estimator is obtained by minimize the ML fit function:

$$F_{ML}(\boldsymbol{\Sigma}, \mathbf{S}) = log|\boldsymbol{\Sigma}| + tr(\mathbf{S}\boldsymbol{\Sigma}^{-1}) - log|\mathbf{S}| - m \quad (6)$$

where $m$ is the number of observed variables. The ML estimator does require the observed variables to be multivariate normally distributed.

An estimator with minimum variance will be considered optimal if it is consistent and asymptotically unbiased. It is known that estimators such as ML, Generalized Least Squares (GLS) or normal theory Weighted Least Squares (WLS) with correct weights are the optimal estimators among all linear asymptotically unbiased estimators, with the latter two, GLS and WLS, equivalent to ML asymptotically (Browne, 1974).

Take WLS as an example. Let $\boldsymbol{\sigma}(\boldsymbol{\theta}) = vech(\boldsymbol{\Sigma}(\boldsymbol{\theta}))$, where $vech(\cdot)$ is the half-vectorization operator that stacks the unique elements in $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ into a vector $\boldsymbol{\sigma}(\boldsymbol{\theta})$**.** Similarly, let $\boldsymbol{s} = vech(\mathbf{S})$**.** The WLS fit function is defined as:

$$F_{WLS}(\boldsymbol{\theta}) = [\boldsymbol{s} - \boldsymbol{\sigma}(\boldsymbol{\theta})]'\mathbf{W}^{-1}[\boldsymbol{s} - \boldsymbol{\sigma}(\boldsymbol{\theta})], \tag{7}$$

where **W** is a positive definite matrix. We can rewrite the above WLS fit function to show that it indeed involves a form of weighted least squares:

$$F_{WLS}(\boldsymbol{\theta}) = \sum_{i=1}^{m}\sum_{j=1}^{i}\sum_{k=1}^{m}\sum_{l=1}^{k} w^{ij,kl}[s_{ij} - \sigma_{ij}(\theta)][s_{kl} - \sigma_{kl}(\theta)], \tag{8}$$

where $w^{ij,kl}$ denotes a typical element of $\mathbf{W}^{-1}$. The WLS estimator is obtained by minimizing $F_{WLS}(\boldsymbol{\theta})$.

The optimal weights turn out to be directly related to the variances and covariances of the sample moments (see, e.g., Browne, 1984; Hansen, 1982). Let a typical element of **W** be denoted by $w_{ij,kl}$ and the subscripts *ij* and *kl* highlight that $w_{ij,kl}$ represents the covariance between $s_{ij}$ (the covariance between observed variables *i* and *j*) and $s_{kl}$ (the covariance between observed variables *k* and *l*). Thus the weights come from covariances of covariances.

Let $m^* = m(m+1)/2$ denote the number of unique elements in the covariance matrix of the observed variables. One of the key theoretical results about WLS was developed by Browne (1984). He noted that $\boldsymbol{s}$ is asymptotically normal with mean vector $\boldsymbol{\sigma}$ and asymptotic covariance matrix **V**:

$$\sqrt{N}(\boldsymbol{s} - \boldsymbol{\sigma}) \xrightarrow{D} \mathcal{N}_{m^*}(\mathbf{0}, \mathbf{V}). \tag{9}$$

A typical element of $\mathbf{V}$ is given by $v_{ij,kl} = \sigma_{ijkl} - \sigma_{ij}\sigma_{kl}$ , with $\sigma_{ijkl}$ being the fourth order moment in the population, i.e., $\sigma_{ijkl} = E(y_i - \mu_i)\big(y_j - \mu_j\big)(y_k - \mu_k)(y_l - \mu_l)$, and $\sigma_{ij} = E(y_i - \mu_i)(y_j - \mu_j)$, with $\mu_i = E(y_i)$ (see Equation 2.1 in Browne, 1984).

Browne (1984) noted that with a sample of size *N*, $v_{ij,kl}$ can be consistently estimated from the data as

$$\hat{v}_{ij,kl} = s_{ijkl} - s_{ij}s_{kl}, \tag{10}$$

where $s_{ijkl}$ is the sample fourth order moment for observed variables *i*, *j*, *k*, *l*, i.e.,

$$s_{ijkl} = N^{-1}\sum_{r=1}^{N}(y_{ir} - \bar{y}_i)\big(y_{jr} - \bar{y}_j\big)(y_{kr} - \bar{y}_k)(y_{lr} - \bar{y}_l), \tag{11}$$

and $s_{ij}$ is the sample variance (biased estimate), i.e.,

$$s_{ij} = N^{-1}\sum_{r=1}^{N}(y_{ir} - \bar{y}_i)\big(y_{jr} - \bar{y}_j\big), \tag{12}$$

with $\bar{y}_i = N^{-1}\sum_{r=1}^{N} y_{ir}$ denoting the sample mean for *i*th observed variable. In other words, let $\mathbf{V}_S$ be a positive definite matrix that is a sample-based consistent estimate of the asymptotic covariance matrix $\mathbf{V}$, then a typical element of $\mathbf{V}_S$ is given by $\hat{v}_{ij,kl}$. This result holds under the mild condition that the eighth order moments of the observed variables are finite, and it does not require that the observed variables be multivariate normally distributed.

Taken together, as long as an asymptotically correct weight matrix can be provided, e.g., when $\mathbf{V}_S$ is used as the weight matrix, one can proceed with WLS

estimation to obtain an asymptotically efficient estimator. This is Browne's (1984) Asymptotically Distribution Free (ADF) estimator, which minimizes the WLS discrepancy function in Equation (7) with sample-based weights given by $\mathbf{V}_S$.

### 2.1.2. Model Evaluation

After minimization of the ML discrepancy function, we have obtained parameter estimates $\widehat{\boldsymbol{\theta}}$ and a minimum discrepancy function value $\hat{F}_{ML}$. When distributional assumptions are met, Jöreskog (1969) showed that $(N-1)$ times the minimized value of $F_{ML}$ is asymptotically distributed as a central chi-square variable with $d = m^* - q$ degrees of freedom under the null hypothesis that model fits exactly in the population. This is referred to as the minimum discrepancy function chi-square statistic:

$$c_1 = (N-1)\hat{F}_{ML}. \tag{13}$$

Under non-normality, $c_1$ is distributed as a mixture of one degree of freedom chi-square variates (see Yuan & Bentler, 2007).

In cases when the model is not exactly correctly specified, Steiger, Shapiro and Browne (1985) showed that the model fit chi-square statistic is distributed as a noncentral chi-square variable. This is particularly useful for model-fit indices that directly depend on the minimized discrepancy function values, such as the Root Mean Square Error of Approximation (RMSEA; Steiger & Lind, 1980). When observed

variables are not normally distributed, *N* times the minimized ADF/WLS discrepancy function value is distributed in large samples as a chi-square random variable.

With non-normal data, the ADF estimator appears to be a very useful choice over traditional ML, since it does not require strong distributional assumptions. However, the ADF estimator does not work well in practical settings, because it requires extremely large sample size (sometimes thousands, unless the model is very small) in order to estimate the weight matrix. Sample higher-order moments can be highly unstable themselves, leading to biased estimates and inflated chi-squares (e.g., see results reported in Hu, Bentler, & Kano, 1992).

Compared to ADF estimator which is often impractical in empirical studies, the traditional normal theory based ML remains a consistent estimator even under non-normality. The main problem of using ML for non-normal data comes in the form of incorrect standard errors and test statistics. They may, however, be adjusted. Satorra and Bentler (1994) proposed such an adjustment. This is referred to as robust ML, which is a practical alternative to ADF in which one continues to estimate the structural parameters with standard estimator such as ML, and then adjusts the test statistic and the standard errors so that they become robust to non-normality. It is known that when model is correctly specified, the usual standard errors for ML estimates can be obtained from the inverse of Fisher information. However, when there is distributional misspecification, the covariance matrix of the ML estimates is of a

"sandwich" form (White, 1982). If we retain the consistent point estimates and replace the usual standard errors with adjusted standard errors, statistical inference will be robust against non-normality.

Satorra and Bentler's (1994) theoretical development can be illustrated as follows. When the observed variables are multivariate normally distributed, a typical element in the asymptotic covariance matrix $\mathbf{V}$ simplifies to $v_{ij,kl} = \sigma_{ik}\sigma_{jl} - \sigma_{il}\sigma_{jk}$. This is because the fourth order moment in the population $\sigma_{ijkl}$ simplifies to the product of second order moments $\sigma_{ik}$ and $\sigma_{jl}$ when there is no excess kurtosis.

Let us use $\mathbf{V}_N$ to denote asymptotic covariance matrix of $\boldsymbol{s}$ derived under normality to distinguish it from the sample-based (ADF) weight matrix $\mathbf{V}_S$. In other words, $\mathbf{V}_N$ has as a typical element $\tilde{v}_{ij,kl} = s_{ik}s_{jl} - s_{il}s_{jk}$. Classical normal theory covariance structure modeling with the GLS or ML discrepancy function relies on the fact that the sample moments $\boldsymbol{s}$ have an asymptotic covariance matrix of the above form. However, with non-normal data, using the ML or GLS discrepancy function amounts to using WLS estimation with an incorrectly specified weight matrix. Although the parameter estimates will still be consistent, the standard errors and goodness-of-fit statistics will be incorrect.

Given a consistent estimator $\widehat{\boldsymbol{\theta}}$, let us denote the model-implied covariances evaluated at the parameter estimates as $\widehat{\boldsymbol{\sigma}} = \boldsymbol{\sigma}(\widehat{\boldsymbol{\theta}})$**.** Denote the Jacobian matrix evaluated at the parameter estimates as

$$\dot{\boldsymbol{\sigma}} = \frac{\partial \boldsymbol{\sigma}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}. \tag{14}$$

The following residual-based statistic $c_2$ is asymptotically chi-square distributed if the normality assumption holds:

$$c_2 = (N-1)(\boldsymbol{s} - \hat{\boldsymbol{\sigma}})' \dot{\boldsymbol{\sigma}}_c (\dot{\boldsymbol{\sigma}}_c' \mathbf{V}_N \dot{\boldsymbol{\sigma}}_c)^{-1} \dot{\boldsymbol{\sigma}}_c' (\boldsymbol{s} - \hat{\boldsymbol{\sigma}}). \tag{15}$$

where $\dot{\boldsymbol{\sigma}}_c$ is an orthogonal complement of $\dot{\boldsymbol{\sigma}}$ so that $\dot{\boldsymbol{\sigma}}_c' \dot{\boldsymbol{\sigma}} = \mathbf{0}$. This is referred to as the normal theory WLS chi-square statistic.

The Satorra-Bentler (1994) correction is called a single-moment adjustment that re-scales the $c_2$ statistic so that when the weight matrix is incorrectly specified, the re-scaled statistic can be better approximated by a chi-square variable in large samples. Specifically, it will have a correct asymptotic expected value. The correction takes the following form:

$$c_3 = \frac{d}{tr[(\dot{\boldsymbol{\sigma}}_c' \mathbf{V}_S \dot{\boldsymbol{\sigma}}_c)(\dot{\boldsymbol{\sigma}}_c' \mathbf{V}_N \dot{\boldsymbol{\sigma}}_c)^{-1}]} c_2, \tag{16}$$

The statistic $c_3$ is referred to as the Satorra-Bentler scaled chi-square statistic. Note that this corrects $c_2$, the residual based statistic, while the typical application of the Satorra-Bentler is to correct $c_1$, the ML chi-square statistic.

The robust standard errors of the parameter estimates are given by the square roots of the diagonal elements of the following sandwich covariance matrix estimator:

$$\left(\dot{\boldsymbol{\sigma}}' \mathbf{V}_N{}^{-1} \dot{\boldsymbol{\sigma}}\right)^{-1} \left(\dot{\boldsymbol{\sigma}}' \mathbf{V}_N{}^{-1} \mathbf{V}_S \mathbf{V}_N{}^{-1} \dot{\boldsymbol{\sigma}}\right) \left(\dot{\boldsymbol{\sigma}}' \mathbf{V}_N{}^{-1} \dot{\boldsymbol{\sigma}}\right)^{-1}. \tag{17}$$

Sandwich expressions identical to Equation (17) can be derived from the WLS discrepancy function (7), as in, e.g., Browne's (1984) Proposition 2.

As an aside, since ML or GLS can be understood as members of the WLS family with normal theory weight matrices, the WLS family provides the link between robust ML and ADF. In the case of ADF estimation, the weight matrix is $\mathbf{W} = \mathbf{V}_S$**.** If we plug $\mathbf{V}_S$ into equation (17), the entire expression simplifies to $(\dot{\boldsymbol{\sigma}}'\mathbf{W}^{-1}\dot{\boldsymbol{\sigma}})^{-1}$. In this case, we say that the weight is correctly specified. Otherwise, the more general sandwich formula will lead to asymptotically correct standard errors as long as a consistent estimator of $\boldsymbol{\theta}$ is used.

When the data are distributed non-normally, we propose to use the robust ML method, which is to use ML parameter estimates in conjunction with the robust standard errors obtained from the sandwich estimator in equation (17) and the Satorra-Bentler Scaled Chi-Square $c_3$ statistic. This approach turns out to work well for practical sample and model sizes and is recommended over distribution-free methods such as ADF.

Finally, the following residual-based statistic $c_4$ is asymptotically chi-square distributed for any consistent and asymptotically normal estimator, whether the observed variables are normally distributed or not:

$$c_4 = (N-1)(\boldsymbol{s} - \widehat{\boldsymbol{\sigma}})'\dot{\boldsymbol{\sigma}}_c(\dot{\boldsymbol{\sigma}}_c'\mathbf{V}_S\dot{\boldsymbol{\sigma}}_c)^{-1}\dot{\boldsymbol{\sigma}}_c'(\boldsymbol{s} - \widehat{\boldsymbol{\sigma}}). \quad (18)$$

This is referred to as Browne's residual-based statistic corrected for non-nomality (see Browne, 1984, Proposition 4). Unfortunately, as with the ADF-based minimum WLS fit function statistic, $c_4$ does not work well at smaller sample sizes. Yuan and Bentler (1997) proposed an adjustment to $c_4$ that works well when the sample size is small:

$$T_{YB} = \frac{c_4}{\left(1 + \frac{c_4}{N-1}\right)}. \tag{19}$$

As sample size increases, $T_{YB}$ becomes equal to $c_4$. At the smallest sample size, the Yuan-Bentler *F*-statistic (Yuan & Bentler, 1999) in EQS is probably better. However, in situations when the sample size is too small and where ADF cannot be resurrected, the robust test statistics and standard errors are recommended as an alternative to ADF. See Bentler (2006) for suggestions on how to obtain the standard errors and test statistics under different distribution assumptions.

### 2.2. Partial Least Squares for CFA Models

Schneeweiss (1990) raised the idea that PLS is not just an estimation method, but rather a particular way of defining latent variables and their relations to the observed variables without assuming any specific model for the observed variables. PLS, as an estimation method, then estimates these variables and relations from given data. Although this idea is not shared by many adherents of the PLS approach, the basic

approach of using PLS for SEM is, in general, to create proxies that are constructed as linear composites of observed variables.

This dissertation study focuses on the use of PLS for Confirmatory Factor Analysis (CFA) type of models. Apart from the many approaches to implement PLS, Wold's (1982) basic design, as the origin of PLS for SEM is well suited for the CFA type of models. We use the procedure as described by Dijkstra (2011) to show how to obtain the PLS estimates in the basic design. This also makes it easier to explain how to follow up the procedure with the new PLS-consistent estimators proposed by Dijkstra (2011) and subsequently the TSLS method as implemented in Dijkstra (2012).

In the basic design by Wold (1982), there are two basic types of algorithms called mode A and mode B, and a third type, mode C which is a mixture of the two (Dijkstra, 2010). For each mode, the goal is to estimate a weight vector $\widehat{\boldsymbol{w}}_i$ in order to construct proxies from the observed variables. Based on the estimates of the weights, factor loading and factor correlations can then be obtained. The current approach falls under Mode A.

For the purpose of model identification, we assume latent variables have been standardized. Since we make no differentiation between endogenous variables and exogenous variables at this stage, let us use $\boldsymbol{z}$ to represent all observed variables and $\boldsymbol{u}$ to represent all latent variables: $\boldsymbol{z} = \begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{x} \end{pmatrix}$, $\boldsymbol{u} = \begin{pmatrix} \boldsymbol{\eta} \\ \boldsymbol{\xi} \end{pmatrix}$. Let there be $p$ latent variables and $m$

manifest variables in total. In discussions of PLS and PLSc, the observed variables are assumed to be standardized. Then from equation (5), we have

$$V(\boldsymbol{z}) = \mathbf{P} = \boldsymbol{\Lambda}\boldsymbol{\Upsilon}\boldsymbol{\Lambda}' + \boldsymbol{\Omega}, \tag{20}$$

where **P** standards for a correlation matrix with unit diagonals.

Since we assume unit variance of the factor, the diagonal elements of the $\boldsymbol{\Upsilon}$ matrix are 1. We also assume no correlated residuals. Thus $\boldsymbol{\Omega}$ is a diagonal matrix. When considering models in which every variable is influenced by only one factor, we can reorder the columns and rows of the loading matrix so that $\boldsymbol{\Lambda}$ has a form of "independent clusters," a specific factor pattern in CFA:

$$\boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{\lambda}_1 & & & & \\ & \boldsymbol{\lambda}_2 & & & \\ & & \ddots & & \\ & & & \boldsymbol{\lambda}_i & \\ & & & & \ddots \end{pmatrix}.$$

We use *i* as a notational device to index all manifest variables in the *i*th block. For example, $\boldsymbol{\Omega}_{ii}$ represents the covariance matrix of the *i*th block of unique factors. The variances of the *i*th block of observed variables can be written as $\mathbf{P}_{ii} = V(\boldsymbol{z}_i) = \boldsymbol{\lambda}_i\boldsymbol{\lambda}_i' + \boldsymbol{\Omega}_{ii}$, where the first part represent the common variances and the second part represent the unique variances. Since we assume all the manifest variables are standardized before being analyzed, there is no need to estimate the unique variance, as they can be easily obtained from 1 minus the estimated common variances. Similarly, let us use *j* to index all such variables in the *j*th block, where $i \neq j$. The covariances of the observed

variables can then be written as $\mathbf{P}_{ij} = V(\boldsymbol{z}_i, \boldsymbol{z}_j) = (\boldsymbol{\lambda}_i \boldsymbol{\lambda}_j')v_{ij}$, where $v_{ij}$ is the covariance between latent variable *i* and latent variable *j*, i.e., the (*i,j*)th element in $\boldsymbol{\Upsilon}$.

PLS produces estimates of the loadings and the factor correlations, all of which depend on the estimation of the weight vector $\widehat{\boldsymbol{w}}_i$, which has to be the same order as $\boldsymbol{\lambda}_i$ and $\boldsymbol{z}_i$. With these estimated weights, proxies can be estimated as inner product of $\widehat{\boldsymbol{w}}_i$ and $\boldsymbol{z}_i$: $\hat{u}_i = \widehat{\boldsymbol{w}}_i' \boldsymbol{z}_i$ . Suppose we have a sample of size *N* and let **R** represent the sample correlation matrix. **R** is a consistent estimator of the population correlation matrix **P**, i.e., $plim(\mathbf{R}) = \mathbf{P}$ (Rice, 1995). A side condition for the weights is that $\widehat{\boldsymbol{w}}_i' \mathbf{P}_{ii} \widehat{\boldsymbol{w}}_i = 1$ as the sample data are assumed to be standardized before being analyzed.

There are several iterative fixed-point algorithms for PLS. According to Dijkstra (2011), the co-called "Mode A algorithm" is in general the most numerically stable algorithm and it typically converges quickly. Mode A is briefly described below. For technical details, see Wold (1982) or Dijkstra (1981, 1983).

Recall that *p* is the number of latent variables. Let $\mathbf{R}_{ij}$ represent the corresponding sub-matrix consisting of correlations of the indicators related to latent variables *i* and *j*. Let $\widehat{\boldsymbol{w}}^{(t)}$ represent a vector containing all the weights estimated at iteration *t*, i.e., $\widehat{\boldsymbol{w}}^{(t)} = (\widehat{\boldsymbol{w}}_1^{(t)}, \widehat{\boldsymbol{w}}_2^{(t)}, \dots, \widehat{\boldsymbol{w}}_i^{(t)}, \dots, \widehat{\boldsymbol{w}}_p^{(t)})$.

Iteration *t* of the Mode A algorithm consists of the following steps:

(1) For all $i \neq j, i = 1, 2, \dots, p$ and $j = 1, 2, \dots, p$, define a sign factor $s_{ij}^{(t)} = sgn(\widehat{\boldsymbol{w}}_i^{(t)\prime}\mathbf{R}_{ij}\widehat{\boldsymbol{w}}_j^{(t)})$, which returns the sign function of the sample covariances between the estimated proxies $\hat{u}_i = \widehat{\boldsymbol{w}}_i'\boldsymbol{z}_i$ and $\hat{u}_j = \widehat{\boldsymbol{w}}_j'\boldsymbol{z}_j$ in iteration $t$.

(2) Update the weights:

$$\widetilde{\boldsymbol{w}}_i^{(t)} = \sum_{j=1, i\neq j}^{p} s_{ij}^{(t)}\,\mathbf{R}_{ij}\widehat{\boldsymbol{w}}_j^{(t)}.$$

(3) Normalize the weights and take absolute values:

$$\widehat{\boldsymbol{w}}_i^{(t+1)} = \left| \frac{\widetilde{\boldsymbol{w}}_i^{(t)}}{\sqrt{\widetilde{\boldsymbol{w}}_i^{(t)\prime}\mathbf{R}_{ij}\widetilde{\boldsymbol{w}}_i^{(t)}}} \right|.$$

Now we can go back to step (1) with $t$ incrementing by 1. Iterate until the convergence criterion is met (e.g. absolute weight change smaller than a certain number) or maximum number of iterations exceeded. The starting values for the weight vector are essentially arbitrary. In general, $\widetilde{\boldsymbol{w}}^{(0)} = (1, \dots, 1)$ is sufficient, such that after initial normalization, the starting values become

$$\widehat{\boldsymbol{w}}_i^{(0)} = \left| \frac{\widetilde{\boldsymbol{w}}_i^{(0)}}{\sqrt{\widetilde{\boldsymbol{w}}_i^{(0)\prime}\mathbf{R}_{ij}\widetilde{\boldsymbol{w}}_i^{(0)}}} \right|.$$

PLS iterations yields sample proxies $\hat{u}_i$ and weights $\widehat{\boldsymbol{w}}_i$ for all the latent variables. Traditionally, we use the sample proxies to replace the latent variables to estimate the latent variable correlation matrix. However, this replacement can never be exhaustive

unless there are no measurement errors in the proxies (Dijkstra, 2010). In the PLS literature, Jöreskog and Wold (1982) discussed the issue of PLS being consistent "at large." That is, the inconsistency due to measurement error will tend to zero if more indicators of sufficient quality can be introduced, and when sample size tends to infinity. According to Dijkstra (2011), unless there are a large number of high quality indicators, the PLS-proxies will tend to underestimate the correlations between the latent variables. In addition, PLS also often overestimates the loadings in absolute value. Hence, some corrections are in order to make the PLS estimates consistent.

### 2.3. Consistent Partial Least Squares (PLSc)

Based on the PLS estimated weights $\widehat{\boldsymbol{w}}_i$, Dijkstra (2011) proposed a method to rescale the PLS estimates to reproduce sample correlation matrix **R** as well as possible. This leads to consistent estimates of factor loadings and latent variable correlations. Note that Mode A has the property that in the probability limit, the weights are proportional to the factor loadings:

$$\bar{\boldsymbol{w}}_i = plim(\widehat{\boldsymbol{w}}_i) = \frac{\boldsymbol{\lambda}_i}{\sqrt{\boldsymbol{\lambda}_i' \mathbf{P}_{ii} \boldsymbol{\lambda}_i}}. \tag{21}$$

That is to say, the factor loadings $\boldsymbol{\lambda}_i = \bar{c}_i \bar{\boldsymbol{w}}_i$, where $\bar{c}_i$ may be understood as a correction factor. According to Dijkstra (2011), the correction factor $\bar{c}_i$ can be consistently estimated by

$$\hat{c}_i^2 = \frac{\widehat{\boldsymbol{w}}_i'(\mathbf{R}_{ii} - diag(\mathbf{R}_{ii}))\widehat{\boldsymbol{w}}_i}{\widehat{\boldsymbol{w}}_i'(\widehat{\boldsymbol{w}}_i\widehat{\boldsymbol{w}}_i' - diag(\widehat{\boldsymbol{w}}_i\widehat{\boldsymbol{w}}_i'))\widehat{\boldsymbol{w}}_i}. \quad (22)$$

The factor loading are corrected consistently as $\hat{\boldsymbol{\lambda}}_i = \hat{c}_i\widehat{\boldsymbol{w}}_i$. The quality of a latent variable proxy can be measured by its squared multiple correlation with the true latent variable, $Q^2(u_i, \bar{u}_i) = (\overline{\boldsymbol{w}}_i'\boldsymbol{\lambda}_i)^2 = (\overline{\boldsymbol{w}}_i'\overline{\boldsymbol{w}}_i)^2\bar{c}_i^2$ . Therefore, we can correct the squared multiple correlations consistently by $\hat{Q}^2(u_i, \bar{u}_i) = (\widehat{\boldsymbol{w}}_i'\widehat{\boldsymbol{w}}_i)^2\hat{c}_i^2$. The correlations between the latent variables can then be corrected because

$$\hat{v}_{ij}^2 = \frac{\hat{Q}^2(\bar{u}_i, \bar{u}_j)}{\hat{Q}^2(u_i, \bar{u}_i)\hat{Q}^2(u_j, \bar{u}_j)} = \frac{(\widehat{\boldsymbol{w}}_i'\mathbf{R}_{ij}\widehat{\boldsymbol{w}}_j)^2}{(\widehat{\boldsymbol{w}}_i'\widehat{\boldsymbol{w}}_i)^2\hat{c}_i^2(\widehat{\boldsymbol{w}}_j'\widehat{\boldsymbol{w}}_j)^2\hat{c}_j^2}.$$

Since corrected estimates are obtained from the PLS estimates, we call them PLS-corrected (PLSc) parameter estimates. The estimated factor loadings and the factor correlations are corrected so that measurement errors are taken into account. With the corrections, PLSc yields a consistent estimate of the population unique factor covariance matrix $\boldsymbol{\Omega}$ as $\widehat{\boldsymbol{\Omega}}_c = diag(\mathbf{R} - \widehat{\boldsymbol{\Lambda}}\widehat{\boldsymbol{\Upsilon}}\widehat{\boldsymbol{\Lambda}}')$**.** This implies that PLSc also yields a consistent estimate of the population correlation matrix as

$$\widehat{\mathbf{P}}_c = \widehat{\boldsymbol{\Lambda}}_c\widehat{\boldsymbol{\Upsilon}}_c\widehat{\boldsymbol{\Lambda}}_c' + \widehat{\boldsymbol{\Omega}}_c. \quad (23)$$

**2.4. Two-Stage Least Squares**

With the above PLSc procedure, the parameters in the measurement model are consistently estimated. The next step is to find the structural relationships between the latent variables, using the PLSc estimates of the factor correlations ($\widehat{\boldsymbol{\Upsilon}}_c$) as the input.

Returning to the structural model: $\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}$, the exogenous latent variable $\boldsymbol{\xi}$ in the equation are assumed to be uncorrelated with the disturbance term $\boldsymbol{\zeta}$. However, since we have multiple simultaneous equations as a system, it is hard to assume that the disturbance terms between the equations are uncorrelated. Thus it is also impossible to assume that the endogenous variables $\boldsymbol{\eta}$ are uncorrelated with the disturbances, especially when they may also be serving as the predictors of the other endogenous latent variables in the equation system. When the predictors are correlated with the disturbances, one cannot use Ordinary Least Squares (OLS) regression to estimate the regression coefficients, because the OLS estimator is not consistent. One potential remedy is Two-Stage Least Squares (TSLS; see Mardia, Kent and Bibby, 1979).

TSLS is a single-equation estimator. It involves two steps. For notational convenience, we will work with the *i*th structural equation. This equation can be written in the form of a regression model:

$$\eta_i = \boldsymbol{\beta}'_{(i)}\boldsymbol{\eta}_{(i)} + \boldsymbol{\gamma}'_i\boldsymbol{\xi} + \zeta_i \tag{24}$$

where $\boldsymbol{\eta}_{(i)}$ is a vector of endogenous latent variables excluding $\eta_i$**.** The vector $\boldsymbol{\beta}_{(i)}$ represents the relationships between $\eta_i$ and the other endogenous latent variables in $\boldsymbol{\eta}_{(i)}$. The vector $\boldsymbol{\gamma}_i$ represents relationships between $\eta_i$ and the exogenous latent variables. In other words, if we define $\boldsymbol{\beta}_i$ as the *i*th row of the **B** matrix, $\boldsymbol{\beta}_{(i)}$ is equal to $\boldsymbol{\beta}_i$ with the *i*th element $\beta_{ii}$ removed. This is because $\beta_{ii}$ is equal to 0. It is clear from

Equation (15) that due to the correlation between $\boldsymbol{\eta}_{(i)}$ and $\zeta_i$, OLS estimation of $\boldsymbol{\beta}_{(i)}$ will lead to inconsistent results.

In the first stage of TSLS, we regress $\boldsymbol{\eta}_{(i)}$ on $\boldsymbol{\xi}$ to obtain fitted values $\boldsymbol{\eta}^*_{(i)}$. This is valid because $\boldsymbol{\xi}$ is exogenous and hence uncorrelated with the random disturbance $\boldsymbol{\zeta}$. The equation implies that it can serve as its own instrument for $\boldsymbol{\eta}_{(i)}$. In the second stage of TSLS, substituting $\boldsymbol{\eta}^*_{(i)}$ for $\boldsymbol{\eta}_{(i)}$ in equation (15), a new regression equation is defined, in which the variables on the right hand side are no longer correlated with the disturbance term

$$\eta_i = \boldsymbol{\beta}'_{(i)}\boldsymbol{\eta}^*_{(i)} + \boldsymbol{\gamma}'_i\boldsymbol{\xi} + \zeta_i. \tag{25}$$

In other words, one may use OLS to solve Equation (25).

Previously with the PLSc procedure, we already estimated the factor correlation matrix $\widehat{\boldsymbol{\Upsilon}}_c$. The matrix $\widehat{\boldsymbol{\Upsilon}}_c$ can be partitioned into four parts as follows:

$$\widehat{\boldsymbol{\Upsilon}}_c = \begin{bmatrix} \widehat{\boldsymbol{\Upsilon}}_{\eta\eta} & \widehat{\boldsymbol{\Upsilon}}_{\eta\xi} \\ \widehat{\boldsymbol{\Upsilon}}_{\xi\eta} & \widehat{\boldsymbol{\Upsilon}}_{\xi\xi} \end{bmatrix},$$

where $\widehat{\boldsymbol{\Upsilon}}_{\xi\xi} = \widehat{\boldsymbol{\Phi}}$, and $\widehat{\boldsymbol{\Upsilon}}_{\xi\eta} = \widehat{\boldsymbol{\Upsilon}}'_{\eta\xi}$. The TSLS estimator of $\boldsymbol{\beta}_{(i)}$ and $\boldsymbol{\gamma}_i$ can be written in terms of the blocks of the factor correlation/covariance matrix as

$$\begin{bmatrix} \boldsymbol{\beta}_{(i)} \\ \boldsymbol{\gamma}_i \end{bmatrix} = \begin{bmatrix} \widehat{\boldsymbol{\Upsilon}}'_{\xi\eta_{(i)}}\widehat{\boldsymbol{\Upsilon}}^{-1}_{\xi\xi}\widehat{\boldsymbol{\Upsilon}}_{\xi\eta_{(i)}} & \widehat{\boldsymbol{\Upsilon}}_{\eta_{(i)}\xi} \\ \widehat{\boldsymbol{\Upsilon}}_{\xi\eta_{(i)}} & \widehat{\boldsymbol{\Upsilon}}_{\xi\xi} \end{bmatrix}^{-1} \begin{bmatrix} \widehat{\boldsymbol{\Upsilon}}'_{\xi\eta_{(i)}}\widehat{\boldsymbol{\Upsilon}}^{-1}_{\xi\xi}\widehat{\boldsymbol{\Upsilon}}_{\xi\eta_i} \\ \widehat{\boldsymbol{\Upsilon}}_{\xi\eta_i} \end{bmatrix}, \tag{26}$$

where $\widehat{\boldsymbol{\Upsilon}}_{\xi\eta_{(i)}}$ is equal to $\widehat{\boldsymbol{\Upsilon}}_{\xi\eta}$ with the *i*th column removed, and $\widehat{\boldsymbol{\Upsilon}}_{\eta_{(i)}\xi} = \widehat{\boldsymbol{\Upsilon}}'_{\xi\eta_{(i)}}$.

## 2.5. Issues Pertaining to Rescaling

It is worth noting that PLS and PLSc requires sample correlation matrices as input. However, in structural equation modeling, use of sample correlation matrices for covariance structural analysis can be potentially problematic. Cudeck (1989) discussed whether it would be appropriate to use sample correlation matrix for a covariance structure with three kinds of CFA models. Due to the model specification required by PLS estimation, PLSc uses common factor analysis models that fall into the third category where all the loadings are freely estimated and all the factors have variances of 1. This kind of models, together those from the first category (a slight variation of category 3 where the factor variances are freely estimated and some of the loadings are constrained to be 1 for scaling indicators) are scale invariant and one could replace sample covariance matrix with sample correlation matrix with no loss of generality.

The caveat, of course, has to do with how the PLSc and TSLS estimates are going to be used subsequently. In the new approaches that will be discussed in sections 2.6 and 2.7, the theory of covariance structure analysis is used. Thus, some rescaling will be necessary. Recall that from Equation (4), the covariance matrix of the latent variables is

$$V\begin{pmatrix}\boldsymbol{\eta}\\ \boldsymbol{\xi}\end{pmatrix} = \boldsymbol{\Upsilon} = \begin{pmatrix}\boldsymbol{\Upsilon}_{\eta\eta} & \\ \boldsymbol{\Upsilon}_{\xi\eta} & \boldsymbol{\Upsilon}_{\xi\xi}\end{pmatrix}.$$

And the covariance matrix of the observed variables is $\boldsymbol{\Sigma} = \boldsymbol{\Lambda\Upsilon\Lambda}' + \boldsymbol{\Omega}$. As described in section 2.2 and 2.3, PLS and subsequently PLSc procedures require that all observed

variables be standardized, all the loadings freely estimated and all the latent factors (regardless whether they are exogenous or endogenous) have unit variances. That is, PLSc treats as its input a correlation matrix $\mathbf{P} = \mathbf{D}_{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\mathbf{D}_{\boldsymbol{\Sigma}}^{-1}$, where $\mathbf{D}_{\boldsymbol{\Sigma}}$ is a diagonal matrix containing the observed variables' standard deviations, i.e., the square roots of diagonal elements of $\boldsymbol{\Sigma}$. As such, the PLS- or PLSc-obtained factor loadings are equivalent to standard regression coefficients. From Equation (23), $\widehat{\mathbf{P}}_c = \widehat{\boldsymbol{\Lambda}}_c\widehat{\boldsymbol{\Upsilon}}_c\widehat{\boldsymbol{\Lambda}}_c' + \widehat{\boldsymbol{\Omega}}_c$, and $\widehat{\mathbf{P}}_c$ has unit diagonals.

Let $\mathbf{D}_S$ be a diagonal matrix containing the observed variables' sample standard deviations, i.e., the square roots of the diagonal elements of the sample covariance matrix $\mathbf{S}$. $\mathbf{D}_{\mathbf{S}}$ is a consistent estimate of $\mathbf{D}_{\boldsymbol{\Sigma}}$. Hence, $\mathbf{D}_{\mathbf{S}}\widehat{\mathbf{P}}_c\mathbf{D}_{\mathbf{S}}$ is a consistent estimate of the the covariance matrix of manifest variables $\boldsymbol{\Sigma}$. This can also be rewritten as

$$\widehat{\boldsymbol{\Sigma}}_c = \mathbf{D}_{\mathbf{S}}\widehat{\mathbf{P}}_c\mathbf{D}_{\mathbf{S}} = \mathbf{D}_{\mathbf{S}}\widehat{\boldsymbol{\Lambda}}_c\widehat{\boldsymbol{\Upsilon}}_c\widehat{\boldsymbol{\Lambda}}_c'\mathbf{D}_{\mathbf{S}}' + \mathbf{D}_{\mathbf{S}}\widehat{\boldsymbol{\Omega}}_c\mathbf{D}_{\mathbf{S}}'. \quad (27)$$

Alternatively, if we take $\widehat{\boldsymbol{\Lambda}} = \mathbf{D}_{\mathbf{S}}\widehat{\boldsymbol{\Lambda}}_c$ as the rescaled factor loading matrix estimate, PLSc can provide a consistent estimate of the observed variables' covariance matrix even if the observed variables are not standardized. This will subsequently be important (see Section 2.6) as one of the two approaches requires $\widehat{\boldsymbol{\Sigma}}_c$.

On the other hand, TSLS takes as its input a PLSc-estimated factor correlation matrix $\widehat{\boldsymbol{\Upsilon}}_c$. From Equation (26), the estimated regression coefficients are effectively standardized coefficients. While this may be convenient for interpretive purposes, it is

not entirely consistent with how identification conditions in typical covariance structure analysis software programs are implemented. Take Equation (24) as an example:

$$\eta_i = \boldsymbol{\beta}'_{(i)}\boldsymbol{\eta}_{(i)} + \boldsymbol{\gamma}'_i\boldsymbol{\xi} + \zeta_i.$$

TSLS estimation treats $\eta_i$ and $\boldsymbol{\xi}$ as standardized variables, leaving open the possibility of estimating the residual variance of $\zeta_i$, which, in the current context, $V(\zeta_i)$ is effectively the proportion of unexplained variance in $\eta_i$. In most covariance structure analysis software programs, it is customary to fix the variances of the equation disturbance terms in $\boldsymbol{\zeta}$, to, say, 1.0, and leaving open the possibility that the endogenous latent variables in $\boldsymbol{\eta}$ may be unstandardized. This requirement calls for another rescaling, of the TSLS estimates of regression parameters, as well as the factor loadings associated with the endogenous latent variables.

Recall that in applying TSLS to PLSc-estimated factor correlation matrix, because the assumption is that $V(\eta_i) = 1$, we have

$$V(\zeta_i) = V(\eta_i) - V_M = \frac{V(\eta_i) - V_M}{V(\eta_i)} = 1 - \frac{V_M}{V(\eta_i)} = 1 - R_i^2,$$

where $V_M$ represents modeled or explained variance in $\eta_i$, and $R_i^2$ is the proportion of modeled or explained variance in the *i*th structural equation. As such, one might propose the use of the reciprocal of the proportion of unexplained variance as a scaling factor to rescale variance of the outcome variable $\eta_i$, in which case, the variance of the equation disturbance term becomes 1.

More formally, let $\psi_{ii}$ be the disturbance variance associated with equation $i$. Let

$$\mathbf{D}_{\psi} = \begin{pmatrix} \text{diag}(\boldsymbol{\Psi}) & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \tag{28}$$

be a $p \times p$ diagonal matrix where the upper-left block contains the equation disturbance variances $\psi_{ii}$ on the diagonal, and the remaining diagonal elements corresponding to the exogenous latent variables are identically equal to 1 (because the $\xi$'s are standardized). Let $\mathbf{D}_{\psi}^{1/2}$ be a symmetric square root matrix of $\mathbf{D}_{\psi}$, i.e., $\mathbf{D}_{\psi} = \mathbf{D}_{\psi}^{1/2}\mathbf{D}_{\psi}^{1/2}$. The rescaling is first directly applied to the PLSc-estimated factor correlation:

$$\widehat{\boldsymbol{\Upsilon}}_{*} = \mathbf{D}_{\psi}^{-1/2}\widehat{\boldsymbol{\Upsilon}}_{c}\mathbf{D}_{\psi}^{-1/2}, \tag{29}$$

and the resulting matrix $\widehat{\boldsymbol{\Upsilon}}_{*}$ will have non-unitary diagonals for the endogenous latent variables $\boldsymbol{\eta}$. Running TSLS estimation for a second time using $\widehat{\boldsymbol{\Upsilon}}_{*}$ will produce regression coefficient estimates that are appropriately adjusted and the equation error variances will be unitary.

The adjustment in (29) does change the scale of the endogenous latent variables, so for their corresponding factor loadings, another (final) rescaling is necessary. Recall that the rescaled PLSc-estimated factor loadings are given by $\widehat{\boldsymbol{\Lambda}} = \mathbf{D}_{\mathbf{S}}\widehat{\boldsymbol{\Lambda}}_{c}$. Using the same notation, the following rescaling

$$\widehat{\boldsymbol{\Lambda}}_{*} = \widehat{\boldsymbol{\Lambda}}\mathbf{D}_{\psi}^{-1/2} = \mathbf{D}_{\mathbf{S}}\widehat{\boldsymbol{\Lambda}}_{c}\mathbf{D}_{\psi}^{-1/2}, \tag{30}$$

will produce a factor loading matrix that is adjusted to correspond to the usual identification condition in covariance structure analysis that the equation disturbance

terms have unit variance. Note that the scaling of the exogenous latent variables does not change because they already have unit variances, i.e., $\widehat{\boldsymbol{\Upsilon}}_{\boldsymbol{\xi\xi}}$ remains unaltered.

### 2.6. One-Step Improvement (PLSe1)

In this section and the following one, we describe two efficient PLS estimation methodologies. They are based on the PLSc estimates for the measurement model and the regression coefficients obtained via TSLS.

The first efficient PLS estimator (PLSe1) utilizes a one-step improvement estimation methodology to obtain an efficient estimator and the associated parameter and model fit test statistics. Due to the consistency of functions of consistent estimators (Rao, 1973), this is equivalent to a single iteration of Newton-Raphson.

There are different estimation methods one could use for the one-step improvement, such as ML, WLS, robust ML and ADF. As shown in Van der Vaart (1998), a one-step Newton-Raphson improvement can be written generally as:

$$\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}_1 - \left(H(\boldsymbol{\theta}_1)\right)^{-1} g(\boldsymbol{\theta}_1) \tag{31}$$

where the starting values $\boldsymbol{\theta}_1$ should be a consistent but not necessarily efficient estimator. $H(\boldsymbol{\theta})$ is the Hessian (second-order derivative) matrix of the discrepancy function, e.g., $\partial^2 F_{ML}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}^2$ or $\partial^2 F_{WLS}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}^2$, and $g(\boldsymbol{\theta})$ is the gradient of the discrepancy function. The gradient and the Hessian are evaluated at $\boldsymbol{\theta}_1$.

The one step improvement can also be motivated using Bentler and Dijkstra's (1985) theory, which involves linearization. We assume that the random vector $\boldsymbol{\theta}_1$ is not necessarily asymptotic normal, but is consistent and have the property that $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_+ + O_P(1/\sqrt{N})$, where $\boldsymbol{\theta}_+$ represents the true parameter, $N$ is sample size and the symbol $O_P$ means that the sequence $\sqrt{N}\boldsymbol{\theta}_1$ is bounded in probability. Overall, $\boldsymbol{\theta}_1$ is a $\sqrt{N}$-consistent estimator of $\boldsymbol{\theta}_+$ .

We can approximate a function $\boldsymbol{\sigma}(\boldsymbol{\theta})$ by using a finite number of terms in its Taylor-series expansion. We can expand $\boldsymbol{\sigma}(\boldsymbol{\theta})$ around $\boldsymbol{\theta}_1$, which yields, approximately, $\boldsymbol{\sigma}(\boldsymbol{\theta}) \approx \boldsymbol{\sigma}(\boldsymbol{\theta}_1) + \dot{\boldsymbol{\sigma}}(\boldsymbol{\theta}_1)(\boldsymbol{\theta} - \boldsymbol{\theta}_1)$, where $\dot{\boldsymbol{\sigma}}(\boldsymbol{\theta}) = \partial\boldsymbol{\sigma}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}'$ is the Jacobian matrix of the model. Let $\boldsymbol{\sigma}_1 = \boldsymbol{\sigma}(\boldsymbol{\theta}_1)$ and $\dot{\boldsymbol{\sigma}}_1 = \dot{\boldsymbol{\sigma}}(\boldsymbol{\theta}_1)$**.** The covariance structure model becomes $\boldsymbol{\sigma}(\boldsymbol{\theta}) \approx \boldsymbol{\sigma}_1 + \dot{\boldsymbol{\sigma}}_1(\boldsymbol{\theta} - \boldsymbol{\theta}_1)$. Substituting the approximation $\boldsymbol{\sigma}_1 + \dot{\boldsymbol{\sigma}}_1(\boldsymbol{\theta} - \boldsymbol{\theta}_1)$ for $\boldsymbol{\sigma}(\boldsymbol{\theta})$, the following WLS fit function is obtained:

$$F_{WLS}(\boldsymbol{\theta}) = (\boldsymbol{s} - \boldsymbol{\sigma}_1 - \dot{\boldsymbol{\sigma}}_1\boldsymbol{\theta} + \dot{\boldsymbol{\sigma}}_1\boldsymbol{\theta}_1)'\mathbf{W}^{-1}(\boldsymbol{s} - \boldsymbol{\sigma}_1 - \dot{\boldsymbol{\sigma}}_1\boldsymbol{\theta} + \dot{\boldsymbol{\sigma}}_1\boldsymbol{\theta}_1). \quad (32)$$

We can take the first derivative of $F_{WLS}(\boldsymbol{\theta})$ and set it to 0 to minimize the fit function:

$$\frac{dF_{WLS}}{d\boldsymbol{\theta}} = 2(\boldsymbol{s} - \boldsymbol{\sigma}_1 - \dot{\boldsymbol{\sigma}}_1\boldsymbol{\theta} + \dot{\boldsymbol{\sigma}}_1\boldsymbol{\theta}_1)'\mathbf{W}^{-1}\dot{\boldsymbol{\sigma}}_1 = \mathbf{0}.$$

This equation can be solved analytically, which shows that the one step improvement estimator can be written as:

$$\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}_1 - (\dot{\boldsymbol{\sigma}}_1'\mathbf{W}^{-1}\dot{\boldsymbol{\sigma}}_1)^{-1}\dot{\boldsymbol{\sigma}}_1'\mathbf{W}^{-1}(\boldsymbol{s} - \boldsymbol{\sigma}_1) \quad (33)$$

If $\mathbf{W}$ is correctly specified, e.g., $\mathbf{W} = \mathbf{V}_S$, the covariance matrix of $\widehat{\boldsymbol{\theta}}$ can be estimated from the following result on the asymptotic normality of $\widehat{\boldsymbol{\theta}}$:

$$\sqrt{N}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_+\right) \xrightarrow{D} N(\mathbf{0}, (\dot{\boldsymbol{\sigma}}_1' \mathbf{W}^{-1} \dot{\boldsymbol{\sigma}}_1)^{-1}). \tag{34}$$

Bentler and Dijkstra (1985, section 3.4) showed that one-step linearization estimator is asymptotically equivalent to ML under normality in which the optimal $\mathbf{W}$ depends on $\boldsymbol{\sigma}$. If $\boldsymbol{s}$ is replaced by $\boldsymbol{\sigma}$ during the iteration, the ML estimator results. They also noted (see Bentler & Dijkstra, 1985, equation 3.4.4) that with elliptical distributed data, the efficient estimators remain efficient, but the estimates of the asymptotic covariance matrices must be multiplied by $(1 + \hat{\kappa})$ and the $\chi^2$ statistics must be divided by $(1 + \hat{\kappa})$ Here, $\hat{\kappa}$ is a consistent estimator of $\kappa$, a measure of common kurtosis of the variables.

For LISREL type of SEM models, it is often difficult to obtain cheap initial estimators of the parameters. One approach was proposed by Jöreskog and Sörbom (1981). Bentler and Dijkstra (1985) proposed alternative approaches. With the PLSc and TSLS approaches we proposed, we can use these consistent estimators as the initial values for the one-step improvement procedure. This procedure can be implemented in the EQS or LISREL, which will also provide appropriate standard error estimates and chi-square test statistics according to the distribution assumption of the data.

For example, if we assume arbitrary distribution of the continuous observed variables, we can implement the arbitrary distribution GLS method (this is the ADF

method described earlier). The WLS fit function in equation (8) is minimized to obtain the optimal estimator $\widehat{\boldsymbol{\theta}}$. Based on $F_{WLS}(\boldsymbol{\theta})$ evaluated at $\widehat{\boldsymbol{\theta}}$, the standard errors can be obtained with the sandwich estimator from equation (17) and the ADF/AGLS test statistic can be obtained as $T_{AGLS} = (N-1)\hat{F}_{WLS}(\boldsymbol{\theta})$.

In our simulation study, we use Satorra and Bentler's (1994) robust ML estimation. The parameter estimates are the same regardless of whether the data are distributed normally or not. With non-normal data, robust ML adjusts the standard error estimates and the model evaluation test statistic. For example, instead of $c_1$ (i.e. the minimum discrepancy function chi-square statistic), one should use $c_3$ the Satorra-Bentler scaled chi-square statistic to evaluation model fit. One could also look into $c_4$ (Browne's residual-based statistic) or $T_{YB}$ when the sample size is small.

### 2.7. Optimal Generalized Least Squares (PLSe2)

The second efficient PLS estimation method (PLSe2) utilizes Browne's (1974) generalized least squares (GLS) covariance structure estimation methodology to obtain an efficient estimator and the associated parameter and model evaluation methodology. We can also use Bentler's (2006) adaptation of Jennrich's (1970) generalized least squares correlation structure methodology for the parallel development of correlation structure statistics.

It was shown earlier that Dijkstra's (2011) consistent PLS estimator PLSc yields an estimator $\widehat{\boldsymbol{\Sigma}}_c$ of the population correlation matrix $\boldsymbol{\Sigma}$. Assuming that a sample of size $N$ is drawn from a multivariate normal population and $\mathbf{S}$ is the sample covariance matrix, Browne (1974, Propositions 1-3) proved that when a consistent estimator $\mathbf{W}_2$ of the population inverse covariance matrix $\boldsymbol{\Sigma}^{-1}$ is used and the normal theory GLS fit function:

$$F_{GLS}(\boldsymbol{\theta}) = \frac{1}{2} tr\{[\mathbf{S} - \boldsymbol{\Sigma}(\boldsymbol{\theta})]\mathbf{W}_2\}^2 \tag{35}$$

is minimized with respect to $\boldsymbol{\theta}$ to yield the minimizing value $\hat{F}_{GLS}$ under the assumption of multivariate normality of variables, the test statistic $(N-1)\hat{F}_{GLS}$ is asymptotically distributed as a $\chi^2_{m^*-q}$ variate. With $m^* = m(m+1)/2$ nonredundant elements of $\mathbf{S}$ and $q$ free parameters in the model, this statistic can be used to test the validity of the hypothesized model.

Furthermore Browne (1974) showed that the estimator resulting from the minimization of $F_{GLS}(\boldsymbol{\theta})$ is consistent, asymptotically normally distributed, and asymptotically efficient. He showed that estimates of the variances of the estimator can be obtained from the diagonal of $(\dot{\boldsymbol{\sigma}}'\mathbf{W}_N^{-1}\dot{\boldsymbol{\sigma}})^{-1}$ evaluated at $\widehat{\boldsymbol{\theta}}$, where $\dot{\boldsymbol{\sigma}}$ is the Jacobian matrix, $\mathbf{W}_N = .5\mathbf{D}_m'(\boldsymbol{W}_2 \otimes \boldsymbol{W}_2)\mathbf{D}_m$, and $\mathbf{D}_m$ is the duplication matrix.

In practical implementations of Browne's theory, $\mathbf{W}_2 = \mathbf{S}^{-1}$ (for GLS) and $\mathbf{W}_2 = \widehat{\boldsymbol{\Sigma}}_{ML}^{-1}$ (for normal theory GLS) are used, where the latter is based on an iteratively

updated estimator $\widehat{\boldsymbol{\Sigma}}$ under the model that leads to the normal theory maximum likelihood (ML) estimator (Lee & Jennrich, 1979). We propose to make use of Browne's results by taking $\mathbf{W}_2 = \widehat{\boldsymbol{\Sigma}}_c^{-1}$, the consistent PLS-based weight matrix. This provides a new member for the class of normal theory generalized least squares estimators, along with the associated statistical results summarized above. When the assumption of multivariate normality is inappropriate, the test statistics and standard errors can again be corrected using the variety of robust methods available in the field.

# CHAPTER 3

MONTE CARLO SIMULATION STUDIES

## 3.1. Previous Simulation Studies

Past research has been focusing on comparing the PLS estimator with the ML-based estimators (MLE) for structural equation modeling. Vilares, Almeida and Coelho (2010) conducted simulation studies to compare the performance of PLS with MLE for structural equation models with symmetric and skewed response data. Their results indicate that globally PLS is better than MLE in terms of bias and precision, particularly with skewed response data. Regarding the issue of "consistency at large", their result showed that PLS seems to be robust even with a small number of indicators (2, 3 or 5 indicators per latent variable) or with a sample sizes as small as 250 which is typical in customer satisfaction research.

Other simulation studies have also shown the robustness of PLS estimators under various assumption violations. Hulland, Ryan, and Rayner (2010) found that PLS' accuracy does not vary much when the multivariate normality assumption is violated, whereas MLE breaks down more frequently especially when sample size is small and when there are only two measures per latent variable. Cassel, Hackl and Westlund (1999) also showed with Monte Carlo simulations that PLS was robust with

regard to skewness, multicollinearity of the indicators, and misspecification of the structural model.

Despite its robustness shown in previous simulation studies, PLS estimators still suffer from one problem. That is, it is only *consistent at large*, Dijkstra (2011) showed that traditional PLS tends to yield inconsistent estimates for the parameters of the underlying covariance structure. He proposed a consistent (PLSc) estimator to correct the biases associated with traditional PLS. The results of his simulation studies showed that PLSc estimates for the factor loadings and correlations are almost unbiased. Based on either PLS- or PLSc-generated factor correlations, Dijkstra (2011) used TSLS to estimate the structural parameters (e.g., regression coefficients). He compared estimates that are based on PLS estimated factor correlations and those that are based on PLSc-corrected factor loadings. His results showed that PLS leads to structure parameters that deviate strongly from the true values even with normally distributed data. PLSc, on the other hand, can generate parameter estimates that are consistent.

Dijkstra (2011) also used full-information maximum likelihood (FIML) estimator in further comparisons of the performance of PLSc (in conjunction with TSLS) for nonlinear and polynomial factor models. As with traditional ML, FIML generates estimates of the loadings and the structural parameters at the same time. Unlike traditional ML that use sample covariance matrix for estimation, one has to use raw data with FIML. This enables the use of direct numerical integration during the FIML

estimation process to handle nonlinear terms in latent variables. Dijkstra's simulation results showed that both FIML and PLSc (in conjunction with TSLS) are virtually unbiased with normally distributed data.

Dijkstra (2011) also studied the case of non-normal data by rescaling the indicators so that the kurtosis of the indicators were increased to 6. He found that both PLSc and FIML generate unbiased estimates of loadings. For estimates of the structural parameters, however, FIML estimates are slightly more biased than PLSc estimates from TSLS. The level of bias is not severe (raw bias < .02) even at sample size of 300 and the bias tends to decrease with less severe degree of non-normality (e.g., when kurtosis is increased by 1.7 as opposed to 6).

One of the drawbacks of traditional PLS estimation is the lack of standard error estimates and model fit statistics. Based on PLSc estimation, Dijkstra (2011) proposed the use of the bootstrap method for estimating standard errors. It appears to work well with normal data (with FIML's standard error estimates being consistently smaller). With non-normal data, however, he pointed out that the bootstrap method may require fine-tuning as in Yuan and Hayashi (2003), and that adjustments are needed for FIML's standard error estimates as well.

The criticism against traditional covariance-based SEM methodology in previous research was largely based on normal theory ML estimation. Since methods designed specifically for handling with non-normality in SEM do exist, such as robust ML and

ADF (see section 2.1.2.), these methods (conveniently implemented in standard SEM software programs such as EQS and LISREL) should perform better than the normal theory counterpart when the multivariate normality assumption is violated.

### 3.2. Simulation Study Design

We use Monte Carlo simulations to study the performance of PLSe (PLSe1 and PLSe2) estimates for structural equation models. We used the software program R to simulate normally and non-normally distributed data. We also used R to obtain PLSc and TSLS estimates. We made use of an existing SEM software program (LISREL 8.8) to obtain PLSe parameter estimates, standard errors and model fit statistics. The PLSe estimators are compared against traditional ML estimation.

For PLSe1, we use Satorra and Bentler's (1994) robust ML estimator to derive the point estimates, standard errors, and fit statistics. Robust ML retains ML parameter estimates but corrects standard errors and model fit statistics for the impact of non-normality. Previous studies (Hu, Bentler & Kano, 1992; Curran, West & Finch, 1996) have established its favorable performance under normality and non-normality.

Since PLSe2 involves normal theory GLS estimation, we did not investigate its performance under non-normality in this simulation study. If PLSe2 were to be used with non-normal data, we would expect the parameter estimates to be correct, but the standard error estimates as well as the model fit test statistics would be incorrect.

We have three research questions. The first one is about the quality of the estimators. That is, whether the PLSe estimators are consistent and asymptotically unbiased. PLSe1 will be studied under both normal data and under non-normal data, whereas PLSe2 will be studied under normal data only. The second research question is about efficiency of the estimators, which will be measured with the variance of the estimates. We would like to see whether the PLSe estimators could attain minimum variance just as the traditional ML estimator under the ideal condition of normally distributed data. Under non-normal data, we would like to see how PLSe1 (with robust standard errors and fit statistics) would perform in comparison with standard ML. The third research question is to test whether the PLSe estimators could yield fit statistics that are chi-square distributed when the distribution assumptions are met and the models fits exactly.

### 3.3. Generating Models

The data generating model resembles the one from Maruyama and McGarvey (1980). Due to requirements imposed by PLS on the estimation of the measurement model, all latent variables and observed variables are assumed to be standardized and the free loadings are estimated without constraints (i.e., there is no scaling indicator). We took their sample correlation matrix and estimated a model as shown below. The parameter estimates are taken as true values to generate normally and non-normally

distributed data sets at sample sizes of 200, 500 and 1000. Under non-normality, we also generate data sets at sample size of 2500.

Since PLS does not differentiate between endogenous latent variables and exogenous latent variables, the measurement model can be more compactly written as

$$\begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{x} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Lambda}_y & \\ & \boldsymbol{\Lambda}_x \end{pmatrix} \begin{pmatrix} \boldsymbol{\eta} \\ \boldsymbol{\xi} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\varepsilon} \\ \boldsymbol{\delta} \end{pmatrix},$$

where the factor loading matrix is 13 × 5 and the true values are given by

$$\begin{pmatrix} \boldsymbol{\Lambda}_y & \\ & \boldsymbol{\Lambda}_x \end{pmatrix} = \begin{pmatrix} \lambda_{1,1} & & & & \\ \lambda_{2,1} & & & & \\ \lambda_{3,1} & & & & \\ & \lambda_{4,2} & & & \\ & \lambda_{5,2} & & & \\ & & \lambda_{6,3} & & \\ & & \lambda_{7,3} & & \\ & & \lambda_{8,3} & & \\ & & & \lambda_{9,4} & \\ & & & \lambda_{10,4} & \\ & & & & \lambda_{11,5} \\ & & & & \lambda_{12,5} \\ & & & & \lambda_{13,5} \end{pmatrix} = \begin{pmatrix} .61 & & & & \\ .69 & & & & \\ .61 & & & & \\ & .48 & & & \\ & .36 & & & \\ & & .66 & & \\ & & .84 & & \\ & & .17 & & \\ & & & .63 & \\ & & & .51 & \\ & & & & .52 \\ & & & & .79 \\ & & & & .40 \end{pmatrix}.$$

Note that there are 2 endogenous latent variables and 3 exogenous ones. There are two structural equations in the model:

$$\eta_1 = \beta_1\eta_2 + \beta_2\xi_3 + \zeta_1,$$

$$\eta_2 = \beta_3\eta_1 + \beta_4\xi_1 + \beta_5\xi_2 + \zeta_2.$$

Written in matrix form, the structural model is:

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta},$$

where

$$\mathbf{B} = \begin{pmatrix} 0 & \beta_1 \\ \beta_3 & 0 \end{pmatrix} = \begin{pmatrix} 0 & .21 \\ .30 & 0 \end{pmatrix}, \mathbf{\Gamma} = \begin{pmatrix} 0 & 0 & \beta_2 \\ \beta_4 & \beta_5 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & .25 \\ .27 & 1.03 & 0 \end{pmatrix}.$$

This model is non-recursive.

The covariance matrix of the equation disturbance terms $V(\boldsymbol{\zeta}) = \mathbf{\Psi}$ is an identity matrix, and the covariance matrix of the exogenous factors is given by

$$V(\boldsymbol{\xi}) = \mathbf{\Phi} = \begin{pmatrix} 1 & & \\ \phi_{21} & 1 & \\ & & 1 \end{pmatrix} = \begin{pmatrix} 1 & & \\ .38 & 1 & \\ & & 1 \end{pmatrix}.$$

The three exogenous latent variables are standardized, and only $\xi_1$ and $\xi_2$ covary, with correlation equal to $\phi_{21}$. The covariance matrix of the unique factors $\boldsymbol{\epsilon}$ and $\boldsymbol{\delta}$ is a diagonal matrix $\mathbf{\Omega}$ whose diagonal elements are .51, .37, .50, .35, .63, .57, .29, .97, .60, .74, .73, .37 and .84. The generating model has 32 parameters.

While we did not alter the degree of complexity of the structural model – the current non-recursive model is complex and is expected to be challenging for most SEM estimators – we instead studied the impact of the size of the measurement model on estimation and inferential quality. The size of the measurement model presented above is considered small with only 2 to 3 indicators per latent variable. We also generated a large model where we doubled the number of indicators per latent variable. The 26 × 5 factor loading matrix for the large model is:

$$\begin{pmatrix}
\lambda_{1,1} & & & & & \\
\lambda_{2,1} & & & & & \\
\lambda_{3,1} & & & & & \\
\lambda_{4,1} & & & & & \\
\lambda_{5,1} & & & & & \\
\lambda_{6,1} & & & & & \\
 & \lambda_{7,2} & & & & \\
 & \lambda_{8,2} & & & & \\
 & \lambda_{9,2} & & & & \\
 & \lambda_{10,2} & & & & \\
 & & \lambda_{11,3} & & & \\
 & & \lambda_{12,3} & & & \\
 & & \lambda_{13,3} & & & \\
 & & \lambda_{14,3} & & & \\
 & & \lambda_{15,3} & & & \\
 & & \lambda_{16,3} & & & \\
 & & & \lambda_{17,4} & & \\
 & & & \lambda_{18,4} & & \\
 & & & \lambda_{19,4} & & \\
 & & & \lambda_{20,4} & & \\
 & & & & \lambda_{21,5} & \\
 & & & & \lambda_{22,5} & \\
 & & & & \lambda_{23,5} & \\
 & & & & \lambda_{24,5} & \\
 & & & & \lambda_{25,5} & \\
 & & & & \lambda_{26,5}
\end{pmatrix} = \begin{pmatrix}
.61 & & & & \\
.69 & & & & \\
.61 & & & & \\
.61 & & & & \\
.69 & & & & \\
.61 & & & & \\
 & .48 & & & \\
 & .36 & & & \\
 & .48 & & & \\
 & .36 & & & \\
 & & .66 & & \\
 & & .84 & & \\
 & & .17 & & \\
 & & .66 & & \\
 & & .84 & & \\
 & & .17 & & \\
 & & & .63 & \\
 & & & .51 & \\
 & & & .63 & \\
 & & & .51 & \\
 & & & & .52 \\
 & & & & .79 \\
 & & & & .40 \\
 & & & & .52 \\
 & & & & .79 \\
 & & & & .40
\end{pmatrix},$$

The covariance matrix of the unique factors $\boldsymbol{\epsilon}$ and $\boldsymbol{\delta}$ is a diagonal matrix $\boldsymbol{\Omega}$ whose diagonal elements are .51, .37, .50, .51, .37, .50, .36, .63, .36, .63, .57, .29, .97, .57, .29, .97, .60, .74, .60, .74, .73, .37, .84, .73, .37 and .84. The large model has 58 parameters.

### 3.4. Data Generation

We compute the model implied covariance matrix using the true parameters and model specification presented in Section 3.3 as $\boldsymbol{\Sigma} = \boldsymbol{\Lambda\Upsilon\Lambda'} + \boldsymbol{\Omega}$**.** For the normal conditions, we generated multivariate normally distributed data using R's mvtnorm package. The "rmvnorm" function in mvtnorm uses the Cholesky decomposition method to convert a matrix of independent and identically distributed standard normal variates into a sample with the desired population covariance structure**.** For a small model, 13 observed variables are generated and for a large model, 26 observed variables are generated.

To generate non-normal data, we follow a procedure outlined in Hu, Bentler and Kano (1992). Among the non-normal conditions studied in Hu et al. (1992), we adopted a similar set up as condition 7. In this condition both common and unique factors are non-normally distributed, and in addition, the factors are not independently distributed, thereby (intentionally) violating the conditions for asymptotic robustness of ML estimation.

We first generate non-normally distributed common factors and unique factors separately in R using the classical Vale and Maurelli (1983) method. Vale and Maurelli's (1983) method improves upon Fleishman's (1978) method and it can produce multivariate random numbers with pre-specified intercorrlations and univariate means, variances, skewness, and kurtosis. The 2 equation disturbance terms in $\boldsymbol{\zeta}$ and the 3 exogenous latent variables $\boldsymbol{\xi}$ are generated with means of 0 and unit variances. Their

true skewness coefficients are 1.00, 0.75, 0.75, 0, and -0.25, respectively. The true (excess) kurtosis coefficients are 3.75, 3.75, 2.00, 3.75 and 1.00, respectively. These random numbers are then inserted into the structural equations to generate the non-normally distributed endogenous variables in $\boldsymbol{\eta}$. The unique factors are independently generated. They have means of 0, variances of 1 and skewness values of 0. The true kurtosis values for the 13 unique factors in the small model are -1.00, -0.75, 2.75, 0.25, 1.00, 1.75, 2.00, 2.25, -0.50, 3.00, 5.00, 3.25, and 3.50, respectively. For the large model, the true kurtosis values for the 26 unique factors are -1.00, -0.75, 2.75, -1.00, -0.75, 2.75, 0.25, 1.00, 0.25, 1.00, 1.75, 2.00, 2.25, 1.75, 2.00, 2.25, -0.50, 3.00, -0.50, 3.00, 5.00, 3.25, 3.50, 5.00, 3.25, and 3.50, respectively.

The second step is to make the common factors and unique factors dependent by dividing them with a common random variable $Z$ so that the asymptotic robustness of normal theory statistics is not to be expected under the non-normal condition. The common random variable $Z = \sqrt{\chi^2_{(5)}/3}$ is generated independent of both the common and unique factors. The division by $\sqrt{3}$ is made so that $E(Z^{-2}) = 1$, i.e., the variances and covariances of the factors remains unchanged by the division (Kano, 1990), and only the kurtosis is modified.

To verify that the data are generated as expected, the averages of sample means, sample variances, sample skewness, and sample kurtosis coefficients are obtained for

the observed and latent variables across Monte Carlo replications. The sampling distributions all correspond well with the true moments in the data generating model. In particular, the observed variables have a wide range of skewness and kurtosis values. The empirical skewness of the observed variables lies between the range of .74 and -.21. The kurtosis is between the range of 6.38 and 13.03, with 6 of the 13 variables having high kurtosis that exceeds 10. This latter observation is important since excess kurtosis is expected to severely impact the quality of normal theory standard error estimates and test statistics.

For normal data, we will compare the performance of four estimators (PLSc with TSLS, PLSe1, PLSe2 and standard ML) at 3 sample sizes (200, 500 and 1000) and 2 model size conditions (a small model with 13 observed variables and a large model with 26 observed variables). Thus there are a total of 6 conditions with normal data.

For non-normal data, we will compare the performance of three estimators (PLSc with TSLS, PLSe1 and standard ML) at 4 sample sizes (200, 500, 1000 and 2500) and 2 model size conditions, leading to a total of 8 conditions.

We ran 300 replication for each condition, except for the non-normal condition with small model at N=200 where we encountered substantial convergence problems wherein we ran 600 replications instead. For each replication, we fitted the same model as the data generating model in section 3.3 (i.e., no model misspecification) with different estimators. If any one of the estimators should encounter any convergence or

numerical problem, we discard the results of that replication, even if the other estimators produced adequate estimates. This ensures the comparison across different estimators is based on the same set of replications and that the comparison would be fair under that specific condition.

### 3.5. Methods for Summarizing Results

For the evaluation of the quality of the estimators, we record the Monte Carlo average of the point estimates for each of the estimators so that they may be compared with the true value.

To evaluate the standard errors, the mean of the estimated standard errors for that parameter across the Monte Carlo replications should be compared with the Monte Carlo standard deviation of a given parameter estimate. Let

$$SE(\hat{\theta}) = M^{-1} \sum_{i=1}^{M} SE(\hat{\theta}_i)$$

represent the mean of the estimated standard errors, where $SE(\hat{\theta}_i)$ is the estimated standard error from replication $i$, and $M$ is the number of Monte Carlo replications. Let

$$SD(\hat{\theta}) = \left( \frac{1}{M-1} \sum_{i=1}^{M} (\hat{\theta}_i - \bar{\theta}) \right)^{1/2}$$

be the Monte Carlo standard deviation of the point estimates, where $\bar{\theta}$ is the mean of point estimates.

We will also calculate Root Mean Square Error (RMSE), which can provide information of both the deviation of each parameter estimate from the true value and the variability of such distances:

$$RMSE = \left[\frac{1}{M}\sum_{i=1}^{M}(\hat{\theta}_i - \theta)^2\right]^{\frac{1}{2}}.$$

In addition to the accuracy of parameter estimates, we will also record chi-square fit statistics and model convergence rate. Factors such as sample size and number of indicators per latent variable could contribute to the performance of these estimators. The model fit chi-square statistics we cover are the ones we discussed in section 2.1.2, including: $c_1$ minimum fit function chi-square statistic, $c_3$ Satorra-Bentler scaled chi-square statistic, $c_4$ Browne's residual-based statistic, and lastly $T_{YB}$ which is Yuan and Bentler's (1997) adjustment to $c_4$.

# CHAPTER 4

RESULTS

## 4.1. Convergence and Proper Solutions

A theoretically good estimator has little empirical value if it runs into convergence problems too frequently. Non-convergence may happen with PLSc, ML or GLS given the number of iterations. The converged results, on the other hand, may still include improper parameter estimates, such as Heywood cases where estimates of unique factor variances are zero or negative. In addition, other parameter estimates could be out of their legitimate bounds. For example, the estimated equation disturbances' variances are negative or correlations not between -1 and 1. When these improper solutions occur in any replication with any of the estimators, we count the replication as an invalid case for that estimator. Table 1 compares the number of invalid cases for each estimator. The total invalid case is the union of all the invalid cases. That is, if any estimator had an invalid case in any replication, that specific replication is not going to be counted for the total number of valid cases that we later use for summaries of point estimates, standard errors, or chi-square statistics.

We observed no convergence issue with PLS and subsequently PLSc. However, we did encounter a number of invalid cases at the stage of TSLS when we estimate the structural parameters at the smallest sample size ($N$=200).

For a small model with normal data, TSLS produced improper estimates in 78 out of 300 replications. In contrast, ML, PLSe1 and PLSe2 have roughly half the number of invalid cases. At $N$=500, the performance of TSLS is about the same with PLSe1 and PLSe2, with ML showing the fewest invalid cases. At $N$=1000, all four estimators are performing well and TSLS produced valid solutions for all 300 replications.

For large model with normal data, all four estimators seem to benefit from having more indicators. At $N$=200, we observe much fewer invalid cases with TSLS than the corresponding condition for the small model. ML and PLSe1 seem to perform the best, with only 1 or 2 invalid cases. PLSe1 uses starting values obtained from PLSc and TSLS. We suspect that even with inferior starting values, PLSe1 with 1-step improvement should be able to recover from the poor initial estimates. PLSe2 seem to have more difficulty when compared with PLSe1 or ML. Such difference is obvious at $N$=200 with large model where PLSe2 generated 11 invalid cases, whereas ML and PLSe1 only resulted in 1 or 2 invalid ones. We suspect that with a large model, the size of the covariance matrix of the observed variables ($26 \times 26$) is significantly larger than that of the small model ($13 \times 13$). At the smallest sample size, the inherent instability in the covariance matrix (with 351 unique elements relative to sample size) might lead to a weight matrix that is less stable for PLSe2 estimation. However, when samples size increases to 1000, all four estimators produced valid cases in all replications.

Table 1: Comparison of convergence problems between different approaches

*Normal data*

| | Small model | | | Large model | | |
|---|---|---|---|---|---|---|
| | *N*=200 | *N* =500 | *N*=1000 | *N*=200 | *N* =500 | *N*=1000 |
| TSLS | 78 | 14 | 0 | 24 | 2 | 0 |
| ML | 35 | 6 | 1 | 1 | 0 | 0 |
| PLSe1 | 21 | 13 | 3 | 2 | 0 | 0 |
| PLSe2 | 49 | 11 | 2 | 11 | 0 | 0 |
| Total #invalid | 128 | 32 | 5 | 37 | 2 | 0 |
| Total #valid out of 300 | 172 | 268 | 295 | 263 | 298 | 300 |

*Non-normal data*

| | Small model | | | | Large model | | | |
|---|---|---|---|---|---|---|---|---|
| | *N*=200 | *N* =500 | *N*=1000 | *N*=2500 | *N*=200 | *N* =500 | *N*=1000 | *N*=2500 |
| TSLS | 251* | 58 | 19 | 3 | 61 | 15 | 5 | 0 |
| ML | 241* | 42 | 17 | 1 | 34 | 11 | 2 | 1 |
| PLSe1 | 52* | 15 | 12 | 0 | 1 | 0 | 0 | 1 |
| Total #invalid | 387* | 92 | 44 | 4 | 85 | 24 | 5 | 2 |
| Total #valid out of 300 | 213* | 208 | 256 | 296 | 215 | 276 | 295 | 298 |

*Note:* * At *N*=200 for non-normal data with small model, the numbers are based on 600 replications.

With non-normal data, we saw the same pattern as in normal data. We added two more conditions at *N*=2500, because we expected non-normality will pose more challenges to the estimation process. Note that the condition of small model at *N*=200 seems to be the one that poses the most severe challenges for TSLS and ML, where more than 40% of the replications are invalid. PLSe1 seems to turn out better due to the 1-step nature, with less than 10% invalid cases. PLSe2 is not studied for the non-normal scenarios. Similar to the normal data conditions, TSLS, ML and PLSe1 all benefit from a larger model with more indicators. Unlike normal data, however, TSLS and ML still have a few invalid cases at *N*=1000.

For both normal and non-normal data, PLSe1 seems to encounter the least difficulty at the smallest sample size in terms of the frequency of invalid solutions. However, we should examine the quality of the parameter estimates and model fit statistics to determine whether it truly outperforms PLSc and TSLS sufficiently to justify the complexities involved in the one-step improvement. At larger sample sizes, all estimators perform increasingly well in terms of the ability to produce valid solutions. When there are fewer convergence problems, we should be able to tell whether PLSe2 can improve PLSc estimates and how it compares to ML.

### 4.2. Parameter Recovery and Standard Errors

We will first show results with normal data (Tables 2 and 3 for small model; Tables 4 and 5 for large model) and then non-normal data (Tables 6 and 7 for small model; Tables 8 and 9 for large model). We focus our discussion on RMSE, which can provide information on bias and variability.

#### 4.2.1. Results under Normality

Table 2 compares the parameter estimates of factor loadings using PLSc, ML, PLSe1 and PLSe2 with normal data and the small model. At $N$=200, ML has the smallest bias and the lowest RMSE. When we compare RMSE across different estimators, RMSE for PLSc are higher than ML (almost twice as high for some loadings). PLSe1 and PLSe2

seem to perform as well as ML in terms of RMSE, although PLSe1 shows slightly higher bias than ML and PLSe2.

When estimating loadings at a small sample size of $N$=200, the two PLSe estimators seem to improve PLSc estimated loadings. The quality of estimates is comparable to ML. When sample size increases to 500 and 1000, all four estimators show improvements in parameter estimates and standard error estimates, where ML, PLSe1 and PLSe2 are almost identical at $N$=1000. PLSc improved drastically at larger $N$, and the RMSE values are only slightly worse than ML. Thus at a large sample size, PLSc seems to perform well with small model and normal data. PLSe estimators do improve upon PLSc to the extent that the PLSe1 and PLSe2 estimates are as good as ML. In addition, PLSe produces standard error estimates that are as good as ML.

Table 2: Comparison of results on loadings using PLSc, ML, PLSe1 and PLSe2 for normal data with small model

| *PLSc* | | *N*=200 | | | | *N* =500 | | | | *N*=1000 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\theta$ | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE |
| $\lambda_{1,1}$ | .61 | .60 | N/A | .11 | .11 | .61 | N/A | .06 | .06 | .61 | N/A | .05 | .05 |
| $\lambda_{2,1}$ | .69 | .67 | N/A | .10 | .10 | .68 | N/A | .06 | .06 | .69 | N/A | .05 | .05 |
| $\lambda_{3,1}$ | .61 | .60 | N/A | .10 | .10 | .61 | N/A | .07 | .07 | .61 | N/A | .05 | .05 |
| $\lambda_{4,2}$ | .48 | .56 | N/A | .12 | .15 | .51 | N/A | .11 | .11 | .50 | N/A | .07 | .07 |
| $\lambda_{5,2}$ | .36 | .42 | N/A | .10 | .12 | .39 | N/A | .08 | .08 | .38 | N/A | .05 | .05 |
| $\lambda_{6,3}$ | .66 | .66 | N/A | .11 | .11 | .65 | N/A | .07 | .07 | .65 | N/A | .05 | .05 |
| $\lambda_{7,3}$ | .84 | .79 | N/A | .10 | .12 | .82 | N/A | .08 | .08 | .83 | N/A | .06 | .06 |
| $\lambda_{8,3}$ | .17 | .20 | N/A | .13 | .14 | .19 | N/A | .11 | .11 | .18 | N/A | .08 | .09 |
| $\lambda_{9,4}$ | .63 | .64 | N/A | .10 | .10 | .63 | N/A | .06 | .06 | .63 | N/A | .05 | .05 |
| $\lambda_{10,4}$ | .51 | .52 | N/A | .10 | .10 | .51 | N/A | .06 | .06 | .50 | N/A | .04 | .04 |
| $\lambda_{11,5}$ | .52 | .53 | N/A | .20 | .20 | .51 | N/A | .19 | .19 | .51 | N/A | .16 | .16 |
| $\lambda_{12,5}$ | .79 | .63 | N/A | .21 | .26 | .69 | N/A | .18 | .21 | .72 | N/A | .16 | .17 |
| $\lambda_{13,5}$ | .40 | .46 | N/A | .21 | .22 | .43 | N/A | .20 | .20 | .44 | N/A | .15 | .16 |
| *ML* | | *N*=200 | | | | *N* =500 | | | | *N*=1000 | | | |
| | $\theta$ | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE |
| $\lambda_{1,1}$ | .61 | .61 | .07 | .07 | .07 | .61 | .04 | .04 | .04 | .61 | .03 | .03 | .03 |
| $\lambda_{2,1}$ | .69 | .67 | .07 | .07 | .07 | .69 | .04 | .04 | .04 | .69 | .03 | .03 | .03 |
| $\lambda_{3,1}$ | .61 | .61 | .07 | .07 | .07 | .61 | .04 | .04 | .04 | .61 | .03 | .03 | .03 |
| $\lambda_{4,2}$ | .48 | .49 | .13 | .10 | .10 | .46 | .08 | .08 | .08 | .47 | .05 | .05 | .05 |
| $\lambda_{5,2}$ | .36 | .37 | .09 | .08 | .08 | .35 | .05 | .05 | .06 | .36 | .04 | .04 | .04 |
| $\lambda_{6,3}$ | .66 | .67 | .09 | .09 | .10 | .66 | .06 | .06 | .06 | .66 | .04 | .04 | .04 |
| $\lambda_{7,3}$ | .84 | .83 | .11 | .09 | .09 | .84 | .07 | .07 | .07 | .84 | .05 | .05 | .05 |
| $\lambda_{8,3}$ | .17 | .18 | .08 | .07 | .07 | .17 | .05 | .05 | .05 | .17 | .04 | .04 | .04 |
| $\lambda_{9,4}$ | .63 | .64 | .10 | .10 | .10 | .63 | .06 | .06 | .06 | .63 | .04 | .05 | .05 |
| $\lambda_{10,4}$ | .51 | .52 | .09 | .09 | .10 | .50 | .06 | .05 | .05 | .50 | .04 | .04 | .04 |
| $\lambda_{11,5}$ | .52 | .54 | .10 | .09 | .09 | .53 | .06 | .06 | .06 | .52 | .04 | .05 | .05 |
| $\lambda_{12,5}$ | .79 | .79 | .12 | .11 | .11 | .79 | .08 | .07 | .07 | .80 | .06 | .06 | .06 |
| $\lambda_{13,5}$ | .40 | .41 | .09 | .08 | .09 | .40 | .06 | .06 | .06 | .40 | .04 | .04 | .04 |

(continued on the next page)

Table 2 (cont.): Comparison of results on loadings using PLSc, ML, PLSe1 and PLSe2 for normal data with small model

| *PLSe1* | | *N*=200 | | | | *N* =500 | | | | *N*=1000 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\theta$ | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE |
| $\lambda_{1,1}$ | .61 | .60 | .07 | .08 | .08 | .61 | .04 | .05 | .05 | .61 | .03 | .03 | .03 |
| $\lambda_{2,1}$ | .69 | .66 | .07 | .07 | .08 | .68 | .04 | .05 | .05 | .68 | .03 | .03 | .03 |
| $\lambda_{3,1}$ | .61 | .60 | .07 | .07 | .07 | .61 | .04 | .04 | .04 | .61 | .03 | .03 | .03 |
| $\lambda_{4,2}$ | .48 | .54 | .12 | .12 | .13 | .51 | .07 | .09 | .10 | .50 | .05 | .07 | .07 |
| $\lambda_{5,2}$ | .36 | .41 | .08 | .09 | .10 | .38 | .05 | .07 | .07 | .38 | .04 | .05 | .05 |
| $\lambda_{6,3}$ | .66 | .67 | .09 | .09 | .09 | .66 | .06 | .06 | .06 | .66 | .04 | .04 | .04 |
| $\lambda_{7,3}$ | .84 | .82 | .10 | .09 | .10 | .83 | .07 | .07 | .07 | .84 | .05 | .05 | .05 |
| $\lambda_{8,3}$ | .17 | .18 | .08 | .07 | .07 | .17 | .05 | .05 | .05 | .17 | .04 | .04 | .04 |
| $\lambda_{9,4}$ | .63 | .63 | .10 | .10 | .10 | .62 | .06 | .06 | .06 | .63 | .04 | .05 | .05 |
| $\lambda_{10,4}$ | .51 | .52 | .09 | .09 | .09 | .50 | .06 | .05 | .05 | .50 | .04 | .04 | .04 |
| $\lambda_{11,5}$ | .52 | .52 | .10 | .11 | .11 | .51 | .06 | .09 | .09 | .51 | .04 | .06 | .06 |
| $\lambda_{12,5}$ | .79 | .73 | .12 | .12 | .14 | .75 | .08 | .10 | .11 | .78 | .06 | .08 | .08 |
| $\lambda_{13,5}$ | .40 | .40 | .09 | .10 | .10 | .39 | .06 | .07 | .07 | .39 | .04 | .05 | .05 |
| *PLSe2* | | *N*=200 | | | | *N* =500 | | | | *N*=1000 | | | |
| | $\theta$ | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE |
| $\lambda_{1,1}$ | .61 | .60 | .07 | .07 | .07 | .61 | .04 | .04 | .04 | .61 | .03 | .03 | .03 |
| $\lambda_{2,1}$ | .69 | .67 | .07 | .07 | .08 | .69 | .04 | .05 | .05 | .68 | .03 | .03 | .03 |
| $\lambda_{3,1}$ | .61 | .60 | .07 | .07 | .07 | .61 | .04 | .04 | .04 | .61 | .03 | .03 | .03 |
| $\lambda_{4,2}$ | .48 | .49 | .13 | .10 | .10 | .46 | .08 | .08 | .08 | .47 | .05 | .05 | .05 |
| $\lambda_{5,2}$ | .36 | .36 | .09 | .08 | .08 | .35 | .06 | .05 | .06 | .36 | .04 | .04 | .04 |
| $\lambda_{6,3}$ | .66 | .65 | .10 | .10 | .10 | .66 | .06 | .06 | .06 | .66 | .04 | .04 | .04 |
| $\lambda_{7,3}$ | .84 | .82 | .11 | .10 | .10 | .84 | .07 | .07 | .07 | .84 | .05 | .05 | .05 |
| $\lambda_{8,3}$ | .17 | .17 | .08 | .07 | .08 | .17 | .05 | .05 | .05 | .16 | .04 | .04 | .04 |
| $\lambda_{9,4}$ | .63 | .62 | .10 | .11 | .11 | .62 | .06 | .06 | .06 | .63 | .04 | .05 | .05 |
| $\lambda_{10,4}$ | .51 | .51 | .09 | .10 | .10 | .50 | .06 | .05 | .06 | .50 | .04 | .04 | .04 |
| $\lambda_{11,5}$ | .52 | .52 | .10 | .09 | .09 | .52 | .06 | .06 | .06 | .52 | .04 | .05 | .05 |
| $\lambda_{12,5}$ | .79 | .78 | .14 | .12 | .12 | .78 | .08 | .08 | .08 | .80 | .06 | .06 | .06 |
| $\lambda_{13,5}$ | .40 | .40 | .09 | .08 | .08 | .40 | .06 | .06 | .06 | .40 | .04 | .04 | .04 |

Table 3 presents results on structural regression coefficients and the factor correlation. Note that TSLS is listed instead of PLSc, because these estimates are based on TSLS applied to PLSc generated factor correlations. We noticed that the estimates of

the regression coefficients are generally more biased than the estimates of the loadings or the factor correlation. This happens to all four estimators. We suspect this is probably due to the complexity of the non-recursive model.

At $N$=200, RMSE of the loading estimates are generally less than 0.10 whereas RMSE of the regression coefficients estimates are all larger than 0.10. For example, we observe a more pronounced bias in $\beta_3$ at $N$=200, where RMSEs for ML and the two PLSe estimators are around 0.22. For this specific parameter at $N$=200, TSLS seems to produce an estimate with RMSE that is twice as large as the other estimators. However, when sample size increases to 1000, we no longer observe such a pronounced difference, and all estimators are performing well in terms of point estimates and standard error estimates. ML and PLSe2 produce almost identical results with smaller RMSE compared to TSLS and PLSe1. PLSe1 still improves TSLS-generated regression coefficients, but to a lesser extent than PLSe2.

Table 3. Comparison of results on structural regression coefficients and the factor correlation using TSLS, ML, PLSe1 and PLSe2 for normal data with small model

| *TSLS* | | *N*=200 | | | | *N* =500 | | | | *N*=1000 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\theta$ | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE |
| $\beta_1$ | .21 | .27 | N/A | .13 | .15 | .23 | N/A | .09 | .09 | .22 | N/A | .06 | .06 |
| $\beta_2$ | .25 | .29 | N/A | .11 | .11 | .26 | N/A | .07 | .07 | .25 | N/A | .05 | .05 |
| $\beta_3$ | .30 | .28 | N/A | .44 | .44 | .31 | N/A | .39 | .39 | .32 | N/A | .29 | .29 |
| $\beta_4$ | .27 | .26 | N/A | .19 | .19 | .25 | N/A | .13 | .13 | .27 | N/A | .09 | .09 |
| $\beta_5$ | 1.03 | .89 | N/A | .30 | .33 | 1.01 | N/A | .26 | .26 | 1.01 | N/A | .18 | .18 |
| $\phi_{2,1}$ | .38 | .39 | N/A | .11 | .11 | .39 | N/A | .07 | .07 | .38 | N/A | .05 | .05 |
| *ML* | | *N*=200 | | | | *N* =500 | | | | *N*=1000 | | | |
| | $\theta$ | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE |
| $\beta_1$ | .21 | .23 | .12 | .12 | .12 | .20 | .07 | .08 | .08 | .21 | .05 | .05 | .05 |
| $\beta_2$ | .25 | .29 | .11 | .10 | .10 | .25 | .07 | .07 | .07 | .25 | .05 | .05 | .05 |
| $\beta_3$ | .30 | .28 | .24 | .21 | .21 | .32 | .16 | .17 | .17 | .31 | .10 | .11 | .11 |
| $\beta_4$ | .27 | .26 | .22 | .21 | .21 | .27 | .12 | .12 | .12 | .28 | .08 | .09 | .09 |
| $\beta_5$ | 1.03 | 1.04 | .47 | .34 | .34 | 1.12 | .30 | .28 | .30 | 1.06 | .18 | .18 | .18 |
| $\phi_{2,1}$ | .38 | .39 | .11 | .11 | .11 | .38 | .07 | .07 | .07 | .38 | .05 | .05 | .05 |
| *PLSe1* | | *N*=200 | | | | *N* =500 | | | | *N*=1000 | | | |
| | $\theta$ | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE |
| $\beta_1$ | .21 | .28 | .13 | .13 | .15 | .23 | .07 | .09 | .09 | .22 | .05 | .06 | .06 |
| $\beta_2$ | .25 | .33 | .12 | .11 | .14 | .28 | .07 | .07 | .08 | .27 | .05 | .05 | .05 |
| $\beta_3$ | .30 | .24 | .22 | .23 | .24 | .28 | .14 | .18 | .18 | .29 | .10 | .12 | .12 |
| $\beta_4$ | .27 | .24 | .20 | .19 | .19 | .24 | .11 | .12 | .12 | .26 | .08 | .09 | .09 |
| $\beta_5$ | 1.03 | .94 | .37 | .31 | .32 | 1.03 | .25 | .26 | .26 | 1.01 | .17 | .18 | .18 |
| $\phi_{2,1}$ | .38 | .40 | .11 | .11 | .11 | .39 | .07 | .07 | .07 | .38 | .05 | .05 | .05 |
| *PLSe2* | | *N*=200 | | | | *N* =500 | | | | *N*=1000 | | | |
| | $\theta$ | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE |
| $\beta_1$ | .21 | .22 | .13 | .12 | .12 | .20 | .07 | .08 | .08 | .21 | .05 | .05 | .05 |
| $\beta_2$ | .25 | .29 | .11 | .10 | .11 | .25 | .07 | .07 | .07 | .25 | .05 | .05 | .05 |
| $\beta_3$ | .30 | .29 | .25 | .22 | .22 | .33 | .16 | .17 | .17 | .31 | .10 | .11 | .11 |
| $\beta_4$ | .27 | .24 | .23 | .22 | .22 | .26 | .13 | .12 | .12 | .27 | .08 | .09 | .09 |
| $\beta_5$ | 1.03 | 1.03 | .50 | .35 | .35 | 1.13 | .31 | .29 | .30 | 1.06 | .18 | .18 | .18 |
| $\phi_{2,1}$ | .38 | .39 | .11 | .12 | .12 | .38 | .07 | .07 | .07 | .38 | .05 | .05 | .05 |

Table 4 presents the estimates of factor loadings from PLSc, ML, PLSe1 and PLSe2 when the model size is large. Similar to the case of small model when we compare RMSE across different estimators, RMSE for PLSc estimated loadings are again almost twice as high as the ML estimator at $N$=200. PLSe1 and PLSe2 seem to perform as well as ML at each sample size in terms of RMSE, although PLSe1 again shows slightly higher bias than ML and PLSe2. The two PLSe estimators seem to improve PLSc estimated loadings.

Table 4: Comparison of results on loadings using PLSc, ML, PLSe1 and PLSe2 for normal data with large model

| *PLSc* | | | *N*=200 | | | | *N* =500 | | | | *N*=1000 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\theta$ | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE |
| $\lambda_{1,1}$ | .61 | .59 | N/A | .10 | .11 | .60 | N/A | .06 | .07 | .61 | N/A | .05 | .05 |
| $\lambda_{2,1}$ | .69 | .66 | N/A | .09 | .10 | .68 | N/A | .06 | .06 | .68 | N/A | .04 | .04 |
| $\lambda_{3,1}$ | .61 | .59 | N/A | .11 | .11 | .60 | N/A | .07 | .07 | .61 | N/A | .05 | .05 |
| $\lambda_{4,1}$ | .61 | .59 | N/A | .11 | .11 | .60 | N/A | .07 | .07 | .61 | N/A | .05 | .05 |
| $\lambda_{5,1}$ | .69 | .66 | N/A | .10 | .10 | .68 | N/A | .06 | .06 | .68 | N/A | .05 | .05 |
| $\lambda_{6,1}$ | .61 | .59 | N/A | .10 | .11 | .61 | N/A | .07 | .07 | .61 | N/A | .05 | .05 |
| $\lambda_{7,2}$ | .48 | .51 | N/A | .11 | .11 | .50 | N/A | .07 | .07 | .48 | N/A | .04 | .04 |
| $\lambda_{8,2}$ | .36 | .39 | N/A | .08 | .09 | .38 | N/A | .06 | .06 | .37 | N/A | .03 | .03 |
| $\lambda_{9,2}$ | .48 | .51 | N/A | .11 | .11 | .50 | N/A | .07 | .08 | .48 | N/A | .03 | .04 |
| $\lambda_{10,2}$ | .36 | .39 | N/A | .08 | .09 | .38 | N/A | .06 | .06 | .37 | N/A | .03 | .03 |
| $\lambda_{11,3}$ | .66 | .62 | N/A | .14 | .15 | .65 | N/A | .09 | .09 | .65 | N/A | .06 | .06 |
| $\lambda_{12,3}$ | .84 | .80 | N/A | .11 | .11 | .82 | N/A | .06 | .07 | .84 | N/A | .05 | .05 |
| $\lambda_{13,3}$ | .17 | .23 | N/A | .15 | .16 | .18 | N/A | .11 | .11 | .16 | N/A | .09 | .09 |
| $\lambda_{14,3}$ | .66 | .64 | N/A | .12 | .13 | .64 | N/A | .09 | .09 | .65 | N/A | .06 | .06 |
| $\lambda_{15,3}$ | .84 | .79 | N/A | .10 | .11 | .82 | N/A | .06 | .07 | .83 | N/A | .05 | .05 |
| $\lambda_{16,3}$ | .17 | .22 | N/A | .14 | .15 | .18 | N/A | .11 | .11 | .17 | N/A | .08 | .08 |
| $\lambda_{17,4}$ | .63 | .62 | N/A | .09 | .09 | .62 | N/A | .06 | .06 | .63 | N/A | .04 | .04 |
| $\lambda_{18,4}$ | .51 | .51 | N/A | .09 | .09 | .51 | N/A | .07 | .07 | .50 | N/A | .04 | .04 |
| $\lambda_{19,4}$ | .63 | .63 | N/A | .09 | .09 | .62 | N/A | .06 | .06 | .63 | N/A | .04 | .04 |
| $\lambda_{20,4}$ | .51 | .49 | N/A | .10 | .10 | .51 | N/A | .07 | .07 | .51 | N/A | .05 | .05 |
| $\lambda_{21,5}$ | .52 | .45 | N/A | .21 | .22 | .50 | N/A | .21 | .21 | .49 | N/A | .17 | .18 |
| $\lambda_{22,5}$ | .79 | .65 | N/A | .22 | .26 | .68 | N/A | .18 | .21 | .74 | N/A | .13 | .14 |
| $\lambda_{23,5}$ | .40 | .38 | N/A | .23 | .23 | .37 | N/A | .19 | .19 | .37 | N/A | .18 | .18 |
| $\lambda_{24,5}$ | .52 | .47 | N/A | .24 | .24 | .47 | N/A | .20 | .21 | .49 | N/A | .16 | .17 |
| $\lambda_{25,5}$ | .79 | .65 | N/A | .21 | .25 | .71 | N/A | .17 | .19 | .73 | N/A | .14 | .15 |
| $\lambda_{26,5}$ | .40 | .39 | N/A | .23 | .23 | .39 | N/A | .20 | .21 | .39 | N/A | .18 | .18 |

(continues on the next page)

Table 4 (cont.): Comparison of results on loadings using PLSc, ML, PLSe1 and PLSe2 for normal data with large model

| *ML* | | *N*=200 | | | | *N* =500 | | | | *N*=1000 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\theta$ | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE |
| $\lambda_{1,1}$ | .61 | .60 | .06 | .06 | .06 | .61 | .04 | .04 | .04 | .61 | .03 | .02 | .02 |
| $\lambda_{2,1}$ | .69 | .68 | .06 | .06 | .06 | .69 | .04 | .03 | .03 | .69 | .03 | .03 | .03 |
| $\lambda_{3,1}$ | .61 | .61 | .06 | .06 | .06 | .61 | .04 | .04 | .04 | .61 | .03 | .03 | .03 |
| $\lambda_{4,1}$ | .61 | .60 | .06 | .06 | .06 | .61 | .04 | .04 | .04 | .61 | .03 | .03 | .03 |
| $\lambda_{5,1}$ | .69 | .68 | .06 | .06 | .06 | .69 | .04 | .04 | .04 | .69 | .03 | .03 | .03 |
| $\lambda_{6,1}$ | .61 | .61 | .06 | .06 | .06 | .61 | .04 | .04 | .04 | .61 | .03 | .03 | .03 |
| $\lambda_{7,2}$ | .48 | .47 | .06 | .06 | .06 | .47 | .04 | .04 | .04 | .48 | .03 | .03 | .03 |
| $\lambda_{8,2}$ | .36 | .36 | .06 | .05 | .05 | .36 | .03 | .03 | .04 | .36 | .02 | .03 | .03 |
| $\lambda_{9,2}$ | .48 | .47 | .06 | .06 | .06 | .48 | .04 | .04 | .04 | .48 | .03 | .03 | .03 |
| $\lambda_{10,2}$ | .36 | .36 | .06 | .06 | .06 | .36 | .04 | .04 | .04 | .36 | .02 | .03 | .03 |
| $\lambda_{11,3}$ | .66 | .66 | .07 | .06 | .06 | .66 | .04 | .04 | .04 | .66 | .03 | .03 | .03 |
| $\lambda_{12,3}$ | .84 | .84 | .06 | .06 | .06 | .84 | .04 | .04 | .04 | .84 | .03 | .03 | .03 |
| $\lambda_{13,3}$ | .17 | .17 | .07 | .07 | .07 | .17 | .05 | .04 | .04 | .17 | .03 | .03 | .03 |
| $\lambda_{14,3}$ | .66 | .66 | .07 | .07 | .07 | .66 | .04 | .05 | .05 | .66 | .03 | .03 | .03 |
| $\lambda_{15,3}$ | .84 | .85 | .06 | .06 | .06 | .84 | .04 | .04 | .04 | .84 | .03 | .03 | .03 |
| $\lambda_{16,3}$ | .17 | .17 | .07 | .07 | .07 | .17 | .05 | .04 | .04 | .17 | .03 | .03 | .03 |
| $\lambda_{17,4}$ | .63 | .63 | .08 | .07 | .07 | .63 | .05 | .05 | .05 | .63 | .03 | .03 | .03 |
| $\lambda_{18,4}$ | .51 | .50 | .08 | .08 | .08 | .51 | .05 | .05 | .05 | .50 | .03 | .04 | .04 |
| $\lambda_{19,4}$ | .63 | .64 | .08 | .07 | .07 | .63 | .05 | .05 | .05 | .63 | .03 | .03 | .03 |
| $\lambda_{20,4}$ | .51 | .51 | .08 | .08 | .08 | .51 | .05 | .05 | .05 | .51 | .03 | .03 | .03 |
| $\lambda_{21,5}$ | .52 | .52 | .07 | .07 | .07 | .52 | .05 | .05 | .05 | .52 | .03 | .03 | .03 |
| $\lambda_{22,5}$ | .79 | .80 | .07 | .07 | .07 | .79 | .04 | .05 | .05 | .79 | .03 | .03 | .03 |
| $\lambda_{23,5}$ | .40 | .39 | .07 | .07 | .07 | .40 | .05 | .05 | .05 | .40 | .03 | .04 | .04 |
| $\lambda_{24,5}$ | .52 | .52 | .07 | .08 | .08 | .51 | .05 | .05 | .05 | .52 | .03 | .03 | .03 |
| $\lambda_{25,5}$ | .79 | .79 | .07 | .07 | .07 | .79 | .04 | .04 | .04 | .79 | .03 | .03 | .03 |
| $\lambda_{26,5}$ | .40 | .39 | .07 | .07 | .08 | .40 | .05 | .05 | .05 | .40 | .03 | .03 | .03 |

(continues on the next page)

Table 4 (cont.): Comparison of results on loadings using PLSc, ML, PLSe1 and PLSe2 for normal data with large model

| *PLSe1* | | *N*=200 | | | | *N* =500 | | | | *N*=1000 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\theta$ | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE |
| $\lambda_{1,1}$ | .61 | .58 | .06 | .06 | .07 | .61 | .04 | .04 | .04 | .61 | .03 | .02 | .02 |
| $\lambda_{2,1}$ | .69 | .65 | .06 | .07 | .07 | .68 | .04 | .04 | .04 | .68 | .03 | .03 | .03 |
| $\lambda_{3,1}$ | .61 | .59 | .06 | .07 | .07 | .60 | .04 | .04 | .04 | .60 | .03 | .03 | .03 |
| $\lambda_{4,1}$ | .61 | .58 | .06 | .06 | .07 | .60 | .04 | .04 | .04 | .60 | .03 | .03 | .03 |
| $\lambda_{5,1}$ | .69 | .65 | .06 | .07 | .07 | .68 | .04 | .04 | .04 | .68 | .03 | .03 | .03 |
| $\lambda_{6,1}$ | .61 | .58 | .06 | .07 | .07 | .60 | .04 | .04 | .04 | .61 | .03 | .03 | .03 |
| $\lambda_{7,2}$ | .48 | .49 | .06 | .08 | .08 | .49 | .04 | .05 | .05 | .49 | .03 | .03 | .03 |
| $\lambda_{8,2}$ | .36 | .38 | .06 | .07 | .07 | .37 | .03 | .04 | .05 | .37 | .02 | .03 | .03 |
| $\lambda_{9,2}$ | .48 | .49 | .06 | .08 | .08 | .50 | .04 | .05 | .05 | .49 | .03 | .03 | .03 |
| $\lambda_{10,2}$ | .36 | .38 | .06 | .07 | .07 | .37 | .03 | .05 | .05 | .37 | .02 | .03 | .03 |
| $\lambda_{11,3}$ | .66 | .65 | .07 | .07 | .07 | .66 | .04 | .04 | .04 | .66 | .03 | .03 | .03 |
| $\lambda_{12,3}$ | .84 | .84 | .06 | .07 | .07 | .84 | .04 | .04 | .04 | .84 | .03 | .03 | .03 |
| $\lambda_{13,3}$ | .17 | .17 | .07 | .07 | .07 | .17 | .05 | .04 | .04 | .17 | .03 | .03 | .03 |
| $\lambda_{14,3}$ | .66 | .65 | .07 | .07 | .07 | .65 | .04 | .05 | .05 | .66 | .03 | .03 | .03 |
| $\lambda_{15,3}$ | .84 | .84 | .06 | .06 | .06 | .84 | .04 | .04 | .04 | .84 | .03 | .03 | .03 |
| $\lambda_{16,3}$ | .17 | .17 | .07 | .07 | .07 | .17 | .05 | .04 | .04 | .17 | .03 | .03 | .03 |
| $\lambda_{17,4}$ | .63 | .63 | .07 | .07 | .07 | .63 | .05 | .05 | .05 | .63 | .03 | .03 | .03 |
| $\lambda_{18,4}$ | .51 | .50 | .08 | .07 | .08 | .51 | .05 | .05 | .05 | .50 | .03 | .04 | .04 |
| $\lambda_{19,4}$ | .63 | .63 | .07 | .08 | .08 | .63 | .05 | .05 | .05 | .63 | .03 | .03 | .03 |
| $\lambda_{20,4}$ | .51 | .50 | .08 | .08 | .08 | .51 | .05 | .05 | .05 | .51 | .03 | .03 | .03 |
| $\lambda_{21,5}$ | .52 | .50 | .07 | .09 | .09 | .51 | .05 | .06 | .06 | .51 | .03 | .04 | .04 |
| $\lambda_{22,5}$ | .79 | .76 | .07 | .09 | .09 | .77 | .04 | .06 | .06 | .79 | .03 | .04 | .04 |
| $\lambda_{23,5}$ | .40 | .37 | .08 | .09 | .10 | .39 | .05 | .06 | .06 | .39 | .03 | .04 | .04 |
| $\lambda_{24,5}$ | .52 | .51 | .07 | .10 | .10 | .51 | .05 | .06 | .06 | .51 | .03 | .04 | .04 |
| $\lambda_{25,5}$ | .79 | .77 | .07 | .10 | .10 | .78 | .04 | .06 | .06 | .79 | .03 | .04 | .04 |
| $\lambda_{26,5}$ | .40 | .37 | .08 | .10 | .10 | .38 | .05 | .06 | .07 | .39 | .03 | .04 | .04 |

(continues on the next page)

Table 4 (cont.): Comparison of results on loadings using PLSc, ML, PLSe1 and PLSe2 for normal data with large model

| *PLSe2* | | *N*=200 | | | | *N* =500 | | | | *N*=1000 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\theta$ | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE |
| $\lambda_{1,1}$ | .61 | .61 | .06 | .07 | .07 | .61 | .04 | .04 | .04 | .61 | .03 | .03 | .03 |
| $\lambda_{2,1}$ | .69 | .68 | .06 | .07 | .07 | .69 | .04 | .04 | .04 | .69 | .03 | .03 | .03 |
| $\lambda_{3,1}$ | .61 | .61 | .06 | .07 | .07 | .61 | .04 | .04 | .04 | .61 | .03 | .03 | .03 |
| $\lambda_{4,1}$ | .61 | .60 | .06 | .07 | .07 | .61 | .04 | .04 | .04 | .61 | .03 | .03 | .03 |
| $\lambda_{5,1}$ | .69 | .68 | .06 | .07 | .07 | .69 | .04 | .04 | .04 | .69 | .03 | .03 | .03 |
| $\lambda_{6,1}$ | .61 | .61 | .06 | .07 | .07 | .62 | .04 | .04 | .04 | .61 | .03 | .03 | .03 |
| $\lambda_{7,2}$ | .48 | .45 | .07 | .07 | .07 | .47 | .04 | .04 | .04 | .48 | .03 | .03 | .03 |
| $\lambda_{8,2}$ | .36 | .35 | .06 | .06 | .06 | .36 | .04 | .04 | .04 | .36 | .02 | .03 | .03 |
| $\lambda_{9,2}$ | .48 | .45 | .07 | .07 | .07 | .47 | .04 | .04 | .04 | .48 | .03 | .03 | .03 |
| $\lambda_{10,2}$ | .36 | .34 | .06 | .06 | .06 | .36 | .04 | .04 | .04 | .36 | .02 | .03 | .03 |
| $\lambda_{11,3}$ | .66 | .63 | .07 | .07 | .08 | .65 | .04 | .04 | .04 | .65 | .03 | .03 | .03 |
| $\lambda_{12,3}$ | .84 | .82 | .06 | .07 | .07 | .84 | .04 | .04 | .04 | .84 | .03 | .03 | .03 |
| $\lambda_{13,3}$ | .17 | .16 | .07 | .07 | .07 | .16 | .05 | .05 | .05 | .16 | .03 | .03 | .03 |
| $\lambda_{14,3}$ | .66 | .64 | .07 | .08 | .08 | .65 | .04 | .05 | .05 | .65 | .03 | .03 | .03 |
| $\lambda_{15,3}$ | .84 | .82 | .06 | .07 | .08 | .84 | .04 | .04 | .04 | .84 | .03 | .03 | .03 |
| $\lambda_{16,3}$ | .17 | .16 | .07 | .08 | .08 | .17 | .05 | .04 | .04 | .17 | .03 | .03 | .03 |
| $\lambda_{17,4}$ | .63 | .60 | .08 | .08 | .09 | .62 | .05 | .05 | .05 | .63 | .03 | .03 | .03 |
| $\lambda_{18,4}$ | .51 | .48 | .08 | .08 | .08 | .50 | .05 | .05 | .05 | .50 | .03 | .04 | .04 |
| $\lambda_{19,4}$ | .63 | .61 | .08 | .09 | .09 | .62 | .05 | .05 | .05 | .63 | .03 | .03 | .03 |
| $\lambda_{20,4}$ | .51 | .48 | .08 | .09 | .10 | .50 | .05 | .05 | .05 | .50 | .03 | .04 | .04 |
| $\lambda_{21,5}$ | .52 | .50 | .07 | .08 | .08 | .51 | .04 | .05 | .05 | .52 | .03 | .03 | .03 |
| $\lambda_{22,5}$ | .79 | .77 | .07 | .07 | .08 | .78 | .04 | .05 | .05 | .79 | .03 | .03 | .03 |
| $\lambda_{23,5}$ | .40 | .38 | .07 | .08 | .08 | .40 | .04 | .05 | .05 | .40 | .03 | .04 | .04 |
| $\lambda_{24,5}$ | .52 | .51 | .07 | .09 | .09 | .51 | .04 | .05 | .05 | .52 | .03 | .03 | .03 |
| $\lambda_{25,5}$ | .79 | .77 | .07 | .07 | .08 | .78 | .04 | .05 | .05 | .79 | .03 | .03 | .03 |
| $\lambda_{26,5}$ | .40 | .37 | .07 | .08 | .09 | .39 | .04 | .05 | .05 | .40 | .03 | .03 | .04 |

Table 5 presents the results on structural regression coefficients and the factor correlation with large model and normal data. Overall, we observe the same pattern as in Table 3, where estimates of the regression coefficients are generally more biased than the estimates of the loadings or the factor correlation. The increased number of indicators seems to improve the estimation of structural regression parameter estimates

– as more observed information becomes available for each latent variable, the inferences about the latent structural equations are expected to improve.

Across all sample sizes, RMSEs for all regression coefficient estimates are consistently smaller than their counterparts in the case of a small model, irrespective of the estimator used. At $N$=200, we observe about the same amount of bias with TSLS, ML and the two PLSe estimators. It seems that small sample size is a challenge for all four estimators. At $N$=1000, all estimators' performance improves, with ML and PLSe2 showing the smallest RMSEs. Similar to what we observed with a small model at $N$=1000, TSLS and PLSe1 are only slightly worse off than ML and PLSe2. We observed that PLSe1 improves the estimation of $\beta_3$ over TSLS. As with the small model, standard error estimates from ML and the two PLSe estimators improve as sample size increases. They are well-calibrated at $N$=1000.

Table 5. Comparison of results on structural regression coefficients and factor correlations using TSLS, ML, PLSe1 and PLSe2 for normal data with large model

| *TSLS* | | | *N*=200 | | | | *N* =500 | | | | *N*=1000 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\theta$ | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE |
| $\beta_1$ | .21 | .25 | N/A | .10 | .11 | .23 | N/A | .07 | .07 | .21 | N/A | .04 | .04 |
| $\beta_2$ | .25 | .29 | N/A | .09 | .09 | .25 | N/A | .06 | .06 | .25 | N/A | .04 | .04 |
| $\beta_3$ | .30 | .30 | N/A | .38 | .38 | .32 | N/A | .30 | .30 | .33 | N/A | .20 | .20 |
| $\beta_4$ | .27 | .26 | N/A | .14 | .14 | .26 | N/A | .08 | .08 | .26 | N/A | .06 | .06 |
| $\beta_5$ | 1.03 | .99 | N/A | .23 | .23 | .99 | N/A | .15 | .15 | 1.02 | N/A | .10 | .10 |
| $\phi_{2,1}$ | .38 | .39 | N/A | .08 | .08 | .39 | N/A | .05 | .05 | .38 | N/A | .04 | .04 |
| *ML* | | | *N*=200 | | | | *N* =500 | | | | *N*=1000 | | |
| | $\theta$ | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE |
| $\beta_1$ | .21 | .21 | .09 | .09 | .09 | .21 | .05 | .06 | .06 | .21 | .04 | .04 | .04 |
| $\beta_2$ | .25 | .27 | .09 | .08 | .09 | .25 | .06 | .06 | .06 | .25 | .04 | .04 | .04 |
| $\beta_3$ | .30 | .32 | .16 | .17 | .17 | .31 | .10 | .10 | .10 | .30 | .07 | .07 | .07 |
| $\beta_4$ | .27 | .27 | .13 | .14 | .14 | .27 | .08 | .08 | .08 | .27 | .06 | .06 | .06 |
| $\beta_5$ | 1.03 | 1.09 | .23 | .23 | .24 | 1.04 | .14 | .14 | .14 | 1.04 | .10 | .09 | .10 |
| $\phi_{2,1}$ | .38 | .37 | .08 | .09 | .09 | .38 | .05 | .05 | .05 | .38 | .04 | .04 | .04 |
| *PLSe1* | | | *N*=200 | | | | *N* =500 | | | | *N*=1000 | | |
| | $\theta$ | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE |
| $\beta_1$ | .21 | .27 | .10 | .11 | .12 | .24 | .06 | .06 | .07 | .22 | .04 | .04 | .04 |
| $\beta_2$ | .25 | .34 | .10 | .10 | .13 | .28 | .06 | .06 | .07 | .27 | .04 | .04 | .04 |
| $\beta_3$ | .30 | .26 | .15 | .19 | .19 | .28 | .10 | .11 | .11 | .29 | .07 | .07 | .07 |
| $\beta_4$ | .27 | .26 | .13 | .15 | .15 | .26 | .08 | .08 | .08 | .26 | .06 | .06 | .06 |
| $\beta_5$ | 1.03 | 1.04 | .22 | .23 | .23 | 1.00 | .13 | .15 | .15 | 1.03 | .09 | .10 | .10 |
| $\phi_{2,1}$ | .38 | .40 | .08 | .09 | .09 | .39 | .05 | .05 | .06 | .38 | .04 | .04 | .04 |
| *PLSe2* | | | *N*=200 | | | | *N* =500 | | | | *N*=1000 | | |
| | $\theta$ | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE |
| $\beta_1$ | .21 | .16 | .08 | .11 | .13 | .19 | .05 | .06 | .07 | .20 | .04 | .04 | .04 |
| $\beta_2$ | .25 | .23 | .08 | .11 | .11 | .23 | .05 | .06 | .06 | .24 | .04 | .04 | .04 |
| $\beta_3$ | .30 | .37 | .20 | .22 | .24 | .32 | .10 | .11 | .12 | .31 | .07 | .07 | .08 |
| $\beta_4$ | .27 | .26 | .27 | .31 | .31 | .26 | .08 | .09 | .09 | .26 | .06 | .06 | .06 |
| $\beta_5$ | 1.03 | 1.07 | .42 | .38 | .38 | 1.03 | .14 | .14 | .14 | 1.04 | .10 | .09 | .10 |
| $\phi_{2,1}$ | .38 | .33 | .08 | .11 | .12 | .36 | .05 | .06 | .06 | .37 | .04 | .04 | .04 |

Overall with either small or large model, what we observed with normal data shows the following trend. All estimators including PLSc (in conjunction with TSLS), ML, PLSe1 and PLSe2 seem to improve as sample size increases. Another general trend

is that the parameter estimates of the structural regression coefficients show more bias than the parameter estimates of the loadings. Specifically, at $N$=200, we observe some large biases from all four estimators. The bias from TSLS is more pronounced with a small model than with a large model. At $N$=1000, performance of the four estimators is similar, where ML and PLSe2 are the best, followed by PLSe1 and TSLS.

The conclusion with normal data is that PLSe1 and PLSe2 can improve PLSc-estimated loadings and can also improve TSLS-estimated regression coefficients, though to a lesser degree. Their performance is shown to be comparable to ML across all sample sizes, with PLSe2 closer to ML than PLSe1. Based on results from PLSc (in conjunction with TSLS), the two proposed PLSe estimators not only provide improvement to PLSc and TSLS point estimates, they also provide adequate estimation of standard errors.

### 4.2.2. Results under Non-normality

Under the non-normal conditions, we compared PLSc (in conjunction with TSLS), ML and PLSe1. PLSe2 is based on normal theory. Thus it is not studied for the non-normal conditions.

Table 6 compares the estimates of the loadings from a small model. Holding sample size constant, we observed that RMSE of the loading estimates are consistently

higher with non-normal data than under normality. A sample size of roughly 2500 is perhaps needed with non-normal data to achieve the same level of RMSE with normal data at $N$=1000.

When we compare RMSE across different estimators, RMSE for PLSc-estimated loadings are slightly higher than those from ML or PLSe1 at each sample size (with the majority of RMSEs from PLSc higher than ML by about 0.02). We observe a similar pattern of performance in PLSe1 with non-normal data as with normal data, where PLSe1 performs almost as well as ML at each sample size in terms of RMSE (with the majority of RMSEs from PLSe1 higher than ML by about 0.01). At $N$=2500, all three estimators are performing well in terms of parameter recovery, with ML having the lowest RMSE followed by PLSe1 and then PLSc. It seems that at each sample size PLSe1 is able to improve parameter estimates from PLSc.

Table 7 compares the results on structural regression coefficients and the factor correlation from a small model. Overall, we observe the same pattern as in Table 3 with normal data where estimates of the regression coefficients are generally more biased than the estimates of the loadings or the factor correlation (indicated by the higher RMSEs). Non-normality poses a challenge for the estimation of the regression coefficients, especially at $N$=200 where large RMSEs are observed with all three estimators. It takes a larger sample with non-normal data (e.g. $N$=500) for these

estimators to have similar performance as normal data when $N$=200. Comparing ML against TSLS and PLSe1 at sample sizes of 500, 1000 and 2500, all three estimators' performance improves as sample size increases. Similar to the results of the loading estimates, ML performs the best with the smallest RMSEs. PLSe1 seems to improve PLSc with slightly better RMSEs.

Table 6. Comparison of results on loadings using PLSc, ML and PLSe1 for non-normal data with small model

| *PLSc* | | | *N*=200 | | | | *N* =500 | | | | *N*=1000 | | | | *N*=2500 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\theta$ | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE |
| $\lambda_{1,1}$ | .61 | .59 | N/A | .14 | .14 | .59 | N/A | .10 | .10 | .60 | N/A | .08 | .08 | .61 | N/A | .05 | .05 |
| $\lambda_{2,1}$ | .69 | .64 | N/A | .14 | .15 | .67 | N/A | .10 | .10 | .66 | N/A | .07 | .07 | .68 | N/A | .05 | .05 |
| $\lambda_{3,1}$ | .61 | .58 | N/A | .16 | .16 | .60 | N/A | .10 | .10 | .61 | N/A | .07 | .07 | .61 | N/A | .05 | .05 |
| $\lambda_{4,2}$ | .48 | .54 | N/A | .16 | .17 | .53 | N/A | .14 | .14 | .52 | N/A | .12 | .13 | .50 | N/A | .07 | .07 |
| $\lambda_{5,2}$ | .36 | .41 | N/A | .12 | .13 | .40 | N/A | .10 | .11 | .39 | N/A | .09 | .09 | .38 | N/A | .05 | .05 |
| $\lambda_{6,3}$ | .66 | .61 | N/A | .14 | .15 | .63 | N/A | .12 | .12 | .64 | N/A | .08 | .08 | .65 | N/A | .05 | .05 |
| $\lambda_{7,3}$ | .84 | .75 | N/A | .15 | .17 | .79 | N/A | .12 | .13 | .81 | N/A | .09 | .09 | .83 | N/A | .06 | .06 |
| $\lambda_{8,3}$ | .17 | .29 | N/A | .16 | .20 | .22 | N/A | .13 | .14 | .21 | N/A | .12 | .12 | .17 | N/A | .08 | .08 |
| $\lambda_{9,4}$ | .63 | .66 | N/A | .15 | .15 | .64 | N/A | .10 | .10 | .63 | N/A | .07 | .07 | .64 | N/A | .05 | .05 |
| $\lambda_{10,4}$ | .51 | .52 | N/A | .13 | .13 | .52 | N/A | .10 | .10 | .52 | N/A | .07 | .07 | .51 | N/A | .05 | .05 |
| $\lambda_{11,5}$ | .52 | .55 | N/A | .26 | .26 | .55 | N/A | .21 | .21 | .50 | N/A | .20 | .20 | .51 | N/A | .15 | .15 |
| $\lambda_{12,5}$ | .79 | .62 | N/A | .22 | .28 | .64 | N/A | .24 | .29 | .68 | N/A | .20 | .23 | .74 | N/A | .15 | .16 |
| $\lambda_{13,5}$ | .40 | .48 | N/A | .25 | .26 | .44 | N/A | .20 | .20 | .45 | N/A | .19 | .20 | .41 | N/A | .15 | .15 |
| *ML* | | | *N*=200 | | | | *N* =500 | | | | *N*=1000 | | | | *N*=2500 | | | |
| | $\theta$ | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE |
| $\lambda_{1,1}$ | .61 | .58 | .07 | .11 | .11 | .60 | .04 | .08 | .08 | .60 | .03 | .05 | .05 | .60 | .02 | .03 | .03 |
| $\lambda_{2,1}$ | .69 | .66 | .07 | .11 | .11 | .67 | .04 | .07 | .08 | .68 | .03 | .05 | .05 | .68 | .02 | .04 | .04 |
| $\lambda_{3,1}$ | .61 | .59 | .07 | .11 | .11 | .60 | .04 | .07 | .07 | .60 | .03 | .05 | .05 | .61 | .02 | .03 | .03 |
| $\lambda_{4,2}$ | .48 | .47 | .14 | .15 | .15 | .47 | .08 | .10 | .10 | .47 | .06 | .08 | .08 | .48 | .03 | .06 | .06 |
| $\lambda_{5,2}$ | .36 | .35 | .10 | .10 | .11 | .35 | .06 | .08 | .08 | .35 | .04 | .06 | .06 | .36 | .02 | .04 | .04 |
| $\lambda_{6,3}$ | .66 | .66 | .09 | .13 | .13 | .66 | .06 | .10 | .10 | .66 | .04 | .07 | .07 | .66 | .03 | .05 | .05 |
| $\lambda_{7,3}$ | .84 | .81 | .10 | .13 | .13 | .82 | .07 | .10 | .10 | .85 | .05 | .07 | .07 | .84 | .03 | .05 | .05 |
| $\lambda_{8,3}$ | .17 | .20 | .08 | .11 | .12 | .17 | .05 | .07 | .07 | .17 | .04 | .06 | .06 | .16 | .02 | .04 | .04 |
| $\lambda_{9,4}$ | .63 | .65 | .10 | .14 | .14 | .64 | .06 | .11 | .11 | .63 | .04 | .08 | .08 | .64 | .03 | .05 | .05 |
| $\lambda_{10,4}$ | .51 | .51 | .09 | .14 | .14 | .51 | .06 | .10 | .10 | .51 | .04 | .07 | .07 | .51 | .03 | .05 | .05 |
| $\lambda_{11,5}$ | .52 | .56 | .09 | .14 | .15 | .53 | .06 | .09 | .10 | .52 | .04 | .06 | .06 | .52 | .03 | .05 | .05 |
| $\lambda_{12,5}$ | .79 | .75 | .11 | .14 | .15 | .78 | .08 | .11 | .11 | .79 | .06 | .09 | .09 | .80 | .04 | .06 | .06 |
| $\lambda_{13,5}$ | .40 | .43 | .09 | .13 | .13 | .41 | .06 | .08 | .08 | .40 | .04 | .07 | .07 | .40 | .02 | .04 | .04 |

Table 6 (cont.) Comparison of results on loadings using PLSc, ML and PLSe1 for non-normal data with small model

| *PLSe1* | *N*=200 | | | | | *N* =500 | | | | *N*=1000 | | | | *N*=2500 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\theta$ | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE |
| $\lambda_{1,1}$ | .61 | .57 | .10 | .11 | .12 | .59 | .07 | .08 | .08 | .59 | .05 | .05 | .06 | .60 | .03 | .04 | .04 |
| $\lambda_{2,1}$ | .69 | .63 | .11 | .11 | .13 | .65 | .07 | .08 | .08 | .67 | .05 | .05 | .06 | .68 | .04 | .04 | .04 |
| $\lambda_{3,1}$ | .61 | .57 | .10 | .12 | .13 | .59 | .07 | .07 | .08 | .60 | .05 | .06 | .06 | .61 | .03 | .03 | .03 |
| $\lambda_{4,2}$ | .48 | .51 | .16 | .15 | .15 | .51 | .11 | .12 | .12 | .51 | .08 | .11 | .11 | .50 | .05 | .07 | .07 |
| $\lambda_{5,2}$ | .36 | .39 | .12 | .11 | .12 | .39 | .08 | .09 | .10 | .38 | .06 | .08 | .08 | .38 | .04 | .05 | .05 |
| $\lambda_{6,3}$ | .66 | .65 | .13 | .12 | .12 | .66 | .09 | .11 | .11 | .66 | .07 | .07 | .07 | .66 | .04 | .05 | .05 |
| $\lambda_{7,3}$ | .84 | .78 | .14 | .12 | .14 | .81 | .10 | .10 | .10 | .84 | .08 | .07 | .07 | .83 | .05 | .05 | .05 |
| $\lambda_{8,3}$ | .17 | .21 | .11 | .12 | .13 | .18 | .07 | .20 | .20 | .17 | .05 | .06 | .06 | .16 | .03 | .04 | .04 |
| $\lambda_{9,4}$ | .63 | .63 | .14 | .14 | .14 | .64 | .10 | .10 | .10 | .63 | .07 | .08 | .08 | .64 | .05 | .05 | .05 |
| $\lambda_{10,4}$ | .51 | .51 | .13 | .13 | .13 | .51 | .08 | .10 | .10 | .51 | .06 | .07 | .07 | .51 | .04 | .05 | .05 |
| $\lambda_{11,5}$ | .52 | .55 | .13 | .15 | .15 | .52 | .09 | .12 | .12 | .51 | .07 | .09 | .09 | .51 | .04 | .05 | .05 |
| $\lambda_{12,5}$ | .79 | .69 | .15 | .16 | .19 | .72 | .11 | .15 | .16 | .76 | .08 | .11 | .12 | .79 | .06 | .07 | .07 |
| $\lambda_{13,5}$ | .40 | .42 | .12 | .14 | .15 | .40 | .08 | .09 | .09 | .38 | .06 | .08 | .08 | .39 | .04 | .04 | .05 |

Table 7. Comparison of results on structural regression coefficients and factor correlations using TSLS, ML and PLSe1 for non-normal data with small model

| *TSLS* | | *N*=200 | | | | *N* =500 | | | | *N*=1000 | | | | *N*=2500 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\theta$ | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE |
| $\beta_1$ | .21 | .27 | N/A | .18 | .19 | .26 | N/A | .13 | .13 | .24 | N/A | .10 | .10 | .23 | N/A | .06 | .06 |
| $\beta_2$ | .25 | .32 | N/A | .15 | .16 | .27 | N/A | .09 | .09 | .27 | N/A | .08 | .08 | .25 | N/A | .05 | .05 |
| $\beta_3$ | .30 | .27 | N/A | .45 | .45 | .33 | N/A | .40 | .40 | .29 | N/A | .39 | .39 | .32 | N/A | .28 | .28 |
| $\beta_4$ | .27 | .30 | N/A | .31 | .32 | .26 | N/A | .20 | .20 | .25 | N/A | .13 | .13 | .26 | N/A | .10 | .10 |
| $\beta_5$ | 1.03 | .93 | N/A | .45 | .46 | .97 | N/A | .37 | .37 | 1.00 | N/A | .30 | .30 | 1.00 | N/A | .18 | .18 |
| $\phi_{2,1}$ | .38 | .40 | N/A | .15 | .15 | .39 | N/A | .12 | .12 | .40 | N/A | .09 | .09 | .38 | N/A | .05 | .05 |
| *ML* | | *N*=200 | | | | *N* =500 | | | | *N*=1000 | | | | *N*=2500 | | | |
| | $\theta$ | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE |
| $\beta_1$ | .21 | .24 | .13 | .18 | .18 | .22 | .08 | .11 | .11 | .22 | .05 | .09 | .09 | .22 | .03 | .05 | .05 |
| $\beta_2$ | .25 | .32 | .11 | .17 | .18 | .27 | .07 | .09 | .09 | .26 | .05 | .08 | .08 | .25 | .03 | .05 | .05 |
| $\beta_3$ | .30 | .35 | .49 | .40 | .41 | .30 | .16 | .20 | .20 | .32 | .11 | .17 | .17 | .30 | .06 | .11 | .11 |
| $\beta_4$ | .27 | .31 | .36 | .35 | .35 | .27 | .13 | .18 | .18 | .27 | .09 | .12 | .12 | .27 | .05 | .10 | .10 |
| $\beta_5$ | 1.03 | 1.19 | 1.26 | .81 | .83 | 1.10 | .32 | .39 | .40 | 1.10 | .21 | .33 | .33 | 1.05 | .11 | .18 | .18 |
| $\phi_{2,1}$ | .38 | .38 | .11 | .14 | .14 | .38 | .07 | .11 | .11 | .39 | .05 | .08 | .08 | .38 | .03 | .05 | .05 |
| *PLSe1* | | *N*=200 | | | | *N* =500 | | | | *N*=1000 | | | | *N*=2500 | | | |
| | $\theta$ | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE |
| $\beta_1$ | .21 | .29 | .19 | .19 | .21 | .27 | .12 | .12 | .13 | .25 | .08 | .09 | .10 | .23 | .05 | .06 | .06 |
| $\beta_2$ | .25 | .38 | .17 | .17 | .22 | .31 | .10 | .10 | .11 | .29 | .08 | .08 | .09 | .26 | .05 | .05 | .05 |
| $\beta_3$ | .30 | .27 | .32 | .34 | .34 | .26 | .20 | .21 | .21 | .28 | .15 | .19 | .20 | .28 | .09 | .11 | .11 |
| $\beta_4$ | .27 | .28 | .34 | .32 | .32 | .25 | .18 | .19 | .20 | .25 | .12 | .12 | .13 | .26 | .08 | .09 | .09 |
| $\beta_5$ | 1.03 | 1.01 | .65 | .48 | .48 | 1.01 | .37 | .37 | .37 | 1.02 | .27 | .31 | .31 | 1.01 | .16 | .18 | .18 |
| $\phi_{2,1}$ | .38 | .41 | .14 | .15 | .15 | .40 | .10 | .12 | .12 | .40 | .07 | .08 | .08 | .39 | .05 | .05 | .05 |

Table 8 and Table 9 present parameter estimates with the large model and non-normal data using PLSc (in conjunction with TSLS), ML and PLSe1. We observe largely similar patterns as the case of the small model. At $N$=200, all three estimators seem to produce estimates with high RMSEs, but their performance improves as sample size increases. Similar to the results with small model at $N$=2500, ML performs the best with the smallest RMSEs. PLSe1 seems to improve PLSc.

We also observe, as before, that a model with more indicators seems to help with the estimation of structural parameter estimates (see Table 9). At $N$=2500, RMSEs for all regression parameter estimates are consistently smaller with a large model than a small model, irrespective of the estimator. Non-normality again poses difficulty in estimation (regardless of the size of the model), where it takes a much larger sample size ($N$=2500) to maintain the same level of accuracy achievable with normal data at $N$=1000.

Table 8. Comparison of results on loadings using PLSc, ML and PLSe1 for non-normal data with large model

| *PLSc* | *N*=200 | | | | | *N* =500 | | | | *N*=1000 | | | | *N*=2500 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\theta$ | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE |
| $\lambda_{1,1}$ | .61 | .58 | N/A | .14 | .15 | .59 | N/A | .11 | .11 | .59 | N/A | .08 | .08 | .61 | N/A | .05 | .05 |
| $\lambda_{2,1}$ | .69 | .65 | N/A | .13 | .14 | .67 | N/A | .09 | .09 | .67 | N/A | .07 | .07 | .68 | N/A | .04 | .04 |
| $\lambda_{3,1}$ | .61 | .57 | N/A | .14 | .15 | .59 | N/A | .11 | .11 | .60 | N/A | .08 | .08 | .61 | N/A | .05 | .05 |
| $\lambda_{4,1}$ | .61 | .57 | N/A | .14 | .15 | .59 | N/A | .10 | .11 | .60 | N/A | .07 | .08 | .61 | N/A | .05 | .05 |
| $\lambda_{5,1}$ | .69 | .64 | N/A | .13 | .13 | .66 | N/A | .09 | .09 | .68 | N/A | .06 | .07 | .68 | N/A | .04 | .04 |
| $\lambda_{6,1}$ | .61 | .56 | N/A | .14 | .15 | .60 | N/A | .10 | .11 | .60 | N/A | .08 | .08 | .60 | N/A | .05 | .05 |
| $\lambda_{7,2}$ | .48 | .52 | N/A | .16 | .17 | .52 | N/A | .12 | .13 | .50 | N/A | .08 | .09 | .49 | N/A | .04 | .04 |
| $\lambda_{8,2}$ | .36 | .40 | N/A | .13 | .14 | .40 | N/A | .10 | .10 | .39 | N/A | .07 | .07 | .37 | N/A | .03 | .03 |
| $\lambda_{9,2}$ | .48 | .52 | N/A | .15 | .15 | .52 | N/A | .11 | .12 | .50 | N/A | .08 | .09 | .49 | N/A | .04 | .04 |
| $\lambda_{10,2}$ | .36 | .38 | N/A | .12 | .12 | .40 | N/A | .10 | .11 | .38 | N/A | .07 | .07 | .37 | N/A | .03 | .03 |
| $\lambda_{11,3}$ | .66 | .60 | N/A | .18 | .19 | .64 | N/A | .12 | .12 | .64 | N/A | .10 | .10 | .64 | N/A | .06 | .06 |
| $\lambda_{12,3}$ | .84 | .76 | N/A | .15 | .17 | .81 | N/A | .09 | .10 | .82 | N/A | .07 | .08 | .84 | N/A | .05 | .05 |
| $\lambda_{13,3}$ | .17 | .25 | N/A | .17 | .19 | .21 | N/A | .14 | .14 | .19 | N/A | .12 | .12 | .17 | N/A | .09 | .09 |
| $\lambda_{14,3}$ | .66 | .63 | N/A | .17 | .17 | .63 | N/A | .13 | .13 | .64 | N/A | .11 | .11 | .66 | N/A | .07 | .07 |
| $\lambda_{15,3}$ | .84 | .77 | N/A | .15 | .17 | .80 | N/A | .09 | .10 | .82 | N/A | .08 | .08 | .83 | N/A | .05 | .05 |
| $\lambda_{16,3}$ | .17 | .24 | N/A | .17 | .19 | .21 | N/A | .13 | .14 | .18 | N/A | .11 | .11 | .17 | N/A | .08 | .08 |
| $\lambda_{17,4}$ | .63 | .62 | N/A | .13 | .13 | .62 | N/A | .09 | .09 | .62 | N/A | .07 | .07 | .63 | N/A | .05 | .05 |
| $\lambda_{18,4}$ | .51 | .49 | N/A | .15 | .15 | .50 | N/A | .11 | .11 | .51 | N/A | .08 | .08 | .50 | N/A | .05 | .05 |
| $\lambda_{19,4}$ | .63 | .62 | N/A | .14 | .14 | .61 | N/A | .09 | .09 | .62 | N/A | .07 | .07 | .63 | N/A | .05 | .05 |
| $\lambda_{20,4}$ | .51 | .49 | N/A | .13 | .13 | .49 | N/A | .10 | .10 | .51 | N/A | .08 | .08 | .50 | N/A | .05 | .05 |
| $\lambda_{21,5}$ | .52 | .49 | N/A | .24 | .24 | .47 | N/A | .23 | .23 | .49 | N/A | .20 | .20 | .50 | N/A | .16 | .16 |
| $\lambda_{22,5}$ | .79 | .61 | N/A | .24 | .30 | .65 | N/A | .20 | .25 | .67 | N/A | .18 | .21 | .73 | N/A | .13 | .15 |
| $\lambda_{23,5}$ | .40 | .42 | N/A | .25 | .25 | .40 | N/A | .24 | .24 | .41 | N/A | .22 | .22 | .39 | N/A | .18 | .18 |
| $\lambda_{24,5}$ | .52 | .44 | N/A | .23 | .25 | .49 | N/A | .22 | .22 | .47 | N/A | .20 | .21 | .50 | N/A | .17 | .17 |
| $\lambda_{25,5}$ | .79 | .60 | N/A | .23 | .30 | .65 | N/A | .21 | .26 | .67 | N/A | .18 | .22 | .72 | N/A | .13 | .15 |
| $\lambda_{26,5}$ | .40 | .43 | N/A | .25 | .25 | .38 | N/A | .22 | .22 | .41 | N/A | .21 | .21 | .39 | N/A | .17 | .17 |

Table 8 (cont.) Comparison of results on loadings using PLSc, ML and PLSe1 for non-normal data with large model

| *ML* | | | *N*=200 | | | | *N* =500 | | | | *N*=1000 | | | | *N*=2500 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\theta$ | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE |
| $\lambda_{1,1}$ | .61 | .59 | .06 | .10 | .10 | .61 | .04 | .07 | .07 | .60 | .03 | .05 | .05 | .61 | .02 | .03 | .03 |
| $\lambda_{2,1}$ | .69 | .67 | .06 | .09 | .10 | .69 | .04 | .08 | .08 | .68 | .03 | .05 | .05 | .69 | .02 | .04 | .04 |
| $\lambda_{3,1}$ | .61 | .59 | .06 | .09 | .09 | .61 | .04 | .07 | .07 | .61 | .03 | .05 | .05 | .61 | .02 | .03 | .03 |
| $\lambda_{4,1}$ | .61 | .59 | .06 | .09 | .09 | .60 | .04 | .07 | .07 | .60 | .03 | .05 | .05 | .61 | .02 | .03 | .03 |
| $\lambda_{5,1}$ | .69 | .66 | .06 | .10 | .11 | .69 | .04 | .07 | .07 | .68 | .03 | .05 | .05 | .69 | .02 | .04 | .04 |
| $\lambda_{6,1}$ | .61 | .59 | .06 | .09 | .09 | .61 | .04 | .07 | .07 | .61 | .03 | .05 | .05 | .61 | .02 | .04 | .04 |
| $\lambda_{7,2}$ | .48 | .46 | .06 | .10 | .10 | .47 | .04 | .06 | .07 | .48 | .03 | .05 | .05 | .48 | .02 | .03 | .03 |
| $\lambda_{8,2}$ | .36 | .36 | .06 | .08 | .08 | .35 | .04 | .06 | .06 | .36 | .02 | .04 | .04 | .37 | .02 | .03 | .03 |
| $\lambda_{9,2}$ | .48 | .47 | .06 | .10 | .10 | .47 | .04 | .07 | .07 | .48 | .03 | .05 | .05 | .48 | .02 | .03 | .03 |
| $\lambda_{10,2}$ | .36 | .35 | .06 | .09 | .09 | .36 | .04 | .06 | .06 | .36 | .02 | .04 | .04 | .37 | .02 | .03 | .03 |
| $\lambda_{11,3}$ | .66 | .65 | .07 | .11 | .11 | .66 | .04 | .06 | .06 | .66 | .03 | .05 | .05 | .66 | .02 | .03 | .03 |
| $\lambda_{12,3}$ | .84 | .83 | .06 | .10 | .10 | .84 | .04 | .06 | .06 | .84 | .03 | .05 | .05 | .84 | .02 | .04 | .04 |
| $\lambda_{13,3}$ | .17 | .17 | .07 | .10 | .10 | .16 | .05 | .07 | .07 | .17 | .03 | .06 | .06 | .17 | .02 | .03 | .03 |
| $\lambda_{14,3}$ | .66 | .65 | .07 | .09 | .09 | .65 | .04 | .07 | .07 | .65 | .03 | .05 | .05 | .66 | .02 | .04 | .04 |
| $\lambda_{15,3}$ | .84 | .82 | .06 | .10 | .10 | .84 | .04 | .07 | .07 | .84 | .03 | .05 | .05 | .84 | .02 | .04 | .04 |
| $\lambda_{16,3}$ | .17 | .18 | .07 | .11 | .11 | .16 | .05 | .07 | .07 | .16 | .03 | .05 | .05 | .17 | .02 | .04 | .04 |
| $\lambda_{17,4}$ | .63 | .62 | .07 | .12 | .12 | .62 | .05 | .09 | .09 | .62 | .03 | .06 | .06 | .63 | .02 | .04 | .04 |
| $\lambda_{18,4}$ | .51 | .50 | .08 | .12 | .12 | .50 | .05 | .08 | .08 | .50 | .03 | .06 | .06 | .51 | .02 | .04 | .04 |
| $\lambda_{19,4}$ | .63 | .61 | .07 | .11 | .12 | .62 | .05 | .08 | .08 | .63 | .03 | .06 | .07 | .63 | .02 | .04 | .04 |
| $\lambda_{20,4}$ | .51 | .50 | .08 | .10 | .10 | .50 | .05 | .08 | .08 | .51 | .03 | .06 | .06 | .51 | .02 | .04 | .04 |
| $\lambda_{21,5}$ | .52 | .51 | .07 | .11 | .11 | .51 | .05 | .07 | .07 | .52 | .03 | .06 | .06 | .52 | .02 | .04 | .04 |
| $\lambda_{22,5}$ | .79 | .78 | .07 | .10 | .10 | .79 | .04 | .07 | .07 | .79 | .03 | .05 | .05 | .79 | .02 | .04 | .04 |
| $\lambda_{23,5}$ | .40 | .40 | .07 | .11 | .11 | .40 | .05 | .09 | .09 | .39 | .03 | .05 | .05 | .40 | .02 | .04 | .04 |
| $\lambda_{24,5}$ | .52 | .51 | .07 | .10 | .11 | .51 | .05 | .07 | .08 | .52 | .03 | .05 | .05 | .52 | .02 | .03 | .04 |
| $\lambda_{25,5}$ | .79 | .77 | .07 | .11 | .11 | .78 | .04 | .07 | .07 | .80 | .03 | .06 | .06 | .79 | .02 | .04 | .04 |
| $\lambda_{26,5}$ | .40 | .40 | .07 | .12 | .12 | .39 | .05 | .07 | .07 | .40 | .03 | .05 | .05 | .40 | .02 | .03 | .03 |

Table 8 (cont.) Comparison of results on loadings using PLSc, ML and PLSe1 for non-normal data with large model

| *PLSe1* | | *N*=200 | | | | *N* =500 | | | | *N*=1000 | | | | *N*=2500 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\theta$ | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE |
| $\lambda_{1,1}$ | .61 | .56 | .09 | .11 | .12 | .59 | .06 | .07 | .08 | .59 | .05 | .05 | .05 | .61 | .03 | .03 | .03 |
| $\lambda_{2,1}$ | .69 | .62 | .09 | .12 | .14 | .67 | .07 | .08 | .08 | .67 | .05 | .05 | .05 | .68 | .03 | .04 | .04 |
| $\lambda_{3,1}$ | .61 | .55 | .09 | .10 | .12 | .59 | .06 | .08 | .08 | .60 | .05 | .05 | .05 | .61 | .03 | .03 | .03 |
| $\lambda_{4,1}$ | .61 | .55 | .09 | .10 | .12 | .58 | .06 | .07 | .08 | .59 | .05 | .05 | .05 | .61 | .03 | .03 | .03 |
| $\lambda_{5,1}$ | .69 | .62 | .09 | .11 | .13 | .67 | .06 | .08 | .08 | .67 | .05 | .05 | .05 | .68 | .03 | .03 | .03 |
| $\lambda_{6,1}$ | .61 | .55 | .09 | .11 | .12 | .59 | .06 | .07 | .07 | .60 | .05 | .05 | .05 | .61 | .03 | .04 | .04 |
| $\lambda_{7,2}$ | .48 | .48 | .09 | .13 | .13 | .50 | .06 | .09 | .09 | .50 | .04 | .06 | .07 | .49 | .03 | .04 | .04 |
| $\lambda_{8,2}$ | .36 | .37 | .08 | .11 | .11 | .38 | .05 | .07 | .07 | .38 | .04 | .05 | .06 | .37 | .03 | .03 | .03 |
| $\lambda_{9,2}$ | .48 | .48 | .09 | .13 | .13 | .50 | .06 | .09 | .09 | .50 | .04 | .07 | .07 | .49 | .03 | .04 | .04 |
| $\lambda_{10,2}$ | .36 | .36 | .08 | .11 | .11 | .38 | .05 | .07 | .08 | .38 | .04 | .05 | .06 | .37 | .03 | .03 | .03 |
| $\lambda_{11,3}$ | .66 | .64 | .10 | .13 | .13 | .65 | .06 | .06 | .06 | .65 | .05 | .05 | .05 | .66 | .03 | .03 | .03 |
| $\lambda_{12,3}$ | .84 | .83 | .10 | .12 | .13 | .84 | .06 | .06 | .06 | .84 | .05 | .05 | .05 | .84 | .03 | .04 | .04 |
| $\lambda_{13,3}$ | .17 | .16 | .10 | .12 | .12 | .16 | .07 | .07 | .07 | .17 | .05 | .06 | .06 | .17 | .03 | .03 | .03 |
| $\lambda_{14,3}$ | .66 | .65 | .10 | .11 | .11 | .65 | .07 | .07 | .07 | .65 | .05 | .06 | .06 | .66 | .03 | .04 | .04 |
| $\lambda_{15,3}$ | .84 | .82 | .10 | .12 | .13 | .84 | .06 | .07 | .07 | .84 | .05 | .06 | .06 | .84 | .03 | .04 | .04 |
| $\lambda_{16,3}$ | .17 | .18 | .10 | .12 | .12 | .16 | .07 | .07 | .07 | .16 | .05 | .06 | .06 | .17 | .03 | .04 | .04 |
| $\lambda_{17,4}$ | .63 | .62 | .11 | .12 | .12 | .61 | .08 | .09 | .09 | .62 | .06 | .06 | .06 | .63 | .04 | .04 | .04 |
| $\lambda_{18,4}$ | .51 | .50 | .11 | .12 | .12 | .49 | .07 | .08 | .08 | .50 | .06 | .06 | .06 | .50 | .04 | .04 | .04 |
| $\lambda_{19,4}$ | .63 | .61 | .11 | .12 | .12 | .62 | .08 | .08 | .09 | .62 | .06 | .07 | .07 | .63 | .04 | .04 | .04 |
| $\lambda_{20,4}$ | .51 | .49 | .11 | .10 | .11 | .49 | .07 | .08 | .08 | .51 | .06 | .06 | .06 | .50 | .04 | .04 | .04 |
| $\lambda_{21,5}$ | .52 | .49 | .10 | .14 | .15 | .50 | .07 | .09 | .10 | .51 | .05 | .07 | .07 | .51 | .03 | .04 | .04 |
| $\lambda_{22,5}$ | .79 | .74 | .10 | .14 | .15 | .77 | .07 | .10 | .10 | .78 | .05 | .07 | .07 | .79 | .03 | .04 | .04 |
| $\lambda_{23,5}$ | .40 | .38 | .10 | .15 | .15 | .38 | .07 | .10 | .10 | .38 | .05 | .07 | .07 | .39 | .03 | .04 | .04 |
| $\lambda_{24,5}$ | .52 | .50 | .10 | .13 | .13 | .50 | .07 | .10 | .10 | .51 | .05 | .07 | .07 | .51 | .03 | .04 | .04 |
| $\lambda_{25,5}$ | .79 | .74 | .10 | .15 | .16 | .76 | .07 | .10 | .11 | .78 | .05 | .07 | .07 | .78 | .03 | .04 | .04 |
| $\lambda_{26,5}$ | .40 | .38 | .11 | .14 | .14 | .37 | .07 | .09 | .09 | .39 | .05 | .07 | .07 | .40 | .03 | .04 | .04 |

Table 9. Comparison of results on structural regression coefficients and factor correlations using TSLS, ML and PLSe1 for non-normal data with large model

| *TSLS* | | *N*=200 | | | | *N* =500 | | | | *N*=1000 | | | | *N*=2500 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\theta$ | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE |
| $\beta_1$ | .21 | .28 | N/A | .14 | .16 | .25 | N/A | .10 | .11 | .23 | N/A | .08 | .08 | .22 | N/A | .04 | .04 |
| $\beta_2$ | .25 | .31 | N/A | .13 | .14 | .28 | N/A | .09 | .09 | .26 | N/A | .06 | .06 | .26 | N/A | .04 | .04 |
| $\beta_3$ | .30 | .28 | N/A | .45 | .45 | .27 | N/A | .39 | .39 | .30 | N/A | .30 | .30 | .31 | N/A | .21 | .21 |
| $\beta_4$ | .27 | .25 | N/A | .20 | .21 | .24 | N/A | .13 | .13 | .26 | N/A | .09 | .09 | .26 | N/A | .06 | .06 |
| $\beta_5$ | 1.03 | 1.00 | N/A | .35 | .35 | .99 | N/A | .22 | .22 | .98 | N/A | .16 | .17 | 1.01 | N/A | .10 | .10 |
| $\phi_{2,1}$ | .38 | .41 | N/A | .11 | .12 | .39 | N/A | .07 | .07 | .38 | N/A | .06 | .06 | .38 | N/A | .04 | .04 |
| *ML* | | *N*=200 | | | | *N* =500 | | | | *N*=1000 | | | | *N*=2500 | | | |
| | $\theta$ | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE |
| $\beta_1$ | .21 | .23 | .09 | .14 | .14 | .21 | .06 | .09 | .09 | .22 | .04 | .06 | .06 | .21 | .02 | .04 | .04 |
| $\beta_2$ | .25 | .29 | .09 | .13 | .14 | .26 | .06 | .09 | .09 | .25 | .04 | .06 | .06 | .25 | .03 | .04 | .04 |
| $\beta_3$ | .30 | .29 | .17 | .24 | .24 | .31 | .10 | .15 | .15 | .30 | .07 | .11 | .11 | .29 | .04 | .07 | .07 |
| $\beta_4$ | .27 | .26 | .14 | .21 | .21 | .26 | .08 | .13 | .13 | .27 | .06 | .09 | .09 | .27 | .04 | .06 | .06 |
| $\beta_5$ | 1.03 | 1.11 | .26 | .37 | .38 | 1.08 | .15 | .21 | .22 | 1.04 | .10 | .15 | .15 | 1.03 | .06 | .10 | .10 |
| $\phi_{2,1}$ | .38 | .39 | .08 | .11 | .11 | .38 | .05 | .08 | .08 | .37 | .04 | .06 | .06 | .38 | .02 | .04 | .04 |
| *PLSe1* | | *N*=200 | | | | *N* =500 | | | | *N*=1000 | | | | *N*=2500 | | | |
| | $\theta$ | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE | $\hat{\theta}$ | $SE(\hat{\theta})$ | $SD(\hat{\theta})$ | RMSE |
| $\beta_1$ | .21 | .32 | .14 | .21 | .23 | .27 | .09 | .10 | .11 | .24 | .06 | .07 | .08 | .23 | .04 | .04 | .04 |
| $\beta_2$ | .25 | .38 | .14 | .16 | .21 | .32 | .08 | .10 | .12 | .29 | .06 | .07 | .07 | .27 | .04 | .04 | .05 |
| $\beta_3$ | .30 | .24 | .24 | .35 | .35 | .25 | .14 | .18 | .19 | .27 | .10 | .12 | .12 | .28 | .07 | .07 | .07 |
| $\beta_4$ | .27 | .26 | .21 | .23 | .23 | .24 | .12 | .13 | .13 | .26 | .08 | .09 | .09 | .26 | .06 | .06 | .06 |
| $\beta_5$ | 1.03 | 1.09 | .36 | .40 | .40 | 1.03 | .20 | .22 | .22 | 1.00 | .14 | .16 | .16 | 1.02 | .09 | .10 | .10 |
| $\phi_{2,1}$ | .38 | .44 | .11 | .12 | .13 | .41 | .07 | .07 | .08 | .39 | .06 | .06 | .06 | .38 | .04 | .04 | .04 |

We observe some interesting patterns in standard error estimates with non-normal data. From the earlier conditions with normal data (as presented in Tables 2-5), the standard error estimates from ML, PLSe1 and PLSe2 are well-calibrated across all sample sizes regardless of the size of the model. In the non-normal conditions, we intentionally used ML instead of robust ML to have an opportunity to see how much ML will underestimate the standard errors under non-normality.

As we can see from Tables 6 to 9 for all the non-normal conditions, ML consistently underestimated standard errors across all sample sizes. PLSe1, on the hand, shows some instability in standard error estimation at $N$=200 (slight underestimation of the loadings and some occasionally more severe underestimation of regression coefficients). When sample size reaches 500, its performance improves markedly. Although it still occasionally underestimates, the scale of such underestimation is not as large as is observed with ML. At $N$=2500 with large model (see Table 9), PLSe1 produces standard errors for the structural regression coefficients that are consistent with the observed Monte Carlo variability.

Across the normal and non-normal conditions, we can make the following conclusions. First, the performance of all estimators improves as sample size increases. Second, the parameter estimates of the structural regression coefficients show more bias than the loading estimates, especially at the smallest sample size. Third, a large model can improve the estimation of the regression coefficients, leading to consistently smaller

RMSEs than those from a small model. PLSc (in conjunction with TSLS) benefits more from a large measurement model than other estimators, especially when $N$ is small. Fourth, non-normality poses challenge for the estimation of the regression coefficients especially in conjunction with small $N$. With non-normal data, we found that it routinely takes a larger sample size for TSLS, ML and PLSe1 to maintain a comparable level of performance achievable with normal data.

To summarize, PLSe1 and PLSe2 can improve PLSc-estimated loadings and can also improve TSLS-estimated regression coefficients, although to a lesser degree. Their performance is comparable to ML across all sample sizes, with PLSe2 closer to ML than PLSe1. In addition, the proposed estimators also provide adequate estimation of the standard errors under normality (PLSe1 and PLSe2) and non-normality (PLSe1 only).

### 4.3. Model Fit Test Statistics

One additional advantage of the proposed PLSe estimators is that they produce formal model fit test statistics. We recorded the minimum fit function chi-square statistic ($c_1$) for ML, PLSe1 and PLSe2 (under normality only). Several robust chi-square statistics are computed under PLSe1. These include the Satorra-Bentler scaled chi-square statistic ($c_3$), Browne's residual-based statistic ($c_4$) and Yuan and Bentler's (1997) adjustment statistic ($T_{YB}$) which is based on $c_4$.

The model we estimated is correctly specified, thus under normality, the expected value for the $c_1$ chi-square statistic in a small model is equal to the model's degrees of freedom ($d = m^* - q = 59$) and the expected variance of this test statistic is $2d = 118$. Similarly with a large model, the expected value of $c_1$ is 293 ($d = 293$) and the expected variances is 586 ($2d = 585$). With non-normal data, we resort to the robust statistics, such as $c_3$, $c_4$ and $T_{YB}$ with known expected means (equal to the model's degrees of freedom).

First, we present results from the normal condition in Table 10. The upper panel of the table compares the test statistics for ML, PLSe1 and PLSe2 with a small model. At $N$=200, the ML $c_1$ test statistic has a mean of 60.54 and variance of 115.27, which are very close to the expected values (58 and 118, respectively). The PLSe2 $c_1$ test statistic has a mean of 58.44 and a slightly elevated variance of 142.02, but the elevated variance gradually decreases as sample size increases. At $N$=1000, both ML and PLSe2 produce test statistics with empirical distributions closely following the theoretical expectation (mean of 58.98 and variance of 128.29 for ML; mean of 58.39 and variance of 127.05 for PLSe2). With a correctly specified model, the expected model rejection rate at $\alpha = .05$ should be controlled at around 5%. Given 300 replications, the 95% confidence interval for the observed percentage of rejected model is between 2.5% and 7.5%. For both ML and PLSe2, the observed model rejection rates at all sample sizes are between 4% and 6%, well within the 95% confidence interval.

While the minimum fit function chi-square $c_1$ is performing as expected with ML and PLSe2, PLSe1's $c_1$ test statistic based on a 1-step improvement from LISREL is much too large (mean around 70) for the expected degrees of freedom (59). We also found that different programs (e.g. EQS) give different values for $c_1$.

While in theory that starting from a consistent estimator, a single step Newton-Raphson improvement is going to make the resulting estimator asymptotically efficient, the resulting PLSe1 parameter estimates do not necessarily minimize the fit function at a finite sample size. Most software programs that implement ML estimation (which is used in PLSe1) are highly sophisticated and routinely employ different numerical methods for line-search and for controlling step-length (e.g. LISREL uses quasi-Newton methods). Even if the step-size constants are the same, and that the gradient values are also the same, the fact that quasi-Newton methods only gradually build up the second-derivative matrix (as opposed to fully computing the Hessian at each cycle) again gives no assurance that the single-iteration adjustments made by LISREL and EQS will be comparable.

Although we do not suggest using $c_1$ for PLSe1 model evaluation purpose, we can employ the robust test statistics with PLSe1. The Satorra-Bentler scaled chi-square statistic ($c_3$), Browne's residual-based statistic ( $c_4$), and Yuan and Bentler's adjustment to $c_4$ ($T_{YB}$) were originally developed under the non-normality context. However, they can be used for normal data as well. This is because they do not explicitly require the

estimates to minimize a particular fitting function (which is not possible with the 1-step improvement of PLSe1), and they only require that the estimator is consistent and asymptotically normal, which PLSe1 satisfies.

Returning to Table 10, the means of the Satorra-Bentler scaled chi-square statistic ($c_3$) across each sample size are close to the expected mean of 59. The observed model rejection rate is well controlled at 6%. In comparison, Browne's residual-based statistic ($c_4$) does not perform well at $N$=200 (with an inflated mean of 90.29 and variance of 367.11). This is to be expected at small sample size given the lack of stability in the sample based asymptotic covariance matrix required to compute $c_4$. As sample sizes increases (e.g. at $N$=1000), the performance of $c_4$ statistic improves slowly. It seems that $c_4$ requires a sample size substantially larger than 1000 to achieve the expected chi-square distribution. Due to the known disadvantage of $c_4$, Yuan and Bentler (1998) proposed a modification of $c_4$ ( $T_{YB}$ ). At large sample sizes, $T_{YB}$ converges to $c_4$. This is exactly what we observed. With a small model, $T_{YB}$ indeed improves $c_4$ by reducing its model rejection rate from 74% to 3% at $N$=200. At $N$=1000, $T_{YB}$ still improves $c_4$ by reducing its mean (63.52) to 59.59 and by bringing down $c_4$'s variance (158.03) to 122.30, values that are much closer to the expected mean of 59 and variance of 118. The model rejection rate from $T_{YB}$ is 6% instead of 11% from $c_4$.

While a large model can help improve parameter estimates, we find that a more complex model leads to less stable model fit statistics. The bottom panel in Table 10

presents the results from the large model conditions under normality. First, the $c_1$ statistic for ML has inflated model rejection rates across all samples sizes (17% at $N$=200, 9% at $N$=500 and 8% at $N$=1000, all outside of the theoretical 95% confidence interval of 2.5% and 7.5%). A similar pattern is observed again with $c_1$ from PLSe2. Although the means are well controlled around the expected values across all sample sizes, the variance and the observed model rejection rates are even more inflated when compared to $c_1$ from ML. With a large model, it would probably take a larger sample size for $c_1$ to have adequate performance.

While one should not use $c_1$ from PLSe1 for the same reasons as we discussed earlier, the robust statistics can be used to test models. The Satorra-Bentler scaled chi-square statistic ($c_3$) performs well again across sample sizes, with means very close to the expected value of 293. We observe slightly elevated variance, especially at larger sample size of $N$=1000. The model rejection rates (6% to 7%) are higher than the 5% theoretical level, although they are still controlled within the 95% confidence interval (between 2.5% and 7.5%). This is the same pattern as noted by Hu, Bentler and Kano (1992). The disadvantage of $c_4$ statistic from PLSe1 seems more obvious with a large model. For example, at $N$=200, we cannot even obtain $c_4$ (and consequently there is no $T_{YB}$ either), because our sample size is too small to accommodate such a large model (and the requisite estimation of a large asymptotic covariance matrix). At $N$=1000, both the mean (430.46) and variance (1895.89) of $c_4$ are much more inflated compared to the

expected values (mean of 293 and variance of 586) and we observe a 99% model rejection rate. As noted previously, $T_{YB}$ improves $c_4$. However, it seems to over-correct $c_4$ at $N$=500 with an underestimated variance of 269.41 and rather low model rejection rate of 2%. $T_{YB}$'s performance seems to improve at $N$=1000, but the variance is still low (450.25). Thus, it appears that a large model would require a larger sample size for stable performance of the chi-square model fit statistics.

Table 10. Comparison of observed $\chi^2$, variance of observed $\chi^2$ and observed percentage of model rejection with normal data

*Small model* (expected mean = $d = 59$; expected variance = $2d = 118$)

| | | $N$=200 | | | $N$ =500 | | | $N$=1000 | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Mean | Variance | % Reject | Mean | Variance | % Reject | Mean | Variance | % Reject |
| ML: | $c_1$ | 60.54 | 115.27 | 5% | 58.71 | 129.53 | 4% | 58.98 | 128.29 | 6% |
| PLSe1: | $c_1$ | 69.65 | 208.31 | 28% | 72.11 | 701.61 | 28% | 71.00 | 602.61 | 24% |
| | $c_3$ | 59.78 | 114.31 | 5% | 58.45 | 129.03 | 4% | 58.73 | 128.01 | 6% |
| | $c_4$ | 90.29 | 367.11 | 74% | 68.08 | 205.69 | 24% | 63.52 | 158.03 | 11% |
| | $T_{YB}$ | 61.43 | 78.75 | 3% | 59.62 | 120.80 | 7% | 59.59 | 122.30 | 6% |
| PLSe2: | $c_1$ | 58.44 | 142.02 | 6% | 57.59 | 133.14 | 4% | 58.39 | 127.05 | 6% |

*Large model* (expected mean = $d = 293$; expected variance = $2d = 586$)

| | | $N$=200 | | | $N$ =500 | | | $N$=1000 | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Mean | Variance | % Reject | Mean | Variance | % Reject | Mean | Variance | % Reject |
| ML: | $c_1$ | 308.43 | 671.79 | 17% | 300.28 | 628.39 | 9% | 296.38 | 704.97 | 8% |
| PLSe1: | $c_1$ | 338.17 | 1374.45 | 52% | 328.52 | 2100.48 | 38% | 320.83 | 2412.18 | 31% |
| | $c_3$ | 296.08 | 607.50 | 7% | 295.24 | 600.59 | 6% | 293.98 | 711.24 | 7% |
| | $c_4$ | N/A* | N/A* | N/A* | 775.93 | 11765.01 | 100% | 430.46 | 1895.98 | 99% |
| | $T_{YB}$ | N/A* | N/A* | N/A* | 301.94 | 269.41 | 2% | 300.10 | 450.25 | 6% |
| PLSe2: | $c_1$ | 307.56 | 1807.02 | 23% | 300.98 | 917.40 | 13% | 297.42 | 860.15 | 12% |

Note: The $\chi^2$ statistics include the minimum fit function chi-square statistic ($c_1$), the Satorra-Bentler scaled chi-square statistic ($c_3$), Browne's residual-based statistic ($c_4$) and $T_{YB}$ which is Yuan and Bentler's (1997) adjustment to $c_4$.

* Note: Browne's residual chi-square $c_3$ and subsequently $c_4$ are not printed at $N$=200 with large model. This is because the number of rows and columns of the asymptotic covariance matrix for the large model is equal to $\frac{m(m+1)}{2} = 351$, which is larger than the sample size.

Now we turn to the non-normal conditions presented in Table 11. Recall that in these conditions, the common and unique factors are dependent which makes the asymptotic robustness properties of normal-theory based methods irrelevant. According to Hu, Bentler and Kano (1992), the normal-theory methods (e.g., ML) under these conditions will always reject the true model even at the largest sample sizes. Our observation with both small and large model showed that $c_1$ from ML indeed performs poorly under the non-normal conditions and the performance deteriorates as sample size increases. With a small model, observed model rejection rate increased from 92% at $N$=200 to 100% at $N$=2500. With large model, observed model rejection rate is 100% for all sample sizes.

For the robust statistics from PLSe1, we observe a similar pattern as in the normal conditions. The $c_3$ statistic seems to behave well in terms of the mean being close to the expected values across all sample sizes. Unlike the normal data conditions where its variance is slightly elevated as sample size increases, under non-normality, its variances seems to be lower than the expected value. We also observe a slight over-rejection at $N$=200 with large model. However, based on only 300 replications, we are unwilling to over-interpret these results of $c_3$, except noting the general fact that its mean is well controlled – as expected for a single-moment adjustment method.

Essentially the same patterns for $c_4$ and $T_{YB}$ are observed in the non-normal conditions as with the normal conditions, where $c_4$'s means and variances are inflated

across all sample sizes, but $c_4$'s performance improves as sample size increases (albeit slowly). $T_{YB}$ is able to correct $c_4$ by bringing its means, variances, and model rejection rates to more tolerable levels. At $N$=2500 with a small model, we observe a decent performance with $c_4$ that requires little correction from $T_{YB}$. Similar to the case of normal data, we found that $T_{YB}$ sometimes over-corrected $c_4$ in terms of its variances and model rejection rates, especially with the large model and small sample size.

Overall, we find that the minimum fit function chi-square statistic ($c_1$) for ML and PLSe2 works well with normal data, but under non-normality, $c_1$ for ML cannot be trusted. Due to a lack of clarity on PLSe1's one-step implementation in LISREL, we do not recommend the use of $c_1$. Instead one can use the robust statistics for model evaluation which works well with both normal and non-normal data. The Satorra-Bentler scaled chi-square statistic ($c_3$) performs well under virtually all conditions with estimated means very close to the expected values. It has a tendency to over-reject the model at the smallest sample size, but that quickly goes away as $N$ increases. Browne's residual-based statistic ( $c_4$) performs well only when sample size becomes extremely large with a relatively small model. The inflated means, variances and model rejection rates of $c_4$ statistic can be corrected with Yuan and Bentler's (1997) adjustment ( $T_{YB}$), we also observed that $T_{YB}$ tends to over-correct $c_4$'s variances and model rejection rate at smaller sample sizes.

Table 11. Comparison of observed $\chi^2$, variance of observed $\chi^2$ and observed percentage of model rejection with non-normal data

*Small model* (expected mean = $d = 59$; expected variance = $2d = 118$)

| | | $N$=200 | | | $N$ =500 | | | $N$=1000 | | | $N$=2500 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Mean | Variance | % Reject | Mean | Variance | % Reject | Mean | Variance | % Reject | Mean | Variance | % Reject |
| ML: | $c_1$ | 110.12 | 868.00 | 92% | 127.09 | 1267.63 | 95% | 136.18 | 2029.34 | 98% | 144.16 | 2510.63 | 100% |
| PLSe1: | $c_1$ | 130.35 | 1787.85 | 96% | 160.10 | 17717.86 | 98% | 166.81 | 4735.25 | 100% | 166.63 | 4802.72 | 100% |
| | $c_3$ | 60.17 | 100.03 | 5% | 59.95 | 119.28 | 5% | 59.86 | 87.61 | 4% | 59.00 | 89.72 | 2% |
| | $c_4$ | 90.49 | 314.63 | 75% | 70.61 | 156.34 | 26% | 64.63 | 124.54 | 13% | 61.44 | 105.99 | 6% |
| | $T_{YB}$ | 61.60 | 68.96 | 1% | 61.63 | 90.58 | 5% | 60.59 | 95.94 | 5% | 59.93 | 95.81 | 4% |

*Large model* (expected mean = $d = 293$; expected variance = $2d = 586$)

| | | $N$=200 | | | $N$ =500 | | | $N$=1000 | | | $N$=2500 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Mean | Variance | % Reject | Mean | Variance | % Reject | Mean | Variance | % Reject | Mean | Variance | % Reject |
| ML: | $c_1$ | 560.94 | 10668.63 | 100% | 611.48 | 15720.93 | 100% | 658.92 | 23377.85 | 100% | 737.00 | 39994.19 | 100% |
| PLSe1: | $c_1$ | 625.77 | 18029.58 | 100% | 674.40 | 24405.45 | 100% | 729.28 | 45989.79 | 100% | 788.88 | 51212.48 | 100% |
| | $c_3$ | 302.23 | 509.61 | 7% | 297.18 | 507.83 | 5% | 294.35 | 555.33 | 4% | 293.12 | 577.7827 | 4% |
| | $c_4$ | N/A* | N/A* | N/A* | 757.73 | 7290.28 | 100% | 428.37 | 1501.87 | 100% | 338.57 | 646.7516 | 57% |
| | $T_{YB}$ | N/A* | N/A* | N/A* | 299.60 | 180.41 | 1% | 299.21 | 356.78 | 3% | 297.98 | 388.7942 | 3% |

Note: The $\chi^2$ statistics include the minimum fit function chi-square statistic ($c_1$), the Satorra-Bentler scaled chi-square statistic ($c_3$), Browne's residual-based statistic ($c_4$) and $T_{YB}$ which is Yuan and Bentler's (1997) adjustment to $c_4$.

* Note: Browne's residual chi-square $c_3$ and subsequently $c_4$ are not printed at $N$=200 with large model. This is because the number of rows and columns of the asymptotic covariance matrix for the large model is equal to $\frac{m(m+1)}{2} = 351$, which is larger than the sample size.

# CHAPTER 5

DISCUSSION

## 5.1. Summation

In this dissertation, we have proposed two efficient PLS estimation methods. Both estimators take Dijkstra's (2011) PLS consistent estimator as the starting point. In PLSe1, we use a 1-step improvement based on PLSc-estimated factor loadings and TSLS-estimated structural parameters. The 1-step improvement leads to a more efficient estimator. It also leads to standard error estimates and model fit test statistics. In PLSe2, we use PLSc-implied covariance matrix and combine it with Browne's (1974) GLS estimation. We provide a new member to the general GLS family and derive the standard errors and model fit test statistics under normality.

It is possible to extend PLSe1 to handle data that are non-normally distributed. Specifically, we use a sample-based asymptotic covariance matrix and adopt the Satorra-Bentler (1994) adjustment to correct normal theory test statistics and standard error estimates for the impact of non-normality. We also examined Browne's (1984) residual based test statistic. While Browne's statistic may not perform as well in small sample situations, we adopt Yuan and Bentler's (1997) adjustment that improves its performance.

The primary motivation for the PLSe estimators comes from the recognition that PLS remains a useful approach either for obtaining starting values for more elaborate estimation methods, such as ML; and, in practical settings, PLS often provides, particularly with Dijkstra's (2011) adjustment, remarkably good point estimates of factor loadings and structural parameters. The major drawbacks of PLS and PLSc were the unavailability of convenient standard errors (the bootstrap not withstanding as it can be computationally demanding and does not work well under non-normality) and the lack of model fit chi-square statistics. We have addressed both deficiencies in the current research. We further demonstrated that the approach can be extended to non-normal situations.

We empirically validated the methods using Monte Carlo simulations. We generated data under a non-recursive structural equation model with five latent variables with either 13 or 26 observed variables. We investigated the performance of the proposed estimators relative to the classical ML estimator under a variety of sample sizes for both normal data and non-normal data. We showed that under normality, the proposed PLSe1 and PLSe2 estimators provide estimates that are almost as good as the classical Maximum Likelihood estimator, which is theoretically asymptotically optimal under normality. We also demonstrated that the standard error estimates for both of these approaches closely correspond to the empirical Monte Carlo variation.

We simulated data under non-normality. As expected, Maximum Likelihood estimators perform well, but the standard error estimates and model test statistics do not perform well under non-normality. Its standard errors are too small relative to Monte Carlo variability. Its model fit test statistic rejects the model far too often. With non-normal adjustments, PLSe1 performs favorably. While there are some variation in terms of the relative merits of Browne's (1984) residual-based statistic, Yuan and Bentler's (1997) adjustment and Satorra-Bentler's (1994) scaled chi-square statistic under different settings, we note that all three appear to produce asymptotically correct inference – with Satorra-Bentler's (1994) scaled chi-square statistic clearly standing out. We did not investigate the performance of PLSe2 under non-normality, because it was not derived under non-normal theory.

### 5.2. Limitations and Future Directions

We only investigated the conditions where the model fits the population covariance structure exactly. In empirical research, the scientist's model can never be identical to true model in the population (MacCallum, 2003; MacCallum & Tucker, 1991). The performance of PLSc under model error and the subsequent PLSe improvements under model error remains a topic for future investigations.

Theoretically, the PLSe1 estimator we propose is given in eq. (31). However, practical implementations of (31) can be done in different ways, so that (31) really

describes a class of methods. We accepted the LISREL approach to approximating $H(\boldsymbol{\theta})$ , the Hessian matrix and in the use of a possible step size to determine the 1-step improved estimates. Research is needed to determine which of three possibilities -- an estimated information matrix, a symbolic second derivative matrix, or a computationally approximated second derivative matrix -- might improve the behavior of (31) in small samples.

In our simulations, we evaluated PLSe2 only under normal theory conditions for which it is well-justified. In principle, it should be possible to "robustify" PLSe2 in the same way that ML is typically made robust. We would expect that its point estimator should be consistent under non-normality, its standard errors should be correctable via the usual sandwich computations, and the test statistic should correctable via Satorra-Bentler and related corrections. Whether these expectations hold in practice will be the focus of followup research.

In this research, we only investigated one kind of structural equation model. Specifically, the model has a fixed set of latent variables where the only variability in terms of model complexity stems from the number of observed variables per latent variable. In future studies, we could investigate different types of models. For example, performance of PLSe under standard correlated factors model may be quite different when compared to a higher-order model.

The smallest sample size we have examined is 200. This lower limit is partly restricted by numerical stability and the proportion of invalid solutions that we empirically observed. However, with a different kind of covariance structural model, the relative importance of model size and sample size needs to be addressed and different conclusions about smallest $N$ for stable estimation might arise.

The level of non-normality investigated in the current study is somewhat contrived. While the observed variables vary in terms of skewness and kurtosis, those were held constant across all conditions in the simulation. Whether the advantages of PLSc or PLSe estimators remain in other non-normal situations, e.g. when data are categorical, is still an open question.

Finally, we only examined models with a linear structural model, i.e. the model is linear in the latent variables. We did not examine polynomials, interactions or nonlinear relationships among the latent variables. Dijkstra (2011, 2012) examined models with nonlinear and polynomial effects. It is well known in structural equation modeling that when latent variables interactions exist (Marsh, Wen & Hau, 2004) estimation of structural equation models with nonlinear terms becomes more complex. Whether PLSc and the subsequent PLSe improvements continue to hold advantages remains to be seen. While the current model is based on an early and classical study involving structural equation modeling, it remains a task of model builders to make such tools available to researchers analyzing psychological data. The ultimate value of

the current approaches can only be validated with time and empirical applications in solving substantive psychological problems.

## References


Bagozzi, B. (1980). *Causal Models in Marketing*. New York: John Wiley & Sons, Inc.

Bentler, P. M., & Weeks, D. G. (1980). Linear structural equations with latent variables. *Psychometrika*, *45*, 289-308.

Bentler, P. M. (1983). Simultaneous equation systems as moment structure models: With an introduction to latent variable models. *Journal of Econometrics*, *22*, 13-42.

Bentler, P. M. (2006). *EQS 6 for Windows program manual.* Encino, CA: Multivariate Software.

Bentler, P. M. (2007). Can scientifically useful hypotheses be tested with correlations? *American Psychologist*, *62*, 772-782.

Bentler, P. M., & Dijkstra, T. (1985). Efficient estimation via linearization in structural models. In P. R. Krishnaiah (Ed.), *Multivariate analysis VI* (pp. 9-42). Amsterdam: North-Holland.

Browne, M. W. (1974). Generalized least squares estimators in the analysis of covariance structures. *South African Statistical Journal*, *8*, 1–24.

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

Browne, M.W. & Arminger, G. (1995). Specification and estimation of mean- and covariance-structure models. In G. Arminger, C.C. Clogg & M.E. Sobel (Eds.), *Handbook of statistical modeling for the social and behavioral sciences* (pp. 311-359). New York: Plenum Press.

Cassel, C., Hackl, P., & Westlund, A. (1999). Robustness of partial least-squares method for estimating latent variable quality structures. *Journal of Applied Statistics, 26*, 435–446.

Cassel, C., Hackl, P., & Westlund, A. (2000). On measurement of intangible assets: a study of robustness of partial least squares. *Total Quality Management, 11*, 897–907.

Chin, W. W. (1998). The partial least squares approach for structural equation modeling. *in* G. A. Marcoulides (Ed.), *Modern methods for business research* (pp. 295–236). London: Lawrence Erlbaum Associates.

Chin, W. W., & Newsted, P. R. (1999). Structural equation modeling analysis with small samples using partial least squares. In R. H. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 307–341). Thousand Oaks, CA: Sage.

Cudeck, R. (1989). Analysis of correlation matrices using covariance structure models. *Psychological Bulletin, 105*, 317-327.

Curran, P. J. , West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods, 1*, 16-29.

Dijkstra, T. K. (1981). "Latent Variables in Linear Stochastic Models", PhD-thesis, its second edition was published in 1985 by Sociometric Research Foundation, Amsterdam.

Dijkstra, T. K. (1983). Some comments on Maximum Likelihood and Partial Least Squares Methods, *Journal of Econometrics, 22*, 67-90.

Dijkstra, T. (1985). *Latent Variables in Linear Stochastic Models: Reflections on Maximum Likelihood and Partial Least Squares Methods.* Second Edition, Amsterdam, The Netherlands: Sociometric Research Foundation.

Dijkstra, T. K. (2010). Latent variables and indices. In: Vinzi, V. E., Chin, W. W., Henseler, J., Wang, H. (eds.), *Handbook of Partial Least Squares*, chapter 1, pp. 23-46. Springer-Verlag, Heidelberg.

Dijkstra, T. K. (2011). Consistent Partial Least Squares estimators for linear and polynomial factor models. A report of a belated, serious and not even unsuccesful attempt. *Unpublished manuscript.*

Dijkstra, T. K. (2012). Consistent Partial Least Squares for Nonlinear Structural Equation Models. *Unpublished manuscript.*

Fornell, C., & Bookstein, F. L. (1982). Two structural equation models: LISREL and PLS applied to consumer exit-voice theory. *Journal of Marketing Research, 19*, 440–452.

Fleishman, A. I. (1978). A method for simulating non-normal distributions. *Psychometrika, 43(4)*, 521-532.

Frank, I. E. & Friedman, J. H. (1993). A Statistical View of Some Chemometric Regression Tools, *Technometrics 35*, 109-135.

Hastie, T., Tibshirani, R., & Friedman, J. (2002). The elements of statistical learning, Springer Verlag, New York.

Hansen, L. P. (1982). Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica, 50(4)*, 1029-054.

Hu, L., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin*, *112*, 351-362.

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. Esposito Vinzi et al. (eds.), *Handbook of partial least squares: Concepts, methods and applications*. *Springer handbooks of computational statistics*. New York.

Jennrich, R. I. (1970). An asymptotic chi-square test for the equality of two correlation matrices. *Journal of the American Statistical Association*, *65*, 904-912.

Jöreskog, K.G. (1969). A general approach to confirmatory maximum-likelihood factor analysis. *Psychometrika, 34*, 183-202.

Jöreskog, K. G. (1973). A general method for estimating a linear structural equation system, in: A.S. Goldberger and O.D. Duncan, eds., *Structural equation model in the social sciences* (pp. 85-l 12). New York: Seminar Press.

Jöreskog, K. G. (1977). Structural equation models in the social sciences: Specification, estimation and testing, in: P.R. Krishnaiah, ed., *Applications of statistics*. North-Holland, Amsterdam.

Jöreskog, K. G. (1978). Structural analysis of covariance and correlation matrices. *Psychometrika*, *43*(4):443–477.

Jöreskog, K. G., & Sörbom, D. (1994). *LISREL 8 user's reference guide.* Mooresville, IN: Scientific Software.

Jöreskog, K. G., & Sörbom, D. (1981). LISREL V. Analysis of linear structural relationships by maximum likelihood and least squares method. Research reports 81-8. Department of Statistics, University of Uppsala.

Jöreskog, K. G., & Wold, H. (1982). The ML and PLS techniques for modeling with latent variables: Historical and comparative aspects. In K. G. Jöreskog & H. Wold (Eds.), *Systems under indirect observation: Causality, structure, prediction, Part I,* Amsterdam: North-Holland, 263–270.

Keesling, J. W. (1972), "Maximum Likelihood Approaches to Causal Analysis," Ph.D. dissertation, University of Chicago, 1972.

Lee, S. –Y., & Jennrich, R. I. (1979). A study of algorithms for covariance structure analysis with specific comparisons using factor analysis. *Psychometrika*, *44*, 99-113.

Mann, H.B. & Wald, A. (1943). On stochastic limit and order relationships. *The Annals of Mathematical Statistics, 14* (3), 217-226.

Mardia, K.V., Kent, J.T. & Bibby, J.M. (1979). *Multivariate Analysis. Chapter 7 Econometrics.* London: Academic Press. 185-212.

Marsh, H.W., Wen, Z., & Hau, K-T. (2004). Structural equation models of latent interactions: Evaluation of alternative estimation strategies and indicator construction. *Psychological Methods, 9*, 275-300.

Maruyama, G., & McGarvey, B. (1980). Evaluating causal models: An application of maximum-likelihood analysis of structural equations. *Psychological Bulletin, 87*(3), 502-512.

MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology, 51,* 201–226.

MacCallum, R. C. (2003). Working with imperfect models. *Multivariate Behavioral Research, 38(1)*, 113-139.

MacCallum, R. C., & Tucker, L. R (1991). Representing sources of error in the common factor model: Implications for theory and practice. *Psychological Bulletin*, *109*, 502-511.

McArdle, J. J. & McDonald, R. P. (1984). Some Algebraic Properties of the Reticular Action Model. *British Journal of Mathematical and Statistical Psychology*, *37*, 234–251.

McDonald, R. P. (1996). Path analysis with composite variables. *Multivariate Behavioral Research, 31*, 239–270.

Rao, C.R. (1973). *Linear Statistical Inference and its Application.* Wiley, New York.

Rice, J. A. (1995). *Mathematical Statistics and Data Analysis.* Belmont, CA: Duxbury Press.

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399-419). Thousand Oaks, CA: Sage.

Savalei, V., & Bentler, P. M. (2006). Structural equation modeling. In R. Grover & M. Vriens (Eds.), *The handbook of marketing research: Uses, misuses, and future advances* (pp. 330-364). Thousand Oaks, CA: Sage.

Schneeweiss, H. (1990). Models with latent variables: LISREL versus PLS. *Contemporary Mathematics, 112*, 33-40.

Steiger, J. H., & Lind, J. C. (1980). Statistically based tests for the number of common factors. Paper presented at the annual meeting of the Psychometric Society. Iowa City, Iowa.

Steiger, J. H., Shapiro, A., & Browne, M. W. (1985). On the multivariate asymptotic distribution of sequential chi-square statistics. *Psychometrika, 50*, 253–264.

Stone, M., & Brooks, R. J. (1990). Continuum regression: cross-validated sequentially constructed prediction embracing ordinary least squares, partial least squares and principal components regression. *Journal of the Royal Statistical Society, Series B (Methodological) 52*: 237-269.

Tenenhaus, M.,Vinzi, V. E., Chatelin, Y-M. & Lauro, C. (2005). PLS path modeling. *Computational Statistics & Data Analysis, 48*, 159-205.

van der Vaart, A.W., (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge. 1–496.

Vale, C. D. & Maurelli, V. A. (1983). Simulating multivariate nonnormal distributions. *Psychometrika, 48(3*), 465-471.

Vilares, M.J., Almeida, M.H. & Coelho, P. S. (2010) Comparison of likelihood and PLS estimators for structural equation modeling: a simulation with customer satisfaction data. . In V. Esposito Vinzi et al. (eds.), *Handbook of partial least squares: Concepts, methods and applications*. *Springer handbooks of computational statistics*. New York.

Vinzi, V.E., Trinchera, L. & Amato, S. (2010). PLS Path Modeling: From Foundations to Recent Developments and Open Issues for Model Assessment and Improvement. In: Vinzi, V. E., Chin, W. W., Henseler, J., Wang, H. (eds.), *Handbook of Partial Least Squares*, chapter 2, pp. 47-82. Springer-Verlag, Heidelberg.

White, H. (1982). Maximum Likelihood Estimation of Misspecified Models. *Econometrica, 50*, 1-25.

Wiley, D. E. (1973). The Identification Problem for Structural Equation Models with Unmeasured Variables. In A. S. Goldberger and O. D. Duncan (eds.), *Structural Equation Models in the Social Sciences*. New York: Seminar Press, 69–83.

Wold, H. O. A. (1966). Nonlinear estimation by iterative least squares procedures, in David, F. N. (ed.), *Research Papers in Statistics, Festschrift for J. Neyman,* Wiley, New York, 411-444.

Wold, H. O. A. (1975). Path Models with Latent Variables: The NIPALS Approach, in Blalock, H. M., Aganbegian, A., Borodkin, F. M., Boudon, R., Capecchi, V., Quantitative Sociology, Academic Press, New York, 307-359.

Wold, H. O. A. (1982). Soft modeling: the basic design and some extensions, in Joreskog, K. G., Wold, H. O. A. (eds), *Systems under indirect observation, Part II*, North-Holland, Amsterdam, 1-55.

Yuan, K. –H., & Bentler, P. M. (1997). Mean and covariance structure analysis: Theoretical and practical improvements. *Journal of the American Statistical Association*, *92*, 767-774.

Yuan, K. –H., & Bentler, P. M. (1998). Normal theory based test statistics in structural equation modeling. *British Journal of Mathematical and Statistical Psychology*, *51*, 289-309.

Yuan, K. –H. & Bentler, P. M. (1999). F tests for mean and covariance structure analysis. *Journal of Educational and Behavioral Statistics*, *24*, 225-243.

Yuan, K.-H., & Hayashi, K. (2003). Bootstrap approach to inference and power analysis based on three test statistics for covariance structure models. *British Journal of Mathematical and Statistical Psychology 56 (1)*, 93-110.