NON-LINEAR LATENT VARIABLE MODELS: A STUDY OF FACTOR SCORE APPROACHES

Submitted by
Johan Vegelius

A thesis submitted to the Department of Statistics in partial fulfillment of the requirements for a two-year Master degree in Statistics in the Faculty of Social Sciences

Supervisors
Shaobo Jin & Fan Yang Wallentin

Spring, 2017
Abstract

Non-linear latent variable models are associated with problems which are difficult to handle in applied sciences. Four methods for estimating factor scores, with the purpose of estimating latent variable models with an interaction term, were investigated. The LISREL procedure provided inconsistent estimates of the interaction term for all sample sizes and distributions of the latent exogenous variables. The Bartlett-Thompson approach yielded consistent estimates only when the distribution of the latent exogenous variables was normal, whereas the Hoshino-Bentler and adjusted LISREL approaches yielded consistency for all distributions of the latent exogenous variables. In the Bartlett-Thompson and LISREL approaches the interaction term is formed from multiplying latent variable scores, whereas in the Hoshino-Bentler and adjusted LISREL approaches the interaction term is treated as yet another factor which is freely estimated. It was, hence, concluded that the methods treating the interaction term as a factor were more appropriate (in terms of consistency and robustness) than those using products of factor scores for estimating the latent variable model.

Keywords: Structural Equation Model, Interaction, Monte Carlo
Contents

1 Introduction ............................................ 1
  1.1 Structural Equation Models (SEMs) ...................... 1
  1.2 Non-Linear Latent Variable Models (Non-Linear LVMs) .......... 1
  1.3 Problems with Non-Linear LVMs ......................... 2
  1.4 Research Aim ........................................ 3

2 Method .................................................. 4
  2.1 Factor Score Regression in Linear SEM .................... 4
    2.1.1 The Bartlett-Thompson Approach ...................... 4
    2.1.2 The LISREL Approach ................................ 5
    2.1.3 The Hoshino-Bentler Approach ....................... 6
  2.2 Factor Score Regression in Non-Linear LVM ................. 6
    2.2.1 The Bartlett-Thompson Approach ...................... 6
    2.2.2 The LISREL Approach ................................ 7
    2.2.3 The Hoshino-Bentler Approach ....................... 7
    2.2.4 The Adjusted LISREL Approach ....................... 9

3 Simulation Study ....................................... 10
  3.1 Simulation Design .................................... 10
  3.2 Measure of Consistency .................................. 11

4 Results ................................................ 12

5 Discussion and Conclusions .............................. 16
1 Introduction

1.1 Structural Equation Models (SEMs)

Structural equation modelling (SEM) is a powerful tool in the study of relations between latent variables (factors), particularly in social sciences (Ajzen, 1987; Bollen, 2002). When a set of indicators \((x, y)\) are measured, it is possible to imagine an underlying structure consisting of latent variables \((\xi, \eta)\). A commonly considered SEM model is the following:

\[
x = \mu_x + \Lambda_x \xi + \delta,
\]
\[
y = \mu_y + \Lambda_y \eta + \epsilon,
\]
\[
\eta = \gamma' \xi + \zeta,
\]

where \(\Lambda_x\) and \(\Lambda_y\) are factor loading matrices, \(\gamma\) is a parameter vector defining a linear relation between \(\xi\) and \(\eta\). \(\delta\), \(\epsilon\) and \(\zeta\) are error terms. Equations 1 and 2 are measurement models and equation 3 is the structural model.

1.2 Non-Linear Latent Variable Models (Non-Linear LVMs)

Equation 3 is valid only if the relationship between \(\xi\) and \(\eta\) is linear. It is, however, easy to imagine situations where it is appropriate to consider non-linear relationships between latent variables. For example in the Theory of Planned Behavior (Ajzen, 1987), the two latent variables Attitude Towards the Behavior (Attitude) and the Perceived Behavioral Control (PBC) affect the latent variable Intention, which in turn, affects the Behavior which is observed. Further, the Theory of Planned Behavior suggests that the effect of Attitude on Intention depends on the Perceived Behavioral Control. In order to incorporate such an effect, the interaction term \(\text{Attit} : \text{PBC}\) is introduced in the structural model. Figure 1 illustrates the path diagram of the Theory of Planned Behavior and the corresponding structural equation is

\[
\eta = \gamma_1 \xi_1 + \gamma_2 \xi_2 + \omega_{21} \xi_1 \xi_2 + \zeta,
\]

where \(\xi_1\) is Attitude and \(\xi_2\) is PBC. Equation 4 can be rewritten as

\[
\eta = \gamma_1 \xi_1 + (\gamma_2 + \omega_{21} \xi_1) \xi_2 + \zeta,
\]

\(^1\)In this investigation only models with one latent endogenous variable will be investigated for the purpose of clarity, although the results are applicable to more general situations.
which indicates that the effect of $\xi_2$ on $\eta$ depends on the value of $\xi_1$. If, for example, $\omega_{21} \geq 0$, the effect of $\xi_2$ on $\eta$ will increase as $\xi_1$ increases. In order to incorporate such relations between latent variables, equation 3 can be reformulated as

$$\eta = \gamma'\xi + \xi'\Omega\xi + \zeta,$$

where parameters corresponding to squared and interaction terms are contained in the lower-triangular matrix $\Omega$. For example, in equation 4

$$\Omega = \begin{pmatrix} 0 & 0 \\ \omega_{21} & 0 \end{pmatrix}.$$

Figure 1: Structural model of the Theory of Planned Behavior.

### 1.3 Problems with Non-Linear LVMs

Non-linear LVMs are associated with some issues, which make LVMs more demanding than linear SEMs. Firstly, even if $\xi$ and $x$ are normally distributed, $\eta$ and $y$ are not. Secondly, $E(\eta) = \text{tr}(\Omega\Phi) \neq 0$, in the non-linear LVMs, where $\Phi$ is the covariance matrix of $\xi$. Hence, the mean vector of the measurement model and the sample covariance matrix need to be estimated simultaneously. Accordingly, many methods have been proposed for estimating non-linear LVMs. Kenny and Judd (1984) first proposed a product indicator method, further developed by Yang-Jonsson (1997) and Kelava and Brandt (2009). Products of indicators are
used as indicators of quadratic and interaction terms of latent variables. An advantage with these methods is that standard SEM packages can be used for estimation of non-linear LVMs. However, due to the non-normal distributions of $y$ and $\eta$ the distributional assumption is violated as recognized by Jöreskog and Yang (1996) and, hence, standard errors of interaction and quadratic terms will in general be underestimated. Further, Yang-Jonsson (1997) derived all constraints induced by the distributional assumption and the model, which are difficult for applied researchers. Consequently, Marsh et al. (2004) suggested to lift those assumptions. It has been shown that the approach by Marsh et al. (2004) produces consistent estimates and is robust against distributional misspecification.

Bollen and Paxton (1998) developed a two-stage least squares procedure of estimating LVMs with quadratic and interaction terms. This approach is, however, inefficient and it is difficult to find instruments for more general functions of the latent variables.

There are method of moments approaches (Wall and Amemiya, 2000), where more general polynomials of latent variables can be incorporated. However, higher order moments need to be estimated leading to unstable small sample properties. Furthermore the procedure is difficult to implement and needs to be investigated further in terms of efficiency.

Distribution analytic methods were developed by Klein and Moosbrugger (2000) and further optimized in terms of computational efficiency by Klein and Muthén (2007). In the former, Gauss-Hermite quadrature is used to model the non-normal distribution of indicator variables, thus it is not applicable when the dimension of $\xi$ is large. Both methods rely on normally distributed $\xi$ and are subject to bias if the normal assumption is violated.

There are also Bayesian approaches (Zhu and Lee, 1999; Arminger and Muthén, 1998) which are computationally heavy and not fully investigated in terms of sensitivity to violations of distributional assumptions.

To summarize, there are no robust (against distributional misspecification) and simply usable approaches to estimate non-linear LVMs available at the present stage, except Marsh et al. (2004). The need for such a method is substantial and addressed in this thesis.

1.4 Research Aim

The aim of this thesis is to investigate methods of estimating $\gamma$ and $\Omega$ in equation 6. We restrict our attention to the factor score regression approach. The factor score regression approach can be divided into three steps:
1. Estimate parameters in the measurement model.

2. Use the estimated parameters in the previous step to calculate the factor scores.

3. Estimate the parameters in the structural model by ordinary least squares (OLS) regression analysis, using the factor scores.

As demonstrated by Skrondal and Laake (2001), factor score regression, using OLS, does not always provide consistent estimates of $\gamma$ in equation 3. Thus, in this study we wish to provide a method, easily implementable, in existing softwares for consistent estimation of the structural parameters in equations 3 and 6.

The rest of the paper is organized as follows. In the next section, factor score regression in linear SEM is discussed, followed by non-linear LVM. Next, the simulation design is discussed, followed by results and discussions.

2  

Method

In this section, different factor score regression approaches are presented.

2.1  

Factor Score Regression in Linear SEM

2.1.1  

The Bartlett-Thompson Approach

In the case of the linear model 3, Skrondal and Laake (2001) proposed a method to consistently estimate $\gamma$. In the first step, the parameters in equation 1 are estimated and Thompson scores of $\xi$ are obtained as

$$
\hat{\xi}_i = \left( \hat{\Phi}_{\xi}^{-1} + \hat{\Lambda}_x \hat{\Theta}_\delta^{-1} \hat{\Lambda}_x \right)^{-1} \hat{\Lambda}_x \hat{\Theta}_\delta^{-1} (x_i - \hat{\mu}_x),
$$

for individual $i$, where $\hat{\Phi}_{\xi}$ is the estimated covariance matrix of $\xi$, $\hat{\Lambda}_x$ the estimated factor loading matrix of the measurement model in equation 1, $\hat{\Theta}_\delta$ the estimated covariance matrix of the error term $\delta$, $\hat{\mu}_x$ the estimated mean vector in equation 1 and $x_i$ the observed indicator vector of individual $i$ in equation 1. In the second step, the unknown parameters in equation 2 are estimated and the Bartlett scores of $\eta$ are obtained as

$$
\hat{\eta}_i = \left( \hat{\Lambda}_x \hat{\Theta}_\delta^{-1} \hat{\Lambda}_x \right)^{-1} \hat{\Lambda}_x \hat{\Theta}_\delta^{-1} (y_i - \hat{\mu}_y),
$$

(9)
for individual \( i \), where \( \hat{\Lambda}_y \) is the estimated factor loading matrix of the measurement model in equation 2, \( \hat{\Theta}_e \) the estimated covariance matrix of the error term \( e \), \( \hat{\mu}_y \) the estimated mean vector in equation 2 and \( y_i \) the observed indicator vector of individual \( i \) in equation 2. Both Bartlett and Thompson scores can be readily obtained from existing SEM packages. In the third step, \( \hat{\eta} \) is regressed on \( \hat{\xi} \) using OLS. Skrondal and Laake (2001) proved that the Bartlett-Thompson score approach provides consistent estimation of \( \gamma \). Using Bartlett scores alone produce biased estimates, as well as using Thompson scores alone. However, the respective biases have opposite signs and when the combination is used, the biases cancel.

### 2.1.2 The LISREL Approach

A procedure for calculating factor scores, implemented in LISREL (Jöreskog and Söderbom, 1988), was proposed by Anderson and Rubin (1956). The idea behind this approach is to produce factor scores with equal means and covariances as the corresponding estimated parameters. Consider the new measurement model

\[
x^* = \mu + \Lambda \xi^* + \delta^*,
\]

where \( x^* = (x', y') \), \( \xi^* = (\xi', \eta') \), \( \mu = E(x^*) \) and \( \delta^* = (\delta', e') \). Minimizing

\[
\sum_{i=1}^{n} (x^*_{i} - \mu - \Lambda \xi^*_{i})' \Theta^{-1} (x^*_{i} - \mu - \Lambda \xi^*_{i}),
\]

subject to

\[
\frac{1}{n} \sum_{i=1}^{n} \xi^*_i \xi^*_i' = \hat{\Phi}^*,
\]

where \( \Theta \) is the covariance matrix of \( \delta^* \) and \( \hat{\Phi}^* \) is the estimated covariance matrix of \( \xi \), leads to the factor scores in LISREL. Following the derivation by Jöreskog and Yang (2000) such factor scores can be calculated by introducing

\[
C = D^\frac{1}{2} U' \Lambda' \Theta^{-1} \left( \sum_{i=1}^{n} (x^*_{i} - \mu)(x^*_{i} - \mu)' \right) \Theta^{-1} \Lambda U D^\frac{1}{2},
\]

where \( \hat{\Phi}^* = U D U' \) is the singular value decomposition of \( \Phi^* \). Let \( Z \) be the symmetric square root of \( C \). Then the factor scores are calculated as

\[
\hat{\xi}^*_i = U D^\frac{1}{2} Z^{-1} D^\frac{1}{2} U' \Lambda' \Theta^{-1} (x^*_i - \mu),
\]

with the properties \( \text{Cov}(\hat{\xi}^*_i) = \hat{\Phi}^* \) and \( E(\hat{\xi}^*_i) = 0 \). Since the factor scores, by construction, have the same means and covariances as the true scores (asymptotically), this method provides consistent estimation of \( \gamma \) in equation 3.
2.1.3 The Hoshino-Bentler Approach

Hoshino and Bentler (2013) proposed a method using a combination of the estimated covariance matrices and Bartlett scores (or regression scores) to consistently estimate a linear structural model. Each measurement model is estimated separately. In a SEM model with $J$ measurement models, they suggest to estimate the latent variable covariance matrix by

$$
\hat{\Phi}_P = \begin{pmatrix}
\hat{\Phi}_{11} & \hat{\Phi}_{12} & \ldots & \hat{\Phi}_{1J} \\
\hat{\Phi}_{21} & \hat{\Phi}_{22} & \ldots & \hat{\Phi}_{2J} \\
\vdots & \vdots & \ddots & \vdots \\
\hat{\Phi}_{J1} & \hat{\Phi}_{J2} & \ldots & \hat{\Phi}_{JJ}
\end{pmatrix},
$$

where $\hat{\Phi}_{kk}$ is the estimated covariance matrix of the $k$:th measurement model extracted from the SEM package output. $\tilde{\Phi}_{kl}$ is the sample covariance of factors between measurement model $k$ and $l$, defined as

$$
\tilde{\Phi}_{kl} = \frac{1}{n} \sum_{i=1}^{n} (\hat{\xi}_{ik} - \hat{\nu}_k)(\hat{\xi}_{il} - \hat{\nu}_l)' = \text{Cov}(\hat{\xi}_k, \hat{\xi}_l),
$$

where $\hat{\nu}_k$ and $\hat{\nu}_l$ are factor score mean vectors. The authors have shown that this approach consistently estimates $\gamma$ in equation 3. The reason for consistency is that the factor scores are constructed as to yield the same means and covariances as the true scores.

2.2 Factor Score Regression in Non-Linear LVM

2.2.1 The Bartlett-Thompson Approach

The Bartlett-Thompson approach can be easily generalized to the non-linear model 6. The first two steps remain the same as those described above. Notice that $E(\eta) = \text{tr}(\Omega\Phi) \neq 0$. The factor scores produced by the existing software are actually $\hat{\eta} - E(\eta)$. Thus, it is necessary to introduce an intercept and consider

$$
\eta - \hat{E}(\eta) = \alpha + \gamma'\hat{\xi} + \tilde{\omega}'\hat{\xi} + \zeta.
$$

$\eta - \hat{E}(\eta)$ is now regressed on the intercept $\alpha$, $\hat{\xi}$ and $\hat{\xi}$, where $\hat{\xi}$ is a vector containing only those squares and cross products of the elements in $\hat{\xi}$ considered in the investigation. The corresponding parameters are collected in $\tilde{\omega}$. For example, in equation 4, $\tilde{\omega} = \omega_{21}$ and $\hat{\xi} = \hat{\xi}_1\hat{\xi}_2$. 

6
2.2.2 The LISREL Approach

Similar to the Bartlett-Thompson approach, products and interaction terms can be formed by multiplication among factor scores. Using the same argument as in the Bartlett-Thompson approach, an intercept is required in the structural model to be estimated according to equation 17. Various studies have shown that the quasi-maximum likelihood estimator, based on the normal assumption, in a Confirmatory Factor Analysis (CFA) model is still consistent under distributional misspecification (See among others Amemiya and Anderson, 1990; Amemiya et al., 1987; Anderson and Amemiya, 1988; Browne, 1974; Satorra, 2002). Thus, \( \mu, \Lambda, \Phi^* \) and \( \Theta \) can always be consistently estimated. However, the third-order moments, such as \( \text{Cov}(\xi_1 \xi_2, \eta) \) cannot be consistently estimated by the sample covariance of \( \hat{\xi}_1 \hat{\xi}_2 \) and \( \hat{\eta} \), so we expect this approach to produce inconsistent estimates of interaction and squared terms. Harring et al. (2012) found that the corresponding approach with a quadratic effect produced estimates, for the quadratic effect, with negative bias in small and moderate samples.

2.2.3 The Hoshino-Bentler Approach

In this approach, the elements in \( \tilde{\xi} \) are treated as new factors alongside \( \xi \) and \( \eta \). Every element in \( \tilde{\xi} \) represents quadratic and cross terms of elements in \( \xi \). The measurement models are

\[
\begin{pmatrix}
x_c \\
\tilde{x}_c
\end{pmatrix} = 
\begin{pmatrix}
0 \\
\tilde{\mu}_x
\end{pmatrix} + 
\begin{pmatrix}
\Lambda_x & 0 \\
0 & \tilde{\Lambda}_x
\end{pmatrix} 
\begin{pmatrix}
\xi \\
\tilde{\xi}
\end{pmatrix} + 
\begin{pmatrix}
\delta \\
\tilde{\delta}
\end{pmatrix},
\]

(18)

\[
y = \mu_y + \Lambda_y \eta + \epsilon,
\]

(19)

where \( x_c = (x - \mu_x) \), i.e., the mean centered vector of indicators, \( \tilde{x}_c \) is a vector of products of elements in \( x_c \) corresponding to two (for interaction terms) or one (for squared terms) of the latent variables in \( \xi \), \( \tilde{\mu}_x \) is the mean vector of \( \tilde{x}_c \), \( \tilde{\Lambda}_x \) is a matrix containing factor loadings corresponding to the effect of the latent product vector \( \tilde{\xi} \) on the indicator product vector \( \tilde{x}_c \), and lastly, \( \tilde{\delta} \) is the error vector of the product equations. For example, the measurement model for equations 1, 2 and 4 is illustrated in figure 2. The indicator variables associated with the new factor \( \tilde{\xi} = \xi_3 = \xi_1 \xi_2 \) are indicators corresponding to \( \xi_1 \) multiplied by indicators corresponding to \( \xi_2 \) such as \((x_1 - \mu_{x_1})(x_4 - \mu_{x_4}), (x_2 - \mu_{x_2})(x_5 - \mu_{x_5}) \) and \((x_3 - \mu_{x_3})(x_6 - \mu_{x_6}) \). Such a mean centering is suggested by Algina and Moulder (2001). In order for the error covariance matrix \( \Theta \) to be diagonal, each indicator can only occur once in products for indicating an interaction or squared term. Otherwise, errors between different product indicators will be correlated and
the model is misspecified. Alternatively, all possible combinations of product indicators could be included simultaneously, but then the diagonal assumption of $\Theta$ must be relaxed. Ignoring the constraints is similar to the unconstrained approach introduced by Marsh et al. (2004). The reader is referred to Marsh et al. (2004) for more discussions on choosing the cross products. Note that $\tilde{\xi}$ is treated as a factor of its own by ignoring the nature that $\tilde{\xi}$ is formed from $\xi$. Despite the possible improvements, associated with such considerations, it is more attractable for applied researchers not being forced to calculate such constraints.

![Diagram of factor models to estimate in the Hoshino-Bentler approach.](image)

**Figure 2:** Factor models to estimate in the Hoshino-Bentler approach.

The latent variable vectors are now

$$\xi^+ = \begin{pmatrix} \xi \\ \tilde{\xi} - E(\tilde{\xi}) \end{pmatrix} \text{ and } \eta - E(\eta),$$

with corresponding Bartlett-type factor scores

$$\hat{\xi}^+ = \begin{pmatrix} \hat{\xi} \\ \hat{\xi} - E(\hat{\xi}) \end{pmatrix} \text{ and } \eta - E(\eta).$$

Notice that $\hat{\xi}_3 = \hat{\xi}_1 \hat{\xi}_2$ is not necessarily the same as $\hat{\xi}_1 \hat{\xi}_2$. Using the Hoshino-Bentler estimator of the covariance matrix of latent variables for the non-linear LVM (1, 2 and 6), yields

$$\hat{\Phi}^P = \begin{pmatrix} \hat{\Phi}_{\xi \xi} & \hat{\Phi}_{\xi \eta} \\ \hat{\Phi}_{\eta \xi} & \hat{\Phi}_{\eta \eta} \end{pmatrix},$$

where $\hat{\Phi}_{\xi \xi}$ is the estimator of $\text{Cov}(\xi^+)$ directly extracted from the software outputs, $\hat{\Phi}_{\xi \eta}$ is the estimator of $\text{Cov}(\xi^+, \eta)$ estimated by the sample covariance and $\hat{\Phi}_{\eta \eta}$ is the estimator of $\text{Var}(\eta)$.
also obtained from the software outputs. Further, $\Phi_{\eta\xi} = \Phi_{\xi\eta}'$. Now, the structural parameter vector $(\gamma', \tilde{\omega}')'$ can be estimated using OLS as

$$
\begin{pmatrix}
\hat{\gamma} \\
\hat{\tilde{\omega}}
\end{pmatrix} = \Phi_{\xi\xi}^{-1} \Phi_{\xi\eta}.
$$

(23)

We expect this approach to be consistent with respect to $\gamma$ and $\Omega$, even under non-normal $\xi$ since the cross products and interactions between latent variables are now factors on their own which are part of the model and the constraints on the factor covariances and error covariances are relaxed. Means and variances of those factors will be consistently estimated.

### 2.2.4 The Adjusted LISREL Approach

In the same way as described in the previous section, $\tilde{\xi}$ is considered as a factor and contains quadratic and cross products of elements in $\xi$. The measurement model becomes

$$
\begin{pmatrix}
\tilde{x}_c^* \\
\tilde{x}_c
\end{pmatrix} = \begin{pmatrix}
0 \\
\tilde{\mu}
\end{pmatrix} + \begin{pmatrix}
\Lambda & 0 \\
0 & \tilde{\Lambda}
\end{pmatrix} \begin{pmatrix}
\xi_* \\
\tilde{\xi}
\end{pmatrix} + \begin{pmatrix}
\delta_* \\
\tilde{\delta}
\end{pmatrix},
$$

(24)

where $x_*^c = (x^* - \mu)$, i.e., the mean centred vector of indicators. Once the factor scores have been calculated in line with the procedure presented in section 2.1.2, the following equation is estimated using OLS:

$$
\eta - \overline{E(\eta)} = \gamma' \hat{\tilde{\xi}} + \tilde{\omega}' \left( \tilde{\xi} - \overline{E(\xi)} \right) + \zeta.
$$

(25)

It is worth mentioning that there is no intercept in equation 25 since the factors $\tilde{\xi} - \overline{E(\xi)}$ and $\eta - \overline{E(\eta)}$ will be constructed as to have mean zero. We expect this approach to be consistent for the same reason as the Hoshino-Bentler procedure.
3 Simulation Study

3.1 Simulation Design

The measurement model is given by

$$
\begin{pmatrix}
  x_1 \\
  x_2 \\
  x_3 \\
  x_4 \\
  x_5 \\
  x_6
\end{pmatrix} =
\begin{pmatrix}
  0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 \\
  0.5 & 0.8 & 0.7 & 0 & 0 & 0 \\
  0.5 & 0.7 & 0 & 0.8 & 0 & 0 \\
  0.5 & 0 & 1 & 0 & 0 & 0 \\
  0.5 & 0 & 0 & 0.7 & 0 & 0 \\
  0.5 & 0 & 0 & 0 & 0.7 & 0 \\
\end{pmatrix}
\begin{pmatrix}
  1 & 0 \\
  0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 \\
  0.5 & 0.8 & 0.7 & 0 & 1 & 1.5 \\
  0.5 & 0.7 & 0 & 0.8 & 0 & 1.5 \\
  0.5 & 0 & 1 & 0 & 0 & 1.5 \\
  0.5 & 0 & 0 & 0.7 & 0 & 0.7 \\
\end{pmatrix}
\begin{pmatrix}
  \xi_1 \\
  \xi_2 \\
  \xi_3 \\
  \xi_4 \\
  \xi_5 \\
  \xi_6
\end{pmatrix} +
\begin{pmatrix}
  \delta_1 \\
  \delta_2 \\
  \delta_3 \\
  \delta_4 \\
  \delta_5 \\
  \delta_6
\end{pmatrix},
$$

(26)

$$
\begin{pmatrix}
  y_1 \\
  y_2 \\
  y_3
\end{pmatrix} =
\begin{pmatrix}
  1.5 & 1.5 & 1.5 \\
  1 & 0.8 & 0.7 \\
  1 & 0.7 & 0
\end{pmatrix}
\begin{pmatrix}
  \xi_1 \\
  \xi_2 \\
  \xi_3
\end{pmatrix} +
\begin{pmatrix}
  \epsilon_1 \\
  \epsilon_2 \\
  \epsilon_3
\end{pmatrix},
$$

(27)

The covariance matrix of $\xi = (\xi_1, \xi_2)'$ is

$$
\Phi = \begin{pmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{pmatrix},
$$

(28)

and the covariance matrices of the error terms $\delta$ and $\epsilon$ are

$$
\Theta_\delta = \begin{pmatrix}
  0.25 & 0 & 0 & 0 & 0 & 0 \\
  0 & 0.25 & 0 & 0 & 0 & 0 \\
  0 & 0 & 0.25 & 0 & 0 & 0 \\
  0 & 0 & 0 & 0.25 & 0 & 0 \\
  0 & 0 & 0 & 0 & 0.25 & 0 \\
  0 & 0 & 0 & 0 & 0 & 0.25 \\
\end{pmatrix},
$$

(29)

and

$$
\Theta_\epsilon = \begin{pmatrix}
  0.35 & 0 & 0 \\
  0 & 0.35 & 0 \\
  0 & 0 & 0.35 \\
\end{pmatrix}.
$$

(30)

Finally, the structural model is

$$
\eta = 0.5\xi_1 + 0.5\xi_2 + 0.25\xi_1\xi_2 + \zeta,
$$

(31)
where the covariance of $\zeta$ is given by $\psi = 0.3$. All error terms are sampled from normal distributions. In the notation of equation 6 $\gamma' = (0.5 \ 0.5)$ and

$$\Omega = \begin{pmatrix} 0 & 0 \\ 0.25 & 0 \end{pmatrix}. \quad (32)$$

Three sample sizes were considered. $n = 200, 400$ and $3200$. To investigate the robustness against distributional misspecification, both normal and non-normal $\xi$ were considered. Non-normal $\xi$ were generated using the method in Fleishman (1978) and Vale and Maurelli (1983). All error terms $\delta, \epsilon$ and $\zeta$ were generated from normal distributions. In summary, all simulation settings are summarized in table 1. The crossproducts, used as indicators of the latent variable interaction, are $(x_1 - \mu_{x_1})(x_4 - \mu_{x_4}), (x_2 - \mu_{x_2})(x_5 - \mu_{x_5})$ and $(x_3 - \mu_{x_3})(x_6 - \mu_{x_6})$.

Table 1: Combinations of distribution properties of latent exogenous variables and sample sizes.

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Distribution Properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 200$</td>
<td>$s = 0, k = 0$</td>
</tr>
<tr>
<td>$n = 400$</td>
<td>$s = 0, k = 0$</td>
</tr>
<tr>
<td>$n = 3200$</td>
<td>$s = 0, k = 0$</td>
</tr>
<tr>
<td></td>
<td>$s = 0.4, k = 2$</td>
</tr>
<tr>
<td></td>
<td>$s = 0.4, k = 2$</td>
</tr>
<tr>
<td></td>
<td>$s = 0.8, k = 7$</td>
</tr>
<tr>
<td></td>
<td>$s = 0.8, k = 7$</td>
</tr>
</tbody>
</table>

Note: $s$ stands for skewness, $k$ for excess kurtosis and $n$ for sample size.

### 3.2 Measure of Consistency

Once $x$ and $y$ were generated, the measurement model was estimated and the factor scores were calculated according to the respective methods. Finally the factor scores were used in the structural model for estimation of the parameters $\gamma_1, \gamma_2$ and $\omega_{21}$. For each combination of sample size and $\xi$ distribution, $R = 2000$ samples were generated. The consistency of the structural parameters were investigated by examining the relative bias (RB) and the root mean square error (RMSE). If $\hat{\theta}_i$ is the estimated parameter value ($\hat{\gamma}_1, \hat{\gamma}_2$ or $\hat{\omega}_{21}$) in sample $i$ of the corresponding true parameter $\theta$, the two measures are given by

$$\text{RB} = \frac{100}{R} \sum_{i=1}^{R} \left( \frac{\hat{\theta}_i - \theta}{\theta} \right), \quad (33)$$

$$\text{RMSE} = \frac{1}{R} \sum_{i=1}^{n} \left| \frac{\hat{\theta}_i - \theta}{\theta} \right|. \quad (34)$$

11
Table 2: Percentage of samples with non-convergent solutions or estimated covariance matrices which were not positive definite.

<table>
<thead>
<tr>
<th></th>
<th>Bartlett-Thompson</th>
<th>LISREL</th>
<th>Hoshino-Bentler</th>
<th>Adj. LISREL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( s = 0, k = 0 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( n = 200 )</td>
<td>0.00</td>
<td>0.00</td>
<td>9.45</td>
<td>5.20</td>
</tr>
<tr>
<td>( n = 400 )</td>
<td>0.00</td>
<td>0.00</td>
<td>1.10</td>
<td>0.35</td>
</tr>
<tr>
<td>( n = 3200 )</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>( s = 0.4, k = 2 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( n = 200 )</td>
<td>0.05</td>
<td>0.00</td>
<td>9.75</td>
<td>5.45</td>
</tr>
<tr>
<td>( n = 400 )</td>
<td>0.00</td>
<td>0.00</td>
<td>1.80</td>
<td>0.35</td>
</tr>
<tr>
<td>( s = 3200 )</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>( s = 0.8, k = 7 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( n = 200 )</td>
<td>0.05</td>
<td>0.00</td>
<td>12.00</td>
<td>6.65</td>
</tr>
<tr>
<td>( n = 400 )</td>
<td>0.00</td>
<td>0.00</td>
<td>3.40</td>
<td>0.85</td>
</tr>
<tr>
<td>( n = 3200 )</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

*Note:* \( s \) stands for skewness, \( k \) for excess kurtosis and \( n \) for sample size.

4 Results

Each sample was checked for the following four conditions:

1. The model converges,
2. the estimated factor covariance matrix is positive definite,
3. the estimated error covariance matrix is positive definite,
4. the estimated model-implied covariance matrix is positive definite.

If any of the above conditions were not satisfied the sample was disregarded. Table 2 shows the percentage of samples which were disregarded for each method, sample size and distribution. As can be seen in table 2 very few samples were disregarded using the Bartlett-Thompson and LISREL approaches for estimating factor scores. However, a substantial fraction of the samples was disregarded using the Hoshino-Bentler and adjusted LISREL approaches when the sample

---

Note: In the Bartlett-Thompson and Bentler approaches there are two measurement models and all six matrices \( \Phi_x \), \( \Phi_y \), \( \Theta_x \), \( \Theta_y \), \( \Sigma_x \) and \( \Sigma_y \) were investigated.
size was \( n = 200 \). The fraction of disregarded samples decreases rapidly with increasing sample size. The fraction is largest for the Hoshino-Bentler procedure. When \( \xi \) departs from normality the fraction of disregarded samples increases for the Hoshino-Bentler and adjusted LISREL procedures. It is reasonable that both the Hoshino-Bentler and the adjusted LISREL approaches become problematic in terms of model estimation for small sample sizes, due to the existence of the additional factor \( \xi_1 \xi_2 \).

Figure 3: Relative bias (RB) as a function of sample size. \( s \) stands for skewness and \( k \) stands for excess kurtosis.
Figure 4: Root mean square error (RMSE) as a function of sample size. $s$ stands for skewness and $k$ stands for excess kurtosis.

Figures 3 and 4 show RB and RMSE respectively, of the three structural parameters $\gamma_1$, $\gamma_2$ and $\omega_{21}$, for three different sample sizes, three different distributions of $\xi$, for each of the four methods for estimating factor scores. Dashed lines corresponding to RB $\pm 5\%$ are included in figure 3 indicating limits of what can be considered as a low bias. All methods provide estimates of $\gamma_1$ and $\gamma_2$ with low relative bias, well within the the $\pm 5\%$ limits, regardless of sample size and distribution of $\xi$. Figure 4 indicates that the RMSE of $\gamma_1$ and $\gamma_2$ approaches zero as the sample size increases for all four methods.
When the distribution of $\xi$ is normal the Bartlett-Thompson, Hoshino-Bentler and adjusted LISREL approaches provide estimates of $\omega_{21}$ with low bias. The Bartlett-Thompson approach gives estimates with low bias for all investigated sample sizes when the distribution is normal but when departing from normality $\omega_{21}$ becomes over estimated. However, even though the adjusted LISREL approach gives slightly over estimated values of $\omega_{21}$ when the sample size is small (typically $n = 200$), the bias becomes small when the sample size increases, and this is true even when $\xi$ departs from normality. The same holds for the Hoshino-Bentler approach but the bias is smaller than with the adjusted LISREL approach for all sample sizes. Using the LISREL approach, the estimates of $\omega_{21}$ are heavily underestimated for all sample sizes and distributions, but the absolute value of the bias decreases somewhat when the sample size increases or the more the distribution of $\xi$ departs from normality.

Using the LISREL approach, the factor scores have the same mean vector and covariance matrix as the true factors on average. Only the mean and covariances affect the estimation of the linear parameters $\gamma_1$ and $\gamma_2$, but when estimating the interaction term $\omega_{21}$, higher order moments become important, which are not necessarily the same as those of the true factors, yielding inconsistent estimates of $\omega_{21}$ as is apparent from the simulation results.

The RMSE of $\omega_{21}$ decreases and possibly approaches zero with increasing sample size for the Bartlett-Thompson, Hoshino-Bentler and adjusted LISREL procedures, when $\xi$ is normally distributed (see figure 4). When the distribution departs from normality, only the Hoshino-Bentler and adjusted LISREL approaches yield RMSE possibly approaching zero, although their RMSE values are larger than with the Bartlett-Thompson and LISREL approaches when $n = 200$. Thus, those are the procedures which provide consistent estimates of $\omega_{21}$, regardless of distribution of $\xi$ and could, hence, be considered as robust against distributional misspecification for large sample sizes. An advantage with the Hoshino-Bentler approach, as compared to the adjusted LISREL approach, is lower bias for all investigated sample sizes. A disadvantage with the Hoshino-Bentler approach is, however, that it is less stable in terms of the model estimation algorithm than the adjusted LISREL approach. This is apparent from the fraction of disregarded samples which is substantially larger for the Hoshino-Bentler approach.
5 Discussion and Conclusions

The aim of this study was to compare methods for calculating factor scores to be used for estimation of a structural model with an interaction term and, if possible, propose a method, easily implementable in existing software. All methods provide estimates of the linear terms with low bias. When the distribution of $\xi$ is normal the Bartlett-Thompson, Hoshino-Bentler and adjusted LISREL approaches provide consistent estimation of $\omega_{21}$, whereas, the LISREL approach yield inconsisent estimates. When $\xi$ departs from normality the Hoshino-Bentler and adjusted LISREL approaches still provide consistent estimation of $\omega_{21}$, but the Bartlett-Thompson approach yields bias. Hence, only the Hoshino-Bentler and adjusted LISREL approaches are robust against distributional misspecification in terms of estimation of $\gamma$ and $\Omega$.

Although, both the Hoshino-Bentler and adjusted LISREL procedures seem to provide consistent estimation of $\omega_{21}$, the bias is smaller when using the Hoshino-Bentler approach. However, the probability for disregarding of samples is larger for the Hoshino-Bentler approach. An appropriate guideline would be to use the Hoshino-Bentler approach if there are no problems with negative covariances or non-convergent solutions using this approach; and if there are such problems, use the adjusted LISREL approach with less probability of such phenomena, but a somewhat larger bias. Some existing approaches are superior to these approaches, although their simplicity makes them attractive in for example social sciences.

It is worth mentioning that our unconstrained approach in the Hoshino-Bentler and adjusted LISREL approaches is similar to the unconstrained approach of Marsh et al. (2004) where the interaction term is treated as another factor. The difference is that the authors estimate the model directly without using factor scores, as in the case of this study. We leave a thorough comparison of the factor score regression approaches with Marsh et al. (2004) as a future topic.

Yet another possible topic for future research is to investigate the standard error estimates. When factor scores are used, it is necessary to consider the fact that the factor scores are obtained conditioned on the fixed parameter estimates. This could lead to underestimation of standard errors.
References


