Title
MM versus ML estimates of structural equation models with interaction terms: robustness to non-normality of the consistency property

Authors
Mooijaart, Ab
Satorra, Albert

Publication Date
2011-02-05

Peer reviewed

MM versus ML estimates of structural equation models with interaction terms: robustness to non-normality of the consistency property *

Ab Mooijaart

Leiden University

&

Albert Satorra

Universitat Pompeu Fabra and Barcelona GSE

January 29, 2011

*Requests for reprints should be sent to Ab Mooijaart, Institute of Psychology, Unit Methodology and Statistics, Leiden University, P.O. Box 9555, 2300 RB, Leiden, the Netherlands. Research of the second author is supported by grant SEJ2006-13537 from the Spanish Ministry of Science and Technology. The second author's work was partially carried out while visiting UCLA campus.

MM versus ML estimates of structural equation models with interaction terms: robustness to non-normality of the consistency property

Abstract

A standard assumption in structural equation models with interaction terms is the normality of all the random constituents of the model. In applications, however, there may be predictors, disturbance terms of equations, or measurement errors, deviating from normality. The present paper investigates how deviation from normality affects the consistency of two alternative estimators; namely, the maximum likelihood (ML) and the method of moments (MM) estimators. The ML approach requires full specification of the distribution of observable variables while this is not required in the MM approach. It will be seen that while the MM estimator is insensitive to departures from normality of all the random constituents of the model not involved in the interaction terms, such deviation from normality distorts considerably the consistency property of the ML estimator. The paper provides analytical results showing the consistency of MM when using a proper selection of moments up to order three, and presents a Monte Carlo illustration showing how the consistency of the ML estimator breaks down when there is deviation from normality. It is concluded that for a variety of distributions of the data, the MM method gives consistent estimates while ML does not.

KEYWORDS: maximum-likelihood, method of moments, third-order moment, asymptotic bias, latent-variable, structural relation

1 INTRODUCTION

Structural equation models are widely used in many disciplines for analyzing observational multivariate data (see, for example, the review paper of Sánchez, Budtz-Jorgenson, Ryan and Hu (2005), and Yuan and Bentler (2007)). The typical form of these models are linear relations among latent variables. In recent years, however, models with interaction effects among latent variables have become of increasing importance in research in psychometrics, education, marketing, and related disciplines, since they can account for more interesting features of the relations among variables under study. Even though a common assumption for analyzing these non-linear models is that all latent variables are normally distributed, the manifest variables are non-normally distributed due to the presence of the underlying interaction (non-linear) terms. So standard methods for linear structural equation models can not be used. There are several ways for analyzing these kinds of models, however here we refer to just a few key papers. For a more detailed overview see Mooijaart and Bentler (2010). Relevant literature in the area of interaction models are the edited book of Schumacker and Marcoulides (1998), and the seminal papers of Kenny and Judd (1984) and Jöreskog and Yang (1996). In assessing models with interaction terms, the necessity of using more information than first- and second-order moments (for instance, higher-order moments) is shown in Mooijaart and Satorra (2009).

Presently, a dominant approach to analyze these kind of models is maximum likelihood (ML) under the assumption of normality of all the stochastic components of the model; see Klein and Moosbrugger (2000), Lee and Zhu (2002), and the computer package Mplus (Muthén and Muthén, 2007). Related to this ML approach is the Bayesian approach in which the likelihood function is augmented with a prior density function in order to deal with small samples; see Arminger and Muthén (1998), Lee (2007), Lee, Song and Tang (2007). Recently, however, Mooijaart and Bentler (2010) advocate the alternative of using the method of moments (MM) to analyze structural equation models with interaction terms. In their paper they give examples where the MM is a practical alternative to the classical ML method. In the present paper we compare the virtue of the two methods in the often encountered set-up where the assumption of normality for stochastic components of the model is violated. In specifically, we investigate whether the basic – and essential – property of consistency of estimates do still hold when specific stochastic components of the model deviate from normality. Since the ML approach requires full specification of the distribution of the observable variables, while the MM method does not involve distributional specification, one may wonder whether the two types of estimators may differ with respect to the properties of robustness to a distributional specification such as normality. It will be seen that while the MM approach is insensitive to deviations form the normality assumption, the ML method fails dramatically to comply with the consistency property of the estimator when non-normality arise.

The paper is structured as follows. We next present a simulation example that gives Monte Carlo evidence of the lack of robustness of ML to violation of normality assumption of stochastic components of the model. A subsequent section gives a theorem that provides theoretical foundation for the consistency of the MM estimator under deviation from normality. A discussion of the implications of the findings concludes this paper. An appendix provides analytical details of the first- second- and higher-order moment structure involved in the theoretical part of the paper.

2 EXAMPLE

In this section we compare the results of the ML and the MM method for a non-linear factor model where some factors are normally distributed and one factor is non-normally distributed. Interaction of normally distributed factors are included in the model. Furthermore, it is assumed that the measurement errors and disturbance deviate highly from normality. In EQS formulation the model can be written as:

$$V_1 = F_1 + E_1 \quad (1)$$

$$V_2 = .6 * F_1 + E_2 \quad (2)$$

$$V_3 = F_2 + E_3 \quad (3)$$

$$V_4 = .7 * F_2 + E_4 \quad (4)$$

$$V_5 = F_3 + E_5 \quad (5)$$

$$V_6 = .8 * F_3 + E_6 \quad (6)$$

$$V_7 = .1 * V_{999} + .2 * F_1 + .4 * F_2 + .6 * F_3 + .4 * F_4 + E_7 \quad (7)$$

In these regression equations a "$*$" means that the corresponding parameter is a free parameter which has to be estimated and V_{999} is a vector with unit elements only. Factor F_1 and F_2 are normally distributed and factor F_4 is an interaction factor defined as $F_4 = F_1F_2$. Factor F_3 is standardized chi-square distributed with one degree of freedom, i.e. highly deviant from the normality assumption. The errors terms, E_j, $j = 1, \ldots, 7$, are scaled chi-square distributions with one degree of freedom. The specification of the variances/covariances of the factors and the errors is given in Table 1. It is obvious from the specification of the model above that, except for the factors F_1 and F_2, all factors and error variables are highly skewed variables, with the consequence that all observed variables are also highly skewed.

In this example a small Monte Carlo study is carried out in which the sample size is $n = 500$ and the number of replications is 200. The main objective is to compare the performance of ML versus MM method with respect to parameter estimates. In the MM method all means, variances and covariances are fitted and in addition just one third-order moment; specifically, the third-order moment V_1, V_3, V_7. The MM estimation have been carried out for the interaction model using EQS (Bentler and Wu, 2010) with the specification of LS method.

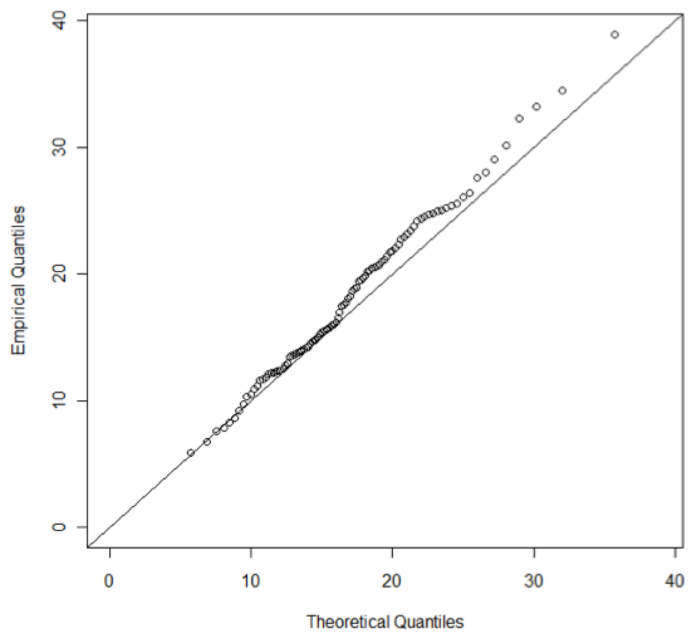

Figure 1: qq-plot for the AGLS corrected chi-square test statistic

2.1 Results

As an indication of the non-normality of the six observed variables, for an arbitrary sample of $n = 500$, the univariate skewness of the seven observable variables were found to be 0.97, 1.59, 0.65, 1.08, 2.67, 2.62, 1.99; i.e., the variables are highly skewed. For both ML and MM methods, MPlus converges 197 out of the 200 replications; for the MM method, EQS converges in all 200 replications.

Table 1 gives the true values of the population parameters and for the ML and MM method: the bias of the parameter estimates, the standard deviations of the estimates, and the means of the estimated standard errors. Lines in bold in this table are related to the structural part of the model. Comparing ML with MM we see that there is a substantial difference between the bias of the two methods. For the MM method, estimates of the bias seem to be reasonable, whereas the bias of ML is unacceptable high for some parameters. In particular, for the structural part of the model (see bold lines in the table), the regression equation, the estimates are far off from the true values. For instance, the mean (across replicates) of the most critical parameter, the interaction effect, is .83 (a bias of .43, i.e., a 107.5% deviation from the true value), whereas for the MM method the mean across replications is .41 (a bias of .01, i.e. a deviation of 2.5% from the true value). From Table 1 we also see that the standard deviation of estimates across replications and the means (across replications) of the standard errors of estimates (computed in each replication run) are close to each other for the MM method, which indicates that the standard errors of estimates are estimated in the MM method. For the ML method we see larger differences between the empirical (Monte Carlo) standard errors and the means (across replications) of the standard errors computed for each sample run.

Table 2 shows information on the distribution of goodness-of-fit test statistics in the case of the MM method (unfortunately, such a goodness-of-fit test statistics do not exist for ML). From this table, it is seen that the goodness-of-fit test statistics delivered by the MM method are approximately chi-square distributed as predicted by the asymptotic theory. The same conclusion can be drawn from Figure 1 that shows a qq-plot of the fit of the AGLS corrected chi-square to a chi-square distribution with 17 degrees of freedom (the value of the Kolmogorov-Smirnov test for the fit to a chi -square distribution is $D = .12$, p-value=.33).

Based on this small Monte Carlo study, in particular based on the estimates of the bias, we conclude that when basic stochastic components of the model deviate from normality, the ML method breaks down severely whereas the MM approach behaves with the bounds that are expected were the variables normally distributed. Robustness of the MM method against non-normality refers to consistency of estimates (i.e. lack of asymptotic bias), estimates of standard errors and chi-squareness of the goodness-of-fit test.

3 THEORETICAL FOUNDATIONS

This section describes a general structural equation model (SEM) with interaction terms for which the robustness of MM and ML estimators against non-normality of the factors not involved in the interaction terms is investigated. The model distinguishes two types of stochastic constituents of the model: an independent (possible latent) vector variable ξ that is assumed to be normally distributed; vector variables ϵ, δ and ζ that are assumed to be uncorrelated among them, independent of ξ, and free from the normality assumption (except for finiteness of second-order moments). The model considered is an extension of the LISREL model to include interaction terms; that is (this extends the model considered in Klein and Moosbrugger, 2000)

$$\begin{aligned} \eta &= \alpha + B_0\eta + \Gamma_1\xi + \Gamma_2(\xi \otimes \xi) + \zeta \qquad (8) \\ y &= \nu_y + \Lambda_y\eta + \epsilon \\ x &= \nu_x + \Lambda_x\xi + \delta \end{aligned}$$

where $y\,(p \times 1)$, $x\,(q \times 1)$, $\eta\,(m \times 1)$ and $\xi\,(n \times 1)$, with the other vector and matrices of conformably dimensions. The matrix Γ_2 contains the regression coefficients for the interactions or quadratic factors on ξ. Some or all of the elements of Γ_2 can of course be set to zero. In practice just a few of the elements of Γ_2 will be free parameters. Model (8) can be written as

$$
\begin{aligned}
y &= \nu_y + \Lambda_y B^{-1}\alpha + \Lambda_y B^{-1}\Gamma_1\xi + \Lambda_y B^{-1}\Gamma_2(\xi \otimes \xi) + \Lambda_y B^{-1}\zeta + \epsilon \quad (9) \\
x &= \nu_x + \Lambda_x \xi + \delta
\end{aligned}
$$

where the matrix $B = I - B_0$ is assumed to be nonsingular. By defining $\tilde{y} = y - E[y]$ and $\tilde{x} = x - E[x]$ and $\phi = E[\xi \otimes \xi]$, where $E[]$ denotes expectation operator, the model with variables in deviation from the means is

$$
\begin{aligned}
\tilde{y} &= \Lambda_y B^{-1}\Gamma_1\xi + \Lambda_y B^{-1}\Gamma_2(\xi \otimes \xi - \phi) + \Lambda_y B^{-1}\zeta + \epsilon \quad (10) \\
\tilde{x} &= \Lambda_x \xi + \delta
\end{aligned}
$$

Using the linear equations (9) and (10), the population first- second- and third-order moments of the observable variables can be expressed (using algebra of moments of linear equations) in terms of model parameters. The first- second- and third-order moments as a function of the model parameters of equations (9) and (10) are listed in Lemma 1 of the Appendix.

From the moment structures expressed in Lemma 1, the following theorem applies.

Theorem 1

Assume model (8) holds, and ξ is normally distributed and independent of δ, ϵ and ζ. Then

1. *The first- and second-order moments of observable variables x and y (see (11) to (15) of Lemma 1 in the Appendix) are free of the distribution of δ, ϵ and ζ.*

2. *The third-order moments of x (the components of σ_{xxx}, see (16) of Lemma 1 in the Appendix) are NOT free of the distribution of δ.*

3. *The third-order moments of the vector $z = (x', y')'$ involving at least on component of x (the components of σ_{xxy} and σ_{xyy}, see (17) and (18) of Lemma 1 in the Appendix) are free of the distribution of δ, ϵ and ζ.*

4. *The third-order moments of y (the components of σ_{yyy}, see (19) of Lemma 1 in the Appendix) are NOT free of the distribution of ϵ and ζ.*

We now consider the MM estimation of the model. Assume the model (8) holds with the loadings, regression coefficients and variance matrices of independent variables being a function of a parameter vector θ that varies in Θ, a compact subset of R^q. Let $z_1, \ldots, z_n$ be iid observations (a sample of size n) from the vector of observable variables $z = (x', y')'$, and let s_n be the sample vector of first- second- and (a selection of) third-order moments of z.

Condition 1 *The selected third-order moments included in s_n do involve at least one x and at least one y (for example, s_{xyy} and s_{xxy})*

Clearly, since the sample is iid, it holds that $s_n \xrightarrow{P} \sigma_0$, when sample size $n \to \infty$. Assume z has finite sixth-order moment; then, from the iid assumption, we have the asymptotic distribution result $\sqrt{n}(s_n - \sigma_0) \xrightarrow{L} \mathcal{N}(0, \Gamma)$ where the matrix Γ is a function of moments of z up to sixth-order. Since s_n can be expressed (asymptotically) as iid sum of product variables vector t_i of order up to three, a consistent estimator $\hat{\Gamma}$ of Γ is obtained simply as the sampling variance of the t_is.

Given the results of the Lemma 1, that express the moments up to order three as a function of parameters, when the model holds we have $\sigma_0 = \sigma(\theta_0)$ where $\theta_0 \in \Theta$. We assume that $\sigma(\theta)$ is a continuously differentiable vector-valued function of θ with full column-rank Jacobian $\partial\sigma(\theta)/\partial\theta' \mid_{\theta=\theta_0}$. The WLS-MM estimator $\hat{\theta}$ is defined as the minimizer of the function $f(\theta)$ defined by the quadratic form

$$f = (s_n - \sigma(\theta))' W_n (s_n - \sigma(\theta))$$

associated to the model $\sigma(\theta)$. Here W_n is a weight matrix that may be sample dependent and which converges in distribution to $W > 0$. Often W_n is simply an identity matrix, leading to a LS- MM estimator. Under this set-up, the following corollary applies.

Corollary 1 *Under the assumptions of Theorem 1 and Condition 1, the MM estimator $\hat{\theta}$ is a consistent estimator of θ_0, i.e. $\hat{\theta} \xrightarrow{P} \theta_0$ when $n \to \infty$.*

Corollary 2 *Under the assumptions of Theorem 1 and Condition 1, the* *quadratic form statistic*

$$T = n(s-\hat{\sigma})\left(\hat{\Gamma}^{-1} - \hat{\Gamma}^{-1}\hat{\Delta}(\hat{\Delta}'\hat{\Gamma}^{-1}\hat{\Delta})^{-1}\hat{\Delta}'\hat{\Gamma}^{-1}\right)(s-\hat{\sigma}),$$

where $\hat{\sigma} = \sigma(\hat{\theta})$, $\hat{\Delta} = \partial\sigma(\theta)/\partial\theta' \mid_{\theta=\hat{\theta}}$ *and* $\hat{\Gamma}$ *is a consistent estimator of* Γ*, has an asymptotic chi-square distribution with degrees of freedom equal to* $p^* - q$*,* p^* *being the number of moments involved in* s *and* q *the length of* θ*.*

Both corollaries are direct application of result related to WLS of moment structures (see e.g., Satorra, 1989).

4 CONCLUSIONS

Corollaries 1 and 2 guaranty the validity of the regular MM estimators for the general class of models (8). The results of the Monte Carlo study in section above, corroborates the results put forward by Corollaries 1 and 2 regarding the robustness of the MM method to non-normality. The non-robustness of the ML method shown by the Monte Carlo study supports the conjecture that the ML approach is sensitive to non-normality. That is, the ML method, which assumes normality for all stochastic components of the model, does not ensure consistency of estimates of interaction effects and other parameters when exogenous variables, disturbance terms of equations, and measurement errors deviate from normality.

REFERENCES

Arminger, G. & Muthén, B. (1998). A Bayesian approach to nonlinear latent variable models using the Gibbs sampler and the Metropolis-Hastings algorithm. *Psychometrika, 63*, 271-300.

Bentler, P. M. (2000-08). *EQS 6 structural equations program manual.* Encino, CA: Multivariate Software, Inc. (www.mvsoft.com).

Klein, A. G., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika, 65*, 457-474.

Jöreskog, K.G. & Yang, F. (1996). Nonlinear structural equation models: The Kenny-Judd model with interaction effects. In: R.E. Marcoulides & G.A. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 57-88). Mahwak, NJ: Erlbaum.

Kenny, D.A. & Judd, C.M. (1984). Estimating the nonlinear and interactive effects of latent variables. *Psychological Bulletin, 96,* 201-210.

Lee, S.-Y. (2007). *Structural equation modeling: a Bayesian approach.* West Sussex: Wiley.

Lee, S.Y., & Zhu, H.T. (2002). Maximum likelihood estimation of nonlinear structural equation models. *Psychometrika, 67,* 189-210.

Lee, S. -Y., Song, X. -Y, & Tang, N. -S. (2007). Bayesian methods for analyzing structural equation models with covariates, interaction, and quadratic latent variables. *Structural Equation Modeling, 14(3)*, 404-434.

Magnus, J.R., & Neudecker, H. (1999). Matrix differential calculus with applications in statistics and econometrics(2nd edit.). Chichester: John Wiley & Sons.

Meijer, E. (2005). Matrix algebra for higher order moments. *Linear Algebra and its Applications, 410*, 112134.

Mooijaart, A., & Bentler, P.M. (2010). An alternative approach for non-linear latent variable models. *Structural Equation Modeling, 17*, 357-373.

Mooijaart, A. and A. Satorra (2009), "On insensitivity of the chi-square model test to non-linear misspecification in structural equation models", *Psychometrika, 74* , 443-455

Muthén, L.K., & Muthén, B.O. (1998-2007). Mplus users guide (5th ed.). Los Angeles, CA: Muthén & Muthén

Sánchez, B.N. E. Budtz-Jorgenson, L.M. Ryan y H. Hu (2005) Structural Equation Models: A Review With Applications to Enviromental Epidemiology, *Journal of the American Statistical Association, 100*, 472, 1443-1455.

Satorra, A. (1989). Alternative test criteria in covariance structure analysis: A unified approach. *Psychometrika, 54(1)*, 131-151.

Schumacker, R. E., & Marcoulides, G. A. (Eds.). (1998). *Interaction and nonlinear effects in structural equation modeling.* Mahwah, NJ: Erlbaum.

Yuan, K-H. and P. M. Benter (2007). Structural Equation Modeling, in *Handbook of Statistics, Vol. 26*, pp. 297-358, Elsevier: Holland.

5 APPENDIX: PROOF OF LEMMA 1

This appendix proves Lemma 1 which is prior to derive the main theorem of the manuscript. The following notation will be used for second-, fourth- and sixth-order moments respectively of the observable variables and the vector variable ξ.

$$\sigma_{xxx} = E[x \otimes x \otimes x]$$

$$\sigma_{xxy} = E[x \otimes x \otimes y];\ \sigma_{xyx} = E[x \otimes y \otimes x];\ \sigma_{yxx} = E[y \otimes x \otimes x]$$

$$\sigma_{xyy} = E[x \otimes y \otimes y];\ \sigma_{yxy} = E[y \otimes x \otimes y];\ \sigma_{yyx} = E[y \otimes y \otimes x]$$

$$\sigma_{yyy} = E[y \otimes y \otimes y]$$

$$\phi_2 = E[\xi \otimes \xi]$$

$$\phi_4^{(1)} = E[(\xi \otimes \xi - \phi_2) \otimes \xi \otimes \xi]$$

$$\phi_4^{(2)} = E[\xi \otimes (\xi \otimes \xi - \phi_2) \otimes \xi]$$

$$\phi_4^{(3)} = E[\xi \otimes \xi \otimes (\xi \otimes \xi - \phi_2)]$$

$$\phi_4 = E[(\xi \otimes \xi - \phi_2) \otimes (\xi \otimes \xi - \phi_2)]$$

$$\phi_6 = E[(\xi \otimes \xi - \phi_2) \otimes (\xi \otimes \xi - \phi_2) \otimes (\xi \otimes \xi - \phi_2)]$$

Clearly, the moments σ_{xxy}, σ_{xyx} and σ_{yxx} can be related linearly among them using a permutation matrix, and the same for σ_{yxy}, σ_{xyy} and σ_{yyx}, so all the properties that apply to one form of the third-order moments do apply to the other equivalent forms.

Lemma 1 *Assume model (8) holds, and ξ is normally distributed and independent of δ, ϵ and ζ. Then the following expression for the first- second- and third-order moments apply:*

$$\mu_x = \nu_x \tag{11}$$

$$\mu_y = \sigma + A_2\phi_2 \tag{12}$$

$$\sigma_{xx} = (\Lambda_x \otimes \Lambda_x)\phi_2 + E[\delta \otimes \delta] \tag{13}$$

$$\sigma_{xy} = (\Lambda_x \otimes A_1)\phi_2 \tag{14}$$

$$\sigma_{yy} = (A_1 \otimes A_1)\phi_2 + (A_2 \otimes A_2)\phi_4 + E[w \otimes w] \tag{15}$$

$$\sigma_{xxx} = E(\delta \otimes \delta \otimes \delta) \tag{16}$$

$$\sigma_{xxy} = (\Lambda_x \otimes \Lambda_x \otimes A_2)\phi_4^{(3)} \tag{17}$$

$$\sigma_{xyy} = (\Lambda_x \otimes A_1 \otimes A_2)\phi_4^{(3)} + (\Lambda_x \otimes A_2 \otimes A_1)\phi_4^{(2)} \tag{18}$$

$$\begin{aligned} \sigma_{yyy} = {} & (A_1 \otimes A_1 \otimes A_2)\phi_4^{(3)} + (A_2 \otimes A_1 \otimes A_1)\phi_4^{(1)} \\ & +(A_1 \otimes A_2 \otimes A_1)\phi_4^{(2)} + (A_2 \otimes A_2 \otimes A_2)\phi_6 + E[w \otimes w \otimes w] \end{aligned} \tag{19}$$

PROOF: Since ξ is assumed to be stochastically independent of all the other random components of the model, it follows that ξ is stochastically independent of $w = \Lambda_y B^{-1}\zeta + \epsilon$. We remain also of the notation $a = \nu_y + \Lambda_y B^{-1}\alpha$; $A_1 = \Lambda_y B^{-1}\Gamma_1$, $A_2 = \Lambda_y B^{-1}\Gamma_2$, so the general model can be written as (see (9))

$$\begin{aligned} x &= \nu_x + \Lambda_x \xi + \delta \\ y &= a + A_1 \xi + A_2(\xi \otimes \xi) + w \end{aligned}$$

which, in deviation of the variable means, implies

$$\begin{aligned} \tilde{x} &= \Lambda_x \xi + \delta \\ \tilde{y} &= A_1 \xi + A_2(\xi \otimes \xi - \phi_2) + w \end{aligned}$$

Clearly, the <u>first-order moments</u> are

$$\begin{aligned} \sigma_x &= E[x] = \nu_x \\ \sigma_y &= E[y] = a + A_2 \phi_2 \end{aligned}$$

where $\phi_2 = E[\xi \otimes \xi]$.

To derive the <u>second-order moments</u> we make use of the following property of Kronecker product: $(A \otimes B)(C \otimes D) = (AC \otimes BD)$. Clearly

$$\sigma_{xx} = E[\tilde{x} \otimes \tilde{x}] = E[(\Lambda_x \xi + \delta) \otimes (\Lambda_x \xi + \delta)] = (\Lambda_x \otimes \Lambda_x)\phi_2 + \sigma_{\delta\delta}$$

where we used $E[\delta] = 0$, the un-correlation of ξ and δ (due to the independence of the two random variables), in conjunction with the notation

$\sigma_{\delta\delta} = E[\delta \otimes \delta]$.

Further

$$\sigma_{xy} = E[\tilde{x} \otimes \tilde{y}] = E[(\Lambda_x \xi + \delta) \otimes (A_1 \xi + A_2(\xi \otimes \xi - \phi_2) + w)] = (\Lambda_x \otimes A_1)\phi_2$$

where we used the normality assumption of ξ to infer $E[(\xi \otimes \xi - \phi_2)] = 0$, the independence of ξ of δ and w to infer that $E[(\delta \otimes (\xi \otimes \xi - \phi_2)] = 0$ and $E[(\xi \otimes w)] = 0$ and the uncorrelation among δ and w to infer $E[\delta \otimes w] = 0$.

Finally

$$\sigma_{yy} = E[\tilde{y} \otimes \tilde{y}] = E[(A_1 \xi + A_2(\xi \otimes \xi - \phi_2) + w) \otimes (A_1 \xi + A_2(\xi \otimes \xi - \phi_2) + w)]$$

$$= (A_1 \otimes A_1)\phi_2 + (A_2 \otimes A_2)\phi_4 + \psi$$

where we again used the normality assumption of ξ, the independence between ξ and w, and the notation $\psi = E[w \otimes w]$.

Now we derived the <u>third-order moments</u>.

$$\sigma_{xxx} = E[\tilde{x} \otimes \tilde{x} \otimes \tilde{x}] = E[(\Lambda_x \xi + \delta) \otimes (\Lambda_x \xi + \delta) \otimes (\Lambda_x \xi + \delta)] = E[\delta \otimes \delta \otimes \delta]$$

where we used the normality assumption of ξ (to infer the zero third-order moments) and the uncorrelated with the centered vector variable δ.

Using again the normality assumption of ξ and its independence of δ and w, we easily obtain

$$\sigma_{xxy} = E[\tilde{x} \otimes \tilde{x} \otimes \tilde{y}] = E[(\Lambda_x \xi + \delta) \otimes (\Lambda_x \xi + \delta) \otimes (A_1 \xi + A_2(\xi \otimes \xi - \phi_2)]$$

$$= (\Lambda_x \otimes \Lambda_x \otimes A_2) E[\xi \otimes \xi \otimes (\xi \otimes \xi - \phi_2)] = (\Lambda_x \otimes \Lambda_x \otimes A_2)\phi_4^{(3)}$$

$$\sigma_{xyy} = E[\tilde{x} \otimes \tilde{y} \otimes \tilde{y}] = E[(\Lambda_x \xi + \delta) \otimes (A_1 \xi + A_2(\xi \otimes \xi - \phi_2) + w) \otimes (A_1 \xi + A_2(\xi \otimes \xi - \phi_2) + w$$

$$= (\Lambda_x \otimes A_1 \otimes A_2) E[\xi \otimes \xi \otimes (\xi \otimes \xi - \phi_2)] + (\Lambda_x \otimes A_2 \otimes A_1) E[\xi \otimes (\xi \otimes \xi - \phi_2) \otimes \xi]$$

$$= (\Lambda_x \otimes A_1 \otimes A_2)\phi_4^{(3)} + (\Lambda_x \otimes A_2 \otimes A_1)\phi_4^{(2)}$$

and

$$\sigma_{yyy} = E[\tilde{y} \otimes \tilde{y} \otimes \tilde{y}]$$

$$= E[(A_1 \xi + A_2(\xi \otimes \xi - \phi_2) + w) \otimes (A_1 \xi + A_2(\xi \otimes \xi - \phi_2) + w) \otimes (A_1 \xi + A_2(\xi \otimes \xi - \phi_2) + w]$$

$$= (A_1 \otimes A_1 \otimes A_2) E[\xi \otimes \xi \otimes (\xi \otimes \xi - \phi_2)]$$

$$+ (A_2 \otimes A_1 \otimes A_1) E[(\xi \otimes \xi - \phi_2) \otimes \xi \otimes \xi]$$

$$+ (A_1 \otimes A_2 \otimes A_1) E[\xi \otimes (\xi \otimes \xi - \phi_2) \otimes \xi]$$

$$+ (A_2 \otimes A_2 \otimes A_2) E[(\xi \otimes \xi - \phi_2) \otimes (\xi \otimes \xi - \phi_2) \otimes (\xi \otimes \xi - \phi_2)]$$

$$+ E[w \otimes w \otimes w]$$

$$= (A_1 \otimes A_1 \otimes A_2)\phi_4^{(3)} + (A_2 \otimes A_1 \otimes A_1)\phi_4^{(1)} + (A_1 \otimes A_2 \otimes A_1)\phi_4^{(2)}$$

$$+(A_2 \otimes A_2 \otimes A_2)\phi_6 + E[w \otimes w \otimes w]$$

Note that it was used that $E[\xi \otimes \xi \otimes \xi]$, $E[\xi \otimes (\xi \otimes \xi - \phi_2) \otimes (\xi \otimes \xi - \phi_2)]$, $E[(\xi \otimes \xi - \phi_2) \otimes \xi \otimes (\xi \otimes \xi - \phi_2)]$ and $E[(\xi \otimes \xi - \phi_2) \otimes (\xi \otimes \xi - \phi_2) \otimes \xi]$ are zero, since the normality assumption for ξ. ■

A more compact expression for σ_{xyy} and σ_{yyy} can be given using commutations matrices as follows (derivations of those expressions can be obtained by the first author):

$$\sigma_{xyy} = E[\tilde{x} \otimes \tilde{y} \otimes \tilde{y}] = T_{xyy}(\Lambda_x \otimes A_1 \otimes A_2)\phi_4^{(3)}$$

where $T_{xyy} = I_{p^2 q} + I_q \otimes K_{p,p}$ and $K_{p,p}$ is the commutation matrix of order p (see Magnus and Neudecker, 1999).

$$\sigma_{yyy} = E[\tilde{y} \otimes \tilde{y} \otimes \tilde{y}] = T_{yyy}(A_1 \otimes A_1 \otimes A_2)(A_2 \otimes A_2 \otimes A_2)\phi_6 + E[w \otimes w \otimes w]$$

where $T_{yyy} = I_{p^3} + I_p \otimes K_{p,p} + K_{p,p^2}$ is a triplication matrix (see Meijer, 2005).

Table 1: Results of Simulation for ML and MM method: bias, standard deviations and standard errors

		ML			MM		
		Intercepts					
pars	true	bias	sd	se	bias	sd	se
V7	**1.00**	**-.04**	**.04**	**.04**	**.00**	**.05**	**.04**
		Factor variances					
pars	true	bias	sd	se	bias	sd	se
F1	.49	-.11	.09	.09	.02	.13	.12
F2	.64	-.08	.08	.09	-.01	.08	.08
F3	1.0	-.01	.19	.17	-.03	.16	.13
		Factor covariances					
pars	true	bias	sd	se	bias	sd	se
F1,F2	.2352	-.07	.05	.06	-.01	.04	.04
		Error variances					
pars	true	bias	sd	se	bias	sd	se
E1	.51	.11	.11	.11	-.05	.14	.13
E2	.64	.01	.12	.11	-.03	.10	.07
E3	.36	.08	.09	.08	-.02	.08	.08
E4	.51	.01	.08	.08	-.03	.07	.06
E5	.60	.00	.12	.10	-.03	.09	.08
E6	.36	.01	.07	.07	-.01	.06	.05
E7	**.20**	**-.14**	**.05**	**.11**	**-.02**	**.06**	**.05**
		Factor Loadings, main					
pars	true	bias	sd	se	bias	sd	se
V2,F1	.60	.06	.10	.11	-.01	.14	.13
V4,F2	.70	.04	.07	.07	.01	.08	.08
V6,F3	.80	.00	.06	.05	.00	.06	.05
		Regression equation					
pars	true	bias	sd	se	bias	sd	se
V7,F1	**.20**	**.15**	**.12**	**.11**	**.01**	**.09**	**.08**
V7,F2	**.40**	**.10**	**.08**	**.08**	**.01**	**.07**	**.06**
V7,F3	**.60**	**-.01**	**.04**	**.04**	**.01**	**.05**	**.04**
V7,F4	**.40**	**.43**	**.28**	**.24**	**.01**	**.18**	**.15**

The elapsed time for 200 replications is

EQS: 00:00:58

Mplus: 00:45:30

Table 2: Goodness-of-fit summaries for the MM method

degrees of freedom	17
chi-square statistic	18.85
% of rejections, $\alpha = 5\%$	10.5
AGLS corrected chi-square statistic	18.15
% of rejections, $\alpha = 5\%$	5.6
AGLS F-statistic	1.08
% of rejections, $\alpha = 5\%$	6.1