British Journal of Mathematical and Statistical Psychology (2003), 56, 47–63

www.bps.org.uk

A method of moments technique for fitting interaction effects in structural equation models

Melanie M. Wall[1] and Yasuo Amemiya[2]
[1]Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, USA
[2]Department of Statistics, Iowa State University, Ames, USA

The desire to fit structural equation models containing an interaction term has received much methodological attention in the social science literature. This paper presents a technique for the cross-product structural model that utilizes factor score estimates and results in closed-form moments-type estimators. The technique, which does not require normality for the underlying factors, was originally introduced in a very general form by Wall and Amemiya (2000) for any polynomial structural model. In this paper, the practical implementation of this method, including standard error estimation, is presented specifically for the cross-product model. The procedure is applied to an example from social/behavioural epidemiology where the flexibility of the cross-product model provides a useful description of the underlying theory. A simulation study is also presented comparing the method of moments for the cross-product model with three other procedures.

1. Introduction

Structural equation modelling (SEM) is a common tool in the social and behavioural sciences for estimating and testing linear relationships among linear latent variables. But in many cases the restriction to linearity is not adequate or flexible enough to explain the phenomena of interest. For example, if the slope between two continuous latent variables is directly affected or 'moderated' by a third continuous latent variable, this relationship cannot be estimated via the traditional SEM. A natural way to model this type of relationship is to include a cross-product of the 'interacting' latent variables into the structural model. Thus, the structural model

$$f_1 = \alpha_0 + \alpha_1 f_2 + \alpha_2 f_3 + \alpha_3 f_2 f_3 + \zeta \quad (1)$$

is considered, where $\mathbf{f} = (f_1, f_2, f_3)'$ is a vector of continuous latent variables,

*Requests for reprints should be addressed to Dr Melanie Wall, Division of Biostatistics, School of Public Health, University of Minnesota, 420 Delaware Street SE, Minneapolis, MN 55455, USA (email: melanie@biostat.umn.edu).

$\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \alpha_2, \alpha_3)$ is a vector of fixed unknown parameters, and ζ is random error. A structural model (1) is referred to as nonlinear since it is nonlinear in the factors—the $f_2 f_3$ term makes it nonlinear. Nonlinear structural models are not straightforward to fit using the usual SEM software packages (LISREL, EQS, PROC CALIS, AMOS, Mplus) since these packages are designed for models containing only linear latent variable terms.

Interaction effects in structural equation models such as (1) have recently received a large amount of methodological attention. For a collection of papers on such models, see Schumacker and Marcoulides (1998). One of the first methods for fitting an interaction effect in a structural model was proposed by Kenny and Judd (1984). Their method uses products of observed indicators as indicators of the cross-product latent term while imposing appropriate nonlinear constraints on the associated parameters. Although this technique attracted methodological discussions and alterations by a number of papers, including Hayduk (1987), Ping (1995, 1996), Jaccard and Wan (1995, 1996), and Jöreskog and Yang (1996, 1997), it produces inconsistent estimators when the observed indicators are not normally distributed.

Also relying on the multivariate normality assumption of the observed indicators of the exogenous variables, Arminger and Muthén (1998) proposed a fully Bayesian approach and Klein and Moosbrugger (2000) proposed fitting models like (1) by the EM algorithm. Apart from the computational burden of these techniques, it is not clear how robust they are when the normality assumption is violated.

An adaptation of the Kenny–Judd technique that does not rely on the normality of the observed indicators was introduced by Wall and Amemiya (2001). Their generalized appended product indicator (GAPI) procedure produces consistent estimators for virtually any distribution of the observed indicator variables. Likewise, a method introduced by Bollen (1995) and then presented specifically for the cross-product structural model by Bollen and Paxton (1998) produces consistent estimators for non-normal data as well. Bollen's procedure uses the instrumental variables technique where instruments are formed by taking products of the observed indicators. Although this technique has simple closed-form solutions for the estimators, it has been shown to have lower efficiency.

Wall and Amemiya (2000) introduced a two-stage method of moments (2SMM) procedure for fitting the general polynomial structural equation model. Like the GAPI procedure and Bollen's instrumental variable technique, the 2SMM procedure produces consistent and asymptotically normal estimators for the structural model parameters for virtually any distribution of the observed indicator variables. The procedure uses factor score estimates in a form of nonlinear errors-in-variables regression and produces closed-form method of moments type estimators as well as asymptotically correct standard errors. Its ability to generalize to any type of polynomial gives it a clear advantage over the GAPI procedure (which can only be used for low-degree polynomials) and in simulation studies it has performed better in efficiency and coverage probability than both the GAPI and the Bollen instrumental variable technique.

The emphasis on theoretical development in Wall and Amemiya (2000) of the 2SMM procedure for the most general-case polynomial may not lend itself to practical implementation for users. Thus, the purpose of this paper is to outline the 2SMM procedure straightforwardly and specifically for the cross-product structural model (1). In Section 2 we give step-by-step instructions for calculating the 2SMM estimator $\hat{\boldsymbol{\alpha}}$ of the coefficients in (1). Section 3 gives the formulation for the standard errors of $\hat{\boldsymbol{\alpha}}$ so that confidence intervals can be formed. In Section 4 we apply the 2SMM procedure to an example from behavioural epidemiology where the cross-product model provides a useful match to the theory. Finally, Section 5 presents a simulation study comparing the 2SMM procedure to other procedures described above. Additional details for standard error estimation are given in the Appendix.

2. The two-stage method of moments estimator

Consider the cross-product structural model (1) with a linear measurement model that has a general form. Thus, for independent individuals $i = 1, \ldots, n$ we model the $p \times 1$ observed vector $\mathbf{z}_i$ as

$$\mathbf{z}_i = \begin{pmatrix} \boldsymbol{\beta}_0 \\ \mathbf{0}_{3\times 1} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\beta}_1 \\ \mathbf{I}_3 \end{pmatrix} \begin{pmatrix} f_{1i} \\ f_{2i} \\ f_{3i} \end{pmatrix} + \boldsymbol{\epsilon}_i, \tag{2}$$

$$f_{1i} = \alpha_0 + \alpha_1 f_{2i} + \alpha_2 f_{3i} + \alpha_3 f_{2i} f_{3i} + \zeta_i. \tag{3}$$

The factors $\mathbf{f}_i = (f_{1i}, f_{2i}, f_{3i})'$ can be treated as fixed or random, and if treated as random their distribution is left unspecified. The measurement errors $\boldsymbol{\epsilon}_i$ are assumed to be independent and identically distributed, with $E\{\boldsymbol{\epsilon}_i\} = \mathbf{0}$ and $\text{Var}\{\boldsymbol{\epsilon}_i\} = \boldsymbol{\Psi}$, and to be independent of $\mathbf{f}_i$. Although the 2SMM method and theory can be extended to include non-diagonal $\boldsymbol{\Psi}$, we assume for simplicity that the p elements of $\boldsymbol{\epsilon}_i$ are independent and thus $\boldsymbol{\Psi}$ is diagonal. The equation errors ζ_i are assumed to be independent and identically distributed with mean zero and variance σ_ζ and independent of $\mathbf{f}_i$ and $\boldsymbol{\epsilon}_i$. The $(p-3)\times 1$ vector $\boldsymbol{\beta}_0$ and $(p-3)\times 3$ matrix $\boldsymbol{\beta}_1$ contain coefficients which may be fixed or free. The notation $\mathbf{0}_{a\times b}$ represents a matrix of zeros with dimension $a \times b$ and $\mathbf{I}_k$ represents an identity matrix of dimension k. Depending on the dimension p, some restriction may need to be placed on the elements of $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_1$ to ensure that $\boldsymbol{\beta}_0$, $\boldsymbol{\beta}_1$ and $\boldsymbol{\Psi}$ can be meaningfully estimated. That is, we make the assumption that the measurement model, when taken on its own, is identified irrespective of the structural model.

The 2SMM procedure for estimation and inference of the model given by (2) and (3) is considered. The first stage of the procedure uses the measurement model (2) alone to estimate the underlying factor scores as well as their variability. The second stage uses these factor score estimates as 'observations' of the factors, and, while incorporating the estimated measurement error, performs an errors-in-variables regression for the structural model (3). An advantage of this two-stage procedure is its natural relation to the way many researchers attack SEM model building and checking—first get the measurement model to fit and then explore the direct relationship among the underlying factors. Below we give a detailed outline of the 2SMM technique for model (2)-(3) under the very general case where the measurement model does not have simple structure, and no distributional form is assumed for the factors or the measurement errors. We also provide a simpler form of the estimator under scenarios where normality is assumed for the measurement errors or the measurement model has simple structure.

Stage 1. Consider the measurement model alone.

(i) Obtain $\hat{\boldsymbol{\beta}}_0$, $\hat{\boldsymbol{\beta}}_1$ and $\hat{\boldsymbol{\Psi}}$ using a standard SEM software package.

(ii) Calculate the Bartlett factor score estimates for each individual, i.e.,

$$\hat{\mathbf{f}}_i = \left[(\hat{\boldsymbol{\beta}}_1', \mathbf{I}_3)\hat{\boldsymbol{\Psi}}^{-1}\begin{pmatrix} \hat{\boldsymbol{\beta}}_1 \\ \mathbf{I}_3 \end{pmatrix}\right]^{-1} (\hat{\boldsymbol{\beta}}_1', \mathbf{I}_3)\hat{\boldsymbol{\Psi}}^{-1}\left[\mathbf{z}_i - \begin{pmatrix} \hat{\boldsymbol{\beta}}_0 \\ \mathbf{0}_{3\times 1} \end{pmatrix}\right] \tag{4}$$

Wall and Amemiya (2000) use an alternative form of (4) which does not explicitly contain $\hat{\boldsymbol{\Psi}}^{-1}$:

$$\hat{\mathbf{f}}_i = [-\hat{\boldsymbol{\Gamma}}, (\mathbf{I}_3 + \hat{\boldsymbol{\Gamma}}\hat{\boldsymbol{\beta}}_1)][\mathbf{z}_i - (\hat{\boldsymbol{\beta}}_0', \mathbf{0}_{1\times 3})'], \tag{5}$$

where

$$\hat{\boldsymbol{\Gamma}} = (\mathbf{0}_{3\times(p-3)}, \mathbf{I}_3)\,\hat{\boldsymbol{\Psi}}\begin{pmatrix}\mathbf{I}_{(p-3)}\\ -\hat{\boldsymbol{\beta}}_1'\end{pmatrix}\left[(\mathbf{I}_{(p-3)}, -\hat{\boldsymbol{\beta}}_1)\hat{\boldsymbol{\Psi}}\begin{pmatrix}\mathbf{I}_{(p-3)}\\ -\hat{\boldsymbol{\beta}}_1'\end{pmatrix}\right]^{-1}. \tag{6}$$

This form, which incorporates the $\hat{\boldsymbol{\Gamma}}$ notation, was used originally because it provided a convenient expression for deriving theoretical properties. When some variables are measured without error and $(\mathbf{I}_{(p-3)}, -\hat{\boldsymbol{\beta}}_1)\hat{\boldsymbol{\Psi}}(\mathbf{I}_{(p-3)}, -\hat{\boldsymbol{\beta}}_1)'$ is singular, a similar expression holds which incorporates a generalized inverse.

(iii) Estimate the variance of the estimation error for the factor score estimates, that is, estimate $\mathrm{Var}\{\mathbf{e}_i\} = \boldsymbol{\Sigma}_{\mathbf{ee}}$ where $\mathbf{e}_i = \hat{\mathbf{f}}_i - \mathbf{f}_i$. The familiar form of this variance is $[(\hat{\boldsymbol{\beta}}_1', \mathbf{I}_3)\hat{\boldsymbol{\Psi}}^{-1}(\hat{\boldsymbol{\beta}}_1', \mathbf{I}_3)']^{-1}$. Wall and Amemiya use the $\hat{\boldsymbol{\Gamma}}$ matrix and equivalently form

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{ee}} = \begin{pmatrix}\hat{\sigma}_{11} & \hat{\sigma}_{12} & \hat{\sigma}_{13}\\ \hat{\sigma}_{12} & \hat{\sigma}_{22} & \hat{\sigma}_{23}\\ \hat{\sigma}_{13} & \hat{\sigma}_{23} & \hat{\sigma}_{33}\end{pmatrix} = [-\hat{\boldsymbol{\Gamma}}, (\mathbf{I}_3 + \hat{\boldsymbol{\Gamma}}\hat{\boldsymbol{\beta}}_1)]\hat{\boldsymbol{\Psi}}\begin{pmatrix}\mathbf{0}_{(p-3)\times 3}\\ \mathbf{I}_3\end{pmatrix}. \tag{7}$$

(iv) Estimate the necessary higher-order moments of $\mathbf{e} = (e_1, e_2, e_3)'$ under one of the following three scenarios. For the cross-product model, the higher moments which are needed in stage 2(ii) are $E(e_1 e_2 e_3) = \mu_1^3$, $E(e_2^2 e_3) = \mu_2^3$, $E(e_2 e_3^2) = \mu_3^3$ and $E(e_2^2 e_3^2) = \mu_1^4$.

Scenario 1. Distribution of $\boldsymbol{\epsilon}$ is unspecified:

Wall and Amemiya (2000) provide a general method for estimating the higher-order moments, $\hat{\mu}_1^3$, $\hat{\mu}_2^3$, $\hat{\mu}_3^3$ and $\hat{\mu}_1^4$ by taking moments of residuals from the measurement model.

Scenario 2. $\boldsymbol{\epsilon}$ is normally distributed:

This implies the odd moments of $\mathbf{e}$ are zero and the fourth moment is a simple function of second moments, i.e., $\mu_1^3 = 0$, $\mu_2^3 = 0$, $\mu_3^3 = 0$, $\mu_1^4 = \sigma_{22}\sigma_{33} + 2\sigma_{23}$. Thus, use $\hat{\mu}_1^3 = 0$, $\hat{\mu}_2^3 = 0$, $\hat{\mu}_3^3 = 0$, $\hat{\mu}_1^4 = \hat{\sigma}_{22}\hat{\sigma}_{33} + 2\hat{\sigma}_{23}$.

Scenario 3. Measurement model (2) has simple structure:

This implies elements of $\mathbf{e}$ are independent and so $\mu_1^3 = 0$, $\mu_2^3 = 0$, $\mu_3^3 = 0$, $\mu_1^4 = \sigma_{22}\sigma_{33}$. Thus, use $\hat{\mu}_1^3 = 0$, $\hat{\mu}_2^3 = 0$, $\hat{\mu}_3^3 = 0$, $\hat{\mu}_1^4 = \hat{\sigma}_{22}\hat{\sigma}_{33}$.

Stage 2. Using estimates from stage 1, fit the structural model.

(i) Rewrite model (2)–(3) as an errors-in-variables model

$$\begin{aligned} f_{1i} &= \alpha_0 + \alpha_1 f_{2i} + \alpha_2 f_{3i} + \alpha_3 f_{2i} f_{3i} + \zeta_i,\\ \hat{f}_{1i} &= f_{1i} + e_{1i},\\ \hat{f}_{2i} &= f_{2i} + e_{2i},\\ \hat{f}_{3i} &= f_{3i} + e_{3i}. \end{aligned} \tag{8}$$

(ii) Let $\mathbf{X}_i = (1, f_{2i}, f_{3i}, f_{2i}f_{3i})$. The key to this step is to find $\mathbf{M}$ and $\mathbf{m}$ such that $E(\mathbf{M}\,|\,\mathbf{X}_1 \ldots \mathbf{X}_n) = (1/n)\Sigma_{i=1}^n(\mathbf{X}_i'\mathbf{X}_i)$ and $E(\mathbf{m}\,|\,\mathbf{X}_1 \ldots \mathbf{X}_n, f_{11} \ldots f_{1n}) = (1/n)\Sigma_{i=1}^n(\mathbf{X}_i' f_{1i})$ so that $\boldsymbol{\alpha}$ can be estimated without bias from the estimating equation $\mathbf{M}\boldsymbol{\alpha} = \mathbf{m}$. Approximately unbiased moment estimators for $\mathbf{M}$ and $\mathbf{m}$ are formed by regarding $\hat{\mathbf{f}}_i$ as 'observations' of $\mathbf{f}_i$ and utilizing $\hat{\boldsymbol{\Sigma}}_{\mathbf{ee}}$ as an estimate of the measurement error variance in the errors-in-variables model. An algorithm for deriving **M** and **m** is described in Wall and Amemiya (2000) for the general polynomial structural model. Here we give $\hat{\mathbf{M}}$ and $\hat{\mathbf{m}}$ for the cross-product model:

$$\hat{\mathbf{M}} = \frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} 1 & \hat{f}_{2i} & \hat{f}_{3i} & \hat{f}_{2i}\hat{f}_{3i}-\hat{\sigma}_{23} \\ \hat{f}_{2i} & \hat{f}_{2i}^2-\hat{\sigma}_{22} & \hat{f}_{2i}\hat{f}_{3i}-\hat{\sigma}_{23} & a_{42} \\ \hat{f}_{3i} & \hat{f}_{2i}\hat{f}_{3i}-\hat{\sigma}_{23} & \hat{f}_{31}^2-\hat{\sigma}_{33} & a_{43} \\ \hat{f}_{2i}\hat{f}_{3i}-\hat{\sigma}_{23} & a_{42} & a_{43} & a_{44} \end{pmatrix}, \tag{9}$$

$$\begin{aligned} a_{42} &= \hat{f}_{2i}^2\hat{f}_{3i} - 2\hat{\sigma}_{23}\hat{f}_{2i} - \hat{\sigma}_{22}\hat{f}_{3i} - \hat{\mu}_2^3, \\ a_{43} &= \hat{f}_{2i}\hat{f}_{3i}^2 - 2\hat{\sigma}_{23}\hat{f}_{3i} - \hat{\sigma}_{33}\hat{f}_{2i} - \hat{\mu}_3^3, \\ a_{44} &= \hat{f}_{2i}^2\hat{f}_{3i}^2 - \hat{\sigma}_{33}\hat{f}_{2i}^2 - \hat{\sigma}_{22}\hat{f}_{3i}^2 - 4\hat{\sigma}_{23}\hat{f}_{2i}\hat{f}_{3i} + 2\hat{\sigma}_{22}\hat{\sigma}_{33} \\ &\quad + 4\hat{\sigma}_{23}^2 - 2\hat{\mu}_3^3\hat{f}_{2i} - 2\hat{\mu}_2^3\hat{f}_{3i} - \hat{\mu}_1^4, \end{aligned}$$

$$\hat{\mathbf{m}} = \frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} \hat{f}_{1i} \\ \hat{f}_{1i}\hat{f}_{2i}-\hat{\sigma}_{12} \\ \hat{f}_{1i}\hat{f}_{3i}-\hat{\sigma}_{13} \\ \hat{f}_{1i}\hat{f}_{2i}\hat{f}_{3i}-\hat{\sigma}_{12}\hat{f}_{3i}-\hat{\sigma}_{13}\hat{f}_{2i}-\hat{\sigma}_{23}\hat{f}_{1i}-\hat{\mu}_1^3 \end{pmatrix}. \tag{10}$$

To understand the form of (9) and (10), consider, for example, the specific element a_{42}. It is derived so that $E(a_{42} \mid f_{2i}, f_{3i}) = f_{2i}^2 f_{3i}$ (when the measurement model parameters are known). Verifying this, we see

$$\begin{aligned} E(a_{42} \mid f_{2i}, f_{3i}) &= E(\hat{f}_{2i}^2\hat{f}_{3i} - 2\hat{\sigma}_{23}\hat{f}_{2i} - \hat{\sigma}_{22}\hat{f}_{3i} - \hat{\mu}_2^3 \mid f_{2i}, f_{3i}) \\ &= E((f_{2i}+e_{2i})^2(f_{3i}+e_{3i}) - 2\hat{\sigma}_{23}(f_{2i}+e_{2i}) \\ &\quad - \hat{\sigma}_{22}(f_{3i}+e_{3i}) - \hat{\mu}_2^3 \mid f_{2i}, f_{3i}) \\ &= f_{2i}^2 f_{3i} + 2f_{2i}E(e_{2i}e_{3i}) + f_{3i}E(e_{2i}^2) + E(e_{2i}^2 e_{3i}) \\ &\quad - 2\sigma_{23}f_{2i} - \sigma_{22}f_{3i} - \mu_2^3 \\ &= f_{2i}^2 f_{3i}. \end{aligned}$$

(iii) Put $\hat{\mathbf{M}}$ and $\hat{\mathbf{m}}$ together to obtain the 2SMM estimator $\hat{\boldsymbol{\alpha}}' = (\hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3)$, that is,

$$\hat{\boldsymbol{\alpha}} = \hat{\mathbf{M}}^{-1}\hat{\mathbf{m}}. \tag{11}$$

This procedure does not require the factors $\mathbf{f}_i$ to be normally distributed, which is very useful since obviously f_{1i}, f_{2i} and f_{3i} cannot all be normal if (3) is the correct model. It also does not require the measurement errors $\boldsymbol{\epsilon}_i$ to be normally distributed. Wall and Amemiya (2000) show that $\hat{\boldsymbol{\alpha}}$ is consistent and asymptotically normally distributed without assuming normality for the factors or the measurement errors.

3. The 2SMM standard errors

In Section 2 we outlined the 2SMM procedure for obtaining the consistent estimator $\hat{\boldsymbol{\alpha}}$. Although normality was not needed in the derivation of $\hat{\boldsymbol{\alpha}}$, normality of the measurement errors (but not the factors) is needed in the following derivation of its asymptotic standard errors. That is, we consider standard errors for $\hat{\boldsymbol{\alpha}}$ formed under scenario 2. The asymptotic standard errors are obtained from the following estimator of the asymptotic covariance matrix given by Wall and Amemiya (2000),

$$\hat{V}\{\hat{\boldsymbol{\alpha}}\} = \frac{1}{n}\hat{\mathbf{M}}^{-1}\hat{\boldsymbol{\Omega}}\hat{\mathbf{M}}^{-1}. \tag{12}$$

If the $\hat{f}_i$ in (8) were actually observed and $\boldsymbol{\Sigma}_{ee}$ was known, then $\hat{\boldsymbol{\Omega}}$ in (12) would have the form $(1/n)\Sigma_{i=1}^{n}\boldsymbol{\ell}_i(\hat{\boldsymbol{\alpha}})\boldsymbol{\ell}_i(\hat{\boldsymbol{\alpha}})'$, where $\boldsymbol{\ell}_i(\hat{\boldsymbol{\alpha}}) = \hat{\mathbf{m}}_i - \hat{\mathbf{M}}_i\hat{\boldsymbol{\alpha}}$ and $\hat{\mathbf{M}}_i$ and $\hat{\mathbf{m}}_i$ are the ith terms in the sums (9) and (10), respectively. But, since $\hat{f}_i$ and $\hat{\boldsymbol{\Sigma}}_{ee}$ appearing in $\hat{\mathbf{M}}_i$ and $\hat{\mathbf{m}}_i$ have been estimated instead of being known quantities, we need to incorporate additional variability into $\hat{\boldsymbol{\Omega}}$.

To do this, the asymptotic covariance matrix for the vector of all measurement model estimators which were used in forming $\hat{\mathbf{m}}$ and $\hat{\mathbf{M}}$ must be evaluated. Let $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}_0', (\mathrm{vec}\,\hat{\boldsymbol{\beta}}_1)', (\mathrm{vec}\,\hat{\boldsymbol{\Gamma}})', (\mathrm{vech}\,\hat{\boldsymbol{\Sigma}}_{ee})')'$ where vec stacks the columns of any matrix and vech stacks the non-duplicated elements of a symmetric matrix. The estimator of the asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}$ is denoted by $\hat{\boldsymbol{\Upsilon}}$. The part of $\hat{\boldsymbol{\Upsilon}}$ directly related to $\hat{\boldsymbol{\beta}}_0$ and $\hat{\boldsymbol{\beta}}_1$, can be obtained from the output in standard SEM software packages, while the part related to $\hat{\boldsymbol{\Gamma}}$ and $\hat{\boldsymbol{\Sigma}}_{ee}$ can be obtained using the delta method, shown in the Appendix. Actually, the part of $\hat{\boldsymbol{\Upsilon}}$ corresponding to $\hat{\boldsymbol{\Gamma}}$ does not need to be estimated since it is eventually multiplied by a vector of zeros.

Since $\hat{\mathbf{f}}_i$ and $\hat{\boldsymbol{\Sigma}}_{ee}$ are functions of $\hat{\boldsymbol{\theta}}$, the following estimator $\hat{\boldsymbol{\Omega}}$ which properly incorporates variability due to $\hat{\boldsymbol{\theta}}$ is used:

$$\hat{\boldsymbol{\Omega}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\ell}_i(\hat{\boldsymbol{\alpha}})\boldsymbol{\ell}_i(\hat{\boldsymbol{\alpha}})' + \hat{\mathbf{C}}\hat{\boldsymbol{\Upsilon}}\hat{\mathbf{C}}', \tag{13}$$

where $\boldsymbol{\ell}_i(\boldsymbol{\alpha}) = \hat{\mathbf{m}}_i - \hat{\mathbf{M}}_i\boldsymbol{\alpha}$, $\hat{\mathbf{M}}_i$ and $\hat{\mathbf{m}}_i$ are the ith terms in the sums (9) and (10), respectively, and $\hat{\mathbf{C}}$ is derived from the first-order Taylor series expansions $\boldsymbol{\ell}_i(\boldsymbol{\alpha})$ with respect to $\hat{\boldsymbol{\theta}}$. In order to define $\hat{\mathbf{C}}$ the following notation for derivatives is needed. Let $\mathbf{a} = (a_1, a_2, \ldots, a_I)'$ be an $I \times 1$ vector and $\mathbf{b} = (b_1, b_2, \ldots, b_J)'$ be a $J \times 1$ vector. Define

$$\frac{\partial \mathbf{a}}{\partial \mathbf{b}'} = \begin{bmatrix} \frac{\partial a_1}{\partial b_1} & \frac{\partial a_1}{\partial b_2} & \cdots & \frac{\partial a_1}{\partial b_J} \\ \frac{\partial a_2}{\partial b_1} & \frac{\partial a_2}{\partial b_2} & \cdots & \frac{\partial a_2}{\partial b_J} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial a_I}{\partial b_1} & \frac{\partial a_I}{\partial b_2} & \cdots & \frac{\partial a_I}{\partial b_J} \end{bmatrix}.$$

Notice that $\boldsymbol{\ell}_i(\boldsymbol{\alpha})$ is a direct function of $\hat{\mathbf{f}}_i = (\hat{f}_{1i}, \hat{f}_{2i}, \hat{f}_{3i})'$ and $\mathrm{vech}\,\hat{\boldsymbol{\Sigma}}_{ee} = (\hat{\sigma}_{11}, \hat{\sigma}_{12}, \hat{\sigma}_{13}, \hat{\sigma}_{22}, \hat{\sigma}_{23}, \hat{\sigma}_{33})'$. Moreover, $\hat{\mathbf{f}}_i$ is a direct function of the first three parts of $\hat{\boldsymbol{\theta}}$, i.e. $(\hat{\boldsymbol{\beta}}_0', (\mathrm{vec}\,\hat{\boldsymbol{\beta}}_1)', \mathrm{vec}\,\hat{\boldsymbol{\Gamma}})')$. Recognizing this, we can use the chain rule to write

$$\frac{\partial \boldsymbol{\ell}_i(\boldsymbol{\alpha})}{\partial \hat{\boldsymbol{\theta}}'} = \left[\frac{\partial \boldsymbol{\ell}_i(\boldsymbol{\alpha})}{\partial \hat{\mathbf{f}}_i'}\frac{\partial \hat{\mathbf{f}}_i}{\partial \hat{\boldsymbol{\beta}}_0'}, \frac{\partial \boldsymbol{\ell}_i(\boldsymbol{\alpha})}{\partial \hat{\mathbf{f}}_i'}\frac{\partial \hat{\mathbf{f}}_i}{\partial (\mathrm{vec}\,\hat{\boldsymbol{\beta}}_1)'}, \frac{\partial \boldsymbol{\ell}_i(\boldsymbol{\alpha})}{\partial \hat{\mathbf{f}}_i'}\frac{\partial \hat{\mathbf{f}}_i}{\partial (\mathrm{vec}\,\hat{\boldsymbol{\Gamma}})'}, \frac{\partial \boldsymbol{\ell}_i(\boldsymbol{\alpha})}{\partial [\mathrm{vech}\,\hat{\boldsymbol{\Sigma}}_{ee}]'}\right].$$

Thus $\hat{\mathbf{C}}$ is constructed as a consistent estimator for the expected value of $\partial \boldsymbol{\ell}_i(\boldsymbol{\alpha})/\partial \hat{\boldsymbol{\theta}}'$:

$$\hat{\mathbf{C}} = \frac{1}{n}\sum_{i=1}^{n}\left[\left.\frac{\partial \boldsymbol{\ell}_i(\boldsymbol{\alpha})}{\partial \hat{\mathbf{f}}_i'}\right|_{\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}}}\hat{\boldsymbol{\Gamma}}, \left.\frac{\partial \boldsymbol{\ell}_i(\boldsymbol{\alpha})}{\partial \hat{\mathbf{f}}_i'}\right|_{\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}}}\hat{\boldsymbol{\Gamma}}(\mathbf{x}_i \otimes \mathbf{I}_{(p-3)}), \mathbf{0}_{4\times 3(p-3)}, \left.\frac{\partial \boldsymbol{\ell}_i(\boldsymbol{\alpha})}{\partial [\mathrm{vech}\,\hat{\boldsymbol{\Sigma}}_{ee}]'}\right|_{\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}}}\right], \tag{14}$$

where $\mathbf{x}_i = [(\mathbf{0}_{3\times(p-3)}, \mathbf{I}_3)\mathbf{z}_i]$, that is, $\mathbf{x}_i$ is the last three elements of $\mathbf{z}_i$ in the measurement model (2) identified to equal $\mathbf{f}_i$ plus measurement error. The part of $\hat{\mathbf{C}}$ corresponding to $(\partial \boldsymbol{\ell}_i(\boldsymbol{\alpha})/\partial \hat{\mathbf{f}}_i')(\partial \hat{\mathbf{f}}_i/\partial(\text{vec}\,\hat{\boldsymbol{\Gamma}})')$ is fixed to zero in (14). This is so because $\partial \hat{\mathbf{f}}_i/\partial(\text{vec}\,\hat{\boldsymbol{\Gamma}})'$ is a matrix containing only zeros and residual like terms that have asymptotically zero means. Since the residual-like terms are also approximately independent of the terms appearing in $(\partial \boldsymbol{\ell}_i(\boldsymbol{\alpha})/\partial \hat{\mathbf{f}}_i')|_{\boldsymbol{\theta},\hat{\boldsymbol{\alpha}}}$, the expected value of their product is approximated by zero.

Notice that each column of $\hat{\mathbf{C}}$ in (14) corresponds to an element in $\hat{\boldsymbol{\theta}}$ and they are in the same order. Thus when $\hat{\mathbf{C}}\hat{\boldsymbol{\Upsilon}}\hat{\mathbf{C}}'$ is formed in (13) there will be no contribution from the covariance of vec $\hat{\boldsymbol{\Gamma}}$ in $\hat{\boldsymbol{\Upsilon}}$ since it corresponds to the columns of zeros in (14). Hence the derivation of $\hat{\boldsymbol{\Upsilon}}$ is simplified because there is no need to estimate the asymptotic covariances corresponding to vec $\hat{\boldsymbol{\Gamma}}$. For this reason, we only show the delta method for the covariances of vech $\hat{\boldsymbol{\Sigma}}_{\mathbf{ee}}$ in the Appendix.

To specify $\hat{\mathbf{C}}$ in (14) completely, we express $\partial \boldsymbol{\ell}_i(\boldsymbol{\alpha})/\partial \hat{\mathbf{f}}_i'$ and $\partial \boldsymbol{\ell}_i(\boldsymbol{\alpha})/\partial[\text{vech}\,\hat{\boldsymbol{\Sigma}}_{\mathbf{ee}}]'$ for the cross-product model (2)–(3). Recall that $\boldsymbol{\ell}_i(\boldsymbol{\alpha}) = \hat{\mathbf{m}}_i - \hat{\mathbf{M}}_i\boldsymbol{\alpha}$, where $\hat{\mathbf{M}}_i$ and $\hat{\mathbf{m}}_i$ are the ith terms in the sums (9) and (10), respectively. So we have

$$\frac{\partial \boldsymbol{\ell}_i(\boldsymbol{\alpha})}{\partial \hat{\mathbf{f}}_i'} = \left[\frac{\partial \boldsymbol{\ell}_i(\boldsymbol{\alpha})}{\partial \hat{f}_{1i}}, \frac{\partial \boldsymbol{\ell}_i(\boldsymbol{\alpha})}{\partial \hat{f}_{2i}}, \frac{\partial \boldsymbol{\ell}_i(\boldsymbol{\alpha})}{\partial \hat{f}_{3i}}\right],$$

$$\frac{\partial \boldsymbol{\ell}_i(\boldsymbol{\alpha})}{\partial \hat{f}_{1i}} = \begin{bmatrix} 1 \\ \hat{f}_{2i} \\ \hat{f}_{3i} \\ \hat{f}_{2i}\hat{f}_{3i} - \hat{\sigma}_{23} \end{bmatrix},$$

$$\frac{\partial \boldsymbol{\ell}_i(\boldsymbol{\alpha})}{\partial \hat{f}_{2i}} = \begin{bmatrix} 0 \\ \hat{f}_{1i} \\ 0 \\ \hat{f}_{1i}\hat{f}_{3i} - \hat{\sigma}_{13} \end{bmatrix} - \begin{bmatrix} 0 & 1 & 0 & \hat{f}_{3i} \\ 1 & 2\hat{f}_{2i} & \hat{f}_{3i} & 2\hat{f}_{2i}\hat{f}_{3i} - 2\hat{\sigma}_{23} \\ 0 & \hat{f}_{3i} & 0 & \hat{f}_{3i}^2 - \hat{\sigma}_{33} \\ \hat{f}_{3i} & 2\hat{f}_{2i}\hat{f}_{3i} - 2\hat{\sigma}_{23} & \hat{f}_{3i}^2 - \hat{\sigma}_{33} & 2\hat{f}_{2i}\hat{f}_{3i}^2 - 2\hat{\sigma}_{33}\hat{f}_{2i} - 4\hat{\sigma}_{23}\hat{f}_{3i} \end{bmatrix}\boldsymbol{\alpha},$$

$$\frac{\partial \boldsymbol{\ell}_i(\boldsymbol{\alpha})}{\partial \hat{f}_{3i}} = \begin{bmatrix} 0 \\ 0 \\ \hat{f}_{1i} \\ \hat{f}_{1i}\hat{f}_{2i} - \hat{\sigma}_{12} \end{bmatrix} - \begin{bmatrix} 0 & 0 & 1 & \hat{f}_{2i} \\ 0 & 0 & \hat{f}_{2i} & \hat{f}_{2i}^2 - \hat{\sigma}_{22} \\ 1 & \hat{f}_{2i} & 2\hat{f}_{3i} & 2\hat{f}_{2i}\hat{f}_{3i} - 2\hat{\sigma}_{23} \\ \hat{f}_{2i} & \hat{f}_{2i}^2 - \hat{\sigma}_{22} & 2\hat{f}_{2i}\hat{f}_{3i} - 2\hat{\sigma}_{23} & 2\hat{f}_{2i}^2\hat{f}_{3i} - 2\hat{\sigma}_{22}\hat{f}_{3i} - 4\hat{\sigma}_{23}\hat{f}_{2i} \end{bmatrix}\boldsymbol{\alpha},$$

$$\frac{\partial \boldsymbol{\ell}_i(\boldsymbol{\alpha})}{\partial[\text{vech}\,\hat{\boldsymbol{\Sigma}}_{\mathbf{ee}}]'} = \begin{bmatrix} 0 & 0 & 0 & 0 & \alpha_3 & 0 \\ 0 & -1 & 0 & \alpha_1 + \alpha_3\hat{f}_{3i} & \alpha_2 + 2\alpha_3\hat{f}_{2i} & \alpha_1 + \alpha_3\hat{f}_{2i} \\ 0 & 0 & -1 & 0 & \alpha_1 + 2\alpha_3\hat{f}_{3i} & 0 \\ 0 & -\hat{f}_{3i} & -\hat{f}_{2i} & \alpha_1 + \alpha_3(\hat{f}_{3i}^2 - \hat{\sigma}_{33}) & a_{45} & \alpha_1 + \alpha_3(\hat{f}_{2i}^2 - \hat{\sigma}_{22}) \end{bmatrix},$$

$$a_{45} = -\hat{f}_{1i} + \alpha_0 + 2\alpha_1\hat{f}_{2i} + 2\alpha_2\hat{f}_{3i} + 4\alpha_3(\hat{f}_{2i}\hat{f}_{3i} - \hat{\sigma}_{23}).$$

Now with all the pieces of (14) and the necessary terms in $\hat{\boldsymbol{\Upsilon}}$, we can form $\hat{\boldsymbol{\Omega}}$ in (13). The standard errors of $\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3)'$ under scenario 2 are calculated as the square root of the diagonal of (12). We will label the resulting standard errors as $s_{\hat{\alpha}_0}$, $s_{\hat{\alpha}_1}$, $s_{\hat{\alpha}_2}$, and $s_{\hat{\alpha}_3}$. Wall and Amemiya (2000) show that the limiting distribution of $\hat{\boldsymbol{\alpha}}$ is normal so that these standard errors can be used to form $(1-c)100\%$ confidence intervals by taking

$$\hat{\alpha}_j \pm z_{c/2}s_{\hat{\alpha}_j} \qquad j = 0, 1, 2, 3 \tag{15}$$

where $z_{c/2}$ is the $1 - c/2$ quantile from the normal distribution. These confidence intervals can also be used to examine if the parameters α_0, α_1, α_2 and α_3 differ from zero by chance or not.

We recognize that the construction of the standard errors given in this section may seem a bit tedious. This is undoubtedly due to the second term in (13) involving $\hat{\mathbf{C}}$ and $\hat{\mathbf{T}}$. If the measurement model parameters were known, the second term in (13) would disappear, leaving

$$\hat{\Omega}^* = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\ell}_i(\hat{\boldsymbol{\alpha}})\boldsymbol{\ell}_i(\hat{\boldsymbol{\alpha}})'.$$

This $\hat{\Omega}^*$ underestimates var($\hat{\boldsymbol{\alpha}}$). Nevertheless to those interested in testing whether the coefficient α_3 of the cross-product term is different from zero, we recommend first using $\hat{\Omega}^*$ in place of $\hat{\Omega}$ in forming the standard errors. If the confidence interval formed using these smaller standard errors surrounds zero, there is no need to compute the correct standard error using the more complicated $\hat{\Omega}$ since it will just yield a wider confidence interval which will still surround zero. On the other hand, if the confidence interval using $\hat{\Omega}^*$ does not surround zero, then the correct (and consequently larger) standard errors detailed in this section should be used.

The normality assumption of the measurement errors required throughout this derivation of the standard errors was used only at two points.

(1) Because of normality, it was not necessary to estimate the asymptotic covariance matrix for the higher-order moments of $\mathbf{e}$ since under scenario 2 the higher-order moments are just $\mathbf{0}$ or a function of the second-order moments.
(2) Because of normality, the components in the first and second terms of the Taylor series expansion of $\boldsymbol{\ell}$ are independent, and thus (13) only involves estimates of the variances and not the covariances of the elements in the Taylor series.

Notice under scenario 3, when the distribution of the measurement errors is unspecified but the measurement model has simple structure, that the standard error derivation above only breaks down with respect to point 2. That is, we cannot be sure the covariance of the elements in the Taylor series is zero, and hence (13) should theoretically contain an extra term estimating this covariance. It is our recommendation that under scenario 3, standard errors can be reasonably estimated using (13) because this omitted covariance is expected to be quite small (albeit difficult to estimate).

4. An illustrative application

As an example of how an interaction effect between latent variables can be modelled and tested, we consider a problem that comes from a large study designed to examine socio-environmental, personal and behavioural factors associated with nutritional intake and weight status among adolescents. The study, entitled Project EAT (Neumark-Sztainer, Wall, Story, & Perry, 2000), examines the relationships among factors using a social cognitive theory framework (Bandura, 1986) with the intention of developing more effective interventions aimed at improved eating behaviours among youth.

Here we focus on just one of the many eating behaviours of interest in Project EAT: fruit and vegetable intake. The study population included a self-report survey of 3826 adolescents from 31 public middle schools and high schools in urban and suburban school districts in the St. Paul/Minneapolis area of Minnesota. As in studies of adult populations, Project EAT found that the two main correlates of fruit and vegetable intake in adolescents are the personal factor, taste preference, and the socio-environmental factor, availability. Nevertheless the traditional linear model is clearly limited here since it requires the effect that availability has on fruit and vegetable intake to be the same for all levels of taste preference and vice versa. The cross-product model allows the appropriate flexibility. For example, the cross-product model allows students with low taste preference to have minimal increase in fruit and vegetable intake when availability is increased, whereas students with higher taste preference can realize much larger gains in fruit and vegetable intake when availability increases. Treating the two factors, taste preference (*TASTE*) and availability (*AVAIL*), as latent, underlying five self-reported questionnaire responses $Z_1, \ldots, Z_5$, we fitted the cross-product model (2)-(3) with a single observed indicator for fruit and vegetable intake (*FRVEG*) as the outcome variable:

$$\begin{pmatrix} Z_{1i} \\ Z_{2i} \\ Z_{3i} \\ Z_{4i} \\ Z_{5i} \end{pmatrix} = \begin{pmatrix} \beta_{01} \\ \beta_{02} \\ \beta_{03} \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \\ \beta_{31} & \beta_{32} \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} AVAIL_i \\ TASTE_i \end{pmatrix} + \begin{pmatrix} \epsilon_{1i} \\ \epsilon_{2i} \\ \epsilon_{3i} \\ \epsilon_{4i} \\ \epsilon_{5i} \end{pmatrix}$$

$$FRVEG_i = \alpha_0 + \alpha_1 AVAIL_i + \alpha_2 TASTE_i + \alpha_3 AVAIL_i * TASTE_i + \zeta_i.$$

The $Z_{1i}, \ldots, Z_{5i}$ represent the answers by individual i to the following five questionnaire items: Z_1 = 'Fruits and vegetables are available in my home', and Z_4 = 'Vegetables are served at dinner in my home', both with responses 1 = Never, 2 = Sometimes, 3 = Usually, 4 = Always; Z_2 = 'Most vegetables taste bad', Z_3 = 'Most healthy foods just don't taste that great' and Z_5 = 'Most unhealthy foods taste better than healthy foods', all three with responses 4 = Strongly Disagree, 3 = Disagree, 2 = Agree, 1 = Strongly Agree (note the reversed direction of scoring). The outcome variable fruit and vegetable intake (*FRVEG*) is a standard scale, computed directly via the Youth and Adolescent Questions (Rockett, Wolf, & Colditz, 1995), which is a weighted sum of reported standard serving sizes from a comprehensive list of fruits and vegetables.

Following the steps outlined in Section 2, we first obtain the parameter estimates associated with the measurement model. The exploratory measurement model suggests the possibility of simple structure, thus a confirmatory model is subsequently fitted with $\beta_{21} = \beta_{31} = \beta_{12} = 0$. The confirmatory model yields a chi-square value of 31.76 on 4 degrees of freedom, an RMSEA of 0.0426 with a 90% confidence interval [0.0296, 0.0569] and an NFI of 0.999. Because the sample size is large, we ignore the significant chi-square test and rely on the other goodness-of-fit statistics to conclude that the confirmatory model gives a reasonable fit to the data. The estimated parameters for the measurement model are

$$\hat{\boldsymbol{\beta}}_0 = \begin{pmatrix} 1.02 \\ 0.57 \\ 0.64 \end{pmatrix}, \quad \hat{\boldsymbol{\beta}}_1 = \begin{pmatrix} 0.75 & 0 \\ 0 & 0.84 \\ 0 & 0.59 \end{pmatrix}, \hat{\boldsymbol{\Upsilon}} = \mathrm{diag}(0.34, 0.41, 0.57, 0.32, 0.24). \quad (16)$$

The factor score estimates for *AVAIL* and *TASTE* are obtained using (5). Since *FRVEG*, the dependent variable in the structural model, is a directly observed variable, it is not necessary to create its factor score estimate. That is, the factor score estimate for *FRVEG* is simply taken to be the observed value for *FRVEG*. Thus, letting $FRVEG = f_1$, *AVAIL* $= f_2$, and *TASTE* $= f_3$ and following the notation in (7), we have

$$\hat{\Sigma}_{ee} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \hat{\sigma}_{22} & 0 \\ 0 & 0 & \hat{\sigma}_{33} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0.21 & 0 \\ 0 & 0 & 0.15 \end{pmatrix}. \tag{17}$$

Because of the simple structure of the measurement model, this example falls under scenario 3 and thus the higher-order moments are estimated simply as $\hat{\mu}_1^3 = 0$, $\hat{\mu}_2^3 = 0$, $\hat{\mu}_3^3 = 0$, and $\hat{\mu}_1^4 = \hat{\sigma}_{22}\hat{\sigma}_{33}$.

To begin Stage 2 of the 2SMM procedure, we rewrite the problem as

$$FRVEG_i = \alpha_0 + \alpha_1 AVAIL_i + \alpha_2 TASTE_i + \alpha_3 AVAIL_i * TASTE_i + \zeta_i,$$

$$\widehat{FRVEG}_i = FRVEG_i,$$

$$\widehat{AVAIL}_i = AVAIL_i + e_{2i},$$

$$\widehat{TASTE}_i = TASTE_i + e_{3i},$$

where $\widehat{\text{Var}}((e_2, e_3)')$ is found in (17) and no measurement error is associated with *FRVEG*. Substituting *FRVEG*, $\widehat{AVAIL}$, $\widehat{TASTE}$, and $\hat{\Sigma}_{ee}$ into (9) and (10) and combining as in (11), we obtain $\hat{\alpha} = (0.80, 0.0002, -0.07, 0.09)$. Standard errors are obtained using (12). We first calculate standard errors using $\hat{\Omega}^*$ in (15) and obtain $se_{\hat{\alpha}_0} = 0.231$, $se_{\hat{\alpha}_1} = 0.080$, $se_{\hat{\alpha}_2} = 0.095$, $se_{\hat{\alpha}_3} = 0.027$. As mentioned in Section 3, these values do not contain the influence of having estimated the parameters in the first stage of the procedure and are therefore too small. But, as this additional component is usually quite small they give a good first glance at the magnitude of each parameter. Now we calculate $\hat{\Omega}$ in (13) and obtain the standard errors $se_{\hat{\alpha}_0} = 0.260$, $se_{\hat{\alpha}_1} = 0.082$, $se_{\hat{\alpha}_2} = 0.099$ and $se_{\hat{\alpha}_3} = 0.029$. We see that the a 95% confidence interval for the cross-product term α_3 is $(0.033, 0.147)$; thus we conclude that it is significant in the model.

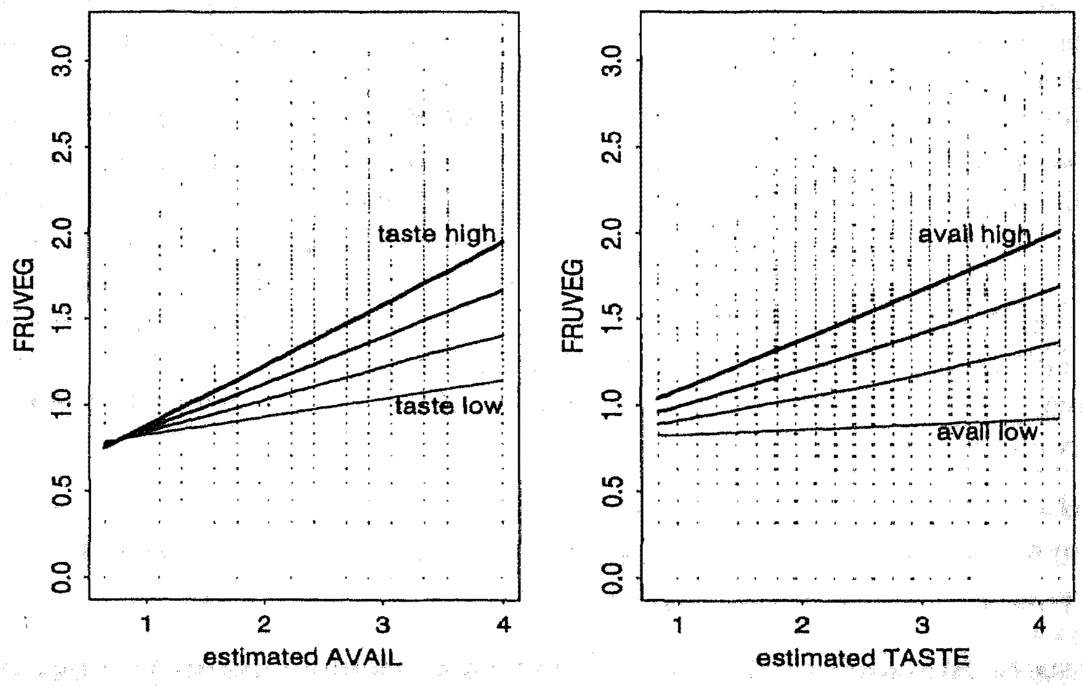

Figure 1. Scatter plot of *FRVEG* against factor score estimates of (a) *AVAIL* and (b) *TASTE* with cross-product model overlaid for four different levels of (a) *TASTE* and (b) *AVAIL* ranging from low to high (thin to thick lines).

Figure 1 presents the resulting cross-product model relating these three variables. The scatter plot in Fig. 1(a) shows the relationship between number of fruit and vegetable servings (*FRVEG*) and the factor score estimate of availability (*AVAIL*) by four different levels of taste preference ranging from low preference for vegetables (thin line) to high preference to vegetables (thick line). Likewise, the scatter plot in Fig. 1(b) shows the relationship between *FRVEG* and *TASTE* at different levels of *AVAIL* ranging from low to high (thin to thick lines, respectively). Note how the interaction term in the model is realized by the non-parallel lines drawn in each plot. For example, students with low fruit and vegetable availability have hardly any gains in fruit and vegetable intake despite their low or high taste preference for vegetables. This relationship would go unnoticed using a strictly linear model and shows the usefulness of incorporating a cross-product into structural equation models in general.

5. Simulation study

In this section we demonstrate the performance of the 2SMM procedure compared to three existing methods briefly described in the introduction: the generalized appended product indicator (GAPI) procedure introduced by Wall and Amemiya (2001), the instrumental variable (B-IV) method of Bollen (1995), and the Kenny–Judd (KJ) approach proposed by Kenny and Judd (1984).

We consider model (2)–(3) for a nine-dimensional vector of observations $\mathbf{z}$ where the measurement model has a simple structure. Although the 2SMM procedure can be applied to a general measurement model without simple structure, we consider the simple block-diagonal measurement model here because the other existing procedures have only been defined for measurement models of this type. Consider the model

$$f_{1i} = \alpha_0 + \alpha_1 f_{2i} + \alpha_2 f_{3i} + \alpha_3 f_{2i} f_{3i} + \zeta_i,$$

$$\begin{pmatrix} z_{1i} \\ z_{2i} \\ z_{3i} \\ z_{4i} \\ z_{5i} \\ z_{6i} \\ z_{7i} \\ z_{8i} \\ z_{9i} \end{pmatrix} = \begin{pmatrix} \beta_{01} \\ \beta_{02} \\ \beta_{03} \\ \beta_{04} \\ \beta_{05} \\ \beta_{06} \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \beta_{11} & 0 & 0 \\ \beta_{21} & 0 & 0 \\ 0 & \beta_{32} & 0 \\ 0 & \beta_{42} & 0 \\ 0 & 0 & \beta_{53} \\ 0 & 0 & \beta_{63} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} f_{1i} \\ f_{2i} \\ f_{3i} \end{pmatrix} + \begin{pmatrix} \epsilon_{1i} \\ \epsilon_{2i} \\ \epsilon_{3i} \\ \epsilon_{4i} \\ \epsilon_{5i} \\ \epsilon_{6i} \\ \epsilon_{7i} \\ \epsilon_{8i} \\ \epsilon_{9i} \end{pmatrix}. \tag{18}$$

For simulation, we use the model (18) with $\alpha_0 = 2$, $\alpha_1 = 1$, $\alpha_2 = 1$, and $\alpha_3 = 2$. The true measurement model parameters were $\beta_{0j} = j, j = 1, 2, \ldots, 6$, $\beta_{11} = 0.5$, $\beta_{21} = 0.4$, $\beta_{32} = 0.7$, $\beta_{42} = 0.3$, $\beta_{53} = 0.4$, and $\beta_{63} = 0.8$. We generated f_{2i} and f_{3i} as uniform random variables with $\mu_{f2} = -0.5$, $\mu_{f3} = 0.5$, $\mathrm{Var}(f_{2i}) = \mathrm{Var}(f_{3i}) = 1$, and $\mathrm{Cov}(f_{2i}, f_{3i}) = 0.5$. The error terms ζ_i and $\epsilon_{ji}, j = 1, \ldots, 9$, were independent normal random variables with $\mathrm{Var}\{\zeta_i\} = 0.3$ and $\mathrm{Var}\{\epsilon_{ji}\}$ chosen so that the reliability for each observed z_{ji} was a constant 0.75. In Fig. 2 we present an example of the generated data which contains the observed outcome variables $z_1, \ldots, z_9$ from a sample of size 200. Since f_1 is a nonlinear function of f_2 and f_3, we can see in the scatter-plot matrix that the

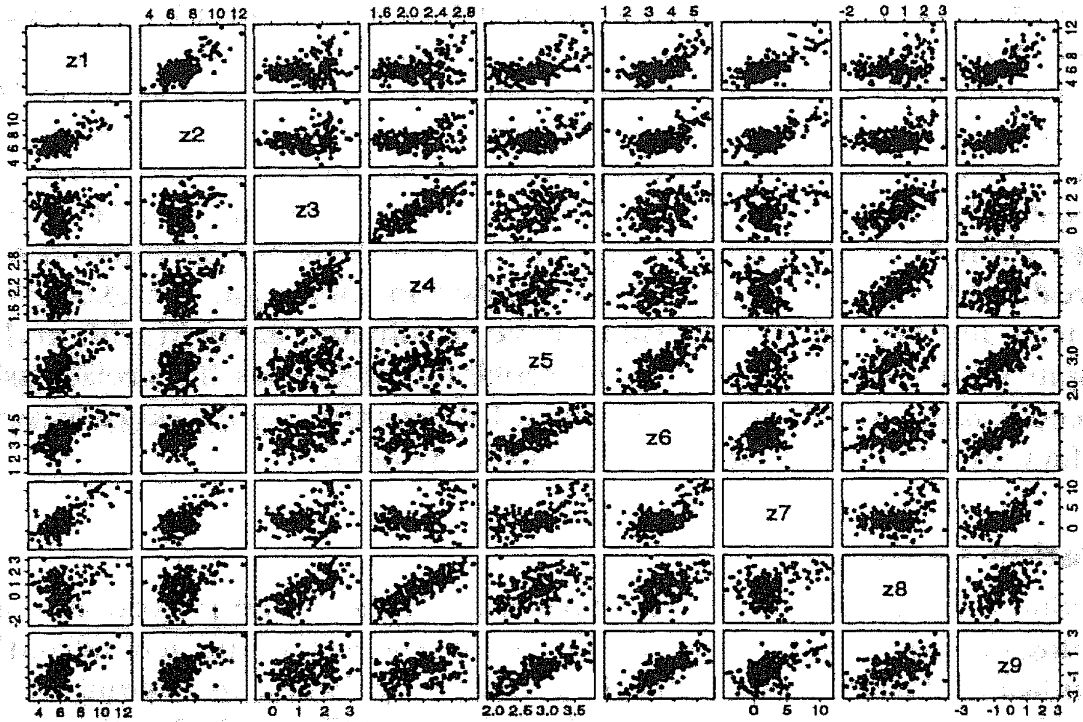

Figure 2. Scatterplot matrix of one realization of size $n = 200$.

indicators of f_1, that is, z_1, z_2, and z_7, appear to have nonlinear relationships with the other indicators.

For comparing the different procedures, we considered three sample sizes, $n = 200$, 500 and 1000. For each sample size, we generated 1000 samples and applied all four estimation procedures. While, the 2SMM procedure is described uniquely, the other estimators, GAPI, B-IV and KJ, involve some choices. For the KJ and GAPI methods, we used the normal likelihood as the discrepancy function, and included all available product indicators, $z_{3i}z_{5i}$, $z_{3i}z_{6i}$, $z_{3i}z_{9i}$, $z_{4i}z_{5i}$, $z_{4i}z_{6i}$, $z_{4i}z_{9i}$, $z_{8i}z_{5i}$, $z_{8i}z_{6i}$ and $z_{8i}z_{9i}$, based on the general recommendation made by Wall and Amemiya (2001). For the B-IV procedure, following the rules described in Bollen and Paxton (1998), we used z_{8i}, z_{9i} and $z_{8i}z_{9i}$ as the explanatory variables measured with error, and z_{3i}, z_{4i}, z_{5i}, z_{6i}, $z_{3i}z_{5i}$, $z_{3i}z_{6i}$, $z_{4i}z_{5i}$ and $z_{4i}z_{6i}$ as instruments. The observed variable z_{7i} was used as the dependent variable in place of f_{1i}. The asymptotic standard errors of these three estimators were as suggested in Wall and Amemiya (2001) for the GAPI estimator, Bollen (1995) for the B-IV, and Jaccard and Wan (1995) for the KJ.

The general pattern of the estimator comparison is depicted in Fig. 3, which presents boxplots of the four estimators of the cross-product coefficient α_3 for $n = 1000$. Table 1 gives the empirical bias and root mean squared error (RMSE) of the estimators for all four estimation procedures. The inconsistency of the KJ estimator for non-normally distributed factors, as pointed out by Wall and Amemiya (2001), can be seen in Fig. 2 and Table 1. Note that Jöreskog and Yang (1996) proposed a slightly different form of the KJ estimator than the one used here, but their form shares the inconsistency property with the original KJ estimator for non-normal data. The empirical distribution of the 2SMM estimator is more tightly concentrated around the true value than that for the B-IV procedure. The empirical distributions for the 2SMM and GAPI procedures look very similar, although, as can be seen in Table 1, the empirical RMSE is always smaller for 2SMM than for GAPI. This observed, albeit slight, improvement in RMSE found in the 2SMM procedure over the GAPI procedure is consistent with results found in Wall and Amemiya (2000) for other polynomial structural models.

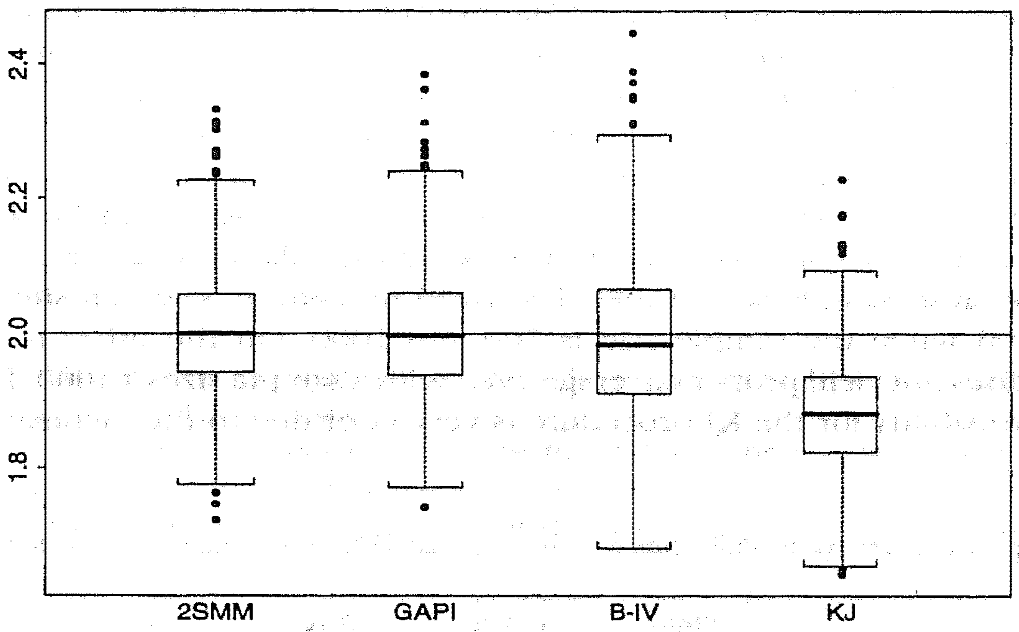

Figure 3. Boxplots for four estimators of $\alpha_3 = 2$ ($n = 1000$).

Table 1. Empirical bias and root mean squared error for four estimators of α (1000 samples). Note that values equal to 0.000 in table represent empirical results with absolute value less than 0.0005

	n		2SMM	GAPI	B-IV	KJ
α_0	200	Bias	0.002	0.003	0.013	0.166
		RMSE	0.193	0.217	0.293	0.252
	500	Bias	0.002	0.004	0.012	0.163
		RMSE	0.127	0.138	0.185	0.205
	1000	Bias	0.001	0.002	0.005	0.163
		RMSE	0.089	0.098	0.130	0.184
α_1	200	Bias	0.000	0.008	−0.017	−0.032
		RMSE	0.191	0.192	0.283	0.183
	500	Bias	0.004	0.006	−0.004	−0.031
		RMSE	0.117	0.118	0.178	0.124
	1000	Bias	−0.002	−0.004	0.004	−0.036
		RMSE	0.082	0.083	0.125	0.093
α_2	200	Bias	0.007	0.001	0.007	0.051
		RMSE	0.184	0.185	0.269	0.192
	500	Bias	0.003	0.000	0.004	0.049
		RMSE	0.113	0.114	0.168	0.125
	1000	Bias	0.000	−0.002	−0.001	0.047
		RMSE	0.081	0.083	0.122	0.097
α_3	200	Bias	0.001	−0.002	−0.040	−0.121
		RMSE	0.205	0.206	0.268	0.225
	500	Bias	0.004	0.000	−0.020	−0.116
		RMSE	0.127	0.128	0.162	0.167
	1000	Bias	0.000	−0.002	−0.012	−0.117
		RMSE	0.090	0.092	0.114	0.145

The apparent improvement of the 2SMM over the others is also seen in the coverage probabilities. Table 2 presents the empirical coverage probabilities of the nominal 95% confidence intervals for the cross-product coefficient α_3 using the four methods. For each method, the interval was obtained by taking an estimate plus or minus 1.96 times the corresponding estimated standard error. In Table 2 we see that for all sample sizes, the 2SMM interval gives the empirical coverage closest to the nominal level. In fact, its coverage is not just closest to nominal, it is within simulation error from the nominal even for the small sample size of 200. The GAPI procedure is within simulation error from nominal when the sample size is 500 and 1000. On the other hand, the B-IV procedure does not yield proper coverage even with a sample size of 1000. Likewise, the coverage probability for the KJ procedure is very poor due to the inconsistency of the estimator.

Table 2. Empirical coverage probabilities of four nominal 95% confidence intervals for α_3

n	2SMM	GAPI	B-IV	KJ
200	94.7%	91.9%	89.7%	73.4%
500	95.1%	93.8%	91.9%	64.2%
1000	94.8%	94.2%	92.2%	50.6%

6. Conclusion

In this paper, we have presented the 2SMM procedure as it applies to the special case of the cross-product structural model. Unlike its competitors, it can in general be applied to any polynomial structural model. The 2SMM has consistently better RMSE than the others. In addition, the accuracy of confidence intervals formed using the 2SMM procedure, even with samples of size 200, is very good. The complexity of the computation of the standard errors that go into these confidence intervals is probably the main drawback of the 2SMM procedure. Because of this, we have tried in Section 3 and in the Appendix to give all the details of how to calculate the standard errors with all the intermediate steps. The standard errors in the GAPI procedure are also very tedious to calculate as they come from a sandwich formula estimator with comparably tedious parts. Although the standard errors using the B-IV procedure are relatively easy to calculate, they do not perform very well even with sample size 1000.

A problem associated with any of the product indicator procedures, GAPI, B-IV and KJ, is the arbitrariness of which products of observed indicators to use. These and related (model selection) issues have been discussed by virtually all of the papers that describe these procedures, but they provide no ready solution. An advantage of the 2SMM procedure, and perhaps an explanation of its apparent improved efficiency in simulation studies, is that it uses the factor score estimates rather than arbitrary combinations of products of indicators to fit the nonlinear structural model. Wall and Amemiya (2000) show that these factor score estimates are statistically sufficient for the structural model parameters and thus incorporate all the appropriate information in the data for estimating the coefficients of the nonlinear structural model.

We hope that this paper has provided enough computational detail and explanation for researchers to implement the 2SMM method so that they will be able to include cross-product terms into their structural equation models. The SAS program implementing the 2SMM method used in this paper's simulation study can be obtained from the authors.

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.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.

Bollen, K. A. (1995). Structural equation models that are nonlinear in latent variables: A least squares estimator. *Sociological Methodology*, *25*, 223-251.

Bollen, K. A., & Paxton, P. (1998). Interactions of latent variables in structural equation models. *Structural Equation Modeling*, *5*, 266-293.

Hayduk, L. A. (1987). *Structural equation modeling with LISREL: Essentials and advances*, Baltimore, MD: Johns Hopkins University Press.

Jaccard, J., & Wan, C. K. (1995). Measurement error in the analysis of interaction effects between continuous predictors using multiple regression: Multiple indicator and structural equation approaches. *Psychological Bulletin*, *117*, 348-357.

Jaccard, J., & Wan, C. K. (1996). *LISREL approaches to interaction effects in multiple regression*. Thousand Oaks, CA: Sage.

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

Jöreskog, K. G., & Yang, F. (1997). Estimation of interaction models using the augmented moment matrix: Comparison of asymptotic standard errors. In W. Bandilla and F. Faulbaum (Eds.), *SoftStat '97 Advances in Statistical Software 6* (pp. 467-478). Stuttgart: Lucius & Lucius.

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

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

Neumark-Sztainer, D., Wall, M. M., Story, M., & Perry, C. (2000). *Correlates of fruit and vegetable intake among adolescents: Findings from the Project EAT.* Manuscript submitted for publication.

Rockett, H., Wolf, A. M., & Colditz, G. A. (1995). Development and reproducibility of a food frequency questionnaire to assess diets of older children and adolescents. *Journal of the American Dietary Association*, *95*, 336-340.

Ping, R. A. (1995). A parsimonious estimating technique for interaction and quadratic latent variables. *Journal of Marketing Research*, *32*, 336-347.

Ping, R. A. (1996). Latent variable interaction and quadratic effect estimation: A two-step technique using structural equation analysis. *Psychological Bulletin*, *119*, 166-175.

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

Wall, M. M., & Amemiya, Y. (2000). Estimation for polynomial structural equations. *Journal of the American Statistical Association*, *95*, 929-940.

Wall, M. M., & Amemiya, Y. (2001). Generalized appended product indicator procedure for nonlinear structural equation analysis. *Journal of Educational and Behavioral Statistics*, *26*, 1-29.

Received 10 July 2000; revised version received 18 December 2001

Appendix

In this appendix we give the formulae needed to construct the asymptotic covariance estimator $\hat{\Upsilon}$. Recall that $\hat{\Upsilon}$ is the estimator of the asymptotic covariance of $\hat{\theta} = (\hat{\beta}_0', (\mathrm{vec}\,\hat{\beta}_1)', (\mathrm{vec}\,\hat{\Gamma})', (\mathrm{vech}\,\hat{\Sigma}_{ee})')'$ that will be used in (13). Explicitly $\hat{\Upsilon}$ has the form

$$\begin{pmatrix} \widehat{\mathrm{Var}}(\hat{\beta}_0) & '' & '' & '' \\ \widehat{\mathrm{Cov}}(\mathrm{vec}\,\hat{\beta}_1, \hat{\beta}_0) & \widehat{\mathrm{Var}}(\mathrm{vec}\,\hat{\beta}_1) & '' & '' \\ \widehat{\mathrm{Cov}}(\mathrm{vec}\,\hat{\Gamma}, \hat{\beta}_0) & \widehat{\mathrm{Cov}}(\mathrm{vec}\,\hat{\Gamma}, \hat{\beta}_1) & \widehat{\mathrm{Var}}(\mathrm{vec}\,\hat{\Gamma}) & '' \\ \widehat{\mathrm{Cov}}(\mathrm{vech}\,\hat{\Sigma}_{ee}, \hat{\beta}_0) & \widehat{\mathrm{Cov}}(\mathrm{vech}\,\hat{\Sigma}_{ee}, \hat{\beta}_1) & \widehat{\mathrm{Cov}}(\mathrm{vech}\,\hat{\Sigma}_{ee}, \mathrm{vec}\,\hat{\Gamma}) & \widehat{\mathrm{Var}}(\mathrm{vech}\,\hat{\Sigma}_{ee}) \end{pmatrix}.$$

Recall from the argument given in Section 3 that the parts of $\hat{\Upsilon}$ containing $\mathrm{vec}\,\hat{\Gamma}$ can be ignored since they will be multiplied by zero in (13). Thus we fix $\widehat{\mathrm{Cov}}(\mathrm{vec}\,\hat{\Gamma}, \hat{\beta}_0)$, $\widehat{\mathrm{Cov}}(\mathrm{vec}\,\hat{\Gamma}, \hat{\beta}_1)$, $\widehat{\mathrm{Var}}(\mathrm{vec}\,\hat{\Gamma})$, and $\widehat{\mathrm{Cov}}(\mathrm{vech}\,\hat{\Sigma}_{ee}, \mathrm{vec}\,\hat{\Gamma})$ to zero. Recall that $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\Psi}$ are estimators obtained in the first stage of the 2SMM procedure when the measurement model is fitted. Standard SEM software packages will output the estimate for the asymptotic covariance matrix of these measurement model estimators (e.g., the OUTEST data set in SAS CALIS). This output is very useful since it will immediately give us the upper 2×2 block of $\hat{\Upsilon}$. Note also that $\hat{\Sigma}_{ee}$ can be written as a function of $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\Psi}$, thus we can obtain the covariances involving $\hat{\Sigma}_{ee}$ by incorporating the delta method. We will use the following as the estimator $\hat{\Upsilon}$:

$$\hat{\Upsilon} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ \dfrac{\partial\,\mathrm{vech}\,\hat{\Sigma}_{ee}}{\partial \hat{\beta}_0'} & \dfrac{\partial\,\mathrm{vech}\,\hat{\Sigma}_{ee}}{\partial (\mathrm{vec}\,\hat{\beta}_1)'} & \dfrac{\partial\,\mathrm{vech}\,\hat{\Sigma}_{ee}}{\partial (\mathrm{diag}\,\hat{\Psi})'} \end{pmatrix} \widehat{\mathrm{Var}}\begin{pmatrix} \hat{\beta}_0 \\ \mathrm{vec}\,\hat{\beta}_1 \\ \mathrm{diag}\,\hat{\Psi} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & \dfrac{\partial\,\mathrm{vech}\,\hat{\Sigma}_{ee}}{\partial \hat{\beta}_0'} \\ 0 & 1 & 0 & \dfrac{\partial\,\mathrm{vech}\,\hat{\Sigma}_{ee}}{\partial (\mathrm{vec}\,\hat{\beta}_1)'} \\ 0 & 0 & 0 & \dfrac{\partial\,\mathrm{vech}\,\hat{\Sigma}_{ee}}{\partial (\mathrm{diag}\,\hat{\Psi})'} \end{pmatrix}. \tag{19}$$

Now we are left with determining the derivatives of $\hat{\Sigma}_{ee}$ with respect to $\hat{\beta}_0$, $\mathrm{vec}\,\hat{\beta}_1$ and $\mathrm{diag}\,\hat{\Psi}$. The formula for $\hat{\Sigma}_{ee}$ given in (7) can be rewritten as

$$\hat{\Sigma}_{ee} = \hat{\Psi}_{22} - \hat{\Psi}_{22}\hat{\beta}_1'(\hat{\Psi}_{11} + \hat{\beta}_1\hat{\Psi}_{22}\hat{\beta}_1')^{-1}\hat{\beta}_1\hat{\Psi}_{22}$$

where $\hat{\Psi}_{22}$ is the $k \times k$ block of $\hat{\Psi}$ corresponding to the last k observations $\mathbf{z}$ defined in (2) and $\hat{\Psi}_{11}$ is the $(p-k) \times (p-k)$ block of $\hat{\Psi}$ corresponding to the first $p-k$ observations z defined in (2). First we note that $\hat{\Sigma}_{ee}$ does not depend on $\hat{\beta}_0$, so

$$\frac{\partial\,\mathrm{vech}\,\hat{\Sigma}_{ee}}{\partial \hat{\beta}_0'} = \mathbf{0}_{[k(k+1)/2] \times (p-k)}.$$

Define $\hat{\Sigma}_{vv}^{-1} = (\hat{\Psi}_{11} + \hat{\beta}_1\hat{\Psi}_{22}\hat{\beta}_1')^{-1}$ and label the columns of $\hat{\beta}_1$ by $\hat{\beta}_1 = (\hat{\mathbf{b}}_1, \ldots, \hat{\mathbf{b}}_k)$. Note that the derivatives of $\hat{\Sigma}_{ee}$ involve the derivatives of $\hat{\Sigma}_{vv}^{-1}$. The derivative of the inverse of a non-singular matrix is

$$\frac{\partial \hat{\Sigma}_{vv}^{-1}}{\partial x} = -\hat{\Sigma}_{vv}^{-1} \frac{\partial \hat{\Sigma}_{vv}}{\partial x} \hat{\Sigma}_{vv}^{-1}.$$

Thus, we first consider the derivatives of $\hat{\Sigma}_{vv}$. Let $\mathbf{E}_{ab}^{c \times d}$ represent a $c \times d$ matrix containing all zeros except for the (a, b)th element which is set equal to 1. Let $\hat{\Psi}_{11j}$

be the (j, j)th element of $\hat{\boldsymbol{\Psi}}_{11}$, $\hat{\boldsymbol{\Psi}}_{22\ell}$ be the (ℓ, ℓ)th element of $\hat{\boldsymbol{\Psi}}_{22}$ and $\hat{\boldsymbol{\beta}}_{1(j\ell)}$ be the (j, ℓ)th element of $\hat{\boldsymbol{\beta}}_1$, where $j = 1, \ldots, p-k$ and $\ell = 1, \ldots, k$. Then we have

$$\frac{\partial \hat{\boldsymbol{\Sigma}}_{\mathbf{vv}}}{\partial \boldsymbol{\Psi}_{11j}} = \mathbf{E}_{jj}^{(p-k)\times(p-k)},$$

$$\frac{\partial \hat{\boldsymbol{\Sigma}}_{\mathbf{vv}}}{\partial \boldsymbol{\Psi}_{22\ell}} = \hat{\mathbf{b}}_\ell \hat{\mathbf{b}}_\ell',$$

$$\frac{\partial \hat{\boldsymbol{\Sigma}}_{\mathbf{vv}}}{\partial \hat{\boldsymbol{\beta}}_{1(j\ell)}} = (\hat{\mathbf{b}}_\ell \mathbf{E}_{1j}^{1\times(p-k)} + \mathbf{E}_{j1}^{(p-k)\times 1} \hat{\mathbf{b}}_\ell') \hat{\boldsymbol{\Psi}}_{22\ell},$$

and thus,

$$\frac{\partial \hat{\boldsymbol{\Sigma}}_{\mathbf{ee}}}{\partial \hat{\boldsymbol{\Psi}}_{11j}} = \hat{\boldsymbol{\Psi}}_{22} \hat{\boldsymbol{\beta}}_1' \hat{\boldsymbol{\Sigma}}_{\mathbf{vv}}^{-1} \mathbf{E}_{jj}^{(p-k)\times(p-k)} \hat{\boldsymbol{\Sigma}}_{\mathbf{vv}}^{-1} \hat{\boldsymbol{\beta}}_1 \hat{\boldsymbol{\Psi}}_{22},$$

$$\begin{aligned}\frac{\partial \hat{\boldsymbol{\Sigma}}_{\mathbf{ee}}}{\partial \hat{\boldsymbol{\Psi}}_{22\ell}} = {} & \mathbf{E}_{\ell\ell}^{k\times k} - \mathbf{E}_{\ell\ell}^{k\times k} \hat{\boldsymbol{\beta}}_1' \hat{\boldsymbol{\Sigma}}_{\mathbf{vv}}^{-1} \hat{\boldsymbol{\Psi}}_{22} - \hat{\boldsymbol{\Psi}}_{22} \hat{\boldsymbol{\beta}}_1' \hat{\boldsymbol{\Sigma}}_{\mathbf{vv}}^{-1} \hat{\boldsymbol{\beta}}_1 \mathbf{E}_{\ell\ell}^{k\times k} \\ & + \hat{\boldsymbol{\Psi}}_{22} \hat{\boldsymbol{\beta}}_1' \hat{\boldsymbol{\Sigma}}_{\mathbf{vv}}^{-1} \hat{\mathbf{b}}_\ell \hat{\mathbf{b}}_\ell' \hat{\boldsymbol{\Sigma}}_{\mathbf{vv}}^{-1} \hat{\boldsymbol{\beta}}_1 \hat{\boldsymbol{\Psi}}_{22},\end{aligned}$$

$$\begin{aligned}\frac{\partial \hat{\boldsymbol{\Sigma}}_{\mathbf{ee}}}{\partial \hat{\boldsymbol{\beta}}_{1(j\ell)}} = {} & -\hat{\boldsymbol{\Psi}}_{22\ell} \mathbf{E}_{\ell j}^{k\times(p-k)} \hat{\boldsymbol{\Sigma}}_{\mathbf{vv}}^{-1} \hat{\boldsymbol{\beta}}_1 \hat{\boldsymbol{\Psi}}_{22} - \hat{\boldsymbol{\Psi}}_{22} \hat{\boldsymbol{\beta}}_1' \hat{\boldsymbol{\Sigma}}_{\mathbf{vv}}^{-1} \mathbf{E}_{j\ell}^{(p-k)\times k} \hat{\boldsymbol{\Psi}}_{22\ell} \\ & + \hat{\boldsymbol{\Psi}}_{22} \hat{\boldsymbol{\beta}}_1' \boldsymbol{\Sigma}_{\mathbf{vv}}^{-1} (\hat{\mathbf{b}}_\ell \mathbf{E}_{1j}^{1\times(p-k)} + \mathbf{E}_{j1}^{(p-k)\times 1} \hat{\mathbf{b}}_\ell') \hat{\boldsymbol{\Psi}}_{22\ell} \hat{\boldsymbol{\Sigma}}_{\mathbf{vv}}^{-1} \hat{\boldsymbol{\beta}}_1 \hat{\boldsymbol{\Psi}}_{22}.\end{aligned}$$

Taking the vech of each of these matrices of derivatives, we can then plug them into the formula for $\hat{\boldsymbol{\Upsilon}}$ given in (19), and we are done.