RC22955 (W0305-021) May 6, 2003
Mathematics

IBM Research Report

Correlated Populations with Fixed and Non-normal Latent Variables

Savas Papadopoulos
Bank of Greece
University of Indianapolis at Athens

Yasuo Amemiya
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

Research Division
Almaden - Austin - Beijing - Haifa - India - T. J. Watson - Tokyo - Zurich

CORRELATED POPULATIONS WITH FIXED AND NON-NORMAL LATENT VARIABLES[1]

BY SAVAS PAPADOPOULOS AND YASUO AMEMIYA

Bank of Greece, University of Indianapolis at Athens,

and IBM T. J. Watson Research Center

Latent variable or structural equation modeling (SEM) is used widely in applications, especially in social and behavioral sciences. Since the normality based model fitting procedures are simple and broadly available, and since such procedures are often applied to non-normal data or non-random samples, it is important to investigate the appropriateness of such practice and to suggest simple remedies. This paper addresses these issues for the analysis of multiple populations. For a very general class of latent variable models, a particular parameterization is proposed for meaningful and interpretable analysis of several populations. It is shown that under this parameterization the large-sample statistical inferences based on the assumption of normal and independent populations are valid for non-normal and dependent populations. This result is also shown to be valid when some latent variables are treated as fixed instead of random, or when a group of individuals is measured over several time points longitudinally. More precisely, the paper shows how to get robust asymptotic standard errors (a.s.e's) and overall-fit measures. The proposed a.s.e's are shown to have less variability than the robust a.s.e's computed by the so-called sandwich estimator. Simulation studies are conducted to verify the theoretical results, assess the use of asymptotic results in finite samples, show the robustness of the power for tests, and demonstrate the efficiency of the method relative to the full-likelihood estimation method that includes all the covariances of the variables over populations.

[1] KEY WORDS: Structural Equation Modeling SEM, asymptotic robustness, multivariate analysis, panel data, repeated measures, nonlinear in parameters.

1. Introduction. Latent variable analysis has been used widely in social and behavioral sciences as well as in economics, and its use in medical and business applications is becoming popular. Structural equation modeling, confirmatory factor analysis, and errors-in-variables regression are examples of latent variable analysis. In latent variable models, underlying subject-matter concepts are represented by unobservable latent variables, and their relationships with each other and with the observed variables are specified. The models that express observed variables as a linear function of latent variables are extensively used, because of their simple interpretation and of the existence of computer packages such as EQS (Bentler (1989)), LISREL (Jöreskog and Sörbom (1989)), and SAS (PROC CALIS (1990)). The standard procedures in the existing computer packages assume that all the variables are normally distributed. The normality and linearity assumptions make the analysis and the interpretation simple, but their applicability in practice is often questionable. In fact, it is rather common in many applications to use the normality-based standard errors and model-fit test procedures when observed variables are highly discrete, bounded, skewed, or generally non-normal. Thus, it is of practical and theoretical interest to examine the extent of the validity of the normality-based inference procedures for non-normal data, and to explore possible ways to parameterize and formulate a model to attain the wide applicability. In the structural equation analysis literature, this type of research is often referred to as asymptotic robustness study. Most existing results on this topic have been for a single sample from one population. This paper addresses the problem for multiple samples or multiple populations, and provides a unified and comprehensive treatment of the so-called asymptotic robustness. The emphasis here is the suggestion of proper parameterization and modeling leading to practical usefulness and to a meaningful interpretation. It is the first study that shows robust asymptotic standard errors (a.s.e's) and overall-fit measures for correlated samples with fixed factors for models with latent variables. Novel formulas are provided for the

computation of the a.s.e's for the means and variances of the fixed correlated factors. Also, in the case for random correlated factors we proved that the a.s.e's of the means for the factors are robust. It is shown numerically the superiority of the suggested a.s.e's to the existing robust a.s.e's that involve the computation of 3rd and 4th moments. The computation of such moments increases the variability of the a.s.e's.

A general latent variable model for a multivariate observation vector $\boldsymbol{v}_j^{(i)}$ with dimension $p^{(i)} \times 1$ that is an extension of those models considered by Anderson (1987, 1989), Browne and Shapiro (1988), and Satorra (1992, 1993, 1994, 1995, and 1997), is

(1) $\boldsymbol{v}_j^{(i)} = \boldsymbol{\beta}^{(i)} + \boldsymbol{B}^{(i)}\boldsymbol{\xi}_j^{(i)}$ with $\boldsymbol{\xi}_j^{(i)} = \begin{pmatrix} \boldsymbol{\zeta}_j^{(i)} \\ \boldsymbol{\varepsilon}_j^{(i)} \end{pmatrix}$, and $i = 1,\ldots,I;\ j = 1,\ldots,n^{(i)}$.

under the following set of assumptions. The model is extended with fixed and correlated-over-populations latent variables.

ASSUMPTION 1

<u>*i) Case A:*</u> $\boldsymbol{\zeta}_j^{(i)}$ *is*

a) random with vector mean $\boldsymbol{\mu}_{\boldsymbol{\zeta}^{(i)}}$ *and covariance matrix* $\boldsymbol{\Sigma}_{\boldsymbol{\zeta}^{(i)}}$

b) correlated over i (that is, the measurements of the *j-th individual of the* i_1*-th population are correlated with the corresponding measurements* of the *j-th individual of the* i_2*-th population).*

c) independent over j (for each population the measurements of the observed individuals are indepoendent)

<u>*or*</u> <u>*Case B:*</u> $\boldsymbol{\zeta}_j^{(i)}$ *is*

a) fixed with limiting vector mean $\boldsymbol{\mu}_{\boldsymbol{\zeta}^{(i)}}$ *and limiting covariance*

matrix $\boldsymbol{\Sigma}_{\boldsymbol{\zeta}^{(i)}}$

b) correlated over i (see comments in Case A-b))

ii) $\boldsymbol{\varepsilon}_j^{(i)} = \left(\boldsymbol{\varepsilon}_{0j}^{(i)\prime}, \quad \boldsymbol{\varepsilon}_{1j}^{(i)\prime}, \quad \ldots \quad \boldsymbol{\varepsilon}_{L^{(i)}j}^{(i)\prime}\right)'$ *where*

a) $\boldsymbol{\varepsilon}_{0j}^{(i)} \sim \boldsymbol{N}(0, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}_0^{(i)}})$

b) $\boldsymbol{\varepsilon}_{\ell j}^{(i)}$ $\ell = 1, \ldots, L^{(i)}$ *are independent over i,* ℓ *and j with mean 0 and covariance matrix* $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}_\ell^{(i)}}$.

iii) The intercepts $\boldsymbol{\beta}^{(i)}$, *the coefficients* $\boldsymbol{B}^{(i)}$, *and the variance matrices of the normally distributed errors* $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}_0^{(i)}}$ *can be restricted. Thus, they are assumed to be functions of a vector* $\boldsymbol{\tau}$.

iv) The mean vectors $\boldsymbol{\mu}_{\boldsymbol{\zeta}^{(i)}}$, *the variance matrices* $\boldsymbol{\Sigma}_{\boldsymbol{\zeta}^{(i)}}$ *of the correlated factors, and the variance matrices of the non-normal vectors* $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}_\ell^{(i)}}$, $\ell = 1, \ldots, L^{(i)}$ *are assumed to be unrestricted.*

A common approach to verifying the identification and fitting the model is to assume hypothetically that all $\boldsymbol{\xi}_j^{(i)\prime}s$ are normally distributed and to concentrate on the first two moments of the observed vector $\boldsymbol{v}_j^{(i)}$. The issue for the so-called asymptotic robustness study is to assess the validity of such procedures based on the assumed normality, in terms of inference for unknown parameters, for a wide class of distributional assumptions on $\boldsymbol{\xi}_j^{(i)}$. It turns out that the type of parameterization used for model , restricting the coefficient $\boldsymbol{B}^{(i)}(\boldsymbol{\tau})$ but keeping the variances $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}_\ell^{(i)}}$ of the non-normal latent variables $\boldsymbol{\varepsilon}_{\ell j}^{(i)}$ unrestricted, plays a key role in the study.

The model, the notation, and the assumptions are explained by the following example,

EXAMPLE 1. A two-population (I=2) recursive system of simultaneous equations with errors in the explanatory variables is considered. The model is shown in (2). The system in (2) can be written in the following matrix form $\boldsymbol{\nu}_j^{(i)} = \boldsymbol{\alpha}^{(i)} + \boldsymbol{\Gamma}^{(i)}\boldsymbol{\nu}_j^{(i)} + \boldsymbol{\Delta}^{(i)}\boldsymbol{\zeta}_j^{(i)} + \boldsymbol{e}_j^{(i)}$ which has the form of model (1) with, $\boldsymbol{\beta}^{(i)} = (\boldsymbol{I}^{(i)} - \boldsymbol{\Gamma}^{(i)})^{-1}\boldsymbol{\alpha}^{(i)}$, $\boldsymbol{B}^{(i)} = (\boldsymbol{I}^{(i)} - \boldsymbol{\Gamma}^{(i)})^{-1}[\boldsymbol{\Delta}^{(i)}, \boldsymbol{I}^{(i)}]$, and $\boldsymbol{\varepsilon}_j^{(i)} = \boldsymbol{e}_j^{(i)}$. The model is also a special case of the LISREL model with no latent variables in the dependent variables $\boldsymbol{y}^{(i)}$, that is $\boldsymbol{y}^{(i)} = \boldsymbol{\eta}^{(i)}$, in the LISREL notation. The latent variables $\zeta_j^{(1)}$ and $\zeta_j^{(2)}$ are correlated for each $j = 1,\ldots,500$, with correlation 0.4. That is, the measurements of each individual from the second population are correlated with the measurements of one individual from the first population. The first population also has 500 individuals that are independent from all the individuals of the second population. Note that the number of observed variables is different for the two populations. Four measurements $x_j^{(1)}$, $y_{1j}^{(1)}$, $y_{2j}^{(1)}$, $y_{3j}^{(1)}$ are taken from the first population $(p^{(1)} = 4)$ and three measurements $x_j^{(2)}$, $y_{1j}^{(2)}$, $y_{2j}^{(2)}$ from the second $(p^{(2)} = 3)$.

$$
\begin{aligned}
x_j^{(1)} &= \zeta_j^{(1)} + e_{0j}^{(1)} \\
y_{1j}^{(1)} &= \beta_1 + \delta_1\zeta_j^{(1)} + e_{1j}^{(1)} \\
y_{2j}^{(1)} &= \beta_2 + \gamma_1 y_{1j}^{(1)} + \delta_2\zeta_j^{(1)} + e_{2j}^{(1)} \\
y_{3j}^{(1)} &= \beta_3 + \gamma_2 y_{2j}^{(1)} + e_{3j}^{(1)}
\end{aligned}
\qquad (2)
$$

$$
\begin{aligned}
x_j^{(2)} &= \zeta_j^{(2)} + e_{0j}^{(2)} \\
y_{1j}^{(2)} &= \beta_1 + \delta_1\zeta_j^{(2)} + e_{1j}^{(2)} \\
y_{2j}^{(2)} &= \beta_2 + \gamma_1 y_{1j}^{(2)} + \delta_2\zeta_j^{(2)} + e_{2j}^{(2)}
\end{aligned}
$$

For $j = 1,\ldots n^{(i)}$, with $n^{(1)}$=1000 and $n^{(2)}$=500.

The parameters β_1, β_2, γ_1, δ_1, and δ_2 do not depend on *i*. That is, they are common for the two populations. These parameters belong to the vector $\boldsymbol{\tau}$. The variables $\zeta_j^{(1)}$ and $\zeta_j^{(2)}$ can be fixed or non-normal according to Cases A and B of Assumption 1. If all the errors are normal in accordance of the notation of Assumption 1 we have $\boldsymbol{\varepsilon}_{0j}^{(i)} = \boldsymbol{e}_j^{(i)}$, while if $\boldsymbol{e}_{0j}^{(i)}$ is normal and all the other errors are non-normal then $\boldsymbol{\varepsilon}_{0j}^{(i)} = e_{0j}^{(i)}$, $\boldsymbol{\varepsilon}_{\ell j}^{(i)} = e_{\ell j}^{(i)}$, for $i = 1,2;\ \mathrm{j} = 1,\ldots,\mathrm{n}^{(\mathrm{i})};\ \ell = 1,\ldots,L^{(i)}$ with $L^{(1)} = 3$ and $L^{(2)} = 2$. According to Assumption 1 only the variances of the normal errors can be restricted to be the same over populations and these variances belong to the vector $\boldsymbol{\tau}$.

Further discussion about the model in (1) is given in Section 2. The model in (2) of Example 1 is simulated in Section 4 and used as an example in this paper to explain the theory.

Latent variable analysis of multiple populations was discussed by Jöreskog (1971), Lee and Tsui (1982), Muthén (1989), and Satorra (1993a, 1993b). Consider multiple samples from several populations, as the model in (1) holds for each population. Then, the interest may be in making inferences about the similarities and differences among populations. If similar variables are measured from each population, then the parameters or characteristics associated with a measurement process are assumed to be common over the samples. Even in such a case, some of the latent variables being measured can have different characteristics (in terms of different distributional parameters) over populations. The existing computer packages, LISREL and EQS, can analyze multiple populations simultaneously under the assumption that the populations (samples) are independent. Another type of the multi-population problem is concerned with the so-called correlated populations. The multiple samples may in fact come from the same population over different time periods (multivariate repeated measures) or may be spatially correlated. See, e.g., Papadopoulos and Amemiya (1995). Such correlated or dependent

population (or sample) cases have not been fully discussed in the literature, but are treated in this paper. For both types of multi-population problems, this paper considers simple model fitting and robust procedures that can be readily carried out using the existing packages and provides asymptotic standard errors (a.s.e's) that have less variability than the existing robust a.s.e's. We discuss a general multi-population model possibly containing fixed, non-normal, and normal components, and introduce a way to formulate and parameterize the model so that the multi-population analysis can be conducted and interpreted meaningfully in practice, and so that the so-called asymptotic robustness is achieved for inferences concerning parameters.

The so-called asymptotic robustness of normal-based methods for the latent variable analysis has been extensively studied in the last fefteen years. For, exploratory (unrestricted) factor analysis, Amemiya, Fuller, and Pantula (1987) proved that the limiting distribution of some estimators is the same for fixed, non-normal, and normal factors under the assumption that the errors are normally distributed. Amemiya (1986) treated functional and structural relationships with error covariance matrix as a function of an unknown parameter vector. The robustness of goodness-of-fit tests was studied by Amemiya (1985). Browne (1987) showed that the above results hold for a more general class of latent variable models assuming finite eighth moments for the factors and normal errors. Anderson and Amemiya (1988), and Amemiya and Anderson (1990) extended the above results to confirmatory factor analysis and non-normal errors; they assume finite second moments for the factors and errors. Browne and Shapiro (1988) introduced a general linear model, and used an approach based on the finite fourth moments that differs from that of Anderson and Amemiya. Considering the model of Browne and Shapiro, Anderson (1987, 1989) included non-stochastic latent variables, and assumed only finite second moments for the non-normal latent variables. Latent variable models with mean and covariance structures were studied by Browne (1990) and Satorra (1992). Satorra (1993a, 1993b, 1993c, 1994, 1995, 1997)

considered the multi-sample analysis of augmented-moment structures. Additional studies on the asymptotic robustness of latent variable analysis were conducted by Shapiro (1987), Mooijaart and Bentler (1991), and Satorra and Bentler (1990).

For the one-sample problem, asymptotic distribution-free (ADF) methods for the latent variable analysis were proposed to deal with non-normal data. See, e.g., Bentler (1983), Browne (1984), and Muthén (1989). The ADF methods turned out to be problematic in practice, since the fourth-order sample moments are very variable. See, e.g., Chou, Bentler, and Satorra (1991) and Muthén and Kaplan (1992). Robust estimation procedures resistant to outliers and contaminated model distributions were examined for structural models from one population. Krishnakumar and Ronchetti (1977) used a robust estimator that is a generalization of MLE for linear and non-linear simultaneous equations. Zamar (1989, 1992) defined orthogonal regression S and M estimates for error-in-variables models. Huggins (1996) also considered robust estimators for the parameters from a bifurcating autoregressive process for cell lineage data.

In this paper, mean and covariance structures are considered for a general multi-population model which contains fixed, normal and non-normal variables; some of the non-normal variables are allowed to be correlated over populations. We use the approach of Anderson and Amemiya (1988) to show that the normal-based methods are applicable for non-normal and non-random data assuming finite second-order moments.

Section 2 explains the suggested parameterization and the estimation procedure. The theoretical results are derived and discussed in Section 3. Section 4 reports results from simulation studies.

2. Model, parameterization, and procedure. In this paper we study the model in (1) introduced in Section 1. We consider I populations and we assume that $n^{(i)}$ individuals are sampled from

the i^{th} population, $i = 1,\ldots,I$, and that $p^{(i)}$ measurements are taken from each sampled individual in the i^{th} population. Denote the multi-sample data set by $\boldsymbol{v}_j^{(i)} : \; i = 1,\ldots,I; j = 1,\ldots,n^{(i)}$ where $\boldsymbol{v}_j^{(i)}$ is the $p^{(i)} \times 1$ measurement vector from the *j-th* individual in the *i-th* population. We consider a very general latent variable model that includes models widely used in single population cases and covers a large class of distributional situations in one form. To cover various distributional settings, it is convenient to assume that the observed vector $\boldsymbol{v}_j^{(i)}$ can be written as a linear combination of $L^{(i)} + 2$ independent latent vectors, and that the latent vectors can be divided into three groups; a fixed or non-normal vector that is correlated over populations $\boldsymbol{\zeta}_j^{(i)}$,a random vector $\boldsymbol{\varepsilon}_{0j}^{(i)}$ assumed to be normally distributed, and $L^{(i)}$ non-normal vectors $\boldsymbol{\varepsilon}_{\ell j}^{(i)}$ $\ell = 1,\ldots,L^{(i)}$. Note that the sample size $n^{(i)}$, the number of measured variables $p^{(i)}$, and the number of latent vectors $L^{(i)}$ generally differ over populations (depend on *i*). This generality of the model allows us to deal with cases where slightly different variables are measured from different populations with possibly different structures.

All normally distributed latent variables are included in $\boldsymbol{\varepsilon}_{0j}^{(i)}$, and their distribution may possibly be related through $\boldsymbol{\tau}$ over populations, $i = 1,\ldots,I$. Other unspecified or non-normal random latent variables are divided into independent parts $\ell = 1,\ldots,L^{(i)}$ with unrestricted covariance matrices. In case A of Assumption 1 with fixed $\boldsymbol{\zeta}_j^{(i)}$ can represent a situation where the interest is in the model fitting and estimation only for a given set of individuals and not for the populations. In addition, the fixed $\boldsymbol{\zeta}_j^{(i)}$ can be used in an analysis conducted conditionally on a given set of $\boldsymbol{\zeta}_j^{(i)}$ values. Such a conditional analysis may be appropriate when the individuals $j = 1,\ldots,n^{(i)}$ do not form a random sample from the *i-th* population and/or when a component of

$\boldsymbol{\nu}_j^{(i)}$ represents some dependency over *I* populations. For example, the *I* populations may actually correspond to a single population at I different time points. With $\boldsymbol{\zeta}_j^{(i)}$ being latent and fixed, the limits of the unobservable sample mean, $\boldsymbol{\mu}_{\boldsymbol{\zeta}^{(i)}}$, and of the sample covariance matrix, $\boldsymbol{\Sigma}_{\boldsymbol{\zeta}^{(i)}}$, are assumed to be unknown and unrestricted. All $\boldsymbol{\beta}^{(i)}(\boldsymbol{\tau})$ and $\boldsymbol{B}^{(i)}(\boldsymbol{\tau})$ are expressed in terms of $\boldsymbol{\tau}$ representing known or restricted elements and allowing functional relationships over *I* populations. Even though $\boldsymbol{\tau}$ also appears in $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}_0^{(i)}}(\boldsymbol{\tau})$, the elements of $\boldsymbol{\tau}$ are usually divided into two groups; one for $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}_0^{(i)}}(\boldsymbol{\tau})$ and another for $\boldsymbol{\beta}^{(i)}(\boldsymbol{\tau})$ and $\boldsymbol{B}^{(i)}(\boldsymbol{\tau})$. Assumptions 1 iii) and iv) provide a particular identifiable parameterization for the model in (1). For a single population case with *I*=1, various equivalent parameterization have been used in practice. Some place restrictions on covariance matrices, e.g., by standardizing latent variables, and leave the coefficients unrestricted. The parameterization that leaves the covariance matrices (and possibly some mean vectors) of latent variables unrestricted and that places identification restrictions only on the coefficients and intercepts is referred to as the errors-in-variables parameterization. For the single population case, a parameterization with restricted covariance matrices generally has an equivalent errors-in-variables parameterization, and the two parameterizations with one-to-one correspondence lead to an equivalent interpretation. The one sample asymptotic robustness results have shown that the asymptotic standard errors for the parameters in the errors-in-variables formulation computed under the normality assumption are valid for non-normal data, but that the same does not hold under parameterization with restricted covariance matrices. For the multi-sample the model in (1), we will show that the errors-in-variables type parameterization, given in Assumption 1, provides the asymptotic robustness. However, for the multi-sample case, there are other reasons for considering the parameterization specified in Assumptions 1 *iii)* and i*v)*. As

mentioned earlier, a multi-population study is conducted because the populations are thought to be different, but certain aspects of the structure generating data are believed to be common over populations. Suppose that the same or similar measurements are taken from different populations. For example, a similar set of psychological tests may be given to a number of different groups, e.g., two gender groups, groups with different occupations or educational backgrounds, groups in varying socio-economic or cultural environments, or different time points in the growth of a group. The subject matter or scientific interest exists in making inferences about some general assertion that holds commonly for various populations. Such interest is usually expressed as relationships among latent (and observed) variables that hold regardless of the location and variability of the variables. Then, a relevant analysis is to estimate and test the relationships, and to explore the range of populations for which the relationships hold. The parameterization in Assumptions 1 *iii)* and i*v)* with

unrestricted $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}_\ell^{(i)}}$ and generally structured $\boldsymbol{B}^{(i)}(\boldsymbol{\tau})$ corresponds very well with the scientific interest of the study, and allows the interpretation consistent with the practical meaning of the problem. Note that $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}_\ell^{(i)}}$, $i = 1,\ldots,I$, $\ell = 1,\ldots,L^{(i)}$, are unrestricted covariance matrices and do not have any relationships over *i* or ℓ, and that $\boldsymbol{\beta}^{(i)}(\boldsymbol{\tau})$ and $\boldsymbol{B}^{(i)}(\boldsymbol{\tau})$ can have known elements and elements with relationships over *i* and ℓ. On the other hand, the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}_0^{(i)}}$ of the normal latent vector $\boldsymbol{\varepsilon}_{0j}^{(i)}$ can have restrictions or equality over populations through $\boldsymbol{\tau}$. This gives the generality of the model in (1) with only one normal latent vector, because a block diagonal $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}_0^{(i)}}$ corresponds to a number of independent sub-vectors in the normal $\boldsymbol{\varepsilon}_{0j}^{(i)}$. In addition, the possibility of restrictions on $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}_0^{(i)}}$ over populations can also be important in applications. For example, if the same measurement instruments are applied to different samples,

then the variances of pure measurement errors may be assumed to be the same over the samples. However, the normal assumption for pure measurement errors is reasonable in most situations, and such errors can be included in $\boldsymbol{\varepsilon}_{0j}^{(i)}$. Assumptions 1 *iv)* and *v)* do not rule out latent variable variances and covariances with restrictions across populations, but does require the latent variables with restricted variances to be normally distributed. This requirement is not very restrictive in most applications, as discussed above, but it is needed to obtain the asymptotic robustness results given in the next section. The general form of $\boldsymbol{\beta}^{(i)}(\boldsymbol{\tau})$ and the inclusion of fixed latent vector allow virtually any structure for the means of the observed $\boldsymbol{\nu}_j^{(i)}$. Hence, the errors-in-variables type parameterization in Assumption 1 *iii)* can solve the identification problem, provides a general and a convenient way to represent the subject-matter theory and concepts, and produces asymptotic robustness results presented in the next section.

For the multi-sample data $\boldsymbol{\nu}_j^{(i)}$ in (1), let $\bar{\boldsymbol{\nu}}^{(i)}$ and $\boldsymbol{S}_{\boldsymbol{\nu}}^{(i)}$ be the sample mean vector and sample covariance matrix (unbiased) for the *i-th* population $i = 1,\ldots,I$. It is assumed that the sample covariance matrices $\boldsymbol{S}_{\boldsymbol{\nu}}^{(i)}$ are non-singular with probability one. Define

$$\boldsymbol{c}^{(i)} = \begin{pmatrix} \bar{\boldsymbol{\nu}}^{(i)} \\ vec(\boldsymbol{S}_{\boldsymbol{\nu}}^{(i)}) \end{pmatrix}, \qquad \boldsymbol{c} = \begin{pmatrix} \boldsymbol{c}^{(1)} \\ \vdots \\ \boldsymbol{c}^{(I)} \end{pmatrix}. \tag{3}$$

We consider model fitting and estimation based only on $\boldsymbol{c}$, because such procedures are simple and have some useful properties. Also note that Assumption 1 does not specify a particular distributional form of observations beyond the first two moments, and that no particular correspondence or relationship between samples is specified in Assumption 1. Let $\boldsymbol{\theta}$ be a $d_{\boldsymbol{\theta}} \times 1$ vector containing all unknown parameters in $E(\boldsymbol{c}) = \boldsymbol{\gamma}(\boldsymbol{\theta})$ under the model in (1) and Assumption 1, and let $\boldsymbol{\theta} = (\boldsymbol{\tau}', \boldsymbol{\upsilon}')'$ where $\boldsymbol{\tau}$ and $\boldsymbol{\upsilon}$ contain the parameters mentioned in

Assumptions 1 *iv)* and 1 *v)*, respectively. That is, $\boldsymbol{\tau}$ contains parameters that can be restricted while $\boldsymbol{\upsilon}$ contains that parameters that can not be restricted over populations. Under the model in (1) and Assumption 1, we compute the following expected means,

$$\boldsymbol{\mu}_{\boldsymbol{\nu}}^{(i)}(\boldsymbol{\theta}) = E(\bar{\boldsymbol{v}}^{(i)}) \text{ and } \boldsymbol{\Sigma}_{\boldsymbol{\nu}}^{(i)}(\boldsymbol{\theta}) = E(\boldsymbol{S}_{\boldsymbol{\nu}}^{(i)}).$$

For the estimation of $\boldsymbol{\theta}$, we consider an estimator $\hat{\boldsymbol{\theta}}$ obtained by minimizing over the parameter space

$$(4) \qquad Q(\boldsymbol{\theta}) = \sum_{i=1}^{I} n^{(i)} \{ tr[\boldsymbol{S}_{\boldsymbol{\nu}}^{(i)} \boldsymbol{\Sigma}_{\boldsymbol{\nu}}^{(i)-1}(\boldsymbol{\theta})] - log\,|\boldsymbol{S}_{\boldsymbol{\nu}}^{(i)} \boldsymbol{\Sigma}_{\boldsymbol{\nu}}^{(i)-1}(\boldsymbol{\theta})| - p^{(i)} +$$

$$+ [\bar{\boldsymbol{v}}^{(i)} - \boldsymbol{\mu}_{\boldsymbol{\nu}}^{(i)}(\boldsymbol{\theta})]' \boldsymbol{\Sigma}_{\boldsymbol{\nu}}^{(i)-1}(\boldsymbol{\theta}) [\bar{\boldsymbol{v}}^{(i)} - \boldsymbol{\mu}_{\boldsymbol{\nu}}^{(i)}(\boldsymbol{\theta})]$$

The obtained estimator $\hat{\boldsymbol{\theta}}$ is a slight modification of the normal maximum likelihood estimator (MLE). The exact normal MLE can be obtained if $[(n^{(i)}-1)/n^{(i)}]\boldsymbol{S}_{\boldsymbol{\nu}}^{(i)}$ is used in place of $\boldsymbol{S}_{\boldsymbol{\nu}}^{(i)}$. Asymptotic results are equivalent for the two estimators. We consider $\hat{\boldsymbol{\theta}}$ because it can be computed by the existing computer packages. The form of $Q(\boldsymbol{\theta})$ corresponds to the so-called mean and covariance structure analysis. But, the existing covariance structure computer packages without mean structure can be used to carry out the minimization of $Q(\boldsymbol{\theta})$ using a certain technique. See, e.g., the manuals of EQS and LISREL. In the next section, asymptotic distribution results for $\hat{\boldsymbol{\theta}}$ are derived for a broad range of situations.

3. Theoretical results. The main results of this paper are presented in Theorem 1. We now define the following set of assumptions for the model in (1) that assumes normal and independent variables over populations under the same parameterization as in Assumption 1,

ASSUMPTION 1B)

i) $\boldsymbol{\zeta}_j^{(i)} \sim \boldsymbol{N}(\boldsymbol{\mu}_{\boldsymbol{\zeta}^{(i)}}, \boldsymbol{\Sigma}_{\boldsymbol{\zeta}^{(i)}})$ *and independent for all i and j* $i = 1,\ldots,I;\ j = 1,\ldots,n^{(i)}$

ii) $\boldsymbol{\varepsilon}_\ell^{(i)} \sim \boldsymbol{N}(0, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}_\ell^{(i)}})$, *for all* $\ell = 0,1,\ldots,L^{(i)}$.

iii) $\boldsymbol{\beta}^{(i)}$, $\boldsymbol{B}^{(i)}$, *and* $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}_0^{(i)}}$ *can be restricted and they are assumed to be functions of a vector* $\boldsymbol{\tau}$.

iv) $\boldsymbol{\mu}_{\boldsymbol{\zeta}^{(i)}}$, $\boldsymbol{\Sigma}_{\boldsymbol{\zeta}^{(i)}}$ *and* $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}_\ell^{(i)}}$, $\ell = 1,\ldots,L^{(i)}$ *are assumed to be unrestricted.*

Theorem 1 shows similarities and differences of the limiting results for the two different sets of Assumptions 1 and 1B).

THEOREM 1. *Assume that the model in (1) holds under Assumption 1, in addition we assume that*

ASSUMPTION 2. $lim_{n_m \to \infty} \dfrac{n^{(i)}}{n} = r^{(i)}$, where $n_m = min\{n^{(1)},\ldots,n^{(I)}\}$ and

$n = \sum_{i=1}^{I} n^{(i)}$,

ASSUMPTION 3. $(\forall \varepsilon > 0)(\exists \delta > 0)\ \ni |\boldsymbol{\gamma}(\boldsymbol{\theta}) - \boldsymbol{\gamma}(\boldsymbol{\theta}_0)| < \boldsymbol{\delta} \Rightarrow \|\theta - \theta_0\| < \varepsilon$

where $\|\boldsymbol{x}\| = \sqrt{\boldsymbol{x}'\boldsymbol{x}}$ *and* $\boldsymbol{\theta}_0$ *is the limiting true value of* $\boldsymbol{\theta}$.

ASSUMPTION 4. For all $i = 1,\ldots,I$, $\boldsymbol{\beta}^{(i)}(\boldsymbol{\tau})$, $\boldsymbol{B}^{(i)}(\boldsymbol{\tau})$, *and* $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}_0^{(i)}}(\boldsymbol{\tau})$ *are twice continuously differentiable in the parameter space of* $\boldsymbol{\tau}$. *The columns of the matrix* $\partial\boldsymbol{\gamma}(\boldsymbol{\theta}_0)/\partial\boldsymbol{\tau}'$ *are linearly independent. Then,*

i) $\boldsymbol{V}_G^{(\boldsymbol{\tau})} = \boldsymbol{V}_{NI}^{(\boldsymbol{\tau})}$

where $\boldsymbol{V}_G^{(\boldsymbol{\tau})}$ *and* $\boldsymbol{V}_{NI}^{(\boldsymbol{\tau})}$ *the asymptotic covariance matrices of* $\widehat{\boldsymbol{\tau}}$ *under the general Assumption 1 and under the standard Assumption 1B), respectively, (the initials NI stand for Normality and Independence over populations and G for the general set of Assumptions 1). The matrix* $\boldsymbol{V}_G^{(\boldsymbol{\tau})}$ *is a part of the matrix* $\boldsymbol{V}_G^{(\boldsymbol{\theta})}$ *that is the asymptotic covariance matrix for the estimated vector* $\boldsymbol{\theta}$ *.*

ii) for the asymptotic covariance matrices for the vector means $\widehat{\boldsymbol{\mu}}_{\boldsymbol{\zeta}^{(i)}}$ *it holds that*

1) In Case A of Assumption 1 with fixed $\boldsymbol{\zeta}_j^{(i)}$ *it holds that*

$$\boldsymbol{V}_G^{(\boldsymbol{\mu}_{\boldsymbol{\zeta}^{(i)}})} = \boldsymbol{V}_{NI}^{(\boldsymbol{\mu}_{\boldsymbol{\zeta}^{(i)}})} - \boldsymbol{\Sigma}_{\boldsymbol{\zeta}^{(i)}} \tag{5}$$

2) In Case B of Assumption 1 with random $\boldsymbol{\zeta}_j^{(i)}$ *it holds that*

$$\boldsymbol{V}_G^{(\boldsymbol{\mu}_{\boldsymbol{\zeta}^{(i)}})} = \boldsymbol{V}_{NI}^{(\boldsymbol{\mu}_{\boldsymbol{\zeta}^{(i)}})} \tag{6}$$

iii) for the asymptotic covariance matrices for the vector $vec(\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\zeta}^{(i)}})$

1) In Case A of Assumption 1 with fixed $\boldsymbol{\zeta}_j^{(i)}$ *it holds that*

$$\boldsymbol{V}_G^{(vec(\boldsymbol{\Sigma}_{\boldsymbol{\zeta}^{(i)}}))} = \boldsymbol{V}_{NI}^{(vec(\boldsymbol{\Sigma}_{\boldsymbol{\zeta}^{(i)}}))} - \frac{2}{n^{(i)}}(\boldsymbol{\Sigma}_{\boldsymbol{\zeta}^{(i)}} \otimes \boldsymbol{\Sigma}_{\boldsymbol{\zeta}^{(i)}}) \tag{7}$$

2) In Case B of Assumption 1 with random $\boldsymbol{\zeta}_j^{(i)}$*, and assuming that* $\boldsymbol{\zeta}_j^{(i)}$ *have finite fourth moments, it holds that*

$$\boldsymbol{V}_G^{(vec(\boldsymbol{\Sigma}_{\boldsymbol{\zeta}^{(i)}}))} = \boldsymbol{V}_{NI}^{(vec(\boldsymbol{\Sigma}_{\boldsymbol{\zeta}^{(i)}}))} + \frac{1}{n^{(i)}} Var[vec(\boldsymbol{\zeta}^{(i)}\boldsymbol{\zeta}^{(i)'})] - \frac{2}{n^{(i)}}(\boldsymbol{\Sigma}_{\boldsymbol{\zeta}^{(i)}} \otimes \boldsymbol{\Sigma}_{\boldsymbol{\zeta}^{(i)}}) \tag{8}$$

iv) The function $Q(\boldsymbol{\theta})$*, defined in (4), evaluated on its minimum* $\widehat{\boldsymbol{\theta}}$ *converges to a chi-square distribution,* $Q(\widehat{\boldsymbol{\theta}}) \xrightarrow{d} \boldsymbol{\chi}_q^2$*, with* $q = \sum_{i=1}^{I} [\, p^{(i)} + p^{(i)}(\, p^{(i)} + 1)/2\,] - d_{\boldsymbol{\theta}}$ *.*

PROOF OF THEOREM 1. For the proof we need the following three lemmas.

LEMMA 1. *Assume that the model in (1) holds. If Assumptions 1, 2 and 3 hold then as* $n_m \to \infty$

(9) $$\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0 .$$

PROOF OF LEMMA 1: From Assumption 1 and the law of large numbers it holds that $\bar{\boldsymbol{c}} \xrightarrow{p} \boldsymbol{\gamma}(\boldsymbol{\theta}_0)$ which implies $Q(\boldsymbol{\theta}_0) \xrightarrow{p} 0$. Since $Q(\boldsymbol{\theta}) > 0 \; \forall \boldsymbol{\theta}$, and $\hat{\boldsymbol{\theta}}$ minimizes *Q*, we have $Q(\hat{\boldsymbol{\theta}}) \xrightarrow{p} 0$. From the last result and Assumption 2 we get $\boldsymbol{\gamma}(\hat{\boldsymbol{\theta}}) \xrightarrow{\mathrm{p}} \boldsymbol{\gamma}(\boldsymbol{\theta}_0)$ and (9) holds from Assumption 3.

LEMMA 2. *Let* $\bar{\boldsymbol{\theta}}_n = \left(\boldsymbol{\tau}_0', \quad \boldsymbol{\upsilon}_n'\right)'$ *where* $\boldsymbol{\tau}_0$ *is the true value of* $\boldsymbol{\tau}$, $\boldsymbol{\upsilon}_n$ *contains the vectors* $\bar{\boldsymbol{\zeta}}^{(i)}$, $vec(\boldsymbol{S}_{\boldsymbol{\zeta}^{(i)}})$, *and* $vec(\boldsymbol{S}_{\boldsymbol{\varepsilon}_\ell^{(i)}})$, $\ell = 1,\ldots,\mathrm{L}^{(\mathrm{i})}$, *for all* $\mathrm{i} = 1,\ldots,I$. *Then, under the model and assumptions considered in Lemma 1, and under Assumption 4 it holds that*

(10) $$\sqrt{\mathrm{n}}(\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}_{\mathrm{n}}) = \mathbf{A}_0 \sqrt{\mathrm{n}}[\mathbf{c} - \boldsymbol{\gamma}(\bar{\boldsymbol{\theta}}_{\mathrm{n}})] + \mathrm{o}_{\mathrm{p}}(1)$$

where $\boldsymbol{A}_0$ *is free of* $\mathrm{n}^{(\mathrm{i})}$, *and equal to*

(11) $$\boldsymbol{A}_0 = (\boldsymbol{J}_0' \boldsymbol{\Omega}_0^{-1} \boldsymbol{J}_0)^{-1} \boldsymbol{J}_0' \boldsymbol{\Omega}_0^{-1},$$

where $\boldsymbol{J}_0 = \boldsymbol{J}(\boldsymbol{\gamma}(\boldsymbol{\theta}_0))$ *is the Jacobean of* $\boldsymbol{\gamma}(\boldsymbol{\theta})$ *evaluated at* $\boldsymbol{\theta}_0$,

$\boldsymbol{\Omega}_0^{-1} = \boldsymbol{\Omega}^{-1}(\boldsymbol{\theta}_0) = [r^{(1)} \boldsymbol{\Omega}^{(1)-1}(\boldsymbol{\theta}_0)] \oplus \cdots \oplus [r^{(I)} \boldsymbol{\Omega}^{(I)-1}(\boldsymbol{\theta}_0)]$, *and*

$$\boldsymbol{\Omega}^{(i)-1}(\boldsymbol{\theta}) = \boldsymbol{\Sigma}^{(i)-1}(\boldsymbol{\theta}) \oplus \{\frac{1}{2}[\boldsymbol{\Sigma}^{(i)-1}(\boldsymbol{\theta}) \otimes \boldsymbol{\Sigma}^{(i)-1}(\boldsymbol{\theta})]\}.$$

Recall that the ratios $r^{(i)}$ *were defined in Assumption 2 and* $\boldsymbol{c}$ *in (3). The symbol* $\oplus$ *is the direct sum for matrices.*

ii) Also, it hold that

(12) $$Q(\hat{\boldsymbol{\theta}}) = n[\boldsymbol{c} - \boldsymbol{\gamma}(\bar{\boldsymbol{\theta}}_n)]' \boldsymbol{M}_0 [\boldsymbol{c} - \boldsymbol{\gamma}(\bar{\boldsymbol{\theta}}_n)] + o_p(1)$$

with $\boldsymbol{M}_0 = \boldsymbol{\Omega}_0^{-1}(\boldsymbol{I} - \boldsymbol{A}_0)$.

PROOF OF LEMMA 2. i) From Taylor's expansion and Lemma 1 it turns out that there exists $\boldsymbol{\theta}^*$ on the line segment between $\hat{\boldsymbol{\theta}}$ and $\bar{\boldsymbol{\theta}}_n$ such that

(13) $$\boldsymbol{J}[Q(\hat{\boldsymbol{\theta}})] = \boldsymbol{J}[Q(\bar{\boldsymbol{\theta}}_n)] + \boldsymbol{H}[Q(\boldsymbol{\theta}^*)](\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}_n)$$

where $\boldsymbol{J}$ and $\boldsymbol{H}$ are the Jacobean and Hessian matrices, respectively.

Now for the Jacobean and Hessian matrices we proved that,

(14) $$\boldsymbol{J}[Q(\bar{\boldsymbol{\theta}}_n)] = -2\boldsymbol{J}_{\boldsymbol{0}}' \boldsymbol{\Omega}_0^{-1} [\boldsymbol{c} - \boldsymbol{\gamma}(\bar{\boldsymbol{\theta}}_n)] + o_p(n_m^{-1/2})$$

(15) $$\boldsymbol{H}[Q(\boldsymbol{\theta}^*)] \xrightarrow{p} 2\boldsymbol{J}_{\boldsymbol{0}}' \boldsymbol{\Omega}_0^{-1} \boldsymbol{J}_{\boldsymbol{0}}$$

The result in (10) follows if we use (14), (15), and the fact that $\boldsymbol{J}[Q(\hat{\boldsymbol{\theta}})] = 0$ in (13).

ii) After doing several matrix modifications we get the following quadratic form,

(16) $$Q(\hat{\boldsymbol{\theta}}) = n[\boldsymbol{c} - \boldsymbol{\gamma}(\hat{\boldsymbol{\theta}})]' \boldsymbol{\Omega}^{-1}(\hat{\boldsymbol{\theta}})[\boldsymbol{c} - \boldsymbol{\gamma}(\hat{\boldsymbol{\theta}})] + o_p(1).$$

Also, there exists $\boldsymbol{\theta}^*$ on the line segment between $\hat{\boldsymbol{\theta}}$ and $\bar{\boldsymbol{\theta}}_n$ such that

(17) $$\boldsymbol{\gamma}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\gamma}(\bar{\boldsymbol{\theta}}_n) = \boldsymbol{J}[\boldsymbol{\gamma}(\boldsymbol{\theta}^*)](\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}_n).$$

From equations (17) and (10) we get that

(18) $$\boldsymbol{c} - \boldsymbol{\gamma}(\hat{\boldsymbol{\theta}}) = [\boldsymbol{I} - \boldsymbol{J}_0 \boldsymbol{A}_0][\boldsymbol{c} - \boldsymbol{\gamma}(\bar{\boldsymbol{\theta}}_n)] + o_p\left(\frac{1}{\sqrt{n}}\right).$$

and the result follows from (16) and (18).

LEMMA 3. *For the model in (1) under Assumption 1 it holds that*

i)

(19) $\boldsymbol{c} - \boldsymbol{\gamma}(\bar{\boldsymbol{\theta}}_n) = \boldsymbol{E}\,\boldsymbol{w}$

where $\boldsymbol{E}$ *is a constant matrix,* $\boldsymbol{w}$ *consists of the subvectors* $\boldsymbol{w}^{(i)}, \mathrm{i} = 1,\ldots,I$, *and* $\boldsymbol{w}^{(i)}$ *consist of the subvectors* $\bar{\boldsymbol{\varepsilon}}^{(i)}$, $vec(\boldsymbol{S}_{\boldsymbol{\varepsilon}_0^{(i)}\boldsymbol{\varepsilon}_0^{(i)}})$, $vec(\boldsymbol{S}_{\boldsymbol{x}^{(i)}\boldsymbol{y}^{(i)}})$, *for all* $\boldsymbol{x}^{(i)}$ *and* $\boldsymbol{y}^{(i)}$ *such that*

$\boldsymbol{x}^{(i)} \neq \boldsymbol{y}^{(i)}, \quad i = 1,\ldots,I; \quad \boldsymbol{x}^{(i)}, \boldsymbol{y}^{(i)} = \boldsymbol{\zeta}^{(i)}, \boldsymbol{\varepsilon}_0^{(i)}, \boldsymbol{\varepsilon}_1^{(i)}, \ldots, \boldsymbol{\varepsilon}_{L^{(i)}}^{(i)}$.

ii) The limiting distribution of $\sqrt{\mathrm{n}}\mathbf{w}$ *is the same under Assumptions 1 and under 1B).*

PROOF OF LEMMA 3. i) We proved that the components of $\boldsymbol{c} - \boldsymbol{\gamma}(\bar{\boldsymbol{\theta}}_n)$ are written in the following form

(20) $$\bar{\boldsymbol{v}}^{(i)} - \boldsymbol{\mu}_{\boldsymbol{v}}^{(i)}(\bar{\boldsymbol{\theta}}_n) = \boldsymbol{B}^{(i)}\begin{bmatrix} \boldsymbol{0} \\ \bar{\boldsymbol{\varepsilon}}^{(i)} \end{bmatrix}$$

(21) $$\boldsymbol{S}_{\boldsymbol{v}^{(i)}} - \boldsymbol{\Sigma}_{\boldsymbol{v}^{(i)}}(\bar{\boldsymbol{\theta}}_n) = \boldsymbol{B}^{(i)}\begin{bmatrix} \boldsymbol{0} & \boldsymbol{S}_{\boldsymbol{\zeta}^{(i)}\boldsymbol{\varepsilon}^{(i)}} \\ \boldsymbol{S}_{\boldsymbol{\varepsilon}^{(i)}\boldsymbol{\zeta}^{(i)}} & \boldsymbol{S}_{\boldsymbol{\varepsilon}^{(i)}\boldsymbol{\varepsilon}^{(i)}} - \boldsymbol{D}_{\boldsymbol{\varepsilon}^{(i)}} \end{bmatrix}\boldsymbol{B}^{(i)\prime}$$

where $\boldsymbol{D}_{\boldsymbol{\varepsilon}^{(i)}} = \boldsymbol{0} \oplus \boldsymbol{S}_{\boldsymbol{\varepsilon}_1^{(i)}} \oplus \ldots \oplus \boldsymbol{S}_{\boldsymbol{\varepsilon}_{L^{(i)}}^{(i)}}$. The result in (19) follows by noting in equations (20) and (21) that the components of $\boldsymbol{c} - \boldsymbol{\gamma}(\bar{\boldsymbol{\theta}}_n)$ are products of constant matrices (functions of $\boldsymbol{B}^{(i)}$) and the subvectors of $\boldsymbol{w}^{(i)}$.

ii) Note that the matrix $\boldsymbol{S}_{\boldsymbol{\varepsilon}^{(i)}\boldsymbol{\varepsilon}^{(i)}} - \boldsymbol{D}_{\boldsymbol{\varepsilon}^{(i)}}$ does not depend on $\boldsymbol{S}_{\boldsymbol{\varepsilon}_\ell^{(i)}\boldsymbol{\varepsilon}_\ell^{(i)}}$ for $\ell = 1,\ldots,L^{(i)}$. Also note that within the populations for each *(i)* the subvectors of $\sqrt{n}\boldsymbol{w}^{(i)}$ are independent and their limiting distributions do not depend on the non-normality of the latent variables and on the fixed latent variables in Case A (See Anderson (1989), Theorem 5.1). Now between the populations, the limiting covariance between $\boldsymbol{w}^{(i)}$ and $\boldsymbol{w}^{(m)}$ for $i \neq m$ is 0 despite the correlation of $\boldsymbol{\zeta}_j^{(i)}$ and

$\boldsymbol{\zeta}_j^{(m)}$ for each *j*. That holds because the limiting covariance between $\sqrt{n}vec(\boldsymbol{S}_{\boldsymbol{\zeta}^{(i)}\boldsymbol{\varepsilon}^{(i)}})$ and $\sqrt{n}vec(\boldsymbol{S}_{\boldsymbol{\zeta}^{(m)}\boldsymbol{\varepsilon}^{(m)}})$ is 0 since the errors are assumed to be independent over populations.

Now, we return to the proof of Theorem 1. For,

i)Lemmas *2i)* and *3i)* show that $\sqrt{n}(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0)$ is a linear combination of $\sqrt{n}\boldsymbol{w}$ and thus the result follows from Lemma *3ii)*.

ii) and iii) For cases *ii)* and *iii)* we use the following two equations, respectively,

(22) $$\sqrt{n}(\hat{\boldsymbol{\mu}}_{\boldsymbol{\zeta}^{(i)}} - \boldsymbol{\mu}^0_{\boldsymbol{\zeta}^{(i)}}) = \sqrt{n}(\hat{\boldsymbol{\mu}}_{\boldsymbol{\zeta}^{(i)}} - \bar{\boldsymbol{\zeta}}^{(i)}) + \sqrt{n}(\bar{\boldsymbol{\zeta}}^{(i)} - \boldsymbol{\mu}^0_{\boldsymbol{\zeta}^{(i)}})$$

(23) $$\sqrt{n}vec(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\zeta}^{(i)}} - \boldsymbol{\Sigma}^0_{\boldsymbol{\zeta}^{(i)}}) = \sqrt{n}vec(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\zeta}^{(i)}} - \boldsymbol{S}_{\boldsymbol{\zeta}^{(i)}}) + \sqrt{n}vec(\boldsymbol{S}_{\boldsymbol{\zeta}^{(i)}} - \boldsymbol{\Sigma}^0_{\boldsymbol{\zeta}^{(i)}})$$

where $\boldsymbol{\mu}^0_{\boldsymbol{\zeta}^{(i)}}$ and $\boldsymbol{\Sigma}^0_{\boldsymbol{\zeta}^{(i)}}$ are the true values of the corresponding parameters. In both *ii)* and *iii)*, for Case A, with fixed factors, we need the limiting distributions of the first vectors in the second parts of equations (22) and (23). For Case B with random factors we need the limiting distributions of the vectors in the first parts of equations (22) and (23). Since the procedure is the same for *ii)* and *iii)* we are going to explain the proof only for part *iii)*. So for Case A in (23) we compute the limiting covariance matrices of all three vectors under the Assumption 1B),

(24) $$\boldsymbol{V}_{NI}^{(vec(\boldsymbol{\Sigma}_{\boldsymbol{\zeta}^{(i)}}))} = \boldsymbol{V}_2 + \frac{2}{n^{(i)}}(\boldsymbol{\Sigma}_{\boldsymbol{\zeta}^{(i)}} \otimes \boldsymbol{\Sigma}_{\boldsymbol{\zeta}^{(i)}}).$$

From Lemmas 2*i*) and 3 it follows that the first vector of the second part of equation (23) has the same limiting distribution under Assumption 1, with fixed factors, and under Assumption 1B). Thus $\boldsymbol{V}_2 = \boldsymbol{V}_G^{(vec(\boldsymbol{\Sigma}_{\boldsymbol{\zeta}^{(i)}}))}$, and the result follows by solving (24) for $\boldsymbol{V}_G^{(vec(\boldsymbol{\Sigma}_{\boldsymbol{\zeta}^{(i)}}))}$.

Now for Case B in iii) we compute the limiting covariance matrices under Assumption 1B) and under Assumption 1 and we get, respectively,

(25) $$\boldsymbol{V}_{NI}^{(vec(\boldsymbol{\Sigma}_{\boldsymbol{\zeta}^{(i)}}))} = \boldsymbol{V}_{NI}^{*} + \frac{2}{n^{(i)}}(\boldsymbol{\Sigma}_{\boldsymbol{\zeta}^{(i)}} \otimes \boldsymbol{\Sigma}_{\boldsymbol{\zeta}^{(i)}})$$

(26) $$\boldsymbol{V}_{G}^{(vec(\boldsymbol{\Sigma}_{\boldsymbol{\zeta}^{(i)}}))} = \boldsymbol{V}_{G}^{*} + \frac{1}{n^{(i)}}\mathrm{Var}[\mathrm{vec}(\boldsymbol{\zeta}^{(i)}\boldsymbol{\zeta}^{(i)\prime})]$$

Again, from Lemmas 2 and 3 it follows that $\boldsymbol{V}_{G}^{*} = \boldsymbol{V}_{NI}^{*}$. The result follows by solving (25) for $\boldsymbol{V}_{NI}^{*}$ and substituting the result in (26).

iv) Lemmas 2*ii*) and 3*i*) show that $Q(\hat{\boldsymbol{\theta}})$ is a quadratic function of $\sqrt{n}\boldsymbol{w}$ and the result follows from Lemma 3*ii*) and the known result that $Q(\hat{\boldsymbol{\theta}}) \xrightarrow{d} \boldsymbol{\chi}_{q}^{2}$ under Assumption 1B).

To derive large sample results for $\hat{\boldsymbol{\theta}}$ minimizing (4) under the model in (1) and Assumption 1, we consider the case where all $n^{(i)}$ increase to infinity at a common rate, and use n_m as the index for taking a limit in Assumption 2. Assumption 3 is a standard identification condition used in Lemma 1. Note that the true value of $\boldsymbol{\theta}$ in Case A of Assumption 1 with fixed variables depends on $n^{(i)}$, since it contains $\bar{\boldsymbol{\zeta}}^{(i)}$ and $\boldsymbol{S}_{\boldsymbol{\zeta}^{(i)}}$. Thus, we denote the limit of the true value as $\boldsymbol{\theta}_0$. Lemma 1 gives the consistency of the estimator $\hat{\boldsymbol{\theta}}$ minimizing (4) for the model in (1). Hence, under very weak distributional specifications in Assumption 1, the estimator $\hat{\boldsymbol{\theta}}$ is consistent for the limiting true value $\boldsymbol{\theta}_0$. In fact, it is clear from the proof that the consistency of $\hat{\boldsymbol{\theta}}$ holds for any general mean and covariance structure model $\boldsymbol{\gamma}(\boldsymbol{\theta}) = E(\boldsymbol{c})$ satisfying $\bar{\boldsymbol{c}} \xrightarrow{p} \boldsymbol{\gamma}(\boldsymbol{\theta}_0)$. To characterize the limiting behavior of $\hat{\boldsymbol{\theta}}$ in more detail, especially for the assessment of the so-called asymptotic robustness properties, it is convenient to consider an expansion of $\hat{\boldsymbol{\theta}}$, not around the true value or the limiting true value $\boldsymbol{\theta}_0$ but around some other quantity $\bar{\boldsymbol{\theta}}_n$ defined in Lemma 2, that depends on the unobservable sample moments of the non-

normal latent variables $\boldsymbol{\zeta}^{(i)}$ and $\boldsymbol{\varepsilon}_{\ell}^{(i)}$ $\ell = 1,\ldots,L^{(i)}$. Thus, the limiting true value $\boldsymbol{\upsilon}_0$ consisting of the true covariance matrices of the random latent variables is replaced in $\bar{\boldsymbol{\theta}}_n$ by $\boldsymbol{\upsilon}_n$ consisting of the unobservable sample moments. While statistical inference is to be made for the true value of $\boldsymbol{\theta}$, $\bar{\boldsymbol{\theta}}_n$ with an artificial quantity $\boldsymbol{\upsilon}_n$ plays a useful role in assessing the property of $\hat{\boldsymbol{\tau}}$ in $\hat{\boldsymbol{\theta}}$, as well as in characterizing the limiting distribution of the whole $\hat{\boldsymbol{\theta}}$ without specifying any moments for $\boldsymbol{\zeta}^{(i)}$ and $\boldsymbol{\varepsilon}_{\ell}^{(i)}$ $\ell = 1,\ldots,\mathrm{L}^{(\mathrm{i})}$, higher than the second order. To obtain an expansion of $\hat{\boldsymbol{\theta}}$ around $\bar{\boldsymbol{\theta}}_n$, we need some smoothness conditions for $\boldsymbol{\beta}^{(i)}(\boldsymbol{\tau})$, $\boldsymbol{B}^{(i)}(\boldsymbol{\tau})$, and $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}_0^{(i)}}(\boldsymbol{\tau})$, and the full column rank of the Jacobean matrix $\boldsymbol{J}[\boldsymbol{\gamma}(\boldsymbol{\theta}_0)]$ that are stated in Assumption 4. Since the linear independency of the columns of $\boldsymbol{J}[\boldsymbol{\gamma}(\boldsymbol{\theta}_0)]$ associated with the $\boldsymbol{\upsilon}$ part of $\boldsymbol{\theta}$ is trivial, we need to assume only that the $\boldsymbol{\tau}$ part of the model is specified without any redundancy. Thus in Assumption 4 we just assume that $\partial\boldsymbol{\gamma}(\boldsymbol{\theta}_0)/\partial\boldsymbol{\tau}'$ is of full-column rank, Lemma 2 expresses the leading term of $\sqrt{n}(\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}_n)$ in terms of $\boldsymbol{c} - \boldsymbol{\gamma}(\bar{\boldsymbol{\theta}}_n)$. Note that the use of $\bar{\boldsymbol{\theta}}_n$ in Lemma 2 produced an expansion of $\hat{\boldsymbol{\theta}}$ around $\bar{\boldsymbol{\theta}}_n$ with the existence of only second moments of $\boldsymbol{\zeta}^{(i)}$ and $\boldsymbol{\varepsilon}_{\ell}^{(i)}$ $\ell = 1,\ldots,L^{(i)}$. It can be shown from the proof that the expansion in Lemma 2 holds for the general model $\boldsymbol{\gamma}(\boldsymbol{\theta}) = E(\boldsymbol{c})$ and for any $\bar{\boldsymbol{\theta}}_n$ with $\bar{\boldsymbol{\theta}}_n \xrightarrow{p} \boldsymbol{\theta}_0$ provided that $\sqrt{n}[\boldsymbol{c} - \boldsymbol{\gamma}(\bar{\boldsymbol{\theta}}_n)]$ converges in distribution. But, the special choice of $\bar{\boldsymbol{\theta}}_n$ for the model in (1) makes the result of Lemma 2 practically meaningful. Lemma 3 is actually the key tool in the proof that shows asymptotic robustness. It expresses $\sqrt{n}[\boldsymbol{c} - \boldsymbol{\gamma}(\bar{\boldsymbol{\theta}}_n)]$ in terms of $\sqrt{n}\boldsymbol{w}$ which has the same limiting distributions under Assumptions 1 and 1B). Thus, the main difficulty in the proof of Theorem 1-*i*) is to express $\sqrt{n}(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0)$ in terms of a vector $\sqrt{n}\boldsymbol{w}$ that its limiting

distribution does not depend on the existence of fixed, non-normal, and correlated-over-population variables. Similarly, we proved Theorem 1 *iv*) by expressing $Q(\widehat{\boldsymbol{\theta}})$ as a quadratic function of $\sqrt{n}\boldsymbol{w}$. The formulas in (5) and (7) in Theorem 1 show what corrections should be made when we have fixed variables in order to get correct asymptotic standard errors for $\widehat{\boldsymbol{\mu}}_{\boldsymbol{\zeta}^{(i)}}$ and $vec(\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\zeta}^{(i)}})$. These results are novel even for the case with one population. The formula (6) in Theorem 1-*ii*)-2) shows that the asymptotic standard errors for $\widehat{\boldsymbol{\mu}}_{\boldsymbol{\zeta}^{(i)}}$ are robust. Equation (8) in Theorem 1-*iii*)-2) gives the limiting covariance matrix for $vec(\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\zeta}^{(i)}})$ when $\boldsymbol{\zeta}^{(i)}$ are random. The formula (8) involves the computation of 4-th order cumulants of the latent variables $\boldsymbol{\zeta}^{(i)}$ in practice. This is possible in practice and we obtain satisfactory results when we use the errors-in variables parameterization and have normal errors. For instance in Example 1, the model in (2) with normal errors the 4th order cumulants for $\boldsymbol{\zeta}^{(i)}$ equal to 4th order cumulants of the observed variables for $x^{(i)}$, since the 4th order cumulants of the normal errors are equal to 0. This technique was used in our simulation study and the results are illustrated in the next section. Note that in most practical cases the measurement errors follow a normal distribution.

Although the paper is referred to the multi-sample case the same theory and methodology can be applied to longitudinal data. That is, two different applications, correlated populations and panel data, can be considered by fitting the same kind of modeling and applying the results presented in this paper. A similar method developed for longitudinal data, known as the General Estimating Equation (GEE) method, was proposed by Liang and Zeger (1986). The GEE method was proposed for generalized linear models with univariate outcome variables. In this paper, several response variables are observed and their relationships are explained by a few latent variables within the time points. It can be shown that a special case of the GEE method, using the

identity matrix as the 'working' correlation matrix, is a special case of the model considered in this paper. This can be done by treating the outcome variable and the covariates of the generalized linear models as observed variables in the model considered in this paper and setting latent variables equal to covariates by fixing error variances equal to zero. Thus, the results presented in this paper can be also applied to simpler models such as generalized linear models for longitudinal data. On the other hand, the use of a 'working' correlation matrix as the one used in the GEE method, could be also used in this methodology in order to increase the efficiency of the method.

Now we define a generalized version of the so-called sandwich estimator used by the GEE method for generalized linear models with the identity matrix as the 'working' correlation matrix, and also used by Satorra (1992, 1993A,B, 1994, 1997) for latent variable models. We generalize this matrix for correlated populations and we are going to compare it with our proposed matrix $\boldsymbol{V}_G^{(\boldsymbol{\theta})}$ defined in Theorem 1 theoretically and numerically. A generalized version of the sandwich (*S*) estimator is

$$\boldsymbol{V}_S^{(\boldsymbol{\theta})} = \boldsymbol{A}_0 E(\boldsymbol{S}_{\boldsymbol{d}}) \boldsymbol{A}_0' \tag{27}$$

where $\boldsymbol{A}_0$ is defined in (11) and $E(\boldsymbol{S}_{\boldsymbol{d}})$ is the expected mean of the sample matrix $\boldsymbol{S}_{\boldsymbol{d}}$ that involves 3-rd and 4-th order sample moments defined as,

$$\boldsymbol{S}_{\boldsymbol{d}} = \begin{pmatrix} \frac{1}{n^{(11)}} \boldsymbol{S}_d^{(11)} & \cdots & \frac{1}{n^{(1I)}} \boldsymbol{S}_d^{(1I)} \\ \vdots & \ddots & \vdots \\ \frac{1}{n^{(I1)}} \boldsymbol{S}_d^{(I1)} & \cdots & \frac{1}{n^{(II)}} \boldsymbol{S}_d^{(II)} \end{pmatrix}, \text{ with}$$

$$\boldsymbol{S}_d^{(ik)} = \frac{1}{n^{(ik)} - 1} \sum_{j=1}^{n^{(ik)}} (\boldsymbol{d}_j^{(i)} - \bar{\boldsymbol{d}}^{(i)})(\boldsymbol{d}_j^{(k)} - \bar{\boldsymbol{d}}^{(k)})' \text{ and}$$

$$\boldsymbol{d}_j^{(i)} = \begin{pmatrix} \boldsymbol{v}_j^{(i)} \\ vec[(\boldsymbol{v}_j^{(i)} - \bar{\boldsymbol{v}}^{(i)})(\boldsymbol{v}_j^{(i)} - \bar{\boldsymbol{v}}^{(i)})'] \end{pmatrix}, \text{ where } i,k = 1,\ldots,I;\ j = 1,\ldots,n^{(i)} \text{ and}$$

$n^{(ik)} = \#$ *of correlated individuals between the i-th and the k-th populations*. Note the form of the matrix $\boldsymbol{V}_S^{(\boldsymbol{\theta})}$ in (27) can be derived from Lemma 2. The equation (12) in Lemma 2 also holds if we replace $\bar{\boldsymbol{\theta}}_n$ by the true value of $\boldsymbol{\theta}$ and the result follows by noting that $Var[\boldsymbol{c} - \boldsymbol{\gamma}(\boldsymbol{\theta}_0)] = E(\boldsymbol{S}_d)$. Theorem 1 actually gives an alternative form of some of the parts of the matrix $\boldsymbol{V}_S^{(\boldsymbol{\theta})}$. The parts of the matrix $\boldsymbol{V}_G^{(\boldsymbol{\theta})}$ defined in Theorem 1 are actually theoretically exactly the same as the corresponding parts of the matrix $\boldsymbol{V}_S^{(\boldsymbol{\theta})}$. In practice the matrix $\boldsymbol{A}_0 = \boldsymbol{A}(\boldsymbol{\theta}_0)$ is estimated by $\hat{\boldsymbol{A}}_0 = \boldsymbol{A}(\hat{\boldsymbol{\theta}})$ and the matrix $E(\boldsymbol{S}_d)$ is estimated by $\boldsymbol{S}_d$. Despite the two matrices $\boldsymbol{V}_G^{(\boldsymbol{\theta})}$ and $\boldsymbol{V}_S^{(\boldsymbol{\theta})}$ are theoretically equal in practice the asymptotic standard errors (a.s.e.'s) computed by the matrix $\boldsymbol{V}_G^{(\boldsymbol{\theta})}$ have less variability than the a.s.e.'s computed by the matrix $\boldsymbol{V}_S^{(\boldsymbol{\theta})}$. This happens because the estimation of $\boldsymbol{V}_S^{(\boldsymbol{\theta})}$ involves third and fourth order moments that are more variable than the second moments of the matrix $\boldsymbol{V}_G^{(\boldsymbol{\theta})}$. The matrix $\boldsymbol{V}_G^{(\boldsymbol{\theta})}$ involves *4-th* moments only in the formula of Theorem 1-*iii*)-2 but these moments do not affect the computation of the other a.s.e.'s. This advantage of using the matrix $\boldsymbol{V}_G^{(\boldsymbol{\theta})}$ is shown in the simulation study in the next section.

4 Simulation Study. We simulate the model in (2) of Example 1. A sample from both populations was generated 1,000 times. The simulation was done twice, one time with fixed $\zeta^{(i)}$ and the other time with random $\zeta^{(i)}$, Cases A and B of Assumption 1, respectively. In both cases $\zeta_j^{(1)}$ and $\zeta_j^{(2)}$ are related (correlated over populations), and were generated as linear

combinations of chi-square random variables with 10 degrees of freedom. In Case A, a sample of $(\zeta_j^{(1)}, \zeta_j^{(2)})$ was generated with sample means, variances, and covariance: $\bar{\zeta}^{(1)} = 4.95$, $\bar{\zeta}^{(2)} = 9.95$, $s^2_{\zeta^{(1)}} = 1.97$, $s^2_{\zeta^{(2)}} = 1.95$, and $s_{\zeta^{(1)}\zeta^{(2)}} = 1.36$, respectively, and the set of$(\zeta_j^{(1)}, \zeta_j^{(2)})$ was used in all 1,000 Monte Carlo samples. In Case B, 1,000 independent samples were generated for $\{\zeta_{\mathrm{j}}^{(1)}, \mathrm{j}=1,\ldots,1000;\ \zeta_{\mathrm{j}}^{(2)}, \mathrm{j}=1,\ldots,500\}$. The true means, variances, and covariance of $\zeta_j^{(1)}$ and $\zeta_j^{(2)}$ are $\mu_{\zeta^{(1)}} = 5$, $\mu_{\zeta^{(2)}} = 10$, $\sigma^2_{\zeta^{(1)}} = 2$, $\sigma^2_{\zeta^{(2)}} = 2$, and $\sigma_{\zeta^{(1)}\zeta^{(2)}} = 1.4.$ Note that the above means and variances are estimated but not the covariance $\sigma_{\zeta^{(1)}\zeta^{(2)}}$, according to the estimation method that we suggest. Note that we suggest this method for several populations with quite unbalanced data. In this study, it is easy to use the full likelihood and estimate the covariance $\sigma_{\zeta^{(1)}\zeta^{(2)}}$, but this is not always true in more complicated studies. By not estimating some of the covariances between the two populations we loose some efficiency, e.g., larger asymptotic standard errors (a.s.e.'s). We discuss the efficiency of the method in more detail later in this section.

In both Cases A and B, 1,000 samples were generated for independent $e_\ell^{(i)}$, $i = 1,2;\ \ell = 0,1,\ldots,L^{(i)}$ with $L^{(1)} = 3$ and $L^{(2)} = 2$. The errors $e_{0j}^{(i)}$, i=1,2, are normally distributed with mean *0* and unknown variance $\sigma^2_{e_0^{(i)}}$ while all the other errors $e_{\ell j}^{(i)}$ for $i = 1,2;\ \ell = 1,\ldots,L^{(i)}$ were generated from a chi-square distribution with 10 degrees of freedom, χ^2_{10}, with adjusted mean 0 and variance $\sigma^2_{e_\ell^{(i)}}$. The variance for $e_{0j}^{(i)}$ is common for the two populations, $\sigma^2_{e_0} = \sigma^2_{e_0^{(1)}} = \sigma^2_{e_0^{(2)}}$. The true values, in both cases with fixed and random $\zeta_j^{(i)}$, for the error variances are $\sigma^{o2}_{e_0^{(i)}} = \sigma^{o2}_{e_1^{(i)}} = 0.1$ and $\sigma^{o2}_{e_2^{(i)}} = \sigma^{o2}_{e_{13}^{(i)}} = 0.2$ and for the vector $\boldsymbol{\tau}$ is,

$\boldsymbol{\tau}^0 = (1, 2, -1, -0.1, 0.1, -0.01, 1, 0.1)$. The parameters of $\boldsymbol{\tau}$ are shown in the first column of Table 1. In accordance with the notation of this paper, the vector $\boldsymbol{\theta} = (\boldsymbol{\tau}', \boldsymbol{\upsilon}')'$ where $\boldsymbol{\upsilon}$ contains $\sigma^2_{e_\ell^{(i)}} \quad i = 1,2; \ \ell = 1, \ldots, L^{(i)}$ and the means and variances of $\boldsymbol{\zeta}_j^{(i)} \quad i = 1,2$. To estimate $\boldsymbol{\theta}$ we use normal MLE by minimizing (4) despite the appearance of fixed and non-normal variables and when we estimate the parameters we are pretending that we do not know the true values of the parameters.

Some of the results in the simulation study are shown in Table 1. Columns 2, 4, and 6, show results from Case A with fixed $\boldsymbol{\zeta}_j^{(i)}$ while Columns 3, 5, and 7, show results from Case B with random $\boldsymbol{\zeta}_j^{(i)}$. Columns 2 and 3, of Table 1 compare the a.s.e.'s, G-se, computed by the matrix $\boldsymbol{V}_G^{(\boldsymbol{\tau})}$ in Theorem 1-i), with the Monte Carlo standard errors, MC-se. All the ratios are 1 or very close to 1 and that means that the proposed a.s.e.'s have very small bias. Bias exists because we use the a.s.e.'s as estimates for the true s.e.'s of the parameters in finite samples. Actually Lemma 1 proves that the bias converges to 0 as the sample sizes increase to infinite. In this study for sample sizes $n^{(1)} = 1{,}000$ and $n^{(2)} = 500$ the bias is negligible.

Now we compute Monte Carlo standard errors for the a.s.e.'s computed by the matrix $\boldsymbol{V}_G^{(\boldsymbol{\theta})}$, (G-MCse), and for the a.s.e.'s computed by the matrix $\boldsymbol{V}_S^{(\boldsymbol{\theta})}$, (S-MCse), defined in (27). The ratio (S-MCse)/(G-MCse) compares the variability of the two different estimates of the a.s.e.'s. This ratio is computed for the parameters in $\boldsymbol{\tau}$ and the results are shown in Columns 4 and 5 of Table 1 for both cases with fixed and random $\zeta_j^{(i)}$. All the ratios are significantly larger than 1 and this fact indicates that the a.s.e.'s computed by the sandwich estimator $\boldsymbol{V}_S^{(\boldsymbol{\theta})}$ have larger variability than the a.s.e.'s computed by our suggested estimator $\boldsymbol{V}_G^{(\boldsymbol{\theta})}$.

Now, as to the efficiency of the method, we computed the a.s.e.'s under the full likelihood (FL) and under the reduced likelihood (RL) given in (4). The ratio of the two a.s.e.'s,

$$\text{efficiency} = \frac{\text{RL - MCse}}{\text{FL - MCse}} \tag{28}$$

is given for all the parameters in $\boldsymbol{\tau}$ in the last two columns of Table 1. These ratios actually show the efficiency of the method relative to the FL. In both cases the efficiency is very satisfactory since the ratios are close to 1. The efficiency loss is very small for Case A with fixed $\boldsymbol{\zeta}_j^{(i)}$ and relatively small for Case B with random $\boldsymbol{\zeta}_j^{(i)}$.

For the parameters $\mu_{\zeta^{(1)}}\ \mu_{\zeta^{(2)}}\ \sigma^2_{\zeta^{(1)}}\ \sigma^2_{\zeta^{(2)}}$ we used the formulas in (5), (6), (7), and (8) provided in Theorem 1-*ii*) and -*iii*) and we derived similar results to the previous ones. It should be pointed out that the sandwich estimator does not provide correct a.s.e.'s for Case A with fixed $\boldsymbol{\zeta}_j^{(i)}$ for the parameters $\mu_{\zeta^{(1)}}\ \mu_{\zeta^{(2)}}\ \sigma^2_{\zeta^{(1)}}\ \sigma^2_{\zeta^{(2)}}$. Our novel formulas in (5) and (7) show what corrections should be made in order to obtain correct a.s.e.'s in this case. The a.s.e.'s are evaluated at the estimated value of $\boldsymbol{\theta}$, $\widehat{\boldsymbol{\theta}}$. Note that all the a.s.e.'s are functions of $\boldsymbol{\theta}$ except the ones for $\widehat{\sigma}^2_{\zeta^{(1)}}\ \widehat{\sigma}^2_{\zeta^{(2)}}$ (elements of the matrix $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\zeta}^{(i)}}$ in Theorem 1) that require *4-th* moments (or cumulants) for $\boldsymbol{\zeta}_j^{(i)}$. In general, the fourth order cumulants, ψ, follow the following property: if $x = y + z$, with *y* and *z* independent random variables, then $\psi_x = \psi_y + \psi_z$. Thus, in the model used in the simulation it holds that $\psi_{x^{(i)}} = \psi_{\zeta^{(i)}} + 0$, since the errors, $e_{0j}^{(i)}$, are assumed to be normal having *4th* order cumulants equal to *0*. Thus, the sample *4th* order cumulants of $x^{(i)}$ used for the computation of the a.s.e.'s for $\widehat{\sigma}^2_{\zeta^{(1)}}\ \widehat{\sigma}^2_{\zeta^{(2)}}$.

The a.s.e.'s can be used for hypothesis testing of the parameters. The power of the tests is also robust when the sample sizes are quite large due to the applicability of the multivariate central limit theorem. In the above simulation study we use, as an example, H_0: $\delta_1 = 0$ versus H_1: $\delta_1 < 0$ in Case A with fixed $\zeta_j^{(i)}$. Using level of significance $\alpha = 0.05$, H_0 is rejected when z < -1.645 where $z = \hat{\delta}_1 / \hat{\sigma}^2_{\delta_1}$. Thus, the expected power (EP) is approximately

$$\text{EP}(\delta_1^*) = \Phi(-1.645 + \frac{\delta_1^*}{\text{MC} - \text{se for } \delta_1}) = 0.956, \tag{29}$$

where the Φ function gives the standard cumulative normal distribution and we compute the power for the actual value of δ_1, δ_1^*=-0.01. We also compute the simulated power (SP) as

$$\text{SP} = \frac{\#\text{ of times that } [\hat{\delta}_1 / (\text{a.s.e. of } \hat{\delta}_1)] < -1.645}{1{,}000} = 0.967\ . \tag{30}$$

Thus, the results support the robustness of power for non-normal and correlated populations.

The robustness of the chi-square test statistic is shown in Table 2 for Case A with fixed $\zeta_j^{(i)}$. The mean and the variance of the 1,000 simulated values of $Q(\hat{\theta})$ in (4) are close to the expected 6 and 12 respectively. Also, the simulated percentiles, in the second row, are close to the expected ones given in the first row of Table 2. For similar studies using simper models, see Satorra (1993b, 1997) and Papadopoulos and Amemiya (1994).

In summary, the model in (1) with the errors-in-variables parameterization can formulate the multi-population analysis in a meaningful fashion. The corresponding statistical analysis under the pseudo normal-independence model gives a simple and correct way to conduct statistical inferences about the parameter vector τ without specifying a distributional form or dependency structure over populations. In practice, τ contains all the parameters of direct interest. The asymptotic covariance matrix and standard errors can be readily computed using the existing

procedures, and provide a good approximation in moderately sized samples. The proposed a.s.e.'s have smaller variability than the variability of the robust sandwich estimator, provide high efficiency relative to the full-likelihood method, and can be used for hypothesis testing with robust power. For instance, in the simulation study for one of the most important parameters δ_1, in Case A with fixed $\zeta_j^{(i)}$, the variability ratio is 1.65 (see Table 1), the efficiency ratio is 1.00 (see Table 1), and the power of the test H_0: $\delta_1 = 0$ versus H_1: $\delta_1 < 0$ is 0.967. That is, if the standard deviation of our proposed a.s.e. for δ_1 is 1 then the standard deviation of the a.s.e. for δ_1 computed by the robust sandwich estimator is 1.65. Also, our proposed a.s.e. for δ_1 is close enough to the a.s.e. for δ_1 using the full likelihood, and the power of the test is very high, 0.967, and very close to the expected power, 0.960.

Acknowledgments. The authors wish to thank the reviewers and editors for their insightful and constructive comments. Departments of Statistics at Iowa State University and Rice University provided research support for this paper.

TABLE 1. Monte Carlo standard errors (MCse) for the estimated parameters in $\boldsymbol{\tau}$ versus the proposed asymptotic standard errors (a.s.e.'s) (Gse) of $\widehat{\boldsymbol{\tau}}$, computed by $\boldsymbol{V}_G^{(\boldsymbol{\tau})}$ defined in Theorem 1. Comparison between the MCse for Gse, ($G-\text{MCse}$), and the MCse for the a.s.e.'s computed by the sandwich estimator, $\boldsymbol{V}_S^{(\boldsymbol{\theta})}$, given in (27), ($S$ - MCse). MCse computed under the full-likelihood (FL) and under the reduced likelihood (RL). Results are shown for Cases A and B of Assumption 1 with fixed and random $\boldsymbol{\zeta}_j^{(i)}$.

Para meters	Bias of Gse		Variability of Gse		Efficiency of the method relative to the full-likelihood (FL)	
	Fixed	Random	Fixed	Random	Fixed	Random
$\boldsymbol{\tau}$	$\frac{\text{Gse}}{\text{MCse}}$	$\frac{\text{Gse}}{\text{MCse}}$	$\frac{S-\text{MCse}}{G\text{-MCse}}$	$\frac{S-\text{MCse}}{G\text{-MCse}}$	$\frac{RL-\text{MCse}}{FL\text{-MCse}}$	$\frac{RL-\text{MCse}}{FL\text{-MCse}}$
β_1	1.01	1.01	1.63	1.56	0.99	1.03
β_2	1.01	0.99	1.78	1.68	1.01	1.05
β_3	0.97	1.00	1.84	1.50	1.00	1.06
γ_1	1.00	0.99	1.44	1.47	1.00	1.04
γ_2	0.97	0.99	2.02	1.56	1.01	1.05
δ_1	1.00	1.00	1.65	1.57	1.00	1.03
δ_2	1.00	0.98	1.60	1.44	1.02	1.06
$\sigma_{e_0}^2$	0.99	0.99	2.68	1.56	1.00	1.03

TABLE 2. Monte Carlo mean, variance and percentiles for the chi-square test statistics with 6 degrees of freedom

Mean	Variance	10%	25%	50%	75%	90%	95%	99%
6.0	11.7	9.2	23.6	49.7	75.9	90.5	96.3	98.9

REFERENCES

AMEMIYA, Y (1985). On the goodness-of-fit tests for linear statistical relationships. *Technical Report No. 10, Econometric Workshop, Stanford University* Stanford, California.

AMEMIYA, Y (1986). Multivariate functional and structural relationships with general error covariance structure. *Preprint Series No. 86-3, Department of Statistics, Iowa State University,* Ames, Iowa.

AMEMIYA, Y., FULLER W.A. and PANTULA, S.G. (1987). The asymptotic distributions of some estimators for a factor analysis model. *Journal of Multivariate Analysis*, **22**, 51-64.

AMEMIYA, Y. and ANDERSON, T.W. (1990). Asymptotic chi-square tests for a large class of factor analysis models. *The Annals of Statistics*, **18**, 1453-1463.

ANDERSON, T.W. (1987). Multivariate linear relations. *Proceedings of the Second International Conference in Statistics*, edited by Pukkila T. and Puntanen S., University of Tampere, Finland. 9-36.

ANDERSON, T.W. (1989). Linear latent variable models and covariance structures. *Journal of Econometrics*, **41**, 91-119.

ANDERSON, T.W. and AMEMIYA, Y.(1988). The asymptotic normal distribution of estimators in factor analysis under general conditions. *The Annals of Statistics*, **16**, 759-771.

BENTLER, P.M. (1983). Some contributions to efficient statistics in structural models: Specification and estimation of moment structures. Psychometrika, 48, 493-517.

BENTLER, P.M. (1989). *EQS Structural Equations Program Manual*. BMDP Statistical Software, Inc., Los Angeles.

BOLLEN, P.A. (1989). *Structural Equations with Latent Variables*. John Wiley and Sons, New York.

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

BROWNE, M.W. (1987). Robustness in statistical inference in factor analysis and related models. *Biometrika*, **74**, 375-384.

BROWNE, M.W. (1990). Asymptotic robustness of normal theory methods for the analysis of latent curves. *Contemporary Mathematics*, **112**, 211-225.

BROWNE, M.W. and SHAPIRO, A. (1988). Robustness of normal theory methods in the analysis of linear latent variate models. *British Journal of Mathematical and Statistical Society*, **41**, 193-208.

CHOU, C.P., BENTLER, P.M., and SATORRA, A. (1991). Scaled test statistics and robust standard errors for non-normal data in covariance structure analysis: A Monte Carlo study. *The British Journal of Mathematical and Statistical Psychology*. **44**, 347-358.

FULLER, W.A. (1987). Measurement Error Models. John Wiley and Sons, New York.

HUGGINS, R.M. (1996). Robust inference for variance components models for single trees of cell lineage data. *The Annals of Statistics*, **24**, 1145-1160.

JÖRESKOG, K. (1971). Simultaneous factor analysis in several populations. *Psychometrika*. *47*, 297-308.

JÖRESKOG, K. and SÖRBOM, D. (1989). *LISREL 7; A Guide to the Program and Applications*. 2nd ed., SPSS INC., Chicago.

KRISHNAKUMAR, J., and RONCHETTI, E. (1997). Robust estimators for simultaneous equations models. *Journal of Econometric.* , **78**, 295-314.

LIANG, K.Y., and ZEGER, S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, **73**, 13-22.

LEE, S.Y. and TSUI, K.L. (1982). Covariance structure analysis in several populations. *Psychometrika*, **55**, 107-122.

MAGNUS, J. and NEUDECKER, H. (1988). *Matrix Differential Calculus*. John Wiley and Sons, New York.

MOOIJAART, A. and BENTLER, P.M. (1991). Robustness of normal theory statistics in structural equation models. *Statistica Neerlandica* **45**, 159-171.

MUTHÉN, B. (1989). Multiple group structural modeling with non-normal continuous variables. *British Journal of Mathematical and Statistical Psychology*, **42**, 55-62.

MUTHÉN, B. and KAPLAN, D. (1992). A comparison of some methodologies for the factor analysis of non-normal Likert variables: A note on the size of the model. *The British Journal of Mathematical and Statistical Psychology*. **45**, 19-30.

PAPADOPOULOS, S. and AMEMIYA, Y. (1994). Asymptotic robustness for the structural equation analysis of several populations. *ASA Proceedings*,

Business and Economics Statistics Section, 65-70.

PAPADOPOULOS, S. and AMEMIYA, Y. (1995). On factor analysis of longitudinal data. *ASA Proceedings*, Biometrics Statistics Section, 130-135.

SAS INSTITUTE Inc. (1990). *SAS/STAT User's Guide, Version 6, Fourth Edition, Volume 1*, Cary, NC: SAS Institute Inc.

SATORRA, A. (1992). Asymptotic robust inferences in the analysis of mean and covariance structures. *Sociological Methodology*, edited by Marsden, P.V., 249-278.

SATORRA, A. (1993a). Multi-sample analysis of moment-structures: Asymptotic validity of inferences based on second-order moments.

Statistical Modeling and Latent Variables, edited by Haagen, K., Bartholomew, D.J. and Deistler, M.. Amsterdam: Elsevier. Forthcoming.

SATORRA, A. (1993b). Asymptotic robust inferences multi-sample analysis of augmented-moment structures. *Multivariate Analysis: Future Directions 2*, edited by Cuadras, C.M. and Rao, C.R., 211-229.

SATORRA, A. (1994). On asymptotic robustness in multiple-group analysis of multivariate relations. Paper presented at *Latent Variable Modeling with Applications to Causality*, March 19-20, Los Angeles.

SATORRA, A. (1995). Asymptotic robust inferences in multi-sample analysis of multivariate linear relations, *Economics Working Paper 126,* Universitat Pompeu Fabra.

SATORRA, A. (1997). Fusion of data sets in multivariate linear regression with errors-in-variables, *Classification and Knowledge Organization* pp. 195-207, edited by R. Klar and O. Opitz, Springer Verlag: London 1997.

SATORRA, A and BENTLER, P.M. (1990). Model conditions for asymptotic robustness in the analysis of linear relations. *Computational Statistics and Data Analysis*, **10**, 235-249.

SHAPIRO, A. (1987). Robustness properties of the MDF analysis of moment structures. *South African Statistical Journal*, **21**, 39-62.

ZAMAR, R. (1989). Robust estimation in the errors-in-variables model. *Biometrika,* **76**, 149-160.

ZAMAR, R. (1992). Bias robust estimation in orthogonal regression. *The Annals of Statistics*, **20**, 1875-1888.