Testing structural equation models: The impact of error variances in the data generating process

Randi Hammervold
Sør-Trøndelag University College

Ulf Henning Olsson
BI Norwegian Business School

This is the authors' final, accepted and refereed manuscript to the article published in

***Quality and Quantity,* 46(2012)5: 1547-1570**

DOI: http://dx.doi.org/10.1007/s11135-011-9466-5

Testing Structural Equation Models: The impact of error variances in the data generating process.

Randi Hammervold

Sør-Trøndelag University College

Trondheim Business School,

N-7004 Trondheim, Norway

Randi.hammervold@hist.no

Ulf Henning Olsson

Norwegian School of Management, BI

0442 Oslo, Norway

Ulf.h.olsson@bi.no

Abstract

Yet another paper on fit measures? To our knowledge, very few papers discuss how fit measures are affected by error variance in the Data Generating Process (DGP). The present paper deals with this. Based upon an extensive simulation study, this paper shows that the effects of increased error variance differ significantly for various fit measures. In addition to error variance the effects depend on sample size and severity of misspecification. The findings confirm the general notion that good fit as measured by the Chi-Square, RMSEA and GFI etc. does not necessarily mean that the model is correctly specified and reliable.

One finding is that the chi square test may give support to misspecified models in situations with a high level of error variance in the DGP, for small sample sizes. Another finding is that the chi-square test looses power also for large sample sizes when the model is negligible misspecified. Other results include incremental fit indices as NFI and RFI which prove to be more informative indicators under these circumstances. At the end of the paper we formulate some guidelines for use of different fit measures.

1 Introduction

A continuous stride has been fought between different researcher communities on the use and interpretation of fit indices. It is obvious that this is a difficult issue. Our contribution will focus on the indices RMSEA, GFI, NFI, RFI, CFI and SRMR, in addition to the chi-square statistics, and on a specific source of “noise in the data”, namely variance in the error term. Two important questions regarding what constitutes a good index (Bollen & Long, 1993) are which indices to use, and how to use them. Central issues are normed vs. non-normed indices, sample size dependency vs. independency, and how to interpret the values of indices to distinguish a good from a fair or poor fit (cutoff values). Hu and Bentler (1999) examine the adequacy of the rules of thumb conventional cutoff criteria and several new alternatives for various fit indices used to evaluate model fit in practice. Jöreskog (1993), Hayduk (1996), and Schermelleh-Engel et al (2003) provide, among many others, some guidelines that help applied researchers to evaluate the adequacy of a given structural equation model.

Many earlier Monte Carlo studies have focused on properties and use of fit indices of SEMs (such as Boomsma, 1983; Anderson & Gerbing, 1984; March et al, 1988; La Du & Tanaka, 1989; Bentler, 1990; Hoogland, 1999; and Chen et al, 2008).

In this paper we will focus on the properties of goodness of fit measures when data are normally distributed, the sample size is small to medium, and where error variance in the data generating process increases. This is studied both for correctly specified models, and misspecified models. The properties of the goodness of fit measures are analyzed by means of a simulation study, for different sample sizes, levels of error variance for normally distributed variables. The estimation method is Maximum Likelihood (ML) throughout the whole study.

In the simulation study we use three main models, seven sample sizes ranging from 25 to 800 and five (six for the orthogonal factor model) different levels of error variance. The error variance levels are numbered as follows: Level 1 is low and level 5 and 6 is high. The true model will in this paper be referred to as the Data Generating Process, DGP, while the model fitted to the data will be referred to as the Assumed Model, AM. 400 replications are used for each of these combinations of sample size, level of error variance, and model type.

The study contains both a small theoretical part and a simulation study. In the theoretical part we discuss the asymptotic properties of some fit indices, namely GFI, NFI, RFI, CFI, RMSEA, and SRMR. For correctly specified models we also deduce formulas for approximate expected values and variances of these fit indices using Taylor series expansions. A comprehensive simulation study, using the computer program PRELIS to generate data and LISREL 8.80 to estimate the models, is presented.

The paper is organized as follows: The goodness of fit indices are presented in section 2, and some theoretic properties of the fit indices are discussed in section 3. This includes a short discussion of the effects of level of error variance in the data generating process. The design of the simulation study is presented in section 4. In section 5 we focus on how the fit measures works for different sample sizes and different levels of error variances. This is done both for correctly and misspecified models. We give a summary of the statistical properties of the fit indices in section 6. Last but not least we draw some guidelines about how to use the fit measures simultaneously to get information about both the data and the model under investigation.

2 Goodness of fit measures and their notation

Let $\mathbf{x}$ be a $q+1$ random vector of observables with a sample covariance matrix $\mathbf{S}$ and let $N=n+1$ denote the sample size. In SEM the model is estimated by minimizing a fit function $F(\mathbf{S},\mathbf{\Sigma(\theta)})$ with respect to a parameter vector $\boldsymbol{\theta}$ and a model implied covariance matrix, $\mathbf{\Sigma(\theta)}$.

The fit function of ML can be expressed as

$$F(\mathbf{S},\mathbf{\Sigma(\theta)}) = \log\left|\Sigma(\theta)\right| + tr\left\{S\Sigma(\theta)^{-1}\right\} - \log\left|S\right| - q \qquad (1)$$

Following the tradition, we let $\hat{F}$ denote the minimum value of $F(\mathbf{S},\mathbf{\Sigma(\theta)})$. There are various ways to measure if the model fits the data or not, but the basic measure or test, is the "chi-square test" $T=n\hat{F}$. Whether it is a real Chi-square depends on several (strict) assumptions which will be briefly discussed below. There are also various competing chi-squares based on different fit functions, but this is not the topic of this study. We will focus on ML.

The chi-square test is often referred to as the test of exact fit, where the null hypothesis $H_0:\mathbf{\Sigma}=\mathbf{\Sigma(\theta)}$ is tested, and where $\mathbf{\Sigma}$ is the population covariance matrix.

If the data comes from a multivariate normal distribution and the model holds, it is well known that $T=n\hat{F}$ has an approximate (central) χ_{ν}^{2} distribution (ν is the degrees of freedom) when *n* is large. But in most empirical research it is unreasonable to assume that the model holds exactly in the population. A consequence of this assumption is that models which hold approximately in the population will be rejected in large samples. Browne and Cudeck (1993) proposed a number of fit measures which take particular account of the error of approximation in the population and the precision of the fit measure itself. They define an estimate of the population discrepancy function as $\hat{F}_0 = \max\left\{\hat{F} - \frac{\nu}{n}, 0\right\}$ (cf. Steiger, Shapiro and Browne, 1985; McDonald; Browne and Cudeck, 1993; MacCallum, Browne and Sugawara, 1996). Since $\hat{F}_0$ generally decreases when parameters are added in the model, Browne & Cudeck (1993) suggest using Steigers's (1990) Root Mean Square Error of Approximation (RMSEA)

$$RMSEA = \sqrt{\frac{\hat{F}_0}{\nu}} = \sqrt{\frac{Max(\hat{F} - \frac{\nu}{n}, 0)}{\nu}} \qquad (2)$$

As a measure of discrepancy per degree of freedom.

We will now discuss some fit indices present in most computer programs for structural equation modeling, e.g., LISREL, and divide them into three different groups in terms of the criteria they are defined to assess;

- **Goodness of fit indices:** Directly assess how well the model accounts for the covariances. Examples are GFI, AGFI, RMR and SRMR.
- **Incremental fit indices:** Assess fit by the degree to which the model accounts for the sample covariances relative to a more restricted null-model. Examples here are NFI, RFI, NNFI, IFI, and CFI.
- **Parsimonious fit indices:** Take parsimony into account as well as fit. Examples here are PGFI, PNFI, AIC, CAIC, ECVI, and RMSEA.

As mentioned earlier, in this paper we focus on the indices GFI, NFI, RFI, CFI, RMSEA, and SRMR because they are widely used and because they give us useful information about the model fit and *error variance* in the data generating process (DGP) when they are interpreted simultaneously. The chosen fit measures represent each of the three above mentioned groups. Properties of the other fit indices are presented in Hammervold (1998).

The most widely used criterion of fit is the degree to which the model accounts for the sample covariances; the first group of fit indices satisfies this criterion. The goodness of fit index (GFI) of Jöreskog & Sörbom (1981) is an example of such an index. The GFI is a measure of the discrepancy between the sample covariance matrix ($\mathbf{S}$) and the estimated covariance matrix $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$, and measures how much better the model fits as compared to no model at all. Jöreskog and Sörbom (1984) note that the GFI is a measure of the relative amount of variances and covariances jointly accounted for by the model. The formula under maximum likelihood estimation for GFI is (Jöreskog & Sörbom, 1981; Tanaka & Huba, 1985):

$$GFI = 1 - \frac{tr(S\Sigma(\hat{\theta})^{-1} - \mathrm{I})^2}{tr(S\Sigma(\hat{\theta})^{-1})^2} \tag{3}$$

Alternatively (3) can be written as (see e.g., Yuan, 2005):

$$GFI = 1 - \frac{T_A}{T_0} \tag{4}$$

where T_A is the statistic T evaluated at the assumed model. And T_0 is the T statistic for testing $\boldsymbol{\Sigma} = \mathbf{0}$.

The Root Mean Square Residuals RMR, of Jöreskog & Sörbom (1981), is a measure of the average of the fitted residuals:

$$RMR = \sqrt{\frac{2\Sigma\Sigma(s_{ij} - \hat{\sigma}_{ij})^2}{(p+q)(p+q+1)}} \tag{5}$$

This index can only be interpreted in relations to the size of the observed variances and covariances in S. Standardized residuals, SRMR, on the other hand, are independent of the units of measurement of the variances and covariances, and provides a statistical metric for judging the size of a residual. A standardized residual is a residual divided by it’s estimated standard error. This index can be used to compare the fit of models for different data.

The second group of indices (incremental fit indices) assesses fit by the degree to which the model accounts for the sample covariances relative to a more restricted model – usually the independence model in which all variables are specified to be uncorrelated. The incremental fit indices compare the chi square for the assumed model and the chi square for the independence model in different ways. The indices are supposed to lie between 0 and 1, but for some of these indices values outside this interval can occur, and since the independence model almost always has a huge chi-square, one often obtains values very close to one. The first incremental fit indices were developed by Tucker & Lewis (1973) and Bentler & Bonett (1980). Other variations of these have been proposed and discussed by Bollen (1986, 1989) and Bentler (1990).

The formula for the normed fit index NFI, is as follows (Bentler & Bonett, 1980):

$$NFI = 1 - \frac{\hat{F}}{\hat{F}_i} \tag{6}$$

where $\hat{F}_i$ is the estimated minimum value of the fit function for the independence model.

Equivalently (6) can be written as (see e.g., Yuan, 2005):

$$NFI = 1 - \frac{T_A}{T_i} \tag{7}$$

where T_i is the T-statistics for the independence model.

NFI is dependent of sample size, and therefore Bollen (1986) proposed a simple alternative fit measure, RFI, that was supposed to remove the dependency of sample size:

$$RFI = \frac{\frac{\hat{F}_i}{\nu_i} - \frac{\hat{F}}{\nu}}{\frac{\hat{F}_i}{\nu_i}} \tag{8}$$

Another popular member of this group is the comparative fit index, CFI (Bentler, 1990) which attempts to measure the relative reduction in the non-centrality parameters of the estimated model and the independence model:

$$CFI = 1 - \frac{\max(n\hat{F} - \nu, 0)}{\max(n\hat{F}_i - \nu_i, n\hat{F} - \nu, 0)} \quad (9)$$

However, these two foregoing families of indices do not address the problem that good fit can be obtained simply by using a very large number of parameters relative to the degrees of freedom in the model. This leads us to the third group, the parsimonious fit indices, which reflect the degrees of freedom available. The parsimonious fit indices fall into three sub-groups: Those based on adjusting general goodness of fit indices (PNFI and PGFI), those based on the chi-square measure (AIC, CAIC, ECVI), and those based on the discrepancy due to approximation (RMSEA). In this paper results for RMSEA will be presented.

3 Theoretical results for GFI, NFI, RFI, and RMSEA

In this section we present some theoretical results useful for evaluation of the goodness of fit indices. In section 3.1 we discuss the effect of sample size, in section 3.2 and 3.3 we deduce the asymptotic properties of these fit indices, including approximate expected values and variances. The influence of increased error variance on the fir measures will be discussed in section 3.4. The theoretical results from this section will be applied and compared with the results from the simulation study in sections 5.

3.1 Sample size dependency

In the literature, two main sample size influences on goodness of fit indices are discussed (cf. Bollen, 1990). The first is when sample size enters directly into the computation of fit indices (e.g. AIC), and the second is when the sampling distribution of the fit measure is affected by sample size (e.g. NFI).

The discussion about sample size dependency has led many researchers to require that the fit indices should be independent of sample size. Gerbing & Anderson (1993) have summarized the results from major simulation studies of fit indices. They argue that the ideal fit index should be independent of sample size. Higher or lower values of the fit indices will then not be obtained simply because the sample size is large or small.

On the other hand one may argue that estimation in small samples is more uncertain, and that the fit indices should reflect this. Cudeck & Henly (1991), for example, argue that the influence of sample size is not necessarily undesirable. We will however not enter into the discussion of sample size dependency, but look into the interaction between sample size and the error variance in the DGP. We analyze how this affects RMSEA, NFI, GFI, RFI, CFI, and SRMR differently and also the chi-square statistic, here denoted T.

3.2 Asymptotic results

If the model is correctly specified we asymptotically have that:

$$S \xrightarrow{P} \Sigma_0 \text{ and } \hat{\Sigma} \xrightarrow{P} \Sigma_0 \qquad (10)$$

where $\xrightarrow{P}$ means convergence in probability. Therefore

$$\hat{F} \xrightarrow{P} 0 \text{ as } n \to \infty . \qquad (11)$$

Referring to the formulas for RMSEA, NFI, RFI, CFI, GFI and SRMR (formulas 2-9), it is obvious that GFI, NFI, RFI and CFI will converge in probability towards 1 as $n \to \infty$.

RMSEA and SRMR will converge in probability towards zero as $n \to \infty$.

3.3 Approximated expected values and variances of GFI, NFI, RFI and RMSEA

In this section we will compute approximated expected values and standard deviations of some of the actual fit indices, using Taylor series expansions. Here we assume correctly specified models. In section 5 these formulas will be applied and compared with the simulation results.

Let $(x_1,....,x_n)$ be a stochastic vector with $E(x_j)=\mu_j$ and $Var(x_j)=\sigma_j{}^2$, j=1, …,n. Consider the function $y=g(x_1,....,x_n)$, where we assume that there exists continuous first and second order partial derivatives around $(\mu_1,...,\mu_n)$. If σ_j^2 is sufficiently small we may yield an approximated expected value and variance for the function $y=g(x_1,....,x_n)$ as follows (here we have applied first order Taylor series):

$$E\left[g(x_1,\dots,x_n)\right] \approx g(\mu_1,....\mu_n) \tag{12}$$

$$Var\left[g(x_1,...,x_n)\right] \approx \sum_j \left(\frac{\partial g(\mu_1,...,\mu_n)}{\partial \mu_j}\right)^2 \sigma_j{}^2 + 2\sum_{j<k} \frac{\partial g}{\partial \mu_j}\frac{\partial g}{\partial \mu_k} Cov(x_j, x_k) \tag{13}$$

We will now apply these formulas for the actual fit indices. Applying these formulas for the actual fit indices gives the following results for the expectation and variance:

$$E(GFI) \approx \frac{p+q}{p+q+\dfrac{2\nu}{n}} \tag{14}$$

$$Var(GFI) \approx \frac{n^2 8\nu(p+q)^2}{(n(p+q)+2\nu)^4} \tag{15}$$

$$E(NFI) \approx 1 - \frac{\nu}{nF_0^{\ i} + \nu_i} \tag{16}$$

$$Var(NFI) \approx \left(\frac{2\nu}{(nF_0^{\ i} + \nu_i)^2}\right) + \left(\frac{2n\nu^2}{(nF_0^{\ i} + \nu_i)^4}\right)\left(2F_0^{\ i} + \frac{\nu_i}{n}\right) \tag{17}$$

$$E(RFI) \approx \frac{F_0^{\ i}}{F_0^{\ i} + \frac{\nu_i}{n}} \tag{18}$$

$$Var(RFI) \approx \frac{2\nu_i^{\ 2}}{\left(nF_0^{\ i} + \nu_i\right)^2 \nu} + \frac{2\nu_i^{\ 2}\left(2F_0^{\ i} + \frac{\nu_i}{n}\right)}{n^3\left(F_0^{\ i} + \frac{\nu_i}{n}\right)^4} \tag{19}$$

$$E(RMSEA) \approx \sqrt{\frac{P(\chi^2_{\nu+2} > \nu) - P(\chi^2_{\nu} > \nu)}{n}} \tag{20}$$

$$Var(RMSEA) \approx \frac{Var(F_0)}{2\nu\sqrt{\frac{1}{n}\left(P(\chi^2_{\nu+2} > \nu) - P(\chi^2_{\nu} > \nu)\right)}} \tag{21}$$

$$Var(F_0) \approx \frac{\nu}{n^2}\Big(\left((\nu+2)P(\chi^2_{\nu+4} > \nu) - 2\nu P(\chi^2_{\nu+2} > \nu) + \nu P(\chi^2_{\nu} > \nu)\right) -$$

$$\nu\left(P(\chi^2_{\nu+2} > \nu) - P(\chi^2_{\nu} > \nu)\right)^2\Big)$$

Where:

p	Number of observed y variables
q	Number of observed x variables
$\hat{F}_0$	Estimated population discrepancy function
F_0^i	Fit function value for the independence model fitted to the population covariance matrix.
ν	Degrees of freedom for the estimated model
ν_i	Degrees of freedom for the independence model

$n = N - 1$ N is the sample size

3.4 The impact of increased error variance in the DGP

MacCallum & Tucker (1991) discusses several sources of model error in common factor models. They identify several sources of error both in the model and in the sample. Error variance in the Data Generating Process (i.e., in “the true model”) is not a model error but a kind of “noise” arising from the fact that the selected variables has a large unique variance. Since very little has been written in the context of fit indices and the influence of large unique variances in the variables, we will here make a short introduction into the subject by two simple examples. Through a simple simulation example we show how a goodness of fit index and an incremental fit index, GFI and NFI, behave differently under different levels of error variance.

In the following example we increase the error variance in the data by increasing the variance of the error term in the DGP, which here is a simple system of simultaneous regression models.

Example 1: Multiple regression models. The effect of increased error variance in DGP on GFI and NFI.

In this example we simulate data from the following simultaneous equation model (DGP), with sample size 200 and only one replication:

$$y_1 = 0.5x_1 + \zeta_1$$

$$y_2 = 0.7x_2 + \zeta_2$$

where $x_1 \sim 0.7$ N(0,1), $x_2 \sim 0.5$ N(0,1), and ζ_1 and ζ_2 are independent normally distributed with standard deviations 0.4 and 0.3 respectively.

For the simulated data we estimate the following model which is correctly specified:

$$y_1 = \gamma_{11} x_1 + \zeta_1$$

$$y_2 = \gamma_{22} x_2 + \zeta_2 .$$

We start with standard deviations $SD(\zeta_1) = 0.4$ and $SD(\zeta_2) = 0.3$, and increasing the error variance by multiplying these values with 2, 4, 8 and 16 respectively. The results are presented in table 1.

Insert table 1 about here.

As we observe, the GFI values are all close to one for all levels of error variance, indicating good fit. On the other hand NFI shows a significant decrease with increasing error variance. So, NFI seems to be affected by error variance in the DGP, while GFI is not affected at all. Chi-square seems to slightly increase. However, we can not draw any general conclusion based on these four single samples.

From formulas (3) and (14), and from (6) and (16), for GFI and NFI respectively, we see that the “test statistic” for the independence model only is present for NFI. This is why NFI is affected by error variance, in contrast to GFI, RMSEA, and the chi-square statistic.

A simple example will illustrate this.

Example 2: A CFA- model.

Let us assume that the DGP is CFA model (measurement model) given by

$$\mathbf{x} = \mathbf{\Lambda\xi} + \mathbf{\delta} \tag{22}$$

where the model implied matrix is:

$$\mathbf{\Sigma} = \mathbf{\Lambda\Phi\Lambda}' + \mathbf{\Theta}_{\delta} \tag{23}$$

If we now let all λ_i 's in the DGP approach zero simultaneously, the model implied matrix in (23) will approach a diagonal matrix, namely $\mathbf{\Theta}_{\delta}$. If we, in this extreme situation, fit a correctly specified model (i.e., AM is correctly specified) the chi square, GFI and RMSEA will indicate perfect fit, but NFI will be close to zero. This is due to the fact that DGP (in the limit, when all λ_i 's are approaching zero) is the independence model, and T_A=T_i in equation (7).

In formula (16) we have the expected approximated value for NFI given by:

$E(NFI) \approx 1 - \frac{\nu}{nF_0^{\,i} + \nu_i}$. Since the assumed model is correctly specified, $F_0^i = 0$ and $\nu = \nu_i$, the expected value of NFI will be zero. This is in accordance with Bentler and Bonett (1980) and Bentler (1990) who made it clear that relative fit indices aim to measure the improvement of a substantive model over the independence model.

Based on the discussion above, we hypothesize:

1.

If there is a DGP where the variance of the error terms are large, fitting a model with a <u>*correctly specified structure*</u> *will result in a small chi-square and a small RMSEA, a GFI approaching 1, and a decreasing NFI, which approaches zero.*

Due to the fact that RFI is very similar to NFI we believe that RFI will decrease and approach zero. When it comes to CFI and due to it's complex form it is hard to predict how it *will behave. But, belonging to the group of Incremental fit indices it is reasonable that it behaves similar to NFI and RFI. Further we also hypothesize that SRMR will be unaffected by increased error variance being a descriptive goodness of fit measure only measuring the residuals. Note that the residuals, in this setting, are the differences between the observed and the estimated (predicted) covariances.*

2.

If there is high degree of error variance in the data, fitting a misspecified model, will result in the same behaviour for the actual fit statistic and fit indices as hypothesized *above* (1).

If the AM is a CFA-model where the error variance in DGP is large, this is obviously a non-reliable measurement model, but it can still fit the data very well, reflected in the chi-square and GFI. Low reliability is at the same time indicated by the low NFI.

4 Models and design

Perhaps the most obvious decision facing the Monte Carlo researcher is the choice of representative models. The choice of models certainly has implications for the generalization of results. Given the broadness of models suited for a LISREL-analysis (Bollen, 1989), restrictive choices are necessary. The following models were incorporated in our study: 26 data generating processes denoted DGPR 1 -5, DGPFU 1- 6, DGPFC 1 – 15. Two assumed models (AM) which are correctly specified for DGPR 1-5 and DGPFU 1-6 respectively, and one assumed model which is misspecified, for DGPFC 1-15. I.e., a total of 26 assumed models.

DGPR 1 -5 are five multivariate regression models (path models), with uncorrelated x variables, where there are five levels of error variance. DGPU 1 – 6 are six factor models with uncorrelated factors for six different levels of error variance. DGPFC 1 -15 are 15 factor models with two common and correlated factors where the correlation is 0.5, 0.8 and 0.95 respectively. For each of theses three, there are five levels of error variance. See tables 2-4 for the mathematical representation of the models.

Insert table 2 about here.

Insert table 3 about here.

Insert table 4 about here.

For the data generating processes described above, we will estimate the following models:

T*he assumed Path Models for DGPR 1-5*:

$$y_1 = \gamma_{11} x_1 + \zeta_1$$
$$y_2 = \gamma_{22} x_2 + \gamma_{23} x_3 + \zeta_2$$
$$y_3 = \gamma_{31} x_1 + \gamma_{33} x_3 + \zeta_3$$

The assumed factor models with uncorrelated factors for DGPFU 1-6:

$$x_1 = \xi_1 + \delta_1$$
$$x_2 = \lambda_{2,1}\xi_1 + \delta_2$$
$$x_3 = \lambda_{3,1}\xi_1 + \delta_3$$
$$x_4 = \lambda_{4,1}\xi_1 + \lambda_{4,2}\xi_2 + \delta_4$$
$$x_5 = \lambda_{5,1}\xi_1 + \xi_2 + \delta_5$$
$$x_6 = \lambda_{6,2}\xi_2 + \delta_6$$
$$x_7 = \lambda_{7,2}\xi_2 + \delta_7$$
$$x_8 = \lambda_{8,2}\xi_2 + \delta_8$$

The assumed one factor model for the two factor models DGPFC 1-15. This model is misspecified. When $Cor(\xi_1, \xi_2)$ in DGP approaches 1, the misspecification decreases towards zero:

$$x_1 = \xi_1 + \delta_1$$
$$x_2 = \lambda_{2,1}\xi_1 + \delta_2$$
$$x_3 = \lambda_{3,1}\xi_1 + \delta_3$$
$$x_4 = \lambda_{4,2}\xi_1 + \delta_4$$
$$x_5 = \lambda_{5,2}\xi_1 + \delta_5$$
$$x_6 = \lambda_{6,2}\xi_1 + \delta_6$$

Sample sizes

We have chosen sample sizes as follows: 25, 50, 75, 100, 200, 400, and 800 for DGPR 1 -5 and DGPFU 1- 6. For DGPFC 1 – 15, sample size is 100, 200, 400, and 800, respectively.

Number of replications

We have chosen the number of replications to be 400 throughout the whole simulation study. For the simulated values of the fit indices we compute means and standard deviations.

Reliability

Tables 5, 6 and 7 present some reliability measures, namely R-square, composite reliability measure (Bagozzi and Yi, 1988), and average variance extracted (Fornell and Larcker, 1981).

Insert table 5 about here.

Insert table 6 about here.

Insert table 7 about here.

In table 5 we observe that the R-square for the three equations in the path model, decrease dramatically for increased level of error variance. In table 6 we observe that the R-square for the factor model also is low for the highest level of error variance (level 6); at the same time the composite reliability for factor one and two in table 7 is 0.83 and 0.81, which is regarded as an indication of high reliability. On the other hand the average variance extracted is only 0.50 and 0.49, which is relatively low (Fornell and Larcker, 1981).

5 Results from the simulation study

5.1 Chi-square test

The tables below shows the proportion of the 400 replications where the chi-square test rejects (at the 0.05 level) the correctly specified path model 1, the correctly specified factor model 2, and the misspecified factor model 3, 4 and 5.

Insert table 8 about here.

Insert table 9 about here.

Correctly specified models (model 1 and 2) for DGPR 1 -5 *and* DGPFU 1- 6

For the correctly specified path model 1 the effect of sample size on the chi-square is substantial (table 8). For small sample sizes (25 and 50) the test seems to reject the model too often. It is interesting to note that the rejection rate seems to decrease as the error variance increases. For the level 1, and N = 25, 50 and 75, the rejection rate is 15.2 %, 8.2 %, and 6.5 % respectively, for level 5 it is 9.0 %, 6.0 %, and 3.0 % respectively. We also note the high rejection rate at level 1 for sample size 400. For the other sample sizes, it appears to be an insignificant effect on the chi-square test for increasing error variance (see table 8). These results are in accordance with e.g. Boomsma (1983) who showed that for small sample sizes (25 and 50) the chi square values tend to be too large. Boomsma (1983) also found the chi square statistics to be close to the theoretical values for sample sizes above 100.

For the factor model 2 (model 2 in table 8), the chi-square test rejects the correctly specified model too often for sample sizes 25, 50 and 75. For N = 50 we observe that the rejecting rate decrease from 9% for level 1 to 3.3 % for level 5. For the other sample sizes there are no clear patterns.

Misspecified models (model 3, 4 and 5) for DGPFC 1 – 15

Model 3, model 4 and model 5 represent three misspecified versions of the DGP, where the degree of misspecification decreases from model 3 to 5. Model 3, where the "true" correlation between the common factors is 0.5, is the most severely misspecified model, model 4, where the "true" correlation is 0.8 is less misspecified and model 5, where the true correlation is 0.95, is minor misspecified. Table 9 shows that for all models the rejection rate decreases as the error variance increases.

For the most severely misspecified models 3 and 4 the rejecting rate is high and near 100 for the lowest error variances. When the error variance increases we observe that the rejection rate decreases rapidly, especially for small and moderate samples sizes (100 and 200). In other words: Too few misspecified models will be rejected, when the error variance is high.

We also find the same pattern for the less misspecified model 4 and 5. It is interesting to note that for model 4, for N = 400, the rejection rate is only 40.1 %, 39.0 %, and 17.5 % respectively for level 3, 4 and 5. The same numbers for model 5 are: 17.5 %, 6.25 % and 5.5 %. I.e.: <u>A significant decrease in the power of the test as a function of increasing error variance.</u> For model 5 we observe that the rejection rates are lower for larger sample sizes. So models with small misspecifications and high error variance, <u>will not be rejected even if the sample size is large</u>.

Concluding remarks: 1) When the error variance in the data generating process is high, which for CFA models (measurement models) will mean low reliability, misspecified models tend to be accepted too often. I.e., *the chi-square test looses power as a function of error variance in the DGP*. This is most significant for small sample sizes. The less misspecified the model is, the less the effect from sample size. 2) When the error variance in the data generating process is low, misspecified models will correctly be rejected. In addition the model will be rejected when the model is negligible misspecified and the sample size is large,. 3) When the model is correctly specified, the rejection rate is close to 5% and seems to be relative independent of sample size and level of error variance.

So, for a correctly specified model, the chi-square test does not give any information about error variance in the DGP. But if the model is misspecified, the test tends to loose power as a function of increasing error variance in the DGP.

In the next chapter, we will see that it is important to supplement the chi square test with other goodness of fit measures. These fit measures will provide valuable information about the model fit and the data, which we do not find in the chi-square test.

5.2 Results for the indices GFI, NFI, RFI, CFI, RMSEA, and SRMR

For each fit index, we present tables that show the mean values and standard deviations of the 400 replications, for all sample sizes and different levels of error variance in the DGP.

GFI, NFI and RMSEA for the correctly specified models (model 1 and 2):

Insert table 10 about here.

Insert table 11 about here.

Insert table 12 about here.

For the correctly specified path models GFI is increasing from 0.91 to 1.00, when sample size increases from 25 to 800 (table 10). Our findings are in accordance with the results of the simulation study of e.g. Anderson & Gerbing (1984). For low error variance, NFI is increasing from 0.89 to 1.00 when sample size is increasing from 25 to 800 (table 11). Our results agree with earlier simulation studies, e.g. NFI in simulation studies of Bearden et al (1982). RMSEA decreases from 0.093 to 0.008 as the sample size increases, for low error variance (table 12).

The mean values of GFI are not affected by the increased error variance at all. This is valid for all sample sizes (see table 10). Hence GFI is very robust against increased error variance. NFI decreases with increasing error variance (see table 11). These observations are in accordance with our theoretical assumptions (see section 3.4). A related phenomenon (but not so obvious as in our results) is noted by Brown et all (2002). RMSEA is not affected by increased error variance at all (table 12). For all levels of error variance RMSEA is decreasing towards zero when sample size increases. Therefore RMSEA seems to be unaffected by the increasing of error variance. This is in accordance with our theoretical assumptions (see section 3).

The fit index with the smallest standard deviation is GFI. NFI has somewhat larger standard deviations than GFI. The index with the largest standard deviations is the RMSEA. Here standard deviations are approximately 25 times as large as for GFI. For RMSEA the standard deviations are about the same size as the mean values.

For increasing error variance, the standard deviations of GFI and RMSEA appear to be relatively constant. The standard deviations for NFI are increasing with increased error variance.

In section 3 we derived formulas for expected values and standard deviations of the fit indices (see equations 14 – 21). For the correctly specified regression models in tables 10, 11, and 12 we note the close correspondence between the simulation results and the computations of approximated expected values and standard deviations.

To conclude: For correctly specified models, mean values and standard deviations of GFI are robust against increased error variance. This index gives information solely about whether the structure of AM fits the structure of the DGP, and tells nothing about the error variance in the data generating process. The structure of the model is the same even if the error variance increases. On the other hand, the mean values of NFI decrease as the error variance increases. We observe that GFI and NFI have different properties: For high levels of error variance, GFI will indicate good fit, while NFI will indicate bad fit. In this situation these two indices simultaneously interpreted will give useful information.

Misspecified models (model 3, 4 and 5) for DGPFC 1 – 15

The mean values of GFI for model 3 and 4 are far from 1 when the error variance is low, indicating that the models are misspecified (table 13). The mean values of GFI are increasing for increasing error variance. The index therefore indicates a better fit for the misspecified model as the error variance increases. This means that for a high level of error variance in the DGP, *the index GFI will not necessarily indicate misspecification for a misspecified model*. For example, for model 3, level 1 of error variance and sample size 200, the mean value for GFI is 0.693. The corresponding value for level 5 of error variance is 0.974!

For model 5 (table 13) where the misspecification is small, GFI tends to have high values for all levels of error variance.

The index GFI has low values for misspecified models, but the mean values increase for increasing error variance. This is the same pattern as for chi-square test. Significant misspecified models with high error variance will get misleading support from the GFI index.

Insert table 13 about here.

NFI behave differently (table 14). For severely misspecified models NFI has low values, indicating bad fit. For increasing error variance, the mean values increase significantly less than for GFI. When error variance is high, NFI still indicates misspecification.
For the models where the misspecification is negligible (model 5), the NFI seems to decrease with increasing error variance. This is the same pattern as described for the correctly specified regression and factor models (model 1 and 2).

Insert table 14.

The incremental fit indices RFI and CFI (tables 15 and 16) behave as NFI. In fact RFI seems to be more sensitive to increased error variance than NFI. CFI also behaves as NFI, but the mean values are higher than for NFI.

Insert table 15.
Insert table 16.

The simulations indicate that RMSEA (table 17) behaves similar to GFI, which is not surprising given its mathematical relation to the chi-square statistic. For significant misspecified models RMSEA is high, indicating bad fit. For increasing error variance, RMSEA decreases, indicating better fit for misspecified models. So a misspecified model where the error variance in DGP is high may get support from RMSEA!

Insert table 17 about here.

SRMR seems to decrease for increasing error variance even if the model is severely misspecified (table 18). On the other hand if the misspecification is negligible (model 5), the SRMR seems to increase, indicating worse fit.

Insert table 18 about here.

6 Concluding remarks and recommendations

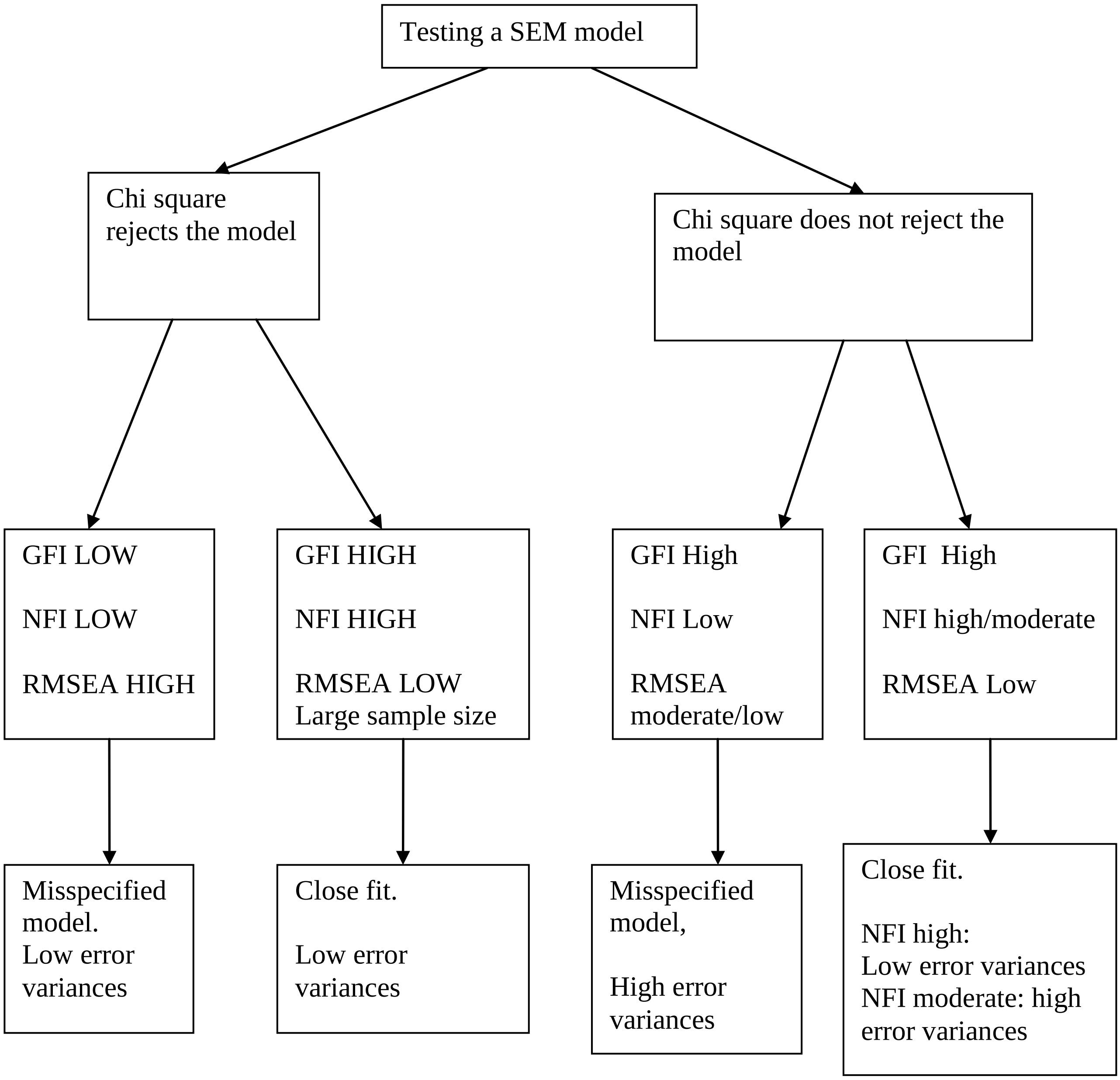
All fit indices attempt to measure, in different ways, the discrepancy between the sample covariance matrix S and the model implied matrix $\Sigma(\theta)$. When the error variance in DGP increases, the S matrix will more and more look like a diagonal matrix. The fitted model will therefore approach the independence model. It is therefore reasonable that large amount of error variance, give rice to the very different behavior of the fit indices, which we have observed. To conclude:

1) When the error variance in the data generating process is high, which for CFA models (measurement models) will mean low reliability; misspecified models tend to be accepted too often by the chi-square test. I.e., *the chi-square test looses power as a function of error variance in the DGP:* For severely misspecified models the decrease in power is most pronounced for small sample sizes. On the other hand if the model is minor misspecified the decrease in power is only present for large sample sizes. At the same time the incremental fit indices (NFI, RFI, and CFI) tend to have low values, while GFI tends to have high values. This effect will increase with increased error variance and higher misspecification. Of the incremental fit indices, RFI seems to be

most sensitive to the combination of misspecification and high error variance, while CFI seems to be less sensitive.

2) When the error variance in the data generating process is low, misspecified models will correctly be rejected by the chi-square test. As we know, the power of the chi-square test increases with increasing sample size implying rejection of models where the misspecification is negligible. In this situation we can consult GFI and NFI, which is common practice.
3) When the model is correctly specified, the rejection rate for the chi square test is close to 5% and seems to be independent of sample size and the level of error variance.

For the researchers testing SEM models, the following guidelines can be useful to interpret the fit measures more correctly:



Our simulation study shows that neither the chi-square test, RMSEA, GFI, nor NFI can reliable give information about the model fit and the variance of the error terms as a single measure or fit. However, they behave differently under the impact of the level of error variance, misspecification, and sample size. Consequently the fit indices should favorably be interpreted simultaneously as demonstrated in the scheme above.

The problem is misspecification: If the model is misspecified, the chi-square test, the RMSEA or the GFI will loose power as a function of increased error variance in the DGP. This can imply acceptance of severely misspecified models. If the estimated error variances are high and at the same time the effect parameters (i.e., factor loadings) are low one should inspect the incremental fit indices NFI, RFI, and CFI. They can give valuable information when the chi-square, the GFI, and the RMSEA do not work.

Admittedly our results and implications may not be generalized across all types of misspecification, models and estimation methods. We have only focused on ML and on three different models. Misspecification is discussed in one of the three models, namely a relatively simple two factor CFA – model with correlated factors. Further research should investigate more complex models, and also include other estimation methods e.g., the GLS and ULS. Likewise studies should be performed to look into the effects of non-normal data on the simultaneous interpretation of fit measures as proposed in this paper.

7 Literature

Anderson, J. C., *&* Gerbing, D. W. (1984). The effects of sampling error on convergence, improper solutions and goodness of fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika*, 49: 155-173.

Bagozzi R.P. and Yi, Y. (1988) On the evaluation of structural equation models. *Journal of the Academy of Marketing Science,* 16: 74-94.

Bearden, W. O, Sharma, S., & Teel, J.R. (1982). Sample size effects on chi-square and other statistics used in evaluating causal models. *Journal of Marketing Research*, 19, 425-430.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological bulletin,* 107, 238-246.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin.* 88, 588-606.

Bollen, K. A. (1986) . Sample size and Bentler and Bonett's non-normed fit index. *Psychometrika,* 51, 375-377.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Bollen, K. A. (1990). Overall fit in covariance structure models: Two types of sample size effects. *Psychological Bulletin,* 107, 256-259.

Bollen, K. A. & Long, J. S. (1993). *Testing structural equation models*. Sage Publications.

Boomsma, A. (1983*). On the robustness of LISREL (maximum likelihood estimation) against small sample size and nonnormality*. Amsterdam: Sociometric Research Foundation.

Browne, M. W. & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), Testing structural equation models (pp. 136-162). Sage publications

Browne, M. W. et al (2002). When Fit Indices and Residuals are Incompatible. *Psychological Methods*, 2002, Vol. 7, No. 4, 403-421.

Chen, F., Curran P. C., Bollen, K. A, Kirby J., Paxton P. (2008). An Empirical Evaluation of the Use of Fixed Cutoff Points in RMSEA Test Statistic in Structural Equation Models. *Sociological Methods & Research*. Volume 36 number 4, May 2008 462-494.

Cudeck, R. & Henly, S. J. (1991). Model selection in covariance structures analysis and the problem of sample size: A clarification. *Psychological Bulletin, 109, 512-519.*

Fornell, C. and Larcker, D.F. (1981). Structural equation models with unobservable variables and Measurement error. *Journal of Marketing Research*, 18: 39-50.

Gerbing D. W. & Anderson J. C. (1993). Monte Carlo Evaluation of Goodness of fit indices for structural equation models, in *Testing structural equation models*, edited by Bollen & Long, 1993. Sage publications.

Hayduk, L. A. (1996). *LISREL issues, debates, and strategies*. Baltimore: The Johns Hopkins University Press.

Hammervold, R. (1998). *Properties of goodness of fit statistics for structural equation models*. NTNU Trondheim Norges tekniske naturvitenskapelige universitet. Dr. Scient thesis.

Hoogland, J.J. (1999). *The robustness of estimation methods for covariance structure analysis.* Rijksuniversiteit Groningen. Thesis.

Hu, Li-tze, Peter M. Bentler (1999). Cutoff Criteria for Fit Indexes in Covariance Structures Analysis: Conventional Criteria Versus New Alternatives. Structural Equation Modelling, 6(1), 1-55.

Jöreskog, K. G. (1993). Testing structural equation models, in *Testing Structural Equation Models*, edited by Bollen & Long, 1993. Sage publications.

Jöreskog, K. G. & Sörbom, D. (1981). *LISREL V. Analysis of linear structural relationships by maximum likelihood and least squares methods*. Department of statistics. University of Uppsala.

Jöreskog, K. G. & Sörbom, D. (1984). *LISREL VI – Users guide*, 3.rd. ed. Mooresville, IN: Scientific Software.

La Du T. J. & Tanaka J. S. (1989). Influence of sample size, estimation method, and model specification on goodness of fit assessment in structural equation models. *Journal of Applied Psychology,* 74, 625-635.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). *Power analysis and determination of sample size for covariance structure modeling*. Psychological Methods, 1, 130-149.

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

Marsh, H. W., Balla, J. R. & McDonald, R. P. (1988). Goodness of fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological bulletin*, 103, 391-410.

McDonald, R. P. (1989) . An index of goodness of fit based on non-centrality. *Journal of classification*, 6, 97-103.

Schermelleh-Engel, K., Moosbrugger, H. & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, 8, 23-74. Available: http://www.mpr-online.de/

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, 25, 173-180.

Steiger, J.H., & Lind, J.C. (1980). Statistically-based tests for the number of factors. Paper presented at the *Annual Spring Meeting of the Psychometric Society*. Iowa City, Iowa.

Steiger, J. H., Shapiro, A. & Browne. M. W. (1985). On the multivariate asymptotic distribution of sequential chi-square statistics. *Psychometrika*, 50, 253-263.

Tanaka, J. S. & Huba, G. J. (1985). A fit index for covariance structure models under arbitrary GLS estimation. *British Journal of Mathematical and Statistical Psychology*, 38, 197-201.

Tucker, L. R. & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika,* 38, 1-10

Yuan, Ke-Hai. (2005). Fit indices versus Test Statistics. *Multivariate Behavioral Research*. 40(1), 115-148.

Table 1: Simultaneous equation model. The goodness of fit indices for different levels of error variance (NR=1).

Goodness of fit measures	LEVELS OF ERROR VARIANCE – Standard deviations of the error terms			
	$SD(\zeta_1) = 0.8$ $SD(\zeta_2) = 0.6$	$SD(\zeta_1) = 1.6$ $SD(\zeta_2) = 1.2$	$SD(\zeta_1) = 3.2$ $SD(\zeta_2) = 2.4$	$SD(\zeta_1) = 6.4$ $SD(\zeta_2) = 4.8$
Chi square statistic (df=3)	3.6	4.7	6.9	4.1
RMSEA	0.03	0.038	0.057	0.031
GFI	0.99	0.99	0.99	0.99
NFI	0.95	0.89	0.84	0.69

Table 2: Path models (model 1). DGPR 1-5

DGPR 1	DGPR 2	DGPR3	DGPR4	DGPR5
$y_1 = 0.15x_1 + \zeta_1$ $y_2 = 0.25x_2 + 0.35x_3 + \zeta_2$ $y_3 = 0.40x_1 + 0.10x_3 + \zeta_3$	The same as for DGPR1	The same as for DGPR1	The same as for DGPR1	The same as for DGPR1

Where we assume: x_1, x_2 and x_3 are independent, and ζ_1, ζ_2 and ζ_3 are uncorrelated.

- $x_i \sim N(0,1)$ for i = 1,2,3
- $\zeta_1 \sim 0.3\ N(0,1)$, $\zeta_2 \sim 0.4\ N(0,1)$ and $\zeta_3 \sim 0.2\ N(0,1)$
- Five levels of error variance (measured by the standard deviations):

$SD(\zeta_1) = 0.3$	$SD(\zeta_1) = 0.6$	$SD(\zeta_1) = 0.9$	$SD(\zeta_1) = 1.2$	$SD(\zeta_1) = 1.5$
$SD(\zeta_2) = 0.4$	$SD(\zeta_2) = 0.8$	$SD(\zeta_2) = 1.2$	$SD(\zeta_2) = 1.6$	$SD(\zeta_2) = 2.0$
$SD(\zeta_3) = 0.2$	$SD(\zeta_3) = 0.4$	$SD(\zeta_3) = 0.6$	$SD(\zeta_3) = 0.8$	$SD(\zeta_3) = 1.0$

Table 3: Factor model with uncorrelated factors (model 2), DGPFU 1-6

DGPFU1	DGPFU2	DGPFU3	DGPFU4	DGPFU5	DGPFU6
$x_1 = \xi_1 + \delta_1$ $x_2 = 2.0\,\xi_1 + \delta_2$ $x_3 = 1.6\,\xi_1 + \delta_3$ $x_4 = 2.5\xi_1 + 1.6\xi_2 + \delta_4$ $x_5 = 1.9\,\xi_1 + \xi_2 + \delta_5$ $x_6 = 1.8\,\xi_2 + \delta_6$ $x_7 = 2.65\,\xi_2 + \delta_7$ $x_8 = 3.5\,\xi_2 + \delta_8$	The same as for DGPFU1	The same as for DGPFU1	The same as for DGPFU1	The same as for DGPFU1	The same as for DGPFU1

Where we assume:

- $\xi_1 \sim 0.75\ N(0,1)$, $\xi_2 \sim 0.87\ N(0,1)$ and independent of each other.
- $\delta_1 \sim 0.3\ N(0,1)$ $\delta_2 \sim 0.4\,N(0,1)$ $\delta_3 \sim 0.5\ N(0,1)$
  $\delta_4 \sim 0.45\ N(0,1)$ $\delta_5 \sim 0.65\ N(0,1)$ $\delta_6 \sim 0.43\ N(0,1)$
  $\delta_7 \sim 0.78\ N(0,1)$ $\delta_8 \sim 0.52\ N(0,1)$
- $\delta_1, \delta_2, ..., \delta_8$ are independent of each other.
- Six levels of error variance:

$SD(\delta_1) = 0.3$	$SD(\delta_1) = 0.6$	$SD(\delta_1) = 0.9$	$SD(\delta_1) = 1.2$	$SD(\delta_1) = 1.5$	$SD(\delta_1) = 1.8$
$SD(\delta_2) = 0.4$	$SD(\delta_2) = 0.8$	$SD(\delta_2) = 1.2$	$SD(\delta_2) = 1.6$	$SD(\delta_2) = 2.0$	$SD(\delta_2) = 2.4$
$SD(\delta_3) = 0.5$	$SD(\delta_3) = 1.0$	$SD(\delta_3) = 1.5$	$SD(\delta_3) = 2.0$	$SD(\delta_3) = 2.5$	$SD(\delta_3) = 3.0$
$SD(\delta_4) = 0.45$	$SD(\delta_4) = 0.9$	$SD(\delta_4) = 1.35$	$SD(\delta_4) = 1.8$	$SD(\delta_4) = 2.25$	$SD(\delta_4) = 2.7$
$SD(\delta_5) = 0.65$	$SD(\delta_5) = 1.3$	$SD(\delta_5) = 1.95$	$SD(\delta_5) = 2.6$	$SD(\delta_5) = 3.25$	$SD(\delta_5) = 3.9$
$SD(\delta_6) = 0.43$	$SD(\delta_6) = 0.86$	$SD(\delta_6) = 1.29$	$SD(\delta_6) = 1.72$	$SD(\delta_6) = 2.15$	$SD(\delta_6) = 2.58$
$SD(\delta_7) = 0.78$	$SD(\delta_7) = 1.56$	$SD(\delta_7) = 2.34$	$SD(\delta_7) = 3.12$	$SD(\delta_7) = 3.9$	$SD(\delta_7) = 4.68$
$SD(\delta_8) = 0.52$	$SD(\delta_8) = 1.04$	$SD(\delta_8) = 1.56$	$SD(\delta_8) = 2.08$	$SD(\delta_8) = 2.6$	$SD(\delta_8) = 3.12$

Table 4: A two-factor model with correlated factors, correlation = 0.5, 0.8 and 0.95 respectively (model 3, 4 and 5): DGPFC 1-15

Two factor model. Decreasing factor loadings and increasing error variances.				
1	2	3	4	5
$x_1 = \xi_1 + \delta_1$	$x_1 = \xi_1 + \delta_1$	$x_1 = \xi_1 + \delta_1$	$x_1 = \xi_1 + \delta_1$	$x_1 = \xi_1 + \delta_1$
$x_2 = 0.8\xi_1 + \delta_2$	$x_2 = 0.7\xi_1 + \delta_2$	$x_2 = 0.6\xi_1 + \delta_2$	$x_2 = 0.4\xi_1 + \delta_2$	$x_2 = 0.3\xi_1 + \delta_2$
$x_3 = 0.8\xi_1 + \delta_3$	$x_3 = 0.7\xi_1 + \delta_3$	$x_3 = 0.6\xi_1 + \delta_3$	$x_3 = 0.4\xi_1 + \delta_3$	$x_3 = 0.3\xi_1 + \delta_3$
$x_4 = \xi_2 + \delta_4$	$x_4 = \xi_2 + \delta_4$	$x_4 = \xi_2 + \delta_4$	$x_4 = \xi_2 + \delta_4$	$x_4 = \xi_2 + \delta_4$
$x_5 = 0.8\xi_2 + \delta_5$	$x_5 = 0.7\xi_2 + \delta_5$	$x_5 = 0.6\xi_2 + \delta_5$	$x_5 = 0.4\xi_2 + \delta_5$	$x_5 = 0.3\xi_2 + \delta_5$
$x_6 = 0.8\xi_2 + \delta_6$	$x_6 = 0.7\xi_2 + \delta_6$	$x_6 = 0.6\xi_2 + \delta_6$	$x_6 = 0.4\xi_2 + \delta_6$	$x_6 = 0.3\xi_2 + \delta_6$

- DGPFC 1–5: $\xi_1 \sim N(0,1)$ and ξ_2 = 0,5 * ξ_1 + 0.866 $N(0,1)$. Cor(ξ_1, ξ_2) = 0.5
- DGPFC 6–10: $\xi_1 \sim N(0,1)$ and ξ_2 = 0,8 * ξ_1 + 0.6 $N(0,1)$. Cor(ξ_1, ξ_2) = 0.8
- DGPFC 11–15: $\xi_1 \sim N(0,1)$ and ξ_2 = 0,95 * ξ_1 + 0.31 $N(0,1)$. Cor(ξ_1, ξ_2) = 0.95
- Five levels of error variance:

$Var(\delta_i) = 0.3$ for i=1,...,6	$Var(\delta_i) = 0.5$ for i=1,...,6	$Var(\delta_i) = 0.64$ for i=1,...,6	$Var(\delta_i) = 0.86$ for i=1,...,6	$Var(\delta_i) = 0.91$ for i=1,...,6

Table 5: R squared for the three equations in the path model (model 1).

Level of error variance	R_1^2	R_2^2	R_3^2
Level 1-low	0.2	0.54	0.81
Level 2	0.06	0.22	0.52
Level 3	0.03	0.11	0.32
Level 4	0.02	0.07	0.21
Level 5-high	0.01	0.04	0.15

Table 6: R square for the eight equations in the factor model with uncorrelated factors (model 2).

Level Of error variance	Item reliability, R squared:							
	R_1^2	R_2^2	R_3^2	R_4^2	R_5^2	R_6^2	R_7^2	R_8^2
Level 1	0.86	0.93	0.85	0.96	0.87	0.93	0.90	0.97
Level 2	0.61	0.78	0.59	0.87	0.62	0.77	0.69	0.90
Level 3	0.41	0.61	0.39	0.75	0.42	0.60	0.49	0.79
Level 4	0.28	0.47	0.26	0.63	0.29	0.45	0.35	0.68
Level 5	0.2	0.36	0.19	0.52	0.21	0.35	0.26	0.58
Level 6	0.15	0.28	0.14	0.43	0.15	0.27	0.20	0.49

Table 7: Reliability measures of the two factor model with uncorrelated factors (model 2). Average variance extracted and composite reliability measure.

Level of error variance	Factors	Average variance Extracted	Composite Reliability
Level 1-low	ξ_1	0.88	0.97
	ξ_2	0.90	0.97
Level 2	ξ_1	0.72	0.93
	ξ_2	0.75	0.92
Level 3	ξ_1	0.62	0.89
	ξ_2	0.64	0.87
Level 4	ξ_1	0.56	0.86
	ξ_2	0.57	0.84
Level 5	ξ_1	0.53	0.85
	ξ_2	0.52	0.82
Level 6-high	ξ_1	0.50	0.83
	ξ_2	0.49	0.81

Table 8: The proportion of the 400 repetitions which rejects the mode. Chi-square test. Path model 1 and factor model 2. Different sample sizes and different levels of error variance.

LEVELS OF ERROR VARIANCES	SAMPLE SIZE 25	50	75	100	200	400	800
			CORRECTLY SPECIFIED PATHMODEL 1				
L-1 Low	15.2	8.2	6.5	3.8	5.2	9.8	3.8
L-2	14.2	10.5	6.0	6.2	5.5	6.0	5.5
L-3	8.2	10.0	3.8	6.2	5.0	7.2	6.0
L-4	12.5	8.5	3.5	7.0	7.2	3.5	4.0
L-5 High	9.0	6.0	3.0	7.8	3.0	7.0	2.8
			CORRECTLY SPECIFIED FACTOR MODEL 2				
L-1 Low	13.2	9.0	8.5	5.0	8.0	6.8	3.0
L-2	12.7	6.2	6.2	9.0	2.0	6.5	5.2
L-3	8.2	9.9	7.4	8.2	5.2	7.5	7.2
L-4	13.0	7.7	5.2	10.4	6.1	3.5	5.0
L-5	5.8	4.5	6.8	5.1	2.5	8.0	2.8
L-6 High	12.9	3.3	5.8	7.5	7.1	9.5	5.0

Table 9: The proportion of the 400 repetitions which rejects the model. Chi-square test. Factor models 3, 4, and 5. Different sample sizes and different levels of error variances.

LEVELS OF ERROR VARIANCES.	SAMPLE SIZE			
	100	200	400	800
		Model 3		
L-1 Low	100	100	100	100
L-2	100	100	100	100
L-3	0,973	100	100	100
L-4	40	78,25	98,75	100
L-5 High	**15**	40,25	77,25	99,75
		Model 4		
L-1 Low	99,0	100	100	100
L-2	69,25	95	100	100
L-3	**16,53**	24,22	40,1	71,0
L-4	**09,00**	**18,00**	39,00	69,25
L-5 High	**7,5**	**10,25**	**17,5**	34,75
		Model 5		
L-1 Low	23,75	45,25	78,25	100,00
L-2	**9,00**	**14,5**	28,25	53,75
L-3	**8,00**	**9,25**	**17,5**	26,25
L-4	**5,0**	**6,25**	**6,25**	**7,25**
L-5 High	**4,00**	**5,75**	**5,5**	**5,75**

Table 10: Mean values and standard deviations of the GFI index for the 400 replications. Path model 1 and factor model 2. Different sample sizes and different levels of error variance.

LEVELS OF ERROR VARIANCE	SAMPLE SIZE 25	50	75	100	200	400	800
	CORRECTLY SPECIFIED PATH MODEL - MEAN VALUES						
L-1 Low	0.907	0.955	0.970	0.977	0.989	0.994	0.997
L-2	0.905	0.951	0.970	0.977	0.988	0.994	0.997
L-3	0.911	0.953	0.971	0.977	0.988	0.994	0.997
L-4	0.905	0.956	0.970	0.976	0.988	0.994	0.997
L-5 High	0.914	0.954	0.970	0.977	0.989	0.994	0.997
	STANDARD DEVIATIONS						
L-1 Low	0.039	0.023	0.016	0.010	0.006	0.003	0.001
L-2	0.048	0.023	0.016	0.011	0.006	0.003	0.002
L-3	0.037	0.024	0.014	0.012	0.006	0.003	0.002
L-4	0.039	0.022	0.016	0.012	0.006	0.003	0.001
L-5 High	0.039	0.021	0.013	0.013	0.005	0.003	0.001
	APPROXIMATED EXPECTED VALUES						
L-1 E(GFI)	0.915	0.955	0.970	0.977	0.988	0.994	0.997
L-1 SD(GFI)	0.042	0.023	0.016	0.012	0.006	0.003	0.002
	CORRECTLY SPECIFIED FACTOR MODEL 2 - MEAN VALUES						
L-1 Low	0.848	0.921	0.946	0.960	0.978	0.989	0.995
L-2	0.853	0.921	0.945	0.960	0.979	0.990	0.995
L-3	0.858	0.921	0.944	0.960	0.980	0.989	0.995
L-4	0.858	0.923	0.946	0.958	0.979	0.989	0.995
L-5	0.862	0.923	0.947	0.960	0.979	0.989	0.995
L-6 High	0.857	0.923	0.949	0.959	0.979	0.989	0.995
	STANDARD DEVIATIONS						
L-1 Low	0.037	0.023	0.018	0.012	0.007	0.004	0.002
L-2	0.037	0.024	0.015	0.013	0.006	0.003	0.002
L-3	0.036	0.024	0.017	0.014	0.007	0.004	0.002
L-4	0.036	0.023	0.017	0.014	0.007	0.003	0.002
L-5	0.035	0.021	0.016	0.013	0.006	0.004	0.002
L-6 High	0.036	0.020	0.015	0.012	0.007	0.004	0.002

Table 11: Mean values and standard deviations of the NFI index for the 400 replications. Path model 1 and factor model 2. Different sample sizes and levels of error variance.

LEVELS OF ERROR VARIANCE	SAMPLE SIZE						
	25	50	75	100	200	400	800
	CORRECTLY SPECIFIED MODEL - MEAN VALUES						
L-1 Low	0.891	0.949	0.966	0.974	0.988	0.993	0.997
L-2	0.788	0.876	0.919	0.937	0.966	0.983	0.992
L-3	0.717	0.814	0.871	0.889	0.940	0.968	0.984
L-4	0.642	0.756	0.810	0.834	0.905	0.950	0.976
L-5 High	0.627	0.702	0.767	0.808	0.876	0.926	0.960
	STANDARD DEVIATIONS						
L-1	0.057	0.029	0.019	0.012	0.007	0.004	0.002
L-2	0.110	0.059	0.041	0.032	0.017	0.009	0.005
L-3	0.120	0.094	0.064	0.057	0.033	0.018	0.009
L-4	0.142	0.116	0.094	0.088	0.047	0.024	0.012
L-5	0.136	0.140	0.108	0.093	0.055	0.038	0.020
	APPROXIMATED EXPECTED VALUES						
L-1 E(NFI)	0.913	0.953	0.967	0.975	0.987	0.993	0.997
L-1 SD(NFI)	0.050	0.026	0.018	0.014	0.007	0.0035	0.0018
L-5 E(NFI)	0.669	0.729	0.786	0.815	0.879	0.929	0.963
L-5 SD(NFI)	0.211	0.171	0.132	0.112	0.071	0.041	0.021
	CORRECTLY SPECIFIED MODEL 2 – MEAN VALUES - df=17						
L-1 Low	0.937	0.971	0.981	0.986	0.993	0.997	0.998
L-2	0.882	0.941	0.960	0.972	0.986	0.993	0.996
L-3	0.819	0.897	0.926	0.950	0.975	0.987	0.993
L-4	0.751	0.846	0.889	0.910	0.956	0.977	0.988
L-5	0.710	0.782	0.837	0.873	0.931	0.962	0.981
L-6 high	0.629	0.733	0.794	0.817	0.894	0.941	0.971
	STANDARD DEVIATIONS						
L-1	0.022	0.009	0.007	0.004	0.002	0.001	0.001
L-2	0.046	0.020	0.011	0.011	0.004	0.002	0.001
L-3	0.061	0.037	0.024	0.018	0.009	0.005	0.002
L-4	0.075	0.054	0.039	0.032	0.015	0.007	0.004
L-5	0.083	0.069	0.053	0.046	0.023	0.014	0.007
L-6	0.092	0.071	0.068	0.058	0.035	0.022	0.009

Table 12: Mean values and standard deviations of the RMSEA for the 400 replications. Path model 1 and factor model 2, different sample sizes and different levels of error variance.

LEVELS OF ERROR VARIANCE	SAMPLE SIZE						
	25	50	75	100	200	400	800
	CORRECTLY SPECIFIED MODEL – MEAN VALUES						
L-1 Low	0.093	0.052	0.032	0.028	0.016	0.014	0.008
L-2	0.094	0.055	0.038	0.030	0.023	0.016	0.010
L-3	0.080	0.052	0.031	0.031	0.020	0.014	0.010
L-4	0.095	0.044	0.035	0.032	0.023	0.013	0.008
L-5 High	0.071	0.050	0.035	0.028	0.019	0.015	0.011
	STANDARD DEVIATIONS						
L-1	0.102	0.060	0.045	0.035	0.025	0.020	0.012
L-2	0.107	0.063	0.047	0.038	0.027	0.019	0.014
L-3	0.092	0.062	0.043	0.039	0.028	0.019	0.013
L-4	0.098	0.058	0.046	0.041	0.027	0.017	0.012
L-5	0.096	0.057	0.042	0.042	0.023	0.019	0.013
	APPROXIMATED EXPECTED VALUES						
E(RMSEA)	0.091	0.065	0.053	0.046	0.032	0.023	0.016
SD(RMSEAI)	0.092	0.055	0.041	0.033	0.019	0.012	0.007
	CORRECTLY SPECIFIED MODEL 2 - MEAN VALUES – df=15						
L-1 Low	0.083	0.039	0.033	0.022	0.020	0.013	0.009
L-2	0.072	0.041	0.032	0.024	0.017	0.011	0.008
L-3	0.070	0.040	0.035	0.023	0.016	0.012	0.008
L-4	0.065	0.039	0.032	0.030	0.018	0.011	0.009
L-5	0.054	0.038	0.028	0.022	0.016	0.013	0.007
L-6 High	0.077	0.037	0.025	0.025	0.018	0.013	0.008
	STANDARD DEVIATIONS						
L-1	0.074	0.045	0.039	0.030	0.023	0.015	0.010
L-2	0.072	0.046	0.035	0.031	0.019	0.015	0.010
L-3	0.067	0.048	0.037	0.031	0.020	0.015	0.011
L-4	0.070	0.043	0.035	0.032	0.021	0.014	0.010
L-5	0.062	0.044	0.034	0.030	0.019	0.016	0.010
L-6	0.070	0.041	0.033	0.029	0.022	0.016	0.010

Table 13: Mean values of the GFI index for the 400 replications. Factor models 3, 4 and 5. Different sample sizes and different levels of error variance.

LEVELS OF ERROR VARIANCE	SAMPLE SIZE 100	200	400	800
		Model 3		
L-1 Low	0.700	0.693	0.689	0.688
L-2	0.820	0.820	0.820	0.816
L-3	0.875	0.879	0.880	0.880
L-4	0.950	0.960	0.965	0.968
L-5 High	0.963	0.974	0.980	0.983
		Model 4		
L-1 Low	0.841	0.846	0.849	0.851
L-2	0.935	0.948	0.954	0.958
L-3	0.946	0.958	0.964	0.968
L-4	0.967	0.980	0.987	0.990
L-5 High	0.970	0.983	0.990	0.994
		Model 5		
L-1 Low	0.957	0.971	0.978	0.981
L-2	0.967	0.981	0.988	0.992
L-3	0.969	0.984	0.991	0.994
L-4	0.971	0.985	0.992	0.996
L-5 High	0.971	0.986	0.993	0.996

Table 14: Mean values of the NFI index for the 400 replications. Factor models 3, 4 and 5. Different sample sizes and different levels of error variance.

LEVELS OF ERROR VARIANCE	SAMPLE SIZE 100	200	400	800
	Model 3			
L-1 Low	0.710	0.716	0.717	0.718
L-2	0.767	0.778	0.784	0.787
L-3	0.781	0.797	0.806	0.811
L-4	0.760	0.804	0.827	0.842
L-5 High	0.730	0.792	0.827	0.851
	Model 4			
L-1 Low	0.918	0.926	0.930	0.932
L-2	0.935	0.948	0.954	0.958
L-3	0.932	0.951	0.960	0.964
L-4	0.884	0.931	0.955	0.967
L-5 High	0.843	0.909	0.944	0.964
	Model 5			
L-1 Low	0.981	0.988	0.990	0.993
L-2	0.975	0.987	0.992	0.995
L-3	0.967	0.983	0.990	0.994
L-4	0.916	0.958	0.978	0.988
L-5 High	0.877	0.936	0.966	0.982

Table 15: Mean values of the RFI index for the 400 replications. Factor models 3, 4 and 5. Different sample sizes and different levels of error variance.

LEVELS OF ERROR VARIANCE	SAMPLE SIZE 100	200	400	800
		Model 3		
L-1 Low	0.517	0.526	0.529	0.530
L-2	0.612	0.631	0.639	0.644
L-3	0.635	0.662	0.676	0.684
L-4	0.600	0.673	0.711	0.736
L-5 High	0.550	0.654	0.712	0.752
		Model 4		
L-1 Low	0.864	0.877	0.884	0.887
L-2	0.891	0.913	0.923	0.929
L-3	0.887	0.918	0.932	0.940
L-4	0.807	0.886	0.925	0.946
L-5 High	0.739	0.849	0.907	0.940
		Model 5		
L-1 Low	0.968	0.980	0.985	0.988
L-2	0.959	0.978	0.986	0.991
L-3	0.945	0.972	0.984	0.990
L-4	0.860	0.929	0.963	0.980
L-5 High	0.795	0.893	0.944	0.970

Table 16: Mean values of the CFI index for the 400 replications. Factor models 3, 4 and 5. Different sample sizes and different levels of error variance.

LEVELS OF ERROR VARIANCE	SAMPLE SIZE 100	200	400	800
		Model 3		
L-1 Low	0.722	0.722	0.720	0.720
L-2	0.791	0.790	0.790	0.790
L-3	0.816	0.815	0.815	0.815
L-4	0.864	0.861	0.856	0.856
L-5 High	0.879	0.886	0.879	0.878
		Model 4		
L-1 Low	0.932	0.933	0.934	0.934
L-2	0.958	0.960	0.960	0.960
L-3	0.966	0.968	0.968	0.969
L-4	0.966	0.976	0.979	0.980
L-5 High	0.960	0.975	0.981	0.984
		Model 5		
L-1 Low	0.993	0.994	0.994	0.994
L-2	0.994	0.996	0.997	0.997
L-3	0.992	0.996	0.997	0.998
L-4	0.982	0.991	0.995	0.997
L-5 High	0.973	0.987	0.993	0.997

Table 17: Mean values of the RMSEA index for the 400 replications. Factor models 3, 4 and 5. Different sample sizes and different levels of error variance.

LEVELS OF ERROR VARIANCE	SAMPLE SIZE 100	200	400	800
			Model 3	
L-1 Low	0.368	0.378	0.384	0.388
L-2	0.250	0.261	0.267	0.271
L-3	0.190	0.201	0.207	0.210
L-4	0.077	0.089	0.096	0.099
L-5 High	0.047	0.055	0.063	0.067
			Model 4	
L-1 Low	0.227	0.235	0.238	0.239
L-2	0.123	0.131	0.135	0.136
L-3	0.084	0.092	0.097	0.098
L-4	0.035	0.035	0.038	0.041
L-5 High	0.027	0.025	0.025	0.026
			Model 5	
L-1 Low	0.059	0.063	0.068	0.070
L-2	0.034	0.031	0.032	0.034
L-3	0.029	0.024	0.023	0.023
L-4	0.025	0.018	0.014	0.011
L-5 High	0.024	0.018	0.013	0.009

Table 18: Mean values of the SRMR index for the 400 replications. Factor models 3, 4 and 5. Different sample sizes and different levels of error variance.

LEVELS OF ERROR VARIANCE	SAMPLE SIZE 100	200	400	800
		Model 3		
L-1 Low	0.172	0.171	0.170	0.170
L-2	0.126	0.124	0.121	0.120
L-3	0.106	0.102	0.100	0.098
L-4	0.072	0.063	0.058	0.055
L-5 High	0.065	0.053	0.046	0.042
		Model 4		
L-1 Low	0.065	0.062	0.060	0.059
L-2	0.056	0.051	0.048	0.047
L-3	0.054	0.046	0.042	0.040
L-4	0.054	0.040	0.032	0.027
L-5 High	0.055	0.040	0.030	0.024
		Model 5		
L-1 Low	0.025	0.020	0.017	0.016
L-2	0.032	0.024	0.019	0.015
L-3	0.037	0.027	0.020	0.016
L-4	0.049	0.034	0.025	0.018
L-5 High	0.053	0.037	0.026	0.019