Biases of parameter estimates in misspecified structural equation models

Stanislav Kolenikov*

December 12, 2009

Acknowledgements

Support from NSF grant SES-0617193 with funds from Social Security Administration is gratefully acknowledged. The author would like to thank Ken Bollen, Oksana Loginova, Cam McIntosh for helpful comments, and Ni Li for research assistance on the project.

*Department of Statistics, University of Missouri, Columbia, MO. Corresponding author: send email to kolenikovs@missouri.edu.
Biases of parameter estimates in misspecified structural equation models

Abstract

We propose a framework to handle lack of fit problems in structural equation models by analyzing misspecified moments. This framework allows to relate misspecification of the moment equations to biases of parameter estimates and non-centrality of the model. Misspecification of a single moment is shown to lead to misfit of all moments as well as biases in most parameter estimates in full information estimation procedures such as FIML and WLS, unless such misspecification is absorbed by a corresponding error variance or error covariance parameter. We provide illustrative examples to quantify biases of parameter estimates and non-centralities of the model due to misfitting moments.

Keywords: Holzinger-Swineford data, model structure misspecification, structural equation models, structural misspecification, parameter estimate bias.

1 Introduction

Structural equation models (SEMs) are a common multivariate statistical technique used in a variety of social science disciplines to analyze the relations between explanatory (exogenous) variables and dependent (endogenous) variables (Bollen 1989, MacCallum & Austin 2000, Yuan & Bentler 2007). SEMs are especially useful in situations when multiple endogenous variables and/or latent variables (Bollen 2002) are present in the model. A number of assumptions, implicit rather than explicit ones, need to be made for the machinery of SEMs to work. These assumptions can be broadly classified into distributional and structural. Distributional assumptions describe the features of the data such as multivariate normality or lack of excess kurtosis in the data. Structural assumptions relate to the issues such as what variables are to be used in the model (including the latent ones), which variables have regression relations between themselves, and linearity of such relations.\footnote{While the term “structural model” is often used to describe the relations between latent variables, we use it to refer to the model as a whole, i.e., both the latent variable part and the measurement part.}

In reality, these assumptions may not be tenable. Extremely rare are the cases when they hold in their entirety. Any model is only an approximation to reality: “Essentially, all models are wrong, but some are useful” (Box & Draper 1987). Applied researchers should be prepared to ask themselves: If my model is not perfectly accurate, what are the consequences? Or, more formally: What are the effects of misspecification on the estimation results?

Distributional misspecifications have been widely addressed in the literature on asymptotic robustness (Anderson & Amemiya 1988, Satorra 1990, Satorra 1992). The effects of distributional misspecification include increase in the parameter estimate variances and inflation of the overall chi-square and difference chi-square statistics. However, under certain assumptions, the standard estimation procedures such as maximum likelihood continue to produce estimates with good properties (consistency, asymptotic efficiency, asymptotic normality) and easily available variance estimates (inverse of the information matrix) for most parameters (except variances and covariances of non-normal variables). Also, goodness of fit tests continue to have familiar \( \chi^2 \) distributions. Under less favorable conditions, the estimates, although no longer efficient,
continue to be asymptotically normal with estimable variances, and goodness of fit tests can be corrected to account for non-normality (Satorra & Bentler 1994).

Different effects of structural misspecifications have been studied to different extents. A well-known (and desirable) consequence of structural misspecifications is an increase in overall fit test statistic. Because of this, the overall fit test has a power to detect structural misspecifications. The overall fit test is known to have differential power for different types of misspecifications. Saris, Satorra & Sörbom (1987) and Kaplan (1988) demonstrate that it tends to be more sensitive to misplaced indicators of latent variables (i.e., when a wrong latent variable is loaded on an observed variable), and less sensitive to small problems in the measurement model such as correlated measurement errors. As an extreme example of lack of sensitivity to structural misspecifications, Mooijaart & Satorra (2009) show that the overall chi-square test has zero power in detecting certain nonlinearities in the model.

The effect of misspecification on parameter biases have been studied in passing. Typically, structural misspecification studies (Saris et al. 1987, Kaplan 1988, Bollen, Kirby, Curran, Paxton & Chen 2007) assume an encompassing model from which certain paths have been omitted, or restrictions imposed on the parameter values. A general treatment of this situation is given by Yuan, Marshall & Bentler (2003). They introduce the concept of parameter orthogonality and derive conditions under which a parameterizable structural misspecification does not affect estimation of other parameters.

In this paper, we provide a model-free approach to tracking the effects of misspecification on the parameter estimates and non-centrality. Hence the information that we operate with is exactly what is available to a researcher upon estimation of a structural equation model: parameter estimates, derivatives of the fit function and the matrices of sample, implied and residual moments. We take the residual moments as they are, without trying to build a better fitting model. Our procedure consists of three steps. First, we compute the matrices describing the transfer of the sample moments to the parameter estimates and residual moments. Second, we identify the entries of the residual matrix that are incompatible with the model. Third, these two pieces of information are combined to indicate which parameters are potentially biased, and to infer the direction and approximate magnitude of biases. In the end, we can make statements like “Provided that this error variance is still interpretable in the current model, it is underestimated, and hence the reliability of this indicator is overstated” or “Bias introduced to this parameter because of misspecification is negligible”.

2 Moment misfit, parameter biases and non-centrality

In this section, we provide a formal framework to relate variations in the input covariance matrix to variations in the parameter estimates, implied matrix, and overall fit statistic.

Neudecker & Satorra (1991) utilize matrix calculus (Magnus & Neudecker 1999) to obtain the first two derivatives of the fitting function

\[ F(\theta) = (s - \sigma(\theta))'W(s - \sigma(\theta)) \]

of the structural equation model

\[ z = \Lambda \eta + \epsilon, \]  
\[ \eta = B_0 \eta + \xi. \]
Here, $z$ is a $p$-vector of observed variables, $\eta$ is an $m$-vector of latent variables, $\epsilon$ and $\xi$ are $p \times 1$ and $m \times 1$ vectors of error terms with covariance matrices $\Psi$ and $\Phi$ ($\text{Cov}[\epsilon, \xi] = 0$), $\Lambda$ is a $p \times m$ matrix of loadings in the measurement model, $B_0$ is an $m \times m$ matrix of coefficients of regression of the latent variables on themselves, $W$ is the weight matrix, and $s$ and $\sigma(\theta)$ are non-redundant vectorizations of the sample covariance matrix $S$ and the implied covariance matrix

$$\Sigma(\theta) = \Lambda B^{-1} \Phi B^{-T} A' + \Psi.$$  \hfill (4)

Exogenous observed variables can be incorporated into $\eta$ vector with zero error variance, if needed. The matrix $B = I_p - B_0$ is assumed to be invertible; otherwise, the model would be underidentified.

Equation (1) gives the weighted least squares criterion. Estimates that are asymptotically equivalent to MLE can be obtained with the normal theory weight matrix

$$W_{NT} = \frac{1}{2} D_p' (\Sigma^{-1} \otimes \Sigma^{-1}) D_p$$  \hfill (5)

where $D_p$ is the $p^* \times p^2$ duplication matrix (Magnus & Neudecker 1999), where $p^* = p(p + 1)/2$ is the number of non-redundant elements of the covariance matrix of size $p$.

Neudecker & Satorra (1991) define the complete $t^* \times 1$ vector of all parameters in the model

$$\delta = [(\text{vec } \Lambda)', (\text{vech } \Phi)', (\text{vec } B)', (\text{vech } \Psi)']'$$  \hfill (6)

and obtain the $t \times 1$ vector of relevant parameters $\theta$ by a mapping $\delta = \delta(\theta)$ with the $t^* \times t$ selection matrix

$$A = \frac{\partial \delta}{\partial \theta}.$$  \hfill (7)

This is usually a constant matrix of zeroes and ones (unless the model is parameterized with correlations instead of covariances).

The estimating equations of the model (1)–(3) are

$$\frac{\partial F}{\partial \delta'} = -2(s - \sigma)' W G = 0,$$  \hfill (8)

where the matrix derivative of the implied covariance matrix

$$d \sigma \equiv d \text{vech } \Sigma = G \; d \delta$$  \hfill (9)

is given in the Appendix.

Let us take the full differential of (8) multiplied by $-1/2$: 

$$0 = d[(s - \sigma)' W G] = d s' W G - d \sigma' W G + (s - \sigma)' W \; d G$$  \hfill (10)

The second term is equal to $d \delta' G' W G$. Neudecker & Satorra (1991) express the last term as

$$(s - \sigma)' W \; d G = d \delta' \sum_{l=1}^{p^*} [(s - \sigma)' W]_l H_l,$$
where $H_l$ is given by their formula (4) (reproduced in the Appendix), and index $l$ enumerates the moments $(i, j), 1 \leq i \leq j \leq p$.

Hence,

$$0 = d's'WG - d'\delta'G'WG + d'\delta'\sum_{l=1}^{p^*}[(s - \sigma)'W]_lH_l$$

from which it follows that

$$\left\{G'WG - \sum_{l=1}^{p^*}[(s - \sigma)'W]_lH_l\right\}d\delta = G'Wd's.$$

This expression relates the variations in the input sample covariance matrix $S$ (and its vectorization $s$) into variations in the complete parameter vector $\delta$. To obtain the variations in the modeled parameters $\theta$, one needs to replace $d\delta$ with $A d\theta$, so

$$d\theta = \left(A'\left\{G'WG - \sum_{l=1}^{p^*}[(s - \sigma)'W]_lH_l\right\}A\right)^{-1}A'G'Wd's \equiv Q d's.$$

The matrix in the parentheses is $-1/2$ times the Hessian of the model and is invertible if the model is locally identified. Matrix $Q$ shows how variations in $S$ (and $s$) translate into variations in $\theta$. We shall call it a transfer matrix.

The differentials in (12) are interpreted as deviations of the relevant quantities from those in the population for the correctly specified model. Depending on whether the actual model is correctly specified or not, equation (12) has multiple meanings. In finite samples and for properly specified models, $d's$ represents the sampling fluctuations of order $O_p(n^{-1/2})$. These sampling fluctuations in the input covariance matrix lead to sampling fluctuations in the parameter estimates $\hat{\theta}$, so the transfer matrix $Q$ is the derivative of the transformation used in the delta method of computing the standard errors (Browne 1984).

For misspecified models, some entries of $d's$ may be of order $O_p(1)$, i.e., do not disappear asymptotically. If this is the case, the transfer matrix $Q$ also shows how non-zero entries of $\mathbb{E}[s - \sigma]$ map to biased estimates of $\theta$.

If misspecifications are small, so that the term $s - \sigma$ in (12) can be ignored, all or some parameters of the model may be affected depending on whether the element $q_{k,ij} = \partial \theta_k / \partial s_{ij}$ that maps $i, j$-th moment to $k$-th parameter is zero or not. For larger misspecifications, however, all parameters will be affected. Indeed, the misfitting moments $s - \sigma$ enter the matrix $Q$ through a matrix inversion operation. A generic element of the inverse matrix

$$[C^{-1}]_{ij} = (-1)^{i+j}|C_{ji}|/|C|,$$

where $|C_{ji}|$ is the cofactor of the element $c_{ji}$, contains $s - \sigma$ for all elements through the determinant in the denominator. Thus, $s - \sigma$ will be present in all elements of $Q$, and hence larger misspecifications will affect all parameters.
Let us introduce the residual matrix \( R = S - \Sigma(\theta) \) with vectorization \( r = \text{vech} \ R = s - \sigma(\theta) \). Then
\[
\begin{align*}
\text{d} \ r &= \text{d} \ s - \text{d} \sigma = \text{d} \ s - G \text{d} \delta = \text{d} \ s - GA \text{d} \theta \\
&= \left[ I_p r^* - GA \left( A' \left\{ G' W G - \sum_{i=1}^{p^*} \left[ (s - \sigma)' W \right] H_i \right\} A \right)^{-1} A' G' W \right] \text{d} \ s \\
&= [I_p^* - GAQ] \, d \ s \equiv \tilde{H} \, d \ s.
\end{align*}
\]

(13)

This expression shows how variations in the input covariance matrix are translated into variations of the residual moments \( r_{ij} = s_{ij} - \sigma_{ij}(\theta), 1 \leq i \leq j \leq p \). For any given sample moment, the corresponding residual moment is dampened by adjustment of parameter estimates. This adjustment is described by the second term in the parentheses. The multiplier in front of \( \text{d} \ s \) can be thought as a nonlinear model analogue of the hat matrix \( H = I - X(X'X)^{-1}X' \) in linear regression models, which shows how the regression errors affect the regression residuals.

Non-centrality of the model can also be decomposed into misfit of individual moments using its differential:
\[
\lambda = F(\theta) = r' W r, \quad \text{d} \lambda = \text{d} F = 2(\text{d} \ r)' W r = 2(\text{d} \ s)' \tilde{H}' W \tilde{H} \, \text{d} \ s.
\]

(14)

The effect of misspecification of the \( i, j \)-th moment is then
\[
\lambda_{ij} = 2r_{ij} e_{ij} W r.
\]

(15)

Here, \( e_{ij} \) is the \( p^* \times 1 \) unit vector corresponding to the \( i, j \)-th moment: \( e_{ij} = \text{vech}(u_i u_j' + u_j u_i') \) where the \( p \times 1 \) unit vector \( u_i \) has 1 in \( i \)-th position and zeroes elsewhere. For a misspecified model, \( \lambda_{ij} \) is the contribution to non-centrality due to the \( i, j \)-th moment misspecification.

Let us now provide the reverse link and relate the observed non-centrality to a plausible misspecification size, and hence to possible biases of parameter estimates. The bias of the \( k \)-th parameter of the model is approximately
\[
\text{d} \ theta_k = Q_k \, \text{d} \ s, \quad \left| \text{d} \ theta_k \right| \leq \|Q_k\| \| \text{d} \ s \|
\]

(16)

where \( Q_k \) is the \( k \)-th row of the transfer matrix \( Q \), and \( \|AB\| \leq \|A\| \|B\| \) by Schwarz inequality.

To relate \( \| \text{d} \ s \| \) to \( \text{d} \lambda \), a couple of different approaches can be used. From (14),
\[
\text{d} \lambda \leq 2 \| \text{d} \ s \| \| H' W r \|.
\]

and a “typical” size of misspecification is
\[
\| \text{d} \ s \| \sim \frac{\text{d} \lambda}{2 \| H' W r \|}.
\]

(17)

Two approximations are made here. First, the non-centrality is assumed to be approximated by its first differential. Second, the vectors \( \text{d} \ s \) and \( H' W r \) are assumed to be parallel, which can only happen if \( \text{d} \ s \) is an eigenvector of \( H' W H \).

Alternatively, using (13),
\[
\lambda \sim s' \tilde{H}' W \tilde{H} \, s \sim \|s\|^2 \| \tilde{H}' W \tilde{H} \|,
\]

6
again making an extremely strong assumption that \( s \) is the largest eigenvector of the matrix \( \tilde{H}'W\tilde{H} \) to obtain the order of magnitude comparisons. From this,

\[
\| ds \| \sim \sqrt{\frac{d \lambda}{\|H'W\|}}.
\] (18)

We can now combine the two approximations and define maximum plausible bias as

\[
\text{MPB}[\theta_k] = \|Q_k\| \max \left\{ \frac{\lambda}{2\|W\|}, \sqrt{\frac{d \lambda}{\|\tilde{H}'W\tilde{H}\|}} \right\}.
\] (19)

This expression gives a generous upper bound on the magnitude of bias of parameter estimates caused by misspecification.

3 Effect of misspecification on parameter estimates

As discussed in the previous section, variations in the input sample covariance matrix, transferred into variations of parameter estimates, are associated with two sources. First, sampling fluctuations inevitably cause the observed sample covariance matrix to differ from the population covariance matrix. In this case, we would expect the variations in both moments and parameter estimates to have zero means and finite variances (provided that the fourth moments of the data are finite). Second, the model assumed by the researcher may be incorrect, in which case the means of the sampling fluctuations are not zero. As we said in the introduction, it is to difficult to believe that all models are absolutely correct, even if they pass the chi-square goodness of fit test and have good fit indices. Thus, we need to know how misspecifications affect parameter estimates.

In this section, we give an example demonstrating how the results of the previous section can be applied to the analysis of biases and non-centralities of a SEM. We compute the necessary matrices based on a working model, including the transfer matrix \( Q \) and the hat matrix \( \tilde{H} \). We then analyze the residual matrix \( R \) to identify the main source(s) of misfit.

3.1 Correct and misspecified models

Suppose a researcher studying socio-economic status (SES) builds a confirmatory factor analysis model to describe the phenomenon. In reality, some indicators are causal (formative) rather than effect (reflective) indicators, so the model is misspecified. For instance, education and occupational prestige may cause SES rather than be determined by it. A graphical representation of the two models is given on Fig. 1. The left panel is the correct model, and the right panel is the working model. Even though the latter is misspecified, it still describes the same conceptual latent variable of SES, and the remaining three indicators may be valid effect indicators of it (e.g., dwelling, health, or outcomes in children). Hence we can still interpret parameters \( \lambda_3, \lambda_4, \psi_3, \psi_4 \) and \( \psi_5 \), and discuss their biases.

Parameter values for the two models are given in Table 1. The first column reports the parameters of the true Model 1, and the second column reports the population parameters of the incorrect Model 2 fitted by maximum likelihood. Other estimation methods can produce slightly different results (Yuan & Chan 2005).
Table 1: Parameter values of the two models in Fig. 1.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Exact bias</th>
<th>Bias from (12) using ( r_{1,2} )</th>
<th>Bias from (12) using ( v_{1,2} )</th>
<th>MPB</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \phi_{11} )</td>
<td>1.00</td>
<td>3.1410</td>
<td>–</td>
<td>0.0440</td>
<td>0.0538</td>
<td>0.2259</td>
</tr>
<tr>
<td>( \phi_{22} )</td>
<td>1.00</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>( \phi_{12} )</td>
<td>0.55</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>( \beta_{31} )</td>
<td>1.00</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>( \beta_{32} )</td>
<td>1.00</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>( \phi_{33} )</td>
<td>0.50</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>( \lambda_3 )</td>
<td>0.60</td>
<td>0.5975</td>
<td>–0.0025</td>
<td>–0.0020</td>
<td>–0.0024</td>
<td>0.0390</td>
</tr>
<tr>
<td>( \lambda_4 )</td>
<td>0.80</td>
<td>0.7973</td>
<td>–0.0027</td>
<td>–0.0021</td>
<td>–0.0026</td>
<td>0.0525</td>
</tr>
<tr>
<td>( \lambda_5 )</td>
<td>1 (fixed)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>( \psi_{33} )</td>
<td>0.70</td>
<td>0.6947</td>
<td>–0.0053</td>
<td>–0.0042</td>
<td>–0.0052</td>
<td>0.1369</td>
</tr>
<tr>
<td>( \psi_{44} )</td>
<td>0.70</td>
<td>0.6873</td>
<td>–0.0127</td>
<td>–0.0103</td>
<td>–0.0126</td>
<td>0.1777</td>
</tr>
<tr>
<td>( \psi_{55} )</td>
<td>0.70</td>
<td>0.6589</td>
<td>–0.0411</td>
<td>–0.0360</td>
<td>–0.0440</td>
<td>0.2438</td>
</tr>
<tr>
<td>( \lambda_1 )</td>
<td>–</td>
<td>0.4667</td>
<td>–</td>
<td>–0.0120</td>
<td>–0.0147</td>
<td>0.0313</td>
</tr>
<tr>
<td>( \lambda_2 )</td>
<td>–</td>
<td>0.4394</td>
<td>–</td>
<td>–0.0135</td>
<td>–0.0165</td>
<td>0.0305</td>
</tr>
<tr>
<td>( \psi_{11} )</td>
<td>–</td>
<td>0.3158</td>
<td>–</td>
<td>0.0194</td>
<td>0.0238</td>
<td>0.1200</td>
</tr>
<tr>
<td>( \psi_{22} )</td>
<td>–</td>
<td>0.3934</td>
<td>–</td>
<td>0.0221</td>
<td>0.0271</td>
<td>0.1178</td>
</tr>
<tr>
<td>Non-centrality</td>
<td>0</td>
<td>0.04358</td>
<td>–</td>
<td>0.02158</td>
<td>0.03231</td>
<td>–</td>
</tr>
</tbody>
</table>
The models come quite close, with the non-centrality of Model 2 equal to \( \lambda = 0.04358 \) per observation. The power to detect misfit is 23% for the sample of size \( N = 100 \), 46% for the sample of size \( N = 200 \), and more than 90% for a sample of size more than \( N \geq 500 \). However, applied researchers can attribute a significant overall chi-square to non-normality if the causal indicators \( x_1, x_2 \) have non-normal distributions.

The matrices of implied moments are as follows:

\[
\Sigma(\text{Model 1}) = \begin{pmatrix}
1 & 0.55 & 0.864 & 1.152 & 1.44 \\
0.55 & 1 & 0.81 & 1.08 & 1.35 \\
0.864 & 0.81 & 1.7872 & 1.4496 & 1.812 \\
1.152 & 1.08 & 1.4496 & 2.6328 & 2.416 \\
1.44 & 1.35 & 1.812 & 2.416 & 3.72
\end{pmatrix},
\]

\[
\Sigma(\text{Model 2}) = \begin{pmatrix}
1 & 0.6134 & 0.8485 & 1.1324 & 1.4220 \\
0.6134 & 1 & 0.7902 & 1.0546 & 1.3243 \\
0.8485 & 0.7902 & 1.7872 & 1.4588 & 1.8317 \\
1.1324 & 1.0546 & 1.4588 & 2.6328 & 2.4448 \\
1.4220 & 1.3243 & 1.8317 & 2.4448 & 3.72
\end{pmatrix},
\]

\[
R = \begin{pmatrix}
0 & -0.0634 & 0.0155 & 0.0196 & 0.0180 \\
-0.0634 & 0 & 0.0198 & 0.0254 & 0.0257 \\
0.0155 & 0.0198 & 0 & -0.0092 & -0.0197 \\
0.0196 & 0.0254 & -0.0092 & 0 & -0.0288 \\
0.0180 & 0.0257 & -0.0197 & -0.0288 & 0
\end{pmatrix}.
\]

Not surprisingly, the worst fitted element of the residual matrix is \((1, 2)\).

### 3.2 Biases of parameter estimates

In the actual data analysis, the researcher may not know that the true model is Model 1, but observes lack of fit as evidenced by both significant chi-square and non-zero residual matrix. If the researcher decides to use the model anyway (e.g., by concluding that significant chi-square is an artefact of a large sample or non-normality), some comments can still be made about biases of meaningful parameter estimates. To calculate an approximation to these biases, the researcher needs to find matrices \( Q \) and \( \tilde{H} \) derived in Section 2. The analytical expressions for them (reproduced from Neudecker & Satorra (1991)), as well as the numeric values for the current example, are given in the Appendix.

Let us analyze the biases due to the primary source of misfit, the element \( r_{1,2} = -0.0634 \). The covariance between \( x_1 \) and \( x_2 \) is the second element in vectorizations \( s, \sigma \) and \( r \). Mapping of this moment variation to parameter estimates is provided by the second column of matrix \( Q \):

\[
(0.1899, 0.2135, 0.0311, 0.0335, -0.6933, -0.3062, -0.3487, 0.0668, 0.1630, 0.5671)',
\]

where the entries correspond to the four freely estimating loadings \( \lambda_1 - \lambda_4 \), the variance of the latent variable, and the variances of the unique error terms. Multiplying the second column of \( Q \) by \( r_{1,2} = -0.0634 \), we obtain the first order approximation for the parameter estimates biases. The numeric values reported in the
third column of Table 1 show good agreement with the exact results (column 2 of Table 1): the estimated bias is about 80% of the exact bias, for most parameters.

Maximum plausible biases are reported in the last two columns of Table 1. In this particular case, the second part of the maxima in (19) are greater than the first part. The MPBs are overly conservative for the parameters that are interpretable under the working model: the numeric values are too large by a factor of at least 5. MPBs correctly indicate $\psi_5$ to suffer the greatest bias.

As a way to characterize robustness of the estimates to misspecification, Yuan et al. (2003) introduced the concept of parameter orthogonality. This idea can be extended to our approach with moments-based diagnostics. If the derivative $\partial \theta_k / \partial s_{ij}$ is zero, then the model parameter $\theta_k$ can be said to be orthogonal to the sample moment $s_{ij}$. Of course this derivative is an element of the $Q$ matrix: row $k$ corresponds to the parameter $\theta_k$, and column $ij$ corresponds to the sample moment $s_{ij}$. If this element is zero, the parameter $\theta_k$ is not sensitive to either sampling fluctuations or misfit of the covariance entry $s_{ij}$. In this example, orthogonality is present between the diagonal entries of $S$ and all parameters except the relevant error variance. The off-diagonal entries of $S$ exert influence on all parameter estimates.

3.3 Analysis of the residual matrix

Recall that the observed residual moments have misspecification propagated through them via the action of the non-linear hat matrix $\hat{H}$ given by (13). Several observations can be made on this matrix. First, $\hat{H}$ has zero columns corresponding to the diagonal entries of the covariance matrix. Indeed, any changes in the diagonal entries are fully absorbed by the corresponding estimates of $\psi_j$, and the residual moment is identically zero. If there are correlated measurement errors in the model, $\hat{H}$ will have zero columns corresponding to the parameters they represent. Second, since the non-zero columns have all elements different from zero, the variations in the sample moments propagate through the whole model. Variations in any of the off-diagonal entries of the covariance matrix will be reflected in every implied moment, and hence in every residual matrix entry. As a result, misspecification of a single off-diagonal moment will result in non-zero values of all residual moments. Third, the diagonal elements of $\hat{H}$ are greater than off-diagonal elements for the variables with freely estimated loadings, and less or comparable to other elements for the variable with the fixed loading $\lambda_5 = 1$. This means that misspecifications due to the off-diagonal moments of the scaling variable tend to appear somewhere else in the model, while misspecifications related to the variables with freely estimated loadings will likely stay where they are, although somewhat dampened. This may also explain a larger bias of $\psi_5$ compared to biases of $\psi_3$ and $\psi_4$: since the loading of $x_5$ cannot change, misspecifications have to be reflected in the error variance estimate.

Since we observe that the largest misspecification occurs in the off-diagonal element $r_{12}$ that does not involve the scaling variable, it can be hypothesized that misfit in this covariance matrix entry is the main source of the overall misfit in the model, and other residuals are actually a mere reflection of that single misfitted value. Unfortunately, we cannot solve for the vector of the moment errors using the vector $r$ of the matrix residuals, since the matrix $\hat{H}$ is not invertible (and neither is the matrix obtained from it by selecting the non-zero columns and corresponding rows). We can, however, obtain an approximate solution $v$ by

\[^2\text{We assume estimation procedures to be unconstrained, and Heywood cases to be allowed. For discussion of constrained and unconstrained estimation, see Savalei \\& Kolenikov (2008).}\]
assuming that it has the only non-zero entry corresponding to the (1,2)-th covariance. The least squares approximation is obtained by minimizing the norm $\|Hv - r\|$, and is equal to $v = (0, -0.0776, 0, \ldots, 0)'$.

This value is used to calibrate biases in the third column of Table 1. It shows almost perfect agreement with the biases actually observed.

### 3.4 Model non-centrality

If misspecification were reflected in $r_{12}$ only, its contribution to the overall non-centrality could be computed as $r_{12}^2 W_{12,12} = 0.02158$. This is approximately half of the overall non-centrality of Model 2. However, this calculation does not account for propagation of misspecification through the system. A more appropriate estimator might be (15) that gives $2r_{1,2}e_{1,2}W r = 0.05496$. Another way to account for the misspecification propagation is to use the estimate of the error vector obtained in the end of the previous section, yielding $v_{12}^2 W_{12,12} = 0.03231$. Both estimates are in a better agreement with the observed non-centrality $\lambda = 0.04358$.

### 3.5 Accuracy of the first order approximation

How accurate is the first order approximation discussed above? Let us vary the main source of misfit, the covariance of the two external variables $\phi_{12}$. The plots of the corresponding population parameters of Model 2 are given on Fig. 2. The target values are shown by the horizontal dotted lines. If $\phi_{12} = 0.7352$, the non-centrality of Model 2 is zero, the two models are chi-square-equivalent, and the estimates are asymptotically unbiased. While linearity appears to be a reasonable assumption for the loadings in a wide range of misspecification sizes, the error variance parameters are more heavily affected by higher order terms, as evidenced by greater curvature and flat areas when $\phi_{12} < 0.6$. Biases of $\psi_3$ and $\psi_4$ achieve minima near $\phi_{12} = 0.48$. The estimates reported in Table 1 are obtained for the selected value of $\phi_{12} = 0.55$, somewhere between these minima and the point of zero bias. The linear approximations produce reasonable, but not exact results.

### 4 Empirical example

In this section, we will analyze the benchmark data set for confirmatory factor analysis, the Holzinger-Swineford data (Jöreskog 1969, Yuan et al. 2003, Yuan & Bentler 2007). The data set consists of 301 observations on nine test variables (visual perception, cubes, lozenges, paragraph comprehension, sentence completion, word meaning, addition, counting dots, and straight-curved capitals) that are measuring three latent factors (spatial, verbal and speed factors, each loading on the three consecutive variables). The scale of the model is identified by setting the factor variances to one. Maximum likelihood parameter estimates
Figure 2: Parameter biases as functions of $\phi_{12}$. 
are reported in Table 2 and the residual matrix is

\[
R = \begin{pmatrix}
0.0000 & -0.0408 & -0.0104 & 0.0970 & -0.0138 & 0.0770 & -0.1778 & -0.0451 & 0.1747 \\
-0.0408 & 0.0000 & 0.1246 & -0.0171 & -0.0406 & 0.0384 & -0.2421 & -0.0615 & 0.0874 \\
-0.0104 & 0.1246 & 0.0000 & -0.0899 & -0.2199 & -0.0318 & -0.1030 & -0.0130 & 0.1671 \\
0.0970 & -0.0171 & -0.0899 & 0.0000 & 0.0075 & -0.0117 & 0.0486 & -0.0798 & 0.0556 \\
-0.0138 & -0.0406 & -0.2199 & 0.0075 & 0.0000 & 0.0048 & -0.0478 & -0.0479 & 0.0858 \\
0.0770 & 0.0384 & -0.0318 & -0.0117 & 0.0048 & 0.0000 & -0.0155 & -0.0246 & 0.0614 \\
-0.1778 & -0.2421 & -0.1030 & 0.0486 & -0.0478 & -0.0155 & 0.0550 & 0.0821 & -0.0415 \\
-0.0451 & -0.0615 & -0.0130 & -0.0798 & -0.0479 & -0.0246 & 0.0821 & 0.0000 & -0.0324 \\
0.1747 & 0.0874 & 0.1671 & 0.0556 & 0.0858 & 0.0614 & -0.0415 & -0.0324 & 0.0000
\end{pmatrix}
\]

A number of residual moments are quite sizeable, and two of them exceed 0.2 in absolute value. To purge the possible effects of multivariate non-normality on model non-centrality, Satorra & Bentler (1994) adjusted statistic can be used: \( T_{adj} = 72.812 \) with \( d = 21.294 \) degrees of freedom. The estimate of non-centrality per observation is \( \hat{\lambda} = (T_{adj} - \hat{d})/N = 0.1712 \).

Following the procedure of the previous section, we computed the derivative matrices, the normal weight matrix, the transfer matrix \( Q = \partial \theta / \partial s' \) and the hat matrix \( H \).

To gauge which moments were misfitted, we computed the studentized moment residuals

\[
t_{ij} = \frac{r_{ij}}{\sqrt{(\kappa_{ijij} - s_{ij}^2)/n}}, \quad \kappa_{ijij} = \frac{1}{n-1} \sum_{k=1}^{n} (x_{ki} - \bar{x}_i)^2(x_{kj} - \bar{x}_j)^2.
\]

The largest five absolute values of the studentized moments were \( t_{2,7} = -3.5756, t_{3,5} = -2.5586, t_{3,9} = 2.4368, t_{1,9} = 2.4060 \) and \( t_{1,7} = -2.2701 \), corresponding also to the largest absolute value of the residual moments reported above. The two-sided 10% Bonferroni-corrected critical value is \( z_{1-0.05/36} = 2.9913 \), so only \( \text{Cov}(x_2, x_7) \) is significantly misfitted.

The likely biases due to \( r_{2,7} \) are shown in the third column of Table 2. The parameters of the second factor measurement model are virtually unaffected by this misspecification. The largest biases are shown by \( \psi_1 \) and \( \phi_{13} \). The former is the variance of \( x_1 \), and apparently is not related to the variables \( x_2 \) and \( x_7 \) whose covariance is misfitted. The latter is the covariance/correlation of the factors loading on \( x_2 \) and \( x_7 \), and is directly responsible for modeling their covariance. None of the biases exceeds 0.03 in absolute value, or 5%, in relative terms. Arguably, those are rather minor biases. The non-centrality attributable to this misfitted moment is \( nr_{2,7}^2 W_{27,27} = 15.775 \) or \( 2nr_{2,7} W r = 23.593 \), which apparently fails to explain a sufficiently large fraction of the overall non-centrality, 85.172, or its adjusted counterpart, 72.812. Hence, other misfitted moments are likely to be present in the model.

The maximum plausible biases are reported in the last column of Table 2. In this case, the first part of the maximum in (19) provides the greater estimate of \( \|s\| \). The estimate of non-centrality based on multivariate kurtosis-adjusted test statistic was used. The reported MPBs are at least an order of magnitude larger than the biases reported in the second column of Table 2. The corresponding relative biases are sometimes as large as 50%, and cannot be ignored.

Another scale to judge the magnitudes of the biases in columns 3 and 4 of Table 2 is to compare them to the standard errors. Huber (1967) sandwich standard errors (robust to both distributional and structural
Table 2: Analysis of Holzinger-Swineford data.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>MLE estimate</th>
<th>S.e.</th>
<th>Bias due to $r_{2,7}$</th>
<th>MPB</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_{11}$</td>
<td>0.9010</td>
<td>0.1005</td>
<td>0.0148</td>
<td>0.1628</td>
</tr>
<tr>
<td>$\lambda_{21}$</td>
<td>0.4989</td>
<td>0.0879</td>
<td>-0.0183</td>
<td>0.1485</td>
</tr>
<tr>
<td>$\lambda_{31}$</td>
<td>0.6573</td>
<td>0.0807</td>
<td>-0.0101</td>
<td>0.1702</td>
</tr>
<tr>
<td>$\lambda_{42}$</td>
<td>0.9914</td>
<td>0.0614</td>
<td>0.0013</td>
<td>0.1166</td>
</tr>
<tr>
<td>$\lambda_{52}$</td>
<td>1.1034</td>
<td>0.0548</td>
<td>0.0010</td>
<td>0.1290</td>
</tr>
<tr>
<td>$\lambda_{62}$</td>
<td>0.9181</td>
<td>0.0582</td>
<td>0.0001</td>
<td>0.1070</td>
</tr>
<tr>
<td>$\lambda_{73}$</td>
<td>0.6199</td>
<td>0.0857</td>
<td>0.0028</td>
<td>0.2807</td>
</tr>
<tr>
<td>$\lambda_{83}$</td>
<td>0.7318</td>
<td>0.0922</td>
<td>0.0097</td>
<td>0.2605</td>
</tr>
<tr>
<td>$\lambda_{93}$</td>
<td>0.6717</td>
<td>0.0985</td>
<td>-0.0083</td>
<td>0.3095</td>
</tr>
<tr>
<td>$\phi_{12}$</td>
<td>0.4585</td>
<td>0.0734</td>
<td>-0.0006</td>
<td>0.0995</td>
</tr>
<tr>
<td>$\phi_{13}$</td>
<td>0.4699</td>
<td>0.0853</td>
<td>-0.0247</td>
<td>0.3184</td>
</tr>
<tr>
<td>$\phi_{23}$</td>
<td>0.2837</td>
<td>0.1183</td>
<td>-0.0020</td>
<td>0.1222</td>
</tr>
<tr>
<td>$\psi_{1}$</td>
<td>0.5511</td>
<td>0.1568</td>
<td>-0.0287</td>
<td>0.3057</td>
</tr>
<tr>
<td>$\psi_{2}$</td>
<td>1.1375</td>
<td>0.1121</td>
<td>0.0157</td>
<td>0.1809</td>
</tr>
<tr>
<td>$\psi_{3}$</td>
<td>0.8471</td>
<td>0.1005</td>
<td>0.0069</td>
<td>0.2174</td>
</tr>
<tr>
<td>$\psi_{4}$</td>
<td>0.3724</td>
<td>0.0504</td>
<td>-0.0003</td>
<td>0.2613</td>
</tr>
<tr>
<td>$\psi_{5}$</td>
<td>0.4477</td>
<td>0.0568</td>
<td>-0.0001</td>
<td>0.3069</td>
</tr>
<tr>
<td>$\psi_{6}$</td>
<td>0.3574</td>
<td>0.0466</td>
<td>0.0003</td>
<td>0.2366</td>
</tr>
<tr>
<td>$\psi_{7}$</td>
<td>0.7992</td>
<td>0.0968</td>
<td>0.0058</td>
<td>0.2606</td>
</tr>
<tr>
<td>$\psi_{8}$</td>
<td>0.4899</td>
<td>0.1184</td>
<td>-0.0118</td>
<td>0.3415</td>
</tr>
<tr>
<td>$\psi_{9}$</td>
<td>0.5694</td>
<td>0.1181</td>
<td>0.0047</td>
<td>0.3380</td>
</tr>
<tr>
<td>Non-centrality</td>
<td>85.172</td>
<td>15.775</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
misspecification) are reported in the second column of the table. They show that the sampling variability greatly exceeds the plausible biases due to \( r_{2,7} \), but the maximum plausible biases in the last column are generally at least as large as the standard errors. The most accurately estimated parameters are \( \phi_{12} \) and \( \phi_{23} \): their maximum plausible biases are about one standard error.

Least squares calibration of residual moments similar to one conducted in the end of Section 3.3 leads to a small corrective factor of 1.0357. The results agree very closely to the second column of Table 2, and not reported.

5 Discussion

One of the biggest problems arising in structurally misspecified models is how the results should be interpreted. The researcher needs to be sure that his or her model is at least approximately correct, and the parameters are still interpretable. Caution should be exercised. At best, one can say, “If the parameter is still interpretable, misspecification has probably made it smaller”, as in the case of \( \psi_5 \) in our first example. On the other hand, it should be reassuring for the researcher to find out that a certain parameter is virtually unaffected by potential misspecification, as in the case of loadings \( \lambda_3 \) and \( \lambda_4 \) in our first example, the verbal factor insensitive to the major source of misspecification in the Holzinger-Swineford example, and the factor correlations \( \phi_{12} \) and \( \phi_{23} \) that have little bias even by the most generous bounds.

By looking at the residuals of the moment matrix, one can explain some of the results in Saris et al. (1987) and Kaplan (1988) regarding different sensitivity of the overall chi-square to different types of misspecification. For instance, Saris et al. (1987) reported high sensitivity to improper linkage of an observed variable to a wrong latent variable: the non-centrality was 18. On the other hand, when a measurement error correlation of two indicators of the same latent variables was omitted, the non-centrality was only 0.07, so sensitivity to this type misspecification was very low. In the latter case only one moment was affected, so the quadratic form in (14) describing the non-centrality consisted of a single term. In the former case the whole set of moments related to the incorrectly loading variable was affected. Several elements of \( r \) contributed to the quadratic form in (14), resulting in much greater overall non-centrality.

In both our examples the zero entries of the matrix \( Q \) corresponded trivially to the diagonal entries of the sample covariance matrix that are fully absorbed by the variance parameter estimates. In more complex models, more interesting patterns might be possible. Further research into this issue may be of interest. Orthogonalities are undoubtedly present in the limited information estimation procedures such as 2SLS-IV that ignores everything beyond the currently estimated equation and its instruments.

The theoretical developments in this paper followed population-based parameters, in which non-zero residual moments should be attributed to misspecification. In finite samples, the residual moments will be non-zero due to both sampling fluctuations and (potentially) misspecifications. We have employed a simple Bonferroni correction procedure in Section 4, but other procedures to account for multiple testing and find the largest misfitting moments can be used. Providing reliable tests in this situation can also be an area of future research.

The analytical example of Section 3 showed that the proposed maximum plausible bias measures were rather conservative. It also showed that misspecification of a single moment, once correctly identified, predicted the biases almost perfectly. More extensive studies of the two procedures to estimate the biases
are called for.

Let us conclude the paper with some practical considerations. The framework we proposed allows the researcher to identify misfitting moments and gauge the approximate magnitudes of resulting biases. There are several choices the researcher can make to proceed with this information.

First, one can try to conceptualize substantive effects that could have resulted in a misspecification detected, and try to correct it by building a different model. A popular way of making changes to poorly fitting SEM model is by using modifications indices (Sörbom 1989, MI). However, as shown by Saris et al. (1987) and Kaplan (1988), the largest MI does not necessarily point to the correct problem. Besides, multiple testing nature of the MIs lead to capitalization on chance (MacCallum, Roznowski & Necowitz 1992). In our example, a better model can be obtained by incorporating the correlation between error variances of $x_1$ and $x_2$. However, it will not yield the correct model, and the resulting model will violate the asymptotic robustness conditions when these variables are non-normal.

Alternatively, the researcher may judge the biases to be sufficiently small for practical reasons, and maintain the working model. One would still need to proceed with caution about interpretability of parameters in a misfitting model. In our first example, the measurement model for the last three variables could be judged as adequate (and the biases as reasonably small), while the measurement model for the variables $x_1$ and $x_2$ is nonsensical as, unbeknownst to the researcher, these variables are the explanatory variables for the latent variable of SES. In this case, the misfitting covariance between $x_1$ and $x_2$ could prevent the researcher from interpreting the part of the model concerning $x_1$ and $x_2$.

Finally, the researcher may want to use additional and complementary estimation and testing procedures to identify the substantive source of misfit. Tetrad analysis (Bollen & Ting 1993) is one possible extension. In fact, for the two models that we have outlined in Section 3 causal test based on the tetrads (Bollen & Ting 2000) will likely be the best solution that will reject Model 2 in favor of Model 1. Another alternative way to localize the source of misfit is to use two-stage least squares estimation with instrumental variables (Bollen 1996, 2SLS-IV). Overidentification tests associated with these procedures also help locating problematic parts of a model.

References


3 For further discussion of the models with causal indicators, see Bollen & Davis (2009).
specification: two-stage least squares (2SLS) and maximum likelihood (ML) estimators’, Sociological
Methods and Research 36(1), 48–86.


Browne, M. W. (1984), ‘Asymptotically distribution-free methods for the analysis of the covariance struc-

Huber, P. (1967), The behavior of the maximum likelihood estimates under nonstandard conditions, in

Jöreskog, K. (1969), ‘A general approach to confirmatory maximum likelihood factor analysis’, Psychome-
trika 34(2), 183–202.

Kaplan, D. (1988), ‘The impact of specification error on the estimation, testing, and improvement of struc-

research’, Annual Review of Psychology 51(1), 201–226.

analysis: The problem of capitalization on chance.’, 111(3), 490–504.

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

Mooijaart, A. & Satorra, A. (2009), ‘On insensitivity of the chi-square model test to nonlinear misspecifi-
cation in structural equation models’, Psychometrika 74(3), 443–455.

Neudecker, H. & Satorra, A. (1991), ‘Linear structural relations: Gradient and Hessian of the fitting func-

tural equation models’, Sociological Methodology 17, 105–129.

Quality and Quantity 24, 367–386.

Satorra, A. (1992), ‘Asymptotic robust inference in the analysis of mean and covariance structures’, Soci-
ological Methodology 22, 249–278.
The first derivative of the implied moments is given by

$$G = \left[ L_p(I_p^2 + K_p)(\Lambda B^{-1}\Phi B^{-T} \otimes I_p), L_p(\Lambda B^{-1} \otimes \Lambda B^{-1})D_m, -L_p(I_p^2 + K_p)(\Lambda B^{-1}\Phi B^{-T} \otimes \Lambda B^{-1}), I_p^* \right].$$

(20)

Here, $D_m$ is the duplication matrix, $L_p$ is the elimination matrix, and $K_p$ is the commutation matrix of appropriate orders (Magnus & Neudecker 1999).

The second derivative of implied moments is given by expression (4) in Neudecker & Satorra (1991):

$$G = \left[ L_p(I_p^2 + K_p)(\Lambda B^{-1}\Phi B^{-T} \otimes I_p), L_p(\Lambda B^{-1} \otimes \Lambda B^{-1})D_m, -L_p(I_p^2 + K_p)(\Lambda B^{-1}\Phi B^{-T} \otimes \Lambda B^{-1}), I_p^* \right].$$

(21)

where

$$H_{11}(r) = B^{-1}\Phi B^{-T} \otimes T_{ij},$$
$$H_{12}(r) = (B^{-1} \otimes T_{ij}\Lambda B^{-1})D_m,$$
$$H_{13}(r) = -(B^{-1}\Phi B^{-T} \otimes T_{ij}\Lambda B^{-1}) - K_p(T_{ij}\Lambda B^{-1}\Phi B^{-T} \otimes B^{-1}),$$
$$H_{23}(r) = -D_{ij}^t(B^{-T} \otimes B^{-T} \Lambda T_{ij}\Lambda B^{-1}),$$
$$H_{33}(r) = (B^{-1}\Phi B^{-T} \otimes T_{ij}\Lambda B^{-1} \otimes B^{-T})K_p + K_p(B^{-T} \Lambda T_{ij}\Lambda B^{-1}\Phi B^{-T} \otimes B^{-1})$$
$$+ (B^{-1}\Phi B^{-T} \otimes B^{-T} \Lambda T_{ij}\Lambda B^{-1}),$$
$$T_{ij} = u_i u_j' + u_j u_i'.$$

A The derivatives of the implied moments

The first derivative of the implied moments is given by

$$G = \left[ L_p(I_p^2 + K_p)(\Lambda B^{-1}\Phi B^{-T} \otimes I_p), L_p(\Lambda B^{-1} \otimes \Lambda B^{-1})D_m, -L_p(I_p^2 + K_p)(\Lambda B^{-1}\Phi B^{-T} \otimes \Lambda B^{-1}), I_p^* \right].$$

(20)

Here, $D_m$ is the duplication matrix, $L_p$ is the elimination matrix, and $K_p$ is the commutation matrix of appropriate orders (Magnus & Neudecker 1999).

The second derivative of implied moments is given by expression (4) in Neudecker & Satorra (1991):

$$G = \left[ L_p(I_p^2 + K_p)(\Lambda B^{-1}\Phi B^{-T} \otimes I_p), L_p(\Lambda B^{-1} \otimes \Lambda B^{-1})D_m, -L_p(I_p^2 + K_p)(\Lambda B^{-1}\Phi B^{-T} \otimes \Lambda B^{-1}), I_p^* \right].$$

(21)

where

$$H_{11}(r) = B^{-1}\Phi B^{-T} \otimes T_{ij},$$
$$H_{12}(r) = (B^{-1} \otimes T_{ij}\Lambda B^{-1})D_m,$$
$$H_{13}(r) = -(B^{-1}\Phi B^{-T} \otimes T_{ij}\Lambda B^{-1}) - K_p(T_{ij}\Lambda B^{-1}\Phi B^{-T} \otimes B^{-1}),$$
$$H_{23}(r) = -D_{ij}^t(B^{-T} \otimes B^{-T} \Lambda T_{ij}\Lambda B^{-1}),$$
$$H_{33}(r) = (B^{-1}\Phi B^{-T} \otimes T_{ij}\Lambda B^{-1} \otimes B^{-T})K_p + K_p(B^{-T} \Lambda T_{ij}\Lambda B^{-1}\Phi B^{-T} \otimes B^{-1})$$
$$+ (B^{-1}\Phi B^{-T} \otimes B^{-T} \Lambda T_{ij}\Lambda B^{-1}),$$
$$T_{ij} = u_i u_j' + u_j u_i'.$$
For Model 2 parameters, the relevant matrices are as follows. $G$ is a $15 \times 21$ matrix whose rows correspond to the non-redundant entries of the $5 \times 5$ covariance matrix $\Sigma$ (Model 2). The first five columns correspond to the loadings $\lambda_1, \ldots, \lambda_5$ (the fifth column will be eliminated by the matrix $A$, as $\lambda_5$ is fixed to 1 for scaling); the sixth column corresponds to the variance of the latent variable $\phi_{11}$; and the remaining entries correspond to vectorization of the variance of the unique errors matrix $\Psi$, of which only the diagonal will be used in estimation.

$$
G = (G_1, I_{15}), \quad G_1 = 
\begin{pmatrix}
2.844 & 0 & 0 & 0 & 0.2146 \\
1.324 & 1.422 & 0 & 0 & 0.1998 \\
1.832 & 0 & 1.422 & 0 & 0.2764 \\
2.445 & 0 & 0 & 1.422 & 0.3689 \\
3.07 & 0 & 0 & 0 & 1.422 & 0.4632 \\
0 & 2.649 & 0 & 0 & 0 & 0.1861 \\
0 & 1.832 & 1.324 & 0 & 0.2574 \\
0 & 2.445 & 0 & 1.324 & 0 & 0.3435 \\
0 & 3.07 & 0 & 0 & 1.324 & 0.4314 \\
0 & 0 & 3.663 & 0 & 0 & 0.356 \\
0 & 0 & 2.445 & 1.832 & 0 & 0.4752 \\
0 & 0 & 3.07 & 0 & 1.832 & 0.5967 \\
0 & 0 & 0 & 4.89 & 0 & 0.6342 \\
0 & 0 & 0 & 3.07 & 2.445 & 0.7964 \\
0 & 0 & 0 & 0 & 6.14 & 1
\end{pmatrix}
$$

The relevant parameter selection matrix $A$ is a $21 \times 10$ matrix, with all entries equal to zero except the entries $(1,1), (2,2), (3,3), (4,4), (5,5), (6,6), (7,6), (12,7), (16,8), (19,9),$ and $(21,10)$ equal to one.

The matrix $Q = \partial \theta / \partial s$ is a $10 \times 15$ matrix, with rows corresponding to the parameters of the model, and columns corresponding to the entries of $\text{vech} \ R$. Its transpose is

$$
Q' = 
\begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
.1899 & .2135 & .0311 & .0335 & -.6933 & -.3062 & -.3487 & .0668 & .163 & .5671 \\
.128 & .0445 & .1721 & .0288 & -.4962 & -.2522 & .0216 & -.4813 & .1363 & .4757 \\
.1755 & .0623 & .0425 & .243 & -.7633 & -.3316 & .0363 & .0883 & -.7688 & .7534 \\
.0563 & -.0823 & -.1541 & -.2299 & 1.022 & -.3855 & .0896 & .194 & .4741 & -.1096 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
.0325 & .1082 & .1264 & .02 & -.3558 & .0199 & -.216 & -.3592 & .0953 & .3326 \\
.0458 & .1479 & .0299 & .1785 & -.5496 & .0336 & -.2866 & .0623 & -.5778 & .5314 \\
-.0701 & .0736 & -.1164 & -.1728 & .7796 & .0844 & -.3488 & .1397 & .3413 & -.8488 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
.0098 & .0083 & .1501 & .1526 & -.3873 & .0368 & .0286 & -.4074 & -.496 & .4465 \\
-.0986 & -.0934 & .0655 & -.1474 & .7748 & .0828 & .0642 & -.4909 & .2882 & -.7349 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-.1303 & -.1237 & -.1269 & .0323 & .9413 & .1221 & .0947 & .1739 & -.6907 & -.8673 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\end{pmatrix}
$$
The $p^* \times p^*$ hat matrix $\hat{H} = \partial r / \partial s'$ has columns corresponding to the sample moments and rows corresponding to the residual moments. The columns corresponding to the diagonal entries are $15 \times 1$ zero vectors. The non-zero columns of $\hat{H}$ are:

$$
\begin{pmatrix}
-0.054 & -0.054 & -0.038 & 0.062 & -0.0359 & -0.0458 & -0.0524 & 0.0184 & 0.0315 & 0.0466 \\
0.5834 & -0.1336 & -0.1685 & -0.1617 & -0.1258 & -0.1611 & -0.1677 & 0.0526 & 0.1086 & 0.1604 \\
-0.2005 & 0.5979 & -0.1709 & -0.1664 & -0.1409 & 0.0256 & 0.0784 & -0.1243 & -0.1266 & 0.1590 \\
-0.2561 & -0.1708 & 0.5069 & -0.1876 & 0.0234 & -0.1629 & 0.1294 & -0.0981 & 0.1649 & -0.0745 \\
-0.2519 & -0.1631 & -0.1852 & 0.3539 & 0.0652 & 0.1141 & -0.1460 & 0.1493 & -0.0561 & -0.0359 \\
-0.0878 & -0.0472 & -0.0593 & -0.0619 & -0.0044 & -0.0028 & 0.0087 & 0.0215 & 0.0390 & 0.0577 \\
-0.2539 & -0.1817 & 0.0261 & 0.0918 & 0.7259 & -0.1690 & -0.1814 & -0.1143 & -0.1151 & 0.1523 \\
-0.3282 & 0.0236 & -0.2120 & 0.1546 & -0.1690 & 0.5909 & -0.2190 & -0.0894 & 0.1574 & -0.0638 \\
-0.3565 & 0.0774 & 0.1379 & -0.1883 & -0.1788 & -0.2169 & 0.4377 & 0.1415 & -0.0475 & -0.0264 \\
0.0662 & 0.0275 & 0.0279 & 0.0067 & 0.0228 & 0.0239 & 0.0092 & 0.0044 & -0.0250 & 0.0442 \\
0.1921 & -0.2377 & -0.1863 & 0.3123 & -0.1767 & -0.1388 & 0.2307 & 0.5377 & -0.2584 & -0.1962 \\
0.3182 & -0.2323 & 0.3251 & -0.1368 & -0.1758 & 0.2362 & -0.1078 & -0.2295 & 0.3366 & -0.1722 \\
0.1130 & 0.0377 & 0.0645 & 0.0019 & 0.0323 & 0.0537 & 0.0092 & -0.0045 & -0.0589 & 0.0520 \\
0.4494 & 0.3068 & -0.1382 & -0.1081 & 0.2218 & -0.1102 & -0.0904 & -0.1601 & -0.1645 & 0.1513 \\
0.1262 & 0.0205 & 0.0099 & 0.0740 & 0.0232 & 0.0182 & 0.0692 & -0.0592 & -0.0399 & 0.0740 \\
\end{pmatrix}
$$

The diagonal entries are underlined to guide the eye.