SATISFYING A SIMPLEX STRUCTURE IS SIMPLER THAN IT SHOULD BE

DAVID ROGOSA and JOHN B. WILLET
Stanford University

KEY WORDS. Longitudinal data analysis, simplex model, growth curves.

ABSTRACT. Markedly different types of growth (learning) curves may generate indistinguishable covariance structures. We illustrate with an example of a $5 \times 5$ covariance matrix representing longitudinal measurements at five occasions. This example appears to conform closely to a simplex correlation pattern, and a simplex covariance structure provides an excellent fit (using LISREL V) to this covariance matrix. However, the (known) structure of this example differs greatly from the simplex model. In addition to indicating that an excellent fit of a simplex structure can be misleading, this example provides an opportunity to question common uses of covariance structure models for the study of growth.

Since the introduction of the simplex model by Guttman (1954), simplex structures have been posited for many types of ordered variables (such as hierarchical tests or longitudinal measurements). In many analyses of longitudinal data, a simplex covariance or correlation structure serves as a basic model for statistical analysis (e.g., Humphreys, 1960; Jöreskog, 1970a, 1978; Jöreskog & Sörbom, 1977; Werts & Hilton, 1977; Werts, Linn, & Jöreskog, 1977, 1978). As Werts, Linn, and Jöreskog assert, "The simplex model appears to be particularly appropriate for studies of academic growth (Humphreys, 1960, 1968; Lunneborg & Lunneborg, 1970)" (p. 745).

Simplex Growth Model

Following Jöreskog (1970a), a discrete-time autoregressive model of lag 1, which gives rise to the simplex structure, has the form:

\[ \eta_{t+1,p} = \beta_t \eta_{tp} + \delta_{t+1,p}, \]

where \( \eta_{tp} \) indicates the value of \( \eta \) for the \( p \)th individual at time \( t \), and where the \( \delta_{t+1,p} \) are independent disturbances (increments to growth) with expectation zero and \( \eta_{tp} = \delta_{tp} \). (See also the stochastic models considered in Anderson, 1960, Section 3.) An important property of Equation 1 is that the partial correlation is zero between \( \eta \) at two time points with \( \eta \) at an intervening time partialled out. This condition, which was used by Guttman (1954, Equation 15) to characterize the simplex structure, can be written

\[ \rho_{\eta_{t}, \eta_{t+t}} = 0 \quad (t_1 < t_2 < t_k). \]
"Unsimplex" Growth Models

An alternative simple model for growth specifies that each individual has a constant rate of change. This constant rate of change model yields a straight line growth curve for each individual:

\[ \eta_p = \eta_0 + \theta_p(t_i - t_i). \]  

(3)

A striking consequence of the constant rate of change model is that it represents the most extreme violation of Guttman's condition for the simplex. That is, for straight-line growth,

\[ p_{\eta_0, \eta_0} = -1 \quad (t_i < t_j < t_k). \]  

(4)

A related property of straight-line growth is that \( \beta_{\eta_0, \eta_0} = (t_j - t_i)/(t_j - t_i) \) (e.g., with equally spaced time points \( \beta_{\eta_0, \eta_0} = -1 \)). In contrast, the autoregressive model in Equation 1 specifies that \( \beta_{\eta_0, \eta_0} = 0 \) (for \( t_i < t_j < t_k \)). Also, for straight-line growth, \( \beta_{\eta_0, \eta_0} = (t_k - t_i)/(t_j - t_i) \) (e.g., with equally spaced time points \( \beta_{\eta_0, \eta_0} = 2 \)).

A second type of growth (learning) curve for which Equation 4 holds is the function that specifies that individuals approach an asymptote \( \lambda_0 \) exponentially with rate parameter \( \gamma \) (where \( \gamma \) is identical for all individuals). This exponential growth curve can be written as

\[ \eta_p = \lambda_0 - (\lambda_0 - \eta_0) e^{-\gamma(t_i - t_i)}. \]  

(5)

Among the applications of the growth curve in Equation 5 to the study of learning are Hicklin (1976), Keats (1983), and Sørensen and Hallinan (1977).

The Numerical Example

A numerical example consisting of a covariance matrix of observed scores over five occasions of measurement is given in Table I. The relation between the observed score \( Y_{i\rho} \) and the true score \( \eta_{i\rho} \) is specified by the measurement model:

\[ Y_{i\rho} = \eta_{i\rho} + \epsilon_{i\rho}, \]  

(6)

where \( Y_{i\rho} \) and \( \eta_{i\rho} \) indicate the values of observed and true scores for the \( p \)th individual \( (p = 1, \ldots, 500, \text{as a sample size of 500 is used in this example}) \) at the time \( t_i \) (for discrete times \( t_1, t_2, \ldots, t_5 \)) and where \( \epsilon_{i\rho} \) denotes the error of measurement in \( Y_{i\rho} \) at time \( t_i \).

The corresponding correlation matrix for the \( Y_{i\rho} \) in Table II displays the typical properties of a simplex correlation structure with the correlations decreasing for elements farther away from the main diagonal (and with nearly equal elements along the subdiagonals). Furthermore, Table II is similar to correlation matrices obtained in large-scale longitudinal studies of student
Satisfying a Simplex Structure

TABLE I

*Observed Score Covariance Matrix*

<table>
<thead>
<tr>
<th></th>
<th>(Y_1)</th>
<th>(Y_2)</th>
<th>(Y_3)</th>
<th>(Y_4)</th>
<th>(Y_5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Y_1)</td>
<td>.619</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Y_2)</td>
<td>.453</td>
<td>.595</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Y_3)</td>
<td>.438</td>
<td>.438</td>
<td>.587</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Y_4)</td>
<td>.422</td>
<td>.430</td>
<td>.438</td>
<td>.595</td>
<td></td>
</tr>
<tr>
<td>(Y_5)</td>
<td>.406</td>
<td>.422</td>
<td>.438</td>
<td>.453</td>
<td>.619</td>
</tr>
</tbody>
</table>

TABLE II

*Observed Score Correlation Matrix*

<table>
<thead>
<tr>
<th></th>
<th>(Y_1)</th>
<th>(Y_2)</th>
<th>(Y_3)</th>
<th>(Y_4)</th>
<th>(Y_5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Y_1)</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Y_2)</td>
<td>.746</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Y_3)</td>
<td>.727</td>
<td>.741</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Y_4)</td>
<td>.695</td>
<td>.723</td>
<td>.741</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>(Y_5)</td>
<td>.656</td>
<td>.695</td>
<td>.727</td>
<td>.746</td>
<td>1.000</td>
</tr>
</tbody>
</table>

achievement (cf., the ETS Growth Study data in Werts & Hilton, 1977, Table 1), both in structure and in the magnitude of the correlations.

The covariance matrix in Table I was not generated using Equations 1 and 6. Instead, the example was generated by the constant rate of change model in Equation 3. (Appendix A provides the details of the construction of Table I using the constant rate of change model.) Also, covariance matrices having structure identical to the example in Table I are easily constructed for collections of exponential growth curves having the form of Equation 5. Thus, the key features of our example are not limited to the particular case of the constant rate of change model in Equation 3.

**Examining the Example with LISREL**

The correlation matrix in Table II appears to the eye to conform exceptionally well to a simplex structure. A more powerful scrutiny of the example may be obtained by using covariance structure analysis to fit a simplex model to the example. A quasi-simplex structure for the observed-score covariance matrix posits that the true-score covariance matrix has a simplex structure. Equations 1 and 6 constitute a specification of the quasi-simplex model, which can also be represented by the “path analysis” diagram in Figure 1 of Jöreskog (1970a) or in Figure 3.5 of Jöreskog and Sörbom (1981). The exact specification of the covariance structure that is fit to the observed score covariance matrix in Table I is given in Appendix B.
The quasi-simplex covariance structure model in Appendix B provides an excellent fit to the covariance matrix in Table I. The elements of the “reproduced” and sample covariance matrices are almost identical (compare Tables III and I). Alternatively, compare the elements of the reproduced and observed correlation matrices in Tables IV and II. In Table IV, the boldface entries are the differences between the elements of the sample and reproduced correlation matrices.

The root mean square residual (or root mean square difference) is also used as a summary statistic of the discrepancy between the observed and reproduced matrices from the LISREL analysis (Jöreskog & Sörbom, 1981; Werts & Hilton, 1977). (The root mean square residual for the covariances is computed by LISREL V; the root mean square residual for the correlations was computed by adapting the formula in Jöreskog and Sörbom, 1981, p. I.41 for the off-diagonal elements of the reproduced and sample correlation matrices). In our analysis, the root mean square residual is .003 for the reproduced covariance matrix and .006 for the reproduced correlation matrix. (For root mean square differences of this magnitude, Werts & Hilton concluded, “Such small differences are clearly not meaningful,” p. 141.)

Furthermore, the χ² goodness-of-fit statistic for the model has five degrees of freedom and a value of 2.13 (probability level of .831). The adjusted goodness-of-fit index is .995. Thus, considering any commonly used criteria, this LISREL analysis provides no indication that the data were generated from a model that is maximally “unsimplex.”

Consequences of Concluding “Simplex”

The fit of the covariance matrix in Table I by the quasi-simplex covariance structure demonstrates that a LISREL analysis may not be sensitive to extreme violations of the posited growth model. Moreover, this example suggests that, even when the fit of the simplex model is excellent, neither substantive interpretations of model parameter estimates nor assessments of reliability can be made with assurance.

Growth statistics. In empirical studies of growth that fit quasi-simplex covariance structures to the longitudinal data, a number of “growth statistics” are computed and given substantive interpretation (e.g., Werts & Hilton, 1977; Werts et al., 1977). Foremost among these growth statistics are estimates of

---

1To protect against the possibility of being misled by local minima, we repeated the LISREL V analysis with different sets of initial estimates. Identical fit statistics and parameter estimates were obtained when (a) the initial estimates were computed by LISREL V (with values of σ² of .57), and (b) the initial estimate for σ² was specified as .145, a value obtained from the reliability estimation procedure of Humphreys (1960).
the correlation between “true” change and “true” initial status. The direct relation between the $\beta_i$ of Equation 1 and the correlation of $\eta_i$ and $\eta_{i-1} - \eta_i$ is easily obtained (see Werts & Hilton, 1977, Equation 4). The estimates of the $\beta_i$ for the example in Table I are all less than 1 (values .937, .956, .971, and .985 for $i = 1, \ldots, 4$, respectively). Consequently, the estimates of $\rho_{\eta_i\eta_{i-1} - \eta_i}$ obtained from the LISREL V analysis are all negative with values of $-.225$, $-.164$, $-.107$, and $-.052$ (for $i = 1, \ldots, 4$, respectively). These estimates differ, especially for $i = 3$ and 4, from the actual values of the correlation between true change and true initial status (based on Equation 4 and Appendix A), which are $-.258$, $-.132$, $0$, and $.132$ (for $i = 1, \ldots, 4$, respectively).\footnote{An important misconception in many analyses of growth is the idea of a single correlation between true gain and true initial status. For the autoregressive model in Equation 1, unequal $\beta_i$ cause the correlation between true change and true initial status to differ for different $t$. For growth models such as Equation 3, the correlation between change and initial status systematically depends on the time of initial status (Rogosa & Willett, in press). Consequently, averaging estimates of this correlation for different times of initial status (as is done in the quasi-simplex covariance structure analysis of Werts et al., 1977) is unwise.}

\begin{table}
\centering
\caption{Reproduced Covariance Matrix from the Quasi-Simplex Analysis}
\begin{tabular}{rrrrr}
\hline
 & $Y_1$ & $Y_2$ & $Y_3$ & $Y_4$ & $Y_5$
\hline
$Y_1$ & .619 & & & & \\
$Y_2$ & .453 & .595 & & & \\
$Y_3$ & .434 & .440 & .588 & & \\
$Y_4$ & .421 & .427 & .440 & .595 & \\
$Y_5$ & .415 & .421 & .434 & .453 & .619 \\
\hline
\end{tabular}
\end{table}

\begin{table}
\centering
\caption{Reproduced Correlations and Residuals From the Quasi-Simplex Analysis}
\begin{tabular}{rrrrr}
\hline
 & $Y_1$ & $Y_2$ & $Y_3$ & $Y_4$ & $Y_5$
\hline
$Y_1$ & 1.000 & & & & \\
$Y_2$ & .746 & 1.000 & & & \\
$Y_3$ & .719 & .744 & 1.000 & & \\
$Y_4$ & .694 & .718 & .744 & 1.000 & \\
$Y_5$ & .670 & .694 & .719 & .746 & 1.000 \\
\hline
\end{tabular}
\end{table}
Another commonly reported growth statistic is an estimate of the variance of \( \eta_{i+1} - \eta_i \), representing individual differences in true change between adjacent time points. The LISREL V analysis yields estimates of .038, .033, .033, and .038 (for \( i = 1, \ldots, 4 \), respectively), which are nearly five times larger than the actual value of .0078 (for all \( i \)).

Reliability estimation. The population values of the reliability of the observed measures, written as \( \rho(Y_i) \), are .758, .748, .745, .748, and .758 for \( i = 1, \ldots, 5 \) respectively. The quasi-simplex covariance structure model can be used to obtain estimates of the \( \rho(Y_i) \), as in Werts et al. (1977, p. 751) and Werts et al. (1978, p. 91). The estimates of \( \rho(Y_i) \) obtained from our LISREL analysis are .782, .773, .770, .773, and .782 for \( i = 1, \ldots, 5 \). A traditional estimate of the reliability of \( Y_i \) (Humphreys, 1960) is \( r_{ni} / r_{nj} \) (for \( t_k < t_i < t_j \)), which is derived from Guttman's condition for a simplex in Equation 2. Using \( h = i - 1 \) and \( j = i + 1 \), the estimates of \( \rho(Y_i) \) from Table II are .761, .760, and .761 for \( i = 2, 3, 4 \), respectively (estimates cannot be obtained for the first and last times). When Equation 2 does not hold, this reliability estimate will be systematically biased; for \( r_{ni} / r_{nj} < 0 \) (e.g., Equation 4) the bias is positive. If the estimate of \( \rho(Y_i) \) is constructed by averaging over all available \( h \) and \( j \) (as in Humphreys, 1960), the reliability estimates move further from \( \rho(Y_i) \) and approach the reliability estimates obtained from the LISREL analysis.

Also, estimates of the reliability of the gain score can be obtained from the quasi-simplex covariance structure analysis (see Werts et al., 1972; Werts et al., 1977, Equation 11). For our example, the LISREL V analysis yields estimates of the reliability of differences between \( Y_i \) and \( Y_{i-1} \) of .123, .109, and .123 for \( i = 1, \ldots, 4 \), respectively. (These reliability estimates are similar in magnitude to those obtained by Werts and Hilton, 1977, for the STEP subtests from the ETS Growth Study.) However, the actual values of these reliabilities are all .025; the LISREL estimates are too large by at least a factor of 4. In addition, the LISREL reliability estimates for gains over longer intervals are too large; for example, for differences between \( Y_i \) and \( Y_{i+3} \), the LISREL V analysis yields an estimated reliability of .274 for both \( i = 1, 2 \), whereas the actual reliability in each case is .189.

Discussion

The covariance matrix is a severe summary of the longitudinal data, one that may discard crucial information about growth. Although the covariance matrix in Table I was generated using a growth model that is maximally disparate from a simplex model (compare Equations 2 and 4), inspection of the correlation matrix and analysis of the covariance structure both indicate that the example "satisfies" a simplex structure. This would seem, at the very least, to challenge the automatic use of the simplex model in covariance structure analyses that seek to investigate growth or change.
The example (which is just one of many possible examples) provides one indication that very different types of individual growth curves may yield indistinguishable covariance or correlation structures. The example is not intended as a "contest" between the growth models, and hypothesis testing is not featured. The surprise in the example is the difficulty in detecting that the covariance matrix in Table I is not generated by Equations 1 and 6, given that the growth curves generated by the autoregressive model in Equation 1 are very distinct from straight-line growth. (Individual growth trajectories will differ greatly for the two models.) The purpose of this paper is less to attack covariance structure analysis than to appeal for explicit description and modeling of growth curves as a foundation for the analysis of longitudinal data.

**APPENDIX A**

**Construction of the Example**

For straight-line growth in Equation 3, the elements of the covariance matrix of the $\eta_i$ can be expressed:

$$\sigma_{\eta_i \eta_j} = \sigma_{\eta_i}^2 + (t_i + t_j - 2t_i)\sigma_{\eta_i \eta_j} + (t_i - t_j)(t_j - t_i)\sigma^2_{\eta_i},$$  \hspace{1cm} (A1)

where $t_i = k$ ($k = 1, \ldots, 5$). Our example used the values $\sigma_{\eta_i}^2 = .4686$, $\sigma_{\eta_i \eta_j} = -.0156$, and $\sigma^2_{\eta_i} = .0078$. Finally, Table 1 was obtained by adding "measurement error variance" of .15 to the diagonal elements of the covariance matrix of the $\eta_i$. Also, this specification determines the values of the following quantities included in the text: $\rho_{\eta_1 \eta_5}$, the variance of $\eta_{t_1} - \eta_{t_5}$, and the reliabilities of $Y_i$ and $Y_{t_5} - Y_{t_1}$.

**APPENDIX B**

**Specification of the Covariance Structure for the Quasi-Simplex**

The standard covariance structure model with "no X" is written (see Jöreskog & Sörbom, 1981, Equation 1.4 and Section 1.3):

$$\Sigma = \Lambda_j (I - B)^{-1} \Psi (I - B')^{-1} \Lambda_j' + \Theta_j. \hspace{1cm} (B1)$$

The covariance structure used in the LISREL analysis further specifies (B1) as follows:

$$\Lambda_j = I,$$

$$B_{5 \times 5} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ \beta_1 & 0 & 0 & 0 & 0 \\ 0 & \beta_2 & 0 & 0 & 0 \\ 0 & 0 & \beta_3 & 0 & 0 \\ 0 & 0 & 0 & \beta_4 & 0 \end{bmatrix},$$  \hspace{1cm} (B2)

$$\Psi = \text{diag}(\sigma_{\eta_1}^2, \sigma_{\eta_2}^2, \sigma_{\eta_3}^2, \sigma_{\eta_4}^2, \sigma_{\eta_5}^2),$$

$$\Theta_j = \sigma^2 I.$$
The specification of $B$ corresponds to Equation 1. The specification of $A_i$ corresponds to Equation 6, and the specification of $\Theta_e$ restricts the measurement error variances in $Y$ to be equal at all times, a restriction consistent with the structure built into the example (see Appendix A). This restriction of equal measurement error variances also has been used in previous quasi-simplex covariance structure analyses (see Jöreskog, 1970b, Equation 33 and p. 247: 1979, Equation 34).

References


**Authors**

DAVID ROGOSA, Assistant Professor. School of Education, Stanford University, Stanford, CA 94305. *Specializations*: Statistical methods, longitudinal research.

JOHN B. WILLET, Assistant Professor. Graduate School of Education, Harvard University, Cambridge, MA 02138. *Specialization*: Statistical methods in educational research.