Psychological Methods
2006, Vol. 11, No. 4, 439–455

1082-989X/06/$12.00 DOI: 10.1037/1082-989X.11.4.439

On the Likelihood Ratio Test in Structural Equation Modeling When Parameters Are Subject to Boundary Constraints

Reinoud D. Stoel, Francisca Galindo Garre, Conor Dolan, and Godfried van den Wittenboer
University of Amsterdam

The authors show how the use of inequality constraints on parameters in structural equation models may affect the distribution of the likelihood ratio test. Inequality constraints are implicitly used in the testing of commonly applied structural equation models, such as the common factor model, the autoregressive model, and the latent growth curve model, although this is not commonly acknowledged. Such constraints are the result of the null hypothesis in which the parameter value or values are placed on the boundary of the parameter space. For instance, this occurs in testing whether the variance of a growth parameter is significantly different from 0. It is shown that in these cases, the asymptotic distribution of the chi-square difference cannot be treated as that of a central chi-square-distributed random variable with degrees of freedom equal to the number of constraints. The correct distribution for testing 1 or a few parameters at a time is inferred for the 3 structural equation models mentioned above. Subsequently, the authors describe and illustrate the steps that one should take to obtain this distribution. An important message is that using the correct distribution may lead to appreciably greater statistical power.

Keywords: structural equation modeling, likelihood ratio test, boundary parameters, inequality constraints, asymptotic distribution

Supplemental material: http://dx.doi.org/10.1037/1082-989X.11.4.439.supp

Structural equation modeling (SEM) has become an important statistical technique in behavioral and educational sciences. In the past few decades, the number of quantitative research publications incorporating SEM has grown exponentially. With roots dating back to work on path analysis (Wright, 1918), factor analysis (Thurstone, 1947), and their combination (Jöreskog, 1973; Keesling, 1972; Wiley, 1973), SEM's attractiveness is largely due to its flexibility in specifying and testing hypotheses about linear relations among both observed and latent variables in one or more groups.

A key aspect of SEM is the assessment of the overall fit of the proposed model. Consequently, considerable theoretical and empirical effort has been devoted to the investigation and development of measures of overall model fit. In contrast, the test of specific model parameters has received relatively little attention in the SEM literature (but see Gonzalez & Griffin, 2001; Neale & Miller, 1997). If the model with a constraint on one or more parameters may be regarded as being nested within the model without the constraint, a chi-square-difference test (i.e., a likelihood ratio [LR] test) is often performed to test the tenability of the constraint.[1] The value of the test statistic is compared with a central chi-square distribution with degrees of freedom equal to the number of parameter constraints. Although this standard practice does produce correct results for many parameters (i.e., regression coefficients, factor loadings, factor covariances), it may lead to incorrect results, specifically, incorrect p values, for certain classes of parameters.

These classes consist of parameters that are implicitly or explicitly constrained. Variance parameters, with admissible values being equal to or greater than zero, constitute one such class; correlations with admissible values between –1 and +1 constitute another. Such so-called boundary parameters are of special interest in estimation and testing of commonly applied

Reinoud D. Stoel and Godfried van den Wittenboer, Department of Education, University of Amsterdam, Amsterdam, the Netherlands; Francisca Galindo Garre, Department of Clinical Epidemiology and Biostatistics, University of Amsterdam, Amsterdam, the Netherlands; Conor Dolan, Department of Psychology, University of Amsterdam, Amsterdam, the Netherlands.

Correspondence concerning this article should be addressed to Reinoud D. Stoel, Department of Education, University of Amsterdam, PO-Box 94208, NL-1090 GE Amsterdam, the Netherlands. E-mail: r.d.stoel@uva.nl

[1] Alternatively, the Wald test may be used to test the statistical significance of a set of parameters. However, as pointed out by Neale and Miller (1997; see also Azzalini, 1996; Gonzalez & Griffin, 2001), the Wald test has the drawback that the results depend on the parameterization. Here, we consider only the LR test as it does not have this deficiency.

Table 1
Parameter Estimates of Three Growth Curve Models on the Reading Comprehension Data of Bast and Reitsma (1997)

Parameter	Model 1	Model 2	Model 3
Var(level)	10.490 (1.52)	10.277 (1.44)	10.047 (1.43)
Var(growth rate)	.020 (.01)	.018 (.01)	0
Cov(level, growth rate)	−.043 (.10)	0	0
$\varepsilon_1 = \varepsilon_2$	20.876 (1.73)	21.132 (1.73)	22.008 (1.69)
$\varepsilon_3 = \varepsilon_4$	17.739 (1.68)	17.687 (1.68)	19.346 (1.52)
$\chi^2(df)$	6.577 (5), $p = .25$	6.793 (6), $p = .34$	9.552 (7), $p = .22$
$\Delta\chi^2(df)$		.22 (1), $p = .66$	2.76 (1), $p = .10$

Note. These models were refitted to the covariance matrix only. $N = 235$. Standard errors are presented in parentheses. Var = variance; Cov = covariance.

structural equation models such as the common factor (CF) model (Van der Sluis, Dolan, & Stoel, 2005), the quasi-simplex model (Jöreskog, 1970; Rovine & Molenaar, 2005), and the latent growth curve (LGC) model (Laird & Ware, 1982; McArdle, 1986, 1988; Meredith & Tisak, 1984, 1990; Willett & Sayer, 1994). On the other hand, boundary parameters may be the result of explicit inequality constraints imposed by the researcher to test a specific hypothesis of interest, for instance, the constraint that an unstandardized regression coefficient is smaller than one. Many of the current SEM software packages allow for such user-specified bounds.

The possible effects of inequality constraints in the psychological SEM literature have not, to our knowledge, been discussed in any detail. Gonzalez and Griffin (2001, p. 263) did note that "the likelihood ratio test should not be performed when one parameter is tested at a boundary," but they did not go into any detail. In the econometric and behavior genetic literature, on the other hand, the effects of parameter bounds, usually on variance components, upon the null distribution of the LR test have received considerable attention (e.g., Carey, 2005; Chernoff, 1954; Crainiceanu & Ruppert, 2004; Dominicus, Skrondal, Gjessing, Pedersen, & Palmgren, 2006; Self & Liang, 1987; Sham, 1998; Shapiro, 1985; Stram & Lee, 1994). Stram and Lee (1994), in particular, discussed the asymptotic behavior of LR tests of variance components in the longitudinal mixed effect model described by Laird and Ware (1982). Although this longitudinal mixed effect model of Laird and Ware is very similar to LGC models (Rovine & Molenaar, 2000), Stram and Lee's results do not seem to have been picked up in the SEM literature, except for the single remark in Gonzalez and Griffin (2001). Moreover, even if the problems of boundary parameters in SEM were fully recognized, to determine the exact nature of and solution to the problem is no simple matter. The aim of the present article is, therefore, to demonstrate how the imposition of inequality constraints on parameters may affect the asymptotic distribution of the LR test in the classes of commonly used models mentioned above.

On the basis of the work of Self and Liang (1987) and Stram and Lee (1994, 1995), the correct distribution for testing one or a few boundary parameters at the same time is inferred, and it is shown that the asymptotic distribution of the LR is often not the standard central chi-square distribution. It is shown that this distribution is a mixture of central chi-square distributions. For complicated multiparameter tests, simulation procedures are discussed, by means of which the (mixture) distribution of the LR statistic may be inferred. The major message of this article is that the traditional distribution of the test statistic (with degrees of freedom equal to the number of constraints, i.e., the naïve test) has too heavy a tail in all situations. In other words, traditional p values tend to be greater than the p values of the true distribution, leading to too-conservative hypothesis tests. In certain complicated testing situations, such as those encountered in CF analysis and LGC analysis, the asymptotic bias of the naïve test and the effect on the statistical power may be quite substantial.

Illustration

To illustrate the problem, we consider the data provided by Bast and Reitsma (1997) in their article on the comparison of growth curve models and quasi-simplex models. The data consist of measurements on reading comprehension measured on four occasions with 235 children. Bast and Reitsma fitted a simple linear growth curve model using normal theory maximum-likelihood estimation. From their results, it can be inferred that there is substantial variation in the level factor (i.e., intercept) but that both the covariance between level and growth rate (i.e., slope) and the variance of the growth rate factor do not differ significantly from zero. Table 1 contains the results of a reanalysis of the covariance matrix provided in the article by Bast and Reitsma.[2] The chi-square difference

[2] Please note that, to ensure that our reanalysis of the data corresponded exactly with the analysis of Bast and Reitsma (1997), only the covariance structure was modeled and that the mean structure was saturated. In a strict sense, this model is not true a growth curve model. It can, however, be easily shown that including the mean structure in the model would not change the distribution of the likelihood statistic for testing a variance component because the means can be regarded as nuisance parameters in this test (see the section entitled Asymptotic Distributions of the LR Test Statistic in Specific Situations, below).

between Model 2 and Model 1, testing the covariance between the level and the growth rate, equals 0.22. The common practice of relating this chi-square difference to a chi-square random variable with one degree of freedom does not lead to a rejection of the null hypothesis (p = .64). Subsequently, the chi-square difference between Model 2 and Model 3, in which the variance of the growth rate is fixed to zero, equals 2.76 (p = .097). Given an alpha level of .05, it may be concluded that the covariance between level and growth rate and the variance of the growth rate are not significantly different from zero. In other words, the null hypothesis that individual differences in growth are absent cannot be rejected.

However, as mentioned above, the distribution to which the chi-square difference should be referred need not be that of a chi-square random variable with degrees of freedom equal to the number of constraints. To show what the correct distribution of the chi-square difference looks like in the situation of Bast and Reitsma (1997), we performed two small simulation studies, one for each test. In the first simulation study, we generated 5,000 datasets.[3] The parameter estimates under Model 2 were used as the population values of the model according to which the data were generated. This implied that the covariance between level and shape factor would be equal to zero in the population. Subsequently, both Model 1 and Model 2 were estimated for each simulated data set, and the chi-square difference between the two models was computed. Figure 1A displays the empirical density of the 5,000 simulated chi-square differences for testing whether the covariance between level and growth rate would equal zero. This distribution is very similar to the $\chi^2(1)$ distribution. The mean and variance of the empirical distribution are 1.03 and 2.08, close to the expected values of the $\chi^2(1)$ distribution of 1 and 2, respectively. So, here, we do not encounter a problem. The second simulation study consisted of a similar procedure, in which 5,000 datasets were generated but now under the null hypothesis that both the variance of the growth rate and the covariance between level and growth rate were equal to zero. The parameter estimates of Model 3 were thus used as the population values, and Model 2 and Model 3 were estimated on each of these 5,000 datasets. The empirical distribution of the chi-square difference is displayed in Figure 1B; the mean and variance of the empirical distribution are 0.51 and 1.23. It is immediately apparent that this distribution does not correspond to the distribution of a $\chi^2(1)$ random variable. Below, we explain why the tail of this distribution is lighter than that of a $\chi^2(1)$ distribution.

In conceptual terms, these results can be understood as follows. An estimated parameter in SEM is assumed to have an asymptotic normal sampling distribution. This means that a parameter estimate can take any value. For instance, an estimated covariance by itself has no constraints regarding the value it can take because it can be negative, zero, or positive,[4] so there appears to be no violation of the normality assumption. Supposing that the true covariance in a population is zero, one then has a .5 chance of finding a negative estimate of the covariance. If, in a simulation study, the difference in likelihood of the model with the covariance fixed to zero and the model with the covariance

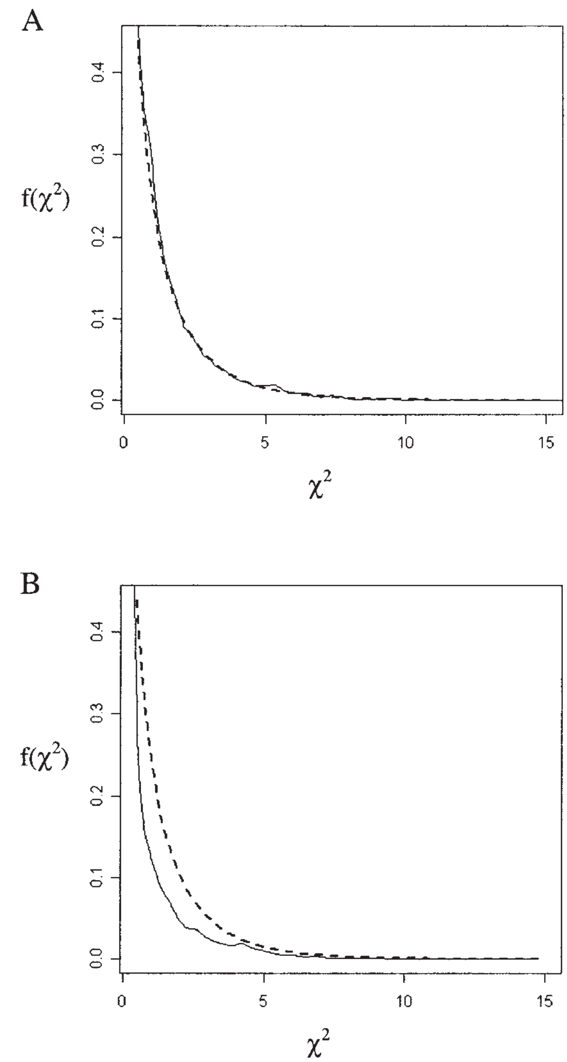

Figure 1. A: Density of the 5,000 simulated chi-square differences for testing whether the covariance between level and growth rate equals zero (solid line) and density of the $\chi^2(1)$ distribution (dashed line). B: Density of the 5,000 simulated chi-square differences for testing whether the variance of the growth rate equals zero (solid line) and density of the $\chi^2(1)$ distribution (dashed line).

[3] All corresponding scripts (for Mplus3 [Muthén & Muthén, 2004] and R [R Development Core Team, 2005]) can be downloaded from the *Psychological Methods* Web site. Although we used Mplus3 for our Monte Carlo simulations and SEM analyses, in principle, any other SEM package could have been used.

[4] Note that a covariance parameter may be subject to constraints originating from the requirement that the covariance matrix, which contains the covariance parameter, be positive (semi-)definite.

freely estimated is computed in each sample, one would expect a negative covariance in 50% of the samples. Both negative and positive covariances contribute to this distribution. Thus, given a sufficiently large sample size and number of replications, this distribution would approach a central $\chi^2(1)$ distribution.

If the parameter represents a variance, on the other hand, the estimate of this parameter should not be negative.[5] The implied censoring of the parameter's distribution, implicit in the definition of a variance parameter, results in a violation of the asymptotic normality assumption. Without the imposition of a constraint, one has a .5 probability of obtaining a negative estimate if the true value of the variance in the population is zero. This can be illustrated with the Monte Carlo simulation in which Model 2 (i.e., H_1) is estimated while Model 3 is the true model (i.e., H_0). A negative estimate of the variance of the growth rate was obtained in 2,529 of the 5,000 samples.[6] Because a negative variance is inadmissible, this has to be corrected by constraining the parameter estimate to the closest value of the admissible parameter space, which is zero,[7] the same value it has under H_0. If the chi-square difference between H_0 and H_1 is computed in each sample of the simulation, it will be equal to zero in approximately 50% of the samples because the likelihood of the model under H_1 is then equal to that under H_0. Consequently, the distribution of the test statistic for testing the variance has half of its mass (i.e., the area under the curve) at the value of zero, and the remaining half follows a $\chi^2(1)$ distribution. In other words, the test statistic follows a .5:.5 mixture of a $\chi^2(0)$ and a $\chi^2(1)$ distribution.

Note that, as a consequence, the common practice of relating the chi-square difference to a $\chi^2(1)$ distribution leads to a relatively conservative test in the analysis of the Bast and Reitsma (1997) data. At a significance level of .05, the critical chi-square in the correct distribution is equal to 2.71, instead of 3.84 in the $\chi^2(1)$ distribution. To emphasize this further, the chi-square difference in the analysis of the true data set is equal to 2.76. In other words, had the correct distribution been used, one would have concluded that the variance of the slope factor was significantly different from zero. Although the analysis as reported by Bast and Reitsma leads to the conclusion that there are no interindividual differences in growth rate of reading comprehension, our results suggest the contrary, namely, that there are differences between individuals in the amount of growth of reading comprehension across time.

In the following sections, we show that the empirical distribution obtained in our simulation study is the correct distribution to test a single variance parameter, and we discuss methods to obtain the parameters of this distribution. The next section provides a technical account of the problem, and the section after that, Asymptotic Distributions of the LR Test Statistic in Specific Situations, provides the correct distributions in commonly encountered cases. The subsection entitled *Testing the Variances of the Growth Parameters in the Latent Growth Curve Model* focuses on cases in which the aim is to test the statistical significance of variances in growth curve models. The subsection entitled *Testing the Latent Correlation in the Common Factor Model* focuses on testing whether a simpler CF model should be preferred to a more complex CF model by means of constraints on factor correlations, and the subsection entitled *Testing a Quasi-Simplex Model Versus a 1-CF Model* focuses on testing the quasi-simplex model against the CF model. The subsequent section provides a worked example of the required steps to obtain the correct results in a given situation. This article concludes with a discussion.

A Formal Description of the Problem

In the case of the linear growth curve model above, it is apparent that the value of variance of the growth rate under H_0 (i.e., the variance is equal to zero) lies on the boundary of the parameter space because a variance cannot assume negative values. To state this more formally, the null hypothesis places the parameter value on the boundary of the parameter space defined by the alternative hypothesis, and this is the reason that the asymptotic distribution of the chi-square difference (henceforth, the LR statistic) is not that of a central chi-square-distributed random variable with one degree of freedom. Because this issue has been neglected in SEM, this section provides a short overview of the underlying theory of the problem of boundaries. For a more technical treatment, we refer the reader to Self and Liang (1987), Shapiro (1985), Stram and Lee (1994, 1995), and the appendix.

To obtain the asymptotic distribution of the LR statistic, several parameter sets need to be defined. These sets represent all possible situations in which parameters may or may not be on the boundary, and they are associated with a specific chi-square distribution. For example, in the case of two boundary parameters, $\boldsymbol{\theta} = (\theta_1, \theta_2)$, the sets of θ_1 and θ_2 may include no parameters on the boundary (e.g., $\theta_1 > 0$ and

[5] For certain classes of models, the problem may or may not be present depending on the context (see Stram & Lee, 1994, p. 1176).

[6] Corresponding to standard model specification in most commonly applied SEM software packages, we did not place any constraint on the variance parameters that may result in negative estimates of variance parameters (so-called inadmissible or improper solutions).

[7] This so-called ad hoc procedure of constraining a negative estimate of a variance to zero when it turns out to be negative in a given sample provides the same results as a model in which the variance parameter is explicitly constrained to be equal to or greater than zero. These models can be easily analyzed with a slight modification in model specification in any SEM software package. Alternatively, a Cholesky decomposition could also be used.

$\theta_2 > 0$), one parameter on the boundary (e.g., $\theta_1 > 0$ and $\theta_2 = 0$, or $\theta_1 = 0$ and $\theta_2 > 0$), or both parameters on boundary (e.g., $\theta_1 = 0$ and $\theta_2 = 0$). Each set is associated with a $\chi^2(2)$, a $\chi^2(1)$, and a $\chi^2(0)$ distribution, respectively. Because it is unknown in practice which set obtains in a given sample, the asymptotic distribution of the LR statistic is a mixture of chi-square distributions, with weights equal to the probability that a specific set is the true set in a particular sample.

If parameters of interest in the model take boundary values under the null hypothesis, the distribution is called the *chi-bar-square* ($\bar{\chi}^2$) distribution and can be represented as follows (Shapiro, 1985, Theorem 3.1):

$$pr(\bar{\chi}^2 \geq c^2) = \sum_{i=0}^{q} w_i pr(\chi^2(i) \geq c^2), \qquad (1)$$

where $\chi^2(i)$ is a chi-square random variable with i degrees of freedom, $\chi^2(0) \equiv 0$ (i.e., a point mass at zero also called a $\chi^2(0)$ distribution), c^2 is a critical value, and w_i are nonnegative weights such that $w_0 + \ldots + w_q = 1$. The chi-bar-square distribution thus depends on the number of boundary parameters in the population because these parameters define the number of sets as described above. With respect to the simulation example of the section Illustration, above, where the interest is in testing whether the variance of the slope is zero, there are two distributions—one in which the parameter is set to zero (in 50% of the samples) and one in which it is positive (in 50% of the samples)—so both $w_0 = .5$ and $w_1 = .5$. If the significance level $\alpha = pr(\bar{\chi}^2 \geq c^2)$ is set to .05, this implies that $\sum_{i=0}^{1} w_i pr(\chi^2(i) \geq c^2) = .05$. If $i = 0$, $pr(\chi^2(0) \geq c^2) = 0$ because zero is the only value that exists in this distribution; if $i = 1$, $pr(\chi^2(1) \geq c^2)$ is equal to the right tail probabilities of the standard $\chi^2(1)$ distribution. So, we have to solve $\sum_{i=0}^{1} w_i pr(\chi^2(i) \geq c^2) = .5 \times 0 + .5 \times pr(\chi^2(1) \geq c^2) = .05$, which subsequently leads to $pr(\chi^2(1) \geq c^2) = .10$ and to a critical value $c^2 = 2.71$.

Often, the null hypothesis may also contain unconstrained parameters of interest as well as nuisance parameters. Nuisance parameters are parameters that are part of the model but do not feature in the test under consideration.[8] For example, in case of a simultaneous test on the variance of the growth rate and the covariance between level and growth rate, the nuisance parameters are the variance of the level and time-specific residuals, and the unconstrained parameter of interest is the covariance. If u is the number of unconstrained parameters of interest, the expression for the distribution starts at $i = u$ rather than $i = 0$. Equation 1 then becomes

$$pr(\bar{\chi}^2 \geq c^2) = \sum_{i=u}^{q} w_i pr(\chi^2(i) \geq c^2). \qquad (2)$$

Instead of a $\chi^2(0)$ distribution, the first term of Equation 2 is a $\chi^2(u)$ distribution.

Determining the Mixture of the Distributions

Suppose that we wish to test a case more specific to SEM concerning a set of elements of the positive semi-definite latent variable covariance matrix $\mathbf{\Psi}$. Let $\mathbf{\Psi}^{ij}$ be a block matrix, and let the hypotheses be

$$H_0\colon \mathbf{\Psi} = \begin{pmatrix} \mathbf{\Psi}^{11} & \\ 0 & 0 \end{pmatrix} \text{ versus } H_1\colon \mathbf{\Psi} = \begin{pmatrix} \mathbf{\Psi}^{11} & \\ \mathbf{\Psi}^{21} & \mathbf{\Psi}^{22} \end{pmatrix}.$$

Suppose, furthermore, that $\mathbf{\Psi}^{11}$ is an $n \times n$ positive definite matrix unconstrained under the null hypothesis and that $\mathbf{\Psi}^{22}$ is a $k \times k$ matrix.

In a test of variances (such as in an LGC model), the diagonal elements of $\mathbf{\Psi}^{22}$ are constrained to zero under H_0. In that case, a set of restrictions should be applied to the elements of $\mathbf{\Psi}^{21}$ and $\mathbf{\Psi}^{22}$ to ensure that $\mathbf{\Psi}$ is positive semi-definite under the alternative hypothesis. The presence of an element equal to zero in the main diagonal of $\mathbf{\Psi}^{22}$ under the alternative hypothesis implies that this element is on the boundary of the parameter space under both the null and the alternative hypotheses. There are $u = kn + \binom{k}{2}$ unconstrained parameters of interest, representing all the elements of $\mathbf{\Psi}^{21}$, the subdiagonal elements of $\mathbf{\Psi}^{22}$, and k parameters of interest whose values may be on the boundary of the parameter space under the null and the alternative hypotheses. Note that the u parameters represent elements of both $\mathbf{\Psi}^{21}$ and $\mathbf{\Psi}^{22}$, and it can be shown that the effect of the u parameters on the asymptotic distribution of the LR vanishes asymptotically (Stram & Lee, 1994, 1995). The resulting asymptotic distribution, in this situation, will thus be a mixture of $k + 1$ chi-square distributions with u, $u + 1, \ldots, u + k$ degrees of freedom.

In the case of a CF model, the situation is slightly different from above because the subdiagonal elements of $\mathbf{\Psi}^{22}$ are constrained under the null hypothesis in such a way that the correlation between the corresponding factors is equal to one. If the corresponding factors have a unit variance, the constraint is simply to set the subdiagonal elements of $\mathbf{\Psi}^{22}$ to one. As before, the only condition under the alternative hypothesis is for $\mathbf{\Psi}$ to be at least positive semi-definite. Having an element equal to one (or minus one) in the subdiagonal of $\mathbf{\Psi}^{22}$ under the alternative hypothesis implies that $\mathbf{\Psi}$ has an element on the boundary of the parameter space under both the null and the alternative hypotheses.

The range of values of the $n(n + 1)/2$ nuisance parameters in $\mathbf{\Psi}^{11}$ is not constrained under either the null or the

[8] In general, the nuisance parameters have no effect on the large sample distribution of the LR statistic, but the unconstrained parameters of interest do have an effect. However, if a nuisance parameter is on the boundary, this may have an effect on the distribution of the LR (see Self & Liang, 1987, Case 8).

alternative hypothesis. In this case, there are $\binom{k}{2}$ parameters in $\mathbf{\Psi}^{22}$ whose values may be on the boundary of the parameter space under the null hypothesis and the alternative hypothesis (i.e., the subdiagonal elements of $\mathbf{\Psi}^{22}$). If u is again the number of unconstrained parameters of interest, again the resulting asymptotic distribution, in this situation, will be a mixture of $\binom{k}{2} + 1$ chi-square distributions with degrees of freedom equal to u, $u + 1, \ldots, u + (k - 1)/2$, respectively.

Determining the Weights

Once the number of distributions and corresponding degrees of freedom of the mixture is known, the probabilities of having exactly $l(l = 0, \ldots, k)$ parameter estimates on the boundary in a particular sample must be computed. These probabilities, mixture proportions, or, henceforth, weights depend on unknown parameter values and on the information matrix of the model under the null hypothesis. Two methods have been developed to compute the weights: analytical derivation and Monte Carlo simulation.

Analytical Derivation

If the number of boundary parameters is four or smaller, analytical methods can be used to compute the weights (Shapiro, 1985). If the number of boundary parameters exceeds four, analytical computation of the weights becomes difficult because the weights can no longer be expressed easily in closed form. We refer the reader to Shapiro (1985) and Self and Liang (1987) for a detailed explanation of analytical computation.

In a number of situations, the analytical derivation of the weights is straightforward. In the commonly encountered case with only one boundary parameter, for example, the weights can be determined fairly easily by making use of the general property stated by Shapiro (1985, p. 141): "Weights of components with an even number of df, as well as weights of components with an odd number of df, always sum up to 1/2." As a consequence, the asymptotic distribution is a .5:.5 mixture of two chi-square distributions. If more than one boundary parameter is involved and the block in the information matrix associated with these parameters is diagonal (i.e., if the parameter estimates are not correlated), the weights of the mixture distribution follow a binomial distribution with the weight of each component equal to

$$\binom{k}{df - n} 2^{-k},$$

where k equals the number of boundary parameters and n equals the number of nuisance parameters. So, for $k = 1$, the weights are .5:.5; for $k = 2$, the weights are .25:.5:.25; for $k = 3$, the weights are .125:.375:.375:.125; and for $k = 4$, the weights are .0625:.25:.375:.25:.0625; and so on. Note also that the general property of Shapiro (1985) holds indeed (e.g., for $k = 4$, $.0625 + .375 + .0625 = .5$). Unfortunately, the binomial distribution of the weights is not very useful in SEM because parameter estimates are often correlated, that is, the relevant block in the information matrix is not a diagonal matrix.

Estimation of Weights by Monte Carlo Simulation

When analytical computation of the weights is difficult, Monte Carlo simulation can be used. Dardanoni and Forcina (1998, p. 1117) proposed a procedure to obtain fairly accurate estimates by means of Monte Carlo simulation. Their procedure involves drawing a great number of parameter vectors from a multivariate normal distribution with mean equal to the hypothesized parameter values and covariance matrix equal to the information matrix under the null hypothesis. These simulated parameter estimates may contain values that lie outside the admissible parameter space. This procedure is sometimes referred to as projecting the simulated values into the parameter space. Practically, this means that negative variances are constrained to zero or that correlations larger than one are constrained to one.

In addition to this procedure, it is also possible to simulate multiple data sets based on a population model and to count the number of times that specific configuration of inadmissible parameter estimates occurs. This last type of Monte Carlo simulation was used above, in the section entitled Illustration. Both simulation procedures result asymptotically in the same estimates of the weights. However, in some situations, the Monte Carlo simulation of data (rather than parameter values) may be preferable because it does not assume that the parameters follow an asymptotic multivariate normal distribution. This assumption may be questionable in the case of small samples.

In the example above, in the section called Illustration, we used Monte Carlo simulation of data and found the variance of the growth rate to be negative in 2,529 of the 5,000 cases. This corresponds to weights of .5058 for the $\chi^2(0)$ distribution and .4942 for the $\chi^2(1)$ distribution. These weights correspond well to the .5:.5 weights by means of the general property of Shapiro (1985).

Asymptotic Distributions of the LR Test Statistic in Specific Situations

In the section entitled A Formal Description of the Problem, above, we showed that the true asymptotic distribution of the LR is a mixture of chi-square distributions (i.e., the chi-bar-square distribution) and that this distribution depends on (a) the number of parameters that are placed on the boundary of the parameter space in the population and (b) the number of unconstrained parameters of interest. With respect to the CF model, the boundary parameters are the correlations between the factors. In the LGC model, the boundary parameters are the variances of the growth factors. With reference to the quasi-simplex model, the boundary parameters are variances of the time-specific residuals.

In this section, we deduce the mixture distributions of the LR statistic for the LGC model, the CF model, and the quasi-simplex model by means of the rules given above. We refer the reader to Bollen (1989) for a detailed treatment of SEM in general.

Testing the Variances of the Growth Parameters in the Latent Growth Curve Model

The LGC model (Laird & Ware, 1982; McArdle, 1986, 1988; Meredith & Tisak, 1984, 1990; Willett & Sayer, 1994) is expressed in Equations 3 and 4:

$$\mathbf{y} = \mathbf{\Lambda}\boldsymbol{\eta} + \boldsymbol{\varepsilon}, \text{ and} \tag{3}$$

$$\boldsymbol{\eta} = \boldsymbol{\alpha} + \boldsymbol{\zeta}, \tag{4}$$

with covariance and mean structure

$$\mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{\Psi}\mathbf{\Lambda}' + \mathbf{\Theta}_{\varepsilon}, \tag{5}$$

$$\boldsymbol{\mu}_{y} = \mathbf{\Lambda}\boldsymbol{\alpha}, \text{ and} \tag{6}$$

$$\boldsymbol{\mu}_{\eta} = \boldsymbol{\alpha}, \tag{7}$$

where $\mathbf{y}$ denotes a $p \times 1$ vector of repeated measurements of the variable $\mathbf{Y}$, $\mathbf{\Lambda}$ is a $p \times q$ matrix of factor loadings, $\boldsymbol{\eta}$ is a $q \times 1$ vector of latent variables, and $\boldsymbol{\varepsilon}$ is a $q \times 1$ vector of residuals. The $q \times q$ matrix $\mathbf{\Psi}$ is the covariance matrix of $\boldsymbol{\eta}$, and the $p \times p$ matrix $\mathbf{\Theta}_{\varepsilon}$ contains the residual variances. In the case of a simple linear LGC model, $\mathbf{\Lambda}$ is a $p \times 2$ matrix constrained in such a way that it contains constants (e.g., 1, 1, 1, 1, . . .) in the first column and known times of measurement (e.g., 0, 1, 2, 3, . . .) in the second column. Vector $\boldsymbol{\alpha}$ in Equation 4 thus contains the population averages of latent growth parameters (level and growth rate), and the random vector $\boldsymbol{\zeta}$ contains the deviations of the individual growth parameters, level and growth rate, from their respective population means. The random vector $\boldsymbol{\varepsilon}$ contains time-specific deviations from the mean growth curve. $\mathbf{\Sigma}$ is the model-implied covariance matrix, $\boldsymbol{\mu}_y$ is the model-implied vector of observed means, and $\boldsymbol{\mu}_\eta$ is a vector of latent means.

Several useful extensions of the basic LGC model have been proposed in the literature. For instance, nonlinear growth can be accommodated by either estimating the factor loadings for the growth rate in $\mathbf{\Lambda}$ (McArdle, 1986; Meredith & Tisak, 1990) or by introducing additional latent variables representing quadratic (or higher order) growth parameters. Other important extensions of the basic LGC model are the combination of two or more growth processes in a multivariate LGC model (MacCallum, Kim, Malarkey, & Kiecolt-Glaser, 1997). In this case, relations of the growth parameters across processes can be modeled. Below, important cases in which boundary parameters affect the asymptotic distribution of the LR statistic of the LGC models are discussed.

Case 1

If the LGC model includes only a level factor, we have a situation in which $n = 0$ and $k = 1$ because the covariance matrix of the latent variables is a positive scalar under H_1 ($\mathbf{\Psi} = \psi_{11}$, i.e., $k = 1$ parameter on the boundary, and $u = 0$ unconstrained parameters of interest). The test of individual differences thus implies the test of hypotheses H_0: $\psi_{11} = 0$ against H_1: $\psi_{11} > 0$.[9]

Distribution. Because k is equal to one, the asymptotic distribution of the LR statistic is a mixture of $k + 1 = 2$ distributions with $u = 0$ and $u + 1 = 1$ degrees of freedom, respectively. In other words, instead of being a $\chi^2(1)$ distribution, the correct asymptotic distribution is a mixture of a point mass at zero and a $\chi^2(1)$ distribution.

Weights. Because there is only one boundary parameter, the general property of Shapiro (1985) can be applied, which results in equal weights of .5:.5. Indeed, Monte Carlo simulation provides a negative estimate of ψ_{11} in 50% of the cases if the number of simulations is large enough. Subsequently projecting the negative estimates of ψ_{11} into the parameter space implies that they are constrained to zero in 50% of the simulations. Because the remaining simulations provide $\psi_{11} > 0$, the resulting weights of the mixture distribution are equal to .5:.5. The asymptotic distribution thus becomes a .5:.5 mixture of a point mass at zero and a $\chi^2(1)$ distribution.

Consequences. Figure 2A presents a graph of both the mixture distribution and the standard $\chi^2(1)$ distribution. For a significance level of $\alpha = .05$, the correct critical value is equal to 2.71—that is, $P(\bar{\chi}^2 \geq 2.71) = .05$ —whereas the corresponding critical value of the $\chi^2(1)$ distribution has a nominal p value of .025—that is, $P(\bar{\chi}^2 \geq 3.84) = .025$. Clearly, the correct asymptotic distribution leads to greater power to reject the null hypothesis if it is not correct. The statistical package R can be used to compute both critical values and the nominal p values. The corresponding scripts can be downloaded from the Web site of *Psychological Methods.*

Case 2

To test individual differences in growth rate in a LGC model, we have $n = 1$ and $k = 1$ because this test implies the presence of $u = 1$ unconstrained parameter of interest (ψ_{21}) and $k = 1$ parameter on the boundary (ψ_{22}) under H_0. The hypotheses to be tested in this case are

[9] See Case 1 of Stram and Lee (1994).

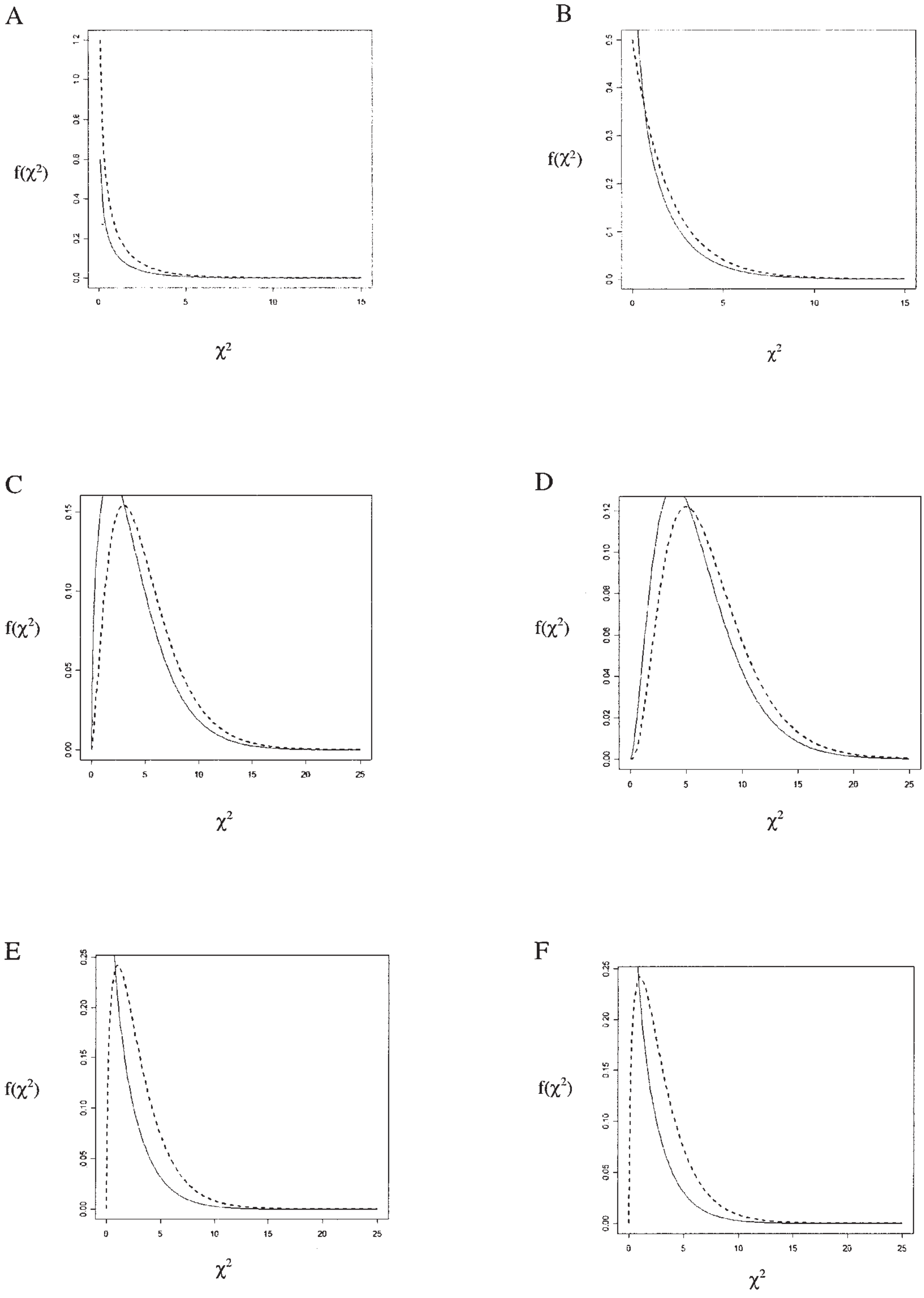

Figure 2. Correct asymptotic distributions of the likelihood ratio with reference to the standard distributions. A: Solid line—.5:.5 mixture of $\chi^2(0)$ and $\chi^2(1)$; dashed line—$\chi^2(1)$ distribution. B: Solid line—.5:.5 mixture of $\chi^2(1)$ and $\chi^2(2)$; dashed line—$\chi^2(2)$ distribution. C: Solid line—.470:.087:.443 mixture of $\chi^2(3)$, $\chi^2(4)$, and $\chi^2(5)$; dashed line—$\chi^2(5)$ distribution. D: Solid line—.243:.506:.251 mixture of a $\chi^2(5)$, $\chi^2(6)$, and $\chi^2(7)$; dashed line—$\chi^2(7)$ distribution. E: Solid line—.209:.288:.291:.212 mixture of $\chi^2(0)$, $\chi^2(1)$, $\chi^2(2)$, and $\chi^2(3)$; dashed line—$\chi^2(3)$ distribution. F: Solid line—.169:.343:.318:.170 mixture of $\chi^2(0)$, $\chi^2(1)$, $\chi^2(2)$, and $\chi^2(3)$; dashed line—$\chi^2(3)$ distribution.

$$H_0\colon \boldsymbol{\Psi} = \begin{bmatrix} \psi_{11} & \\ 0 & 0 \end{bmatrix} \text{ against } H_1\colon \boldsymbol{\Psi} = \begin{bmatrix} \psi_{11} & \\ \psi_{21} & \psi_{22} \end{bmatrix},$$

with $\boldsymbol{\Psi}$ being positive semi-definite.[10]

Distribution. The correct asymptotic distribution of the LR statistic is a mixture of $k + 1 = 2$ distributions in this case with $u = 1$ and $u + 1 = 2$ degrees of freedom, respectively, instead of the standard $\chi^2(2)$ distribution.

Weights. Because there is only one boundary parameter, the weights are the same as in Case 1 (i.e., .5:.5), and the asymptotic distribution is a .5:.5 mixture of a $\chi^2(1)$ and a $\chi^2(2)$ distribution, respectively.

Consequences. Figure 2B presents a graph of both the mixture distribution and the $\chi^2(2)$ distribution. The correct critical value is equal to 5.14 ($\alpha = .05$)—that is, $P(\bar{\chi}^2 \geq 5.14) = .05$—whereas the corresponding critical value of the $\chi^2(2)$ distribution has a nominal p value of .032—that is, $P(\bar{\chi}^2 \geq 5.99) = .032$.

Case 3

A simultaneous test of individual differences in both linear and quadratic growth implies $n = 1$ and $k = 2$, as well as hypotheses

$$H_0\colon \boldsymbol{\Psi} = \begin{bmatrix} \psi_{11} & & \\ 0 & 0 & \\ 0 & 0 & 0 \end{bmatrix} \text{ against}$$

$$H_1\colon \boldsymbol{\Psi} = \begin{bmatrix} \psi_{11} & & \\ \psi_{21} & \psi_{22} & \\ \psi_{31} & \psi_{32} & \psi_{33} \end{bmatrix},$$

with $\boldsymbol{\Psi}$ being positive semi-definite.

Distribution. Given $u = 3$ unconstrained parameters of interest (ψ_{21}, ψ_{31}, and ψ_{32}) and $k = 2$ parameters on the boundary, the correct asymptotic distribution of the LR statistic is a mixture of $\chi^2(3)$, $\chi^2(4)$, and $\chi^2(5)$ distributions, instead of a $\chi^2(5)$ distribution, because there can be 0, 1, or 2 parameters on the boundary.

Weights. The weights depend on the data and can be calculated by means of analytical computation (e.g., integrating the multivariate normal distribution of the parameter estimates based on the correlations between the parameter estimates) or by means of Monte Carlo simulation for each specific situation.

Consequences. The critical values depend on the information matrix and thus on the covariances between the parameters. To illustrate how such a mixture distribution might look in a specific sample, Figure 2C presents a graph of the .470:.087:.443 mixture of $\chi^2(3)$, $\chi^2(4)$, and $\chi^2(5)$ distributions, respectively, as well as the $\chi^2(5)$ distribution obtained on an analysis of a quadratic LGC on the Curran (1997a) data (see A Worked Example, below). The critical value in the correct distribution is equal to 9.85 ($\alpha = .05$), compared with 11.07 in the standard $\chi^2(5)$ distribution.

Case 4

In the case of a simultaneous test of individual differences in growth in two processes, $n = 2$, and $k = 2$, so there are two parameters on the boundary and $u = 5$ unconstrained parameters of interest (ψ_{21}, ψ_{32}, ψ_{41}, ψ_{42}, and ψ_{43}). The hypotheses to be tested are

$$H_0\colon \boldsymbol{\Psi} = \begin{bmatrix} \psi_{11} & & & \\ 0 & 0 & & \\ \psi_{31} & 0 & \psi_{33} & \\ 0 & 0 & 0 & 0 \end{bmatrix} \text{ against}$$

$$H_1\colon \boldsymbol{\Psi} = \begin{bmatrix} \psi_{11} & & & \\ \psi_{21} & \psi_{22} & & \\ \psi_{31} & \psi_{32} & \psi_{33} & \\ \psi_{41} & \psi_{42} & \psi_{43} & \psi_{44} \end{bmatrix},$$

with $\boldsymbol{\Psi}$ being positive semi-definite.

Distribution. The correct asymptotic distribution of the LR statistic is therefore a mixture of $k + 1 = 3$ chi-square distributions with $u = 5$, $u + 1 = 6$, and $u + 2 = 7$ degrees of freedom, respectively, instead of the standard $\chi^2(7)$ distribution.

Weights. The weights and thus the consequences depend on the data (see Case 3). Figure 2D presents a graph of the .243:.506:.251 mixture of $\chi^2(5)$, $\chi^2(6)$, and $\chi^2(7)$ distributions and the $\chi^2(7)$ distribution obtained on an analysis of a bivariate LGC on the Curran (1997a) data. The critical value in the correct distribution is equal to 12.74 ($\alpha = .05$), compared with 14.07 in the standard $\chi^2(7)$ distribution.

Testing the Latent Correlation in the Common Factor Model

The CF model is obtained by

$$\mathbf{y} = \boldsymbol{\tau} + \boldsymbol{\Lambda\eta} + \boldsymbol{\varepsilon}, \text{ and} \tag{8}$$

$$\boldsymbol{\eta} = \boldsymbol{\alpha} + \boldsymbol{\zeta}, \tag{9}$$

with covariance and mean structure

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda\Psi\Lambda}' + \boldsymbol{\Theta}_{\varepsilon}, \tag{10}$$

$$\boldsymbol{\mu}_{y} = \boldsymbol{\tau} + \boldsymbol{\Lambda\alpha}, \text{ and} \tag{11}$$

$$\boldsymbol{\mu}_{\eta} = \boldsymbol{\alpha}, \tag{12}$$

and we define an m-CF model to contain m common factors. Without further restrictions, the mean structure and covariance structure implied by the model cannot be identified. Identification of the mean structure may be obtained either by constraining the latent variable means $\boldsymbol{\alpha}$ to zero or by constraining a different intercept in vector $\boldsymbol{\tau}$ to zero for each

[10] See Case 2 of Stram and Lee (1994).

latent variable in the model. Identification of the covariance structure may be obtained either by constraining the latent variable variances in $\boldsymbol{\Psi}$ to one or by constraining a factor-loading in $\boldsymbol{\Lambda}$ for each factor to one. Because the mean structure is usually not of interest in this type of single-group CF model, we arbitrarily obtain identification by constraining the latent means $\boldsymbol{\alpha}$ to zero. To keep the subsequent discussion on parameter bounds simple, we obtain identification of the covariance structure by constraining the latent factor variances to one. This constraint implies $\boldsymbol{\Psi}$ to be a $q \times q$ correlation matrix.

Given a 2-CF model, for instance, it may be of interest to determine whether a 1-CF model might have been sufficient to explain the relations among the observed variables. This may be done by testing whether the latent factor correlation is significantly smaller than one. Implicit in this test is the assumption that a 1-CF model is statistically equivalent to a 2-CF model with a perfect correlation between the factors. Under appropriate identification conditions, Van der Sluis et al. (2005) showed that a q-CF model with k independent factor correlations constrained to one is statistically equivalent, in general, to a $(q - k)$-CF model, and they provided a detailed exposition of the appropriate constraints to specify such models. Another example of a commonly encountered test situation is the test of a 3-CF model against a 1-CF or 2-CF model.

Case 5

In the case of a test of a 2-CF model against a 1-CF model, there are $u = 0$ unconstrained parameters of interest and $k(k - 1)/2 = 1$ boundary parameter. The hypotheses to be tested are

$$H_0\text{: } \boldsymbol{\Psi} = \begin{bmatrix} 1 & \\ 1 & 1 \end{bmatrix} \text{ against } H_1\text{: } \boldsymbol{\Psi} = \begin{bmatrix} 1 & \\ \psi_{21} & 1 \end{bmatrix},$$

with $\boldsymbol{\Psi}$ being positive semi-definite (i.e., H_0: $\psi_{21} = 1$ against H_1: $-1 \leq \psi_{21} < 1$; see Van der Sluis et al., 2005).

Distribution. The correct asymptotic distribution is a mixture of a point mass at zero and a $\chi^2(u + 1 = 1)$, instead of a $\chi^2(1)$, distribution.

Weights. The weights and consequences are the same as in Case 1.

Case 6

In the case of a test of a 3-CF versus a 1-CF model, the situation contains $k(k - 1)/2 = 3$ boundary parameters and $u = 0$ unconstrained parameters of interest. In other words, there can be 3, 2, 1, or 0 parameters on the boundary under H_1. The hypotheses to be tested are

$$H_0\text{: } \boldsymbol{\Psi} = \begin{bmatrix} 1 & & \\ 1 & 1 & \\ 1 & 1 & 1 \end{bmatrix} \text{ against } H_1\text{: } \boldsymbol{\Psi} = \begin{bmatrix} 1 & & \\ \psi_{21} & 1 & \\ \psi_{31} & \psi_{32} & 1 \end{bmatrix}$$

with $\boldsymbol{\Psi}$ again being positive semi-definite (i.e., H_0: $\psi_{21} = 1$, $\psi_{31} = 1$, $\psi_{32} = 1$, against H_1: $-1 \leq \psi_{21}, \psi_{31}, \psi_{32} < 1$).

Distribution. The correct asymptotic distribution is a mixture of $k(k - 1)/2 + 1 = 4$ distributions with degrees of freedom equal to $u = 0$, $u + 1 = 1$, $u + 2 = 2$, and $u + 3 = 3$.

Weights. The weights and thus the consequences depend on the data (see Case 3). Figure 2E presents a graph of a .209:.288:.291:.212 mixture of $\chi^2(0)$, $\chi^2(1)$, $\chi^2(2)$, and $\chi^2(3)$ distributions and the $\chi^2(3)$ distribution obtained from a test of a 3-CF versus a 1-CF model on a covariance matrix provided in the article by Miyake, Friedman, Emerson, Witzki, and Howerter (2000). The critical value in the correct distribution is equal to 5.64 ($\alpha = .05$), compared with 7.81 in the standard $\chi^2(3)$ distribution.

Case 7

In the case of a test of a 3-CF versus a 2-CF model, the situation is more complicated than in Case 6 because the restriction of the correlation between one pair of factors to the value of one under the H_0 has restrictive implications for the correlation of each of these two factors with the third factor: The two correlations should be constrained to be equal (cf. Van der Sluis et al., 2005). In this case, there is $k(k - 1)/2 = 1$ boundary parameter, and because of the equality constraint on two correlations, there is $u = nk/2 = 1$ unconstrained parameter of interest. The hypotheses to be tested are

$$H_0\text{: } \boldsymbol{\Psi} = \begin{bmatrix} 1 & & \\ \psi_{21} & 1 & \\ \psi_{21} & 1 & 1 \end{bmatrix} \text{ against}$$

$$H_1\text{: } \boldsymbol{\Psi} = \begin{bmatrix} 1 & & \\ \psi_{21} & 1 & \\ \psi_{31} & \psi_{32} & 1 \end{bmatrix},$$

with $\boldsymbol{\Psi}$ being positive semi-definite (i.e., H_0: $\psi_{21} = 1$, $\psi_{31} = \psi_{32}$, against H_1: $-1 \leq \psi_{21}, \psi_{31}, \psi_{32} < 1$).

Distribution. Given $u = 1$ unconstrained parameter of interest and one boundary parameter, the correct distribution of the LR statistic is a mixture of a $\chi^2(1)$ and a $\chi^2(2)$ distribution.

Weights. The weights and consequences are identical to those of Case 2.

Testing a Quasi-Simplex Model Versus a 1-CF Model

The quasi-simplex model (Jöreskog, 1970) represents a covariance structure generated by a nonstationary first-order autoregressive process that can be obtained by Equations 13 and 14.

$$\eta_{t+1} = \beta_{t+1,t}\eta_t + \zeta_{t+1}, \text{ and} \tag{13}$$

$$y_t = \lambda_t \eta_t + \varepsilon_t, \tag{14}$$

with covariance structure

$$\mathbf{\Sigma} = \mathbf{\Lambda}(\mathbf{I} - \mathbf{B})^{-1}\mathbf{\Psi}(\mathbf{I} - \mathbf{B})^{-1'}\mathbf{\Lambda}' + \mathbf{\Theta}_\varepsilon. \tag{15}$$

Here, η_t is a latent variable at time t, $\beta_{t+1,t}$ represents the regression coefficient between adjacent occasions $t + 1$ and t, and ζ_{t+1} is the normally distributed equation residual added to the model at each occasion. Furthermore, λ_t is the factor loading that links the latent variable η_t to the observed variable y_t, and it is fixed to 1.0 in case of single indicators. The measurement errors, ε_t, are independent and normally distributed.

Given the quasi-simplex model, it may be of interest to test whether a 1-CF model might have been sufficient to explain the relations among the observed variables. Rovine and Molenaar (2005) have shown that such a test may be performed by constraining the residual variances at all time points next to the initial time point, $t = 1$, at zero. They showed that η_t is then reduced to a single η, whereas the $\beta_{t+1,t}$ coefficients are absorbed into the factor loadings.

Case 8

In the case of a test of a simplex model against a 1-CF model, there is $u = 1$ unconstrained parameter of interest, because the residual at the first occasions is not constrained, and $k = t - 1$ boundary parameters. The hypotheses to be tested are

$$H_0\text{: } \mathbf{\Psi} = \begin{bmatrix} \psi_{11} & & & \\ 0 & 0 & & \\ \cdots & \cdots & \cdots & \\ 0 & 0 & 0 & 0 \end{bmatrix} \text{ against}$$

$$H_1\text{: } \mathbf{\Psi} = \begin{bmatrix} \psi_{11} & & & \\ 0 & \psi_{22} & & \\ \cdots & \cdots & \cdots & \\ 0 & 0 & 0 & \psi_{tt} \end{bmatrix},$$

with $\mathbf{\Psi}$ being positive semi-definite.

Distribution. Because $k = t - 1$, the asymptotic distribution of the LR statistic is a mixture of t distributions with $u = 0$, $u + 1 = 1, \ldots, u + k = k$ degrees of freedom, respectively. In other words, instead of being a $\chi^2(k)$ distribution, the correct asymptotic distribution is a mixture of a point mass at zero, that is, $\chi^2(0)$, and $\chi^2(1)$, $\chi^2(2)$, . . . , $\chi^2(k)$ distributions.

Weights. The weights and thus the consequences depend on the data (see Case 3). Figure 2F presents a graph of the .169:.343:.318:.170 mixture of $\chi^2(0)$, $\chi^2(1)$, $\chi^2(2)$, and $\chi^2(3)$ distributions that is obtained below, in the section entitled A Worked Example, in testing a quasi-simplex model with four occasions against a 1-CF model. The critical value in the correct distribution is equal to 5.52 (α = .05), compared with 7.81 in the standard $\chi^2(3)$ distribution. Table 2 provides a short overview of values of k and u, the number of distributions, and the consequences for each of the eight cases described in this section.

A Worked Example

To illustrate the different steps one should take if a null hypothesis places the value of the parameter on the boundary of the parameter space, we analyze data taken from the National Longitudinal Survey of Youth of Labor Market Experience in Youth, a study initiated in 1979 by the U.S. Department of Labor to examine the transition of young people into the labor force. The data were collected using face-to-face interviews of both mother and child taken in 2-year intervals between 1986 and 1992, making the measurement unit of time equal to 2 years. A detailed description of the data can be found in Baker, Keck, Mott, and Quinlan (1993) and in Curran (1997a). The measurements in the present example are from a battery of assessments by Curran, who presented the complete data of 261 children on four consecutive measures of children's antisocial behavior, reading ability, and one (time-invariant) measure of the degree of cognitive stimulation provided to the child at home.[11] In this example, we use the data on antisocial behavior.

Assume that we are interested in testing, first, whether a quasi-simplex model fits the data well and, second, whether, if this model holds, it can be constrained to a 1-CF model by constraining the residual variances to zero (see Rovine & Molenaar, 2005). This test thus corresponds to Case 8 above with $t = 4$ measurement occasions.

The first step consists of testing whether a quasi-simplex model fits the data. To identify the model, the residuals are constrained to be equal. The parameter estimates of this model are presented in the second column of Table 3; the quasi-simplex model fits the data quite well. The second step is to estimate the 1-CF model. This can easily be performed by constraining the residual variances of the quasi-simplex model to zero (Rovine & Molenaar, 2005). This constrained quasi-simplex model seems to fit the data also quite well, and the LR statistic (i.e., $\Delta\chi^2$) is equal to 6.09.

Standard practice of comparing the LR statistic with a chi-square distribution with three degrees of freedom with a critical value of 7.81 (α = .05) would not lead to a rejection of the null hypothesis that the residual variances are equal to zero. In other words, we would conclude that the 1-CF model should not be rejected and that subsequent substantive conclusions should be based on this model.

In the prior sections, we have shown, however, that the asymptotic distribution of the LR statistic is not the chi-

[11] The data can be found in Curran (1997b).

Table 2
Values of k and u, the Number of Distributions, and the Consequences for Each of the Eight Cases

Case	No. of boundary parameters (k)	No. of unconstrained parameters of interest (u)	No. of distributions in the mixture	No. of *df*	Mixture distribution	Consequence
1	1	0	$k + 1 = 2$	$u = 0$ *df* and $u + 1 = 1$ *df*	$.5\ \chi^2(0) + .5\ \chi^2(1)$	$P(\chi^2 \geq 2.71) = .05$
2	1	1	$k + 1 = 2$	$u = 1$ *df* and $u + 1 = 2$ *df*	$.5\ \chi^2(1) + .5\ \chi^2(2)$	$P(\chi^2 \geq 5.14) = .05$
3	2	3	$k + 1 = 3$	$u = 3$ *df*, $u + 1 = 4$ *df*, and $u + 2 = 5$ *df*	Weights depend on the data	R script
4	2	5	$k + 1 = 3$	$u = 5$ *df*, $u + 1 = 6$ *df*, and $u + 2 = 7$ *df*	Weights depend on the data	R script
5	1	0	$k(k - 1)/2 + 1 = 2$	$u = 0$ *df* and $u + 1 = 1$ *df*	$.5\ \chi^2(0) + .5\ \chi^2(1)$	$P(\chi^2 \geq 2.71) = .05$
6	3	0	$k(k - 1)/2 + 1 = 4$	$u = 0$ *df*, $u + 1 = 1$ *df*, $u + 2 = 2$ *df*, and $u + 3 = 3$ *df*	Weights depend on the data	R script
7	1	1	$k(k - 1)/2 + 1 = 2$	$u = 1$ *df* and $u + 1 = 2$ *df*	$.5\ \chi^2(1) + .5\ \chi^2(2)$	$P(\chi^2 \geq 5.14) = .05$
8	$t - 1$	1	$t - 1 + 1 = t$	$u = 0$ *df*, $u + 1 = 1$ *df*, . . . , $u + k = k$ *df*	Weights depend on the data	R script

Table 3
Parameter Estimates of Quasi-Simplex Model Using the Data of Curran (1997b)

Parameter	Quasi-simplex model	Constrained quasi-simplex model
$\beta_{2,1}$	1.022 (.167)	1.409 (.156)
$\beta_{3,2}$	0.924 (.093)	1.035 (.092)
$\beta_{4,3}$	1.167 (.100)	1.208 (.097)
ψ_{11}	1.370 (.278)	0.924 (.185)
ψ_{22}	0.653 (.308)	0
ψ_{33}	0.343 (.196)	0
ψ_{44}	0.298 (.310)	0
$\varepsilon_1 = \varepsilon_2 = \varepsilon_3 = \varepsilon_4$	1.326 (.147)	1.621 (.082)
$\chi^2(df)$	1.094 (2), $p = .58$	7.193 (5), $p = .21$
$\Delta\chi^2(df)$		6.099 (3)

Note. These models were refitted to the covariance matrix only. $N = 261$. Standard errors are presented in parentheses.

square distribution with three degrees of freedom because the null hypothesis places the values of three parameters on the boundary of the parameter space. In this example, the distribution of the LR statistic can be obtained by using Case 8. Because this model consists of $t = 4$ measurement occasions, there are $k = 3$ boundary parameters and no unconstrained parameters of interest, and the asymptotic distribution is consequently a mixture of a point mass at zero and $\chi^2(1)$, $\chi^2(2)$, and $\chi^2(3)$ distributions.

To compute the weights, we used Monte Carlo simulation to generate 10,000 samples with the parameter estimates of the constrained quasi-simplex model (H_0) as the population values. Subsequently, the quasi-simplex model (H_1) was estimated in each data set, and because the parameter estimates were saved in a separate file, the weights could be obtained easily by counting the number of parameters that were negative in each sample. All three residual variances were negative, for instance, in 16.9% of the cases, leading to a weight of .169 for the point mass of zero. The other weights were .343, .318, and .170 for the $\chi^2(1)$, $\chi^2(2)$, and $\chi^2(3)$ distributions, respectively. So, the true asymptotic distribution of the LR statistic in this situation is a .169:.343:.318:.170 mixture of $\chi^2(0)$, $\chi^2(1)$, $\chi^2(2)$, and $\chi^2(3)$ distributions. The critical value of this distribution at a significance level of $\alpha = .05$ is equal to 5.52. The p value of the LR test statistic is equal to $P(\bar{\chi}^2 > 6.09) = .038$. In other words, the null hypothesis of zero residual variances (the 1-CF model) should have been rejected if the correct asymptotic distribution of the LR statistic had been used.

Because sample size was small, we preferred Monte Carlo simulation of data to obtain the weights. In principle, however, Monte Carlo simulation of parameters as proposed by Dardanoni and Forcina (1998) could also be used. To obtain the weights using this procedure, the quasi-simplex model was fitted to the data, and the correlation matrix of parameters (i.e., the information matrix) was used to simulate 10,000 parameter vectors. Subsequently, counting the times that ψ_{22}, ψ_{33}, and ψ_{44} were less than zero resulted in weights of .167, .336, .333, and .165 for the $\chi^2(0)$, $\chi^2(1)$, $\chi^2(2)$, and $\chi^2(3)$ distributions, respectively. These weights do not differ significantly from the weights obtained with Monte Carlo simulation of data as tested by the standard chi-square test ($p = .17$). As a consequence, if we compare the critical values at several significance levels and the p value of the LR test statistic—$P(\bar{\chi}^2 > 6.09)$—the differences appear not to be that large (see Table 4).

A general procedure of the subsequent steps to be taken if the null hypothesis places the value of the parameter on the boundary of the parameter space is presented in Table 5. With the information and explanation provided in this article, this figure will be of help in obtaining the correct distribution of the LR statistic in any situation.

Discussion

The LR test, or chi-square-difference test, is commonly applied in SEM for testing specific parameters or sets of parameters. In this article, we have shown, on the basis of the theory of Shapiro (1985) and Self and Liang (1987), that there are several common test situations in social research in which the standard theory underlying the LR statistic does not hold. In particular, the standard theory does not hold in SEM if the null hypothesis places values of parameters on the boundary of the parameter space, as may occur in testing variances (e.g., in the LGC model or quasi-simplex model) and in testing correlations (e.g., in the CF model). We have discussed several important cases of boundary problems in SEM, and a general procedure to obtain the correct asymptotic distribution of the LR statistic has been provided.

It is important to use the correct distribution. If the standard chi-square distribution is incorrect, this will result in too conservative a test. To illustrate this, consider Figures 2A and 2B. These figures show that the true distributions of the test statistic in testing variances or correlations have a lighter tail than the standard distributions that are commonly applied. If these standard but incorrect distributions are used to obtain p values or critical values, there will be a higher probability of committing a Type II error (i.e., a failure to reject the incorrect null hypotheses). Consequently, the statistical power of the standard distributions will be too low. This can be verified easily by comparing the critical values of the chi-bar-square distribution for Cases 1–8 with those of the standard chi-square distributions. The critical values of the chi-bar-square distribution are lower in all cases.

In the simple case where the information matrix of the relevant parameters is diagonal (this includes the special case of a single parameter as the factor correlation in the 2-CF model), the weights of the mixture distribution can be obtained by means of the binomial expansion. Unfortunately, parameter estimates are rarely uncorrelated. Generally, the weights may be obtained by means of analytical derivation or by means of Monte Carlo simulation. Of these methods, the latter is more practical because it can be applied regardless of the number of components in the mixture distributions.

We have also discussed two methods of Monte Carlo simulation, namely, the simulation of parameter values (Dardanoni & Forcina, 1998) and the actual simulation of data. The former is simple to implement, but it is based on the assumption of multivariate normality of the parameter estimates. The latter is based on the assumption of multivariate normality of the data and does not include the assumption that the parameter estimates are normally distributed (although, given a large *N,* this is likely to be the case). For another use of data simulation in the context of SEM, we refer the reader to Muthén and Muthén (2002), who applied data simulation in the context of power calculations given missing data.

It is difficult to make general statements concerning the bias incurred in using the incorrect distribution because this depends on the specific values of the weights and the weights depend on unknown parameter values and on the information matrix of the model under the null hypothesis. In Case 3, for instance, the correct asymptotic distribution of the LR statistic is a $w_3{:}w_4{:}w_5$ mixture of $\chi^2(3)$, $\chi^2(4)$, and $\chi^2(5)$ distributions, instead of a $\chi^2(5)$ distribution. If w_5 approaches one, however, in an extreme situation, then w_3 and w_4 must approach zero, and there will be no bias

Table 4
Comparison of the Weights, Critical Values, and p Values

	Monte Carlo simulation of data	Monte Carlo simulation of parameters
Weights of the χ^2 distribution with 0, 1, 2, and 3 *df*	.169:.343:.318:.170	.167:.336:.333:.165
$P(\chi^2 > 6.09)$	.038	.037
$P(\chi^2 > c) = .10$	$c = 4.06$	$c = 4.05$
$P(\chi^2 > c) = .05$	$c = 5.52$	$c = 5.48$
$P(\chi^2 > c) = .01$	$c = 8.88$	$c = 8.68$

Note. The differences between the weights are not significant at an alpha level of .05 ($p = .17$).

Table 5
Steps to Be Taken in Deriving the Asymptotic Distribution of the Likelihood Ratio (LR) Statistic

Step	
1.	Estimate the model under the alternative hypothesis (H_1), and estimate the model under the null hypothesis (H_0).
2.	Compute the LR statistic of testing H_0 against H_1.
3.	Decide whether this test contains boundary parameters.
4.	Determine the number of boundary parameters and unconstrained parameters of interest and, subsequently, the number of degrees of freedom of the mixture distribution.
5.	Simulate 10,000 replicate datasets under the null hypothesis (H_0).
6.	Estimate H_1 in each replicate.
7.	Compute the weights by counting the number of boundary parameters that hit the boundary in each replicate analysis.
8.	Compute the *p* value of the LR test statistic, or compute critical values.
9.	Perform the hypothesis test.

because the mixture distribution is equivalent to a $\chi^2(5)$ distribution. The smaller w_5 becomes, the larger the bias will be.

As the true asymptotic distribution of the LR statistic in a test containing boundary parameters, the chi-bar-square distribution could be said to have greater statistical power compared with the traditional distribution of the LR statistic with degrees of freedom equal to differences in the number of constraints. Strictly speaking, though, the power is not greater but, rather, simply more accurate. However, because it has never been applied in the practice of SEM before, tests of interindividual differences in the growth parameters of LGC models are likely to have been too conservative in the past. Stated otherwise, it may have been concluded too often that no interindividual differences existed where they truly did exist. Similarly, the standard LR test is biased toward the 1-CF model compared with a quasi-simplex model or a more complex CF model. The world is likely to be more complex than standard LR tests have suggested. It must be stressed that these conclusions are based on the assumptions of perfectly multivariate normally distributed data and large enough sample sizes, conditions that are often violated in practice. Under such nonidealized conditions, the LR is not generally chi-square distributed, which may also render the chi-bar-square distribution inappropriate.

Several issues remain to be discussed at this point. First, an important issue that may pop up is whether it makes a difference if a boundary constraint is implicit in the comparison of two models versus when it is explicitly imposed by means of an inequality constraint. An implicit constraint may result in an inadmissible solution (e.g., it may provide a negative estimate of a variance parameter), and one has to constrain such a parameter estimate to the closest value of the admissible parameter space (the ad hoc approach) and reestimate the model before interpretation of its parameters. On the other hand, if an inequality constraint is modeled explicitly, the parameter estimate will always be within the parameter space. In practice, both approaches provide approximately the same parameter estimates, and therefore, the value of the LR statistic for testing the specific parameter will also be the same. Most software packages, such as Mplus (Muthén & Muthén, 2004) and LISREL (Jöreskog & Sörboom, 2001; with `ad=off`), do allow for negative variances, whereas other packages do not allow for negative variances by default. If a package allows for negative estimates of a variance parameter, this information can be easily used in a Monte Carlo simulation of data to estimate the weights of the chi-bar-square distribution, as we have illustrated in the prior sections. Of course, the weights can also be obtained by Monte Carlo simulation of parameters based on the information matrix. Some software packages (e.g., the MIXED procedure of SAS) stop running when an inadmissible parameter estimate is encountered during estimation. The consequence of this is that the asymptotic distribution is somewhat different because the user is forced to simplify the model by dropping the inadmissible estimate and all associated parameters from the model (Stram & Lee, 1994, p. 1176). In our Case 2, for instance, the asymptotic distribution would be a .5:.5 mixture of a $\chi^2(0)$ and a $\chi^2(2)$ distribution because both the variance and the covariance are dropped.

A second issue is that the problem of boundary parameters is not specific to the LR test but may also affect other tests. Because of their asymptotic equivalence to the LR test, the Wald test and score tests will suffer the same problem if boundary parameters are present. Also, as noted by Dominicus et al. (2006), other goodness-of-fit measures, such as the Akaike information criterion (AIC; Akaike, 1987), the Bayesian information criterion (BIC; Schwartz, 1978), and the root-mean-square error of approximation (Browne & Cudeck, 1992), will be affected. AIC and BIC cannot be computed because the number of degrees of freedom in the model is not known. How these fit measures are affected precisely, as well as the relationship with conditional testing (Moreira, 2003), will be the topic of future work.

Third, it is important to note that the use of the Lagrangian multiplier, or modification index (MI), also needs a slight adaptation for boundary parameters. It is very important to always use the expected parameter change (EPC) together with an MI. If the EPC is such that the value of the parameter falls outside the admissible parameter space (e.g., a negative EPC in the case of a variance parameter), the MI provided by the program is not meaningful and should be set to zero. If the EPC indicates an admissible change, the MI is meaningful but should be related to a .5:.5 mixture of a $\chi^2(0)$ and a $\chi^2(1)$ distribution (corresponding to Case 1 of Section 3), with a critical value of 2.71 instead of 3.84 (α = .05).

Fourth, we have focused here on the asymptotic distribution of the LR for testing one or more boundary parameters and not on the LR as a test of overall goodness of fit. The reason for this is the following. In the case of a test of overall model fit, the null hypothesis represents the model that we have posited and states, among other things, that all variance parameters are greater than or equal to zero. However, we do not know the true value of the parameters under the null hypothesis, and the theory of Self and Liang (1987) and Stram and Lee (1994, 1995) does not apply. The value of LR for overall goodness of fit will be chi-bar-square distributed, but the nature of this distribution is not known and cannot be inferred easily by means of standard theory. In the commonly encountered situation of a test of overall model fit, our advice is to be conservative by constraining parameter estimates that lie outside the admissible parameter space by means of the ad hoc approach, and to test the overall goodness of fit of this constrained model with the degrees of freedom of the unconstrained model. Alternatively, the parametric bootstrap (Bollen & Stine, 1993) could be used for overall model fit testing. In a simulation study, Galindo-Garre and Vermunt (2004) showed that the parametric bootstrap provides correct *p* values for log-linear models with explicit inequality constraints. The issue of overall model fit with inequality constraints calls for further research.

The general issue discussed in this article is not isolated but is part of a larger family of order-constrained inferences. In general, inequality constraints make a difference in every aspect of statistical testing (Geyer, 1995). Many procedures thought to be well understood become problematic when inequality constraints are introduced. Fortunately, sophisticated statistical inference has been developed to address these problems. Whereas this methodology should be commonplace, it is rarely used in practice (Iverson, 2005). Here, we have shown how inequality constraints make a difference in testing relatively standard structural equation models. These models contained inequality constraints from the start, albeit this fact has not been commonly acknowledged. Apart from this, we believe that inequality constraints may be of great utility in SEM in general because such constraints may express prior information explicitly (see Klugkist, Laudy, & Hoijtink, 2005). In factor analysis, for instance, factor loadings may be required to be nonnegative or may be constrained to be greater than others on the basis of the item content. Including such information in SEM may lead to appreciable increases in statistical power.

References

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika, 52,* 317–332.

Azzalini, A. (1996). *Statistical inference: Based on the likelihood.* London: Chapman & Hall.

Baker, P. C., Keck, C. K., Mott, F. L., & Quinlan, S. V. (1993). *NLSY child handbook: A guide to the 1986–1990 National Longitudinal Survey of Youth child data.* Columbus, OH: Center for Human Resource Research.

Bast, J., & Reitsma, P. (1997). Matthew effects in reading: A comparison of latent growth curve models and simplex models with structured means. *Multivariate Behavioral Research, 32,* 135–167.

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

Bollen, K. A., & Stine, R. A. (1993). Bootstrapping goodness-of-fit measures in structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 111–135). Newbury Park, CA: Sage.

Browne, M. W., & Arminger, G. (1995). Specification and estimation of mean- and covariance-structure models. In G. Arminger, C. C. Clogg, & M. E. Sobal (Eds.), *Handbook of statistical modeling for the social and behavioral sciences* (pp. 185–249). New York: Plenum Press.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research, 21,* 230–258.

Carey, G. C. (2005). *Cholesky problems.* Unpublished manuscript. Retrieved June 14, 2005, from http://ibgwww.colorado.edu~carey/CholeskyProblems2004–14-07.pdf

Chernoff, H. (1954). On the distribution of the likelihood ratio. *Annals of Mathematical Statistics, 25,* 573–578.

Crainiceanu, C. M., & Ruppert, D. (2004). Likelihood ratio tests in linear mixed models with one variance component. *Journal of the Royal Statistical Society, Series B, 66,* 165–185.

Curran, P. J. (1997a, April). The bridging of quantitative methodology and applied developmental research. In P. J. Curran (Chair), *Comparing three modern approaches to longitudinal data analysis.* Symposium conducted at the meeting of the Society for Research in Child Development, Washington, DC.

Curran, P. J. (1997b). *Growth curve example.* Retrieved October 12, 2006, from http://www.unc.edu/~curran/example.htm

Dardanoni, V., & Forcina, A. (1998). A unified approach to likelihood inference on stochastic ordering in a nonparametric context. *Journal of the American Statistical Association, 93,* 1112–1123.

Dominicus, A., Skrondal, A., Gjessing, H. K., Pedersen, N. L., & Palmgren, J. (2006). Likelihood ratio tests in behavioral genetics: Problems and solutions. *Behavior Genetics, 36,* 331–340.

Galindo-Garre, F., & Vermunt, J. K. (2004). The order-restricted association model: Two estimation algorithms and issues in testing. *Psychometrika, 69,* 641–654.

Geyer, C. J. (1995). *Likelihood ratio tests and inequality constraints* (Tech. Rep. No. 610). Minneapolis: University of Minnesota, Twin Cities Campus, School of Statistics.

Gonzalez, R., & Griffin, D. (2001). Testing parameters in structural equation modeling: Every "one" matters. *Psychological Methods, 3,* 258–269.

Iverson, G. (2005, July). *Order-restricted inference and Bayesian model selection illustrated via Tversky's study of intransitive preference.* Paper presented at the annual meeting of the Psychometric Society, Tilburg, the Netherlands.

Jöreskog, K. G. (1970). Estimation and testing of simplex models. *British Journal of Mathematical and Statistical Psychology, 23,* 121–145.

Jöreskog, K. G. (1973). A general method for estimating a linear structural equation system. In A. S. Goldberger & O. D. Duncan (Eds.), Structural equation models in the social sciences (pp. 85–112). New York: Wiley.

Jöreskog, K. G., & Sörboom, D. (2001). LISREL (Version 8.51) [Computer software]. Lincolnwood, IL: Scientific Software International.

Keesling, J. W. (1972). *Maximum likelihood approaches to causal analysis.* Unpublished doctoral dissertation, University of Chicago.

Klugkist, I., Laudy, O., & Hoijtink, H. (2005). Inequality constrained analysis of variance: A Bayesian approach. *Psychological Methods, 10,* 477–493.

Laird, N. M., & Ware, J. H. (1982). Random effects models for longitudinal data. *Biometrics, 38,* 963–974.

MacCallum, R. C., Kim, C., Malarkey, W. B., & Kiecolt-Glaser, J. K. (1997). Studying multivariate change using multilevel models and latent growth curve models. *Multivariate Behavioral Research, 32,* 215–253.

McArdle, J. J. (1986). Latent variable growth within behavior genetic models. *Behavior Genetics, 16,* 163–200.

McArdle, J. J. (1988). Dynamic but structural equation modeling of repeated measures data. In R. B. Cattel & J. Nesselroade (Eds.), *Handbook of multivariate experimental psychology* (2nd ed., pp. 561–614). New York: Plenum Press.

Meredith, W. M., & Tisak, J. (1984, June). *"Tuckerizing" curves.* Paper presented at the annual meeting of the Psychometric Society, Santa Barbara, CA.

Meredith, W. M., & Tisak, J. (1990). Latent curve analysis. *Psychometrika, 55,* 107–122.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., & Howerter, A. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41,* 49–100.

Moreira, M. J. (2003). A conditional ratio test for structural models. *Econometrica, 71,* 1027–1048.

Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling, 9,* 599–620.

Muthén, L. K., & Muthén, B. O. (2004). Mplus3 (Version 12) [Computer software]. Los Angeles: Muthén & Muthén.

Neale, M. C., & Miller, M. B. (1997). The use of likelihood-based confidence intervals in genetic models. *Behavior Genetics, 27,* 113–120.

R Development Core Team. (2005). R: A language and environment for statistical computing [Computer software]. Vienna: R Foundation for Statistical Computing.

Rovine, M. J., & Molenaar, P. C. M. (2000). A structural modeling approach to a multilevel random coefficients model. *Multivariate Behavioral Research, 35,* 51–88.

Rovine, M. J., & Molenaar, P. C. M. (2005). Relating factor models for longitudinal data to quasi-simplex and NARMA models. *Multivariate Behavioral Research, 40,* 83–113.

Schwartz, G. (1978). Estimating the dimensions of a model. *Annals of Statistics, 6,* 461–464.

Self, S. G., & Liang, K. Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association, 82,* 605–610.

Sham, P. C. (1998). *Statistics in human genetics.* New York: Wiley.

Shapiro, A. (1985). Asymptotic distribution of test statistics in the analysis of moment structures under inequality constraints. *Biometrika, 72,* 133–144.

Stram, D. O., & Lee, J. W. (1994). Variance components testing in the longitudinal mixed effects model. *Biometrics, 50,* 1171–1177.

Stram, D. O., & Lee, J. W. (1995). Correction to "Variance components testing in the longitudinal mixed effects model." *Biometrics, 51,* 1196.

Thurstone, L. L. (1947). *Multiple-factor analysis.* Chicago: University of Chicago Press.

Van der Sluis, S., Dolan, C. V., & Stoel, R. D. (2005). A note on testing perfect correlations in SEM. *Structural Equation Modeling, 12,* 551–577.

Widaman, K. F., & Thompson, J. S. (2003). On specifying the null model for incremental fit indices in structural equation modeling. *Psychological Methods, 8,* 16–37.

Wiley, D. E. (1973). The identification problem for structural equation models with unmeasured variables. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences* (pp. 69–83). New York: Academic Press.

Willett, J. B., & Sayer, A. G. (1994). Using covariance structure analysis to detect correlates and predictors of individual change over time. *Psychological Bulletin, 116,* 363–381.

Wright, S. (1918). On the nature of size factor. *Genetics, 3,* 367–374.

Appendix

The Asymptotic Distribution of the Likelihood Ratio

Estimation of the model parameters in structural equation modeling is often performed using the maximum-likelihood (ML) method under the assumption of independence of the cases and of multivariate normality. Estimates are obtained by minimizing the log-likelihood ratio function $F_{ml} = (N - 1) \times \{[(\overline{\mathbf{Y}} - \boldsymbol{\mu}_y)'\Sigma^{-1}(\overline{\mathbf{Y}} - \boldsymbol{\mu}_y)] + [\log|\boldsymbol{\Sigma}| - \log|\mathbf{S}| + \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S}) - p]\}$ (cf. Browne & Arminger, 1995; see also Widaman & Thompson, 2003, p. 17), where N is the sample size, $\overline{\mathbf{Y}}$ is a column vector of sample means of the p indicators, $\mathbf{S}$ is the $p \times p$ sample covariance matrix, $|.|$ denotes the determinant of a matrix, and tr is the trace operator that returns the sum of diagonal elements of a matrix.[A1] The $p \times p$ covariance matrix Σ and the p-dimensional vector $\boldsymbol{\mu}_y$ are the implied covariance matrix and mean vector, respectively. F_{ml} is bounded below by zero and equals this value only if both $\overline{\mathbf{Y}} = \boldsymbol{\mu}_y$ and $\Sigma = \mathbf{S}$.

To show the distribution of the LR statistic in the case of a test that contains boundary parameters, let $\boldsymbol{\theta}$ be a parameter vector that belongs to the parameter space $\boldsymbol{\Omega}$. Suppose that the true parameter value, denoted by $\boldsymbol{\theta}_0$, lies on the boundary of the parameter space under the null hypothesis $\boldsymbol{\Omega}_0$ and under the alternative hypothesis $\boldsymbol{\Omega}_1$. To find the asymptotic distribution of the likelihood ratio (LR) test statistic $-2\ln(F_{ml})$, it is assumed that the parameter space can be approximated by a convex cone. This is a set of points in which any linear combination of points of the cone also belongs to the cone. We need our set of parameters to be a convex cone to define the distance function. The theorem of Self and Liang (1987) provides the following general solution.

Theorem 1: Let $\mathbf{Z}$ be a random variable with a multivariate normal distribution $N(\theta, I^{-1}\theta_0)$, and let $C_{\Omega o}$ and $C_{\Omega 1}$ be nonempty cones approximating $\boldsymbol{\Omega}_0$ and $\boldsymbol{\Omega}_1$ at $\boldsymbol{\theta}_0$, respectively. Then, under usual regularity conditions (Self & Liang, 1987, p. 605), the asymptotic distribution of the LR statistic, $-2\ln(F_{ml})$, is the same as the asymptotic distribution of the LR test of $\boldsymbol{\theta} \in C_{\Omega_1}$ (i.e., the model under H_0) versus the alternative $\boldsymbol{\theta} \in C_{\Omega_1}$ (i.e., the model under H_1) based on a single realization of $\mathbf{Z}$ if $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ (i.e., if H_0 is true).

Note that the regularity conditions mentioned in this theorem ensure that the test statistic, $-2\ln(F_{ml})$, is asymptotically chi-square distributed only if there are no parameters on the boundary under H_0. The asymptotic distribution of the LR mentioned in Theorem 1 may be written as

$$\sup_{\theta \in C\Omega - \theta_0} \{-(\mathbf{Z} - \boldsymbol{\theta})' I^{-1}[\boldsymbol{\theta}_0](\mathbf{Z} - \boldsymbol{\theta})\} - \sup_{\theta \in C_{\Omega 0} - \theta_0} \{-(\mathbf{Z} - \boldsymbol{\theta})' I^{-1}[\boldsymbol{\theta}_0](\mathbf{Z} - \boldsymbol{\theta})\}. \quad \text{(A1)}$$

The first term in Equation A1 is related to the likelihood of the model under the alternative hypothesis (i.e., the unconstrained model) and the second term to the likelihood of the model under the null hypothesis. In this article, $\boldsymbol{\theta}$ contains the variances and covariances of the latent variables in a structural equation model (in standard LISREL language, they are the elements of matrix $\boldsymbol{\Psi}$).

[A1] We limit our presentation to the single-group ML function based on summary statistics. The extension to a multigroup model and the formulation of the ML function for raw data analysis (e.g., in the case of missing data) are trivial.

Received September 15, 2005
Revision received July 5, 2006
Accepted July 17, 2006 ■