# Chapter 14
Introduction to Structural Equations with Latent Variables

## Chapter Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>OVERVIEW</td>
<td>189</td>
</tr>
<tr>
<td>COMPARISON OF THE CALIS AND SYSLIN PROCEDURES</td>
<td>189</td>
</tr>
<tr>
<td>MODEL SPECIFICATION</td>
<td>190</td>
</tr>
<tr>
<td>ESTIMATION METHODS</td>
<td>191</td>
</tr>
<tr>
<td>STATISTICAL INFERENCE</td>
<td>191</td>
</tr>
<tr>
<td>GOODNESS-OF-FIT STATISTICS</td>
<td>193</td>
</tr>
<tr>
<td>OPTIMIZATION METHODS</td>
<td>193</td>
</tr>
<tr>
<td>SPECIFYING STRUCTURAL EQUATION MODELS</td>
<td>194</td>
</tr>
<tr>
<td>Identification of Models</td>
<td>199</td>
</tr>
<tr>
<td>PATH DIAGRAMS AND THE RAM MODEL</td>
<td>203</td>
</tr>
<tr>
<td>SOME MEASUREMENT MODELS</td>
<td>209</td>
</tr>
<tr>
<td>A COMBINED MEASUREMENT-STRUCTURAL MODEL WITH RECIPROCAL INFLUENCE AND CORRELATED RESIDUALS</td>
<td>222</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>250</td>
</tr>
</tbody>
</table>
Chapter 14
Introduction to Structural Equations with Latent Variables

Overview

You can use the CALIS procedure for analysis of covariance structures, fitting systems of linear structural equations, and path analysis. These terms are more or less interchangeable, but they emphasize different aspects of the analysis. The analysis of covariance structures refers to the formulation of a model for the variances and covariances among a set of variables and the fitting of the model to an observed covariance matrix. In linear structural equations, the model is formulated as a system of equations relating several random variables with assumptions about the variances and covariances of the random variables. In path analysis, the model is formulated as a path diagram, in which arrows connecting variables represent (co)variances and regression coefficients. Path models and linear structural equation models can be converted to models of the covariance matrix and can, therefore, be fitted by the methods of covariance structure analysis. All of these methods allow the use of hypothetical latent variables or measurement errors in the models.


Comparison of the CALIS and SYSLIN Procedures

The SYSLIN procedure in the SAS/ETS product can also fit certain kinds of path models and linear structural equation models. PROC CALIS differs from PROC SYSLIN in that PROC CALIS allows more generality in the use of latent variables in the models. Latent variables are unobserved, hypothetical variables, as distinct from manifest variables, which are the observed data. PROC SYSLIN allows at most one latent variable, the error term, in each equation. PROC CALIS allows several latent variables to appear in an equation—in fact, all the variables in an equation can be latent as long as there are other equations that relate the latent variables to manifest variables.

 Both the CALIS and SYSLIN procedures enable you to specify a model as a system of linear equations. When there are several equations, a given variable may be a dependent variable in one equation and an independent variable in other equations. Therefore, additional terminology is needed to describe unambiguously the roles of variables in the system. Variables with values that are determined jointly and simultaneously by the system of equations are called endogenous variables. Variables
with values that are determined outside the system, that is, in a manner separate from
the process described by the system of equations, are called exogenous variables. The
purpose of the system of equations is to explain the variation of each endogenous vari-
able in terms of exogenous variables or other endogenous variables or both. Refer to
Loehlin (1987, p. 4) for further discussion of endogenous and exogenous variables.
In the econometric literature, error and disturbance terms are usually distinguished
from exogenous variables, but in systems with more than one latent variable in an
equation, the distinction is not always clear.

In PROC SYSLIN, endogenous variables are identified by the ENDOGENOUS state-
ment. When you specify structural equations in PROC CALIS, endogenous variables
are assumed to be those that appear on the left-hand sides of the equations; a given
variable may appear on the left-hand side of at most one equation.

PROC SYSLIN provides many methods of estimation, some of which are applica-
ble only in special cases. For example, ordinary least-squares estimates are suitable
in certain kinds of systems but may be statistically biased and inconsistent in other
kinds. PROC CALIS provides three methods of estimation that can be used with
most models. Both the CALIS and SYSLIN procedures can do maximum likelihood
estimation, which PROC CALIS calls ML and PROC SYSLIN calls FIML. PROC
SYSLIN can be much faster than PROC CALIS in those special cases for which it
provides computationally efficient estimation methods. However, PROC CALIS has
a variety of sophisticated algorithms for maximum likelihood estimation that may be
much faster than FIML in PROC SYSLIN.

PROC CALIS can impose a wider variety of constraints on the parameters, including
nonlinear constraints, than can PROC SYSLIN. For example, PROC CALIS can con-
strain error variances or covariances to equal specified constants, or it can constrain
two error variances to have a specified ratio.

## Model Specification

PROC CALIS provides several ways to specify a model. Structural equations can
be transcribed directly in the LINEQS statement. A path diagram can be described
in the RAM statement. You can specify a first-order factor model in the FACTOR
and MATRIX statements. Higher-order factor models and other complicated models
can be expressed in the COSAN and MATRIX statements. For most applications,
the LINEQS and RAM statements are easiest to use; the choice between these two
statements is a matter of personal preference.

You can save a model specification in an OUTRAM= data set, which can then be
used with the INRAM= option to specify the model in a subsequent analysis.
Estimation Methods

The CALIS procedure provides three methods of estimation specified by the METHOD= option:

- **ULS** unweighted least squares
- **GLS** generalized least squares
- **ML** maximum likelihood for multivariate normal distributions

Each estimation method is based on finding parameter estimates that minimize a badness-of-fit function that measures the difference between the observed sample covariance matrix and the predicted covariance matrix given the model and the parameter estimates. See the section “Estimation Methods” on page 462 in Chapter 19, “The CALIS Procedure,” for formulas, or refer to Loehlin (1987, pp. 54–62) and Bollen (1989, pp. 104–123) for further discussion.

The default is METHOD=ML, which is the preferred method for most applications with respect to statistical considerations. The option METHOD=GLS usually produces very similar results to METHOD=ML. Both methods are suitable regardless of the scaling of the covariance matrix. The ULS method is appropriate for a covariance matrix only if the variables are measured on comparable scales; otherwise, METHOD=ULS should be applied to the correlation matrix. PROC CALIS cannot compute standard errors or test statistics with the ULS method.

You should not specify METHOD=ML or METHOD=GLS if the observed or predicted covariance matrix is singular—you should either remove variables involved in the linear dependencies or specify METHOD=ULS.

PROC CALIS should not be used if your data are extremely nonnormal data, especially if they have high kurtosis. You should remove outliers and try to transform variables that are skewed or heavy-tailed. This applies to all three estimation methods, since all the estimation methods depend on the sample covariance matrix, and the sample covariance matrix is a poor estimator for distributions with high kurtosis (Bollen 1989, pp. 415–418; Huber 1981; Hampel et. al 1986). PROC CALIS displays estimates of univariate and multivariate kurtosis (Bollen 1989, pp. 418–425) if you specify the KURTOSIS option in the PROC CALIS statement.

Statistical Inference

When you specify the ML or GLS estimates with appropriate models, PROC CALIS can compute

- a chi-square goodness-of-fit test of the specified model versus the alternative that the data are from a multivariate normal distribution with unconstrained covariance matrix (Loehlin 1987, pp. 62–64; Bollen 1989, pp. 110, 115, 263–269)
• approximate standard errors of the parameter estimates (Bollen 1989, pp. 109, 114, 286), displayed with the STDERR option

• various modification indices, requested via the MODIFICATION or MOD option, that give the approximate change in the chi-square statistic that would result from removing constraints on the parameters or constraining additional parameters to zero (Bollen 1989, pp. 293–303)

If you have two models such that one model results from imposing constraints on the parameters of the other, you can test the constrained model against the more general model by fitting both models with PROC CALIS. If the constrained model is correct, the difference between the chi-square goodness-of-fit statistics for the two models has an approximate chi-square distribution with degrees of freedom equal to the difference between the degrees of freedom for the two models (Loehlin 1987, pp. 62–67; Bollen 1989, pp. 291–292).

All of the test statistics and standard errors computed by PROC CALIS depend on the assumption of multivariate normality. Normality is a much more important requirement for data with random independent variables than it is for fixed independent variables. If the independent variables are random, distributions with high kurtosis tend to give liberal tests and excessively small standard errors, while low kurtosis tends to produce the opposite effects (Bollen 1989, pp. 266–267, 415–432).

The test statistics and standard errors computed by PROC CALIS are also based on asymptotic theory and should not be trusted in small samples. There are no firm guidelines on how large a sample must be for the asymptotic theory to apply with reasonable accuracy. Some simulation studies have indicated that problems are likely to occur with sample sizes less than 100 (Loehlin 1987, pp. 60–61; Bollen 1989, pp. 267–268). Extrapolating from experience with multiple regression would suggest that the sample size should be at least five to twenty times the number of parameters to be estimated in order to get reliable and interpretable results.

The asymptotic theory requires that the parameter estimates be in the interior of the parameter space. If you do an analysis with inequality constraints and one or more constraints are active at the solution (for example, if you constrain a variance to be nonnegative and the estimate turns out to be zero), the chi-square test and standard errors may not provide good approximations to the actual sampling distributions.

Standard errors may be inaccurate if you analyze a correlation matrix rather than a covariance matrix even for sample sizes as large as 400 (Boomsma 1983). The chi-square statistic is generally the same regardless of which matrix is analyzed provided that the model involves no scale-dependent constraints.

If you fit a model to a correlation matrix and the model constrains one or more elements of the predicted matrix to equal 1.0, the degrees of freedom of the chi-square statistic must be reduced by the number of such constraints. PROC CALIS attempts to determine which diagonal elements of the predicted correlation matrix are constrained to a constant, but it may fail to detect such constraints in complicated models, particularly when programming statements are used. If this happens, you should add parameters to the model to release the constraints on the diagonal elements.
Goodness-of-fit Statistics

In addition to the chi-square test, there are many other statistics for assessing the goodness of fit of the predicted correlation or covariance matrix to the observed matrix.

Akaike’s (1987) information criterion (AIC) and Schwarz’s (1978) Bayesian criterion (SBC) are useful for comparing models with different numbers of parameters—the model with the smallest value of AIC or SBC is considered best. Based on both theoretical considerations and various simulation studies, SBC seems to work better, since AIC tends to select models with too many parameters when the sample size is large.

There are many descriptive measures of goodness of fit that are scaled to range approximately from zero to one: the goodness of fit index (GFI) and GFI adjusted for degrees of freedom (AGFI) (Jöreskog and Sörbom 1988), centrality (McDonald 1989), and the parsimonious fit index (James, Mulaik, and Brett 1982). Bentler and Bonett (1980) and Bollen (1986) have proposed measures for comparing the goodness of fit of one model with another in a descriptive rather than inferential sense.

None of these measures of goodness of fit are related to the goodness of prediction of the structural equations. Goodness of fit is assessed by comparing the observed correlation or covariance matrix with the matrix computed from the model and parameter estimates. Goodness of prediction is assessed by comparing the actual values of the endogenous variables with their predicted values, usually in terms of root mean squared error or proportion of variance accounted for ($R^2$). For latent endogenous variables, root mean squared error and $R^2$ can be estimated from the fitted model.

Optimization Methods

PROC CALIS uses a variety of nonlinear optimization algorithms for computing parameter estimates. These algorithms are very complicated and do not always work. PROC CALIS will generally inform you when the computations fail, usually by displaying an error message about the iteration limit being exceeded. When this happens, you may be able to correct the problem simply by increasing the iteration limit (MAXITER= and MAXFUNC=). However, it is often more effective to change the optimization method (OMETHOD=) or initial values. For more details, see the section “Use of Optimization Techniques” on page 551 in Chapter 19, “The CALIS Procedure,” and refer to Bollen (1989, pp. 254–256).

PROC CALIS may sometimes converge to a local optimum rather than the global optimum. To gain some protection against local optima, you can run the analysis several times with different initial estimates. The RANDOM= option in the PROC CALIS statement is useful for generating a variety of initial estimates.
Specifying Structural Equation Models

Consider fitting a linear equation to two observed variables, \( Y \) and \( X \). Simple linear regression uses the model of a particular form, labeled for purposes of discussion, as Model Form A.

\[
Y = \alpha + \beta X + E_Y
\]

where \( \alpha \) and \( \beta \) are coefficients to be estimated and \( E_Y \) is an error term. If the values of \( X \) are fixed, the values of \( E_Y \) are assumed to be independent and identically distributed realizations of a normally distributed random variable with mean zero and variance \( \text{Var}(E_Y) \). If \( X \) is a random variable, \( X \) and \( E_Y \) are assumed to have a bivariate normal distribution with zero correlation and variances \( \text{Var}(X) \) and \( \text{Var}(E_Y) \), respectively. Under either set of assumptions, the usual formulas hold for the estimates of the coefficients and their standard errors (see Chapter 3, “Introduction to Regression Procedures”).

In the REG or SYSLIN procedure, you would fit a simple linear regression model with a MODEL statement listing only the names of the manifest variables:

```sas
proc reg;
  model y=x;
run;
```

You can also fit this model with PROC CALIS, but you must explicitly specify the names of the parameters and the error terms (except for the intercept, which is assumed to be present in each equation). The linear equation is given in the LINEQS statement, and the error variance is specified in the STD statement.

```sas
proc calis cov;
  lineqs y=beta x + ex;
  std ex=vex;
run;
```

The parameters are the regression coefficient BETA and the variance VEX of the error term EX. You do not need to type an * between BETA and X to indicate the multiplication of the variable by the coefficient.

The LINEQS statement uses the convention that the names of error terms begin with the letter E, disturbances (error terms for latent variables) in equations begin with D, and other latent variables begin with F for “factor.” Names of variables in the input SAS data set can, of course, begin with any letter.
If you leave out the name of a coefficient, the value of the coefficient is assumed to be 1. If you leave out the name of a variance, the variance is assumed to be 0. So if you tried to write the model the same way you would in PROC REG, for example,

```
proc calis cov;
    lineqs y=x;
```

you would be fitting a model that says $Y$ is equal to $X$ plus an intercept, with no error.

The COV option is used because PROC CALIS, like PROC FACTOR, analyzes the correlation matrix by default, yielding standardized regression coefficients. The COV option causes the covariance matrix to be analyzed, producing raw regression coefficients. See Chapter 3, “Introduction to Regression Procedures,” for a discussion of the interpretation of raw and standardized regression coefficients.

Since the analysis of covariance structures is based on modeling the covariance matrix and the covariance matrix contains no information about means, PROC CALIS neglects the intercept parameter by default. To estimate the intercept, change the COV option to UCov, which analyzes the uncorrected covariance matrix, and use the AUGMENT option, which adds a row and column for the intercept, called INTERCEPT, to the matrix being analyzed. The model can then be specified as

```
proc calis ucov augment;
    lineqs y=alpha intercept + beta x + ex;
    std ex=vex;
run;
```

In the LINEQS statement, intercept represents a variable with a constant value of 1; hence, the coefficient alpha is the intercept parameter.

Other commonly used options in the PROC CALIS statement include

- MODIFICATION to display model modification indices
- RESIDUAL to display residual correlations or covariances
- STDERR to display approximate standard errors
- TOTEFF to display total effects

For ordinary unconstrained regression models, there is no reason to use PROC CALIS instead of PROC REG. But suppose that the observed variables $Y$ and $X$ are contaminated by error, and you want to estimate the linear relationship between their true, error-free scores. The model can be written in several forms. A model of Form B is as follows.
Model Form B

\[ Y = \alpha + \beta F_X + E_Y \]
\[ X = F_X + E_X \]
\[ \text{Cov}(F_X, E_X) = \text{Cov}(F_X, E_Y) = \text{Cov}(E_X, E_Y) = 0 \]

This model has two error terms, \( E_Y \) and \( E_X \), as well as another latent variable \( F_X \) representing the true value corresponding to the manifest variable \( X \). The true value corresponding to \( Y \) does not appear explicitly in this form of the model.

The assumption in Model Form B is that the error terms and the latent variable \( F_X \) are jointly uncorrelated is of critical importance. This assumption must be justified on substantive grounds such as the physical properties of the measurement process. If this assumption is violated, the estimators may be severely biased and inconsistent.

You can express Model Form B in PROC CALIS as follows:

```
proc calis cov;
  lineqs y=beta fx + ey,
    x=fx + ex;
  std fx=vfx,
    ey=vey,
    ex=vex;
run;
```

You must specify a variance for each of the latent variables in this model using the STD statement. You can specify either a name, in which case the variance is considered a parameter to be estimated, or a number, in which case the variance is constrained to equal that numeric value. In general, you must specify a variance for each latent exogenous variable in the model, including error and disturbance terms. The variance of a manifest exogenous variable is set equal to its sample variance by default. The variances of endogenous variables are predicted from the model and are not parameters. Covariances involving latent exogenous variables are assumed to be zero by default. Covariances between manifest exogenous variables are set equal to the sample covariances by default.

Fuller (1987, pp. 18–19) analyzes a data set from Voss (1969) involving corn yields \( Y \) and available soil nitrogen \( X \) for which there is a prior estimate of the measurement error for soil nitrogen \( \text{Var}(E_X) \) of 57. You can fit Model Form B with this constraint using the following SAS statements.
data corn(type=cov);
   input _type_ $ _name_ $ y x;
   datalines;
   n . 11 11
   mean . 97.4545 70.6364
   cov y 87.6727 .
cov x 104.8818 304.8545 ;

proc calis data=corn cov stderr;
   lineqs y=beta fx + ey,
   x=fx + ex;
   std ex=57,
   fx=vfx,
   ey=vey;
run;

In the STD statement, the variance of EX is given as the constant value 57. PROC CALIS produces the following estimates.

The SAS System
The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation

\[ y = 0.4232 \times \text{fx} + 1.0000 \times \text{ey} \]
\[ \text{Std Err} = 0.1658 \times \text{beta} \]
\[ \text{t Value} = 2.5520 \]
\[ x = 1.0000 \times \text{fx} + 1.0000 \times \text{ex} \]

<table>
<thead>
<tr>
<th>Variable Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>t Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>fx</td>
<td>247.85450</td>
<td>136.33508</td>
<td>1.82</td>
</tr>
<tr>
<td>ey</td>
<td>43.29105</td>
<td>23.92488</td>
<td>1.81</td>
</tr>
<tr>
<td>ex</td>
<td>57.00000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 14.1.** Measurement Error Model for Corn Data

PROC CALIS also displays information about the initial estimates that can be useful if there are optimization problems. If there are no optimization problems, the initial estimates are usually not of interest; they are not be reproduced in the examples in this chapter.

You can write an equivalent model (labeled here as Model Form C) using a latent variable \( F_Y \) to represent the true value corresponding to \( Y \).
Model Form C

\[
Y = F_Y + E_Y \\
X = F_X + E_X \\
F_Y = \alpha + \beta F_X \\
\text{Cov}(F_X, E_X) = \text{Cov}(F_X, E_X) = \text{Cov}(E_X, E_Y) = 0
\]

The first two of the three equations express the observed variables in terms of a true score plus error; these equations are called the measurement model. The third equation, expressing the relationship between the latent true-score variables, is called the structural or causal model. The decomposition of a model into a measurement model and a structural model (Keesling 1972; Wiley 1973; Jöreskog 1973) has been popularized by the program LISREL (Jöreskog and Sörbom 1988). The statements for fitting this model are

```sas
proc calis cov;
  lineqs y=fy + ey,
    x=fx + ex,
    fy=beta fx;
  std fx=vfx,
    ey=vey,
    ex=vex;
  cov ey ex=ceyex;
run;
```

You do not need to include the variance of $F_Y$ in the STD statement because the variance of $F_Y$ is determined by the structural model in terms of the variance of $F_X$, that is, $\text{Var}(F_Y) = \beta^2 \text{Var}(F_X)$.

Correlations involving endogenous variables are derived from the model. For example, the structural equation in Model Form C implies that $F_Y$ and $F_X$ are correlated unless $\beta$ is zero. In all of the models discussed so far, the latent exogenous variables are assumed to be jointly uncorrelated. For example, in Model Form C, $E_Y$, $E_X$, and $F_X$ are assumed to be uncorrelated. If you want to specify a model in which $E_Y$ and $E_X$, say, are correlated, you can use the COV statement to specify the numeric value of the covariance $\text{Cov}(E_Y, E_X)$ between $E_Y$ and $E_X$, or you can specify a name to make the covariance a parameter to be estimated. For example,

```sas
proc calis cov;
  lineqs y=fy + ey,
    x=fx + ex,
    fy=beta fx;
  std fy=vfy,
    fx=vfx,
    ey=vey,
    ex=vex;
  run;
```
This COV statement specifies that the covariance between EY and EX is a parameter named CEYEX. All covariances that are not listed in the COV statement and that are not determined by the model are assumed to be zero. If the model contained two or more manifest exogenous variables, their covariances would be set to the observed sample values by default.

Identification of Models

Unfortunately, if you try to fit models of Form B or Form C without additional constraints, you cannot obtain unique estimates of the parameters. These models have four parameters (one coefficient and three variances). The covariance matrix of the observed variables Y and X has only three elements that are free to vary, since Cov(Y,X)=Cov(X,Y). The covariance structure can, therefore, be expressed as three equations in four unknown parameters. Since there are fewer equations than unknowns, there are many different sets of values for the parameters that provide a solution for the equations. Such a model is said to be underidentified.

If the number of parameters equals the number of free elements in the covariance matrix, then there may exist a unique set of parameter estimates that exactly reproduce the observed covariance matrix. In this case, the model is said to be just identified or saturated.

If the number of parameters is less than the number of free elements in the covariance matrix, there may exist no set of parameter estimates that reproduces the observed covariance matrix. In this case, the model is said to be overidentified. Various statistical criteria, such as maximum likelihood, can be used to choose parameter estimates that approximately reproduce the observed covariance matrix. If you use ML or GLS estimation, PROC CALIS can perform a statistical test of the goodness of fit of the model under the assumption of multivariate normality of all variables and independence of the observations.

If the model is just identified or overidentified, it is said to be identified. If you use ML or GLS estimation for an identified model, PROC CALIS can compute approximate standard errors for the parameter estimates. For underidentified models, PROC CALIS obtains approximate standard errors by imposing additional constraints resulting from the use of a generalized inverse of the Hessian matrix.

You cannot guarantee that a model is identified simply by counting the parameters. For example, for any latent variable, you must specify a numeric value for the variance, or for some covariance involving the variable, or for a coefficient of the variable in at least one equation. Otherwise, the scale of the latent variable is indeterminate, and the model will be underidentified regardless of the number of parameters and the size of the covariance matrix. As another example, an exploratory factor analysis with two or more common factors is always underidentified because you can rotate the common factors without affecting the fit of the model.

PROC CALIS can usually detect an underidentified model by computing the approximate covariance matrix of the parameter estimates and checking whether any estimate is linearly related to other estimates (Bollen 1989, pp. 248–250), in which case PROC CALIS displays equations showing the linear relationships among the estimates. An-
other way to obtain empirical evidence regarding the identification of a model is to run the analysis several times with different initial estimates to see if the same final estimates are obtained.

Bollen (1989) provides detailed discussions of conditions for identification in a variety of models.

The following example is inspired by Fuller (1987, pp. 40–41). The hypothetical data are counts of two types of cells, cells forming rosettes and nucleated cells, in spleen samples. It is reasonable to assume that counts have a Poisson distribution; hence, the square roots of the counts should have a constant error variance of 0.25.

You can use PROC CALIS to fit a model of Form C to the square roots of the counts without constraints on the parameters, as displayed in following statements. The option OMETHOD=QUANEW is used in the PROC CALIS statement because in this case it produces more rapid convergence than the default optimization method.

data spleen;
  input rosette nucleate;
  sqrtrose=sqrt(rosette);
  sqrlnucl=sqrt(nucleate);
  datalines;
  4  62
  5  87
  5 117
  6 142
  8 212
  9 120
 12 254
 13 179
 15 125
 19 182
 28 301
 51 357;

proc calis data=spleen cov omethod=quanew;
  lineqs sqrtrose=factrose + err_rose,
     sqrlnucl=factnucl + err_nucl,
     factrose=beta factnucl;
  std err_rose=v_rose,
    err_nucl=v_nucl,
    factnucl=v_factnu;
run;
This model is underidentified. PROC CALIS displays the following warning:

```
WARNING: Problem not identified: More parameters to estimate ( 4 )
than given values in data matrix ( 3 ).
```

and diagnoses the indeterminacy as follows:

```
NOTE: Hessian matrix is not full rank. Not all parameters are identified.
Some parameter estimates are linearly related to other parameter
estimates as shown in the following equations:

\[
  v_{nucl} = -10.554977 - 0.036438 \times \beta + 1.00000 \times v_{factnu} + 0.149564 \times v_{rose}
\]
```

The constraint that the error variances equal 0.25 can be imposed by modifying the STD statement:

```
proc calis data=spleen cov stderr;
  lineqs sqrtrose=factrose + err_rose,
             sqrtnucl=factnucl + err_nucl,
             factrose=beta factnucl;
  std err_rose=.25,
              err_nucl=.25,
              factnucl=v_factnu;
run;
```

The resulting parameter estimates are displayed in Figure 14.2.

```
The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation

factrose = 0.4034*factnucl
Std Err  0.0508 beta
t Value  7.9439

Variances of Exogenous Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>t Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>factnucl</td>
<td>v_factnu</td>
<td>10.45846</td>
<td>4.56608</td>
<td>2.29</td>
</tr>
<tr>
<td>err_rose</td>
<td></td>
<td>0.25000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>err_nucl</td>
<td></td>
<td>0.25000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

**Figure 14.2.** Spleen Data: Parameter Estimates for Overidentified Model

This model is overidentified and the chi-square goodness-of-fit test yields a \( p \)-value of 0.0219, as displayed in Figure 14.3.
Figure 14.3. Spleen Data: Fit Statistics for Overidentified Model

The sample size is so small that the \( p \)-value should not be taken to be accurate, but to get a small \( p \)-value with such a small sample indicates it is possible that the model is seriously deficient. The deficiency could be due to any of the following:

- The error variances are not both equal to 0.25.
- The error terms are correlated with each other or with the true scores.
- The observations are not independent.
- There is a disturbance in the linear relation between \texttt{factrose} and \texttt{factnucl}.
- The relation between \texttt{factrose} and \texttt{factnucl} is not linear.
- The actual distributions are not adequately approximated by the multivariate normal distribution.

A simple and plausible modification to the model is to add a “disturbance term” or “error in the equation” to the structural model, as follows.
proc calis data=spleen cov stderr;
   lineqs sqrtrose=factrose + err_rose,
            sqrtnucl=factnucl + err_nucl,
            factrose=beta factnucl + disturb;
   std err_rose=.25,
             err_nucl=.25,
             factnucl=v_factnu,
             disturb=v_dist;
run;

The following parameter estimates are produced.

```
   The CALIS Procedure
   Covariance Structure Analysis: Maximum Likelihood Estimation

   factrose = 0.3907*factnucl + 1.0000 disturb
   Std Err  0.0771 beta
   t Value  5.0692

   Variances of Exogenous Variables

   Variable  Parameter  Estimate  Standard  t Value
              Error      
   factnucl v_factnu 10.50458 4.58577   2.29
   err_rose          0.25000
   err_nucl          0.25000
   disturb v_dist   0.38153 0.28556   1.34
```

**Figure 14.4.** Spleen Data: Parameter Estimated for Just Identified Model

This model is just identified, so there are no degrees of freedom for the chi-square goodness-of-fit test.

**Path Diagrams and the RAM Model**

Complicated models are often easier to understand when they are expressed as path diagrams. One advantage of path diagrams over equations is that variances and covariances can be shown directly in the path diagram. Loehlin (1987) provides a detailed discussion of path diagrams.

It is customary to write the names of manifest variables in rectangles and names of latent variables in ovals. The coefficients in each equation are indicated by drawing arrows from the independent variables to the dependent variable. Covariances between exogenous variables are drawn as two-headed arrows. The variance of an exogenous variable can be displayed as a two-headed arrow with both heads pointing to the exogenous variable, since the variance of a variable is the covariance of the variable with itself. Here is a path diagram for the spleen data, explicitly showing all latent variables and variances of exogenous variables.
There is an easier way to draw the path diagram based on McArdle’s reticular action model (RAM) (McArdle and McDonald 1984). McArdle uses the convention that a two-headed arrow that points to an endogenous variable actually refers to the error or disturbance term associated with that variable. A two-headed arrow with both heads pointing to the same endogenous variable represents the error or disturbance variance for the equation that determines the endogenous variable; there is no need to draw a separate oval for the error or disturbance term. Similarly, a two-headed arrow connecting two endogenous variables represents the covariance between the error of disturbance terms associated with the endogenous variables. The RAM conventions allow the previous path diagram to be simplified, as follows.

**Figure 14.5.** Path Diagram: Spleen
The RAM statement in PROC CALIS provides a simple way to transcribe a path diagram based on the reticular action model. Assign the integers 1, 2, 3, . . . to the variables in the order in which they appear in the SAS data set or in the VAR statement, if you use one. Assign subsequent consecutive integers to the latent variables displayed explicitly in the path diagram (excluding the error and disturbance terms implied by two-headed arrows) in any order. Each arrow in the path diagram can then be identified by two numbers indicating the variables connected by the path. The RAM statement consists of a list of descriptions of all the arrows in the path diagram. The descriptions are separated by commas. Each arrow description consists of three or four numbers and, optionally, a name in the following order:

1. The number of heads the arrow has.
2. The number of the variable the arrow points to, or either variable if the arrow is two-headed.
3. The number of the variable the arrow comes from, or the other variable if the arrow is two-headed.
4. The value of the coefficient or (co)variance that the arrow represents.

Figure 14.6. Path Diagram: Spleen
5. A name if the arrow represents a parameter to be estimated, in which case the previous number is taken to be the initial estimate of the parameter. Omit the name if the arrow represents a constant. If you specify a name, the fourth number may be omitted.

The model for the spleen data can be specified with the RAM statement, as follows:

```sas
/* 1 sqrtrose */
/* 2 sqrtnucl */
/* 3 factrose */
/* 4 factnucl */
proc calis data=spleen cov stderr method=ml outram=splram1;
  var sqrtrose sqrtnucl;
  ram 1 1 3 1, /* sqrtrose <- factrose */
       1 2 4 1, /* sqrtnucl <- factnucl */
       1 3 4 beta, /* factrose <- factnucl */
       2 1 1 .25, /* error variance for sqrtrose */
       2 2 2 .25, /* error variance for sqrtnucl */
       2 3 3 v_dist, /* disturbance variance for factrose */
       2 4 4 v_factnu; /* variance of factnucl */
run;
```

The resulting output in RAM form is displayed in Figure 14.7.

**Figure 14.7.** Spleen Data: RAM Model

You can request an output data set containing the model specification by using the OUTRAM= option in the PROC CALIS statement. Names for the latent variables can be specified in a VNAMES statement.
proc calis data=spleen cov stderr method=ml outram=splram1;
  var sqrtrose sqrtnucl;
  vnames 1 factrose factnucl,
         2 err_rose err_nucl disturb factnucl;
  ram 1 1 3 1,   /* sqrtrose <- factrose */
       1 2 4 1,   /* sqrtnucl <- factnucl */
       1 3 4 beta, /* factrose <- factnucl */
       2 1 1 .25, /* error variance for sqrtrose */

       2 2 2 .25,  /* error variance for sqrtnucl */
       2 3 3 v_dist, /* disturbance variance for factrose */
       2 4 4 v_factnu; /* variance of factnucl */
run;

proc print;
run;

The RAM output is displayed in Figure 14.8.

The OUTRAM= data set contains the RAM model as you specified it in the RAM statement, but it contains the final parameter estimates and standard errors instead of the initial values.

Figure 14.8. Spleen Data: RAM Model with Names for Latent Variables

The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation

<table>
<thead>
<tr>
<th>Term Matrix</th>
<th>---Row-----</th>
<th>---Column---</th>
<th>Parameter</th>
<th>Estimate</th>
<th>Error</th>
<th>t Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2 sqrtrose</td>
<td>1 factrose</td>
<td>3</td>
<td>1.00000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2 sqrtnucl</td>
<td>2 factnucl</td>
<td>4</td>
<td>1.00000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2 factrose</td>
<td>3 factnucl</td>
<td>4 beta</td>
<td>0.39074</td>
<td>0.07708</td>
<td>5.07</td>
</tr>
<tr>
<td>1</td>
<td>3 err_rose</td>
<td>1 err_rose</td>
<td>1</td>
<td>0.25000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3 err_nucl</td>
<td>2 err_nucl</td>
<td>2</td>
<td>0.25000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3 disturb</td>
<td>3 disturb</td>
<td>3 v_dist</td>
<td>0.38153</td>
<td>0.28556</td>
<td>1.34</td>
</tr>
<tr>
<td>1</td>
<td>3 factnucl</td>
<td>4 factnucl</td>
<td>4 v_factnu</td>
<td>10.50458</td>
<td>4.58577</td>
<td>2.29</td>
</tr>
</tbody>
</table>
### Figure 14.9.  Spleen Data: OUTRAM= Data Set with Final Parameter Estimates

This data set can be used as input to another run of PROC CALIS with the INRAM= option in the PROC CALIS statement. For example, if the iteration limit is exceeded, you can use the RAM data set to start a new run that begins with the final estimates from the last run. Or you can change the data set to add or remove constraints or modify the model in various other ways. The easiest way to change a RAM data set is to use the FSEDIT procedure, but you can also use a DATA step. For example, you
could set the variance of the disturbance term to zero, effectively removing the disturbance from the equation, by removing the parameter name _v_dist_ in the _NAME_ variable and setting the value of the estimate to zero in the _ESTIM_ variable:

```sas
data splram2(type=ram);
   set splram1;
   if _name_='v_dist' then
      do;
         _name_=' ';
         _estim_=0;
      end;
run;

proc calis data=spleen inram=splram2 cov stderr;
run;
```

The resulting RAM output is displayed in Figure 14.10.

<table>
<thead>
<tr>
<th>Term Matrix</th>
<th>Row</th>
<th>Column</th>
<th>Parameter</th>
<th>Estimate</th>
<th>Error</th>
<th>t Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>sqrtrose</td>
<td>factrose 3</td>
<td>1.00000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>sqrtnucl</td>
<td>factnucl 4</td>
<td>1.00000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>factrose</td>
<td>factnucl 4</td>
<td>beta</td>
<td>0.40340</td>
<td>0.05078</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>err_rose</td>
<td>err_rose 1</td>
<td>.</td>
<td>0.25000</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>err_nucl</td>
<td>err_nucl 2</td>
<td>.</td>
<td>0.25000</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>disturb</td>
<td>disturb 3</td>
<td>.</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>factnucl</td>
<td>factnucl 4</td>
<td>v_factnu</td>
<td>10.45846</td>
<td>4.56608</td>
</tr>
</tbody>
</table>

**Figure 14.10.** Spleen Data: RAM Model with INRAM= Data Set

### Some Measurement Models

Psychometric test theory involves many kinds of models relating scores on psychological and educational tests to latent variables representing intelligence or various underlying abilities. The following example uses data on four vocabulary tests from Lord (1957). Tests W and X have 15 items each and are administered with very liberal time limits. Tests Y and Z have 75 items and are administered under time pressure. The covariance matrix is read by the following DATA step:

```sas
data lord(type=cov);
   input _type_ $ _name_ $ w x y z;
   datalines;
   n    649  .  .  .  .
   cov  w  86.3979  .  .  .
   cov  x  57.7751  86.2632  .  .
   cov  y  56.8651  59.3177  97.2850  .
   cov  z  58.8986  59.6683  73.8201  97.8192
;```
The psychometric model of interest states that $W$ and $X$ are determined by a single common factor $F_{WX}$, and $Y$ and $Z$ are determined by a single common factor $F_{YZ}$. The two common factors are expected to have a positive correlation, and it is desired to estimate this correlation. It is convenient to assume that the common factors have unit variance, so their correlation will be equal to their covariance. The error terms for all the manifest variables are assumed to be uncorrelated with each other and with the common factors. The model (labeled here as Model Form D) is as follows.

**Model Form D**

$$
W = \beta_W F_{WX} + E_W \\
X = \beta_X F_{WX} + E_X \\
Y = \beta_Y F_{YZ} + E_Y \\
Z = \beta_Z F_{YZ} + E_Z \\
\text{Var}(F_{WX}) = \text{Var}(F_{YZ}) = 1 \\
\text{Cov}(F_{WX}, F_{YZ}) = \rho \\
\text{Cov}(E_W, E_X) = \text{Cov}(E_W, E_Y) = \text{Cov}(E_W, E_Z) = \text{Cov}(E_X, E_Y) = \text{Cov}(E_X, E_Z) = \text{Cov}(E_W, F_{WX}) = \text{Cov}(E_X, F_{WX}) = \text{Cov}(E_Y, F_{WX}) = \text{Cov}(E_Z, F_{WX}) = 0 \\
\text{Cov}(E_Y, F_{YZ}) = \text{Cov}(E_Y, F_{YZ}) = \text{Cov}(E_Y, F_{YZ}) = \text{Cov}(E_Z, F_{YZ}) = 0
$$

The corresponding path diagram is as follows.
Figure 14.11. Path Diagram: Lord

This path diagram can be converted to a RAM model as follows:

```sas
/* 1=w 2=x 3=y 4=z 5=fwx 6=fyz */
title 'H4: unconstrained';
proc calis data=lord cov;
  ram 1 1 5 betaw,
       1 2 5 betax,
       1 3 6 betay,
       1 4 6 betaz,
       2 1 1 vew,
       2 2 2 vex,
       2 3 3 vey,
       2 4 4 vez,
       2 5 5 1,
       2 6 6 1,
       2 5 6 rho;
run;
```

Here are the major results.
Chapter 14. Introduction to Structural Equations with Latent Variables

Figure 14.12. Lord Data: Major Results for RAM Model, Hypothesis H4

H4: unconstrained

The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation

Fit Function 0.0011
Goodness of Fit Index (GFI) 0.9995
GFI Adjusted for Degrees of Freedom (AGFI) 0.9946
Root Mean Square Residual (RMR) 0.2720
 Parsimonious GFI (Mulaik, 1989) 0.1666
Chi-Square 0.7030
Chi-Square DF 1
Pr > Chi-Square 0.4018
Independence Model Chi-Square 1466.6
Independence Model Chi-Square DF 6
RMSEA Estimate 0.0000
RMSEA 90% Lower Confidence Limit .
RMSEA 90% Upper Confidence Limit 0.0974
ECVI Estimate 0.0291
ECVI 90% Lower Confidence Limit .
ECVI 90% Upper Confidence Limit 0.0391
Probability of Close Fit 0.6854
Bentler’s Comparative Fit Index 1.0000
Normal Theory Reweighted LS Chi-Square 0.7026
Akaike’s Information Criterion -1.2970
Bozdogan’s (1987) CAIC -6.7725
Schwarz’s Bayesian Criterion -5.7725
McDonald’s (1989) Centrality 1.0002
Bentler & Bonett’s (1980) Non-normed Index 1.0012
Bentler & Bonett’s (1980) NFI 0.9995
James, Mulaik, & Brett (1982) Parsimonious NFI 0.1666
Z-Test of Wilson & Hilferty (1931) 0.2363
Bollen (1986) Normed Index Rho1 0.9971
Bollen (1988) Non-normed Index Delta2 1.0002
Hoelter’s (1983) Critical N 3543

H4: unconstrained

The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation

RAM Estimates

<table>
<thead>
<tr>
<th>Term</th>
<th>Matrix</th>
<th>--Row--</th>
<th>--Column--</th>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>t Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2 w</td>
<td>1 F1</td>
<td>5 betaw</td>
<td>7.50066</td>
<td>0.32339</td>
<td>23.19</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2 x</td>
<td>2 F1</td>
<td>5 betax</td>
<td>7.70266</td>
<td>0.32063</td>
<td>24.02</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2 y</td>
<td>3 F2</td>
<td>6 betay</td>
<td>8.50947</td>
<td>0.32694</td>
<td>26.03</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2 z</td>
<td>4 F2</td>
<td>6 betaz</td>
<td>8.67505</td>
<td>0.32560</td>
<td>26.64</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3 E1</td>
<td>1 E1</td>
<td>1 vew</td>
<td>30.13796</td>
<td>2.47037</td>
<td>12.20</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3 E2</td>
<td>2 E2</td>
<td>2 vex</td>
<td>26.93217</td>
<td>2.43065</td>
<td>11.08</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3 E3</td>
<td>3 E3</td>
<td>3 vey</td>
<td>24.87396</td>
<td>2.35986</td>
<td>10.54</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3 E4</td>
<td>4 E4</td>
<td>4 vez</td>
<td>22.56264</td>
<td>2.35028</td>
<td>9.60</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3 D1</td>
<td>5 D1</td>
<td>5 rho</td>
<td>0.89855</td>
<td>0.01865</td>
<td>48.18</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3 D2</td>
<td>6 D2</td>
<td>6 .</td>
<td>1.00000</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The same analysis can be performed with the LINEQS statement. Subsequent analyses are illustrated with the LINEQS statement rather than the RAM statement because it is slightly easier to understand the constraints as written in the LINEQS statement without constantly referring to the path diagram. The LINEQS and RAM statements may yield slightly different results due to the inexactness of the numerical optimization; the discrepancies can be reduced by specifying a more stringent convergence criterion such as GCONV=1E-4 or GCONV=1E-6. It is convenient to create an OUT-RAM= data set for use in fitting other models with additional constraints.

```sas
title 'H4: unconstrained';
proc calis data=lord cov outram=ram4;
  lineqs w=betaw fwx + ew,
    x=betax fwx + ex,
    y=betay fyz + ey,
    z=betaz fyz + ez;
  std fwx fyz=1,
    ew ex ey ez=vew vex vey vez;
  cov fwx fyz=rho;
run;
```

The LINEQS displayed output is as follows.
### Variances of Exogenous Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>t Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>fwx</td>
<td></td>
<td>1.00000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>fyz</td>
<td></td>
<td>1.00000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ew</td>
<td>vew</td>
<td>30.13796</td>
<td>2.47037</td>
<td>12.20</td>
</tr>
<tr>
<td>ex</td>
<td>vex</td>
<td>26.93217</td>
<td>2.43065</td>
<td>11.08</td>
</tr>
<tr>
<td>ey</td>
<td>vey</td>
<td>24.87396</td>
<td>2.35986</td>
<td>10.54</td>
</tr>
<tr>
<td>ez</td>
<td>vez</td>
<td>22.56264</td>
<td>2.35028</td>
<td>9.60</td>
</tr>
</tbody>
</table>

### Covariances Among Exogenous Variables

<table>
<thead>
<tr>
<th>Var1 Var2</th>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>t Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>fwx fyz</td>
<td>rho</td>
<td>0.89855</td>
<td>0.01865</td>
<td>48.18</td>
</tr>
</tbody>
</table>

**Figure 14.13.** Lord Data: Using LINEQS Statement for RAM Model, Hypothesis H4

In an analysis of these data by Jöreskog and Sörbom (1979, pp. 54–56; Loehlin 1987, pp. 84–87), four hypotheses are considered:

1. **H1:** \( \rho = 1 \),
   \( \beta_W = \beta_X \), \( \Var(E_W) = \Var(E_X) \),
   \( \beta_Y = \beta_Z \), \( \Var(E_Y) = \Var(E_Z) \)
2. **H2:** same as H1: except \( \rho \) is unconstrained
3. **H3:** \( \rho = 1 \)
4. **H4:** Model Form D without any additional constraints
The hypothesis $H_3$ says that there is really just one common factor instead of two; in the terminology of test theory, $W$, $X$, $Y$, and $Z$ are said to be congeneric. The hypothesis $H_2$ says that $W$ and $X$ have the same true-scores and have equal error variance; such tests are said to be parallel. The hypothesis $H_2$ also requires $Y$ and $Z$ to be parallel. The hypothesis $H_1$ says that $W$ and $X$ are parallel tests, $Y$ and $Z$ are parallel tests, and all four tests are congeneric.

It is most convenient to fit the models in the opposite order from that in which they are numbered. The previous analysis fit the model for $H_4$ and created an OUTRAM= data set called ram4. The hypothesis $H_3$ can be fitted directly or by modifying the ram4 data set. Since $H_3$ differs from $H_4$ only in that $\rho$ is constrained to equal 1, the ram4 data set can be modified by finding the observation for which _NAME_ = ’rho’ and changing the variable _NAME_ to a blank value (meaning that the observation represents a constant rather than a parameter to be fitted) and setting the variable _ESTIM_ to the value 1. Both of the following analyses produce the same results:

```

title 'H3: W, X, Y, and Z are congeneric';
proc calis data=lord cov;
   lineqs w=betaw f + ew,
      x=betax f + ex,
      y=betay f + ey,
      z=betaz f + ez;
   std f=1,
      ew ex ey ez=vew vex vey vez;
run;

data ram3(type=ram);
set ram4;
   if _name_ = ’rho’ then
      do;
         _name_ = ’ ’;
         _estim_ = 1;
      end;
run;

proc calis data=lord inram=ram3 cov;
run;

```

The resulting output from either of these analyses is displayed in Figure 14.14.
H3: W, X, Y, and Z are congeneric

The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation

<table>
<thead>
<tr>
<th>Metric</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fit Function</td>
<td>0.0559</td>
</tr>
<tr>
<td>Goodness of Fit Index (GFI)</td>
<td>0.9714</td>
</tr>
<tr>
<td>GFI Adjusted for Degrees of Freedom (AGFI)</td>
<td>0.8570</td>
</tr>
<tr>
<td>Root Mean Square Residual (RMR)</td>
<td>2.4636</td>
</tr>
<tr>
<td>Parsimonious GFI (Mulaik, 1989)</td>
<td>0.3238</td>
</tr>
<tr>
<td>Chi-Square</td>
<td>36.2095</td>
</tr>
<tr>
<td>Chi-Square DF</td>
<td>2</td>
</tr>
<tr>
<td>Pr &gt; Chi-Square</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Independence Model Chi-Square</td>
<td>1466.6</td>
</tr>
<tr>
<td>Independence Model Chi-Square DF</td>
<td>6</td>
</tr>
<tr>
<td>RMSEA Estimate</td>
<td>0.1625</td>
</tr>
<tr>
<td>RMSEA 90% Lower Confidence Limit</td>
<td>0.1187</td>
</tr>
<tr>
<td>RMSEA 90% Upper Confidence Limit</td>
<td>0.2108</td>
</tr>
<tr>
<td>ECVI Estimate</td>
<td>0.0808</td>
</tr>
<tr>
<td>ECVI 90% Lower Confidence Limit</td>
<td>0.0561</td>
</tr>
<tr>
<td>ECVI 90% Upper Confidence Limit</td>
<td>0.1170</td>
</tr>
<tr>
<td>Probability of Close Fit</td>
<td>0.0000</td>
</tr>
<tr>
<td>Bentler’s Comparative Fit Index</td>
<td>0.9766</td>
</tr>
<tr>
<td>Normal Theory Reweighted LS Chi-Square</td>
<td>38.1432</td>
</tr>
<tr>
<td>Akaike’s Information Criterion</td>
<td>32.2095</td>
</tr>
<tr>
<td>Bozdogan’s (1987) CAIC</td>
<td>21.2586</td>
</tr>
<tr>
<td>Schwarz’s Bayesian Criterion</td>
<td>23.2586</td>
</tr>
<tr>
<td>McDonald’s (1989) Centrality</td>
<td>0.9740</td>
</tr>
<tr>
<td>Bentler &amp; Bonett’s (1980) Non-normed Index</td>
<td>0.9297</td>
</tr>
<tr>
<td>Bentler &amp; Bonett’s (1980) NFI</td>
<td>0.9753</td>
</tr>
<tr>
<td>James, Mulaik, &amp; Brett (1982) Parsimonious NFI</td>
<td>0.3251</td>
</tr>
<tr>
<td>Z-Test of Wilson &amp; Hilferty (1931)</td>
<td>5.2108</td>
</tr>
<tr>
<td>Bollen (1986) Normed Index Rho1</td>
<td>0.9259</td>
</tr>
<tr>
<td>Bollen (1988) Non-normed Index Delta2</td>
<td>0.9766</td>
</tr>
<tr>
<td>Hoelter’s (1983) Critical N</td>
<td>109</td>
</tr>
</tbody>
</table>

Figure 14.14.  Lord Data: Major Results for Hypothesis H3
H3: W, X, Y, and Z are congeneric

The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation

\[
\begin{align*}
  w &= 7.1047 \times fwx + 1.0000 \times ew \\
  t \text{ Value} &= 22.0802 \\
  x &= 7.2691 \times fwx + 1.0000 \times ex \\
  t \text{ Value} &= 22.8397 \\
  y &= 8.3735 \times fyz + 1.0000 \times ey \\
  t \text{ Value} &= 25.7316 \\
  z &= 8.5106 \times fyz + 1.0000 \times ez \\
  t \text{ Value} &= 26.2598
\end{align*}
\]

Variances of Exogenous Variables

<table>
<thead>
<tr>
<th>Variable Parameter</th>
<th>Estimate</th>
<th>Std Error</th>
<th>t Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>fwx</td>
<td>1.00000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>fyz</td>
<td>1.00000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ew</td>
<td>35.92087</td>
<td>2.41466</td>
<td>14.88</td>
</tr>
<tr>
<td>ex</td>
<td>33.42397</td>
<td>2.31038</td>
<td>14.47</td>
</tr>
<tr>
<td>ey</td>
<td>27.16980</td>
<td>2.24619</td>
<td>12.10</td>
</tr>
<tr>
<td>ez</td>
<td>25.38948</td>
<td>2.20839</td>
<td>11.50</td>
</tr>
</tbody>
</table>

The hypothesis $H_2$ requires that several pairs of parameters be constrained to have equal estimates. With PROC CALIS, you can impose this constraint by giving the same name to parameters that are constrained to be equal. This can be done directly in the LINEQS and STD statements or by using PROC FSEDIT or a DATA step to change the values in the ram4 data set:

```sas
title 'H2: W and X parallel, Y and Z parallel';
proc calis data=lord cov;
  lineqs w=betawx fwx + ew,
    x=betawx fwx + ex,
    y=betayz fyz + ey,
    z=betayz fyz + ez;
  std fwx fyz=1,
    ew ex ey ez=vewx vewx veyz veyz;
  cov fwx fyz=rho;
run;

data ram2(type=ram);
  set ram4;
  if _name_='betaw' then _name_='betawx';
  if _name_='betax' then _name_='betawx';
  if _name_='betay' then _name_='betayz';
  if _name_='betaz' then _name_='betayz';
  if _name_='vew' then _name_='vewx';
```

SAS OnlineDoc™: Version 8
The resulting output from either of these analyses is displayed in Figure 14.15.

```
if _name_='vex' then _name_='vewx';
if _name_='vey' then _name_='veyz';
if _name_='vez' then _name_='veyz';
run;
proc calis data=lord inram=ram2 cov;
run;
```

Figure 14.15. Lord Data: Major Results for Hypothesis H2
H2: W and X parallel, Y and Z parallel

The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation

\[
\begin{align*}
\text{w} &= 7.6010 \times \text{fwx} + 1.0000 \times \text{ew} \\
\text{std err} &= 0.2684 \times \text{betawx} \\
\text{t value} &= 28.3158 \\
\text{x} &= 7.6010 \times \text{fwx} + 1.0000 \times \text{ex} \\
\text{std err} &= 0.2684 \times \text{betawx} \\
\text{t value} &= 28.3158 \\
\text{y} &= 8.5919 \times \text{fyz} + 1.0000 \times \text{ey} \\
\text{std err} &= 0.2797 \times \text{betayz} \\
\text{t value} &= 30.7215 \\
\text{z} &= 8.5919 \times \text{fyz} + 1.0000 \times \text{ez} \\
\text{std err} &= 0.2797 \times \text{betayz} \\
\text{t value} &= 30.7215
\end{align*}
\]

Variances of Exogenous Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter</th>
<th>Estimate</th>
<th>Error</th>
<th>t Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>fwx</td>
<td>1.00000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fyz</td>
<td>1.00000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ew</td>
<td>vewx</td>
<td>28.55545</td>
<td>1.58641</td>
<td>18.00</td>
</tr>
<tr>
<td>ex</td>
<td>vewx</td>
<td>28.55545</td>
<td>1.58641</td>
<td>18.00</td>
</tr>
<tr>
<td>ey</td>
<td>veyz</td>
<td>23.73200</td>
<td>1.31844</td>
<td>18.00</td>
</tr>
<tr>
<td>ez</td>
<td>veyz</td>
<td>23.73200</td>
<td>1.31844</td>
<td>18.00</td>
</tr>
</tbody>
</table>

Covariances Among Exogenous Variables

<table>
<thead>
<tr>
<th>Var1</th>
<th>Var2</th>
<th>Parameter</th>
<th>Estimate</th>
<th>Error</th>
<th>t Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>fwx</td>
<td>fyz</td>
<td>rho</td>
<td>0.89864</td>
<td>0.01865</td>
<td>48.18</td>
</tr>
</tbody>
</table>

The hypothesis $H1$ requires one more constraint in addition to those in $H2$:

```
title 'H1: W and X parallel, Y and Z parallel, all congeneric';
proc calis data=lord cov;
  lineqs w=betawx f + ew,
  x=betawx f + ex,
  y=betayz f + ey,
  z=betayz f + ez;
  std f=1,
    ew ex ey ez=vewx vewx veyz veyz;
run;

data raml(type=ram);
  set ram2;
  if _name_='rho' then
    do;
      _name_='   ';
      _estim_=1;
    end;
run;
```
proc calis data=lord inram=ram1 cov;
run;

The resulting output from either of these analyses is displayed in Figure 14.16.

Figure 14.16. Lord Data: Major Results for Hypothesis H1
H1: W and X parallel, Y and Z parallel, all congeneric

The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation

\[
w = 7.1862 \times f_{wx} + 1.0000 \times ew
\]
\[
w = 7.1862 \times f_{wx} + 1.0000 \times ew
\]
\[
x = 7.1862 \times f_{wx} + 1.0000 \times ex
\]
\[
x = 7.1862 \times f_{wx} + 1.0000 \times ex
\]
\[
y = 8.4420 \times f_{yz} + 1.0000 \times ey
\]
\[
y = 8.4420 \times f_{yz} + 1.0000 \times ey
\]
\[
z = 8.4420 \times f_{yz} + 1.0000 \times ez
\]
\[
z = 8.4420 \times f_{yz} + 1.0000 \times ez
\]

Variances of Exogenous Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter Estimate</th>
<th>Error</th>
<th>t Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>fwx</td>
<td>1.00000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>fyz</td>
<td>1.00000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ew</td>
<td>vewx 34.68865</td>
<td>1.64634</td>
<td>21.07</td>
</tr>
<tr>
<td>ex</td>
<td>vewx 34.68865</td>
<td>1.64634</td>
<td>21.07</td>
</tr>
<tr>
<td>ey</td>
<td>veyz 26.28513</td>
<td>1.39955</td>
<td>18.78</td>
</tr>
<tr>
<td>ez</td>
<td>veyz 26.28513</td>
<td>1.39955</td>
<td>18.78</td>
</tr>
</tbody>
</table>

Covariances Among Exogenous Variables

<table>
<thead>
<tr>
<th>Var1</th>
<th>Var2</th>
<th>Parameter</th>
<th>Estimate</th>
<th>Error</th>
<th>t Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>fwx</td>
<td>fyz</td>
<td></td>
<td>1.00000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The goodness-of-fit tests for the four hypotheses are summarized in the following table.

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Number of Parameters</th>
<th>Degrees of Freedom</th>
<th>( \chi^2 )</th>
<th>p-value</th>
<th>( \hat{\rho} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( H_1 )</td>
<td>4</td>
<td>6</td>
<td>37.33</td>
<td>0.0000</td>
<td>1.0</td>
</tr>
<tr>
<td>( H_2 )</td>
<td>5</td>
<td>5</td>
<td>1.93</td>
<td>0.8583</td>
<td>0.8986</td>
</tr>
<tr>
<td>( H_3 )</td>
<td>8</td>
<td>2</td>
<td>36.21</td>
<td>0.0000</td>
<td>1.0</td>
</tr>
<tr>
<td>( H_4 )</td>
<td>9</td>
<td>1</td>
<td>0.70</td>
<td>0.4018</td>
<td>0.8986</td>
</tr>
</tbody>
</table>

The hypotheses \( H_1 \) and \( H_3 \), which posit \( \rho = 1 \), can be rejected. Hypotheses \( H_2 \) and \( H_4 \) seem to be consistent with the available data. Since \( H_2 \) is obtained by adding four constraints to \( H_1 \), you can test \( H_2 \) versus \( H_4 \) by computing the differences of the chi-square statistics and their degrees of freedom, yielding a chi-square of 1.23 with four degrees of freedom, which is obviously not significant. So hypothesis \( H_2 \) is consistent with the available data.

The estimates of \( \rho \) for \( H_2 \) and \( H_4 \) are almost identical, about 0.90, indicating that the speeded and unspeeded tests are measuring almost the same latent variable, even
though the hypotheses that stated they measured exactly the same latent variable are rejected.

### A Combined Measurement-Structural Model with Reciprocal Influence and Correlated Residuals

To illustrate a more complex model, this example uses some well-known data from Haller and Butterworth (1960). Various models and analyses of these data are given by Duncan, Haller, and Portes (1968), Jöreskog and Sörbom (1988), and Loehlin (1987).

The study is concerned with the career aspirations of high-school students and how these aspirations are affected by close friends. The data are collected from 442 seventeen-year-old boys in Michigan. There are 329 boys in the sample who named another boy in the sample as a best friend. The observations to be analyzed consist of the data from these 329 boys paired with the data from their best friends.

The method of data collection introduces two statistical problems. First, restricting the analysis to boys whose best friends are in the original sample causes the reduced sample to be biased. Second, since the data from a given boy may appear in two or more observations, the observations are not independent. Therefore, any statistical conclusions should be considered tentative. It is difficult to accurately assess the effects of the dependence of the observations on the analysis, but it could be argued on intuitive grounds that since each observation has data from two boys and since it seems likely that many of the boys will appear in the data set at least twice, the effective sample size may be as small as half of the reported 329 observations.

The correlation matrix is taken from Jöreskog and Sörbom (1988).

```sas
title 'Peer Influences on Aspiration: Haller & Butterworth (1960)';
data aspire(type=corr);
  _type_='corr';
  input _name_ $ riq rpa rses roa rea fiq fpa fses foa fea;
  label riq='Respondent: Intelligence'
    rpa='Respondent: Parental Aspiration'
    rses='Respondent: Family SES'
    roa='Respondent: Occupational Aspiration'
    rea='Respondent: Educational Aspiration'
    fiq='Friend: Intelligence'
    fpa='Friend: Parental Aspiration'
    fses='Friend: Family SES'
    foa='Friend: Occupational Aspiration'
    fea='Friend: Educational Aspiration';
datalines;
  riq    1.  . . . . . . . . . . .
rpa .1839  1. . . . . . . . . . . .
rses .2220 .0489  1. . . . . . . . . .
roa .4105 .2137 .3240  1. . . . . . . .
rea .4043 .2742 .6247  1. . . . . . . .
fiq .3355 .0782 .2302 .2995 .2863  1. . . . . .
fpa .1021 .1147 .0931 .0760 .0702 .2087  1. . . . . .
fses .1861 .0186 .2707 .2930 .2407 .2950 -0.0438  1. . . .
```

SAS OnlineDoc™: Version 8
The model analyzed by Jöreskog and Sörbom (1988) is displayed in the following path diagram:

Figure 14.17. Path Diagram: Career Aspiration, Jöreskog and Sörbom

Two latent variables, \( f_{-ramb} \) and \( f_{-famb} \), represent the respondent’s level of ambition and his best friend’s level of ambition, respectively. The model states that the respondent’s ambition is determined by his intelligence and socioeconomic status, his perception of his parents’ aspiration for him, and his friend’s socioeconomic status and ambition. It is assumed that his friend’s intelligence and socioeconomic status affect the respondent’s ambition only indirectly through his friend’s ambition. Ambition is indexed by the manifest variables of occupational and educational aspiration, which are assumed to have uncorrelated residuals. The path coefficient from ambition to occupational aspiration is set to 1.0 to determine the scale of the ambition latent variable.
This model can be analyzed with PROC CALIS using the LINEQS statement as follows, where the names of the parameters correspond to those used by Jöreskog and Sörbom (1988). Since this TYPE=CORR data set does not contain an observation with _TYPE_='N' giving the sample size, it is necessary to specify the degrees of freedom (sample size minus one) with the EDF= option in the PROC CALIS statement.

```sas
title2 'Joreskog-Sorbom (1988) analysis 1';
proc calis data=aspire edf=328;
  lineqs /* measurement model for aspiration */
    rea=lambda2 f_ramb + e_rea,
    roa=f_ramb + e_roa,
    fera=lambda3 f_famb + e_fea,
    foa=f_famb + e_foa,
    /* structural model of influences */
    f_ramb=gam1 rpa + gam2 riq + gam3 rses +
    gam4 fses + beta1 f_famb + d_ramb,
    f_famb=gam8 fpa + gam7 fiq + gam6 fses +
    gam5 rses + beta2 f_ramb + d_famb;
  std d_ramb=psi11,
    d_famb=psi22,
    e_rea e_roa e_fea e_foa=theta;
  cov d_ramb d_famb=psi12,
    rpa riq rses fpa fiq fses=cov;
run;
```

Specify a name followed by a colon to represent a list of names formed by appending numbers to the specified name. For example, in the COV statement, the line

```
rpa riq rses fpa fiq fses=cov:
```

is equivalent to

```
rpa riq rses fpa fiq fses=cov1-cov15;
```

The results from this analysis are as follows.
A Combined Measurement-Structural Model with Reciprocal Influence and Correlated Residuals

Peer Influences on Aspiration: Haller & Butterworth (1960)
Joereskog-Sorbom (1988) analysis 1

The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation

Fit Function 0.0814
Goodness of Fit Index (GFI) 0.9844
GFI Adjusted for Degrees of Freedom (AGFI) 0.9428
Root Mean Square Residual (RMR) 0.0202
Parsimonious GFI (Mulaik, 1989) 0.3281
Chi-Square 26.6972
Chi-Square DF 15
Pr > Chi-Square 0.0313
Independence Model Chi-Square 872.00
Independence Model Chi-Square DF 45
RMSEA Estimate 0.0488
RMSEA 90% Lower Confidence Limit 0.0145
RMSEA 90% Upper Confidence Limit 0.0783
ECVI Estimate 0.2959
ECVI 90% Lower Confidence Limit 0.2823
ECVI 90% Upper Confidence Limit 0.3721
Probability of Close Fit 0.4876
Bentler’s Comparative Fit Index 0.9859
Normal Theory Reweighted LS Chi-Square 26.0113
Akaike’s Information Criterion -3.3028
Bozdogan’s (1987) CAIC -75.2437
Schwarz’s Bayesian Criterion -60.2437
McDonald’s (1989) Centrality 0.9824
Bentler & Bonett’s (1980) Non-normed Index 0.9576
Bentler & Bonett’s (1980) NFI 0.9694
James, Mulaik, & Brett (1982) Parsimonious NFI 0.3231
Z-Test of Wilson & Hilferty (1931) 1.8625
Bollen (1986) Normed Index Rho1 0.9082
Bollen (1988) Non-normed Index Delta2 0.9864
Hoelter’s (1983) Critical N 309

Figure 14.18. Career Aspiration Data: J&S Analysis 1
Joereskog and Sörbom (1988) present more detailed results from a second analysis in which two constraints are imposed:

- The coefficients connecting the latent ambition variables are equal.
- The covariance of the disturbances of the ambition variables is zero.

This analysis can be performed by changing the names beta1 and beta2 to beta and omitting the line from the COV statement for psi12:

title2 'Joereskog-Sorbom (1988) analysis 2';
proc calis data=aspire edf=328;
  lineqs /* measurement model for aspiration */
    rea=lambda2 f_ramb + e_rea,
    roa=f_ramb + e_roa,
    fega=lambda3 f_famb + e_fea,
    foa=f_famb + e_foa,
  /* structural model of influences */
    f_ramb=gam1 rpa + gam2 riq + gam3 rses +
      gam4 fses + beta f_famb + d_ramb,
The results are displayed in Figure 14.19.


### The CALIS Procedure

**Covariance Structure Analysis: Maximum Likelihood Estimation**

- **Fit Function**: 0.0820
- **Goodness of Fit Index (GFI)**: 0.9843
- **GFI Adjusted for Degrees of Freedom (AGFI)**: 0.9492
- **Root Mean Square Residual (RMR)**: 0.0203
- ** Parsimonious GFI (Mulaik, 1989)**: 0.3718
- **Chi-Square**: 26.8987
- **Chi-Square DF**: 17
- **Pr > Chi-Square**: 0.0596
- **Independence Model Chi-Square**: 872.00
- **Independence Model Chi-Square DF**: 45
- **RMSEA Estimate**: 0.0421
- **RMSEA 90% Lower Confidence Limit**
- **RMSEA 90% Upper Confidence Limit**: 0.0710
- **ECVI Estimate**: 0.3592
- **ECVI 90% Lower Confidence Limit**: 0.2839
- **ECVI 90% Upper Confidence Limit**: 0.3592
- **Probability of Close Fit**: 0.6367
- **Bentler’s Comparative Fit Index**: 0.9880
- **Normal Theory Reweighted LS Chi-Square**: 26.1595
- **Akaike’s Information Criterion**: -88.6343
- **Bozdogan’s (1987) CAIC**: -88.6343
- **Schwarz’s Bayesian Criterion**: -71.6343
- **McDonald’s (1989) Centrality**: 0.9851
- **Bentler & Bonett’s (1980) Non-normed Index**: 0.9683
- **Bentler & Bonett’s (1980) NFI**: 0.9692
- **James, Mulaik, & Brett (1982) Parsimonious NFI**: 0.3661
- **Z-Test of Wilson & Hilferty (1931)**: 1.5599
- **Bollen (1986) Normed Index Rho1**: 0.9183
- **Bollen (1988) Non-normed Index Delta2**: 0.9884
- **Hoelter’s (1983) Critical N**: 338

**Figure 14.19.** Career Aspiration Data: J&S Analysis 2
A Combined Measurement-Structural Model with Reciprocal Influence and Correlated Residuals

Peer Influences on Aspiration: Haller & Butterworth (1960)
Joreskog-Sorbom (1988) analysis 2

The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation

\[ \begin{align*}
  \text{roa} & = 1.0000 \ f_{\text{ramb}} + 1.0000 \ e_{\text{roa}} \\
  \text{rea} & = 1.0610 \ f_{\text{ramb}} + 1.0000 \ e_{\text{rea}} \\
  \text{foa} & = 1.0000 \ f_{\text{famb}} + 1.0000 \ e_{\text{foa}} \\
  \text{fea} & = 1.0736 \ f_{\text{famb}} + 1.0000 \ e_{\text{fea}}
\end{align*} \]

\[ \begin{align*}
  \text{Std Err} & = 0.0892 \ \lambda_2 \\
  \text{t Value} & = 11.8923
\end{align*} \]

\[ \begin{align*}
  \text{foa} & = 1.0000 \ f_{\text{famb}} + 1.0000 \ e_{\text{foa}} \\
  \text{fea} & = 1.0736 \ f_{\text{famb}} + 1.0000 \ e_{\text{fea}}
\end{align*} \]

\[ \begin{align*}
  \text{Std Err} & = 0.0806 \ \lambda_3 \\
  \text{t Value} & = 13.3150
\end{align*} \]

Peer Influences on Aspiration: Haller & Butterworth (1960)
Joreskog-Sorbom (1988) analysis 2

The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation

\[ \begin{align*}
  f_{\text{ramb}} & = 0.1801 \ f_{\text{famb}} + 0.2540 \ r_{\text{riq}} + 0.1637 \ r_{\text{rpa}} + 0.2211 \ r_{\text{rses}} + 0.0773 \ f_{\text{fses}} + 1.0000 \ d_{\text{ramb}} \\
  \text{Std Err} & = 0.0391 \ \beta \ 0.0419 \ \gamma_2 \ 0.0387 \ \gamma_1 \\
  \text{t Value} & = 4.6031 \ 6.0673 \ 4.2274
\end{align*} \]

\[ \begin{align*}
  f_{\text{famb}} & = 0.1801 \ f_{\text{ramb}} + 0.0684 \ r_{\text{rses}} + 0.3306 \ r_{\text{fiq}} + 0.1520 \ r_{\text{rpa}} + 0.2184 \ f_{\text{fses}} + 1.0000 \ d_{\text{famb}} \\
  \text{Std Err} & = 0.0391 \ \beta \ 0.0387 \ \gamma_5 \ 0.0412 \ \gamma_7 \\
  \text{t Value} & = 4.6031 \ 1.7681 \ 8.0331
\end{align*} \]
The difference between the chi-square values for the two preceding models is 26.8987 - 26.6972 = 0.2015 with 2 degrees of freedom, which is far from significant. However, the chi-square test of the restricted model (analysis 2) against the alternative of a completely unrestricted covariance matrix yields a \( p \)-value of 0.0596, which indicates that the model may not be entirely satisfactory (\( p \)-values from these data are probably too small because of the dependence of the observations).

Loehlin (1987) points out that the models considered are unrealistic in at least two aspects. First, the variables of parental aspiration, intelligence, and socioeconomic status are assumed to be measured without error. Loehlin adds uncorrelated measurement errors to the model and assumes, for illustrative purposes, that the reliabilities of these variables are known to be 0.7, 0.8, and 0.9, respectively. In practice, these reliabilities would need to be obtained from a separate study of the same or a very similar population. If these constraints are omitted, the model is not identified. However,
constraining parameters to a constant in an analysis of a correlation matrix may make
the chi-square goodness-of-fit test inaccurate, so there is more reason to be skeptical
of the \( p \)-values. Second, the error terms for the respondent’s aspiration are assumed
to be uncorrelated with the corresponding terms for his friend. Loehlin introduces a
correlation between the two educational aspiration error terms and between the two
occupational aspiration error terms. These additions produce the following path dia-
gram for Loehlin’s model 1.

Figure 14.20. Path Diagram: Carrer Aspiration – Loehlin

The statements for fitting this model are as follows:

```sas
title2 'Loehlin (1987) analysis: Model 1';
proc calis data=aspire edf=328;
lineqs /* measurement model for aspiration */
  rea=lambda2 f_ramb + e_rea,
  roa=f_ramb + e_roa,
  fpa=lambda3 f_famb + e_fpa,
  fiq=f_famb + e_fiq,
  fse=.837 f_fpa + e_fpa,
  fpiq=.894 f_fiq + e_fiq,
  fses=.949 f_fses + e_fses,
```
Chapter 14. Introduction to Structural Equations with Latent Variables

/* structural model of influences */

f_ramb=gam1 f_rpa + gam2 f_riq + gam3 f_rses +
gam4 f_fses + bet1 f_famb + d_ramb,
f_famb=gam8 f_fpa + gam7 f_fiq + gam6 f_fses +
gam5 f_rses + bet2 f_ramb + d_famb;

std d_ramb=psi11,
d_famb=psi22,
f_rpa f_riq f_rses f_fpa f_fiq f_fses=1,
e_rea e_roa e_fea e_foa=theta:,
e_rpa e_riq e_rses e_fpa e_fiq e_fses=err:;
cov d_ramb d_famb=psi12,
e_rea e_fea=covea,
e_roa e_foa=covoa,
f_rpa f_riq f_rses f_fpa f_fiq f_fses=cov:;

run;

The results are displayed in Figure 14.21.

Peer Influences on Aspiration: Haller & Butterworth (1960)
Loehlin (1987) analysis: Model 1

The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation

Fit Function 0.0366
Goodness of Fit Index (GFI) 0.9927
GFI Adjusted for Degrees of Freedom (AGFI) 0.9692
Root Mean Square Residual (RMR) 0.0149
Parsimonious GFI (Mulaik, 1989) 0.2868
Chi-Square 12.0132
Chi-Square DF 13
Pr > Chi-Square 0.5266
Independence Model Chi-Square 872.00
Independence Model Chi-Square DF 45
RMSEA Estimate 0.0000
RMSEA 90% Lower Confidence Limit .
RMSEA 90% Upper Confidence Limit 0.0512
ECVI Estimate 0.3016
ECVI 90% Lower Confidence Limit .
ECVI 90% Upper Confidence Limit 0.3392
Probability of Close Fit 0.9435
Bentler’s Comparative Fit Index 1.0000
Normal Theory Reweighted LS Chi-Square 12.0168
Akaike’s Information Criterion -13.9868
Bozdogan’s (1987) CAIC -76.3356
Schwarz’s Bayesian Criterion -63.3356
McDonald’s (1989) Centrality 1.0015
Bentler & Bonett’s (1980) Non-normed Index 1.0041
Bentler & Bonett’s (1980) NFI 0.9862
James, Mulaik, & Brett (1982) Parsimonious NFI 0.2849
Z-Test of Wilson & Hilferty (1931) -0.0679
Bollen (1986) Normed Index Rho1 0.9523
Bollen (1988) Non-normed Index Delta2 1.0011
Hoelter’s (1983) Critical N 612

Figure 14.21. Career Aspiration Data: Loehlin Model 1
Peer Influences on Aspiration: Haller & Butterworth (1960)
Loehlin (1987) analysis: Model 1

The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation

\[
\begin{align*}
\text{riq} & = 0.8940 \ f_{\text{riq}} + 1.0000 \ e_{\text{riq}} \\
\text{rpa} & = 0.8370 \ f_{\text{rpa}} + 1.0000 \ e_{\text{rpa}} \\
\text{rses} & = 0.9490 \ f_{\text{rses}} + 1.0000 \ e_{\text{rses}} \\
\text{roa} & = 1.0000 \ f_{\text{roa}} + 1.0000 \ e_{\text{roa}} \\
\text{rea} & = 1.0840* f_{\text{ramb}} + 1.0000 \ e_{\text{rea}} \\
\text{Std Err} & = 0.0942 \ lambda2 \\
\text{t Value} & = 11.5105
\end{align*}
\]

Peer Influences on Aspiration: Haller & Butterworth (1960)
Loehlin (1987) analysis: Model 1

The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation

\[
\begin{align*}
\text{f_ramb} & = 0.1190* f_{\text{famb}} + 0.1837* f_{\text{rpa}} + 0.2800* f_{\text{riq}} \\
\text{Std Err} & = 0.1140 \ \text{bet1} 0.0504 \ \text{gam1} 0.0614 \ \text{gam2} \\
\text{t Value} & = 1.0440 3.6420 4.5618 \\
+ & 0.2262* f_{\text{rses}} + 0.0870* f_{\text{fses}} + 1.0000 \ d_{\text{ramb}} \\
0.0522 \ \text{gam3} & = 0.0548 \ \text{gam4} \\
4.3300 & = 1.5884 \\
\text{f_famb} & = 0.1302* f_{\text{famb}} + 0.0633* f_{\text{rses}} + 0.1688* f_{\text{fpa}} \\
\text{Std Err} & = 0.1207 \ \text{bet2} 0.0522 \ \text{gam5} 0.0493 \ \text{gam8} \\
\text{t Value} & = 1.0792 1.2124 3.4205 \\
+ & 0.3529* f_{\text{fiq}} + 0.2154* f_{\text{fses}} + 1.0000 \ d_{\text{famb}} \\
0.0674 \ \text{gam7} & = 0.0512 \ \text{gam6} \\
5.2497 & = 4.2060
\end{align*}
\]
Since the $p$-value for the chi-square test is 0.5266, this model clearly cannot be rejected. However, Schwarz’s Bayesian Criterion for this model ($SBC = -63.3356$) is somewhat larger than for Jöreskog and Sörbom’s (1988) analysis 2 ($SBC = -71.6343$), suggesting that a more parsimonious model would be desirable.

Since it is assumed that the same model applies to all the boys in the sample, the path diagram should be symmetric with respect to the respondent and friend. In particular, the corresponding coefficients should be equal. By imposing equality constraints
on the 15 pairs of corresponding coefficients, this example obtains Loehlin’s model 2. The LINEQS model is as follows, where an OUTRAM= data set is created to facilitate subsequent hypothesis tests:

```
title2 'Loehlin (1987) analysis: Model 2';
proc calis data=aspire edf=328 outram=ram2;
  lineqs /* measurement model for aspiration */
    rea= lambda f_ramb + e_rea, /* 1 ec! */
    roa= f_ramb + e_roa,
    fبا= lambda f_famb + e_fea,
    foa= f_famb + e_foa,
    /* measurement model for intelligence and environment */
    rpa=.837 f_rpa + e_rpa,
    riq=.894 f_riq + e_riq,
    rses=.949 f_rses + e_rses,
    fpa=.837 f_fpa + e_fpa,
    fiq=.894 f_fiq + e_fiq,
    fses=.949 f_fses + e_fses,
    /* structural model of influences */ /* 5 ec! */
    f_ramb= gam1 f_rpa + gam2 f_riq + gam3 f_rses +
           gam4 f_fses + beta f_famb + d_ramb,
    f_famb= gam1 f_fpa + gam2 f_fiq + gam3 f_fses +
           gam4 f_fses + beta f_ramb + d_famb;
  std d_ramb=psi, /* 1 ec! */
  std d_famb=psi,
  f_rpa f_riq f_rses f_fpa f_fiq f_fses=1,
  e_rea e_fea=thetaea thetaea, /* 2 ec! */
  e_roa e_foa=thetaoa thetaoa,
  e_rpa e_fpa=errpa1 errpa2,
  e_riq e_fiq=errriq1 errriq2,
  e_rses e_fses=errses1 errses2;
  cov d_ramb d_famb=psi12,
  e_rea e_fea=covea,
  e_roa e_foa = covoa,
  f_rpa f_riq f_rses=cov1-cov3, /* 3 ec! */
  f_fpa f_fiq f_fses=cov1-cov3,
  f_rpa f_riq f_rses * f_fpa f_fiq f_fses = /* 3 ec! */
    cov4 cov5 cov6
    cov5 cov7 cov8
    cov6 cov8 cov9;
run;
```

The results are displayed in Figure 14.22.
Peer Influences on Aspiration: Haller & Butterworth (1960)
Loehlin (1987) analysis: Model 2

The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation

Fit Function 0.0581
Goodness of Fit Index (GFI) 0.9884
GFI Adjusted for Degrees of Freedom (AGFI) 0.9772
Root Mean Square Residual (RMR) 0.0276
 Parsimonious GFI (Mulaik, 1989) 0.6150
Chi-Square 19.0697
Chi-Square DF 28
Pr > Chi-Square 0.8960
Independence Model Chi-Square 872.00
Independence Model Chi-Square DF 45
RMSEA Estimate 0.0000
RMSEA 90% Lower Confidence Limit .
RMSEA 90% Upper Confidence Limit 0.0194
ECVI Estimate 0.2285
ECVI 90% Lower Confidence Limit .
ECVI 90% Upper Confidence Limit 0.2664
Probability of Close Fit 0.9996
Bentler’s Comparative Fit Index 1.0000
Normal Theory Reweighted LS Chi-Square 19.2372
Akaike’s Information Criterion -36.9303
Bozdogan’s (1987) CAIC -171.2200
Schwarz’s Bayesian Criterion -143.2200
McDonald’s (1989) Centrality 1.0137
Bentler & Bonett’s (1980) Non-normed Index 1.0174
Bentler & Bonett’s (1980) NFI 0.9781
James, Mulaik, & Brett (1982) Parsimonious NFI 0.6086
Z-Test of Wilson & Hilferty (1931) -1.2599
Bollen (1986) Normed Index Rho1 0.9649
Bollen (1988) Non-normed Index Delta2 1.0106
Hoelter’s (1983) Critical N 713

Figure 14.22. Career Aspiration Data: Loehlin Model 2
A Combined Measurement-Structural Model with Reciprocal Influence and Correlated Residuals

Peer Influences on Aspiration: Haller & Butterworth (1960)
Loehlin (1987) analysis: Model 2

The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation

\[
\begin{align*}
  r_{iq} &= 0.8940 f_{riq} + 1.0000 e_{riq} \\
  r_{pa} &= 0.8370 f_{rpa} + 1.0000 e_{rpa} \\
  r_{ses} &= 0.9490 f_{rses} + 1.0000 e_{rses} \\
  r_{oa} &= 1.0000 f_{ramb} + 1.0000 e_{roa} \\
  r_{ea} &= 1.1007 f_{ramb} + 1.0000 e_{rea} \\
  \text{Std Err} &= 0.0684 \lambda \\
  \text{t Value} &= 16.0879
\end{align*}
\]

Peer Influences on Aspiration: Haller & Butterworth (1960)
Loehlin (1987) analysis: Model 2

The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation

\[
\begin{align*}
  f_{ramb} &= 0.1158 f_{famb} + 0.1758 f_{rpa} + 0.3223 f_{riq} \\
  \text{Std Err} &= 0.0839 \beta + 0.0351 \gamma_1 + 0.0470 \gamma_2 \\
  \text{t Value} &= 1.3801 5.0130 6.8557 \\
  &+ 0.2227 f_{rses} + 0.0756 f_{fses} + 1.0000 d_{ramb} \\
  &0.0363 \gamma_3 + 0.0375 \gamma_4 \\
  &6.1373 2.0170
\end{align*}
\]

\[
\begin{align*}
  f_{famb} &= 0.1158 f_{ramb} + 0.0756 f_{rses} + 0.1758 f_{fpa} \\
  \text{Std Err} &= 0.0839 \beta + 0.0375 \gamma_4 + 0.0351 \gamma_1 \\
  \text{t Value} &= 1.3801 2.0170 5.0130 \\
  &+ 0.3223 f_{fiq} + 0.2227 f_{fses} + 1.0000 d_{famb} \\
  &0.0470 \gamma_2 + 0.0363 \gamma_3 \\
  &6.8557 6.1373
\end{align*}
\]
Peer Influences on Aspiration: Haller & Butterworth (1960)
Loehlin (1987) analysis: Model 2

The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter</th>
<th>Estimate</th>
<th>Error</th>
<th>t Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>f_rpa</td>
<td></td>
<td>1.00000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>f_riq</td>
<td></td>
<td>1.00000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>f_rses</td>
<td></td>
<td>1.00000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>f_fpa</td>
<td></td>
<td>1.00000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>f_fiq</td>
<td></td>
<td>1.00000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>e_rea</td>
<td>thetaea</td>
<td>0.30662</td>
<td>0.03726</td>
<td>8.23</td>
</tr>
<tr>
<td>e_roa</td>
<td>thetaoa</td>
<td>0.42295</td>
<td>0.03651</td>
<td>11.58</td>
</tr>
<tr>
<td>e_fea</td>
<td>thetaea</td>
<td>0.30662</td>
<td>0.03726</td>
<td>8.23</td>
</tr>
<tr>
<td>e_roa</td>
<td>thetaoa</td>
<td>0.42295</td>
<td>0.03651</td>
<td>11.58</td>
</tr>
<tr>
<td>e_rpa</td>
<td>errpa1</td>
<td>0.30758</td>
<td>0.07511</td>
<td>4.09</td>
</tr>
<tr>
<td>e_fpa</td>
<td>errpa2</td>
<td>0.28834</td>
<td>0.07369</td>
<td>3.91</td>
</tr>
<tr>
<td>e_fiq</td>
<td>errfpa1</td>
<td>0.15573</td>
<td>0.06700</td>
<td>2.32</td>
</tr>
<tr>
<td>e_fses</td>
<td>errses1</td>
<td>0.11467</td>
<td>0.07267</td>
<td>1.58</td>
</tr>
<tr>
<td>d_ramb</td>
<td>psi</td>
<td>0.22456</td>
<td>0.02971</td>
<td>7.56</td>
</tr>
<tr>
<td>d_famb</td>
<td>psi</td>
<td>0.22456</td>
<td>0.02971</td>
<td>7.56</td>
</tr>
</tbody>
</table>

Covariances Among Exogenous Variables

<table>
<thead>
<tr>
<th>Var1</th>
<th>Var2</th>
<th>Parameter</th>
<th>Estimate</th>
<th>Error</th>
<th>t Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>f_rpa</td>
<td>f_riq</td>
<td>cov1</td>
<td>0.26470</td>
<td>0.05442</td>
<td>4.86</td>
</tr>
<tr>
<td>f_rpa</td>
<td>f_rses</td>
<td>cov2</td>
<td>0.00176</td>
<td>0.04996</td>
<td>0.04</td>
</tr>
<tr>
<td>f_riq</td>
<td>f_rses</td>
<td>cov3</td>
<td>0.31129</td>
<td>0.05057</td>
<td>6.16</td>
</tr>
<tr>
<td>f_rpa</td>
<td>f_fpa</td>
<td>cov4</td>
<td>0.18784</td>
<td>0.07872</td>
<td>2.39</td>
</tr>
<tr>
<td>f_riq</td>
<td>f_fpa</td>
<td>cov5</td>
<td>0.11837</td>
<td>0.05447</td>
<td>2.17</td>
</tr>
<tr>
<td>f_rses</td>
<td>f_fpa</td>
<td>cov6</td>
<td>0.06910</td>
<td>0.04996</td>
<td>1.38</td>
</tr>
<tr>
<td>f_rpa</td>
<td>f_fiq</td>
<td>cov7</td>
<td>0.43061</td>
<td>0.07258</td>
<td>5.93</td>
</tr>
<tr>
<td>f_rses</td>
<td>f_fiq</td>
<td>cov8</td>
<td>0.24967</td>
<td>0.05060</td>
<td>4.93</td>
</tr>
<tr>
<td>f_fpa</td>
<td>f_fiq</td>
<td>cov9</td>
<td>0.26470</td>
<td>0.05442</td>
<td>4.86</td>
</tr>
<tr>
<td>f_rpa</td>
<td>f_fses</td>
<td>cov10</td>
<td>0.06910</td>
<td>0.04996</td>
<td>1.38</td>
</tr>
<tr>
<td>f_rses</td>
<td>f_fses</td>
<td>cov11</td>
<td>0.24967</td>
<td>0.05060</td>
<td>4.93</td>
</tr>
<tr>
<td>f_fpa</td>
<td>f_fses</td>
<td>cov12</td>
<td>0.30190</td>
<td>0.06362</td>
<td>4.75</td>
</tr>
<tr>
<td>f_fiq</td>
<td>f_fses</td>
<td>cov13</td>
<td>0.00176</td>
<td>0.04996</td>
<td>0.04</td>
</tr>
<tr>
<td>e_rea</td>
<td>e_fea</td>
<td>covea</td>
<td>0.01260</td>
<td>0.03144</td>
<td>0.04</td>
</tr>
<tr>
<td>e_roa</td>
<td>e_foa</td>
<td>covoa</td>
<td>0.11208</td>
<td>0.03257</td>
<td>3.44</td>
</tr>
<tr>
<td>d_ramb</td>
<td>d_famb</td>
<td>psi12</td>
<td>-0.00344</td>
<td>0.04931</td>
<td>-0.07</td>
</tr>
</tbody>
</table>

The test of Loehlin’s model 2 against model 1 yields a chi-square of 19.0697 - 12.0132 = 7.0565 with 15 degrees of freedom, which is clearly not significant. Schwarz’s Bayesizn Criterion (SBC) is also much lower for model 2 (-143.2200) than model 1 (-63.3356). Hence, model 2 seems preferable on both substantive and statistical grounds.

A question of substantive interest is whether the friend’s socioeconomic status (SES) has a significant direct influence on a boy’s ambition. This can be addressed by
omitting the paths from $f_{rse}$ to $f_{ramb}$ and from $f_{rse}$ to $f_{fam}$ designated by the parameter name $gam4$, yielding Loehlin's model 3:

```sas
title2 'Loehlin (1987) analysis: Model 3';
data ram3(type=ram);
set ram2;
if _name_='gam4' then
  do;
    _name_=' ';
    _estim_=0;
  end;
run;
proc calis data=aspire edf=328 inram=ram3;
run;
```

The output is displayed in Figure 14.23.

---

**Peer Influences on Aspiration: Haller & Butterworth (1960)**

Loehlin (1987) analysis: Model 3

The CALIS Procedure

Covariance Structure Analysis: Maximum Likelihood Estimation

<table>
<thead>
<tr>
<th>Fit Function</th>
<th>0.0702</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goodness of Fit Index (GFI)</td>
<td>0.9858</td>
</tr>
<tr>
<td>GFI Adjusted for Degrees of Freedom (AGFI)</td>
<td>0.9731</td>
</tr>
<tr>
<td>Root Mean Square Residual (RMR)</td>
<td>0.0304</td>
</tr>
<tr>
<td>Parsimonious GFI (Mulaik, 1989)</td>
<td>0.6353</td>
</tr>
<tr>
<td>Chi-Square</td>
<td>23.0365</td>
</tr>
<tr>
<td>Chi-Square DF</td>
<td></td>
</tr>
<tr>
<td>Pr &gt; Chi-Square</td>
<td>0.7749</td>
</tr>
<tr>
<td>Independence Model Chi-Square</td>
<td>872.00</td>
</tr>
<tr>
<td>Independence Model Chi-Square DF</td>
<td>45</td>
</tr>
<tr>
<td>RMSEA Estimate</td>
<td>0.0000</td>
</tr>
<tr>
<td>RMSEA 90% Lower Confidence Limit</td>
<td></td>
</tr>
<tr>
<td>RMSEA 90% Upper Confidence Limit</td>
<td>0.0295</td>
</tr>
<tr>
<td>ECVI Estimate</td>
<td>0.2343</td>
</tr>
<tr>
<td>ECVI 90% Lower Confidence Limit</td>
<td></td>
</tr>
<tr>
<td>ECVI 90% Upper Confidence Limit</td>
<td>0.2780</td>
</tr>
<tr>
<td>Probability of Close Fit</td>
<td>0.9984</td>
</tr>
<tr>
<td>Bentler's Comparative Fit Index</td>
<td>1.0000</td>
</tr>
<tr>
<td>Normal Theory Reweighted LS Chi-Square</td>
<td>23.5027</td>
</tr>
<tr>
<td>Akaike's Information Criterion</td>
<td>-34.9635</td>
</tr>
<tr>
<td>Bozdogan’s (1987) CAIC</td>
<td>-174.0492</td>
</tr>
<tr>
<td>Schwarz’s Bayesian Criterion</td>
<td>-145.0492</td>
</tr>
<tr>
<td>McDonald’s (1989) Centrality</td>
<td>1.0091</td>
</tr>
<tr>
<td>Bentler &amp; Bonett’s (1980) Non-normed Index</td>
<td>1.0112</td>
</tr>
<tr>
<td>Bentler &amp; Bonett’s (1980) NFI</td>
<td>0.9736</td>
</tr>
<tr>
<td>James, Mulaik, &amp; Brett (1982) Parsimonious NFI</td>
<td>0.6274</td>
</tr>
<tr>
<td>Z-Test of Wilson &amp; Hilferty (1931)</td>
<td>-0.7563</td>
</tr>
<tr>
<td>Bollen (1986) Normed Index Rho1</td>
<td>0.9590</td>
</tr>
<tr>
<td>Bollen (1988) Non-normed Index Delta2</td>
<td>1.0071</td>
</tr>
<tr>
<td>Hoelter’s (1983) Critical N</td>
<td>607</td>
</tr>
</tbody>
</table>

Figure 14.23. Career Aspiration Data: Loehlin Model 3

The chi-square value for testing model 3 versus model 2 is $23.0365 - 19.0697 = 3.9668$ with 1 degree of freedom and a $p$-value of 0.0464. Although the parameter is of marginal significance, the estimate in model 2 (0.0756) is small compared to the other coefficients, and SBC indicates that model 3 is preferable to model 2.
Another important question is whether the reciprocal influences between the respondent’s and friend’s ambitions are needed in the model. To test whether these paths are zero, set the parameter $\beta$ for the paths linking \(f \rightarrow ramb\) and \(f \rightarrow famb\) to zero to obtain Loehlin’s model 4:

\[
\text{title2 'Loehlin (1987) analysis: Model 4';}
\text{data ram4(type=ram);}
\text{set ram2;}
\text{if _name_='beta' then}
\text{do;}
\text{  _name_=' ';}
\text{  _estim_=0;}
\text{end;}
\text{run;}
\]

\[
\text{proc calis data=aspire edf=328 inram=ram4;}
\text{run;}
\]

The output is displayed in Figure 14.24.

![Figure 14.24. Career Aspiration Data: Loehlin Model 4](SAS OnlineDoc™: Version 8)
A Combined Measurement-Structural Model with Reciprocal Influence and Correlated Residuals

Peer Influences on Aspiration: Haller & Butterworth (1960) 
Loehlin (1987) analysis: Model 4

The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation

\[ \begin{align*}
\text{riq} &= 0.8940 \ f_{\text{riq}} + 1.0000 \ e_{\text{riq}} \\
\text{rpa} &= 0.8370 \ f_{\text{rpa}} + 1.0000 \ e_{\text{rpa}} \\
\text{rses} &= 0.9490 \ f_{\text{rses}} + 1.0000 \ e_{\text{rses}} \\
\text{roa} &= 1.0000 \ f_{\text{roa}} + 1.0000 \ e_{\text{roa}} \\
\text{rea} &= 1.1051 \ f_{\text{roa}} + 1.0000 \ e_{\text{rea}} \\
\end{align*} \]

\[ \text{Std Err} = 0.0680 \text{ lambda} \]
\[ \text{t Value} = 16.2416 \]

\[ \begin{align*}
\text{fiq} &= 0.8940 \ f_{\text{fiq}} + 1.0000 \ e_{\text{fiq}} \\
\text{fpa} &= 0.8370 \ f_{\text{fpa}} + 1.0000 \ e_{\text{fpa}} \\
\text{fses} &= 0.9490 \ f_{\text{fses}} + 1.0000 \ e_{\text{fses}} \\
\text{foa} &= 1.0000 \ f_{\text{foa}} + 1.0000 \ e_{\text{foa}} \\
\text{fea} &= 1.1051 \ f_{\text{foa}} + 1.0000 \ e_{\text{fea}} \\
\end{align*} \]

\[ \text{Std Err} = 0.0680 \text{ lambda} \]
\[ \text{t Value} = 16.2416 \]

The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation

\[ \begin{align*}
f_{\text{ramb}} &= 0 \ f_{\text{ramb}} + 0.1776 \ f_{\text{rpa}} + 0.3486 \ f_{\text{riq}} \\
\text{Std Err} &= 0.0301 \text{ gam1} + 0.0361 \text{ gam2} \\
\text{t Value} &= 4.9195 + 7.5362 \\
&+ 0.2383 \ f_{\text{rses}} + 0.1081 \ f_{\text{fses}} + 1.0000 \ d_{\text{ramb}} \\
&0.0355 \text{ gam3} + 0.0299 \text{ gam4} \\
&6.7158 + 3.6134 \\
\text{f_famb} &= 0 \ f_{\text{ramb}} + 0.1081 \ f_{\text{rses}} + 0.1776 \ f_{\text{fpa}} \\
\text{Std Err} &= 0.0299 \text{ gam4} + 0.0361 \text{ gam1} \\
\text{t Value} &= 3.6134 + 4.9195 \\
&+ 0.3486 \ f_{\text{fiq}} + 0.2383 \ f_{\text{fses}} + 1.0000 \ d_{\text{famb}} \\
&0.0463 \text{ gam2} + 0.0355 \text{ gam3} \\
&7.5362 + 6.7158
\end{align*} \]
Peer Influences on Aspiration: Haller & Butterworth (1960)  
Loehlin (1987) analysis: Model 4

The CALIS Procedure  
Covariance Structure Analysis: Maximum Likelihood Estimation

### Variances of Exogenous Variables

<table>
<thead>
<tr>
<th>Variable Parameter</th>
<th>Standard Error</th>
<th>t Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>f_rpa</td>
<td>1.00000</td>
<td></td>
</tr>
<tr>
<td>f_riq</td>
<td>1.00000</td>
<td></td>
</tr>
<tr>
<td>f_rses</td>
<td>1.00000</td>
<td></td>
</tr>
<tr>
<td>f_fpa</td>
<td>1.00000</td>
<td></td>
</tr>
<tr>
<td>f_fiq</td>
<td>1.00000</td>
<td></td>
</tr>
<tr>
<td>f_fses</td>
<td>1.00000</td>
<td></td>
</tr>
<tr>
<td>e_rea thetaea</td>
<td>0.30502</td>
<td>8.18</td>
</tr>
<tr>
<td>e_roa thetaoa</td>
<td>0.42429</td>
<td>11.64</td>
</tr>
<tr>
<td>e_fea thetaea</td>
<td>0.30502</td>
<td>8.18</td>
</tr>
<tr>
<td>e_foa thetaoa</td>
<td>0.42429</td>
<td>11.64</td>
</tr>
<tr>
<td>e_rpa errp1</td>
<td>0.31354</td>
<td>4.16</td>
</tr>
<tr>
<td>e_riq erriq1</td>
<td>0.29611</td>
<td>4.06</td>
</tr>
<tr>
<td>e_rses errses1</td>
<td>0.12320</td>
<td>1.69</td>
</tr>
<tr>
<td>e_fpa errpa2</td>
<td>0.29051</td>
<td>3.94</td>
</tr>
<tr>
<td>e_fiq erriq2</td>
<td>0.18181</td>
<td>2.75</td>
</tr>
<tr>
<td>e_fses errses2</td>
<td>0.09873</td>
<td>1.39</td>
</tr>
<tr>
<td>d_ramb psi</td>
<td>0.22738</td>
<td>7.24</td>
</tr>
<tr>
<td>d_famb psi</td>
<td>0.22738</td>
<td>7.24</td>
</tr>
</tbody>
</table>

### Covariances Among Exogenous Variables

<table>
<thead>
<tr>
<th>Var1</th>
<th>Var2</th>
<th>Parameter</th>
<th>Standard Error</th>
<th>t Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>f_rpa</td>
<td>f_riq</td>
<td>cov1</td>
<td>0.27241</td>
<td>4.94</td>
</tr>
<tr>
<td>f_rpa</td>
<td>f_rses</td>
<td>cov2</td>
<td>0.00476</td>
<td>0.09</td>
</tr>
<tr>
<td>f_riq</td>
<td>f_rses</td>
<td>cov3</td>
<td>0.32463</td>
<td>6.38</td>
</tr>
<tr>
<td>f_rpa</td>
<td>f_fpa</td>
<td>cov4</td>
<td>0.16949</td>
<td>2.16</td>
</tr>
<tr>
<td>f_riq</td>
<td>f_fpa</td>
<td>cov5</td>
<td>0.13539</td>
<td>2.50</td>
</tr>
<tr>
<td>f_rses</td>
<td>f_fpa</td>
<td>cov6</td>
<td>0.07362</td>
<td>1.46</td>
</tr>
<tr>
<td>f_rpa</td>
<td>f_fiq</td>
<td>cov5</td>
<td>0.13539</td>
<td>2.50</td>
</tr>
<tr>
<td>f_riq</td>
<td>f_fiq</td>
<td>cov7</td>
<td>0.46893</td>
<td>6.72</td>
</tr>
<tr>
<td>f_rses</td>
<td>f_fiq</td>
<td>cov8</td>
<td>0.26289</td>
<td>5.16</td>
</tr>
<tr>
<td>f_fpa</td>
<td>f_fiq</td>
<td>cov1</td>
<td>0.27241</td>
<td>4.94</td>
</tr>
<tr>
<td>f_rpa</td>
<td>f_fses</td>
<td>cov6</td>
<td>0.07362</td>
<td>1.46</td>
</tr>
<tr>
<td>f_riq</td>
<td>f_fses</td>
<td>cov8</td>
<td>0.26289</td>
<td>5.16</td>
</tr>
<tr>
<td>f_rses</td>
<td>f_fses</td>
<td>cov9</td>
<td>0.30880</td>
<td>4.82</td>
</tr>
<tr>
<td>d_ramb d_famb</td>
<td>psi12</td>
<td>0.05479</td>
<td>2.03</td>
<td></td>
</tr>
</tbody>
</table>
The chi-square value for testing model 4 versus model 2 is $20.9981 - 19.0697 = 1.9284$ with 1 degree of freedom and a $p$-value of 0.1649. Hence, there is little evidence of reciprocal influence.

Loehlin’s model 2 has not only the direct paths connecting the latent ambition variables $f_{\text{ramb}}$ and $f_{\text{famb}}$ but also a covariance between the disturbance terms $d_{\text{ramb}}$ and $d_{\text{famb}}$ to allow for other variables omitted from the model that might jointly influence the respondent and his friend. To test the hypothesis that this covariance is zero, set the parameter $\psi_{12}$ to zero, yielding Loehlin’s model 5:

```sas

title2 'Loehlin (1987) analysis: Model 5';
data ram5(type=ram);
   set ram2;
   if _name_='psi12' then
      do;
         _name_=' ;
         _estim_=0;
      end;
   run;

   proc calis data=aspire edf=328 inram=ram5;
   run;
```

The output is displayed in Figure 14.25.
### Peer Influences on Aspiration: Haller & Butterworth (1960)
Loehlin (1987) analysis: Model 5

The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation

<table>
<thead>
<tr>
<th>Metric</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fit Function</td>
<td>0.0582</td>
</tr>
<tr>
<td>Goodness of Fit Index (GFI)</td>
<td>0.9884</td>
</tr>
<tr>
<td>GFI Adjusted for Degrees of Freedom (AGFI)</td>
<td>0.9780</td>
</tr>
<tr>
<td>Root Mean Square Residual (RMR)</td>
<td>0.0276</td>
</tr>
<tr>
<td>Parsimonious GFI (Mulaik, 1989)</td>
<td>0.6370</td>
</tr>
<tr>
<td>Chi-Square</td>
<td>19.0745</td>
</tr>
<tr>
<td>Chi-Square DF</td>
<td>29</td>
</tr>
<tr>
<td>Pr &gt; Chi-Square</td>
<td>0.9194</td>
</tr>
<tr>
<td>Independence Model Chi-Square</td>
<td>872.00</td>
</tr>
<tr>
<td>Independence Model Chi-Square DF</td>
<td>45</td>
</tr>
<tr>
<td>RMSEA Estimate</td>
<td>0.0000</td>
</tr>
<tr>
<td>RMSEA 90% Lower Confidence Limit</td>
<td>0.0152</td>
</tr>
<tr>
<td>ECVI Estimate</td>
<td>0.2222</td>
</tr>
<tr>
<td>ECVI 90% Lower Confidence Limit</td>
<td>0.2592</td>
</tr>
<tr>
<td>Probability of Close Fit</td>
<td>0.9998</td>
</tr>
<tr>
<td>Bentler’s Comparative Fit Index</td>
<td>1.0000</td>
</tr>
<tr>
<td>Normal Theory Reweighted LS Chi-Square</td>
<td>19.2269</td>
</tr>
<tr>
<td>Akaike’s Information Criterion</td>
<td>-38.9255</td>
</tr>
<tr>
<td>Bozdogan’s (1987) CAIC</td>
<td>-178.0111</td>
</tr>
<tr>
<td>Schwarz’s Bayesian Criterion</td>
<td>-149.0111</td>
</tr>
<tr>
<td>McDonald’s (1989) Centrality</td>
<td>1.0152</td>
</tr>
<tr>
<td>Bentler &amp; Bonett’s (1980) Non-normed Index</td>
<td>1.0186</td>
</tr>
<tr>
<td>Bentler &amp; Bonett’s (1980) NFI</td>
<td>0.9781</td>
</tr>
<tr>
<td>James, Mulaik, &amp; Brett (1982) Parsimonious NFI</td>
<td>0.6303</td>
</tr>
<tr>
<td>Z-Test of Wilson &amp; Hilferty (1931)</td>
<td>-1.4014</td>
</tr>
<tr>
<td>Bollen (1986) Normed Index Rho1</td>
<td>0.9661</td>
</tr>
<tr>
<td>Bollen (1988) Non-normed Index Delta2</td>
<td>1.0118</td>
</tr>
<tr>
<td>Hoelter’s (1983) Critical N</td>
<td>733</td>
</tr>
</tbody>
</table>

**Figure 14.25.** Career Aspiration Data: Loehlin Model 5
A Combined Measurement-Structural Model with Reciprocal Influence and Correlated Residuals

Peer Influences on Aspiration: Haller & Butterworth (1960)
Loehlin (1987) analysis: Model 5

The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation

\[
\begin{align*}
riq &= 0.8940 f_{riq} + 1.0000 e_{riq} \\
rpa &= 0.8370 f_{rpa} + 1.0000 e_{rpa} \\
rses &= 0.9490 f_{rses} + 1.0000 e_{rses} \\
roa &= 1.0000 f_{ramb} + 1.0000 e_{roa} \\
rea &= 1.1009 f_{ramb} + 1.0000 e_{rea} \\
\end{align*}
\]

Std Err 0.0684 lambda

\[
\begin{align*}
fiq &= 0.8940 f_{fiq} + 1.0000 e_{fiq} \\
fpa &= 0.8370 f_{fpa} + 1.0000 e_{fpa} \\
fses &= 0.9490 f_{fses} + 1.0000 e_{fses} \\
foa &= 1.0000 f_{famb} + 1.0000 e_{foa} \\
fea &= 1.1009 f_{famb} + 1.0000 e_{fea} \\
\end{align*}
\]

Std Err 0.0684 lambda

\[
\begin{align*}
\text{t Value} &= 16.1041 \\
\text{t Value} &= 16.1041 \\
\end{align*}
\]

Peer Influences on Aspiration: Haller & Butterworth (1960)
Loehlin (1987) analysis: Model 5

The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation

\[
\begin{align*}
f_{ramb} &= 0.1107 f_{famb} + 0.1762 f_{rpa} + 0.3235 f_{riq} \\
\text{Std Err} &= 0.0428 \beta 0.0350 \gamma1 0.0435 \gamma2 \\
\text{t Value} &= 2.5854 5.0308 7.4435 \\
+ 0.2233 f_{rses} + 0.0770 f_{fses} + 1.0000 d_{ramb} \\
& 6.3215 2.3870 \\
f_{famb} &= 0.1107 f_{ramb} + 0.0770 f_{rses} + 0.1762 f_{fpa} \\
\text{Std Err} &= 0.0428 \beta 0.0323 \gamma4 0.0350 \gamma1 \\
\text{t Value} &= 2.5854 2.3870 5.0308 \\
+ 0.3235 f_{fiq} + 0.2233 f_{fses} + 1.0000 d_{famb} \\
& 7.4435 6.3215 \\
\end{align*}
\]
Peer Influences on Aspiration: Haller & Butterworth (1960)
Loehlin (1987) analysis: Model 5

The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation

Variances of Exogenous Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>t Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>f_rpa</td>
<td></td>
<td>1.00000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>f_riq</td>
<td></td>
<td>1.00000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>f_rses</td>
<td></td>
<td>1.00000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>f_fpa</td>
<td></td>
<td>1.00000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>f_fiq</td>
<td></td>
<td>1.00000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>f_fses</td>
<td></td>
<td>1.00000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>e_rea</td>
<td>thetaea</td>
<td>0.30645</td>
<td>0.03721</td>
<td>8.24</td>
</tr>
<tr>
<td>e_roa</td>
<td>thetaoa</td>
<td>0.42304</td>
<td>0.03650</td>
<td>11.59</td>
</tr>
<tr>
<td>e_fea</td>
<td>thetaea</td>
<td>0.30645</td>
<td>0.03721</td>
<td>8.24</td>
</tr>
<tr>
<td>e_foa</td>
<td>thetaoa</td>
<td>0.42304</td>
<td>0.03650</td>
<td>11.59</td>
</tr>
<tr>
<td>e_rpa</td>
<td>errpa1</td>
<td>0.30781</td>
<td>0.07510</td>
<td>4.10</td>
</tr>
<tr>
<td>e_riq</td>
<td>errriq1</td>
<td>0.26748</td>
<td>0.07295</td>
<td>3.67</td>
</tr>
<tr>
<td>e_rses</td>
<td>errses1</td>
<td>0.11477</td>
<td>0.07265</td>
<td>1.58</td>
</tr>
<tr>
<td>e_fpa</td>
<td>errpa2</td>
<td>0.28837</td>
<td>0.07366</td>
<td>3.91</td>
</tr>
<tr>
<td>e_fiq</td>
<td>errriq2</td>
<td>0.15653</td>
<td>0.06614</td>
<td>2.37</td>
</tr>
<tr>
<td>e_fses</td>
<td>errses2</td>
<td>0.08832</td>
<td>0.07088</td>
<td>1.25</td>
</tr>
</tbody>
</table>

Covariances Among Exogenous Variables

<table>
<thead>
<tr>
<th>Var1</th>
<th>Var2</th>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>t Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>f_rpa</td>
<td>f_riq</td>
<td>cov1</td>
<td>0.26494</td>
<td>0.05436</td>
<td>4.87</td>
</tr>
<tr>
<td>f_rpa</td>
<td>f_rses</td>
<td>cov2</td>
<td>0.00185</td>
<td>0.04995</td>
<td>0.04</td>
</tr>
<tr>
<td>f_riq</td>
<td>f_rses</td>
<td>cov3</td>
<td>0.31164</td>
<td>0.05039</td>
<td>6.18</td>
</tr>
<tr>
<td>f_rpa</td>
<td>f_fpa</td>
<td>cov4</td>
<td>0.15828</td>
<td>0.07846</td>
<td>2.02</td>
</tr>
<tr>
<td>f_riq</td>
<td>f_fpa</td>
<td>cov5</td>
<td>0.11895</td>
<td>0.05383</td>
<td>2.21</td>
</tr>
<tr>
<td>f_rses</td>
<td>f_fpa</td>
<td>cov6</td>
<td>0.06924</td>
<td>0.04993</td>
<td>1.39</td>
</tr>
<tr>
<td>f_rpa</td>
<td>f_fiq</td>
<td>cov5</td>
<td>0.11895</td>
<td>0.05383</td>
<td>2.21</td>
</tr>
<tr>
<td>f_riq</td>
<td>f_fiq</td>
<td>cov7</td>
<td>0.43180</td>
<td>0.07084</td>
<td>6.10</td>
</tr>
<tr>
<td>f_rses</td>
<td>f_fiq</td>
<td>cov8</td>
<td>0.25004</td>
<td>0.05039</td>
<td>4.96</td>
</tr>
<tr>
<td>f_fpa</td>
<td>f_fiq</td>
<td>cov1</td>
<td>0.26494</td>
<td>0.05436</td>
<td>4.87</td>
</tr>
<tr>
<td>f_rpa</td>
<td>f_fses</td>
<td>cov6</td>
<td>0.06924</td>
<td>0.04993</td>
<td>1.39</td>
</tr>
<tr>
<td>f_riq</td>
<td>f_fses</td>
<td>cov8</td>
<td>0.25004</td>
<td>0.05039</td>
<td>4.96</td>
</tr>
<tr>
<td>f_rses</td>
<td>f_fses</td>
<td>cov9</td>
<td>0.30203</td>
<td>0.06360</td>
<td>4.75</td>
</tr>
<tr>
<td>f_fpa</td>
<td>f_fses</td>
<td>cov2</td>
<td>0.00185</td>
<td>0.04995</td>
<td>0.04</td>
</tr>
<tr>
<td>f_fiq</td>
<td>f_fses</td>
<td>cov3</td>
<td>0.31164</td>
<td>0.05039</td>
<td>6.18</td>
</tr>
<tr>
<td>e_rea</td>
<td>e_fea</td>
<td>covrea</td>
<td>0.02120</td>
<td>0.03094</td>
<td>0.69</td>
</tr>
<tr>
<td>e_roa</td>
<td>e_foa</td>
<td>covoa</td>
<td>0.11197</td>
<td>0.03254</td>
<td>3.44</td>
</tr>
<tr>
<td>d_ramb</td>
<td>d_famb</td>
<td>d_famb</td>
<td>0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The chi-square value for testing model 5 versus model 2 is 19.0745 - 19.0697 = 0.0048 with 1 degree of freedom. Omitting the covariance between the disturbance terms, therefore, causes hardly any deterioration in the fit of the model.

These data fail to provide evidence of direct reciprocal influence between the respondent’s and friend’s ambitions or of a covariance between the disturbance terms when these hypotheses are considered separately. Notice, however, that the covariance psi12 between the disturbance terms increases from -0.003344 for model 2 to
0.05479 for model 4. Before you conclude that all of these paths can be omitted from the model, it is important to test both hypotheses together by setting both $\beta$ and $\psi_{12}$ to zero as in Loehlin’s model 7:

```plaintext
title2 'Loehlin (1987) analysis: Model 7';
data ram7(type=ram);
   set ram2;
   if _name_='psi12'|_name_='beta' then
do;
   _name_=' ';
   _estim_=0;
end;
run;

proc calis data=aspire edf=328 inram=ram7;
run;
```

The relevant output is displayed in Figure 14.26.

![Figure 14.26. Career Aspiration Data: Loehlin Model 7](SAS OnlineDoc™: Version 8)
Peer Influences on Aspiration: Haller & Butterworth (1960)
Loehlin (1987) analysis: Model 7

The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation

\[
\begin{align*}
\text{riq} & = 0.8940 \ f_{\text{riq}} + 1.0000 \ e_{\text{riq}} \\
\text{rpa} & = 0.8370 \ f_{\text{rpa}} + 1.0000 \ e_{\text{rpa}} \\
\text{rses} & = 0.9490 \ f_{\text{rses}} + 1.0000 \ e_{\text{rses}} \\
\text{roa} & = 1.0000 \ f_{\text{roa}} + 1.0000 \ e_{\text{roa}} \\
\text{rea} & = 1.1037 \ f_{\text{rea}} + 1.0000 \ e_{\text{rea}} \\
\end{align*}
\]

Std Err 0.0678 lambda
t Value 16.2701

\[
\begin{align*}
\text{fiq} & = 0.8940 \ f_{\text{fiq}} + 1.0000 \ e_{\text{fiq}} \\
\text{fpa} & = 0.8370 \ f_{\text{fpa}} + 1.0000 \ e_{\text{fpa}} \\
\text{fses} & = 0.9490 \ f_{\text{fses}} + 1.0000 \ e_{\text{fses}} \\
\text{foa} & = 1.0000 \ f_{\text{foa}} + 1.0000 \ e_{\text{foa}} \\
\text{fea} & = 1.1037 \ f_{\text{rea}} + 1.0000 \ e_{\text{fea}} \\
\end{align*}
\]

Std Err 0.0678 lambda
t Value 16.2701

Peer Influences on Aspiration: Haller & Butterworth (1960)
Loehlin (1987) analysis: Model 7

The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation

\[
\begin{align*}
f_{\text{famb}} = & 0 \ f_{\text{famb}} + 0.1765 \ f_{\text{rpa}} + 0.3573 \ f_{\text{riq}} \\
& + 0.2419 \ f_{\text{rses}} + 0.1109 \ f_{\text{fses}} + 1.0000 \ d_{\text{famb}} \\
\text{Std Err} & = 0.0360 \ \text{gam1} \ 0.0461 \ \text{gam2} \\
\text{t Value} & = 4.8981 \ 7.7520 \\
\end{align*}
\]

\[
\begin{align*}
f_{\text{famb}} = & 0 \ f_{\text{famb}} + 0.1109 \ f_{\text{rses}} + 0.1765 \ f_{\text{fpa}} \\
& + 0.3573 \ f_{\text{fiq}} + 0.2419 \ f_{\text{fses}} + 1.0000 \ d_{\text{famb}} \\
\text{Std Err} & = 0.0461 \ \text{gam2} \ 0.0363 \ \text{gam3} \\
\text{t Value} & = 7.7520 \ 6.6671 \\
\end{align*}
\]
A Combined Measurement-Structural Model with Reciprocal Influence and Correlated Residuals


The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation

<table>
<thead>
<tr>
<th>Variable Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>t Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>f_rpa</td>
<td>1.00000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>f_riq</td>
<td>1.00000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>f_rses</td>
<td>1.00000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>f_fpa</td>
<td>1.00000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>f_fiq</td>
<td>1.00000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>f_fses</td>
<td>1.00000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>e_rea</td>
<td>0.31633</td>
<td>0.03648</td>
<td>8.67</td>
</tr>
<tr>
<td>e_roa</td>
<td>0.42656</td>
<td>0.03610</td>
<td>11.82</td>
</tr>
<tr>
<td>e_fea</td>
<td>0.31633</td>
<td>0.03648</td>
<td>8.67</td>
</tr>
<tr>
<td>e_foa</td>
<td>0.42656</td>
<td>0.03610</td>
<td>11.82</td>
</tr>
<tr>
<td>e_rpa</td>
<td>0.31329</td>
<td>0.07538</td>
<td>4.16</td>
</tr>
<tr>
<td>e_riq</td>
<td>0.30776</td>
<td>0.07307</td>
<td>4.21</td>
</tr>
<tr>
<td>e_rses</td>
<td>0.14303</td>
<td>0.07313</td>
<td>1.96</td>
</tr>
<tr>
<td>e_fpa</td>
<td>0.29286</td>
<td>0.07389</td>
<td>3.96</td>
</tr>
<tr>
<td>e_fiq</td>
<td>0.19193</td>
<td>0.06613</td>
<td>2.90</td>
</tr>
<tr>
<td>e_fses</td>
<td>0.11804</td>
<td>0.07147</td>
<td>1.65</td>
</tr>
<tr>
<td>d_ramb</td>
<td>0.21011</td>
<td>0.02940</td>
<td>7.15</td>
</tr>
<tr>
<td>d_famb</td>
<td>0.21011</td>
<td>0.02940</td>
<td>7.15</td>
</tr>
</tbody>
</table>

Covariances Among Exogenous Variables

<table>
<thead>
<tr>
<th>Var1</th>
<th>Var2</th>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>t Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>f_rpa</td>
<td>f_riq</td>
<td>cov1</td>
<td>0.27533</td>
<td>0.05552</td>
<td>4.96</td>
</tr>
<tr>
<td>f_rpa</td>
<td>f_rses</td>
<td>cov2</td>
<td>0.00611</td>
<td>0.05085</td>
<td>0.12</td>
</tr>
<tr>
<td>f_riq</td>
<td>f_rses</td>
<td>cov3</td>
<td>0.33510</td>
<td>0.05150</td>
<td>6.51</td>
</tr>
<tr>
<td>f_rpa</td>
<td>f_fpa</td>
<td>cov4</td>
<td>0.17099</td>
<td>0.07872</td>
<td>2.17</td>
</tr>
<tr>
<td>f_riq</td>
<td>f_fpa</td>
<td>cov5</td>
<td>0.13859</td>
<td>0.05431</td>
<td>2.55</td>
</tr>
<tr>
<td>f_rses</td>
<td>f_fpa</td>
<td>cov6</td>
<td>0.07563</td>
<td>0.05077</td>
<td>1.49</td>
</tr>
<tr>
<td>f_rpa</td>
<td>f_fiq</td>
<td>cov7</td>
<td>0.13859</td>
<td>0.05431</td>
<td>2.55</td>
</tr>
<tr>
<td>f_riq</td>
<td>f_fiq</td>
<td>cov8</td>
<td>0.48105</td>
<td>0.06993</td>
<td>6.88</td>
</tr>
<tr>
<td>f_rses</td>
<td>f_fiq</td>
<td>cov9</td>
<td>0.27235</td>
<td>0.05157</td>
<td>5.28</td>
</tr>
<tr>
<td>f_fpa</td>
<td>f_fiq</td>
<td>cov10</td>
<td>0.27533</td>
<td>0.05552</td>
<td>4.96</td>
</tr>
<tr>
<td>f_rpa</td>
<td>f_fses</td>
<td>cov11</td>
<td>0.07563</td>
<td>0.05077</td>
<td>1.49</td>
</tr>
<tr>
<td>f_riq</td>
<td>f_fses</td>
<td>cov12</td>
<td>0.27235</td>
<td>0.05157</td>
<td>5.28</td>
</tr>
<tr>
<td>f_rses</td>
<td>f_fses</td>
<td>cov13</td>
<td>0.32046</td>
<td>0.06517</td>
<td>4.92</td>
</tr>
<tr>
<td>f_fpa</td>
<td>f_fses</td>
<td>cov14</td>
<td>0.00611</td>
<td>0.05085</td>
<td>0.12</td>
</tr>
<tr>
<td>f_fiq</td>
<td>f_fses</td>
<td>cov15</td>
<td>0.33510</td>
<td>0.05150</td>
<td>6.51</td>
</tr>
<tr>
<td>e_rea</td>
<td>e_fea</td>
<td>cov16</td>
<td>0.04535</td>
<td>0.02918</td>
<td>1.55</td>
</tr>
<tr>
<td>e_roa</td>
<td>e_foa</td>
<td>cov17</td>
<td>0.12085</td>
<td>0.03214</td>
<td>3.76</td>
</tr>
<tr>
<td>d_ramb</td>
<td>d_famb</td>
<td>cov18</td>
<td>0.0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

When model 7 is tested against models 2, 4, and 5, the $p$-values are respectively 0.0433, 0.0370, and 0.0123, indicating that the combined effect of the reciprocal influence and the covariance of the disturbance terms is statistically significant. Thus, the hypothesis tests indicate that it is acceptable to omit either the reciprocal influences or the covariance of the disturbances but not both.

It is also of interest to test the covariances between the error terms for educational (COVEA) and occupational aspiration (COVOA), since these terms are omitted from
Jöreskog and Sörbom’s models. Constraining COVEA and COVOA to zero produces Loehlin’s model 6:

```
title2 'Loehlin (1987) analysis: Model 6';
data ram6(type=ram);
  set ram2;
  if _name_='covea'|_name_='covoa' then
    do;
      _name_=' ';
      _estim_=0;
    end;
run;
proc calis data=aspire edf=328 inram=ram6;
run;
```

The relevant output is displayed in Figure 14.27.

```
Peer Influences on Aspiration: Haller & Butterworth (1960)
Loehlin (1987) analysis: Model 6

The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation

Fit Function 0.1020
Goodness of Fit Index (GFI) 0.9802
GFI Adjusted for Degrees of Freedom (AGFI) 0.9638
Root Mean Square Residual (RMR) 0.0306
 Parsimonious GFI (Mulaik, 1989) 0.6535
Chi-Square 33.4475
Chi-Square DF 30
Pr > Chi-Square 0.3035
 Independence Model Chi-Square 872.00
 Independence Model Chi-Square DF 45
RMSEA Estimate 0.0187
RMSEA 90% Lower Confidence Limit .
RMSEA 90% Upper Confidence Limit 0.0471
ECVI Estimate 0.2597
ECVI 90% Lower Confidence Limit .
ECVI 90% Upper Confidence Limit 0.3164
Probability of Close Fit 0.9686
Bentler’s Comparative Fit Index 0.9958
Normal Theory Reweighted LS Chi-Square 32.9974
Akaike’s Information Criterion -26.5525
Bozdogan’s (1987) CAIC -170.4342
Schwarz’s Bayesian Criterion -140.4342
Z-Test of Wilson & Hilferty (1931) 0.5151
Bollen (1986) Normed Index Rho1 0.9425
Bollen (1988) Non-normed Index Delta2 0.9959
Hoelter’s (1983) Critical N 431
```

Figure 14.27. Career Aspiration Data: Loehlin Model 6

The chi-square value for testing model 6 versus model 2 is 33.4476 - 19.0697 = 14.3779 with 2 degrees of freedom and a p-value of 0.0008, indicating that there is considerable evidence of correlation between the error terms.
The following table summarizes the results from Loehlin’s seven models.

<table>
<thead>
<tr>
<th>Model</th>
<th>$\chi^2$</th>
<th>df</th>
<th>p-value</th>
<th>SBC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Full model</td>
<td>12.0132</td>
<td>13</td>
<td>0.5266</td>
<td>-63.3356</td>
</tr>
<tr>
<td>2. Equality constraints</td>
<td>19.0697</td>
<td>28</td>
<td>0.8960</td>
<td>-143.2200</td>
</tr>
<tr>
<td>3. No SES path</td>
<td>23.0365</td>
<td>29</td>
<td>0.7749</td>
<td>-145.0492</td>
</tr>
<tr>
<td>4. No reciprocal influence</td>
<td>20.9981</td>
<td>29</td>
<td>0.8592</td>
<td>-147.0876</td>
</tr>
<tr>
<td>5. No disturbance correlation</td>
<td>19.0745</td>
<td>29</td>
<td>0.9194</td>
<td>-149.0111</td>
</tr>
<tr>
<td>6. No error correlation</td>
<td>33.4475</td>
<td>30</td>
<td>0.3035</td>
<td>-140.4342</td>
</tr>
<tr>
<td>7. Constraints from both 4 &amp; 5</td>
<td>25.3466</td>
<td>30</td>
<td>0.7080</td>
<td>-148.5351</td>
</tr>
</tbody>
</table>

For comparing models, you can use a DATA step to compute the differences of the chi-square statistics and p-values.

```
title 'Comparisons among Loehlin’s models';
data _null_
  array adf[7] _temporary_ (13 28 29 29 29 30 30);
  retain indent 16;
  file print;
  input ho ha @@;
  chisq = achiq[ho] - achiq[ha];
  df = adf[ho] - adf[ha];
  p = 1 - probchi( chisq, df);
  if _n_ = 1 then put
      / +indent 'model comparison  chi**2  df  p-value'
      / +indent '---------------------------------------';
  put +indent +3 ho ' versus ' ha @18 +indent chisq 8.4 df 5. p 9.4;
datalines;
2 1 3 2 4 2 5 2 7 2 7 4 7 5 6 2 ;
```

The DATA step displays the following table.

```
Comparisons among Loehlin’s models

<table>
<thead>
<tr>
<th>model comparison</th>
<th>chi**2</th>
<th>df</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 versus 1</td>
<td>7.0565</td>
<td>15</td>
<td>0.9561</td>
</tr>
<tr>
<td>3 versus 2</td>
<td>3.9668</td>
<td>1</td>
<td>0.0464</td>
</tr>
<tr>
<td>4 versus 2</td>
<td>1.9284</td>
<td>1</td>
<td>0.1649</td>
</tr>
<tr>
<td>5 versus 2</td>
<td>0.0048</td>
<td>1</td>
<td>0.9448</td>
</tr>
<tr>
<td>7 versus 2</td>
<td>6.2769</td>
<td>2</td>
<td>0.0433</td>
</tr>
<tr>
<td>7 versus 4</td>
<td>4.3485</td>
<td>1</td>
<td>0.0370</td>
</tr>
<tr>
<td>7 versus 5</td>
<td>6.2721</td>
<td>1</td>
<td>0.0123</td>
</tr>
<tr>
<td>6 versus 2</td>
<td>14.3778</td>
<td>2</td>
<td>0.0008</td>
</tr>
</tbody>
</table>
```

Although none of the seven models can be rejected when tested against the alternative of an unrestricted covariance matrix, the model comparisons make it clear that...
there are important differences among the models. Schwarz’s Bayesian Criterion indicates model 5 as the model of choice. The constraints added to model 5 in model 7 can be rejected ($p=0.0123$), while model 5 cannot be rejected when tested against the less-constrained model 2 ($p=0.9448$). Hence, among the small number of models considered, model 5 has strong statistical support. However, as Loehlin (1987, p. 106) points out, many other models for these data could be constructed. Further analysis should consider, in addition to simple modifications of the models, the possibility that more than one friend could influence a boy’s aspirations, and that a boy’s ambition might have some effect on his choice of friends. Pursuing such theories would be statistically challenging.

References


