SAS/STAT® 9.2 User's Guide

Introduction to Structural Equation Modeling with Latent Variables

(Book Excerpt)

SAS® Documentation

This document is an individual chapter from *SAS/STAT® 9.2 User's Guide*.

The correct bibliographic citation for the complete manual is as follows: SAS Institute Inc. 2008. *SAS/STAT® 9.2 User's Guide*. Cary, NC: SAS Institute Inc.

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513.

1st electronic book, March 2008
2nd electronic book, February 2009

Chapter 17

Introduction to Structural Equation Modeling with Latent Variables

Contents

Overview: Structural Equation Modeling with Latent Variables

You can use the CALIS or TCALIS 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 variances, covariances, and regression (or path) 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 support the use of hypothetical latent variables and measurement errors in the models.

Loehlin (1987) provides an excellent introduction to latent variable models by using path diagrams and structural equations. A more advanced treatment of structural equation models with latent variables is given by Bollen (1989). Fuller (1987) provides a highly technical statistical treatment of measurement-error models.

TCALIS and CALIS Procedures

In SAS 9.2, the newly developed TCALIS procedure serves as an enhanced version of the older CALIS procedure. The TCALIS procedure has most of the capabilities of the CALIS procedure. In addition, it has some new capabilities that are not available in PROC CALIS. These new capabilities include multiple-group analysis, mean structures analysis, and a whole spectrum of modeling languages, among others. See Chapter 88, "The TCALIS Procedure (Experimental)," for more details about the new features available in PROC TCALIS. In this chapter, statistical analyses in the examples are done using PROC TCALIS.

Comparison of the TCALIS and SYSLIN Procedures

The SYSLIN procedure in SAS/ETS software can also fit certain kinds of path models and linear structural equation models. PROC TCALIS differs from PROC SYSLIN in that PROC TCALIS is more general 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 TCALIS 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 TCALIS and SYSLIN procedures enable you to specify a model as a system of linear equations. When there are several equations, a given variable might 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 variable 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 statement. In PROC TCALIS, endogenous variables are identified by the procedure automatically after you specify the model. With different modeling languages, the identification of endogenous variables by PROC TCALIS is done by different sets of rules. For example, when you specify structural equations by using the LINEQS modeling language in PROC TCALIS, endogenous variables are assumed to be those that appear on the left-hand sides of the equations; a given variable can appear on the left-hand side of at most one equation. When you specify your model by using the PATH modeling language in PROC TCALIS, endogenous variables are those variables pointed to by arrows at least once in the path specifications.

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

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

Model Specification

PROC TCALIS provides several modeling languages to specify a model. Different modeling languages in PROC TCALIS are signified by the main model specification statement used. In the TCALIS procedure, FACTOR, LINEQS, LISMOD, MSTRUCT, PATH, and RAM are the main modeling specification statements. Each of these statements invokes a specific modeling language. Depending on your modeling philosophy and the type of the model, you can choose a modeling language that is most suitable for your application. Models specified using structural equations can be transcribed directly into the LINEQS statement. Models that are hypothesized using path diagrams can be described easily in the PATH or RAM statement. First-order confirmatory or exploratory factor models are most conveniently specified using the FACTOR and MATRIX statements. Traditional LISREL models are supported through the LISMOD and MATRIX statements. Finally, patterned covariance and mean models can be specified directly by the MSTRUCT and MATRIX statements.

For most applications, the PATH and LINEQS statements are the easiest to use. In other cases, the FACTOR, LISMOD, MSTRUCT, or RAM statement might be more suitable. See the section “Which Modeling Language?” on page 6706 in Chapter 88, “The TCALIS Procedure (Experimental),” for a more detailed discussion.

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

Estimation Methods

The TCALIS procedure provides five methods of estimation specified by the METHOD= option:

DWLS	diagonally weighted least squares
ULS	unweighted least squares
GLS	normal theory generalized least squares
ML	maximum likelihood for multivariate normal distributions
WLS	weighted least squares for arbitrary distributions

Each estimation method is based on finding parameter estimates that minimize a badness-of-fit or discrepancy function, which measures the difference between the observed sample covariance matrix and the fitted (or predicted) covariance matrix, given the model and the parameter estimates. See the section "Estimation Criteria" on page 6880 in Chapter 88, "The TCALIS Procedure (Experimental)," 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 most popular method for applications. The option METHOD=GLS usually produces very similar results to those produced by METHOD=ML. Asymptotically, ML and GLS are the same. Both methods assume a multivariate normal distribution in the population. The WLS method with default weight matrix, which is equivalent to the asymptotically distribution free (ADF) method, yields asymptotically normal estimates regardless of the distribution in the population. When the multivariate normal assumption is in doubt, especially if the variables have high kurtosis, you should seriously consider the WLS method. When a correlation matrix is analyzed, only the WLS can produce correct standard error estimates. However, in order to use the WLS method with the expected statistical properties, sample size must be large. Several thousands might be a minimum requirement.

The ULS and DWLS methods yield reasonable estimates under less restrictive assumptions. You can apply these methods to normal or nonnormal situations, or to covariance or correlation matrices. The drawback is that the statistical qualities of the estimates seem to be unknown. For this reason, PROC TCALIS does not provide standard errors or test statistics with these two methods.

You cannot use METHOD=ML or METHOD=GLS if the observed covariance matrix is singular. You could either remove variables involved in the linear dependencies or use less restrictive estimation methods like ULS. Specifying METHOD=ML assumes that the predicted covariance matrix is nonsingular. If ML fails because of a singular predicted covariance matrix, you need to examine whether the model specification leads to the singularity. If so, modify the model specification to eliminate the problem. If not, you probably need to use other estimation methods.

You should remove outliers and try to transform variables that are skewed or heavy-tailed. This applies to all 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 TCALIS displays estimates of univariate and multivariate kurtosis (Bollen 1989, pp. 418–425) if you specify the KURTOSIS option in the PROC TCALIS statement.

Statistical Inference

When you specify the ML, GLS, or WLS estimation with appropriate models, PROC TCALIS can compute the following:

- a chi-square goodness-of-fit test of the specified model versus the alternative that the data are from a population 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 TCALIS. 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 under ML and GLS 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).

All test statistics and standard errors computed by PROC TCALIS are 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 5 to 20 times the number of parameters to be estimated in order to get reliable and interpretable results. The WLS method might even require that the sample size be over several thousand.

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 might not provide good approximations to the actual sampling distributions.

For modeling correlation structures, the only theoretically correct method is the WLS method with the default ASYCOV=CORR option. For other methods, standard error estimates for modeling correlation structures might be inaccurate even for sample sizes as large as 400. The chi-square statistic is generally the same regardless of which matrix is analyzed, provided that the model involves no scale-dependent constraints. However, if the purpose is to obtain reasonable parameter estimates for the correlation structures only, then you might find other estimation methods useful as well.

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 TCALIS attempts to determine which diagonal elements of the predicted correlation matrix are constrained to a constant, but it might 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.

The root mean squared error approximation (RMSEA) proposed by Steiger and Lind (1980) does not assume a true model being fitted to the data. It measures the discrepancy between the fitted model and the covariance matrix in the population. For samples, RMSEA and confidence intervals can be estimated. Statistical tests for determining whether the population RMSEA's fall below certain specified values are available (Browne and Cudeck 1993). In the same vein, Browne and Cudeck (1993) propose the expected cross validation index (ECVI) that measures how good a model is for predicting future sample covariances. Point estimate and confidence intervals for ECVI are also developed.

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 TCALIS uses a variety of nonlinear optimization algorithms for computing parameter estimates. These algorithms are very complicated and do not always work. PROC TCALIS will generally inform you when the computations fail, usually by displaying an error message about the iteration limit being exceeded. When this happens, you might 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 6915 in Chapter 88, "The TCALIS Procedure (Experimental)," and refer to Bollen (1989, pp. 254–256).

PROC TCALIS might 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 TCALIS statement is useful for generating a variety of initial estimates.

Structural Equation Models and the LINEQS Modeling Language

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.

Model Form A

$$Y = \alpha + \beta X + E_Y$$

with the following assumption:

$$\text{Cov}(X, E_Y) = 0$$

where α and β 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 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 Var(X) and 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 4, "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, as shown in the following statements:

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

You can also fit this model with PROC TCALIS, but you must explicitly specify the error terms and the parameter name for the regression coefficient (except for the intercept, which is assumed to be present in each equation). The following specification in PROC TCALIS is equivalent to the preceding regression model:

```
proc tcalis;
   lineqs
      y = beta x + ey;
run;
```

where beta is the parameter name for the regression coefficient and ey is the error term of the equation. You do not need to type an "*" between beta and x to indicate the multiplication of the variable by the coefficient.

You might use other names for the parameters and the error terms, but there are rules to follow in the LINEQS model specification. The LINEQS statement uses the convention that the names of error terms begin with the letter E or e, disturbances (errors terms for latent variables) in equations begin with D or d, and other latent variables begin with F or f for "factor." Names of variables in the input SAS data set can, of course, begin with any letter.

Optionally, you can specify the variance parameters of exogenous variables explicitly by using the STD statement as follows:

```
proc tcalis;
   lineqs
      y  = beta x + ey;
   std
      x  = vx,
      ey = vey;
run;
```

where vx and vey represent the variance of x and ey, respectively. Explicitly giving these variance parameters names would be useful when you need to set constraints on these parameters by referring to their names. In this simple case without constraints, however, naming these variance parameters is not necessary.

Although naming variance parameters for exogenous variables is optional, naming the regression coefficients when they should be free parameters is not. Consider the following statements:

```
proc tcalis;
   lineqs
      y = x + ey;
run;
```

In this specification, you leave out the regression coefficient beta from the preceding LINEQS model. Instead of being a parameter to estimate, the regression coefficient in this specification is a fixed constant 1.

In certain special situations where you need to specify a regression equation without an error term, you can explicitly set the variance of the error term to zero. For example, the following statements, in effect, will fit an equation without an error term:

```
proc tcalis;
   lineqs
      y = beta x + ey;
   std
      ey = 0;
run;
```

In this specification, the mean of the error term ey is presumably zero, as all error terms in the PROC TCALIS are set to have fixed zero means. Together with the zero variance specification for ey in the STD statement, ey is essentially a zero constant in the equation.

By default, PROC TCALIS analyzes the covariance matrix. This is exactly the opposite of PROC CALIS or PROC FACTOR, which analyzes the correlation matrix by default. Most applications that use PROC TCALIS should employ the default covariance structure analysis.

Since the analysis of covariance structures is based on modeling the covariance matrix and the covariance matrix contains no information about means, PROC TCALIS neglects the intercept parameter by default. To estimate the intercept, you can add the intercept term explicitly into the LINEQS statement. For example, the following statements fit a regression model with the estimation of the intercept alpha:

```
proc tcalis;
   lineqs
      y = alpha intercept + beta x + ey;
run;
```

In the LINEQS statement, intercept represents a "variable" with a constant value of 1; hence, the coefficient alpha is the intercept parameter. Notice that with the simultaneous analysis of mean and covariance structures, you have to provide either the raw data or the means and covariances in your input data set.

Other commonly used options in the PROC TCALIS statement include the following:

- MODIFICATION to display model modification indices
- NOBS to specify the number of observations
- NOSE to suppress the display of approximate standard errors
- RESIDUAL to display residual correlations or covariances
- TOTEFF to display total and indirect effects

For ordinary unconstrained regression models, there is no reason to use PROC TCALIS instead of PROC REG. But suppose that the observed variables Y and X are contaminated by errors (especially measurement errors), 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

$$
\begin{aligned}
Y &= \alpha + \beta F_X + E_Y \\
X &= F_X + E_X
\end{aligned}
$$

with the following assumption:

$$\mathrm{Cov}(F_X, E_X) = \mathrm{Cov}(F_X, E_Y) = \mathrm{Cov}(E_X, E_Y) = 0$$

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 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 might be severely biased and inconsistent.

You can express model form B in the LINEQS statement as follows:

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

In this specification, you specify a variance for each of the latent variables in this model by 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 this model, vfx, vey, and vex are variance parameters to estimate.

The variances of endogenous variables are predicted from the model and hence are not parameters. Covariances involving latent exogenous variables are assumed to be zero 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 $\mathrm{Var}(E_X)$ of 57. You can fit model form B with this constraint to the data by using the following 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 tcalis data=corn;
   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 TCALIS produces the estimates shown in Figure 17.1.

Figure 17.1 Measurement Error Model for Corn Data

```
                          Linear Equations

                 y        =     0.4232*fx    +  1.0000 ey
                 Std Err        0.1658 beta
                 t Value        2.5520
                 x        =     1.0000 fx    +  1.0000 ex


            Estimates for Variances of Exogenous Variables

 Variable                                                 Standard
 Type           Variable    Parameter      Estimate          Error         t Value

 Error          ex                         57.00000
 Latent         fx          vfx           247.85450      136.33508         1.81798
 Error          ey          vey            43.29105       23.92488         1.80946
```

PROC TCALIS 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 reproduced in the examples in this chapter.

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

Model Form C

$$
\begin{aligned}
Y &= F_Y + E_Y \\
X &= F_X + E_X \\
F_Y &= \alpha + \beta F_X
\end{aligned}
$$

with the following assumption:

$$\mathrm{Cov}(F_X, E_X) = \mathrm{Cov}(F_X, E_X) = \mathrm{Cov}(E_X, E_Y) = 0$$

The first two equations express the observed variables in terms of a true score plus error; these two 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 shown in the following:

```
proc tcalis;
   lineqs
      y   = fy + ey,
      x   = fx + ex,
      fy  = beta fx + dfy;
   std
      fx  = vfx,
      ey  = vey,
      ex  = vex,
      dfy = 0;
run;
```

As a syntactic requirement, each equation in the LINEQS statement should have an error term. As discussed before, because dfy has a fixed variance 0 in the STD statement, in effect, there is no error term in the structural equation with the outcome variable fy.

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, Var(F_Y)=β^2 Var(F_X).

Correlations or covariances 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 β 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 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:

```
proc tcalis;
   lineqs
      y   = fy + ey,
      x   = fx + ex,
      fy  = beta fx + dfy;
   std
      fx  = vfx,
      ey  = vey,
      ex  = vex,
      dfy = 0;
   cov
      ey ex = ceyex;
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 as free parameters 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 might 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 might 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, GLS, or WLS estimation, PROC TCALIS can perform a statistical test of the goodness of fit of the model under the certain statistical assumptions.

If the model is just identified or overidentified, it is said to be identified. If you use ML, GLS, or WLS estimation for an identified model, PROC TCALIS can compute approximate standard errors for the parameter estimates. For underidentified models, PROC TCALIS 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 TCALIS 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 TCALIS displays equations showing the linear relationships among the estimates. Another 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 TCALIS to fit a model of form C to the square roots of the counts without constraints on the parameters, as displayed in the following statements:

```
data spleen;
   input rosette nucleate;
   sqrtrose=sqrt(rosette);
   sqrtnucl=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 tcalis data=spleen;
   lineqs sqrtrose = factrose + err_rose,
          sqrtnucl = factnucl + err_nucl,
          factrose = beta factnucl + disturb;
   std err_rose = v_rose,
       err_nucl = v_nucl,
       factnucl = v_factnu,
       disturb  = 0;
run;
```

This model is underidentified. PROC TCALIS displays the following warning:

```
WARNING: Estimation problem not identified: More parameters to
         estimate ( 4 ) than the total number of mean and
         covariance elements ( 3 ).
```

Then it diagnoses the indeterminacy as follows:

```
NOTE: Covariance matrix for the estimates is not full rank.
NOTE: The variance of some parameter estimates is zero or
      some parameter estimates are linearly related to other
      parameter estimates as shown in the following equations:

v_rose     =       0.207718    +       0.108978   *   beta
                               +       0.916873   *   v_nucl
                               -       0.916873   *   v_factnu
```

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

```
proc tcalis data=spleen;
   lineqs sqrtrose = factrose + err_rose,
          sqrtnucl = factnucl + err_nucl,
          factrose = beta factnucl + disturb;
   std err_rose = .25,
       err_nucl = .25,
       factnucl = v_factnu,
       disturb  = 0;
run;
```

This model is overidentified and the chi-square goodness-of-fit test yields a *p*-value of 0.0219, as displayed in Figure 17.2.

Figure 17.2 Spleen Data: Fit Statistics for Overidentified Model

Fit Summary

Modeling Info	N Observations	12
	N Variables	2
	N Moments	3
	N Parameters	2
	N Active Constraints	0
	Independence Model Chi-Square	13.2732
	Independence Model Chi-Square DF	1
Absolute Index	Fit Function	0.4775
	Chi-Square	5.2522
	Chi-Square DF	1
	Pr > Chi-Square	0.0219
	Z-Test of Wilson & Hilferty	2.0375
	Hoelter Critical N	10
	Root Mean Square Residual (RMSR)	0.1785
	Standardized RMSR (SRMSR)	0.0745
	Goodness of Fit Index (GFI)	0.7274
Parsimony Index	Adjusted GFI (AGFI)	0.1821
	Parsimonious GFI	0.7274
	RMSEA Estimate	0.6217
	RMSEA Lower 90% Confidence Limit	0.1899
	RMSEA Upper 90% Confidence Limit	1.1869
	Probability of Close Fit	0.0237
	ECVI Estimate	0.9775
	ECVI Lower 90% Confidence Limit	.
	ECVI Upper 90% Confidence Limit	2.2444
	Akaike Information Criterion	3.2522
	Bozdogan CAIC	1.7673
	Schwarz Bayesian Criterion	2.7673
	McDonald Centrality	0.8376
Incremental Index	Bentler Comparative Fit Index	0.6535
	Bentler-Bonett NFI	0.6043
	Bentler-Bonett Non-normed Index	0.6535
	Bollen Normed Index Rho1	0.6043
	Bollen Non-normed Index Delta2	0.6535
	James et al. Parsimonious NFI	0.6043

PROC TCALIS arranges the fit statistics according to their types: absolute, parsimony, and incremental. After displaying some important modeling information in the fit summary table, the absolute fit indices are printed. Absolute indices are those model fit statistics that assess the model fit without comparing to a "null" model. The most typical fit index of the this type is the chi-square fit statistic. Nonsignificance of the chi-square indicates good model fit. Other popular absolute fit indices include the root mean square residual (RMSR), the standardized RMSR (SRMSR), and the goodness-of-fit index (GFI). By convention, a good model should have an SRMSR smaller than 0.05 and a GFI larger than 0.90.

Parsimony indices are those fit indices that assess the model fit without comparing with a null model, but with the number of parameters in the model taking into account. These fit indices favor precise models. If two models have the same chi-square value for the same data set but have different number of parameters in the models, the model with fewer parameters will have a better parsimony fit statistic. The most popular parsimony index displayed in the table is perhaps the root mean squared error of approximation, or RMSEA (Steiger and Lind 1980). An RMSEA below 0.05 is recommended for a good model fit (Browne and Cudeck 1993). Another popular index in this category is the adjusted GFI (AGFI). By convention, an AGFI above 0.90 is required for a good model fit.

Finally, the incremental fit indices are those indices that measure model fit by comparing with a null model. A null model is usually the independence model that assumes the measured variables are all uncorrelated. The most popular incremental index is Bentler's CFI. By convention, a CFI above 0.90 is required for a good model fit.

After the model fit summary, the parameter estimates are displayed in Figure 17.3.

Figure 17.3 Spleen Data: Parameter Estimates for Overidentified Model

```
                              Linear Equations

               sqrtrose =    1.0000 factrose +  1.0000 err_rose
               sqrtnucl =    1.0000 factnucl +  1.0000 err_nucl
               factrose =    0.4034*factnucl +  1.0000 disturb
               Std Err       0.0508 beta
               t Value       7.9439

                Estimates for Variances of Exogenous Variables

Variable                                                 Standard
Type           Variable    Parameter      Estimate          Error        t Value

Error          err_rose                    0.25000
               err_nucl                    0.25000
Latent         factnucl    v_factnu       10.45846        4.56608        2.29047
Disturbance    disturb                           0
```

Overall, the model does not provide a good fit. The sample size is so small that the *p*-value of the chi-square test 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 nonzero disturbance in the linear relation between factrose and factnucl.
- The relation between factrose and 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 make the "disturbance term" disturb a real random variable with nonzero variance in the structural model. This can be done by giving the variance of disturb a parameter name in the STD statement, as shown in the following statements:

```
proc tcalis data=spleen;
   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;
```

In the STD statement, v_dist is now specified as a free variance parameter to be estimated. The parameter estimates are produced in Figure 17.4.

Figure 17.4 Spleen Data: Parameter Estimated for Just Identified Model

```
                              Linear Equations

              sqrtrose =    1.0000 factrose +  1.0000 err_rose
              sqrtnucl =    1.0000 factnucl +  1.0000 err_nucl
              factrose =    0.3907*factnucl +  1.0000 disturb
              Std Err       0.0771 beta
              t Value       5.0692

               Estimates for Variances of Exogenous Variables
```

Variable Type	Variable	Parameter	Estimate	Standard Error	t Value
Error	err_rose		0.25000		
	err_nucl		0.25000		
Latent	factnucl	v_factnu	10.50458	4.58577	2.29069
Disturbance	disturb	v_dist	0.38153	0.28556	1.33607

As shown in Figure 17.4, the variance of disturb is estimated at 0.382. Due to the inclusion of this new parameter, estimates for beta and v_factnu also shift a little bit from the previous analysis.

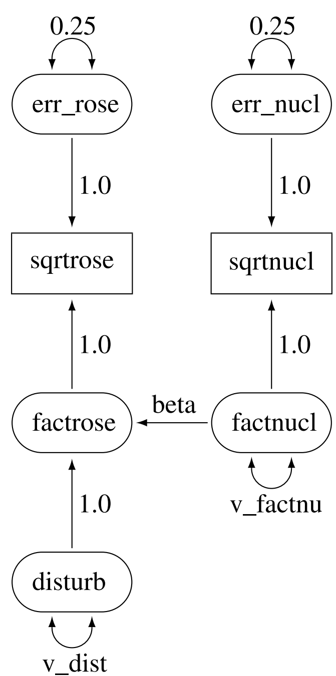
Because this model is just identified or saturated, there are no degrees of freedom for the chi-square goodness-of-fit test.

Path Diagrams and the PATH Modeling Language

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. Another advantage is that the path diagram can be translated easily into the PATH modeling language supported by PROC TCALIS.

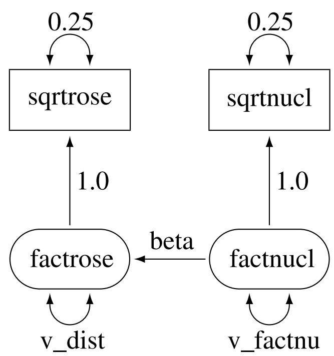
It is customary to write the names of manifest variables in rectangles and the names of latent variables in ovals. The coefficients in each equation are indicated by drawing arrows from the independent variables to the dependent variable. Covariances between exogenous variables are drawn as two-headed 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. Figure 17.5 displays a path diagram for the spleen data, explicitly showing all latent variables (including error terms) and variances of exogenous variables.

Figure 17.5 Path Diagram: Spleen



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 enable the previous path diagram to be simplified, as shown in Figure 17.6.

Figure 17.6 Simplified Path Diagram: Spleen



The PATH modeling language in PROC TCALIS provides a simple way to transcribe a path diagram based on the reticular action model. In the PATH modeling languages, there are three statements to capture the specifications in path diagrams:

- The PATH statement enables you to specify each of the one-headed arrows (paths). The parameters specified in the PATH statement are the path (regression) coefficients.
- The PVAR statement enables you to specify each of the double-headed arrows with both heads pointing to the ***same*** variable. In general, you specify *partial* (or total) variance parameters in the PVAR statement. If the variable being pointed at is *exogenous*, a (total) variance parameter is specified. If the variable being pointed at is *endogenous*, a partial or an error variance parameter is specified.
- The PCOV statement enables you to specify each of the double-headed arrows with its heads pointing to ***different*** variables. In general, you specify (*partial*) covariance parameters in the PCOV statement. The two most common cases are as follows: (1) If the heads of a double-headed arrow are connecting two *exogenous* variables, a covariance parameter between the two variables is specified; and (2) If the heads of a double-headed arrow are connecting two *endogenous* variables, an error covariance parameter for the two variables is specified. This error covariance is also a partial covariance between the endogenous variables.

For example, the path diagram for the spleen data in Figure 17.6 can be specified with the PATH modeling language as follows:

```
proc tcalis data=spleen outmodel=splmod1;
   path
      sqrtrose <- factrose   1.0,
      sqrtnucl <- factnucl   1.0,
      factrose <- factnucl   beta;
   pvar
      sqrtrose = 0.25,      /* error variance for sqrtrose */
      sqrtnucl = 0.25,      /* error variance for sqrtnucl */
      factrose = v_dist,    /* disturbance variance for factrose */
      factnucl = v_factnu;  /* variance of factnucl */
run;
```

One notable item in the specification is that each of the single-headed or double-headed arrows in the path diagram is transcribed into an entry in either the PATH or PVAR statement:

- **PATH** statement:
  The paths "sqrtrose <– factrose" and "sqrtnucl <– factnucl" in the PATH statement are followed by the constant 1, indicating fixed path coefficients. The path "factrose <– factnucl" is followed by a parameter named beta, indicating a free path coefficient to estimate in the model.

- **PVAR** statement:
  A fixed value 0.25 is specified after the equal signs of sqrtrose and sqrtnucl in the PVAR statement. Because sqrtrose and sqrtnucl are endogenous in the model, you are fixing the error variances of sqrtrose and sqrtnucl to 0.25 in the specification.
  In the last two entries of the PVAR statement, you are putting parameter names after the equal signs. Because factrose and factnucl are exogenous in the model, v_dist and v_factnu are variance parameters of factrose and factnucl, respectively.

Because there are no double-headed arrows each pointing to different variables in the path diagram, the PCOV statement is not needed in the model specification. The resulting output of the PATH model is displayed in Figure 17.7.

Figure 17.7 Spleen Data: RAM Model

PATH List

----------Path----------			Parameter	Estimate	Standard Error	t Value
sqrtrose	<-	factrose		1.00000		
sqrtnucl	<-	factnucl		1.00000		
factrose	<-	factnucl	beta	0.39074	0.07708	5.06920

Variance Parameters

Variance Type	Variable	Parameter	Estimate	Standard Error	t Value
Error	sqrtrose		0.25000		
	sqrtnucl		0.25000		
	factrose	v_dist	0.38153	0.28556	1.33607
Exogenous	factnucl	v_factnu	10.50458	4.58577	2.29069

In the PROC TCALIS statement, the OUTMODEL=SPLMOD1 option is used. This will save the model specification, together with final estimates in a SAS data set called SPLMOD1. This special type of SAS data set is called “CALISMDL.” The following statements are used to display the contents of this OUTMODEL= data set:

```
proc print data=splmod1;
run;
```

As displayed in Figure 17.8, the first record saves the model type, which is the PATH model specification in this case. The next seven records save the information about the PATH model: 3 paths and 4 partial variances specifications.

Figure 17.8 Spleen Data: OUTMODEL= Data Set with Final Parameter Estimates

Obs	_TYPE_	_NAME_	_VAR1_	_VAR2_	_ESTIM_	_STDERR_
1	MDLTYPE	PATH			.	.
2	LEFT		sqrtrose	factrose	1.0000	.
3	LEFT		sqrtnucl	factnucl	1.0000	.
4	LEFT	beta	factrose	factnucl	0.3907	0.07708
5	PVAR		sqrtrose		0.2500	.
6	PVAR		sqrtnucl		0.2500	.
7	PVAR	v_dist	factrose		0.3815	0.28556
8	PVAR	v_factnu	factnucl		10.5046	4.58577

In each record, the variables involved, the parameter name, the final estimate, and the standard error estimate are stored. For records with fixed parameters, the parameter names entries are blanks and the standard error estimates are indicated by missing values. This data set can be used as input to another run of the TCALIS procedure with the INMODEL= option in the PROC TCALIS statement. For example, if the iteration limit is exceeded, you can use the CALISMODEL 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 CALISMDL 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:

```
data splmod2(type=calismdl);
   set splmod1;
   if _name_='v_dist' then
      do;
         _name_=' ';
         _estim_=0;
      end;
run;
```

Hence, due to the fixed zero error variance for factrose, a model with perfect prediction of factrose from factnucl is specified in the new CALISMDL data set SPLMOD2. This data set serves as the INMODEL= data set in the following statements for another PROC TCALIS run:

```
proc tcalis data=spleen inmodel=splmod2;
run;
```

The main estimation results are displayed in Figure 17.9.

Figure 17.9 Spleen Data: PATH Model Estimates with INMODEL= Data Set

PATH List

----------Path----------			Parameter	Estimate	Standard Error	t Value
sqrtrose	<-	factrose		1.00000		
sqrtnucl	<-	factnucl		1.00000		
factrose	<-	factnucl	beta	0.40340	0.05078	7.94391

Variance Parameters

Variance Type	Variable	Parameter	Estimate	Standard Error	t Value
Error	sqrtrose		0.25000		
	sqrtnucl		0.25000		
	factrose		0		
Exogenous	factnucl	v_factnu	10.45846	4.56608	2.29047

As can be seen in Figure 17.9, error variance for factrose is a fixed zero in the final results. Because of this modification in the current model specified in the SPLMOD2 data set, the estimates of beta and v_factnu are different from those of the previous model results, which are stored in the SPLMOD1 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:

```
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

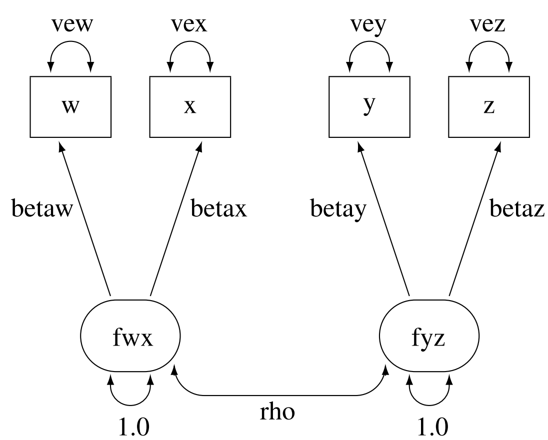
$$
\begin{aligned}
W &= \beta_W F_{WX} + E_W \\
X &= \beta_X F_{WX} + E_X \\
Y &= \beta_Y F_{YZ} + E_Y \\
Z &= \beta_Z F_{YZ} + E_Z
\end{aligned}
$$

with the following assumptions:

$$
\begin{aligned}
\mathrm{Var}(F_{WX}) &= \mathrm{Var}(F_{YZ}) = 1 \\
\mathrm{Cov}(F_{WX}, F_{YZ}) &= \rho \\
\mathrm{Cov}(E_W, E_X) &= \mathrm{Cov}(E_W, E_Y) = \mathrm{Cov}(E_W, E_Z) = \mathrm{Cov}(E_X, E_Y) \\
&= \mathrm{Cov}(E_X, E_Z) = \mathrm{Cov}(E_Y, E_Z) = \mathrm{Cov}(E_W, F_{WX}) \\
&= \mathrm{Cov}(E_W, F_{YZ}) = \mathrm{Cov}(E_X, F_{WX}) = \mathrm{Cov}(E_X, F_{YZ}) \\
&= \mathrm{Cov}(E_Y, F_{WX}) = \mathrm{Cov}(E_Y, F_{YZ}) = \mathrm{Cov}(E_Z, F_{WX}) \\
&= \mathrm{Cov}(E_Z, F_{YZ}) = 0
\end{aligned}
$$

The corresponding path diagram is shown in Figure 17.10.

Figure 17.10 Path Diagram: Lord



With the following rules, the conversion from the path diagram to the PATH model specification is very straightforward:

- Each single-headed arrow in the path diagram is specified in the PATH statement.
- Each double-headed arrow that points to a single variable is specified in the PVAR statement.
- Each double-headed arrow that points to two distinct variables is specified in the PCOV statement.

Hence, this path diagram can be converted easily to a PATH model as follows:

```
title 'H4: Unconstrained';
proc tcalis data=lord outmodel=model4;
   path
      w <- fwx    betaw,
      x <- fwx    betax,
      y <- fyz    betay,
      z <- fyz    betaz;
   pvar
      fwx fyz = 2 * 1.0,
      w x y z = vew vex vey vez;
   pcov
      fwx fyz = rho;
run;
```

The major results are displayed in Figure 17.11 and Figure 17.12.

Figure 17.11 Lord Data: Fit Summary, Hypothesis H4

Fit Summary

Modeling Info	N Observations	649
	N Variables	4
	N Moments	10
	N Parameters	9
	N Active Constraints	0
	Independence Model Chi-Square	1466.5524
	Independence Model Chi-Square DF	6
Absolute Index	Fit Function	0.0011
	Chi-Square	0.7030
	Chi-Square DF	1
	Pr > Chi-Square	0.4018
	Z-Test of Wilson & Hilferty	0.2363
	Hoelter Critical N	3543
	Root Mean Square Residual (RMSR)	0.2720
	Standardized RMSR (SRMSR)	0.0030
	Goodness of Fit Index (GFI)	0.9995
Parsimony Index	Adjusted GFI (AGFI)	0.9946
	Parsimonious GFI	0.1666
	RMSEA Estimate	0.0000
	RMSEA Lower 90% Confidence Limit	.
	RMSEA Upper 90% Confidence Limit	0.0974
	Probability of Close Fit	0.6854
	ECVI Estimate	0.0291
	ECVI Lower 90% Confidence Limit	.
	ECVI Upper 90% Confidence Limit	0.0391
	Akaike Information Criterion	-1.2970
	Bozdogan CAIC	-6.7725
	Schwarz Bayesian Criterion	-5.7725
	McDonald Centrality	1.0002
Incremental Index	Bentler Comparative Fit Index	1.0000
	Bentler-Bonett NFI	0.9995
	Bentler-Bonett Non-normed Index	1.0012
	Bollen Normed Index Rho1	0.9971
	Bollen Non-normed Index Delta2	1.0002
	James et al. Parsimonious NFI	0.1666

Figure 17.12 Lord Data: Estimation Results, Hypothesis H4

PATH List

-------Path-------			Parameter	Estimate	Standard Error	t Value
w	<-	fwx	betaw	7.50066	0.32339	23.19390
x	<-	fwx	betax	7.70266	0.32063	24.02354
y	<-	fyz	betay	8.50947	0.32694	26.02730
z	<-	fyz	betaz	8.67505	0.32560	26.64301

Figure 17.12 *continued*

Variance Parameters

Variance Type	Variable	Parameter	Estimate	Standard Error	t Value
Exogenous	fwx		1.00000		
	fyz		1.00000		
Error	w	vew	30.13796	2.47037	12.19979
	x	vex	26.93217	2.43065	11.08021
	y	vey	24.87396	2.35986	10.54044
	z	vez	22.56264	2.35028	9.60000

Covariances Among Exogenous Variables

Var1	Var2	Parameter	Estimate	Standard Error	t Value
fwx	fyz	rho	0.89855	0.01865	48.17998

It is convenient to create the OUTMODEL= data set called model4 for use in fitting other models with additional constraints. The same analysis can be performed with the LINEQS statement, as specified in the following:

```
title 'H4: Unconstrained; LINEQS Specification';
proc tcalis data=lord;
   lineqs w = betaw fwx + ew,
          x = betax fwx + ex,
          y = betay fyz + ey,
          z = betaz fyz + ez;
   std fwx fyz = 2 * 1.,
       ew ex ey ez = vew vex vey vez;
   cov fwx fyz = rho;
run;
```

Unlike the PATH model specification, in the LINEQS specification you need to specify the error terms explicitly in the LINEQS statement. In the STD statement, you would need to specify the variance parameters for the exogenous variables, including both of the factors and the error terms. However, using the PATH model specification, no explicit names for error or disturbance terms are needed. As a result, the exogenous variance and error variance parameters are both specified in the PVAR statement. This treatment generalizes to the following useful rule about the PATH model specification:

- Each variable in the PATH model specification or path diagram should have a variance or partial variance parameter specified in the PVAR statement—as either an exogenous variance or a partial variance due to error.

The main results from the LINEQS model specification are displayed in Figure 17.13.

Figure 17.13 Lord Data: Using LINEQS Statement for RAM Model, Hypothesis H4

```
                          Linear Equations

          w        =     7.5007*fwx   +  1.0000 ew
          Std Err        0.3234 betaw
          t Value       23.1939
          x        =     7.7027*fwx   +  1.0000 ex
          Std Err        0.3206 betax
          t Value       24.0235
          y        =     8.5095*fyz   +  1.0000 ey
          Std Err        0.3269 betay
          t Value       26.0273
          z        =     8.6751*fyz   +  1.0000 ez
          Std Err        0.3256 betaz
          t Value       26.6430

             Estimates for Variances of Exogenous Variables

Variable                                              Standard
Type          Variable    Parameter      Estimate        Error       t Value

Latent        fwx                         1.00000
              fyz                         1.00000
Error         ew          vew            30.13796      2.47037      12.19979
              ex          vex            26.93217      2.43065      11.08021
              ey          vey            24.87396      2.35986      10.54044
              ez          vez            22.56264      2.35028       9.60000

                 Covariances Among Exogenous Variables

                                                Standard
     Var1     Var2     Parameter      Estimate        Error       t Value

     fwx      fyz      rho             0.89855      0.01865      48.17998
```

Aside from the output format, all estimates in the LINEQS model results in Figure 17.13 match those of the PATH model results in Figure 17.12. In some situations, the PATH and LINEQS statements might 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.

Subsequent analyses are illustrated with the PATH statement rather than the LINEQS statement because it is easier to translate the path diagram to the PATH model specification.

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

H_1: $\rho = 1,$
$\beta_W = \beta_X, \ \mathrm{Var}(E_W) = \mathrm{Var}(E_X)(or \mathrm{vew} = \mathrm{vex}),$
$\beta_Y = \beta_Z, \ \mathrm{Var}(E_Y) = \mathrm{Var}(E_Z)(or \mathrm{vey} = \mathrm{vez})$

H_2: same as H_1: except ρ is unconstrained

H_3: $\rho = 1$

H_4: 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 OUTMODEL= data set called model4. The hypothesis H_3 can be fitted directly or by modifying the model4 data set. Since H_3 differs from H_4 only in that ρ is constrained to equal 1, the model4 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 by setting the variable _ESTIM_ to the value 1. The following statements create a new model stored in the model3 data set that is modified from the model4 data set:

```
data model3(type=calismdl);
   set model4;
   if _name_='rho' then
      do;
         _name_=' ';
         _estim_=1;
      end;
run;
```

In other words, the model information stored in data set model3 is specified exactly as hypothesis H_3 requires. This data set is then read as an INMODEL= data set for the following PROC TCALIS run:

```
title 'H3: W, X, Y, and Z are congeneric';
proc tcalis data=lord inmodel=model3;
run;
```

Another way to specify the model under hypothesis H_3 is to specify the entire PATH model anew, such as in the following statements:

```
title 'H3: W, X, Y, and Z are congeneric';
proc tcalis data=lord;
   path  w <- f  betaw,
         x <- f  betax,
         y <- f  betay,
         z <- f  betaz;
   pvar
      f = 1,
      w x y z = vew vex vey vez;
run;
```

This would produce essentially the same results as those of the analysis based on the model stored in the data set model3. The main results from the analysis with the INMODEL=MODEL3 data set are displayed in Figure 17.14.

Figure 17.14 Lord Data: Major Results for Hypothesis H3

Fit Summary

Modeling Info	N Observations	649
	N Variables	4
	N Moments	10
	N Parameters	8
	N Active Constraints	0
	Independence Model Chi-Square	1466.5524
	Independence Model Chi-Square DF	6
Absolute Index	Fit Function	0.0559
	Chi-Square	36.2095
	Chi-Square DF	2
	Pr > Chi-Square	0.0000
	Z-Test of Wilson & Hilferty	5.2108
	Hoelter Critical N	109
	Root Mean Square Residual (RMSR)	2.4636
	Standardized RMSR (SRMSR)	0.0277
	Goodness of Fit Index (GFI)	0.9714
Parsimony Index	Adjusted GFI (AGFI)	0.8570
	Parsimonious GFI	0.3238
	RMSEA Estimate	0.1625
	RMSEA Lower 90% Confidence Limit	0.1187
	RMSEA Upper 90% Confidence Limit	0.2108
	Probability of Close Fit	0.0000
	ECVI Estimate	0.0808
	ECVI Lower 90% Confidence Limit	0.0561
	ECVI Upper 90% Confidence Limit	0.1170
	Akaike Information Criterion	32.2095
	Bozdogan CAIC	21.2586
	Schwarz Bayesian Criterion	23.2586
	McDonald Centrality	0.9740
Incremental Index	Bentler Comparative Fit Index	0.9766
	Bentler-Bonett NFI	0.9753
	Bentler-Bonett Non-normed Index	0.9297
	Bollen Normed Index Rho1	0.9259
	Bollen Non-normed Index Delta2	0.9766
	James et al. Parsimonious NFI	0.3251

Figure 17.14 *continued*

PATH List

-------Path-------			Parameter	Estimate	Standard Error	t Value
w	<-	fwx	betaw	7.10472	0.32177	22.08019
x	<-	fwx	betax	7.26906	0.31826	22.83965
y	<-	fyz	betay	8.37348	0.32542	25.73160
z	<-	fyz	betaz	8.51057	0.32409	26.25985

Variance Parameters

Variance Type	Variable	Parameter	Estimate	Standard Error	t Value
Exogenous	fwx		1.00000		
	fyz		1.00000		
Error	w	vew	35.92087	2.41466	14.87615
	x	vex	33.42397	2.31038	14.46688
	y	vey	27.16980	2.24619	12.09595
	z	vez	25.38948	2.20839	11.49684

Covariances Among Exogenous Variables

Var1	Var2	Estimate	Standard Error	t Value
fwx	fyz	1.00000		

The hypothesis H_2 requires that several pairs of parameters be constrained to have equal estimates. With PROC TCALIS, 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 PATH and PVAR statements or by using the DATA step to change the values in the model4 data set.

First, you can specify the model directly under the hypothesis H_2; the following PATH model is specified:

```
title 'H2: W and X parallel, Y and Z parallel';
proc tcalis data=lord;
   path
      w <- fwx   betawx,
      x <- fwx   betawx,
      y <- fyz   betayz,
      z <- fyz   betayz;
   pvar
      fwx fyz = 2 * 1.0,
      w x y z = vewx vewx veyz veyz;
   pcov
      fwx fyz = rho;
run;
```

Alternatively, if you use the DATA step to modify from the model4 data set, you would specify a new data set called model2 for storing the model information under the hypothesis H_2, as shown in the following statements:

```
data model2(type=calismdl);
   set model4;
   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';
   if _name_='vex' then _name_='vewx';
   if _name_='vey' then _name_='veyz';
   if _name_='vez' then _name_='veyz';
run;
```

Then you would use model2 as the INMODEL= data set in the following PROC TCALIS run:

```
title 'H2: W and X parallel, Y and Z parallel';
proc tcalis data=lord inmodel=model2;
run;
```

The main results from either of these analyses are displayed in Figure 17.15.

Figure 17.15 Lord Data: Major Results for Hypothesis H2

Fit Summary

Modeling Info	N Observations	649
	N Variables	4
	N Moments	10
	N Parameters	5
	N Active Constraints	0
	Independence Model Chi-Square	1466.5524
	Independence Model Chi-Square DF	6
Absolute Index	Fit Function	0.0030
	Chi-Square	1.9335
	Chi-Square DF	5
	Pr > Chi-Square	0.8583
	Z-Test of Wilson & Hilferty	-1.0768
	Hoelter Critical N	3712
	Root Mean Square Residual (RMSR)	0.6983
	Standardized RMSR (SRMSR)	0.0076
	Goodness of Fit Index (GFI)	0.9985
Parsimony Index	Adjusted GFI (AGFI)	0.9970
	Parsimonious GFI	0.8321
	RMSEA Estimate	0.0000
	RMSEA Lower 90% Confidence Limit	.
	RMSEA Upper 90% Confidence Limit	0.0293
	Probability of Close Fit	0.9936
	ECVI Estimate	0.0185
	ECVI Lower 90% Confidence Limit	.
	ECVI Upper 90% Confidence Limit	0.0276
	Akaike Information Criterion	-8.0665
	Bozdogan CAIC	-35.4436
	Schwarz Bayesian Criterion	-30.4436
	McDonald Centrality	1.0024
Incremental Index	Bentler Comparative Fit Index	1.0000
	Bentler-Bonett NFI	0.9987
	Bentler-Bonett Non-normed Index	1.0025
	Bollen Normed Index Rho1	0.9984
	Bollen Non-normed Index Delta2	1.0021
	James et al. Parsimonious NFI	0.8322

PATH List

-------Path-------			Parameter	Estimate	Standard Error	t Value
w	<-	fwx	betawx	7.60099	0.26844	28.31580
x	<-	fwx	betawx	7.60099	0.26844	28.31580
y	<-	fyz	betayz	8.59186	0.27967	30.72146
z	<-	fyz	betayz	8.59186	0.27967	30.72146

Figure 17.15 *continued*

Variance Parameters

Variance Type	Variable	Parameter	Estimate	Standard Error	t Value
Exogenous	fwx		1.00000		
	fyz		1.00000		
Error	w	vewx	28.55545	1.58641	18.00000
	x	vewx	28.55545	1.58641	18.00000
	y	veyz	23.73200	1.31844	18.00000
	z	veyz	23.73200	1.31844	18.00000

Covariances Among Exogenous Variables

Var1	Var2	Parameter	Estimate	Standard Error	t Value
fwx	fyz	rho	0.89864	0.01865	48.18011

The hypothesis H_1 requires one more constraint in addition to those in H_2. Again, there are two ways to do this. First, a direct model specification is shown in the following statements:

```
title 'H1: W and X parallel, Y and Z parallel, all congeneric';
proc tcalis data=lord;
   path
      w <- f   betawx,
      x <- f   betawx,
      y <- f   betayz,
      z <- f   betayz;
   pvar
      f        = 1.0,
      w x y z = vewx vewx veyz veyz;
run;
```

Alternatively, you can modify the model2 data set to create a new data set model2 that stores the model information required by the hypothesis H_1, as shown in the following statements:

```
data model1(type=calismdl);
   set model2;
   if _name_='rho' then
      do;
         _name_=' ';
         _estim_=1;
      end;
run;
```

You can then pass the model information stored in model1 as an INMODEL= data set in the following PROC TCALIS run:

```
title 'H1: W and X parallel, Y and Z parallel, all congeneric';
proc tcalis data=lord inmodel=model1;
run;
```

The main results from either of these analyses are displayed in Figure 17.16.

Figure 17.16 Lord Data: Major Results for Hypothesis H1

Fit Summary

Modeling Info	N Observations	649
	N Variables	4
	N Moments	10
	N Parameters	4
	N Active Constraints	0
	Independence Model Chi-Square	1466.5524
	Independence Model Chi-Square DF	6
Absolute Index	Fit Function	0.0576
	Chi-Square	37.3337
	Chi-Square DF	6
	Pr > Chi-Square	0.0000
	Z-Test of Wilson & Hilferty	4.5535
	Hoelter Critical N	220
	Root Mean Square Residual (RMSR)	2.5430
	Standardized RMSR (SRMSR)	0.0286
	Goodness of Fit Index (GFI)	0.9705
Parsimony Index	Adjusted GFI (AGFI)	0.9509
	Parsimonious GFI	0.9705
	RMSEA Estimate	0.0898
	RMSEA Lower 90% Confidence Limit	0.0635
	RMSEA Upper 90% Confidence Limit	0.1184
	Probability of Close Fit	0.0076
	ECVI Estimate	0.0701
	ECVI Lower 90% Confidence Limit	0.0458
	ECVI Upper 90% Confidence Limit	0.1059
	Akaike Information Criterion	25.3337
	Bozdogan CAIC	-7.5189
	Schwarz Bayesian Criterion	-1.5189
	McDonald Centrality	0.9761
Incremental Index	Bentler Comparative Fit Index	0.9785
	Bentler-Bonett NFI	0.9745
	Bentler-Bonett Non-normed Index	0.9785
	Bollen Normed Index Rho1	0.9745
	Bollen Non-normed Index Delta2	0.9785
	James et al. Parsimonious NFI	0.9745

Figure 17.16 *continued*

```
                         PATH List

                                                     Standard
  --------Path-------      Parameter      Estimate      Error      t Value

  w       <-     fwx       betawx          7.18622    0.26598     27.01798
  x       <-     fwx       betawx          7.18622    0.26598     27.01798
  y       <-     fyz       betayz          8.44198    0.28000     30.14946
  z       <-     fyz       betayz          8.44198    0.28000     30.14946

                    Variance Parameters

 Variance                                              Standard
 Type          Variable     Parameter      Estimate       Error      t Value

 Exogenous     fwx                          1.00000
               fyz                          1.00000
 Error         w            vewx           34.68878     1.64635     21.07013
               x            vewx           34.68878     1.64635     21.07013
               y            veyz           26.28501     1.39954     18.78118
               z            veyz           26.28501     1.39954     18.78118

              Covariances Among Exogenous Variables

                                          Standard
           Var1     Var2      Estimate       Error       t Value

           fwx      fyz        1.00000
```

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

Hypothesis	Number of Parameters	χ^2	Degrees of Freedom	p-value	$\hat{\rho}$
H_1	4	37.33	6	0.0000	1.0
H_2	5	1.93	5	0.8583	0.8986
H_3	8	36.21	2	0.0000	1.0
H_4	9	0.70	1	0.4018	0.8986

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_4, 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 4 degrees of freedom, which is obviously not significant. So hypothesis H_2 is consistent with the available data.

The estimates of ρ 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.

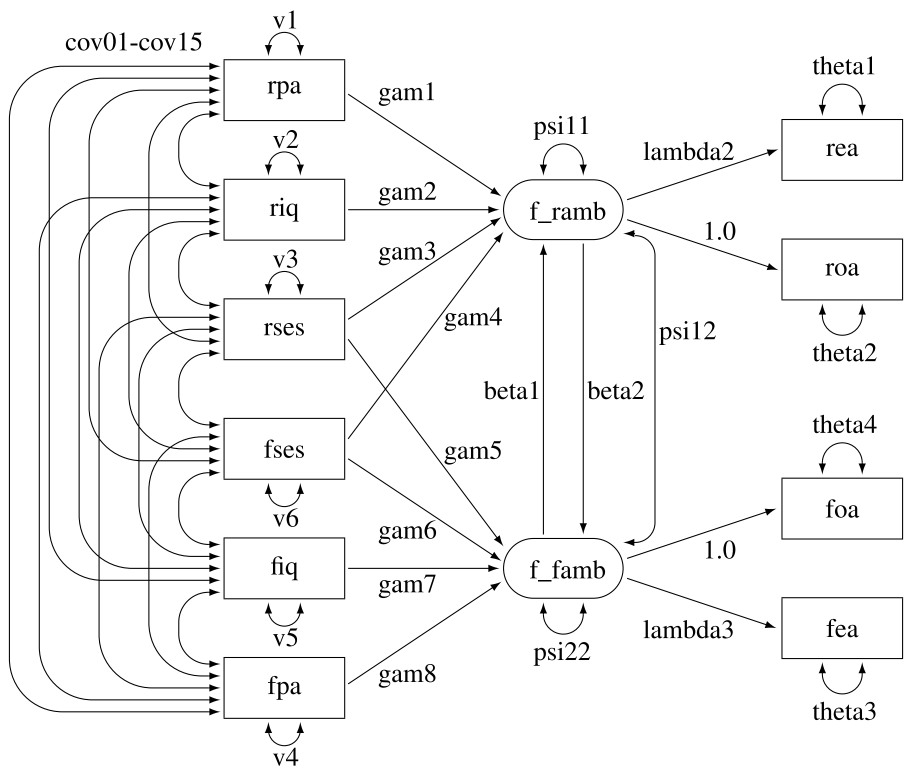
The method of data collection introduces two statistical problems. First, restricting the analysis to boys whose best friends are in the original sample causes the reduced sample to be biased. Second, since the data from a given boy might appear in two or more observations, the observations are not independent. 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 might be as small as half of the reported 329 observations.

The correlation matrix, taken from Jöreskog and Sörbom (1988), is shown in the following DATA step:

```
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  .4047  .6247  1.        .       .       .       .       .
fiq    .3355  .0782  .2302  .2995  .2863  1.        .       .       .       .
fpa    .1021  .1147  .0931  .0760  .0702  .2087    1.       .       .       .
fses   .1861  .0186  .2707  .2930  .2407  .2950  -.0438  1.        .       .
foa    .2598  .0839  .2786  .4216  .3275  .5007   .1988  .3607  1.        .
fea    .2903  .1124  .3054  .3269  .3669  .5191   .2784  .4105  .6404  1.
;
```

The model analyzed by Jöreskog and Sörbom (1988) is displayed in the path diagram in Figure 17.17.

Figure 17.17 Path Diagram: Career Aspiration – Jöreskog and Sörbom (1988)



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 (riq) and socioeconomic status (rses), his perception of his parents' aspiration for him (rpa), and his friend's socioeconomic status (fses) and ambition (f_famb). It is assumed that his friend's intelligence (fiq) and parental aspiration (fpa) affect the respondent's ambition only indirectly through the friend's ambition (f_famb). Ambition is indexed by the manifest variables of occupational (roa) and educational aspiration (rea), 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 TCALIS by using the PATH modeling language, as shown in the following statements:

```
proc tcalis corr data=aspire nobs=329;
   path
      /* measurement model for aspiration */
      rea <-  f_ramb    lambda2,
      roa <-  f_ramb    1.,
      foa <-  f_famb    1.,
      fea <-  f_famb    lambda3,
      /* structural model of influences */
      f_ramb <- rpa     gam1,
      f_ramb <- riq     gam2,
      f_ramb <- rses    gam3,
      f_ramb <- fses    gam4,
      f_famb <- rses    gam5,
      f_famb <- fses    gam6,
      f_famb <- fiq     gam7,
      f_famb <- fpa     gam8,
      f_ramb <- f_famb  beta1,
      f_famb <- f_ramb  beta2;
   pvar
      f_ramb = psi11,
      f_famb = psi22,
      rpa riq rses fpa fiq fses = v1-v6,
      rea roa fea foa = theta1-theta4;
   pcov
      f_ramb f_famb = psi12,
      rpa riq rses fpa fiq fses = 15 * cov__;
run;
```

In this specification, 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 NOBS= option in the PROC TCALIS statement.

Specifying a name followed by double underscores is a quick way to generate unique parameter names. The double underscores are replaced with a unique number each time a new parameter name is generated. For example, in the COV statement, the specification

```
rpa riq rses fpa fiq fses = 15 * cov__;
```

is equivalent to

```
rpa riq rses fpa fiq fses = cov01-cov15;
```

In the PROC TCALIS statement, the CORR option is used to indicate that the correlation matrix is fitted by the model. Fitting correlation matrices by covariance structure modeling method is plagued with some statistical issues. For example, the chi-square statistic might not follow the theoretical distribution well, and the estimates of standard errors might not be accurate. Nonetheless, the correlation matrix is fitted here for illustration and comparison purposes.

The results from this analysis are displayed in Figure 17.18.

Figure 17.18 Career Aspiration Data: Fit Summary of Jöreskog and Sörbom (1988) Analysis 1

Fit Summary

Modeling Info	N Observations	329
	N Variables	10
	N Moments	55
	N Parameters	40
	N Active Constraints	0
	Independence Model Chi-Square	872.0008
	Independence Model Chi-Square DF	45
Absolute Index	Fit Function	0.0814
	Chi-Square	26.6972
	Chi-Square DF	15
	Pr > Chi-Square	0.0313
	Z-Test of Wilson & Hilferty	1.8625
	Hoelter Critical N	309
	Root Mean Square Residual (RMSR)	0.0202
	Standardized RMSR (SRMSR)	0.0202
	Goodness of Fit Index (GFI)	0.9844
Parsimony Index	Adjusted GFI (AGFI)	0.9428
	Parsimonious GFI	0.3281
	RMSEA Estimate	0.0488
	RMSEA Lower 90% Confidence Limit	0.0145
	RMSEA Upper 90% Confidence Limit	0.0783
	Probability of Close Fit	0.4876
	ECVI Estimate	0.3338
	ECVI Lower 90% Confidence Limit	0.3012
	ECVI Upper 90% Confidence Limit	0.3910
	Akaike Information Criterion	-3.3028
	Bozdogan CAIC	-75.2437
	Schwarz Bayesian Criterion	-60.2437
	McDonald Centrality	0.9824
Incremental Index	Bentler Comparative Fit Index	0.9859
	Bentler-Bonett NFI	0.9694
	Bentler-Bonett Non-normed Index	0.9576
	Bollen Normed Index Rho1	0.9082
	Bollen Non-normed Index Delta2	0.9864
	James et al. Parsimonious NFI	0.3231

Jöreskog 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 (that is, beta1 $=$ beta2).
- The covariance of the disturbances of the ambition variables is zero (that is, psi12 $= 0$).

This analysis can be performed by changing the names beta1 and beta2 to beta and omitting the line from the COV statement for psi12, as shown in the following statements:

```
proc tcalis corr data=aspire nobs=329;
   path
      /* measurement model for aspiration */
      rea <-  f_ramb    lambda2,
      roa <-  f_ramb    1.,
      foa <-  f_famb    1.,
      fea <-  f_famb    lambda3,
      /* structural model of influences */
      f_ramb <- rpa     gam1,
      f_ramb <- riq     gam2,
      f_ramb <- rses    gam3,
      f_ramb <- fses    gam4,
      f_famb <- rses    gam5,
      f_famb <- fses    gam6,
      f_famb <- fiq     gam7,
      f_famb <- fpa     gam8,
      f_ramb <- f_famb  beta,
      f_famb <- f_ramb  beta;
   pvar
      f_ramb = psi11,
      f_famb = psi22,
      rpa riq rses fpa fiq fses = v1-v6,
      rea roa fea foa = theta1-theta4;
   pcov
      rpa riq rses fpa fiq fses = 15 * cov__;
 run;
```

The fit summary is displayed in Figure 17.19, and the estimation results are displayed in Figure 17.20.

Figure 17.19 Career Aspiration Data: Fit Summary of Jöreskog and Sörbom (1988) Analysis 2

Fit Summary

Modeling Info	N Observations	329
	N Variables	10
	N Moments	55
	N Parameters	38
	N Active Constraints	0
	Independence Model Chi-Square	872.0008
	Independence Model Chi-Square DF	45
Absolute Index	Fit Function	0.0820
	Chi-Square	26.8987
	Chi-Square DF	17
	Pr > Chi-Square	0.0596
	Z-Test of Wilson & Hilferty	1.5599
	Hoelter Critical N	338
	Root Mean Square Residual (RMSR)	0.0203
	Standardized RMSR (SRMSR)	0.0203
	Goodness of Fit Index (GFI)	0.9843
Parsimony Index	Adjusted GFI (AGFI)	0.9492
	Parsimonious GFI	0.3718
	RMSEA Estimate	0.0421
	RMSEA Lower 90% Confidence Limit	.
	RMSEA Upper 90% Confidence Limit	0.0710
	Probability of Close Fit	0.6367
	ECVI Estimate	0.3218
	ECVI Lower 90% Confidence Limit	.
	ECVI Upper 90% Confidence Limit	0.3781
	Akaike Information Criterion	-7.1013
	Bozdogan CAIC	-88.6343
	Schwarz Bayesian Criterion	-71.6343
	McDonald Centrality	0.9851
Incremental Index	Bentler Comparative Fit Index	0.9880
	Bentler-Bonett NFI	0.9692
	Bentler-Bonett Non-normed Index	0.9683
	Bollen Normed Index Rho1	0.9183
	Bollen Non-normed Index Delta2	0.9884
	James et al. Parsimonious NFI	0.3661

Figure 17.20 Career Aspiration Data: Estimation Results of Jöreskog and Sörbom (1988) Analysis 2

PATH List

---------Path---------			Parameter	Estimate	Standard Error	t Value
rea	<-	f_ramb	lambda2	1.06097	0.08921	11.89233
roa	<-	f_ramb		1.00000		
foa	<-	f_famb		1.00000		
fea	<-	f_famb	lambda3	1.07359	0.08063	13.31498
f_ramb	<-	rpa	gam1	0.16367	0.03872	4.22740
f_ramb	<-	riq	gam2	0.25395	0.04186	6.06725
f_ramb	<-	rses	gam3	0.22115	0.04187	5.28219
f_ramb	<-	fses	gam4	0.07728	0.04149	1.86264
f_famb	<-	rses	gam5	0.06840	0.03868	1.76809
f_famb	<-	fses	gam6	0.21839	0.03948	5.53198
f_famb	<-	fiq	gam7	0.33063	0.04116	8.03314
f_famb	<-	fpa	gam8	0.15204	0.03636	4.18169
f_ramb	<-	f_famb	beta	0.18007	0.03912	4.60305
f_famb	<-	f_ramb	beta	0.18007	0.03912	4.60305

Variance Parameters

Variance Type	Variable	Parameter	Estimate	Standard Error	t Value
Error	f_ramb	psi11	0.28113	0.04640	6.05867
	f_famb	psi22	0.22924	0.03889	5.89393
Exogenous	rpa	v1	1.00000	0.07809	12.80625
	riq	v2	1.00000	0.07809	12.80625
	rses	v3	1.00000	0.07809	12.80625
	fpa	v4	1.00000	0.07809	12.80625
	fiq	v5	1.00000	0.07809	12.80625
	fses	v6	1.00000	0.07809	12.80625
Error	rea	theta1	0.33764	0.05178	6.52039
	roa	theta2	0.41205	0.05103	8.07403
	fea	theta3	0.31337	0.04574	6.85165
	foa	theta4	0.40381	0.04608	8.76428

Figure 17.20 *continued*

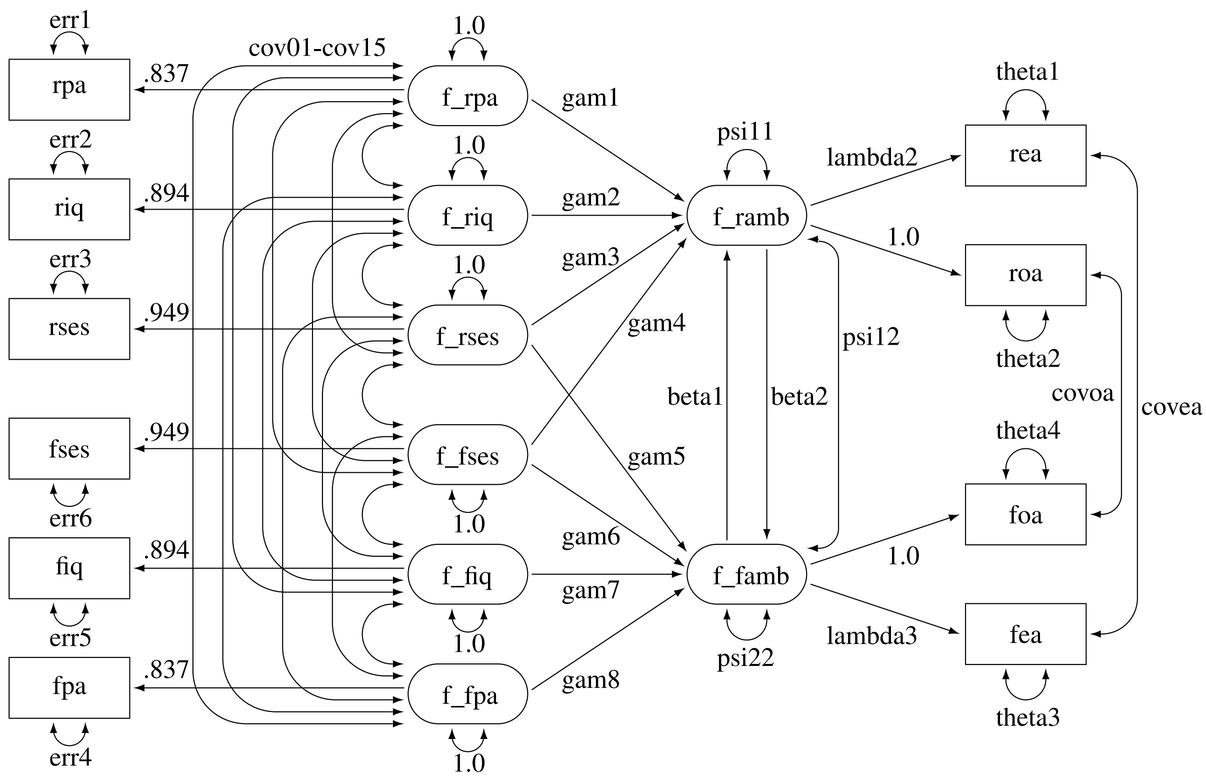
Covariances Among Exogenous Variables

Var1	Var2	Parameter	Estimate	Standard Error	t Value
rpa	riq	cov01	0.18390	0.05614	3.27564
rpa	rses	cov02	0.04890	0.05528	0.88456
riq	rses	cov03	0.22200	0.05656	3.92503
rpa	fpa	cov04	0.11470	0.05558	2.06377
riq	fpa	cov05	0.10210	0.05550	1.83955
rses	fpa	cov06	0.09310	0.05545	1.67885
rpa	fiq	cov07	0.07820	0.05538	1.41195
riq	fiq	cov08	0.33550	0.05824	5.76060
rses	fiq	cov09	0.23020	0.05666	4.06284
fpa	fiq	cov10	0.20870	0.05641	3.70000
rpa	fses	cov11	0.01860	0.05523	0.33680
riq	fses	cov12	0.18610	0.05616	3.31352
rses	fses	cov13	0.27070	0.05720	4.73226
fpa	fses	cov14	-0.04380	0.05527	-0.79249
fiq	fses	cov15	0.29500	0.05757	5.12435

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. This indicates that the restricted model (analysis 2) fits as well as the unrestricted model (analysis 1). However, the chi-square test of the restricted model against the alternative of a completely unrestricted covariance matrix yields a p-value of 0.0596, which indicates that the model might 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 respects. 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 might 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 path diagram for Loehlin's model 1 shown in Figure 17.21.

Figure 17.21 Path Diagram: Career Aspiration – Loehlin (1987)



In Figure 17.21, the observed variables rpa, riq, rses, fses, fiq, and fpa are all treated as measurements with errors. Their purified counterparts f_rpa, f_riq, f_rses, f_fses, f_fiq, and f_fpa are latent variables created in the model to represent measurements without errors. Path coefficients from these latent variables to the observed variables are fixed coefficients, indicating the square roots of the theoretical reliabilities in the model. These latent variables, rather than the observed counterparts, serve as predictors of the ambition variables f_ramb and f_famb. Correlated errors for the occupational aspiration variables and the educational aspiration variables are also shown in Figure 17.21. The error covariance for the educational aspiration variables rea and fea is indicated by the parameter covea, and the error covariance for the occupational aspiration variables roa and foa is indicated by the parameter covoa.

The statements for fitting this model by using the PATH modeling language are as follows:

```
proc tcalis corr data=aspire nobs=329;
   path
      /* measurement model for aspiration */
      rea  <-  f_ramb    lambda2,
      roa  <-  f_ramb    1.,
      foa  <-  f_famb    1.,
      fea  <-  f_famb    lambda3,

      /* measurement model for intelligence and environment */
      rpa  <-  f_rpa     0.837,
      riq  <-  f_riq     0.894,
      rses <-  f_rses    0.949,
      fses <-  f_fses    0.949,
      fiq  <-  f_fiq     0.894,
      fpa  <-  f_fpa     0.837,

      /* structural model of influences */
      f_ramb <- f_rpa     gam1,
      f_ramb <- f_riq     gam2,
      f_ramb <- f_rses    gam3,
      f_ramb <- f_fses    gam4,
      f_famb <- f_rses    gam5,
      f_famb <- f_fses    gam6,
      f_famb <- f_fiq     gam7,
      f_famb <- f_fpa     gam8,
      f_ramb <- f_famb    beta1,
      f_famb <- f_ramb    beta2;
   pvar
      f_ramb = psi11,
      f_famb = psi22,
      f_rpa f_riq f_rses f_fses f_fiq f_fpa = 6 * 1.0,
      rea roa fea foa                       = theta1-theta4,
      rpa riq rses fpa fiq fses             = err1-err6;
   pcov
      f_ramb f_famb   = psi12,
      rea  fea        = covea,
      roa  foa        = covoa,
      f_rpa f_riq f_rses f_fses f_fiq f_fpa = 15 * cov__;
run;
```

The fit summary is displayed in Figure 17.22, and the estimation results are displayed in Figure 17.23.

Figure 17.22 Career Aspiration Data: Fit Summary of Loehlin (1987) Model 1

Fit Summary

Modeling Info	N Observations	329
	N Variables	10
	N Moments	55
	N Parameters	42
	N Active Constraints	0
	Independence Model Chi-Square	872.0008
	Independence Model Chi-Square DF	45
Absolute Index	Fit Function	0.0366
	Chi-Square	12.0132
	Chi-Square DF	13
	Pr > Chi-Square	0.5266
	Z-Test of Wilson & Hilferty	-0.0679
	Hoelter Critical N	612
	Root Mean Square Residual (RMSR)	0.0149
	Standardized RMSR (SRMSR)	0.0149
	Goodness of Fit Index (GFI)	0.9927
Parsimony Index	Adjusted GFI (AGFI)	0.9692
	Parsimonious GFI	0.2868
	RMSEA Estimate	0.0000
	RMSEA Lower 90% Confidence Limit	.
	RMSEA Upper 90% Confidence Limit	0.0512
	Probability of Close Fit	0.9435
	ECVI Estimate	0.3016
	ECVI Lower 90% Confidence Limit	.
	ECVI Upper 90% Confidence Limit	0.3392
	Akaike Information Criterion	-13.9868
	Bozdogan CAIC	-76.3356
	Schwarz Bayesian Criterion	-63.3356
	McDonald Centrality	1.0015
Incremental Index	Bentler Comparative Fit Index	1.0000
	Bentler-Bonett NFI	0.9862
	Bentler-Bonett Non-normed Index	1.0041
	Bollen Normed Index Rho1	0.9523
	Bollen Non-normed Index Delta2	1.0011
	James et al. Parsimonious NFI	0.2849

Figure 17.23 Career Aspiration Data: Estimation Results of Loehlin (1987) Model 1

PATH List

---------Path---------			Parameter	Estimate	Standard Error	t Value
rea	<-	f_ramb	lambda2	1.08398	0.09417	11.51054
roa	<-	f_ramb		1.00000		
foa	<-	f_famb		1.00000		
fea	<-	f_famb	lambda3	1.11631	0.08627	12.93937
rpa	<-	f_rpa		0.83700		
riq	<-	f_riq		0.89400		
rses	<-	f_rses		0.94900		
fses	<-	f_fses		0.94900		
fiq	<-	f_fiq		0.89400		
fpa	<-	f_fpa		0.83700		
f_ramb	<-	f_rpa	gam1	0.18370	0.05044	3.64201
f_ramb	<-	f_riq	gam2	0.28004	0.06139	4.56183
f_ramb	<-	f_rses	gam3	0.22616	0.05223	4.33004
f_ramb	<-	f_fses	gam4	0.08698	0.05476	1.58836
f_famb	<-	f_rses	gam5	0.06327	0.05219	1.21240
f_famb	<-	f_fses	gam6	0.21539	0.05121	4.20600
f_famb	<-	f_fiq	gam7	0.35386	0.06741	5.24971
f_famb	<-	f_fpa	gam8	0.16877	0.04934	3.42051
f_ramb	<-	f_famb	beta1	0.11897	0.11396	1.04397
f_famb	<-	f_ramb	beta2	0.13022	0.12067	1.07921

Variance Parameters

Variance Type	Variable	Parameter	Estimate	Standard Error	t Value
Error	f_ramb	psi11	0.25418	0.04469	5.68738
	f_famb	psi22	0.19698	0.03814	5.16533
Exogenous	f_rpa		1.00000		
	f_riq		1.00000		
	f_rses		1.00000		
	f_fses		1.00000		
	f_fiq		1.00000		
	f_fpa		1.00000		
Error	rea	theta1	0.32707	0.05452	5.99883
	roa	theta2	0.42307	0.05243	8.06948
	fea	theta3	0.28715	0.04804	5.97748
	foa	theta4	0.42240	0.04730	8.93103
	rpa	err1	0.29584	0.07774	3.80573
	riq	err2	0.20874	0.07832	2.66519
	rses	err3	0.09887	0.07803	1.26715
	fpa	err4	0.29987	0.07807	3.84089
	fiq	err5	0.19988	0.07674	2.60475
	fses	err6	0.10324	0.07824	1.31950

Figure 17.23 *continued*

Covariances Among Exogenous Variables

Var1	Var2	Parameter	Estimate	Standard Error	t Value
f_rpa	f_riq	cov01	0.24677	0.07519	3.28204
f_rpa	f_rses	cov02	0.06184	0.06945	0.89034
f_riq	f_rses	cov03	0.26351	0.06687	3.94075
f_rpa	f_fses	cov04	0.02383	0.06952	0.34272
f_riq	f_fses	cov05	0.22135	0.06648	3.32976
f_rses	f_fses	cov06	0.30156	0.06359	4.74205
f_rpa	f_fiq	cov07	0.10853	0.07362	1.47419
f_riq	f_fiq	cov08	0.42476	0.07219	5.88373
f_rses	f_fiq	cov09	0.27250	0.06660	4.09143
f_fses	f_fiq	cov10	0.34922	0.06771	5.15755
f_rpa	f_fpa	cov11	0.15789	0.07873	2.00553
f_riq	f_fpa	cov12	0.13085	0.07418	1.76393
f_rses	f_fpa	cov13	0.11517	0.06978	1.65053
f_fses	f_fpa	cov14	-0.05623	0.06971	-0.80655
f_fiq	f_fpa	cov15	0.27867	0.07530	3.70083

Covariances Among Errors

Error of	Error of	Parameter	Estimate	Standard Error	t Value
f_ramb	f_famb	psi12	-0.00935	0.05010	-0.18669
rea	fea	covea	0.02308	0.03139	0.73543
roa	foa	covoa	0.11206	0.03258	3.43993

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 his 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 (1987) model 2. The PATH model is as follows, where an OUTMODEL= data set is created to facilitate subsequent hypothesis tests:

```
proc tcalis corr data=aspire nobs=329 outmodel=model2;
   path
      /* measurement model for aspiration */
      rea  <-  f_ramb    lambda,             /* 1 ec! */
      roa  <-  f_ramb    1.,
      foa  <-  f_famb    1.,
      fea  <-  f_famb    lambda,

      /* measurement model for intelligence and environment */
      rpa  <-  f_rpa     0.837,
      riq  <-  f_riq     0.894,
      rses <-  f_rses    0.949,
      fses <-  f_fses    0.949,
      fiq  <-  f_fiq     0.894,
      fpa  <-  f_fpa     0.837,

      /* structural model of influences */
      f_ramb <- f_rpa     gam1,              /* 5 ec! */
      f_ramb <- f_riq     gam2,
      f_ramb <- f_rses    gam3,
      f_ramb <- f_fses    gam4,
      f_famb <- f_rses    gam4,
      f_famb <- f_fses    gam3,
      f_famb <- f_fiq     gam2,
      f_famb <- f_fpa     gam1,
      f_ramb <- f_famb    beta,
      f_famb <- f_ramb    beta;
   pvar
      f_ramb = psi,                          /* 1 ec! */
      f_famb = psi,
      f_rpa f_riq f_rses f_fpa f_fiq f_fses = 6 * 1.0,
      rea fea                = 2 * thetaea,  /* 2 ec! */
      roa foa                = 2 * thetaoa,
      rpa fpa                = errpa1 errpa2,
      riq fiq                = erriq1 erriq2,
      rses fses              = errses1 errses2;
   pcov
      f_ramb f_famb          = psi12,
      rea  fea               = covea,
      roa  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 fit summary is displayed in Figure 17.24, and the estimation results are displayed in Figure 17.25.

Figure 17.24 Career Aspiration Data: Fit Summary of Loehlin (1987) Model 2

Fit Summary

Modeling Info	N Observations	329
	N Variables	10
	N Moments	55
	N Parameters	27
	N Active Constraints	0
	Independence Model Chi-Square	872.0008
	Independence Model Chi-Square DF	45
Absolute Index	Fit Function	0.0581
	Chi-Square	19.0697
	Chi-Square DF	28
	Pr > Chi-Square	0.8960
	Z-Test of Wilson & Hilferty	-1.2599
	Hoelter Critical N	713
	Root Mean Square Residual (RMSR)	0.0276
	Standardized RMSR (SRMSR)	0.0276
	Goodness of Fit Index (GFI)	0.9884
Parsimony Index	Adjusted GFI (AGFI)	0.9772
	Parsimonious GFI	0.6150
	RMSEA Estimate	0.0000
	RMSEA Lower 90% Confidence Limit	.
	RMSEA Upper 90% Confidence Limit	0.0194
	Probability of Close Fit	0.9996
	ECVI Estimate	0.2285
	ECVI Lower 90% Confidence Limit	.
	ECVI Upper 90% Confidence Limit	0.2664
	Akaike Information Criterion	-36.9303
	Bozdogan CAIC	-171.2200
	Schwarz Bayesian Criterion	-143.2200
	McDonald Centrality	1.0137
Incremental Index	Bentler Comparative Fit Index	1.0000
	Bentler-Bonett NFI	0.9781
	Bentler-Bonett Non-normed Index	1.0174
	Bollen Normed Index Rho1	0.9649
	Bollen Non-normed Index Delta2	1.0106
	James et al. Parsimonious NFI	0.6086

Figure 17.25 Career Aspiration Data: Estimation Results of Loehlin (1987) Model 2

PATH List

---------Path---------			Parameter	Estimate	Standard Error	t Value
rea	<-	f_ramb	lambda	1.10067	0.06842	16.08795
roa	<-	f_ramb		1.00000		
foa	<-	f_famb		1.00000		
fea	<-	f_famb	lambda	1.10067	0.06842	16.08795
rpa	<-	f_rpa		0.83700		
riq	<-	f_riq		0.89400		
rses	<-	f_rses		0.94900		
fses	<-	f_fses		0.94900		
fiq	<-	f_fiq		0.89400		
fpa	<-	f_fpa		0.83700		
f_ramb	<-	f_rpa	gam1	0.17585	0.03508	5.01299
f_ramb	<-	f_riq	gam2	0.32234	0.04702	6.85568
f_ramb	<-	f_rses	gam3	0.22273	0.03629	6.13725
f_ramb	<-	f_fses	gam4	0.07564	0.03750	2.01699
f_famb	<-	f_rses	gam4	0.07564	0.03750	2.01699
f_famb	<-	f_fses	gam3	0.22273	0.03629	6.13725
f_famb	<-	f_fiq	gam2	0.32234	0.04702	6.85568
f_famb	<-	f_fpa	gam1	0.17585	0.03508	5.01299
f_ramb	<-	f_famb	beta	0.11578	0.08390	1.38007
f_famb	<-	f_ramb	beta	0.11578	0.08390	1.38007

Variance Parameters

Variance Type	Variable	Parameter	Estimate	Standard Error	t Value
Error	f_ramb	psi	0.22456	0.02971	7.55930
	f_famb	psi	0.22456	0.02971	7.55930
Exogenous	f_rpa		1.00000		
	f_riq		1.00000		
	f_rses		1.00000		
	f_fpa		1.00000		
	f_fiq		1.00000		
	f_fses		1.00000		
Error	rea	thetaea	0.30662	0.03726	8.22956
	fea	thetaea	0.30662	0.03726	8.22956
	roa	thetaoa	0.42295	0.03651	11.58311
	foa	thetaoa	0.42295	0.03651	11.58311
	rpa	errpa1	0.30758	0.07511	4.09498
	fpa	errpa2	0.28834	0.07369	3.91289
	riq	erriq1	0.26656	0.07389	3.60730
	fiq	erriq2	0.15573	0.06700	2.32445
	rses	errses1	0.11467	0.07267	1.57800
	fses	errses2	0.08814	0.07089	1.24333

Figure 17.25 *continued*

Covariances Among Exogenous Variables

Var1	Var2	Parameter	Estimate	Standard Error	t Value
f_rpa	f_riq	cov1	0.26470	0.05442	4.86370
f_rpa	f_rses	cov2	0.00177	0.04996	0.03533
f_riq	f_rses	cov3	0.31129	0.05057	6.15553
f_fpa	f_fiq	cov1	0.26470	0.05442	4.86370
f_fpa	f_fses	cov2	0.00177	0.04996	0.03533
f_fiq	f_fses	cov3	0.31129	0.05057	6.15553
f_rpa	f_fpa	cov4	0.15784	0.07872	2.00521
f_rpa	f_fiq	cov5	0.11837	0.05447	2.17325
f_rpa	f_fses	cov6	0.06910	0.04996	1.38303
f_riq	f_fpa	cov5	0.11837	0.05447	2.17325
f_riq	f_fiq	cov7	0.43061	0.07258	5.93255
f_riq	f_fses	cov8	0.24967	0.05060	4.93420
f_rses	f_fpa	cov6	0.06910	0.04996	1.38303
f_rses	f_fiq	cov8	0.24967	0.05060	4.93420
f_rses	f_fses	cov9	0.30190	0.06362	4.74578

Covariances Among Errors

Error of	Error of	Parameter	Estimate	Standard Error	t Value
f_ramb	f_famb	psi12	-0.00344	0.04931	-0.06981
rea	fea	covea	0.02160	0.03144	0.68686
roa	foa	covoa	0.11208	0.03257	3.44076

The test of Loehlin's (1987) 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. This indicates the restricted model 2 fits at least as well as model 1. Schwarz's Bayesian criterion (SBC) is also much lower for model 2 (-143.2200) than for 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_fses to f_ramb and from f_rses to f_famb designated by the parameter name gam4, yielding Loehlin's (1987) model 3:

```
title2 'Loehlin (1987) analysis: Model 3';
data model3(type=calismdl);
   set model2;
   if _name_='gam4' then
      do;
         _name_=' ';
         _estim_=0;
      end;
run;

proc tcalis corr data=aspire nobs=329 inmodel=model3;
run;
```

The fit summary is displayed in Figure 17.26.

Figure 17.26 Career Aspiration Data: Fit Summary of Loehlin (1987) Model 3

Fit Summary

Modeling Info	N Observations	329
	N Variables	10
	N Moments	55
	N Parameters	26
	N Active Constraints	0
	Independence Model Chi-Square	872.0008
	Independence Model Chi-Square DF	45
Absolute Index	Fit Function	0.0702
	Chi-Square	23.0365
	Chi-Square DF	29
	Pr > Chi-Square	0.7749
	Z-Test of Wilson & Hilferty	-0.7563
	Hoelter Critical N	607
	Root Mean Square Residual (RMSR)	0.0304
	Standardized RMSR (SRMSR)	0.0304
	Goodness of Fit Index (GFI)	0.9858
Parsimony Index	Adjusted GFI (AGFI)	0.9731
	Parsimonious GFI	0.6353
	RMSEA Estimate	0.0000
	RMSEA Lower 90% Confidence Limit	.
	RMSEA Upper 90% Confidence Limit	0.0295
	Probability of Close Fit	0.9984
	ECVI Estimate	0.2343
	ECVI Lower 90% Confidence Limit	.
	ECVI Upper 90% Confidence Limit	0.2780
	Akaike Information Criterion	-34.9635
	Bozdogan CAIC	-174.0492
	Schwarz Bayesian Criterion	-145.0492
	McDonald Centrality	1.0091
Incremental Index	Bentler Comparative Fit Index	1.0000
	Bentler-Bonett NFI	0.9736
	Bentler-Bonett Non-normed Index	1.0112
	Bollen Normed Index Rho1	0.9590
	Bollen Non-normed Index Delta2	1.0071
	James et al. Parsimonious NFI	0.6274

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_ramb and f_famb to zero to obtain Loehlin's (1987) model 4:

```
title2 'Loehlin (1987) analysis: Model 4';
data model4(type=calismdl);
   set model2;
   if _name_='beta' then
      do;
         _name_=' ';
         _estim_=0;
      end;
run;

proc tcalis corr data=aspire nobs=329 inmodel=model4;
run;
```

The fit summary is displayed in Figure 17.27, and the estimation results are displayed in Figure 17.28.

Figure 17.27 Career Aspiration Data: Fit Summary of Loehlin (1987) Model 4

Fit Summary

Modeling Info	N Observations	329
	N Variables	10
	N Moments	55
	N Parameters	26
	N Active Constraints	0
	Independence Model Chi-Square	872.0008
	Independence Model Chi-Square DF	45
Absolute Index	Fit Function	0.0640
	Chi-Square	20.9981
	Chi-Square DF	29
	Pr > Chi-Square	0.8592
	Z-Test of Wilson & Hilferty	-1.0780
	Hoelter Critical N	666
	Root Mean Square Residual (RMSR)	0.0304
	Standardized RMSR (SRMSR)	0.0304
	Goodness of Fit Index (GFI)	0.9873
Parsimony Index	Adjusted GFI (AGFI)	0.9760
	Parsimonious GFI	0.6363
	RMSEA Estimate	0.0000
	RMSEA Lower 90% Confidence Limit	.
	RMSEA Upper 90% Confidence Limit	0.0234
	Probability of Close Fit	0.9994
	ECVI Estimate	0.2281
	ECVI Lower 90% Confidence Limit	.
	ECVI Upper 90% Confidence Limit	0.2685
	Akaike Information Criterion	-37.0019
	Bozdogan CAIC	-176.0876
	Schwarz Bayesian Criterion	-147.0876
	McDonald Centrality	1.0122
Incremental Index	Bentler Comparative Fit Index	1.0000
	Bentler-Bonett NFI	0.9759
	Bentler-Bonett Non-normed Index	1.0150
	Bollen Normed Index Rho1	0.9626
	Bollen Non-normed Index Delta2	1.0095
	James et al. Parsimonious NFI	0.6289

Figure 17.28 Career Aspiration Data: Estimation Results of Loehlin (1987) Model 4

PATH List

---------Path---------			Parameter	Estimate	Standard Error	t Value
rea	<-	f_ramb	lambda	1.10505	0.06804	16.24157
roa	<-	f_ramb		1.00000		
foa	<-	f_famb		1.00000		
fea	<-	f_famb	lambda	1.10505	0.06804	16.24157
rpa	<-	f_rpa		0.83700		
riq	<-	f_riq		0.89400		
rses	<-	f_rses		0.94900		
fses	<-	f_fses		0.94900		
fiq	<-	f_fiq		0.89400		
fpa	<-	f_fpa		0.83700		
f_ramb	<-	f_rpa	gam1	0.17760	0.03610	4.91945
f_ramb	<-	f_riq	gam2	0.34856	0.04625	7.53618
f_ramb	<-	f_rses	gam3	0.23834	0.03549	6.71576
f_ramb	<-	f_fses	gam4	0.10810	0.02992	3.61340
f_famb	<-	f_rses	gam4	0.10810	0.02992	3.61340
f_famb	<-	f_fses	gam3	0.23834	0.03549	6.71576
f_famb	<-	f_fiq	gam2	0.34856	0.04625	7.53618
f_famb	<-	f_fpa	gam1	0.17760	0.03610	4.91945
f_ramb	<-	f_famb		0		
f_famb	<-	f_ramb		0		

Variance Parameters

Variance Type	Variable	Parameter	Estimate	Standard Error	t Value
Error	f_ramb	psi	0.22738	0.03140	7.24263
	f_famb	psi	0.22738	0.03140	7.24263
Exogenous	f_rpa		1.00000		
	f_riq		1.00000		
	f_rses		1.00000		
	f_fpa		1.00000		
	f_fiq		1.00000		
	f_fses		1.00000		
Error	rea	thetaea	0.30502	0.03728	8.18091
	fea	thetaea	0.30502	0.03728	8.18091
	roa	thetaoa	0.42429	0.03645	11.64071
	foa	thetaoa	0.42429	0.03645	11.64071
	rpa	errpa1	0.31354	0.07543	4.15664
	fpa	errpa2	0.29051	0.07374	3.93945
	riq	erriq1	0.29611	0.07299	4.05703
	fiq	erriq2	0.18181	0.06611	2.75034
	rses	errses1	0.12320	0.07273	1.69400
	fses	errses2	0.09873	0.07109	1.38881

Figure 17.28 *continued*

Covariances Among Exogenous Variables

Var1	Var2	Parameter	Estimate	Standard Error	t Value
f_rpa	f_riq	cov1	0.27241	0.05520	4.93523
f_rpa	f_rses	cov2	0.00476	0.05032	0.09455
f_riq	f_rses	cov3	0.32463	0.05089	6.37870
f_fpa	f_fiq	cov1	0.27241	0.05520	4.93523
f_fpa	f_fses	cov2	0.00476	0.05032	0.09455
f_fiq	f_fses	cov3	0.32463	0.05089	6.37870
f_rpa	f_fpa	cov4	0.16949	0.07863	2.15559
f_rpa	f_fiq	cov5	0.13539	0.05407	2.50384
f_rpa	f_fses	cov6	0.07362	0.05027	1.46453
f_riq	f_fpa	cov5	0.13539	0.05407	2.50384
f_riq	f_fiq	cov7	0.46893	0.06980	6.71822
f_riq	f_fses	cov8	0.26289	0.05093	5.16164
f_rses	f_fpa	cov6	0.07362	0.05027	1.46453
f_rses	f_fiq	cov8	0.26289	0.05093	5.16164
f_rses	f_fses	cov9	0.30880	0.06409	4.81849

Covariances Among Errors

Error of	Error of	Parameter	Estimate	Standard Error	t Value
f_ramb	f_famb	psi12	0.05479	0.02699	2.03009
rea	fea	covea	0.02127	0.03150	0.67534
roa	foa	covoa	0.11245	0.03258	3.45136

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 (1987) model 2 has not only the direct paths connecting the latent ambition variables f_ramb and f_famb but also a covariance between the disturbance terms d_ramb and d_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 psi12 to zero, yielding Loehlin's (1987) model 5:

```
title2 'Loehlin (1987) analysis: Model 5';
data model5(type=calismdl);
   set model2;
   if _name_='psi12' then
      do;
         _name_=' ';
         _estim_=0;
      end;
run;

proc tcalis corr data=aspire nobs=329 inmodel=model5;
run;
```

The fit summary is displayed in Figure 17.29, and the estimation results are displayed in Figure 17.30.

Figure 17.29 Career Aspiration Data: Fit Summary of Loehlin (1987) Model 5

Fit Summary

Modeling Info	N Observations	329
	N Variables	10
	N Moments	55
	N Parameters	26
	N Active Constraints	0
	Independence Model Chi-Square	872.0008
	Independence Model Chi-Square DF	45
Absolute Index	Fit Function	0.0582
	Chi-Square	19.0745
	Chi-Square DF	29
	Pr > Chi-Square	0.9194
	Z-Test of Wilson & Hilferty	-1.4014
	Hoelter Critical N	733
	Root Mean Square Residual (RMSR)	0.0276
	Standardized RMSR (SRMSR)	0.0276
	Goodness of Fit Index (GFI)	0.9884
Parsimony Index	Adjusted GFI (AGFI)	0.9780
	Parsimonious GFI	0.6370
	RMSEA Estimate	0.0000
	RMSEA Lower 90% Confidence Limit	.
	RMSEA Upper 90% Confidence Limit	0.0152
	Probability of Close Fit	0.9998
	ECVI Estimate	0.2222
	ECVI Lower 90% Confidence Limit	.
	ECVI Upper 90% Confidence Limit	0.2592
	Akaike Information Criterion	-38.9255
	Bozdogan CAIC	-178.0111
	Schwarz Bayesian Criterion	-149.0111
	McDonald Centrality	1.0152
Incremental Index	Bentler Comparative Fit Index	1.0000
	Bentler-Bonett NFI	0.9781
	Bentler-Bonett Non-normed Index	1.0186
	Bollen Normed Index Rho1	0.9661
	Bollen Non-normed Index Delta2	1.0118
	James et al. Parsimonious NFI	0.6303

Figure 17.30 Career Aspiration Data: Estimation Results of Loehlin (1987) Model 5

PATH List

---------Path---------			Parameter	Estimate	Standard Error	t Value
rea	<-	f_ramb	lambda	1.10086	0.06836	16.10408
roa	<-	f_ramb		1.00000		
foa	<-	f_famb		1.00000		
fea	<-	f_famb	lambda	1.10086	0.06836	16.10408
rpa	<-	f_rpa		0.83700		
riq	<-	f_riq		0.89400		
rses	<-	f_rses		0.94900		
fses	<-	f_fses		0.94900		
fiq	<-	f_fiq		0.89400		
fpa	<-	f_fpa		0.83700		
f_ramb	<-	f_rpa	gam1	0.17618	0.03502	5.03081
f_ramb	<-	f_riq	gam2	0.32351	0.04346	7.44351
f_ramb	<-	f_rses	gam3	0.22334	0.03533	6.32150
f_ramb	<-	f_fses	gam4	0.07698	0.03225	2.38702
f_famb	<-	f_rses	gam4	0.07698	0.03225	2.38702
f_famb	<-	f_fses	gam3	0.22334	0.03533	6.32150
f_famb	<-	f_fiq	gam2	0.32351	0.04346	7.44351
f_famb	<-	f_fpa	gam1	0.17618	0.03502	5.03081
f_ramb	<-	f_famb	beta	0.11074	0.04283	2.58539
f_famb	<-	f_ramb	beta	0.11074	0.04283	2.58539

Variance Parameters

Variance Type	Variable	Parameter	Estimate	Standard Error	t Value
Error	f_ramb	psi	0.22453	0.02973	7.55201
	f_famb	psi	0.22453	0.02973	7.55201
Exogenous	f_rpa		1.00000		
	f_riq		1.00000		
	f_rses		1.00000		
	f_fpa		1.00000		
	f_fiq		1.00000		
	f_fses		1.00000		
Error	rea	thetaea	0.30645	0.03721	8.23647
	fea	thetaea	0.30645	0.03721	8.23647
	roa	thetaoa	0.42304	0.03650	11.58877
	foa	thetaoa	0.42304	0.03650	11.58877
	rpa	errpa1	0.30781	0.07510	4.09880
	fpa	errpa2	0.28837	0.07366	3.91467
	riq	erriq1	0.26748	0.07295	3.66672
	fiq	erriq2	0.15653	0.06614	2.36682
	rses	errses1	0.11477	0.07265	1.57975
	fses	errses2	0.08832	0.07088	1.24608

Figure 17.30 *continued*

Covariances Among Exogenous Variables

Var1	Var2	Parameter	Estimate	Standard Error	t Value
f_rpa	f_riq	cov1	0.26494	0.05436	4.87395
f_rpa	f_rses	cov2	0.00185	0.04995	0.03696
f_riq	f_rses	cov3	0.31164	0.05039	6.18460
f_fpa	f_fiq	cov1	0.26494	0.05436	4.87395
f_fpa	f_fses	cov2	0.00185	0.04995	0.03696
f_fiq	f_fses	cov3	0.31164	0.05039	6.18460
f_rpa	f_fpa	cov4	0.15828	0.07846	2.01729
f_rpa	f_fiq	cov5	0.11895	0.05383	2.20978
f_rpa	f_fses	cov6	0.06924	0.04993	1.38664
f_riq	f_fpa	cov5	0.11895	0.05383	2.20978
f_riq	f_fiq	cov7	0.43180	0.07084	6.09540
f_riq	f_fses	cov8	0.25004	0.05039	4.96207
f_rses	f_fpa	cov6	0.06924	0.04993	1.38664
f_rses	f_fiq	cov8	0.25004	0.05039	4.96207
f_rses	f_fses	cov9	0.30203	0.06360	4.74852

Covariances Among Errors

Error of	Error of	Parameter	Estimate	Standard Error	t Value
f_ramb	f_famb		0		
rea	fea	covea	0.02120	0.03094	0.68516
roa	foa	covoa	0.11197	0.03254	3.44068

The chi-square value for testing model 5 versus model 2 is $19.0745 - 19.0697 = 0.0048$ with 1 degree of freedom. This test statistic is insignificant. 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 psi12 to zero as in Loehlin's (1987) model 7:

```
title2 'Loehlin (1987) analysis: Model 7';
data model7(type=calismdl);
   set model2;
   if _name_='psi12'|_name_='beta' then
      do;
         _name_=' ';
         _estim_=0;
      end;
run;

proc tcalis corr data=aspire nobs=329 inmodel=model7;
run;
```

The fit summary is displayed in Figure 17.31, and the estimation results are displayed in Figure 17.32.

Figure 17.31 Career Aspiration Data: Fit Summary of Loehlin (1987) Model 7

Fit Summary

Modeling Info	N Observations	329
	N Variables	10
	N Moments	55
	N Parameters	25
	N Active Constraints	0
	Independence Model Chi-Square	872.0008
	Independence Model Chi-Square DF	45
Absolute Index	Fit Function	0.0773
	Chi-Square	25.3466
	Chi-Square DF	30
	Pr > Chi-Square	0.7080
	Z-Test of Wilson & Hilferty	-0.5487
	Hoelter Critical N	568
	Root Mean Square Residual (RMSR)	0.0363
	Standardized RMSR (SRMSR)	0.0363
	Goodness of Fit Index (GFI)	0.9846
Parsimony Index	Adjusted GFI (AGFI)	0.9718
	Parsimonious GFI	0.6564
	RMSEA Estimate	0.0000
	RMSEA Lower 90% Confidence Limit	.
	RMSEA Upper 90% Confidence Limit	0.0326
	Probability of Close Fit	0.9975
	ECVI Estimate	0.2350
	ECVI Lower 90% Confidence Limit	.
	ECVI Upper 90% Confidence Limit	0.2815
	Akaike Information Criterion	-34.6534
	Bozdogan CAIC	-178.5351
	Schwarz Bayesian Criterion	-148.5351
	McDonald Centrality	1.0071
Incremental Index	Bentler Comparative Fit Index	1.0000
	Bentler-Bonett NFI	0.9709
	Bentler-Bonett Non-normed Index	1.0084
	Bollen Normed Index Rho1	0.9564
	Bollen Non-normed Index Delta2	1.0055
	James et al. Parsimonious NFI	0.6473

Figure 17.32 Career Aspiration Data: Estimation Results of Loehlin (1987) Model 7

PATH List

---------Path---------			Parameter	Estimate	Standard Error	t Value
rea	<-	f_ramb	lambda	1.10371	0.06784	16.27015
roa	<-	f_ramb		1.00000		
foa	<-	f_famb		1.00000		
fea	<-	f_famb	lambda	1.10371	0.06784	16.27015
rpa	<-	f_rpa		0.83700		
riq	<-	f_riq		0.89400		
rses	<-	f_rses		0.94900		
fses	<-	f_fses		0.94900		
fiq	<-	f_fiq		0.89400		
fpa	<-	f_fpa		0.83700		
f_ramb	<-	f_rpa	gam1	0.17653	0.03604	4.89810
f_ramb	<-	f_riq	gam2	0.35727	0.04609	7.75204
f_ramb	<-	f_rses	gam3	0.24187	0.03628	6.66710
f_ramb	<-	f_fses	gam4	0.11087	0.03056	3.62795
f_famb	<-	f_rses	gam4	0.11087	0.03056	3.62795
f_famb	<-	f_fses	gam3	0.24187	0.03628	6.66710
f_famb	<-	f_fiq	gam2	0.35727	0.04609	7.75204
f_famb	<-	f_fpa	gam1	0.17653	0.03604	4.89810
f_ramb	<-	f_famb		0		
f_famb	<-	f_ramb		0		

Variance Parameters

Variance Type	Variable	Parameter	Estimate	Standard Error	t Value
Error	f_ramb	psi	0.21011	0.02940	7.14704
	f_famb	psi	0.21011	0.02940	7.14704
Exogenous	f_rpa		1.00000		
	f_riq		1.00000		
	f_rses		1.00000		
	f_fpa		1.00000		
	f_fiq		1.00000		
	f_fses		1.00000		
Error	rea	thetaea	0.31633	0.03648	8.67106
	fea	thetaea	0.31633	0.03648	8.67106
	roa	thetaoa	0.42656	0.03610	11.81508
	foa	thetaoa	0.42656	0.03610	11.81508
	rpa	errpa1	0.31329	0.07538	4.15589
	fpa	errpa2	0.29286	0.07389	3.96366
	riq	erriq1	0.30776	0.07307	4.21157
	fiq	erriq2	0.19193	0.06613	2.90250
	rses	errses1	0.14303	0.07313	1.95574
	fses	errses2	0.11804	0.07147	1.65171

Figure 17.32 *continued*

Covariances Among Exogenous Variables

Var1	Var2	Parameter	Estimate	Standard Error	t Value
f_rpa	f_riq	cov1	0.27533	0.05552	4.95900
f_rpa	f_rses	cov2	0.00611	0.05085	0.12020
f_riq	f_rses	cov3	0.33510	0.05150	6.50648
f_fpa	f_fiq	cov1	0.27533	0.05552	4.95900
f_fpa	f_fses	cov2	0.00611	0.05085	0.12020
f_fiq	f_fses	cov3	0.33510	0.05150	6.50648
f_rpa	f_fpa	cov4	0.17099	0.07872	2.17210
f_rpa	f_fiq	cov5	0.13859	0.05431	2.55174
f_rpa	f_fses	cov6	0.07563	0.05077	1.48956
f_riq	f_fpa	cov5	0.13859	0.05431	2.55174
f_riq	f_fiq	cov7	0.48105	0.06993	6.87858
f_riq	f_fses	cov8	0.27235	0.05157	5.28154
f_rses	f_fpa	cov6	0.07563	0.05077	1.48956
f_rses	f_fiq	cov8	0.27235	0.05157	5.28154
f_rses	f_fses	cov9	0.32046	0.06517	4.91719

Covariances Among Errors

Error of	Error of	Parameter	Estimate	Standard Error	t Value
f_ramb	f_famb		0		
rea	fea	covea	0.04535	0.02918	1.55444
roa	foa	covoa	0.12085	0.03214	3.75976

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 (1988) models. Constraining covea and covoa to zero produces Loehlin's (1987) model 6:

```
title2 'Loehlin (1987) analysis: Model 6';
data model6(type=calismdl);
   set model2;
   if _name_='covea'|_name_='covoa' then
      do;
         _name_=' ';
         _estim_=0;
      end;
run;

proc tcalis corr data=aspire nobs=329 inmodel=model6;
run;
```

The fit summary is displayed in Figure 17.33.

Figure 17.33 Career Aspiration Data: Loehlin (1987) Model 6

Fit Summary

Modeling Info	N Observations	329
	N Variables	10
	N Moments	55
	N Parameters	25
	N Active Constraints	0
	Independence Model Chi-Square	872.0008
	Independence Model Chi-Square DF	45
Absolute Index	Fit Function	0.1020
	Chi-Square	33.4475
	Chi-Square DF	30
	Pr > Chi-Square	0.3035
	Z-Test of Wilson & Hilferty	0.5151
	Hoelter Critical N	431
	Root Mean Square Residual (RMSR)	0.0306
	Standardized RMSR (SRMSR)	0.0306
	Goodness of Fit Index (GFI)	0.9802
Parsimony Index	Adjusted GFI (AGFI)	0.9638
	Parsimonious GFI	0.6535
	RMSEA Estimate	0.0187
	RMSEA Lower 90% Confidence Limit	.
	RMSEA Upper 90% Confidence Limit	0.0471
	Probability of Close Fit	0.9686
	ECVI Estimate	0.2597
	ECVI Lower 90% Confidence Limit	.
	ECVI Upper 90% Confidence Limit	0.3164
	Akaike Information Criterion	-26.5525
	Bozdogan CAIC	-170.4342
	Schwarz Bayesian Criterion	-140.4342
	McDonald Centrality	0.9948
Incremental Index	Bentler Comparative Fit Index	0.9958
	Bentler-Bonett NFI	0.9616
	Bentler-Bonett Non-normed Index	0.9937
	Bollen Normed Index Rho1	0.9425
	Bollen Non-normed Index Delta2	0.9959
	James et al. Parsimonious NFI	0.6411

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 (1987) seven models.

Model	χ^2	df	p-value	SBC
1. Full model	12.0132	13	0.5266	−63.3356
2. Equality constraints	19.0697	28	0.8960	−143.2200
3. No SES path	23.0365	29	0.7749	−145.0492
4. No reciprocal influence	20.9981	29	0.8592	−147.0876
5. No disturbance correlation	19.0745	29	0.9194	−149.0111
6. No error correlation	33.4475	30	0.3035	−140.4342
7. Constraints from both 4 and 5	25.3466	30	0.7080	−148.5351

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

```
data _null_;
   array achisq[7] _temporary_
      (12.0132 19.0697 23.0365 20.9981 19.0745 33.4475 25.3466);
   array adf[7] _temporary_
      (13 28 29 29 29 30 30);
   retain indent 16;
   file print;
   input ho ha @@;
   chisq = achisq[ho] - achisq[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 table in Figure 17.34.

Figure 17.34 Career Aspiration Data: Model Comparisons

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

      model comparison   chi**2   df  p-value
      ---------------------------------------
         2  versus 1     7.0565   15   0.9561
         3  versus 2     3.9668    1   0.0464
         4  versus 2     1.9284    1   0.1649
         5  versus 2     0.0048    1   0.9448
         7  versus 2     6.2769    2   0.0433
         7  versus 4     4.3485    1   0.0370
         7  versus 5     6.2721    1   0.0123
         6  versus 2    14.3778    2   0.0008
```

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

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

Bentler, P. M. and Bonett, D. G. (1980), "Significance Tests and Goodness of Fit in the Analysis of Covariance Structures," *Psychological Bulletin*, 88, 588–606.

Bollen, K. A. (1986), "Sample Size and Bentler and Bonett's Nonnormed Fit Index," *Psychometrika*, 51, 375–377.

Bollen, K. A. (1989), *Structural Equations with Latent Variables*, New York: John Wiley & Sons.

Browne, M. W. and Cudeck, R. (1993), "Alternative Ways of Assessing Model Fit," in *Testing Structural Equation Models*, ed. K. A. Bollen and S. Long, Newbury Park, CA: Sage Publications.

Duncan, O. D., Haller, A. O., and Portes, A. (1968), "Peer Influences on Aspirations: A Reinterpretation," *American Journal of Sociology*, 74, 119–137.

Fuller, W. A. (1987), *Measurement Error Models*, New York: John Wiley & Sons.

Haller, A. O., and Butterworth, C. E. (1960), "Peer Influences on Levels of Occupational and Educational Aspiration," *Social Forces*, 38, 289–295.

Hampel F. R., Ronchetti E. M., Rousseeuw P. J., and Stahel W. A. (1986), *Robust Statistics*, New York: John Wiley & Sons.

Huber, P. J. (1981), *Robust Statistics*, New York: John Wiley & Sons.

James, L. R., Mulaik, S. A., and Brett, J. M. (1982), *Causal Analysis*, Beverly Hills, CA: Sage Publications.

Jöreskog, K. G. (1973), "A General Method for Estimating a Linear Structural Equation System," in *Structural Equation Models in the Social Sciences*, ed. A. S. Goldberger and O. D. Duncan, New York: Seminar Press.

Jöreskog, K. G. and Sörbom, D. (1979), *Advances in Factor Analysis and Structural Equation Models*, Cambridge, MA: Abt Books.

Jöreskog, K. G. and Sörbom, D. (1988), *LISREL 7: A Guide to the Program and Applications*, Chicago: SPSS.

Keesling, J. W. (1972), "Maximum Likelihood Approaches to Causal Analysis," Ph.D. dissertation, University of Chicago, 1972.

Loehlin, J. C. (1987), *Latent Variable Models*, Hillsdale, NJ: Lawrence Erlbaum Associates.

Lord, F. M. (1957), "A Significance Test for the Hypothesis That Two Variables Measure the Same Trait Except for Errors of Measurement," *Psychometrika*, 22, 207–220.

McArdle, J. J. and McDonald, R. P. (1984), "Some Algebraic Properties of the Reticular Action Model," *British Journal of Mathematical and Statistical Psychology*, 37, 234–251.

McDonald, R. P. (1989), "An Index of Goodness-of-Fit Based on Noncentrality," *Journal of Classification*, 6, 97–103.

Schwarz, G. (1978), "Estimating the Dimension of a Model," *Annals of Statistics*, 6, 461–464.

Steiger, J. H. and Lind, J. C. (1980), "Statistically Based Tests for the Number of Common Factors," paper presented at the annual meeting of the Psychometric Society, Iowa City, IA.

Voss, R. E. (1969), "Response by Corn to NPK Fertilization on Marshall and Monona Soils as Influenced by Management and Meteorological Factors," Ph.D. dissertation, Iowa State University.

Wiley, D. E. (1973), "The Identification Problem for Structural Equation Models with Unmeasured Variables," in *Structural Equation Models in the Social Sciences*, ed. A. S. Goldberger and O. D. Duncan, New York: Seminar Press, 69–83.

Your Turn

We welcome your feedback.

- If you have comments about this book, please send them to **`yourturn@sas.com`**. Include the full title and page numbers (if applicable).
- If you have comments about the software, please send them to **`suggest@sas.com`**.