Creating Path Diagrams That Impress: A New Graphical Capability of the CALIS Procedure

Yiu-Fai Yung, SAS Institute Inc.

ABSTRACT

In structural equation modeling, researchers often use path diagrams to represent their models graphically. Path diagrams enable you to visualize the conceptual models behind the research and to depict statistical results in an intuitive way. In SAS/STAT® 13.1, the CALIS procedure produces high-quality graphical output of path diagrams from model specifications. You can simply use the PLOTS=PATHDIAGRAM option to request path diagrams for the unstandardized model. You can also customize path diagrams by using the options in the PATHDIAGRAM statement. This paper illustrates how you can create and use path diagrams to present your model and statistical results. Data examples demonstrate some important PATHDIAGRAM statement options that you can use to produce path diagrams that have different emphases, styles, or formats. The paper also demonstrates the use of the ODS Graphics Editor to edit path diagrams that are produced from the CALIS procedure. By taking advantage of these new functionalities, you can start to create your own path diagrams to show your research results in an even more impressive way!

INTRODUCTION

The CALIS procedure fits structural equation models that have a wide range of applications in education, psychology, sociology, health science, marketing, and other fields. In a structural equation model, you hypothesize functional or causal relationships among observed and latent variables. For an introduction to structural equation modeling, see Loehlin (2004), Bollen (1989), Everitt (1984), or Long (1983); for more information about manifest variables with measurement errors, see Fuller (1987).

The CALIS procedure provides several types of syntax input for you to specify complex relationships among variables of structural equation models. For example, the most popular syntax that PROC CALIS supports includes the FACTOR, LINEQS, and PATH statements. The procedure interprets these statements and then estimates the model parameters by maximum likelihood or other estimation methods.

In SAS/STAT 13.1, the CALIS procedure adds new functionality that produces path diagrams from input models. Path diagrams are graphical representations of theoretical models. They enable you to visualize the complex relationships among variables. After model parameters are estimated, you can also use path diagrams to graphically show the statistical results. PROC CALIS can produce path diagrams for the initial, standardized, and unstandardized models and provides a comprehensive set of options for customization.

This paper introduces this new path diagram functionality in the CALIS procedure. The following section shows how to use PROC CALIS to produce basic path diagrams. The section after that illustrates how to use the main options available for customizing path diagrams. Data examples demonstrate the features and options. The final section summarizes and demonstrates the capabilities of the ODS Graphics Editor for editing path diagrams. The paper concludes with a summary.

CREATING PATH DIAGRAMS WITH THE PLOTS=PATHDIAGRAM OPTION

To show how path diagrams work, this section starts with regression models, which are special cases of structural equation models, and then proceeds to more general models.

Example 1: A Multiple Regression Model of Sales Data

Consider the sales (in millions) by a company at various locations in the four quarters of its fiscal year, \(q_1\), \(q_2\), \(q_3\), and \(q_4\). The following DATA step reads a fictitious data set of the sales at the various locations of the company in the four quarters:
data sales;
  input q1 q2 q3 q4;
datalines;
  1.03  1.54  1.11  2.22
  1.23  1.43  1.65  2.12
  ... more lines ...
  1.34  1.74  2.16  3.28
;  
The following regression model predicts the fourth-quarter sales, $q_4$, by previous sales, $q_1$, $q_2$, and $q_3$:

$$q_4 = a + b_1 q_1 + b_2 q_2 + b_3 q_3 + e$$

You can use the REG procedure to fit this model to the sales data:

```
proc reg data=sales;
  model q4 = q1 q2 q3;
run;
```

Figure 1 shows the estimates that PROC REG produces. None of these effects are statistically significant at $\alpha = 0.05$.

| Parameter Estimates | Parameter | Standard Error | t Value | Pr > |t| |
|---------------------|-----------|----------------|---------|-------|---|
| Variable            | DF        | Estimate       |         |       |   |
| Intercept           | 1         | -0.27794       | 2.23296 | -0.12 | 0.9034 |
| q1                  | 1         | 0.55980        | 0.74041 | 0.76  | 0.4670 |
| q2                  | 1         | 0.58946        | 0.96411 | 0.61  | 0.5546 |
| q3                  | 1         | 0.88290        | 0.58873 | 1.50  | 0.1646 |

Because the class of structural equation models subsumes the multiple regression model as a special case, you can also use the CALIS procedure to fit the current regression model. To understand the CALIS syntax for model specification, it would be useful to start with a path diagram. Figure 2 shows the path diagram for the multiple regression model of the sales data.
In path diagrams, nodes represent variables, and links (directed paths) represent functional relationships among variables. These path diagram representations are so intuitive that you almost do not need to explain what the model is about—the hypothesized functional or causal relationships among variables are clearly depicted in path diagrams. For example, Figure 2 shows a multiple regression model that has three directional paths from \( q_1 \), \( q_2 \), and \( q_3 \), respectively, to \( q_4 \). Parallel to the use of path diagrams for representing models, the CALIS procedure supports a pathlike syntax for specifying models. For example, the following statements specify the current regression model:

```sas
ods graphics on;
proc calis data=sales plots=pathdiagram;
    path
        q4 <= q1 q2 q3;
run;
```

The PATH statement specification appears just like a syntactic transcription of the path diagram: you specify three paths from \( q_1 \), \( q_2 \), and \( q_3 \), respectively, to \( q_4 \). In addition, the ODS GRAPHICS ON statement enables the graphical output, and the PLOTS=PATHDIAGRAM option requests a path diagram for the (default) unstandardized solution.

Figure 3 shows the output path diagram with parameter estimates. If an estimate is significant at \( \alpha = 0.05 \), PROC CALIS flags it with an asterisk (*). If an estimate is significant at \( \alpha = 0.01 \), PROC CALIS flags it with two asterisks (**). Hence, all effects, which are indicated on the three directed paths, from previous sales to \( q_4 \), are not significant. These effect estimates agree with those that you obtain from the results of PROC REG in Figure 1. Because none of the path effects are statistically significant, you can conclude that the theoretical model is not a good one for the data.

The path diagram in Figure 3 also shows variance estimates on the double-headed paths (or bidirectional links). All these estimates are significant at \( \alpha = 0.05 \). Depending on whether a variable is predicted by at least one other variable in the model, the associated variance estimate can be interpreted as either an error variance or the variance of the variable. However, interpretations of these estimates are not critical at this point. The section “What Are Those Bidirectional Paths in Path Diagrams?” on page 6 explains different types of variances and their interpretations.

Example 2: A Sequential Prediction Model of Sales Data

Besides the multiple regression model, you can hypothesize other models that explain the sales data. These alternative models are best described by using path diagrams. For example, Figure 4 shows a path diagram that hypothesizes sequential functional relationships between adjacent sales in the four quarters.

**Figure 4** A Sequential Prediction Model of Sales Data

In contrast with the preceding regression model, in this diagram \( q_1 \) and \( q_2 \) do not predict the fourth-quarter sales directly—that is, there are no directed paths from the sales in the first two quarters to \( q_4 \). The influences of \( q_1 \) and \( q_2 \) on \( q_4 \) occur only through their indirect and direct effects, respectively, on \( q_3 \). How do you fit the model that this path diagram depicts?

---

1. ODS Graphics remains enabled for the rest of the paper.
You can simply “write out” the three paths in the path diagram when you use the PATH statement to specify the model, as shown in the following specification:

```
proc calis data=sales plots=pathdiagram;
  path
    q4 <= q3,
    q3 <= q2,
    q2 <= q1;
run;
```

Essentially, each directed path in the path diagram in **Figure 4** corresponds to a path entry in the PATH statement. Again, the PLOTS=PATHDIAGRAM option requests a path diagram of the estimated model, which is shown in **Figure 5**.

![Path Diagram](image)

**Figure 5** Path Diagram of the Sequential Prediction Model of Sales Data

Only one of the three path estimates is statistically significant at \( \alpha = 0.05 \), meaning that the effect of \( q1 \) on \( q2 \) and the effect of \( q2 \) on \( q3 \) might be null in the population. Only the effect of \( q3 \) on \( q4 \) is supported by statistical evidence. Again, interpretations of the variance estimates, which are displayed with the double-headed paths, are not made here.

**Example 3: A Confirmatory Factor Model of Test Scores**

Path diagram output is not limited to models that you specify using the PATH statement. PROC CALIS can produce path diagrams for many different types of models, including those that are specified by the FACTOR, LINEQS, LISMOD, and RAM statements. This example shows a path diagram of a confirmatory factor model that is specified by the FACTOR statement. Unlike the preceding two examples, which involve only observed variables, this example also includes latent factors in the model.

The following DATA step reads a fictitious data set of test scores:

```
data scores;
  input x1 x2 x3 y1 y2 y3;
datalines;
 23 17 16 15 14 16
 29 26 23 22 18 19
  ... more lines ...
 29 21 22 13 17 12
;Tests x1–x3 assess individuals’ verbal skills, and tests y1–y3 assess individuals’ math skills. Because of the unambiguous nature of the tests, you hypothesize a confirmatory factor model that contains two factors: one is the verbal ability factor and the other is the math ability factor. You use the following statements to specify the model:
proc calis data=scores plots=pathdiagram;
  factor
    verbal ===> x1-x3,
    math  ===> y1-y3;
  pvar
    verbal = 1.,
    math  = 1.;
run;

The FACTOR statement specifies the hypothesized relationships between the two factors (verbal and math) and the observed variables. The PVAR statement fixes the variances of the factors at 1. You must fix the variances of the factors in order to identify the scales of the factors. Otherwise, the model would be unidentified.

*Figure 6* Path Diagram of a Confirmatory Factor Model of Test Scores

The PLOTS=PATHDIAGRAM option in the PROC CALIS statement creates the path diagram in *Figure 6* for the (default) unstandardized solution. In path diagrams, ovals or circles represent latent variables, and rectangles or squares represent manifest variables.

The path diagram in *Figure 6* includes a fit summary table in the lower right corner. This table shows the most important fit statistics of the model. The chi-square model fit statistic ($\chi^2 = 9.81$, $df = 8$, $p = 0.28$) is not significant, indicating that the hypothesized model cannot be rejected at $\alpha = 0.01$. Other fit statistics, such as the CFI (comparative fit index), SRMR (standardized root mean square residual), RMSEA (root mean square error of approximation), and “Pr Close Fit” (probability for test of close fit), all indicate adequate or good model fit. Only the AGFI (adjusted goodness-of-fit index) shows a less-than-adequate model fit.

In addition to the fit statistics, it is important to examine the significance of the factor loading estimates, which are the path effects of the latent factors on the observed variables. These effects are shown as labels that are attached to the directed paths in the diagram. All of them are significant at $\alpha = 0.01$. Therefore, the hypothesized factor-variable relationships are supported by strong statistical evidence.
What Are Those Bidirectional Paths in Path Diagrams?

In structural equation modeling, variances and covariances are often treated as model parameters. In path diagrams, these parameters are represented by bidirectional paths (or double-headed arrows). Figure 6 shows several examples of these bidirectional paths.

In order to explain these parameters further, first it would be useful to distinguish between endogenous and exogenous variables. Endogenous variables are predicted by at least one other variable in the model. In path diagrams, each endogenous variable has at least one directed path (single-headed arrow) that points to it. For example, the diagram in Figure 6 shows that $x_1 \rightarrow x_3$ and $y_1 \rightarrow y_3$ are all endogenous variables. Simply put, exogenous variables are not endogenous, meaning that they are not predicted by any other variables in the model. In path diagrams, exogenous variables do not have any directed paths that point to them. For example, the diagram in Figure 6 shows that the two latent factors, verbal and math, are exogenous variables.

Covariance parameters apply only to pairs of exogenous variables in the structural equation model. For example, Figure 6 shows that the two latent factors, verbal and math, have a covariance of 0.52, which is significant at $\alpha = 0.01$. Because the variances of these latent factors are fixed at 1, the covariance estimate here is also the correlation between the two factors. It shows a strong correlation between verbal and math.

Other kinds of bidirectional paths in Figure 6 are attached to individual variables. These bidirectional paths (or self-links) represent variances of the variables. For exogenous variables, numerical values for these paths are the total variances, or simply the variances of the variables. For example, the fixed unity variances were set for the two latent factors, verbal and math. For endogenous variables, numerical values for these bidirectional paths are error variances. For example, PROC CALIS estimates and displays six error variances for all the observed variables in Figure 6.

All error variances, except the one for $y_1$, are significant at either the 0.01 or 0.05 $\alpha$-level. This means that these variances are significantly different from zero—a fact that is somewhat noninformative with respect to the validity of the structural equation model. In most practical applications, only the significance of the directed effects (path coefficients) is critical to the validation of the theoretical model.

Although significant error variances might not lend additional support to the theoretical model, too many insignificant error variances without proper justifications might cast doubt on the validity of your model. An insignificant error variance estimate might be due to a nearly perfect prediction of the endogenous variable in the model or simply due to sampling fluctuations, even for a reasonably good model in the population. Even so, you expect this to happen occasionally in one or two error variance estimates only. However, if your estimation results contain a lot of insignificant error variance estimates, it is likely that the model is not appropriate for the data.

The Utilities of Path Diagrams

The preceding examples show the following utilities of path diagrams:

- You can depict the theoretical relationships among variables intuitively.
- You can compare and contrast alternative or competing models visually.
- You can show your theoretical model and statistical results succinctly in a single display.

All you need to do is to use the PLOTS=PATHDIAGRAM option.

CUSTOMIZING PATH DIAGRAMS WITH THE PATHDIAGRAM STATEMENT OPTIONS

Using the PLOTS=PATHDIAGRAM option in the PROC CALIS statement is a good way to create simple default path diagrams. But sometimes you might want to customize your path diagrams for special purposes or to meet particular style requirements. The PATHDIAGRAM statement offers a comprehensive set of options for customization. Moreover, you can use multiple PATHDIAGRAM statements to customize several path diagrams differently for the same model. This section illustrates the options listed in Table 1.
Table 1  Some Important PATHDIAGRAM Statement Options

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIAGRAM=</td>
<td>Requests path diagrams for the initial, unstandardized, or standardized model</td>
</tr>
<tr>
<td>EMPHSTRUCT</td>
<td>Emphasizes the structural components of models</td>
</tr>
<tr>
<td>FITINDEX</td>
<td>Selects the fit statistics to display in fit summary tables</td>
</tr>
<tr>
<td>LABEL=</td>
<td>Specifies labels for variables to be used in path diagrams</td>
</tr>
<tr>
<td>NOTITLE</td>
<td>Suppresses the display of default titles</td>
</tr>
<tr>
<td>NOVARIANCE</td>
<td>Suppresses the display of the variance estimates</td>
</tr>
<tr>
<td>STRUCTURAL</td>
<td>Requests path diagrams for the structural components of models</td>
</tr>
<tr>
<td>TITLE=</td>
<td>Specifies titles to be displayed in path diagrams</td>
</tr>
</tbody>
</table>

To illustrate these options, this section uses two examples that are based on the data from Wheaton et al. (1977). The following DATA step inputs the lower triangle of the covariance matrix of six observed variables:

```sas
data Wheaton(TYPE=COV);
  _type_ = 'cov';
  input _name_ $ 1-11 Anomie67 Powerless67 Anomie71 Powerless71 Education SEI;
  datalines;
  Anomie67 11.834 . . . . .
  Anomie71 6.819 5.091 12.532 . . .
;```

In this data set, the variables Anomie67 and Powerless67 are measures of feelings of social alienation in 1967, which is a latent construct denoted as Alien67 in the model. Similarly, the variables Anomie71 and Powerless71 are measures of alienation in 1971, which is a latent construct denoted as Alien71 in the model. The variables Education and SEI (socioeconomic index) are measures of socioeconomic status in 1967, which is a latent construct denoted as SES in the model.

Example 4: Customizing Path Diagrams to Show the Main Modeling Results

The following statements specify a model of the data:

```sas
proc calis nobs=932 data=Wheaton;
  path
    Anomie67 Powerless67 <=== Alien67 = 1.0 0.833,
    Anomie71 Powerless71 <=== Alien71 = 1.0 0.833,
    Education SEI <=== SES = 1.0 ,
    Alien67 Alien71 <=== SES ,
    Alien71 <=== Alien67 ;
  pathdiagram;
  pathdiagram
    title="Stability of Alienation" novariance
    label=[Powerless71="Powerless 1971" Powerless67="Powerless 1967"
        Alien71="Alienation 1971" Alien67="Alienation 1967"
        SES="SocioEco Status"]
    fitindex=[nobs chisq df probchi srmr rmsea cfi aic];
run;
```

The PATH statement specifies the relationships among the observed variables and the latent constructs. The values in the PATH statement are fixed values for the corresponding path effects. Path effects to which you do not assign numerical values are free parameters in the model.

Two PATHDIAGRAM statements request that PROC CALIS create two separate path diagrams. The first PATHDIAGRAM statement does not specify any options. Effectively, it requests the path diagram for the default unstandardized solution. Figure 7 shows this default diagram.
The second PATHDIAGRAM statement specifies several options to customize the path diagram for the unstandardized solution. Figure 8 shows this customized diagram.

**Figure 8** Customized Path Diagram of the Unstandardized Solution
As compared with the default diagram, the customized diagram in Figure 8 has several distinctive features that result from specifying options as follows:

- **The TITLE= option** provides a more meaningful title that reflects the research problem, instead of the default title “Unstandardized Solution,” which is shown in the default diagram in Figure 7.

- **The NOVARIANCE option** makes the customized diagram in Figure 8 look “cleaner” by suppressing the display of all variance estimates and the bidirectional paths. It also causes the path effect estimates in the directional paths to stand out. All path effect estimates are significant at the 0.01 \( \alpha \)-level. Because the variance estimates in this research are not as important as the path effect estimates—this is usually true of most structural equation models—using the NOVARIANCE option is not only justifiable but also beneficial to the display quality of the path diagram.

- **The LABEL= option** provides more meaningful variable labels in Figure 8 than the default version in Figure 7. The space in labels has also provided desirable potential locations for breaking the labels into separate lines when necessary. For example, Figure 8 shows “Powerless 1971” and “Anomie 1971” in two lines with desirable text breaks.

- **The FITINDEX= option** displays the specified set of fit indices or information in the fit summary table. For example, the customized diagram in Figure 8 displays the number of observations (the NOBS keyword) and Akaike’s information criterion (the AIC keyword), which are not displayed in the default version in Figure 7. In addition, the customized diagram has excluded some default fit indices in the fit summary table. For example, AGFI, which is displayed in the default path diagram shown in Figure 7, is suppressed in Figure 8.

**Example 5: Creating Path Diagrams to Show Your Main Modeling Ideas**

Structural equation modelers distinguish between the structural component and the measurement component in models. Loosely speaking, the **measurement component** (or measurement model) contains functional relationships between latent factors and observed (or manifest) variables. It involves all latent factors and observed variables in the model. The role of the measurement component is to define the latent factors by their relationships with the observed variables.

In contrast, the **structural component** (or structural model) contains only the latent factors and their functional relationships. The role of the structural component is to specify the functional relationships of the latent factors so that hypotheses about a problem domain can be tested statistically.

If the research focuses on validating specific structural relationships rather than measuring latent factors, then path diagrams that highlight or show only the structural components would be more preferable. You can produce this kind of path diagram by using some of the options in the PATHDIAGRAM statement.

For example, using the same data set and the same model as in Example 4, the following PATHDIAGRAM statements produce two separate path diagrams that highlight the structural component in the initial model specifications:

```plaintext
pathdiagram
  diagram=initial notitle structural(only)
     SES="SocioEco Status");
pathdiagram
  diagram=initial notitle emphstruct novariance
     SES="SocioEco Status");
```

Figure 9 shows the customized diagram that the first PATHDIAGRAM statement specifies. The DIAGRAM=INITIAL option requests the path diagram for the initial specification rather than the default unstandardized solution. The NOTITLE option turns off the display of the title. The STRUCTURAL(ONLY)

---

2In some research studies, the structural component might also include measurement-error-free observed variables that serve as predictors of other variables. But the current definition of structural component is good and simple enough for demonstration purposes.
option displays only the structural component of the model. The LABEL= option specifies more meaningful labels to display in the path diagram. Figure 9 thus shows the essence of structural relationships. That is, socioeconomic status is a predictor of alienation in both 1967 and 1971, and alienation in 1967 also predicts alienation in 1971. The measurement component is omitted altogether in order to highlight the theoretical relationships among the hypothetical constructs.

Figure 9 Structural Component Only

![Figure 9](image)

Figure 10 Emphasizing the Structural Component

![Figure 10](image)

Figure 10 shows the customized diagram that is specified by the second PATHDIAGRAM statement, which differs from the first PATHDIAGRAM specification by omitting the STRUCTURAL(ONLY) option and adding two new options: NOVARIANCE and EMPHSTRUCT. The NOVARIANCE option omits the bidirectional paths for variances so as to make the directional effects stand out.

Unlike the first PATHDIAGRAM statement, which shows only the structural component, the EMPHSTRUCT option in the second PATHDIAGRAM statement still displays the full model but emphasizes the structural component. As a result, the path diagram labels only the latent factors and uses much smaller rectangles to represent observed variables.

One advantage of using the EMPHSTRUCT option instead of the STRUCTURAL option to highlight the structural component is that the path diagram still indicates how many observed variables have served as indicators of the latent factors in the model, but at the same time it displays only the pertinent information about the structural component.

The STRUCTURAL(ONLY) and EMPHSTRUCT options are not limited to the initial specification. For example, you can use the following statement to produce a path diagram for the unstandardized solution that emphasizes the structural component:

```
pathdiagram
    notitle emphstruct=2 novariance
            SES="SocioEco Status"];
```

Figure 11 shows the resulting path diagram, which differs from the one in Figure 10 in that it displays the fit summary table and the estimates of the path effects in the structural component.

---

If you specify the STRUCTURAL option, PROC CALIS produces path diagrams for both the structural component and the full model.

---

3
Figure 11  Emphasizing the Structural Component of the Unstandardized Solution

Another difference is due to the EMPHSTRUCT=2 option. The size of the latent factors relative to the size of the observed variables in Figure 11 is not the same as that in the diagram in Figure 10. For path diagrams that you produce without using the EMPHSTRUCT option, the default size ratio of factors to variables is 1.5:1 (or approximately 3:2 in a single dimension). To highlight the latent factors, the EMPHSTRUCT option sets the ratio to 4:1 by default. This default ratio is used in the path diagram in Figure 10.

In the current example, the EMPHSTRUCT=2 option overwrites the default by setting the size ratio to 2:1. As a result, the factors in Figure 11 are still larger than the observed variables, although the factors-to-variables size ratio is actually smaller than the default ratio.

The ratio from the EMPHSTRUCT=2 option seems to have made the path diagram in Figure 11 more “balanced” and visually pleasing than the one in Figure 10. However, because the “best” EMPHSTRUCT= ratio depends on the number of observed variables that are related to each latent factor in the model, you might need to determine the optimal ratio by trial and error.

MORE OPTIONS FOR CUSTOMIZING PATH DIAGRAMS

This section continues to illustrate some important options in the PATHDIAGRAM statement. For a full list of options, see the CALIS procedure documentation in the *SAS/STAT User's Guide*.

**Suppressing the Display of Variance and Covariance Parameters**

To suppress the display of some sets of variance or covariance parameters, you can use the options in Table 2. These options improve the overall clarity of the output path diagrams by eliminating unnecessary displays. Because variances and covariances are often considered to be less important parameters than the path effects in models, removing them in path diagrams does not compromise the main model interpretations.

<table>
<thead>
<tr>
<th>Option</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOCOVARIANCE</td>
<td>All covariances, including error covariances</td>
</tr>
<tr>
<td>NOERRCOV</td>
<td>All error covariances</td>
</tr>
<tr>
<td>NOERRVAR</td>
<td>All error variances</td>
</tr>
<tr>
<td>NOEXOGCOV</td>
<td>All covariances among exogenous non-error variables</td>
</tr>
<tr>
<td>NOEXOGVAR</td>
<td>All variances of exogenous non-error variables</td>
</tr>
<tr>
<td>NOVARIANCE</td>
<td>All variances, including error variances</td>
</tr>
</tbody>
</table>
Formatting Parameter Estimates

Table 3 shows the options for formatting the display of parameter estimates. These options enable you to specify your own formatting requirements for the parameter estimates.

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DECP=</td>
<td>Specifies the decimal places of the estimates</td>
</tr>
<tr>
<td>NOESTIM</td>
<td>Disables the display of numerical estimates</td>
</tr>
<tr>
<td>NOFLAG</td>
<td>Disables the display of flags for statistical significance</td>
</tr>
<tr>
<td>PARMNAMES</td>
<td>Enables the display of parameter names or labels</td>
</tr>
</tbody>
</table>

For example, suppose that a path effect parameter called “beta” has an estimate of 0.7138, which is significant at $\alpha = 0.01$. Table 4 shows the labels on the path in the unstandardized path diagram that result from different combinations of options.

<table>
<thead>
<tr>
<th>Options</th>
<th>Label for Parameter Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>(default)</td>
<td>0.71**</td>
</tr>
<tr>
<td>PARMNAMES</td>
<td>beta=0.71**</td>
</tr>
<tr>
<td>PARMNAMES NOESTIM</td>
<td>beta**</td>
</tr>
<tr>
<td>NOFLAG DECP=3</td>
<td>0.714</td>
</tr>
</tbody>
</table>

Style Control

Table 5 shows several options for the display styles.

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEANPARM=</td>
<td>Specifies that mean parameters be displayed as either paths or labels</td>
</tr>
<tr>
<td>NOFITTABLE</td>
<td>Suppresses the display of fit statistics</td>
</tr>
<tr>
<td>USEERROR</td>
<td>Requests the use of explicit error variables</td>
</tr>
<tr>
<td>VARPARM=</td>
<td>Specifies that variance parameters be displayed as either paths or labels</td>
</tr>
</tbody>
</table>

You can use the MEANPARM= and VARPARM= options to specify how the mean and variance parameters, respectively, are represented in path diagrams. For example, MEANPARM=PATH represents the means as directed path effects from a unit constant that is represented by a triangle. MEANPARM=LABEL shows the means as labels that are affixed above the corresponding variables. Similarly, VARPARM=PATH represents variances as double-headed paths, and VARPARM=LABEL shows variances as labels.

The NOFITTABLE option omits the fit summary table from the path diagrams for the unstandardized or standardized solution. This is useful for models in which you do not need to see the fit statistics. For example, regression models that are always saturated in terms of mean and covariance structures have trivial fit statistics that are unnecessary to display.

Finally, the USEERROR option shows the error variables (by using small circles) explicitly in path diagrams. The default is not to use error variables. Some researchers might prefer to use explicit error variables, because error variances would then be associated with the error variables directly rather than with the corresponding endogenous variables, which is the default setting. However, because of the additional amount of display (error variables and their directional paths), showing explicit error variables would usually add clutter to path diagrams.
**Layout Algorithms**

The ARRANGE= option specifies the layout algorithm for path diagrams. A layout algorithm determines where the variables are placed in the path diagram. Needless to say, whether a path diagram looks good depends on which layout algorithm you choose. The CALIS procedure provides three layout algorithms for path diagrams: the process-flow algorithm (ARRANGE=FLOW), the grouped-flow algorithm (ARRANGE=GROUPEDFLOW), and the GRIP (Graph dRawing with Intelligent Placement) algorithm (ARRANGE=GRIP).

The process-flow algorithm arranges variables in hierarchical order. It works well for variables that can be ordered unambiguously in a hierarchy. For example, the process-flow algorithm would be appropriate for regression models such as those in Examples 1 and 2.

Better-known latent variable models that can use the process-flow algorithm are confirmatory factor models, higher-order factor models, and hierarchical factor models. For example, a second-order factor model is illustrated in Figure 12.

**Figure 12 A Second-Order Factor Model**

![Second-Order Factor Model Diagram](image)

The grouped-flow algorithm is much like the process-flow algorithm in its assumption of unambiguous hierarchical ordering. However, the grouped-flow algorithm focuses on the hierarchical ordering of factor clusters rather than of individual variables. Figure 13 shows a path diagram example that uses the grouped-flow algorithm. The **Factor1** cluster is at the first level. The **Factor2** and **Factor3** clusters are at the second level. The grouped-flow algorithm not only orders the factor clusters properly but also retains the distinctiveness of the three factor clusters.
Finally, the GRIP algorithm in PROC CALIS is a more general layout algorithm for path diagrams. When no unambiguous hierarchical ordering occurs in either variables or factor clusters in the model, the GRIP algorithm is usually the best choice. For example, the path diagrams in both Figure 7 and Figure 8 use the GRIP algorithm. When you use the GRIP algorithm, the spread of variables in these path diagrams is usually well balanced.

Because the choice of the layout algorithm is critical to the production of good-looking path diagrams, how should you determine which layout algorithm to use for a particular model? Fortunately, the default ARRANGE=AUTOMATIC option in the CALIS procedure solves this problem for you automatically. By examining the properties of interrelationships among variables in the model, PROC CALIS selects the “best” algorithm for the corresponding path diagram. For more information, see the CALIS procedure documentation in the SAS/STAT User's Guide.

You can still select a particular layout algorithm by specifying the ARRANGE= option, especially when you want to use the DESTROYERS= option to define destroyer paths. The discussion of destroyer paths is beyond the scope of this paper. For more information about the DESTROYERS= option and the ideas behind it, see the CALIS procedure documentation in the SAS/STAT User's Guide.

EDITING PATH DIAGRAMS WITH THE ODS GRAPHICS EDITOR

You can edit the path diagrams that are output from PROC CALIS by using the ODS Graphics Editor. The ODS Graphics Editor provides the following useful functionalities for editing path diagrams:

- Moves the nodes around without losing the associated links
- Aligns the selected nodes vertically or horizontally
- Straightens or bends links
- Reshapes or reroutes links
- Relocates the labels of links

Note that the terms “nodes” and “links” used by the ODS Graphics Editor are more general graph-theoretic terminology. In the current context of path diagrams, nodes refer to the two-dimensional shapes (rectangles,
ovals, or triangles) that represent variables, and links refer to the directional paths (single-headed arrows) that represent effects or the bidirectional paths (double-headed arrows) that represent covariances.

To use the ODS Graphics Editor, you must first enable ODS Graphics to create editable graphs. After the editable path diagrams are created, you would need to invoke the ODS Graphics Editor for editing. For step-by-step instructions, see Example 6. For general information about using the ODS Graphics Editor for statistical graphics, see the chapter “Statistical Graphics Using ODS” in the SAS/STAT User’s Guide. For specific information about editing path diagrams, see the chapter “Working with Nodes and Links” in the SAS 9.4 ODS Graphics Editor: User’s Guide.

Example 6: Displaying and Editing Path Diagrams with Means and Intercepts

This example shows how PROC CALIS displays means and intercepts in path diagrams. It also illustrates how you can use the ODS Graphics Editor to improve the path diagram output. The following statements specify the same model of the sales data as in Example 1, but with the addition of the MEANSTR option in the PROC CALIS statement:

```sas
proc calis data=sales meanstr;
  path
     q4 <=== q3,
     q3 <=== q2,
     q2 <=== q1;
  pathdiagram notitle nofittable;
run;
```

The MEANSTR option requests that PROC CALIS estimate the mean and the intercept parameters in the model, even though adding these parameters would not affect the original model fit. The NOTITLE and NOFITTABLE options in the PATHDIAGRAM statement suppress the printing of the title and the fit summary table, respectively.

Figure 14 shows the output path diagram. This path diagram is similar to the one in Figure 4, except that the current path diagram displays the means and intercepts in addition to the variances and path effects.

Figure 14 Displaying the Means and Intercepts as Labels in a Path Diagram

By default, PROC CALIS shows the means (or intercepts) and variances as labels. For example, the label (1.37**, 0.34*), which is located above variable q1, indicates that q1 has a mean of 1.37 (significant at \( \alpha = 0.01 \)) and a variance of 0.34 (significant at \( \alpha = 0.05 \)). Because all other variables in this model are endogenous, similar labels above these variables represent their intercepts and error variances, respectively.

Some researchers prefer to represent means and intercepts as path effects from the unit constant, which is conventionally represented by a triangle-shaped node in path diagrams. You can request such a representation scheme by using the MEAN=PATH option in the PATHDIAGRAM statement, as shown in the following specification:

```sas
ods listing sge=on;
proc calis data=sales meanstr;
  path data=sales meanstr;
    q4 <=== q3,
    q3 <=== q2,
    q2 <=== q1;
    pathdiagram notitle nofittable mean=path;
run;
```
In order to use the ODS Graphics Editor, you specify the SGE=ON option in the ODS LISTING statement before running PROC CALIS. Essentially, the SGE=ON option requests the creation of SGE files that contain graphical information about the path diagrams. The ODS Graphics Editor is then able to open these SGE files for editing. Figure 15 displays the resulting path diagram in the SAS windowing environment.

**Figure 15** Locating the Icon for the SGE File of the Path Diagram

The path diagram in Figure 15 now displays the means and intercepts just like path effects from the triangular node, which is labeled with “1,” representing the unit constant.

In the Results panel shown on the left side of Figure 15, two different icons are labeled as **Unstandardized** under the **Path Diagram** tree item. The first icon represents the (static) path diagram in the SAS Output panel. The second icon, which is circled in red, represents the SGE file for the same path diagram. If you double-click this icon, SAS invokes the ODS Graphics Editor to open this editable file. Figure 16 shows a pop-up window in which the ODS Graphics Editor displays the path diagram that is encoded in the SGE file.
To enable the editing of the objects in the path diagram, click the data icon (circled in red in Figure 16) in the editing toolbar. The following sections illustrate some features of the ODS Graphics Editor.

### Relocating a Node or Variable

You can select a node or variable by clicking the node. Figure 17 shows that the triangular node is now selected for moving.

**Figure 17** Selecting the Triangular Node

With the mouse pointer positioned over the triangular node, you can hold the mouse button down and drag the node to whatever place you want in the editing area. The node stays in the place where you release the mouse button. Figure 18 shows that you put the triangular node somewhere in the middle of the path diagram. The path diagram now looks more balanced. However, this move also causes some twisting in the links that connect the triangular node to other nodes.
Changing the Shapes of Links

The curved links from the triangular node seem distracting. It would be better to replace them with straight links. To do this for a single link, right-click the link. A property box for the link opens, as shown in Figure 19. It shows that the “Curved” property has been enabled for this link.

To disable the Curved property, move the mouse pointer over this property. This toggles off the Curved property, and the link is now straight, as shown in Figure 20.
Repeat the previous steps for other curved links that you want to straighten. Figure 21 show the final path diagram.

When you are satisfied with the editing, you can save the path diagram as a static image file (for example, a PNG file) by using the Save As option in the File menu. You can also save the path diagram information in an SGE file for future editing.

**SUMMARY**

This paper introduces the new path diagram features of the CALIS procedure available in SAS/STAT 13.1. Examples demonstrate how to output the path diagram and how to customize the path diagram to suit specific purposes. You can use the PLOTS=PATHDIAGRAM option to request the default path diagram for
the unstandardized solution. You can also use the PATHDIAGRAM statement to customize the output path diagrams. Each output path diagram can be customized individually in separate PATHDIAGRAM statements. PROC CALIS analyzes the interrelationships among the variables in your model and can automatically find the most suitable layout algorithm for producing path diagrams. Editing of path diagrams is also supported by the ODS Graphics Editor.

REFERENCES


ACKNOWLEDGMENTS

The author wishes to thank the entire ODS Graphics team at SAS for making the production of path diagrams possible. The contributions from the following individuals are deeply appreciated: Zachariah Cox, Pratik Phadke, and Xiao Le Xu for implementing the process-flow, grouped-flow, and GRIP algorithms at different stages of development; Jyoti Yakovenko for implementing the features for path diagrams editing; and Dan Heath, Prashant Hebbar, Lingxiao Li, and Brad Morris for general help in setting up the GTL and other graphics-related aspects. The author would also like to thank Anne Baxter, Ed Huddleston, Warren Kuhfeld, Jill Tao, and Catherine Truxillo for technical and editorial review of the paper. Last but not least, the author is grateful for the continuous support of the path diagram project by Sanjay Matange and Bob Rodriguez.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author:

Yiu-Fai Yung
SAS Institute Inc.
SAS Campus Drive
Cary, NC 27513
Yiu-Fai.Yung@sas.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.