

Latent-Variable, Structural Equation Modeling with PROC CALIS

Larry Hatcher, Winthrop University, Rock Hill, SC

ABSTRACT

This paper shows how to use the SAS® System's PROC CALIS to perform path analysis with latent variables. This is a type of structural equation modeling that is popularly associated with the LISREL program (Joreskog & Sorbom, 1989). This paper describes a two-phase procedure recommended by Anderson and Gerbing (1988) in which you (a) first use confirmatory factor analysis to develop an adequate measurement model, and (b) then follow up by performing a latent-variable path analysis. For each phase, it shows how to draw a program figure that identifies all of the parameters to be estimated, and how to convert this program figure into the required PROC CALIS program.

INTRODUCTION

This paper shows how to use the CALIS procedure to perform path analysis with latent variables. Specifically, it will show how to test the type of latent-variable path model that appears in Figure 1 (all figures are presented at the end of this paper). This figure illustrates Rusbult's (1980) investment model, a theory that identifies the determinants of commitment in romantic associations and friendships.

In this model, **commitment** refers to an individual's intention to remain in the relationship. The model in Figure 1 predicts that a person's commitment to a relationship is directly affected by his or her **satisfaction** with the relationship, the size of the **investments** that he or she has put into the relationship (i.e., the amount of time and energy put into it), and the attractiveness of his or her **alternatives** to the relationship (e.g., how attractive it would be to be with a different romantic partner).

Satisfaction, in turn, is said to be affected by the **rewards** associated with the relationship (i.e., the number of "good things" come from the relationship), and the **costs** associated with the relationship (i.e., the number of "bad things" or hardships that come with the relationship).

"V" VARIABLES AND "F" VARIABLES

Assume that you administer a 19-item questionnaire to a group of individuals who are engaged in romantic associations, and that the 19 items constituting the questionnaire are designed to measure the six investment model constructs described above.

Specifically, items 1-4 assess commitment, items 5-7 assess satisfaction, items 8-10 assess rewards, items 11-13 assess costs, items 14-16 assess investments, and items 17-19 assess alternatives.

This paper uses Bentler's (1989) conventions of giving manifest variables (observed variables) short variable names beginning with the letter "V," and of representing manifest variables as rectangles in figures. It also uses Bentler's convention of giving latent variables (unobserved variables) short variable names beginning with the letter "F" (for Factor), and of representing latent variables as ovals in figures. Given these conventions, you can see that the 19 rectangles in Figure 1 (variables V1-V19) correspond to the 19 questionnaire items, and that the six ovals in Figure 1 (factors F1-F6) correspond to the six latent factors.

THE 18 RULES OF STRUCTURAL EQUATION MODELING

To perform the analyses described here, you will (a) prepare program figures and (b) then convert these program figures into PROC CALIS programs. The following 18 rules will guide you in this process:

- Rule 1:** In general, only exogenous variables are allowed to have covariances.
- Rule 2:** A residual term must be identified for each endogenous variable in the model.
- Rule 3:** Exogenous variables do not have residual terms.
- Rule 4:** Variances should be estimated for every exogenous variable in the model, including residual terms.
- Rule 5:** In most cases, covariances should be estimated for every possible pair of manifest exogenous variables; covariances are not estimated for endogenous variables.
- Rule 6:** For simple recursive models, covariances should not be estimated for residual terms.
- Rule 7:** One equation should be created for each endogenous variable, with that variable's name to the left of the equals sign.
- Rule 8:** Variables that have a direct effect on that

endogenous variable are listed to the right of the equals sign.

- Rule 9:** Exogenous variables, including residual terms, are never listed to the left of the equals sign.
- Rule 10:** To estimate a path coefficient for a independent variable, a unique path coefficient name should be created for the path coefficient associated with that independent variable.
- Rule 11:** The last term in each equation should be the residual (disturbance) term for that endogenous variable; this E (or D) term will have no name for its path coefficient.
- Rule 12:** To *estimate* a parameter, create a name for that parameter.
- Rule 13:** To *fix* a parameter at a given numerical value, insert that value in the place of the parameter's name.
- Rule 14:** To *constrain* two or more parameters to be equal, use the same name for those parameters.
- Rule 15:** In confirmatory factor analysis, the variances of the latent F variables are usually fixed at 1.
- Rule 16:** In path analysis with latent variables, the variances of the exogenous F variables are free parameters to be estimated.
- Rule 17:** In path analysis with latent variables, one factor loading for each F variable should be fixed at 1.
- Rule 18:** In a confirmatory factor analysis of a nonstandard model, the variances of a manifest structural variable should be a free parameter to be estimated.

A TWO-PHASE PROCEDURE

To test the latent-variable path model of Figure 1, this paper recommends that you follow a two-phase procedure recommended by Anderson and Gerbing (1988). In Phase 1, you will use confirmatory factor analysis (CFA) to develop an acceptable measurement model. In Phase 2, you will use path analysis with latent variables to test the model in Figure 1.

DRAWING THE PROGRAM FIGURE FOR CONFIRMATORY FACTOR ANALYSIS

You will facilitate the process of performing a CFA if you first prepare a program figure that identifies all of the parameters to be estimated in the analysis. This process begins by drawing the (uncompleted) program figure presented here as Figure 2. You can see that it is identical to the path model presented in Figure 1, except that each F variable is now connected to every other F variable by a curved, double-headed arrow (this indicates that every F variable is allowed to freely covary with every other F variable in this CFA).

The completed program figure presented as Figure 3 identifies all of the parameters to be estimated in the CFA. The following steps show you how to prepare the various components of this program figure:

Step 1: Identify a residual term for each endogenous variable. An endogenous variable is a variable that has a straight, single-headed arrow pointing at it. All V variables in Figure 3 are endogenous variables. This paper follows Bentler's (1989) convention of using the letter "E" to represent residual terms for the V variables. You can see that an E term has been created for each V variable in Figure 3.

Step 2: Identifying variances to be estimated for the exogenous E variables. An exogenous variable is a variable that does not have a straight, single-headed arrow pointing at it. All E terms in Figure 3 are exogenous variables. Rule 4 states that variances should be estimated for all exogenous variables, and so the symbol "VAR?" appears below each E term in Figure 3.

Step 3: Fixing the variances of the F variables at 1. Rule 15 advises that you fix the variances of all F variables at 1 in a CFA. This is illustrated by placing the symbol "VAR=1" in each oval in Figure 3.

Step 4: Identifying the covariances to be estimated. The symbol "C?" appears on each curved arrow in Figure 3, signifying that PROC CALIS will estimate the covariance between every possible pair of F variables.

Step 5: Identifying the factor loadings to be estimated. In Figure 3, the symbol "L?" appears on each arrow going from an F variable to a V variable, identifying the factor loadings that are to be estimated in the analysis.

PREPARING THE PROC CALIS PROGRAM TO PERFORM CFA

Below is the entire PROC CALIS program (minus the data step) to estimate the model appearing in Figure 3 (line numbers have been added on the left):

```

1  PROC CALIS  COVARIANCE
2          CORR
3          RESIDUAL
4          MODIFICATION ;
5  LINEQS
6      V1 = LV1F1 F1 + E1,
7      V2 = LV2F1 F1 + E2,
8      V3 = LV3F1 F1 + E3,
9      V4 = LV4F1 F1 + E4,
10     V5 = LV5F2 F2 + E5,
11     V6 = LV6F2 F2 + E6,
12     V7 = LV7F2 F2 + E7,
13     V8 = LV8F3 F3 + E8,
14     V9 = LV9F3 F3 + E9,
15     V10 = LV10F3 F3 + E10,
16     V11 = LV11F4 F4 + E11,
17     V12 = LV12F4 F4 + E12,
18     V13 = LV13F4 F4 + E13,
19     V14 = LV14F5 F5 + E14,
20     V15 = LV15F5 F5 + E15,
21     V16 = LV16F5 F5 + E16,
22     V17 = LV17F6 F6 + E17,
23     V18 = LV18F6 F6 + E18,
24     V19 = LV19F6 F6 + E19;
25  STD
26     F1 = 1,
27     F2 = 1,
28     F3 = 1,
29     F4 = 1,
30     F5 = 1,
31     F6 = 1,
32     E1-E19 = VARE1-VARE19;
33  COV
34     F1 F2 = CF1F2,
35     F1 F3 = CF1F3,
36     F1 F4 = CF1F4,
37     F1 F5 = CF1F5,
38     F1 F6 = CF1F6,
39     F2 F3 = CF2F3,
40     F2 F4 = CF2F4,
41     F2 F5 = CF2F5,
42     F2 F6 = CF2F6,
43     F3 F4 = CF3F4,
44     F3 F5 = CF3F5,
45     F3 F6 = CF3F6,
46     F4 F5 = CF4F5,
47     F4 F6 = CF4F6,
48     F5 F6 = CF5F6;
49  VAR V1-V19 ;
50  RUN;
```

See Hatcher (1994) for a detailed discussion of each statement constituting the preceding program. Some of the more important points regarding this program are summarized below.

The PROC CALIS statement appears on lines 1-4 of the preceding program. This statement calls up the CALIS procedure, and requests a number of options (e.g., the COVARIANCE option requests that the analysis be performed on the covariance matrix, rather than the correlation matrix).

The LINEQS statement appears on lines 5-24. This statement requests that the factor loadings be estimated in the CFA. For example, line 6 requests that CALIS estimate the factor loading for the manifest variable V1. The fact that the name "V1" appears to the left of the equals sign, and the name "F1" appears to the right of the equals sign means that CALIS will estimate the factor loading for the manifest variable V1 on the latent factor F1. The parameter name "LV1F1" appears in the equation, meaning that this loading will be given the name "LV1F1".

The STD statement appears on lines 25-32, and this statement requests that variances either be estimated or fixed. Lines 26-31 fix the variances of all of the F variables at 1. Line 32 requests that variances be estimated for E1 through E19. It requests that these variances be given the variance estimate names VARE1 through VARE19.

The COV statement appears on lines 33-48, and it requests that covariances be estimated for every possible combination of F variables. For example, line 34 requests that CALIS estimate the covariance between F1 and F2, and that CALIS give this covariance the name "CF1F2".

The VAR statement on line 49 merely lists the manifest variables to be analyzed.

SUBMITTING THE PROGRAM AND REVIEWING THE RESULTS

In conducting this analysis, you would next submit the preceding program, and review the output of PROC CALIS to determine whether your measurement model provides an adequate fit to the data. If not, you would use the output of CALIS to modify the model so as to attain a better fit. Space constraints preclude an adequate discussion of the steps to follow when doing this; see Hatcher (1994) for a detailed discussion of these steps.

DRAWING THE PROGRAM FIGURE FOR A LATENT-VARIABLE PATH ANALYSIS

Once you have used CFA to develop an adequate measurement model, you will modify the program figure presented in Figure 3 so that it is transformed into a model appropriate for path analysis with latent variables. Figure 4 presents such a model: It is a completed program figure for path analysis with latent variables.

Notice that, in Figure 4, the manifest variable V4 no longer appears. Assume that this is because the CFA of the measurement model (discussed above) showed that V4 was doing a poor job of assessing latent factor F1. Because of this, V4 was dropped from all subsequent analyses.

You should complete the following steps when transforming the program figure for CFA (presented in Figure 3) into the program figure for latent-variable path analysis (presented in Figure 4):

Step 1: Delete the variance estimates for the endogenous F variables. In Figure 4, the symbol "VAR?" no longer appears in the ovals for F1 and F2. This is because they are now endogenous variables, and you do not estimate the variances of endogenous variables.

Step 2: Create D terms for the endogenous F variables. Rule 2 indicates that you must create a residual term for each endogenous variable in the figure, and since F1 and F2 are now endogenous variables, you must create residual terms for them. Notice how this is done in Figure 4: The residual term D1 is created for F1, and D2 is created for F2. Since D1 and D2 are exogenous variables, the symbol "VAR?" appears below their names in the figure, consistent with Rule 4.

Step 3: Estimate the variances of the exogenous F variables. Because F3-F6 are now exogenous variables in a path model, you must now estimate their variances. This is done in Figure 4 by placing the symbol "VAR?" in the ovals for F3-F6.

Step 4: Fix one factor loading at 1 for each F variable. To solve the scale indeterminacy problem, Rule 17 indicates that you must fix one factor loading at one for each F variable in your model. Notice how this is done in Figure 4: With respect to F1, the factor loading for V3 has been fixed at 1 by writing the number "1" next to the appropriate causal arrow; with respect to F2, the factor loading for V7 has been fixed at 1 in the same way (and so forth, for the remaining F variables).

Step 5: Identifying the path coefficients to be estimated. In Figure 4, you can see that a straight, single-headed arrow goes from F2 to F1. This reflects the prediction that F2 (satisfaction) will have a causal effect

on F1 (commitment). To test this prediction, you will request that CALIS estimate the path coefficient for this path. In the program figure, this is symbolized by placing the symbol "P?" on the appropriate causal arrow. Similarly, you can see that the symbol "P?" appears on each straight arrow going from one F variable to another F variable in Figure 4.

PREPARING THE PROC CALIS PROGRAM TO PERFORM LATENT-VARIABLE PATH ANALYSIS

Below is the entire PROC CALIS program (minus the data step) to estimate the model appearing in Figure 4 (line numbers have been added on the left):

```

1   PROC CALIS
2       COVARIANCE
3       CORR
4       RESIDUAL
5       MODIFICATION ;
6   LINEQS
7       V1 = LV1F1 F1 + E1,
8       V2 = LV2F1 F1 + E2,
9       V3 = F1 + E3,
10      V5 = LV5F2 F2 + E5,
11      V6 = LV6F2 F2 + E6,
12      V7 = F2 + E7,
13      V8 = LV8F3 F3 + E8,
14      V9 = LV9F3 F3 + E9,
15      V10 = F3 + E10,
16      V11 = LV11F4 F4 + E11,
17      V12 = F4 + E12,
18      V13 = LV13F4 F4 + E13,
19      V14 = F5 + E14,
20      V15 = LV15F5 F5 + E15,
21      V16 = LV16F5 F5 + E16,
22      V17 = LV17F6 F6 + E17,
23      V18 = F6 + E18,
24      V19 = LV19F6 F6 + E19,
25      F1 = PF1F2 F2 +
26           PF1F5 F5 +
27           PF1F6 F6 + D1,
28      F2 = PF2F3 F3 +
29           PF2F4 F4 + D2;
30   STD
31      E1-E3 = VARE1-VARE3,
32      E5-E19 = VARE5-VARE19,
33      F3-F6 = VARF3-VARF6,
34      D1-D2 = VARD1-VARD2;
35   COV
36      F3 F4 = CF3F4,
37      F3 F5 = CF3F5,
38      F3 F6 = CF3F6,
39      F4 F5 = CF4F5,
40      F4 F6 = CF4F6,
41      F5 F6 = CF5F6;
42   VAR V1 V2 V3 V5-V19 ;
43   RUN;
```

Many of the equations in the preceding program are similar to those used in the program that performed CFA. Because those equations were discussed earlier, they will not be discussed again in this section. Instead, this section will describe how some equations from the earlier program had to be modified in order to estimate the model presented in Figure 4.

First, notice that, on line 9, the name for the factor loading "LV3F1" has been dropped from the equation for V3. This has the effect of fixing that factor loading at 1. This is appropriate, because Figure 4 shows that the factor loading for the path from F1 to V3 was to be fixed at 1. In the same way, notice that factor loadings have also been fixed at 1 in the equations for V7, V10, V12, V14, and V18. These are the same factor loadings that, according to Figure 4, were supposed to be fixed at 1.

Next, you can see that lines 25-27 present the equation that identifies the F variables that are predicted to have causal effects on F1 (you know that this is the equation for F1, because the short name "F1" appears to the left of the equals sign). The short variable names F2, F5, and F6 appear to the right of the equals sign in this equation, indicating that they are the variables expected to have direct effects on F1. Line 25 shows that the path coefficient name "PF1F2" has been given to the path coefficient that represents the effect of F2 on F1, line 26 shows that the path coefficient name "PF1F5" has been given to the path coefficient that represents the effect of F5 on F1, and so forth. In the same way, lines 28-29 identify the F variables predicted to have direct effects on F2.

The revised STD statement appears on lines 30-34. You can see that the variances for the E terms E1 through E3 are now estimated on line 31. You know this because the names for these E terms ("E1-E3") appear to the left of the equals sign, and the names for the corresponding variance estimates ("VARE1-VARE3") appear to the right of the equals sign. In the same way, line 32 contains the equation that causes variances to be estimated for the E terms E5 through E19. Similarly, you can see that the variances for the exogenous F variables are now estimated on line 33, and the variances of the D terms are estimated on line 34.

Finally, the COV statement appears on lines 35-41. It has been revised so that F1 and F2 are no longer allowed to have covariances with any of the other F variables. Instead, CALIS will now only estimate the covariance between every possible combination of F3, F4, F5, and F6. This is appropriate, because only these four F variables now are connected to one another by curved arrows in Figure 4.

SUBMITTING THE PROGRAM AND REVIEWING THE RESULTS

After the program has been submitted for analysis, you will review the results for PROC CALIS to again assess the extent to which the model provides an acceptable fit to the data. Because this is a path analysis, however, you will also review output to (a) determine whether the path coefficients depicted in Figure 4 were significant and in the predicted direction, and (b) assess the parsimony of your model. Hatcher (1994) provides detailed guidelines for these activities.

CONCLUSION

The purpose of this paper was to introduce a structured approach for using PROC CALIS to perform path analysis with latent variables. Due to space constraints, this paper was not able to cover many issues that are important in structural equation modeling. Such issues include the problem of model identification, appropriate methods for interpreting the results of the analysis, and problems associated with model modification. For a more detailed treatment of these and related issues, see Bollen (1989), Hatcher (1994), and Loehlin (1997).

ACKNOWLEDGMENTS

Reprinted with permission of SAS Institute Inc. from: Hatcher, Larry, *A Step-by-Step Approach to Using the SAS® System for Factor Analysis and Structural Equation Modeling*, Cary NC: 1994, Copyright © SAS Institute Inc. 588 pp.

SAS is a registered trademark or trademark of SAS Institute Inc. in the USA and other countries. ®indicates USA registration.

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

REFERENCES

- Anderson, J.C. & Gerbing, D.W. (1988), "Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach." *Psychological Bulletin*, 103, 411-423.
- Bentler, P.M. (1989), *EQS Structural Equations Program Manual*. Los Angeles: BMDP Statistical Software.
- Bollen, K.A. (1989), *Structural Equations with Latent Variables*. New York: John Wiley & Sons.
- Hatcher, L (1994), *A Step-by-Step Approach to Using the SAS® System for Factor Analysis and Structural Equation Modeling*. Cary, NC: SAS Institute Inc..

Statistics, Data Analysis, and Modeling

Joreskog, K.G. & Sorbom, D. (1989), *LISREL 7: A Guide to the Program and Applications, 2nd Edition*. Chicago: SPSS Inc.

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

Rusbult, C.E. (1980), "Commitment and Satisfaction in Romantic Associations: A Test of the Investment Model." *Journal of Experimental Social Psychology*, 16, 172-186.

Larry Hatcher, Psychology Department, Winthrop University, Rock Hill, SC 29733. (803) 323-2117.

Figure 1:
Theoretical Model
To Be Tested

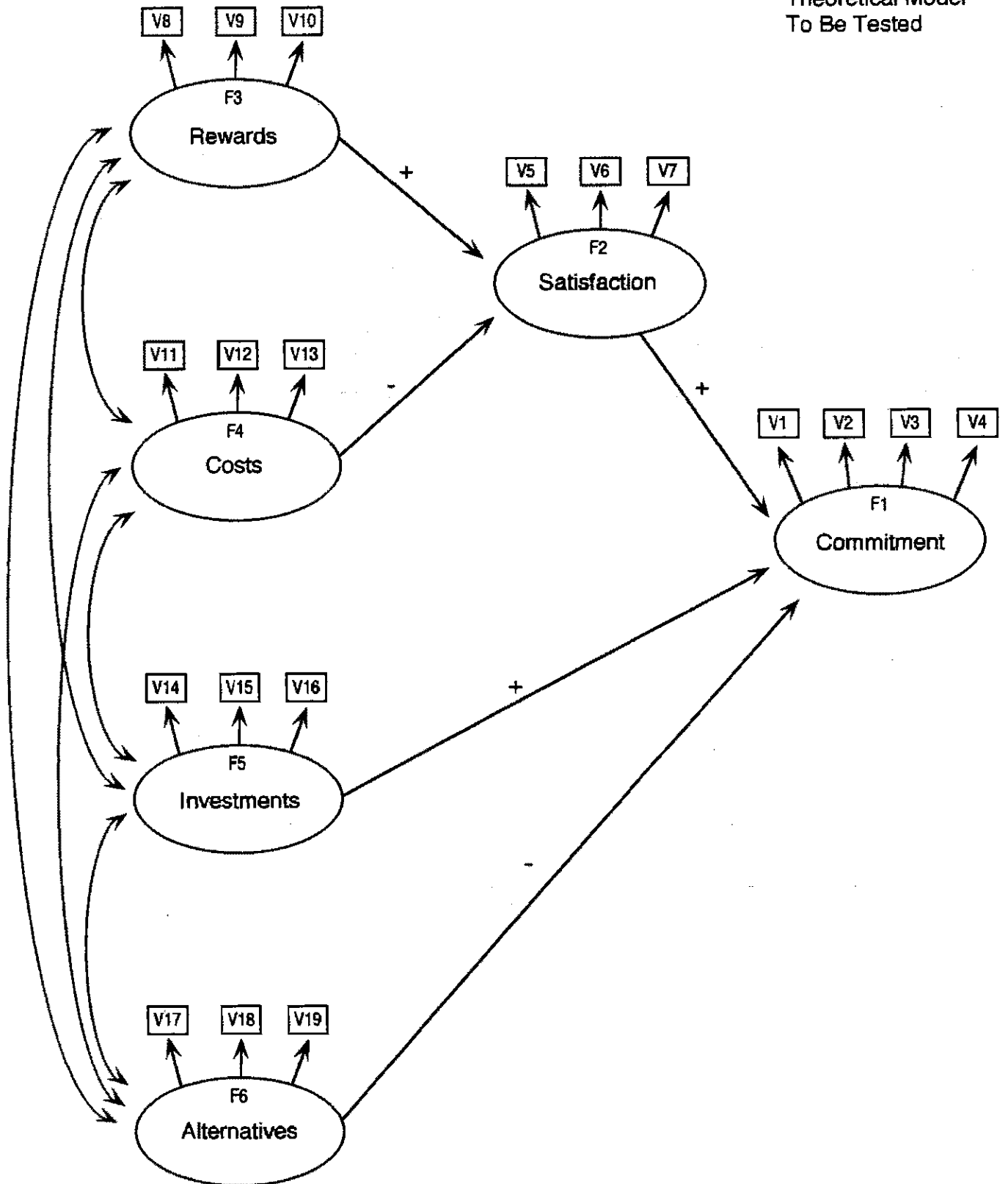


Figure 2:
Initial Program
Figure for
Confirmatory Factor
Analysis

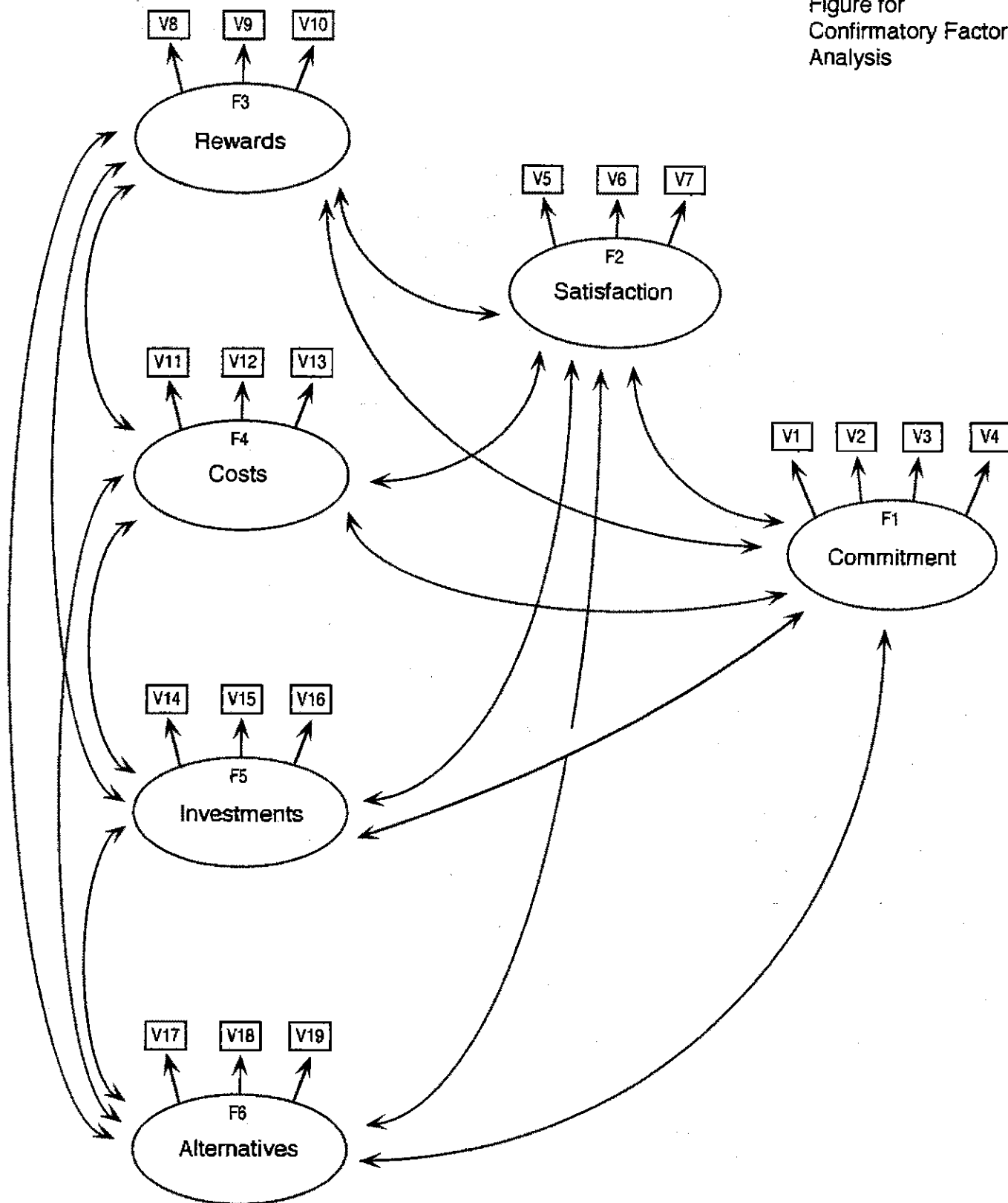


Figure 3:
Completed
Program Figure for
Confirmatory Factor
Analysis

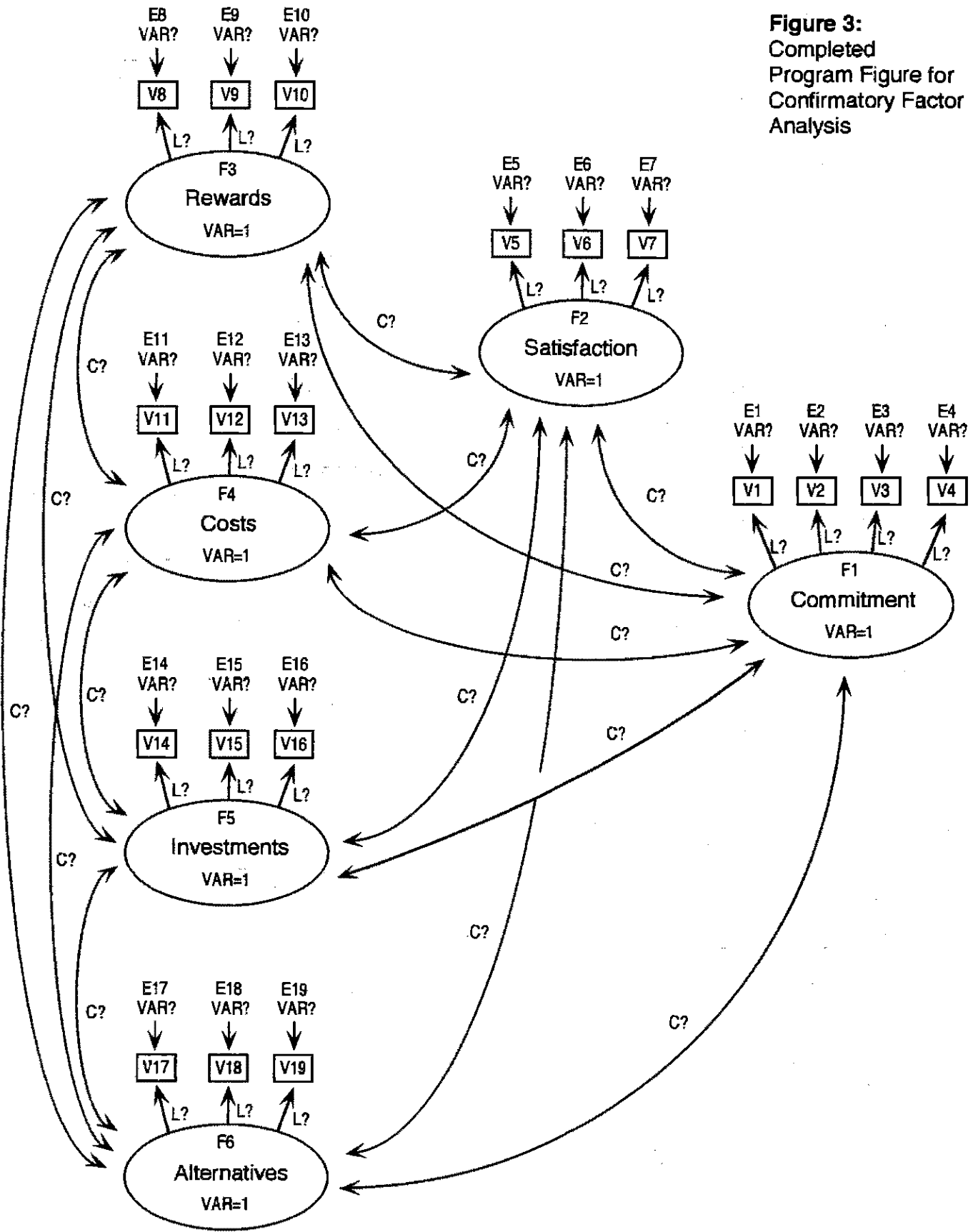


Figure 4:
Completed
Program Figure for
Latent-Variable
Path Analysis

