Paper 384-2008

Structural Equation Modeling and Path Analysis Using PROC TCALIS in SAS® 9.2

Yiu-Fai Yung, SAS Institute Inc., Cary NC

ABSTRACT

The TCALIS procedure, which is new and experimental in SAS 9.2, is a major enhancement of the CALIS procedure. Both the TCALIS and CALIS procedures provide statistical tools for analyzing structural equations and related models, although the new features are available only in PROC TCALIS. In this paper, practical examples are used to illustrate some PROC TCALIS features: the PATH modeling language, customization of the fit summary table, effects partitioning, the multiple-group model, analysis of mean structures, and simultaneous tests of parametric functions. Other important features of PROC TCALIS are also described.

INTRODUCTION

Structural equation modeling is a sophisticated statistical method that can model complicated functional or "causal" relationships among variables, whether the variables are observed (that is, manifest variables) or not (that is, latent variables). Structural equation modeling has a wide range of applications. In health science research, structural equation modeling has been used to identify factors associated with substance abuse and personality disorders. In marketing research, structural equation models have been used to suggest ways to improve Web-page design to enhance the Web-browsing experience. For an introduction to structural equation modeling, see Bollen (1989) or Loehlin (2004).

PROC CALIS fits structural equation models. In SAS 9.2, PROC TCALIS provides a major update of PROC CALIS. Most of the popular features of PROC CALIS are available in PROC TCALIS. This paper highlights some new features in PROC TCALIS. Specifically, the following new features are illustrated with two examples:

- PATH modeling language—enables you to specify your path model easily
- effect analysis with the EFFPART statement—enables you to look at direct, indirect, and total effects in customizable ways
- multiple-group analysis—enables you to study the similarities and differences among independent groups
- enhanced mean structure analysis—enables you to analyze the mean structures simultaneously with the covariance structures
- simultaneous tests of parametric functions with the SIMTEST statement—enables you to test specific a priori hypotheses simultaneously

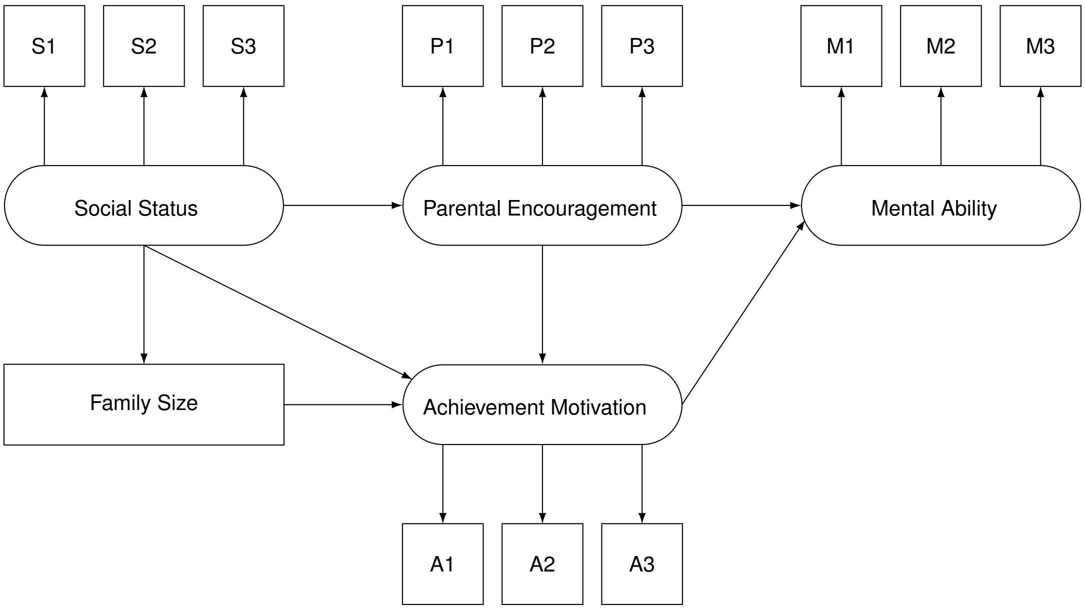
Other important new features of PROC TCALIS are summarized in the CONCLUSION section.

STRUCTURAL EQUATION MODELS AND PATH DIAGRAMS

Many structural equation models are represented by path diagrams, with which researchers describe their theories about the relationships among variables. For example, a researcher has a complicated theory about how mental ability, achievement motivation, and the academic success of a student depend on his or her social background, family size, and parental encouragement. In his theory, these variables are related by a series of "causal chains." The researcher starts with a representation of his causal theory and hopes to see that the data lend support to his theory.

First, the researcher explicates his theory by using a path diagram. Inspired by the theoretical model of Marjoribanks (1974), Figure 1 shows an example of such a path diagram representation of the researcher's causal theory about the mental abilities of students.

Figure 1 Factors Affecting Mental Abilities: Path Diagram



The path diagram shown in Figure 1 is modified from Marjoribanks (1974), and it is not intended to perfectly represent the theoretical model of Marjoribanks (1974). The data in the current illustration are fictitious, so no part of the current analysis is meant to be comparable to the results of Marjoribanks (1974). Here, the purpose is to illustrate the new PROC TCALIS features in an interesting substantive context.

In the path diagram, Mental Ability is predicted directly by Parental Encouragement and Achievement Motivation. In turn, Parental Encouragement is predicted by Social Status, and Achievement Motivation is predicted by Social Status, Family Size, and Parental Encouragement. The path diagram also implies that the effects of Social Status and Family Size on Mental Ability are only indirect: they affect Mental Ability only through their effects on Parental Encouragement and Achievement Motivation. Furthermore, some variables serve dual roles as predictors and outcomes in the path diagram. For example, Parental Encouragement is a predictor variable of Mental Ability and Achievement Motivation; but at the same time it is an outcome variable of Social Status. These important relationships among the theoretical variables are referred to as structural relationships. The corresponding model for these relationships is called a *structural model*.

The path-diagram representation of complicated structural relationships is much more than a description of a few predictors and outcome blocks. This means that you cannot simply use multiple regression techniques to analyze such a structural model. For example, consider a regression of Mental Ability on Parental Encouragement and Achievement Motivation by using the following statements:

```
proc reg;
   model MentalAbility = ParentalEncouragement AchievementMotivation;
run;
```

The problem with this regression analysis is that it ignores the common causes of the two predictors and the regression (or dependency) of Achievement Motivation on Parental Encouragement.

Consider now a regression of Mental Ability on Social Status and Family Size as follows:

```
proc reg;
   model MentalAbility = SocialStatus FamilySize;
run;
```

The problem with this regression analysis is that it incorrectly models the *indirect* effects of Social Status and Family Size on Mental Ability as *direct* effects.

In addition to the limitations in treating complicated causal relationships, multiple regression techniques cannot treat latent variables, which are represented by ovals in the path diagram. The variables Social Status, Parental Encouragement, Achievement Motivation, and Mental Ability are all latent variables, and hence are not available in the input data set.

All the previously described issues call for structural equation modeling techniques, which can handle observed and latent variables and analyze the entire complicated path model. PROC TCALIS provides you with such structural equation modeling techniques. Specifically, you can use the newly developed PATH modeling language in PROC TCALIS to input your path model with ease.

Before diving into the PATH modeling specification, some more details about the the path diagram are needed. The measurement indicators in Figure 1 have been ignored in the exposition of the structural model. These indicators reflect the underlying latent constructs that are linked to them. For example, there are three indicator measures for Mental Ability: M1, M2, and M3. These three variables could be mental tests administered to the individuals in the research. M1 could be a verbal test, M2 could be a math test, and M3 could be a test of reasoning skills. Similarly, other latent variables in the path diagram are associated with different sets of observed indicators. The part of the model that links the latent variables to their observed indicators is called the *measurement model*.

Conceptually, the indicator and its associated latent variable take the following form of relationship in the measurement model:

indicator = latent construct + measurement error

In the path diagram, this translates into the following path:

indicator ← latent construct

In the current path-diagram notation, the measurement error in the path is implicitly assumed. There is no need to depict the error terms explicitly in the path diagram. Hence, the path diagram shown in Figure 1 provides a complete account of the researcher's causal theory, which contains the structural model and the measurement model.

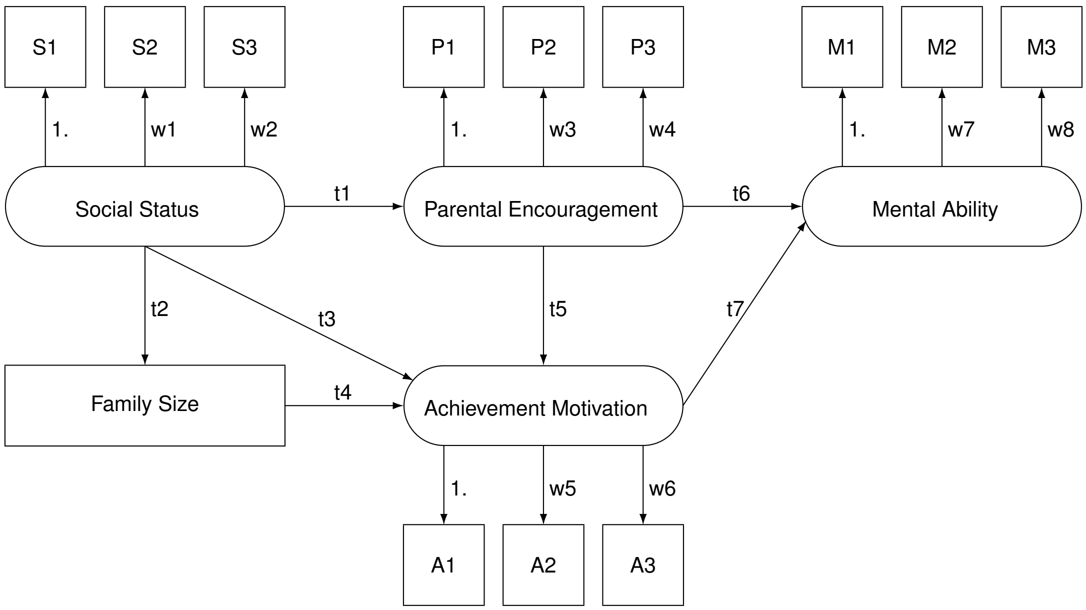
THE PATH MODELING LANGUAGE IN PROC TCALIS

Given the path diagram shown in Figure 1, you are almost ready to translate it into the path modeling language supported by PROC TCALIS. But first, you need to define some important parameters in the model.

TRANSLATING YOUR PATH DIAGRAM INTO THE PATH STATEMENT SPECIFICATION

To make the correspondence between the path diagram and the PATH statement specification in PROC TCALIS more obvious, you insert the path effect or coefficient parameters into the path diagram in Figure 1. Figure 2 shows such an updated version of the path diagram.

Figure 2 Factors Affecting Mental Abilities: Path Diagram with Parameters



In Figure 2, either a parameter name or a value of 1.0 is added to each path in the diagram. Parameters t1, t2, . . . , t7 represent the direct path effects or regression coefficients in the structural model. Parameters w1, w2, . . . , w9 represent the effects or regression coefficients of the latent variables on the associated observed indicators in the measurement model. All of these path coefficients or effects are free parameters to estimate. For each latent variable, there is exactly one regression coefficient fixed at 1.0 for an indicator variable. These values are fixed parameters in the measurement model and are used to identify the scale of the latent variables. The scale identification is needed because latent variables are unmeasured and can have arbitrary scales of measurement. For each latent variable, you fix its scale with one of its indicator variables so as to eliminate the arbitrariness of the latent variable scale. This is necessary for estimating the rest of the free parameters.

With the updated path diagram, you are ready to translate your path diagram into the PATH modeling language. This is shown in the following statements:

```
data mental(type=cov);
   _type_='cov';
   input _name_ $10. FamilySize s1 s2 s3 p1 p2 p3 a1 a2 a3 m1 m2 m3;
   datalines;
FamilySize 12.25 -1.05 -0.85 -1.84 -0.82 -0.06 -1.12 -2.46 -4.04 -0.46 -0.59 -5.20 -4.84
s1         -1.05  5.29  4.91  2.90  0.83  0.78  1.20  2.42  3.23  1.82  2.99  3.11  4.27
s2         -0.85  4.91  7.29  3.57  1.26  1.06  1.72  3.03  3.39  2.20  2.95  3.16  3.38
s3         -1.84  2.90  3.57  4.41  1.13  0.75  1.40  2.41  2.74  1.48  1.76  3.02  3.45
p1         -0.82  0.83  1.26  1.13  3.24 -1.47  0.89  1.81  1.45  0.44  0.68  0.81  1.24
p2         -0.06  0.78  1.06  0.75 -1.47  2.89  0.52  1.19  1.11  1.31  0.29  1.15  1.54
p3         -1.12  1.20  1.72  1.40  0.89  0.52  2.10  2.17  1.86  1.31  0.55  1.57  2.19
a1         -2.46  2.42  3.03  2.41  1.81  1.19  2.17  5.48  4.23  2.37  2.16  4.32  5.26
a2         -4.04  3.23  3.39  2.74  1.45  1.11  1.86  4.23  6.30  2.21  2.95  7.45  7.16
a3         -0.46  1.82  2.20  1.48  0.44  1.31  1.31  2.37  2.21  4.84  0.92  3.37  3.99
m1         -0.59  2.99  2.95  1.76  0.68  0.29  0.55  2.16  2.95  0.92 17.64  7.18  6.17
m2         -5.20  3.11  3.16  3.02  0.81  1.15  1.57  4.32  7.45  3.37  7.18 20.25 12.83
m3         -4.84  4.27  3.38  3.45  1.24  1.54  2.19  5.26  7.16  3.99  6.17 12.83 18.66
;

proc tcalis data=mental nobs=115;
   path
      /* Structural Model */
      SocialStatus           ->  ParentalEncouragement     t1,
      SocialStatus           ->  FamilySize                t2,
      SocialStatus           ->  AchievementMotivation     t3,
      FamilySize             ->  AchievementMotivation     t4,
      ParentalEncouragement  ->  AchievementMotivation     t5,
      ParentalEncouragement  ->  MentalAbility             t6,
      AchievementMotivation  ->  MentalAbility             t7,

      /* Measurement Model */
      S1     <-     SocialStatus              1.0,
      S2     <-     SocialStatus              w1,
      S3     <-     SocialStatus              w2,
      P1     <-     ParentalEncouragement     1.0,
      P2     <-     ParentalEncouragement     w3,
      P3     <-     ParentalEncouragement     w4,
      A1     <-     AchievementMotivation     1.,
      A2     <-     AchievementMotivation     w5,
      A3     <-     AchievementMotivation     w6,
      M1     <-     MentalAbility             1.0,
      M2     <-     MentalAbility             w7,
      M3     <-     MentalAbility             w8;
run;
```

You use the DATA step to input the covariance matrix of the observed variables. The set of variables includes the FamilySize variable and all measurement indicators of the latent variables in the model. The default PROC TCALIS method, maximum likelihood (METHOD=ML), does not require the raw data input, so you can just input the covariance matrix.

Next, the TCALIS procedure is invoked. You specify the number of observations (`NOBS=115`) in this statement because this information is not provided in the input data set. The PATH statement signifies the PATH modeling language for specifying the model. In the PATH statement, you specify the paths of the model as entries, which are separated by commas. Each PATH statement entry corresponds to a path in the path diagram. In an entry, you specify a path from

Variable1 to Variable2, followed by the path effect or coefficient specification parameter_spec, as shown in the following format:

```
Variable1  ->  Variable2         parameter_spec
```

The effect specification parameter_spec can take any of the following formats:

- a parameter name without an initial estimate (for example, **w1**)
- a parameter name with an initial estimate (for example, **w1 (0.5)**)
- a fixed parameter value (for example, **1.0**)

Notice that path specifications Variable1–>Variable2 and Variable2<–Variable1 are equivalent. Although path entries in the PATH statement are not required to be ordered in a particular way, you can group the entries for more clarity.

For this particular model, the 19 path entries in the PATH statement are the only specification you need to define the entire model. PROC TCALIS sets other parameters such as error variances in the model automatically. For example, by default the regression of Mental Ability on Parental Encouragement and Achievement Motivation is not a perfect one, and so PROC TCALIS sets a nontrivial error term in this regression automatically. As a result, the corresponding error variance for Mental Ability is a free parameter by default, making this kind of routine specification unnecessary. Nonetheless, you can specify these error variances as free (or fixed) parameters explicitly in the PVAR statement, as shown in a later example. This becomes necessary in situations where you want to constrain the error variance parameters.

OUTPUT FOR THE PATH MODEL

This section illustrates some output from running PROC TCALIS for the current example. PROC TCALIS obtains a converged solution for the current analysis. To conserve space, optimization results are not presented here. The fit summary is shown in Figure 3.

Figure 3 Fit Summary of the Path Model about Mental Abilities

```
                                Fit Summary

      Modeling Info        N Observations                              115
                           N Variables                                  13
                           N Moments                                    91
                           N Parameters                                 32
                           N Active Constraints                          0
                           Independence Model Chi-Square          797.8656
                           Independence Model Chi-Square DF             78
      Absolute Index       Fit Function                             1.7258
                           Chi-Square                             196.7455
                           Chi-Square DF                                59
                           Pr > Chi-Square                          0.0000
                           Z-Test of Wilson & Hilferty              8.1107
                           Hoelter Critical N                           47
                           Root Mean Square Residual (RMSR)         0.5579
                           Standardized RMSR (SRMSR)                0.0936
                           Goodness of Fit Index (GFI)              0.8276
      Parsimony Index      Adjusted GFI (AGFI)                      0.7341
                           Parsimonious GFI                         0.6260
                           RMSEA Estimate                           0.1431
                           RMSEA Lower 90% Confidence Limit         0.1213
                           RMSEA Upper 90% Confidence Limit         0.1655
                           Probability of Close Fit                 0.0000
                           ECVI Estimate                            2.3658
                           ECVI Lower 90% Confidence Limit          2.0099
                           ECVI Upper 90% Confidence Limit          2.7982
                           Akaike Information Criterion            78.7455
                           Bozdogan CAIC                         -142.2055
                           Schwarz Bayesian Criterion             -83.2055
                           McDonald Centrality                      0.5494
      Incremental Index    Bentler Comparative Fit Index            0.8087
                           Bentler-Bonett NFI                       0.7534
                           Bentler-Bonett Non-normed Index          0.7470
                           Bollen Normed Index Rho1                 0.6740
                           Bollen Non-normed Index Delta2           0.8136
                           James et al. Parsimonious NFI            0.5699
```

Figure 3 shows the fit indices and modeling information. There is so much information that you might wonder why it is called a fit "summary" table. In practice, not all of these indices are considered by all researchers. Some of them are more popular than others and different researchers might prefer to look at different sets of fit indices. To make your fit summary concise and tailored to your needs, you can use the FITINDEX statement to customize the fit summary table. For example, you can select a subset of fit information to display by the following statement:

```
fitindex on(only)=[agfi srmsr rmsea bentlercfi] noindextype;
```

In the FITINDEX statement, you use the ON(ONLY)= option to select your set of fit indices or information to display. In addition, you suppress the display of fit index types by using the NOINDEXTYPE option. This avoids the superfluous organization of just a few fit indices. By including the preceding statement in your PROC TCALIS run, you obtain the fit summary table shown in Figure 4 instead.

Figure 4 Customized Fit Summary of the Path Model about Mental Abilities

Fit Summary

Standardized RMSR (SRMSR)	0.0936
Adjusted GFI (AGFI)	0.7341
RMSEA Estimate	0.1431
Bentler Comparative Fit Index	0.8087

This fit summary table is much more concise than the original one shown in Figure 3. Both standardized RMSR (root mean square residual) and RMSEA (root mean square error of approximation) are well above 0.05, indicating a bad model fit. According to Browne and Cudeck (1993), an RMSEA under 0.05 indicates a good model fit. Also, both AGFI (adjusted goodness-of-fit index) and Bentler's CFI (comparative fit index) are well below 0.90, confirming a less than satisfactory model fit. Despite the bad model fit, it is still worthwhile to proceed to the estimation results to illustrate the new features of PROC TCALIS.

The estimates, standard errors, and t values for path coefficients or effects are shown in the Figure 5.

Figure 5 Estimation of the Path Coefficients of the Model about Mental Abilities

PATH List

Path			Parameter	Estimate	Standard Error	t Value
SocialStatus	->	ParentalEncouragement	t1	0.25157	0.06920	3.63558
SocialStatus	->	FamilySize	t2	-0.29486	0.17120	-1.72228
SocialStatus	->	AchievementMotivation	t3	0.21931	0.11467	1.91246
FamilySize	->	AchievementMotivation	t4	-0.12875	0.03599	-3.57686
ParentalEncouragement	->	AchievementMotivation	t5	1.57011	0.51756	3.03370
ParentalEncouragement	->	MentalAbility	t6	-1.73425	0.90472	-1.91690
AchievementMotivation	->	MentalAbility	t7	1.31958	0.41244	3.19945
s1	<-	SocialStatus		1.00000		
s2	<-	SocialStatus	w1	1.19873	0.10518	11.39712
s3	<-	SocialStatus	w2	0.74996	0.08640	8.68011
p1	<-	ParentalEncouragement		1.00000		
p2	<-	ParentalEncouragement	w3	0.58396	0.26679	2.18888
p3	<-	ParentalEncouragement	w4	1.46540	0.37668	3.89035
a1	<-	AchievementMotivation		1.00000		
a2	<-	AchievementMotivation	w5	1.18205	0.10843	10.90119
a3	<-	AchievementMotivation	w6	0.59866	0.10832	5.52685
m1	<-	MentalAbility		1.00000		
m2	<-	MentalAbility	w7	2.10940	0.50924	4.14222
m3	<-	MentalAbility	w8	2.06794	0.49730	4.15835

Because the estimation is based on the asymptotic theory, the t value is assumed to behave approximately like a standardized normal variate. Therefore, at the 0.05 α-level, you can use 1.96 as a critical value for judging statistical significance of these path coefficients. Hence, the SocialStatus –> FamilySize path is not significant, and the SocialStatus –> AchievementMotivation and the ParentalEncouragement –> MentalAbility paths are only marginally significant. All other path coefficients are significant.

There are three path coefficients with negative estimates. First, the negative path coefficient for the SocialStatus –> FamilySize path means that higher social status predicts smaller family size. However, this coefficient is not significant at the 0.05 α-level, and this interpretation is not definite. Second, the negative coefficient for the FamilySize –> AchievementMotivation path means that students in large families tend to have less achievement motivation. Finally, the negative coefficient for the ParentalEncouragement –> MentalAbility path indicates that ParentalEncouragement has negative impact on MentalAbility. This seems to be counterintuitive, although the effect is only marginally significant.

The estimates for the variance parameters are shown in Figure 6.

Figure 6 Estimation of Variance Parameters of the Model about Mental Abilities

Variance Parameters

Variance Type	Variable	Parameter	Estimate	Standard Error	t Value
Exogenous	SocialStatus	_Add01	4.02630	0.71941	5.59670
Error	FamilySize	_Add02	11.89995	1.58165	7.52377
	s1	_Add03	1.26370	0.28745	4.39621
	s2	_Add04	1.50437	0.38816	3.87566
	s3	_Add05	2.14545	0.32421	6.61750
	p1	_Add06	2.66808	0.37568	7.10195
	p2	_Add07	2.69497	0.36413	7.40105
	p3	_Add08	0.87184	0.22081	3.94845
	a1	_Add09	1.85531	0.30644	6.05431
	a2	_Add10	1.23540	0.28150	4.38869
	a3	_Add11	3.54095	0.48630	7.28135
	m1	_Add12	14.65867	2.01101	7.28921
	m2	_Add13	6.98437	1.42401	4.90471
	m3	_Add14	5.91079	1.29816	4.55322
	ParentalEncouragement	_Add15	0.31711	0.16217	1.95539
	AchievementMotivation	_Add16	0.81382	0.28730	2.83261
	MentalAbility	_Add17	0.42560	0.42053	1.01204

By default, PROC TCALIS sets all of the variance parameters in Figure 6. PROC TCALIS also generates the corresponding parameter names. All these names start with a prefix '_Add.' The first one is a variance parameter for the exogenous variable SocialStatus. This variance parameter is significant at the 0.05 α-level. The remaining parameters are error variances for the variables. Only the estimate of error variance for MentalAbility is not significant (t=1.01); all others are either significant or marginally significant at the 0.05 α-level.

ANALYZING TOTAL, DIRECT, AND INDIRECT EFFECTS BY USING THE EFFPART STATEMENT

With complicated causal paths in structural equation models, researchers usually want to assess how a focal set of variables affect each other. The focal set of effects includes not only the direct path effects as shown in the path diagram, but also the indirect and total effects even when the pairs of variables of interest are not linked directly by paths. For example, in the current example although SocialStatus does not have a direct effect on MentalAbility, it clearly is a remote "cause" of MentalAbility due to its effects on other variables that are immediate predictors (that is, ParentalEncouragement and AchievementMotivation) of MentalAbility. In PROC TCALIS, you can perform this kind of customized effect analysis by using the EFFPART statement. For example, to study the effects of all determinants on mental ability in the model, you can specify the following statement in your PROC TCALIS step:

```
effpart MentalAbility <- SocialStatus FamilySize ParentalEncouragement AchievementMotivation;
```

This statement requests an analysis of the effects of SocialStatus, FamilySize, ParentalEncouragement, and AchievementMotivation on MentalAbility. The results are shown in Figure 7.

Figure 7 Stability Coefficient and Effects Partitioning of the Determinants of Mental Abilities

```
Stability Coefficient of Reciprocal Causation = 0

Stability Coefficient < 1

Total and Indirect Effects Converge
```

Figure 7 *continued*

Effects on MentalAbility
Effect / Std Error / tValue / pValue

	Total	Direct	Indirect
SocialStatus	0.4244	0	0.4244
	0.1280		0.1280
	3.3159		3.3159
	0.000914		0.000914
FamilySize	-0.1699	0	-0.1699
	0.0572		0.0572
	-2.9696		-2.9696
	0.002982		0.002982
ParentalEncouragement	0.3376	-1.7343	2.0719
	0.4045	0.9047	1.0483
	0.8346	-1.9169	1.9763
	0.4039	0.0553	0.0481
AchievementMotivation	1.3196	1.3196	0
	0.4124	0.4124	
	3.1995	3.1995	
	0.001377	0.001377	

Figure 7 shows the total, direct, and indirect effects of various determinants of MentalAbility. These effects are computed with standard errors, t values, and approximate p values. Before they are computed, PROC TCALIS checks the convergence of total and indirect effects. The so-called stability coefficient is computed for the estimated model. If this coefficient is less than one, the convergence of the total and indirect effects is guaranteed and the computations of the total and indirect effects are valid for interpretations. For the current example, the stability coefficient shows the convergence of the total and indirect effects.

As shown in Figure 7, the effect of SocialStatus on MentalAbility is indirect only, because the direct effect is zero and the total effect is the same as the indirect effect. Even though SocialStatus is not a direct predictor of MentalAbility in the model, this table shows that the total effect of SocialStatus on MentalAbility is indeed significant ($p < 0.01$). Overall, social status does have a positive impact on the mental abilities of the students. Similarly, the effect of FamilySize on MentalAbility is all *indirect* and is significant ($p < 0.01$). Overall, family sizes negatively affect the mental abilities of the students. In contrast, the effect of AchievementMotivation on MentalAbility is all *direct* and is significant ($p < 0.01$). Higher achievement motivation predicts better mental abilities. Because only direct effect is involved in the relationship between AchievementMotivation and MentalAbility, this interpretation could have been drawn merely by looking at the estimation results of the path diagram. The effect analysis provided in this section does not add more information about the relationship between this pair of variables. For other pairs of variables, however, the effect analysis does provide more information and interesting interpretations. As explained in the following, the partitioning of the effect of ParentalEncouragement on MentalAbility would provide such an example.

The effects of ParentalEncouragement on MentalAbility as revealed in Figure 7 are quite interesting (or surprising if your interpretations do not match your anticipations). While the indirect effect of ParentalEncouragement on MentalAbility is positive, the direct effect of ParentalEncouragement on MentalAbility is actually negative. Both of the direct and indirect effects are marginally significant. Adding these two effects together results in a small total effect that is not statistically significant. Looking back at the path diagram, you can see that the indirect effect of Parental Encouragement on Mental Ability is only through the Parental Encouragement→Achievement Motivation→Mental Ability track. Therefore, Parental Encouragement affects Mental Ability positively only through its effects on Achievement Motivation, while its direct effect on Mental Ability is actually negative. This dual role of Parental Encouragement is quite interesting.

Another possible interpretation is that Parental Encouragement actually has nothing to do with Mental Ability. The model specification itself might have created such an artifact so that the direct and indirect effects must cancel each other out in the model. With limited substantive background knowledge about the nature of the problem, you cannot make a definite conclusion about which interpretation is more plausible in the current analysis. However, the usefulness of effect analysis by using the EFFPART statement is demonstrated. That is, effect analysis reveals additional information that might not be gleaned from the estimation results of the path model.

Like many other output tables from PROC TCALIS, the effect analysis tables can be customized in certain ways by the EFFPART statement. In the preceding EFFPART statement, effects on Mental Ability are put into a *single* table with all determinant variables listed. If you want to emphasize the effects of the determinant variables and create effects tables for each of them, you can switch the order of specification as in the following statement:

```
effpart  SocialStatus FamilySize ParentalEncouragement AchievementMotivation -> MentalAbility;
```

This statement creates four tables, each with a determinant variable's effects on MentalAbility. This arrangement would be especially useful if you have more outcome variables on the right side of the arrow. For example, if you want to emphasize how SocialStatus affects each of the four endogenous variables (FamilySize, MentalAbility, ParentalEncouragement, and AchievementMotivation), the following specification would achieve your purpose better than if you specify the other way around:

```
effpart  SocialStatus -> FamilySize MentalAbility ParentalEncouragement AchievementMotivation;
```

MULTIPLE-GROUP ANALYSIS

The ability to analyze several independent groups of data simultaneously is another important new feature of PROC TCALIS. Consider a group of individuals where a structural equation model has been established. The model could be a theory about the individuals' Web-surfing behaviors, drug-addictive behaviors, consumer spending patterns, or any area that might postulate a complicated set of causal relationships among variables. Consider now another group of individuals who are put into a different computing environment, an intervention program, or a different marketing location. Would the established structural equation model apply to this new group of individuals? Even if the same model structures apply to both groups, would the parameter estimates be the same for the two groups? These kinds of research questions motivate the multiple-group structural equation modeling (see, for example, Jöreskog 1971), which PROC TCALIS supports with an easy syntax for model specifications. In this section, an artificial example is used to illustrate the use of PROC TCALIS for a multiple-group analysis of purchase behavior. In addition, analysis of mean structures is demonstrated.

MULTIPLE-GROUP PURCHASE BEHAVIOR

In this example, data were collected from customers who made purchases from a retail company during years 2002 and 2003. A two-group structural equation model is fitted to the data.

The measured variables are:

Spend02:	total purchase amount in 2002
Spend03:	total purchase amount in 2003
Courtesy:	rating of the courtesy of the customer service
Responsive:	rating of the responsiveness of the customer service
Helpful:	rating of the helpfulness of the customer service
Delivery:	rating of the timeliness of the delivery
Pricing:	rating of the product pricing
Availability:	rating of the product availability
Quality:	rating of the product quality

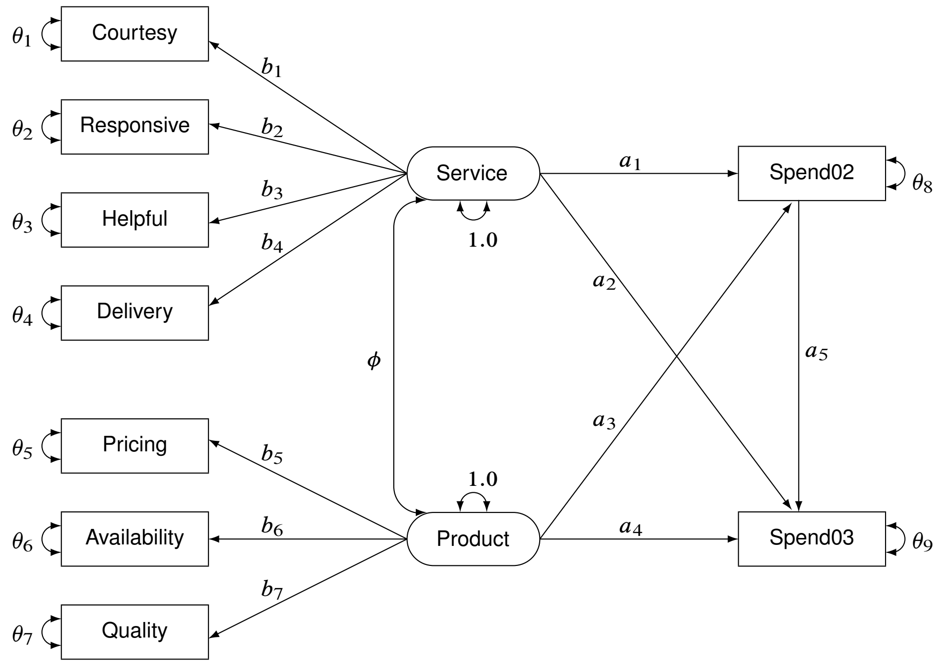
Nine-point rating scales were used, with 1 representing "extremely unsatisfied" and 9 representing "extremely satisfied." Data were collected from two different regions, which are labeled as Region 1 (N=378) and Region 2 (N=423), respectively. These two regions are the two independent groups in the multiple-group structural equation model analysis. The ratings were collected at the end of year 2002 so that they represent customers' purchasing experience in year 2002.

The central questions of the study are:

- How do customer service quality and product quality predict purchases?
- Do previous purchases predict subsequent purchases?
- Do the two regions have different structural models for predicting purchases?

In asking these research questions, you use several constructs that might or might not correspond directly to the actual data obtained. You can operationalize previous purchases and subsequent purchases by the amounts of money spent by the customers in 2002 (Spend02) and 2003 (Spend03), respectively. Spend02 and Spend03 are measured variables in the data set. For measuring customer service quality and product quality, however, you can only measure them indirectly by using the observed indicators, as shown in the measurement model of Figure 8.

Figure 8 Path Diagram of Purchasing Behavior



In Figure 8, the left part of the path diagram shows the measurement model for the latent factors Service (customer service quality) and Product (product quality). The Service factor has four measured indicators: Courtesy, Responsive, Helpful, and Delivery. The associated effects of the Service factor on these indicators are b_1, b_2, b_3, and b_4, respectively. The Product factor has three indicators: Pricing, Availability, and Quality, with associated path coefficients b_5, b_6, and b_7, respectively.

The two latent factors are predictors of the purchase amounts Spend02 (previous purchase) and Spend03 (subsequent purchase). In addition, Spend02 also serves as a predictor of Spend03. The path coefficients or effects for this structural model are represented by a1–a5 in the path diagram.

Unlike the previous example where you let PROC TCALIS set the default variance and covariance parameters in your model, in this example you depict them explicitly in the path diagram. This introduces new notation for the path diagram and new specification types of the PATH modeling language in PROC TCALIS.

First, consider the variance parameters in the model. A very simple principle is that each variable in the path diagram should associate with a variance parameter in the path diagram. The nature of this variance parameter depends on whether the variable is endogenous (or dependent) or exogenous (or independent).

For endogenous variables, which serve as outcome variables (those being pointed at by at least one single-headed arrow) at least once in the model, the associated variance parameters are their error variances that are not accounted for by the predictors. For example, all observed indicator variables in the current model are endogenous variables. In the path diagram, the double-headed arrows attached to these variables represent error variances. The associated error variance parameters are θ_1 to θ_9, respectively. Notice that these error variance parameters are just *partial* variances of the endogenous variables. The total variance of an endogenous variable has another additive component: the systematic variance accounted for by the predictors. The systematic variance components of the endogenous variables are implied by the model structures such as the paths and the variances of the predictors, and so you do not need to specify them explicitly in either the PATH diagram or the PATH model input of PROC TCALIS.

For exogenous variables, which never serve as outcome variables in the model, the variance parameters are the total variances, or simply, variances of these variables. For example, in the path diagram, the double-headed arrows attached to Service and Product represent the variances of these two latent variables. In the current model, both of these variances are fixed at 1.0 for the identification of the latent variable scales.

Now, consider the covariance parameters in the model. Which pairs of variables should you specify as the covariance parameters? The basic principle is that you usually need to specify covariances between all pairs of exogenous variables (excluding error terms). This is needed because these covariances are not functions of other parameters in the model. Nonzero covariances between exogenous variables (excluding error terms), no matter how small they might be, are the norm rather than the exception. In this model, there are only two non-error type exogenous variables: Service and Product. Their covariance is represented by a double-headed arrow pointing to the two variables in Figure 8. The corresponding covariance parameter is ϕ.

A MACRO FOR THE BASIC PATH MODEL

For the moment, it is hypothesized that both Region 1 and Region 2 data are fitted by the same model as shown in Figure 8. You use the PATH modeling language to specify the model. In addition, you define the entire model specification as a macro for later use. The macro is defined as follows:

```
%macro BasePathModel;
path
   Service -> Spend02         a1,
   Service -> Spend03         a2,
   Product -> Spend02         a3,
   Product -> Spend03         a4,
   Spend02 -> Spend03         a5,
   Service -> Courtesy        b1,
   Service -> Responsive      b2,
   Service -> Helpful         b3,
   Service -> Delivery        b4,
   Product -> Pricing         b5,
   Product -> Availability    b6,
   Product -> Quality         b7;
pvar
   Courtesy Responsive Helpful
   Delivery Pricing
   Availability Quality = theta01-theta07,
   Spend02 = theta08,
   Spend03 = theta09,
   Service Product = 2 * 1.;
pcov
   Service Product = phi;
mean
   Courtesy Responsive Helpful
   Delivery Pricing
   Availability Quality  = intercept01-intercept07,
   Spend02  Spend03      = InterSpend02 InterSpend03,
   Service Product       = 2 * 0.;
%mend;
```

The name of this macro is BasePathModel. Defining a model by a SAS macro is not essential. You could have inserted the PATH model code directly into a PROC TCALIS step to get the same results. A macro is used to facilitate the presentation of the subsequent specification.

As explained in the previous example, you can translate each path in the path diagram to the path entries in the PATH statement. In each path entry, you specify the path and then the associated parameter for the effect or regression coefficient. Three additional statements are used in the current specification for the variance, covariance, and mean (or intercept) parameters.

In the PVAR statement, you specify the variance parameters: either error variances or variances of exogenous variables. These variance parameters correspond to those double-headed arrows that are pointing to *single* variables in the path diagram. In the first specification, you specify free parameters theta01–theta07 for the error variances of the seven rating variables. In the next two specifications, you specify free parameters theta08 and theta09 for the error variances of Spend02 and Spend03. In the last specification, you specify fixed values 1.0 for the variances of the exogenous variables Service and Product. This fixes (or identifies) the scales of the latent variables.

In the PCOV statement, you specify the covariance parameters: either error covariances or covariances between exogenous variables. These covariance parameters correspond to those double-headed arrows that are pointing to pairs of *distinct* variables in the path diagram. In the current example, there is only one covariance parameter phi (ϕ in the path diagram) for the covariance between exogenous variables Service and Product.

In the MEAN statement, you specify the means or intercepts of the variables. Unlike some other representation schemes proposed in the field, the mean parameters are not depicted in the path diagram shown in Figure 8. The reason is that representing the mean and intercept parameters in the path diagram might obscure the “causal” paths, which are of primary interest. In addition, it is a simple matter to specify the mean and intercept parameters in the MEAN statement without the help of a path diagram if you follow these principles:

- Each variable in the path diagram has an associated mean parameter that you can specify in the MEAN statement.
- For exogenous variables, the specifications in the MEAN statement are for the means of these variables.
- For endogenous variables, the specifications in the MEAN statement are for the intercepts of these variables.
- For variables that are not specified in the MEAN statement, their means or intercepts are zero by default.
- The total number of free parameters for means or intercepts should not exceed the number of observed variables.

In the model, there are nine observed variables. Because all of them are endogenous, you specify intercept01–intercept07, InterSpend02, and InterSpend03 for their intercepts in the MEAN statement. The remaining two variables in the model are Service and Product. Both of them are exogenous, and therefore you specify their mean parameters in the MEAN statement. In this example, these means are zeros, which is necessary for fixing the locations of the latent variables.

A RESTRICTIVE MODEL WITH INVARIANT MEAN AND COVARIANCE STRUCTURES

Consider a very restrictive structural model for the two groups (regions). That is, the two groups are fitted by exactly the same model with the same set of parameter estimates, as shown in the following specification:

```
data region1(type=cov);
   input _type_ $6. _name_ $12. Spend02 Spend03 Courtesy Responsive
         Helpful Delivery Pricing Availability Quality;
   datalines;
COV    Spend02      14.428   2.206   0.439 0.520 0.459 0.498 0.635 0.642 0.769
COV    Spend03       2.206  14.178   0.540 0.665 0.560 0.622 0.535 0.588 0.715
COV    Courtesy      0.439   0.540   1.642 0.541 0.473 0.506 0.109 0.120 0.126
COV    Responsive    0.520   0.665   0.541 2.977 0.582 0.629 0.119 0.253 0.184
COV    Helpful       0.459   0.560   0.473 0.582 2.801 0.546 0.113 0.121 0.139
COV    Delivery      0.498   0.622   0.506 0.629 0.546 3.830 0.120 0.132 0.145
COV    Pricing       0.635   0.535   0.109 0.119 0.113 0.120 2.152 0.491 0.538
COV    Availability  0.642   0.588   0.120 0.253 0.121 0.132 0.491 2.372 0.589
COV    Quality       0.769   0.715   0.126 0.184 0.139 0.145 0.538 0.589 2.753
MEAN      .        183.500 301.921 4.312 4.724 3.921 4.357 6.144 4.994 5.971
;

data region2(type=cov);
   input _type_ $6. _name_ $12. Spend02 Spend03 Courtesy Responsive
         Helpful Delivery Pricing Availability Quality;
   datalines;
COV    Spend02       14.489    2.193 0.442 0.541 0.469 0.508 0.637 0.675 0.769
COV    Spend03        2.193   14.168 0.542 0.663 0.574 0.623 0.607 0.642 0.732
COV    Courtesy       0.442    0.542 3.282 0.883 0.477 0.120 0.248 0.283 0.387
COV    Responsive     0.541    0.663 0.883 2.717 0.477 0.601 0.421 0.104 0.105
COV    Helpful        0.469    0.574 0.477 0.477 2.018 0.507 0.187 0.162 0.205
COV    Delivery       0.508    0.623 0.120 0.601 0.507 2.999 0.179 0.334 0.099
COV    Pricing        0.637    0.607 0.248 0.421 0.187 0.179 2.512 0.477 0.423
COV    Availability   0.675    0.642 0.283 0.104 0.162 0.334 0.477 2.085 0.675
COV    Quality        0.769    0.732 0.387 0.105 0.205 0.099 0.423 0.675 2.698
MEAN      .         156.250 313.670 2.412 2.727 5.224 6.376 7.147 3.233 5.119
;
proc tcalis maxiter=1000 omethod=nrr;
   group 1 / data=region1 label="Region 1" nobs=378;
   group 2 / data=region2 label="Region 2" nobs=423;
   model 1 / group=1,2;
      %BasePathModel
   run;
```

In the PROC TCALIS specification, you use the GROUP statements to specify the data for the two regions. With the DATA= options in the GROUP statements, you assign the Region 1 data to group 1 and the Region 2 data to group 2. You also label the two groups by the LABEL= options. Because the numbers of observations are not specified in the data sets, you use the NOBS= options in the GROUP statements to provide this information for the two groups.

In the MODEL statement, you use the GROUP= option to specify which groups are fitted by the associated model. For the current restrictive model, both group 1 and group 2 are fitted by the same model—model 1. Next, the BasePathModel macro is included for defining model 1, the model for the two groups. Again, you could have inserted all the PATH model specifications there instead of the macro. But you use the macro to show the organization of the specification more clearly.

In the PROC TCALIS statement, the option OMETHOD=NRR (Newton-Raphson ridge optimization) is used because the default Levenberg-Marquardt optimization did not converge in 1000 iterations. Also, the MAXITER=1000 option is used for allowing enough iterations to get to a converged solution.

The fit summary table is presented in Figure 9.

Figure 9 Fit Summary of the Restrictive Multiple-Group Model about Purchase Behavior

Fit Summary

Modeling Info	N Observations	801
	N Variables	9
	N Moments	108
	N Parameters	31
	N Active Constraints	0
	Independence Model Chi-Square	399.7468
	Independence Model Chi-Square DF	72
Absolute Index	Fit Function	3.5310
	Chi-Square	2821.2798
	Chi-Square DF	77
	Pr > Chi-Square	0.0000
	Z-Test of Wilson & Hilferty	43.2651
	Hoelter Critical N	29
	Root Mean Square Residual (RMSR)	28.2211
	Standardized RMSR (SRMSR)	2.1370
	Goodness of Fit Index (GFI)	0.9996
Parsimony Index	Adjusted GFI (AGFI)	0.9995
	Parsimonious GFI	1.0690
	RMSEA Estimate	0.2987
	RMSEA Lower 90% Confidence Limit	0.2893
	RMSEA Upper 90% Confidence Limit	0.3082
	Probability of Close Fit	.
	Akaike Information Criterion	2667.2798
	Bozdogan CAIC	2229.4685
	Schwarz Bayesian Criterion	2306.4685
	McDonald Centrality	0.1803
Incremental Index	Bentler Comparative Fit Index	-7.3732
	Bentler-Bonett NFI	-6.0577
	Bentler-Bonett Non-normed Index	-6.8295
	Bollen Normed Index Rho1	-5.5994
	Bollen Non-normed Index Delta2	-7.5029
	James et al. Parsimonious NFI	-6.4783

The model chi-square statistic is 2821.28. With df=77 and $p < 0.0001$, the null hypothesis for the mean and covariance structures is rejected. All incremental fit indices are negative, indicating a very bad model fit as compared with the independence model. The RMSEA for the structural model is 0.2987, which also indicates a bad model fit. However, GFI, AGFI, and parsimonious GFI indicate good model fit, which is a little surprising given the fact that all other indices indicate the opposite and the overall model is pretty restrictive in the first place.

A MODEL WITH UNCONSTRAINED PARAMETERS FOR THE TWO REGIONS

With all the bad model fit indications, it is easy to conclude that an overly restricted model has been fit. Region 1 and Region 2 might not share exactly the same set of parameters. How about fitting a model at the other extreme with all parameters unconstrained for the two regions? Such a model can be easily specified, as shown in the following statements:

```
proc tcalis omethod=nrr;
   group 1 / data=region1 label="Region 1" nobs=378;
   group 2 / data=region2 label="Region 2" nobs=423;
   model 1 / groups=1;
      %BasePathModel
   model 2 / groups=2;
      refmodel 1/ AllNewParms;
   run;
```

In the current specification, unlike the previous specification, group 2 is now fitted by a new model designated as model 2 in the MODEL statement. This model is based on model 1, as indicated by the model number in the REFMODEL statement. The ALLNEWPARMS option in the REFMODEL statement renames all of the parameters specified in model 1 so that they become new parameters in model 2. This results in different sets of parameter estimates for model 1 and model 2, although the two models have the same path structure and comparable sets of parameters.

Figure 10 displays the fit summary of this unconstrained multiple-group model. The chi-square statistic is 29.61 (df=46, p=0.97). The theoretical model is not rejected. Many other fit indices also indicate very good model fit. For example, GFI, AGFI, Bentler CFI, Bentler-Bonett NFI, and Bollen nonnormed index delta2 are all close to one, and RMSEA is close to zero.

Figure 10 Fit Summary of the Unconstrained Multiple-Group Model about Purchase Behavior

Fit Summary

Modeling Info	N Observations	801
	N Variables	9
	N Moments	108
	N Parameters	62
	N Active Constraints	0
	Independence Model Chi-Square	399.7468
	Independence Model Chi-Square DF	72
Absolute Index	Fit Function	0.0371
	Chi-Square	29.6131
	Chi-Square DF	46
	Pr > Chi-Square	0.9710
	Z-Test of Wilson & Hilferty	-1.8950
	Hoelter Critical N	1697
	Root Mean Square Residual (RMSR)	0.0670
	Standardized RMSR (SRMSR)	0.0220
	Goodness of Fit Index (GFI)	1.0000
Parsimony Index	Adjusted GFI (AGFI)	1.0000
	Parsimonious GFI	0.6389
	RMSEA Estimate	0.0000
	RMSEA Lower 90% Confidence Limit	.
	RMSEA Upper 90% Confidence Limit	.
	Probability of Close Fit	1.0000
	Akaike Information Criterion	-62.3869
	Bozdogan CAIC	-323.9365
	Schwarz Bayesian Criterion	-277.9365
	McDonald Centrality	1.0103
Incremental Index	Bentler Comparative Fit Index	1.0000
	Bentler-Bonett NFI	0.9259
	Bentler-Bonett Non-normed Index	1.0783
	Bollen Normed Index Rho1	0.8840
	Bollen Non-normed Index Delta2	1.0463
	James et al. Parsimonious NFI	0.5916

Notice that because there are no constraints between the two models for the groups, you could have fit the two sets of data by the respective models separately and obtained exactly the same results as in the current analysis. You get two model fit chi-square values from separate analyses. Adding up these two chi-squares gives you the same overall chi-square as shown in Figure 10.

Despite a very good fit, the current model is not intended to be the final model. It was fitted mainly for showing how you can define a new model by using the REFMODEL statement and its options. For multiple-group analysis, cross-group constraints are of primary interest and should be explored whenever appropriate. Whereas the first fitting with a single model for the two groups has been shown to be too restrictive, the current fitting with no cross-group constraints fits too well—so well that it might have overfit unnecessarily. In addition, the fact that the two groups are fitted by two completely distinct models without constraints renders the multiple-group analysis trivial, if not totally useless. A multiple-group model between these extremes is explored in the next section.

A MODEL WITH CONSTRAINED COVARIANCE PARAMETERS ONLY

The following step fits a multiple-group structural equation model with cross-group (cross-model) constraints applied to the covariance parameters only:

```
proc tcalis omethod=nrr;
   group 1 / data=region1 label="Region 1" nobs=378;
   group 2 / data=region2 label="Region 2" nobs=423;
   model 1 / groups=1;
      %BasePathModel
   model 2 / groups=2;
      refmodel 1;
      mean
         Courtesy Responsive Helpful
         Delivery Pricing
         Availability Quality = G2_intercept01-G2_intercept07,
         Spend02 Spend03 = G2_InterSpend02 G2_InterSpend03;
run;
```

In this specification, Region 1 is fitted by model 1, which is essentially unchanged from the previous analyses. Region 2 is fitted by model 2, which, again, is modified from model 1, as specified in the REFMODEL statement. In the MEAN statement, nine new mean parameters are specified again for model 2. All these parameters are prefixed with G2_, to be distinguished from the parameters in model 1. Parameters G2_intercept01–G2_intercept07 are intercepts for the seven indicator measures. Parameters G2_InterSpend02 and G2_InterSpend03 are intercepts for variables Spend02 and Spend03, respectively. Because the latent variables in the model have fixed zero means, the intercept parameters G2_intercept01–G2_intercept07 and G2_InterSpend02 are also the means for the corresponding variables. To summarize, covariance structure parameters for the two regions are completely constrained in this multiple-group model, while the means structure parameters are allowed to be different for the two groups.

Figure 11 shows the fit summary of this multiple-group model.

Figure 11 Fit Summary of the Multiple-Group Model with Constrained Covariance Structures Only

Fit Summary

Modeling Info	N Observations	801
	N Variables	9
	N Moments	108
	N Parameters	40
	N Active Constraints	0
	Independence Model Chi-Square	399.7468
	Independence Model Chi-Square DF	72
Absolute Index	Fit Function	0.1346
	Chi-Square	107.5461
	Chi-Square DF	68
	Pr > Chi-Square	0.0016
	Z-Test of Wilson & Hilferty	2.9452
	Hoelter Critical N	657
	Root Mean Square Residual (RMSR)	0.1577
	Standardized RMSR (SRMSR)	0.0678
	Goodness of Fit Index (GFI)	1.0000
Parsimony Index	Adjusted GFI (AGFI)	0.9999
	Parsimonious GFI	0.9444
	RMSEA Estimate	0.0382
	RMSEA Lower 90% Confidence Limit	0.0237
	RMSEA Upper 90% Confidence Limit	0.0514
	Probability of Close Fit	0.9275
	Akaike Information Criterion	-28.4539
	Bozdogan CAIC	-415.0924
	Schwarz Bayesian Criterion	-347.0924
	McDonald Centrality	0.9756
Incremental Index	Bentler Comparative Fit Index	0.8793
	Bentler-Bonett NFI	0.7310
	Bentler-Bonett Non-normed Index	0.8722
	Bollen Normed Index Rho1	0.7151
	Bollen Non-normed Index Delta2	0.8808
	James et al. Parsimonious NFI	0.6904

The chi-square value is 107.55 (df=68, p=0.0016), which is statistically significant. The null hypothesis of the mean and covariance structures is rejected if the α-level is chosen at 0.01 or larger. However, in practical structural equation modeling, the chi-square test is not the only criterion, or even an important criterion, for evaluating model fit. Some other fit indices are used more often for gauging model fit. For example, the RMSEA estimate for the current model is 0.0382, which indicates a good fit. The probability level of close fit is 0.9969, indicating that a good population fit (RMSEA < 0.05) hypothesis cannot be rejected. The GFI, AGFI, and parsimonious GFI all indicate good model fit. However, the incremental indices show only respectable model fit.

PROC TCALIS also provides a table for comparing relative model fit of the groups. Figure 12 shows basic modeling information and some measures of fit for the two groups along with the corresponding overall measures.

Figure 12 Fit Comparison among Groups

Fit Comparison Among Groups

		Overall	Region 1	Region 2
Modeling Info	N Observations	801	378	423
	N Variables	9	9	9
	N Moments	108	54	54
	N Parameters	40	31	31
	N Active Constraints	0	0	0
	Independence Model Chi-Square	399.7468	173.4482	226.2986
	Independence Model Chi-Square DF	72	36	36
Fit Index	Fit Function	0.1346	0.1261	0.1422
	Percent Contribution to Chi-Square	100	44	56
	Root Mean Square Residual (RMSR)	0.1577	0.1552	0.1599
	Standardized RMSR (SRMSR)	0.0678	0.0792	0.0557
	Goodness of Fit Index (GFI)	1.0000	1.0000	1.0000
	Bentler-Bonett NFI	0.7310	0.7260	0.7348

Note that generally the group statistics are not independent. Therefore, when you compare the group fits by using the statistics in Figure 12, you should treat those as descriptive measures only. Looking at the percentage contribution to the chi-square, the Region 2 shows a worse fit. However, this might be due to the larger sample size in Region 2. When you compare the fit of the two regions by using the RMSR (which does not take the sample size into account), you see that the fits of the two groups are about the same. The standardized RMSR even shows that Region 2 is fitted better. Therefore, it is safe to conclude that the models fit almost equally well (or badly) for the two regions.

Constrained parameter estimates for the two regions are shown in Figure 13.

Figure 13 Estimates of Path Coefficients and Other Covariance Parameters

Model 1. PATH List

-----------Path------------			Parameter	Estimate	Standard Error	t Value
Service	->	Spend02	a1	0.37475	0.21318	1.75795
Service	->	Spend03	a2	0.53851	0.20840	2.58401
Product	->	Spend02	a3	0.80372	0.21939	3.66347
Product	->	Spend03	a4	0.59879	0.22144	2.70409
Spend02	->	Spend03	a5	0.08952	0.03694	2.42326
Service	->	Courtesy	b1	0.72418	0.07989	9.06482
Service	->	Responsive	b2	0.90452	0.08886	10.17972
Service	->	Helpful	b3	0.64969	0.07683	8.45574
Service	->	Delivery	b4	0.64473	0.09021	7.14677
Product	->	Pricing	b5	0.63452	0.07916	8.01600
Product	->	Availability	b6	0.76737	0.08265	9.28516
Product	->	Quality	b7	0.79716	0.08922	8.93470

Figure 13 *continued*

Model 1. Variance Parameters

Variance Type	Variable	Parameter	Estimate	Standard Error	t Value
Error	Courtesy	theta01	1.98374	0.13169	15.06379
	Responsive	theta02	2.02152	0.16159	12.51005
	Helpful	theta03	1.96535	0.12263	16.02727
	Delivery	theta04	2.97542	0.17049	17.45184
	Pricing	theta05	1.93952	0.12326	15.73583
	Availability	theta06	1.63156	0.13067	12.48646
	Quality	theta07	2.08849	0.15329	13.62464
	Spend02	theta08	13.47066	0.71842	18.75051
	Spend03	theta09	13.02883	0.68682	18.96966
Exogenous	Service		1.00000		
	Product		1.00000		

Model 1. Covariances Among Exogenous Variables

Var1	Var2	Parameter	Estimate	Standard Error	t Value
Service	Product	phi	0.33725	0.07061	4.77599

All parameter estimates but one are statistically significant at α=0.05. The parameter a1, which represents the path coefficient from Service to Spend02, has a t value of 1.76. This is only marginally significant. Although all these results bear the title of "Model 1," these estimates are the same for "Model 2," for which the corresponding results are not shown here.

The mean and intercept parameters for the two models (regions) are shown in Figure 14 and Figure 15.

Figure 14 Estimates of Means and Intercepts for Model 1

Model 1. Means and Intercepts

Type	Variable	Parameter	Estimate	Standard Error	t Value
Intercept	Courtesy	intercept01	4.31200	0.08157	52.86519
	Responsive	intercept02	4.72400	0.08679	54.43096
	Helpful	intercept03	3.92100	0.07958	49.27201
	Delivery	intercept04	4.35700	0.09484	45.93968
	Pricing	intercept05	6.14400	0.07882	77.94992
	Availability	intercept06	4.99400	0.07674	65.07315
	Quality	intercept07	5.97100	0.08500	70.24543
	Spend02	InterSpend02	183.50000	0.19585	936.95628
	Spend03	InterSpend03	285.49480	6.78127	42.10048
Mean	Service		0		
	Product		0		

Figure 15 Estimates of Means and Intercepts for Model 2

Model 2. Means and Intercepts

Type	Variable	Parameter	Estimate	Standard Error	t Value
Intercept	Courtesy	G2_intercept01	2.41200	0.07709	31.28628
	Responsive	G2_intercept02	2.72700	0.08203	33.24350
	Helpful	G2_intercept03	5.22400	0.07522	69.45319
	Delivery	G2_intercept04	6.37600	0.08964	71.12697
	Pricing	G2_intercept05	7.14700	0.07450	95.93427
	Availability	G2_intercept06	3.23300	0.07254	44.57020
	Quality	G2_intercept07	5.11900	0.08034	63.71500
	Spend02	G2_InterSpend02	156.25000	0.18511	844.09015
	Spend03	G2_InterSpend03	299.68311	5.77478	51.89515
Mean	Service		0		
	Product		0		

All the mean and intercept estimates are statistically significant at α=0.01. Except for the fixed zero means for Service and Product, these mean and intercepts estimates show a different pattern for the two models. Do these estimates truly differ beyond chance? To tackle this problem, simultaneous and individual tests of the pairs of estimates in the two models are conducted in the next section.

TESTING A PRIORI HYPOTHESES INDIVIDUALLY OR SIMULTANEOUSLY

PROC TCALIS enables you to test a priori hypotheses about any parametric functions, either individually or simultaneously. These hypotheses take the following form:

$$H_0 : t_i = 0 \quad (i = 1 \ldots k)$$

where t_i is the ith parametric function (either linear or nonlinear) and k is the number of hypotheses.

For example, in the current analysis you might want to test whether each pair of intercepts for the spending in 2002 and 2003 is the same for the two models (groups). You set up the following hypotheses:

$$\begin{aligned} H_0 : \text{G2\_InterSpend02} - \text{InterSpend02} &= 0 \\ H_0 : \text{G2\_InterSpend03} - \text{InterSpend03} &= 0 \end{aligned}$$

Similarly, if you want to test whether each pair of the seven measurement intercepts is the same for the two models (groups), you can set up the following hypotheses:

$$\begin{aligned} H_0 : \text{G2\_intercept01} - \text{intercept01} &= 0 \\ H_0 : \text{G2\_intercept02} - \text{intercept02} &= 0 \\ H_0 : \text{G2\_intercept03} - \text{intercept03} &= 0 \\ H_0 : \text{G2\_intercept04} - \text{intercept04} &= 0 \\ H_0 : \text{G2\_intercept05} - \text{intercept05} &= 0 \\ H_0 : \text{G2\_intercept06} - \text{intercept06} &= 0 \\ H_0 : \text{G2\_intercept07} - \text{intercept07} &= 0 \end{aligned}$$

You can use the new TESTFUNC or SIMTEST statement in PROC TCALIS to test these a priori hypotheses about parametric functions, . You might also need to use the SAS programming statements to define the parametric functions. This is illustrated in the following statements:

```
simtest
   SpendDiff       = (Spend02Diff Spend03Diff)
   MeasurementDiff = (CourtesyDiff ResponsiveDiff
                      HelpfulDiff DeliveryDiff
                      PricingDiff AvailabilityDiff
                      QualityDiff);

/* SAS Programming Statements for Defining the Parametric Functions */
Spend02Diff      = G2_InterSpend02 - InterSpend02;
Spend03Diff      = G2_InterSpend03 - InterSpend03;
CourtesyDiff     = G2_intercept01  - intercept01;
ResponsiveDiff   = G2_intercept02  - intercept02;
HelpfulDiff      = G2_intercept03  - intercept03;
DeliveryDiff     = G2_intercept04  - intercept04;
PricingDiff      = G2_intercept05  - intercept05;
AvailabilityDiff = G2_intercept06  - intercept06;
QualityDiff      = G2_intercept07  - intercept07;
```

The SIMTEST statement can test parametric functions simultaneously, so it is chosen over the TESTFUNC statement in this example. In the SIMTEST statement, two simultaneous hypotheses are tested. One is labeled SpendDiff, which includes two parametric functions Spend02Diff and Spend03Diff; the other is labeled MeasurementDiff, which includes seven parametric functions: CourtesyDiff, ResponsiveDiff, HelpfulDiff, DeliveryDiff, PricingDiff, AvailabilityDiff, and QualityDiff.

All these parametric functions are then defined by the SAS programming statements as the differences between the pairs of the parameters in the two models.

Inserting these statements into your preceding PROC TCALIS step, you obtain the output for the simultaneous tests as shown in Figure 16.

Figure 16 Simultaneous Tests of Differences in Means and Intercepts

Simultaneous Tests

Simultaneous Test	Parametric Function	Function Value	DF	Chi-Square
SpendDiff			2	10458
	Spend02Diff	-27.25000	1	10225
	Spend03Diff	14.18831	1	185.86725
MeasurementDiff			7	1610
	CourtesyDiff	-1.90000	1	286.58605
	ResponsiveDiff	-1.99700	1	279.63659
	HelpfulDiff	1.30300	1	141.59942
	DeliveryDiff	2.01900	1	239.35318
	PricingDiff	1.00300	1	85.52567
	AvailabilityDiff	-1.76100	1	278.09360
	QualityDiff	-0.85200	1	53.06240

In Figure 16, the exceedingly large chi-square value 10458 for the first simultaneous test suggests that the null hypothesis should be rejected. The two groups differ in their intercepts for the spending in 2002 and 2003. Individual tests for these intercepts suggest that each of the individual hypotheses should be rejected. The chi-square values for individual tests are 10225 and 185.87, respectively.

Similarly, the simultaneous and individual tests of the measurement intercepts suggest that the two models (groups) differ significantly in the intercepts (or means in the current case because the associated latent factors have zero means) of the measured variables. Region 2 has significantly higher means in variables Helpful, Delivery, and Pricing, but significantly lower means in variables Courtesy, Responsive, Availability, and Quality.

CONCLUSION OF THE MULTIPLE-GROUP ANALYSIS OF PURCHASE BEHAVIOR

Based on the estimation results of the current model, you are now ready to answer the main research questions. The overall customer service (Service) does affect purchases in 2003 (Spend03). However, it might not affect purchases in 2002 (Spend02) because the corresponding path effect a1 is only marginally significant. This lack of significant relationship might be due to the fact that the customer service ratings were done after the purchases in 2002. That is, purchases in 2002 had been completed before the impression about customer service was fully formed. However, this argument cannot explain why overall product quality (Product) shows a strong and significant relationship with purchases in 2002 (Spend02). Nonetheless, customer service and product quality do affect the purchases in 2003 (Spend03) in an expected way, even after partialling out the effect of the previous purchase amount (Spend02). Apart from the mean differences of the variables, the common measurement and prediction (or structural) models fit the two regions very well.

CONCLUSION

In this paper, several important features of PROC TCALIS are illustrated with practical applications. With the PATH modeling language, you can specify structural equation models by using a syntax that is closely related to the path-diagram representations. You can use the PATH, PVAR, and PCOV statements to specify the paths and their effects, the variance parameters, and the covariance parameters, respectively. With the EFFPART statement, you can analyze customized sets of important effects in your model. With the FITINDEX statement, you can customize your fit summary table to report only those desirable fit indices and modeling information. Multiple-group structural equation modeling is supported in PROC TCALIS. With the REFMODEL and its associated options, you can define models by referencing and by renaming parameters. You can also analyze mean structures in PROC TCALIS. For example, for the PATH modeling language you can use the MEAN statement to specify the intercept or mean parameters. PROC TCALIS also enables you to test a priori hypotheses about linear or nonlinear parametric functions by using the SIMTEST or TESTFUNC statement.

Not all new features of PROC TCALIS have been illustrated in this paper, however. Those more important new features and enhancements not mentioned in this paper are summarized as follows:

- The PATH modeling language is only one of the modeling languages supported by PROC TCALIS. Other modeling languages include:
  - LINEQS—a model specification method based on specifying equations, as proposed by Bentler (1995)
  - RAM—a model specification method based on the RAM model proposed by McArdle (see, for example, McArdle and McDonald 1984)
  - FACTOR—confirmatory or exploratory factor models with simplified syntax
  - MSTRUCT—a matrix model specification method that is most suitable for direct testing of matrix structures
  - LISMOD—a matrix model specification method that is based on the LISREL model (see, for example, Jöreskog and Sörbom 1988)
- Standard error estimates are computed for standardized estimates and standardized effects.
- You can customize the sets of parameters for ranking the Lagrange multiplier statistics. This would be useful if a target set of parameters is being probed for model improvements.

REFERENCES

Bentler, P. M. (1995), *EQS, Structural Equations Program Manual*, Program Version 5.0, Encino, CA: Multivariate Software.

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 K. A. Bollen and S. Long, eds., *Testing Structural Equation Models*, Newbury Park, CA: Sage Publications.

Jöreskog, K. G. (1971), "Simultaneous Factor Analysis in Several Populations," *Psychometrika*, 36, 409–426.

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

Loehlin, J. C. (2004), *Latent Variable Models: An Introduction to Factor, Path, and Structural Analysis*, Fourth Edition, Mahway, NJ: Lawrence Erlbaum Associates.

Marjoribanks, K., ed. (1974), *Environments for Learning*, London: National Foundation for Educational Research Publications.

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.

CONTACT INFORMATION

Yiu-Fai Yung
SAS Institute Inc.
SAS Campus Drive
Cary, NC 27513
Phone: 919-531-4032
Email: yiu-fai.yung@sas.com