Introduction to Structural Equation Modeling

Course Notes

Introduction to Structural Equation Modeling Course Notes was developed by Werner Wothke, Ph.D., of the American Institute of Research. Additional contributions were made by Bob Lucas and Paul Marovich. Editing and production support was provided by the Curriculum Development and Support Department.

Introduction to Structural Equation Modeling Course Notes

Book code E1565, course code LWBAWW, prepared date 20May2010. LWBAWW_003

ISBN 978-1-60764-253-4

Table of Contents

Course Description

This lecture focuses on structural equation modeling (SEM), a statistical technique that combines elements of traditional multivariate models, such as regression analysis, factor analysis, and simultaneous equation modeling. SEM can explicitly account for less than perfect reliability of the observed variables, providing analyses of attenuation and estimation bias due to measurement error. The SEM approach is sometimes also called causal modeling because competing models can be postulated about the data and tested against each other. Many applications of SEM can be found in the social sciences, where measurement error and uncertain causal conditions are commonly encountered. This presentation demonstrates the structural equation modeling approach with several sets of empirical textbook data. The final example demonstrates a more sophisticated re-analysis of one of the earlier data sets.

To learn more…

For information on other courses in the curriculum, contact the SAS Education Division at 1-800-333-7660, or send e-mail to training@sas.com. You can also find this information on the Web at support.sas.com/training/ as well as in the Training Course Catalog.

For a list of other SAS books that relate to the topics covered in this Course Notes, USA customers can contact our SAS Publishing Department at 1-800-727-3228 or send e-mail to sasbook@sas.com. Customers outside the USA, please contact your local SAS office.

Also, see the Publications Catalog on the Web at support.sas.com/pubs for a complete list of books and a convenient order form.

Prerequisites

Before attending this course, you should be familiar with using regression analysis, factor analysis, or both.

Chapter 1 Introduction to Structural Equation Modeling

1.1 Introduction

Course Outline

1. Welcome to the Webcast
2. Structural Equation Modeling – Overview
3. Two Easy Examples
   a. Regression Analysis
   b. Factor Analysis
4. Confirmatory Models and Assessing Fit
5. More Advanced Examples
   a. Structural Equation Model (Incl. Measurement Model)
   b. Effects of Errors-in-Measurement on Regression
6. Conclusion

3

The course presents several examples of what kind of interesting analyses we can perform with structural equation modeling. For each example, the course demonstrates how the analysis can be implemented with PROC CALIS.

1.01 Multiple Choice Poll

What experience have you had with structural equation modeling (SEM) so far?

a. No experience with SEM
b. Beginner
c. Occasional applied user
d. Experienced applied user
e. SEM textbook writer and/or software developer
f. Other

5

1.02 Multiple Choice Poll

How familiar are you with linear regression and factor analysis?

a. Never heard of either.

b. Learned about regression in statistics class.

c. Use linear regression at least once per year with real data.

d. Use factor analysis at least once per year with real data.

e. Use both regression and factor analysis techniques frequently.

6

1.03 Poll

Have you used PROC CALIS before?

❍ Yes

❍ No

7

1.04 Multiple Choice Poll

Please indicate your main learning objective for this structural equation modeling course.

a. I am curious about SEM and want to find out what it can be used for.
b. I want to learn to use PROC CALIS.
c. My advisor requires that I use SEM for my thesis work.
d. I want to use SEM to analyze applied marketing data.
e. I have some other complex multivariate data to model.
f. What is this latent variable stuff good for anyways?
g. Other.

8

1.2 Structural Equation Modeling—Overview

What Is Structural Equation Modeling?

SEM = General approach to multivariate data analysis!

- aka, Analysis of Covariance Structures,
- aka, Causal Modeling,
- aka, LISREL Modeling.

Purpose: Study complex relationships among variables, where some variables can be hypothetical or unobserved.

Approach: SEM is model based. We try one or more competing models – SEM analytics show which ones fit, where there are redundancies, and can help pinpoint what particular model aspects are in conflict with the data.

Difficulty: Modern SEM software is easy to use. Nonstatisticians can now solve estimation and testing problems that once would have required the services of several specialists.

10

SEM – Some Origins

- Psychology – Factor Analysis:
  Spearman (1904), Thurstone (1935, 1947)
- Human Genetics – Regression Analysis:
  Galton (1889)
- Biology – Path Modeling:
  S. Wright (1934)
- Economics – Simultaneous Equation Modeling:
  Haavelmo (1943), Koopmans (1953), Wold (1954)
- Statistics – Method of Maximum Likelihood Estimation:
  R.A. Fisher (1921), Lawley (1940)
- Synthesis into Modern SEM and Factor Analysis:
  Jöreskog (1970), Lawley & Maxwell (1971), Goldberger & Duncan (1973)

11

Common Terms in SEM

Types of Variables:

Measured, Observed, Manifest ← Variable present in the data file and not missing

versus

Hypothetical, Unobserved, Latent ← Not in data file

Endogenous Variable—Exogenous Variable (in SEM)

but

Dependent Variable—Independent Variable (in Regression)

Where are the variables in the model?

$z \leftarrow y \leftarrow x$

12

1.05 Multiple Choice Poll

An endogenous variable is

a. the dependent variable in at least one of the model equations

b. the terminating (final) variable in a chain of predictions

c. a variable in the middle of a chain of predictions

d. a variable used to predict other variables

e. I'm not sure.

14

1.06 Multiple Choice Poll

A manifest variable is

a. a variable with actual observed data
b. a variable that can be measured (at least in principle)
c. a hypothetical variable
d. a predictor in a regression equation
e. the dependent variable of a regression equation
f. I'm not sure.

16

1.3 Example 1: Regression Analysis

Example 1: Multiple Regression

A. Application: Predicting Job Performance of Farm Managers
B. Use summary data (covariance matrix) by Warren, White, and Fuller 1974
C. Illustrate Covariance Matrix Input with PROC REG and PROC CALIS
D. Illustrate PROC REG and PROC CALIS Parameter Estimates
E. Introduce PROC CALIS Model Specification in LINEQS Format

19

Example 1: Multiple Regression

Warren, White, and Fuller (1974) studied 98 managers of farm cooperatives. Four of the measurements made on each manager were:

- **Performance**: A 24-item test of performance related to “planning, organization, controlling, coordinating and directing.”
- **Knowledge**: A 26-item test of knowledge of “economic phases of management directed toward profit-making ... and product knowledge.”
- **ValueOrientation**: A 30-item test of “tendency to rationally evaluate means to an economic end.”
- **JobSatisfaction**: An 11-item test of “gratification obtained ... from performing the managerial role.”

A fifth measure**, PastTraining**, was reported but will not be employed in this example.

20

Warren, White, and Fuller (1974) Data

Warren 5 Variables: Summary Data, SAS Format

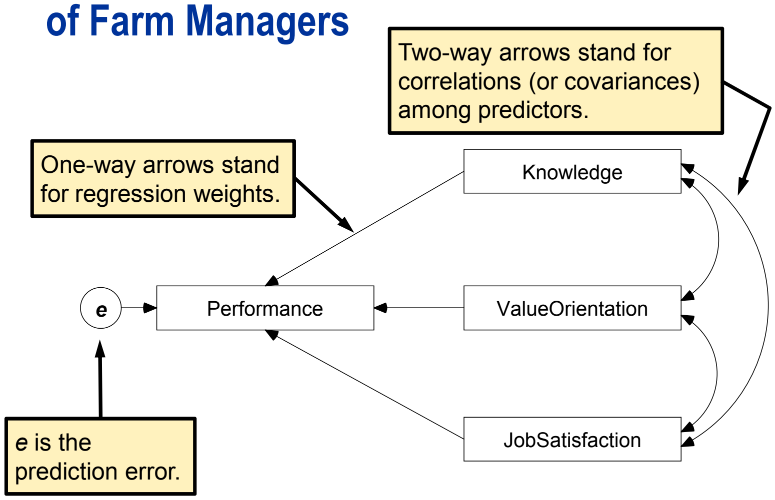
Obs	_type_	_name_	Performance	Knowledge	ValueOrientation	JobSatisfaction	PastTraining
1	n		98.0000	98.0000	98.0000	98.0000	98.0000
2	cov	Performance	0.0209	.	.	.	.
3	cov	Knowledge	0.0177	0.0520	.	.	.
4	cov	ValueOrientation	0.0245	0.0280	0.1212	.	.
5	cov	JobSatisfaction	0.0046	0.0044	-0.0063	0.0901	.
6	cov	PastTraining	0.0187	0.0192	0.0353	-0.0066	0.0946
7	mean		0.0589	1.3796	2.8773	2.4613	2.1174

This SAS file must be saved with attribute TYPE=COV.

21

This file can be found in the worked examples as **`Warren5Variables.sas7bdat`**.

Prediction Model: Job Performance of Farm Managers



22

Linear Regression Model: Using PROC REG

```
TITLE "Example 1a: Linear Regression with PROC REG";

PROC REG DATA=SEMdata.Warren5variables;
  MODEL
    Performance = Knowledge
                  ValueOrientation
                  JobSatisfaction;
RUN;
QUIT;
```

Notice the two-level filename. In order to run this code, you must first define a SAS LIBNAME reference.

PROC REG will continue to run interactively. QUIT ends PROC REG.

23

Parameter Estimates: PROC REG

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-0.83415	0.14238	-5.86	<.0001
Knowledge	1	0.25818	0.05439	4.75	<.0001
ValueOrientation	1	0.14502	0.03562	4.07	<.0001
JobSatisfaction	1	0.04859	0.03873	1.25	0.2128

Prediction Equation for Job Performance:

Performance = -0.83 +
0.26*`Knowledge` +
0.15*`ValueOrientation` +
0.05*`JobSatisfaction`, v(e) = 0.01

JobSatisfaction is **not** an important predictor of Job Performance.

24

1.07 Multiple Choice Poll

How many PROC REG subcommands are required to specify a linear regression with PROC REG?

a. None
b. 1
c. 2
d. 3
e. 4
f. More than 4

26

LINEQS Model Interface in PROC CALIS

```
PROC CALIS DATA=<input-file> <options>;
  VAR <list of variables>;
  LINEQS
    <equation>, … , <equation>;
  STD
    <variance-terms>;
  COV
    <covariance-terms>;
RUN;
```

Easy model specification with PROC CALIS – only five model components are needed.

28

The PROC CALIS statement starts the SAS procedure; the four statements VAR, LINEQS, STD, and COV are subcommands of its LINEQS interface.

PROC CALIS comes with four interfaces for specifying structural equation factor models (LINEQS, RAM, COSAN, and FACTOR). For the purpose of this introductory Webcast, the LINEQS interface is completely general and seemingly the easiest to use.

LINEQS Model Interface in PROC CALIS

```
PROC CALIS DATA=<input-file> <options>;
  VAR <list of variables>;

  LINEQS
    <equation>, … , <equation>;
  STD
    <variance-terms>;
  COV
    <covariance-terms>;
RUN;
```

The PROC CALIS statement begins the model specs. ***<inputfile>*** refers to the data file. ***<options>*** specify computational and statistical methods.

29

LINEQS Model Interface in PROC CALIS

```
PROC CALIS DATA=<input-file> <options>;
  VAR <list of variables>;

  LINEQS
    <equation>, … , <equation>;
  STD
    <variance-terms>;
  COV
    <covariance-terms>;
RUN;
```

VAR (optional statement) to select and reorder variables from *<input-file>*

30

LINEQS Model Interface in PROC CALIS

```
PROC CALIS DATA=<input-file> <options>;

  VAR <list of variables>;

  LINEQS
    <equation>, … , <equation>;
  STD
    <variance-terms>;
  COV
    <covariance-terms>;
RUN;
```

Put all model equations in the **LINEQS** section, separated by commas.

31

LINEQS Model Interface in PROC CALIS

```
PROC CALIS DATA=<input-file> <options>;

  VAR <list of variables>;

  LINEQS
    <equation>, … , <equation>;
  STD
    <variance-terms>;
  COV
    <covariance-terms>;
RUN;
```

Variances of unobserved exogenous variables to be listed here

32

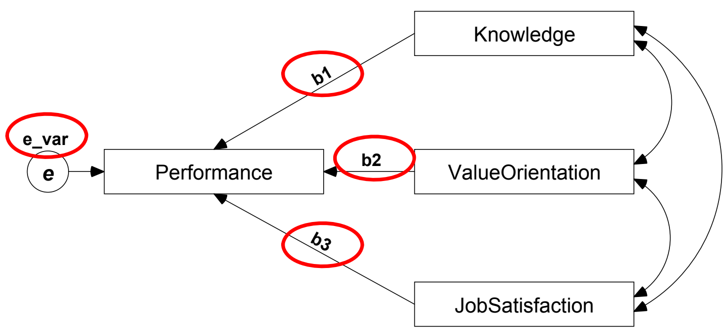
LINEQS Model Interface in PROC CALIS

```
PROC CALIS DATA=<input-file> <options>;

  VAR <list of variables>;

  LINEQS
    <equation>, … , <equation>;
  STD
    <variance-terms>;
  COV
    <covariance-terms>;
RUN;
```

Covariances of unobserved exogenous variables to be listed here

33

Prediction Model of Job Performance of Farm Managers (Parameter Labels Added)



34

In contrast to PROC REG, PROC CALIS (LINEQS) expects the regression weight and residual variance parameters to have their own unique names (b1-b3, e_var).

Linear Regression Model: Using PROC CALIS

```
TITLE "Example 1b: Linear Regression with PROC CALIS";

PROC CALIS DATA=SEMdata.Warren5variables COVARIANCE;
  VAR
    Performance Knowledge
    ValueOrientation JobSatisfaction;
  LINEQS
    Performance = b1 Knowledge +
                  b2 ValueOrientation +
                  b3 JobSatisfaction + e1;
  STD
    e1 = e_var;
RUN;
```

The **COVARIANCE** option picks covariance matrix analysis (default: correlation matrix analysis).

35

Linear Regression Model: Using PROC CALIS

```
TITLE "Example 1b: Linear Regression with PROC CALIS";

PROC CALIS DATA=SEMdata.Warren5variables COVARIANCE;
  VAR
    Performance Knowledge
    ValueOrientation JobSatisfaction;
  LINEQS
    Performance = b1 Knowledge +
                  b2 ValueOrientation +
                  b3 JobSatisfaction + e1;
  STD
    e1 = e_var;
RUN;
```

The regression model is specified in the LINEQS section. The residual term (e1) and the names (b1-b3) of the regression parameters must be given explicitly.
Convention: Residual terms of observed endogenous variables start with the letter e.

36

Linear Regression Model: Using PROC CALIS

```
TITLE "Example 1b: Linear Regression with PROC CALIS";

PROC CALIS DATA=SEMdata.Warren5variables COVARIANCE;
  VAR
    Performance Knowledge
    ValueOrientation JobSatisfaction;
  LINEQS
    Performance = b1 Knowledge +
                  b2 ValueOrientation +
                  b3 JobSatisfaction + e1;
  STD
    e1 = e_var;
RUN;
```

The name of the residual term is given on the left side of STD equation; the label of the variance parameter goes on the right side.

37

This model contains *only one* unobserved exogenous variable (e1). Thus, there a no covariance terms to model, and no COV subcommand is needed.

Parameter Estimates: PROC CALIS

This is the estimated regression equation for a deviation score model. Estimates and standard errors are identical at three decimal places to those obtained with PROC REG. The t-values (> 2) indicate that **`Performance`** is predicted by **`Knowledge`** and **`ValueOrientation`**, not **`JobSatisfaction`**.

Example 1b: Linear Regression Using PROC CALIS

The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation

Manifest Variable Equations with Estimates

```
Performance = 0.2582 * Knowledge + 0.1450 * ValueOrientation + 0.0486 * JobSatisfaction + 1.0000 e1
Std Err       0.0535   b1          0.0351   b2                 0.0381   b3
t Value       4.8223               4.1365                      1.2743
```

38

✎ The standard errors slightly differ from their OLS regression counterparts. The reason is that PROC REG gives *exact* standard errors, even in small samples, while the standard errors obtained by PROC CALIS are asymptotically correct.

Standardized Regression Estimates

The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation

Manifest Variable Equations with Standardized Estimates

Performance = 0.4072 * Knowledge + 0.3492 * ValueOrientation + 0.1009 * JobSatisfaction + 0.7750 e1
b1 b2 b3

Squared Multiple Correlations				
	Variable	Error Variance	Total Variance	R-Square
1	Performance	0.01255	0.02090	0.3994

In standard deviation terms, **`Knowledge`** and **`ValueOrientation`** contribute to the regression with similar weights. The regression equation determines 40% of the variance of **`Performance`**.

39

PROC CALIS computes and displays the standardized solution by default.

1.08 Multiple Choice Poll

How many PROC CALIS subcommands are required to specify a linear regression with PROC CALIS?

a. None
b. 1
c. 2
d. 3
e. 4
f. More than 4

41

LINEQS Defaults and Peculiarities

Some standard assumptions of linear regression analysis are built into LINEQS:

1. Observed exogenous variables (**`Knowledge`**, **`ValueOrientation`** and **`JobSatisfaction)`** are automatically assumed to be correlated with each other.
2. The error term *e1* is treated as independent of the predictor variables.

43

continued...

LINEQS Defaults and Peculiarities

Built-in differences from PROC REG:

1. The error term *e1* must be specified explicitly (CALIS convention: error terms of observed variables must start with the letter e).
2. Regression parameters (b1, b2, b3) must be named in the model specification.
3. As traditional in SEM, the LINEQS equations are for *deviation scores*, in other words, without the intercept term. PROC CALIS centers all variables automatically.
4. The order of variables in the PROC CALIS output is controlled by the **VAR** statement.
5. Model estimation is iterative.

44

Iterative Estimation Process

```
  Vector of Initial Estimates

       Parameter        Estimate     Type
 1    b1                0.25818     _GAMMA_[1:1]
 2    b2                0.14502     _GAMMA_[1:2]
 3    b3                0.04859     _GAMMA_[1:3]
 4    e_var             0.01255     _PHI_[4:4]
```

45

continued...

The iterative estimation process computes stepwise updates of provisional parameter estimates, until the fit of the model to the sample data cannot be improved any further.

Iterative Estimation Process

```
Optimization Start

Active Constraints
Objective Function
Max Abs Gradient Element              1.505748E-14
 ...
              Optimization Results
 Iterations                                      0
 ...
 Max Abs Gradient Element             1.505748E-14
 ...
 ABSGCONV convergence criterion satisfied.
```

This number should be really close to zero.

Important message, displayed in both list output and SAS log. Make sure it is there!

46

Example 1: Summary

Tasks accomplished:

1. Set up a multiple regression model with both PROC REG and PROC CALIS
2. Estimated the regression parameters both ways
3. Verified that the results were comparable
4. Inspected iterative model fitting by PROC CALIS

47

1.09 Multiple Choice Poll

Which PROC CALIS output message indicates that an iterative solution has been found?

a. Covariance Structure Analysis: Maximum Likelihood Estimation
b. Manifest Variable Equations with Estimates
c. Vector of Initial Estimates
d. ABSGCONV convergence criterion satisfied
e. None of the above
f. Not sure

49

1.4 Example 2: Factor Analysis

Example 2: Confirmatory Factor Analysis

A. Application: Studying dimensions of variation in human abilities
B. Use raw data from Holzinger and Swineford (1939)
C. Illustrate raw data input with PROC CALIS
D. Introduce latent variables
E. Introduce tests of fit
F. Introduce modification indices
G. Introduce model-based statistical testing
H. Introduce nested models

53

Factor analysis frequently serves as the measurement portion in structural equation models.

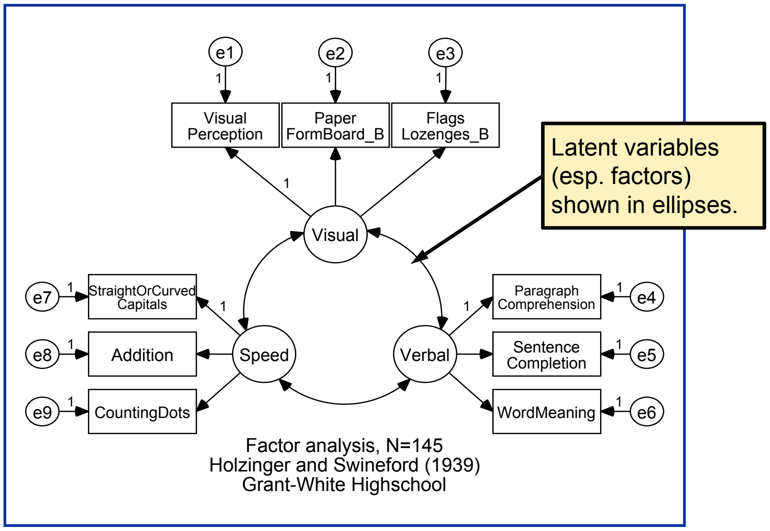
Confirmatory Factor Analysis: Model 1

Holzinger and Swineford (1939) administered 26 psychological aptitude tests to 301 seventh- and eighth-grade students in two Chicago schools. Here are the tests selected for the example and the types of abilities they were meant to measure:

Ability	Test
Visual	VisualPerception
	PaperFormBoard
	FlagsLozenges_B
Verbal	ParagraphComprehension
	SentenceCompletion
	WordMeaning
Speed	StraightOrCurvedCapitals
	Addition
	CountingDots

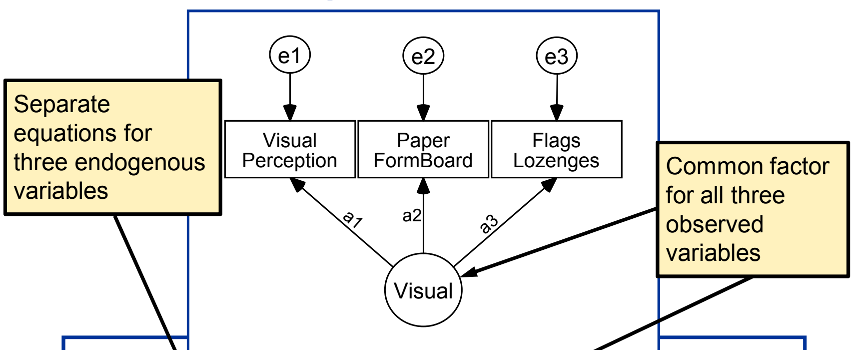
54

CFA, Path Diagram Notation: Model 1



55

Measurement Specification with LINEQS

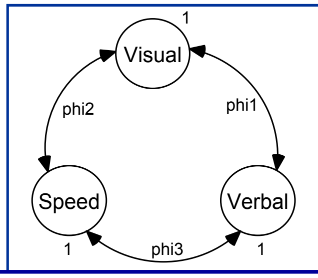



```
LINEQS
   VisualPerception = a1 F_Visual + e1,
   PaperFormBoard   = a2 F_Visual + e2,
   FlagsLozenges    = a3 F_Visual + e3,
```

Factor names start with F in PROC CALIS.

56

Specifying Factor Correlations with LINEQS



Factor variances are 1.0 for correlation matrix.

```
STD
   F_Visual F_Verbal F_Speed = 1.0 1.0 1.0,
   ...;
COV
   F_Visual F_Verbal F_Speed = phi1 phi2 phi3;
```

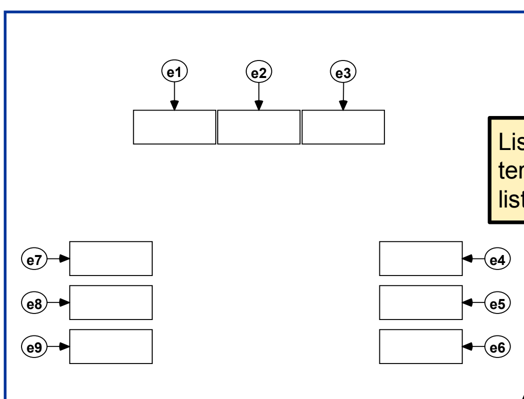
Factor correlation terms go into the COV section.

57

There is some freedom about setting the scale of the latent variable. We need to fix the scale of each somehow in order to estimate the model. Typically, this is either done by fixing one factor loading to a positive constant, or by fixing the variance of the latent variable to unity (1.0).

Here we set the variances of the latent variables to unity. Since the latent variables are thereby standardized, the phi1-phi3 parameters are now correlation terms.

Specifying Measurement Residuals



List of residual terms followed by list of variances

```
STD
  ...,
  e1 e2 e3 e4 e5 e6 e7 e8 e9 = e_var1 e_var2 e_var3
    e_var4 e_var5 e_var6 e_var7 e_var8 e_var9;
```

58

CFA, PROC CALIS/LINEQS Notation: Model 1

```
PROC CALIS DATA=SEMdata.HolzingerSwinefordGW
  COVARIANCE RESIDUAL MODIFICATION;
  VAR <...> ;
  LINEQS
    VisualPerception          = a1 F_Visual + e1,
    PaperFormBoard            = a2 F_Visual + e2,
    FlagsLozenges             = a3 F_Visual + e3,
    ParagraphComprehension    = b1 F_Verbal + e4,
    SentenceCompletion        = b2 F_Verbal + e5,
    WordMeaning               = b3 F_Verbal + e6,
    StraightOrCurvedCapitals  = c1 F_Speed  + e7,
    Addition                  = c2 F_Speed  + e8,
    CountingDots              = c3 F_Speed  + e9
  STD
    F_Visual F_Verbal F_Speed = 1.0 1.0 1.0,
    e1 e2 e3 e4 e5 e6 e7 e8 e9 = e_var1 e_var2
      e_var4 e_var5 e_var6 e_var7 e_var8 e_var9;
  COV
    F_Visual F_Verbal F_Speed = phi1 phi2 phi3;
RUN;
```

Residual statistics and modification indices

Nine ***measurement*** equations

59

1.10 Multiple Choice Poll

How many LINEQS equations are needed for a factor analysis?

a. Nine, just like the previous slide

b. One for each observed variable in the model

c. One for each factor in the model

d. One for each variance term in the model

e. None of the above

f. Not sure

61

Model Fit in SEM

The chi-square statistic is central to assessing fit with Maximum Likelihood estimation, and many other fit statistics are based on it. The standard χ^2 measure in SEM is

$$\chi^2_{\mathrm{ML}} = (N-1)\left[\mathrm{trace}\left(\mathbf{S}\,\hat{\Sigma}^{-1}\right) - p + \ln\left(\left|\hat{\Sigma}\right|\right) - \ln\left(\left|\mathbf{S}\right|\right)\right]$$

Always zero or positive. The term is zero only when the match is exact.

Here, N is the sample size, p the number of observed variables, **S** the sample covariance matrix, and $\hat{\Sigma}$ the fitted model covariance matrix.

This gives the test statistic for the null hypotheses that the predicted matrix $\hat{\Sigma}$ has the specified model structure against the alternative that $\hat{\Sigma}$ is unconstrained.

Degrees of freedom for the model:

df = number of elements in the lower half of the covariance matrix [$p(p+1)/2$] minus number of estimated parameters

63

The χ^2 statistic is a discrepancy measure. It compares the sample covariance matrix with the implied model covariance matrix computed from the model structure and all the model parameters.

Degrees of Freedom for CFA Model 1

```
From General Modeling Information Section...
The CALIS Procedure
Covariance Structure Analysis:
      Maximum Likelihood Estimation
Levenberg-Marquardt Optimization
Scaling Update of More (1978)

Parameter Estimates            21
Functions (Observations)       45
```

DF = 45 – 21
= 24

64

CALIS, CFA Model 1: Fit Table

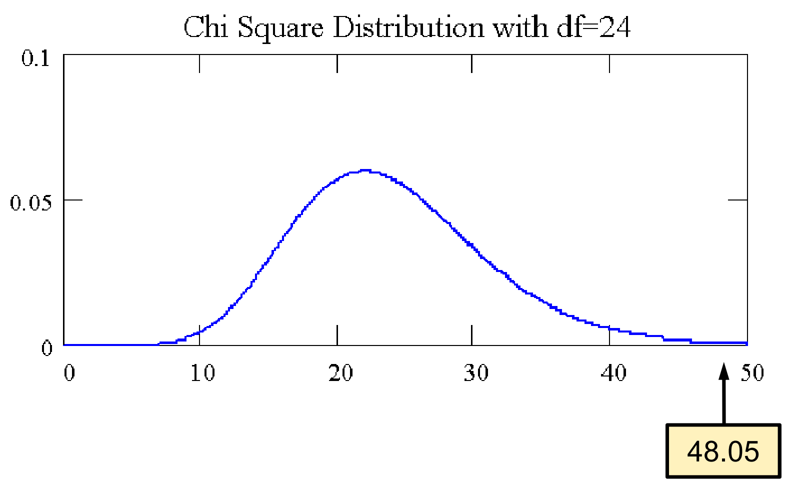
```
Fit Function                                0.3337
Goodness of Fit Index (GFI)                 0.9322
GFI Adjusted for Degrees of Freedom (AGFI)0.8729
Root Mean Square Residual (RMR)             15.9393
Parsimonious GFI (Mulaik, 1989)             0.6215
Chi-Square                                  48.0536
Chi-Square DF                               24
Pr > Chi-Square                             0.0025
Independence Model Chi-Square               502.86
Independence Model Chi-Square DF
RMSEA Estimate
RMSEA 90% Lower Confidence Limit
```

Pick out the chi-square section. This chi-square is significant. What does this mean?

…and many more fit statistics on list output.

65

Chi Square Test: Model 1



66

Standardized Residual Moments: Part 1

Asymptotically Standardized Residual Matrix

	Visual Perception	PaperForm Board	Flags Lozenges_B	Paragraph Comprehension	Sentence Completion
VisualPerc	0.000000000	-0.490645663	0.634454156	-0.376267466	-0.853201760
PaperFormB	-0.490645663	0.000000000	-0.133256120	-0.026665527	0.224463460
FlagsLozen	0.634454156	-0.133256120	0.000000000	0.505250934	0.901260142
ParagraphC	-0.376267466	-0.026665527	0.505250934	0.000000000	-0.303368250
SentenceCo	-0.853201760	0.224463460	0.901260142	-0.303368250	0.000000000
WordMeanin	-0.530010952	0.187307568	0.474116387	0.577008266	-0.268196124
StraightOr	4.098583857	2.825690487	1.450078999	1.811782623	2.670254862
Addition	-3.084483125	-1.069283994	-2.383424431	0.166892980	1.043444072
CountingDo	-0.219601213	-0.619535105	-2.101756596	-2.939679987	-0.642256508

Residual covariances, divided by their approximate standard error

67

Recall that residual statistics were requested on the PROC CALIS command line by the “RESIDUAL” keyword. In the output listing, we need to find the section on *Asymptotically Standardized Residuals*. These are fitted residuals of the covariance matrix, divided by their asymptotic standard errors, essentially *z*-values.

Standardized Residual Moments: Part 2

Asymptotically Standardized Residual Matrix

	WordMeaning	StraightOr Curved Capitals	Addition	CountingDots
VisualPerc	-0.530010952	4.098583857	-3.084483125	-0.219601213
PaperFormB	0.187307568	2.825690487	-1.069283994	-0.619535105
FlagsLozen	0.474116387	1.450078999	-2.383424431	-2.101756596
ParagraphC	0.577008266	1.811782623	0.166892980	-2.939679987
SentenceCo	-0.268196124	2.670254862	1.043444072	-0.642256508
WordMeanin	0.000000000	1.066742617	-0.196651078	-2.124940910
StraightOr	1.066742617	0.000000000	-2.695501076	-2.962213789
Addition	-0.196651078	-2.695501076	0.000000000	5.460518790
CountingDo	-2.124940910	-2.962213789	5.460518790	0.000000000

68

1.11 Multiple Choice Poll

A large chi-square fit statistic means that

a. the model fits well

b. the model fits poorly

c. I'm not sure.

70

Modification Indices (Table)

```
Univariate Tests for Constant Constrai
 Lagrange Multiplier or Wald Index
   / Probability / Approx Change of Val

                   F_Visual F_Verbal   F_Speed
 ...<snip>...
 StraightOr         30.2118   8.0378   76.3854 [c1]
 CurvedCapitals      0.0000   0.0046      .
                    25.8495   9.0906      .
 Addition           10.3031   0.0413   57.7158 [c2]
                     0.0013   0.8390      .
                    -9.2881   0.4163      .
 CountingDots        6.2954   8.5986   83.7834 [c3]
                     0.0121   0.0034      .
                    -6.8744  -5.4114      .
```

Wald Index, or expected chi-square increase if parameter is fixed at 0.

MI's or Lagrange Multipliers, or expected chi-square decrease if parameter is freed.

72

The "MODIFICATION" keyword on the PROC CALIS command line produces two types of diagnostics, Lagrange Multipliers and Wald indices. PROC CALIS prints these statistics in the same table. Lagrange multipliers are printed in place of fixed parameters; they indicate how much better the model would fit if the related parameter was freely estimated. Wald indices are printed in the place of free parameters; these statistics tell how much worse the model would fit if the parameter was fixed at zero.

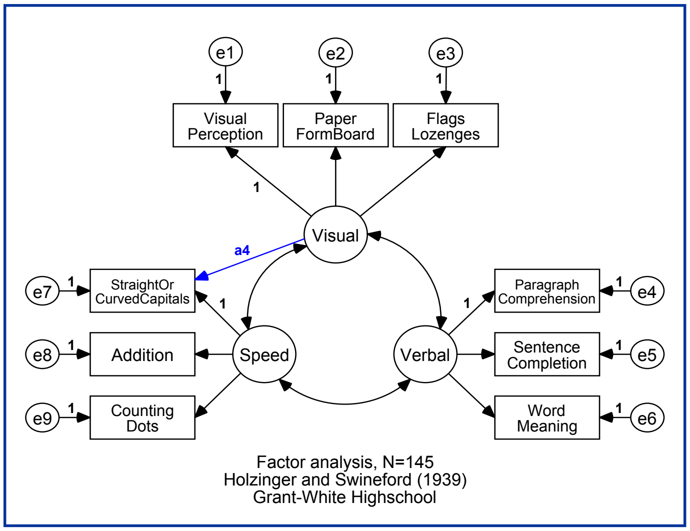
Modification Indices (Largest Ones)

Rank Order of the 9 Largest Lagrange Multipliers in GAMMA

Row	Column	Chi-Square	Pr > ChiSq
StraightOrCurvedCaps	F_Visual	30.21180	<.0001
Addition	F_Visual	10.30305	0.0013
CountingDots	F_Verbal	8.59856	0.0034
StraightOrCurvedCaps	F_Verbal	8.03778	0.0046
CountingDots	F_Visual	6.29538	0.0121
SentenceCompletion	F_Speed	2.69124	0.1009
FlagsLozenges_B	F_Speed	2.22937	0.1354
VisualPerception	F_Verbal	0.91473	0.3389
FlagsLozenges_B	F_Verbal	0.73742	0.3905

73

Modified Factor Model 2: Path Notation



74

CFA, PROC CALIS/LINEQS Notation: Model 2

```
PROC CALIS DATA=SEMdata.HolzingerSwinefordGW
  COVARIANCE RESIDUAL;
  VAR ...;
  LINEQS
    VisualPerception           = a1 F_Visual + e1,
    PaperFormBoard             = a2 F_Visual + e2,
    FlagsLozenges_B            = a3 F_Visual + e3,
    ParagraphComprehension     = b1 F_Verbal + e4,
    SentenceCompletion         = b2 F_Verbal + e5,
    WordMeaning                = b3 F_Verbal + e6,
    StraightOrCurvedCapitals   = a4 F_Visual +
                                 c1 F_Speed  + e7,
    Addition                   = c2 F_Speed  + e8,
    CountingDots               = c3 F_Speed  + e9;
  STD
    F_Visual F_Verbal F_Speed = 1.0 1.0 1.0,
    e1 - e9 = 9 * e_var:;
  COV
    F_Visual F_Verbal F_Speed = phi1 - phi3;
RUN;
```

75

Degrees of Freedom for CFA Model 2

```
CFA of Nine Psychological Variables, Model 2,
Holzinger-Swineford data.
The CALIS Procedure
Covariance Structure Analysis:
      Maximum Likelihood Estimation
Levenberg-Marquardt Optimization
Scaling Update of More (1978)
Parameter Estimates                    22
Functions (Observations)               45
```

One parameter more than Model 1 – one degree of freedom less

76

Degrees of freedom calculation for this model: df = 45 - 22 = 23.

CALIS, CFA Model 2: Fit Table

Fit Function	0.1427
Goodness of Fit Index (GFI)	0.9703
GFI Adjusted for Degrees of Freedom(AGFI)	0.9418
Root Mean Square Residual (RMR)	5.6412
Parsimonious GFI (Mulaik, 1989)	0.6199
Chi-Square	20.5494
Chi-Square DF	23
Pr > Chi-Square	0.6086
Independence Model Chi-Square	502.86
Independence Model Chi-Square DF	36
RMSEA Estimate	0.0000
RMSEA 90% Lo	

The chi-square statistic indicates that this model fits. In 61% of similar samples, a larger chi-square value would be found by chance alone.

77

The χ^2 statistic falls into the neighborhood of the degrees of freedom. This is what should be expected of a well-fitting model.

1.12 Multiple Choice Poll

A modification index (or Lagrange Multiplier) is

a. an estimate of how much fit can be improved if a particular parameter is estimated

b. an estimate of how much fit will suffer if a particular parameter is constrained to zero

c. I'm not sure.

79

Nested Models

Suppose there are two models for the same data:

A. a base model with $q1$ free parameters

B. a more general model with the same $q1$ free parameters, plus an additional set of $q2$ free parameters

Models A and B are considered to be nested. The nesting relationship is in the parameters – Model A can be thought to be a more constrained version of Model B.

81

Comparing Nested Models

Model	Chi-square	DF	P-Value	Comment
Model 1	48.0536	24	0.0025	Base model
Added Path "StraightOrCurvedCapitals <- F_Visual"	20.5494	23	0.6086	More general model
Difference	27.5042	1	0.0000	**"Significance of added parameters"**

If the more constrained model is true, then the difference in chi-square statistics between the two models follows, again, a chi-square distribution. The degrees of freedom for the chi-square difference equals the difference in model *dfs*.

82

Conversely, if the χ^2-difference is significant then the more constrained model is probably incorrect.

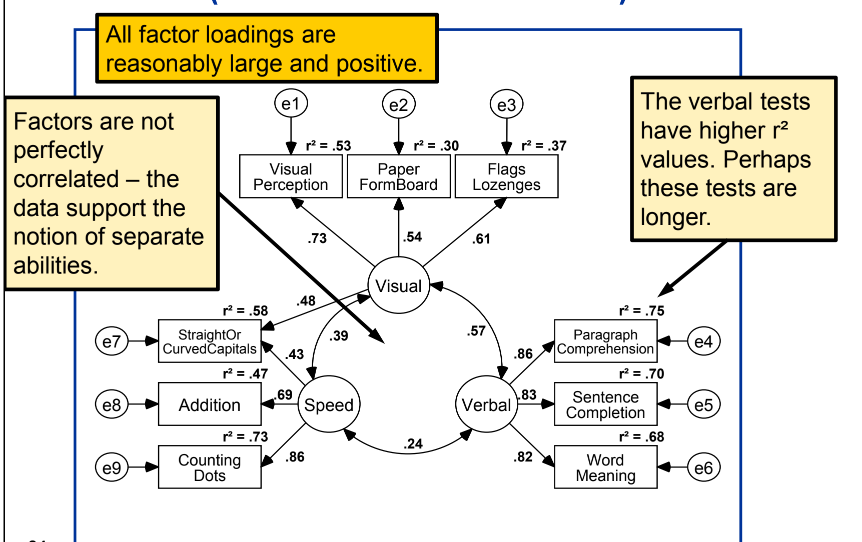
Some Parameter Estimates: CFA Model 2

```
Manifest Variable Equations with Estimates

  VisualPerception        =   5.0319*F_Visual     +  1.0000 e1
  Std Err                     0.5889 a1
  t Value                     8.5441

  PaperFormBoard          =   1.5377*F_Visual     +  1.0000 e2
  Std Err                     0.2499 a2
  t Value                     6.1541

  FlagsLozenges_B         =   5.0830*F_Visual      +  1.0000 e3
  Std Err                     0.7264 a3
  t Value                     6.9974
…
  StraightOrCurvedCaps = 17.7806*F_Visual + 15.9489*F_Speed + 1.0000 e7
  Std Err                3.1673 a4          3.1797 c1
  t Value                5.6139             5.0159
…
```

Estimates should be in the right direction; t-values should be large.

83

Results (Standardized Estimates)



84

Example 2: Summary

Tasks accomplished:

1. Set up a theory-driven factor model for nine variables, in other words, a model containing latent or unobserved variables
2. Estimated parameters and determined that the first model did not fit the data
3. Determined the source of the misfit by residual analysis and modification indices
4. Modified the model accordingly and estimated its parameters
5. Accepted the fit of new model and interpreted the results

85

1.5 Example3: Structural Equation Model

Example 3: Structural Equation Model

A. Application: Studying determinants of political alienation and its progress over time
B. Use summary data by Wheaton, Muthén, Alwin, and Summers (1977)
C. Entertain model with both structural and measurement components
D. Special modeling considerations for time-dependent variables
E. More about fit testing

88

Alienation Data: Wheaton et al. (1977)

Longitudinal Study of 932 persons from 1966 to 1971.

Determination of reliability and stability of alienation, a social psychological variable measured by attitude scales.

For this example, six of Wheaton's measures are used:

Variable	Description
Anomia67	1967 score on the Anomia scale
Anomia71	1971 Anomia score
Powerlessness67	1967 score on the Powerlessness scale
Powerlessness71	1971 Powerlessness score
YearsOfSchool66	Years of schooling reported in 1966
SocioEconomicIndex	Duncan's Socioeconomic index administered in 1966

89

Wheaton et al. (1977): Summary Data

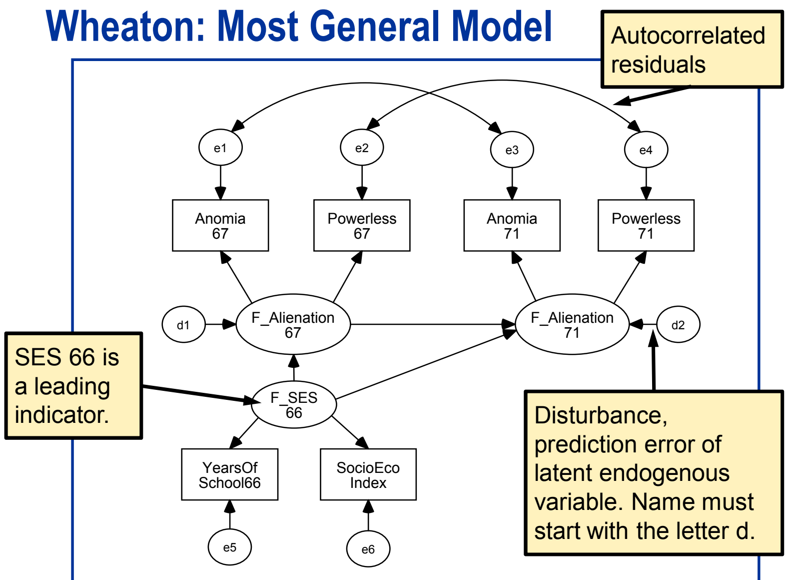
Obs	_type_	Anomia67	Powerlessness67	Anomia71	Powerlessness71	YearsOf School66	Socio Economic Index
1	n	932.00	932.00	932.00	932.00	932.00	932.00
2	corr	1.00	.	.	.	.	.
3	corr	0.66	1.00	.	.	.	.
4	corr	0.56	0.47	1.00	.	.	.
5	corr	0.44	0.52	0.67	1.00	.	.
6	corr	-0.36	-0.41	-0.35	-0.37	1.00	.
7	corr	-0.30	-0.29	-0.29	-0.28	0.54	1.00
8	STD	3.44	3.06	3.54	3.16	3.10	21.22
9	mean	13.61	14.76	14.13	14.90	10.90	37.49

The **_name_** column has been removed here to save space.

90

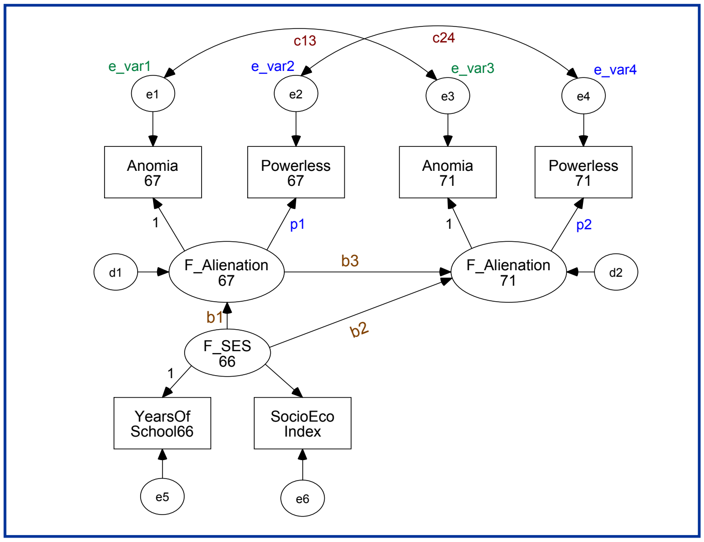
In the summary data file, the entries of “STD” type (in line 8) are really sample standard deviations. Please remember that this is different from the PROC CALIS subcommand “STD”, which is for variance terms.

Wheaton: Most General Model



91

Wheaton: Model with Parameter Labels



92

Wheaton: LINEQS Specification

```
LINEQS
  Anomia67              = 1.0 F_Alienation67 + e1,
  Powerlessness67       =  p1 F_Alienation67 + e2,

  Anomia71              = 1.0 F_Alienation71 + e3,
  Powerlessness71       =  p2 F_Alienation71 + e4,

  YearsOfSchool66       = 1.0 F_SES66 + e5,
  SocioEconomicIndex    =  s1 F_SES66 + e6,

  F_Alienation67        =  b1 F_SES66 + d1,
  F_Alienation71        =
           b2 F_SES66 + b3 F_Alienation67 + d2;
```

93

Wheaton: LINEQS Specification

```
LINEQS
  Anomia67             = 1.0 F_Alienation67 + e1,
  Powerlessness67      =  p1 F_Alienation67 + e2,

  Anomia71             = 1.0 F_Alienation71 + e3,
  Powerlessness71      =  p2 F_Alienation71 + e4,

  YearsOfSchool66      = 1.0 F_SES66 + e5,
  SocioEconomicIndex   =  s1 F_SES66 + e6,

  F_Alienation67       =  b1 F_SES66 + d1,
  F_Alienation71       =
           b2 F_SES66 + b3 F_Alienation67 + d2;
```

Measurement model coefficients can be constrained as time-invariant.

94

Wheaton: STD and COV Parameter Specs

```
STD
  F_SES66 = V_SES,
  e1 e2 e3 e4 e5 e6 =
    e_var1 e_var2 e_var3 e_var4 e_var5 e_var6,
  d1 d2 = d_var1 d_var2;
COV
  e1 e3 = c13,
  e2 e4 = c24;
RUN;
```

95

Wheaton: STD and COV Parameter Specs

```
STD
  F_SES66 = V_SES,
  e1 e2 e3 e4 e5 e6 =
    e_var1 e_var2 e_var3 e_var4 e_var5 e_var6,
  d1 d2 = d_var1 d_var2;
COV
  e1 e3 = c13,
  e2 e4 = c24;
RUN;
```

Some time-invariant models call for constraints of residual variances. These can be specified in the STD section.

96

Wheaton: STD and COV Parameter Specs

```
STD
  F_SES66 = V_SES,
  e1 e2 e3 e4 e5 e6 =
    e_var1 e_var2 e_var3 e_var4 e_var5 e_var6,
  d1 d2 = d_var1 d_var2;
COV
  e1 e3 = c13,
  e2 e4 = c24;
RUN;
```

Some time-invariant models call for constraints of residual variances. These can be specified in the STD section.

For models with uncorrelated residuals, remove this entire COV section.

97

Wheaton: Most General Model, Fit

```
SEM: Wheaton, Most General Model                                    30

The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation

Fit Function                                                0.0051
Chi-Square                                                  4.7701
Chi-Square DF                                                    4
Pr > Chi-Square                                             0.3117
Independence Model Chi-Square                               2131.8
Independence Model Chi-Square DF                                15
RMSEA Estimate                                              0.0144
RMSEA 90% Lower Confidence Limit                                 .
RMSEA 90% Upper Confidence Limit                            0.0533
ECVI Estimate                                               0.0419
ECVI 90% Lower Confidence Limit                                  .
ECVI 90% Upper Confidence Limit                             0.0525
Probability of Close Fit                                    0.9281
```

The most general model fits okay. Let's see what some more restricted models will do.

98

Wheaton: Time-Invariance Constraints (Input)

```
LINEQS
  Anomia67             = 1.0 F_Alienation67 + e1,
  Powerlessness67      = p1  F_Alienation67 + e2,

  Anomia71             = 1.0 F_Alienation71 + e3,
  Powerlessness71      = p1  F_Alienation71 + e4,
     …
STD
     …
  e1 - e6 = e_var1 e_var2 e_var1 e_var2 e_var5 e_var6,
     …
```

99

Wheaton: Time-Invariance Constraints (Output)

```
The CALIS Procedure (Model Specification and Initial Values Section)
Covariance Structure Analysis: Pattern and Initial Values

 Manifest Variable Equations with Initial Estimates
  Anomia67          =   1.0000 F_Alienation67     +  1.0000 e1
  Powerlessness67   =         .*F_Alienation67    +  1.0000 e2
                               p1
  Anomia71          =   1.0000 F_Alienation71     +  1.0000 e3
  Powerlessness71   =         .*F_Alienation71    +  1.0000 e4
                               p1
...
Variances of Exogenous Variables

Variable           Parameter      Estimate
F_SES66            V_SES                 .
e1                 e_var1                .
e2                 e_var2                .
e3                 e_var1                .
e4                 e_var2                .
...
```

100

1.13 Multiple Choice Poll

The difference between the time-invariant and the "most general" model is as follows:

a. The time-invariant model has the same measurement equations in 67 and 71.

b. The time-invariant model has the same set of residual variances in 67 and 71.

c. In the time-invariant model, both measurement equations and residual variances are the same in 67 and 71.

d. The time-invariant model has correlated residuals.

e. I'm not sure.

102

Wheaton: Chi-Square Model Fit and LR Chi-Square Tests

	Uncorrelated Residuals	Correlated Residuals	Difference
Time-Invariant	$\chi^2 = 73.0766, df = 9$	$\chi^2 = 6.1095, df = 7$	$\chi^2 = 66.9671, df = 2$
Time-Varying	$\chi^2 = 71.5438, df = 6$	$\chi^2 = 4.7701, df = 4$	$\chi^2 = 66.7737, df = 2$
Difference	$\chi^2 = 1.5328, df = 3$	$\chi^2 = 1.3394, df = 3$	

This is shown by the large column differences.

Conclusions:

1. There is evidence for autocorrelation of residuals – models with uncorrelated residuals fit considerably worse.
2. There is some support for time-invariant measurement – time-invariant models fit no worse (statistically) than time-varying measurement models.

This is shown by the small row differences.

104

Information Criteria to Assess Model Fit

Akaike's Information Criterion (AIC)
This is a criterion for selecting the best model among a number of candidate models. The model that yields the smallest value of AIC is considered the best.

$$AIC = \chi^2 - 2 \cdot df$$

Consistent Akaike's Information Criterion (CAIC)
This is another criterion, similar to AIC, for selecting the best model among alternatives. CAIC imposed a stricter penalty on model complexity when sample sizes are large.

$$CAIC = \chi^2 - (\ln(N) + 1) \cdot df$$

Schwarz's Bayesian Criterion (SBC)
This is another criterion, similar to AIC, for selecting the best model. SBC imposes a stricter penalty on model complexity when sample sizes are large.

$$SBC = \chi^2 - \ln(N) \cdot df$$

105

The intent of the information criteria is to identify models that replicate better than others. This means first of all that we must actually have multiple models to use these criteria. Secondly, models that fit best to sample data are not always the models that replicate best. Using the information criteria accomplishes a trade-off between estimation bias and uncertainty as they balance model fit on both these criteria.

✎ Information criteria can be used to evaluate models that are not nested.

Wheaton: Model Fit According to Information Criteria

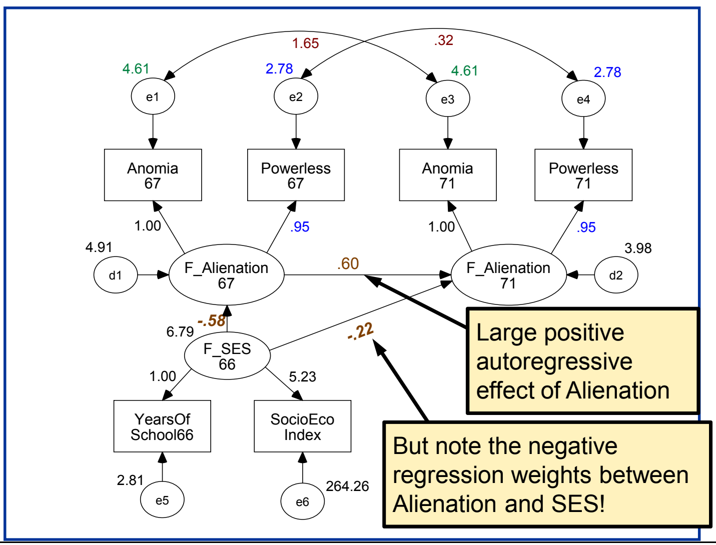
Model	AIC	CAIC	SBC
Most General	-3.222	-26.5972	-22.5792
Time-Invariant	-7.8905	-48.7518	-41.7518
Uncorrelated Residuals	59.5438	24.5198	30.5198
Time-Invariant & Uncorrelated Residuals	55.0766	2.5406	11.5406

Notes:

- Each of the three information criteria favors the time-invariant model.
- We would expect this model to *replicate* or *cross-validate* well with new sample data.

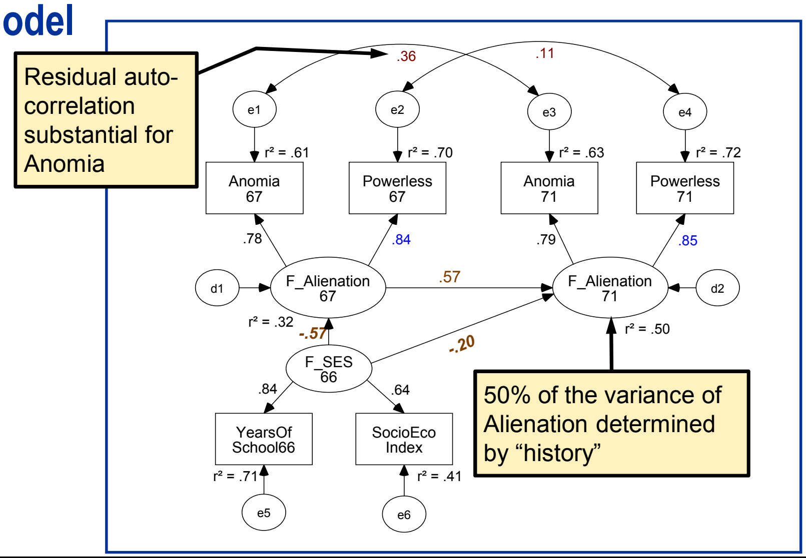
106

Wheaton: Parameter Estimates, Time-Invariant Model



107

Wheaton, Standardized Estimates, Time-Invariant Model



108

PROC CALIS Output (Measurement Model)

```
Covariance Structure Analysis: Maximum Likelihood Estimation
                 Manifest Variable Equations with Estimates

Anomia67           =  1.0000 F_Alienation67 + 1.0000 e1
Powerlessness67    =  0.9544*F_Alienation67 + 1.0000 e2
Std Err               0.0523 p1
t Value              18.2556

Anomia71           =  1.0000 F_Alienation71 + 1.0000 e3
Powerlessness71    =  0.9544*F_Alienation71 + 1.0000 e4
Std Err               0.0523 p1
t Value              18.2556

YearsOfSchool66    =  1.0000 F_SES66        + 1.0000 e5
SocioEconomicIndex =  5.2290*F_SES66        + 1.0000 e6
Std Err               0.4229 s1
t Value              12.3652
```

Is this the time-invariant model? How can we tell?

109

PROC CALIS OUTPUT (Structural Model)

```
Covariance Structure Analysis: Maximum Likelihood Estimation

                   Latent Variable Equations with Estimates

 F_Alienation67     = -0.5833*F_SES66       +  1.0000 d1
 Std Err               0.0560 b1
 t Value             -10.4236

 F_Alienation71     =  0.5955*F_Alienation67 + -0.2190*F_SES66
 Std Err               0.0472 b3                 0.0514 b2
 t Value              12.6240                   -4.2632

                        +  1.0000 d2
```

Cool, regressions among unobserved variables!

110

Wheaton: Asymptotically Standardized Residual Matrix

SEM: Wheaton, Time-Invariant Measurement

	Anomia67	Powerlessness67	Anomia71
Anomia67	-0.060061348	0.729927201	-0.051298262
Powerlessness67	0.729927201	-0.032747610	0.897225295
Anomia71	-0.051298262	0.897225295	0.059113256
Powerlessness71	-0.883389142	0.051352815	-0.736453922
YearsOfSchool66	1.217289084	-1.270143495	0.055115253
SocioEconomicIndex	-1.113169201	1.143759617	-1.413361725

	Powerlessness71	YearsOf School66	Socio Economic Index
Anomia67	-0.883389142	1.217289084	-1.113169201
Powerlessness67	0.051352815	-1.270143495	1.143759617
Anomia71	-0.736453922	0.055115253	-1.413361725
Powerlessness71	0.033733409	0.515612093	0.442256742
YearsOfSchool66	0.515612093	0.000000000	0.000000000
SocioEconomicIndex	0.442256742	0.000000000	0.000000000

Any indication of misfit in this table?

111

Example 3: Summary

Tasks accomplished:

1. Set up several competing models for time-dependent variables, conceptually and with PROC CALIS
2. Models included measurement and structural components
3. Some models were time-invariant, some had autocorrelated residuals
4. Models were compared by chi-square statistics and information criteria
5. Picked a winning model and interpreted the results

112

1.14 Multiple Choice Poll

The preferred model

a. has a small fit chi-square

b. has few parameters

c. replicates well

d. All of the above.

114

1.6 Example4: Effects of Errors-in-Measurement on Regression

Example 4: Warren et al., Regression with Unobserved Variables

A. Application: Predicting Job Performance of Farm Managers.
B. Demonstrate regression with unobserved variables, to estimate and examine the effects of measurement error.
C. Obtain parameters for further "what-if" analysis; for instance,
   a) Is the low r-square of 0.40 in Example 1 due to lack of reliability of the dependent variable?
   b) Are the estimated regression weights of Example 1 true or biased?
D. Demonstrate use of very strict parameter constraints, made possible by virtue of the measurement design.

118

Warren9Variables: Split-Half Versions of Original Test Scores

Variable	Explanation
Performance_1	12-item subtest of Role Performance
Performance_2	12-item subtest of Role Performance
Knowledge_1	13-item subtest of Knowledge
Knowledge_2	13-item subtest of Knowledge
ValueOrientation_1	15-item subtest of Value Orientation
ValueOrientation_2	15-item subtest of Value Orientation
Satisfaction_1	5-item subtest of Role Satisfaction
Satisfaction_2	6-item subtest of Role Satisfaction
past-training	Degree of formal education

119

The Effect of Shortening or Lengthening a Test

Statistical effects of changing the length of a test:
Lord, F.M. and Novick, M.R. 1968. *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley.

Suppose:
Two tests, X and Y, differing only in length, with

LENGTH(Y) = w·LENGTH(X)

Then, by Lord & Novick, chapter 4:

$\sigma^2(X) = \sigma^2(\tau_x) + \sigma^2(\varepsilon_x)$, and

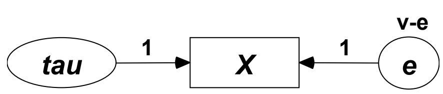
$\sigma^2(Y) = \sigma^2(\tau_y) + \sigma^2(\varepsilon_y)$

$= w^2 \cdot \sigma^2(\tau_x) + w \cdot \sigma^2(\varepsilon_x)$

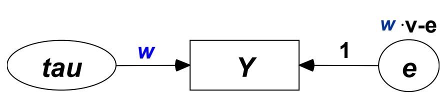
120

Path Coefficient Modeling of Test Length

$\sigma^2(X) = \sigma^2(\tau) + \sigma^2(\varepsilon_x)$

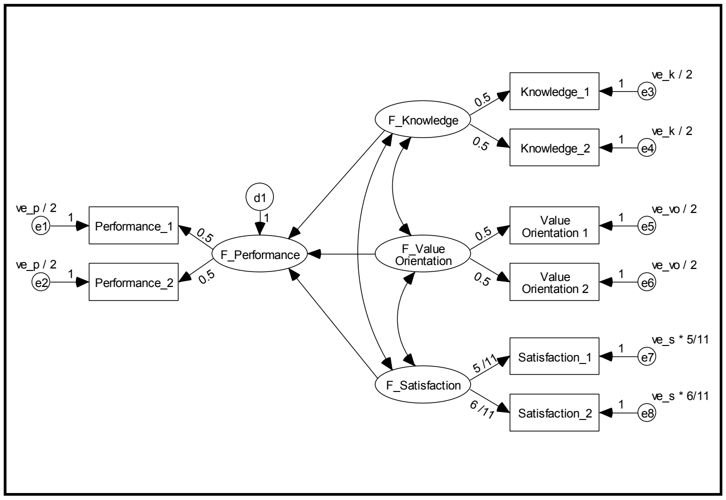


$\sigma^2(Y) = w^2 \cdot \sigma^2(\tau) + w \cdot \sigma^2(\varepsilon_x)$



121

Warren9Variables: Graphical Specification



122

This model is highly constrained, courtesy of the measurement design and formal results of classical test theory (*e.g.,* Lord & Novick, 1968).

Warren9Variables: CALIS Specification (1/2)

```
LINEQS
   Performance_1       = 0.5 F_Performance       + e_p1,
   Performance_2       = 0.5 F_Performance       + e_p2,
   Knowledge_1         = 0.5 F_Knowledge         + e_k1,
   Knowledge_2         = 0.5 F_Knowledge         + e_k2,
   ValueOrientation_1  = 0.5 F_ValueOrientation  + e_vo1,
   ValueOrientation_2  = 0.5 F_ValueOrientation  + e_vo2,
   Satisfaction_1      = 0.454545 F_Satisfaction + e_s1,
   Satisfaction_2      = 0.545454 F_Satisfaction + e_s2,

   F_Performance = b1 F_Knowledge + b2 F_ValueOrientation
                + b3 F_Satisfaction + d1;
 STD
   e_p1 e_p2     e_k1 e_k2    e_vo1 e_vo2    e_s1 e_s2 =
    ve_p1 ve_p2   ve_k1 ve_k2   ve_vo1 ve_vo2  ve_s1 ve_s2,

   d1 F_Knowledge F_ValueOrientation F_Satisfaction =
     v_d1 v_K v_VO v_S;
 COV
   F_Knowledge F_ValueOrientation F_Satisfaction = phi1 - phi3;
```

123

continued...

Warren9Variables: CALIS Specification (2/2)

```
  PARAMETERS /* Hypothetical error variance terms of original  */
             /* scales; start values must be set by modeler    */
       ve_p ve_k ve_vo ve_s = 0.01 0.01 0.01 0.01;

       ve_p1  = 0.5 * ve_p;      /* SAS programming statements */
       ve_p2  = 0.5 * ve_p;      /*   express error variances  */
       ve_k1  = 0.5 * ve_k;      /*   of eight split scales    */
       ve_k2  = 0.5 * ve_k;      /*   as exact functions of    */
       ve_vo1 = 0.5 * ve_vo;     /*   hypothetical error       */
       ve_vo2 = 0.5 * ve_vo;     /*   variance terms of the    */
       ve_s1  = 0.454545 * ve_s; /*   four original scales.    */
       ve_s2  = 0.545454 * ve_s;
RUN;
```

124

Warren9Variables: Model Fit

```
...
Chi-Square                                  26.9670
Chi-Square DF                                    22
Pr > Chi-Square                              0.2125
...
```

Comment:

The model fit is acceptable.

125

1.15 Multiple Answer Poll

How do you fix a parameter with PROC CALIS?

a. Use special syntax to constrain the parameter values.
b. Just type the parameter value in the model specification.
c. PROC CALIS does not allow parameters to be fixed.
d. Both options (a) and (b).
e. I'm not sure.

127

Warren9variables: Structural Parameter Estimates Compared to Example 1

Predictor Variable	Regression Controlled for Error (PROC CALIS)	Regression without Error Model (PROC REG)
Knowledge	0.3899 (0.1393)	0.2582 (0.0544)
Value Orientation	0.1800 (0.0838)	0.1450 (0.0356)
Satisfaction	0.0561 (0.0535)	0.0486 (0.0387)

129

Modeled Variances of Latent Variables

```
PROC CALIS
        DATA=SemLib.Warren9variables
        COVARIANCE PLATCOV;
```

Prints variances and covariances of latent variables.

Variable	Performance	Knowledge	Value Orientation	Satisfaction
$\sigma^2(\tau)$	0.0688	0.1268	0.3096	0.2831
$\sigma^2(e)$	0.0149	0.0810	0.1751	0.0774
r_{xx}	0.82	0.61	0.64	0.79

Reliability estimates for example 1:
$r_{xx} = \sigma^2(\tau) / [\sigma^2(\tau) + \sigma^2(e)]$

130

Warren9variables: Variance Estimates

Variances of Exogenous Variables

Variable	Parameter	Estimate	Standard Error	t Value
F_Knowledge	v_K	0.12680	0.03203	3.96
F_ValueOrientation	v_VO	0.30960	0.07400	4.18
F_Satisfaction	v_S	0.28313	0.05294	5.35
e_p1	ve_p1	0.00745	0.00107	6.96
e_p2	ve_p2	0.00745	0.00107	6.96
e_k1	ve_k1	0.04050	0.00582	6.96
e_k2	ve_k2	0.04050	0.00582	6.96
e_vo1	ve_vo1	0.08755	0.01257	6.96
e_vo2	ve_vo2	0.08755	0.01257	6.96
e_s1	ve_s1	0.03517	0.00505	6.96
e_s2	ve_s2	0.04220	0.00606	6.96
d1	v_d1	0.02260	0.00851	2.66

131

Warren9variables: Standardized Estimates

```
F_Performance   =
   0.5293*F_Knowledge   +   0.3819*F_ValueOrientation
          b1                       b2
      +  0.1138*F_Satisfaction    +   0.5732 d1
               b3

Squared Multiple Correlations

                                    Error          Total
   Variable                       Variance       Variance     R-Square
   Performance_1                   0.00745        0.02465       0.6978
   Performance_2                   0.00745        0.02465       0.6978
   Knowledge_1                     0.04050        0.07220       0.4391
   Knowledge_2                     0.04050        0.07220       0.4391
   ValueOrientation_1              0.08755        0.16495       0.4692
   ValueOrientation_2              0.08755        0.16495       0.4692
   Satisfaction_1                  0.03517        0.09367       0.6245
   Satisfaction_2                  0.04220        0.12644       0.6662
   F_Performance                   0.02260        0.06880       0.6715
```

Considerable measurement error in these "split" variables!

Hypothetical R-square for 100% reliable variables, up from 0.40.

132

1.16 Multiple Choice Poll

In Example 4, the R-square of the factor F_Performance is larger than that of the observed variable Performance of Example 1 because

a. measurement error is eliminated from the structural equation

b. the sample is larger, so sampling error is reduced

c. the reliability of the observed predictor variables was increased by lengthening them

d. I'm not sure.

134

Example 4: Summary

Tasks accomplished:

1. Set up model to study effect of measurement error in regression
2. Used split versions of original variables as multiple indicators of latent variables
3. Constrained parameter estimates according to measurement model
4. Obtained an acceptable model
5. Found that predictability of **`JobPerformance`** could potentially be as high as R-square=0.67

136

1.7 Conclusion

Conclusions

Course accomplishments:

1. Introduced Structural Equation Modeling in relation to regression analysis, factor analysis, simultaneous equations
2. Showed how to set up Structural Equation Models with PROC CALIS
3. Discussed model fit by comparing covariance matrices, and considered chi-square statistics, information criteria, and residual analysis
4. Demonstrated several different types of modeling applications

138

Comments

Several components of the standard SEM curriculum were omitted due to time constraints:

- Model identification
- Non-recursive models
- Other fit statistics that are currently in use
- Methods for nonnormal data
- Methods for ordinal-categorical data
- Multi-group analyses
- Modeling with means and intercepts
- Model replication
- Power analysis

139

Current Trends

Current trends in SEM methodology research:

1. Statistical models and methodologies for missing data
2. Combinations of latent trait and latent class approaches
3. Bayesian models to deal with small sample sizes
4. Non-linear measurement and structural models (such as IRT)
5. Extensions for non-random sampling, such as multi-level models

140

Solutions to Student Activities (Polls/Quizzes)

1.07 Multiple Choice Poll – Correct Answer

How many PROC REG subcommands are required to specify a linear regression with PROC REG?

a. None
b. 1
c. 2
d. 3
e. 4
f. More than 4

27

1.08 Multiple Choice Poll – Correct Answer

How many PROC CALIS subcommands are required to specify a linear regression with PROC CALIS?

a. None
b. 1
c. 2
d. 3
e. 4
f. More than 4

42

1.09 Multiple Choice Poll – Correct Answer

Which PROC CALIS output message indicates that an iterative solution has been found?

a. Covariance Structure Analysis: Maximum Likelihood Estimation
b. Manifest Variable Equations with Estimates
c. Vector of Initial Estimates
d. ABSGCONV convergence criterion satisfied
e. None of the above
f. Not sure

50

1.10 Multiple Choice Poll – Correct Answer

How many LINEQS equations are needed for a factor analysis?

a. Nine, just like the previous slide
b. One for each observed variable in the model
c. One for each factor in the model
d. One for each variance term in the model
e. None of the above
f. Not sure

62

1.11 Multiple Choice Poll – Correct Answer

A large chi-square fit statistic means that

a. the model fits well

b. the model fits poorly

c. I'm not sure.

71

1.12 Multiple Choice Poll – Correct Answer

A modification index (or Lagrange Multiplier) is

a. an estimate of how much fit can be improved if a particular parameter is estimated

b. an estimate of how much fit will suffer if a particular parameter is constrained to zero

c. I'm not sure.

80

1.13 Multiple Choice Poll – Correct Answer

The difference between the time-invariant and the "most general" model is as follows:

a. The time-invariant model has the same measurement equations in 67 and 71.

b. The time-invariant model has the same set of residual variances in 67 and 71.

c. In the time-invariant model, both measurement equations and residual variances are the same in 67 and 71.

d. The time-invariant model has correlated residuals.

e. I'm not sure.

103

1.14 Multiple Choice Poll – Correct Answer

The preferred model

a. has a small fit chi-square

b. has few parameters

c. replicates well

d. All of the above.

115

1.15 Multiple Answer Poll – Correct Answer

How do you fix a parameter with PROC CALIS?

a. Use special syntax to constrain the parameter values.
b. Just type the parameter value in the model specification.
c. PROC CALIS does not allow parameters to be fixed.
d. Both options (a) and (b).
e. I'm not sure.

128

1.16 Multiple Choice Poll – Correct Answer

In Example 4, the R-square of the factor F_Performance is larger than that of the observed variable Performance of Example 1 because

a. measurement error is eliminated from the structural equation
b. the sample is larger, so sampling error is reduced
c. the reliability of the observed predictor variables was increased by lengthening them
d. I'm not sure.

135

1.8 References

Materials Referenced in the Web Lecture

Akaike, H. 1987. “Factor analysis and AIC.” *Psychometrika* 52(3):317-332.

Bozdogan, H. 1987. “Model Selection and Akaike's Information Criterion (AIC): The General Theory and Its Analytical Extensions.” *Psychometrika* 52(3):345-370.

Fisher, R.A. 1921 “On the “probable error” of the coefficient of correlation deduced from a small sample.” *Metron* 1:3-32.

Galton, F. 1889. *Natural Inheritance*. London: Macmillan.

Goldberger, A. S. and O.D. Duncan, eds. 1973. *Structural Equation Models in the Social Sciences*. New York: Seminar Press/Harcourt Brace.

Haavelmo, T. 1943. “The statistical implications of a system of simultaneous equations.” *Econometrica* 11:1-12

Holzinger, K. J. and F. Swineford. 1939. “A study in factor analysis: The stability of a bi-factor solution.” *Supplementary Educational Monographs*. Chicago: University of Chicago

Jöreskog, K. G. 1970. “A general method for analysis of covariance structures.” *Biometrika* 57:239-251.

Koopmans, T.C. 1953. “Identification problems in econometric model construction.” In *Studies in Econometric Method*, eds. W.C. Hood and T.C. Koopmans, 27-48. New York: Wiley

Lawley, D. N. 1940. “The estimation of factor loadings by the method of maximum likelihood.” *Proceedings of the Royal Statistical Society of Edinburgh, Sec. A* 60:64-82.

Lawley, D. N. and A. E. Maxwell. 1971. *Factor Analysis as a Statistical Method*. London: Butterworth and Co.

Lord, F. M. and M.R. Novick. 1968. *Statistical Theories of Mental Test Scores*. Reading MA: Addison-Welsley Publishing Company.

Schwarz, G. 1978. “Estimating the dimension of a model.” *Annals of Statistics* 6:461–464.

Spearman, C. 1904. “General intelligence objectively determined and measured.” *American Journal of Psychology* 15:201-293.

Thurstone, L. L. 1935. *Vectors of the Mind*. Chicago: University of Chicago Press.

Thurstone, L. L. 1947. *Multiple Factor Analysis*. Chicago: University of Chicago Press

Warren, R.D., J.K. White, and W.A. Fuller. 1974. “An Errors-In-Variables Analysis of Managerial Role Performance.” *Journal of the American Statistical Association* 69:886–893.

Wheaton, B., et al. 1977. “Assessing Reliability and Stability in Panel Models.” In *Sociological Methodology*, ed. D. Heise, San Francisco: Jossey-Bass.

Wold, H. 1954. “Causality and Econometrics.” *Econometrica* 22:162–177.

Wright, S. 1934. "The method of path coefficients." *Annals of Mathematical Statistics* 5:161-215.

A Small Selection of Introductory SEM Text Books

Hoyle, Rick. 1995. *Structural Equation Modeling: Concepts, Issues and Applications*. Thousand Oaks, CA: Sage Publications (0-8039-5318-6).

Kline, R. B. 2005. *Principles and Practice of Structural Equation Modeling, 2nd Edition*. New York: Guilford Press.

Loehlin, John C. 1998. *Latent Variable Models: An Introduction to Factor, Path, and Structural Analysis. 3rd Edition.* Mahwah, NJ: Lawrence Erlbaum Associates.

Maruyama, G.M. 1998. *Basics of Structural Equation Modeling*. Thousand Oaks, CA: Sage Publications.

Schumacker, Randall and Richard Lomax. 1996. *A Beginner's Guide to Structural Equation Modeling*. Mahwah, NJ: Lawrence Erlbaum Associates. (0-8058-1766-2).

A Selection of Graduate-Level SEM Books
(Prior training in matrix algebra and statistical sampling theory suggested)

Bollen, K.A. 1989. *Structural Equations with Latent Variables*. New York: Wiley.

Kaplan, D. 2000. *Structural Equation Modeling. Foundations and Extensions*. Thousand Oaks, CA: Sage Publications.

Lee, S.-Y. 2007. *Structural Equation Modeling: A Bayesian Approach*. New York: Wiley.

Skrondal, A. and S. Rabe-Hesketh. 2004. *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models.* Boca Raton, FL: Chapman & Hall/CRC.

Useful Web Links

- SmallWaters Corp. SEM-related links: http://www.smallwaters.com/weblinks/
- Peter Westfall's demonstration of the effect of measurement error in regression analysis (relates to Example 4 of the Web lecture):
http://www2.tltc.ttu.edu/Westfall/images/6348/measurmenterrorbias.htm