An Introduction to Structural Equation Modeling

Course Notes

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

An Introduction to Structural Equation Modeling Course Notes

Book code E70366, course code LWBAWW, prepared date 21-Feb-07 LWBAWW_002

ISBN 978-1-59994-400-5

Table of Contents

Course Description iv

Chapter 1 Introduction to Structural Equation Modeling 1-1

1.1 Introduction 1-2

1.2 Structural Equation Modeling – Overview 1-3

1.3 Example 1: Regression Analysis 1-5

1.4 Example 2: Factor Analysis 1-13

1.5 Example 3: Structural Equation Model 1-22

1.6 Example 4: Effects of Errors-in-Measurement on Regression 1-30

1.7 Conclusions 1-34

Course Description

This session 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…

sas Education

A full curriculum of general and statistical instructor-based training is available at any of the Institute's training facilities. Institute instructors can also provide on-site training.

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.

sas Publishing

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 www.sas.com/pubs for a complete list of books and a convenient order form.

Chapter 1 Introduction to Structural Equation Modeling

1.1 Introduction......1-2

1.2 Structural Equation Modeling – Overview......1-3

1.3 Example 1: Regression Analysis......1-5

1.4 Example 2: Factor Analysis......1-17

1.5 Example 3: Structural Equation Model......1-29

1.6 Example 4: Effects of Errors-in-Measurement on Regression......1-40

1.7 Conclusions......1-45

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

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.

9

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)

10

Common Terms in SEM

<u>Types of Variables:</u>

Measured, Observed, Manifest

versus

Hypothetical, Unobserved, Latent

Variable present in the data file and not missing

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 ← y ← x

1.3 Example 1: Regression Analysis

Example 1: Multiple Regression

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

15

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.

16

Warren, White, and Fuller (1974) Data

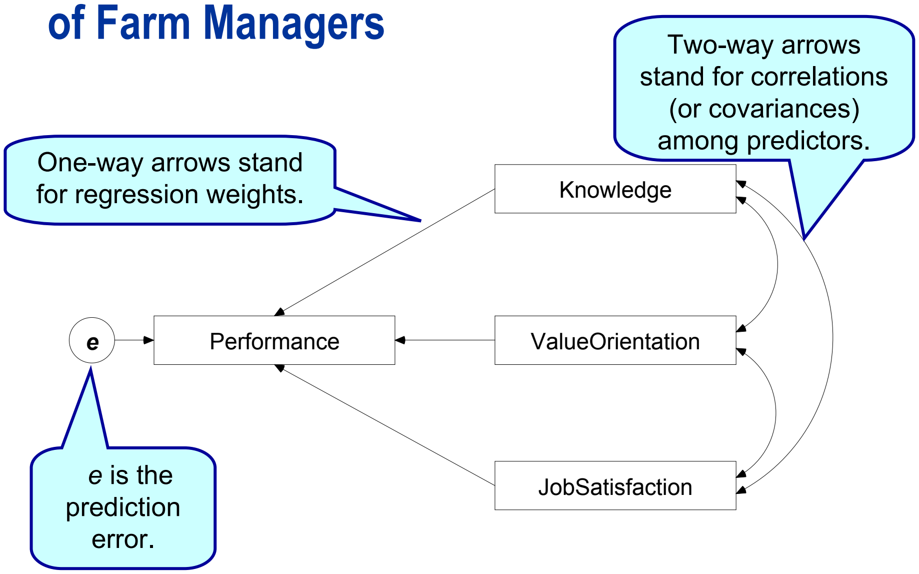
Warren 5 Variables: Summary Data, SAS Format

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.

17

Prediction Model: Job Performance of Farm Managers



18

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.

19

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

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

Prediction Equation for Job Performance:

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

20

LINEQS Model Interface in PROC CALIS

```
PROC CALIS DATA=<inputfile> <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.

22

LINEQS Model Interface in PROC CALIS

```
PROC CALIS DATA=<inputfile> <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.

23

LINEQS Model Interface in PROC CALIS

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

VAR (optional statement) to select and reorder variables from <*inputfile*>

24

LINEQS Model Interface in PROC CALIS

```
PROC CALIS DATA=<inputfile> <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.

25

LINEQS Model Interface in PROC CALIS

```
PROC CALIS DATA=<inputfile> <opt

  VAR <list of variables>;

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

Variances of unobserved exogenous variables to be listed here

26

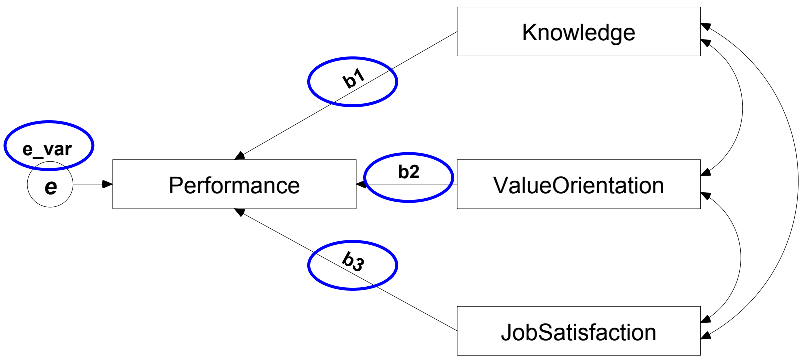
LINEQS Model Interface in PROC CALIS

```
PROC CALIS DATA=<inputfile> <options>;

  VAR <list of variables>;

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

Covariances of unobserved exogenous variables to be listed here

27

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



28

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).

29

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 letter 'e'.*

30

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.

31

Parameter Estimates: PROC CALIS

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
```

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`**.

32

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`**.

33

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.

continued...

35

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.

36

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]
```

37

continued...

Iterative Estimation Process

```
                         Function        Active   ...
Iter    Restarts            Calls   Constraints   ...
   1           0                2             0   ...
   2           0                3             0   ...

                Optimization Results
Iterations                                      2
...
Max Abs Gradient Element             4.348156E-14
...
GCONV2 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!

38

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

39

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

43

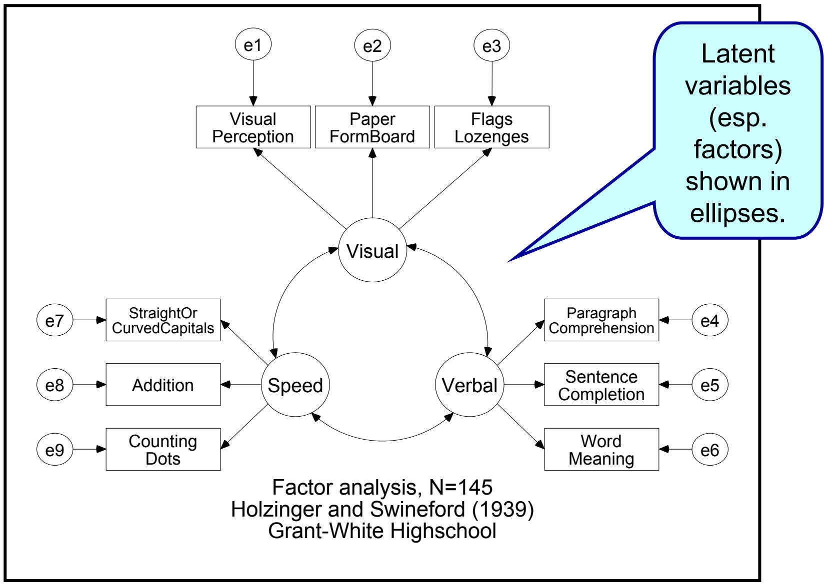
Confirmatory Factor Analysis: Model 1

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

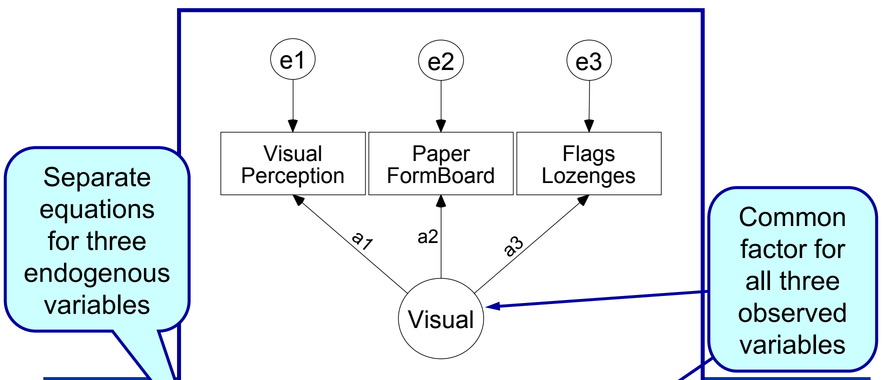
44

CFA, Path Diagram Notation: Model 1



45

Measurement Specification with LINEQS

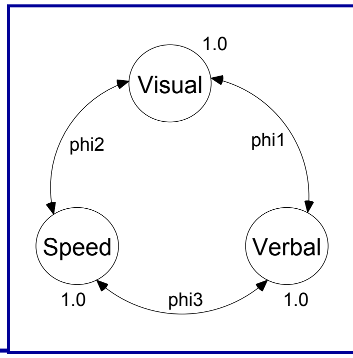



```
LINEQS
   VisualPerception = a1 F_Visual + e1,
   PaperFormBoard   = a2 F_Visual + e2,
   FlagsLozenges    = a3 F_Visual + e3,
```

Factor names start with ‘F’ in PROC CALIS.

46

Specifying Factor Correlations with LINEQS



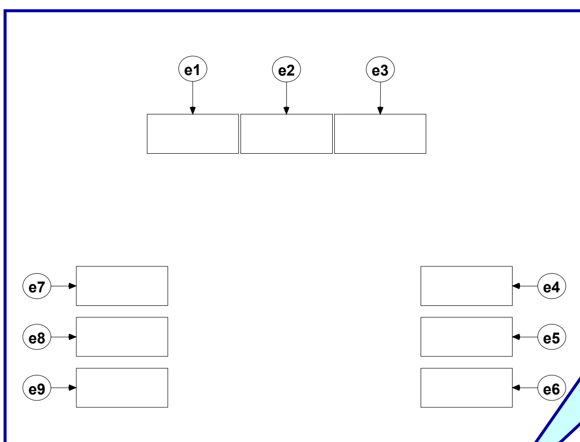
Factor variances are 1.0 for correlation matrix.

Factor correlation terms go into the COV section.

```
STD
   F_Visual F_Verbal F_Speed = 1.0 1.0 1.0
   ...;

COV
   F_Visual F_Verbal F_Speed = phi1 phi2 phi3;
```

47

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;
```

48

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_var3
      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

49

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

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

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

51

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

52

CALIS, CFA Model 1: Fit Table

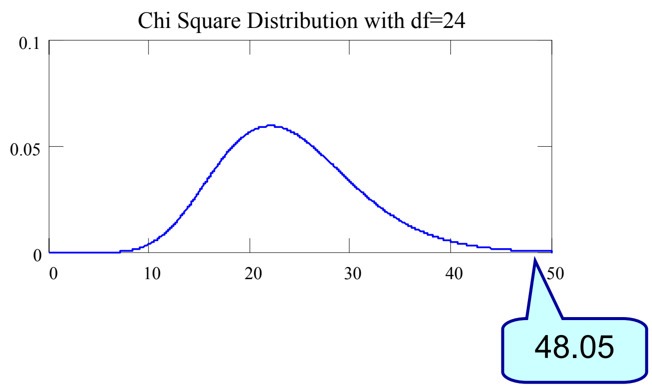
Fit Function	0.3337
Goodness of Fit Index (GFI)	0.9322
GFI Adjusted for Degrees of Freedom (AGFI)	0.8729
Root Mean Square Residual (RMR)	15.9393
Parsimonious GFI (Mulaik, 1989)	0.6215
Chi-Square	48.0536
Chi-Square DF	24
Pr > Chi-Square	0.0025
Independence Model Chi-Square	502.86
Independence Model Chi-Square DF	36
RMSEA Estimate	0.0834
RMSEA 90% Lower Confidence Limit	0.0483

…and many more fit statistics on list output.

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

53

Chi Square Test: Model 1



54

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

55

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

56

Modification Indices (Table)

Univariate Tests for Constant Constraints
Lagrange Multiplier or Wald Index
/ Probability / Approx Change of Value

	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.

58

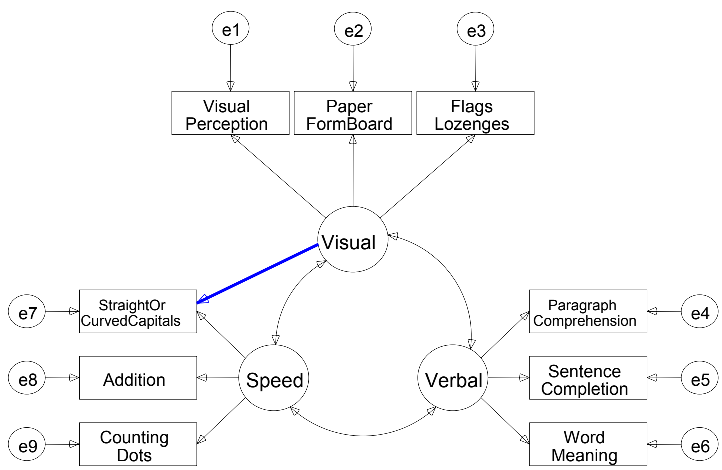
Modification Indices (Largest Ones)

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

59

Modified Factor Model 2: Path Notation



Factor analysis, N=145
Holzinger and Swineford (1939)
Grant-White Highschool

60

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;
```

61

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

62

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% Lower C	

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.

63

Nested Models

Suppose there are two models for the same data:

A. a base model with *q*1 free parameters

B. a more general model with the same *q*1 free parameters, plus an additional set of *q*2 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.

65

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*.

66

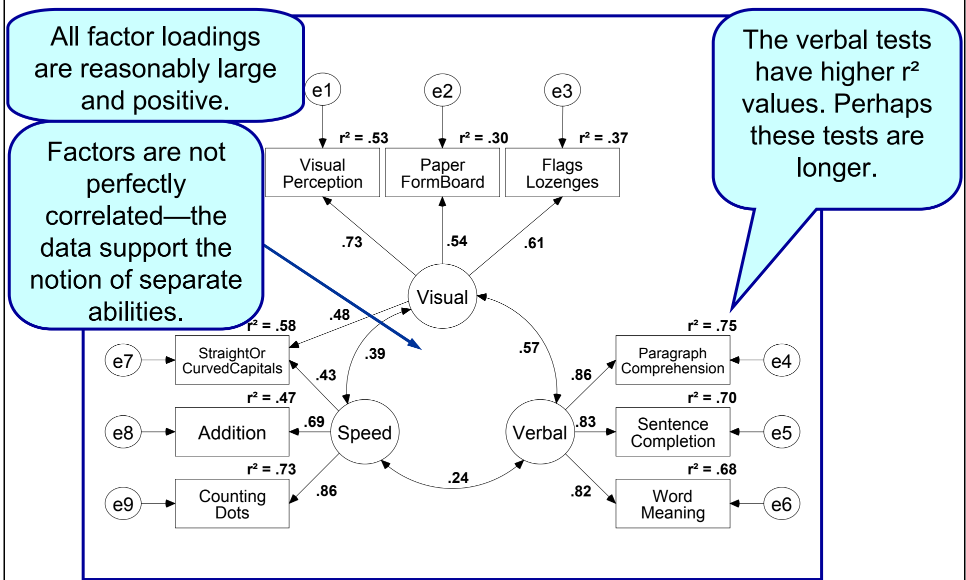
Some Parameter Estimates: CFA Model 2

```
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.

67

Results (Standardized Estimates)



68

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

69

1.5 Example 3: 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

72

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

73

Wheaton et al. (1977): Summary Data

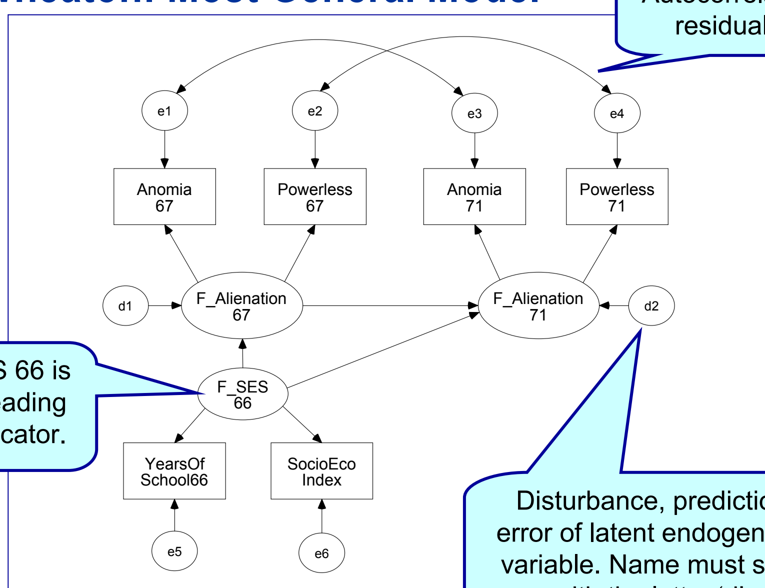
Obs	_type_	Anomia67	Powerlessness67	Anomia71	Powerlessness71	YearsOf School66	Socio Economic Index
1	n	932.00	932.00	932.00	932.00	932.00	932.00
2	corr	1.00	.	.	.	.	.
3	corr	0.66	1.00	.	.	.	.
4	corr	0.56	0.47	1.00	.	.	.
5	corr	0.44	0.52	0.67	1.00	.	.
6	corr	-0.36	-0.41	-0.35	-0.37	1.00	.
7	corr	-0.30	-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.

74

Wheaton: Most General Model

Autocorrelated residuals

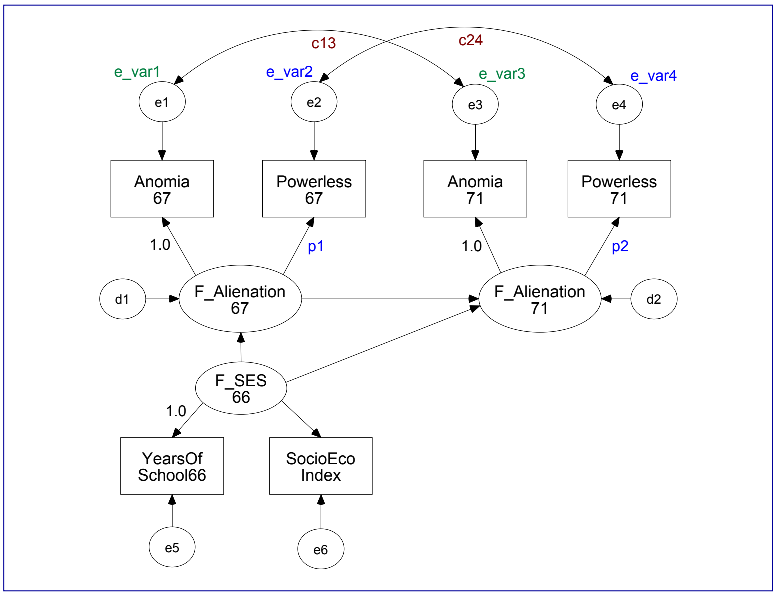


SES 66 is a leading indicator.

Disturbance, prediction error of latent endogenous variable. Name must start with the letter ‘d’.

75

Wheaton: Model with Parameter Labels



76

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;
```

77

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.

78

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;
```

79

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.

80

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.

81

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.

82

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,
     …
```

83

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                .
...
```

84

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

	Uncorrelated Residuals	Correlated Residuals	Difference
Time-Invariant	$\chi^2 = 73.5438, df = 9$	$\chi^2 = 6.1095, df = 7$	$\chi^2 = 67.4343, df = 2$
Time-Varying	$\chi^2 = 71.5138, df = 6$	$\chi^2 = 4.7701, df = 4$	$\chi^2 = 66.7437, df = 2$
Difference	$\chi^2 = 2.0300, df = 3$	$\chi^2 = 1.3394, df = 3$	

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.

86

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.

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

87

Wheaton: Model Fit According to Information Criteria

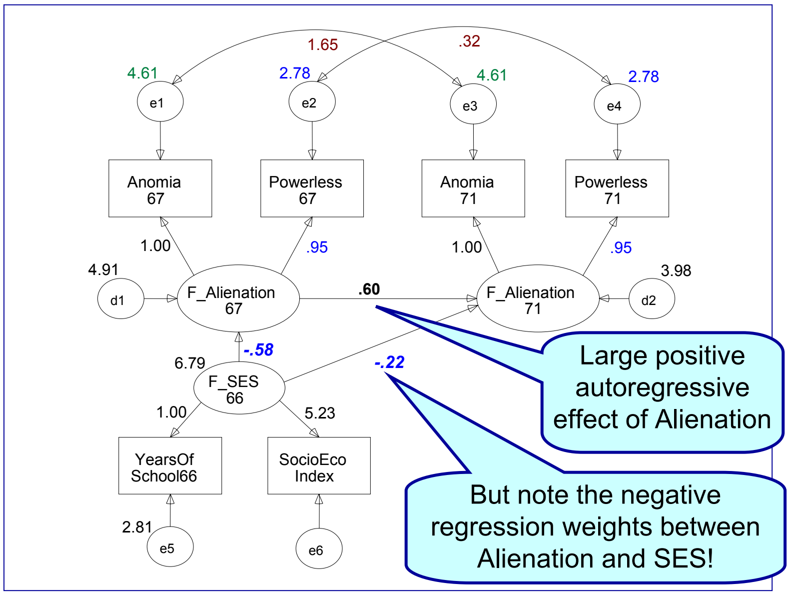
Model	AIC	CAIC	SBC
Most General	-3.2299	-26.5792	-22.5792
Time-Invariant	-7.8905	-48.7518	-41.7518
Uncorrelated Residuals	59.5438	24.5198	30.5198
Time-Invariant & Uncorrelated Residuals	55.0766	2.5406	11.5406

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.

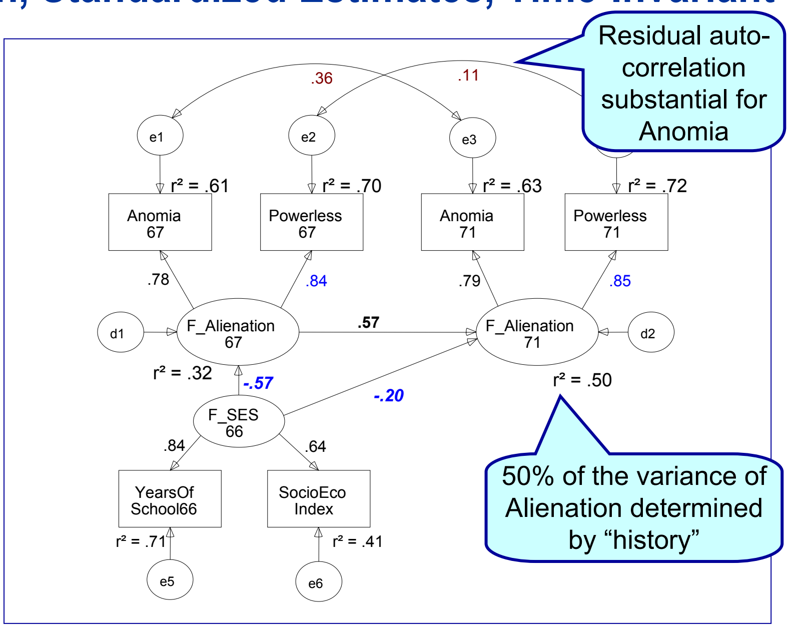
88

Wheaton: Parameter Estimates, Time-Invariant Model



89

Wheaton, Standardized Estimates, Time-Invariant Model



90

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?

91

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!

92

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?

93

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

94

1.6 Example 4: 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.

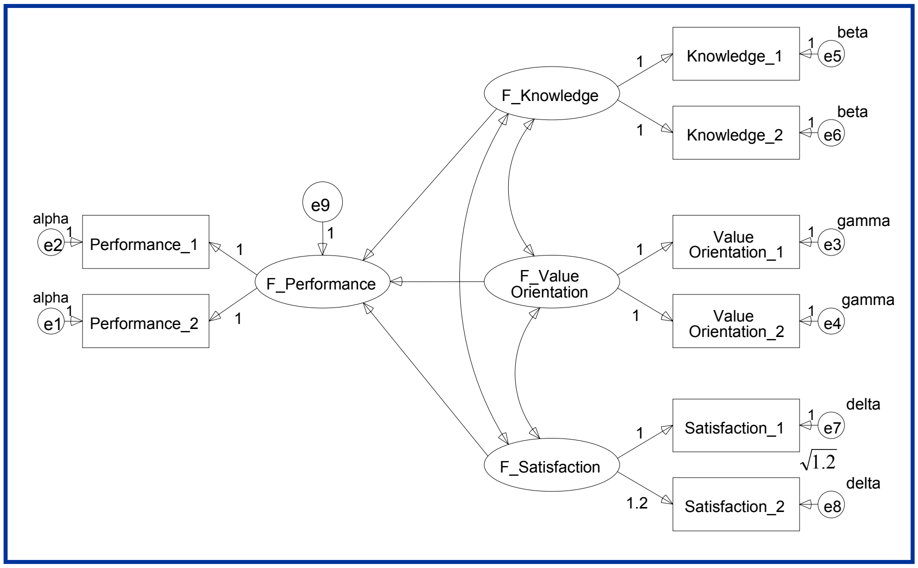
98

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

99

Warren9Variables: Graphical Specification



100

Warren9Variables: CALIS Specification

```
LINEQS
   Performance_1       = 1.0 F_Performance + e1,
   Performance_2       = 1.0 F_Performance + e2,
   Knowledge_1         = 1.0 F_Knowledge   + e3,
   Knowledge_2         = 1.0 F_Knowledge   + e4,
   ValueOrientation_1  = 1.0 F_ValueOrientation  + e5,
   ValueOrientation_2  = 1.0 F_ValueOrientation  + e6,
   Satisfaction_1      = 1.0 F_Satisfaction      + e7,
   Satisfaction_2      = 1.2 F_Satisfaction      + 1.095445 e8,

   F_Performance = b1 F_Knowledge + b2 F_ValueOrientation
                 + b3 F_Satisfaction + d1;
 STD
   e1 e2 e3 e4 e5 e6 e7 e8 =
     alpha alpha beta beta gamma gamma delta delta,
   d1 = v_d1,
   F_Knowledge F_ValueOrientation F_Satisfaction = v_K v_VO v_S;
 COV
   F_Knowledge F_ValueOrientation F_Satisfaction =
     phi1 phi2 phi3;
```

101

Warren9Variables: Model Fit

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

Comment:

The model fit is acceptable.

102

Warren9variables: Structural Parameter Estimates

```
Latent Variable Equations with Estimates

F_Performance  = 0.3899*F_Knowledge  +  0.1800*F_ValueOrientation
Std Err          0.1393 b1              0.0838 b2
t Value          2.7979                 2.1475

               +  0.0617*F_Satisfaction     +  1.0000 d1
                  0.0588 b3
                  1.0490
```

104

Warren9variables: Variance Estimates

Variances of Exogenous Variables

Variable	Parameter	Estimate	Standard Error	t Value
F_Knowledge	v_K	0.03170	0.00801	3.96
F_ValueOrientation	v_VO	0.07740	0.01850	4.18
F_Satisfaction	v_S	0.05850	0.01094	5.35
e1	alpha	0.00745	0.00107	6.96
e2	alpha	0.00745	0.00107	6.96
e3	beta	0.04050	0.00582	6.96
e4	beta	0.04050	0.00582	6.96
e5	gamma	0.08755	0.01257	6.96
e6	gamma	0.08755	0.01257	6.96
e7	delta	0.03517	0.00505	6.96
e8	delta	0.03517	0.00505	6.96
d1	v_d1	0.00565	0.00213	2.66

105

Warren9variables: Standardized Estimates

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

Considerable measurement error in these variables!

Squared Multiple Correlations

	Variable	Error Variance	Total Variance	R-Square
1	Performance_1	0.00745	0.02465	0.6978
2	Performance_2	0.00745	0.02465	0.6978
3	Knowledge_1	0.04050	0.07220	0.4391
4	Knowledge_2	0.04050	0.07220	0.4391
5	ValueOrientation_1	0.08755	0.16495	0.4692
6	ValueOrientation_2	0.08755	0.16495	0.4692
7	Satisfaction_1	0.03517	0.09367	0.6245
8	Satisfaction_2	0.03517	0.12644	0.7218
9	F_Performance	0.00565	0.01720	0.6715

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

106

Example 4: Summary

Tasks accomplished:

1. Set up model to study effect of measurement error in regression
2. Used split-half version of original variables as parallel tests
3. Fixed parameters according to measurement model (just typed in their fixed values)
4. Obtained an acceptable model
5. Found that predictability of `JobPerformance` could potentially be as high as R-square=0.67

108

1.7 Conclusions

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

110

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

111

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

112