Fitting Structural Equation Models by the CALIS Procedure†

Yiu-Fai Yung, PhD
Senior Research Statistician
SAS Institute Inc.

† Presented at the SAS Halifax Regional User Group Meeting, October 24, 2008, Halifax, Nova Scotia

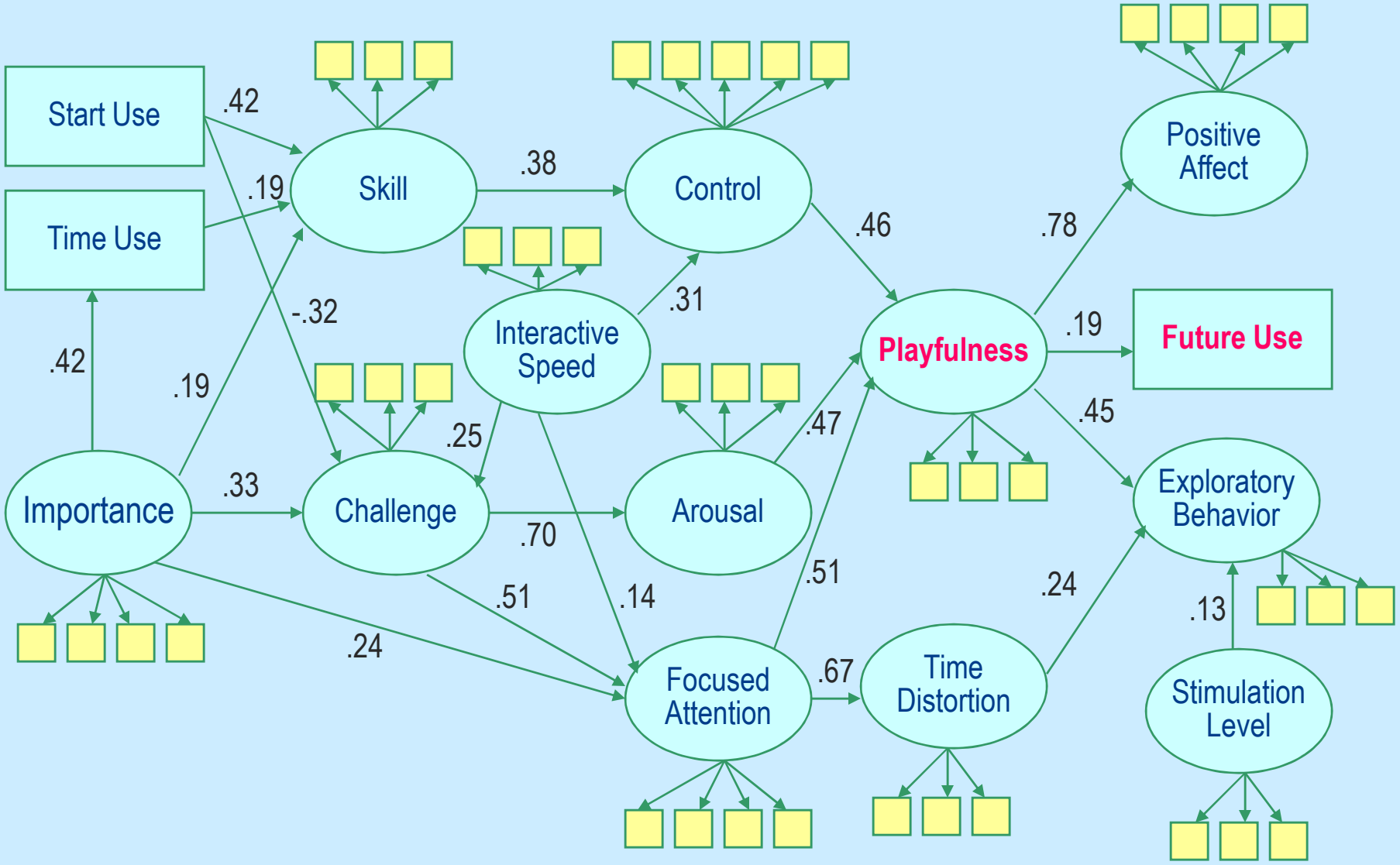
What is Structural Equation Modeling (SEM)?

- Also known as: covariance structure analysis, causal modeling, path analysis, LISREL, etc.
- Statistically: a moment fitting technique (especially variances and covariances) to estimate parameters
- SEM is what PROC CALIS (**C**ovariance **A**nalysis of **LI**near **S**tructural Equations) does

Example. A Structural Equation Model of Web-Surfing Behavior

- Modified from Nowak, Hoffman, and Yung (2000)
- A global theory about web-surfing behavior
- What will make the surfers come back?

The Path Diagram for the Structural Equation Model of Web-Surfing Behavior



Key Features of SEM

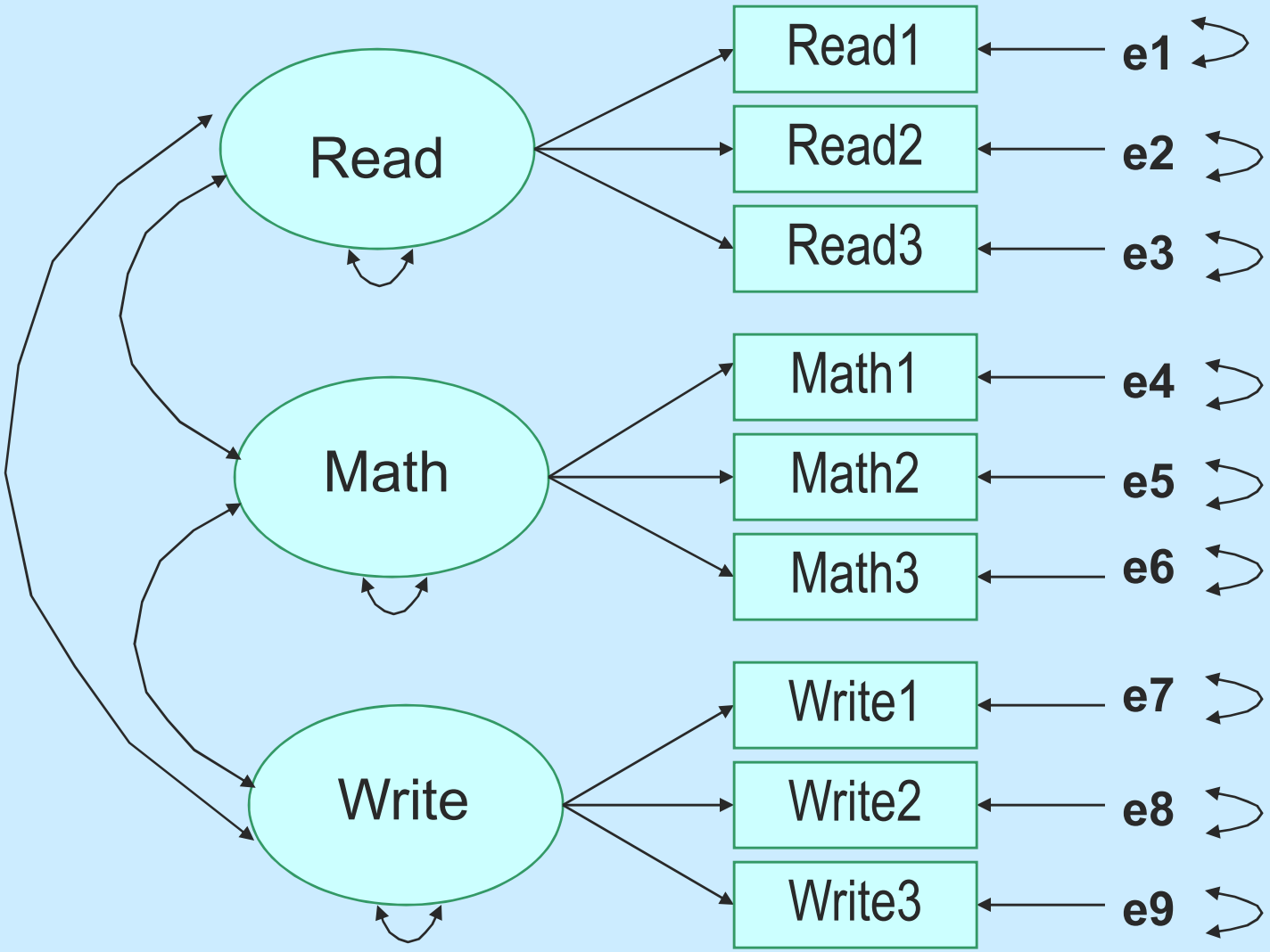
- Analyzing complicated relationships among variables
- Path diagram representations for structural equation models
- The ability to handle latent and observed variables simultaneously
- Testing the model fit and significance of the parameters
- Suggesting ways to improve the model

Confirmatory Factor Analysis

- Confirmatory factor model: precursor of the structural equation model

- Confirming the relationships between factors (latent constructs) and their indicators (observed variables)

Confirmatory Factor Model of Cognitive skills



Example. Consumer Rating and Spending

- Consumer spending in an online store during 2002 and 2003 were recorded
- Two main (latent) factors of spending:

1. service quality
2. product quality

Measuring Service and Product Qualities

- At the end of 2002
- Customers were asked to rate their buying experience
- Seven-point scales on seven items

Seven Items

- Courtesy of customer service
- Responsiveness of customer service
- Helpfulness of customer service
- Delivery of product
- Product pricing
- Product availability
- Product satisfaction

Indicators of Service Quality

Indicators of Product Quality

Samples of Items

- Courtesy of customer service

The customer service representatives are courteous.

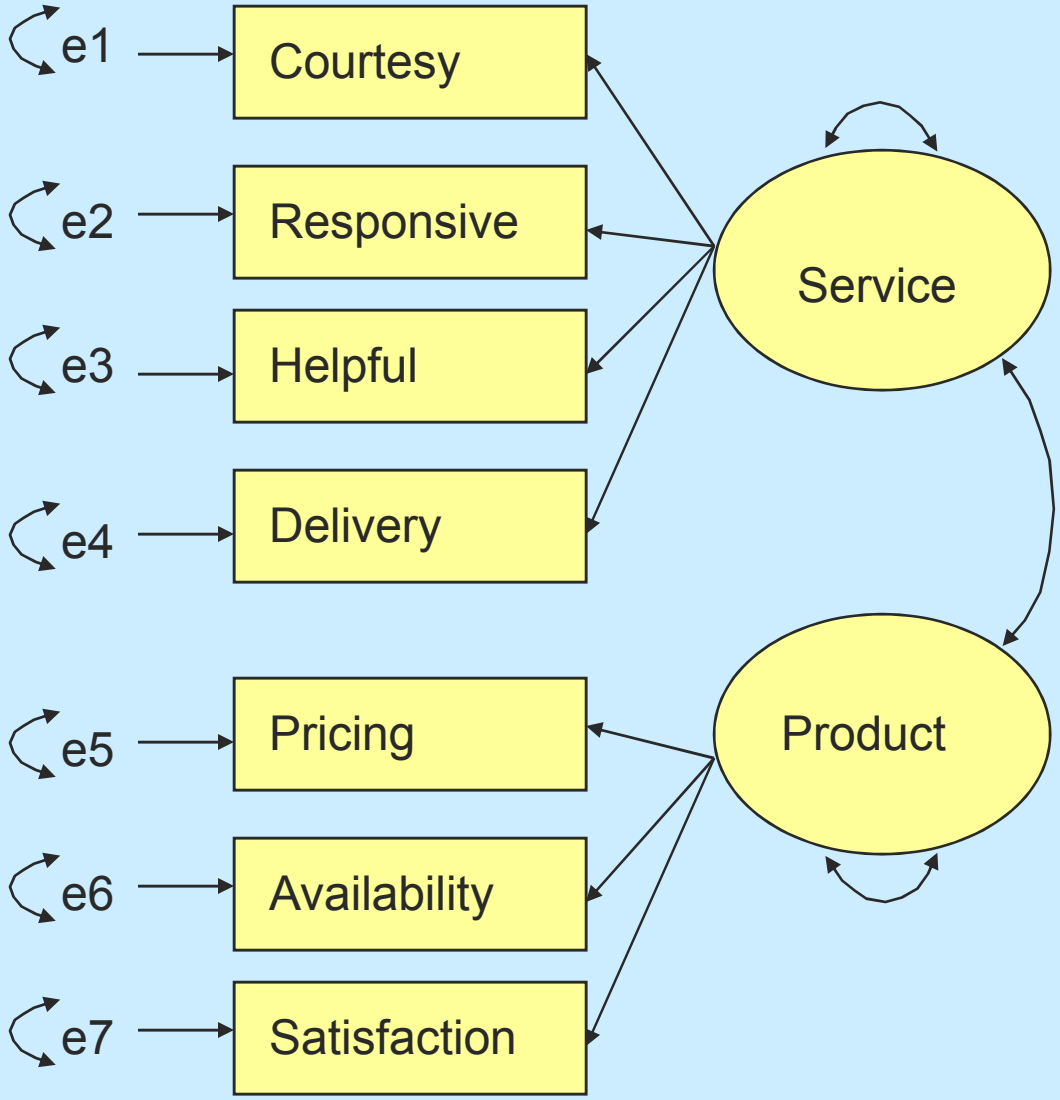
1	2	3	4	5	6	7
Strongly Disagree	Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Agree	Strongly Agree

- Product Pricing

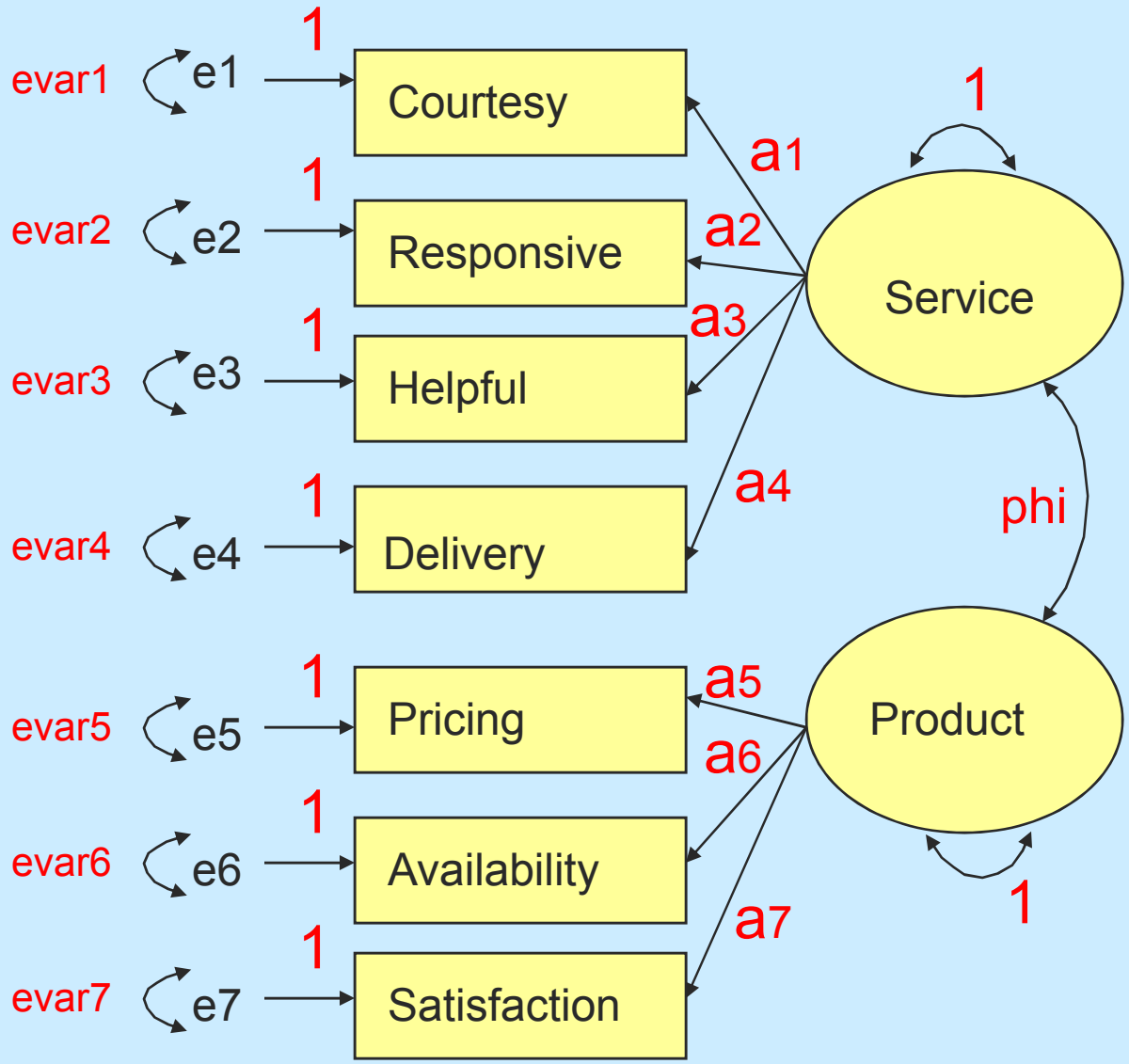
The prices are competitive.

1	2	3	4	5	6	7
Strongly Disagree	Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Agree	Strongly Agree

A Confirmatory Factor Model for Consumer Ratings



What are the Parameters?



Covariance Matrix for the Consumer Data

```
data spending(type=cov);
   input _type_ $6. _name_ $12. Spend02 Spend03 Courtesy Responsive
         Helpful Delivery Pricing Availability Satisfaction;
   datalines;
COV    Spend02       14.428  2.206  0.439 0.520 0.459 0.498 0.635 0.642 0.769
COV    Spend03        2.206 14.178  0.540 0.665 0.560 0.622 0.535 0.588 0.715
COV    Courtesy       0.439  0.540  1.642 0.541 0.473 0.506 0.109 0.120 0.126
COV    Responsive     0.520  0.665  0.541 2.977 0.582 0.629 0.119 0.253 0.184
COV    Helpful        0.459  0.560  0.473 0.582 2.801 0.546 0.113 0.121 0.139
COV    Delivery       0.498  0.622  0.506 0.629 0.546 3.830 0.120 0.132 0.145
COV    Pricing        0.635  0.535  0.109 0.119 0.113 0.120 2.152 0.491 0.538
COV    Availability   0.642  0.588  0.120 0.253 0.121 0.132 0.491 2.372 0.589
COV    Satisfaction   0.769  0.715  0.126 0.184 0.139 0.145 0.538 0.589 2.753
MEAN       .        183.500 301.921 4.312 4.724 3.921 4.357 6.144 4.994 5.971
N                     578     578    578   578   578   578   578   578   578
;
```

You can use raw data as well.

CFA Model for Consumer Rating

```
proc calis cov data=spending;
   lineqs
      Courtesy      = a1 F_Service + e1,
      Responsive    = a2 F_Service + e2,
      Helpful       = a3 F_Service + e3,
      Delivery      = a4 F_Service + e4,
      Pricing       = a5 F_Product + e5,
      Availability  = a6 F_Product + e6,
      Satisfaction  = a7 F_Product + e7;
   std
      F_Service     = 1.,
      F_Product     = 1.,
      e1-e7         = evar1-evar7;
   cov
      F_Service F_Product = phi;
```

Specify variable relations: Single-headed paths and effect parameters

Specify variances: Double-headed arrows pointing to individual variables

Specify covariances: Double-headed arrows pointing to paired-variables

Fit Summary of the Confirmatory Factor Model

Chi-Square	1.1571	
Chi-Square DF	13	
Pr > Chi-Square	1.0000	NS
GFI Adjusted for Degrees of Freedom (AGFI)	0.9988	> .9
Bentler's Comparative Fit Index	1.0000	> .9
Standardized Root Mean Square Residual (SRMR)	0.0062	< .05
RMSEA Estimate	0.0000	< .05

All fit indices show good model fit.

The Measurement Model for the Service Factor

Courtesy	=	0.6581*F_Service	+	1.0000 e1
Std Err		0.0800 a1		
t Value		8.2254		
Responsive	=	0.8283*F_Service	+	1.0000 e2
Std Err		0.1054 a2		
t Value		7.8556		
Helpful	=	0.7100*F_Service	+	1.0000 e3
Std Err		0.0995 a3		
t Value		7.1378		
Delivery	=	0.7631*F_Service	+	1.0000 e4
Std Err		0.1151 a4		
t Value		6.6299		

The Measurement Model for the Product Factor

```
Pricing      =   0.6632*F_Product    +  1.0000 e5
Std Err          0.0985 a5
t Value          6.7353

Availability =   0.7435*F_Product    +  1.0000 e6
Std Err          0.1071 a6
t Value          6.9409

Satisfaction =   0.7993*F_Product    +  1.0000 e7
Std Err          0.1153 a7
t Value          6.9346
```

Squared Multiple Correlations

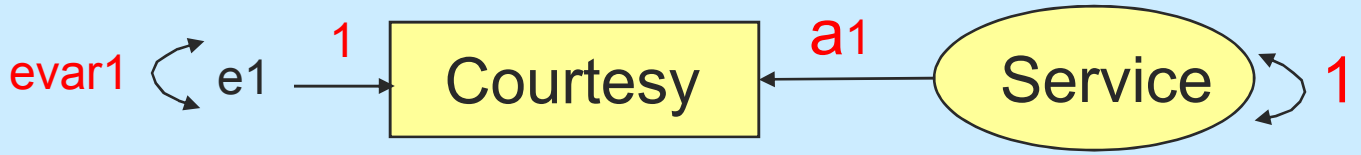
Squared Multiple Correlations

	Variable	Error Variance	Total Variance	R-Square
1	Courtesy	1.20895	1.64200	0.2637
2	Responsive	2.29091	2.97700	0.2305
3	Helpful	2.29683	2.80100	0.1800
4	Delivery	3.24762	3.83000	0.1521
5	Pricing	1.71213	2.15200	0.2044
6	Availability	1.81921	2.37200	0.2330
7	Satisfaction	2.11409	2.75300	0.2321

Key Elements in Specifying a CFA Model in PROC CALIS

- Setting up equations for latent factors:

```
lineqs Courtesy = a1 F_Service + e1;
```

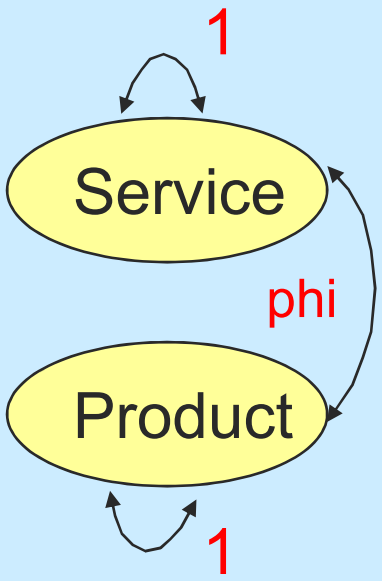


- Use the STD statement to specify variance:

```
std   e1 = evar1,
      F_Service = 1;
```

- Use the COV statement to specify covariance:

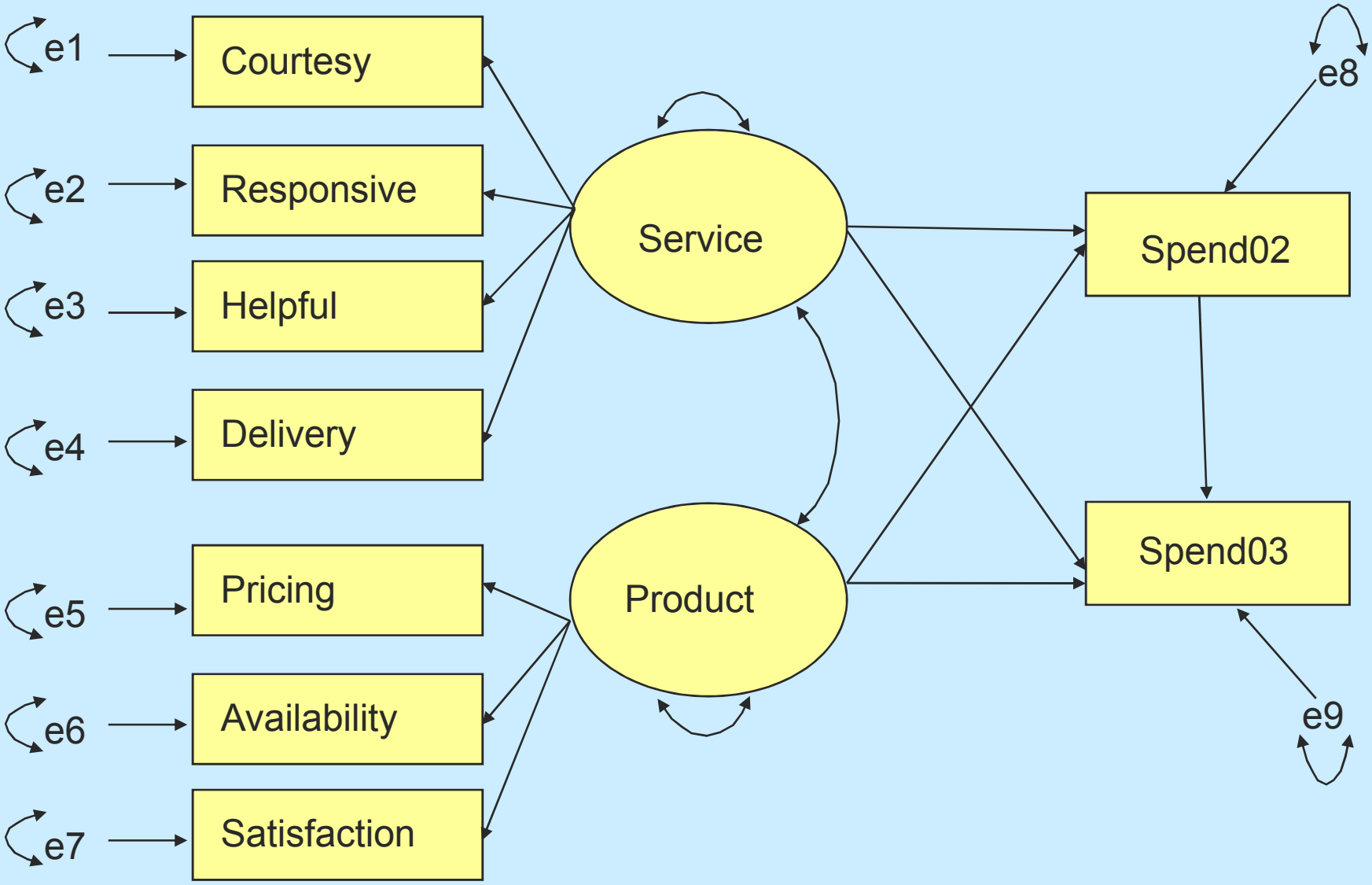
```
cov   F_Service F_Product = phi;
```



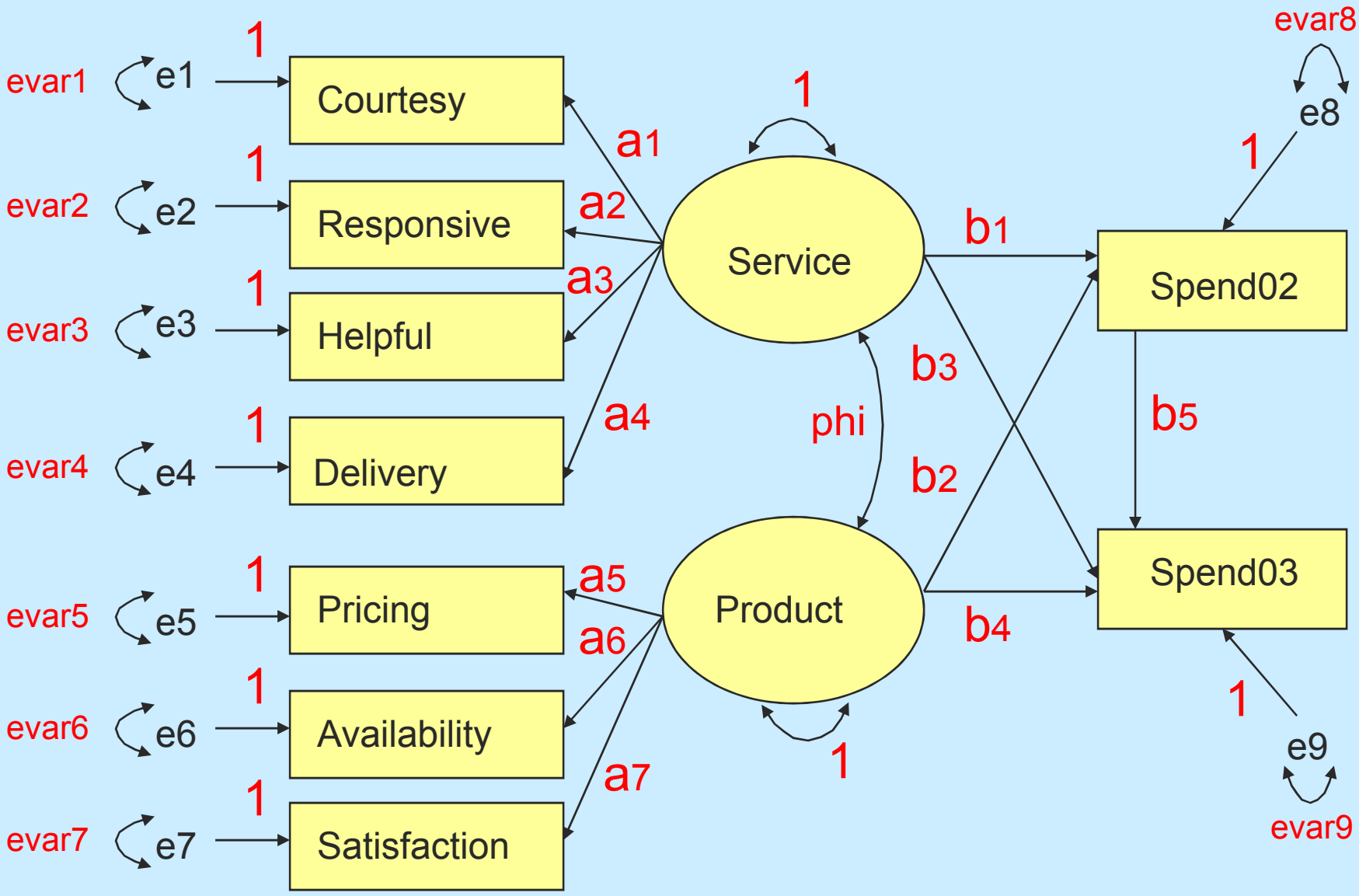
Key Results of a CFA Model

- Fit Summary: Insignificant model fit chi-square is desired; but many other fit indices are more important to look at
- Do the measured indicators have significant relationships with the latent factors?

Factors Affecting Consumer Spending



What are the Parameters?



Model for Consumer Spending

```
proc calis cov data=spending;
   lineqs
      Spend02 = b1 F_Service + b2 F_Product + e8,
      Spend03 = b3 F_Service + b4 F_Product + b5 Spend02 + e9,
      Courtesy      = a1 F_Service + e1,
      Responsive    = a2 F_Service + e2,
      Helpful       = a3 F_Service + e3,
      Delivery      = a4 F_Service + e4,
      Pricing       = a5 F_Product + e5,
      Availability = a6 F_Product + e6,
      Satisfaction = a7 F_Product + e7;
   std
      F_Service    = 1.,
      F_Product    = 1.,
      e1-e9        = evar1-evar9;
   cov
      F_Service F_Product = phi;
```

Fit Summary of the Consumer Spending Model

Chi-Square	1.3265	
Chi-Square DF	23	
Pr > Chi-Square	1.0000	NS
GFI Adjusted for Degrees of Freedom (AGFI)	0.9990	> .9
Bentler's Comparative Fit Index	1.0000	> .9
Standardized Root Mean Square Residual (SRMR)	0.0074	< .05
RMSEA Estimate	0.0000	< .05

All fit indices show good model fit.

Manifest Variable Equations with Estimates

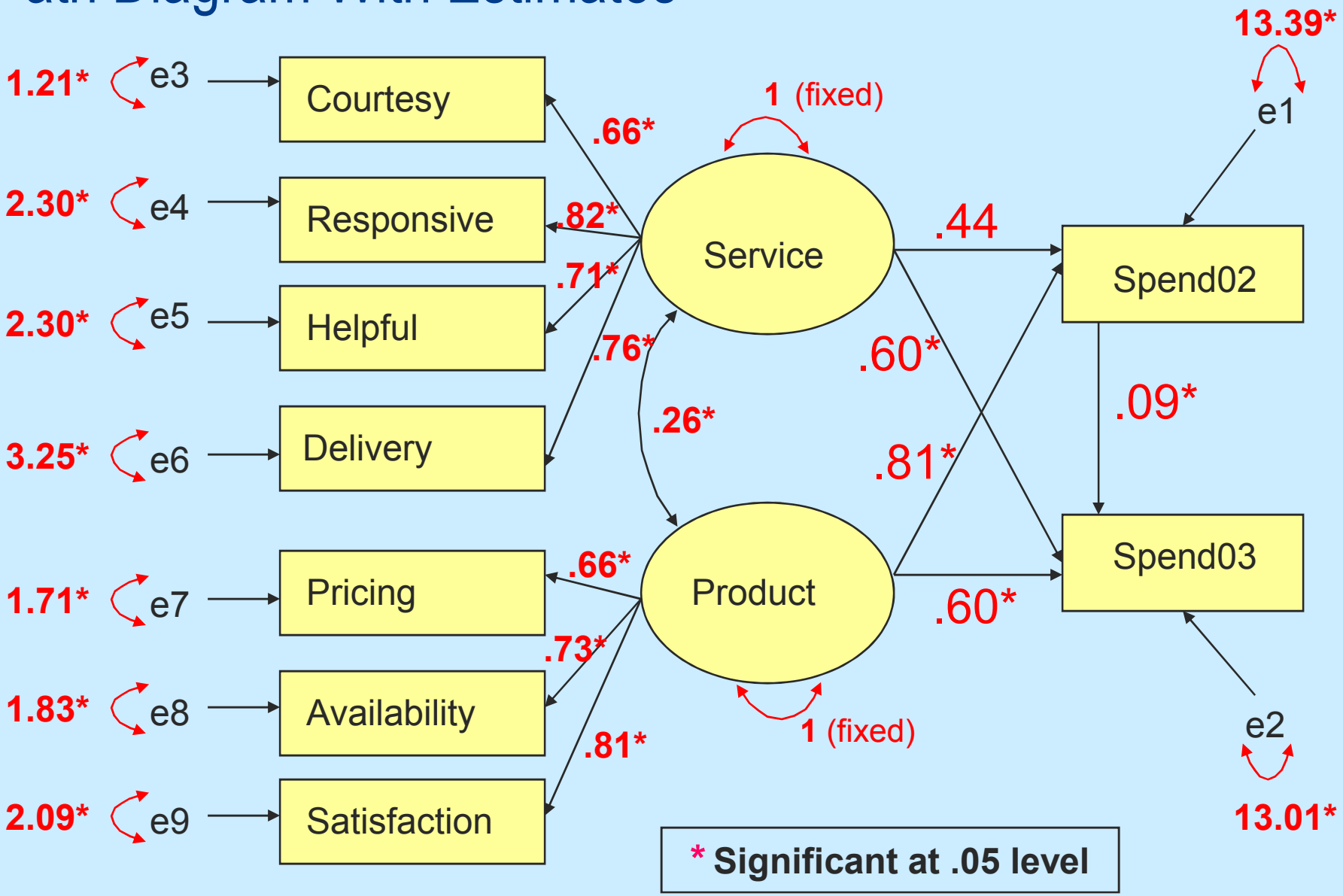
```
Spend02       =    0.4389*F_Service     +   0.8132*F_Product
Std Err            0.2355 b1                 0.2452 b2
t Value            1.8635                    3.3159

        +   1.0000 e8
```

This is "marginally significant."

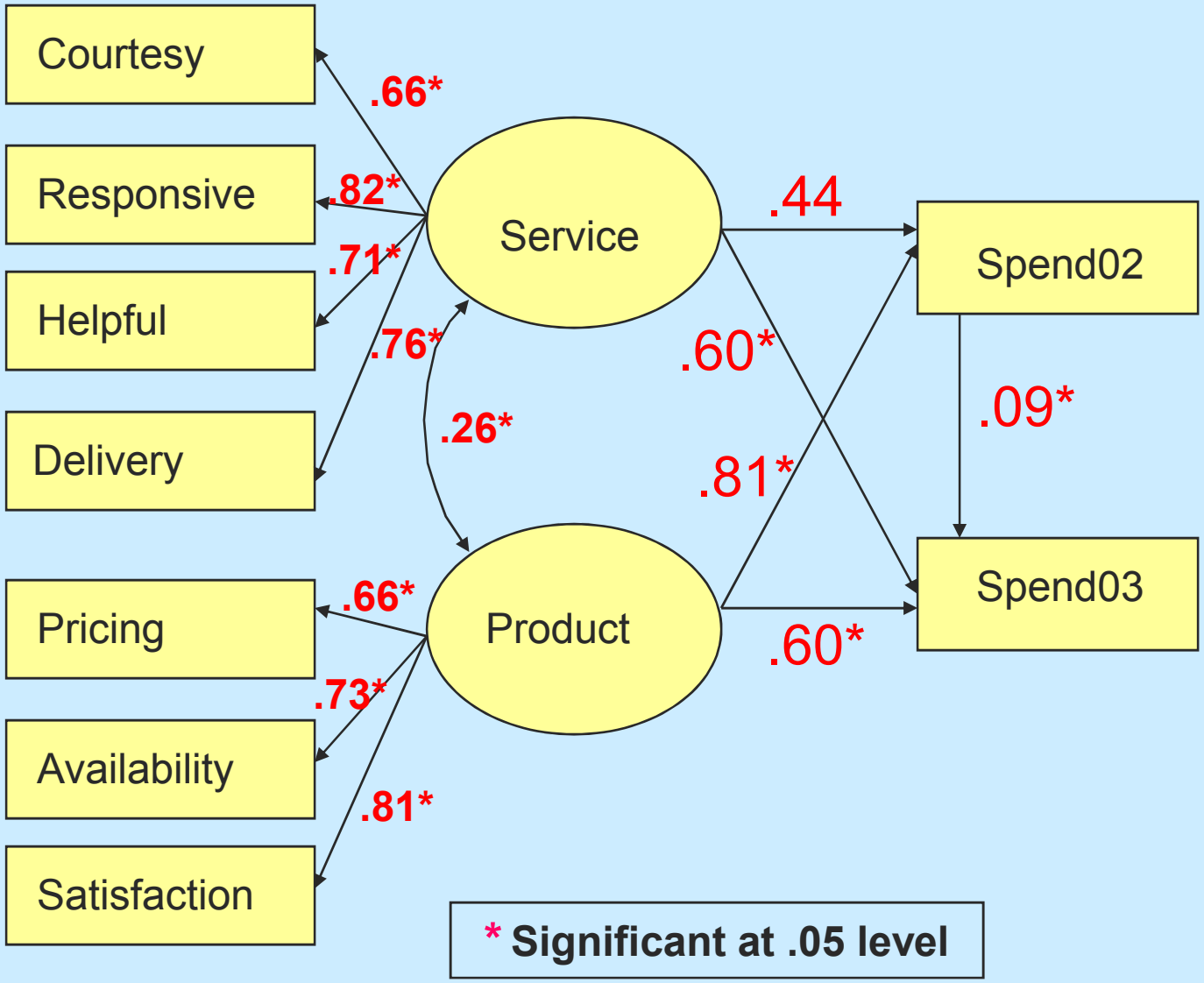
```
Spend03       =    0.0877*Spend02       +   0.5966*F_Service
Std Err            0.0438 b5                 0.2315 b3
t Value            2.0036                    2.5775

        +   0.5969*F_Product     +   1.0000 e9
            0.2491 b4
            2.3963
```

Note: Subsequent results not shown.

Path Diagram With Estimates

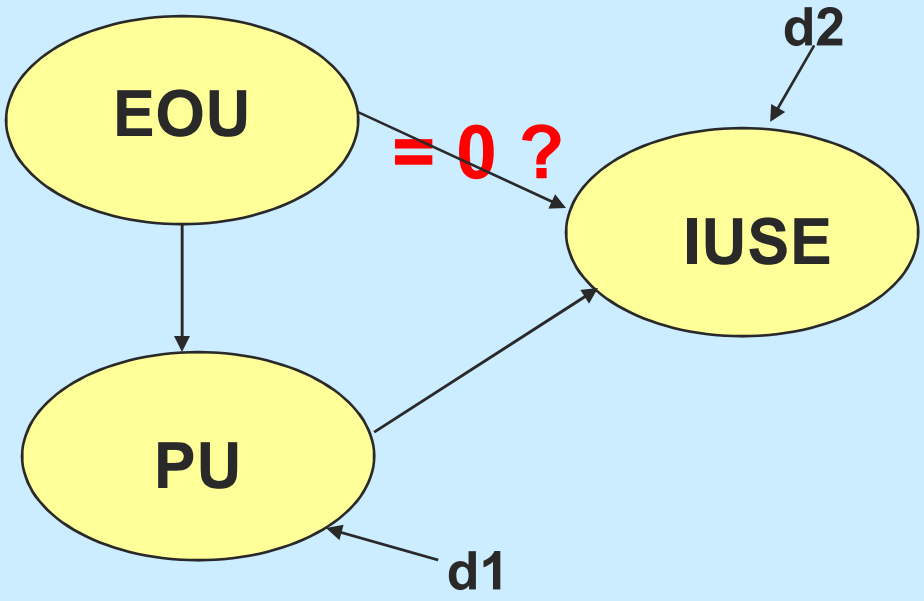


Simplified Path Diagram With Estimates



IS Research Example: Technology Acceptance Model (TAM) (Gefen, Straub, & Boudreau, 2000)

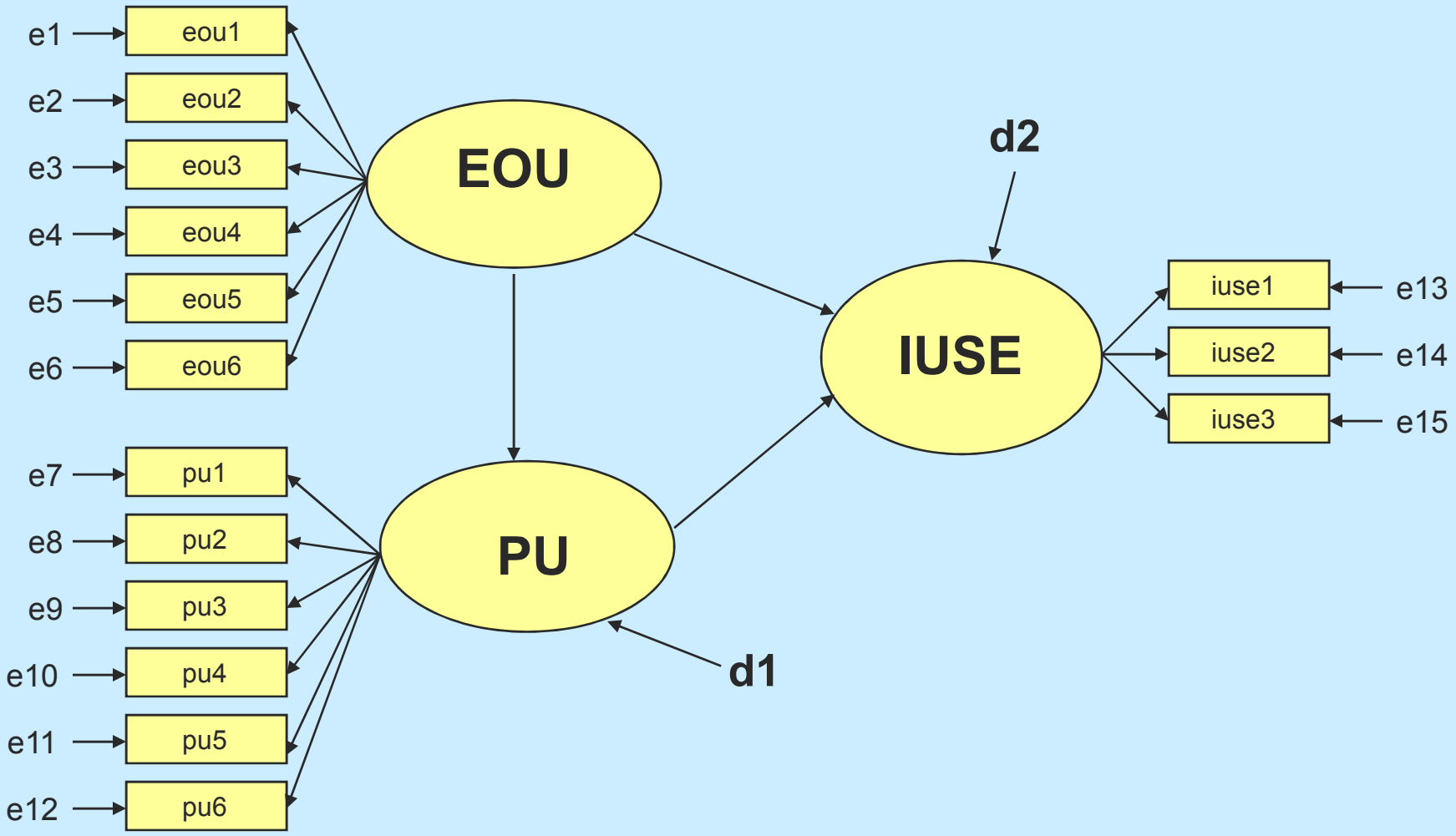
- PU: Perceived Usefulness
- EOU: Perceived Ease of Use
- IUSE: Intention to Use



Data Collection

- 160 Business school students
- Go to www.travelocity.com to search a flight to Heathrow Airport (London) next month
- Fill in a questionnaire containing personal information and rating scales about the web experience
- In this presentation, the data were simulated
- Some results may not correspond to that of the original study

Theoretical TAM Model



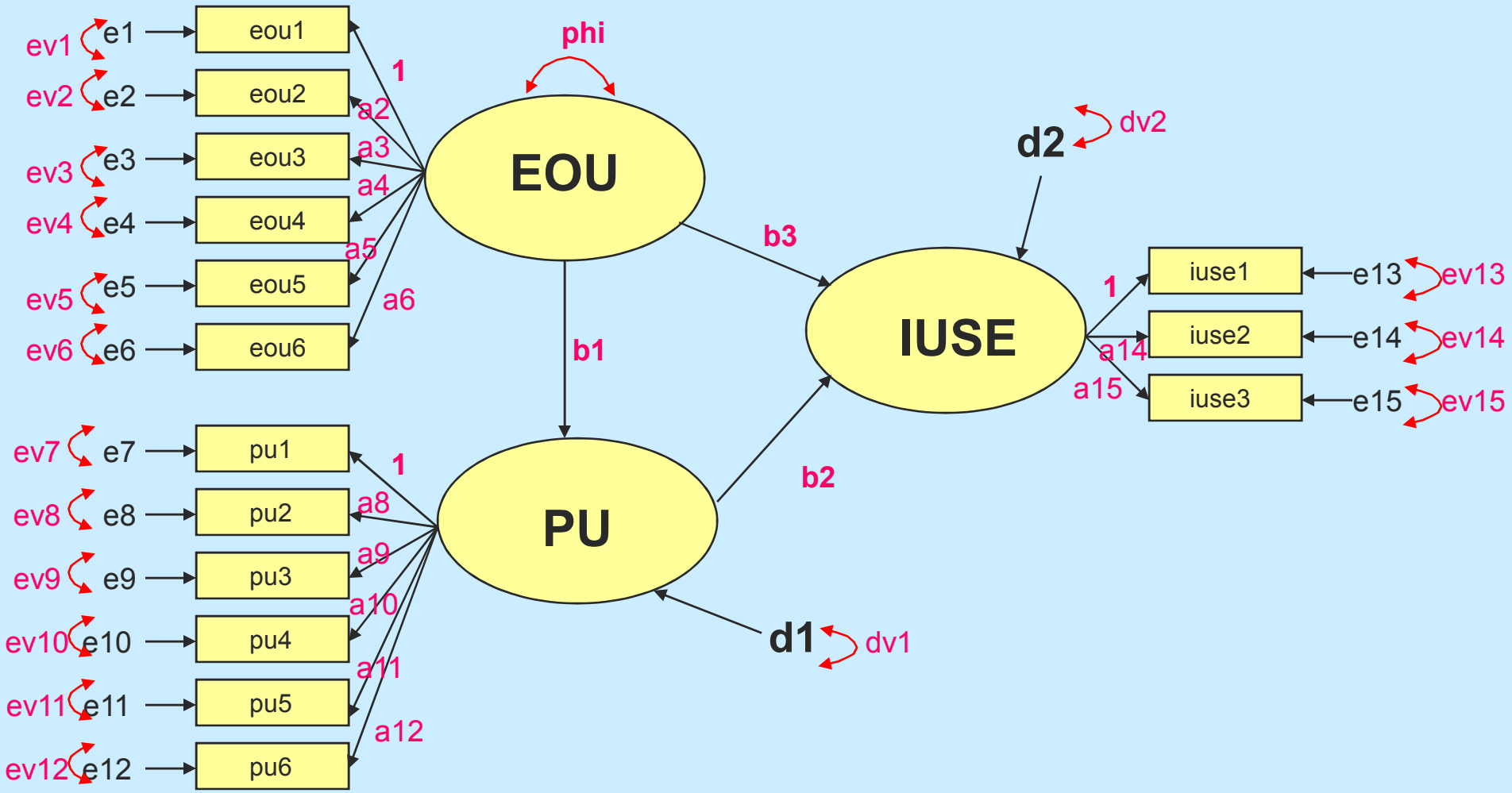
Rating scales

- 1: Strongly Disagree
- 2: Disagree
- 3: Somewhat Disagree
- 4: Neutral
- 5: Somewhat Agree
- 6: Agree
- 7: Strongly Agree

Samples of rating scales:

- EOU1 : Travelocity.com is easy to use.
- EOU4: Travelocity.com is flexible to interact with.
- PU3: Travelocity.com enables me to search and buy flights faster.
- PU6: Travelocity.com increases my productivity in searching and purchasing flights.
- IUSE1: I am very likely to buy books from Travelocity.com.
- IUSE2: I would use my credit card to purchase from Travelocity.com.

A More Complete TAM Path Diagram



TAM Specified by CALIS

```
proc calis data=tam cov;
   lineqs
      eou1   = 1.   f_eou  + e1,
      eou2   = a2   f_eou  + e2,
      eou3   = a3   f_eou  + e3,
      eou4   = a4   f_eou  + e4,
      eou5   = a5   f_eou  + e5,
      eou6   = a6   f_eou  + e6,
      pu1    = 1.   f_pu   + e7,
      pu2    = a8   f_pu   + e8,
      pu3    = a9   f_pu   + e9,
      pu4    = a10  f_pu   + e10,
      pu5    = a11  f_pu   + e11,
      pu6    = a12  f_pu   + e12,
```

Measurement Model for Ease of Use (EOU) and Perceived Usefulness (PU)

TAM Specified by CALIS (Continued)

```
   iuse1  = 1.   f_iuse + e13,
   iuse2  = a14  f_iuse + e14,
   iuse3  = a15  f_iuse + e15,

   f_pu   = b1   f_eou  + d1,
   f_iuse = b2   f_pu   + b3 f_eou  + d2;

std
   f_eou  = phi ,
   e1-e15 = ev1-ev15,
   d1-d2  = dv1-dv2 ;
run;
```

Measurement Model for Intention to Use (IUSE)

Structural Equations

Variance Parameters

Fit Summary of the TAM

Chi-Square	141.2017	
Chi-Square DF	87	
Pr > Chi-Square	0.0002	< .05
GFI Adjusted for Degrees of Freedom (AGFI)	0.8572	< .9
Bentler's Comparative Fit Index	0.9017	> .9
Standardized Root Mean Square Residual (SRMR)	0.0653	≈ .05
RMSEA Estimate	0.0626	≈ .05

Model Fits reasonably well, but can be improved.

Measurement Model for Perceived Ease of Use (EOU)

```
eou1     =   1.0000 f_eou     +  1.0000 e1

eou2     =   0.9909*f_eou     +  1.0000 e2
Std Err      0.1621 a2
t Value      6.1119

eou3     =   1.0154*f_eou     +  1.0000 e3
Std Err      0.1598 a3
t Value      6.3551

eou4     =   1.1128*f_eou     +  1.0000 e4
Std Err      0.1642 a4
t Value      6.7750

eou5     =   0.9054*f_eou     +  1.0000 e5
Std Err      0.1446 a5
t Value      6.2623

eou6     =   0.8757*f_eou     +  1.0000 e6
Std Err      0.1382 a6
t Value      6.3378
```

All EOU indicators are significantly related to the EOU factor

Measurement Model for Perceived Usefulness (PU)

```
pu1      =   1.0000 f_pu      +  1.0000 e7

pu2      =   1.3257*f_pu      +  1.0000 e8
Std Err      0.1988 a8
t Value      6.6686

pu3      =   1.3133*f_pu      +  1.0000 e9
Std Err      0.2091 a9
t Value      6.2821

pu4      =   1.0743*f_pu      +  1.0000 e10
Std Err      0.1663 a10
t Value      6.4584

pu5      =   1.1723*f_pu      +  1.0000 e11
Std Err      0.1771 a11
t Value      6.6197

pu6      =   4.9149*f_pu      +  1.0000 e12
Std Err      4.4952 a12
t Value      1.0934
```

NS

PU6 may not be a good indicator of the PU factor

Measurement Model for Intention to Use (IUSE)

```
iuse1    =    1.0000 f_iuse    +   1.0000 e13

iuse2    =    0.3088*f_iuse    +   1.0000 e14
Std Err       0.1856 a14
t Value       1.6633

iuse3    =    0.5922*f_iuse    +   1.0000 e15
Std Err       0.2397 a15
t Value       2.4712
```

NS

IUSE2 may not be a good indicator of the IUSE factor

Structural Equations

```
f_pu     =    0.5086*f_eou     +  1.0000 d1
Std Err       0.1065 b1
t Value       4.7745

f_iuse   =    0.4580*f_pu      +  0.2221*f_eou     +  1.0000 d2
Std Err       0.1808 b2           0.1514 b3
t Value       2.5335              1.4667
```

NS

Direct effect of Perceived Ease of Use (EOU) on Intention to Use (IUSE) is not significant.

Remove the “Bad” Items and the Weak EOU >>> IUSE Path

```
proc calis data=tam cov;
   lineqs
      eou1   = 1.   f_eou  + e1,
      eou2   = a2   f_eou  + e2,
      eou3   = a3   f_eou  + e3,
      eou4   = a4   f_eou  + e4,
      eou5   = a5   f_eou  + e5,
      eou6   = a6   f_eou  + e6,
      pu1    = 1.   f_pu   + e7,
      pu2    = a8   f_pu   + e8,
      pu3    = a9   f_pu   + e9,
      pu4    = a10  f_pu   + e10,
      pu5    = a11  f_pu   + e11,
```

Remove the "Bad" Items and the Weak EOU >>> IUSE Path (Continued)

```
   /*pu6    = a12  f_pu   + e12,*/
   iuse1   = 1.   f_iuse + e13,
   /*iuse2  = a14  f_iuse + e14, */
   iuse3   = a15  f_iuse + e15,
   f_pu    = b1   f_eou  + d1,
   f_iuse  = b2   f_pu   /*+ b3 f_eou*/  + d2;
std
   f_eou   = phi ,
   e1-e11 e13 e15 = ev1-ev11 ev13 ev15,
   d1-d2   = dv1-dv2 ;
run;
```

Take away the weak indicators

Take away the direct effect of Ease of Use

Take away e12 and e14

Model Fit of the Revised TAM Model

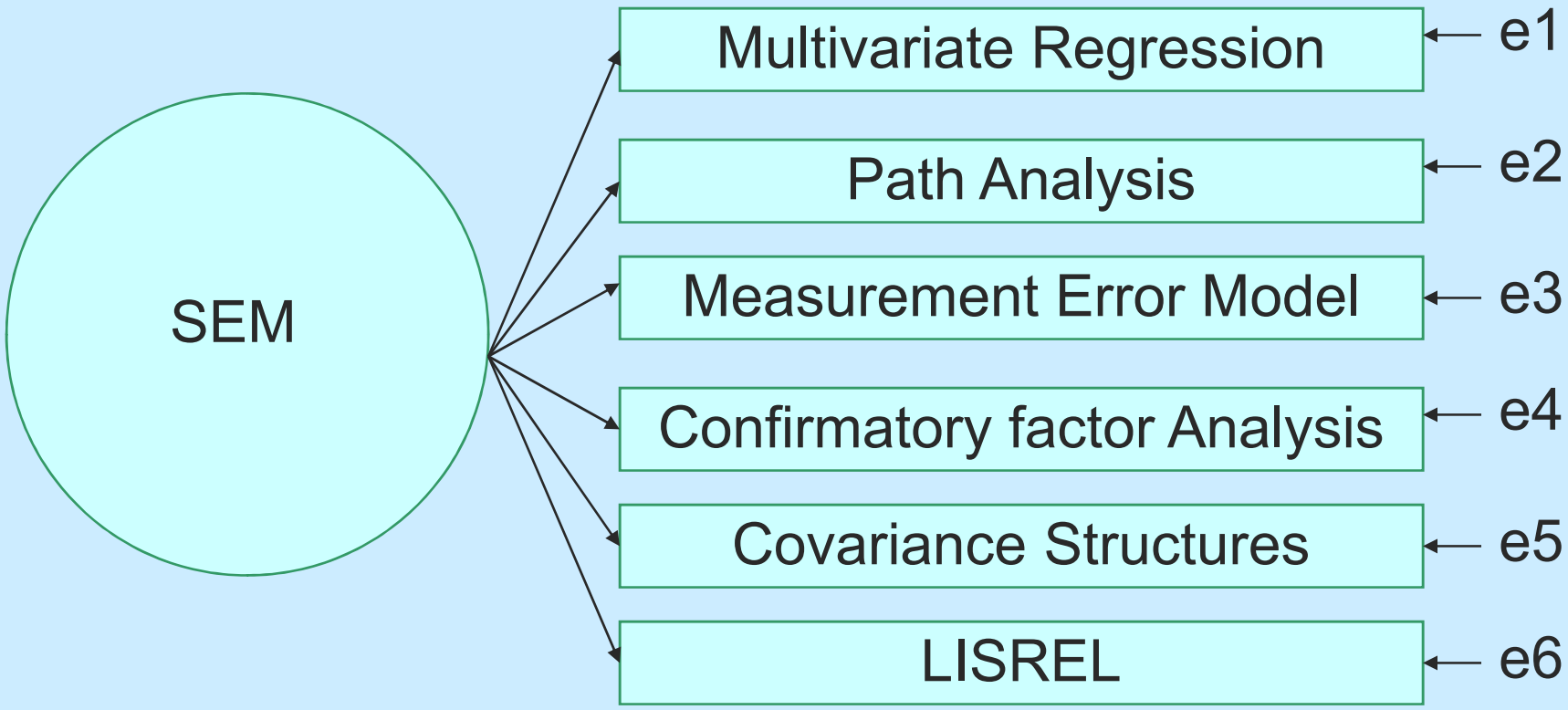
		"Old "
Chi-Square	98.5493	141.2017
Chi-Square DF	63	87
Pr > Chi-Square	0.0028	0.0002
GFI Adjusted for Degrees of Freedom (AGFI)	0.8826	0.8572
Bentler's Comparative Fit Index	0.9328	0.9017
Standardized Root Mean Square Residual (SRMR)	0.0581	0.0653
RMSEA Estimate	0.0596	0.0626

The revised model fits better!

Key SEM Concepts With the Use of CALIS

- Effects from one variable to another can be direct or indirect
- You can test the direct effects using CALIS
- You can use CALIS results to probe weak paths or measurement items (insignificant effects)
- The improved model needs to be validated on new data

What is Structural Equation Modeling?



What Can PROC CALIS Do For You?

- Fit a Large Class of Structural Equation Models
- LINEQS is one of the modeling languages in CALIS for specifying SEM
- Several estimation methods can be chosen
- Fit statistics for assessing model fit
- Estimates with standard errors
- Model modification statistics and residual analysis
- Others …

References

Gefen, D., Straub, D. W., & Boudreau, M. (2000). Structural equation modeling and regression: Guidelines for research practice. *Communications of AIS, 7, 1-78.*

Nowak, T. P., Hoffman, D. L., and Yung, Y. F. (2000). Measuring the customer experience in online environments: A structural modeling approach, *Marketing Science*, *19(1)*, 22-42.

Textbooks

Bollen, K. A. (1989), *Structural Equations with Latent Variables*. New York: John Wiley & Sons.

Loehlin, J. C. (1987), *Latent Variable Models, An Introduction to Factor, Path, and Structural Analysis.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Glossary

MANIFEST – Observed variables (measured variables) in the data set; called indicators when they are related to latent factors.

LATENT – Unobserved variables; factors (F-prefix), errors (E-prefix), or disturbances (D-prefix); factors are evidenced by the manifest indicators.

ENDOGENOUS – Dependent /mediating variables; at least one single-headed arrow points to it; used as an outcome variable in an equation; can also be a predictor variable in other equations.

EXOGENOUS – Independent variables; no single-headed arrows point to it; never used as an outcome variable in the model; used only as a predictor in the model.

FACTOR – A latent (unmeasured) variable that is treated as a hypothetical construct (systematic source) in the model.

ERROR – An exogenous term for uncertainty (unsystematic source) associated with an endogenous *manifest* variable (E-prefix).

DISTURBANCE – An exogenous term for uncertainty (unsystematic source) associated with an endogenous *latent* variable (D-prefix).

PATH DIAGRAM REPRESENTATION

– Rectangles: Observed / Manifest variables.

– Ovals / Circles : latent variables (factors, errors, and disturbances). Errors and disturbances are not necessarily put into ovals/circles.

– Single-headed arrows: directed paths; direct effects; path coefficients; specified in the LINEQS statement.

– Double-headed arrows that point to individual variables: Variance parameters of exogenous variables; specified in the STD statement.

– Double-headed arrows that point to two distinct variables: Covariance parameters between exogenous variables; specified in the COV statement.

FIT ASSESSMENT

– Model fit chi-square statistic: Non-significance means that the theoretical model is supported; not a very practical index because it almost always rejects all approximating models that are practically useful.

– AGFI (Adjusted Goodness-of-fit index) and Bentler's CFI (Comparative fit index): Two popular fit indices which indicate good model fit when their values are above .9.

– SRMR (Standardized Root Mean Square Residual) and RMSEA (Root Mean Squared Error Approximation): Two popular fit indices which indicate good model fit when their values are below .05.