5 Structural Equation Modeling 2: General Structural Models

References:

• Beaujean (2014). Chapter 3.

5.1. Latent Variable Analysis

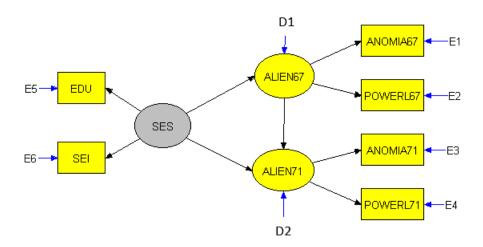
• Example 1. Stability of Alienation (Wheaton et al., 1977)

• Filename: *alien.cov* (N = 932)

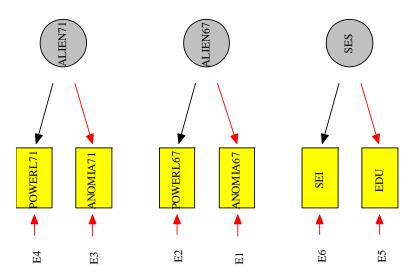
	ANOMIA67	POWERL67	ANOMIA71	POWERL71	EDUC	SEI
ANOMIA67	11.83					
POWERL67	6.95	9.36				
ANOMIA71	6.82	5.09	12.53			
POWERL71	4.78	5.03	7.50	9.99		
EDUC	-3.84	-3.89	-3.84	-3.63	9.61	
SEI	-2.19	-1.88	-2.72	-1.88	3.55	4.50

ANOMIA67, ANOMIA71=a scale measured feeling of isolation (in 1967 and 1971 POWERL67, POWERL71=a scale measured powerlessness in 1967 and 1971 EDUC=education level, SEI=socioeconomic index

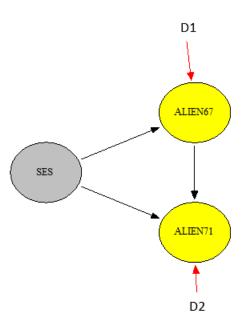
• In general, a latent variable model consists of (1) the measurement part, and (2) the structural part



• The measurement part (CFA) postulates the relationship between the factors and their indicators:



• The structural part postulates the structural relationship among the factors:



- Suppose there is a set of variables v that can be characterized by its population covariance matrix, Σ . The *structural model* represents our belief about how Σ is structured. That is, $\Sigma = \Sigma(\theta)$.
- Basic tasks involve
 - parameter estimation
 - model evaluation
 - model modification
- We follow the same 5-step procedure as in CFA

5.2. Model Specification

• Measurement equation:

$$v = \mu + \Lambda f + e$$

```
v is p \times 1 vector of observed variables f = \begin{pmatrix} f_x \\ f_y \end{pmatrix} is k \times 1 vector of latent factors, k = k_x + k_y e is p \times 1 vector of measurement errors is p \times 1 vector of intercepts of v is p \times k factor loading matrix
```

• Structural equation:

$$\begin{bmatrix} f_x \\ f_y \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ \gamma & \beta \end{bmatrix} \begin{bmatrix} f_x \\ f_y \end{bmatrix} + \begin{bmatrix} f_x \\ d \end{bmatrix}$$
$$f = Bf + z$$

- f_x is $k_x \times 1$ vector of exogenous factors
- f_y is $k_y \times 1$ vector of endogenous factors
- d is $k_y \times 1$ vector of disturbances
- γ is $k_y \times k_x$ matrix of structural coefficients from f_x to f_y
- β is $k_y \times k_y$ matrix of structural coefficients among f_y

• Combining the measurement and structural equations,

$$v = \mu + \Lambda (I - B)^{-1}z + e$$

• Assuming cov(z,e)=0, the covariance matrix of v is

$$\Sigma = \Lambda (I - B)^{-1} \Psi (I - B)^{-1} \Lambda' + \Theta = \Sigma(\theta)$$

where Λ is the $p \times k$ factor loading matrix, B is the $k \times k$ structural coefficient matrix, Ψ is the $k \times k$ variance covariance matrix of f_x and d, and Θ is the $p \times p$ variance covariance matrix of e.

• The variables are:

Name	Type	Cause/Effect	Dimension
v	observed	endogenous	$p \times 1$
$f_{\rm y}$	latent	endogenous	$k_y \times 1$
f_{x}	latent	exogenous	$k_x \times 1$
e	latent	exogenous	$p \times 1$
d	latent	exogenous	$k_{\rm v} \times 1$

• And the parameter matrices are:

Parameter matrix	Symbol	Name	<u>Dimension</u>
factor loading	Λ	lambda	$p \times 1$
variance-covariance matrix of errors, e	Θ	theta	$p \times p$
structural coefficients among the latent variables, f	B	beta	$k \times k$
variance-covariance matrix of f_x and d	Ψ	psi	$k \times k$

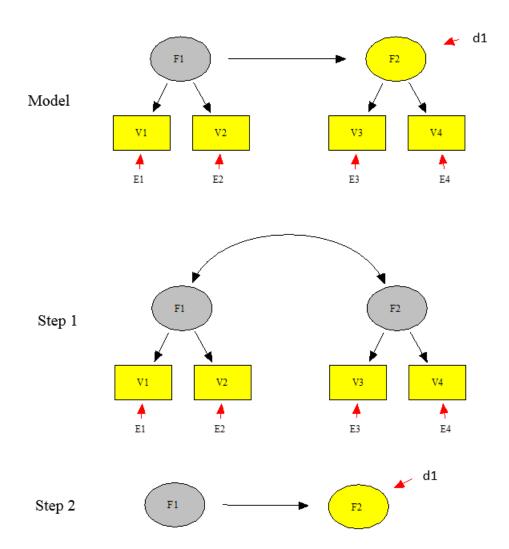
5.3. Identification

- The model $\Sigma = \Sigma(\theta)$ is *identified* if there are no vectors θ^* and θ such that $\Sigma(\theta^*) = \Sigma(\theta)$ unless $\theta^* = \theta$
- As in CFA, we have to fix the scale of the latent variables, which are, f_y , f_x , e, d
- Use ULI for f_y because the variances of f_y do not exist as free parameters in the system
- The *t*-*r*ule remains valid (necessary condition for identification)

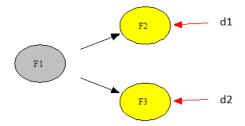
5.3.1. The two-step rule

- Step 1: Reformulate the original model as a measurement (CFA) model and establish identification of measurement model
- Step 2: Establish identification of the latent variable model by treating it as structural equation with observed variables
- A sufficient condition for identification

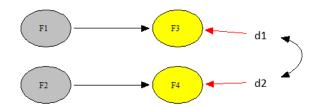
• Example 2



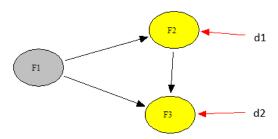
- Some valid Step 2 models (Long, 1983):
- Type I (regression models)



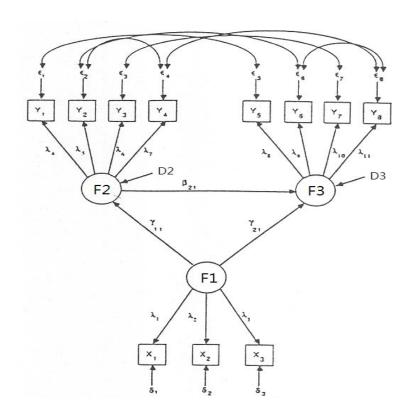
- Type II (seemingly unrelated regression)



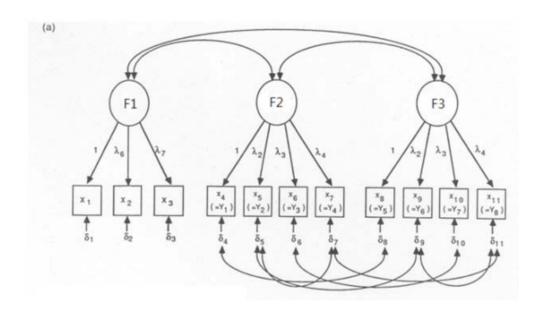
- Type III (recursive models)

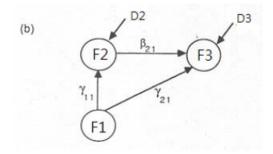


• Example 3: SEM Model of Political Democracy and Industrialization for Developing Countries, 1960 to 1965 (Bollen, 1989):

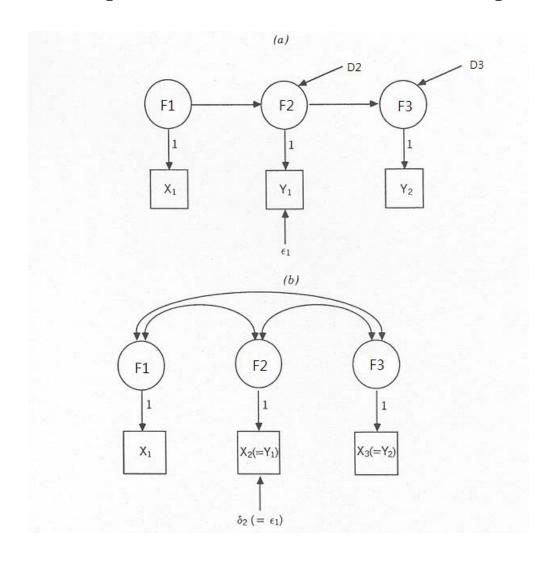


• Same model reformulated for two-step rule:





• Example 4. A model that fails the two-step rule but is identified (Bollen, 1989):



5.4. Estimation

• Estimation methods are basically the same as that in §3.6

5.5. Goodness of Fit Assessment

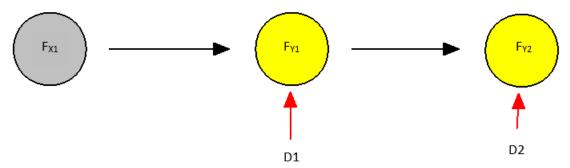
• Same as that in §3.7

5.6. Effect Decomposition

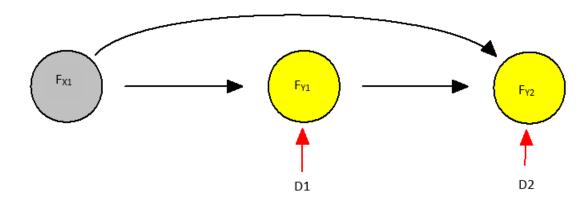
• *Direct effect*: Influence of one variable on another that is unmediated by any other variables in a path model



• *Indirect effect:* Influence of one variable on another is mediated by at least one intervening variable (mediator)



• *Total effect:* The total of the direct effect and all indirect effects



 F_{X1} to F_{Y2} : direct effect = γ_{21} indirect effect = $\gamma_{11}\beta_{21}$ total effect = $\gamma_{21} + \gamma_{11}\beta_{21}$

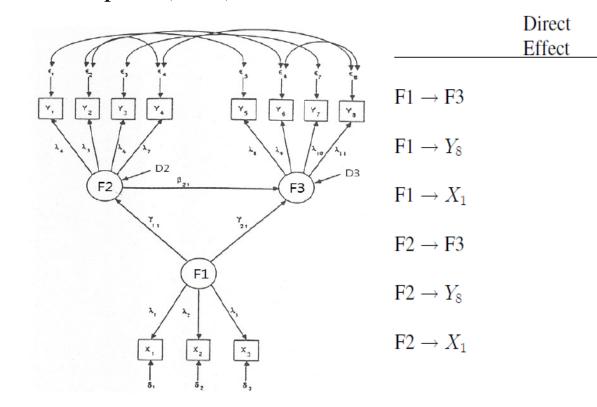
Total

Effect

Indirect

Effect

• Example 3 (cont.):



5.7. Example 1 (cont.)

filenames: alien.R (R script)

```
# Example 1: Stability of Alienation (Wheaton et al., 1977)
# set work directory
setwd("c:/users/wchan/google drive/stat6108/data")
# load the lavaan package
library(lavaan)
# write the input data into a full covariance matrix
alien <- scan("alien.cov")</pre>
alien.cov <- getCov(alien, names=c("anomia67", "powerl67", "anomia71", "powerl71", "educ", "sei"))</pre>
# specify Model 1
model1 <- "
# measurement model
 SES =~ educ + sei
 ALIEN67 =~ anomia67 + power167
 ALIEN71 =~ anomia71 + power171
# structural model
 ALIEN67 ~ a*SES
 ALIEN71 ~ b*ALIEN67 + c*SES
# effect decomposition
 direct := c
 indirect := a*b
 total := direct + indirect
```

```
# Fit Model 1 to data
fit1 <- lavaan(model1, sample.cov=alien.cov, sample.cov.rescale=FALSE, sample.nobs=932, auto.var=TRUE,
auto.fix.first=TRUE)
mil <- modificationIndices(fit1, sort.=TRUE)</pre>
```

```
# specify Model 2 (with error covariances added)
model2 <- "
# measurement model
SES =~ educ + sei
ALIEN67 =~ anomia67 + power167
ALIEN71 =~ anomia71 + power171
# structural model
ALIEN67 ~ a*SES
ALIEN71 ~ b*ALIEN67 + c*SES
# error covariances
anomia67 ~~ anomia71
power167 ~~ power171
# effect decomposition
direct := c
indirect := a*b
total := direct + indirect
# comparing direct vs indirect effect
effect_d := indirect - direct
# fit Model 2 to data
fit2 <- lavaan(model2, sample.cov=alien.cov, sample.cov.rescale=FALSE, sample.nobs=932, auto.var=TRUE,
auto.fix.first=TRUE)
```

```
# save the output
sink("alien.out", split=TRUE)
writeLines("\n Example 1: Stability of Alienation\n")
writeLines("\n Output for Model 1\n")
inspect(fit1)
summary(fit1, fit.measures=TRUE, standardized=TRUE, rsquare=TRUE)
writeLines("\n Modification indices\n")
list(mi1)
writeLines("\n Output for Model 2\n")
inspect(fit2)
summary(fit2, fit.measures=TRUE, standardized=TRUE, rsquare=TRUE)
writeLines("\n Comparing Model 1 and Model 2\n")
lavTestLRT(fit1, fit2)
sink()
```

filename: alien.out (output file)

Example 1: Stability of Alienation

Output for Model 1

\$lambda

	SES	ALIEN6	ALIEN7
educ	0	0	0
sei	1	0	0
anomia67	0	0	0
power167	0	2	0
anomia71	0	0	0
power171	0	0	3

\$theta

	educ	sei	anom67	pwrl67	anom71	pwrl71
educ	7					
sei	0	8				
anomia67	0	0	9			
power167	0	0	0	10		
anomia71	0	0	0	0	11	
power171	0	0	0	0	0	12

\$psi

SES ALIEN6 ALIEN7
SES 13
ALIEN67 0 14
ALIEN71 0 0 15

\$beta

	SES	ALIEN6	ALIEN7
SES	0	0	0
ALIEN67	4	0	0
ALIEN71	6	5	0

lavaan 0.6-3 ended normally after 58 iterations

Optimization method Number of free parameters	:	NLMINB 15
Number of observations		932
Estimator Model Fit Test Statistic Degrees of freedom P-value (Chi-square) Model test baseline model:		ML 88.293 6 0.000
Minimum Function Test Statistic Degrees of freedom P-value	21	59.906 15 0.000
User model versus baseline model:		
Comparative Fit Index (CFI) Tucker-Lewis Index (TLI)		0.962 0.904
Loglikelihood and Information Criteria:		
Loglikelihood user model (H0) Loglikelihood unrestricted model (H1)		98.966 54.819
Number of free parameters Akaike (AIC) Bayesian (BIC) Sample-size adjusted Bayesian (BIC)	263	15 27.931 00.491 52.852
Root Mean Square Error of Approximation:		
RMSEA 90 Percent Confidence Interval P-value RMSEA <= 0.05	0.100	0.121 0.144 0.000

Standardized Root Mean Square Residual:

SRMR	0.025
------	-------

3.246

3.546

Parameter Estimates:

.power167

.anomia71

Parameter Estim	ate	3:					
Information Information saturated (h1) model Standard Errors				St	Expected ructured Standard		
Latent Variable	s:						
		Estimate	Std.Err	z-value	P(> z)	$\mathtt{Std.lv}$	Std.all
SES =~							
educ		1.000				2.492	0.804
sei		0.572	0.044	13.010	0.000	1.425	0.672
ALIEN67 =~							
anomia67		1.000				2.809	
power167		0.881	0.041	21.307	0.000	2.473	0.808
ALIEN71 =~							
anomia71		1.000				2.998	
power171		0.834	0.039	21.385	0.000	2.500	0.791
Regressions:							
		Estimate	Std.Err	z-value	P(> z)	std.lv	Std.all
ALIEN67 ~							
SES	(a)	-0.646	0.057	-11.250	0.000	-0.574	-0.574
ALIEN71 ~							
ALIEN67	(b)	0.682	0.053	12.771	0.000	0.639	0.639
SES	(c)	-0.229	0.058	-3.976	0.000	-0.190	-0.190
Variances:							
		Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
.educ		3.398	0.455	7.462	0.000	3.398	0.354
.sei		2.472	0.181	13.687	0.000	2.472	0.549
.anomia67		3.946	0.346	11.420	0.000	3.946	0.333

0.273

0.372

11.906

9.545

3.246

3.546

0.000

0.000

0.347

0.283

.power171	3.735	0.289	12.910	0.000	3.735	0.374
SES	6.212	0.597	10.413	0.000	1.000	1.000
.ALIEN67	5.292	0.474	11.176	0.000	0.671	0.671
.ALIEN71	3.738	0.389	9.606	0.000	0.416	0.416

R-Square:

	Estimate
educ	0.646
sei	0.451
anomia67	0.667
power167	0.653
anomia71	0.717
power171	0.626
ALIEN67	0.329
ALIEN71	0.584

Defined Parameters:

	Estimate	Std.Err	z-value	P(> z)	$\mathtt{Std.lv}$	Std.all
direct	-0.229	0.058	-3.976	0.000	-0.190	-0.190
indirect	-0.441	0.048	-9.215	0.000	-0.367	-0.367
total	-0.670	0.060	-11.195	0.000	-0.557	-0.557

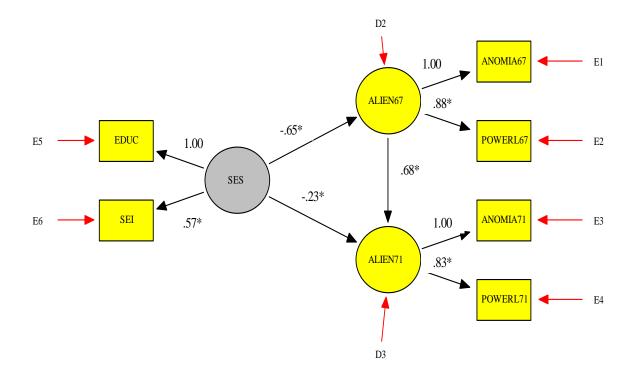
Modification indices

[[1]]

	lhs	op	rhs	mi	epc	sepc.lv	sepc.all	sepc.nox
44	anomia67	~~	anomia71	57.693	1.863	1.863	0.498	0.498
46	power167	~~	anomia71	54.108	-1.597	-1.597	-0.471	-0.471
45	anomia67	~~	power171	45.229	-1.429	-1.429	-0.372	-0.372
47	power167	~~	power171	41.895	1.219	1.219	0.350	0.350
41	sei	~~	anomia71	18.443	-0.647	-0.647	-0.219	-0.219
37	educ	~~	anomia71	16.531	0.917	0.917	0.264	0.264
36	educ	~~	power167	12.065	-0.685	-0.685	-0.206	-0.206
42	sei	~~	power171	6.804	0.353	0.353	0.116	0.116
38	educ	~~	power171	6.744	-0.516	-0.516	-0.145	-0.145
27	ALTEN67	=~	sei	5.182	0.222	0.624	0.294	0.294

26	ALIEN67	=~	educ	5.182	-0.389	-1.092	-0.352	-0.352
30	ALIEN71	=~	educ	5.182	0.296	0.888	0.287	0.287
31	ALIEN71	=~	sei	5.182	-0.169	-0.508	-0.239	-0.239
40	sei	~~	power167	4.840	0.293	0.293	0.103	0.103
32	ALIEN71	=~	anomia67	3.497	0.310	0.931	0.271	0.271
33	ALIEN71	=~	power167	3.497	-0.273	-0.820	-0.268	-0.268
22	SES	=~	anomia67	3.497	0.137	0.341	0.099	0.099
23	SES	=~	power167	3.497	-0.120	-0.300	-0.098	-0.098
35	educ	~~	anomia67	3.276	0.402	0.402	0.110	0.110
28	ALIEN67	=~	anomia71	0.215	0.104	0.293	0.083	0.083
29	ALIEN67	=~	power171	0.215	-0.087	-0.245	-0.077	-0.077
25	SES	=~	power171	0.215	-0.029	-0.073	-0.023	-0.023
24	SES	=~	anomia71	0.215	0.035	0.087	0.025	0.025
39	sei	~~	anomia67	0.064	-0.038	-0.038	-0.012	-0.012

• Model 1



$$T_{ML} = 88.29, df = 6, p < .001$$

 $NNFI = .904, CFI = .962, RMSEA = .121, SRMR = .025$

- From SES to ALIEN71,
- direct effect is c = -0.23, p < .01
- indirect effect is ab = (-0.65)(0.68) = -0.44, p < .01 (this is known as the Sobel test)
- total effect is c + ab = -0.23 0.44 = -0.67, p < .01
- Full mediation occurs when the direct effect is no longer significant given the mediator
- Partial mediation occurs when the magnitude of the direct effect is reduced (but still significant) after controlling for the mediator
- Problems of Sobel test: the asymptotic standard error of the indirect effect may be incorrect for small sample sizes (Bollen & Stine, 1990), because the distribution of the product of the two coefficients (a and b) is skewed

• Use bootstrap standard error to improve the accuracy of statistical inference for a mediation effect

```
# fit Model 1 to data
fit1 <- lavaan(model1, sample.cov=alien.cov, sample.cov.rescale=FALSE,
sample.nobs=932, auto.var=TRUE, auto.fix.first=TRUE, se="bootstrap",
bootstrap=1000)</pre>
```

- Bootstrapping is a resampling technique and it requires (1) raw data and (2) a sufficiently large number of bootstrap replications to work (e.g., brep=1000)
- Model fit can be much improved if error covariances are allowed
- This is easy to justify as the same scale was administered twice at 2 time points ⇒ Model 2
- How do we know which effect, direct or indirect, is stronger?

Output for Model 2

\$lambda

	SES	ALIEN6	ALIEN7
educ	0	0	0
sei	1	0	0
anomia67	0	0	0
power167	0	2	0
anomia71	0	0	0
power171	0	0	3

\$theta

\$psi

SES ALIEN6 ALIEN7

SES 15 ALIEN67 0 16

ALIEN71 0 0 17

\$beta

 SES
 ALIEN6
 ALIEN7

 SES
 0
 0
 0

 ALIEN67
 4
 0
 0

 ALIEN71
 6
 5
 0

lavaan 0.6-3 ended normally after 70 iterations

Optimization method NLMINB
Number of free parameters 17

Number of observations	932
Estimator Model Fit Test Statistic Degrees of freedom P-value (Chi-square)	ML 23.742 4 0.000
Model test baseline model:	
Minimum Function Test Statistic Degrees of freedom P-value	2159.906 15 0.000
User model versus baseline model:	
Comparative Fit Index (CFI) Tucker-Lewis Index (TLI)	0.991 0.965
Loglikelihood and Information Criteria:	
Loglikelihood user model (H0) Loglikelihood unrestricted model (H1)	-13066.690 -13054.819
Number of free parameters Akaike (AIC) Bayesian (BIC) Sample-size adjusted Bayesian (BIC)	17 26167.380 26249.615 26195.624
Root Mean Square Error of Approximation:	
RMSEA 90 Percent Confidence Interval P-value RMSEA <= 0.05	0.073 0.046 0.102 0.076
Standardized Root Mean Square Residual:	
SRMR	0.016

Parameter Estimates:

Information Information Standard Err	rated (h1)	model		Expected ructured Standard				
Latent Variabl	Latent Variables:							
		Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all	
SES =~								
educ		1.000				2.522	0.814	
sei		0.559	0.043	12.948	0.000	1.408	0.664	
ALIEN67 =~								
anomia67		1.000				2.704	0.786	
power167		0.951	0.059	15.981	0.000	2.571	0.840	
ALIEN71 =~								
anomia71		1.000				2.925	0.827	
power171		0.875	0.055	15.926	0.000	2.559	0.810	
Regressions:		Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all	
ALIEN67 ~		<u> </u>	Dografia	2 value	- (> 2)	564.11	bearan	
SES	(a)	-0.614	0.058	-10.630	0.000	-0.572	-0.572	
ALIEN71 ~	()							
ALIEN67	(b)	0.586	0.052	11.330	0.000	0.542	0.542	
SES	(c)	-0.286	0.057	-5.015	0.000	-0.246	-0.246	
Covariances:								
		Estimate	Std.Err	z-value	P(> z)	$\mathtt{Std.lv}$	Std.all	
.anomia67 ~~								
.anomia71		1.412	0.312	4.524	0.000	1.412	0.333	
.power167 ~~								
.power171		0.535	0.251	2.132	0.033	0.535	0.174	
Variances:		_	_	_				
_		Estimate	Std.Err		P(> z)			
.educ		3.250	0.463	7.013	0.000	3.250	0.338	
.sei		2.519	0.180	14.014	0.000	2.519	0.559	

.anomia67	4.533	0.460	9.856	0.000	4.533	0.383
.power167	2.750	0.394	6.983	0.000	2.750	0.294
.anomia71	3.961	0.516	7.677	0.000	3.961	0.316
.power171	3.442	0.408	8.444	0.000	3.442	0.344
SES	6.360	0.606	10.489	0.000	1.000	1.000
.ALIEN67	4.916	0.479	10.259	0.000	0.672	0.672
.ALIEN71	4.221	0.419	10.086	0.000	0.493	0.493

R-Square:

	Estimate
educ	0.662
sei	0.441
anomia67	0.617
power167	0.706
anomia71	0.684
power171	0.656
ALIEN67	0.328
ALIEN71	0.507

Defined Parameters:

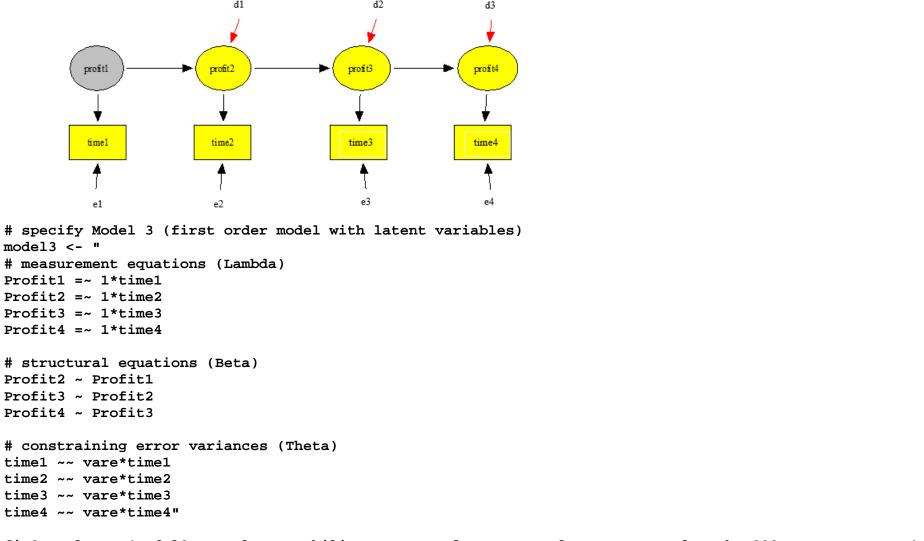
	Estimate	Std.Err	z-value	P(> z)	$\mathtt{Std.lv}$	Std.all
direct	-0.286	0.057	-5.015	0.000	-0.246	-0.246
indirect	-0.360	0.042	-8.587	0.000	-0.310	-0.310
total	-0.645	0.060	-10.753	0.000	-0.556	-0.556
effect_d	-0.074	0.080	-0.927	0.354	-0.064	-0.064

Comparing Model 1 and Model 2

Chi Square Difference Test

```
Df AIC BIC Chisq Chisq diff Df diff Pr(>Chisq)
fit2 4 26167 26250 23.742
fit1 6 26228 26301 88.293 64.551 2 9.615e-15 ***
---
Signif. codes: 0 \***' 0.001 \**' 0.01 \*' 0.05 \.' 0.1 \' 1
```

5.8. Profit Growth: Model 3 (see Example 3 in Chapter 4)



fit3 <- lavaan(model3, sample.cov=ability.cov, sample.cov.rescale=FALSE, sample.nobs=200, auto.var=TRUE)</pre>

```
Output for Model 3
Note: model contains equality constraints:
 lhs op rhs
1 4 ==
          5
2 4 ==
3 4 == 7
$lambda
     Proft1 Proft2 Proft3 Proft4
time1
          0
time2
time3
          0
                 0
                       0
                              0
time4
          0
                0
                       0
$theta
     time1 time2 time3 time4
time1 4
time2 0
time3 0
time4 0
                0
           0
                       7
$psi
       Proft1 Proft2 Proft3 Proft4
Profit1 8
Profit2 0
Profit3 0
                    10
Profit4 0
                  0
                           11
$beta
       Proft1 Proft2 Proft3 Proft4
Profit1
            0
                   0
                         0
                                0
```

Profit2

Profit3

Profit4

1

0

0

0

3

0

0

0

lavaan 0.6-5 ended normally after 54 iterations

Estimator	ML
Optimization method	NLMINB
Number of free parameters	11
Number of equality constraints	3
Row rank of the constraints matrix	3
Number of observations	200
Model Test User Model:	
Test statistic	4.397
Degrees of freedom	2
P-value (Chi-square)	0.111
Model Test Baseline Model:	
Test statistic	822.319
Degrees of freedom	6
P-value	0.000
User Model versus Baseline Model:	
Comparative Fit Index (CFI)	0.997
Tucker-Lewis Index (TLI)	0.991
Loglikelihood and Information Criteria:	
Loglikelihood user model (H0)	-2368.824
Loglikelihood unrestricted model (H1)	-2366.626
Akaike (AIC)	4753.648
Bayesian (BIC)	4780.034
Sample-size adjusted Bayesian (BIC)	4754.689

Root Mean Square Error of Approximation:

RMSEA	0.077
90 Percent confidence interval - lower	0.000
90 Percent confidence interval - upper	0.178
P-value RMSEA <= 0.05	0.234

Standardized Root Mean Square Residual:

Parameter Estimates:

Information	Expected
Information saturated (h1) model	Structured
Standard errors	Standard

Latent Variables:

		Estimate	Sta.Err	z-varue	P(> 2)	Sca. IV	sta.aii
Profit1	=~						
time1		1.000				5.767	0.912
Profit2	=~						
time2		1.000				6.782	0.934
Profit3	=~						
time3		1.000				7.317	0.942
Profit4	=~						
time4		1.000				10.013	0.968

Regressions:

	Estimate	Std.Err	z-value	P(> z)	$\mathtt{Std.lv}$	Std.all
Profit2 ~						
Profit1	1.120	0.065	17.247	0.000	0.952	0.952
Profit3 ~						
Profit2	1.050	0.046	22.710	0.000	0.973	0.973
Profit4 ~						
Profit3	1.303	0.054	24.205	0.000	0.952	0.952

Variances:

		Estimate	Std.Err	z-value	P(> z)	$\mathtt{Std.lv}$	Std.all
.time1	(vare)	6.737	0.881	7.644	0.000	6.737	0.168
.time2	(vare)	6.737	0.881	7.644	0.000	6.737	0.128
.time3	(vare)	6.737	0.881	7.644	0.000	6.737	0.112
.time4	(vare)	6.737	0.881	7.644	0.000	6.737	0.063
Profit1		33.263	4.096	8.121	0.000	1.000	1.000
.Profit2		4.294	2.048	2.097	0.036	0.093	0.093
.Profit3		2.863	1.433	1.998	0.046	0.053	0.053
.Profit4		9.377	2.625	3.572	0.000	0.094	0.094

R-Square:

	Estimate
time1	0.832
time2	0.872
time3	0.888
time4	0.937
Profit2	0.907
Profit3	0.947
Profit4	0.906