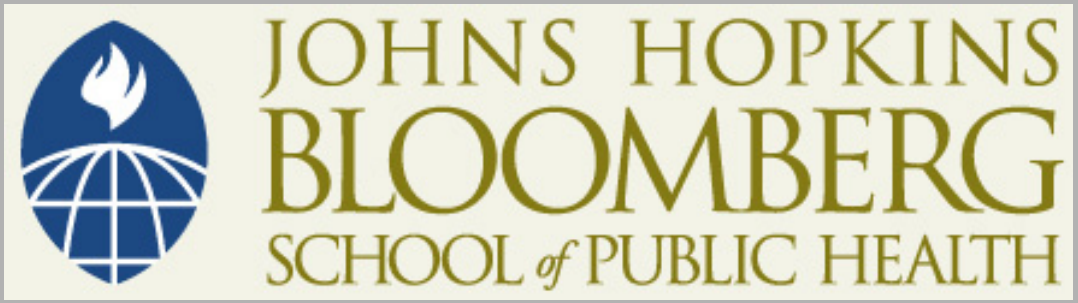
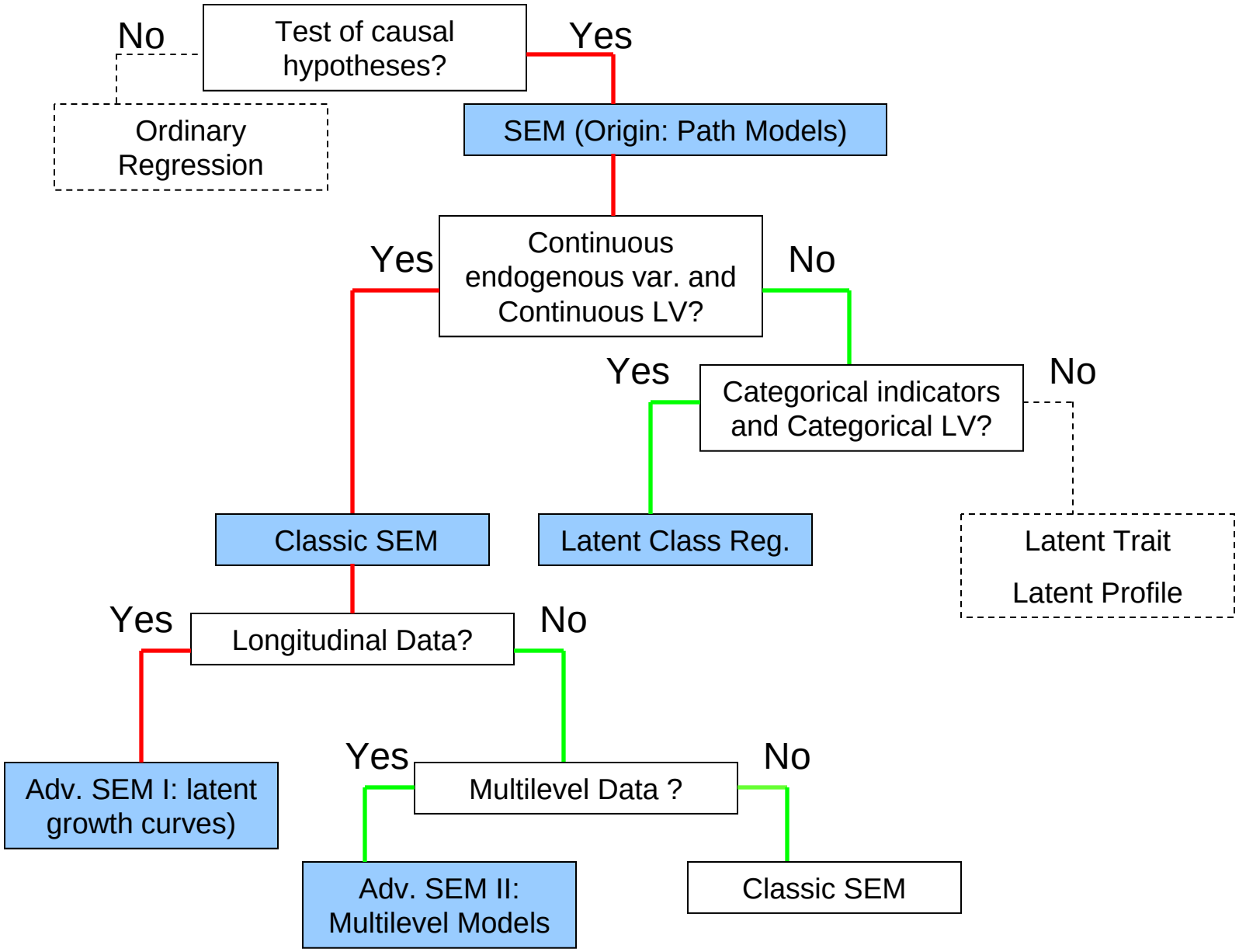


Advanced Structural Equations Models I

Statistics for Psychosocial Research II:
Structural Models

Qian-Li Xue


No
Test of causal hypotheses?
Yes
Ordinary Regression
SEM (Origin: Path Models)
Yes
Continuous endogenous var. and Continuous LV?
No
Yes
Categorical indicators and Categorical LV?
No
Classic SEM
Latent Class Reg.
Latent Trait
Latent Profile
Yes
Longitudinal Data?
No
Adv. SEM I: latent growth curves)
Yes
Multilevel Data ?
No
Adv. SEM II: Multilevel Models
Classic SEM

Outline

1. Estimating means of observed and latent variables
2. Modeling repeated measures of outcome over time
   - The Simplex-Growth Over Time
3. Non-Recursive Models
4. Modeling repeated measures of outcome and covariate over time
   - Cross-Lag Panel Analysis
   - Latent Growth Curve Models (Next Lecture)

1. Estimating Means of Observed and Latent Variables

Estimating Means of Observed and Latent Variables

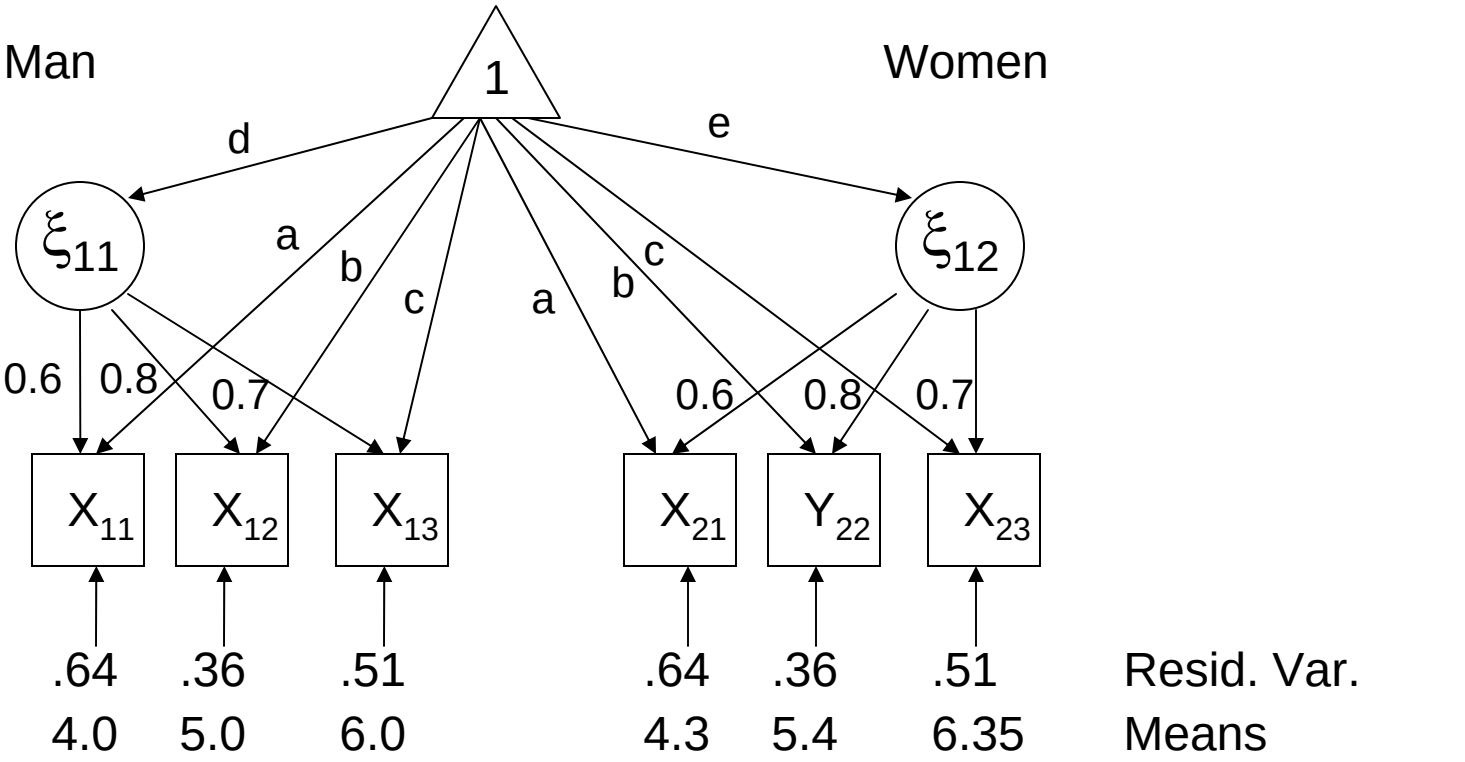
- So far, we have largely ignored intercept terms in our analyses
- What has happened to the alpha coefficient?

Estimating Means of Observed and Latent Variables

- Up to now, information on means and intercepts has not been of interest
  - It is possible to estimate levels of association without information on these parameters
- If of interest, these parameters can be estimated using a “mean model.”
  - In addition to covariances, these models also require information on mean of variables
- These parameters are of key interest in group comparisons and growth curve models

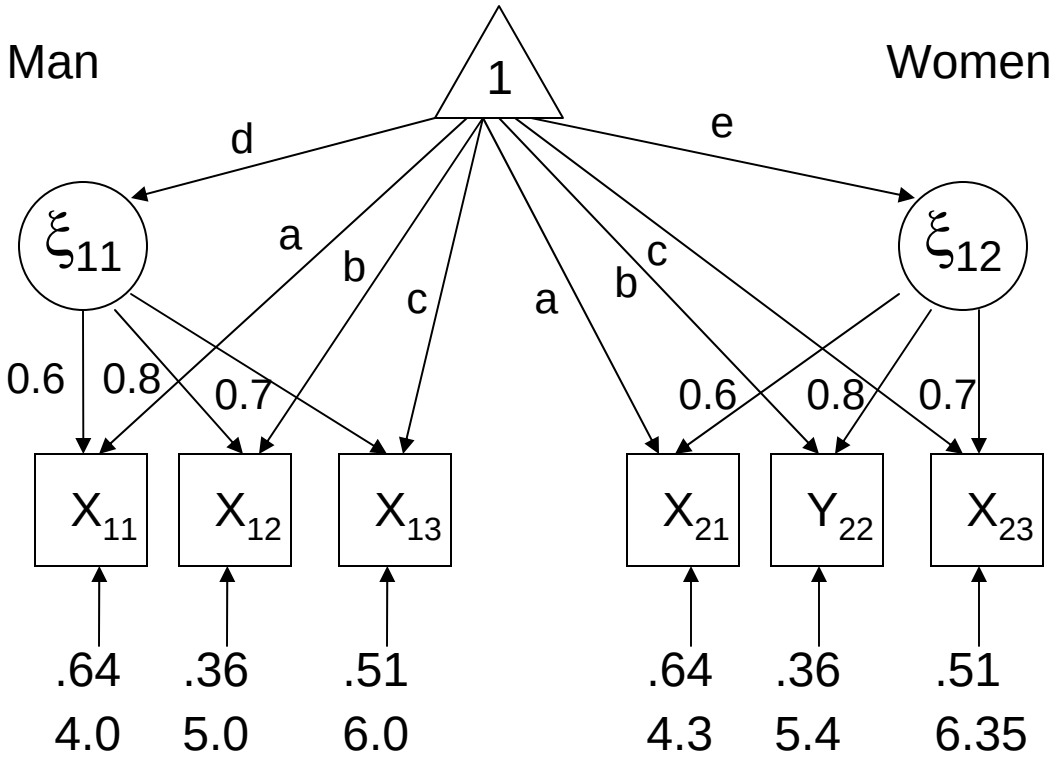
Estimating Means of Observed and Latent Variables

- Does the mean score on the latent variable ξ (e.g. depression) differ between men and women?



(Loehlin p.139)

Estimating Means of Observed and Latent Variables

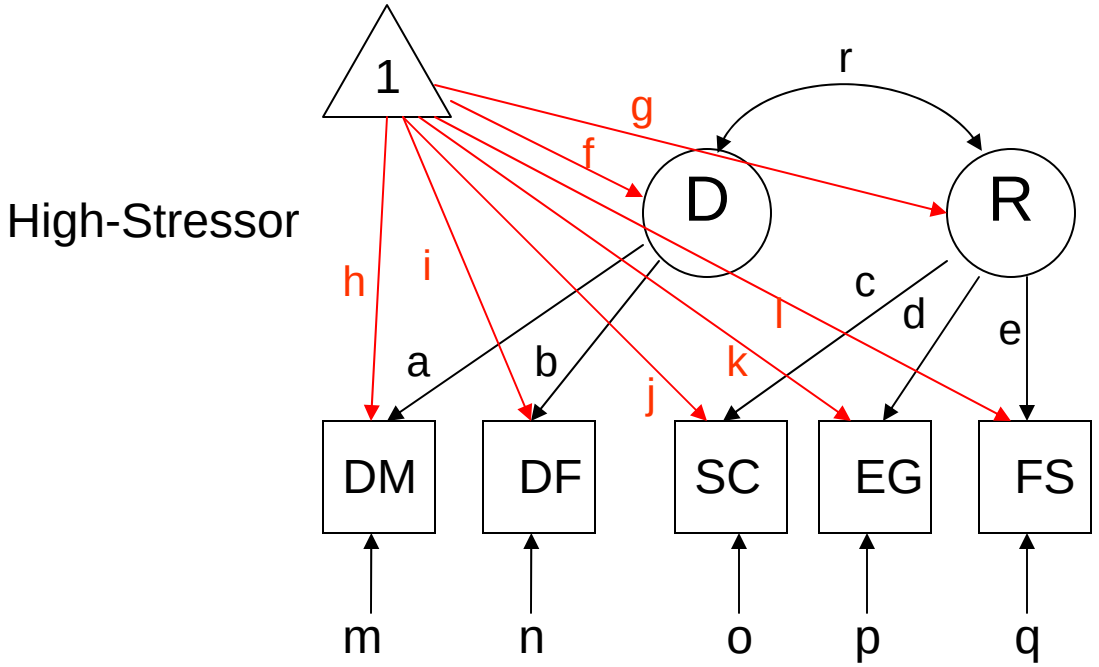


- d=0 (reference)
- a=4.0,b=5.0,c=6.0 (baseline values, same across groups)
- e – difference between the means of the latent variable
- e*0.6+a=4.3 ⇒ e=0.5

(Loehlin p.139)

Example: Stress, Resources, and Depression (Holahan & Moos, 1991)

- "How do the high-stressor and the low-stressor groups compare on the two latent variables: depression (D) and resources (R)



Example: Stress, Resources, and Depression (Holahan & Moos, 1991)

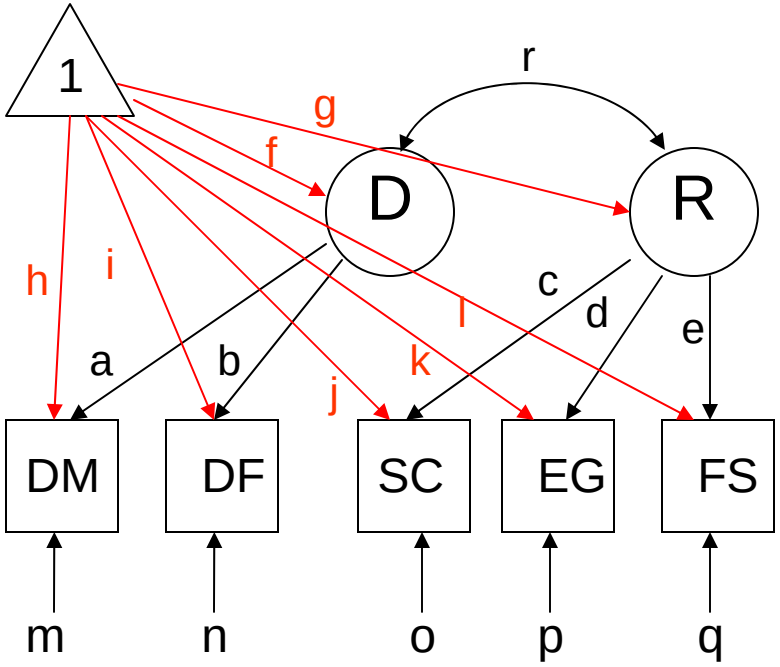
High-stressor group: above diagonal (underlined)

Low-stressor group: below diagonal

	DM	DF	SC	EG	FS	SD	M
Depressed Mood	1	.84	-.36	-.45	-.51	5.97	8.82
Depressive Features	.71	1	-.32	-.41	-.50	7.98	13.87
Self-confidence	-.35	-.16	1	.26	.47	3.97	15.24
Easygoingness	-.35	-.21	.11	1	.34	2.27	7.92
Family support	-.38	-.26	.30	.28	1	4.91	19.03
Standard Deviation	4.84	6.33	3.84	2.14	4.43	N	128
Mean	6.15	9.96	15.14	8.80	20.43	126	

Example: Stress, Resources, and Depression (Holahan & Moos, 1991)

Low-Stressor



- MPLUS code

```
TITLE:  Stress, resources, and depression (Loehlin, p.142)
DATA:   FILE is c:/teaching/140.658.2007/depression.dat;
        TYPE IS CORRELATION MEANS STDEVIATIONS;
        NOBSERVATIONS ARE 126 128;
        NGROUPS=2;
VARIABLE: NAMES ARE DM DF SC EG FS;
        USEVARIABLES ARE DM-FS;
MODEL:
        D BY DM* DF;
        R BY SC* EG FS;
        DM (1);
        DF (2);
        SC (3);
        EG (4);
        FS (5);
MODEL g1:
        [D@0 R@0];
        D@1 R@1;
OUTPUT:
        TECH1;
```

Equate the measurement models across the groups

Set reference group (i.e. low-stressor)

Example: Stress, Resources, and Depression (Holahan & Moos, 1991)

	Measurement Model								
	Latent Variables			Path Coeff.		Residual Var.		Baseline means	
Low-Stressor	Depression: Mean	f	[0]*	a	4.42	m	2.91	h	6.09
	Resources: Mean	g	[0]	b	5.22	n	16.04	i	10.27
	Depression: SD		[1]	c	1.56	o	11.76	j	15.59
	Resources: SD		[1]	d	1.01	p	3.61	k	8.61
	correlation	r	-0.72	e	2.67	q	12.25	l	20.40
High-Stressor	Depression: Mean	f	0.63	Same as above					
	Resources: Mean	g	-0.50						
	Depression: SD		1.30						
	Resources: SD		1.29						
	correlation	r	-0.78						

* Numbers in [ ] are prefixed in order to make the model identified

Example: Stress, Resources, and Depression (Holahan & Moos, 1991)

TESTS OF MODEL FIT

Chi-Square Test of Model Fit

Value	27.245	
Degrees of Freedom	19	
P-Value	0.0991	

CFI/TLI

CFI	0.979	
TLI	0.978	

RMSEA (Root Mean Square Error Of Approximation)

Estimate	0.058	
90 Percent C.I.	0.000	0.104

SRMR (Standardized Root Mean Square Residual)

Value	0.055

The model fits reasonably well to the data!

2. Modeling Repeated Measures of Outcome Over Time

The Simplex-Growth Over Time

- Modeling growth over (e.g. height)
- Measurements taken repeatedly over time
- In general, measurements made closer together in time would be more highly correlated (called “simplex” by Guttman, 1954)
- E.g.

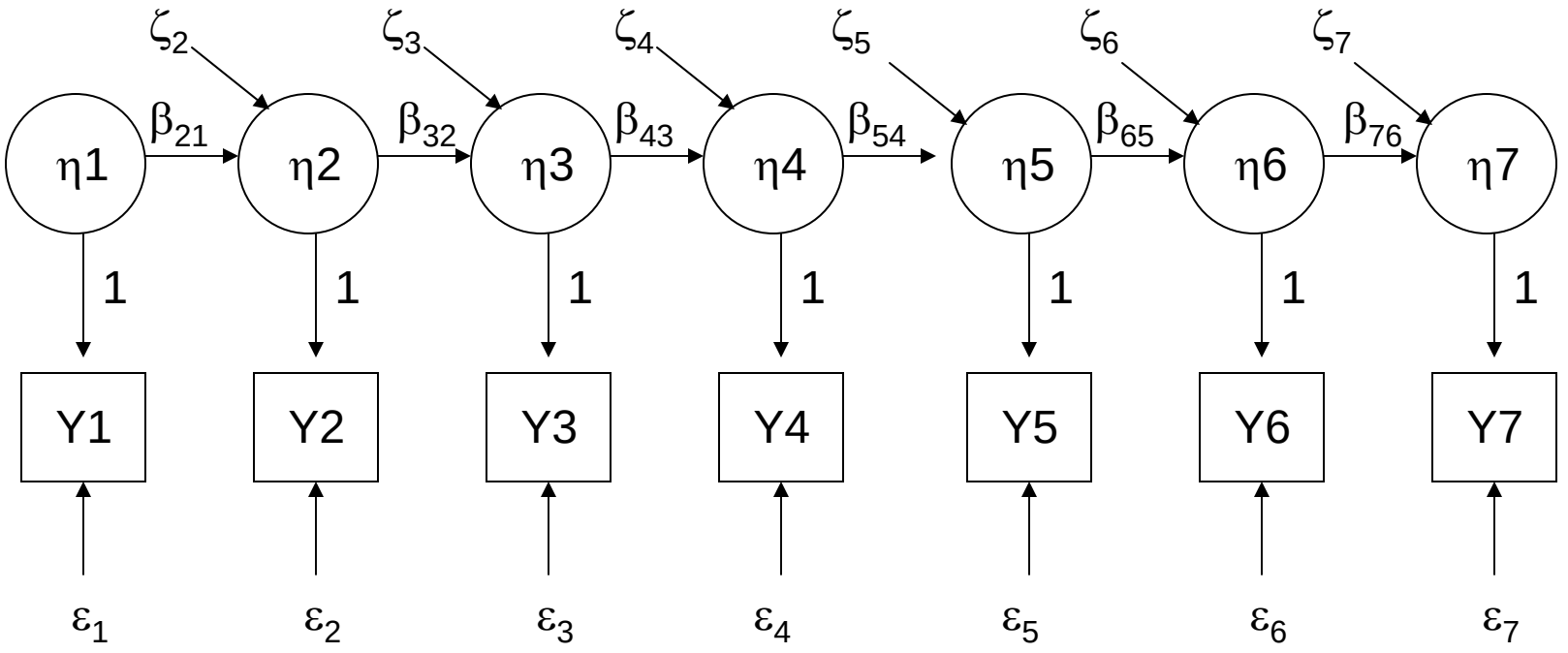
Smaller →

Correlation	1	2	3	4
1	1	0.73	0.72	0.68
2		1	0.79	0.76
3			1	0.84
4				1

↑ Smaller

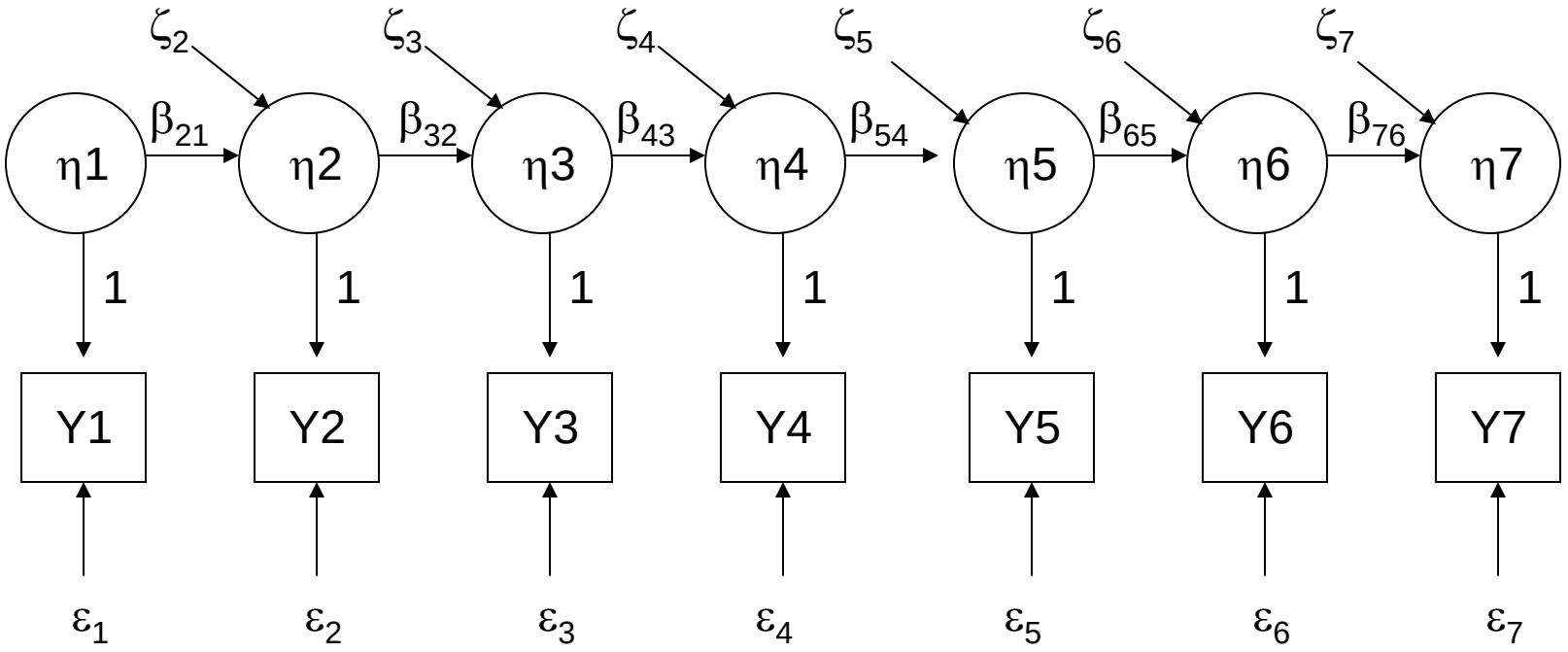
The Simplex-Growth Over Time

- Example: Scores on standardized tests of academic achievement at grades 1-7 (Bracht & Hopkins, 1972)
- Test score (Y) is a measure of the latent academic achievement (η)
- Achievement at grade t is a function of achievement at t-1 via β, and other factors ζ



Loehlin p.125

The Simplex-Growth Over Time



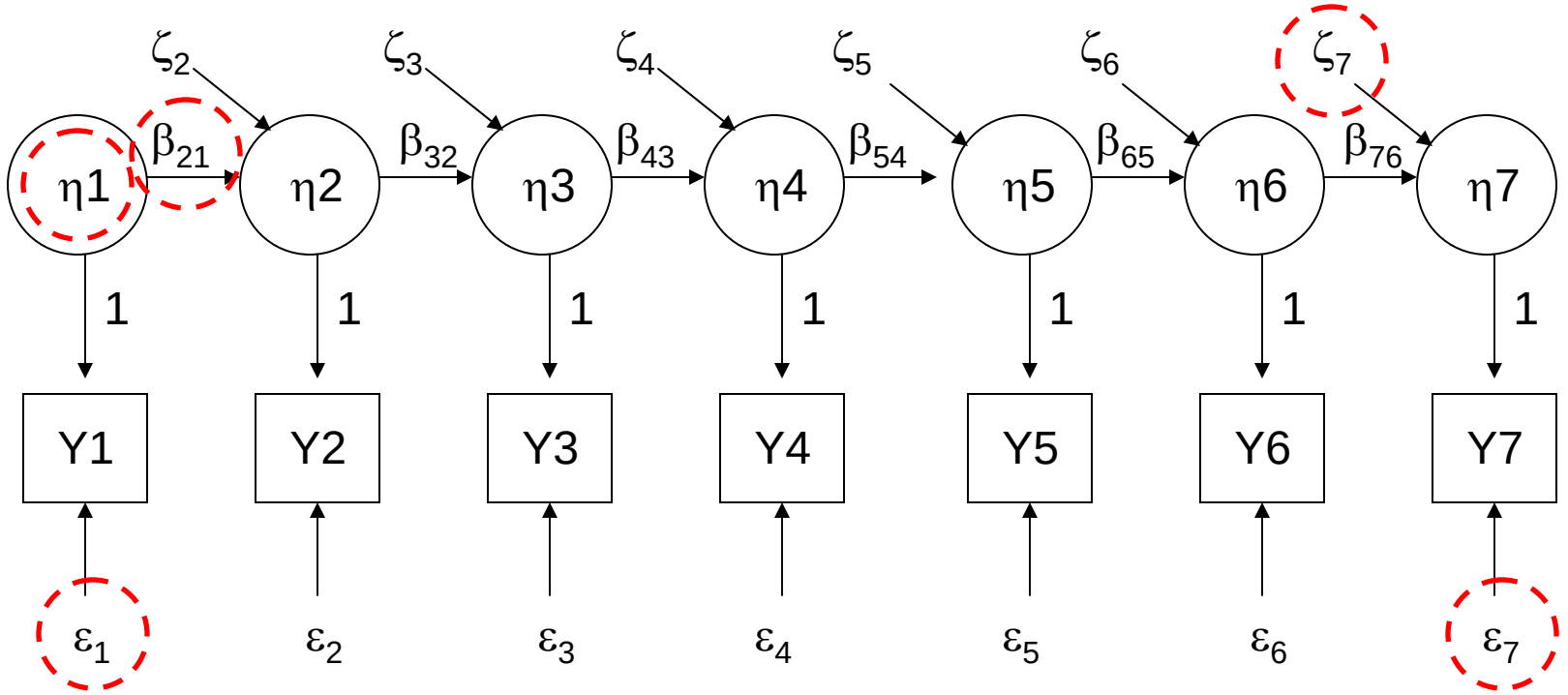
$$Y_i = \eta_i + \varepsilon_i$$

$$\eta_i = \beta_i \eta_{i-1} + \varsigma_i$$

ε_i are uncorrelated, $\varepsilon_i \perp \eta_i$, and $\zeta_i \perp \eta_{i-1}$

The Simplex-Growth Over Time

- $Var(\eta_1)$, $Var(\zeta_7)$, $Var(\varepsilon_1)$, $Var(\varepsilon_2)$, β_{21} are unidentified
- To achieve identification, set $Var(\varepsilon_1)=Var(\varepsilon_2)$ AND $Var(\varepsilon_6)=Var(\varepsilon_7)$, reasonable if Ys are on the same scale
- # free parameters = 3p-3, where p=# of Ys
- For testing a simplex model, p>3 !!!

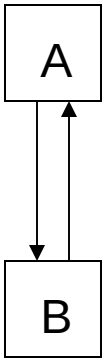


3. Non-Recursive Models

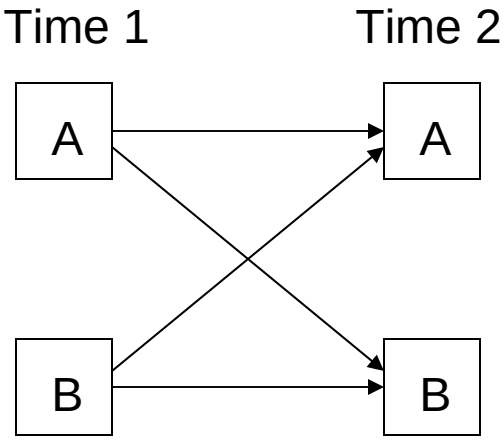
Non-Recursive Models

- So far, there has been little discussion of models with feedback loops
- Non-recursive models deal with reciprocal causal relationships
- Can not be analyzed by ordinary regression analysis due to correlated errors
- Non-recursive models may not be identified even if the T-rule is met

Non-Recursive Models



Reciprocal

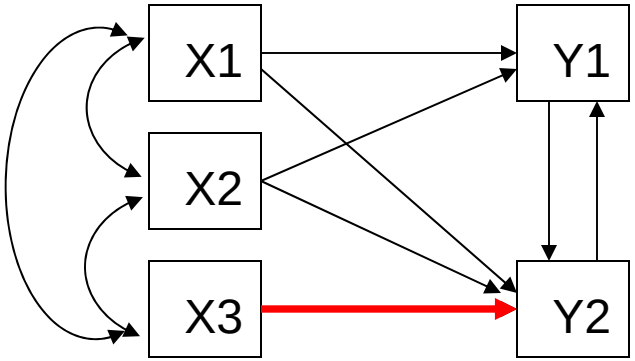


Lagged

- What do you mean by “reciprocal causation”?
- Alternative: Lagged model
  - Assumption: the principal of “finite causal lag”
  - Roles of the variables in the bidirectional relationship change over time (e.g. A is a cause at Time 1, but effect at Time 2)
- The reciprocal causation model becomes the only choice if only cross-sectional data are available

Non-Recursive Models: Model Identification

- Recall: recursive path models without measurement error are always identified
- Not true for non-recursive models
- Definition: Instrumental variable – a predictor is an instrument for an endogenous variable if it has a direct path to other endogenous variables but not the endogenous variable of interest



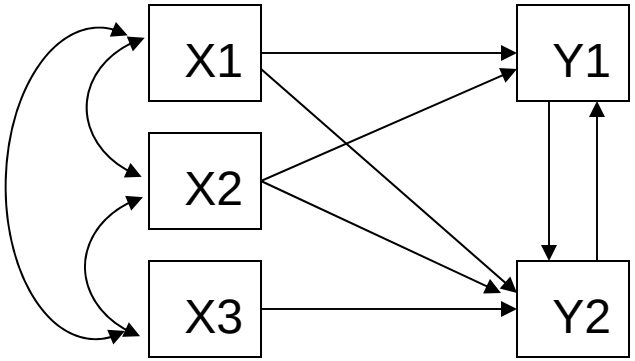
X3 is an instrument for Y1

Maruyama, 1998; p.106

Non-Recursive Models: Model Identification

- Order condition (necessary but not sufficient) – For any system of N endogenous variables, a particular equation is identified only if at least N-1 variables are left out of that equation

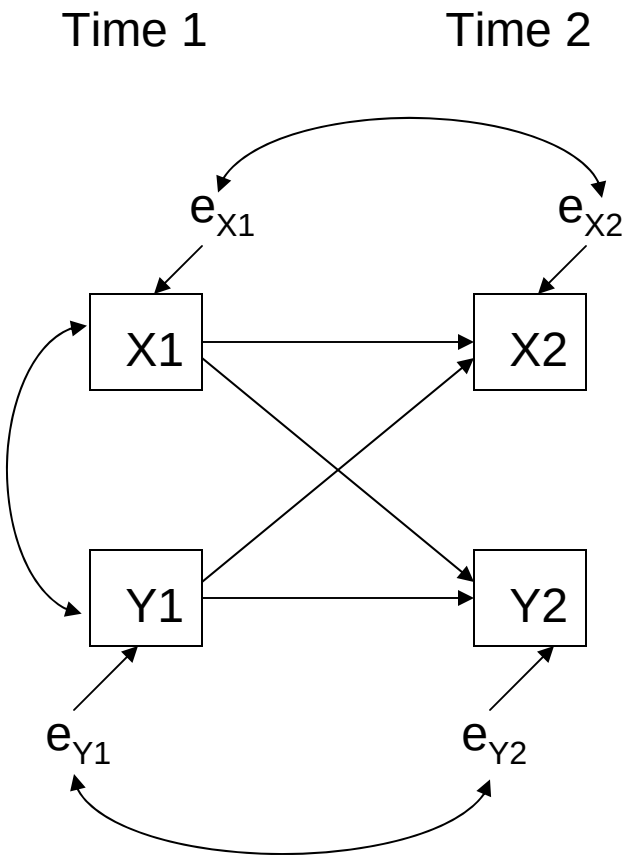
- Rank condition (necessary AND sufficient) – is met for a particular equation if there is at least one non-zero determinant of rank N-1 from the coefficients of the variables omitted from that equation



Maruyama, 1998; p.106

4. Modeling Repeated Measures of Outcome and Covariate Over Time

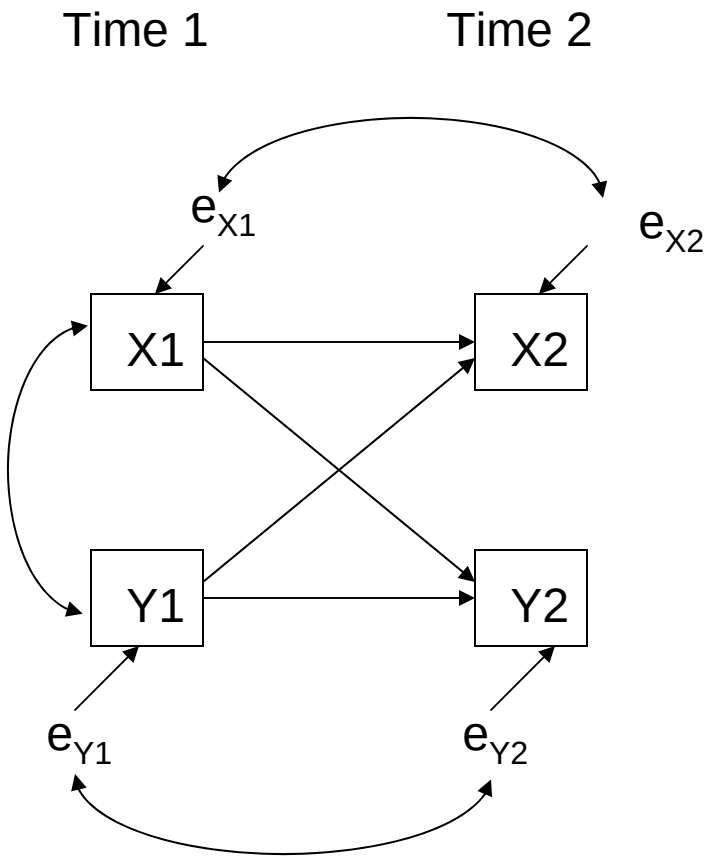
Cross-Lagged Panel Analysis: Terminology



- Synchronous correlations: Corr(X1,Y1) and Corr(X2,Y2)
- Autocorrelations (i.e. stability): Corr(X1,X2) and Corr(Y1,Y2)
- Cross-lagged: Corr(X1,Y2) and Corr(Y1,X2)
- Residual correlations (due to measure-specific variance): Corr(e_{x1},e_{X2}) and Corr(e_{Y1},e_{Y2})

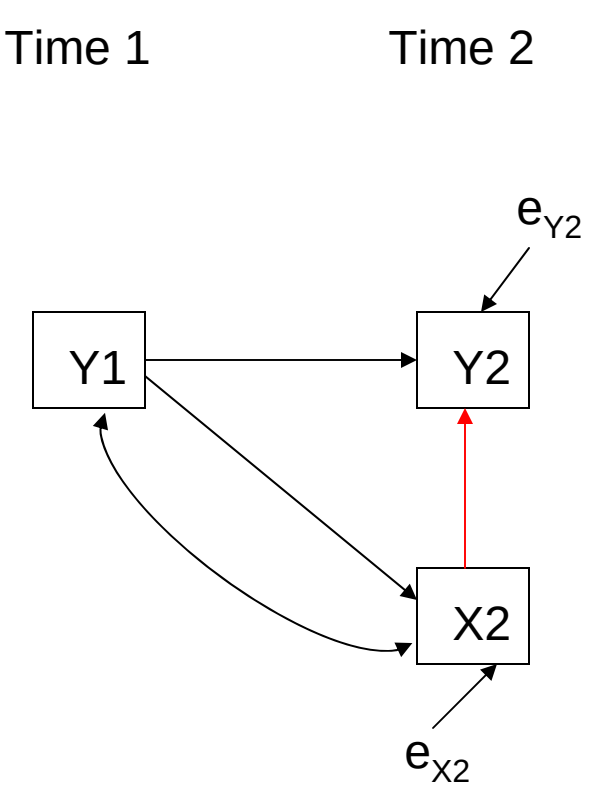
Here Corr. denotes total correlation!

Cross-Lagged Panel Analysis: Identification



- Is this model identified?
- # equations = 4*5/2=10
- # unknowns = 11
- Not identified!
- What is the problem?
  - The repeated assessment of the same measure leads to two sources of common variance
    - construct variance
    - Measure-specific variance
- Model would be identified if delete residual correlations or
- Build multiple-indicator models

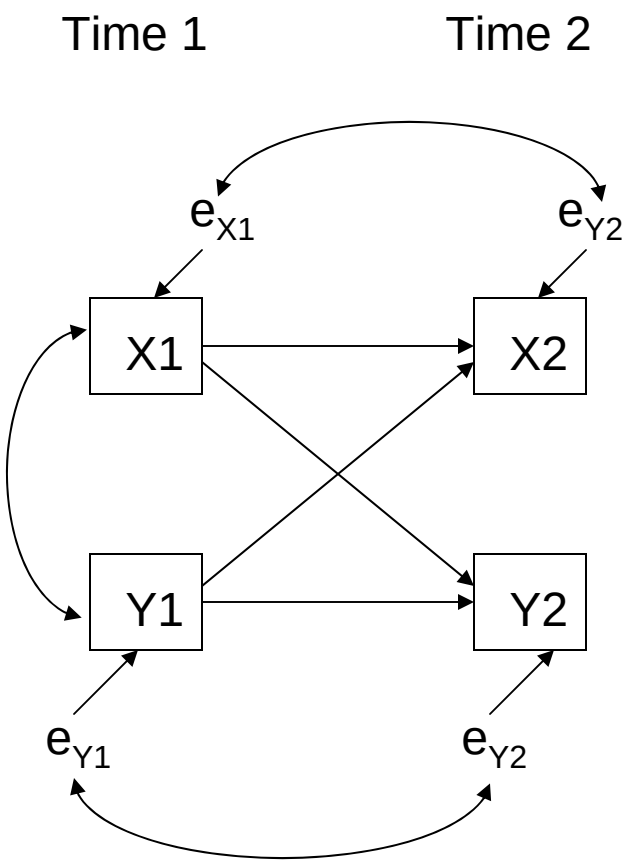
Cross-Lagged Panel Analysis: Key Issues (Maruyama, pp.112-120)



1. Stability of a variable

- For example, if Y is perfectly stable, Y2 is perfectly determined by Y1
- If data is only available at Time 2, then Y1 is not available
- Any variable correlated with Y or caused by Y could be included as predictors, leading to a misspecified model!
- Low stability over time may result from poor reliability (if so, we're in trouble!) or
- Real change in the measure

Cross-Lagged Panel Analysis: Key Issues



2. Temporal Lags

- **How long is the causal lag?**
- **It the sampling interval > causal lag ⇒ attenuated effect**
- **If the sampling interval < causal lag ⇒ no effect or underestimated effect**
- **What if the causal lag from X1 to Y2 is different from Y1 to X2?**
- **Solution: three-wave data with different intervals**

Cross-Lagged Panel Analysis: Key Issues

3. Growth Across Time

- **When to use covariance vs. correlation data in SEM**
  - **Covariance allows for "growth" by focusing on raw scores**
  - **Correlation focuses on standardized relationships**
  - **If no change in variability of any of the variables over time, the results are identical**
  - **Using covariance is highly recommended!**

Cross-Lagged Panel Analysis: Key Issues

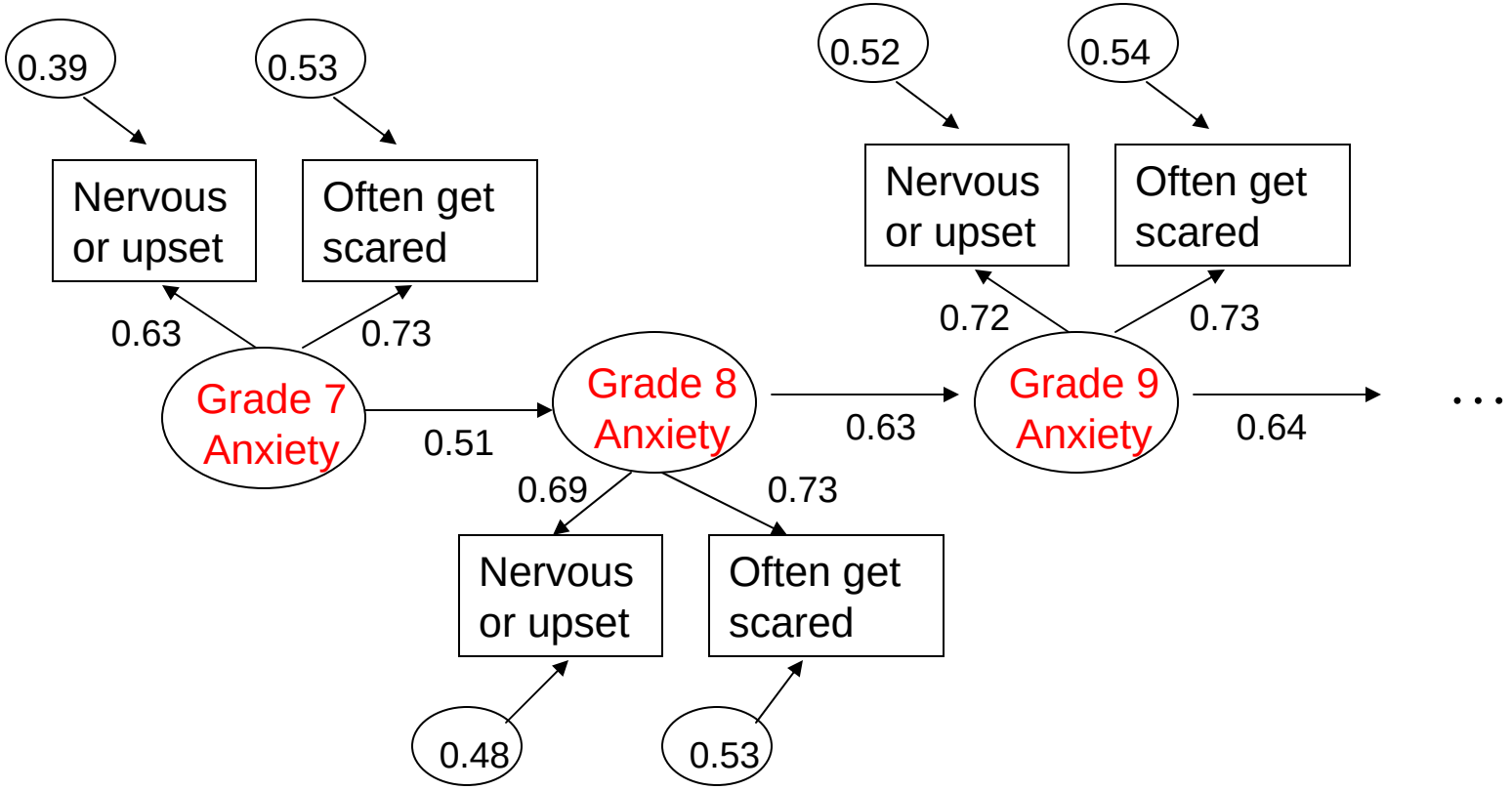
3. Stability of Causal Process

- **Causal dynamics between variables remain stable across time intervals of the same length**

- **If not true, the relationships would differ depending on the particular interval sampled**

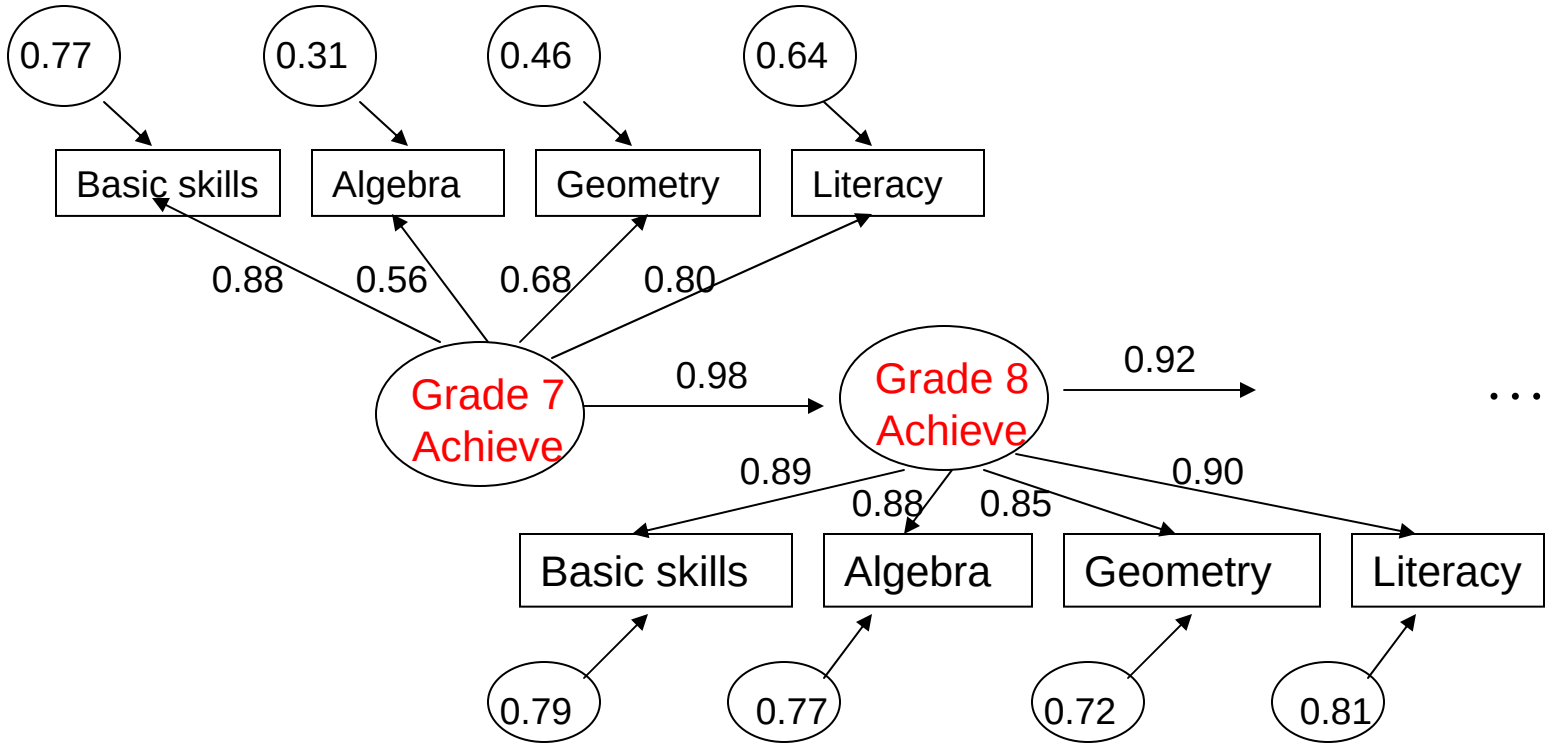
- **On the other hand, modeling unstable processes may be warranted when studying**
  - **Developmental processes**
  - **Time-varying interventions**

Cross-Lagged Panel Analysis with Latent Variables: Example

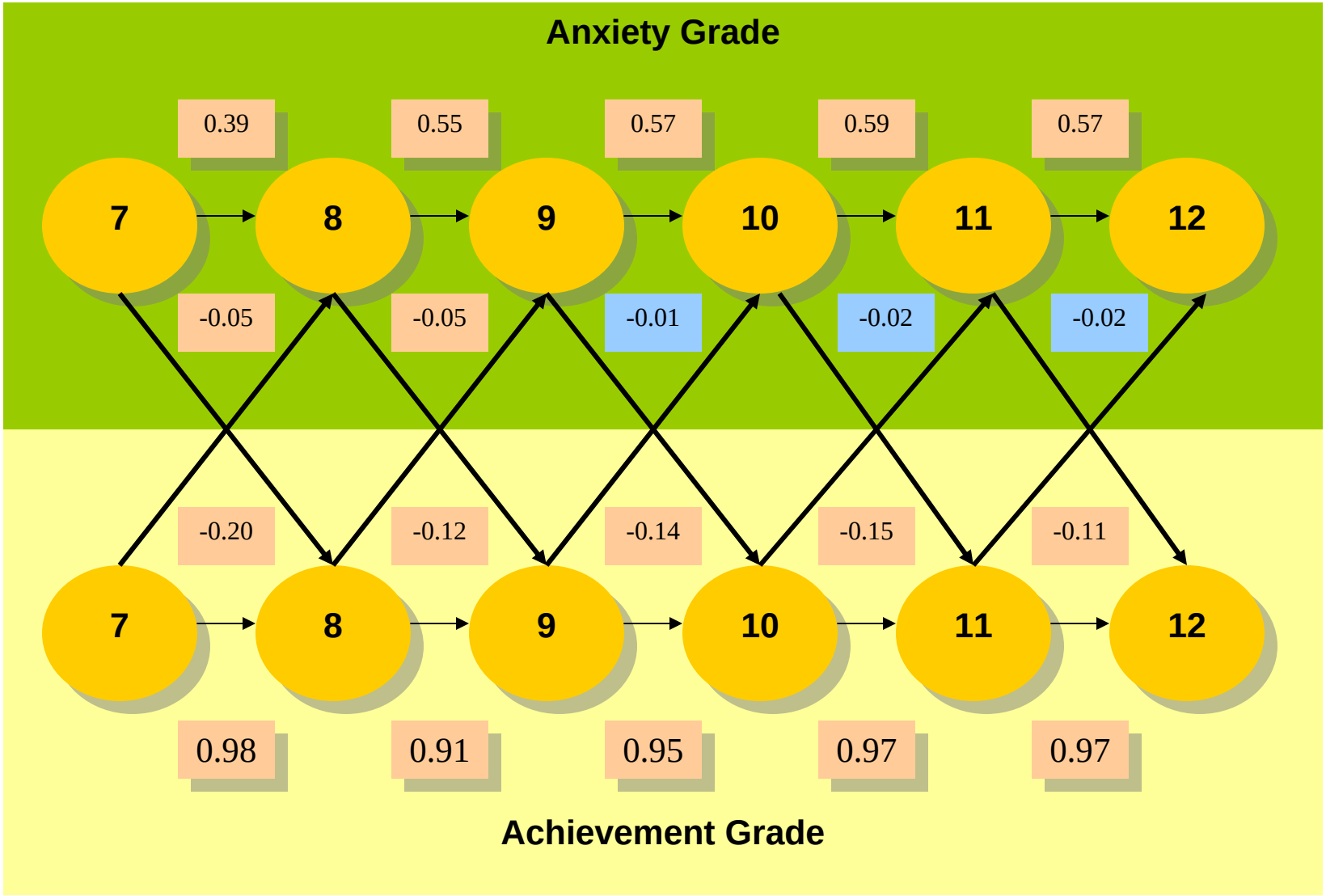


(Ma & Xu, Journal of Adolescence 27 (2): 165-179 APR 2004)

Cross-Lagged Panel Analysis with Latent Variables: Example



(Ma & Xu, Journal of Adolescence 27 (2): 165-179 APR 2004)



Example of cross-lagged panel analysis with latent variables. Structural equation model estimating the causal relationship between mathematics anxiety & mathematics achievement across Grades 7–12. Large ovals represent latent factors & unidirectional arrows represent casual links. All parameter estimates for unidirectional paths are standardized. Pink boxes indicated $P < 0.001$). Adapted from Ma & Xu, *Journal of Adolescence* 2004;27:165-179