Structural Equation Panel Models II: Cross-Lagged and Synchronous Effects Models

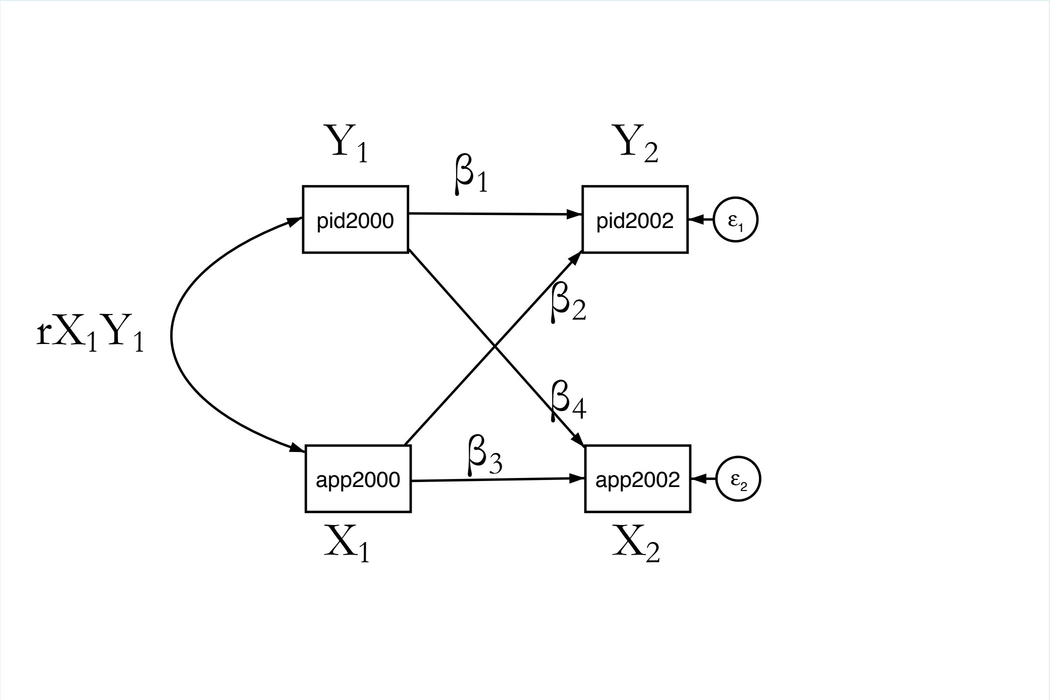
PS 2701-2019

Longitudinal Analysis

Week 6

Professor Steven Finkel

I: Cross-Lagged Panel Models



SEM Diagram in STATA with Party ID and Presidential Approval, 2000-2004

- A *Dynamic* Model of Change in Each Endogenous Variable

$$\text{(a)}\quad Y_2 = \beta_1 Y_1 + \beta_2 X_1 + \varepsilon_1$$

$$\text{(b)}\quad Y_2 - Y_1 = (\beta_1 - 1) Y_1 + \beta_2 X_1 + \varepsilon_1$$

- Goal: Compare relative direction and magnitude of cross-lagged effects to give evidence regarding causal direction and strength
- Each equation is a longitudinal regression model with a lagged DV ("lagged endogenous variable") and a lagged IV
- The estimated effect of X_1 is the same in (a) or (b). The latter is sometimes called the "conditional change" model – the effect of X_1 on changes in some variable, *conditional* on the initial level of that variable. With no lagged DV (here no Y_1), we would have an "unconditional change" model
- Finkel (1995) also called these the "static-score " (a) and "change-score" (b) models
- Differences based on whether and how the researcher can justify including lagged Y as a predictor of the *level* of Y_2 or (equivalently) the *change in Y*

Causal Lags and "Synchronous" Effects

- Cross-lagged model assumes no synchronous or contemporaneous effect at time *t*.
  - Does time period separating waves of measurement correspond to the length of time for causal effect to take place?
  - If no, may need to add an additional effect from X_2 to Y_2 , or have a model that *only* has the contemporaneous effect and no lagged effect. Need to think about how the causal effect operates
- With a unidirectional single equation model, can add contemporaneous effect with no problem. Model is still recursive.

$$Y_2 = \beta_5 X_2 + \beta_1 Y_1 + \beta_2 X_1 + \varepsilon_1$$

- And since the following is true,

$$X_2 = X_1 + \Delta X$$

can express this as:

$$(a) \qquad Y_2 = \beta_1 Y_1 + (\beta_5 + \beta_2) X_1 + \beta_5 \Delta X + \varepsilon_1$$

- Or, in terms of X and ΔY :

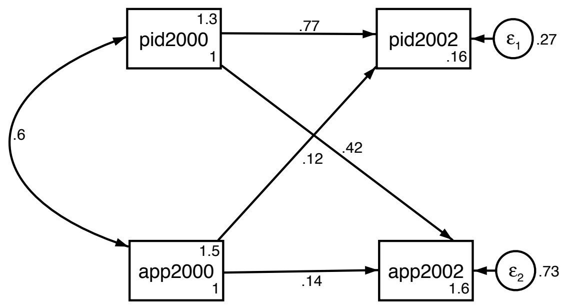
$$(b) \qquad \Delta Y = (\beta_1 - 1) Y_1 + (\beta_5 + \beta_2) X_1 + \beta_5 \Delta X + \varepsilon_1$$

- In Time Series language:
  - Model a: "Autoregressive Distributed Lag Model" (ADL, 1, 1)
  - Model b: "Error Correction Model" (ECM)
- BUT in a multiple causal system, including both synchronous and lagged effects in both directions makes the model "non-recursive," and "underidentified". Need more information, as we will see !!!

Estimating the Two-Wave Cross-Lagged Model

- It is a **recursive** model with no reciprocal effects specified at the same point in time. Therefore, the model is either identified or overidentified.
  - Which is it?
    - Number of knowns: 4 variables=(4*5)/2= 10 covariances
    - Number of unknowns: 9 (4 β, 2 variances of ε structural disturbances, plus necessary to estimate the three variances and covariances of the observed wave 1 "exogenous" variables)
    - Therefore, ***overidentified*** with 1 degree of freedom
    - What is left out of this model? What is the "overidentifying restriction?" Answer: the covariation between wave 2 disturbances – call this $\psi_{21}=0$
    - So the model chi-square tests whether this restriction is correct, i.e., whether the cross-lagged and stability effects in the model *completely* account for the cross-sectional covariation between Y_2 and X_2.
    - Can saturate the model by relaxing this restriction and estimating ψ_{21}
    - The other estimated coefficients in this revised model will be exactly the same.

STATA Results (Standardized)




```
. estat gof, sta(all)
```

Fit statistic	Value	Description
Likelihood ratio		
chi2_ms(**1**)	**90.399**	model vs. saturated
p > chi2	**0.000**	
chi2_bs(**5**)	**1798.754**	baseline vs. saturated
p > chi2	**0.000**	
Population error		
RMSEA	**0.291**	Root mean squared error of approximation
90% CI, lower bound	**0.242**	
upper bound	**0.343**	
pclose	**0.000**	Probability RMSEA <= 0.05
Information criteria		
AIC	**15146.499**	Akaike's information criterion
BIC	**15211.045**	Bayesian information criterion
Baseline comparison		
CFI	**0.950**	Comparative fit index
TLI	**0.751**	Tucker-Lewis index
Size of residuals		
SRMR	**0.034**	Standardized root mean squared residual
CD	**0.753**	Coefficient of determination

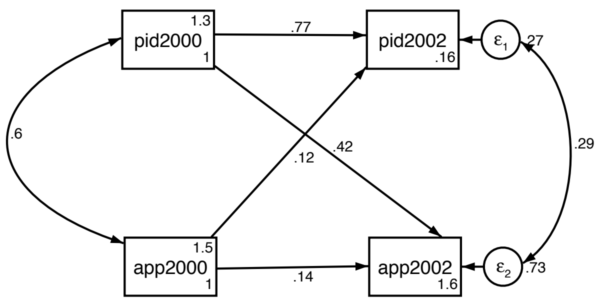
```
. estat mindices
```

Modification indices

	MI	df	P>MI	EPC	Standard EPC
Structural					
pid2002 <-					
app2002	**86.648**	**1**	**0.00**	**.2372708**	**.1744527**
app2002 <-					
pid2002	**86.648**	**1**	**0.00**	**.3448425**	**.4690155**
cov(e.pid2002,e.app2002)	**86.648**	**1**	**0.00**	**.435695**	**.2860438**

EPC = expected parameter change

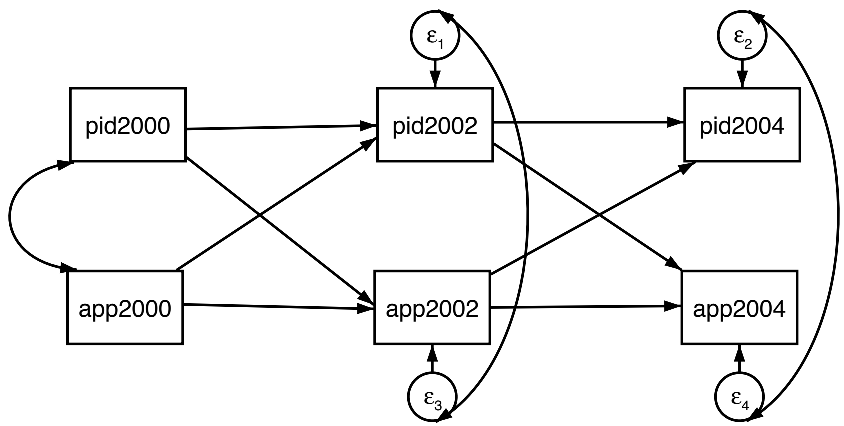
Saturated Version of Two Wave Model



Standardized	Coef.	OIM Std. Err.	z	P>|z|	[95% Conf.	Interval]
Structural						
pid2002 <-						
pid2000	.7729582	.0158919	48.64	0.000	.7418106	.8041057
app2000	.1241617	.0200544	6.19	0.000	.0848557	.1634676
_cons	.1624004	.0322593	5.03	0.000	.0991733	.2256274
app2002 <-						
pid2000	.4239438	.0308988	13.72	0.000	.3633831	.4845044
app2000	.1370873	.0326491	4.20	0.000	.0730961	.2010784
_cons	1.558611	.072866	21.39	0.000	1.415797	1.701426
mean(pid2000)	1.339711	.0423285	31.65	0.000	1.256749	1.422674
mean(app2000)	1.521585	.0451376	33.71	0.000	1.433117	1.610053
var(e.pid2002)	.2722378	.0142733			.245652	.3017009
var(e.app2002)	.7319102	.0232906			.6876559	.7790124
var(pid2000)	1	.			.	.
var(app2000)	1	.			.	.
cov(e.pid2002, e.app2002)	.2860438	.028215	10.14	0.000	.2307435	.3413441
cov(pid2000, app2000)	.5985186	.0197213	30.35	0.000	.5598656	.6371716

LR test of model vs. saturated: chi2(**0**) = **0.00**, Prob > chi2 = .

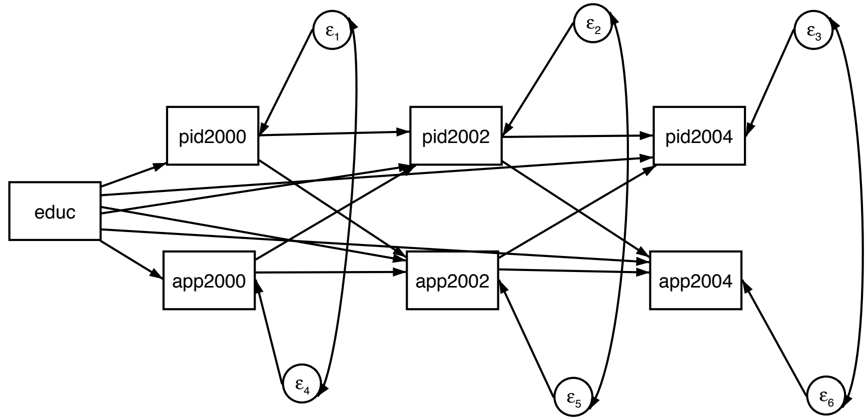
Extension to Three and More Waves



- Flexibility in model and lag specification increases substantially with 3 or more waves. Many more additional "knowns" in the covariance matrix, yields more degrees of freedom to test restrictions in the model and compare alternative nested models

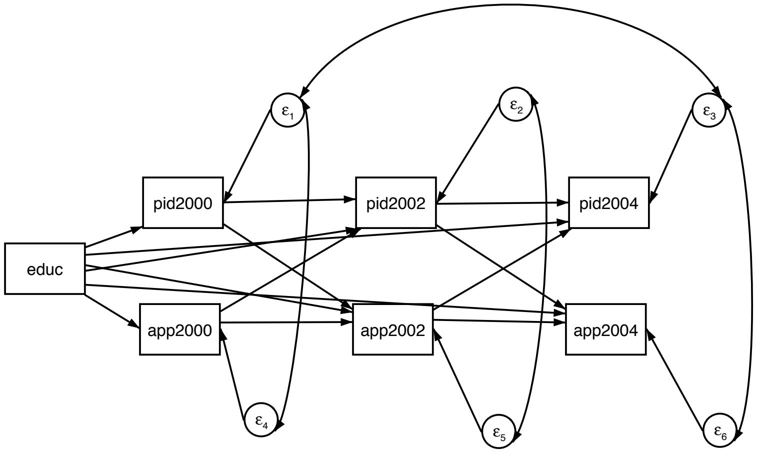
Testing Alternative Multi-wave Models

- Imposing Equality of Causal Effects is Straightforward
  - Equal stabilities across waves
  - Equal cross-lagged effects across waves
  - Equal error covariances in waves 2 and 3
    - Note: error covariance in wave 1 cannot be equated with these two. (Why?)
- Test these equality constraints via difference in chi-square statistics, as each models *with* the respective constraint is nested within the model *without* the respective constraint
- Add exogenous background variables
- Relax constraints regarding error covariances for same variable over time (i.e. autocorrelated disturbances)



No Wave 1-Wave 3 PID Error Term Covariance

Wave 1-Wave 3 PID Error Term Covariance Added



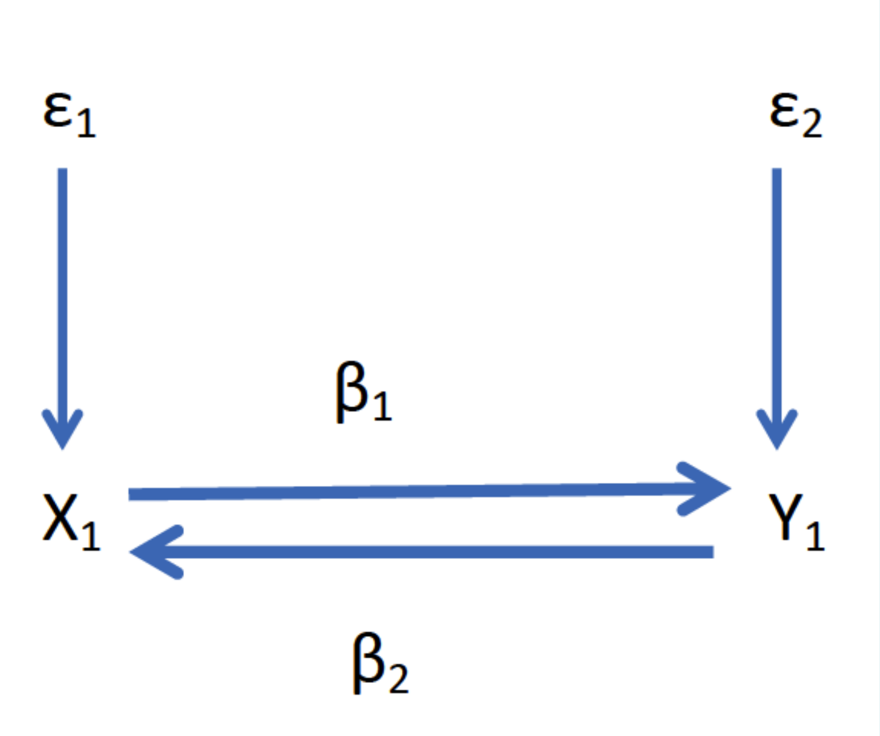
Problems/Omissions in the Cross-Lagged Models So Far

- No synchronous causal effects
- No measurement error
- No controls for stable unmeasured variables (or unit-specific intercepts) that were focus of Unit 1
- This last omission means that the causal effects estimated so far are – in our previous "hybrid" or in multilevel modeling language – a mixture of "within" and "between" unit effects. We should (and will) do better!

II. Non-Recursive SEM Models

- Involve reciprocal causality between variables at a *given* point in time (as opposed to cross-lagged effects)
  - Note difference between a model with a synchronous causal effect from one variable to another (which would not necessarily be non-recursive), and a model with *reciprocal* synchronous effects, which would necessarily be non-recursive
- When appropriate?
  - When theory says the causal lag is short, relative to time between waves
  - When cross-lagged effects are weak or insignificant, with relatively large covariances between the two variables at a given point in time

Reciprocal Effects or Simultaneous Causality



$$Y_1 = \beta_1 X_1 + \varepsilon_2 \qquad (5)$$
$$X_1 = \beta_2 Y_1 + \varepsilon_1$$

$$X_1 = \beta_2(\beta_1 X_1 + \varepsilon_2) + \varepsilon_1$$
$$X_1 = \beta_2 \beta_1 X_1 + \beta_2 \varepsilon_2 + \varepsilon_1$$

So X_1 is related to ε_2, hence endogenous! Same with Y_1 and ε_1 in equation predicting X

Further problem: Underidentification!! 4 unknowns, 3 knowns!!

Identification

- *Model* Identification: Start with the "counting rule".
  - Number of knowns: $t(t+1)/2$, where t is the number of observed vars
  - Number of unknowns: k parameters to be estimated
  - **Model as a whole** is identified if $t(t+1)/2-k\geq 0$
- But still possible for individual equations *not* to be identified
- A 99.9% effective rule for identifying equations within non-recursive models is the "order condition":
- If equation involves p endogenous variables, then there must be at least $(p-1)$ excluded exogenous (Z) variables for the equation to be identified. Another way to put it: for each endogenous variable that is included as a predictor, there must be at least one excluded exogenous variable for the equation to be identified.
- That is, must have a Z that does *not* have an effect on the endogenous variable in question for every endogenous variable that does. This represents an "exclusion" restriction that allow identification of the given equation (i.e. the X is "excluded" from the X or Y equation in question).
- These Z variables will be the "instrumental variables" for the causal system

- In our reciprocal effects case: to identify each equation, we therefore need a variable that is a) *exogenous*, and hence unrelated to all equations' error terms; b) directly affects one of the endogenous variables in the reciprocal effects causal system (and the stronger, the better); but c) is *excluded,* i.e., does not affect the other endogenous variable in question.

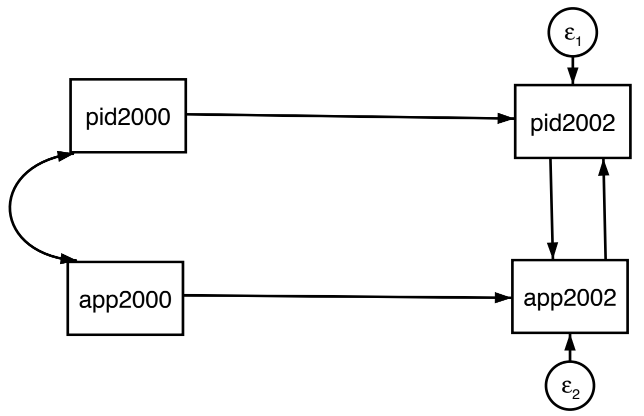
  THESE VARIABLES ARE DIFFICULT TO FIND!!!!

  IMAGINE TWO VARIABLES THAT ARE RECIPROCALLY RELATED. IS IT EASY TO ALSO IMAGINE AN EXOGENOUS VARIABLE THAT AFFECTS ONE AND ONLY ONE OF THOSE TWO VARIABLES??

- As we have discussed, this is one of the main advantages of panel analysis: the ***lag values*** of the variables can, under certain conditions, serve as instruments for reciprocally-related variables.
  - If one can assume **no cross-lagged effects**, and if one can assume **exogeneity** for the lagged values, then panels provide ready-made instruments for causal systems that contain contemporaneous reciprocal effects.
  - But if there are cross-lagged effects, or if there are correlations between the disturbances, these assumptions will not hold. In some cases, moreover, it will not be possible to test all of these assumptions due to the lack of information in the causal system.

General Moral: we always estimate causal effects within the context of models whose assumptions need to be justified, and which may ultimately not hold

The Two-Wave Synchronous Effects Model



$$Y_2 = \beta_3 X_2 + \beta_1 Y_1 + \varepsilon_1$$

$$X_2 = \beta_4 Y_1 + \beta_2 X_1 + \varepsilon_2$$

ML Estimation of the Model

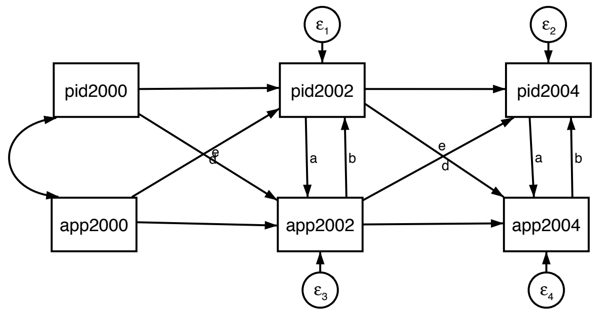
- Each of the equations is just-identified: one included endogenous variable and one excluded exogenous variable.
  - X_1 is an instrument for X_2 in the Y_2 equation
  - Y_1 is an instrument for Y_2 in the X_2 equation.
  - We can arrive at the coefficients through algebraic manipulation – multiply each endogenous equation by each of the wave 1 variables (X_1,Y_1) and solve for the 4 structural effects in terms of the observed covariances. This is tedious algebra but follows the same logic as what we have been doing so far
- ML estimation also follows the same logic as before.
  - 10 observed variances-covariances, 9 unknowns (4 structural effects, 2 disturbances, 3 variances-covariances of wave 1 variables)
  - Model as a whole has ***one*** over-identifying restriction (either one of the cross-lagged effects is zero, or the disturbance covariance between the ε is zero).
  - The model χ2 provides a test of the restriction.

Extensions

- Three wave synchronous effects model provides additional information from third wave, and ability to constrain effects to be equal across waves to gain additional degrees of freedom.
- Two wave **combined** synchronous and cross-lagged model is not identified without including more instruments, but since only two waves exist in the panel, such instruments cannot come from lagged values of the variables. Hence the researcher is in the same difficult position as the cross-sectional analyst in looking for appropriate instrumental variables in a two-wave model with both lagged and synchronous effects.

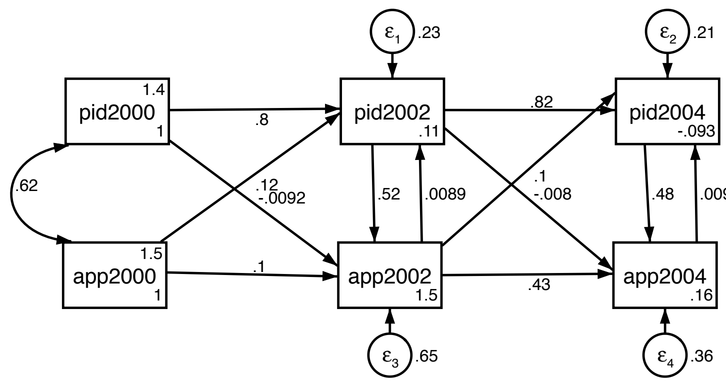
- Three wave combined synchronous and cross-lagged model may be identified in several ways:
  - Use the wave 1 variables to identify the wave 3 equations. This is rarely implemented because one cannot say anything about the wave 2 estimates
  - Impose equality constraints on the coefficients across waves, so that the cross-lagged effects are equal from waves 1-2 and 2-3, and the synchronous effects are also equal between wave 2 and wave 3.

    In the absence of strong theoretical reasons for expecting different lag lengths, the analyst will often estimate cross-lagged only, synchronous only, and various combined models, often with equality constraints as well. Decisions about the likely lag length and the direction and magnitude of the causal effects are made after inspecting the overall pattern of results from these models. In this case, for example, it looks pretty clear that PID has a synchronous effect on APP, and APP has a lagged effect on PID.



Stata estimation of Synch and C-L effect model, with effects equated across waves

Likelihood ratio		
chi2_ms(6)	134.899	model vs. saturated
p > chi2	0.000	
chi2_bs(14)	3458.563	baseline vs. saturated
p > chi2	0.000	
Population error		
RMSEA	0.171	Root mean squared error of approximation
90% CI, lower bound	0.146	
upper bound	0.196	
pclose	0.000	Probability RMSEA <= 0.05
Information criteria		
AIC	14813.960	Akaike's information criterion
BIC	14910.643	Bayesian information criterion
Baseline comparison		
CFI	0.963	Comparative fit index
TLI	0.913	Tucker-Lewis index
Size of residuals		
SRMR	0.029	Standardized root mean squared residual
CD	0.777	Coefficient of determination



. estat mindices

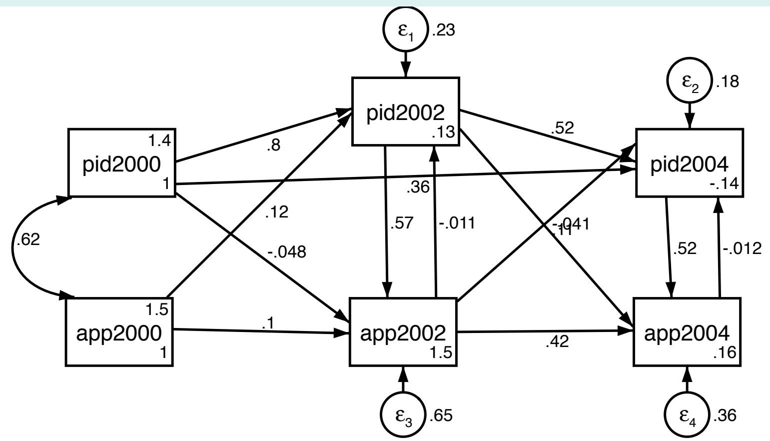
Modification indices

	MI	df	P>MI	EPC	Standard EPC
Structural					
pid2002 <-					
pid2004	102.723	1	0.00	-.398454	-.4252892
app2004	15.564	1	0.00	-.1289866	-.1068166
pid2004 <-					
pid2000	105.332	1	0.00	.3849349	.3561638
app2000	24.988	1	0.00	.1498549	.1073146
app2004 <-					
app2000	6.171	1	0.01	.0705983	.0651619
cov(e.pid2002,e.pid2004)	113.023	1	0.00	-.502823	-.4459717

Modification indices shows some unexplained covariation between PID2000 and PID2004 ("Presidential Year Effect"). Options: add either structural disturbance covariance, or a direct causal effect from PID2000->PID2004

So: final model (?) below:

Likelihood ratio		
chi2_ms(5)	21.595	model vs. saturated
p > chi2	0.001	
chi2_bs(14)	3458.563	baseline vs. saturated
p > chi2	0.000	
Population error		
RMSEA	0.067	Root mean squared error of approximation
90% CI, lower bound	0.040	
upper bound	0.097	
pclose	0.141	Probability RMSEA <= 0.05
Information criteria		
AIC	14702.656	Akaike's information criterion
BIC	14803.943	Bayesian information criterion
Baseline comparison		
CFI	0.995	Comparative fit index
TLI	0.987	Tucker-Lewis index
Size of residuals		
SRMR	0.016	Standardized root mean squared residual
CD	0.801	Coefficient of determination



Excellent GOF measures, though chi-square is still "significant"