Structural Equation Panel Models III: Measurement Error Models

PS 2701-2019

Longitudinal Analysis

Week 7

Professor Steven Finkel

Measurement Error Models

- The assumption in OLS regression that variables are perfectly measured is obviously violated in most research situations, but it is very difficult to overcome the problem in the cross-sectional context.
- With panel data, the information on the same variables over time allows much more flexibility in estimating "true" effects once measurement error is taken into account, and in estimating the measurement properties of the indicators as well. In the SEM framework, all of these effects can be estimated simultaneously

Measurement Error: The Basics

- Indicators of variables may contain error, in that the value that is assigned to a given unit is not the "true" value for that variable for that unit. Errors in variables may be:
  - **Systematic**, in which case we may say that the observed indicator is always off from the true value in one direction or the other, or that some other variable is also systematically influencing indicator, aside from the true variable of interest. In that case, we say the measure is not a ***valid*** indicator.
  - **Random**, in which case the observed indicator is sometimes higher or lower than the true value depending on random factors in the measurement process, such as (among other things):
    - poor record keeping
    - individual coder decisions (e.g. the people at Freedom House deciding on a 2 versus a 3 for some country's civil liberties index).
    - ambiguous questions in surveys
    - mood or other transient factors in the interview or observation process
    - scaling of variables (e.g. where does "3.55" attitude go on a 1-2-3-4-5 scale? Most of the time to 4, but some of the time no doubt to 3).

- Random errors lead indicators to be *unreliable*

- **ALL SOCIAL SCIENCE MEASURES ARE UNRELIABLE TO SOME EXTENT!!!!**
- **SOME ARE ALSO INVALID, THOUGH THIS PROBLEM IS MUCH MORE DIFFICULT TO DETECT AND CORRECT.**

- **SEM methods useful for correcting for reliability problems only!**

The Problem of Measurement Error for OLS Regression

- Assume the following cross-sectional model: an indicator *y** is a function of some "latent construct" Y and some random measurement error w_i:

  (1) $y_i^* = Y_i + w_i$

- Note regression coefficient of "1" for effect of Y on *y**. This is a convenient assumption that simply means that the two variables, the observed indicator *y** and the latent unobserved variable Y, are measured on the same scale (dollars, miles, number of conflicts, etc.). Since we do not observe Y or other latent variables, we always need to set their scale to *something*.

  (This is implicit in the regression coefficient of "1" for any equation's error term as well. The error term is unobserved and we conveniently set its scale to be the same as the endogenous variable in its equation.)

- Equation (1) above is likely to part of an overall causal system that has a *measurement* portion as well as a *structural* portion where the relationships among the Y variables are of theoretical interest as well. So assume that Y is caused by one exogenous X variable, itself with an imperfect "x" indicator:

$$(2)\quad \begin{aligned} &(a)\quad Y_i = \beta_1 X_i + \varepsilon_i \\ &(b)\quad y_i^* = Y_i + w_i \\ &(c)\quad x_i^* = X_i + v_i \end{aligned}$$

- Structural part of model is (2a); measurement part is (2b)-(2c)
- Assumptions:
  - $E(w)=E(v)=E(\varepsilon)=0$ (all errors have a mean of 0)
  - $E(X\varepsilon)=E(Xv)$, $E(Yw)=0$ (all errors are uncorrelated with their own equation's independent variables)
  - $E(\varepsilon v)=E(\varepsilon v)=E(vw)=0$ (all errors are uncorrelated with each other)
  - ε, v and w are normally distributed (as is the standard OLS assumption)

Measurement Error in X

Y is related to True X, but we use fallible x_i*
that contains random error v_i

$$Y_i = \beta_1 X_i + \varepsilon_i$$

(3) $$\text{and } \mathrm{x}_i^* = X_i + v_i$$

$$\text{then } \mathrm{Y}_i = \beta_1(\mathrm{x}_i^* - v_i) + \varepsilon_i$$

$$\text{and } \mathrm{Y}_i = \beta_1 \mathrm{x}_i^* + (\varepsilon_i - \beta_1 v_i)$$

x_i* is related to the error term (since v_i and x_i^* are related). So fallible x_i* is endogenous and we get inconsistent estimates of the causal effect β_1

Measurement Error, Reliablity and the Attenuation of OLS Estimates

- Another way to look at this: the OLS estimate of β_1 with the "fallible' measure x^* is:

$$\beta_{OLS} = \frac{Cov(Y_i x_i *)}{Var(x_i *)}$$

- With a little substitution from (3a), we can arrive at:

$$\beta_{OLS} = \frac{Cov(\beta_{TRUE} X_i + \varepsilon, x_i^*)}{Var(x_i^*)} = \frac{Cov(\beta_{TRUE} X_i x_i^*)}{Var(x_i^*)} = \beta_{TRUE} \frac{Cov(x_i^* X_i)}{Var(x_i^*)}$$

- Multiplying 2c through by X and taking expectations shows that Cov(x_i^*, X)=Var(X), so:

$$(4) \qquad \beta_{OLS} = \beta_{TRUE} \frac{Var(X_i)}{Var(x_i^*)}$$

- **Conclusion: Unless there is no measurement error in *x*, the (bivariate) OLS estimate of β will be less than the true value, and will be *attenuated* by the factor** $\frac{Var(X_i)}{Var(x_i^*)}$

- We call this factor, which is the ratio of "true score variance to observed score variance," the **reliability** of x (denoted as ϱ_{xx}). It is the proportion of the observed variance in x that is composed of the latent true score and *not* the measurement error v_i . So the OLS bivariate β equals the true β multiplied by the reliability of x

- Higher reliability means that the observed score is closely related to the true score and hence the attenuation of the OLS regression coefficient will be small; lower reliability means greater random noise in the indicator and consequently greater attenuation of the OLS regression coefficient in the bivariate case

Notes on Reliability

- The direction of bias due to measurement error in explanatory variables is *always* downward in the bivariate case; in the multivariate case it may be downwards or upwards, depending on the amount of measurement error in particular variables and their intercorrelations
- We can also take equation 2c, square both sides and take expectations to yield:

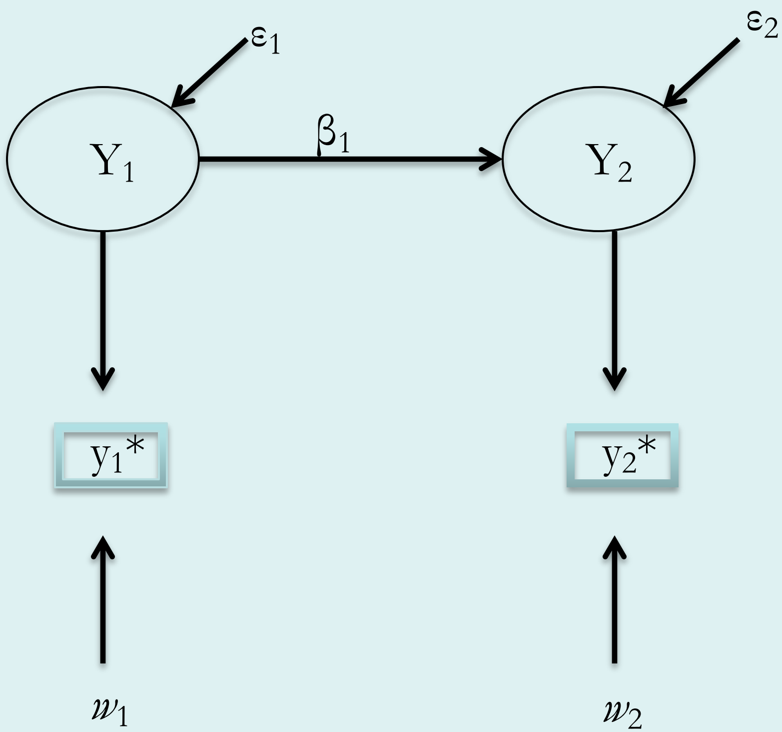
$$Var(x_i^*) = Var(X_i) + Var(v_i)$$

- which expresses the variance in a fallible indicator as composed of two parts: the true score variance and the error variance. So the reliability of *x* is the proportion of its variance being "true score" variance – it is akin to R^2 in that we can say that the higher the reliability, the lower the error variance in an indicator and thus the greater amount of variance 'explained" by the latent true score X

Measurement Error in Panel Models

- For panel analyses, measurement error is especially problematic. It turns out that measurement error in the lagged endogenous variable wreaks havoc with the estimation of a variable's stability, and hence throws off inferences in nearly all of the models that we have examined thus far
- The ability to model the substantive effect of "lagged y" will be compromised, along with the ability to use "lagged y" as a control for regression to the mean effects and for the estimation of the model's causal dynamics and long-range effects of X
- **Thus --- somewhat ironically --- correcting for measurement error is one of the key advantages of panel analyses, but it is also with panel data that measurement error corrections are sorely needed**

Example: Two Wave "Y-Only" Model with Measurement Error



MEASUREMENT MODEL FIGURE 1

NOTE: SEM convention to use circles to represent "latent" variables, squares to represent "observed" variables. Also using wave 1 as "endogenous" for ease of presentation; same concepts apply if wave 1 indicator is "x*" and "latent" variable is X

- We can write the structural portion of the model in equation form as:

$$(5a) \quad Y_1 = \varepsilon_1$$

$$(5b) \quad Y_2 = \beta_1 Y_1 + \varepsilon_2$$

- And the measurement portion as:

$$(6a) \quad y_1^* = Y_1 + w_1$$

$$(6b) \quad y_2^* = Y_2 + w_2$$

- Subtract equation 6(a) from 6(b) to arrive at an expression for the *change in y*:

$$\Delta y^* = \Delta Y + (w_2 - w_1)$$

- Multiply equation by y^*_1 in the first step, substitute (Y_1+w_1) for y^*_1 in one term on the right side of the equation in the second step, and take expectations to yield:

$$(7) \qquad Cov(\Delta y^*, y_1^*) = Cov(\Delta Y, Y_1) - Var(w_1)$$

- **THIS IS A VERY IMPORTANT RESULT!!!**

- It says that the covariance between the initial value and the change in a fallible indicator over time is equal to the "true" covariance between the *latent* initial true score and the change in the *latent* true score over time MINUS the amount of measurement error in the initial (time 1) indicator.

- So the more measurement error in y_1, the more we *overestimate* the amount of negative covariation between the variables initial level and change if we don't take the measurement error into account. That means that, with measurement error in the indicators, we think that there is *more* regression to the mean or speed of the system toward equilibrium than there really is.

- In terms of the *level* of the dependent variable, we think there is *less* stability from one time point to the next when we analyze the fallible indicators than there really is in the true score latent variables. And to the extent that exogenous variables X are related to the initial level of y_1* we will consequently wrongly estimate *their* impact on y_2* as well.

- Big problems, right?

Correcting for Measurement Error in SEM Panel Models: Single Indicators, Multiple Waves

- Consider again the causal system depicted in Figure 1 (Equations 5 and 6):
- The structural portion:

$$(5a) \quad Y_1 = \varepsilon_1$$

$$(5b) \quad Y_2 = \beta_1 Y_1 + \varepsilon_2$$

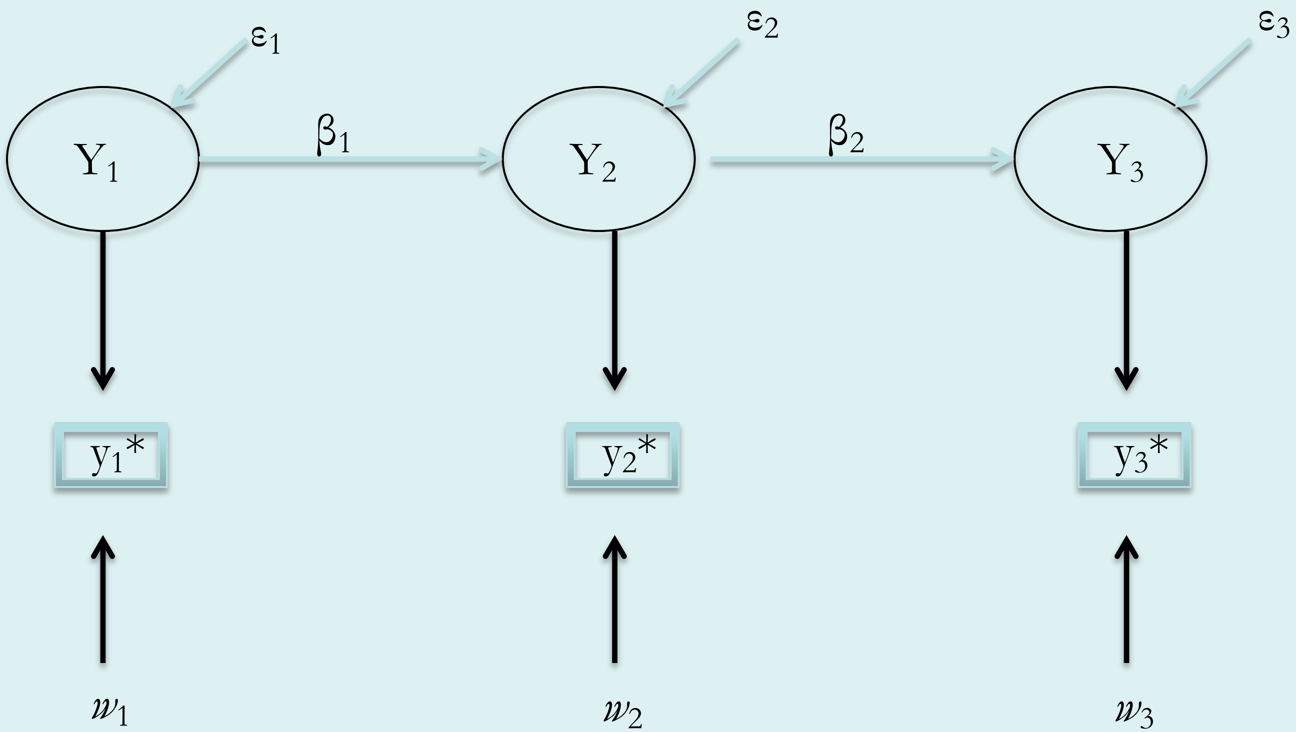
- The measurement portion:

$$(6a) \quad y_1^* = Y_1 + w_1$$

$$(6b) \quad y_2^* = Y_2 + w_2$$

- Are the model parameters identified?
- How many unknowns? Five:
  - One structural effect β_1
  - Two variances of the disturbance terms ε_1 and ε_2
  - Two measurement error variances of the w_1 and w_2
- How many knowns? Three: variance of y^* at time 1, the variance of y^* at time 2, and their covariance.
- So the two wave, single indicator model is underidentified!! We need either more waves or more indicators, or both!

Three Wave, Single Indicator Model



MEASUREMENT MODEL FIGURE 2

Is this model identified? YES, but only with equality constraints!

- **WILEY-WILEY (1970) SOLUTION: ASSUME EQUAL MEASUREMENT ERROR VARIANCES OVER TIME!!!! IF THIS IS THE CASE, WE HAVE A JUST-IDENTIFIED MODEL WITH 5 STRUCTURAL PARAMETERS AND 1 MEASUREMENT ERROR VARIANCE TO ESTIMATE.**

- Structural parameters: 2 "stability" effects and 3 error terms for the Y_t
- Measurement error variance for the *w*

- Finkel (p. 53-54) goes through some of the (tedious) covariance algebra to solve for the unknowns. Key results:

$$\beta_2 = \frac{Cov(y_1^*, y_3^*)}{Cov(y_1^*, y_2^*)}$$

$$\beta_1 = \frac{Cov(y_1^*, y_2^*)}{Var(y_1^*) - Var(w)}$$

$$Var(w) = Var(y_2^*) - \frac{Cov(y_2^*, y_3^*)Cov(y_1^*, y_2^*)}{Cov(y_1^*, y_3^*)}$$

- Pretty straightforward, but note that you first need to solve for Var(w) in order to arrive at the estimate for β_1.

EXAMPLE: PARTY IDENTIFICATION, 2000-2002-2004 NES PANEL STUDY

$$S = \begin{bmatrix} 4.68 & & \\ 4.11 & 4.75 & \\ 4.30 & 4.47 & 5.42 \end{bmatrix}$$

$$Var(w) = 4.75 - ((4.47 * 4.11) / 4.30) = .48$$

$$\beta_1 = 4.11 / (4.75 - .48) = .96$$

$$\beta_2 = 4.30 / 4.11 = 1.05$$

- Can also calculate the reliability of each indicator of *y*, following the formula given on slides 9-10 above:

$$\rho_{yy} = \frac{Var(Y_i)}{Var(y_i^*)} = \frac{Var(y_i^*) - Var(w_i)}{Var(y_i^*)} = 1 - \frac{Var(w_i)}{Var(y_i^*)}$$

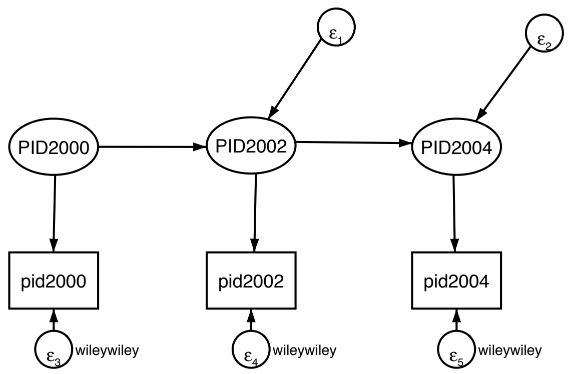
$$\rho_{11} = (4.68 - .48) / 4.68 = .90$$

$$\rho_{22} = (4.75 - .48) / 4.75 = .90$$

$$\rho_{33} = (5.42 - .48) / 5.42 = .91$$

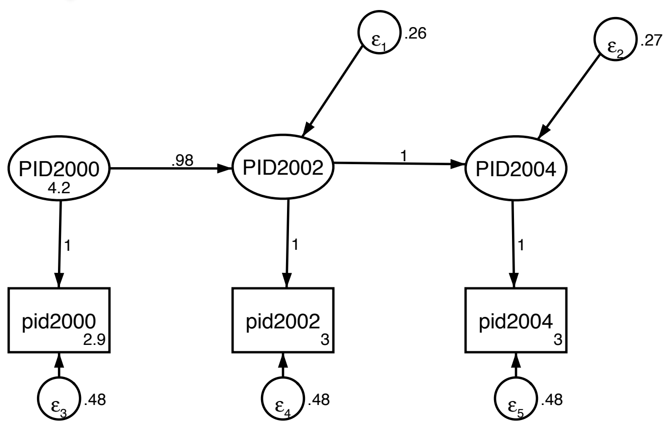
Notes on Example

- We have about 90% true score variance in the indicators and 10% error variance. This shows very high reliability of the partisanship variable, and means that using OLS on the fallible indicators would "not be so bad." Still, correcting for measurement error is necessary to yield unbiased estimates of the β.
- Once measurement error is taken into account, there is nearly perfect stability of partisanship over time. But we have a stability coefficient of 1.05 for waves 2 and 3, which means we probably have something wrong with the model.
- The applicability of the Wiley-Wiley procedure rests on the key assumption of equal measurement error over time. This is plausible in some cases, but it is sometimes the case that errors tend to shrink in systems over time as individuals/units "learn" and respond more consistently to stimuli in repeated measures. Implausible parameter estimates may be an indication of violation of this assumption.
- A closely related procedure for three-wave single indicator models was developed by Heise (1969). In that procedure, we use *standardized* variables for y, so we have 3 known correlations to work with. The model is identified by assuming equal reliabilities for the y indicators (as opposed to equal error variances in Wiley-Wiley)



STATA set-up with Wiley-Wiley constraint on measurement error variances

Results

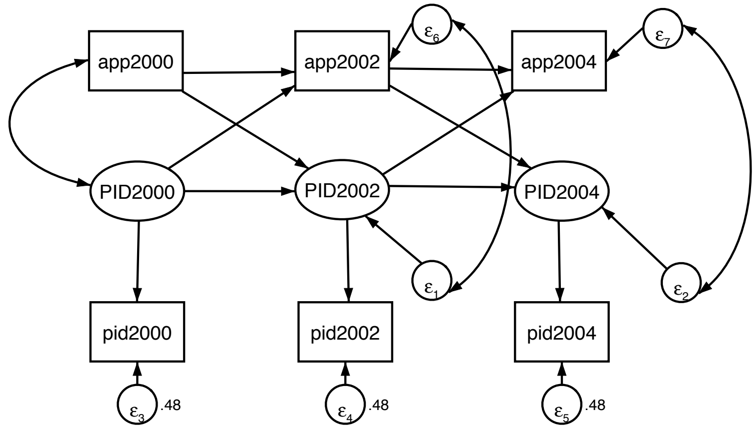


Notes:

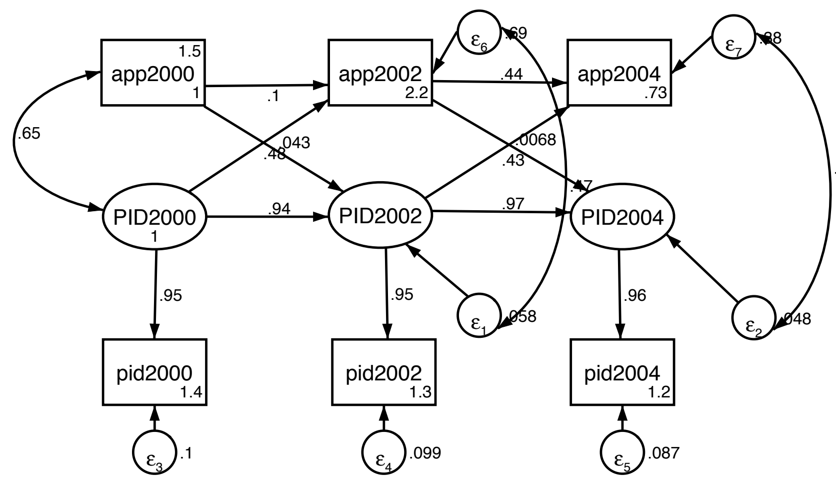
- Zero df, chi-square=0.0
- Estimated error variance of single indicator of PID=.48, as we calculated manually earlier
- Estimated stability=1.045 from wave 2 to wave 3, implausible (and rounded here to 1.0, so be careful on this!)

Additional Notes on Measurement Error Models

- How to include measurement error in more elaborate structural models? Two choices:
  - Estimate measurement error separately, and then plug in estimated values as "fixed parameters" in subsequent structural runs. In this case, we could take our estimate of .48 for the error variance, and use this value whenever we model Party ID in the cross-lagged or synchronous effects models
  - Estimate unconstrained (free) measurement and structural parameters simultaneously
  - Former method is simpler and more "stable", less prone to problems with implausible estimates or lack of convergence; latter method more defensible statistically because of a more properly specified model
- With more waves of observations, can relax Wiley-Wiley assumptions. In a four wave model, e.g., the "inner" elements of the measurement error variances are identified without constraint



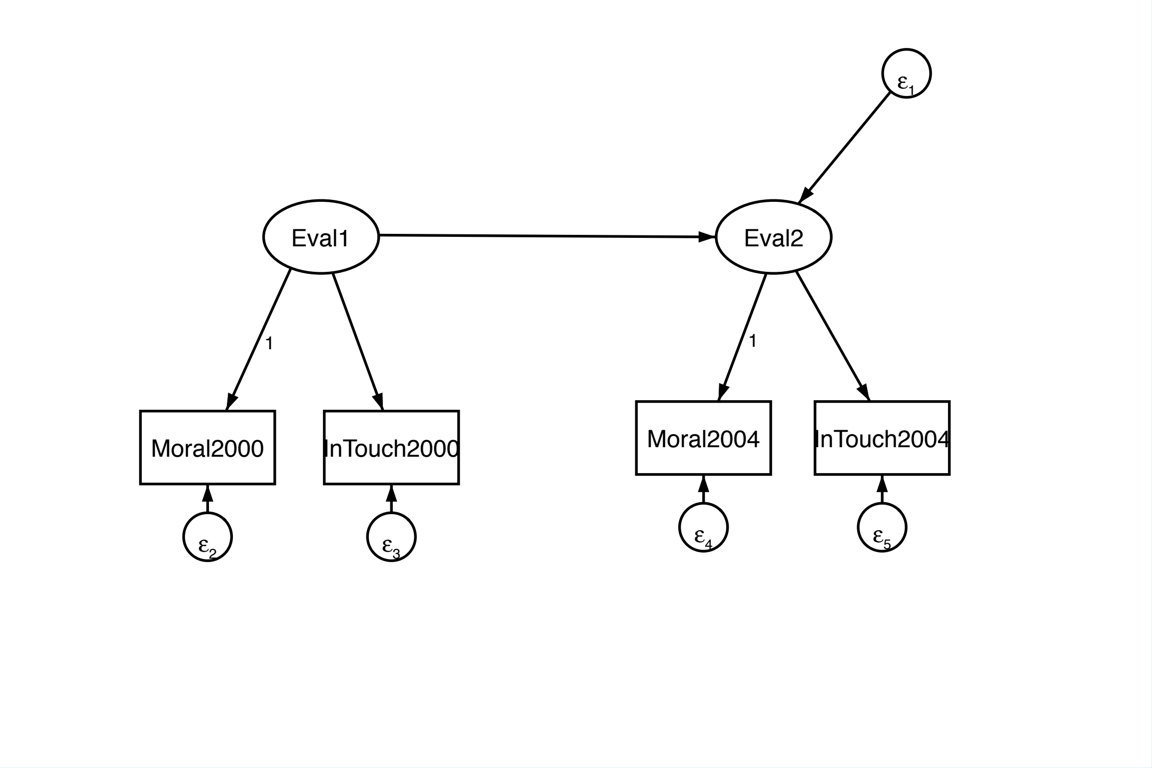
STATA estimation of three-wave single indicator panel model, measurement error in PID only, fixed at .48



LR test of model: chi2(4) = 10.15, Prob > chi2 = 0.0380

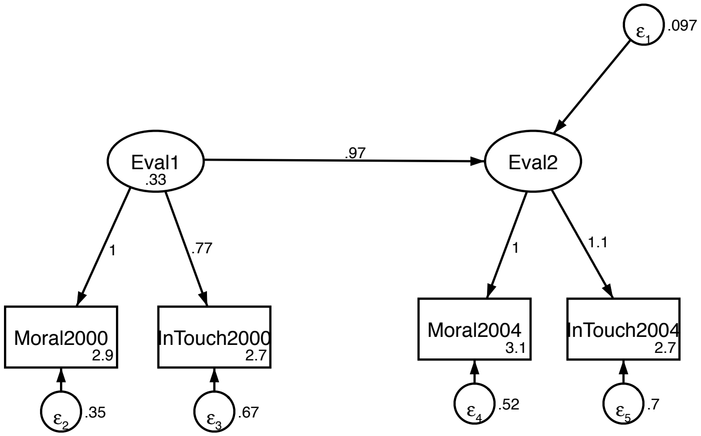
Correcting for Measurement Error in Panel Models: Multiple Indicators

- The last strategy for dealing with measurement error is to add additional indicators of each latent variable. If multiple indicators are available, this strategy is almost always the best, because the additional variances and covariances will often allow identification of measurement models without constraint, and because, logically, having more than one indicator of a latent variable should provide a more precise estimation of that variable for use in subsequent structural modeling.



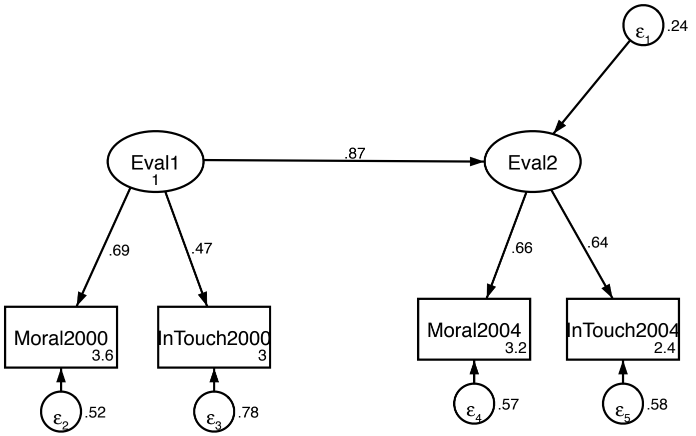
MEASUREMENT MODEL FIGURE 3:
Two Wave, Two Indicator Model

- Two indicators of the latent variable at each point in time.
- First, we scale the latent variable to equal the scale of one of the indicators by setting one of the "factor loadings" at each wave to be equal to 1. STATA does this automatically by setting the loading of the first indicator to 1; you can change this if you'd like. The other factor loading will be free to vary. Our unknowns total 9:
  - 1 β
  - 2 variances of the latent variable disturbances
  - 2 factor loadings (the regression coefficients of Y_1 and Y_2 on the second indicator at times 1 and 2, respectively)
  - 4 error variances of the observed indicators
- Number of knowns? 4*5/2=10 variances and covariances of the y^*. So we have an "overidentified" model!
- **Moral: With multiple indicators you only need two waves to identify measurement models, as opposed to three waves with single indicator models.**



Fit statistic	Value	Description
Likelihood ratio		
chi2_ms(1)	7.443	model vs. saturated
p > chi2	0.006	
chi2_bs(6)	218.035	baseline vs. saturated
p > chi2	0.000	
Population error		
RMSEA	0.135	Root mean squared error of approximation
90% CI, lower bound	0.058	
upper bound	0.233	
pclose	0.037	Probability RMSEA <= 0.05
Information criteria		
AIC	3645.608	Akaike's information criterion
BIC	3695.798	Bayesian information criterion
Baseline comparison		
CFI	0.970	Comparative fit index
TLI	0.818	Tucker-Lewis index
Size of residuals		
SRMR	0.025	Standardized root mean squared residual
CD	0.813	Coefficient of determination

unstandardized

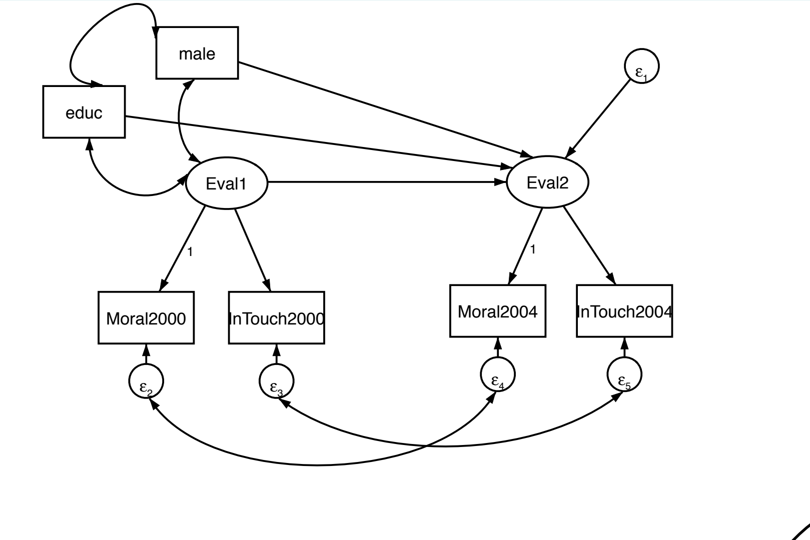


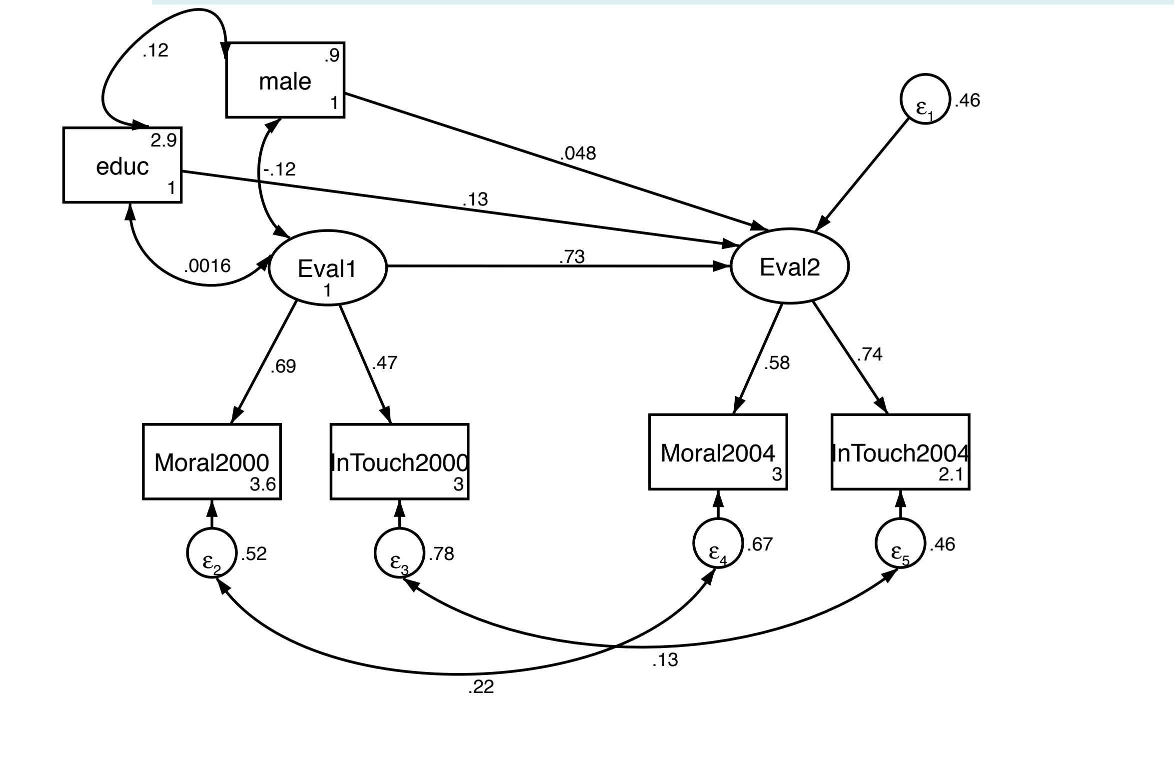
standardized

Notes on Multiple Indicator Models

- One conspicuously absent causal effect in this model is a free covariance term between the measurement errors of a given indicator with its own value over time. But should not the errors of measurement for indicator 1 at time 1 be related to the errors of measurement for indicator 1 at time 2? Such a correlation would be induced, for example, if the indicator was a measure not only of Latent Y but also of another construct that was unobserved in the model. It could also be induced through **method effects**, or **interviewer effects that are consistent across** time but are unrelated to the Latent Y
- It can be argued that these kinds of correlated measurement errors are *essential* to proper estimation of the causal effects in the model, notably the stability effect β_1, which, if these effects are ignored, will be overestimated to the extent that the correlated errors are not zero and thus account for some of the over-time covariation between the indicators. (We may also view the model χ^2 for the model without correlated errors as providing some information about the plausibility of the restrictions that they are zero, although there are many other restrictions in the model as given).

- However, including these correlated errors renders the model underidentified, as there are 11 unknowns and only ten knowns. What can you do?
  - Add exogenous background variables to the model. As long as these exogenous variables affect both Y_1 and Y_2 the model will be identified without constraint.
  - Add additional indicators – the 2 wave, 3 indicator model is identified without constraint.
  - Add additional waves of observations – the 3 wave, 2 indicator model is identified, though not the wave 1-3 error covariances without some constraint.
  - With more waves of measurement you can also include error covariances between indicators within a given wave. So measurement errors ε_1 and ε_2 might be related because they are questions that are right next to one another in a survey, independent of their respective success in serving as indicators of Y. More waves equals more covariances and more flexibility in estimating these effects. There are dangers, though, in overfitting error terms because they may not replicate across other data sets.

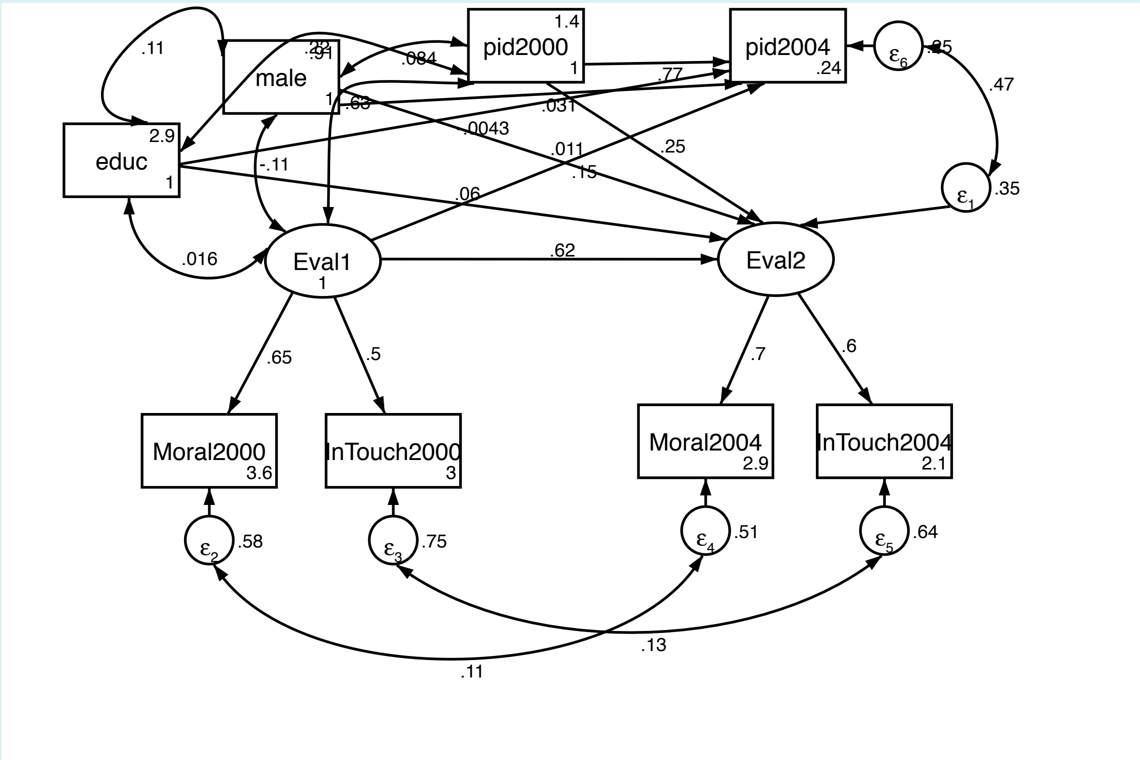




Notes:

- chi2(3) =3.45, P> chi2 =0.3271
- Excellent fit to the data
- Much lower estimated stability through inclusion of the correlated error terms

And of course it is straightforward to estimate measurement error along with structural effects between variables in the same model (especially so in STATA!!!)



Strong effect from PID onto Eval (.25), weaker but still significant effect from Eval to PID (.15)!

Two-wave cross-lagged model with multiple indicators in Eval, perfect measurement assumed in PID.

Chi2(7) = 14.07, Prob > chi2 = 0.05

Not bad!! But remember, the structural portion of the model is saturated (as before), so nothing really revealed here from the chi-square in that sense

- There is an important discussion in the measurement error literature regarding the stability of latent variable measurement structures. It is usually argued that the λ effects from η to a given indicator must be the same over time, or else the latent variable is not the "same" latent variable over time. That is, $\lambda_{21} = \lambda_{42}$ (and if there was a third wave it would also be equal to λ_{63}). This is the assumption of **"measurement invariance"** that many argue is essential for proper causal inference.
- This constraint can easily be imposed via an equality command. It is recommended to start with this model and try **not to reject** it and then no one will criticize you for it. The philosophical issues are not settled, though, so you could *possibly* argue your way to a position whereby η is the "same" but its components change their weights from time to time.
- Estimating the reliabilities of indicators in multiple indicator models is straightforward: use $1 - \frac{Var(\varepsilon_i)}{Var(y_i)}$ as before. Or, in a "standardized solution" (where y and η are both standardized), the reliability of y_i will be λ^2.
- For discussion of procedures for assessing measurement invariance, see Vandenberg and Lance, "A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research." *Organizational Research Methods* 3(1), 4–69 (2000)