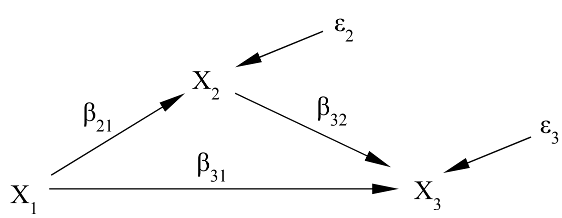
LECTURE 3: A SIMPLE STRUCTURAL MODEL WITH UNOBSERVED VARIABLES

I. WHAT'S SO STRUCTURAL ABOUT STRUCTURAL MODELS?
II. CONSEQUENCES OF MEASUREMENT ERROR IN LINEAR MODELS
III. SPECIFICATION OF A WALKING DOG MODEL.
IV. MOMENTS IN TERMS OF PARAMETERS AND VICE-VERSA.
V. ESTIMATION AND TESTING.

I. WHAT'S SO STRUCTURAL ABOUT STRUCTURAL MODELS?

We can now say something about the term "structure" in structural equation models. Note that the covariance structure models we've been working with could be used in a non-structural way, as merely a useful description of some sample data, or as a description of a set of variables in a population (often regression models are used in this way). But we are hoping to do a little more with our models than mere description of either a sample or a population. We want to get at an underlying structure, represented by our equations, and containing parameters that are in some sense "invariant." Take our Model I from last lecture as a simple example:



X_1 = Father's Income,
X_2 = Education
X_3 = Offspring's Income

The structural equations are:

$X_2 = \beta_{21} X_1 + \varepsilon_2$
$X_3 = \beta_{31} X_1 + \beta_{32} X_2 + \varepsilon_3$

In matrix form, the complete model (moments in terms of parameters) is:

$$\underset{3\times 3}{\underset{\Sigma}{\begin{bmatrix} \sigma_{11} & & \\ \sigma_{21} & \sigma_{22} & \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}}} = \underset{3\times 3}{\underset{\Sigma(\theta)}{\begin{bmatrix} \sigma_{11} & & \\ \beta_{21}\,\sigma_{11} & \beta_{21}^2\,\sigma_{11} + \sigma_{\varepsilon 22} & \\ \beta_{31}\,\sigma_{11} + \beta_{32}\,\sigma_{21} & \beta_{31}\,\sigma_{21} + \beta_{32}\,\sigma_{22} & \beta_{31}\,\sigma_{31} + \beta_{32}\,\sigma_{32} + \sigma_{\varepsilon 33} \end{bmatrix}}}$$

We hope that a structural model holds not for just one sample or one population, but for several or all populations of interest. Thus, if a model only held for a given population at a given moment in time and space, so that we'd have to change the model completely at the next moment in time—or if tiny changes in the population caused the model to fail—the model would not be of too much more utility than using the observed sample covariances to describe the population. We hope that some property of the model would be invariant across some time span and across some populations. Thus, a human capital model that only held for white males in the U.S. in 1950 would be much less useful than an alternative model that held also held for females, minorities, all western nations, and the period of late capitalism.

Note that this logic corresponds to what most social scientists are concerned with—identifying *patterns* of social action and interaction that tends to persist over time and space. Whenever we are specifying and testing models and doing model-based inference—as opposed to pure description, prediction, or forecasting—we are searching for invariant patterns or structure, which corresponds to our theoretical understanding of the phenomenon, and which we assume generated the sample data.

With this in mind, Duncan (1975) points out, if we changed the variance of X_1, which is an unanalyzed quantity in the model, we would change each and every covariance (and variance) in the model. (You can see this by noting that the expressions for σ_{32} and σ_{33} each contain σ_{21}, which is a function of σ_{11}.) Since we would be changing each variance of the model, we would necessarily be changing every standardized parameter and R^2 as well. However, we wouldn't be changing the structural parameters of the model -- particularly the βs. In other words, they would be invariant with respect to changes in σ_{11}. Because σ_{11} is exogenous to the model, it can be changed easily, by manipulating the assignments of subjects to treatments if it's an experiment, or moving to a different population (such as across race) that has less variation on X_2 (father's occupation), or in the case of drawing sample data, stratifying X_1 and varying the sampling from strata. The fact that *the βs do not change* when we change σ_{11} gives them a *property of invariance* that the observed covariances and correlations (and standardized coefficients) lack. The same holds for changing the variances of the disturbances: the covariances and correlations (and standardized coefficients) will change, but the βs will remain invariant.

This implies that R^2s and standardized coefficients *lack the property of invariance*, and are thus limited in describing underlying structure and making cross-population comparisions. Standardized population parameters, $p_{yx} = \beta\, \sigma_x/\sigma_y$, are a function of the variance of the independent and dependent variables, and the estimates are a function of the corresponding sample variances, $\hat{P}_{yx} = \hat{\beta} s_x/s_y$. Similarly, the R^2 for the model is dependent on a combination of moments and parameters $R^2 = 1 - s_{\varepsilon 11}/s_{11}$, where $s_{\varepsilon 11} = (N - 1)/(N - K)\ (s_{11} - \hat{\beta}_{31}\, s_{11} - \hat{\beta}_{12}\, s_{12} - \ldots - \hat{\beta}_{1K}\, s_{1K})$. Note that the R^2 *is not a parameter of the model*, but a summary of parameters and moments, and thus it is unlikely to be invariant across populations. Social scientists are notorious for making the error of comparing standardized coefficients and R^2s across populations. Thus, a difference in unstandardized parameters is often telling us about *differences in underlying structure* across populations, whereas differences in standardized coefficients could be due not to differences in structure, but rather *differences in observed sample moments*. In the applied literature, social scientists also tend to overemphasize the desirability of having large R^2s. It is desirable ceteris paribus –given one has the correct model and has estimated invariant parameters. A low R^2 *may* suggest that a model is misspecified, but if a model is properly-specified, it may simply mean that the phenomenon is not highly-structured. Moreover, a high R^2 may suggest that a model has succeeded in identifying structure, but that is *no*t necessarily so—there are many ways of artifically inflating R^2s in unmeaningful ways without using a structural model (e.g., add correlated regressors, aggregate the data, add regressors that are part of the dependent variable, use aggregate time-series data, etc.). In other words, an R^2 has no *necessary* connection to structure.

As noted above, a model that only fit a single population in a single point in time is much less useful than one that described the population in different time periods, or described a family of similar but different populations. The parameters of our structural model, the βs, would remain relatively invariant across different populations that varied only in the variances of the exogenous variables (e.g., σ_{11}) and the disturbance variances. The covariances would change, but that change is a kind of *"surface"* change, since the structural parameters remain invariant. A *"deeper"* structural change would occur *if the βs themselves changed*. For example, one could envision a simple human capital model of returns to education operating identically in advanced western democracies, but begin to change when the democracies turned to socialism. This would represent *a fundamental structural change* that is theoretically more meaningful than mere changes in observed covariances (see Duncan pp. 51-57 and 149-168).

> Aside: Hierarchical (multi-level) linear models are random coefficient models in which the βs are assumed to be random variables drawn from a known (e.g., normal) distribution, with a certain mean and variance. Thus, the β_is vary across some unit (like nations) according to a probability distribution. As random variables, they are treated as the dependent variables in a second equation where the subscript *i* refers to the unit, like nations: $\beta_i = \gamma\, X_i + \varepsilon_i$.

We can list out different forms of social change, ranging from superficial to deeply structural:

1. Random fluctuations in the data due to sampling error (within an invariant structural model).
2. Variation due to changes in the variance of exogenous variables (within an invariant structural model).
3. Variation due to changes in the above, plus changes in disturbance variances (within an invariant structural model).
4. Structural changes in the values of the structural parameters βs.
5. Structural changes in the specification of the structural model itself, requiring new or different structural parameters (e.g., structural differentiation in a biological organism).

Our simple human capital model was specified on individuals—predicting the income of individuals as a function of father's occupation and offspring's education. But the mechanisms by which father's occupation and education translate into income is *not* an individual process. It implicates a family system, an educational system, and a labor market system. Thus, one could argue that the underlying structure (if indeed, the model captures structure) is a *social* structure.

> Aside: Note the tradeoff in internal and external validity in seeking to identify structure in a randomized experiment versus a structural model on nonexperimental data. In the former, the disturbance term arises solely due to randomization, effects have a causal interpretation, since we've assigned treatments, conditional on the experimental setting, but we often have little confidence that such effects would hold in naturally-occurring populations of interest. In the latter, we have confidence that effects would hold in naturally-occurring populations, since we observe assignment to exogenous variables in natural settings, but must make a leap of faith that the model—and in particular, the disturbance specification—is correctly specified to allow us to talk about structure. In the econometrics and social statistics literatures there is a controversy over the relative merits of using structural (selection) models to estimate program effects versus using randomized field experiments.

But how do we know when we have a structural model with parameters that have some invariance? And how do we construct such a model? The answer to these questions takes us out of the realm of covariance structure analysis. The proper specification of a structural model is a substantive theoretical question. Whether a model is structural model depends on the state of the theoretical knowledge of the phenomenon of interest. As Duncan (1975) notes, often there is no structural model underlying variables of a given substantive area, and the area is waiting for a theory that will *invent* the proper variables that get at structure. Whether your model really represents structure must be judged in the context of the state of theoretical knowledge of the substantive area, including judgements of whether the assumptions of the model seem plausible, whether it is plausible to treat parameters as invariant, and whether more plausible alternative models exist. But the method cannot "reveal" structure to us from sample data—claims of TETRAD advocates (Glymour et al. 1988) notwithstanding (although we can rule out some possibilties). It cannot tell us which way causality flows or which variables are needed or how disturbances should be specified. These questions are answered by the accumulation of empirical knowledge combined with our theoretical understanding of the phenomenon. Structural equation models are only a tool for formalizing such theoretical knowledge in a parsimonious mathematical representation (model), and conditional on the strength of that knowledge and the resulting specification, estimating and testing the parameters of such a model. Duncan (1975, p. 150) offers some sage advice:

> Do not undertake the study of structural equation models (or, for that matter, any topic in sociological methods) in the hope of acquiring a technique that can be applied mechanically to a set of numerical data with the expectation that the result will automatically be "research."

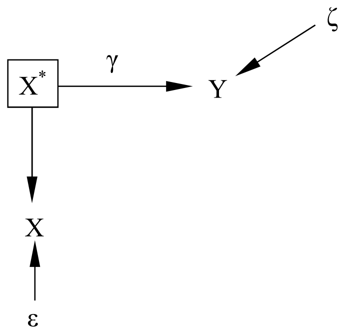
The single most important problem in the use of structural models is having a sufficiently *strong substantive theory* (and using it properly) to get at the underlying structure of a social process. One could make a strong argument (as Freedman 1987, 1991 does) that contemporary social science theory is not strong enough to specify truly structural models, and that our models do not get at invariant causal processes.

II. THE CONSEQUENCES OF MEASUREMENT ERROR IN LINEAR MODELS.

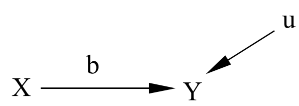
Much of the popularity of covariance structure analysis derives from its ability to estimate linear models with variables measured with error. We will reach two conclusions here: (1) random measurement error in independent variables causes the OLS estimator to be biased (and inconsistent) and inefficient, and estimated standard errors to be biased; (2) random measurement error in dependent variables results in OLS being unbiased but inefficient, with biased estimates of standard errors.

A. RANDOM MEASUREMENT ERROR IN AN INDEPENDENT VARIABLE.

Consider the following two models:



MODEL I: TRUE



MODEL II: WRONG

We've seen Model II before, and we know that $b = \sigma_{xy}/\sigma_x^2$ and $\sigma_u^2 = \sigma_y^2 - b^2\,\sigma_x^2$. Model I assumes that X, our observed independent variable is measured with *random error*. We specify X as a linear combination of X*, an unobservable variable, plus a random measurement error ε. Think of X* as the *"true"* variable that would result if we could measure X *perfectly*. We have two equations (assume we're in deviation scores):

$X = X^* + \varepsilon$

$Y = \gamma\, X^* + \zeta$

The first equation is a linear measurement equation, expressing the observed variable X as a function of the true score plus a random measurement error. The second is the substantive structural model, such as returns to education. Let's assume that $E(\varepsilon) = E(\zeta) = 0$ and $E(\varepsilon\, \zeta) = E(X^*\, \varepsilon) = E(X^*\, \zeta) = E(X\, \zeta) = E(y\, \varepsilon) = 0$. Later, when we try to estimate parameters from sample moments and do inference, we'll need to assume $\varepsilon_i \sim N(0,\sigma_\varepsilon^2)$, $\zeta_i \sim N(0,\sigma_\zeta^2)$, so let's just do so now. We have three moments σ_x^2, σ_y^2, and σ_{yx}, and four parameters (σ_x^2, $\sigma_{x^*}^2$, σ_ζ^2, γ). We have more parameters than moments (and normal equations) and thus the model is *underidentified*. We'll be able to express moments in terms of parameters, but not vice-versa, without additional information on the parameter values. (This also implies that we won't be able to estimate parameters from sample data without additional information.) For now, let's assume we know the values of the population parameters (and so could express parameters uniquely in terms of moments). We'll show what happens if we compute parameters using incorrect Model II instead of the correct Model I. Before doing that, it will be useful to give the standardized version of the first equation:

$X = X^* + \varepsilon \qquad X/\sigma_x = (\sigma_{x^*}/\sigma_x)\, X^*/\sigma_{x^*} + (\sigma_\varepsilon/\sigma_x)\, \varepsilon/\sigma_\varepsilon$

Standardized: $Z_x = P_{xx^*}\, Z_{x^*} + P_{x\varepsilon}\, Z_\varepsilon$ where $Z_x = X/\sigma_x$, $Z_{x^*} = X^*/\sigma_{x^*}$, $Z_\varepsilon = \varepsilon/\sigma_\varepsilon$, $P_{xx^*} = \rho_{xx^*} = \sigma_{x^*}/\sigma_x$, $P_{x\varepsilon} = \sigma_\varepsilon/\sigma_x$.

Back to the unstandardized equations. We can compute moments in terms of parameters for Model I:

1. $\sigma_x^2 = \sigma_{x^*}^2 + \sigma_\varepsilon^2$ which implies that $\sigma_{x^*}^2 = \sigma_x^2 - \sigma_\varepsilon^2$

The true score variance is the observed variance minus the error variance.

2. $\sigma_y^2 = \gamma^2\, \sigma_{x^*}^2 + \sigma_\zeta^2$
3. $\sigma_{yx} = \gamma\, \sigma_{x^*}^2$

Also, note that $\sigma_{yx} = \sigma_{yx^*}$ (by multiplying the first equation by Y) and $\sigma_{xx^*} = \sigma_{x^*}^2$ (by multiplying the first equation by X).
To illustrate the consequence of measurement error, let's solve for just two parameters in the true model (Model I), γ and σ_ζ^2 (and assume we know the population value of ε).

4. $\gamma = \sigma_{yx}/\sigma_{x^*}^2$

$\sigma_\zeta^2 = \sigma_y^2 - \gamma^2 \sigma_{x*}^2$

5. $\sigma_\zeta^2 = \sigma_y^2 - \gamma^2 \sigma_x^2 + \gamma^2 \sigma_\varepsilon^2$

Now, what would we get if we incorrectly computed parameters in terms of moments using the wrong model (Model II)? Let's express Model II's parameters in terms of the parameters of the true model (Model I).

$b = \sigma_{xy}/\sigma_x^2$ For (true) Model I, $\sigma_{yx} = \gamma\ \sigma_{x*}^2$. Therefore, we can replace σ_{yx}

$b = \gamma\ (\sigma_{x*}^2/\sigma_x^2)$ since from standardized results above, $\sigma_{x*}/\sigma_x = \rho_{xx*}$; it follows that

$b = \gamma\ \rho_{xx*}^2$ where ρ_{xx*}^2 is the squared correlation between X* and X, called the *reliability coefficient* (note Bollen labels this ρ_{xx}).

The above implies that in the population, if the true model entails random measurement error in X, and we incorrectly assume a perfectly-measured X, we'll get the wrong value for γ. Note also that if you somehow knew the value of ρ_{xx*}^2 , for example, from an earlier study, you could correct for attenuation in b,

$\gamma = b\ /\ \rho_{xx*}$. This assumes that we are in the population. What if we are relying on sample data and we try to estimate the parameters of Model I using the OLS estimators from Model II? Let's start with the OLS estimator of b for Model II:

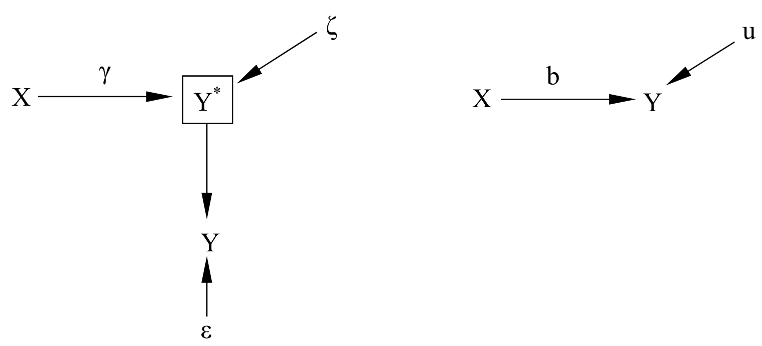
$\hat{b} = s_{xy}/s_x^2$ If we use $\hat{b}$ to estimate γ, we'll get a biased estimate:

$E(\hat{b}) = \gamma\ \rho_{xx*}^2$ The OLS estimator of γ will be biased when random measurement error is ignored: since $\rho_{xx*}^2 <$ 1, it will underestimate γ.

Conclusion: When X is measured with random error, the OLS estimator of the regression coefficient will be downwardly biased, and this holds regardless of sample size.

B. RANDOM MEASUREMENT ERROR IN A DEPENDENT VARIABLE.

Consider the following two models:



MODEL I: TRUE MODEL II: WRONG

Model I assumes that Y, our observed dependent variable is measured with random error: Y is a linear combination of Y*, an unobservable variable, plus a random measurement error ε. Again, assume we're in deviation scores):

$Y = Y^* + \varepsilon$

Standardized: $Z_y = P_{yy*}\ Z_{y*} + P_{y\varepsilon}\ Z_\varepsilon$ where $Z_y = X/\sigma_y$, $Z_{y*} = X^*/\sigma_{y*}$, $Z_\varepsilon = \varepsilon/\sigma_\varepsilon$, $P_{yy*} = \rho_{yy*} = \sigma_{y*}/\sigma_y$, and $P_{y\varepsilon} = \sigma_\varepsilon/\sigma_y$.

The second equation is again our substantive bivariate regression model when Y is corrected for measurement error.

$Y^* = \gamma X + \zeta$

Assume that $E(\varepsilon) = E(\zeta) = 0$ and $E(\varepsilon,\zeta) = 0$. Later, when we try to estimate parameters from sample moments we'll need to assume $\varepsilon_i \sim N(0,\sigma_\varepsilon^2)$, $\zeta_i \sim N(0,\sigma_\zeta^2)$, so again let's just do so now. Again, we have three moments σ_x^2, σ_y^2, and σ_{yx}, and four parameters (σ_x^2, σ_{y*}^2, σ_ζ^2, γ). (Again, without additional information about the parameter values, we would not be able to estimate this model from sample data.) For now, let's assume we know the values of the population parameters, and show what happens if we estimated Model II instead of Model I.

We can compute moments in terms of parameters for Model I:

$\sigma_x^2 = \sigma_x^2$

$\sigma_y^2 = \sigma_{y*}^2 + \sigma_\varepsilon^2$ which implies that $\sigma_{y*}^2 = \sigma_y^2 - \sigma_\varepsilon^2$ The true score variance is the observed variance minus the error variance.

$\sigma_{yx} = \gamma\, \sigma_x^2$

Also, $\sigma_{y*}^2 = \gamma^2\, \sigma_x^2 + \sigma_\zeta^2$

To illustrate the consequence of measurement error, let's solve for just two parameters, γ and σ_ζ^2 (and assume we know the population value of ε).

$\gamma = \sigma_{yx}/\sigma_x^2$

$\sigma_\zeta^2 = \sigma_{y*}^2 - \gamma^2\, \sigma_x^2$ but from above, $\sigma_{y*}^2 = \sigma_y^2 - \sigma_\varepsilon^2$; therefore,

$\sigma_\zeta^2 = \sigma_y^2 - \sigma_\varepsilon^2 - \gamma^2\, \sigma_x^2$

Now, what would we get if we (incorrectly) estimated the wrong model (Model II)? Let's express Model II's parameters in terms of the parameters of the true model (Model I).

$b = \sigma_{xy}/\sigma_x^2$ This is exactly the correct formula for γ

Therefore, ignoring random measurement error in the dependent variable has no effect on our computation of the regression coefficient from observable moments. What about the disturbance variance?

$\sigma_\zeta^2 = \sigma_y^2 - \gamma^2\, \sigma_x^2 - \sigma_\varepsilon^2$ $\sigma_u^2 = \sigma_y^2 - b^2\, \sigma_x^2$ but since $b = \gamma$, $\sigma_u^2 = \sigma_y^2 - \gamma^2\, \sigma_x^2$

$\sigma_\zeta^2 = \sigma_u^2 - \sigma_\varepsilon^2$ Substitute σ_u^2 for the first two terms:

$\sigma_u^2 = \sigma_\zeta^2 + \sigma_\varepsilon^2$

Therefore, if we ignore random measurement error in the dependent variable, our formula for the variance of the disturbance term from observable moments will be wrong—it'll give us a value that is too big, since it pools the true disturbance variance plus the measurement error variance. This means that our formula for the R^2 for the equation will also be wrong (too small):

R^2 (Model I) = $1 - (\sigma_\zeta^2/\sigma_y^2)$ for (true) Model I. But for (wrong) Model II, we get

R^2 (Model II) = $1 - (\sigma_u^2/\sigma_y^2) = 1 - (\sigma_\zeta^2 + \sigma_\varepsilon^2)/\sigma_y^2$ Since σ_u^2 is too big, R^2(II) will be too small (unless $\sigma_\varepsilon^2 = 0$).
Aside: This decomposition of the disturbance is a variance components model, used in random effects ANOVA designs, pooled time-series cross-section econometric models, and random intercept HLM models, where σ_ε^2

represents the error component (assumed orthogonal to regressors) that varies either across time or across cross-sections (depending on the model) but not both, whereas σ_ζ^2 is an error term orthogonal to σ_ε^2 that varies across both time and cross-sections. (For a discussion of LISREL as a variance components model, see Jöreskog 1978 *Annals de l'INSEE no. 30-31*).

What if we are relying on sample data and we try to estimate the parameters of Model I using the estimators from Model II? Let's start with the OLS estimator of b for Model II:

$\hat{b} = s_{xy}/s_x^2$ If we use $\hat{b}$ to estimate γ, we'll get an unbiased estimate: $E(\hat{b}) = \sigma_{xy}/\sigma_x^2 = \gamma$.

However, the variance of $\hat{b}$, $E(\hat{b}^{\,2}) = \sigma_u^2/\Sigma(x_i - \bar{x})^2$ will be bigger than the variance of $\hat{b}$, $E(\hat{b}^{\,2}) = \sigma_\zeta^2/\Sigma(x_i - \bar{x})^2$, making $\hat{b}$ inefficient $E(\hat{b}^2) > E(\hat{b}^2)$. That inefficiency is dependent on the magnitude of σ_ε.

The least squares residual will give a biased estimate of the disturbance variance:

$s_u^2 = s_y^2 - \hat{b}^{\,2} s_x^2$

$E(s_u^2) = \sigma_u^2 = \sigma_\zeta^2 + \sigma_\varepsilon^2$ If $\sigma_\varepsilon^2 > 0$, estimates will be upwardly biased (too big).

This means that the estimated standard errors of $\hat{b}$ will be biased, since they depend on the biased estimate of the disturbance:

$s_b = [s_u^2/\Sigma(x_i - \bar{x})^2]^{|}$

The estimated standard errors will be biased upward because the estimated disturbance variance was biased upward.

Conclusion: When Y is measured with random error, the OLS estimator of the regression coefficient will be *unbiased but inefficient*, and the estimated disturbance variance and estimated standard error will be upwardly biased. This result holds regardless of sample size.

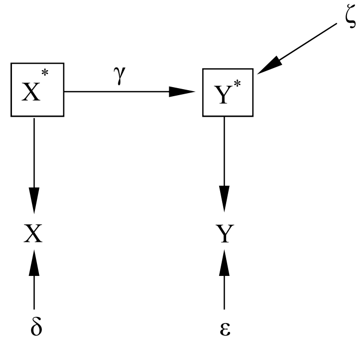
These results, for random measurement error in the independent and dependent variables can be generalized to the multiparameter and multiequation cases. First, in a single-equation multiple regression, when one independent variable, X_1 is estimated with random measurement error, and that error is ignored, OLS estimates of the coefficient for X_1 will be biased and inefficient. Moreover, if X_1 is correlated with other independent variables, their effects will also be estimated with bias (but the direction of bias is difficult to predict.) When the dependent variable Y_1 is measured with random error, all coefficients estimated by OLS will be unbiased but inefficient and estimated standard errors will be upwardly biased. In models with multiple indicators that have correlated measurement errors, the direction of biases are less predictible (see Bollen 1989, pp. 167-171 for a discussion of bias and inconsistency).

C. CORRECTING A CORRELATION FOR ATTENUATION DUE TO UNRELIABILITY.

Here is the classical formula for a correlation coefficient corrected for attenuation due to unreliability:

$$\rho_{x^*y^*} = \rho_{xy}/(\rho_{xx^*}^2\, \rho_{yy^*}^2)^{|}$$

where $\rho_{x^*y^*}$ is the true correlation between X* and Y*, ρ_{xy} is the observed correlation between X and Y, and $\rho_{xx^*}^2$ and $\rho_{yy^*}^2$ are the reliability coefficients for X and Y, respectively. We can derive this using the results above:



$Y = Y^* + \varepsilon$ $P_{yy^*} = \rho_{yy^*} = \sigma_{y^*}/\sigma_y$ (standardized)

$X = X^* + \delta \qquad P_{xx^*} = \rho_{xx^*} = \sigma_{x^*}/\sigma_x$ (standardized)

$Y^* = \gamma X^* + \zeta \quad P_{y^*x^*} = \rho_{y^*x^*} = \gamma\, \sigma_{x^*}/\sigma_{y^*}$ (standardized)

$\sigma_{yx} = \sigma_{y^*x^*} = \gamma\, \sigma_{x^*}^2$

Begin with the correlation between observed Y and X:

$\rho_{yx}^2 = \sigma_{yx}^2/(\sigma_x^2\, \sigma_y^2)$	We can replace $\sigma_{yx} = \gamma\, \sigma_{x^*}^2$
$\rho_{yx}^2 = \gamma^2\, \sigma_{x^*}^2/(\sigma_x^2\, \sigma_y^2)$	Multiply this out
$\rho_{yx}^2 = \gamma^2\, (\sigma_{x^*}^2/\sigma_x^2)\, (\sigma_{x^*}^2/\sigma_y^2)$	but $\rho_{xx^*}^2 = \sigma_{x^*}^2/\sigma_x^2$
$\rho_{yx}^2 = \rho_{xx^*}^2\, \gamma^2\, (\sigma_{x^*}^2/\sigma_y^2)$	Multiply the right-hand side by $\sigma_{y^*}^2/\sigma_{y^*}^2 = 1$
$\rho_{yx}^2 = \rho_{xx^*}^2\, \gamma^2\, (\sigma_{x^*}^2/\sigma_y^2)(\sigma_{y^*}^2/\sigma_{y^*}^2)$	Rearrange
$\rho_{yx}^2 = \rho_{xx^*}^2\, \gamma^2\, (\sigma_{x^*}^2/\sigma_{y^*}^2)(\sigma_{y^*}^2/\sigma_y^2)$	We can replace $\rho_{y^*x^*}^2 = \gamma^2\, \sigma_{x^*}^2/\sigma_{y^*}^2$ from above
$\rho_{yx}^2 = \rho_{xx^*}^2\, \rho_{y^*x^*}^2\, (\sigma_{y^*}^2/\sigma_y^2)$	We can also replace $\rho_{yy^*}^2 = \sigma_{y^*}^2/\sigma_y^2$
$\rho_{yx}^2 = \rho_{xx^*}^2\, \rho_{y^*x^*}^2\, \rho_{yy^*}^2$	Take the square root of both sides

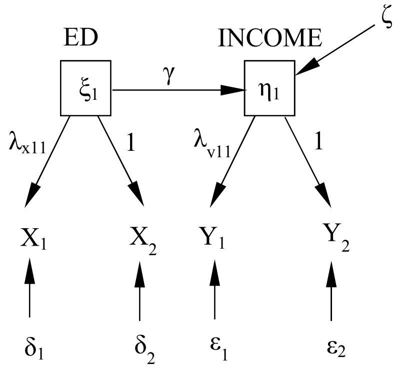
$\rho_{x^*y^*} = \rho_{xy}/(\rho_{xx^*}^2\, \rho_{yy^*}^2)^{1/2}$

Thus, if we know the reliabilities of X and Y, we can correct the observed correlation for attenuation due to unreliability by simply dividing the correlation by the square roots of the reliabilities of X and Y. This technique is often used in psychology and sociology.

But from the standpoint of structural equation modeling, a more important issue is to estimate the reliabilities of indicators within a measurement model from sample data, test such a model, and correct for attenuation due to unreliability in multivariate models. The principal motivation for developing the LISREL approach to covariance structure analysis was to incorporate models of measurement error into structural equation models.

II. SPECIFICATION OF A WALKING DOG MODEL.

Consider the following simple structural model with four observable variables and two unobservable or latent variables. Suppose we are interested in returns to education, but have pairs of fallible indicators of income and education, which we assume are each valid measures but measured with some random measurement error. For example, they could be reports given by the respondent and the respondent's parent, or they could be a self-report and census measures.



There are four measurement equations relating latent (unobserved variables) to observed indicators:

(1) $x_1 = \lambda_{x11}\, \xi_1 + \delta_1$ (3) $y_1 = \lambda_{y11}\, \eta_1 + \varepsilon_1$

(2) $x_2 = \xi_1 + \delta_2$ (4) $y_2 = \eta_1 + \varepsilon_2$

Each measurement equation specifies the observed variable to be a linear combination of a latent true variable plus a random measurement error. There is also a substantive equation relating unobserved variables:

(5) $\eta_1 = \gamma_{11}\,\xi_1 + \zeta_1$

Let's make the usual assumptions about disturbances being uncorrelated with each other, with regressors in their equations, having no serial correlation, having constant variance zero means, etc. Suppose that η_1 is income and ξ_1 is education. Then equation (5) gives us returns to education after controlling for measurement error. The latent variable does not have an inherent metric, so we have to assign it one. We've done this above by specifying that $\lambda_{x21} = \lambda_{y21} = 1$, which fixes the metric of η_1 to be equal to that of x_2, ξ_1 to equal that of y_2. (We could also have specified the other λs = 1, or specified the variances of η_1 and ξ_1 to be some constant, such as 1.0. More on measurement parameters later.) Therefore, we have 10 moments and 9 parameters (λ_{x11}, λ_{y11}, $\sigma_{\delta1}^2$, $\sigma_{\delta2}^2$, $\sigma_{\varepsilon1}^2$, $\sigma_{\varepsilon2}^2$, $\sigma_{\xi1}^2$, γ_{11}, and $\sigma_{\zeta1}^2$). That implies one overidentifying restriction on our sample moments.

We'll simplify the notation by using LISREL notation for structural variances:

$E(\xi_1^2) = \phi_{11}$ $\quad E(\delta_1^2) = \theta_{11}^{\delta}$ $\quad E(\delta_2^2) = \theta_{22}^{\delta}$ $\quad E(\varepsilon_1^2) = \theta_{11}^{\varepsilon}$ $\quad E(\varepsilon_2^2) = \theta_{22}^{\varepsilon}$ $\quad E(\zeta_1^2) = \psi_{11}$

To compute moments in terms of parameters, begin with two preliminary computations:

$E(\eta_1^2) = \gamma_{11}^2\,\phi_{11} + \psi_{11}$ since $E(\eta_1^2) = E(\gamma_{11}\,\xi_1 + \zeta_1)^2 = \gamma_{11}^2\,\underbrace{E(\xi_1^2)}_{\phi_{11}} + \underbrace{E(\zeta_1^2)}_{\psi_{11}} + 2\gamma_{11}\,\underbrace{E(\xi_1\,\zeta_1)}_{0}$

$E(\xi_1\,\eta_1) = \gamma_{11}\,\phi_{11}$ since $E(\xi_1\,\eta_1) = E[\xi_1\,(\gamma_{11}\,\xi_1 + \zeta_1)] = \gamma_{11}\,E(\xi_1^2) + E(\xi_1\,\zeta_1)$

II. MOMENTS IN TERMS OF PARAMETERS AND VICE-VERSA.

Now, let's compute observable moments in terms of parameters:

1. $\sigma_{x1}^2 = \lambda_{x11}^2\,\phi_{11} + \theta_{11}^{\delta}$ $\qquad E(x_1^2) = (\lambda_{x11}\,\xi_1 + \delta_1)^2 = (\lambda_{x11})^2\underbrace{E(\xi_1^2)}_{\phi_{11}} + \underbrace{E(\delta_1^2)}_{\theta_{11}^{\delta}} + 2\lambda_{x11}\underbrace{E(\xi_1\,\delta_1)}_{0}$

2. $\sigma_{x2}^2 = \phi_{11} + \theta_{22}^{\delta}$ $\qquad E(x_2^2) = (\xi_1 + \delta_2)^2 = \underbrace{E(\xi_1^2)}_{\phi_{11}} + \underbrace{E(\delta_2^2)}_{\theta_{22}^{\delta}} + 2\,\underbrace{E(\xi_1\,\delta_2)}_{0}$

3. $\sigma_{x2x1} = \lambda_{x11}\,\phi_{11}$ $\qquad E(x_2x_1) = \lambda_{x11}\,E(_{\xi1}^2) + E(\delta_1\delta_2) = \lambda_{x11}\,\phi_{11} \underbrace{+\,\theta_{21}^{\delta}}_{\text{0 in this model}}$

4. $\sigma_{y1}^2 = \lambda_{y11}^2\,(\gamma_{11}^2\,\phi_{11} + \psi_{11}) + \theta_{11}^{\varepsilon}$ $\qquad E(y_1^2) = \lambda_{y11}^2\,E(\eta_1^2) + E(\varepsilon_1^2) = \lambda_{y11}^2\,(E(\eta_1^2)) + \sigma_{\varepsilon1}^2$

   but from above, $E(\eta_1^2) = \gamma_{11}^2\,\phi_{11} + \psi_{11}$

5. $\sigma_{y2}^2 = \gamma_{11}^2\,\phi_{11} + \psi_{11} + \theta_{22}^{\varepsilon}$ $\qquad E(y_2^2) = E(\eta_1^2) + E(\varepsilon_2^2) = \gamma_{11}^2\,\phi_{11} + \psi_{11} + \theta_{22}^{\varepsilon}$

6. $\sigma_{y2y1} = \lambda_{y11}\,(\gamma_{11}^2\,\phi_{11} + \psi_{11})$ $\qquad E(y_2y_1) = \lambda_{y21}\,\lambda_{y11}\,(\gamma_{11}^2\,\phi_{11} + \psi_{11}) + \theta_{21}^{\varepsilon}$

7. $\sigma_{y1x1} = \lambda_{x11}\,\lambda_{y11}\,\gamma_{11}\,\phi_{11}$ $\qquad E(x_1y_1) = \lambda_{x11}\,\lambda_{y11}\,E(\xi_1\eta_1) + E(\delta_1\varepsilon_1) = \lambda_{x11}\,\lambda_{y11}\,E(\xi_1\eta_1)$

   cross-products = 0, but from above, $E(\xi_1\,\eta_1) = \gamma_{11}\,\phi_{11}$

8. $\sigma_{y2x1} = \lambda_{x11}\, \gamma_{11}\, \phi_{11}$ $E(x_1y_2) = \lambda_{x11}\, E(\xi_1\eta_1) + E(\delta_1\varepsilon_2) = \lambda_{x11}\, \gamma_{11}\, \phi_{11}$

9. $\sigma_{y1x2} = \lambda_{y11}\, \gamma_{11}\, \phi_{11}$ $E(x_2y_1) = \lambda_{y11}\, E(\xi_1\eta_1) + E(\delta_2\varepsilon_1) = \lambda_{y11}\, \gamma_{11}\, \phi_{11}$

10. $\sigma_{y2x2} = \gamma_{11}\, \phi_{11}$ $E(x_2y_2) = E(\xi_1\eta_1) + E(\delta_1\varepsilon_2) = \gamma_{11}\, \phi_{11}$

We can arrange these results into our usual matrices: $\Sigma = \Sigma(\theta)$

$$\underset{4\times 4}{\Sigma} = \begin{bmatrix} \sigma^2_{x_1} & & & \\ \sigma_{x_2x_1} & \sigma^2_{x_2} & & \\ \sigma_{y_1\,x_1} & \sigma_{y_1\,x_2} & \sigma^2_{y_1} & \\ \sigma_{y_2x_1} & \sigma_{y_2x_2} & \sigma_{y_2y_1} & \sigma^2_{y_2} \end{bmatrix}$$

$$\underset{4\times 4}{\Sigma(\theta)} = \begin{bmatrix} \lambda^2_{x_{11}}\Phi_{11} + \theta^{\delta}_{11} & & & \\ \lambda_{x_{11}}\Phi_{11} & \Phi_{11} + \theta^{\delta}_{22} & & \\ \lambda_{y_{11}}\lambda_{x_{11}}\gamma^2_{11}\Phi_{11} & \lambda_{y_{11}}\gamma^2_{11}\Phi_{11} & \lambda^2_{y_{11}}(\gamma^2_{11}\Phi_{11} + \psi_{11}) + \theta^{\varepsilon}_{11} & \\ \lambda_{x_{11}}\gamma^2_{11}\Phi_{11} & \lambda_{y_{11}}\gamma^2_{11}\Phi_{11} & \lambda_{y_{11}}(\gamma^2_{11}\Phi_{11} + \psi_{11}) & \gamma^2_{11}\Phi_{11} + \psi_{11} + \theta^{\varepsilon}_{22} \end{bmatrix}$$

Now, let's compute parameters in terms of moments:

For λ_{y11}, divide (9) by (10): $\sigma_{y1x2}/\sigma_{y2x2} = (\lambda_{y11}\, \gamma_{11}\, \Phi_{11})/\gamma_{11}\, \Phi_{11}$

$\lambda_{y11} = \sigma_{y1x2}/\sigma_{y2x2}$

But we can also divide (7) by (8): $\sigma_{y1x1}/\sigma_{y2x1} = (\lambda_{x11}\, \lambda_{y11}\, \gamma_{11}\, \Phi_{11})/(\lambda_{x11}\, \gamma_{11}\, \Phi_{11})$

$\lambda_{y11} = \sigma_{y1x1}/\sigma_{y2x1}$

1. $\lambda_{y11} = \sigma_{y1x2}/\sigma_{y2x2} = \sigma_{y1x1}/\sigma_{y2x1}$

For λ_{x11}, divide (8) by (10): $\sigma_{y2x1}/\sigma_{y2x2} = (\lambda_{x11}\, \gamma_{11}\, \phi_{11})/\ \phi_{11}\, \gamma_{11}$

2. $\lambda_{x11} = \sigma_{y2x1}/\sigma_{y2x2} = \sigma_{y1x1}/\sigma_{y1x2}$ (the latter from dividing (7) by (9)

For γ_{11}, divide (8) by (3): $\sigma_{y2x1}/\sigma_{x2x1} = (\lambda_{x11}\, \gamma_{11}\, \phi_{11})/(\lambda_{x11}\, \phi_{11})$

3. $\gamma_{11} = \sigma_{y2x1}/\sigma_{x2x1} = \sigma_{y1x1}\, \sigma_{y2x2}/\sigma_{x2x1}\, \sigma_{x2y1}$ (the latter from replacing ϕ_{11} in (7)

We can stop here. For the *anal-retentive*, here are the rest of the parameters in terms of moments:

For ϕ_{11}, take (2) $\sigma_{x2x1} = \lambda_{x11}\, \phi_{11}$; we know that $\lambda_{x11} = \sigma_{y2x1}/\sigma_{y2x2}$; so $\sigma_{x2x1} = \sigma_{y2x1}/\sigma_{y2x2}\, \phi_{11}$

4. $\phi_{11} = \sigma_{x2x1}\, \sigma_{y2x2}/\sigma_{y2x1} = \sigma_{x2x1}\, \sigma_{y1x2}/\sigma_{y1x1}$ (the latter replacing $\lambda_{x11} = \sigma_{y1x1}/\sigma_{y1x2}$)

For $\sigma_{\delta 1}^2$, take (1) $\sigma_{x1}^2 - \lambda_{x11}^2\, \phi_{11} = \sigma_{\delta 1}^2$, and replace ϕ_{11} and $\lambda_{x11}^2 = \sigma_{y2x1}^2/\sigma_{y2x2}^2$

$\theta^{\delta}_{11} = \sigma_{x1}^2 - (\sigma_{y2x1}^2/\sigma_{y2x2}^2)(\sigma_{x2x1}\, \sigma_{y2x2}/\sigma_{y2x1}) = \sigma_{x1}^2 - (\sigma_{y2x1}\, \sigma_{x2x1}/\sigma_{y2x2}) = (\sigma_{x1}^2\, \sigma_{y2x2} - \sigma_{y2x1}\, \sigma_{x2x1})/\sigma_{y2x2}$

5. $\theta^{\delta}_{11} = (\sigma_{x1}^2\, \sigma_{y2x2} - \sigma_{y2x1}\, \sigma_{x2x1})/\sigma_{y2x2} = (\sigma_{x1}^2\, \sigma_{y1x2} - \sigma_{y1x1}\, \sigma_{x1x1})/\sigma_{y1x2}$

(the latter from replacing $\lambda_{x11} = \sigma_{y1x1}/\sigma_{y1x2}$)

For θ^{δ}_{22}, take (2) $\sigma_{x2}{}^2 = \phi_{11} + \theta^{\delta}_{22}$ and replace ϕ_{11} so that $\sigma_{x2}{}^2 = \sigma_{x2x1}\, \sigma_{y2x2}/\sigma_{y2x1} + \theta^{\delta}_{22}$

$\theta^{\delta}_{22} = \sigma_{x2}{}^2 - (\sigma_{x2x1}\, \sigma_{y2x2}/\sigma_{y2x1})$

6. $\theta^{\delta}_{22} = (\sigma_{x2}{}^2\, \sigma_{y2x1} - \sigma_{x2x1}\, \sigma_{y2x2})/\sigma_{y2x1} = (\sigma_{x2}{}^2\, \sigma_{y1x1} - \sigma_{x2x1}\, \sigma_{y1x2})/\sigma_{y1x1}$ (the latter from replacing ϕ_{11} $= \sigma_{x2x1}\, \sigma_{y1x2}/\sigma_{y1x1}$)

For $\sigma_{\zeta1}{}^2$, take (6) $\sigma_{y2y1} = \lambda_{y21}\, \lambda_{y11}\, (\gamma_{11}{}^2\, \phi_{11} + \psi_{11}) = \lambda_{y21}\, \lambda_{y11}\, \gamma_{11}{}^2\, \sigma_{\xi1}{}^2 + \lambda_{y21}\, \lambda_{y11}\, \psi_{11}$

$(\sigma_{y2y1} - \lambda_{y21}\, \lambda_{y11}\, \gamma_{11}{}^2\, \sigma_{\xi1}{}^2)/\lambda_{y21}\, \lambda_{y11} = \psi_{11}$ then replace all the parameters in the left-hand side:

7. $\psi_{11} = \sigma_{y2x2}\, (\sigma_{y2y1}\, \sigma_{x2x1} - \sigma_{y2x1}\, \sigma_{y1x2})/\sigma_{x1x2}\, \sigma_{y1x2} = \sigma_{y2x1}\, (\sigma_{y2y1}\, \sigma_{x2x1} - \sigma_{y2x1}\, \sigma_{y1x2})/\sigma_{x1x2}\, \sigma_{y1x1}$

For $\theta^{\varepsilon}_{11}$, take (4) $\sigma_{y1}{}^2 = \lambda_{y11}{}^2\, (\gamma_{11}{}^2\, \phi_{11} + \psi_{11}) + \theta^{\varepsilon}_{11}$, which gives $\theta^{\varepsilon}_{11} = \sigma_{y1}{}^2 - \lambda_{y11}{}^2\, (\gamma_{11}{}^2\, \phi_{11} + \psi_{11})$

Replace all parameters on the right-hand side with moments, and obtain:

8. $\theta^{\varepsilon}_{11} = (\sigma_{y1}{}^2\, \sigma_{y2x2} - \sigma_{y1x2}\, \sigma_{y2y1})/\sigma_{y2x2} = (\sigma_{y1}{}^2\, \sigma_{y2x1} - \sigma_{y1x1}\, \sigma_{y2y1})/\sigma_{y2x1}$

For $\theta^{\varepsilon}_{22}$, take (5) $\sigma_{y2}{}^2 = \gamma_{11}{}^2\, \phi_{11} + \psi_{11} + \theta^{\varepsilon}_{22}$, which gives $\theta^{\varepsilon}_{22} = \sigma_{y2}{}^2 - \gamma_{11}{}^2\, \phi_{11} + \psi_{11}$

Replace all parameters on the right-hand side with moments, and obtain:

9. $\theta^{\varepsilon}_{22} = (\sigma_{y2}{}^2\, \sigma_{y1x2} - \sigma_{y2x2}\, \sigma_{y2y1})/\sigma_{y1x2} = (\sigma_{y2}{}^2\, \sigma_{y1x1} - \sigma_{y2x1}\, \sigma_{y2y1})/\sigma_{y1x1}$

All this yields a list of parameters in terms of moments (exercise: check my algebra for errors or typos):

1. $\lambda_{y11} = \sigma_{y1x2}/\sigma_{y2x2} = \sigma_{y1x1}/\sigma_{y2x1}$
2. $\lambda_{x11} = \sigma_{y2x1}/\sigma_{y2x2} = \sigma_{y1x1}/\sigma_{y1x2}$
3. $\gamma_{11} = \sigma_{y2x1}/\sigma_{x2x1} = \sigma_{y1x1}\, \sigma_{y2x2}/\sigma_{x2x1}\, \sigma_{x2y1}$
4. $\phi_{11} = \sigma_{x2x1}\, \sigma_{y2x2}/\sigma_{y2x1} = \sigma_{x2x1}\, \sigma_{y1x2}/\sigma_{y1x1}$
5. $\theta^{\delta}_{11} = (\sigma_{x1}{}^2\, \sigma_{y2x2} - \sigma_{y2x1}\, \sigma_{x2x1})/\sigma_{y2x2} = (\sigma_{x1}{}^2\, \sigma_{y1x2} - \sigma_{y1x1}\, \sigma_{x1x1})/\sigma_{y1x2}$
6. $\theta^{\delta}_{22} = (\sigma_{x2}{}^2\, \sigma_{y2x1} - \sigma_{x2x1}\, \sigma_{y2x2})/\sigma_{y2x1} = (\sigma_{x2}{}^2\, \sigma_{y1x1} - \sigma_{x2x1}\, \sigma_{y1x2})/\sigma_{y1x1}$
7. $\psi_{11} = \sigma_{y2x2}\, (\sigma_{y2y1}\, \sigma_{x2x1} - \sigma_{y2x1}\, \sigma_{y1x2})/\sigma_{x1x2}\, \sigma_{y1x2} = \sigma_{y2x1}\, (\sigma_{y2y1}\, \sigma_{x2x1} - \sigma_{y2x1}\, \sigma_{y1x2})/\sigma_{x1x2}\, \sigma_{y1x1}$
8. $\theta^{\varepsilon}_{11} = (\sigma_{y1}{}^2\, \sigma_{y2x2} - \sigma_{y1x2}\, \sigma_{y2y1})/\sigma_{y2x2} = (\sigma_{y1}{}^2\, \sigma_{y2x1} - \sigma_{y1x1}\, \sigma_{y2y1})/\sigma_{y2x1}$
9. $\theta^{\varepsilon}_{22} = (\sigma_{y2}{}^2\, \sigma_{y1x2} - \sigma_{y2x2}\, \sigma_{y2y1})/\sigma_{y1x2} = (\sigma_{y2}{}^2\, \sigma_{y1x1} - \sigma_{y2x1}\, \sigma_{y2y1})/\sigma_{y1x1}$

The important point here is that there are two ways of computing each parameter, which results because the model is overidentified. The model implies one overidentifying restriction on observed moments, which we can obtain by equating any of the pairs of parameters expressed in terms of moments. For example, set the two equations for λ_{y11} equal:

$\lambda_{y11} = \sigma_{y1x2}/\sigma_{y2x2} = \sigma_{y1x1}/\sigma_{y2x1}$ cross-multiply and get:

$$\sigma_{y1x2}\,\sigma_{y2x1} = \sigma_{y2x2}\,\sigma_{y1x1}$$

Or set the two equations for γ_{11} equal:

$\gamma_{11} = \sigma_{y2x1}/\sigma_{x2x1} = \sigma_{y1x1}\,\sigma_{y2x2}/\sigma_{x2x1}\,\sigma_{x2y1}$ cross-multiply and get:

$$\sigma_{y2x1}\,\sigma_{x2x1}\,\sigma_{x2y1} = \sigma_{x2x1}\,\sigma_{y1x1}\,\sigma_{y2x2}$$

$$\sigma_{y2x1}\,\sigma_{x2y1} = \sigma_{y1x1}\,\sigma_{y2x2}$$

In the factor analysis literature, this overidentifying restriction is called a *"tetrad difference" constraint*. The principles we've discussed for an overidentified recursive model in observables hold here. If the model is correct in the population, this overidentifying restriction will hold exactly in the population, and the two different ways of computing parameters in terms of moments will yield the exact same result. If the restriction does not hold exactly in the population, we can conclude that the model is wrong.

III. ESTIMATION AND TESTING.

In empirical applications, of course, we don't have access to population moments, so we have to use moments drawn from a random sample to estimate parameters. We can specify the sample counterparts to our covariance structure matrices: S = $\Sigma(\hat{\theta})$

$$\underset{4\times 4}{S} = \begin{bmatrix} s_{x_1}^2 & & & \\ s_{x_2x_1} & s_{x_2}^2 & & \\ s_{y_1\,x_1} & s_{y_1\,x_2} & s_{y_1}^2 & \\ s_{y_2x_1} & s_{y_2x_2} & s_{y_2x_1} & s_{y_2}^2 \end{bmatrix}$$

$$\underset{4\times 4}{\Sigma(\hat{\theta})} = \begin{bmatrix} \hat{\lambda}_{x_{11}}^2\hat{\Phi}_{11} + \hat{\theta}_{11}^{\delta} & & & \\ \hat{\lambda}_{x_{11}}\hat{\Phi}_{11} & \hat{\Phi}_{11} + \hat{\theta}_{22}^{\delta} & & \\ \hat{\lambda}_{y_{11}}\hat{\lambda}_{x_{11}}\hat{\gamma}_{11}^2\hat{\Phi}_{11} & \hat{\lambda}_{y_{11}}\hat{\gamma}_{11}^2\hat{\Phi}_{11} & \hat{\lambda}_{y_{11}}^2(\hat{\gamma}_{11}^2\hat{\Phi}_{11} + \hat{\Psi}_{11}) + \hat{\theta}_{11}^{\varepsilon} & \\ \hat{\lambda}_{x_{11}}\hat{\gamma}_{11}^2\hat{\Phi}_{11} & \hat{\lambda}_{y_{11}}\hat{\gamma}_{11}^2\hat{\Phi}_{11} & \hat{\lambda}_{y_{11}}(\hat{\gamma}_{11}^2\hat{\Phi}_{11} + \hat{\Psi}_{11}) & \hat{\gamma}_{11}^2\hat{\Phi}_{11} + \hat{\Psi}_{11} + \hat{\theta}_{22}^{\varepsilon} \end{bmatrix}$$

But when we use the method of moments to estimate parameters from sample covariances, we have two ways of estimating. Which do we use?

1. $\hat{\lambda}_{y_{11}} = s_{y1x2}/s_{y2x2} = s_{y1x1}/s_{y2x1}$

2. $\hat{\lambda}_{x_{11}} = s_{y2x1}/s_{y2x2} = s_{y1x1}/s_{y1x2x11}$

3. $\hat{\gamma}_{11} = s_{y2x1}/s_{x2x1} = s_{y1x1}\,s_{y2x2}/s_{x2x1}\,s_{x2y1}$

4. $\hat{\Phi}_{11} = s_{x2x1}\,s_{y2x2}/s_{y2x1} = s_{x2x1}\,s_{y1x2}/s_{y1x1}$

5. $\hat{\theta}_{11}^{\delta} = (s_{x1}^2\,s_{y2x2} - s_{y2x1}\,s_{x2x1})/s_{y2x2} = (s_{x1}^2\,s_{y1x2} - s_{y1x1}\,s_{x1x1})/s_{y1x2}$

6. $\hat{\theta}_{22}^{\delta} = (s_{x2}^2\,s_{y2x1} - s_{x2x1}\,s_{y2x2})/s_{y2x1} = (s_{x2}^2\,s_{y1x1} - s_{x2x1}\,s_{y1x2})/s_{y1x1}$

7. $\hat{\psi}_{11} = s_{y2x2}\,(s_{y2y1}\,s_{x2x1} - s_{y2x1}\,s_{y1x2})/s_{x1x2}\,s_{y1x2} = s_{y2x1}\,(s_{y2y1}\,s_{x2x1} - s_{y2x1}\,s_{y1x2})/s_{x1x2}\,s_{y1x1}$

8. $\hat{\theta}^{\varepsilon}_{11} = (s_{y1}^{\;2}\,s_{y2x2} - s_{y1x2}\,s_{y2y1})/s_{y2x2} = (s_{y1}^{\;2}\,s_{y2x1} - s_{y1x1}\,s_{y2y1})/s_{y2x1}$

9. $\hat{\theta}^{\varepsilon}_{22} = (s_{y2}^{\;2}\,s_{y1x2} - s_{y2x2}\,s_{y2y1})/s_{y1x2} = (s_{y2}^{\;2}\,s_{y1x1} - s_{y2x1}\,s_{y2y1})/s_{y1x1}$

Let's take $\hat{\lambda}_{y_{11}}$ as an example. Let the two estimates be defined as:

$\hat{\lambda}^{a}_{y_{11}} = s_{y1x2}/s_{y2x2}$ $\qquad \hat{\lambda}^{b}_{y_{11}} = s_{y1x1}/s_{y2x1}$

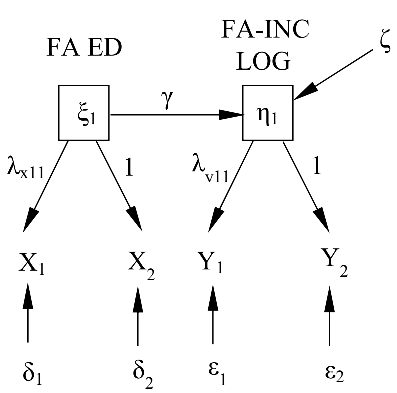
We have two ways of estimating the parameter. Should we choose $\hat{\lambda}^{a}_{y_{11}}$, or $\hat{\lambda}^{b}_{y_{11}}$, or a weighted average of the two $\hat{\lambda}^{c}_{y_{11}} = w_1\,\hat{\lambda}^{a}_{y_{11}} + w_2\hat{\lambda}^{b}_{y_{11}}$, where w_1 and w_2 are weights? They are both *unbiased* (and consistent), and thus, any weighted sum of the two will be unbiased. But they vary in efficiency. Unlike the situation of recursive models in observables, it turns out that the efficient estimator is a weighted sum of the two $\hat{\lambda}^{c}_{y_{11}}$. But given this, how should we determine the weights? Note that in the sixties, some sociological methodologists suggested using an unweighted average of the two estimates $\hat{\lambda}^{c}_{y_{11}} = w_1\,\hat{\lambda}^{a}_{y_{11}} + w_2\hat{\lambda}^{b}_{y_{11}}$, where $w_1 = w_2$), since they knew of no principle to weight them, and they figured using more sample information was better than using less (see Costner). It turns out that the principle of *maximum likelihood (ML)* gives us a set of criteria for determining weights. ML determines weights in such a way as to give optimal asymptotic properties (the behavior of the estimator as N goes to infinity): consistent and asymptotically efficient estimates. This holds for the other parameters as well, such as γ:

$\hat{\gamma}^{a}_{11} = s_{y2x1}/s_{x2x1}$ $\qquad \hat{\gamma}^{b}_{11} = s_{y1x1}\,s_{y2x2}/s_{x2x1}\,s_{x2y1}$

ML would provide a consistent and asymptotically efficient estimator by choosing weights for $\hat{\gamma}^{c}_{11} = m_1\,\hat{\gamma}^{a}_{11} + m_2\hat{\gamma}^{b}_{11}$, where m_1 and m_2 are the weights. It is important to remember that if the model were just-identified—for example, if there were a correlation between ε_1 and ε_2 (which we would parameterize as a covariance in our unstandardized model $\theta^{\varepsilon}_{21}$), then there would be only one way of computing moments in terms of parameters. In that case, the method of moments –simply using sample counterparts to population moments -- will give us optimal finite sample properties: unbiased and efficient estimates. Therefore, ML is needed when models are *overidentified*, which is *almost always* the case with covariance structure models.

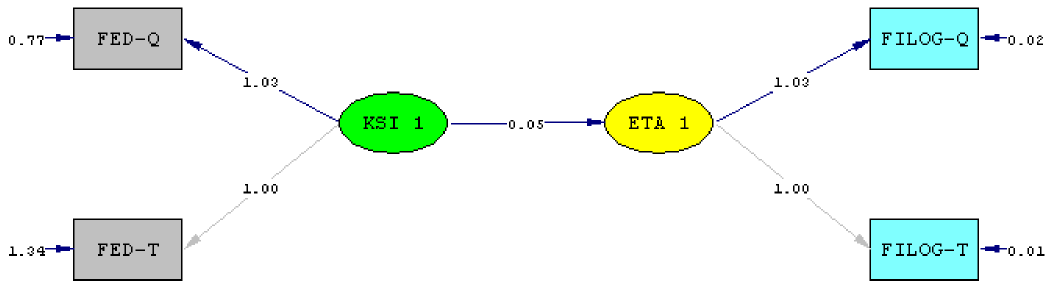
A second question arises: how do we test the model's overidentifying restriction? And how do we get appropriate estimates of standard errors? One way of testing would be to somehow test how close S and $\Sigma(\hat{\theta})$ are to each other, e.g., is $S - \Sigma(\hat{\theta}) = 0$? It turns out that the likelihood ratio test procedure provides a formal way of carrying out such a test. Moreover, in our walking dog model, there are several ways of parameterizing the overidentifying restriction on sample moments. For example, we could specify θ^{δ}_{21} or $\theta^{\varepsilon}_{21}$ or a direct effect from ξ_1 to y_1 or y_2, etc. The likelihood ratio method will allow us to test one of these specific hypotheses about parametric structure. More on this later.

Here is a simple walking dog example estimated in LISREL. First, the path diagram:



Data: Occupational Changes in a Generation
Source: Bielby, Hauser & Featherman (1977) AJS

X_1 = Father's Education Questionnaire
X_2 = Father's Education Telephone
Y_3 = Father's Income Telephone
X_2 = Father's Income Telephone



The following lines were read from file H:\529 New Examples\Example files 1\Walking Dog 1.LS8:

```
!WALKING DOG REGRESSION ON LATENT VARIABLES NO ERR CORR
!Nonblack males OCGR March 1973
DA NI=4 NO=578                              DA(ta) line: I'm inputting four variables (NI=4) with N=578.
SD                                          I'm reading in standard deviations in free format
*
4.19  4.14  0.41  0.39
KM                                          I'm reading in a correlation matrix in free format
*
1.000
.939  1.000
.477   .467  1.000
.486   .478   .913  1.000
LA                                          I'm reading in variable labels in free format
*
FED-Q FED-T FILOG-Q FILOG-T
SE                                          I'm selecting the two income variables first (they
3 4 1 2 /                                   are Ys) followed by two education variables (xs)
MO NX=2 NY=2 NK=1 NE=1 GA=FU,FR BE=ZE PS=DI,FR LX=FU,FI LY=FU,FI TD=SY,FI TE=SY,FI
FR LX 1 1 LY 1 1                            By not freeing these and giving them start values of 1.0, I'm selecting x2
FR TD 1 1 TD 2 2                            to be reference indicator for ξ and y2for  η.
VA 1 LX 2 1 LY 2 1
FR TD 1 1 TD 2 2                            I'm freeing the measurement error variances.
FR TE 1 1 TE 2 2
PD                                          Asking for the path diagram.
OU ME=ML RS EF SC MI                        Using ML, asking for residuals, indirect effects,
                     `                      completely standardized solution, and modification
                                            indices.
```

```
!WALKING DOG REGRESSION ON LATENT VARIABLES NO ERR CORR

                         Number of Input Variables  4
                         Number of Y - Variables    2
                         Number of X - Variables    2
                         Number of ETA - Variables  1
                         Number of KSI - Variables  1
                         Number of Observations   578

!WALKING DOG REGRESSION ON LATENT VARIABLES NO ERR CORR



       Covariance Matrix
                                                           This is S, the sample covariance
            FILOG-Q    FILOG-T      FED-Q      FED-T       matrix
```

```
           --------   --------   --------   --------
 FILOG-Q       0.17
 FILOG-T       0.15       0.15
   FED-Q       0.82       0.79      17.56
   FED-T       0.79       0.77      16.29      17.14
```

```
!WALKING DOG REGRESSION ON LATENT VARIABLES NO ERR CORR

Parameter Specifications

        LAMBDA-Y

              ETA 1
           --------
 FILOG-Q          1
 FILOG-T          0          Reference indicator for ETA is Y2


        LAMBDA-X

              KSI 1
           --------
   FED-Q          2
   FED-T          0          Reference indicator for KSI is X2

        GAMMA

              KSI 1
           --------
   ETA 1          3

        PHI

              KSI 1
           --------
                  4

        PSI

              ETA 1
           --------
                  5

        THETA-EPS

            FILOG-Q    FILOG-T
           --------   --------
                  6          7

        THETA-DELTA

              FED-Q      FED-T
           --------   --------
                  8          9


!WALKING DOG REGRESSION ON LATENT VARIABLES NO ERR CORR

Number of Iterations =  5

LISREL Estimates (Maximum Likelihood)

        LAMBDA-Y

              ETA 1
           --------
 FILOG-Q       1.03
```

```
           (0.04)
            27.90
 FILOG-T     1.00

       LAMBDA-X

            KSI 1
         --------
   FED-Q     1.03
           (0.03)
            33.53
   FED-T     1.00

       GAMMA

            KSI 1
         --------
   ETA 1     0.05                  This is returns to education in log dollars
           (0.00)
            13.28

       Covariance Matrix of ETA and KSI        Covariance matrix of latent variables

            ETA 1      KSI 1
         --------   --------
   ETA 1     0.14
   KSI 1     0.77      15.80

       PHI

            KSI 1
         --------
            15.80                  Variance of latent variable, KSI
           (1.09)
            14.55

       PSI

            ETA 1
         --------
             0.10                  Disturbance variance, PSI
           (0.01)
            13.71

       Squared Multiple Correlations for Structural Equations

            ETA 1
         --------
             0.27                  R-Squared for the income equation

       THETA-EPS

          FILOG-Q    FILOG-T
         --------   --------
             0.02       0.01       Measurement error variances for income measures
           (0.00)     (0.00)
             3.69       2.33

       Squared Multiple Correlations for Y - Variables

          FILOG-Q    FILOG-T
         --------   --------
             0.89       0.93       R-Squareds for income measures—reliability
                                   coefficients (squared standardized LAMBDAs

       THETA-DELTA

            FED-Q      FED-T
         --------   --------
```

```
        0.77        1.34
      (0.44)      (0.42)
        1.75        3.21
```

Measurement error variances for education measures

```
 Squared Multiple Correlations for X - Variables

       FED-Q       FED-T
    --------    --------
        0.96        0.92
```

R-Squareds for education measures—reliability coefficients (squared standardized LAMBDAs

```
                     Goodness of Fit Statistics

                        Degrees of Freedom = 1
        Minimum Fit Function Chi-Square = 0.13 (P = 0.72)
 Normal Theory Weighted Least Squares Chi-Square = 0.13 (P = 0.72)
        Estimated Non-centrality Parameter (NCP) = 0.0
     90 Percent Confidence Interval for NCP = (0.0 ; 3.60)

                Minimum Fit Function Value = 0.00022
          Population Discrepancy Function Value (F0) = 0.0
       90 Percent Confidence Interval for F0 = (0.0 ; 0.0062)
       Root Mean Square Error of Approximation (RMSEA) = 0.0
     90 Percent Confidence Interval for RMSEA = (0.0 ; 0.079)
       P-Value for Test of Close Fit (RMSEA < 0.05) = 0.86

          Expected Cross-Validation Index (ECVI) = 0.033
     90 Percent Confidence Interval for ECVI = (0.033 ; 0.039)
                  ECVI for Saturated Model = 0.035
                 ECVI for Independence Model = 2.64

 Chi-Square for Independence Model with 6 Degrees of Freedom = 1514.96
                      Independence AIC = 1522.96
                          Model AIC = 18.13
                        Saturated AIC = 20.00
                     Independence CAIC = 1544.40
                          Model CAIC = 66.37
                        Saturated CAIC = 73.60

                    Normed Fit Index (NFI) = 1.00
                  Non-Normed Fit Index (NNFI) = 1.00
               Parsimony Normed Fit Index (PNFI) = 0.17
                  Comparative Fit Index (CFI) = 1.00
                  Incremental Fit Index (IFI) = 1.00
                   Relative Fit Index (RFI) = 1.00

                      Critical N (CN) = 29550.26


              Root Mean Square Residual (RMR) = 0.00062
                      Standardized RMR = 0.00037
                  Goodness of Fit Index (GFI) = 1.00
             Adjusted Goodness of Fit Index (AGFI) = 1.00
            Parsimony Goodness of Fit Index (PGFI) = 0.100
```

There is one tetrad-difference overidentifying restriction

Likelihood Ratio test

Asymptotically equivalent test

Great fit!

This is a useful fit statistic. Rule of thumb < .05 is good

```
!WALKING DOG REGRESSION ON LATENT VARIABLES NO ERR CORR

        Fitted Covariance Matrix

             FILOG-Q    FILOG-T      FED-Q      FED-T
            --------   --------   --------   --------
 FILOG-Q        0.17
 FILOG-T        0.15       0.15
   FED-Q        0.82       0.79      17.56
   FED-T        0.79       0.77      16.29      17.14

        Fitted Residuals
```

This is $\Sigma(\hat{\theta})$, the covariance matrix implied by the parameter estimates.

```
            FILOG-Q    FILOG-T      FED-Q      FED-T
              --------   --------   --------   --------
  FILOG-Q       0.00
 FILOG-T       0.00       0.00
   FED-Q       0.00       0.00       0.00
   FED-T       0.00       0.00       0.00       0.00
```

This is $S - \Sigma(\widehat{\theta})$, the residual matrix that subtracts the moments implied by the model from the sample moments.

```
Summary Statistics for Fitted Residuals

Smallest Fitted Residual =    0.00
  Median Fitted Residual =    0.00
 Largest Fitted Residual =    0.00

Stemleaf Plot

- 1|5
- 1|
- 0|5
- 0|000000
  0|
  0|89

        Standardized Residuals

            FILOG-Q    FILOG-T      FED-Q      FED-T
           --------   --------   --------   --------
 FILOG-Q        - -
 FILOG-T        - -        - -
   FED-Q       0.36      -0.36        - -
   FED-T      -0.36       0.36        - -        - -

Summary Statistics for Standardized Residuals

Smallest Standardized Residual =   -0.36
  Median Standardized Residual =    0.00
 Largest Standardized Residual =    0.36


!WALKING DOG REGRESSION ON LATENT VARIABLES NO ERR CORR

Modification Indices and Expected Change

No Non-Zero Modification Indices for LAMBDA-Y

No Non-Zero Modification Indices for LAMBDA-X

No Non-Zero Modification Indices for GAMMA

No Non-Zero Modification Indices for PHI

No Non-Zero Modification Indices for PSI


        Modification Indices for THETA-DELTA-EPS
```

This are univariate LaGrangian multiplier test Statistics, distributed as a Z-statistic.

```
            FILOG-Q    FILOG-T
           --------   --------
   FED-Q       0.13       0.13
   FED-T       0.13       0.13

        Expected Change for THETA-DELTA-EPS
```

This is the amount a fixed parameter is Estimated to chang4 if it were freed.

```
            FILOG-Q    FILOG-T
           --------   --------
   FED-Q       0.00       0.00
   FED-T       0.00       0.00

        Completely Standardized Expected Change for THETA-DELTA-EPS
```

This standardizes expected change, putting

```
           FILOG-Q    FILOG-T
          --------   --------
   FED-Q      0.00       0.00
   FED-T      0.00       0.00
```

them in standard deviation units.

```
 Maximum Modification Index is    0.13 for Element ( 1, 1) of THETA DELTA-EPSILON

 !WALKING DOG REGRESSION ON LATENT VARIABLES NO ERR CORR

 Standardized Solution
```

This only partially standardizes coefficients by the latent variables, η and ξ

```
        LAMBDA-Y

               ETA 1
            --------
 FILOG-Q       0.39
 FILOG-T       0.38
```

This standardizes λy by σ_η only.

$\lambda_{11}\,\sigma_\eta$

$\lambda_{21}\,\sigma_\eta$

```
        LAMBDA-X

               KSI 1
            --------
   FED-Q       4.10
   FED-T       3.98
```

This standardizes λx by σξ only.

$\lambda_{11}\,\sigma_\xi$

$\lambda_{21}\,\sigma_\xi$

```
        GAMMA

               KSI 1
            --------
   ETA 1       0.52
```

This is fully standardized by η and ξ

$\gamma_{11}\frac{\sigma_\xi}{\sigma_\eta}$

```
        Correlation Matrix of ETA and KSI

               ETA 1      KSI 1
            --------   --------
   ETA 1       1.00
   KSI 1       0.52       1.00

        PSI

               ETA 1
            --------
               0.73
```

` **This is $P^2_{\eta\zeta}$, therefore the square root of this is the standardized disturbance path coefficient ($P_{\eta\zeta}$)= .85**

```
        Regression Matrix ETA on KSI (Standardized)

               KSI 1
            --------
   ETA 1       0.52
```

This is the standardized regression of η on ξ (structural form = reduced form)

```
 !WALKING DOG REGRESSION ON LATENT VARIABLES NO ERR CORR

 Completely Standardized Solution

        LAMBDA-Y

               ETA 1
            --------
 FILOG-Q       0.95
 FILOG-T       0.97
```

$\lambda_{y11}\frac{\sigma_\eta}{\sigma_{y1}}$

$\lambda_{y21}\frac{\sigma_\eta}{\sigma_{y2}}$

```
        LAMBDA-X

               KSI 1
            --------
   FED-Q       0.98
```

$\lambda_{x11}\frac{\sigma_\xi}{\sigma_{x1}}$

```
    FED-T        0.96
```
$\lambda_{x21}\frac{\sigma_\xi}{\sigma_{x2}}$

```
         GAMMA

               KSI 1
            --------
    ETA 1        0.52
```
$\gamma_{11}\frac{\sigma_\xi}{\sigma_\eta}$

```
         Correlation Matrix of ETA and KSI

               ETA 1      KSI 1
            --------   --------
    ETA 1        1.00
    KSI 1        0.52       1.00

         PSI

               ETA 1
            --------
                 0.73
```
This is $P^2_{\eta\zeta}$, therefore the square root of this is the standardized disturbance path coefficient ($P_{\eta\zeta}$)= .85

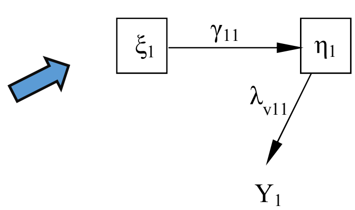
```
         THETA-EPS

             FILOG-Q    FILOG-T
            --------   --------
                 0.11       0.07
```
This is $P^2_{y\eta}$, therefore the square root of this is the standardized disturbance path coefficient ($P_{y\eta}$)= .33 and .26

```
         THETA-DELTA

               FED-Q      FED-T
            --------   --------
                 0.04       0.08
```
This is $P^2_{x\eta}$, therefore the square root of this is the standardized disturbance path coefficient ($P_{x\eta}$)= .20 and .28

```
         Regression Matrix ETA on KSI (Standardized)

               KSI 1
            --------
    ETA 1        0.52
```
This is the standardized reduced form regression of η on ξ (structural form = reduced form)

```
 !WALKING DOG REGRESSION ON LATENT VARIABLES NO ERR CORR

 Total and Indirect Effects

         Total Effects of KSI on Y

               KSI 1
            --------
  FILOG-Q        0.05
               (0.00)
                12.95

  FILOG-T        0.05
               (0.00)
                13.28
```
This is $\gamma_{11}\lambda_{y11}$

This is $\gamma_{11}\lambda_{y21}$




```
 !WALKING DOG REGRESSION ON LATENT VARIABLES NO ERR CORR

 Standardized Total and Indirect Effects

         Standardized Total Effects of KSI on Y

               KSI 1
```

```
           --------
FILOG-Q        0.20
FILOG-T        0.19
```

Standardizes the total effect only by ξ and η -- $\boldsymbol{\gamma_{11}(\sigma_\xi/\sigma_\eta)\,\lambda_{y11}(\sigma_\eta)}$ (FILOG-Q)

$\boldsymbol{\gamma_{11}(\sigma_\xi/\sigma_\eta)\,\lambda_{y21}(\sigma_\eta)}$ (FILOG-T)

```
       Completely Standardized Total Effects of KSI on Y

              KSI 1
           --------
FILOG-Q        0.49
FILOG-T        0.50
```

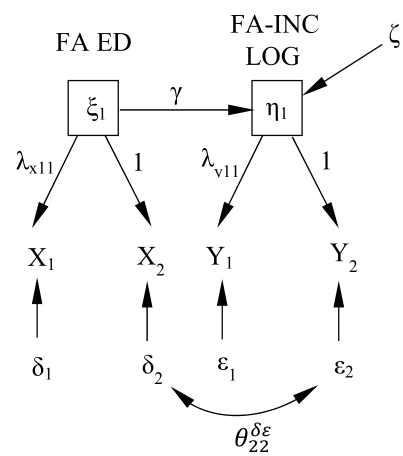
Completely standardized: $\boldsymbol{\gamma_{11}\left(\frac{\sigma_\xi}{\sigma_\eta}\right)\lambda_{y11}\left(\frac{\sigma_\eta}{\sigma_{y1}}\right)}$ (FILOG-Q)

$\boldsymbol{\gamma_{11}\left(\frac{\sigma_\xi}{\sigma_\eta}\right)\lambda_{y21}\left(\frac{\sigma_\eta}{\sigma_{y1}}\right)}$ (FILOG-T)

```
                        Time used:    0.016 Seconds
```

Although this model fit the data extremely well, let's try estimating a model that includes a measurement error correlation between the two telephone measures:

For error covariances between measurement errors of X and Y, LISREL refers to the matrix as TH.




```
The following lines were read from file H:\529 New Examples\Example files 1\Walking Dog 2.LS8:

 !WALKING DOG REGRESSION ON LATENT VARIABLES ONE ERR CORR
 !Nonblack males OCGR March 1973
 DA NI=4 NO=578
 SD
 *
 4.19  4.14  0.41  0.39
 KM
 *
 1.000
 .939  1.000
 .477   .467  1.000
 .486   .478   .913  1.000
 LA
 *
 FED-Q FED-T FILOG-Q FILOG-T
 SE
 3 4 1 2 /
 MO NX=2 NY=2 NK=1 NE=1 GA=FU,FR BE=ZE PS=DI,FR LX=FU,FI LY=FU,FI TD=SY,FI TE=SY,FI
 FR LX 1 1 LY 1 1
 VA 1 LX 2 1 LY 2 1
 FR TD 1 1 TD 2 2
 FR TE 1 1 TE 2 2
 FR TH 2 2                Here, we're freeing the error correlation between Delta 2 and Epsilon 2
 PD
 OU ME=ML RS EF SC MI

 !WALKING DOG REGRESSION ON LATENT VARIABLES ONE ERR CORR
```

```
                           Number of Input Variables  4
                           Number of Y - Variables    2
                           Number of X - Variables    2
                           Number of ETA - Variables  1
                           Number of KSI - Variables  1
                           Number of Observations   578

 !WALKING DOG REGRESSION ON LATENT VARIABLES ONE ERR CORR

        Covariance Matrix

            FILOG-Q    FILOG-T      FED-Q      FED-T
           --------   --------   --------   --------
  FILOG-Q      0.17
  FILOG-T      0.15       0.15
    FED-Q      0.82       0.79      17.56
    FED-T      0.79       0.77      16.29      17.14


 !WALKING DOG REGRESSION ON LATENT VARIABLES ONE ERR CORR

 Parameter Specifications

        LAMBDA-Y

              ETA 1
           --------
  FILOG-Q         1
  FILOG-T         0

        LAMBDA-X

              KSI 1
           --------
    FED-Q         2
    FED-T         0

        GAMMA

              KSI 1
           --------
    ETA 1         3

        PHI

              KSI 1
           --------
                  4

        PSI

              ETA 1
           --------
                  5

        THETA-EPS

            FILOG-Q    FILOG-T
           --------   --------
                  6          7

        THETA-DELTA-EPS

            FILOG-Q    FILOG-T
           --------   --------
    FED-Q         0          0
    FED-T         0          9
```

**We're adding another parameter to be estimated, so
Now we're estimating 10 parameters. With 10**

moments, this model is now just-identified.

```
      THETA-DELTA

             FED-Q      FED-T
          --------   --------
                 8         10


!WALKING DOG REGRESSION ON LATENT VARIABLES ONE ERR CORR

Number of Iterations =  5

LISREL Estimates (Maximum Likelihood)

        LAMBDA-Y

              ETA 1
           --------
 FILOG-Q       1.03
             (0.04)
              27.70
 FILOG-T       1.00

        LAMBDA-X

              KSI 1
           --------
   FED-Q       1.03
             (0.03)
              32.40
   FED-T       1.00

        GAMMA

              KSI 1
           --------
   ETA 1       0.05
             (0.00)
              13.28

        Covariance Matrix of ETA and KSI

              ETA 1      KSI 1
           --------   --------
   ETA 1       0.14
 KSI 1       0.77      15.76

        PHI

              KSI 1
           --------
              15.76
             (1.09)
              14.42

        PSI

              ETA 1
           --------
               0.10
             (0.01)
              13.71

        Squared Multiple Correlations for Structural Equations

              ETA 1
           --------
```

```
              0.26

         THETA-EPS

             FILOG-Q    FILOG-T
            --------   --------
                0.02       0.01
              (0.00)     (0.00)
                3.61       2.36

         Squared Multiple Correlations for Y - Variables

             FILOG-Q    FILOG-T
            --------   --------
                0.90       0.93

         THETA-DELTA-EPS

             FILOG-Q    FILOG-T
            --------   --------
    FED-Q        - -        - -
    FED-T        - -       0.00
                         (0.01)
                           0.36
```

Here's our new parameter estimate: it is tiny and not signficant. This is to be expected because the more restrictive model fit very well. This parameter adds little to the model.

```
         THETA-DELTA

               FED-Q      FED-T
            --------   --------
                0.72       1.38
              (0.46)     (0.43)
                1.57       3.18

         Squared Multiple Correlations for X - Variables

               FED-Q      FED-T
            --------   --------
                0.96       0.92

                           Goodness of Fit Statistics
```

Zero degrees of freedom because it is just idenfied (10 moments minus 10 parameters = 0 degrees of freedom)

```
                              Degrees of Freedom = 0
                Minimum Fit Function Chi-Square = 0.00 (P = 1.00)
        Normal Theory Weighted Least Squares Chi-Square = 0.00 (P = 1.00)
                 The Model is Saturated, the Fit is Perfect !
```