LECTURE 1: MOMENTS AND PARAMETERS IN A BIVARIATE LINEAR STRUCTURAL MODEL

I PRELIMINARIES: STRUCTURAL EQUATION MODELS.
II. POPULATION MOMENTS AND PARAMETERS.
III. CORRELATIONS AND STANDARDIZED COEFFICIENTS.
IV. ESTIMATION AND TESTING.

I. PRELIMINARIES: STRUCTURAL EQUATION MODELS

This course is about the use of structural equation models for examining social phenomena. There are four basic principles involved:

1. Assume we can characterize a social phenomenon by a set of random variables.
    - A random variable is one whose values or outcomes are governed by a probability distribution.
    - A *set* of random variables have joint outcomes, which are governed by a *joint* probability distribution.
    - We characterize this joint distribution by observable moments (means, variances, and covariances).

2. We assume that some *underlying causal structure or model* generates observed moments (covariances)
    - The underlying structural model represents our (parsimonious) *theory* of the phenomenon.
    - The model generates the data (joint distribution of random variables), which we characterize by observed moments.
    - The parameters of the model are assumed to be *invariant*: they are somewhat stable over time and only change when there is true social change.
    - The causal structure can be expressed in linear structural equations and path diagrams.
    - Hence the terms, "covariance structure analysis," "path analysis" and "structural equation models."

3. Given observed population moments, we could compute population parameters of our structural model (assuming the model is identified).

4. However, we lack access to *population moments*, and therefore, we cannot compute population parameters.
    - Instead, we have access only to *sample moments*.
    - We can use them to *estimate* population parameters.
    - We want estimators with desirable properties (unbiased and efficient).
    - We will need to test hypotheses about parameter estimates to rule out chance (sampling variability).

We begin by assuming a phenomenon of interest can be characterized by a set of random variables, whose joint distribution can be described by their observed moments. But first we need to review the algebra of expectations of random variables. The expected value of a random variable is a weighted sum of the values, each weighted by the probability of its occurence. The expected value of a random variable is the population mean.

Aside: Review of algebra of expectations rules (see Winkler & Hayes 1975 pp. 141-47).

1. $E(c) = c$ where c is a constant.
2. $E(cX) = cE(X)$ where X is a random variable with expectation E(X).
3. $E(X + c) = c + E(X)$.
4. $E(X + Y) = E(X) + E(Y)$ where Y is a random variable with expectation E(Y).
5. $E(X + Y + ... + Z) = E(X) + E(Y) + ... + E(Z)$.

Technically, moments refer to the expectations of different powers of random variables. We're concerned with two kinds of moments—moments about the origin and moments about the mean. Moments about the origin describe the central tendency of the distribution; we're interested in the first moment about the origin, the mean $\mu = E(X)$.

Moments about the mean describe the shape of the distribution according to $\mu_r = E[X - E(X)]^r$. Note that this gives the expectations of powers of the random variable after subtracting the mean. The first moment about the mean is zero—$\mu_1 = E[X - E(X)] = 0$—t he second refers to the variance (covariance), the third to a measure of skewness (symmetric or not), and the fourth to kurtosis (how peaked is the distribution):

1. First moment about the origin:

   $\mu = E(X)$ Mean

2. Second moment about the mean:

   $\sigma^2 = E[X - E(X)]^2$ Variance

   $\sigma_{xy} = E[X - E(X)]\, E[Y - E(Y)]$ Covariance

3. Higher-order moments about the mean:

   $\mu_3 = E[X - E(X)]^3$ Related to Skewness

   $\mu_4 = E[X - E(X)]^4$ Related to Kurtosis

In this course we will be concerned mostly with the *second moments*, the variances and covariances. We'll leave the means unanalyzed, and assume that variables are *normally-distributed*, which implies that we can safely ignore higher-order moments. Later in the semester, we'll have reason to come back to the mean (when we look at the difference between means across groups) and higher-order moments (when we consider what happens when multivariate normality fails to hold).

II. POPULATION MOMENTS AND POPULATION STRUCTURAL PARAMETERS.

Let's begin with the simplest possible example: just two variables, Y and X. As an example, assume we have census data on all individuals in the U.S., and Y is income measured in dollars, while X is years of education. For simplicity, let's assume the means of the two variables is zero: $E(X) = 0$; $E(Y) = 0$. We can accomplish this by deviating each variable from its mean: $x_i' = x_i - \mu_x$ and $y_i' = y_i - \mu_y$. In general, we'll work with x' and y', but omit the prime. We begin by characterizing these two variables by computing their second moments:

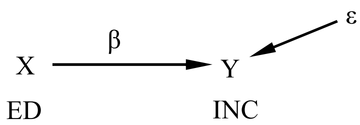
$\sigma_x^2 = E[X - E(X)]^2$ $\sigma_{xy} = E\{[X - E(X)][Y - E(Y)]\} = E(XY)$ because $E(X) = E(Y) = 0$

$\sigma_y^2 = E[Y - E(Y)]^2$

Note that these are *population* moments, not sample moments; so they are quantities of the population we're interested in—the U.S.—and not estimates. Let's arrange these in a convenient matrix of population covariances of observed variables, call it Σ:

$$\underset{2 \times 2}{\Sigma} = \begin{bmatrix} \sigma_x^2 & \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix}$$

Let's assume that Σ, the population moment matrix of our variables, provides an adequate description of the variables X and Y. Our goal is to specify a structural model that underlies these population moments—that is, that *generates* the moments. That model should be based on *scientific theory*. Let's say our theory is some variant on human capital theory, and assumes that income is a linear function of years of education:



$y_i = \beta x_i + \varepsilon_i$ where i = 1, . . ,

Remember that this is a representation of the *population* model of X and Y. (The subscripts, i, represent individuals. From now on, I will omit the subscripts for convenience, but remember that they are implicit.) Note that in this model Y, X, and ε are *random variables*, but β is a *constant*. Let's say Y is income measured in dollars and X education measured in years. Following a simple human capital argument, we assume that income is determined by education plus an error or disturbance term, ε. Variation in Y, income, is generated by two components: βX is the deterministic component; and ε is the stochastic component. ε is a random variable called the error or disturbance because it disturbs what would otherwise be a stable relationship. The disturbancecaptures the probabilistic (random, chance, stochastic) source of variation in Y. Think of it as the sum of all other sources of variation in Y—excluded minor variables, which taken together act like a *random shock* on each value of Y. At least we hope (and pray) that this is correct.

Let's make the usual regression assumptions of this model: (1) $E(\varepsilon_i X_i) = 0$; (2) $\varepsilon_i \sim N(0,\sigma^2)$; and (3) $E(\varepsilon_i \varepsilon_j) = 0$. The first assumption implies that the model is properly specified, and we haven't left out any important variables. It is critical: if wrong, we'll get wrong (biased) results. The second assumption implies that ε is zero on average, which makes E(Y) = βE(X), and homoscedasticity (constant error variance across observations). This means that each *observation*, *i*, is assumed to be drawn from a probability distribution of possible observations. And each observation is drawn from a probability distribution that has the same mean (0) and same variance σ^2 (as opposed to σ_i^2, which would allow the error variance to differ across observations). The third assumption is that the errors are not serially correlated. We also assume that the model is properly and completely specified, effects are linear, and variables are perfectly-measured.

There are three population parameters in this model: β, σ_x^2, and σ_ε^2. These three parameters govern the joint distribution between X and Y. They tell us everything we need to know about our model, and thus our two variables.

Given this, we assume that the *observable* population moments were generated by the population model (which we *cannot observe*). We can use a little algebra to compute the values of population moments from the population parameters. This entails using the algebra of expectations on our original structural equation model:

$Y = \beta X + \varepsilon$

Now, we want to know how this model generated our observable moments (σ_x^2, σ_{xy}, σ_y^2). Multiply our equation through by a variable, then take the expected value:

X is exogenous (predetermined and unanalyzed) in the model; so we treat its variance as both a population moment and parameter:

$E(X^2) = E(X^2) =$

1. $\boldsymbol{\sigma_x^2 = \sigma_x^2}$

To compute the covariance of X and Y, we multiply our equation by X, then take expectations:

$XY = X(\beta X + \varepsilon) = \beta X^2 + X\varepsilon$

$E(XY) = \beta E(X^2) + E(X\varepsilon) = \beta\, \sigma_x^2$

2. $\boldsymbol{\sigma_{xy} = \beta\, \sigma_x^2}$

And we do the same for the variance of Y:

$Y^2 = (\beta X + \varepsilon)(\beta X + \varepsilon) = \beta^2 X^2 + 2\beta X\varepsilon + \varepsilon^2)$

$E(Y^2) = \beta^2 E(X^2) + 2\beta E(X\varepsilon) + E(\varepsilon^2) = \beta^2 \sigma_x^2 + \sigma_\varepsilon^2$

3. $\boldsymbol{\sigma_y^2 = \beta^2 \sigma_x^2 + \sigma_\varepsilon^2}$

Let's arrange these population parameters into a matrix, call it $\Sigma(\theta)$, that corresponds to our population moment matrix (Σ).

$$\underset{2\times 2}{\underset{\Sigma}{\begin{bmatrix} \sigma_x^2 & \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix}}} = \underset{2\times 2}{\underset{\Sigma(\theta)}{\begin{bmatrix} \sigma_x^2 & \\ \beta\sigma_x^2 & \beta^2\sigma_x^2 + \sigma_\varepsilon^2 \end{bmatrix}}}$$

The fundamental theorem of covariance structure analysis is: $\Sigma = \Sigma(\theta)$. That is, the covariance (moment) matrix in the population is a function of an underlying structure—a "covariance structure" or "structural equation model" in the population. Again, the structure of $\Sigma(\theta)$ reflects our social science *theory* about the phenomenon. Once we specify our model based on theory, we assume that the population model generated the observable moments, so if we knew the values of the parameters, we could compute the moments using $\Sigma(\theta)$. Here's an example. Let's play *God* and assume we're omniscient, capable of "seeing" (without estimating from sample data) population parameters of our simple model:

$\beta = 0.8$ \$1,000/yr, $\sigma_x = 3$yrs, and $\sigma_\varepsilon = 3 \times .\$1,000$

Return to education is \$800 per year, but we rescale this variable into the metric of one thousand dollars..

Our simple population model becomes:

$Y = 0.8\, X + \varepsilon$.

Using the above $\Sigma(\theta)$, we can compute the population moments implied by this model:

$$\underset{\Sigma(\theta)_a}{\begin{bmatrix} 9\,yrs^2 & \\ (0.8\,\$1{,}000/yr)(9yrs^2) & (0.8^2\,\$1{,}000^2/yrs^2)9\,yrs^2 + 9\,1{,}000\$^2 \end{bmatrix}} = \underset{\Sigma_a}{\begin{bmatrix} 9\,yrs^2 & \\ (7.2\,\$1{,}000/yrs) & 14.76\ \$1{,}000^2 \end{bmatrix}}$$

Of course, one *never* knows the values of parameters specified in $\Sigma(\theta)$. But if we knew Σ, we could use the formula of $\Sigma(\theta)$ to compute the values of population parameters. (Note that we rarely observe the population moments directly, but instead must rely on their sample counterparts as estimates.) In our simple model, there are three population parameters, σ_x^2, β, and σ_ε^2. We also have three equations given above. Three equations in three unknowns will allow us to solve the unknowns.

1. $\boldsymbol{\sigma_x^2 = \sigma_x^2}$(from 1 above)

$\sigma_{xy} = \beta\ \sigma_x^2$(from 2 above)

2. $\boldsymbol{\beta = \sigma_{xy}/\sigma_x^2}$

$\sigma_y^2 = \beta^2\ \sigma_x^2 + \sigma_\varepsilon^2$(from 3 above)$\sigma_\varepsilon^2 = \sigma_y^2 - \beta^2\ \sigma_x^2 = \sigma_y^2 - [(\sigma_{xy})^2/\sigma_x^2)^2]\ \sigma_x^2$ (given 2)

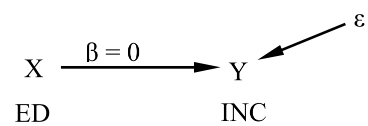
3. $\boldsymbol{\sigma_\varepsilon^2 = \sigma_y^2 - (\sigma_{xy}^2/\sigma_x^2)}$

This is the situation we would normally find ourselves in. We can characterize our variables with observable population moments; then we specify a structural model we assume generated those moments. We can then compute the values of all parameters using these equations. (Note that we are still in the population –no need to estimate anything or worry about samples.) Let's compute the parameters from moments in our example.

1. $\sigma_x^2 = 9$ years
2. $\beta = \sigma_{xy}/\sigma_x^2 = 7.2/9 = 0.8$ \$1,000/yr
3. $\sigma_\varepsilon^2 = \sigma_y^2 - (\sigma_{xy}^2/\sigma_x^2) = 14.76 - (7.2^2/9) = 9$ $\$1,000^2$

Note that these computations assume that the model proposed is correct -- that is, $\Sigma = \Sigma(\theta)$. If the model is wrong, the equivalence won't hold. For example, suppose the "true" population model is exactly as above, but with one exception, $\beta = 0$. (Aside: What does this difference imply substantively?) If this is the true model, then we would necessarily observe the following population moments:

$$\underset{\Sigma(\theta)_b}{\begin{bmatrix} \sigma_x^2 & \\ 0\sigma_x^2 & 0^2\sigma_x^2 + \sigma_\varepsilon^2 \end{bmatrix}} = \underset{\Sigma_b}{\begin{bmatrix} 9\ yrs^2 & \\ 0\ \$\ yrs & 14.76\ \ \$1{,}000^2 \end{bmatrix}}$$



So, if we observe Σ_b, we know that $\Sigma(\theta)_a$ is wrong and $\Sigma(\theta)_b$is correct. But if we observe Σ_a, then we know that $\Sigma(\theta)_a$ is correct and $\Sigma(\theta)_b$ is wrong. Since we're in the population, truth is easy to determine. We're not relying on sample information, so we have no need for estimation or statistical inference.

What if we *had not* assumed that the population mean of X (μ_x) and Y (μ_y) were zero? We could model the *means* of our variables as well. In our simple example, there are two additional moments, μ_x and μ_y, and two additional parameters, μ_x and α:

$Y = \alpha + \beta X + \varepsilon$ (Neither Y nor X are deviated from their means)

$E(Y) = \alpha + \beta\ E(X) + E(\varepsilon)$ Since $E(Y) = \mu_y$, $E(X) = \mu_x$, and $E(\varepsilon) = 0$,

$\mu_y = \alpha + \beta\ \mu_x$ where α is a constant representing the Y-intercept.

Again, we can express moments in terms of parameters

$\mu_x = \mu_x$ or in vector form: $\underset{1\times1}{\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}} = \underset{1\times1}{\begin{bmatrix} \mu_1 \\ \alpha + \beta\mu_1 \end{bmatrix}}$

$\mu_y = \alpha + \beta\ \mu_x$

and parameters in terms of moments:

4. $\boldsymbol{\mu_x = \mu_x}$

$\mu_y = \alpha + \beta\ \mu_x \alpha = \mu_y - \beta\ \mu_x$

5. $\alpha = \mu_y - (\mu_x \sigma_{xy}/\sigma_x^2)$

Let's be *God* again and know the population parameters α = \$4,000 and μ_x = 12 years. Because our model generates the means of X and Y, we can compute them:

μ_x = 12 yrs

$\mu_y = \alpha + \beta\, \mu_x$ = 0.4 \$1,000 + (0.8 \$1,000/yr)(12yrs) = 10.0 \$1,000

Or, if we began with the means of Y and X—μ_y = 10 \$1,000 and μ_x = 12 years—we could then solve for α:

$\alpha = \mu_y - (\mu_x\, \sigma_{xy}/\sigma_x^2)$ = 10.0 \$1,000 - (12yrs)(7.2 \$1,000 yrs/9yrs^2 = 0.4 \$1,000

Exercise: Give a substantive interpretation for α and β. How might this model be untenable in the real world?

III. CORRELATIONS AND STANDARDIZED COEFFICIENTS.

Thus far, we have been modeling covariance structures, developing models in terms of the original metrics of random variables, such as dollars and years of schooling. This can also be done analyzing correlation matrices and working with standardized coefficients. Historically, sociologists' use of path analysis focused on standardized coefficients (sometimes with disastrous results) which simplifies certain algebraic manipulations. Recall the formula for the population correlation:

$\rho = \sigma_{xy}/(\sigma_x \sigma_y)$

Recall that this is also equal to the covariance of X and Y after each has been standardized ($\mu = 0$; $\sigma^2 = 1$). Assume we're in deviation scores, E(X) = E(Y) = 0. We can standardize X and Y:

$Z_x = X/\sigma_x$ and $Z_y = Y/\sigma_y$

$E\{[Z_x - E(Z_x)][Z_y - E(Z_y)]\} = E(Z_x Z_y) = E[(X/\sigma_x)(Y/\sigma_y)] = E(XY)/\sigma_x\sigma_y = \sigma_{xy}/(\sigma_x\sigma_y) = \rho$

Then, our population moment matrix becomes

$$\underset{2\times 2}{\Sigma_\rho} = \begin{bmatrix} 1 & \\ \rho_{xy} & 1 \end{bmatrix}$$

Again, we assume that some underlying structure generated this population correlation matrix, but now the model must be standardized. Again, assume we've deviated from means E(X) = E(Y) = 0 and now standardize our bivariate regression model:

$Y = \beta X + \varepsilon$ divide through by σ_y

$Y/\sigma_y = \beta/\sigma_y X + \varepsilon/\sigma_y$ multiply second term by 1 (σ_x/σ_x), third term by 1 $(\sigma_\varepsilon/\sigma_\varepsilon)$

$Y/\sigma_y = (\beta/\sigma_y)(\sigma_x/\sigma_x) X + (\sigma_\varepsilon/\sigma_\varepsilon)\, \varepsilon/\sigma_y$ rearrange terms

$$\underbrace{Y/\sigma_y}_{Z_y} = \underbrace{(\beta\, \sigma_x/\sigma_y)}_{P_{yx}} \underbrace{(X/\sigma_x)}_{Z_x} + \underbrace{(\sigma_\varepsilon/\sigma_y)}_{P_{y\varepsilon}} \underbrace{(\varepsilon/\sigma_\varepsilon)}_{Z_\varepsilon}$$

Therefore, the standardized form of our model is:

$Z_y = P_{yx} Z_x + P_{y\varepsilon} Z_\varepsilon$

where $Z_y = Y/\sigma_y$, $Z_x = X/\sigma_x$, $Z_\varepsilon = \varepsilon/\sigma_x$, $P_{yx} = \beta\, \sigma_x/\sigma_y$, and $P_{y\varepsilon} = \sigma_\varepsilon/\sigma_y$

(Note that Duncan (1975) forgets to replace X and Y with Z_x and Z_y.) Just as before, we can express moments in terms of parameters:

Begin with $Z_y = P_{yx} Z_x + P_{y\varepsilon} Z_\varepsilon$, multiply through and take expectations

1. $E(Z_x^2) = E(Z_x^2) = 1$

2. $E(Z_y^{2)} = P_{yx}^2 E(Z_x^2) + P_{y\varepsilon}^2 E(Z_\varepsilon^2) = P_{yx}^2 + P_{y\varepsilon}^2 = 1$

3. $E(Z_y Z_{x)} = P_{yx} E(Z_x^2) + P_{y\varepsilon} E(Z_\varepsilon Z_x) = P_{yx} = \rho_{xy}$

Note that, in a bivariate equation, $P_{yx} = \rho_{xy}$. This does not necessarily hold in the multivariate case. Now express parameters in terms of moments:

1. $E(Z_x^2) = 1$

2. $P_{yx} = \rho_{xy}$

3. $P_{y\varepsilon} = \%(1 - P_{yx}^2)$

We can also express the R^2 for this bivariate equation:

$R^2 = 1 - P_{y\varepsilon}^2$

$= P_{yx}^2$ in the bivariate case only

Now we can again express Σ_ρ in terms of $\Sigma_\rho(\theta)$:

$$\begin{bmatrix} 1 & \\ \rho_{xy} & 1 \end{bmatrix} = \begin{bmatrix} 1 & \\ P_{xy} & 1 \end{bmatrix}$$

$$\underset{2\times 2}{\Sigma_\rho} = \underset{2\times 2}{\Sigma_\rho(\theta)}$$

Let's return to our example:

$Y = \beta X + \varepsilon$, Unstandardized form, where Y = Income and X = Years of Education

$Z_y = P_{yx} Z_x + P_{y\varepsilon} Z_\varepsilon$ Standardized form, where Z_y = Income (standardized), Z_x = Years of Education (std.), and Z_ε = disturbance term (std.)

$P_{y\varepsilon} = (1 - P_{yx}^2)^{1/2} = [1 - (.625)^2]^{1/2} = (1 - .391)^{1/2} = .609^{1/2} = .780$

$R^2 = 1 - P_{y\varepsilon}^2 = 1 - .609 = .391$

What do the standardized parameters tell us?

Exercise: Due next time. Assume that the above population moments and parameters are for a population of white males in the U.S. Consider the following population moments on African-Americans males in the U.S.: $\mu_y = 7.0$ \$1,000; $\mu_x = 10$ years; $\sigma_x^2 = 4$ years^2; $\sigma_y^2 = 15.0$ 1,000\$^{2}; $\sigma_{xy} = 1.6$ \$1,0001yrs. Compute the unstandardized and standardized parameters above and briefly contrast the substantive interpretations of each parameter with that of whites.

IV. ESTIMATION AND TESTING.

So far, we have been describing our little model assuming we had access both to population moments and parameters. In such a case, we would have no reason for estimation or inferential statistics. In fact, however, we rarely observe the population directly; instead we only obtain data on a sample drawn from the population. The sample moments and parameter estimates are direct counterparts to population moments and parameters:

$\Sigma = \Sigma(\theta)$ Population moments in terms of population parameters
$S = \Sigma(\hat{\theta})$ Sample moments in terms of estimated parameters

We begin with sample data, which we characterize by the sample moments (sample covariance matrix and sample means):

$\bar{x} = 3x_i/n$
$\bar{y} = 3y_i/n$

$s_x^2 = \frac{1}{n-1}\sum (x_i - \bar{x})^2$

$s_{xy} = \frac{1}{n-1}\sum (x_i - \bar{x})\ (y_i - \bar{y})$

$s_y^2 = \frac{1}{n-1}\sum (y_i - \bar{y})^2$

$$\underset{2\times 2}{S} = \begin{bmatrix} s_x^2 & \\ s_{xy} & s_y^2 \end{bmatrix}$$

S, the sample moments of observed variables, is an unbiased and efficient estimator of Σ. We can express our sample moments in terms of parameters:

$$\underset{2\times 2}{\begin{bmatrix} s_x^2 & \\ s_{xy} & s_y^2 \end{bmatrix}} = \underset{2\times 2}{\begin{bmatrix} s_x^2 & \\ \hat{\beta} s_x^2 & \hat{\beta} s_x^2 + s_\varepsilon^2 \end{bmatrix}}$$

$$\underset{2\times 2}{S} = \underset{2\times 2}{\Sigma(\hat{\theta})}$$

This is precisely what we did above, except we've replaced the population covariance matrix with the sample covariance matrix. $\Sigma(\hat{\theta})$ is the matrix of moments *implied* by the *estimated* parameters of the model. We can now express estimated parameters in terms of sample moments:

1. $s_x^2 = s_x^2$
2. $\hat{\beta} = s_{xy}/s_x^2$
3. $s_\varepsilon^2 = s_y^2 - (s_{xy}^2/s_x^2)$

(Note that $\hat{\beta}$, the sample estimator, is a *random variable*; whereas β is a *constant*.) This method of estimation is called the method of moments (see Goldberger 1991, *A Course in Econometrics*,) which simply takes the equation expressing population parameters in terms of population moments and replaces population moments with sample moments. The parameter estimates obtained from sample moments are sample analogs of parameters expressed as population moments. Hence, in econometrics, the method of moments falls under the general term, "analog estimators."

Important fact: In recursive linear models in observables, estimation by the method of moments is identical to the OLS estimator, which gives desirable finite (small) sample properties: unbiased $E[\Sigma(\hat{\theta})] = \Sigma(\theta)$ and efficient $E\{\Sigma(\hat{\theta}) - E\,\Sigma(\hat{\theta})]\}^2$ is smaller than that of any other linear unbiased estimator. If we drew numerous samples, and constructed a distribution of OLS estimates, the average of the distribution would be the population parameter β, and the variance of the sampling distribution would be smaller than that of other unbiased linear estimators.

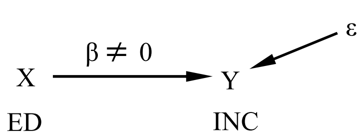
For any given sample, because of sampling variability, we cannot expect $S = \Sigma$, or $\Sigma(\hat{\theta}) = \Sigma(\theta)$. This implies that our parameter estimates will be subject to sampling variability. This means we need to assess the effects of chance on our estimates, and subject our estimates to formal hypothesis testing. We can do this using the usual principles of classical statistical inference.

Assuming that Y|X is normally distributed (that is ε is normally distributed), we can assume that $\hat{\beta}$ is normally distributed with mean β and variance $\sigma_\beta^2 = \sigma_\varepsilon^2/\sum (x_i - \bar{x})^2$. The standard error of $\hat{\beta}$ is the square root of the variance, $(\sigma_\beta^2)^{1/2}$. But because we typically do not know the population variance σ_ε^2 of the disturbance, we have to estimate it from the sample $s_\varepsilon^2 = s_y^2 - (s_{xy}^2/s_x^2)$. The estimated standard error of $\hat{\beta}$ is $s_\beta = [s_\varepsilon^2/\sum (x_i - \bar{x})^2]^{1/2}$. To test the null hypothesis, H_0: $\beta = \beta_0$ versus the alternative hypothesis H_1: $\beta \neq 0$, construct the t-statistic:

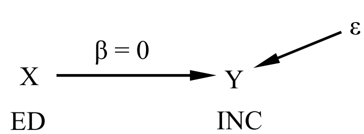
$$t_{n-1} = (\hat{\beta} - \beta_0)/s_\beta$$

If $t_{n-1} > t_{critical}$, (e.g., $t_{critical} > 1.96$) then we reject the null hypothesis. [Aside: note that when the intercept is included in the model, the degrees of freedom become n - 2, and the standard error for β is $s_\beta(n - 1)/(n - 2)$ (see Maddala 1988 *Introduction to Econometrics*)].

The two models that correspond to the two hypotheses illustrate an important principle: the second model (corresponding to the null hypothesis) is nested within the first model. Nested means that the second model is a special case of the first: by *constraining* one or more parameters in the first model, one can obtain the second. Here the second model constrains β to be zero, whereas the first model allows it to be any number. We can test the null hypothesis H_0: $\beta = 0$, against the alternative hypothesis, H_0: $\beta \neq 0$:



Model 1: Alternative Hypothesis: $\beta \neq 0$



Model 2: Null Hypothesis: $\beta = 0$

Let's run this bivariate regression model in LISREL 8.8:

```
BIVARIATE REGRESSION MODEL                  Title of the run

 DA NI=2 NO=500                             DA = data NI= No.of input vars. NO = no.cases

 CM                                         Read in a covariance matrix = CM
 *                                          * = Read data in free format (starting on next line)
 9
 7.2 14.76
 LA                                         LA = labels
 *                                          * = Read labels in free format (next line)
```

SE	SE = select variables (Ys first, then Xs)
2 1 /	The 2nd variable is Y and the first X. Slash /
MO NX=1 NY=1 GA=FU,FR PS=DI, FR	MO = Model Parameters NX= Xs, NY = Ys GA = gamma PS is the disturbance
VA 0.0 GA 1 1	VA = start value, GA 1 1 is gamma
PD	PD = Path Diageram
OU ME=ML SC	OU = Output ME = method of estimation, ML = maximum likeilhoood, SC = completely standardized solution

LISREL prints out the path diagram;

Here is the LISREL output:

The following lines were read from file H:\529 examples\bivariate regression.LS8:

```
 BIVARIATE REGRESSION MODEL
 DA NI=2 NO=500

 CM
 *
 9
 7.2 14.76
 LA
 *
 ED INC
 SE
 2 1 /
 MO NX=1 NY=1 GA=FU,FR PS=DI, FR
 VA 0.0 GA 1 1
 PD
 OU ME=ML  SC

 BIVARIATE REGRESSION MODEL

                          Number of Input Variables  2
                          Number of Y - Variables    1
                          Number of X - Variables    1
                          Number of ETA - Variables  1
                          Number of KSI - Variables  1
                          Number of Observations   500
 BIVARIATE REGRESSION MODEL

         Covariance Matrix
                INC         ED
            --------   --------
      INC      14.76
       ED       7.20       9.00

 BIVARIATE REGRESSION MODEL
```

= S, the sample covariance matrix

```
 Parameter Specifications

         GAMMA
                  ED
            --------
      INC          1
         PHI
                  ED
            --------
                   2
         PSI
                 INC
            --------
                   3
```

= our regression coefficient

= the variance of X (ED)

= the disturbance variance

```
BIVARIATE REGRESSION MODEL

 Number of Iterations =  0

 LISREL Estimates (Maximum Likelihood)

         GAMMA

                  ED
            --------
      INC       0.80
              (0.04)
               17.85
```

= $\hat{\beta}$
= Standard Error
= t-value (or Z-statistic)

```
         Covariance Matrix of Y and X

                 INC         ED
            --------   --------
      INC      14.76
       ED       7.20       9.00

         PHI

                  ED
            --------
                9.00
              (0.57)
               15.78
```

= estimated variance of X

```
         PSI

                 INC
            --------
                9.00
              (0.57)
               15.78
```

= Estimated variance of the disturbance

```
         Squared Multiple Correlations for Structural Equations

                 INC
            --------
                0.39
```

R-squared

```
                           Goodness of Fit Statistics
```

```
                              Degrees of Freedom = 0
                 Minimum Fit Function Chi-Square = 0.0 (P = 1.00)
        Normal Theory Weighted Least Squares Chi-Square = 0.00 (P = 1.00)

                  The Model is Saturated, the Fit is Perfect !


 BIVARIATE REGRESSION MODEL

 Standardized Solution

         GAMMA

                  ED
            --------
      INC       0.62
```

Standardized regression estimate $\widehat{P}_{xy}$

```
         Correlation Matrix of Y and X

                 INC         ED
            --------   --------
      INC       1.00
       ED       0.62       1.00

         PSI

                 INC
            --------
                0.61

         Regression Matrix Y on X (Standardized)

                  ED
            --------
      INC       0.62

                           Time used:    0.062 Seconds
```