Lecture 2: Econometrics Primer

This lecture is very terse review of econometrics from ECON 141: Introduction to Econometrics. We will not cover this material in class, but this outline is designed to help guide you with reading the Studenmund chapters assigned as background reading.

I. What is econometrics?
- Measurement – quantitative measurement of economic phenomena
- Uses
  - Describe/quantify economic reality
  - Testing hypotheses of economic theory
  - Forecasting future economic activity
- Explain/quantify the behavior of dependent variable using explanatory variables.

II. Fundamentals

A. Correlation vs. causation
- Empirically, we can only quantify relationships in terms of correlation
- Causation is based on our underlying model.
- Example: Relationship between real GDP, capital, and labor.
  Empirically, we can see how these variables are related
  We assume a causal relationship based on the production function – we assume capital and labor are determined exogenously.

B. Single-equation linear model

1) Variables
   - X – explanatory variable (independent variable)
   - Y – dependent variable

2) Function
   - Theoretical model: $Y = \beta_0 + \beta_1 X$
   - Data: $Y = \beta_0 + \beta_1 X + e$

3) Residual, e
   1. Specification error:
      a. Influences omitted from the equation
      b. Nonlinear relationship
   2. Measurement error
   3. Unpredictable variation

C. Cross section vs. time series data

1) Notation
   - common to use “i” to denote cross section (for individual)
   - common to use “t” to denote time series (time interval)

2) Example: Consumption Function
   - $C = \beta_0 + \beta_1 DI$
   - Cross section data on individual household consumption and household disposable income
   - Time series data on aggregate consumption and disposable income

D. Estimated relationship
   1) Predicted value of Y :: Yhat = E(Y_i | X_i)
   2) Estimated value of e :: ehat = Yhat – Y

III. OLS
A. Overview
   1) Objective
      1. Summation Notation
      Minimize the sum of squared errors
      Choose β_0 and β_1 to minimize the sum of squared residuals:

$$\min \sum_{i=1}^{N} (e_i)^2 = \min \sum_{i=1}^{N} (Y_i - \beta_o - \beta_1 X_i)^2$$

First-order conditions:

$$\frac{\partial}{\partial \beta_0} = 0 : -2\left[\sum Y_i - N\beta_0 - \beta_1 \sum X_i\right] = 0$$

$$\Rightarrow \beta_0 = \frac{\sum Y_i}{N} - \beta_1 \frac{\sum X_i}{N}$$

$$\Rightarrow \beta_0 = \overline{Y} - \beta_1 \overline{X}$$

$$\frac{\partial}{\partial \beta_1} = 0 : -2\left[\sum (Y_i - \beta_0 - \beta_1 X_i) \sum X_i\right] = 0$$

$$\Rightarrow \left[\sum \left(Y_i - (\overline{Y} - \beta_1 \overline{X}) - \beta_1 X_i\right) \sum X_i\right] = 0$$

$$\Rightarrow \sum \left[(Y_i - \overline{Y})(X_i - \overline{X})\right] - \beta_1 \sum (X_i - \overline{X})^2 = 0$$

$$\beta_1 = \frac{\sum \left[(Y_i - \overline{Y})(X_i - \overline{X})\right]}{\sum (X_i - \overline{X})^2}$$

      2. Matrix Notation $Y = X\beta + e$
      where

Y is an n x 1 vector	e is an n x 1 vector
X is an n x k matrix	β is a k x 1 vector

$$\min(e'e) = \min(Y - X\beta)'(Y - X\beta) = \min(Y'Y - 2\beta X'Y + \beta X'X\beta)$$

$$\frac{\partial}{\partial \beta} = 0 : -2X'Y + 2X'X\beta = 0$$

$$\Rightarrow (X'X)\beta = X'Y$$

$$\Rightarrow \beta = (X'X)^{-1} X'Y$$

B. Features
   1) Ybar = b0 + b1Xbar
   2) Sum(e)=0
   3) OLS is BLUE

C. Evaluating the regresion
   1) R^2
   2) F-test
   3) T-test of individual coefficients

IV. CLM

A. Assumptions (let T = sample size)
- Linear and correctly specified
- E(e)=0
- E(X,e)=0
- $E(e_t,e_{t-s})=0$ – no autocorrelation
- $E(e^2)=\sigma^2$ – no heteroskedasticity
- No perfect correlation
- $e \sim N(0,\sigma^2)$

B. OLS is BLUE – Gauss Markov Theorem implies the following for the estimated β:

Unbiased:	E[βhat] = β
Minimum variance (efficient)	var[βhat] is as small as possible
Consistent	lim E[βhat] $\rightarrow \beta$ as $T \rightarrow \infty$
Normal distribution	βhat ~ N(β,var(βhat))

V. Hypothesis Testing

A. Overview
- It's almost never possible to prove something (theory) correct or "true" in the same way we can in mathematics.
- Strategy: Take a random sample to see whether it conforms to a hypothesis
  - If we can reject a hypothesis with a certain degree of confidence, then it's very unlikely the sample result would have been observed if the theory were correct.

B. Steps
   1) Set up null and alternative hypothesis
      H_0: $\beta_1 = 0$
      H_a: $\beta_1 = 0$
      Analogy: guilt or innocence of the accused

   2) Determine degree of confidence w/ which to reject the null
      Type I error: reject a true null hypothesis = find an innocent person guilty
      Type II error: fail to reject a false null = find the guilty person innocent

      We face a tradeoff, which type of error do we want to avoid?

Eliminate Type I error – never reject the null = find everyone innocent
Reject the null only when we're very confident

3) Decision rule
Common decision rule is to use a 95% confidence level – 90% and 99% also common

C. t-test
Used for hypothesis testing of individual regression coefficients
Standardize the βhat to use a common distribution – the t-distribution.

In the CLM, βhat ~ N(β,var(βhat)).

t = (βhat – β_0)/se(βhat) t ~ N(0, s2)

This is why the assumptions of the CLM are so important. If we violate any of the assumptions, like normality, no autocorrelation, no heteroskedasiticity, then the se(βhat) are incorrectly estimated using OLS and all of our statistical inference is incorrect.

D. F-test
Used for hypothesis testing of multiple regression coefficients at once.