Introductory Econometrics

Lecture 1

Xiaoxia Shi

University of Wisconsin - Madison

09/02/2010

What is Econometrics?

What is Econometrics?

- Econometrics employs statistical methods

What is Econometrics?

What is Econometrics?

- Econometrics employs statistical methods
- to analyze data

What is Econometrics?

What is Econometrics?

- Econometrics employs statistical methods
- to analyze data
- in order to

What is Econometrics?

- Econometrics employs statistical methods
- to analyze data
- in order to
    1. estimate economic relationships,

What is Econometrics?

What is Econometrics?

- Econometrics employs statistical methods
- to analyze data
- in order to
  1. estimate economic relationships,
  2. evaluate government and business policies,

What is Econometrics?

What is Econometrics?

- Econometrics employs statistical methods
- to analyze data
- in order to
  1. estimate economic relationships,
  2. evaluate government and business policies,
  3. test economic theories, and

What is Econometrics?

What is Econometrics?

- Econometrics employs statistical methods
- to analyze data
- in order to
  1. estimate economic relationships,
  2. evaluate government and business policies,
  3. test economic theories, and
  4. make predictions and forecasts.

Data that Econometricians Use

Four types of Data:

- Cross-Sectional Data

Data that Econometricians Use

Four types of Data:

- Cross-Sectional Data
- Time Series Data

Data that Econometricians Use

Four types of Data:

- Cross-Sectional Data
- Time Series Data
- Pooled Cross Sections

Data that Econometricians Use

Four types of Data:

- Cross-Sectional Data
- Time Series Data
- Pooled Cross Sections
- Panel Data

Random Variables

- What is a random variable?
- In the linear regression:

$$Y = \beta_0 + \beta_1 X + U$$

Random Variables

- What is a random variable?
- In the linear regression:

$$Y = \beta_0 + \beta_1 X + U$$

- are Y, X, U random variables?

Random Variables

- What is a random variable?
- In the linear regression:

$$Y = \beta_0 + \beta_1 X + U$$

- are Y, X, U random variables?
- are β_0, β_1 random variables?

Random Variables

- What is a random variable?
- In the linear regression:

$$Y = \beta_0 + \beta_1 X + U$$

- are Y, X, U random variables?
- are β_0, β_1 random variables?
- Are the OLS estimators of β_0, β_1 random variables?

Expectation

- What is Expectation?
- What's the difference between expectation and sample average?

Expectation

- What is Expectation?
- What's the difference between expectation and sample average?
- Properties of expectation:

$$E\left(aX+bY+c\right)=?$$

Expectation

- What is Expectation?
- What's the difference between expectation and sample average?
- Properties of expectation:

$$E\left(aX + bY + c\right) = ?$$

- What's conditional Expectation?

Expectation

- What is Expectation?
- What's the difference between expectation and sample average?
- Properties of expectation:

$$E\left(aX + bY + c\right) = ?$$

- What's conditional Expectation?
- Law of Iterated Expectation:

$$E\left[E\left(Y|X\right)\right] = E\left(Y\right)$$

Variance, Covariance and Correlation

- What's the variance of a random variable X?

Variance, Covariance and Correlation

- What's the variance of a random variable X?
- What's the covariance of random variable X and Y?

Variance, Covariance and Correlation

- What's the variance of a random variable X?
- What's the covariance of random variable X and Y?
- Properties of variances:

$$Var\left(aX + bY + c\right) = ?$$

Variance, Covariance and Correlation

- What's the variance of a random variable X?
- What's the covariance of random variable X and Y?
- Properties of variances:

$$Var\left(aX + bY + c\right) = ?$$

- What's the correlation between X and Y?

$$\rho\left(X, Y\right) = \frac{?}{?}$$

Simple Regression

- We want to know the parameters β_0 and β_1 in the population equation:

$$Y = \beta_0 + \beta_1 X + U$$

Simple Regression

- We want to know the parameters β_0 and β_1 in the population equation:

$$Y = \beta_0 + \beta_1 X + U$$

- We have a sample $\{X_i, Y_i\}_{i=1}^n$. How do we estimate the parameters?

Simple Regression

- We want to know the parameters β_0 and β_1 in the population equation:

$$Y = \beta_0 + \beta_1 X + U$$

- We have a sample $\{X_i, Y_i\}_{i=1}^n$. How do we estimate the parameters?
- Ordinary least square (OLS):

$$\hat{\beta}_0, \hat{\beta}_1 \text{ minimizes } \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i)^2$$

Simple Regression

- We want to know the parameters β_0 and β_1 in the population equation:

$$Y = \beta_0 + \beta_1 X + U$$

- We have a sample $\{X_i, Y_i\}_{i=1}^n$. How do we estimate the parameters?
- Ordinary least square (OLS):

$$\hat{\beta}_0, \hat{\beta}_1 \text{ minimizes } \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i)^2$$

- What is the formula for the OLS estimators? (hint: take first derivatives and set to zero)

$$\hat{\beta}_0 = ?, \ \hat{\beta}_1 = ?$$

Another Way to Derive OLS Estimators

- Population Moments: the true parameters (β_0, β_1) solve:

$$\begin{aligned} E(Y - \beta_0 - \beta_1 X) &= 0 \\ E\left(YX - \beta_0 X - \beta_1 X^2\right) &= 0 \end{aligned}$$

Another Way to Derive OLS Estimators

- Population Moments: the true parameters (β_0, β_1) solve:

$$\begin{aligned} E(Y - \beta_0 - \beta_1 X) &= 0 \\ E\left(YX - \beta_0 X - \beta_1 X^2\right) &= 0 \end{aligned}$$

- Sample analogue: $\hat{\beta}_0, \hat{\beta}_1$ solves:

$$\begin{aligned} n^{-1} \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i) &= 0 \\ n^{-1} \sum_{i=1}^{n} \left(Y_i X_i - b_0 X_i - b_1 X_i^2\right) &= 0 \end{aligned}$$

Another Way to Derive OLS Estimators

- Population Moments: the true parameters (β_0, β_1) solve:

$$\begin{aligned} E(Y - \beta_0 - \beta_1 X) &= 0 \\ E\left(YX - \beta_0 X - \beta_1 X^2\right) &= 0 \end{aligned}$$

- Sample analogue: $\hat{\beta}_0, \hat{\beta}_1$ solves:

$$\begin{aligned} n^{-1} \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i) &= 0 \\ n^{-1} \sum_{i=1}^{n} \left(Y_i X_i - b_0 X_i - b_1 X_i^2\right) &= 0 \end{aligned}$$

- $\hat{\beta}_0 = ?$, $\hat{\beta}_1 = ?$

Residual and Predicted Value

- What is "regression residual"?

Residual and Predicted Value

- What is "regression residual"?
- $$\hat{U}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$$

Residual and Predicted Value

- What is "regression residual"?
- $$\hat{U}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$$
- What is "predicted value of Y_i"?

Residual and Predicted Value

- What is "regression residual"?
- $$\hat{U}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$$
- What is "predicted value of Y_i"?
- $$\begin{aligned} \hat{Y}_i &= \hat{U}_i \\ &= \hat{\beta}_0 + \hat{\beta}_1 X_i \end{aligned}$$

Residual and Predicted Value

- What is "regression residual"?
- $$\hat{U}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$$
- What is "predicted value of Y_i"?
- $$\begin{aligned} \hat{Y}_i &= \hat{U}_i \\ &= \hat{\beta}_0 + \hat{\beta}_1 X_i \end{aligned}$$
- Now,

$$\sum_{i=1}^{n} \hat{U}_i = ?, \qquad \sum_{i=1}^{n} \hat{U}_i X_i = ?$$

Changing Units of Measurement

- Holding other variables constant, if smoking one more cigarette reducing baby birth weight by 1 pound, what's the effect of smoking one pack (20) more of cigarette?

Changing Units of Measurement

- Holding other variables constant, if smoking one more cigarette reducing baby birth weight by 1 pound, what's the effect of smoking one pack (20) more of cigarette?
- $Y = \beta_0 + \beta_1 X + U = \beta_0 + (20\beta_1)\left(\frac{X}{20}\right) + U$

Changing Units of Measurement

- Holding other variables constant, if smoking one more cigarette reducing baby birth weight by 1 pound, what's the effect of smoking one pack (20) more of cigarette?
- $Y = \beta_0 + \beta_1 X + U = \beta_0 + (20\beta_1)\left(\frac{X}{20}\right) + U$
- What's the effect of smoking one more cigarette on baby birth weight measured by grams?

Changing Units of Measurement

- Holding other variables constant, if smoking one more cigarette reducing baby birth weight by 1 pound, what's the effect of smoking one pack (20) more of cigarette?
- $Y = \beta_0 + \beta_1 X + U = \beta_0 + (20\beta_1)\left(\frac{X}{20}\right) + U$
- What's the effect of smoking one more cigarette on baby birth weight measured by grams?
- $$\begin{aligned} \frac{(450Y)}{450} &= \beta_0 + \beta_1 X + U \qquad \Rightarrow \\ (450Y) &= 450\beta_0 + 450\beta_1 X + 450U \end{aligned}$$

Log-forms

- How to interpret coefficients obtained from the regression:

$$\log(Y) = \beta_0 + \beta_1 X + U?$$

Log-forms

- How to interpret coefficients obtained from the regression:

$$\log(Y) = \beta_0 + \beta_1 X + U?$$

- one unit of change in X changes Y by $\beta_1 \times 100\%$

Log-forms

- How to interpret coefficients obtained from the regression:

$$\log(Y) = \beta_0 + \beta_1 X + U?$$

- one unit of change in X changes Y by $\beta_1 \times 100\%$
- How to interpret coefficients obtained from the regression:

$$Y = \beta_0 + \beta_1 \log(X) + U?$$

Log-forms

- How to interpret coefficients obtained from the regression:

$$\log\left(Y\right) = \beta_0 + \beta_1 X + U?$$

- one unit of change in X changes Y by $\beta_1 \times 100\%$
- How to interpret coefficients obtained from the regression:

$$Y = \beta_0 + \beta_1 \log(X) + U?$$

- one percent change in X changes Y by $\beta_1/100$ units

R-square

- What is the sum of squared total (SST) of the dependent variable Y?

R-square

- What is the sum of squared total (SST) of the dependent variable Y?
- What is the sum of squared explained (SSE) of the regression $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$?

R-square

- What is the sum of squared total (SST) of the dependent variable Y?
- What is the sum of squared explained (SSE) of the regression $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$?
- What is the sum of squared residual (SSR) of that regression?

R-square

- What is the sum of squared total (SST) of the dependent variable Y?
- What is the sum of squared explained (SSE) of the regression $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$?
- What is the sum of squared residual (SSR) of that regression?
- How much of the variation in Y is explained by the model?

$$R^2 = \frac{?}{?} = \frac{1-?}{?}$$

Classical Linear Model Assumptions:

- Assumption SLR.1 (Linear in Parameters) $Y = \beta_0 + \beta_1 X + U$
- Assumption SLR.2 (Random Sampling) (X_i, Y_i), $i = 1, ..., N$, is a random sample from the population.
- Assumption SLR.3 (Sample Variation in the Explanatory Variable) $\{X_i, i = 1, ..., N\}$ are not all the same value.
- Assumption SLR.4 (Zero Conditional Mean) $E(U|X) = 0$.
- Assumption SLR.5 (Homoskedasticity) $Var\,(U|X) = \sigma^2$.

Unbiasedness and Variance

- What is unbiasedness?

$$\text{we call the estimator } \hat{\beta}_1 \text{ unbiased if } E\left(\hat{\beta}_1|X\right) = ?$$

- Under which assumptions unbiasedness hold?
- Under all of the five assumptions,

$$Var\left(\hat{\beta}_1|X\right) = ?$$

- What happens if there is no variation in X (SLR.3 is violated)?
- What happens if there is heteroskedasticity (SLR.5 is violated)?
- What does Gauss-Markov Theorem say?

Omitted Variable Bias

- When we do causal analysis, what is the interpretation for U?
- If X and U are positively correlated, which assumption is violated? What happens to $E\left(\hat{\beta}_1|X\right)$ now?
- What is the effect of omitting a variable that is independent of X?
- What is the effect of omitting a variable that has both direct positive effect on Y and positive effect on X?
- What is the effect of omitting a variable that has both direct negative effect on Y and positive effect on X?

Multiple Regression

$$Y = \beta_0 + \beta_1 X_1 + ... + \beta_K X_K + U$$

- Estimated equation:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + ... + \hat{\beta}_K X_{Ki}$$

- How are the OLS estimators $\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_K$ obtained?

Unbiasedness and Variance

- Under Assumptions MLR.1-4 (Linearity, Random Sampling, No Multicollinearity, Conditional Mean-Zero):

$$E\left(\hat{\beta}_j|X\right) = ?$$

- Under the above assumptions and MLR.5 (Homoskedasticity):

$$Var\left(\hat{\beta}_j|X\right) = \frac{\sigma^2}{SST_X\left(1 - R_j^2\right)}$$

What is σ^2, SST_X or R_j^2?

What Variables to Include?

- How does R^2 change when we add more variables to the regression?

What Variables to Include?

- How does R^2 change when we add more variables to the regression?
- What's the effect on the unbiasedness of the estimators of including irrelevant variables?

What Variables to Include?

- How does R^2 change when we add more variables to the regression?
- What's the effect on the unbiasedness of the estimators of including irrelevant variables?
- What's the effect on the unbiasedness of the estimators of omitting relevant variables?

What Variables to Include?

- How does R^2 change when we add more variables to the regression?
- What's the effect on the unbiasedness of the estimators of including irrelevant variables?
- What's the effect on the unbiasedness of the estimators of omitting relevant variables?
- Are there any reasons for not including a particular variable on the left hand side?

Squares and Interactions

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + ... + U$$

- What's the partial effect (marginal effect) of X_1 on Y?

$$\frac{\partial Y}{\partial X_1} = ?$$

Squares and Interactions

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + ... + U$$

- What's the partial effect (marginal effect) of X_1 on Y?

$$\frac{\partial Y}{\partial X_1} = ?$$

- $$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + ... + U$$

Squares and Interactions

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + \dots + U$$

- What's the partial effect (marginal effect) of X_1 on Y?

$$\frac{\partial Y}{\partial X_1} = ?$$

- $$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \dots + U$$

- What's the partial effect (marginal effect) of X_1 on Y?

$$\frac{\partial Y}{\partial X_1} = ?$$

Confidence Interval

- How is the t-statistic for $\hat{\beta}_j$ formulated?

$$t_n = \frac{\hat{\beta}_j - ?}{?}$$

Confidence Interval

- How is the t-statistic for $\hat{\beta}_j$ formulated?

$$t_n = \frac{\hat{\beta}_j - ?}{?}$$

- What's the 95% confidence interval of β_j?

$$\left[\hat{\beta}_j - 1.96 * ?, \hat{\beta}_j + 1.96 * ?\right]$$

Confidence Interval

- What is type I error of a test?

Confidence Interval

- What is type I error of a test?
- What is type II error of a test?

Confidence Interval

- What is type I error of a test?
- What is type II error of a test?
- What is the significance level of a test?

Confidence Interval

- What is type I error of a test?
- What is type II error of a test?
- What is the significance level of a test?
- Suppose we are doing the two-sided test of $H_0 : \beta_j = 0$ vs. $H_1 : \beta_j \neq 0$.

Confidence Interval

- What is type I error of a test?
- What is type II error of a test?
- What is the significance level of a test?
- Suppose we are doing the two-sided test of $H_0 : \beta_j = 0$ vs. $H_1 : \beta_j \neq 0$.
- If $\hat{\beta}_j = 2$, $se(\hat{\beta}_j) = 0.1$, do we reject the null at 5% level?

A Single Linear Combination of Parameters

- How do we test:

$$H_0 : a\beta_1 + b\beta_2 = 0?$$

A Single Linear Combination of Parameters

- How do we test:

$$H_0 : a\beta_1 + b\beta_2 = 0?$$

- Modified the regression:

$$\begin{aligned} Y &= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \varepsilon \\ &= \beta_0 + (a\beta_1)\frac{X_1}{a} + (b\beta_2)\frac{X_2}{b} + \ldots + \varepsilon \\ &= \beta_0 + (a\beta_1 + b\beta_2)\frac{X_1}{a} + (b\beta_2)\left(\frac{X_2}{b} - \frac{X_1}{a}\right) + \ldots + \varepsilon \end{aligned}$$

A Single Linear Combination of Parameters

- How do we test:

$$H_0 : a\beta_1 + b\beta_2 = 0?$$

- Modified the regression:

$$\begin{aligned} Y &= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \varepsilon \\ &= \beta_0 + (a\beta_1)\frac{X_1}{a} + (b\beta_2)\,\frac{X_2}{b} + ... + \varepsilon \\ &= \beta_0 + (a\beta_1 + b\beta_2)\frac{X_1}{a} + (b\beta_2)\left(\frac{X_2}{b} - \frac{X_1}{a}\right) + ... + \varepsilon \end{aligned}$$

- Run regression of Y on $\frac{X_1}{a}$ and $\frac{X_2}{b} - \frac{X_1}{a}$.

Multiple Hypotheses

- How do we test:

$$H_0 : \beta_1 = 0, \beta_2 = 0?$$

Multiple Hypotheses

- How do we test:

$$H_0 : \beta_1 = 0,\ \beta_2 = 0?$$

- Obtain SSR (SSR_{ur}) from the unrestricted regression:

$$log\left(wage\right) = \beta_0 + \beta_1 jc + \beta_2 univ + \beta_3 exper + \varepsilon$$

Multiple Hypotheses

- How do we test:

$$H_0 : \beta_1 = 0,\ \beta_2 = 0?$$

- Obtain SSR (SSR_{ur}) from the unrestricted regression:

$$log\left(wage\right) = \beta_0 + \beta_1 jc + \beta_2 univ + \beta_3 exper + \varepsilon$$

- Obtain SSR (SSR_r) from the restricted regression:

$$\log\left(wage\right) = \beta_0 + \beta_3 exper + \varepsilon.$$

Multiple Hypotheses

- Form the F-statistic:

$$F \equiv \frac{(SSR_r - SSR_{ur})/q}{SSR_{ur}/(n-K-1)} \sim F_{q,n-K-1}.$$

What is q? What is $F_{q,n-K-1}$?

Multiple Hypotheses

- Form the F-statistic:

$$F \equiv \frac{(SSR_r - SSR_{ur})/q}{SSR_{ur}/(n-K-1)} \sim F_{q,n-K-1}.$$

What is q? What is $F_{q,n-K-1}$?

- Reject H_0 if

$$F > F_{q,n-K-1,1-\alpha}$$

where α is the significance level, $F_{q,n-K-1,1-\alpha}$ is the $1-\alpha$ quantile of $F_{q,n-K-1}$

Read the Table Reported by STATA

Source	SS	df	MS
Model	32001.7271	1	32001.7271
Residual	396383.812	1532	258.73617
Total	428385.539	1533	279.442622

Number of obs =	1534
F(1, 1532) =	123.68
Prob > F =	0.0000
R-squared =	0.0747
Adj R-squared =	0.0741
Root MSE =	16.085

prate	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
mrate	5.861079	.5270107	11.12	0.000	4.82734	6.894818
_cons	83.07546	.5632844	147.48	0.000	81.97057	84.18035