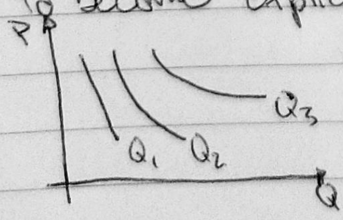
Econometrics

Introduction

- What is econometrics?
  - Economic Measurement
- It is the quantitative measurement and analysis of actual economic & business phenomena - and so it involves:
  - Economic theory, statistics, Math, observation/data collection
- 3 major uses of econometrics
  - Describing economic reality
  - Testing hypothesis about economic theory
  - Forecasting future economic activity
- So Econometrics is all about questions
  - The researcher (you!) first asks questions and then uses econometrics to answer them
- Economics suggests important relationships, often with policy implications, but virtually never suggests quantitative magnitude of causal effects.
  - How does another year of education affect earnings?
  - What is price elasticity of darts?
  - What is the effect on output growth of a 1% point decrease in interest rates by the Fed? etc.
- Consider the general & purely theoretical relationship:
  - $Q = f(P)$
- Econometrics allows this general and purely theoretical relationship to become explicit.
- 

$Q^D = f(P)$
"Theory"

- Example: Demand for iPhone 6 (Price drop).

  Result ⇒ $Q^D = 10 - 0.007P$

  ② $P = \$599$ ① $P = \$699$

  $Q^D = 5.807$ $Q^D = 5.107$

Causal Effects

- The course is about using data to measure causal effects
- Ideally we would like data from an experiment
- But almost always we only have observational (nonexperimental) data.
  - returns to education, cigarette prices, monetary policies
- Most of the course deals with difficulties arising from using observational data to estimate causal effects.
  - Confounding effects (omitted factors), simultaneous causality, "correlation doesn't imply causation".

In this course you will

- Learn methods for estimating causal effects using observation data
-

Statistical Principles.

Probability

- A random variable, x, is a variable whose numerical value is determined by chance, the outcome of a random phenomenon
  - A discrete random variable has a countable number of possible values
  - A continuous random variable can take on any value in an interval
- A probability distribution, P(x), for a discrete random variable, x, assigns probabilities to the possible values x_1, x_2 and so on
-

Mean, Variance, & Standard Deviation

- The expected value (mean) of a discrete random variable, x, is a weighted average of all possible values of x, using the probability of each x values as weights: $\mu_x = E(x) = \sum_{i=1}^{N} x_i P(x)$
- When all weights are equal we can simplify that to: $\mu_x = E(x) = \frac{1}{N}\sum_{i=1}^{N} x$
- The variance of a discrete random variable, x, is a weighted average, for all possible values of x, of the squared difference between x_i and its expected value, using the probability of each x value as weights $\sigma_x^2 = E[(x-\mu_x)^2] = \sum_{i=1}^{N}(x_i-\mu_x)^2 P(x_i)$
- When all weights are equal we can simplify that to:
  $\sigma_x^2 = E[(x_i-\mu_x)^2] = \frac{1}{N}\sum_{i=1}^{N}(x_i-\mu_x)^2$

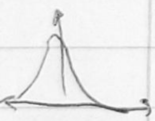
Standardized Variables

- To ~~standardized~~ standardize a random variable, x, we subtract its mean μ and then divide by its standard deviation σ

$$z = \frac{x_i - \mu_x}{\sigma_x}$$

- No matter what the units of x, standardized ~~deviation~~ random variable z has a mean of 0 and a standard deviation of 1.
- The standardized varible, z, measures how many standard deviations x is above or below it mean.
  - If $x = \mu$, $z = 0$
  - If x = 1 stdev above mean, $z = 1$
  - If " " 2 " " ", $z = 2$.

The Normal Distribution



- The density curve of Z for many rolls of a dice approaches the normal distribution
- The central limit theorem (CLT) states:
  - "If z is a standardized sum of N independent, identically distributed random variables with a finite, nonzero standard deviation, the the probability distribution of Z approaches the normal distribution as N increases."

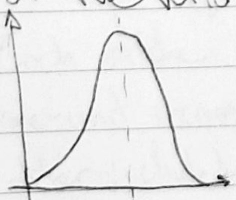
Table B-7 Textbook.

- In other words, CLT says the sum (or mean) of many random variables is distributed as ~~and~~ according to the normal distribution
- Special feature of the normal distribution: the probability that the value of z will be in a specified interval is given by the corresponding area under the density curve
  - The areas can be found in tables or software
  - $P(-1 < z < 1) \cong 68\%$

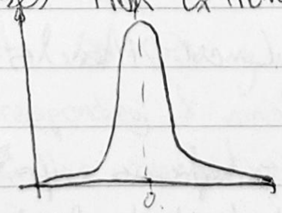
    $P(-2 < z < 2) \cong 95\%$

Moments of a Distribution

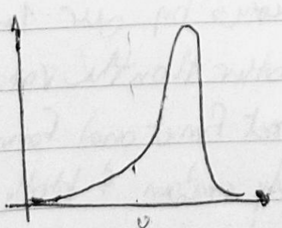
- The mean of a distribution is called the first (central) moment of a distribution. The variance is the 2^{nd} moment of a distribution. When a stats distribution has a strong central tendency, it is useful to characterize it by its moments.
- The 3^{rd} & 4^{th} moments of a distribution are skewness & kurtosis
- Skewness is a measure of symmetry
  - Skewness = $\frac{E[(X-\mu_X)^3]}{\sigma_X^3}$
  - $= 0 \Rightarrow$ symmetric
  - $> 0 \Rightarrow$ long right tail
  - $< 0 \Rightarrow$ long left tail
- Kurtosis measures the mass in the tails of a distribution. It is a measure for the probability of large values.
  - Kurtosis = $\frac{E[(X-\mu_X)^4]}{\sigma_X^4}$
  - $= 3 \Rightarrow$ normal distribution
  - $> 3 \Rightarrow$ heavy tails
- Kurtosis of a distribution is a measure of how much is in the tails & therefore is a measure of how much of the variance of X arises from extreme values
-



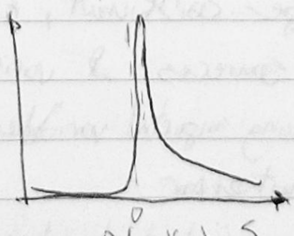
Skew = 0, Kurt = 3



Skew = 0, Kurt = 20



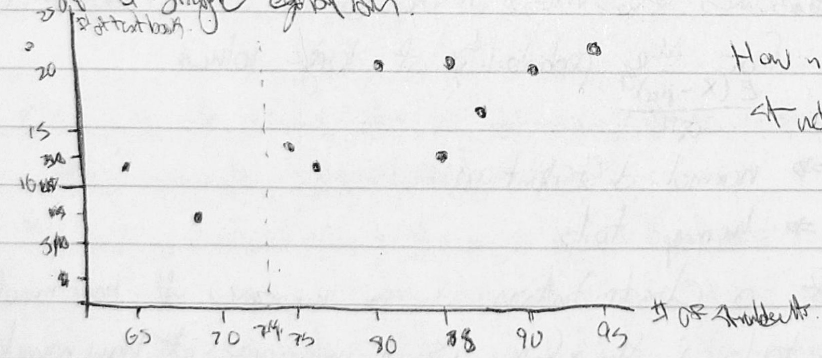
Skew = -0.1, Kurt = 5



Skew = 0.6, Kurt = 5

What is Regression Analysis?

- Economic theory can give us the direction of a change (change in demand, followed by price decrease)
- But what if we want to know not just "how?" but also "how much"
- Then we need: - A sample of data
  - A way to estimate such a relationship
  - Most frequently done by regression analysis.
- Formally, regression analysis is a statistical technique that attempts to "explain" movements in one variable, the dependent variable, as a function of movements in a set of other variables, the independent (or explanatory) variables, through the quantification of a single equation.



How many texts for 74 students

Single-Equation-Linear Models

- The simplest example is: $y = B_0 + B_1 x$
- $B_0 \rightarrow$ constant / intercept of the term
- $B_1 \rightarrow$ slope coefficient; Amount y increases by our 1 unit of x
- There are 4 sources of variation in y other than the variation in x.
  1. Other missing important variables
  2. Measurement error
  3. Incorrect functional form
  4. Purely random & totally unpredictable occurrences
- Inclusion of a "stochastic term" effectively takes care of these sources of variation in y.

- $y = \underbrace{\beta_0 + \beta_1 x}_{\text{deterministic}} + \underbrace{\varepsilon}_{\text{stochastic}}$

- Why is the first component deterministic?
  - Indicates the value of y that is determined by a given value of x (which is non-stochastic
  - Alternatively the deterministic component can be thought of as the expected value of y given x - namely $E(y|x)$. The mean (or average) value of the y's associated with particular value of x.
  - This is also denoted the conditional mean of y
  - Thus stats principles of distribution for conditional mean apply.

Extending the Notation

- Include reference to the number of observations
  - Single-equation linear case: $y_i = \beta_0 + \beta_1 x + \varepsilon_i \; (i = 1, 2, \ldots n)$
  - So really, there are N equations, one for each observation
  - coefficients, β_0 & β_1 stay the same in all N equations
  - The values of y, x, & ε differ across all observations.
- The general case: multivariate regression
  $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots \varepsilon_i \; (i = 1, 2, \ldots N)$
- Each of the slope coefficients $(\beta_1 - \beta_3)$ gives the impact of a one-unit increase in the corresponding x variable on y, holding the other included independent variables constant.

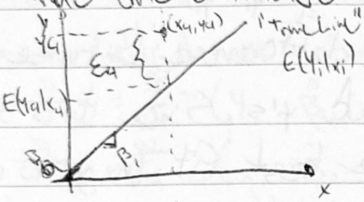
Wage Regression.

- Let wages in \$ depend on:
  - years of work experience (EXP)
  - years of education (EDU)
  - gender (GEN) (Dummy variable)
- Substituting into equation (3) yields:
  - $WAGE_i = \beta_0 + \beta_1 EXP_i + \beta_2 EDU_i + \beta_3 GEN_i + \epsilon_i$
  - $\beta_0 \to$ Average wage of $\to$ female $\to$ no EDU $\to$ no EXP
  - $\beta_1 \to$ The amount of \$ more a person make with one additional year of EXP
  - $\beta_2 \to$ " " " " " " " " " " EDU
  - $\beta_3 \to$ " " " a male makes than female.

The Estimated Regression ~~M~~ Equation

- The regression equation considered ~~so~~ for are the "true" -but unknown- theoretical regression equations/functions
- Instead of "true" one might think about this as the population regression function vs the sample/estimated regression function
- How do we obtain the empirical counterpart of a theoretical regression model? - It has to be estimated.
- The empirical counterpart to eqn (2) is
  - $y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i$
- The estimated error term is e_i. The signs on top of the estimates are denoted "hat".
- For each sample we get a different set of estimated regression coefficients
- $\hat{y}_i$ is the estimated value of y_i; similarly, its the prediction of $E(y_i|x)$ from the regression eqn.
- The closer $\hat{y}_i$ is to the observed y_i, the better the "fit" of eqn
- Similarly, the smaller the estimated error term, the better the "fit"

- This can also been seen in the fact that
  - $e_i = y_i - \hat{y}_i$
- Note, difference with the error term, ε_i, given as
  - $\varepsilon_i = y_i - E(y_i|x_i)$
- This comes together in the next Figure.
- True and Estimated Regressions



- It is important to clarify the different ways of writing the regression model.
  - ① Population "true" regression fxn:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

$$E(y_i|x_i) = \beta_0 + \beta_1 x_i$$

  - ② Estimated "sample" regression fxn.

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i$$

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

Estimating Single-Independent-Variable Models with OLS.

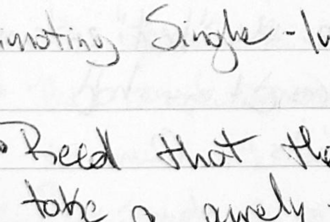
- Recall that the purpose of regression analysis is to take a purely theoretical equation like:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \qquad (4.1)$$

- And through the use of data, to get to:

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i \qquad (4.2).$$

- Recall that equation (4.1) is purely theoretical and can never be observed, while equation (4.2) is its empirical counterpart.
- This lecture is about how to move from (4.1) to (4.2).



$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i$

$e_i = y_i - \hat{y}_i$

observed expected

- One of the most widely used methods is Ordinary Least Squares (OLS).
- OLS minimizes $\sum_{i=1}^{N} e_i^2 = \quad (i = 1, 2, \ldots N)$.
- We also denote this term the "residual sum of squares" (RSS).
- Using $e_i = y_i - \hat{y}_i$ OLS minimizes $\sum_{i=1}^{N} (y_i - \hat{y}_i)^2$
- That means, OLS minimizes the difference between the predicted y's & the observed y's. Since it is squared OLS treats observations below the best fit (negative residual) the same as observations above (positive residuals).
- Why use OLS? Why minimizing the sum of squared residuals?
  1. Relatively easy to use. The calculations can be done by hand
  2. OLS is intuitively / theoretically appealing. It makes sense that we want the estimated regression equation to be as close as possible to the observed data
  3. OLS estimates have at least 2 useful characteristics:
     - OLS can be shown to be the "best" estimator when certain specific conditions hold
     - The sum of the residuals is 0.

Math Reminder.

- $\sum_{i=1}^{N} x_i = x_1 + x_2 \ldots x_N$
- $\sum_{i=1}^{N} c = cN$
- $\sum_{i=1}^{N} c x_i = c \sum_{i=1}^{N} x_i$
- $\sum_{i=1}^{N} (x_i + y_i) = \sum_{i=1}^{N} x_i + \sum_{i=1}^{N} y_i$
- $(a+b)^2 = a^2 + 2ab + b^2$

Estimating Single-Independent-Variable Models with OLS

- How does OLS compute the coefficient? OLS minimizes the squared residual, summed over all the sample data points:

$$\min_{e_i} \sum_{i=1}^{N} e_i^2 = \min_{\beta_0, \beta_1} \sum_{i=1}^{N} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

- To find a minimum we use FOCs (set PDE $=0$). The eqns are called normal equations

$$\frac{\partial}{\partial \beta_0} = \sum_{i=1}^{N} 2(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)(-1) = 0$$

$$\frac{\partial}{\partial \beta_0} = \sum_{i=1}^{N} 2(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)(-x_i) = 0$$

- Next we will derive the 2 OLS estimators with the help of the 2 normal eq'ns.
- Optimization Problem $\min_{e_i} \sum e_i^2$

Using: $y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i$ or $e_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$

$$\min_{\beta_0 \beta_1} \sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$ Objective function.

$$\frac{\partial}{\partial \hat{\beta}_0} = \sum 2(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)(-1) = 0 \quad \text{I}$$

$$\frac{\partial}{\partial \hat{\beta}_1} = \sum 2(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)(-x_i) = 0 \quad \text{II}$$

} Normal Equations.

2 equations 2 unknowns

- The sample variance s_x^2 is an estimator of the population variance σ_x^2 and is given by: $s_x^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2$
- The sample covariance s_{xy} is an estimator of the population covariance, σ_{xy} & is given by: $s_{xy} = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})$
- We divide by $(N-1)$ instead of N, because the mean of X, μ_x and the mean of Y, μ_x and are unknown and have to be estimated themselves with $\bar{x}$ & $\bar{y}$. This creates a bias that needs to be corrected. Note, when N is large, it makes little difference whether we divide by N or by $(N-1)$
- The two estimators are

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\hat{\beta}_1 = \frac{\sum[(x_i - \bar{x})(y_i - \bar{y})]}{\sum (x_i - \bar{x})^2} = \frac{s_{xy}}{s_x^2}$$

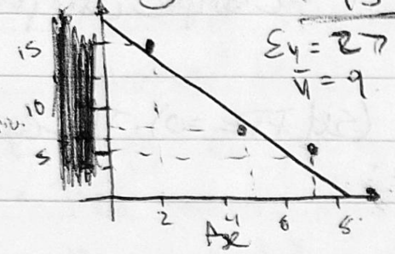
- Note that the slope estimate (the impact x & y) is the sample covariance of x & y, s_{xy} divided by the sample variance of x, s_x^2 (see next slide).

- Ex.

Car#	Price y (in thous)	Age x	$(x_i-\bar{x})$	$(y_i-\bar{y})$	$(x_i-\bar{x})(y_i-\bar{y})$	$(x_i-x)^2$
1	5	7	2.33	4	9.32	5.43
2	7	5				
3	+ 15	+ 2				
	$\Sigma y = 27$	$\Sigma x = 14$				
	$\bar{y} = 9$	$\bar{x} = 4.67$				

$\hat{\beta}_1 = -2.131$

$\hat{\beta}_0 = 18.5$



- Why is the sum of the residuals zero?

$\Sigma e_i = 0$ (4.4.a)

$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i$ or $e_i = -\hat{\beta}_0 - \hat{\beta}_1 x_i + y_i$

Now: $\Sigma e_i = \Sigma y_i - N\hat{\beta}_0 - \hat{\beta}_1 \Sigma x_i$ $\quad N$

$\frac{1}{N}\Sigma e_i = \frac{1}{N}\Sigma y_i - \hat{\beta}_0 - \frac{\hat{\beta}_1}{N}\Sigma x_i$

$\bar{e} = \bar{y} - \hat{\beta}_0 - \hat{\beta}_1\bar{x}$ $\quad \bar{y} - \hat{\beta}_1 \bar{x}$

$\bar{e} = \bar{y} - \bar{y} + \beta_1 \bar{x} - \beta_1 \bar{x}$

$\bar{e} = 0$.

Estimating Multivariate Regression Models with OLS.

- In the 'real world' one explanatory variable is not enough
- The general multivariate regression model with k independent variables is: $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots \beta_k x_{ki} + \varepsilon_i$ $(i = 1, 2, \dots N)$
- Most of the characteristics from single variable model stay valid. The biggest difference is in the interpretation of the slope coefficient:
  - Now a slope coefficient indicates the change in the dependent variable associated with a one-unit increase in the explanatory variable holding the other explanatory variables constant.

Decomposition of Variance

- Econometrics use the squared variation of Y around its mean to be explained by the model, called the total sum of squares (TSS):
$TSS = \sum_{i=1}^{N} (Y_i - \bar{Y})^2$
- The TSS for OLS has 2 components: variation that can be explained by the regression and variation that cannot (decomposition of variance):
$$\underbrace{\sum_{i=1}^{N}(Y_i - \bar{Y})^2}_{TSS} = \underbrace{\sum_{i=1}^{N}(\hat{Y}_i - \bar{Y})^2}_{ESS} + \underbrace{\sum_{i=1}^{N} e_i^2}_{RSS}$$

Evaluating the Quality of a Regression Equation:

- ① Is the equation supported by sound theory?
② How well does the estimated regression fit the data?
③ Is the data set reasonably large and accurate?
④ Is the OLS the best estimator to be used for this equation?
⑤ How well do the estimated coefficients correspond to the expectations developed by the researcher before the data were collected?
⑥ Are all the obviously important variables included in the equation?
⑦ Has the most theoretically logical functional form been used?
⑧ Does the regression appear to be free of major econometric problems?

Describing the Overall Fit of the Estimated Model.

- The commonly used measure of overall fit is the coefficient of determination, R^2:
$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum e_i^2}{\sum (Y_i - \bar{Y})^2}.$$
- Since OLS selects the coefficient estimates that minimizes RSS, OLS provides the largest possible R^2
- If the equation has exactly one independent variable we have the following relation to the correlation coefficient r:
$$r = \sqrt{R^2}.$$

The Adjusted Coefficient of Determination

- A major problem with R^2 is that it can never decrease if another independent variable is added.
- An alternative to R^2 ~~tha~~ that addresses this issue is the adjusted R^2 also called "R-squared-bar":

$$\bar{R}^2 = 1 - \frac{\sum e_i^2/(N-K-1)}{\sum(y_i - \bar{y})^2/(N-1)}$$

where $N-K-1$ = degrees of freedom
$N :=$ # of observation.
$K :=$ # of slope coefficients

- So, $\bar{R}^2$ measures the share of the variation of Y around its mean that is explained by the regression equation, adjusted for degrees of freedom
- $\bar{R}^2$ can be used to compare the fits of regressions with the same dependent variable and different numbers of independent variables
- As a result, most researchers automatically use $\bar{R}^2$ when evaluating the fit of their estimated regressions equations

The Classical Model

The Classical Assumptions

- The classical linear regression is the cornerstone of most econometric theory. The model was developed by Gauss in 1821. The model is based on ~~seven~~ seven assumptions. When these assumptions hold, OLS is considered the "best" estimator available for regression models.
- The seven assumptions are:
  1. The regression is linear, is correctly specified, & has an additive error
     $Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i$
  2. The error term has a zero population mean:
     $E(\varepsilon_i) = 0$
  3. All explanatory variables are uncorrelated with the error term
     $\text{cov}(X_i, \varepsilon_i) = 0$
  4. Observations of the error term are uncorrelated with each other (no serial correlation).
     $\text{cov}(\varepsilon_i, \varepsilon_j) = 0$.
  5. The error term has a constant variance (no heteroskedasticity)
     $\text{var}(\varepsilon_i) = \sigma^2$
  6. No explanatory variable is a perfect linear function of any other explanatory variable(s) (no perfect multicollinearity).
     $X_i = cX_i$ where c is some constant.
  7. the error term is normally distributed (this assumption is optional but usually is invoked)
     $\varepsilon_i \sim N(0, \sigma^2)$.

① Linear, Correctly Specified, Additive Error term

- Consider the following regression model:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots \beta_K x_{Ki} + \varepsilon_i$$

- This model - is linear (in the coefficients)
  - has an additive term
- If we also assume all the relevant explanatory variables are included in ($y_i = \beta_0 + \beta_1 x_{1i} \dots$) then the model is also correctly specified.

② Error Term has a zero population mean.

- Recall the stochastic (random) error term.
- Reason to account for variation in the dependent variable that is not explained by the model
- Assumption ② states that the specific value of the error term for each observation is determined purely by chance

③ All Explanatory Variable are uncorrelated with the error term

- If ③ was violated, the OLS estimates would be likely to attribute to the x some of the variation in y that actually came from the error term
- This assumption is violated most frequently when a researcher omits an important independent variable from an equation. Or when a model is simultaneous in nature

④ No special correlation of the error term

- If, overall the observations of the sample ϵ_{t+1} is correlated with ϵ_t then the error term is said to be serially correlated (or auto-correlated) and ⑤ is violated.
- If a systematic correlation does exist between one observation of the error term and another, then it will be more difficult for OLS to get accurate estimates of the standard errors of the coefficients.
- This assumption is most likely to be violated in time-series models.
  - An increase in the error term in one time period is likely followed by an increase in the next period, also.

⑤ Constant variance / No heteroskedasticity in the Error term

- The error term must have one constant variance, that is, the variance of the error term cannot change for each observation or range of observations.
- If it does, there is heteroskedasticity present in the error term.
- ⑤ is likely to be violated in cross-sectional data sets

⑥ No perfect multicollinearity

- Perfect collinearity between two independent variables implies that:
  - they are really the same variable, or
  - one is a multiple of the other, and/or
  - that a constant has be added to one of the variables.
- Solution - dropping one of the perfectly collinear variables from the equation
- But even imperfect multicollinearity can cause problems for estimation

⑦ The error term is ~~nor~~ normally distributed

- Although ~~when~~ we have already assumed that observations of the error term are drawn independently from a distribution that has a mean of zero and that has a constant variance, we have said little about the shape of that distribution.
- This assumption basically implies that the error term follows a bell-shaped curve.
- Strictly speaking not required for OLS estimation (related to Gauss-Markov Theorem more on this shortly)
- Its major application is in hypothesis testing, which uses the estimated regression coefficient to investigate hypotheses about economic behavior

About the Assumptions

- How realistic are all these assumptions?
- In any scientific study we make certain assumptions because they facilitate the development of the subject matter in gradual steps, not because they are necessarily realistic
- Some argue that it does not matter whether the assumptions are realistic. What matters are the predictions based on those assumptions.

The Sampling Distribution of $\hat{\beta}$

- Just as the error term follows a probability distribution, so too do the estimates of β
  - The probability distribution of these $\hat{\beta}$ values across different samples is called the sampling distribution of $\hat{\beta}$
- We will next look at the properties of the mean, the variance and the standard error of this sampling distribution.

Properties of the Variance

- Just as we wanted the mean of the sampling distribution to be centered around the true population $\hat{\beta}$ so too, it is desirable for the sampling distribution to be as narrow (or precise) as possible
  - Centering around "the truth" but with high variability might be of very little use.
- One way of narrowing the sampling distribution is to increase the sampling size (which therefore also increases the degrees of freedom)

The Sampling Distribution of $\hat{\beta}$.

- The variance and standard error of the estimated coefficients are given by

$$var(\hat{\beta}_0) = \frac{\hat{\sigma}^2 \sum (x_i)^2}{N \sum (x_i - \bar{x})^2} \quad , \quad se(\hat{\beta}_0) = \sqrt{var(\hat{\beta}_0)}$$

$$var(\hat{\beta}_1) = \frac{\hat{\sigma}^2}{\sum (x_i - \bar{x})^2} \quad , \quad se(\hat{\beta}_1) = \sqrt{var(\hat{\beta}_1)}$$

- where $\hat{\sigma}^2$ denotes the variance estimator of the random error

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{N-2} = \frac{\text{residual sum of squares (RSS)}}{\text{number of degrees of freedom}}$$

The Gauss-Markov Theorem and the Properties of OLS Estimators

- Given assumptions ①-⑥, the OLS estimator is the minimum variance estimator from among the set of all linear unbiased estimators (also called BLUE - Best Linear Unbiased Estimator).
- If we add to ⑦, the result of the G-M Theorem is strengthened because of the OLS estimator can be shown to be the best unbiased estimator (BUE) out of all possible estimators
- Unbiased - OLS estimates coefficients are centered around the true population values
- Minimum Variance - no other unbiased estimator has a lower variance for each estimated coefficient than OLS
- Consistent - as the sample size gets larger, the variance gets smaller, and each estimate approaches the true value of the coefficient being estimated
- Normally Distributed - when the error term is normally distributed, so are the estimated coefficients which enables various statistical tests requiring normality to be applied

Notations

- Population Parameter | Estimate

Name	Symbol	Name	Symbol
Regression Coefficient	β_k	Estimated Regression Coeff.	$\hat{\beta}_k$
Expected Value of Estimated Coeff	$E(\beta_k)$		
Variance of Error term	σ^2 or $Var(\epsilon_i)$	Estimated variance of error term	s^2 or $\hat{\sigma}^2$
Stdev of error term	σ	Standard Error of Equation	s or SE
Variance of estimated coeff	$\sigma^2(\beta_k)$ or $Var(\beta_k)$	Estimated variance of estimated coefficient	$s^2(\hat{\beta}_k)$ or $\widehat{VAR}(\beta_k)$
Stdev of estimated coeff	$\sigma_{\hat{\beta}}$ or $\sigma(\hat{\beta}_k)$	Standard Error of estimated coeff	$\hat{\sigma}(\hat{\beta}_k)$ or $SE(\beta_k)$
Error or disturbance term	ϵ_i	Residual	e_i

Hypothesis Testing

What is hypothesis testing?

- Economist are interested in numerous relationships, for example that between demand & price.
- With hypothesis testing we can statistically test these relationships with the help of sample data

Classical Null and Alternative Hypotheses

- Hypotheses to be tested:
  - Null hypothesis ("H_0") - the outcome that the researcher doesn't expect
  - Alternative hypothesis ("H_A") - the outcome the researcher does expect.

Type I and Type II Errors

- Two types of errors possible in hypothesis testing
  - Type I Rejecting a true null hypothesis
  - Type II: Not rejecting a false null hypothesis
- Alternatively its possible to obtain an ~~estimate~~ of β that is close enough to zero (or ~~negative~~) to be considered "not significantly" positive
- Such a result may lead the ~~researcher~~ to "accept" the null hypothesis that $\beta \leq 0$ when in truth $\beta > 0$

Decision Rules of Hypothesis Testing

- To test a hypothesis, we calculate a sample statistic that determines when the null hypothesis can be rejected depending on the magnitude of that sample statistic relative to a pre-selected critical value
- This procedure is referred to as a decision rule
- The range of possible values of the estimates is divided into 2 regions, an "acceptance" (really, non rejection) region & and a rejection region
- The critical value strictly seperates the "acceptance"/non-rejection region from the rejection region when testing a null hypothesis
-

The t-test.

- The t-test is usually used to test hypotheses about individual regression slope coefficients
  - Tests of more than one coefficient at a time (joint hypothesis) are typically done with the F-test, presented shortly
- The appropriate test to use when stochastic error term is normally distributed (Assumption 7) and when the variance of that distribution must be estimated
  - Since these usually are the case, the use of the t test for hypothesis testing has become standard practice in econometrics

The t-statistic

- For a typical multiple regression equation: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i$
we can calculate t-values for each of the estimated coefficients
  - Usually these are only calculated for the slope coefficients
- Specifically, the t-statistic for the kth-coefficient is:
$t_k = \frac{(\hat{\beta}_k - \beta_{H_0})}{SE(\hat{\beta}_k)}$ $(k = 1, 2, \ldots K)$

The Critical t-Value and the t-Test Decision Rule.

- To decide whether to reject or not to reject a null hypothesis based on a calculated t-value, we use a critical t-value
- A critical t-value is the value that distinguishes the "acceptance" region from the rejection region.
- The critical t-value, t_c, is selected from a t-table depending on:
  - whether the test is one-sided or two-sided
  - the level of ~~Error~~ Type I Error specified and
  - the degrees of freedom (defined as the number of observations minus the number of coefficients estimated (including the constant) or $N-K-1$).

p-Values

- This is an alternative to the t-test
- A p-value, or marginal significance level, is the probability of observing a t-score that size or larger (in absolute value) if the null hypothesis were true
- In theory, we could find this by combing through pages and pages of statistics
- GRETL & other software give p-values as the standard output.
- In light of all this, the p-value decision rule therefore is:
Reject H_0 if p-value$_k$ < the level of significance and for one-sided tests:
if $\hat{\beta}_k$ has the sign implied by H_A

Confidence Intervals

- A confidence interval is a range that contains the true value on item a specified percentage of the time
- It is calculated using the estimated regression coefficient, the two-sided critical t-value ~~and~~ the standard error of the estimated coefficient as follows:

$$\text{Confidence interval} = \hat{\beta} \pm t_c \cdot SE(\hat{\beta})$$

- What's the relationship between confidence intervals and two-sided hypothesis testing?
- If a hypothesized value falls within the confidence interval then we cannot reject the null hypothesis (given the same level of significance)

Limitations of the t-Test

- 1- The t-Test Does Not Test Theoretical Validity:
  If you regress the consumer price index on rainfall in a time-series regression and find strong statistical significance it remains a nonsense result
- 2- The t-Test Does Not Test Importance
  The fact that one coefficient is "more statistically significant" than another does not mean that it is also more important in explaining the dependent variable
- 3- The t-Test Is Not Intended for Test of the Entire Population
  From the definition of the t-score, it is seen that as sample size approaches the population, the t-score approaches infinity.

The f-Test

- The f-Test checks for multiple hypothesis jointly such as:

$$\beta_2 = \beta_3 = 0$$

- The f-Statistic: $F = \dfrac{(RSS_M - RSS)/M}{RSS/(N-K-1)}$

RSS = Residual Sum of Squares

RSS_M = Residual Sum of Squares for restricted eqn

M = number of restrictions placed on eqn

(N-K-1) = degrees of freedom

- The decision rule is: Reject H_0 if $F > F_c$

Do not reject H_0 if $F \leq F_c$

Specification - Building a Regression Model

Specifying an Econometric Equation and Specification Error

- Before any equation can be estimated, it must be completely specified
- Specifying an econometric equation consists of 3 parts
  1. independent variables
  2. functional form
  3. form of the stochastic error term.
- Note, this is part of the 1st classical assumption.
- A specification error results when one of these choices is made incorrectly
- ~~This lecture~~

Choosing Independent Variables

Omitted Variables

- Two reasons why an important explanatory variable might have been left out.
  - we forgot or it is not available in the data set.
- Either way, this may lead to omitted variable bias
- The reason for this is that when a variable is not included, it cannot be held constant
- Omitting a relevant variable usually is evidence that the entire equation is a suspect, because of the likely bias of the coefficients

The Consequences of an Omitted Variable

- Suppose the true regression model is

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i$$

- If X_2 is omitted, the equation becomes

$$Y_i = \beta_0 + \beta_1 X_{1i} + \epsilon_i^* \quad , \quad \epsilon_i^* = \epsilon_i + \beta_2 X_{2i}$$

- Hence, the explanatory variables in the estimated regression are not independent of the error term
- But this violates classical assumption 3
- ~~Thus~~ When there is a violation of the classical assumptions, the Gauss-Markov Theorem does not hold, and the OLS estimates are not BLUE
- This results in a bias
- Also, the variance of the estimated coefficients decreases, and this increases the absolute magnitude of their t-scores
  - ⇒ we are more likely to reject H_0 when it is actually true

Correcting for an Omitted Variable

- In theory, the solution to a problem of specification bias seems easy: add the omitted variable to the equation
- Unfortunately, that's easier said than done, for a couple of reasons
  1. Omitted variables bias is hard to detect. the amount of bias introduced can be small and not immediately detectable
  2. Even if it has been decided that a given equation is suffering from omitted variable bias, how to decide exactly which variable to include?
- The best way to prevent the issue of omitted variables is to base the model on sound theory

Irrelevant Variable

- This refers to the case of including a variable in an equation when it does not belong there
- This is the opposite of the omitted variable case - and so the impact can be illustrated using the same model
- Assume that the true regression specification is:

$$y_i = \beta_0 + \beta_1 X_{1i} + \epsilon_i$$

- But for some reason the researcher includes an extra variable

$$y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i^{**}$$

- Where the error then becomes:

$$\epsilon_i^{**} = \epsilon_i - \beta_2 X_{2i}$$

- So, the inclusion of an irrelevant variable will not cause bias (since the true coefficient of the irrelevant variable is zero, and so the second term will drop out of)
- However, the inclusion of an irrelevant variable will
  - increase the variance of ~~the~~ estimated coefficients, and this increased variance will tend to decrease the absolute magnitude of their t-scores ⇒ we are more likely to not reject H_0 when it is untrue (Type II error)

Effects of Omitted Variables and Irrelevant Variables on the Coefficient Estimates

Effect on Coefficient Estimates	Omitted Variables	Irrelevant Variable
Bias	Yes	No
Variance	Decrease	Increases

Four Important Specification Criteria

- We can summarize the previous discussion into 4 criteria to help decide whether a given variable belongs in the equation
    1. Theory - Is the variables place in the equation ambiguous and theoretically sound?
    2. t-Test - Is the variables estimated coefficient significant in the expected direction
    3. $\bar{R}^2$ - Does the overall fit of the equation improve when the variable is added to the equation?
    4. Bias - Do the variables' coefficient change significantly when the variable is added to the equation?
- If all these conditions hold, the variable belongs in the equation
- If none of them hold, it does not belong
- The tricky part is the intermediate cases: use sound judgment

Choosing a Functional form

Alternative Functional Forms

- An equation is linear in the variables if plotting the function in terms of X and Y generates a straight line
- Ex $Y = \beta_0 + \beta_1 X_i + \epsilon_i$, is in linear
  $Y = \beta_0 + \beta_1 X_i^2 + \epsilon_i$, is not linear
- Similarly, an equation is linear in the coefficients only if the coefficients appear in their simplest form - they: aren't raised to any powers (>1), are not multiplied or divided by other coefficients, or are not in a function (log or exp)
- In fact, of all possible equations for a single explanatory variable, only functions of the general form: $f(Y) = \beta_0 + \beta_1 f(X)$ are linear in the coefficients β_0 and β_1

Linear Form

- This is based on the assumption that the slope of the relationship between the independent variable and the dependent variable constant. $\frac{\Delta Y}{\Delta X_k} = \beta_k \quad k = 1, 2, \ldots K$
- For this linear case, the elasticity of Y with respect to X is

$$\text{Elasticity}_{Y,X_k} = \frac{\Delta Y / Y}{\Delta X_k / X_k} = \frac{\Delta Y}{\Delta X_k} \cdot \frac{X_k}{Y} = \beta_k \frac{X_k}{Y}$$

Double-Log Form

- Here, the natural log of Y is the dependent variable and the natural log of X is the independent variable

$$\ln y = \beta_0 + \beta_1 \ln x_{1i} + \beta_2 \ln x_{2i} + \epsilon_i$$

- In a double log equation, an individual regression coefficient can be interpreted as an elasticity because:

$$\beta_k = \frac{\Delta \ln Y}{\Delta \ln X_k} = \ldots = \text{Elasticity}_{Y, X_k}$$

- Note that the elasticities of the model are constant and the slope are not
- This is in contrast to the linear model, in which the slopes are constant but the elasticities are not

Polynomial Form

- Polynomial functional form express Y as a function of independent variables, some of which are raised to powers other than 1
- The slope of Y with the respect to X_1 is

$$\frac{\Delta Y}{\Delta X_1} = \beta_1 + 2\beta_2 X_1$$

- Note the slope depends on X.

Summary of Alternative Functional Forms

- The choice of a functional form should be based on theory.
- Functional form | Eq. | Meaning of β_1

Functional form	Eq.	Meaning of β_1
Linear	$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$	Slope
Double Log	$\ln Y_i = \beta_0 + \beta_1 \ln X_i + \epsilon_i$	Elasticity
Polynomial	$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \epsilon_i$	Roughly slope

Using Dummy Variables

- A dummy variable is a variable that takes on the values of 0 or 1, depending on whether a condition for a qualitative attribute is met
- These conditions take the general form:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \epsilon_i$$

- This is an example of an intercept dummy

Slope Dummy Variable

- The slope dummy is also called interaction term changes both the intercept slope
- The general of a slope dummy equation is:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \beta_3 X_i D_i + \epsilon_i$$

- The slope depends on D

when $D = 0$, $\frac{\Delta Y}{\Delta X} = \beta_1$

when $D = 1$, $\frac{\Delta Y}{\Delta X} = (\beta_1 + \beta_3)$

Multicollinearity

Introduction and Overview

- The next lectures deal with violations of the classical assumptions and remedies for those violations
- The lecture addresses multicollinearity; we will also look into serial correlation and heteroskedasticity (time permitting).
- For each of these problems, we will attempt to answer the following questions:
  1. What is the nature of the problem?
  2. What are the consequences of the problem?
  3. How is the problem diagnosed.
  4. What remedies for the problem are available.

Imperfect Multicollinearity

- We have seen before that perfect multicollinearity can be solved by dropping the collinear variable
- Imperfect multicollinearity occurs when two (or more) explanatory variables are imperfectly related, as in:

  $x_{1i} = \alpha_0 + \alpha_1 x_{2i} + u_i$, notice the stochastic error term
- That means, the relationship between x_1 and x_2 might be fairly strong, but it is not strong enough to allow x_1 to be completely explained by x_2.