Introduction to Econometrics for Finance (AcFn521)

Tigabu Molla (PhD)
bitanya2007@gmail.com
Faculty of Business and Economics
Woldia University
2023

Abstract

- **Overview:**

  Econometrics has become an integral part of teaching & research in modern economics and business. Along with microeconomics and macroeconomics, it has been taught as one of the three core courses in most undergraduate and graduate economic programs. The importance of econometrics has been increasingly recognized and econometric tools and methods have been widely employed in empirical studies. This is a graduate level course and focuses on the classical paradigm of econometrics whereby we use sample to infer about corresponding population. Topics to be studied include nature of econometrics, specification, estimation, and inference in the context of models that include then extend beyond the standard linear multiple regression framework. It incorporates theoretical background and practical skills in econometrics used to answer economic/business/financial questions using data on entities observed at one or multiple points in time.

- **Objective:**

  Preparation to read and carry out empirical social science research using modern econometric methods.

Learning outcomes

At the end of this course you should:

- have knowledge and skills of regression analysis
- know the theoretical background and assumptions for standard econometric methods
- be able to use SPSS/STATA to perform an empirical analyses
- be able to read, understand, evaluate and interpret empirical studies (like journal articles that make use of econometric methods)
- be able to make use of econometric techniques in your own academic and other empirical works, for example in your master's thesis

Course content

Chapter 1: Introduction to econometrics

Chapter 2: Regression with one regressor (simple regreion)

- simple regression model
- The method of OLS
- The classical linear regression model and assumptions
- Properties of least square estimators
- The coefficient of determination
- Hypothesis testing

Chapter 3: Multiple regression model

- The three variable model notation and assumptions
- Hypothesis testing
- Interpretation

Chapter 4: Problems of measurement, specification, estimation and their solutions

- Heteroskedasticity, Autocorrelation, Multicollinearity
- Distributed lag models and expectations

Chapter 5: Basic understanding of econometric software useful for model building (Stata)

Course delivery and evaluation

- Teaching methods:
  - Student-centred
- Practicals:
  - Software
    - SPSS/STATA
    - Questions and review as requested
  - Problem Sets: (more details later)
- Readings:
  1. Books:
     - Wooldridge, Jeffery M: Introductory Econometrics
     - "Introductory Econometrics for Finance" by Chris Brooks. Cambridge.
     - Greene, W., Econometric Analysis, 8th Edition, Prentice Hall, 2017
  2. Notes and materials
  3. A few articles

Cont.

Prerequisites:

- A previous course that used linear regression
- A background in statistics and mathematics

Course Requirements:

- Problem set and article review (30%)
- Midterm (20%)
- Final exam (50%)

Chapter 1
The Nature of Econometrics and Economic Data

Structure:

1.1 The nature of econometrics

- ❑What is econometrics?
- ❑Why econometrics?

1.2 **Definition of financial econometrics**

1.3 Major uses of (financial) econometrics

1.4 Data characteristics (types of data)

1.5 Classical Inference versus Bayesian Inference

1.6 What is Regression Analysis?

1.7 Methodology of econometrics

- Model formulation
- Model estimation
- Model evaluation

1.1 The Nature of Econometrics

❑What is Econometrics?

- Econo-metrics: ”measurement in economics”. (literal meaning)
- **Econometrics** is the science and art of using economic/financial theory, mathematical techniques, statistical techniques and data in the analysis of economic/business/financial issues.
  - Econometrics helps us to use information from financial/management/economic theory and data in appropriate techniques to make economic decisions

Cont.

- Econometrics helps us to identify two things in relationship between variables : **quantification** (how much is the effect) and **attribution** (whether the effect is attributed to the variable assumed to be cause not something else)
- Listing variables in economic relationship is not enough; we need to answer the how much question and what causes it.
- For effective policy we must know the amount of change needed for a policy instrument to bring about the desired effect.
- Example: the demand function for a good Q = f(P, T, Pr, Y, N)

  Where Q is quantity demanded, P own price, T taste/preference, Pr price of related goods, Y income and N number of buyers

  - a) Does this relationship exist, and b) how much is the magnitude for the effect?
  - This is where econometrics comes in.

1.2 Definition of financial econometrics

- **Financial econometrics** is a subset of econometrics and is referred to as the application of econometrics to problems in finance.

- **What makes financial econometrics distinct?**
  - Financial econometrics uses financial theory, math, statistical tools and financial data to address financial issues.
  - That is, it is the use of financial data and financial theory to address financial issues that makes it distinct.

Cont.

❑What are financial issues?

- Finance is concerned with how to allocate scarce resources across assets overtime to earn a return.
  - What should we invest in (capital budgeting)?
  - Should we use cash (equity) or incur debt (capital structure)?
  - Working capital, for example, liquidity issues?
- These financial issues are addressed to maximize return
- However, since the future is unknown, this makes finance difficult.

Returns in financial modeling

Since decisions are made in face of uncertainty, econometric methods are important in finance.

1. Expected return of a share is the sum of the earnings per share and expected percentage capital gain.

$$E(r) = \frac{E(P_{t+1}) - P_t}{P_t} + \frac{E(D_{t+1})}{P_t}$$

2. CAPM

The CAPM states that the expected return on any stock i is equal to the risk-free rate of interest, R_f , plus a risk premium. $E(R_i) = R_f + \beta_i[E(R_m) - R_f]$

Cont.

- We use financial theories and financial data along with the tools of mathematics and statistical principles to address financial issues and earn the best possible return

❑What are financial theories?

These are theories that deal with finance.

❑Examples:

– Optimal asset allocation

– value at risk (a statistic that quantifies the extent of possible financial losses within a firm, portfolio, or position over a specific time frame)

Cont.

- arbitrage pricing model
- price/index forecasting
- volatility modeling
- term structure of interest rate (yield curve) (The term structure of interest rate can be defined as the graphical representation that depicts the relationship between interest rates (or yields on a bond) and a range of different maturities. The graph itself is called a "**yield curve**."
- market efficiency

❑What are financial data?

Financial data have special characteristics.

What are the Special Characteristics of Financial Data?

- *Frequency & quantity of data*

  Stock market prices are measured every time there is a trade or somebody posts a new quote.

- *Quality*

  Recorded asset prices are usually those at which the transaction took place. No possibility for measurement error but financial data are "noisy".

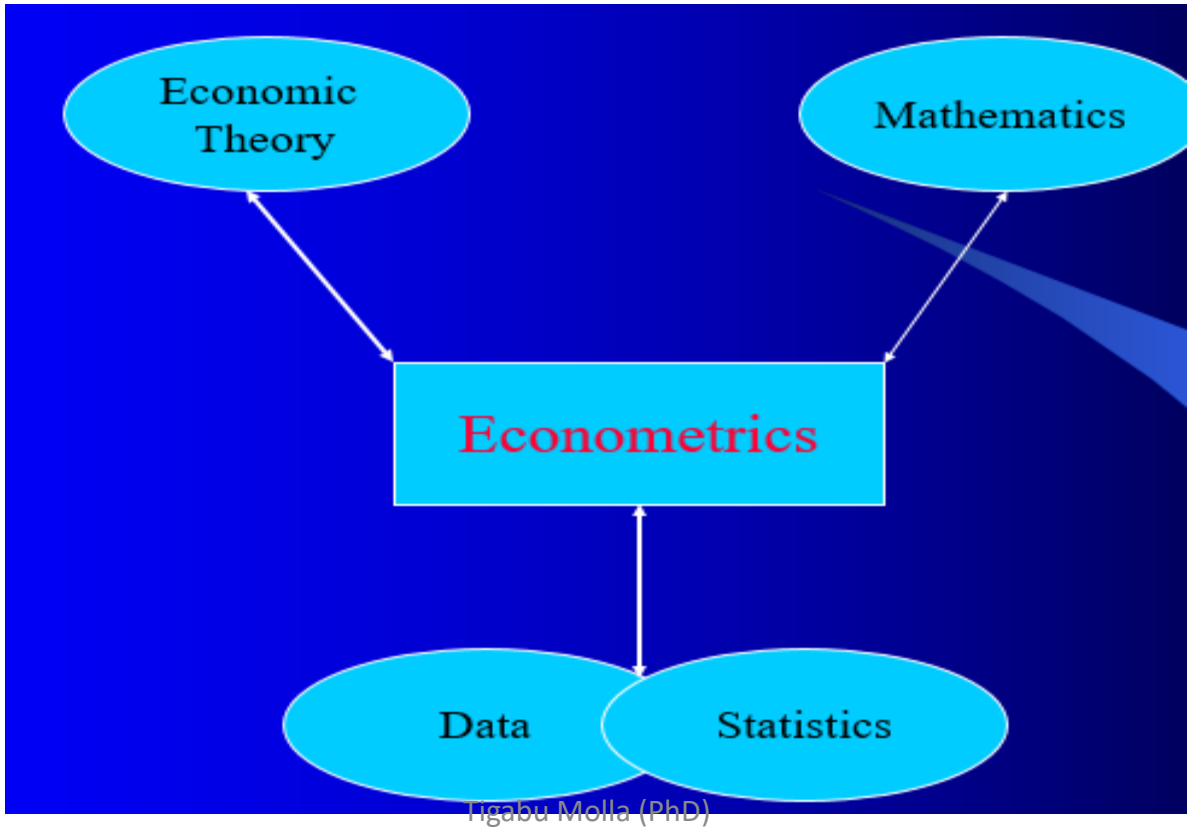
Examples of questions we would like to study:

- What is the effect of advertising on sales?
- What is the effect of risk on return?
- What is the effect of a retraining program on the duration of unemployment?
- What is the effect of an additional year of education on earnings?
- How is stock price volatility related to macroeconomic factors?
- Testing whether financial markets are weak-form informationally efficient.
- Is the CAPM consistent with the reality?
- Do earnings or dividend announcements have effect on stock prices?

How does (financial) econometrics address such questions?

❑How does (financial) econometrics address such questions?

➢It integrates (unifies) economic/financial theory, mathematics, statistics and economic data for a real understanding of economic/business issues



❑Economic/financial theory

- Economic/financial theory only suggests a qualitative economic relationship, while econometrics gives empirical (quantitative) content to most economic theory

❑Mathematics

- Summarizes the essence of a theory in concise manner without empirical measure of the theory.
- Ensures logical consistency and coherency in economic theory itself (checks whether the reasoning process of an economic theory is correct)
- mathematical modeling is a necessary path to empirical verification of an economic theory.
- Econometrics gives consistency between theory and stylized facts

❑Statistics

The tools and methods of statistics provide the operating principles for analyzing data.

❑Data

- Data are realizations of stochastic setting (economy)
- We use economic data to ascertain the consistency of economic theory with reality.

1.3 Major uses of (financial) Econometrics

1. Estimating relationships between variables

   EX: the relationship between education and wages or between risk and return

2. Testing theories in economics/finance.

   Ex: Testing theories of capital structure (static trade-off, pecking order, …)

3. Impact evaluation.

   Ex: Evaluating the impact of community based health insurance or safety net program

4. Forecasting financial variables (asset prices, returns…)

   Ex: What will be the inflation rate in 2015 E.C in Ethiopia?

Uses of financial econometrics

- Testing theories in finance,
- Determining asset prices or returns,
- Testing hypotheses concerning the relationships between variables,
- Examining the effect on financial markets of changes in economic conditions,
- Forecasting future values of financial variables

1.4 Data Characteristics

- All empirical analysis requires data.
- Different techniques are required to analyze different types of data.
- Data characteristics:
  - Quantitative versus qualitative data
  - Cross sectional, time series and panel data
  - Experimental versus non-experimental data
  - Nominal, ordinal and scale variables

Cont.

a. Quantitative versus Qualitative

- Quantitative variables measure "quantities" such as price, sales volume, weight or income.
- Qualitative variables are used to model "either/or" situations and might be used to model membership in one of several groups such as homeowner or non-homeowner, employed/unemployed, male/female, accurate or inaccurate income tax returns
- Dependent and independent variables can be quantitative or qualitative variables.
- Example: Consider a possible relationship between salary, years of employment and gender. This model might be formulated as:

$$\text{Salary} = \beta_0 + \beta_1 \text{ years employed} + \beta_2 \text{ Gender}$$

Cont.

b. Time Series, Cross Sectional and Panel Data

- Cross section data is data collected for multiple entities at one point in time
- Each observation is a new individual, household, firm, etc.. With information at a point in time.
- **Examples:** Data on expenditures, income, hours of work, household composition, assets, investments, employment, etc..
- Cross sectional data is usually a random sample of the underlying population.
  - E.g. Randomly select 500 people from the population of all working people in Ethiopia.
  - Ordering of observations does not matter.
- If the data is not a random sample, we have a sample-selection problem.

Cross sectional data

- Cross section data is data collected for multiple entities at one point in time
- Each observation is a new individual, household, firm, etc.. With information at a point in time.
- **Examples:** Data on expenditures, income, hours of work, household composition, assets, investments, employment, etc..
- Cross sectional data is usually a random sample of the underlying population.
  - E.g. Randomly select 500 people from the population of all working people in Ethiopia.
  - Ordering of observations does not matter.
- If the data is not a random sample, we have a sample-selection problem.

A cross-sectional dataset on wages and other individual characteristics

Obs	Wage	Educ	Exper	Female	Married
1	3.10	11	2	1	0
2	3.24	12	22	1	1
3	3.00	11	2	0	0
4	6.00	8	44	0	1
5	5.30	12	7	0	1
⋮	⋮	⋮	⋮	⋮	⋮
525	11.56	16	5	0	1
526	3.50	14	5	1	0

Time series

- Time series data is data collected for a single entity at multiple points in time
- Separate observation for each time period. Ordering of observations does matter.
- Typically Macroeconomic measures: Yearly GDP of Ethiopia for a period of 20 years, quarterly data on Inflation and unemployment rate in Ethiopia from 1970-2011, Prices, daily Birr/US dollar exchange for the past year, Interest Rates, etc..
- Financial data: Stock Prices, Bonds and other financial instruments at frequencies that range from minute to minute up to annual.
- Since not a random sample, different problems to consider.
- Trends and seasonality will be important. (E.g.: monthly ice cream sales; wheat production.)

Time series dataset on minimum wage, unemployment & related variables for Puerto Rico

Obsno	year	avgmin	avgcov	unemp	gnp
1	1950	0.20	20.1	15.4	878.7
2	1951	0.21	20.7	16	925
3	1952	0.23	22.6	14.8	1015.9
⋮	⋮	⋮	⋮	⋮	⋮
37	1986	3.35	58.1	18.9	4281.6
38	1987	3.35	58.2	16.8	4496.7

Panel (longitudinal) data

- Panel data is data collected for same multiple entities at multiple points in time
- Has both cross sectional and time series dimensions
- Similar to pooled cross sections with one important difference.
  - Responses related to the same cross-sectional unit (individual, firm, country) over time.
  - Consists of time series for each cross-sectional unit.
- Allows to control for unobserved characteristics of the cross-sectional unit (e.g. ability).
- Data is usually structured by cross sectional unit over time.
  - All observations on unit 1 over time, then observations on unit 2 over time, …

Panel data

country	year	Y	X1	X2	X3
1	2000	6.0	7.8	5.8	1.3
1	2001	4.6	0.6	7.9	7.8
1	2002	9.4	2.1	5.4	1.1
2	2000	9.1	1.3	6.7	4.1
2	2001	8.3	0.9	6.6	5.0
2	2002	0.6	9.8	0.4	7.2
3	2000	9.1	0.2	2.6	6.4
3	2001	4.8	5.9	3.2	6.4
3	2002	9.1	5.2	6.9	2.1

Cont.

c. Experimental versus Non-experimental Data

- This classification is based on methods for collection

1) Experimental data: are collected through active intervention by the researcher to produce and measure change or to create difference when a variable is altered.

- Experimental data typically allows the researcher to determine a causal relationship and is typically projectable to a larger population.
- common in natural sciences, but experimental data are becoming more commonly used in economics

2) Non-experimental data (Observational data):- are data drawn from a system not subject to experimental control.

- they are captured through observation of a behavior or activity.
- collected using methods like human observation, open-ended surveys
- typical in the social sciences

d. Nominal, ordinal and scale variables

- Nominal. A variable can be treated as nominal when its values represent categories with no intrinsic ranking

  For example,
  - Department of the company in which an employee works
  - Region, religious affiliation,.....

- Ordinal. A variable can be treated as ordinal when its values represent categories with some intrinsic ranking

  For example,
  - Levels of satisfaction in a certain service giving organization
  - Grade, Attitude scores, preference rating scores,... .

- Scale. A variable can be treated as scale when its values represent ordered categories with a meaningful metric, so that distance comparisons between values are appropriate.
  - Examples of scale variables include weight in kg, age in years , income in thousands of dollars.

Why econometrics?

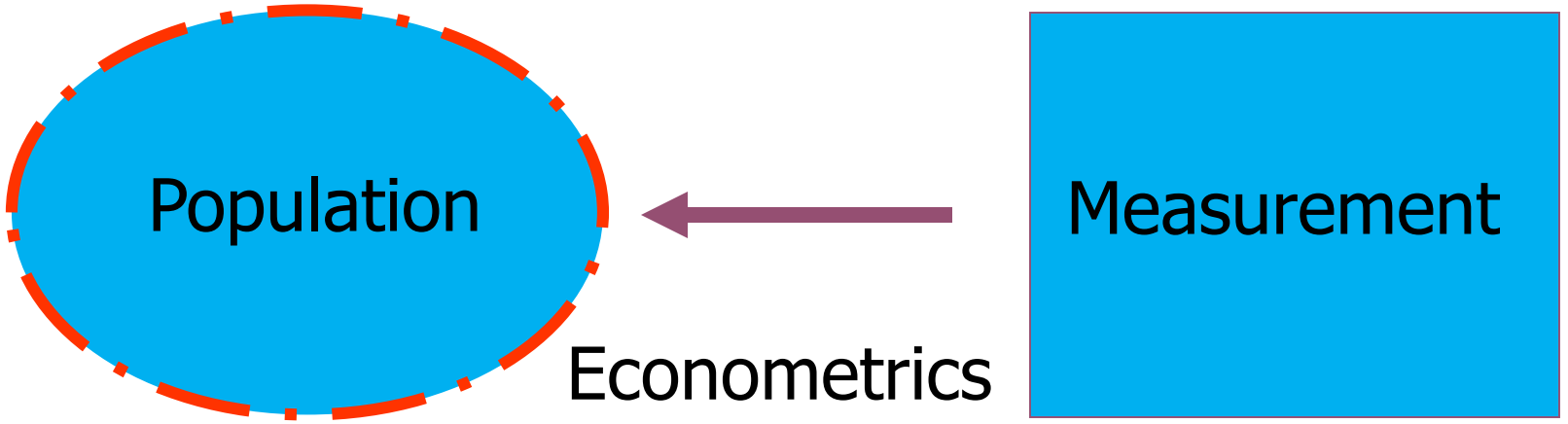
- Economic data is (in the main) non-experimental and hence non-deterministic.
  - It is observed from the world around us; they are outcomes of uncontrolled experiments
  - It is not obtained from 'test tube' experiments.
  - This means we can rarely 'hold constant' all factors in the economic model.
  - This means that all variables must be treated as random

E.g.: What is the effect of an extra year of education on wages, holding all other factors constant (ability, age, sex, experience, industry etc.)?

- It is impossible to hold these factors constant in reality. (We cannot force people to undergo or not undergo extra education. And we certainly cannot keep age constant while adding another year of education...)
- Even if it was possible, it would be unethical in many cases. E.g.: What is the effect of a reduction in unemployment benefits on the probability of finding a job?

1.5 Econometrics: Paradigm
Classical Inference versus Bayesian Inference

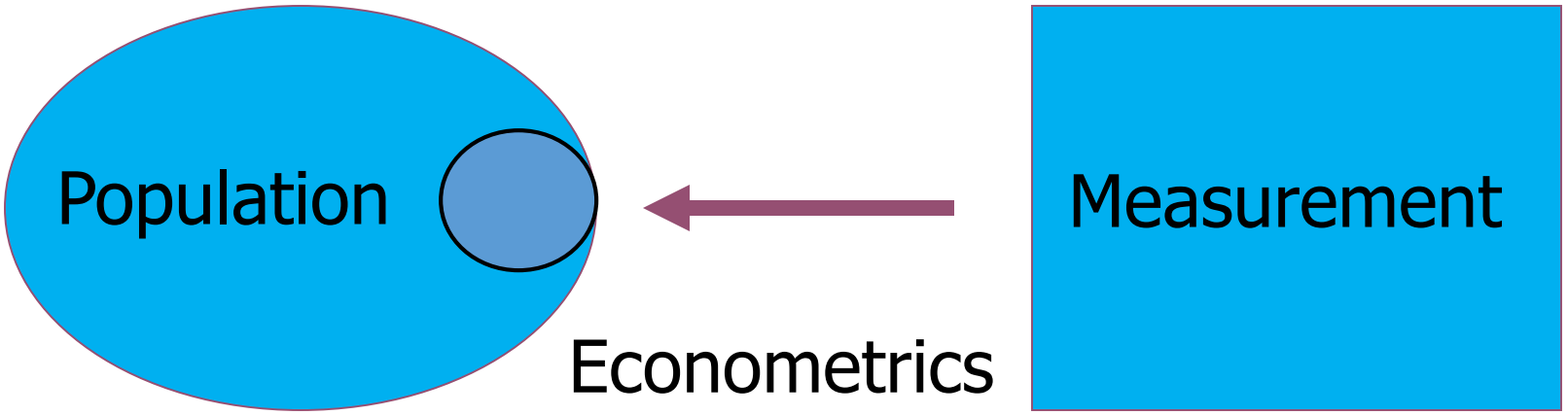
Paradigm: Classical Inference



Imprecise inference about the entire population – sampling theory and asymptotics

Characteristics
Behavior Patterns
Choices

Paradigm: Bayesian Inference



Sharp, 'exact' inference about only the sample – the 'posterior' density.

Characteristics
Behavior Patterns
Choices

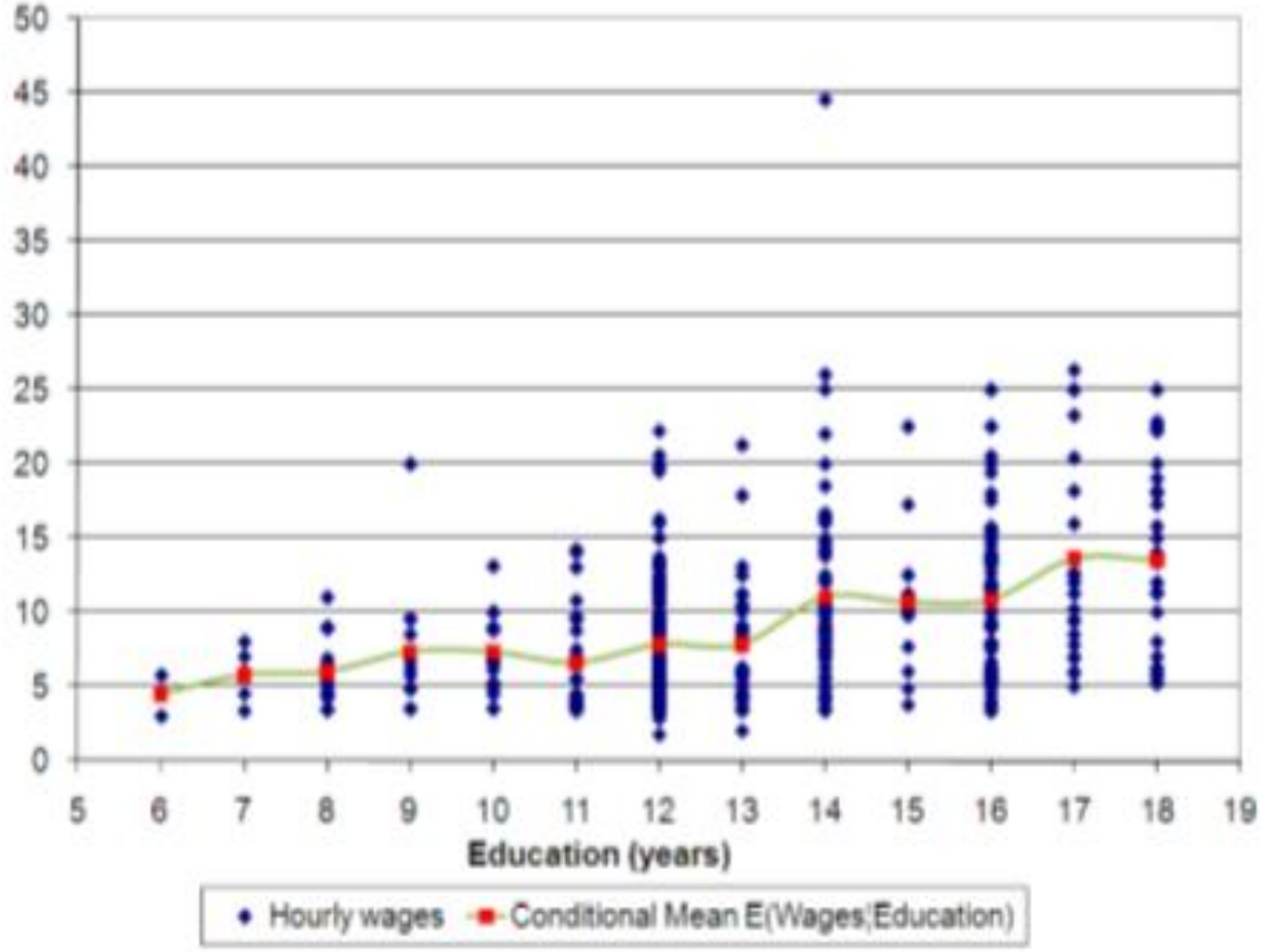
1.6 Regression analysis

- regression analysis is concerned with the conditional expectation (the average value) of the dependent (response) variable given the explanatory variables
- Less commonly, the focus can also be on a quantile, or other location parameter of the conditional distribution of the dependent variable given the explanatory variables.
  - For the case of a conditional quantile, one usually calls it the quantile regression function.
  - In all cases, a function of the explanatory variables is loosely called the regression function.
  - Conditional mean analysis or regression analysis is one of the most popular statistical methods in econometrics

cont

- The term "regression" was coined by Galton (1877, 1885) in the nineteenth century to describe a biological phenomenon. The phenomenon was that the heights of descendants of tall ancestors tend to regress down towards a normal average, a phenomenon also known as regression toward the mean.

- **Conditional Mean Function**: Hourly Wages and Education



- **Hence, the object that we are interested in studying is the mean of wages given the years of Education: E[Wages|Education].**

1.7 Methodology in (financial) econometrics

Stages of econometric analysis:

1. Model formulation (specification)
   - What is the right model to use to analyze economic relationship?
2. Model estimation
3. Model evaluation (diagnosis)

1. Formulating a model

- Identifying an econometric model (population regression model) which is a complete specification of economic and statistical behavior. EX: $Y = \beta_0 + \beta X + \boldsymbol{\varepsilon}$
- In this stage, there are three important tasks to fulfill:

A) Developing hypothesis

- The dependent and independent (explanatory) variables which will be included in the model.
- We form *a priori* theoretical expectations about the ***size*** and ***sign*** of the relationship between variables (parameters of the function).
- Depends on theory, experience, other studies, 'Common sense'

Cont.

B) Specification of the mathematical model of the hypothesis.

- Expressing the hypothesized relationship in mathematical form with which economic phenomena will be explored empirically.
- A mathematical model expresses the relationship in deterministic (exact) way

C) Specification of the econometrics model of the hypothesis.

- Expressing the hypothesized relationship in econometric model
- An econometric model expresses the relationship in stochastic (inexact) way

- An econometric model is a complete specification of economic and statistical behavior.
- We include the stochastic component ε in the mathematical (deterministic) model to form an econometric (stochastic) model

Formulating a model: Example

❑Example: A study on returns to education

A model of human capital investment predicts that getting more education should lead to higher wages.

❑**Formulating a model (first stage in econometrics)**

- Hypothesis: more education leads to higher wages
- Mathematical model

  Wages = $\beta_0 + \beta_1$Education (exact relationship)

- Econometric model

  wages = $\beta_0 + \beta_1$ Education + $\boldsymbol{\varepsilon}$ *(inexact relationship)*

  ➢ε *is called* the disturbance (error, stochastic) term

  ➢$\beta_0 + \beta_1$ Education is called the systematic component and $\boldsymbol{\varepsilon}$ *the random component*

a. Stochastic (econometric) versus non-stochastic (deterministic) relationships

- **A relationship between X and Y**, characterized as Y = f(X) is said to be **non-stochastic** if ***for each value of the independent variable*** (X) there is ***one and only one corresponding value*** of dependent variable (Y). That is

$$Y_i = \alpha + \beta X_i \text{ -------} -(1)$$

- **A relationship between X and Y is said to be** stochastic if for a particular value of X there is a whole probabilistic distribution of values of Y.

$$Y_i = \alpha + \beta X_i + u_i \text{ -------} -(2)$$

b. The Disturbance (Error) Term, *ε or u*

- Disturbance term plays a pivotal role in econometrics.
- It accounts for all the factors that affect the dependent variable that we have not explicitly accounted for in our regression function.
  - Omission of variables from the function. For example, education is not the only determinants of wage.
  - Randomness of human behavior
  - Unavailability of data
  - Wrong or imperfect specification of the functional form
  - Errors of aggregation
  - Errors of measurement

c. Population regression versus sample regression

i) Population regression

- The true econometric model that relates the dependent variable (variable to be explained) with the variable (s) that are expected to explain this effect is called the population regression model (PRM)
- The true function that relates the conditional mean of the dependent variable with the independent variable (s) is called the population regression function (PRF)

Cont.

- Population regression curve is the locus of the conditional means or expectations of the dependent variable for the fixed values of the explanatory variable X
- PRF tells us how the average value of y changes with x. It does not say that $y = \beta_0 + \beta_1 X$ for everyone in the population. It refers to the **average tendency.**
  - Some observations will be below the average and some will be above the average, depending on the influence of u in each specific case.

- Example: $Y = \beta_0 + \beta_1 X + \varepsilon$ *(PRM)*

$Y = E(Y|X) + \varepsilon$ *(regression identity)*

$E(Y/X) = \beta_0 + \beta_1 X$ population regression function

$E(\varepsilon/X) = 0$ expected value of the error term

- Population regression model is a true model but not directly observable or not feasible to compute.

- The random variable ε represents the part of Y that is not captured by E(Y |X). It is usually called a noise or a disturbance, because it "disturbs" an otherwise stable or deterministic relationship between Y and X. On the other hand, the regression function E(Y |X) is called a signal.
- The property that E(ε|X) = 0 implies that the regression disturbance ε contains no systematic information of X that can be used to predict the expected value of Y. In other words, all information of X that can be used to predict the expectation of Y has been completely summarized by E(Y |X).
- The condition E(ε|X) = 0 is crucial for the validity of economic interpretation of model parameters

Parameters and variables

- There are two types of objects in this PRM:
  - Variables
  - Parameters
- A **variable** is a factor, trait or condition which assumes a variety of values in a particular problem (one whose values is not known yet)
  - There are dependent and independent variables in a model.
- **Parameter**:- is a quantity that characterizes a population (true) relationship and that can be estimated from sample data.
  - parameters are also called constants/unknowns/coefficients

- Intercept parameters are parameters which describe the value that the dependent variable will take when independent variables are all equal zero
- Slope parameters (also called slope coefficients) are parameters that describe the impact that independent variable has on the dependent variable. For example, β_1 in the wage model shows the effect on Wages for one additional year of Education.
- The statistical & economic significances are important

- Example: $Y = \beta_0 + \beta_1 X + \varepsilon$
  - Y is dependent variable
  - X is independent variable
  - β_0 is the intercept term
  - β_1 is the slope coefficient

ii) sample regression function (SRF)

- The PRF ($E(Y/X)=\beta_0 + \beta_1 X$) is a true model but not directly observable or not feasible to compute.
  - We therefore estimate it from the SRF

- **Example:**

- Imagine we wish to find the relationship between father's height and son's height.
- Imagine we had data on ALL the world's father's and son's heights.
- Then we can get the **Population Regression Function**!

$$\text{son's height}_i = \alpha + \beta(\text{father's height})_i + \varepsilon_i$$

- But we can't observe the heights of ALL fathers and sons in the world throughout time.
- So we have to do with a sample!

- As we do not have the luxury to have data on the population, we rely on the sample
- SRF is the sample counterpart of PRF:

where

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x \qquad \text{sample regression function}$$

$$Y = \hat{Y} + \hat{\varepsilon} = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\varepsilon}$$

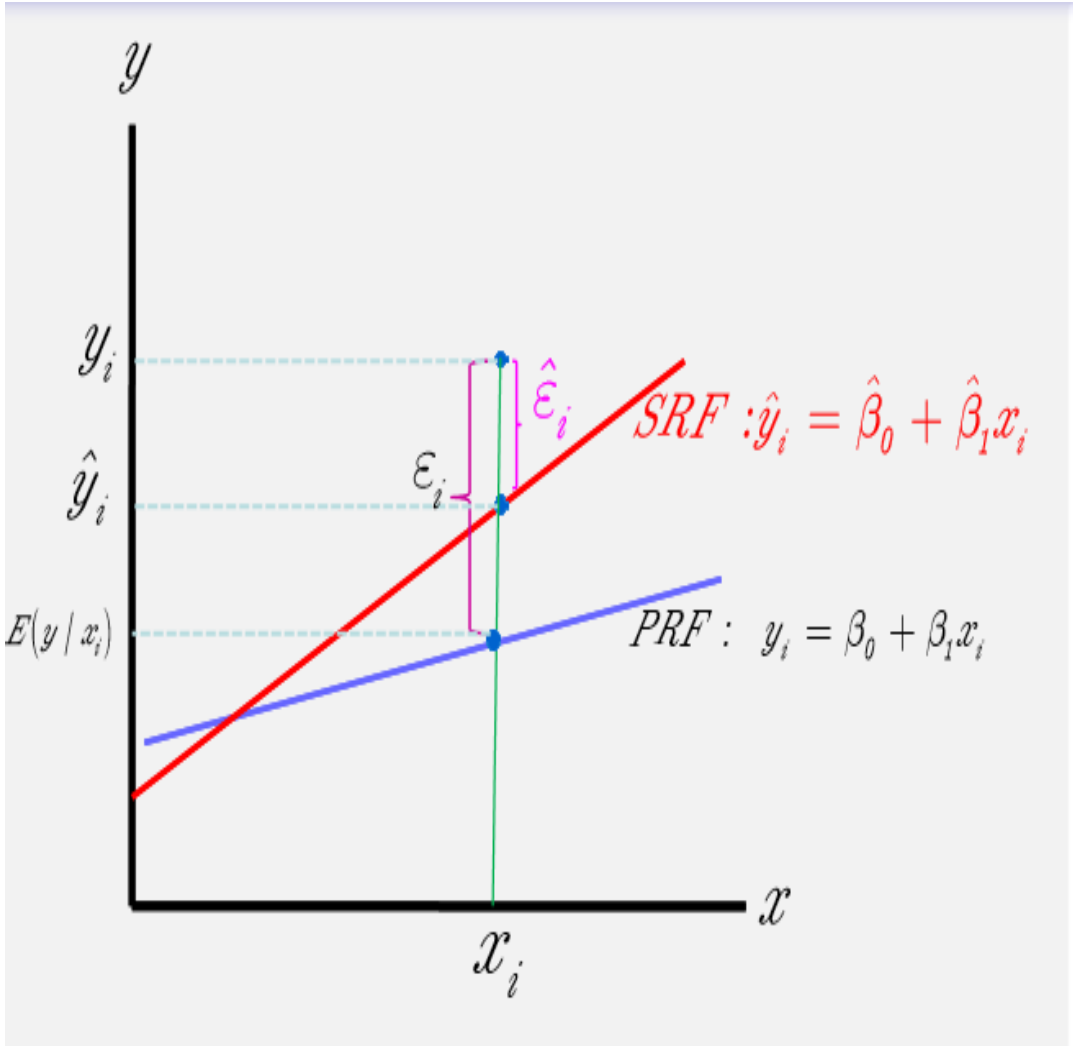
$\hat{Y}$ is called fitted/predicted value of actual value Y

$\hat{\beta}_0$ = estimator of β_0

$\hat{\beta}_1$ = estimator of β_1

$\hat{\varepsilon} = Y - \hat{Y} = actual\ value - predicted\ value$ = estimator of ε & *is called residual.*

- Are results of the SRF the same as that of the PRF? The answer is no due to sampling fluctuation
- The task is to make sure that the estimators are as close as the population parameters

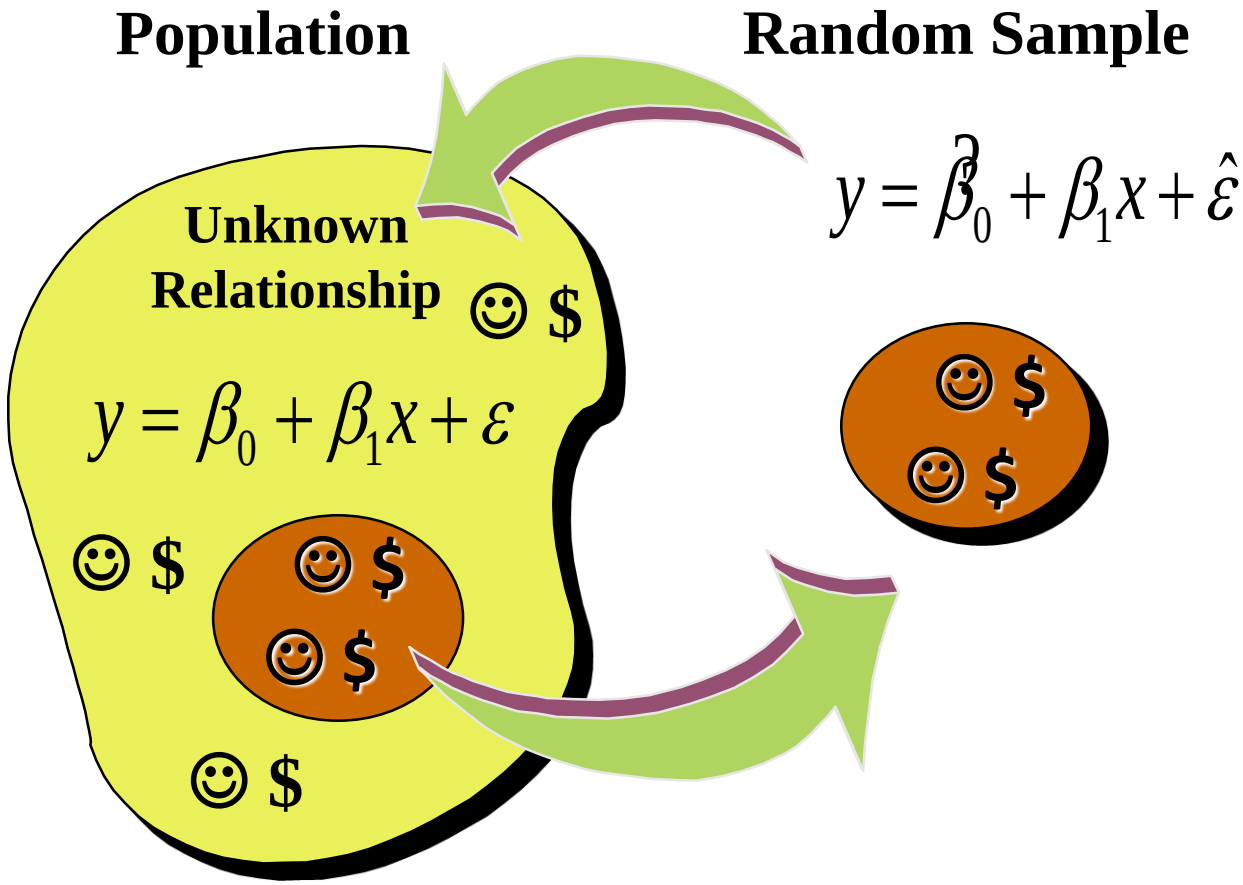


Sample and population regression lines

PRF shows population average tendency whereas

SRF sample average tendency

Population & Sample Regression Models (classical regression)



2. Model estimation (second stage)

- Model estimation refers to estimating (calculating) the values of parameters based on empirical data that has a random component
  - Parameters describe an underlying physical setting but they are unknown and hence should be estimated

- **Two tasks:**

1. Estimating the model using appropriate econometric techniques (estimators).
2. Collection of data on the variables of the model

Cont.

- The choice of estimator (estimation method) depends on:
  a) type of data we use in our study;
  b) nature of the dependent variable
  c) functional form
- We use an **estimator** to calculate parameter **estimates**.
  - An estimator is a **formula** for calculating parameter estimates or for estimating parameters.
  - An estimate is a **numerical value** we obtain after substituting data/sample in the estimator.

- **Types of estimators:**
  - ***Ordinary least squares method***
  - ***Method of moments***
  - ***Maximum likelihood method***

- We typically use Greek letters such as $\beta, \theta \text{ and } \sigma^2$ to denote unknown parameters of an econometric model
- Estimators are typically denoted by putting a hat ^, tilde ~or bar - over the corresponding letter, e.g., $\hat{\beta} \, and \, \tilde{\beta}$ are estimators of β
- Estimators work in relevant environment which is determined by assumptions.

3. Model evaluation (third stage)

- The process of checking the adequacy of the model against a range of criteria and possibly returning to the model formulation stage.
- The evaluation consist of deciding whether the estimates of the parameters are theoretically meaningful and statistically satisfactory.
- The evaluation of the model is based on economic a priori criteria, statistical criteria, econometric criteria and the forecasting ability of the model.

1. Economic a priori criteria:

• These criteria are determined by economic/financial theory and refer to the size and sign of the parameters of economic relationships.

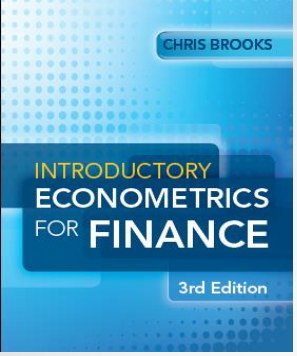
Cont.

ii. Statistical criteria (first-order tests):

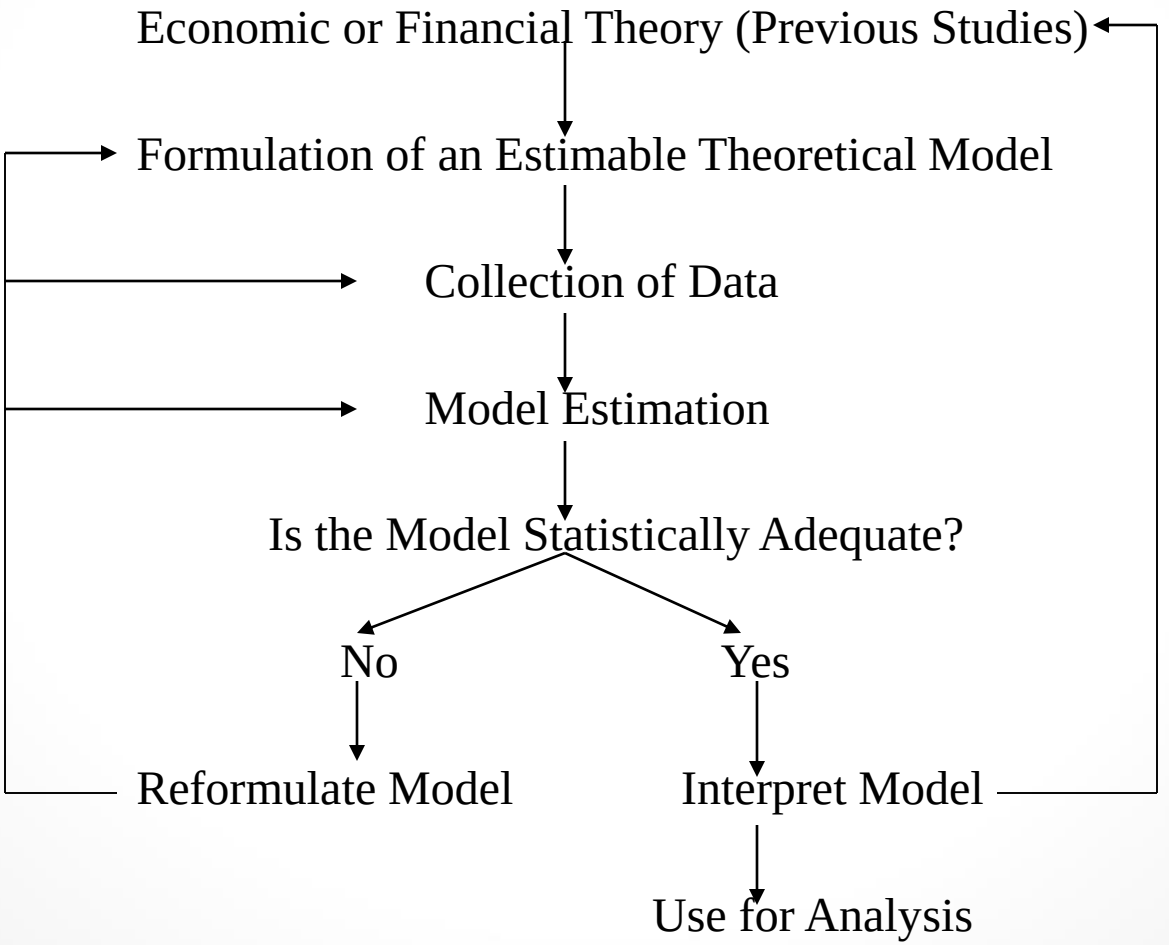
- These are determined by statistical theory
- aim at the evaluation of the statistical reliability of the estimates of the parameters of the model.
- Correlation coefficient test, standard error test, t-test, F-test, and R^2-test are some of the most commonly used statistical tests.

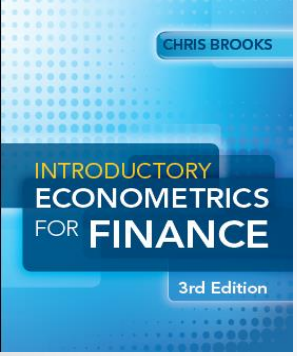
iii. Econometric criteria (second-order tests):

- · aim at the detection of the violation or validity of the assumptions of the various econometric techniques.
- · They serve as a test of the statistical tests i.e. they determine the reliability of the statistical criteria; they help us establish whether the estimates have the desirable properties



Summary: Steps of an Empirical Study



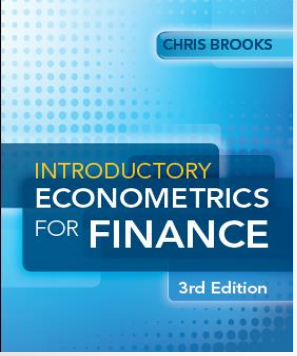


Some Points to Consider when reading papers in the academic finance literature

1. Does the paper involve the development of a theoretical model or is it merely a technique looking for an application, or an exercise in data mining?

2. Is the data of "good quality"? Is it from a reliable source? Is the size of the sample sufficiently large for asymptotic theory to be invoked?

3. Have the techniques been validly applied? Have diagnostic tests been conducted for violations of any assumptions made in the estimation of the model?



Some Points to Consider when reading papers in the academic finance literature (cont'd)

4. Have the results been interpreted sensibly? Is the strength of the results exaggerated? Do the results actually address the questions posed by the authors?

5. Are the conclusions drawn appropriate given the results, or has the importance of the results of the paper been overstated?

Formulating a model: Example

❑Example: A study on returns to education

A model of human capital investment predicts that getting more education should lead to higher wages.

❑**Formulating a model (first stage in econometrics)**

- Hypothesis: more education leads to higher wages
- Mathematical model

  Wages = $\beta_0 + \beta_1$Education (exact relationship)

- Econometric model

  wages = $\beta_0 + \beta_1$ Education + $\boldsymbol{\varepsilon}$ *(inexact relationship)*

  ➢ε *is called* the disturbance (error, stochastic) term

  ➢$\beta_0 + \beta_1$ Education is called the systematic component and $\boldsymbol{\varepsilon}$ *the random component*

Cont.

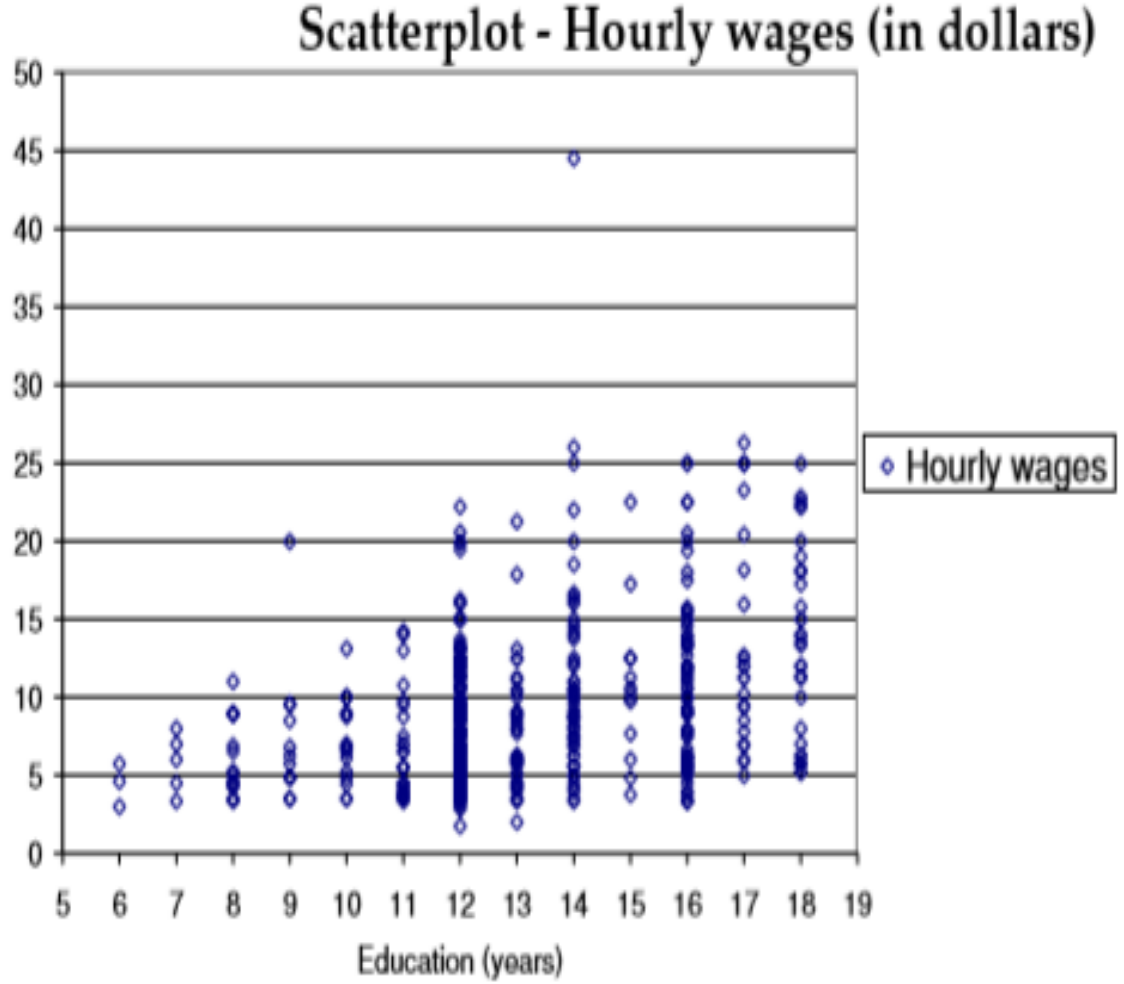
- How can we study if the evidence of the data supports economic theory (the model of human capital investment)?
- How to approach this question?

- Let us look at the following data set as example :

  US national survey of 528 people in the labor force that already completed their education.

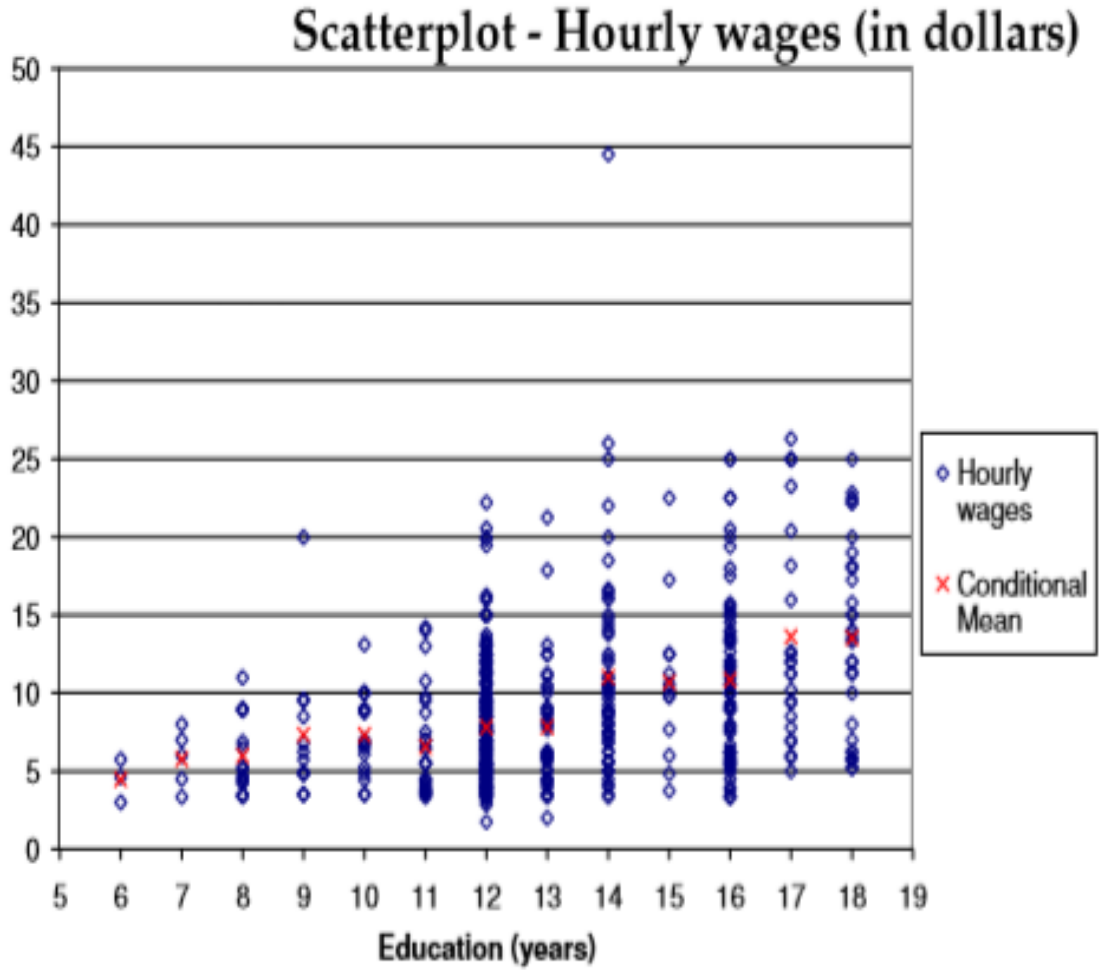
- Scatterplot of this dataset is drawn on next page

Formulating a model

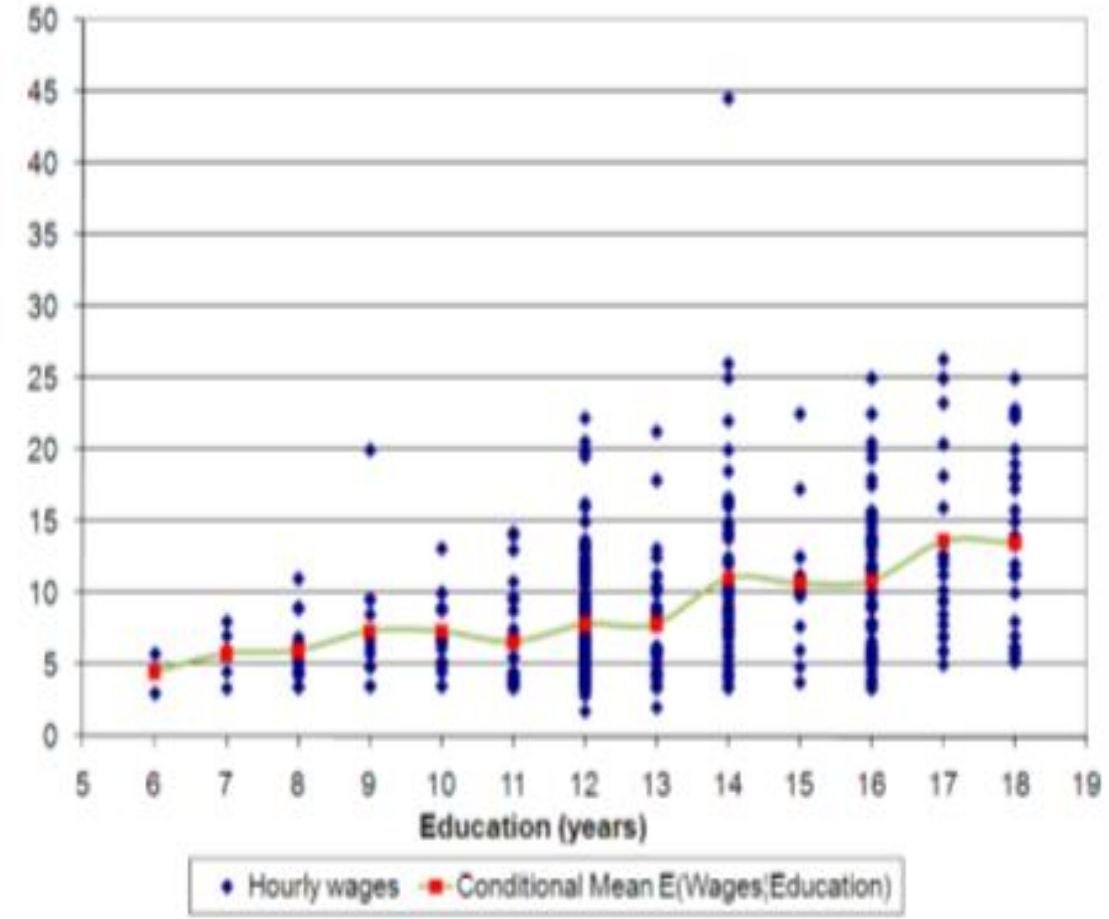


- People with the same years of education have different hourly wages.
- There is a distribution for the hourly wages conditional on the years of education.

- How can we study if the evidence of the data supports Economic Theory?
- A possibility is to look at means of wages conditional on the years of Education.



- **Conditional Mean Function**: Hourly Wages and Education



- We can see that the mean of wages vary with the years of Education.
- **Hence, the object that we are interested in studying is the mean of wages given the years of Education: E[Wages|Education].**

Exercise

Discuss the econometric stages we follow in addressing the following issues:

1. The effect of advertising on sales