INTRODUCTORY ECONOMETRICS SUMMARY 2011

BUSINESSKOOL.COM

BUSINESSKOOL PTY LTD

TABLE OF CONTENTS

Sources used 4

Topic 1: The nature of Econometrics and Economic data 5
1.1 The nature of Econometrics 5
1.1.1 Introduction 5
1.1.2 Causality 5
1.2 Economic Data 6
1.2.1 Cross sectional data 6
Definition: To be an **independent** observation, it must be drawn without influence from another variable. 6
1.2.2 Time series Data 6
1.2.3 Pooled Cross Sections 7
1.2.4 Panel or longitudinal data 7
Topic 2: Simple regression model 7
2.1 Zero Conditional Mean 8
2.2 Regression functions 8
Population regression function 8
Sample regression function 9
2.3 Estimators 9
2.4 Simple regression model statistics 10
2.4.1 Sum of squares 10
2.4.2 Goodness of fit 11
2.5 Functional forms 11
2.6 Assumptions of the simple linear regression model 12
2.7 Variance of OLS estimators 13
Topic 3: Multiple regression model 14
3.1 Assumptions of the multiple regression model 14
3.2 Irrelevant variables 15
3.3 Omitted variable bias 15
Topic 4: Inference 16
4.1 Classical linear model 16
4.2 Hypothesis tests 16
4.2.1 t-tests 16
4.2.2 p-value 18

SAMPLE ONLY

4.2.3 Confidence interval 18
4.2.4 Single Linear combinations (F-tests) 18
Topic 5: Multiple regression with qualitative information 21
5.1 Single dummy independent variables 21
5.2 Comparison of means test 21
5.3 Dummy variables for multiple categories 22
5.4 Interaction effects 22
Topic 6: Heteroskedasticity 23
6.1 Consequences of Heteroskedasticity for ordinary least squares 23
6.3 Testing for Heteroskedasticity 23
6.3.1 Breusch-pagan test for Heteroskedasticity 23
6.3.2 White test for heteroskedasticity 24
6.4 Correcting for heteroskedasticity 24
6.4.1 Feasible generalised least squares 24
6.4.2 Robust standard errors method 25
6.4.3 Weighted least squares 25
Topic 7: Specification issues 26
7.1 Functional form misspecification 26
7.2 Testing for misspecification 26
7.2.1 RESET test 26
7.2.2 Proxy variable 27
7.2.3 Used lagged dependent variables as proxy variables 27
7.3 Measurement error 27
7.3.1 Measurement error in the dependent variable 27
7.3.2 Measurement error in the explanatory variable 28
7.4 Data problems 28
7.4.1 Missing data 28
7.4.2 Non-random samples 28
7.4.3 Outliers 29

SAMPLE ONLY

SOURCES USED

- William H, Greene, *Econometrics analysis, 5th ed.* Upper Saddle River, New Jersey: Pearson Education Inc., 2003
- Damodar N, Gujarati, *Basic Econometrics, 4th ed.* New York: McGraw-Hill 2003
- Jefferey M, Wooldridge, *Introductory Econometrics: A modern approach, 3rd ed.* Mason OH: Thomson Higher Education, 2006.

SAMPLE ONLY

Topic 1: The nature of Econometrics and Economic data

1.1 The nature of Econometrics

1.1.1 Introduction

Econometrics concerns the use of statistical methods in:

1. Estimating economic relationships
2. Testing economic theory
3. Evaluating government and business policy.

Empirical analyses generally consist of:

1. **An economic model** – which may be formally developed (e.g. derivation of consumer demand equations from a model of utility maximization) or based on intuitive reasoning
2. **An econometric model** – which requires specifying the nature of the relationship between variables.

SAMPLE ONLY

Definition: A **variable** is a symbol that may assume any given value or set of values. Often in econometrics this can represent factors such as age, height, gender etc…

1.1.2 Causality

In most tests of economic theory and certainly for evaluating public policy, the economist's goal is to infer that one variable has a casual effect on another variable. Simply finding an association between two or more variables might be suggestive but unless causality can be established it is rarely compelling.

Ceteris Paribus is a Latin phrase which literally means "other relevant factors being equal". Predictions in econometrics must be *ceteris paribus* in order to rule out the possibility that other factors could influence the observed relationship.

1.2 Economic Data

1.2.1 Cross sectional data

A cross sectional data set consists of a sample of individuals, households, firms or other units taken at a given point of time. In pure cross-sectional data we would ignore any minor timing differences in collecting the data.

An important feature of the data is that we can often assume they have been obtained by random sampling. Sometimes it is not appropriate to assume random sampling, such as when people do not disclose the right information; it would not be a random sample.

Definition: **A random sample** is which a sample is collected from a population in such a way that every possible sample that could be selected has a predetermined probability of being selected.

SAMPLE ONLY

There is also a violation of random sampling when the sample is large relative to the population as the population would not be large enough to assume the observations are independent draws.

Definition: An **observation** is a piece of data taken from a particular unit, time or place in a sample

Definition: To be an **independent** observation, it must be drawn without influence from another variable.

1.2.2 Time series Data

A time series data set consists of observations on a variable or several variables over time e.g., stock prices, money supply, GDP. Because past events can influence events and lags in behaviour are prevalent in the social sciences, time is an important dimension in a time series data set. Order conveys important information in a time series data set.

They are difficult to analyse because they can rarely be assumed to be independent across time, another feature that requires special attention is the data frequency at which the data is collected, which will show seasonal trends.

1.2.3 Pooled Cross Sections

Have both cross-sectional and time series features. It combines two sets of cross-sectional data. It increases the sample size by combining the two years and is an effective way of analysing the effects of a new government policy.

The order which we store the data is not important, but keeping track of the year is very important. In addition of increasing the sample size, the point of a pooled cross-sectional analysis is often to see how a key relationship has changed over time

SAMPLE ONLY

1.2.4 Panel or longitudinal data

A panel data set consists of a time series for each cross-sectional member in the data set. The key feature of panel data that distinguishes them from a pooled cross section is the fact that the same cross-sectional units are followed over a given time period. The ordering in the cross section of a panel data set does not matter.

Panel data allows us to control certain unobserved characteristics, and secondly it also allows us to study the behaviour or the result of decision making.

♠♠♠♠

To order the complete version of the Businesskool Introductory Econometric Summary please visit www.businesskool.com