

Contents

1	Introduction and Motivation	3
1	What is econometrics?	3
1.1	Steps in Empirical Economics Analysis	5
2	Types of economic data	6
2	Review of Some Statistical Concepts	7
1	Random variables and distributions	7
1.1	Types of Random Variables	8
1.2	Distribution of a Random Variable (X)	9
1.3	Examples of Some Distributions	10
1.4	Population Distribution vs. Sample Distribution	11
2	Joint Distributions, Conditional Distributions and Independence	12
2.1	Joint Distributions	12
2.2	Conditional Distribution and Marginal Distribution	13
2.3	Independence	16
3	Expectation, variance, covariance and correlation	17
3.1	Expected value of a Random Variable – $E(X)$ or μ_X	17
3.2	Properties of Expected Values	18
3.3	Conditional and Marginal Expectations	19
3.4	Variance and Standard Deviation of a Random Variable	20
3.5	Properties of Variances	21
3.6	Properties of Standard Deviations	21
3.7	Covariance and Correlation of Two Variables	23
4	Estimators and desirable properties of estimators	25
4.1	Desirable properties of estimators	26
4.2	The concept of Mean Squared Error (MSE)	29
3	The Simple Regression Model	31
1	Principle, assumptions and derivation of ordinary least squares (OLS) estimators	32
1.1	Terminology for the Linear Regression	32
1.2	Derivation of Ordinary Least Squares (OLS) Estimators	33
2	Properties of OLS estimators	36
2.1	Algebraic Properties	36
2.2	Properties proving BLUE	37
2.3	Assumptions on the simple linear regression (SLR) model	37
2.4	Homoskedasticity VS. Heteroskedasticity	38
2.5	Expectation of Estimators	39
2.6	Variance of OLS Estimators	40
2.7	Some Concepts to be emphasized (Population vs. Sample; PRF vs. SRF; error vs. residual)	41

2.8	Goodness of Fit (R^2)	43
2.9	Incorporating Nonlinearities in Sample Regression	44
3	Regression Through the Origin	45
4	Multiple Regression Analysis (Estimation)	47
1	Motivation	47
1.1	Assumption SLR 4 ($E(u X) = 0$) in the Multiple Regression Context	48
2	The Model with k Independent Variables	49
2.1	Accounting Nonlinearity	50
3	Estimation of parameters and properties of estimators	51
3.1	Deriving OLS Estimators	51
3.2	Algebraic Properties of OLS estimators	51
3.3	How could the multiple regression analysis enable ceteris paribus analysis?	52
4	Expected Value of the OLS Estimators	53
4.1	Issue #1: Including Irrelevant Variable (Overspecifying the Model)	53
4.2	Issue #2: Excluding Relevant Variable (Underspecifying the Model → omitted variable bias. This is a serious problem!)	54
5	Variance of the OLS Estimators	55
6	Estimator of the OLS Variance	56
5	Stata Lab 1 – Introduction	57
1	What is STATA?	57
1.1	STATA supports	57
1.2	Data files and Do-files	57
2	Tutorial 1: Exploring the Data and Running a Simple Regression	58
2.1	Examples from Wooldridge(2009)	60
6	Matrix Algebra (Quick Review)	61
1	The notation	61
2	Matrix Operation	62
3	OLS Estimation using Matrix Notation	64
4	Variance-Covariance Matrix of $(\hat{\beta})$	66
7	Multiple Regression Analysis (Inference)	67
1	Sampling Distribution of the OLS estimators ($\hat{\beta}_{OLS}$)	68
2	Testing Hypotheses about an individual regression coefficient "the t-test"	69
2.1	Testing Against Two-Sided Alternatives	69
2.2	Testing Against One-Sided Alternatives	71
3	Testing other hypotheses about β_j	72
4	Testing Hypotheses about a Single Linear Combination of the Parameter	73
5	Computing p-Values for t-Tests	74
6	Confidence Intervals (CI)	75
7	Testing Multiple Linear Restrictions: The F-test	77
8	How the Hypothesis Testing is done in Practice	80
8	Multiple Regression Analysis : Further Issues	81
1	Data scaling on OLS statistics	81
2	More on functional forms	82

3	Models with Interaction Terms	84
4	More on the Goodness-of-Fit and Selection of Regressors	85
9	Multiple Regression Analysis with Qualitative Information:	87
1	Outline	87
2	Describing Qualitative Information	87
3	Models with a single dummy independent variable	88
4	It is not possible to include all of the dummy alternatives in the same model	89
5	Using dummy variables for multiple categories	90
6	Interactions involving dummy variables	93
7	Testing for Differences in Regression Functions across Groups	95
	7.1 We can use the "Chow statistics" to test this type of hypothesis as well	96
8	A Binary Dependent Variable (y variable): The Linear Probability Model	98
10	Heteroscedasticity Problem	101
1	Nature and Consequences of heteroscedasticity for OLS	101
	1.1 Nature of Heteroskedasticity	101
	1.2 Consequences of Heteroskedasticity	101
	1.3 How can the estimated value of $Var(\hat{\beta}_{OLS})$ be wrong?	102
	1.4 Two types of remedies	103
2	Testing for heteroskedasticity	104
	2.1 Breusch-Pagan test (BP test)	105
	2.2 The White Test	108
3	Remedial measures	109
	3.1 Weighted Least Squares (WLS)	109
	3.2 Feasible GLS	110
	3.3 What if the assumed h_i function is wrong?	111
11	More on Specification and Data Issues	113
1	Functional Form Misspecification	113
	1.1 The RESET test	114
	1.2 Tests against Nonnested Alternatives	115
2	Using Proxy Variables for Unobserved Explanatory Variables	116
	2.1 Using Proxy variables	116
	2.2 Using Lagged Dependent Variables as Proxy Variables	116
3	OLS with Measurement Error	116
	3.1 Measurement Error in the Dependent Variable	117
	3.2 Measurement Error in the Explanatory Variable	117
4	Missing Data, Nonrandom Samples, Outlying Observations	118
	4.1 Missing Data	118
	4.2 Nonrandom Samples	119
	4.3 Outliers and Influential Observations	119
5	Models with Random Slopes	119
12	Basic Regression Analysis with Time Series Data	121
1	The Nature of Time Series Data	121
2	Examples of Time Series Regression Models	122
	2.1 Static Models	122

2.2	Finite Distributed Lag Models	122
3	Properties of OLS under classical assumptions	124
3.1	The variance of $\hat{\beta}_{OLS}$ (under ass.TS1-TS5)	125
13	Serial Correlation and Heteroskedasticity in Time Series Regressions	127
1	Properties of OLS with Serially Correlated Errors	127
2	Unbiasedness and Consistency	127
3	Efficiency and Inference	127
4	Testing for Serial Correlation	129
4.1	A "t-test" for AR(1) serial correlation with strictly exogeneous regressors	129
4.2	The Durbin-Watson Test (DW test)	130
4.3	Testing for AR(1) serial correlation "without" strictly exogenous regressors	131
4.4	Testing for AR(q) serial correlation "without" strictly exogenous regressors	132
5	Correcting for serial correlation	132
5.1	Passive way	132
5.2	Active way –	132
14	Lab 2 – Dummy, Heteroskedasticity, Specification Issues	133
1	Does "beauty" help increase wage?	133
2	Fixing Heteroskedasticity	135
3	Labor Force Participation of Female	136
15	Instrumental Variables Estimation and Two Stage Least Squares	139
1	Motivation	139
2	Consistency of the estimators	141
3	Properties of IV with a Poor Instrumental Variable	142
4	IV Estimation for the Multiple Regression Model	143
5	Two-Stage Least Squares (2SLS)	144
16	Maximum Likelihood Estimation	145
1	MLE estimator for μ	146
2	MLE estimator for σ	147
3	MLE estimator for β	147
17	Logit and Probit Models for Binary Response	149
1	The Models	149
2	Interpreting the Logit and Probit estimators	150
18	Stata Lab 3 – Instrumental Variables	153