

# EE 325 Introductory Econometrics<sup>1</sup>

By Wanwiphang Manachotphong<sup>2</sup>

Semester 1, 2018

<sup>1</sup>Based on: Wooldridge, J.M. (2016). Introductory Econometrics: A Modern Approach. 4th ed. Mason, Ohio: South-Western Cengage Learning.

<sup>2</sup>Email: wanwiphang@econ.tu.ac.th; Office: 525 Faculty of Economics, Tha Prachan; Office hours: Tuesdays, 11.00 – 12 noon (or by appointment)



# Contents

<b>1</b>	<b>Introduction and Motivation</b>	<b>3</b>
1	What is econometrics? . . . . .	3
1.1	Steps in Empirical Economics Analysis . . . . .	5
2	Types of economic data . . . . .	6
<b>2</b>	<b>Review of Some Statistical Concepts</b>	<b>7</b>
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 $\mu_X$ . . . . .	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
<b>3</b>	<b>The Simple Regression Model</b>	<b>31</b>
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
<b>4</b>	<b>Multiple Regression Analysis (Estimation)</b>	<b>47</b>
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
<b>5</b>	<b>Stata Lab 1 – Introduction</b>	<b>57</b>
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
<b>6</b>	<b>Multiple Regression Analysis (Inference)</b>	<b>61</b>
1	Sampling Distribution of the OLS estimators ( $\hat{\beta}_{OLS}$ ) . . . . .	62
2	Testing Hypotheses about an individual regression coefficient "the t-test"	63
2.1	Testing Against Two-Sided Alternatives . . . . .	63
2.2	Testing Against One-Sided Alternatives . . . . .	65
3	Testing other hypotheses about $\beta_j$ . . . . .	66
4	Testing Hypotheses about a Single Linear Combination of the Parameter	67
5	Computing p-Values for t-Tests . . . . .	68
6	Confidence Intervals (CI) . . . . .	69
7	Testing Multiple Linear Restrictions: The F-test . . . . .	71
8	How the Hypothesis Testing is done in Practice . . . . .	74
<b>7</b>	<b>Multiple Regression Analysis : Further Issues</b>	<b>75</b>
1	Data scaling on OLS statistics . . . . .	75
2	More on functional forms . . . . .	76
3	Models with Interaction Terms . . . . .	78
4	More on the Goodness-of-Fit and Selection of Regressors . . . . .	79
<b>8</b>	<b>Multiple Regression Analysis with Qualitative Information:</b>	<b>81</b>
1	Outline . . . . .	81
2	Describing Qualitative Information . . . . .	81

3	Models with a single dummy independent variable . . . . .	82
4	It is not possible to include all of the dummy alternatives in the same model	83
5	Using dummy variables for multiple categories . . . . .	84
6	Interactions involving dummy variables . . . . .	87
7	Testing for Differences in Regression Functions across Groups . . . . .	89
	7.1 We can use the "Chow statistics" to test this type of hypothesis as well . . . . .	90
8	A Binary Dependent Variable (y variable): The Linear Probability Model	92
<b>9</b>	<b>Heteroscedasticity Problem</b>	<b>95</b>
1	Nature and Consequences of heteroscedasticity for OLS . . . . .	95
	1.1 Nature of Heteroskedasticity . . . . .	95
	1.2 Consequences of Heteroskedasticity . . . . .	95
	1.3 How can the estimated value of $Var(\hat{\beta}_{OLS})$ be wrong? . . . . .	96
	1.4 Two types of remedies . . . . .	97
2	Testing for heteroskedasticity . . . . .	98
	2.1 Breusch-Pagan test (BP test) . . . . .	99
	2.2 The White Test . . . . .	102
3	Remedial measures . . . . .	102
	3.1 Weighted Least Squares (WLS) . . . . .	102
<b>10</b>	<b>Lab 2 – Dummy, Heteroskedasticity, Specification Issues</b>	<b>105</b>
1	The scaling issue. . . . .	105
2	Does "beauty" help increase wage? . . . . .	105
3	Labor Force Participation of Female . . . . .	107
<b>11</b>	<b>Serial Correlation and Heteroskedasticity in Time Series Regressions</b>	<b>109</b>
1	The Nature of Time Series Data . . . . .	109
2	Examples of Time Series Regression Models . . . . .	110
	2.1 Static Models . . . . .	110
	2.2 Finite Distributed Lag Models . . . . .	110
3	Properties of OLS under classical assumptions . . . . .	111
4	Properties of OLS with Serially Correlated Errors . . . . .	112
5	Unbiasedness and Consistency . . . . .	112
6	Efficiency and Inference . . . . .	113
7	Testing for Serial Correlation . . . . .	114
	7.1 A "t-test" for AR(1) serial correlation with strictly exogeneous regressors . . . . .	114
	7.2 The Durbin-Watson Test (DW test) . . . . .	115
8	Correcting for serial correlation . . . . .	116
	8.1 Passive way . . . . .	116
	8.2 Active way – . . . . .	116
<b>12</b>	<b>Multicollinearity</b>	<b>117</b>
1	The Nature of Multicollinearity . . . . .	117
2	Consequences of Multicollinearity . . . . .	117
	2.1 The OLS estimator will still be BLUE. . . . .	117
	2.2 The variances and covariances will be very large. This makes precise estimation difficult. . . . .	117

3	Detection of multicollinearity . . . . .	118
4	Remedial Measures . . . . .	119