Econometrics and data analysis for developing countries

Chandan Mukherjee, Howard White and Marc Wuyts





Contents

Lis Lis Lis Pre	t of t of t of eface	figures tables boxes	ix xiii xvi xvii		
Int	Introduction				
	1	The purpose of this book	1		
	2	The approach of this book: an example	3		
Par	rt I	Foundations of data analysis	21		
1	Мо	del specification and applied research	23		
	1.1	Introduction	23		
	1.2	Model specification and statistical inference	24		
	1.3	The role of data in model specification:			
		traditional modelling	29		
	1.4	The role of data in model specification:			
		modern approaches	32		
	1.5	The time dimension in data	39		
	1.6	Summary of main points	42		
2	Modelling an average				
	2.1	Introduction	44		
	2.2	Kinds of averages	45		
	2.3	The assumptions of the model	50		
	2.4	The sample mean as best linear unbiased			
		estimator (BLUE)	53		
	2.5	Normality and the maximum likelihood principle	58		
	2.6	Inference from a sample of a normal distribution	61		
	2.7	Summary of main points	71		
	Ap	pendix 2.1: Properties of mean and variance	73		
	Appendix 2.2: Standard sampling distributions				

vi	Contents

3	Outli 3.1	iers, skewness and data transformations Introduction	75 75
	3.2	The least squares principle and the concept	
		of resistance	76
	3.3	Mean-based versus order-based sample statistics	80
	3.4	Detecting non-normality in data	90
	3.5	Data transformations to eliminate skewness	97 197
	3.6	Summary of main points	106
Pa	rt II	Regression and data analysis	109
4	Data	analysis and simple regression	111 .
	4.1	Introduction	111
	4.2	Modelling simple regression	112
	4.3	Linear regression and the least squares principle	114
	4.4	Inference from classical normal linear	
		regression model	120
	4.5	Regression with graphics: checking the model	
		assumptions	124
	4.6	Regression through the origin	136
	4.7	Outliers, leverage and influence	137
	4.8	Transformation towards linearity	148
	4.9	Summary of main points	159
5	Parti	al regression: interpreting multiple regression coefficients	163
	5.1	Introduction	163
	5.2	The price of food and the demand for manufactured goods in India	165
	53	Least squares and the sample multiple regression line	173
	54	Partial regression and partial correlation	180
	55	The linear regression model	184
	5.6	The <i>t</i> -test in multiple regression	192
	57	Fragility analysis, making sense of	172
	5.7	regression coefficients	198
	5.8	Summary of main points	206
6	Mod	el selection and misspecification in multiple regression	208
	6.1	Introduction	208
	6.2	Griffin's aid versus savings model: the omitted variable bias	209
	63	Omitted variable bias: the theory	212
	6.4	Testing zero restrictions	219
	6.5	Testing non-zero linear restrictions	229
	6.6	Tests of parameter stability	231
	6.7	The use of dummy variables	237
	6.8	Summary of main points	246
	0.0	- · · · · · · · · · · · · · · · · · · ·	

Contents vii

Pai	t III	Analysing cross-section data	249	
7	Deali	ng with heteroscedasticity	251	
	7.1	Introduction	251	
	7.2	Diagnostic plots: looking for heteroscedasticity	252	
	7.3	Testing for heteroscedasticity	256	
	7.4	Transformations towards homoscedasticity	264	
	7.5	Dealing with genuine heteroscedasticity: weighted		
		least squares and heteroscedastic standard errors	270	
	7.6	Summary of main points	277	
8	Categories, counts and measurements			
	8.1	Introduction	279	
	8.2	Regression on a categorical variable: using	280	
	83	Contingency tables: association between	200	
	0.5	categorical variables	287	
	84	Partial association and interaction	207	
	85	Multiple regression on categorical variables	295	
	8.6	Summary of main points	298	
0	Logit	transformation modalling and rearrangian	202	
2		Introduction	302	
	0.2	The logit transformation	302	
	93	Logit modelling with contingency tables	307	
	94	The linear probability model versus logit regression	313	
	9.5	Estimation and hypothesis testing in logit regression	320	
	9.6	Graphics and residual analysis in logit regression	327	
	9.7	Summary of main points	331	
Pai	rt IV	Regression with time-series data	333	
10	Trend	s spurious regressions and transformations		
10	to sta	tionarity	335	
	10.1	Introduction	335	
	10.2	Stationarity and non-stationarity	335	
	10.3	Random walks and spurious regression	338	
	10.4	Testing for stationarity	349	
	10.5	Transformations to stationarity	356	
	10.6	Summary of main points	363	
	Appendix 10.1: Generated DSP and TSP series for exercises			
11	Misspecification and autocorrelation			
	11.1	Introduction	366	
	11.2	What is autocorrelation and why is it a problem?	366	
	11.3	Why do we get autocorrelation?	370	
	11.4	Detecting autocorrelation	379	

v 111	Con	ients		
	11.5 11.6	What to do about autocorrelation Summary of main points	387 390	
	Appe	Appendix 11.1: Derivation of variance and covariance		
	for A	R(1) model	391	
12	Coint	egration and the error correction model	393	
	12.1	Introduction	393	
	12.2	What is cointegration?	393	
	12.3	Testing for cointegration	399	
	12.4	The error correction model (ECM)	406	
	12.5	Summary of main points	412	
Pai	tV	Simultaneous equation models	413	
12	Micor	posification bigs from single equation estimation	415	
15	12 1	Introduction	415	
	13.1	Simultaneity bias in a supply and demand model	415	
	13.2	Simultaneity bias: the theory	417	
	13.5	The Granger and Sims tests for causality and	422	
	13.4	concepts of exogeneity	425	
	135	The identification problem	428	
	13.6	Summary of main points	434	
		Summing of main points	-101	
14	Estim	ating simultaneous equation models	437	
	14.1	Introduction	437	
	14.2	Recursive models	437	
	14.3	Indirect least squares	439	
	14.4	least squares	442	
	14.5	Estimating the consumption function in a		
		simultaneous system	445	
	14.6	Full information estimation techniques	448	
	14.7	Summary of main points	451	
Ар	pendix	A: The data sets used in this book	455	
Ap	pendix	B: Statistical tables	463	
References				
Index				

-.