

Panel Data Econometrics with R

Yves Croissant

Professor of Economics

CEMOI

Faculté de Droit et d'Economie

Université de La Réunion

France

Giovanni Millo

Senior Economist

Group Insurance Research, Assicurazioni Generali S.p.A.

Trieste, Italy

WILEY

Contents

Preface *xiii*

Acknowledgments *xvii*

About the Companion Website *xix*

1	Introduction	<i>1</i>
1.1	Panel Data Econometrics: A Gentle Introduction	<i>1</i>
1.1.1	Eliminating Unobserved Components	<i>2</i>
1.1.1.1	Differencing Methods	<i>2</i>
1.1.1.2	LSDV Methods	<i>2</i>
1.1.1.3	Fixed Effects Methods	<i>2</i>
1.2	R for Econometric Computing	<i>6</i>
1.2.1	The Modus Operandi of R	<i>7</i>
1.2.2	Data Management	<i>8</i>
1.2.2.1	Outsourcing to Other Software	<i>8</i>
1.2.2.2	Data Management Through Formulae	<i>8</i>
1.3	plm for the Casual R User	<i>8</i>
1.3.1	R for the Matrix Language User	<i>9</i>
1.3.2	R for the User of Econometric Packages	<i>10</i>
1.4	plm for the Proficient R User	<i>11</i>
1.4.1	Reproducible Econometric Work	<i>12</i>
1.4.2	Object-orientation for the User	<i>13</i>
1.5	plm for the R Developer	<i>13</i>
1.5.1	Object-orientation for Development	<i>14</i>
1.6	Notations	<i>17</i>
1.6.1	General Notation	<i>18</i>
1.6.2	Maximum Likelihood Notations	<i>18</i>
1.6.3	Index	<i>18</i>
1.6.4	The Two-way Error Component Model	<i>18</i>
1.6.5	Transformation for the One-way Error Component Model	<i>19</i>
1.6.6	Transformation for the Two-ways Error Component Model	<i>20</i>
1.6.7	Groups and Nested Models	<i>20</i>
1.6.8	Instrumental Variables	<i>20</i>
1.6.9	Systems of Equations	<i>20</i>
1.6.10	Time Series	<i>21</i>
1.6.11	Limited Dependent and Count Variables	<i>21</i>
1.6.12	Spatial Panels	<i>21</i>

2	The Error Component Model	23
2.1	Notations and Hypotheses	23
2.1.1	Notations	23
2.1.2	Some Useful Transformations	24
2.1.3	Hypotheses Concerning the Errors	25
2.2	Ordinary Least Squares Estimators	27
2.2.1	Ordinary Least Squares on the Raw Data: The <i>Pooling</i> Model	27
2.2.2	The <i>between</i> Estimator	28
2.2.3	The <i>within</i> Estimator	29
2.3	The Generalized Least Squares Estimator	33
2.3.1	Presentation of the GLS Estimator	34
2.3.2	Estimation of the Variances of the Components of the Error	35
2.4	Comparison of the Estimators	39
2.4.1	Relations between the Estimators	39
2.4.2	Comparison of the Variances	40
2.4.3	Fixed vs Random Effects	40
2.4.4	Some Simple Linear Model Examples	42
2.5	The Two-ways Error Components Model	47
2.5.1	Error Components in the Two-ways Model	47
2.5.2	Fixed and Random Effects Models	48
2.6	Estimation of a Wage Equation	49
3	Advanced Error Components Models	53
3.1	Unbalanced Panels	53
3.1.1	Individual Effects Model	53
3.1.2	Two-ways Error Component Model	54
3.1.2.1	Fixed Effects Model	55
3.1.2.2	Random Effects Model	56
3.1.3	Estimation of the Components of the Error Variance	57
3.2	Seemingly Unrelated Regression	64
3.2.1	Introduction	64
3.2.2	Constrained Least Squares	65
3.2.3	Inter-equations Correlation	66
3.2.4	SUR With Panel Data	67
3.3	The Maximum Likelihood Estimator	71
3.3.1	Derivation of the Likelihood Function	71
3.3.2	Computation of the Estimator	73
3.4	The Nested Error Components Model	74
3.4.1	Presentation of the Model	74
3.4.2	Estimation of the Variance of the Error Components	75
4	Tests on Error Component Models	83
4.1	Tests on Individual and/or Time Effects	83
4.1.1	F Tests	84
4.1.2	Breusch-Pagan Tests	84
4.2	Tests for Correlated Effects	88
4.2.1	The Mundlak Approach	89
4.2.2	Hausman Test	90
4.2.3	Chamberlain's Approach	90

4.2.3.1	Unconstrained Estimator	91
4.2.3.2	Constrained Estimator	93
4.2.3.3	Fixed Effects Models	93
4.3	Tests for Serial Correlation	95
4.3.1	Unobserved Effects Test	95
4.3.2	Score Test of Serial Correlation and/or Individual Effects	96
4.3.3	Likelihood Ratio Tests for AR(1) and Individual Effects	99
4.3.4	Applying Traditional Serial Correlation Tests to Panel Data	101
4.3.5	Wald Tests for Serial Correlation using <i>within</i> and First-differenced Estimators	102
4.3.5.1	Wooldridge's <i>within</i> -based Test	102
4.3.5.2	Wooldridge's First-difference-based Test	103
4.4	Tests for Cross-sectional Dependence	104
4.4.1	Pairwise Correlation Coefficients	104
4.4.2	CD-type Tests for Cross-sectional Dependence	105
4.4.3	Testing Cross-sectional Dependence in a pseries	107
5	Robust Inference and Estimation for Non-spherical Errors	109
5.1	Robust Inference	109
5.1.1	Robust Covariance Estimators	109
5.1.1.1	Cluster-robust Estimation in a Panel Setting	110
5.1.1.2	Double Clustering	115
5.1.1.3	Panel Newey-west and scc	116
5.1.2	Generic Sandwich Estimators and Panel Models	120
5.1.2.1	Panel Corrected Standard Errors	122
5.1.3	Robust Testing of Linear Hypotheses	123
5.1.3.1	An Application: Robust Hausman Testing	125
5.2	Unrestricted Generalized Least Squares	127
5.2.1	General Feasible Generalized Least Squares	128
5.2.1.1	Pooled GGLS	129
5.2.1.2	Fixed Effects GLS	130
5.2.1.3	First Difference GLS	132
5.2.2	Applied Examples	133
6	Endogeneity	139
6.1	Introduction	139
6.2	The Instrumental Variables Estimator	140
6.2.1	Generalities about the Instrumental Variables Estimator	140
6.2.2	The <i>within</i> Instrumental Variables Estimator	141
6.3	Error Components Instrumental Variables Estimator	143
6.3.1	The General Model	143
6.3.2	Special Cases of the General Model	145
6.3.2.1	The <i>within</i> Model	145
6.3.2.2	Error Components Two Stage Least Squares	146
6.3.2.3	The Hausman and Taylor Model	146
6.3.2.4	The Amemiya-Macurdy Estimator	147
6.3.2.5	The Breusch, Mizon and Schmidt's Estimator	147
6.3.2.6	Balestra and Varadharajan-Krishnakumar Estimator	147
6.4	Estimation of a System of Equations	154
6.4.1	The Three Stage Least Squares Estimator	155

6.4.2	The Error Components Three Stage Least Squares Estimator	156
6.5	More Empirical Examples	158
7	Estimation of a Dynamic Model	161
7.1	Dynamic Model and Endogeneity	163
7.1.1	The Bias of the OLS Estimator	163
7.1.2	The within Estimator	164
7.1.3	Consistent Estimation Methods for Dynamic Models	165
7.2	GMM Estimation of the Differenced Model	168
7.2.1	Instrumental Variables and Generalized Method of Moments	168
7.2.2	One-step Estimator	169
7.2.3	Two-steps Estimator	171
7.2.4	The Proliferation of Instruments in the Generalized Method of Moments Difference Estimator	172
7.3	Generalized Method of Moments Estimator in Differences and Levels	174
7.3.1	Weak Instruments	174
7.3.2	Moment Conditions on the Levels Model	175
7.3.3	The System GMM Estimator	177
7.4	Inference	178
7.4.1	Robust Estimation of the Coefficients' Covariance	178
7.4.2	Overidentification Tests	179
7.4.3	Error Serial Correlation Test	181
7.5	More Empirical Examples	182
8	Panel Time Series	185
8.1	Introduction	185
8.2	Heterogeneous Coefficients	186
8.2.1	Fixed Coefficients	186
8.2.2	Random Coefficients	187
8.2.2.1	The Swamy Estimator	187
8.2.2.2	The Mean Groups Estimator	190
8.2.3	Testing for Poolability	192
8.3	Cross-sectional Dependence and Common Factors	194
8.3.1	The Common Factor Model	195
8.3.2	Common Correlated Effects Augmentation	196
8.3.2.1	CCE Mean Groups vs. CCE Pooled	198
8.3.2.2	Computing the CCEP Variance	199
8.4	Nonstationarity and Cointegration	200
8.4.1	Unit Root Testing: Generalities	201
8.4.2	First Generation Unit Root Testing	204
8.4.2.1	Preliminary Results	204
8.4.2.2	Levin-Lin-Chu Test	205
8.4.2.3	Im, Pesaran and Shin Test	205
8.4.2.4	The Maddala and Wu Test	206
8.4.3	Second Generation Unit Root Testing	207
9	Count Data and Limited Dependent Variables	211
9.1	Binomial and Ordinal Models	213
9.1.1	Introduction	213

9.1.1.1	The Binomial Model	213
9.1.1.2	Ordered Models	214
9.1.2	The Random Effects Model	214
9.1.2.1	The Binomial Model	214
9.1.2.2	Ordered Models	217
9.1.3	The Conditional Logit Model	219
9.2	Censored or Truncated Dependent Variable	223
9.2.1	Introduction	223
9.2.2	The Ordinary Least Squares Estimator	223
9.2.3	The Symmetrical Trimmed Estimator	225
9.2.3.1	Truncated Sample	225
9.2.3.2	Censored Sample	226
9.2.4	The Maximum Likelihood Estimator	226
9.2.4.1	Truncated Sample	226
9.2.4.2	Censored Sample	227
9.2.5	Fixed Effects Model	227
9.2.5.1	Truncated Sample	227
9.2.5.2	Censored Sample	229
9.2.6	The Random Effects Model	233
9.2.6.1	Truncated Sample	233
9.2.6.2	Censored Sample	234
9.3	Count Data	236
9.3.1	Introduction	236
9.3.1.1	The Poisson Model	236
9.3.1.2	The NegBin Model	237
9.3.2	Fixed Effects Model	237
9.3.2.1	The Poisson Model	237
9.3.2.2	Negbin Model	239
9.3.3	Random Effects Models	239
9.3.3.1	The Poisson Model	239
9.3.3.2	The NegBin Model	240
9.4	More Empirical Examples	243
10	Spatial Panels	245
10.1	Spatial Correlation	245
10.1.1	Visual Assessment	245
10.1.2	Testing for Spatial Dependence	246
10.1.2.1	CD p Tests for Local Cross-sectional Dependence	247
10.1.2.2	The Randomized W Test	247
10.2	Spatial Lags	250
10.2.1	Spatially Lagged Regressors	251
10.2.2	Spatially Lagged Dependent Variables	253
10.2.2.1	Spatial OLS	254
10.2.2.2	ML Estimation of the SAR Model	254
10.2.3	Spatially Correlated Errors	255
10.3	Individual Heterogeneity in Spatial Panels	258
10.3.1	Random versus Fixed Effects	258
10.3.2	Spatial Panel Models with Error Components	260
10.3.2.1	Spatial Panels with Independent Random Effects	260

10.3.2.2	Spatially Correlated Random Effects	261
10.3.3	Estimation	261
10.3.3.1	Spatial Models with a General Error Covariance	262
10.3.3.2	General Maximum Likelihood Framework	263
10.3.3.3	Generalized Moments Estimation	267
10.3.4	Testing	269
10.3.4.1	LM Tests for Random Effects and Spatial Errors	269
10.3.4.2	Testing for Spatial Lag vs Error	272
10.4	Serial and Spatial Correlation	277
10.4.1	Maximum Likelihood Estimation	277
10.4.1.1	Serial and Spatial Correlation in the Random Effects Model	277
10.4.1.2	Serial and Spatial Correlation with KKP-Type Effects	278
10.4.2	Testing	281
10.4.2.1	Tests for Random Effects, Spatial, and Serial Error Correlation	281
10.4.2.2	Spatial Lag vs Error in the Serially Correlated Model	284

Bibliography 285

Index 297