

SECOND EDITION

JEFFREY M. WOOLDRIDGE

ECONOMETRIC ANALYSIS
OF CROSS SECTION
AND PANEL DATA

Econometric Analysis of Cross Section and Panel Data

Econometric Analysis of Cross Section and Panel Data
Second Edition

Jeffrey M. Wooldridge

The MIT Press
Cambridge, Massachusetts
London, England

© 2010, 2002, Massachusetts Institute of Technology

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the publisher.

For information about special quantity discounts, please email special_sales@mitpress.mit.edu

This book was set in Times Roman by Asco Typesetters, Hong Kong. Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Wooldridge, Jeffrey M.
Econometric analysis of cross section and panel data / Jeffrey M. Wooldridge.—2nd ed.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-262-23258-6 (hardcover : alk. paper)
1. Econometrics—Asymptotic theory. I. Title.
HB139.W663 2010
330.01'5195—dc22

2010020912

10 9 8 7 6 5 4 3 2 1

Contents

	Preface	xxi
	Acknowledgments	xxix
I	INTRODUCTION AND BACKGROUND	1
1	Introduction	3
1.1	Causal Relationships and Ceteris Paribus Analysis	3
1.2	Stochastic Setting and Asymptotic Analysis	4
1.2.1	Data Structures	4
1.2.2	Asymptotic Analysis	7
1.3	Some Examples	7
1.4	Why Not Fixed Explanatory Variables?	9
2	Conditional Expectations and Related Concepts in Econometrics	13
2.1	Role of Conditional Expectations in Econometrics	13
2.2	Features of Conditional Expectations	14
2.2.1	Definition and Examples	14
2.2.2	Partial Effects, Elasticities, and Semielasticities	15
2.2.3	Error Form of Models of Conditional Expectations	18
2.2.4	Some Properties of Conditional Expectations	19
2.2.5	Average Partial Effects	22
2.3	Linear Projections	25
	Problems	27
	Appendix 2A	30
2.A.1	Properties of Conditional Expectations	30
2.A.2	Properties of Conditional Variances and Covariances	32
2.A.3	Properties of Linear Projections	34
3	Basic Asymptotic Theory	37
3.1	Convergence of Deterministic Sequences	37
3.2	Convergence in Probability and Boundedness in Probability	38
3.3	Convergence in Distribution	40
3.4	Limit Theorems for Random Samples	41
3.5	Limiting Behavior of Estimators and Test Statistics	42
3.5.1	Asymptotic Properties of Estimators	42
3.5.2	Asymptotic Properties of Test Statistics	45
	Problems	47

II	LINEAR MODELS	51
4	Single-Equation Linear Model and Ordinary Least Squares Estimation	53
4.1	Overview of the Single-Equation Linear Model	53
4.2	Asymptotic Properties of Ordinary Least Squares	55
4.2.1	Consistency	56
4.2.2	Asymptotic Inference Using Ordinary Least Squares	59
4.2.3	Heteroskedasticity-Robust Inference	60
4.2.4	Lagrange Multiplier (Score) Tests	62
4.3	Ordinary Least Squares Solutions to the Omitted Variables Problem	65
4.3.1	Ordinary Least Squares Ignoring the Omitted Variables	65
4.3.2	Proxy Variable–Ordinary Least Squares Solution	67
4.3.3	Models with Interactions in Unobservables: Random Coefficient Models	73
4.4	Properties of Ordinary Least Squares under Measurement Error	76
4.4.1	Measurement Error in the Dependent Variable	76
4.4.2	Measurement Error in an Explanatory Variable	78
	Problems	82
5	Instrumental Variables Estimation of Single-Equation Linear Models	89
5.1	Instrumental Variables and Two-Stage Least Squares	89
5.1.1	Motivation for Instrumental Variables Estimation	89
5.1.2	Multiple Instruments: Two-Stage Least Squares	96
5.2	General Treatment of Two-Stage Least Squares	98
5.2.1	Consistency	98
5.2.2	Asymptotic Normality of Two-Stage Least Squares	101
5.2.3	Asymptotic Efficiency of Two-Stage Least Squares	103
5.2.4	Hypothesis Testing with Two-Stage Least Squares	104
5.2.5	Heteroskedasticity-Robust Inference for Two-Stage Least Squares	106
5.2.6	Potential Pitfalls with Two-Stage Least Squares	107
5.3	IV Solutions to the Omitted Variables and Measurement Error Problems	112
5.3.1	Leaving the Omitted Factors in the Error Term	112
5.3.2	Solutions Using Indicators of the Unobservables	112
	Problems	115

6	Additional Single-Equation Topics	123
6.1	Estimation with Generated Regressors and Instruments	123
6.1.1	Ordinary Least Squares with Generated Regressors	123
6.1.2	Two-Stage Least Squares with Generated Instruments	124
6.1.3	Generated Instruments and Regressors	125
6.2	Control Function Approach to Endogeneity	126
6.3	Some Specification Tests	129
6.3.1	Testing for Endogeneity	129
6.3.2	Testing Overidentifying Restrictions	134
6.3.3	Testing Functional Form	137
6.3.4	Testing for Heteroskedasticity	138
6.4	Correlated Random Coefficient Models	141
6.4.1	When Is the Usual IV Estimator Consistent?	142
6.4.2	Control Function Approach	145
6.5	Pooled Cross Sections and Difference-in-Differences Estimation	146
6.5.1	Pooled Cross Sections over Time	146
6.5.2	Policy Analysis and Difference-in-Differences Estimation	147
	Problems	152
	Appendix 6A	157
7	Estimating Systems of Equations by Ordinary Least Squares and Generalized Least Squares	161
7.1	Introduction	161
7.2	Some Examples	161
7.3	System Ordinary Least Squares Estimation of a Multivariate Linear System	166
7.3.1	Preliminaries	166
7.3.2	Asymptotic Properties of System Ordinary Least Squares	167
7.3.3	Testing Multiple Hypotheses	172
7.4	Consistency and Asymptotic Normality of Generalized Least Squares	173
7.4.1	Consistency	173
7.4.2	Asymptotic Normality	175
7.5	Feasible Generalized Least Squares	176
7.5.1	Asymptotic Properties	176
7.5.2	Asymptotic Variance of Feasible Generalized Least Squares under a Standard Assumption	180

7.5.3	Properties of Feasible Generalized Least Squares with (Possibly Incorrect) Restrictions on the Unconditional Variance Matrix	182
7.6	Testing the Use of Feasible Generalized Least Squares	183
7.7	Seemingly Unrelated Regressions, Revisited	185
7.7.1	Comparison between Ordinary Least Squares and Feasible Generalized Least Squares for Seemingly Unrelated Regressions Systems	185
7.7.2	Systems with Cross Equation Restrictions	188
7.7.3	Singular Variance Matrices in Seemingly Unrelated Regressions Systems	189
7.8	Linear Panel Data Model, Revisited	191
7.8.1	Assumptions for Pooled Ordinary Least Squares	191
7.8.2	Dynamic Completeness	194
7.8.3	Note on Time Series Persistence	196
7.8.4	Robust Asymptotic Variance Matrix	197
7.8.5	Testing for Serial Correlation and Heteroskedasticity after Pooled Ordinary Least Squares	198
7.8.6	Feasible Generalized Least Squares Estimation under Strict Exogeneity	200
	Problems	202
8	System Estimation by Instrumental Variables	207
8.1	Introduction and Examples	207
8.2	General Linear System of Equations	210
8.3	Generalized Method of Moments Estimation	213
8.3.1	General Weighting Matrix	213
8.3.2	System Two-Stage Least Squares Estimator	216
8.3.3	Optimal Weighting Matrix	217
8.3.4	The Generalized Method of Moments Three-Stage Least Squares Estimator	219
8.4	Generalized Instrumental Variables Estimator	222
8.4.1	Derivation of the Generalized Instrumental Variables Estimator and Its Asymptotic Properties	222
8.4.2	Comparison of Generalized Method of Moment, Generalized Instrumental Variables, and the Traditional Three-Stage Least Squares Estimator	224

8.5	Testing Using Generalized Method of Moments	226
8.5.1	Testing Classical Hypotheses	226
8.5.2	Testing Overidentification Restrictions	228
8.6	More Efficient Estimation and Optimal Instruments	229
8.7	Summary Comments on Choosing an Estimator	232
	Problems	233
9	Simultaneous Equations Models	239
9.1	Scope of Simultaneous Equations Models	239
9.2	Identification in a Linear System	241
9.2.1	Exclusion Restrictions and Reduced Forms	241
9.2.2	General Linear Restrictions and Structural Equations	245
9.2.3	Unidentified, Just Identified, and Overidentified Equations	251
9.3	Estimation after Identification	252
9.3.1	Robustness-Efficiency Trade-off	252
9.3.2	When Are 2SLS and 3SLS Equivalent?	254
9.3.3	Estimating the Reduced Form Parameters	255
9.4	Additional Topics in Linear Simultaneous Equations Methods	256
9.4.1	Using Cross Equation Restrictions to Achieve Identification	256
9.4.2	Using Covariance Restrictions to Achieve Identification	257
9.4.3	Subtleties Concerning Identification and Efficiency in Linear Systems	260
9.5	Simultaneous Equations Models Nonlinear in Endogenous Variables	262
9.5.1	Identification	262
9.5.2	Estimation	266
9.5.3	Control Function Estimation for Triangular Systems	268
9.6	Different Instruments for Different Equations	271
	Problems	273
10	Basic Linear Unobserved Effects Panel Data Models	281
10.1	Motivation: Omitted Variables Problem	281
10.2	Assumptions about the Unobserved Effects and Explanatory Variables	285
10.2.1	Random or Fixed Effects?	285
10.2.2	Strict Exogeneity Assumptions on the Explanatory Variables	287
10.2.3	Some Examples of Unobserved Effects Panel Data Models	289

10.3	Estimating Unobserved Effects Models by Pooled Ordinary Least Squares	291
10.4	Random Effects Methods	291
10.4.1	Estimation and Inference under the Basic Random Effects Assumptions	291
10.4.2	Robust Variance Matrix Estimator	297
10.4.3	General Feasible Generalized Least Squares Analysis	298
10.4.4	Testing for the Presence of an Unobserved Effect	299
10.5	Fixed Effects Methods	300
10.5.1	Consistency of the Fixed Effects Estimator	300
10.5.2	Asymptotic Inference with Fixed Effects	304
10.5.3	Dummy Variable Regression	307
10.5.4	Serial Correlation and the Robust Variance Matrix Estimator	310
10.5.5	Fixed Effects Generalized Least Squares	312
10.5.6	Using Fixed Effects Estimation for Policy Analysis	315
10.6	First Differencing Methods	315
10.6.1	Inference	315
10.6.2	Robust Variance Matrix	318
10.6.3	Testing for Serial Correlation	319
10.6.4	Policy Analysis Using First Differencing	320
10.7	Comparison of Estimators	321
10.7.1	Fixed Effects versus First Differencing	321
10.7.2	Relationship between the Random Effects and Fixed Effects Estimators	326
10.7.3	Hausman Test Comparing Random Effects and Fixed Effects Estimators	328
	Problems	334
11	More Topics in Linear Unobserved Effects Models	345
11.1	Generalized Method of Moments Approaches to the Standard Linear Unobserved Effects Model	345
11.1.1	Equivalence between GMM 3SLS and Standard Estimators	345
11.1.2	Chamberlain’s Approach to Unobserved Effects Models	347
11.2	Random and Fixed Effects Instrumental Variables Methods	349
11.3	Hausman and Taylor–Type Models	358
11.4	First Differencing Instrumental Variables Methods	361

11.5	Unobserved Effects Models with Measurement Error	365
11.6	Estimation under Sequential Exogeneity	368
11.6.1	General Framework	368
11.6.2	Models with Lagged Dependent Variables	371
11.7	Models with Individual-Specific Slopes	374
11.7.1	Random Trend Model	375
11.7.2	General Models with Individual-Specific Slopes	377
11.7.3	Robustness of Standard Fixed Effects Methods	382
11.7.4	Testing for Correlated Random Slopes	384
	Problems	387
III	GENERAL APPROACHES TO NONLINEAR ESTIMATION	395
12	M-Estimation, Nonlinear Regression, and Quantile Regression	397
12.1	Introduction	397
12.2	Identification, Uniform Convergence, and Consistency	401
12.3	Asymptotic Normality	405
12.4	Two-Step M-Estimators	409
12.4.1	Consistency	410
12.4.2	Asymptotic Normality	411
12.5	Estimating the Asymptotic Variance	413
12.5.1	Estimation without Nuisance Parameters	413
12.5.2	Adjustments for Two-Step Estimation	418
12.6	Hypothesis Testing	420
12.6.1	Wald Tests	420
12.6.2	Score (or Lagrange Multiplier) Tests	421
12.6.3	Tests Based on the Change in the Objective Function	428
12.6.4	Behavior of the Statistics under Alternatives	430
12.7	Optimization Methods	431
12.7.1	Newton-Raphson Method	432
12.7.2	Berndt, Hall, Hall, and Hausman Algorithm	433
12.7.3	Generalized Gauss-Newton Method	434
12.7.4	Concentrating Parameters out of the Objective Function	435
12.8	Simulation and Resampling Methods	436
12.8.1	Monte Carlo Simulation	436
12.8.2	Bootstrapping	438

12.9	Multivariate Nonlinear Regression Methods	442
12.9.1	Multivariate Nonlinear Least Squares	442
12.9.2	Weighted Multivariate Nonlinear Least Squares	444
12.10	Quantile Estimation	449
12.10.1	Quantiles, the Estimation Problem, and Consistency	449
12.10.2	Asymptotic Inference	454
12.10.3	Quantile Regression for Panel Data Problems	459
13	Maximum Likelihood Methods	462
13.1	Introduction	469
13.2	Preliminaries and Examples	470
13.3	General Framework for Conditional Maximum Likelihood Estimation	473
13.4	Consistency of Conditional Maximum Likelihood Estimation	475
13.5	Asymptotic Normality and Asymptotic Variance Estimation	476
13.5.1	Asymptotic Normality	476
13.5.2	Estimating the Asymptotic Variance	479
13.6	Hypothesis Testing	481
13.7	Specification Testing	482
13.8	Partial (or Pooled) Likelihood Methods for Panel Data	485
13.8.1	Setup for Panel Data	486
13.8.2	Asymptotic Inference	490
13.8.3	Inference with Dynamically Complete Models	492
13.9	Panel Data Models with Unobserved Effects	494
13.9.1	Models with Strictly Exogenous Explanatory Variables	494
13.9.2	Models with Lagged Dependent Variables	497
13.10	Two-Step Estimators Involving Maximum Likelihood	499
13.10.1	Second-Step Estimator Is Maximum Likelihood Estimator	499
13.10.2	Surprising Efficiency Result When the First-Step Estimator Is Conditional Maximum Likelihood Estimator	500
13.11	Quasi-Maximum Likelihood Estimation	502
13.11.1	General Misspecification	502
13.11.2	Model Selection Tests	505
13.11.3	Quasi-Maximum Likelihood Estimation in the Linear Exponential Family	509

13.11.4	Generalized Estimating Equations for Panel Data	514
	Problems	517
	Appendix 13A	522
14	Generalized Method of Moments and Minimum Distance Estimation	525
14.1	Asymptotic Properties of Generalized Method of Moments	525
14.2	Estimation under Orthogonality Conditions	530
14.3	Systems of Nonlinear Equations	532
14.4	Efficient Estimation	538
14.4.1	General Efficiency Framework	538
14.4.2	Efficiency of Maximum Likelihood Estimator	540
14.4.3	Efficient Choice of Instruments under Conditional Moment Restrictions	542
14.5	Classical Minimum Distance Estimation	545
14.6	Panel Data Applications	547
14.6.1	Nonlinear Dynamic Models	547
14.6.2	Minimum Distance Approach to the Unobserved Effects Model	549
14.6.3	Models with Time-Varying Coefficients on the Unobserved Effects	551
	Problems	555
	Appendix 14A	558
IV	NONLINEAR MODELS AND RELATED TOPICS	559
15	Binary Response Models	561
15.1	Introduction	561
15.2	Linear Probability Model for Binary Response	562
15.3	Index Models for Binary Response: Probit and Logit	565
15.4	Maximum Likelihood Estimation of Binary Response Index Models	567
15.5	Testing in Binary Response Index Models	569
15.5.1	Testing Multiple Exclusion Restrictions	570
15.5.2	Testing Nonlinear Hypotheses about β	571
15.5.3	Tests against More General Alternatives	571

15.6	Reporting the Results for Probit and Logit	573
15.7	Specification Issues in Binary Response Models	582
15.7.1	Neglected Heterogeneity	582
15.7.2	Continuous Endogenous Explanatory Variables	585
15.7.3	Binary Endogenous Explanatory Variable	594
15.7.4	Heteroskedasticity and Nonnormality in the Latent Variable Model	599
15.7.5	Estimation under Weaker Assumptions	604
15.8	Binary Response Models for Panel Data	608
15.8.1	Pooled Probit and Logit	609
15.8.2	Unobserved Effects Probit Models under Strict Exogeneity	610
15.8.3	Unobserved Effects Logit Models under Strict Exogeneity	619
15.8.4	Dynamic Unobserved Effects Models	625
15.8.5	Probit Models with Heterogeneity and Endogenous Explanatory Variables	630
15.8.6	Semiparametric Approaches	632
	Problems	635
16	Multinomial and Ordered Response Models	643
16.1	Introduction	643
16.2	Multinomial Response Models	643
16.2.1	Multinomial Logit	643
16.2.2	Probabilistic Choice Models	646
16.2.3	Endogenous Explanatory Variables	651
16.2.4	Panel Data Methods	653
16.3	Ordered Response Models	655
16.3.1	Ordered Logit and Ordered Probit	655
16.3.2	Specification Issues in Ordered Models	658
16.3.3	Endogenous Explanatory Variables	660
16.3.4	Panel Data Methods	662
	Problems	663
17	Corner Solution Responses	667
17.1	Motivation and Examples	667
17.2	Useful Expressions for Type I Tobit	671

17.3	Estimation and Inference with the Type I Tobit Model	676
17.4	Reporting the Results	677
17.5	Specification Issues in Tobit Models	680
17.5.1	Neglected Heterogeneity	680
17.5.2	Endogenous Explanatory Models	681
17.5.3	Heteroskedasticity and Nonnormality in the Latent Variable Model	685
17.5.4	Estimating Parameters with Weaker Assumptions	687
17.6	Two-Part Models and Type II Tobit for Corner Solutions	690
17.6.1	Truncated Normal Hurdle Model	692
17.6.2	Lognormal Hurdle Model and Exponential Conditional Mean	694
17.6.3	Exponential Type II Tobit Model	697
17.7	Two-Limit Tobit Model	703
17.8	Panel Data Methods	705
17.8.1	Pooled Methods	705
17.8.2	Unobserved Effects Models under Strict Exogeneity	707
17.8.3	Dynamic Unobserved Effects Tobit Models	713
	Problems	715
18	Count, Fractional, and Other Nonnegative Responses	723
18.1	Introduction	723
18.2	Poisson Regression	724
18.2.1	Assumptions Used for Poisson Regression and Quantities of Interest	724
18.2.2	Consistency of the Poisson QMLE	727
18.2.3	Asymptotic Normality of the Poisson QMLE	728
18.2.4	Hypothesis Testing	732
18.2.5	Specification Testing	734
18.3	Other Count Data Regression Models	736
18.3.1	Negative Binomial Regression Models	736
18.3.2	Binomial Regression Models	739
18.4	Gamma (Exponential) Regression Model	740
18.5	Endogeneity with an Exponential Regression Function	742
18.6	Fractional Responses	748

18.6.1	Exogenous Explanatory Variables	748
18.6.2	Endogenous Explanatory Variables	753
18.7	Panel Data Methods	755
18.7.1	Pooled QMLE	756
18.7.2	Specifying Models of Conditional Expectations with Unobserved Effects	758
18.7.3	Random Effects Methods	759
18.7.4	Fixed Effects Poisson Estimation	762
18.7.5	Relaxing the Strict Exogeneity Assumption	764
18.7.6	Fractional Response Models for Panel Data Problems	766
19	Censored Data, Sample Selection, and Attrition	777
19.1	Introduction	777
19.2	Data Censoring	778
19.2.1	Binary Censoring	780
19.2.2	Interval Coding	783
19.2.3	Censoring from Above and Below	785
19.3	Overview of Sample Selection	790
19.4	When Can Sample Selection Be Ignored?	792
19.4.1	Linear Models: Estimation by OLS and 2SLS	792
19.4.2	Nonlinear Models	798
19.5	Selection on the Basis of the Response Variable: Truncated Regression	799
19.6	Incidental Truncation: A Probit Selection Equation	802
19.6.1	Exogenous Explanatory Variables	802
19.6.2	Endogenous Explanatory Variables	809
19.6.3	Binary Response Model with Sample Selection	813
19.6.4	An Exponential Response Function	814
19.7	Incidental Truncation: A Tobit Selection Equation	815
19.7.1	Exogenous Explanatory Variables	815
19.7.2	Endogenous Explanatory Variables	817
19.7.3	Estimating Structural Tobit Equations with Sample Selection	819
19.8	Inverse Probability Weighting for Missing Data	821

19.9	Sample Selection and Attrition in Linear Panel Data Models	827
19.9.1	Fixed and Random Effects Estimation with Unbalanced Panels	828
19.9.2	Testing and Correcting for Sample Selection Bias	832
19.9.3	Attrition	837
	Problems	845
20	Stratified Sampling and Cluster Sampling	853
20.1	Introduction	853
20.2	Stratified Sampling	854
20.2.1	Standard Stratified Sampling and Variable Probability Sampling	854
20.2.2	Weighted Estimators to Account for Stratification	856
20.2.3	Stratification Based on Exogenous Variables	861
20.3	Cluster Sampling	863
20.3.1	Inference with a Large Number of Clusters and Small Cluster Sizes	864
20.3.2	Cluster Samples with Unit-Specific Panel Data	876
20.3.3	Should We Apply Cluster-Robust Inference with Large Group Sizes?	883
20.3.4	Inference When the Number of Clusters Is Small	884
20.4	Complex Survey Sampling	894
	Problems	899
21	Estimating Average Treatment Effects	903
21.1	Introduction	903
21.2	A Counterfactual Setting and the Self-Selection Problem	904
21.3	Methods Assuming Ignorability (or Unconfoundedness) of Treatment	908
21.3.1	Identification	911
21.3.2	Regression Adjustment	915
21.3.3	Propensity Score Methods	920
21.3.4	Combining Regression Adjustment and Propensity Score Weighting	930
21.3.5	Matching Methods	934

21.4	Instrumental Variables Methods	937
21.4.1	Estimating the Average Treatment Effect Using IV	937
21.4.2	Correction and Control Function Approaches	945
21.4.3	Estimating the Local Average Treatment Effect by IV	951
21.5	Regression Discontinuity Designs	954
21.5.1	The Sharp Regression Discontinuity Design	954
21.5.2	The Fuzzy Regression Discontinuity Design	957
21.5.3	Unconfoundedness versus the Fuzzy Regression Discontinuity	959
21.6	Further Issues	960
21.6.1	Special Considerations for Responses with Discreteness or Limited Range	960
21.6.2	Multivalued Treatments	961
21.6.3	Multiple Treatments	964
21.6.4	Panel Data Problems	968
22	Duration Analysis	975
22.1	Introduction	983
22.2	Hazard Functions	984
22.2.1	Hazard Functions without Covariates	984
22.2.2	Hazard Functions Conditional on Time-Invariant Covariates	988
22.2.3	Hazard Functions Conditional on Time-Varying Covariates	989
22.3	Analysis of Single-Spell Data with Time-Invariant Covariates	991
22.3.1	Flow Sampling	992
22.3.2	Maximum Likelihood Estimation with Censored Flow Data	993
22.3.3	Stock Sampling	1000
22.3.4	Unobserved Heterogeneity	1003
22.4	Analysis of Grouped Duration Data	1010
22.4.1	Time-Invariant Covariates	1011
22.4.2	Time-Varying Covariates	1015
22.4.3	Unobserved Heterogeneity	1017

22.5	Further Issues	1018
22.5.1	Cox’s Partial Likelihood Method for the Proportional Hazard Model	1018
22.5.2	Multiple-Spell Data	1018
22.5.3	Competing Risks Models	1019
	Problems	1019
	References	1025
	Index	1045

To obtain an estimable equation, replace q in equation (4.19) with equation (4.27) to get

$$y = (\beta_0 + \gamma\theta_0) + \beta_1x_1 + \cdots + \beta_Kx_K + \gamma\theta_1z + (\gamma r + v). \quad (4.28)$$

Under the assumptions made, the composite error term $u \equiv \gamma r + v$ is uncorrelated with x_j for all j ; redundancy of z in equation (4.18) means that z is uncorrelated with v and, by definition, z is uncorrelated with r . It follows immediately from Theorem 4.1 that the OLS regression y on $1, x_1, x_2, \dots, x_K, z$ produces consistent estimators of $(\beta_0 + \gamma\theta_0), \beta_1, \beta_2, \dots, \beta_K$, and $\gamma\theta_1$. Thus, we can estimate the partial effect of each of the x_j in equation (4.18) under the proxy variable assumptions.

When z is an **imperfect proxy**, then r in equation (4.27) is correlated with one or more of the x_j . Generally, when we do not impose condition (4.26) and write the linear projection as

$$q = \theta_0 + \rho_1x_1 + \cdots + \rho_Kx_K + \theta_1z + r,$$

the proxy variable regression gives $\text{plim } \hat{\beta}_j = \beta_j + \gamma\rho_j$. Thus, OLS with an imperfect proxy is inconsistent. The hope is that the ρ_j are smaller in magnitude than if z were omitted from the linear projection, and this can usually be argued if z is a reasonable proxy for q ; but see the end of this subsection for further discussion.

If including z induces substantial collinearity, it might be better to use OLS without the proxy variable. However, in making these decisions we must recognize that including z reduces the error variance if $\theta_1 \neq 0$: $\text{Var}(\gamma r + v) < \text{Var}(\gamma q + v)$ because $\text{Var}(r) < \text{Var}(q)$, and v is uncorrelated with both r and q . Including a proxy variable can actually reduce asymptotic variances as well as mitigate bias.

Example 4.3 (Using IQ as a Proxy for Ability): We apply the proxy variable method to the data on working men in NLS80.RAW, which was used by Blackburn and Neumark (1992), to estimate the structural model

$$\begin{aligned} \log(\text{wage}) = & \beta_0 + \beta_1 \text{exper} + \beta_2 \text{tenure} + \beta_3 \text{married} \\ & + \beta_4 \text{south} + \beta_5 \text{urban} + \beta_6 \text{black} + \beta_7 \text{educ} + \gamma \text{abil} + v, \end{aligned} \quad (4.29)$$

where *exper* is labor market experience, *married* is a dummy variable equal to unity if married, *south* is a dummy variable for the southern region, *urban* is a dummy variable for living in an SMSA, *black* is a race indicator, and *educ* is years of schooling. We assume that *IQ* satisfies the proxy variable assumptions: in the linear projection $\text{abil} = \theta_0 + \theta_1 \text{IQ} + r$, where r has zero mean and is uncorrelated with *IQ*, we also assume that r is uncorrelated with experience, tenure, education, and other factors

N very large relative to T , there is no need to downweight correlations between time periods that are far apart, as in the Newey and West (1987) estimator applied to time series problems. Ziliak and Kniesner (1998) do use a Newey-West type procedure in a panel data application with large N . Theoretically, this is not required, and it is not completely general because it assumes that the underlying time series are weakly dependent. (See Wooldridge (1994a) for discussion of weak dependence in time series contexts.) A Newey-West type estimator might improve the finite-sample performance of the GMM estimator.

The asymptotic variance of the optimal GMM estimator is estimated as

$$\left[(\mathbf{X}'\mathbf{Z}) \left(\sum_{i=1}^N \mathbf{Z}_i' \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i' \mathbf{Z}_i \right)^{-1} (\mathbf{Z}'\mathbf{X}) \right]^{-1}, \quad (8.37)$$

where $\hat{\mathbf{u}}_i \equiv \mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}$; asymptotically, it makes no difference whether the first-stage residuals $\hat{\mathbf{u}}_i$ are used in place of \mathbf{u}_i . The square roots of diagonal elements of this matrix are the asymptotic standard errors of the optimal GMM estimator. This estimator is called a **minimum chi-square estimator**, for reasons that will become clear in Section 8.5.2.

When $\mathbf{Z}_i = \mathbf{X}_i$ and the $\hat{\mathbf{u}}_i$ are the system OLS residuals, expression (8.37) becomes the robust variance matrix estimator for SOLS [see expression (7.28)]. This expression reduces to the robust variance matrix estimator for FGLS when $\mathbf{Z}_i = \hat{\boldsymbol{\Omega}}^{-1} \mathbf{X}_i$ and the $\hat{\mathbf{u}}_i$ are the FGLS residuals [see equation (7.52)].

8.3.4 The Generalized Method of Moments Three-Stage Least Squares Estimator

The GMM estimator using weighting matrix (8.36) places no restrictions on either the unconditional or conditional (on \mathbf{Z}_i) variance matrix of \mathbf{u}_i ; we can obtain the asymptotically efficient estimator without making additional assumptions. Nevertheless, it is still common, especially in traditional simultaneous equations analysis, to assume that the conditional variance matrix of \mathbf{u}_i given \mathbf{Z}_i is constant. This assumption leads to a system estimator that is a middle ground between system 2SLS and the always-efficient minimum chi-square estimator.

The **GMM three-stage least squares (GMM 3SLS) estimator** (or just 3SLS when the context is clear) is a GMM estimator that uses a particular weighting matrix. To define the 3SLS estimator, let $\check{\mathbf{u}}_i = \mathbf{y}_i - \mathbf{X}_i \check{\boldsymbol{\beta}}$ be the residuals from an initial estimation, usually system 2SLS. Define the $G \times G$ matrix

$$\hat{\boldsymbol{\Omega}} \equiv N^{-1} \sum_{i=1}^N \check{\mathbf{u}}_i \check{\mathbf{u}}_i'. \quad (8.38)$$

natural than the strict exogeneity assumption, which requires conditioning on future values of \mathbf{x}_{it} as well. As we proceed, it is important to remember that equation (11.52) is what we should have in mind when interpreting the estimates of $\boldsymbol{\beta}$. Estimating equations in first differences, such as (11.35), do not have natural interpretations when the explanatory variables are only sequentially exogenous.

As we will explicitly show in the next subsection, models with lagged dependent variables are naturally analyzed under sequential exogeneity. Keane and Runkle (1992) argue that panel data models with heterogeneity for testing rational expectations hypotheses do not satisfy the strict exogeneity requirement. But they do satisfy sequential exogeneity; in fact, the conditioning set in assumption (11.51) can include all variables observed at time $t - 1$.

As we saw in Section 7.2, in panel data models without unobserved effects, strict exogeneity is sometimes too strong an assumption, even in static and finite distributed lag models. For example, suppose

$$y_{it} = \mathbf{z}_{it}\boldsymbol{\gamma} + \delta h_{it} + c_i + u_{it}, \quad (11.53)$$

where $\{\mathbf{z}_{it}\}$ is strictly exogenous and $\{h_{it}\}$ is sequentially exogenous:

$$E(u_{it} | \mathbf{z}_i, h_{it}, \dots, h_{i1}, c_i) = 0. \quad (11.54)$$

Further, h_{it} is influenced by past y_{it} , say

$$h_{it} = \mathbf{z}_{it}\boldsymbol{\zeta} + \eta y_{i,t-1} + \psi c_i + r_{it}. \quad (11.55)$$

For example, let y_{it} be per capita condom sales in city i during year t , and let h_{it} be the HIV infection rate for city i in year t . Model (11.53) can be used to test whether condom usage is influenced by the spread of HIV. The unobserved effect c_i contains city-specific unobserved factors that can affect sexual conduct, as well as the incidence of HIV. Equation (11.55) is one way of capturing the fact that the spread of HIV depends on past condom usage. Generally, if $E(r_{i,t+1}u_{it}) = 0$, it is easy to show that $E(h_{i,t+1}u_{it}) = \eta E(y_{it}u_{it}) = \eta E(u_{it}^2) > 0$ if $\eta > 0$ under equations (11.54) and (11.55). Therefore, strict exogeneity fails unless $\eta = 0$.

Sometimes in panel data applications one sees variables that are thought to be contemporaneously endogenous appear with a lag, rather than contemporaneously. So, for example, we might use $h_{i,t-1}$ in place of h_{it} in equation (11.53) because we think h_{it} and u_{it} are correlated. As an example, suppose y_{it} is percentage of flights cancelled by airline i in year t , and h_{it} is profits in the same year. We might specify $y_{it} = \mathbf{z}_{it}\boldsymbol{\gamma} + \delta h_{i,t-1} + c_i + u_{it}$ for strictly exogenous \mathbf{z}_{it} . Of course, at $t + 1$, the regressors are $\mathbf{x}_{i,t+1} = (\mathbf{z}_{i,t+1}, h_{it})$, which is correlated with u_{it} if h_{it} is. As we discussed in

where, in most cases, $\mathbf{s}(\mathbf{w})$ is the score of an objective function (evaluated at θ_o) and \mathbf{A} is the expected value of the Jacobian of the score, again evaluated at θ_o . (We suppress an “o” subscript here, as the value of the true parameter is irrelevant.) All M-estimators with twice continuously differentiable objective functions (and even some without) have variance matrices of this form, as do GMM estimators. The following lemma is a useful sufficient condition for showing that one estimator is more efficient than another.

LEMMA 14.1 (Relative Efficiency): Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two \sqrt{N} -asymptotically normal estimators of the $P \times 1$ parameter vector θ_o , with asymptotic variances of the form (14.48) (with appropriate subscripts on \mathbf{A} , \mathbf{s} , and \mathbf{V}). If for some $\rho > 0$,

$$E[\mathbf{s}_1(\mathbf{w})\mathbf{s}_1(\mathbf{w})'] = \rho\mathbf{A}_1, \quad (14.49)$$

$$E[\mathbf{s}_2(\mathbf{w})\mathbf{s}_1(\mathbf{w})'] = \rho\mathbf{A}_2, \quad (14.50)$$

then $\mathbf{V}_2 - \mathbf{V}_1$ is p.s.d.

The proof of Lemma 14.1 is given in the chapter appendix.

Condition (14.49) is essentially the generalized information matrix equality (GIME) we introduced in Section 12.5.1 for the estimator $\hat{\theta}_1$. Notice that \mathbf{A}_1 is necessarily symmetric and positive definite under condition (14.49). Condition (14.50) is new. In most cases, it says that the expected outer product of the scores \mathbf{s}_2 and \mathbf{s}_1 equals the expected Jacobian of \mathbf{s}_2 (evaluated at θ_o). In Section 12.5.1 we claimed that the GIME plays a role in efficiency, and Lemma 14.1 shows that it does so.

Verifying the conditions of Lemma 14.1 is also very convenient for constructing simple forms of the Hausman (1978) statistic in a variety of contexts. Provided that the two estimators are jointly asymptotically normally distributed—something that is almost always true when each is \sqrt{N} -asymptotically normal, and that can be verified by stacking the first-order representations of the estimators—assumptions (14.49) and (14.50) imply that the asymptotic covariance between $\sqrt{N}(\hat{\theta}_2 - \theta_o)$ and $\sqrt{N}(\hat{\theta}_1 - \theta_o)$ is $\mathbf{A}_2^{-1}E(\mathbf{s}_2\mathbf{s}_1')\mathbf{A}_1^{-1} = \mathbf{A}_2^{-1}(\rho\mathbf{A}_2)\mathbf{A}_1^{-1} = \rho\mathbf{A}_1^{-1} = \text{Avar}[\sqrt{N}(\hat{\theta}_1 - \theta_o)]$. In other words, the asymptotic covariance between the (\sqrt{N} -scaled) estimators is equal to the asymptotic variance of the efficient estimator. This equality implies that $\text{Avar}[\sqrt{N}(\hat{\theta}_2 - \hat{\theta}_1)] = \mathbf{V}_2 + \mathbf{V}_1 - \mathbf{C} - \mathbf{C}' = \mathbf{V}_2 + \mathbf{V}_1 - 2\mathbf{V}_1 = \mathbf{V}_2 - \mathbf{V}_1$, where \mathbf{C} is the asymptotic covariance. If $\mathbf{V}_2 - \mathbf{V}_1$ is actually positive definite (rather than just p.s.d.), then $[\sqrt{N}(\hat{\theta}_2 - \hat{\theta}_1)]'(\hat{\mathbf{V}}_2 - \hat{\mathbf{V}}_1)^{-1}[\sqrt{N}(\hat{\theta}_2 - \hat{\theta}_1)] \stackrel{a}{\sim} \chi_P^2$ under the assumptions of Lemma 14.1, where $\hat{\mathbf{V}}_g$ is a consistent estimator of \mathbf{V}_g , $g = 1, 2$. Statistically significant differences between $\hat{\theta}_2$ and $\hat{\theta}_1$ signal some sort of model misspecification. (See Section 6.2.1, where we discussed this form of the Hausman test for comparing

are easily obtained using the panel data bootstrap. See Papke and Wooldridge (2008) for more discussion.

Problems

18.1. a. For estimating the mean of a nonnegative random variable y , the Poisson quasi-log likelihood for a random draw is

$$\ell_i(\mu) = y_i \log(\mu) - \mu, \quad \mu > 0$$

(where terms not depending on μ have been dropped). Letting $\mu_o \equiv E(y_i)$, we have $E[\ell_i(\mu)] = \mu_o \log(\mu) - \mu$. Show that this function is uniquely maximized at $\mu = \mu_o$. This simple result is the basis for the consistency of the Poisson QMLE in the general case.

b. The gamma (exponential) quasi-log likelihood is

$$\ell_i(\mu) = -y_i/\mu - \log(\mu), \quad \mu > 0$$

Show that $E[\ell_i(\mu)]$ is uniquely maximized at $\mu = \mu_o$.

18.2. Carefully write out the robust variance matrix estimator (18.14) when $m(\mathbf{x}, \boldsymbol{\beta}) = \exp(\mathbf{x}\boldsymbol{\beta})$.

18.3. Use the data in `SMOKE.RAW` to answer this question.

a. Use a linear regression model to explain *cigs*, the number of cigarettes smoked per day. Use as explanatory variables $\log(\text{cigpric})$, $\log(\text{income})$, *restaurn*, *white*, *educ*, *age*, and *age*². Are the price and income variables significant? Does using heteroskedasticity-robust standard errors change your conclusions?

b. Now estimate a Poisson regression model for *cigs*, with an exponential conditional mean and the same explanatory variables as in part a. Using the usual MLE standard errors, are the price and income variables each significant at the 5 percent level? Interpret their coefficients.

c. Find $\hat{\sigma}$. Is there evidence of overdispersion? Using the GLM standard errors, discuss the significance of $\log(\text{cigpric})$ and $\log(\text{income})$.

d. Compare the usual MLE *LR* statistic for joint significance of $\log(\text{cigpric})$ and $\log(\text{income})$ with the *QLR* statistic in equation (18.17).

e. Compute the fully robust standard errors, and compare these with the GLM standard errors.

Table 20.1
Salary-Benefits Trade-off for Michigan Teachers

Dependent Variable	log(<i>avgsal</i>)		
	(1)	(2)	(3)
Estimation Method	Pooled OLS	Random Effects	Fixed Effects
Explanatory Variable			
<i>bs</i>	−0.177 (0.122) [0.260]	−0.381 (0.112) [0.150]	−0.495 (0.133) [0.194]
log(<i>staff</i>)	−0.691 (0.018) [0.035]	−0.617 (0.015) [0.036]	−0.622 (0.017) [0.043]
log(<i>enroll</i>)	−0.0292 (0.0085) [0.0257]	−0.0249 (0.0076) [0.0115]	−0.0515 (0.0094) [0.0131]
<i>lunch</i>	−0.00085 (0.00016) [0.00057]	0.00030 (0.00018) [0.00020]	0.00051 (0.00021) [0.00021]
constant	13.724 (0.112) [0.256]	13.367 (0.098) [0.197]	13.618 (0.113) [0.241]
Number of districts	537	537	537
Number of schools	1,848	1,848	1,848

Quantities in parentheses are the nonrobust standard errors; those in brackets are robust to arbitrary within-district correlation as well as heteroskedasticity.
The intercept reported for fixed effects is the average of the estimated district effects.
The fully robust regression based Hausman test, with four degrees-of-freedom in the chi-square distribution, yields $H = 20.70$ and p -value = 0.0004.

percentage of students eligible for the federal free or reduced-price lunch program. Using the approximation $\log(1 + x) \approx x$ for “small” x , it can be shown that a dollar-for-dollar trade-off in salary and benefits is the same as $\beta_1 = -1$.

We estimate the equation using three methods: pooled OLS, random effects, and fixed effects. The results are given in Table 20.1. The table contains the nonrobust standard errors for each method—that is, the standard errors computed under the “ideal” set of assumptions for the particular estimator—along with the standard errors that are robust to arbitrary within-district correlation and heteroskedasticity.

The POLS estimates provide little evidence of a trade-off between salary and benefits. The coefficient is negative, but its value, -0.177 , is pretty small, and not close to -1 (the hypothesized value for a one-for-one trade-off between salary and benefits). Its fully robust t statistic is less than 0.7 in magnitude. Notice that the robust standard error, which properly accounts for the cluster nature of the data, is more than twice as large as the nonrobust one.