

GLOBAL
EDITION



Introduction to Econometrics

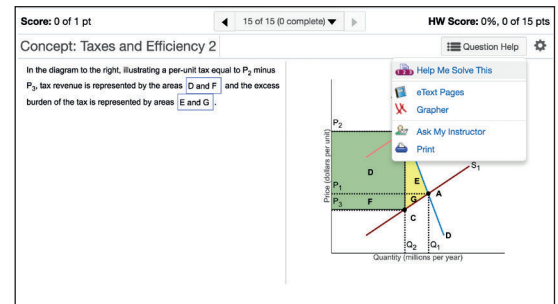
FOURTH EDITION

James H. Stock • Mark W. Watson

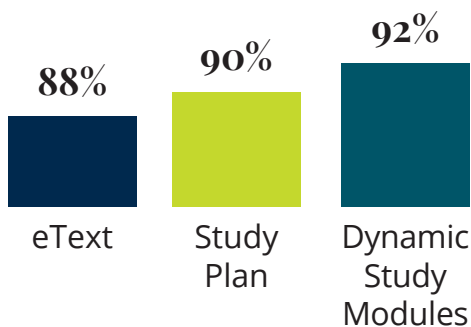


Question Help

MyLab Economics homework and practice questions are correlated to the textbook, and many generate algorithmically to give students unlimited opportunity for mastery of concepts. If students get stuck, Learning Aids including Help Me Solve This and eText Pages walk them through the problem and identify helpful information in the text, giving them assistance when they need it most.



"[MyLab Economics] provides ample practice and explanation of the concepts at hand."
— Heather Burkett, University of Nebraska at Omaha




% of students who
found learning
tool helpful

Dynamic Study Modules help students study chapter topics effectively on their own by continuously assessing their **knowledge application** and performance in real time. These are available as prebuilt Prepare assignments, and are accessible on smartphones, tablets, and computers.

Pearson eText enhances student learning—both in and outside the classroom. Worked examples, videos, and interactive tutorials bring learning to life, while algorithmic practice and self-assessment opportunities test students' understanding of the material. Accessible anytime, anywhere via MyLab or the app.

The **MyLab Gradebook** offers an easy way for students and instructors to view course performance. Item Analysis allows instructors to quickly see trends by analyzing details like the number of students who answered correctly/incorrectly, time on task, and median time spend on a question by question basis. And because it's correlated with the AACSB Standards, instructors can track students' progress toward outcomes that the organization has deemed important in preparing students to be **leaders**.

87% 
of students would tell
their instructor to keep
using MyLab Economics

For additional details visit: www.pearson.com/mylab/economics

Introduction to Econometrics, Global Edition

Table of Contents

Cover

Title Page

Copyright Page

Brief Contents

Contents

Key Concepts

General Interest Boxes

Preface

Acknowledgments

Global Acknowledgments

Chapter 1: Economic Questions and Data

1.1. Economic Questions We Examine

Question #1: Does Reducing Class Size Improve Elementary School Education?

Question #2: Is There Racial Discrimination in the Market for Home Loans?

Question #3: Does Healthcare Spending Improve Health Outcomes?

Question #4: By How Much Will U.S. GDP Grow Next Year?

Quantitative Questions, Quantitative Answers

1.2. Causal Effects and Idealized Experiments

Estimation of Causal Effects

Prediction, Forecasting, and Causality

1.3. Data: Sources and Types

Experimental versus Observational Data

Cross-Sectional Data

Time Series Data

Panel Data

Chapter 2: Review of Probability

Table of Contents

2.1. Random Variables and Probability Distributions

Probabilities, the Sample Space, and Random Variables

Probability Distribution of a Discrete Random Variable

Probability Distribution of a Continuous Random Variable

2.2. Expected Values, Mean, and Variance

The Expected Value of a Random Variable

The Standard Deviation and Variance

Mean and Variance of a Linear Function of a Random Variable

Other Measures of the Shape of a Distribution

Standardized Random Variables

2.3. Two Random Variables

Joint and Marginal Distributions

Conditional Distributions

Independence

Covariance and Correlation

The Mean and Variance of Sums of Random Variables

2.4. The Normal, Chi-Squared, Student t, and F Distributions

The Normal Distribution

The Chi-Squared Distribution

The Student t Distribution

The F Distribution

2.5. Random Sampling and the Distribution of the Sample Average

Random Sampling

The Sampling Distribution of the Sample Average

2.6. Large-Sample Approximations to Sampling Distributions

The Law of Large Numbers and Consistency

The Central Limit Theorem

Appendix 2.1: Derivation of Results in Key Concept 2.3

Appendix 2.2: The Conditional Mean as the Minimum Mean Squared Error Predictor

Chapter 3: Review of Statistics

3.1. Estimation of the Population Mean

Estimators and Their Properties

Table of Contents

Properties of Y

The Importance of Random Sampling

3.2. Hypothesis Tests Concerning the Population Mean

Null and Alternative Hypotheses

The p-Value

Calculating the p-Value When s_Y Is Known

The Sample Variance, Sample Standard Deviation, and Standard Error

Calculating the p-Value When s_Y Is Unknown

The t-Statistic

Hypothesis Testing with a Prespecified Significance Level

One-Sided Alternatives

3.3. Confidence Intervals for the Population Mean

3.4. Comparing Means from Different Populations

Hypothesis Tests for the Difference Between Two Means

Confidence Intervals for the Difference Between Two Population Means

3.5. Differences-of-Means Estimation of Causal Effects Using Experimental Data

The Causal Effect as a Difference of Conditional Expectations

Estimation of the Causal Effect Using Differences of Means

3.6. Using the t-Statistic When the Sample Size Is Small

The t-Statistic and the Student t Distribution

Use of the Student t Distribution in Practice

3.7. Scatterplots, the Sample Covariance, and the Sample Correlation

Scatterplots

Sample Covariance and Correlation

Appendix 3.1: The U.S. Current Population Survey

Appendix 3.2: Two Proofs That Y Is the Least Squares Estimator of μ_Y

Appendix 3.3: A Proof That the Sample Variance Is Consistent

Chapter 4: Linear Regression with One Regressor

4.1. The Linear Regression Model

4.2. Estimating the Coefficients of the Linear Regression Model

The Ordinary Least Squares Estimator

OLS Estimates of the Relationship Between Test Scores and the StudentTeacher

Table of Contents

Ratio

Why Use the OLS Estimator?

4.3. Measures of Fit and Prediction Accuracy

The R^2

The Standard Error of the Regression

Prediction Using OLS

Application to the Test Score Data

4.4. The Least Squares Assumptions for Causal Inference

Assumption 1: The Conditional Distribution of u_i Given X_i Has a Mean of Zero

Assumption 2: (X_i, Y_i) , $i = 1, \dots, n$

Assumption 3: Large Outliers Are Unlikely

Use of the Least Squares Assumptions

4.5. The Sampling Distribution of the OLS Estimators

4.6. Conclusion

Appendix 4.1: The California Test Score Data Set

Appendix 4.2: Derivation of the OLS Estimators

Appendix 4.3: Sampling Distribution of the OLS Estimator

Appendix 4.4: The Least Squares Assumptions for Prediction

Chapter 5: Regression with a Single Regressor: Hypothesis Tests and Confidence Intervals

5.1. Testing Hypotheses About One of the Regression Coefficients

Two-Sided Hypotheses Concerning β_1

One-Sided Hypotheses Concerning β_1

Testing Hypotheses About the Intercept β_0

5.2. Confidence Intervals for a Regression Coefficient

5.3. Regression When X Is a Binary Variable

Interpretation of the Regression Coefficients

5.4. Heteroskedasticity and Homoskedasticity

What Are Heteroskedasticity and Homoskedasticity?

Mathematical Implications of Homoskedasticity

What Does This Mean in Practice?

5.5. The Theoretical Foundations of Ordinary Least Squares

Table of Contents

Linear Conditionally Unbiased Estimators and the GaussMarkov Theorem

Regression Estimators Other Than OLS

5.6. Using the t-Statistic in Regression When the Sample Size Is Small

The t-Statistic and the Student t Distribution

Use of the Student t Distribution in Practice

5.7. Conclusion

Appendix 5.1: Formulas for OLS Standard Errors

Appendix 5.2: The GaussMarkov Conditions and a Proof of the
GaussMarkov Theorem

Chapter 6: Linear Regression with Multiple Regressors

6.1. Omitted Variable Bias

Definition of Omitted Variable Bias

A Formula for Omitted Variable Bias

Addressing Omitted Variable Bias by Dividing the Data into Groups

6.2. The Multiple Regression Model

The Population Regression Line

The Population Multiple Regression Model

6.3. The OLS Estimator in Multiple Regression

The OLS Estimator

Application to Test Scores and the StudentTeacher Ratio

6.4. Measures of Fit in Multiple Regression

The Standard Error of the Regression (SER)

The R^2

The Adjusted R^2

Application to Test Scores

6.5. The Least Squares Assumptions for Causal Inference in Multiple Regression

Assumption 1: The Conditional Distribution of u_i Given X_{1i} , X_{2i} ,

Assumption 2: (X_{1i}, X_{2i}) ,

Assumption 3: Large Outliers Are Unlikely

Assumption 4: No Perfect Multicollinearity

6.6. The Distribution of the OLS Estimators in Multiple Regression

Table of Contents

6.7. Multicollinearity

Examples of Perfect Multicollinearity

Imperfect Multicollinearity

6.8. Control Variables and Conditional Mean Independence

Control Variables and Conditional Mean Independence

6.9. Conclusion

Appendix 6.1: Derivation of Equation (6.1)

Appendix 6.2: Distribution of the OLS Estimators When There Are Two Regressors and Homoskedastic Errors

Appendix 6.3: The FrischWaugh Theorem

Appendix 6.4: The Least Squares Assumptions for Prediction with Multiple Regressors

Appendix 6.5: Distribution of OLS Estimators in Multiple Regression with Control Variables

Chapter 7: Hypothesis Tests and Confidence Intervals in Multiple Regression

7.1. Hypothesis Tests and Confidence Intervals for a Single Coefficient

Standard Errors for the OLS Estimators

Hypothesis Tests for a Single Coefficient

Confidence Intervals for a Single Coefficient

Application to Test Scores and the StudentTeacher Ratio

7.2. Tests of Joint Hypotheses

Testing Hypotheses on Two or More Coefficients

The F-Statistic

Application to Test Scores and the StudentTeacher Ratio

The Homoskedasticity-Only F-Statistic

7.3. Testing Single Restrictions Involving Multiple Coefficients

7.4. Confidence Sets for Multiple Coefficients

7.5. Model Specification for Multiple Regression

Model Specification and Choosing Control Variables

Interpreting the R^2 and the Adjusted R^2 in Practice

7.6. Analysis of the Test Score Data Set

Table of Contents

7.7. Conclusion

Appendix 7.1: The Bonferroni Test of a Joint Hypothesis

Chapter 8: Nonlinear Regression Functions

8.1. A General Strategy for Modeling Nonlinear Regression Functions

Test Scores and District Income

The Effect on Y of a Change in X in Nonlinear Specifications

A General Approach to Modeling Nonlinearities Using Multiple Regression

8.2. Nonlinear Functions of a Single Independent Variable

Polynomials

Logarithms

Polynomial and Logarithmic Models of Test Scores and District Income

8.3. Interactions Between Independent Variables

Interactions Between Two Binary Variables

Interactions Between a Continuous and a Binary Variable

Interactions Between Two Continuous Variables

8.4. Nonlinear Effects on Test Scores of the StudentTeacher Ratio

Discussion of Regression Results

Summary of Findings

8.5. Conclusion

Appendix 8.1: Regression Functions That Are Nonlinear in the Parameters

Appendix 8.2: Slopes and Elasticities for Nonlinear Regression Functions

Chapter 9: Assessing Studies Based on Multiple Regression

9.1. Internal and External Validity

Threats to Internal Validity

Threats to External Validity

9.2. Threats to Internal Validity of Multiple Regression Analysis

Omitted Variable Bias

Misspecification of the Functional Form of the Regression Function

Measurement Error and Errors-in-Variables Bias

Missing Data and Sample Selection

Simultaneous Causality

Sources of Inconsistency of OLS Standard Errors

Table of Contents

9.3. Internal and External Validity When the Regression Is Used for Prediction

9.4. Example: Test Scores and Class Size

External Validity

Internal Validity

Discussion and Implications

9.5. Conclusion

Appendix 9.1: The Massachusetts Elementary School Testing Data

Chapter 10: Regression with Panel Data

10.1. Panel Data

Example: Traffic Deaths and Alcohol Taxes

10.2. Panel Data with Two Time Periods: Before and After Comparisons

10.3. Fixed Effects Regression

The Fixed Effects Regression Model

Estimation and Inference

Application to Traffic Deaths

10.4. Regression with Time Fixed Effects

Time Effects Only

Both Entity and Time Fixed Effects

10.5. The Fixed Effects Regression Assumptions and Standard Errors for
Fixed Effects Regression

The Fixed Effects Regression Assumptions

Standard Errors for Fixed Effects Regression

10.6. Drunk Driving Laws and Traffic Deaths

10.7. Conclusion

Appendix 10.1: The State Traffic Fatality Data Set

Appendix 10.2: Standard Errors for Fixed Effects Regression

Chapter 11: Regression with a Binary Dependent Variable

11.1. Binary Dependent Variables and the Linear Probability Model

Binary Dependent Variables

The Linear Probability Model

11.2. Probit and Logit Regression

Table of Contents

Probit Regression

Logit Regression

Comparing the Linear Probability, Probit, and Logit Models

11.3. Estimation and Inference in the Logit and Probit Models

Nonlinear Least Squares Estimation

Maximum Likelihood Estimation

Measures of Fit

11.4. Application to the Boston HMDA Data

11.5. Conclusion

Appendix 11.1: The Boston HMDA Data Set

Appendix 11.2: Maximum Likelihood Estimation

Appendix 11.3: Other Limited Dependent Variable Models

Chapter 12: Instrumental Variables Regression

12.1. The IV Estimator with a Single Regressor and a Single Instrument

The IV Model and Assumptions

The Two Stage Least Squares Estimator

Why Does IV Regression Work?

The Sampling Distribution of the TSLS Estimator

Application to the Demand for Cigarettes

12.2. The General IV Regression Model

TSLS in the General IV Model

Instrument Relevance and Exogeneity in the General IV Model

The IV Regression Assumptions and Sampling Distribution of the TSLS Estimator

Inference Using the TSLS Estimator

Application to the Demand for Cigarettes

12.3. Checking Instrument Validity

Assumption 1: Instrument Relevance

Assumption 2: Instrument Exogeneity

12.4. Application to the Demand for Cigarettes

12.5. Where Do Valid Instruments Come From?

Three Examples

12.6. Conclusion

Table of Contents

Appendix 12.1: The Cigarette Consumption Panel Data Set

Appendix 12.2: Derivation of the Formula for the TSLS Estimator in Equation (12.4)

Appendix 12.3: Large-Sample Distribution of the TSLS Estimator

Appendix 12.4: Large-Sample Distribution of the TSLS Estimator When the Instrument
Is Not Valid

Appendix 12.5: Instrumental Variables Analysis with Weak Instruments

Appendix 12.6: TSLS with Control Variables

Chapter 13: Experiments and Quasi-Experiments

13.1. Potential Outcomes, Causal Effects, and Idealized Experiments

Potential Outcomes and the Average Causal Effect

Econometric Methods for Analyzing Experimental Data

13.2. Threats to Validity of Experiments

Threats to Internal Validity

Threats to External Validity

13.3. Experimental Estimates of the Effect of Class Size Reductions

Experimental Design

Analysis of the STAR Data

Comparison of the Observational and Experimental Estimates of Class Size Effects

13.4. Quasi-Experiments

Examples

The Differences-in-Differences Estimator

Instrumental Variables Estimators

Regression Discontinuity Estimators

13.5. Potential Problems with Quasi-Experiments

Threats to Internal Validity

Threats to External Validity

13.6. Experimental and Quasi-Experimental Estimates in Heterogeneous
Populations

OLS with Heterogeneous Causal Effects

IV Regression with Heterogeneous Causal Effects

13.7. Conclusion

Appendix 13.1: The Project STAR Data Set

Table of Contents

Appendix 13.2: IV Estimation When the Causal Effect Varies Across Individuals

Appendix 13.3: The Potential Outcomes Framework for Analyzing Data from Experiments

Chapter 14: Prediction with Many Regressors and Big Data

14.1. What Is Big Data?

14.2. The Many-Predictor Problem and OLS

The Mean Squared Prediction Error

The First Least Squares Assumption for Prediction

The Predictive Regression Model with Standardized Regressors

The MSPE of OLS and the Principle of Shrinkage

Estimation of the MSPE

14.3. Ridge Regression

Shrinkage via Penalization and Ridge Regression

Estimation of the Ridge Shrinkage Parameter by Cross Validation

Application to School Test Scores

14.4. The Lasso

Shrinkage Using the Lasso

Application to School Test Scores

14.5. Principal Components

Principals Components with Two Variables

Principal Components with k Variables

Application to School Test Scores

14.6. Predicting School Test Scores with Many Predictors

14.7. Conclusion

Appendix 14.1: The California School Test Score Data Set

Appendix 14.2: Derivation of Equation (14.4) for $k = 1$

Appendix 14.3: The Ridge Regression Estimator When $k = 1$

Appendix 14.4: The Lasso Estimator When $k = 1$

Appendix 14.5: Computing Out-of-Sample Predictions in the Standardized Regression Model

Chapter 15: Introduction to Time Series Regression and Forecasting

Table of Contents

15.1. Introduction to Time Series Data and Serial Correlation

Real GDP in the United States

Lags, First Differences, Logarithms, and Growth Rates

Autocorrelation

Other Examples of Economic Time Series

15.2. Stationarity and the Mean Squared Forecast Error

Stationarity

Forecasts and Forecast Errors

The Mean Squared Forecast Error

15.3. Autoregressions

The First-Order Autoregressive Model

The p th-Order Autoregressive Model

15.4. Time Series Regression with Additional Predictors and the Autoregressive Distributed Lag Model

Forecasting GDP Growth Using the Term Spread

The Autoregressive Distributed Lag Model

The Least Squares Assumptions for Forecasting with Multiple Predictors

15.5. Estimation of the MSFE and Forecast Intervals

Estimation of the MSFE

Forecast Uncertainty and Forecast Intervals

15.6. Estimating the Lag Length Using Information Criteria

Determining the Order of an Autoregression

Lag Length Selection in Time Series Regression with Multiple Predictors

15.7. Nonstationarity I: Trends

What Is a Trend?

Problems Caused by Stochastic Trends

Detecting Stochastic Trends: Testing for a Unit AR Root

Avoiding the Problems Caused by Stochastic Trends

15.8. Nonstationarity II: Breaks

What Is a Break?

Testing for Breaks

Detecting Breaks Using Pseudo Out-of-Sample Forecasts

Avoiding the Problems Caused by Breaks

Table of Contents

15.9. Conclusion

Appendix 15.1: Time Series Data Used in Chapter 15

Appendix 15.2: Stationarity in the AR(1) Model

Appendix 15.3: Lag Operator Notation

Appendix 15.4: ARMA Models

Appendix 15.5: Consistency of the BIC Lag Length Estimator

Chapter 16: Estimation of Dynamic Causal Effects

16.1. An Initial Taste of the Orange Juice Data

16.2. Dynamic Causal Effects

Causal Effects and Time Series Data

Two Types of Exogeneity

16.3. Estimation of Dynamic Causal Effects with Exogenous Regressors

The Distributed Lag Model Assumptions

Autocorrelated u_t , Standard Errors, and Inference

Dynamic Multipliers and Cumulative Dynamic Multipliers

16.4. Heteroskedasticity- and Autocorrelation-Consistent Standard Errors

Distribution of the OLS Estimator with Autocorrelated Errors

HAC Standard Errors

16.5. Estimation of Dynamic Causal Effects with Strictly Exogenous Regressors

The Distributed Lag Model with AR(1) Errors

OLS Estimation of the ADL Model

GLS Estimation

16.6. Orange Juice Prices and Cold Weather

16.7. Is Exogeneity Plausible? Some Examples

U.S. Income and Australian Exports

Oil Prices and Inflation

Monetary Policy and Inflation

The Growth Rate of GDP and the Term Spread

16.8. Conclusion

Appendix 16.1: The Orange Juice Data Set

Appendix 16.2: The ADL Model and Generalized Least Squares in Lag

Table of Contents

Operator Notation

Chapter 17: Additional Topics in Time Series Regression

17.1. Vector Autoregressions

The VAR Model

A VAR Model of the Growth Rate of GDP and the Term Spread

17.2. Multi-period Forecasts

Iterated Multi-period Forecasts

Direct Multi-period Forecasts

Which Method Should You Use?

17.3. Orders of Integration and the Nonnormality of Unit Root Test Statistics

Other Models of Trends and Orders of Integration

Why Do Unit Root Tests Have Nonnormal Distributions?

17.4. Cointegration

Cointegration and Error Correction

How Can You Tell Whether Two Variables Are Cointegrated?

Estimation of Cointegrating Coefficients

Extension to Multiple Cointegrated Variables

17.5. Volatility Clustering and Autoregressive Conditional Heteroskedasticity

Volatility Clustering

Realized Volatility

Autoregressive Conditional Heteroskedasticity

Application to Stock Price Volatility

17.6. Forecasting with Many Predictors Using Dynamic Factor Models and Principal Components

The Dynamic Factor Model

The DFM: Estimation and Forecasting

Application to U.S. Macroeconomic Data

17.7. Conclusion

Appendix 17.1: The Quarterly U.S. Macro Data Set

Chapter 18: The Theory of Linear Regression with One Regressor

18.1. The Extended Least Squares Assumptions and the OLS Estimator

The Extended Least Squares Assumptions

Table of Contents

The OLS Estimator

18.2. Fundamentals of Asymptotic Distribution Theory

Convergence in Probability and the Law of Large Numbers

The Central Limit Theorem and Convergence in Distribution

Slutskys Theorem and the Continuous Mapping Theorem

Application to the t-Statistic Based on the Sample Mean

18.3. Asymptotic Distribution of the OLS Estimator and t-Statistic

Consistency and Asymptotic Normality of the OLS Estimators

Consistency of Heteroskedasticity-Robust Standard Errors

Asymptotic Normality of the Heteroskedasticity-Robust t-Statistic

18.4. Exact Sampling Distributions When the Errors Are Normally Distributed

Distribution of b_{n1} with Normal Errors

Distribution of the Homoskedasticity-Only t-Statistic

18.5. Weighted Least Squares

WLS with Known Heteroskedasticity

WLS with Heteroskedasticity of Known Functional Form

Heteroskedasticity-Robust Standard Errors or WLS?

Appendix 18.1: The Normal and Related Distributions and Moments of Continuous Random Variables

Appendix 18.2: Two Inequalities

Chapter 19: The Theory of Multiple Regression

19.1. The Linear Multiple Regression Model and OLS Estimator in Matrix Form

The Multiple Regression Model in Matrix Notation

The Extended Least Squares Assumptions

The OLS Estimator

19.2. Asymptotic Distribution of the OLS Estimator and t-Statistic

The Multivariate Central Limit Theorem

Asymptotic Normality of b_n

Heteroskedasticity-Robust Standard Errors

Confidence Intervals for Predicted Effects

Asymptotic Distribution of the t-Statistic

Table of Contents

19.3. Tests of Joint Hypotheses

- Joint Hypotheses in Matrix Notation
- Asymptotic Distribution of the F-Statistic
- Confidence Sets for Multiple Coefficients

19.4. Distribution of Regression Statistics with Normal Errors

- Matrix Representations of OLS Regression Statistics
- Distribution of \mathbf{b} with Independent Normal Errors
- Distribution of \mathbf{su}^2
- Homoskedasticity-Only Standard Errors
- Distribution of the t-Statistic
- Distribution of the F-Statistic

19.5. Efficiency of the OLS Estimator with Homoskedastic Errors

- The GaussMarkov Conditions for Multiple Regression
- Linear Conditionally Unbiased Estimators
- The GaussMarkov Theorem for Multiple Regression

19.6. Generalized Least Squares

- The GLS Assumptions
- GLS When \mathbf{O} Is Known
- GLS When \mathbf{O} Contains Unknown Parameters
- The Conditional Mean Zero Assumption and GLS

19.7. Instrumental Variables and Generalized Method of Moments Estimation

- The IV Estimator in Matrix Form
- Asymptotic Distribution of the TSLS Estimator
- Properties of TSLS When the Errors Are Homoskedastic
- Generalized Method of Moments Estimation in Linear Models

Appendix 19.1: Summary of Matrix Algebra

Appendix 19.2: Multivariate Distributions

Appendix 19.3: Derivation of the Asymptotic Distribution of \mathbf{b}

Appendix 19.4: Derivations of Exact Distributions of OLS Test Statistics with Normal Errors

Appendix 19.5: Proof of the GaussMarkov Theorem for Multiple

Table of Contents

Regression

Appendix 19.6: Proof of Selected Results for IV and GMM Estimation

Appendix 19.7: Regression with Many Predictors: MSPE, Ridge Regression,
and Principal Components Analysis

Appendix

References

Glossary

Index