REGRESSION ANALYSIS II: Linear Models
Tuesday, June 27 - Friday, July 21, 2017 — 3:00 - 5:00 p.m.

Tim McDaniel
Buena Vista University — Storm Lake, Iowa USA — mcdaniel@bvu.edu
Joint appointment: The Siebens School of Business and the School of Science & Mathematics

Introduction

This course is intended for social scientists who are comfortable with algebra and basic introductory-level statistics, and now want to learn applied ordinary least squares (OLS) multiple regression analysis for their own research and to understand the work of others.

As social scientists, it is important that we know how to use multiple regression. But it is also important for us to know why and how multiple regression works (and fails) under varying conditions. Given this, we will discuss much of the mathematical and statistical theory behind multiple regression and also some potential drawbacks and circumstantial limitations. So while the presentations will not be purely theoretical, neither will this course be “cookbook” in nature.

The primary goal of the course is to develop an applied and intuitive (as opposed to purely theoretical or mathematical) understanding of the topics. Whenever possible, presentations will be in “Words,” “Picture,” and “Math” languages in order to appeal to a variety of learning styles.

Course Topics

We will begin with a quick review of basic univariate statistics and hypothesis testing. I am assuming that most (but probably not all...) of this material will indeed be a review for you.

After that we will cover various topics in bivariate and then multiple regression, including:

- Model specification and interpretation.
- Diagnostic tests and plots.
- Analysis of residuals.
- Dichotomous (dummy) variables.
- Outliers, influence, and leverage.
- Logit models and analysis, and Probit models.
- Advanced diagnostic plots and graphical techniques.
- Regression models from a matrix perspective.
- Various F-tests (e.g., group significance test; Chow test; relative importance of variables and groups of variables; comparison of overall model performance).
- Categorical independent variables (e.g., new tests for “Intervalness” and “Collapsing”).
- Heteroskedasticity: Definition, consequences, detection, and correction.
- Autocorrelation: Definition, consequences, detection, and correction.
- Generalized Least Squares (GLS) and Weighted Least Squares (WLS).
- Transformations to induce linearity.
- Multicollinearity.
- Multiplicative interaction terms.
- Categorical (e.g., Likert) variables.
Lecture Notes

This course will utilize approximately 890 pages of Lecture Notes. These Lecture Notes are organized in five Booklets. They will serve as the sole required “textbook” for this course and also as an information resource for you after the course ends. In addition, these Lecture Notes will significantly reduce the amount of notes you have to write during class, which means you can concentrate much more on learning and understanding the material itself.

A detailed outline of the Lecture Notes is included in this syllabus. Each of the five Booklets of Lecture Notes also has a Table of Contents with page number information.

Packets 1, 2, and 3 constitute “Booklet #1” of the Lecture Notes. This first Booklet is furnished during the first class to all participants taking the course. It will be used for the first two classes. The remainder of the twenty Lecture Notes Packets (Numbers 4-20, contained in Booklets #2, 3, 4, and 5) are available for purchase at the Photoduplicating Office, which is in Room 160 in the basement of the ISR building. The total cost is approximately $50.00 or so. We will begin using Booklet #2 on the third class day. Further details will be shared the first class day.

I urge you to purchase this year’s edition of the Lecture Notes, as opposed to using an earlier edition; I have made several substantive additions, deletions, and other changes this year.

Although these Lecture Notes are detailed, comprehensive, and self-contained, it is still advisable for you to study the relevant Packets before and (especially!) after each class, to ask questions during class, and to talk to either me or a Teaching Assistant (TA) outside of class so you can maximize your learning and other benefits from this course. The Lecture Notes contain several algebraic derivations and proofs. We will not use class time to work through most of them, but instead they are provided for your information and consideration outside of class.

Textbooks and Other Readings

There is no required textbook (other than the aforementioned five Lecture Notes Booklets, which constitute a de facto textbook) for this course. Later in this syllabus I provide you with information about several optional supplemental readings from various textbooks and journals, organized across each of the twenty Packets of the required Lecture Notes Booklets. See the beginning of the “Some Suggested Readings” section for a discussion of the four textbooks (each a different style) included in these optional readings.

I do not think that any one of these four textbooks is significantly better than the others; instead, the one(s) that you might find best will depend upon a number of personal factors, including your particular learning style. Therefore, instead of just picking one textbook, I have designed the course so that you can experiment and pick-and-choose which style of learning and (therefore) textbook(s) you prefer. Of course, you may decide to read and study (and maybe purchase) none, one, two, three, or even all four of these textbooks; again, that is entirely up to you. See the alphabetized list of “Some Suggested Readings” near the end of this syllabus for the textbook readings corresponding to each Packet of the Lecture Notes. Finally, other textbooks are also appropriate for this course; see me if you have any questions about this.
At the end of this syllabus is a bibliography for the textbooks and all other optional readings. Note that I have included quite a few of the Sage “Little Green Book” monographs. I like these a lot. You may notice that some (though certainly not all) of the other readings are from political science books and journals; however, this is not a cause for either rejoicing or concern since they deal with methodological topics that are easily and broadly generalizable across other social science disciplines (I would not have selected them otherwise!). I was careful to select readings and substantive examples from other areas and journals (e.g., criminal justice, economics, law, sociology, and social science in general). Everything listed in the bibliography is available from the Summer Program Library. Before you actually purchase anything I strongly suggest that you first “check it out” (in more ways than one!) from the Summer Program Library.

The Optional Session

I will hold an Optional Session the evening of our first class day (Tuesday); the time and room will be announced during the first day’s class. During this optional evening session I will continue and finish a review of basic univariate statistics that we will have started earlier that day (in the first class). The Lecture Notes for this review are in the Packets 1 and 2 (in Booklet #1). Some of you will find a continuation and completion of this review useful, while others may decide not to attend this first evening’s session (remember it is optional!). In either case, you at least will have the Lecture Notes for this review of basic statistics. We will start Packet 3—“Basic Regression”—at the beginning of the second (Wednesday’s) class.

Assignments and Software

The first assignment covers some basic regression terminology, notation, and concepts. Then for the remaining five assignments you use data to generate computer output; that will be relatively trivial. Then your substantive task will be to interpret, analyze, and explain that output.

I will distribute an extensive and comprehensive “Tutorial and Answer Key” for each assignment. Each of these will provide you with yet another excellent learning opportunity.

This is a course on Regression Analysis, NOT on statistical software. The emphasis, including in the assignments, is on understanding and interpreting (not generating) output. That said...

• Since it is so quick to learn and easy to use, SPSS is used to generate all in-class output.
• But you can use any computer software you choose to generate your own output!
• The TAs will provide minimal (but sufficient) Tutorial Workshops over Stata and R.
  ◦ These will cover just the basics needed to replicate our in-class examples...
  ◦ ...So that you can generate all of the output required to complete the assignments.
• The TAs are primarily a resource for assistance in explaining substantive material and interpreting output, but also for basic help generating output using SPSS, R, and Stata.

The data files for all of the assignments and in-class examples will be sent to you via email.

IMPORTANT: You should attend some of the “Introduction to Computing” sessions if you are not already comfortable with basic computer operations or with any relevant-to-you software.
Matrix Algebra

I have designed the course so that matrix algebra will not be used during the first three weeks. Instead, you will learn as much of the regression material as possible within a simple scalar algebra environment. Packet 15, which you can study during the weekend between the third and fourth weeks, consists of an introduction to matrices and matrix algebra. Then parts of some of the presentations covering the more advanced topics in the fourth week will involve matrices.

IMPORTANT: You should attend the first two weeks or so of the Lecture Series “Mathematics for Social Scientists II” if you are at all deficient in your understanding of matrix algebra.

Learning, Teaching Assistants, and the Lecture Notes

The primary purpose of the class meetings, Lecture Notes, Assignments, and the Optional Learning Exercise is to help you learn this material. You are not in this alone. Studying and learning with other participants is probably a very good idea for many of you, as is taking advantage of Office Hours opportunities involving myself and the Teaching Assistants.

The Teaching Assistants and I make every effort to be accessible to you. I encourage you to attend our office hours or make an appointment if those hours do not fit your schedule. Early in the course I will give you information (e.g., an office hours schedule) regarding these matters.

You need to read, search, and study the Lecture Notes as you do the homework assignments and before you ask any questions. After that, if you still have a question then that is fine; ask away! Now the resulting discussion will be much better informed, more efficient, and more productive.

Also: The outline in this syllabus, and the Table of Contents (which includes page numbers) in the front of each of the five Lecture Notes booklets, might be useful to you...so use them!

Practical Matters Involving Assignments and Grading

Quantity. There are six Assignments (100 total points) that count toward your course grade:

- Assignment #1 — “Welcome to Regression!” — 10 points.
- Assignment #2 — “Bivariate Regression” — 15 points.
- Assignment #3 — “Inducing Linearity” — 15 points.
- Assignment #4 — “Multiple Regression” — 15 points.
- Assignment #5 — “Interaction Terms” — 25 points.
- Assignment #6 — “Categorical Independent Variables” — 20 points.

Optional Learning Exercise — “Logit Models” — Neither submitted nor graded.

Purpose. The primary purpose of the Assignments is to further enhance your learning of the material. The Assignments also serve as the sole evaluation vehicle for assigning grades.

Answer Keys. Everyone in the class will receive a “Tutorial and Answer Key” for each Assignment, immediately after it is due. Use each of these as an opportunity to learn!
Due Date. You will submit each Assignment at the beginning of class on the day it is due.

Details Matter. Your work will be graded on its quality, clarity, completeness, and accuracy.

- You can either type your work or write it out by hand. In either case, make it very easy to read, follow, and understand (e.g., be organized and if you write then print neatly).
- Clearly label which question you are answering (e.g., “A” or “B” or “C”).
- Show all of your work and supporting evidence, not just your bottom-line answer.
- Neatly write (print!) your name and the Assignment Number at the top of the first page.
- Staple all of the submitted sheets of paper together, in the upper-left-hand corner.

Show Code. If you write code (as opposed to use a “point-and-click” interface)—e.g., perhaps using Stata or R—to generate output, then you need to include that code as part of your answer.

No Late Submissions. Unless we have made specific and explicit arrangements beforehand, no assignment submitted late (i.e., after the due date and time) will be accepted.

Not “For Credit”? Not a Problem! If you are not taking the course “for credit” then you can still submit none, some, or all of the Assignments for grading; it is completely up to you. Any work you decide to submit will be evaluated as if you were taking the course for credit.

Grading Templates. In order to facilitate more efficient and accurate grading by the Teaching Assistants, and to enhance your understanding of partial credit decisions and your total numerical grade, each Assignment will be returned along with a completed Grading Template form.

Grade Appeals. If you ever have a question about your grade (e.g., a partial credit decision) then you need to see me (not a TA!) to discuss the situation. Here are some relevant course policies:

- You must wait until at least the next day after receiving your grade. This “Next Day Rule” gives you time to study and contemplate your work-product, your notes, the Lecture Notes, the Grading Template, and the relevant Tutorial and Answer Key...and then to think about your work and the corresponding grading decision.
- The maximum amount of time you have to appeal a grading decision is three class days after that Assignment is returned; after then no grading appeals will be considered.
- All grading appeals must be made before the end of my last day’s office hours.

Working Together and Doing Your Own Work. I encourage you to work and learn with other participants on all learning activities in this course, including the Assignments. However, all of the work on each Assignment you submit needs to be your work; i.e., you need to produce your own submitted work-product material. For example, you cannot simply copy-paste (or in any other way simply copy) the work of others. See me if you have any questions about this.

Conclusion

Learning the material in this course will require a substantial amount of effort...but that is why you are here! The payoff will be well worth it. Let me know if the TAs or I can be of any assistance to you in this endeavor. Finally, it is an honor to be your instructor for this course.
Course (and Lecture Notes) Outline

Packet I. Basic Statistics Review
1. Summations and Sigma Notation
2. Basic Statistics
   1. Mean
   2. Variance and Standard Deviation
   3. Probability
   4. Random Variables
      a. Continuous versus Discrete
      b. Nominal, Ordinal, and Interval
   5. Standardized Variables
   6. Expected Value
   7. Covariance, Correlation, and Causality
   8. Independence
   9. Normal Distribution: Notation, Skewness & Kurtosis, Areas
10. Central Limit Theorem
11. Student’s t Distribution
12. Hypothesis Testing
13. Prob-Values (“p-Values”)
14. Confidence Intervals
15. Properties of Estimators: Unbiased, Consistent, Efficient, Sufficient

Packet II. Supplement to Basic Statistics Review
2. A Closer Look at Population and Sample Variances
3. Hypothesis Testing: Summary, Flowchart, Protocol, and p-Values
4. Some Abuses and Misuses of Probability and Statistics
5. Symbol Glossary

Packet III. Bivariate Regression
3. Notation
4. Fitting a Line
5. Ordinary Least Squares Assumptions
6. Assignment #1
   D. Deriving the Sample Intercept and the Slope Coefficient Using Algebra
   E. Centered Variables
   F. The Estimated Slope Coefficient (“b”)
      1. Variance and Confidence Interval
      2. Confidence Interval and Hypothesis Testing
   G. The Gauss-Markov Theorem
   H. Appendix: Deriving the Formulas for “a” and “b” Using Calculus
# Residuals

1. **Definition and Estimation**

2. **Population “Error” as a Disturbance, or Stochastic Element**

# Explained, Unexplained, and Total Deviations and Sums of Squares

# Goodness of Fit

1. **Coefficient of Determination (R-Squared)**
   
   a. Correlations (Again...)
   
   b. Why the R-Squared Can Be Inappropriate and Misleading
   
   c. Perils of Maximizing R-Squared: A Monte Carlo Simulation

2. **Standard Error of Regression (SER)**

# Standardized Variables and Beta Weights

# Reporting OLS Regression Results

# Regression Forced Through the Origin

1. **Definition, Illustrations, and Examples**

2. **The Importance of Theoretical and Substantive Justifications**

# Comparison of Centering, Standardizing, and Forcing Through the Origin

# Another Note on the Meaning and Interpretation of “a,” “b,” and Ŷ

# An Analogy: Means, Slopes, Standardization, Samples and Populations

---

# Functional Transformations of Independent Variables

1. **The Need for Transformations**

2. **The Regression is Still Linear**

3. **The Natural Log and the Square Root Transformations**

4. **The X-Squared Transformation (a First Look)**

5. **Presenting Findings with Transformed Independent Variables**

---

# Interpolation, Predictive Intervals, and Extrapolation

# Some Simple Diagnostic Plots

1. **Y vs. X (and an Introduction to Outliers)**

2. **Y vs. Ŷ**

3. **Residual vs. X**

4. **Residual vs. Case Number**

5. **Residual vs. Lagged Residual**

6. **Residual vs. Ŷ**

# Simpson’s Paradox, Aggregation Bias, and the Ecological Fallacy

# Generation and Interpretation of Computer Output Using Real Data

1. **Setting Up the Substantive Example**

2. **Generating and Analyzing the Output**

# The Usefulness of Simple Scatter Plots: An Illustration

# R-Squared, “b,” and SER: A Monte Carlo Simulation
IV. Multiple Regression

A. Limits of Bivariate Regression

B. Trivariate Regression
   1. Visualization
   2. The Residual Term
      a. Definition
      b. The SER and Degrees of Freedom
   3. The (Two) Estimated Slope Coefficients
      a. Partial Effects Equations, Models, and Venn Diagrams
      b. Computing the Two Slope Coefficients
      c. Variance and Confidence Intervals
      d. Meaning and Interpretation
   4. Holding One Variable Constant: What’s That All About?
   5. The Impact on OLS Assumptions

C. Multiple Regression: The General OLS Model
   1. The Slope Coefficient
      a. Partial Effects Equations
      b. Computing the Slope Coefficients
      c. OLS Assumptions and the Gauss-Markov Theorem
      d. Meaning and Interpretation
      e. Variance and Confidence Intervals
   2. The Residual Term
      a. Definition
      b. The SER and Degrees of Freedom
   3. Degrees of Freedom: From One to Many Variables

4. Summary and Review of Partial Effects and the Interpretation of “b”
5. Summary and Review of t Stats, p-Values, and Hypothesis Tests
6. Units of the SER and Comparing SER’s Across Equations
7. R-Squared and Adjusted R-Squared
   a. Review of the R-Squared Statistic and Inappropriate Uses
   b. R-Squared and Functional Transformations
   c. Multiple Independent Variables and the Effect on R-Squared
   d. Understanding and Interpreting the Adjusted R-Squared Value
8. Multiplicative Interactions
   a. Introduction, Description, Analogy to Functional Transformations
   b. Models with an Interaction of Two Dummy Variables
   c. Models with an Interaction of Two Continuous Variables
   d. Interpreting and Understanding: Algebra, Components, and Theory
   e. Conditional Impact: “Can I Exclude a Stand-Alone Term?”
   f. Presenting Results, Graphs, and Significance Tests
   g. Additional Readings and Concluding Remarks
   h. Bivariate Regression and Interaction Term Models: A Comparison
Packet 9

9. Multicollinearity and Multicollinearity Diagnostics
   a. Perfect Multicollinearity: An Example
   b. Auxiliary R-Squared, Tolerance, and Variance Inflation Factor
   c. Explaining Multicollinearity Using Venn Diagrams
   d. Consequences, Including Possible “Backdoor Bias”
   e. When to Suspect Multicollinearity Problems
   f. Do Not Use Bivariate Correlations to Diagnose Multicollinearity
   g. Possible Remedies
   h. A Statistics, Estimation, and Information (Not a Theory!) Problem
   i. Multicollinearity and Models with Interaction Terms
   j. Narrowing Down the Source of the Multicollinearity

10. Dummy and Categorical Independent Variables
    a. Definition and Interpretation
    b. The Importance of “Intervalness” for Independent Variables
    c. Replacing a Categorical Independent Variable with Dummies
        i. Implementation and Interpretation
        ii. Excluding One Dummy Variable from the Model
    d. Graphing Models with Dummy Variables (and Interactions)
    e. Comparison: “Regression Forced Through Origin” to “Having X in an Interaction, but Not as a Stand-Alone, Term”
    f. Interpretation of Category Dummies and “Jumps”

Packet 10

11. Functional Transformations
    a. Things to Consider Regarding Transforming Y
    b. Log Transformations and Constant Elasticity Models
    c. Conditional Impacts and Slopes in Models with Interaction Terms
    d. More on Models with an X-Squared (e.g., Threshold Models)
    e. The Bend Rule

12. Model Specification
    a. Review of Types of Specification Error
    b. Omitting Relevant Variables: Derivation and Consequences
    c. Including Irrelevant Variables: Derivation and Consequences
    d. Variable Selection
    e. Perils of Stepwise Regression
    f. An Alternative to Standardization for Interval-Level Discrete X’s

13. Missing Data
    a. Data Missing at Random: Dependent and Independent Variables
    b. “Solutions” and Their Potential Problems
        i. Casewise (Listwise) Deletion
        ii. Pairwise Deletion
        iii. Mean and Conditional Mean Substitution
        iv. Other Methods (e.g., Profile Models)

14. Measurement Errors: In Y and in an X

15. Partial Effects Plots and Linearity in Multiple Regression Models
Packet 11
16. Summary: The Effects of Multicollinearity and Specification Errors on Slope Coefficient Estimation and Hypothesis Testing
   a. Type I and Type II Errors
   b. Flowcharts: Review of the Logic of Hypothesis Testing
   c. Essay: Hypothesis Testing and American Criminal Trials

Assignment #4
and
Assignment #5
Given
17. Review: Category Dummy Variables
18. Review: Models with Continuous, Dummy, and Interaction Terms
19. Review: Diagnostic Partial Plots
20. Omitted Variable Plots
21. Another Look at Outliers: What They Are and Why They Matter
23. Multiple Regression: A Computer Example Using ANES Data
   a. Setting Up the Example (American National Election Study)
   b. Generation and Interpretation of Computer Output
      i. Control Variables, Multicollinearity, and Diagnostic Plots
      ii. An Illustration of Partial Effects
      iii. Categorical X’s: Intervalness, Jumps, and Collapsing
24. Subgroup Differences: “Dummy*Continuous” Interaction Terms

Packet 12
V. Analysis of Variance and the F Test
   A. The F Distribution
   B. Total, Regression, and Error Sum of Squares...and the ANOVA Table
   C. The F Statistic and the F Test
      1. A Test Involving All of the Regression Coefficients
      2. When the F Test and Student’s t Test Are Identical
   D. Analysis and Demonstrations Using a Regression Simulation
      1. Interpreting the F Test: A “Randomly Generated Values” Illustration
      2. A Closer Look at R-Squared, Adjusted R-Squared, and SER

Packet 13
VI. Categorical Independent Variables (and F Tests): A Closer Look
   A. Problems using Categorical and Non-Interval Independent Variables
   B. Review: Replacing a Categorical X with Dummy Variables in OLS
   C. Various Tests Involving the F Distribution
      1. A Review of the Whole-Model F Test in a Regression Environment
      2. Nested F Tests: Testing Groups (Subsets) of Regression Coefficients
         a. Purpose, Hypotheses, and Procedures
         b. Group Significance of Multicollinear Variables (e.g., Interactions)
         c. Equivalent to Testing the SER’s
      3. The Chow Test
         a. Purpose, Hypotheses, and Procedures
         b. Applications: Structural Shifts and Whole-Model Performance
         c. Generalized (Multiple Subgroup) Chow Test and Aggregation Bias
      4. Overview of Comparing “Safe” versus “Risky” Models

(Packet 13 is continued on the next page)
(Packet 13, continued from the previous page)

5. Comparison of SER’s with No Prior Knowledge of Relative Model Performance (i.e., When There is No “Safe” or “Risky” Model)


7. A General Discussion of F Tests, Using ANOVA Terminology

D. Collapsing Categories of a Categorical Variable
   1. Setup and Notation
   2. Statistical Hypothesis Test for Collapsing Categories

E. Treating a Categorical Variable as Interval-Level
   1. Setup and Notation, and “Contextually Interval”
   2. Statistical Hypothesis Test for Intervalness

F. Monte Carlo Simulation Results: Which Dummy to Exclude

G. Generalizations of the F Test
   1. For Collapsing Categories and for Intervalness
   2. For “Safe” and “Tested” Models in General

H. Optional: Discussion and Analysis of a Substantive Example (Venezuela)
   1. Description of the Study
   2. Statistical Tests Performed
   3. Potential Effects of Multicollinearity
   4. Different Specifications, Interpretations, and Conclusions
   5. Interpretation of Manipulated Coefficients

Packet 14

VII. Dichotomous Dependent Variables and the Logit Model

A. Examples

B. The Binary Choice Model
   1. Description and Illustration
   2. Problems Using OLS with a Dichotomous Dependent Variable
      a. Over- or Under-Estimating Y
      b. Var(e), Heteroskedasticity, and the Goldberger Procedure
      c. Probable Non-Linearity
   3. Using an S-Shaped Curve Instead of a Line
      a. Description and Theory
      b. Fit and Bias

(Note: This)

C. The Logit Model

Optional Learning

1. The Logistic Function and Log-Odds
2. Interpreting Logit Coefficients
3. Interpreting and Presenting the Results of Logit Analysis: An Example
4. How Logit Works: Big Picture and Comparison to Linear OLS
5. Estimating Coefficients: Maximum Likelihood Estimation

Exercise is not part of your grade.)

6. Significance of Coefficients: The Likelihood Ratio Test and the Z Test
7. Using and Interpreting Logit Output: An Education Example

Exercise Given

(Continued on next page)
Packet VIII. Matrices and Matrix Algebra
15  A. General Terms and Definitions
    B. Addition, Subtraction, and Multiplication: How to Do It and Examples
    (For you to review over the weekend)
    C. Inverse Matrices
    D. Determinants and Condition Numbers
    E. Rank and Singularity
    F. Properties of Matrix Algebra
    G. Kronecker (Tensor) Product
    H. Eigenvalues and Eigenvectors
    I. Positive Definite Matrices
    J. Expressing Linear Equations in Matrix Form: A Taste of Things to Come

Packet IX. Outliers and Graphical Techniques (“Weird Points and Pretty Pictures”) 16
A. Outliers: Detection, Classification, and Possible Effects
1. Influence
   a. Definition and Description
   b. Detection Using Statistics
      i. DFBETA and DFBETAS
      ii. Cook’s Distance (D) and DFFITS
2. Leverage
   a. Definition and Description
   b. Hat Values
3. Studentized Residuals
   a. Definition, Calculation, and the Bonferroni Adjustment
   b. A Note on Notation and Terminology (It’s a Mess...)
   c. Case Dummies and Joint Cases
B. Diagnostic Plots and Graphical Techniques
   1. Proportional Leverage Plots: Definition and Different Types
   2. Jointly Influential Observations: Definition and Examples
   3. Conditional Effects Plots
      a. Definition, Description, Presentation, and Interpretation
      b. Models with Interaction Terms (Including Confidence Bounds)
      c. The Importance of “Conditional”
   4. Jittered Data Plots (for Categorical Independent Variables)
   5. Scatterplot Matrix

Packet X. Multiple Regression Using Matrices
17  A. The General Linear Model
    B. OLS Using Matrices
    C. Vectors and Matrices Involving Residuals
    D. OLS Assumptions in Matrix Form
    E. Variance-Covariance Matrices
    F. Deriving “b” and Var(b)
    G. SER, R-Squared, and Adjusted R-Squared in Matrix Form
    H. Review: Multicollinearity

(Packet 17 is continued on the next page)
I. Ordinal and Dummy Variables: An Example Using Matrices
J. Example: A Singular X’X Matrix
K. Transformations and Interactions in Matrix Form: OLS is Still Linear
L. Appendix: A Proof that the Matrix Derivation of “B” Minimizes the SSE

Packet XI. Heteroskedasticity (and Generalized/Weighted Least Squares)
A. Definition and Omega Matrices
B. Variance-Covariance Matrices
C. Detection of Heteroskedasticity Using Diagnostic Plots
D. Types of Heteroskedasticity
E. When to Expect Heteroskedasticity
F. Consequences of Using OLS with Heteroskedastic Errors
G. Detection of Heteroskedasticity Using Statistical Tests
   1. Goldfeld-Quandt Test
   2. Glejser Test
   3. Likelihood Ratio Test
H. Correcting for Heteroskedasticity Using Generalized Least Squares
   1. Introduction, and Deriving the Slope Coefficients using GLS
   2. Benefits of Using GLS When You Have Heteroskedastic Errors
   3. GLS as Weighted Least Squares (WLS)
   4. The Underlying Logic of WLS: An Example Using State-Level Data
   5. Using GLS Given a Model for Var(e): The General Case

Packet XII. Autocorrelation
A. Definition of First-Order Autocorrelation
B. When to Expect Autocorrelation
C. Algebraic Interpretation of First-Order Autocorrelation
D. Examples of Positive and Negative First-Order Autocorrelation
E. Four Common Types of Time Series Models
F. Consequences of Using OLS with Autocorrelation
G. Detection of Autocorrelation
   1. Diagnostic Plots
   2. Using a Bivariate Model
   3. The Durbin-Watson Statistic (First-Order Autocorrelation)
      a. Algebraic Derivation and Implementation
      b. Some Things to Keep in Mind
   4. The Wallis Test (Fourth-Order Autocorrelation)
H. Correcting for First-Order Autocorrelation
   1. The Prais-Winsten GLS Estimator: Algebraic Procedure and Proofs
   2. Some Other GLS Methods
Packet XIII. More on Dichotomous Dependent Variables: The Probit Model

20

A. The Normal Distribution Revisited
   1. The Standard Normal Distribution
   2. The Normal (Gaussian) CDF
B. The Probit Function
C. Finding Probabilities Using the Normal Distribution
D. Interpreting and Presenting Probit Coefficients
E. Maximum Likelihood Estimation of Probit Models
F. Deciding Between Using Logit versus Probit Models

XIV. Chart: Some Potential Problems When Using OLS Regression Models

XV. Final Remarks
   A. More than One Equation: Simultaneous Equations Models
   B. Additional Topics for Further Study
   C. The Importance of Parsimony and Presentation
   D. The Dangers of Over-Reliance on Statistical Procedures

Some (Optional) Suggested Readings

Again: The Lecture Notes that I wrote and we use in each class serve as the de facto required textbook for this course. I advise you to use the outline in this syllabus (along with the Table of Contents at the start of each Booklet) as a guide for these Lecture Notes.

The additional readings in this section of the syllabus (details of which can be found in the Bibliography section that immediately follows) are totally optional for you in this course.

Several of these readings are from the following four traditional textbooks, each of which I like a lot. While there are many similarities between them, I think that each takes a somewhat different (though at times only subtly different) pedagogical approach.

- **Gujarati** takes more of a “Math Language” (but not too advanced or difficult) approach.
- **Hamilton** takes more of a “Picture Language” (lots of graphics, etc.) approach.
- **Kennedy** takes more of an “English Language” (narrative explanations) approach.
- **Wooldridge** takes more of a “Combined” (and accessible and comprehensive) approach.

Also, Fox’s (2017) textbook provides broad coverage, thorough explanations, and creative uses of graphics that are among the best I have ever seen. This book might (?) be a bit too advanced for some participants in this course, but it could be a wonderful post-course reference. So I have included it in the bibliography, but not in the list of suggested optional readings, in this syllabus.

You will notice several “little green books” from the QASS Sage Series and also numerous articles from a number of journals across several academic disciplines. Both the more recent ones and the older classics are very useful as learning vehicles...which is why I included them!

The suggested optional readings for each Packet of the Lecture Notes are on the next four pages.
<table>
<thead>
<tr>
<th>Lecture Packet</th>
<th>Suggested Optional Readings (In Alphabetical Order)</th>
</tr>
</thead>
</table>
| Packets 1 & 2  | Baumgartner, Breunig, et al. (especially pages 606-613)  
Gujarati and Porter: Intro; Sections 5.5-5.8, 6A.2; Appendix A  
Hamilton: Pages 289-296  
Kennedy: Sections 2.5-2.8  
Mohr  
Wooldridge: Appendices A.1-A.4, B, C.2-C.6 |
| Packet 3       | Gujarati and Porter: Chapter 1; Sections 2.1, 2.2., 2.6, 2.7, 3.1-3.4, 3A.1-3A.7, 4.1-4.3, 5.1-5.3  
Hamilton: Pages 29-34, 42-49, 296-297  
Kennedy: Chapter 3; Sections 1.1, 1.4, 2.1-2.3, 2.11  
Lewis-Beck and Lewis-Beck: Pages 1-14, 23-38  
Schroeder, Sjoquist, and Stephan: Pages 11-23, 81-82  
Wooldridge: Chapter 1; Sections 2.1-2.5 |
| Packet 4       | Achen (1982): Pages 73-77  
Achen (1991)  
Gujarati and Porter: Sects. 2.4, 2.5, 3.5, 3.8, 5.4, 5.11, 6.1-6.3, 6A.1  
Hamilton: Pages 37-41, 49-51, 124-125  
Kennedy: Sections 1.2, 2.4, 2.10; Pages 109-110  
King (1986)  
King (1991)  
Lewis-Beck and Skalaban  
Schroeder, Sjoquist, and Stephan: Pages 23-29, 31-32  
Wooldridge: Section 2.6 |
| Packet 5       | Gujarati and Porter: Sections 2.3, 6A.3; Pages 164-166  
Hamilton: Pages 53-58, 148 |
Gujarati and Porter: Sections 3.6, 3.7, 5.10  
Lewis-Beck and Lewis-Beck: Pages 19-20, 39-49, 86-88 |
| Packet 7       | Achen (1982): Pages 7-51  
Asher: Pages 237-248  
Berry: Pages 1-24, 81-83  
Berry and Feldman: Pages 9-15  
Fox (1991): Pages 3-9  
Gujarati and Porter: Sections 7.1-7.4, 7.6, 8.1-8.3, 8.8  
Hamilton: Pages 65-72, 109-113  
Kennedy: Section 1.3  

(Packet 7 is continued on the next page)
<table>
<thead>
<tr>
<th>Lecture Packet</th>
<th>Suggested Optional Readings (In Alphabetical Order)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Packet 7</td>
<td>Lewis-Beck and Lewis-Beck: Pages 55-64</td>
</tr>
<tr>
<td>(Continued from previous page)</td>
<td>Schroeder, Sjoquist, and Stephan: Pages 29-31</td>
</tr>
<tr>
<td></td>
<td>Wooldridge: Chapters 3, 5; Sections 4.1-4.3</td>
</tr>
<tr>
<td>Packet 8</td>
<td>Berry: Pages 24-27</td>
</tr>
<tr>
<td></td>
<td>Berry and Feldman: Pages 15-16</td>
</tr>
<tr>
<td></td>
<td>Brambor, Clark, and Golder</td>
</tr>
<tr>
<td></td>
<td>Braumoeller</td>
</tr>
<tr>
<td></td>
<td>Davenport, Armstrong, and Lichbach</td>
</tr>
<tr>
<td></td>
<td>Friedrich</td>
</tr>
<tr>
<td></td>
<td>Gujarati and Porter: Sections 7.5, 7.8, 9.6</td>
</tr>
<tr>
<td></td>
<td>Hamilton: Pages 77-80, 84-85</td>
</tr>
<tr>
<td></td>
<td>Kennedy: Section 5.5</td>
</tr>
<tr>
<td></td>
<td>Lewis-Beck and Lewis-Beck: Pages 69-71</td>
</tr>
<tr>
<td></td>
<td>Schroeder, Sjoquist, and Stephan: Pages 32-51, 54-56, 58-59</td>
</tr>
<tr>
<td></td>
<td>Wooldridge: Section 6.3</td>
</tr>
<tr>
<td>Packet 9</td>
<td>Asher: Pages 248-250</td>
</tr>
<tr>
<td></td>
<td>Berry: Pages 45-47</td>
</tr>
<tr>
<td></td>
<td>Berry and Feldman: Pages 37-50, 64-70</td>
</tr>
<tr>
<td></td>
<td>Fox (1991): Pages 10-21</td>
</tr>
<tr>
<td></td>
<td>Gujarati and Porter: Chapter 10; Sections 8.5, 9.1-9.4</td>
</tr>
<tr>
<td></td>
<td>Hamilton: Pages 82, 85-92, 133-136</td>
</tr>
<tr>
<td></td>
<td>Hardy: Pages 1-21, 29-48</td>
</tr>
<tr>
<td></td>
<td>Kennedy: Chapter 12; Sections 15.1-15.4</td>
</tr>
<tr>
<td></td>
<td>Lewis-Beck and Lewis-Beck: Pages 64-69, 75-83</td>
</tr>
<tr>
<td></td>
<td>Schroeder, Sjoquist, and Stephan: Pages 56-58, 71-72</td>
</tr>
<tr>
<td></td>
<td>Wooldridge: Sections 7.1-7.3, 7.6; Pages 240-245</td>
</tr>
<tr>
<td>Packet 10</td>
<td>Achen (1982): Pages 51-73; 77-79</td>
</tr>
<tr>
<td></td>
<td>Asher: Pages 250-255</td>
</tr>
<tr>
<td></td>
<td>Baumgartner, Breunig, et al. (especially pages 613-615)</td>
</tr>
<tr>
<td></td>
<td>Berry: Pages 27-41, 60-66</td>
</tr>
<tr>
<td></td>
<td>Berry and Feldman: Pages 18-37, 71-72</td>
</tr>
<tr>
<td></td>
<td>Clarke</td>
</tr>
<tr>
<td></td>
<td>Fox (1991): Pages 53-61</td>
</tr>
<tr>
<td></td>
<td>Gujarati and Porter: Sections 6.4, 6.5, 6.8, 6.9, 7.7, 7.9, 7.10, 13.1-13.3, 13.5, 13A.1-13A.3; Pages 499-500</td>
</tr>
<tr>
<td></td>
<td>Halvorsen and Palmquist</td>
</tr>
<tr>
<td></td>
<td>Hamilton: Pages 72-77, 82-84, 148-153, 163-167, 173-174</td>
</tr>
<tr>
<td></td>
<td>Kennedy: Sections 5.1-5.4, 6.1-6.3; Pages 49 (the first point), 111</td>
</tr>
<tr>
<td></td>
<td>Lewis-Beck and Lewis-Beck: Pages 49-52, 72-73</td>
</tr>
<tr>
<td></td>
<td>Schroeder, Sjoquist, and Stephan: Pages 53, 59-61, 65-71</td>
</tr>
<tr>
<td></td>
<td>Wooldridge: Sections 6.1, 6.2, 9.4; Pages 303-306, 324</td>
</tr>
<tr>
<td>Lecture Packet</td>
<td>Suggested Optional Readings (In Alphabetical Order)</td>
</tr>
<tr>
<td>----------------</td>
<td>----------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td>Achen (2005)</td>
</tr>
<tr>
<td></td>
<td>Berry: Pages 41-45, 49-60</td>
</tr>
<tr>
<td></td>
<td>Eisenberg and Wells</td>
</tr>
<tr>
<td></td>
<td>Hamilton: Pages 158-163</td>
</tr>
<tr>
<td></td>
<td>Lewis-Beck and Lewis-Beck: Pages 88-93</td>
</tr>
<tr>
<td></td>
<td>Schroeder, Sjoquist, and Stephan: Pages 63-65</td>
</tr>
<tr>
<td></td>
<td>Wooldridge: Sections 4.6, 6.4</td>
</tr>
<tr>
<td>Packet 12</td>
<td>Berry and Feldman: Pages 17-18</td>
</tr>
<tr>
<td></td>
<td>Gujarati and Porter: Sections 5.9, 8.4</td>
</tr>
<tr>
<td></td>
<td>Hamilton: Pages 92-101</td>
</tr>
<tr>
<td></td>
<td>Mock and Weisberg</td>
</tr>
<tr>
<td></td>
<td>Schroeder, Sjoquist, and Stephan: Pages 51-52</td>
</tr>
<tr>
<td></td>
<td>Wooldridge: Pages 152-153</td>
</tr>
<tr>
<td></td>
<td>Gujarati and Porter: Sections 8.7, 9.5, 13.7</td>
</tr>
<tr>
<td></td>
<td>Hamilton: Pages 80-82, 101</td>
</tr>
<tr>
<td></td>
<td>Hardy: Pages 21-29, 48-53, 78-82, 84-85</td>
</tr>
<tr>
<td></td>
<td>Kennedy: Sections 4.3, 4.4</td>
</tr>
<tr>
<td></td>
<td>McDaniel</td>
</tr>
<tr>
<td></td>
<td>Wooldridge: Section 4.4; Pages 143-152, 245-248</td>
</tr>
<tr>
<td>Packet 14</td>
<td>Ai and Norton</td>
</tr>
<tr>
<td></td>
<td>Berry: Pages 47-49</td>
</tr>
<tr>
<td></td>
<td>Gujarati and Porter: Sections 15.1-15.8, 15A.1</td>
</tr>
<tr>
<td></td>
<td>Hamilton: Chapter 7</td>
</tr>
<tr>
<td></td>
<td>Hardy: Pages 75-78</td>
</tr>
<tr>
<td></td>
<td>Kennedy: Sections 2.9, 4.5, 16.1</td>
</tr>
<tr>
<td></td>
<td>Markus: Pages 555-556</td>
</tr>
<tr>
<td></td>
<td>Menard: Chapters 1-4; Appendix</td>
</tr>
<tr>
<td></td>
<td>Pampel: Chapters 1-3; Appendix</td>
</tr>
<tr>
<td></td>
<td>Schroeder, Sjoquist, and Stephan: Pages 79-80</td>
</tr>
<tr>
<td></td>
<td>Wooldridge: Sections 7.5, 17.1</td>
</tr>
<tr>
<td>Packet 15</td>
<td>Gujarati and Porter: Appendices B.1-B.3</td>
</tr>
<tr>
<td></td>
<td>Hamilton: Pages 333-337, 344-345</td>
</tr>
<tr>
<td></td>
<td>Wooldridge: Appendices D.1-D.5</td>
</tr>
<tr>
<td>Packet 16</td>
<td>Bollen and Jackman</td>
</tr>
<tr>
<td></td>
<td>Chatterjee and Wiseman</td>
</tr>
<tr>
<td></td>
<td>Fox (1991): Pages 21-40</td>
</tr>
</tbody>
</table>

(Packet 16 is continued on the next page)
<table>
<thead>
<tr>
<th>Lecture Packet</th>
<th>Suggested Optional Readings (In Alphabetical Order)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Packet 16</td>
<td>Gujarati and Porter: Pages 496-497</td>
</tr>
<tr>
<td>(Continued from previous page)</td>
<td>Hamilton: Ch. 1; Pages 113-117, 125-133, 141 (note #4), 154-158</td>
</tr>
<tr>
<td></td>
<td>Kennedy: Sections 21.1, 21.2; Page 347</td>
</tr>
<tr>
<td></td>
<td>Narang and Nelson (especially pages 364-370)</td>
</tr>
<tr>
<td></td>
<td>Wooldridge: Pages 326-330</td>
</tr>
<tr>
<td>Packet 17</td>
<td>Achen (1982): Appendix</td>
</tr>
<tr>
<td></td>
<td>Fox (1991): Pages 80-82, 83-85</td>
</tr>
<tr>
<td></td>
<td>Gujarati and Porter: Appendices C.1-C.10, C.12, CA.4</td>
</tr>
<tr>
<td></td>
<td>Hamilton: Pages 338-344</td>
</tr>
<tr>
<td></td>
<td>Kennedy: Sections 4.1, 4.2</td>
</tr>
<tr>
<td></td>
<td>Wooldridge: Appendices E.1-E.3</td>
</tr>
<tr>
<td>Packet 18</td>
<td>Berry: Pages 67, 72-78</td>
</tr>
<tr>
<td></td>
<td>Berry and Feldman: Pages 73-88</td>
</tr>
<tr>
<td></td>
<td>Gujarati and Porter: Chapter 11; Appendices C.11, CA.3</td>
</tr>
<tr>
<td></td>
<td>Hamilton: Pages 183-189</td>
</tr>
<tr>
<td></td>
<td>Hardy: Pages 53-61</td>
</tr>
<tr>
<td></td>
<td>Kennedy: Sections 8.1-8.3</td>
</tr>
<tr>
<td></td>
<td>Schroeder, Sjoquist, and Stephan: Pages 75-77</td>
</tr>
<tr>
<td></td>
<td>Wooldridge: Sections 8.1-8.4</td>
</tr>
<tr>
<td>Packet 19</td>
<td>Berry: Pages 67-72, 78-81</td>
</tr>
<tr>
<td></td>
<td>Gujarati and Porter: Sections 12.1-12.9, 12A.1, 12A.2</td>
</tr>
<tr>
<td></td>
<td>Hamilton: Pages 118-124</td>
</tr>
<tr>
<td></td>
<td>Hardy: Pages 82-84</td>
</tr>
<tr>
<td></td>
<td>Kennedy: Section 8.4</td>
</tr>
<tr>
<td></td>
<td>Schroeder, Sjoquist, and Stephan: Pages 72-75</td>
</tr>
<tr>
<td></td>
<td>Wooldridge: Sections 10.1-10.3, 11.2, 11.5, 12.1-12.3, 12.6</td>
</tr>
<tr>
<td>Packet 20</td>
<td>Asher: Pages 255-256, 260-261</td>
</tr>
<tr>
<td></td>
<td>Berry and Feldman: Pages 88-89</td>
</tr>
<tr>
<td></td>
<td>Fox (1991): Pages 75-80</td>
</tr>
<tr>
<td></td>
<td>Goertzel (2002)</td>
</tr>
<tr>
<td></td>
<td>Granato</td>
</tr>
<tr>
<td></td>
<td>Gujarati and Porter: Chapter 18; Sections 13.13, 15.9, 15.10</td>
</tr>
<tr>
<td></td>
<td>Hamilton: Pages 246 (note #2), 249-252, 281-282</td>
</tr>
<tr>
<td></td>
<td>Kennedy: Sections 22.1-22.4</td>
</tr>
<tr>
<td></td>
<td>Lewis-Beck and Lewis-Beck: Pages 93-95</td>
</tr>
<tr>
<td></td>
<td>Menard: Chapter 5</td>
</tr>
<tr>
<td></td>
<td>Pampel: Chapters 4, 5</td>
</tr>
<tr>
<td></td>
<td>Schroeder, Sjoquist, and Stephan: Pages 77-79, 80</td>
</tr>
<tr>
<td></td>
<td>Wooldridge: Chapter 19; Sections 16.1-16.4, 17.1</td>
</tr>
</tbody>
</table>
Bibliography


Hamilton, L. *Regression with Graphics: A Second Course in Applied Statistics.* Duxbury, 1992. (Note: This is the same as the edition published by Brooks Cole in 1999.)


