ME104 Linear Regression Analysis

1   Objectives and Learning Outcomes

This course focuses on building a greater understanding, theoretical underpinning, and tools for applying the linear regression model and its generalizations. With a practical focus, it explores the workings of multiple regression and problems that arise in applying it, as well as going deeper into the theory of inference underlying regression and most other statistical methods. The course also covers new classes of models for binary and count data, emphasizing the need to fit appropriate models to the underlying processes generating the data being explained.

This course is primarily about data analysis and developing a deeper understanding of the generalized linear model. The focus is on practice, and this focus is reflected in the choice of texts and in the emphasis on applied coursework. While this course deals to some degree with the generalized linear model on a mathematical and theoretical level, its main focus is practical, the ability to use the techniques when faced with the need in practical research. Consequently the learning method combines lectures and reading with hands-on statistical programming exercises using real datasets.

The learning outcomes associated with this nine-week course are aimed at students being able to:

- Develop a deeper understanding of the linear regression model and its limitations;

- Know how to diagnose and apply corrections to some problems with the generalized linear model found in real data; discussed;

- Use and understand generalizations of the linear model to binary and count data;

- Develop a greater familiarity with a range of techniques and methods through a diverse set of theoretical and applied readings;

- Know where to go to learn more about the techniques in this class and those called for that were not covered in this class.

2   Prerequisites

An introductory course in quantitative methods is a prerequisite for this course, since this course extends rather than introduces linear regression analysis.

The Stata statistical package is used for all exercises, and students in ME104 should be familiar or at least prepared to quickly learn this package.

3 Logistics

Meetings. Classes will meet over ten days for a combination of lectures and lab sessions. Lectures will take place from 10:00–12:00 each day, with labs following at xxx. The lab sessions will consist of guided exercises using Stata to implement practical data analysis.

Teaching assistant. The TA in 2012 is Carolina Plescia, `plesciac@tcd.ie`. Carolina will be leading the lab sessions.

Computer software. Stata 12 will be be used for all exercises.

Reading materials consist of:

- Kennedy, Peter. 2008. *A Guide to Econometrics*. 6th ed. Oxford: Blackwell.

  This text provides a three-level discussion of each topic: first a general discussion, then a technical discussion, and then a very technical discussion. Most students find this quite useful since it permits them to dig as deep as their abilities let them or as their need allows.

- Agresti, Alan and Barbara Finlay. 2009. *Statistical Methods for the Social Sciences* (4th Edition). Prentice Hall.

  Written by statisticians rather than econometricians, this text provides a very accessible treatment of the methods covered in this course in a traditional textbook format. For those who consider this too basic, I recommend Wonnacott and Wonnacott (below) as an alternative.

- Wonnacott, Thomas H. and Ronald J. Wonaccott. 1990. *Introductory Statistics*. 5th Ed. New York: Wiley.

  This text is *recommended, but not required*, as a supplement or alternative to Agresti and Finlay if you should prefer W&W or find the A&F text too basic.

The **home page** for this course will be `http://www.kenbenoit.net/ME104/` and you will be able to find this course handout there, links to exercises, and additional material.

4 Schedule

1. **The Error Term and Properties of Estimators**.

   Required Reading:
   Kennedy Ch 1–2, "Introduction" and "Criteria for Estimators"
   A&F Review by reading Chs 4–5, "Probability Distributions" and "Statistical Inference: Estimation
   Recommended Reading:

W&W Ch 7, "Point Estimation"
Homework:
TBA

2. **The Classical Linear Regression Model**.

   Required Reading:
   Kennedy Ch 3, "The Classical Linear Regression Model"
   A&F Ch 19, "Multiple Regression and Correlation"
   Recommended Reading:
   W&W Chs 11–13, "Fitting a Line", "Simple Regression", and "Multiple Regression"

   William Greene. 2003. *Econometric Analysis*. 5th ed. Upper Saddle River, NJ: Prentice Hall. Ch 2, "The Classical Multiple Linear Regression Model"
   Homework:
   TBA

3. **Inference, Intervals, and Hypothesis Testing**.

   Required Reading:
   Kennedy Ch 4, "Interval Estimation and Hypothesis Testing"
   A&F Ch 6, "Statistical Inference: Significance Tests"
   Recommended Reading:
   W&W Ch 8, "Confidence Intervals" and Ch. 9 "Hypothesis Testing"
   Homework:
   TBA

4. **Diagnosing problems with the CLRM.**

   Required Reading:
   Kennedy Chs 5–6, "Specification" and "Wrong Regressors, Nonlinearities, and Parameter Inconsistency"
   A&F Ch 14, "Model Building with Multiple Regression"
   Recommended Reading:
   Berry, William D. and Stanley Feldman. 1993. "Multiple Regression in Practice." In *Regression Analysis*, ed. Michael Lewis-Beck. London: Sage.
   Fox, John. 1993. "Regression Diagnostics" In *Regression Analysis*, ed. Michael Lewis-Beck. London: Sage.
   Homework:
   TBA

5. **Problems with predictors.**

   Required Reading:
   Kennedy Ch 12, "Multicollinearity"; Kennedy Ch 11.1, "Errors in Variables"
   Recommended Reading:
   Review A&F Ch 14
   Homework:
   TBA

6. **Problems with error assumptions.**

   Required Reading:
   Kennedy Ch 7, "Non-zero expected disturbance"; Ch 8, "Non-spherical disturbances"; Ch 11, "Simultaneous Equations"
   A&F Ch 16.6, "Structural Equation Models" **Recommended Reading:**
   Hanushek, Eric A. and John E. Jackson. 1977. *Statistical Methods for Social Scientists.* San Diego: Academic Press. Ch 8, "Structural Equations."
   Kenneth Benoit and Michael Marsh. 2008. "The Campaign Value of Incumbency: A New Solution to the Puzzle of Less Effective Incumbent Spending." *American Journal of Political Science* 52(4, October): 874-890.
   Homework:
   TBA

7. **Models for binary data: Binary Logistic Regression.**

   Required Reading:
   Kennedy Ch 16, "Qualitative Dependent Variables"
   A&F Ch 15.1–15.3, "Logisitic Reression" **Recommended Reading:**
   Aldrich, John H. and Forrest D. Nelson. 1984. *Linear Probability, Logit, and Probit Models.* Beverly Hills: Sage.
   W&W Ch 18 "Maximum Likelihood Estimation"
   Homework:
   TBA

8. **Models for categorial and ordinal data.**

   Required Reading:
   Kennedy CH 16.2–16.3,
   A&F Chs 15.4–15.6
   Recommended Reading:
   Homework:
   TBA

9. **Models for count data**

   Required Reading:
   Kennedy Ch 16.4, "Count Data"
   Recommended Reading:
   King, Gary. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference.* Cambridge, England and New York: Cambridge University Press, 1989. Reprinted, Ann Arbor: University of Michigan Press, 1998. Chapter 5.7–5.10.
   Homework:
   TBA

10. **Estimating uncertainty in inferential models**.

    Required Reading:
    King, Gary; Michael Tomz; and Jason Wittenberg. 2000. "Making the Most of Statistical Analyses: Improving Interpretation and Presentation." *American Journal of Political Science* 44(2, April): 341-355.

Recommended Reading:
The Zelig package for R — this will allow you to implement the "CLARIFY" techniques in King, Tomz and Wittenberg (2000)
Kennedy Ch 21, "Robust Estimation"
King, Gary. 1986. "How Not to Lie with Statistics: Avoinding Common Mistakes in Quantitative Political Science." *AJPS* 30(3):666-687.
Homework:
TBA