Syllabus and Outline
Applied Regression Analysis 506
Spring 2009
Monday 9-12

Instructor: Matthew Archibald
Office: Tarbutton 213
Hours: Monday 2-4
E-mail: marchib@emory.edu

Course Description:

Applied Regression Analysis 506 is a graduate-level course in the theory and application of regression models as they pertain to social science data. Topics include simple linear and multiple regression, dummy variable techniques, model reduction and diagnostics, violation of classic assumptions, nonlinear and dynamic models, and simultaneous equations. The course will provide the necessary foundation in statistical theory and estimation procedures underlying multivariate regression analysis and give students data analysis tools, which will allow them to undertake their own research using these techniques. In addition, a thorough grounding in this material will facilitate students' understanding and interpretation of regression-based social science research literatures.

Grading:

There will be three exams (55%) and six exercises (45%). The exam questions will be given-due 2/23-3/2, 3/30-4/6 and, 4/27-5/4. Exercises will be due 2/2, 2/19, 2/23, 3/16, 3/30 and, 4/13.

Exercises:

Students will respond to questions provided by the instructor. Questions will be drawn from the daily readings as well as from material located in the library, in library databases and on the Internet.

Texts:

- Gujarati, Damodar N. 2003. *Basic Econometrics*. Fourth edition. McGraw Hill: Boston.
- Allison, Paul D. 1999. *Multiple Regression: A Primer*. Pine Forge: Thousand Oaks, CA.

Course Outline and Calendar

Week 1 – Holiday
1/19

Week 2 – Linear regression models – bivariate regression
1/26 (1) Introduction and basic ideas; Gujarati Pp. 1-36
(2) Linearity, population, sample; Gujarati Pp. 37-52

Goals I. Introduction
II. Evaluation exercise
III. Review scientific method (from Gujarati).
IV. Review of some basic statistical ideas: variables, regression, correlation
V. Terminology and notation
VI. Exercise 1 – due Monday 2/2:
1. Use Gujarati data - Table 1.1 U.S. Egg Production
2. Calculate mean, median, mode variance and standard deviation
3. Write a statement interpreting over-time changes in these statistics
4. Calculate correlation coefficients for all Ys and Xs
5. Write a statement interpreting the coefficients

Week 3 – Linear regression models – bivariate regression
2/2 (1) Estimation; Gujarati Pp. 58-105

(2) More assumptions; Gujarati Pp. 107-118

Goals I. OLS
II. Assumptions of the least-squares models
III. Standard error, goodness of fit
IV. Normality
V. Maximum likelihood
VI. Exercise 2 – due Monday 2/9:
1. Use the Gujarati data - Table 3.6 Productivity and Related Data.
2. Focus on the business sector, and calculate the various components of the regression equation including r, R^2, b_1, and the intercept. Provide a substantial interpretation of these components.
3. Do the same for the non-farm business sector
4. Does the model fit the data better in one sector than the other? How do you know?

Week 4 – Linear regression models – multiple regression

2/9 (1) Franzosi, Roberto (1994) "Outside and Inside the Regression Black Box: A New Approach to Data Analysis." Quality and Quantity, No. 28, pp. 21–53.
(2) Hypothesis testing; Gujarati Pp. 119-140
(3) Estimation and interpretation; Gujarati Pp. 202-225

Goals I. Confidence intervals and significance tests
II. P-values
III. Evaluating results (from pages 146-147)
IV. Multiple regression

Week 5 – Linear regression models – multiple regression

2/16 (1) Inference; Gujarati Pp. 248-296;
(2) Dummy variables; Gujarati Pp. 297-334

Goals I. Multiple regression and drawing inferences
II. Advanced model testing
III. Dummy variable regression
IV. Advanced: Dummy variable interactions
V. Exercise 3 – due Monday 2/23
1. Use Gujarati's example 9.5 on page 311
2. Interpret the coefficients in the two models
3. What variables are significant and what does that imply?
4. How do the results overall match theoretical expectations?

Week 6 –Exam

2/23 (1) Recapitulation – bivariate and multivariate regression
Exam 1 (due March 2)

Week 7 – Violation of underlying assumptions

3/2 (1) Multicollinearity; Gujarati Pp. 335-386
(2) Error variance; Gujarati Pp. 387 – 440

Goals I. What is Multicollinearity?
II. What are its sources?
III. How to solve it
IV. What is error variance – heteroscedasticity
V. What are its sources?
VI. How to solve it

VII. Exercise 4 – due Monday 3/16
1. Use Gujarati's discussion of White on page 417 example 11.8
2. What are robust standard errors?
3. How do they correct for OLS standard errors?
4. What does their use imply with regard to hypothesis testing?
5. What kind error (type I or type II) is involved and why?

Week 8
Spring Recess 3/9-13

Week 9 – Assumptions and model specification errors
3/16 (1) Autocorrelation; Gujarati Pp. 441-505
(2) Specification errors; Gujarati Pp. 506-562
Goals I. What is autocorrelation?
II. What are its sources?
III. How to solve it
IV. Specification errors
V. Model selection

Week 10 – Nonlinear regression models
3/23 (1) Functional form; Gujarati Pp. 226-229, 175-201, 280-281
(2) Interactions; handout
Goals I. Power polynomials – linear in the variables
II. Other transformations - linear in the parameters
III. Interactions
IV. Exercise 5 – due Monday 3/30
1. Use Gujarati's Table 10.14
2. Test the hypothesis that assets are a multiplicative function of number of dependents and school attainment
3. There's multicollinearity present, so make sure that you center number of dependents before creating the interaction term
4. Interpret the outcomes
5. Discuss the multicollinearity issue

Week 11 – Exam
3/30 (1) Recapitulation, exercise 5 due
Exam 2 (due April 6)

Week 12 – Nonlinear regression and longitudinal models
4/6 (1) Categorical outcomes; Gujarati Pp. 563-598, 604-609, 616
(2) Panels; Gujarati Pp. 636-655
Goals I. Qualitative response outcomes
II. What is panel data?
III. What is its use and what problems does its use raise?
IV. How do we solve these problems?
II. Exercise 6 - due Monday 4/13
1. Based on your reading of the Soule and Olzak (2004), summarize the results in Table 2.
2. Pick several key relationships and discuss the regression coefficients
3. Any anomalous results?

Week 13– Structural equation models

4/13 (1) Path analysis; handout - Pedhazur 577-594
(2) Identification problems; Pedhazur 594-623

Goals I. Path Analysis
II. Assumptions of the models
III. Solving for paths with correlations
III. Notation
IV. Introduction - Structural equation models
V. Identification problems

Week 14 – Structural equation models

4/20 (1) Holye 1-15; MacCallum 16-36; Pedhazur 636-644
(2) Models; Pedhazur 670-677, RSO legitimacy example

Goals I. Measurement models – confirmatory factor analysis
II. Structural models
III. Summary
IV. Structural Models
V. M-plus code
VI. Summary

Week 15

4/27 (1) Recapitulation
Exam 3 (due May 4)