Conducting Multivariate Analyses of Social, Economic, and Political Data

Dr. Harold D. Clarke   Dr. Guy D. Whitten
University of Texas, Dallas   Texas A& M
hclarke@utdallas.edu   g-whitten@pols.tamu.edu

Overview: This workshop teaches students how to conduct multivariate analyses of social, economic and political data. Several valuable tools for academic researchers in the social sciences and analysts in the public and private sectors are considered. The course first focuses on discrete choice models, including the mixed logit model, for analyzing factors affecting choices individuals make in social, economic and political environments. The course then considers time series methods for studying policy interventions and the dynamics of policy outcomes. Topics covered include ARIMA models for interventions and forecasting, error correction models of cointegrated processes, dynamic panel models and ARCH models of volatility and risk. Assuming only familiarity with the standard OLS regression model, the course emphasizes practical applications with real data and students will learn how to use major statistical software packages such as Stata, R and Rats in their research. Students are invited to bring their own data sets to lab sessions.

Class Schedule: We will spend as much time as necessary on each topic for this course. Additional topics (such as VAR/VECM models) may be added based on available time and student interests.

Topic 1   Course Introduction and Essential Background Materials

- Course Introduction

- The Place of Methodology in a Research Program

- Some Rules of the Road For Conducting Statistical Analyses

- Essential concepts in OLS

- The joys of simulation for both learning and inference

- Readings:

    - Nagler, Jonathan (1995) "Coding Style and Good Computing Practices." *The Political Methodologist*, 6:2-8.

Topic 2   Regression Essentials

- Review of regression assumptions

- F-tests and model selection criteria

- Omitted variable bias

- Multicollinearity

- Heteroscedasticity

- "Curing" heteroscedasticity with FGLS

- Interactive model specifications

- Readings:

  - Kellstedt, P., & Whitten, G. (2013). *The fundamentals of political science research.* Cambridge University Press. As needed and available.
  - Brambor, Thomas, William Roberts Clark, and Matt Golder. "Understanding interaction models: Improving empirical analyses." Political analysis 14.1 (2006): 63-82.

Topic 3   Dichotomous Response Models

- The Linear Probability Model

- Binomial Logit and Probit

- Estimation of Binary Choice Models

- Heteroscedastic Probit

- Scobit

- Goodness of Fit for Discrete Choice Models

- Readings:

  - Kellstedt and Whitten Chapter 10.
  - Nagler, J. (1994). Scobit: an alternative estimator to logit and probit. *American Journal of Political Science*, 230-255.

Topic 4   Polychotomous Response Models

- Multinomial Logit

- Conditional Logit

- IID and IIA

- Multinomial Probit

- Mixed Logit

- Additional Resources

- Readings:

  - Whitten, G. D., & Palmer, H. D. (1996). Heightening comparativists' concern for model choice: Voting behavior in Great Britain and the Netherlands. *American Journal of Political Science*, 231-260.
  - Dow, J. K., & Endersby, J. W. (2004). Multinomial probit and multinomial logit: a comparison of choice models for voting research. Electoral studies, 23(1), 107-122.
  - Glasgow, G. (2001). Mixed logit models for multiparty elections. Political Analysis, 9(2), 116-136.
  - Chapter 5 of Clarke, Harold D., D. Sanders, M.C. Stewart and P. Whiteley. (2009). *Performance Politics and the British Voter.* Cambridge: Cambridge University Press.

Topic 5 Fundamental Concepts of Time Series Analysis

- Introduction to time series data and notation

- Introduction to using Stata to analyze time series data

- Non-stationarity and unit roots

- Overview of the threats to inference in time series analysis

- Readings:

  - Kellstedt and Whitten Chapter 11.

Topic 6 Univariate ARIMA Models

- The analysis of non-stationary data–a brief look ahead

- The Box-Jenkins approach to time series

- AR and MA processes and how we might know them if we saw them

- PDQ notation and diagnostics

- Model estimation and evaluation

- Univariate ARIMA forecasting

- Readings (at least one of the following):

  - Chapter 13 of Asteriou, D., & Hall, S. (2011). *Applied Econometrics.* Palgrave Macmillan

Topic 7 Intervention and Transfer Functions

- Thinking theoretically about interventions

- Modeling interventions

- Transfer functions

- Recommended Readings:

  - Clarke, H. D., Mishler, W., & Whiteley, P. (1990). "Recapturing the Falklands: Models of conservative popularity, 197983." *British Journal of Political Science*, 20(01), 63-81.

Topic 8 Time Series Regression Models

- Traditional treatments of autocorrelation in regression models

- To lag or not to lag

- Interpreting time series regression models

- Cointegration and error correction models

- Recommended Readings:

  - Beck, N. (1991). "Comparing dynamic specifications: The case of presidential approval." *Political Analysis*, 3(1), 51-87.

  - Keele, L., & Kelly, N. J. (2006). "Dynamic models for dynamic theories: The ins and outs of lagged dependent variables." *Political Analysis*, 14(2), 186-205.

  - De Boef, S., & Keele, L. (2008). "Taking time seriously." *American Journal of Political Science*, 52(1), 184-200.

Topic 9 ARCH and GARCH Models

- Thinking theoretically about heteroscedasticity

- ARCH models

- GARCH models

- Recommended Readings:

  - Chapter 14 of Asteriou, D., & Hall, S. (2011). *Applied Econometrics*. Palgrave Macmillan

  - Bernhard, William and David Leblang. 2006. "Polls and Pounds: Political Expectations and Exchange Rate Volatility in Britain." *Quarterly Journal of Political Science.*

Topic 10   The Basics of Pooling Across Time and Space

- Why pool?

- Examining variation across time and space

- Dummy variable approaches

- Recommended Readings:

  - Stimson, James. 1985. "Regression in Time and Space: A Statistical Essay." *American Journal of Political Science* 29:914-947.
  - Kittel, Bernhard and Hannes Winter. 2005. "How reliable is pooled analysis in political economy? The globalization-welfare state nexus revisited." EJPR.

Topic 11   Models of Pooled Time Series Data

- Panel unit roots

- The ins and outs, and ups and downs of panel corrected standard errors

- Other strategies for dealing with the challenges of pooled time series models

- Recommended Readings:

  - Beck, Nathaniel and Jonathan Katz. 1995. "What To Do (and Not To Do) with Time Series Cross-Section Data." *American Political Science Review* 89:634-47.
  - Williams, Laron K. and Guy D. Whitten. 2012. "But Wait, Theres More! Maximizing Substantive Inferences from TSCS Models." *Journal of Politics.*