

Essex Summer School in Social Science Data Analysis

Panel Data Analysis for Comparative Research

Christopher Adolph

Associate Professor, Department of Political Science
and Center for Statistics and the Social Sciences
University of Washington, Seattle, USA

Summer 2018

23 July–3 August (two week course / 35 hrs)

Goals. This course provides a survey of regression models for time series (TS; also called longitudinal) and time series cross-section (TSCS; also called panel) data, with an emphasis on modeling dynamics and panel structures. Because panel data are ubiquitous in many political science subfields, including political economy, comparative politics, and international relations, training in TSCS analysis is essential preparation for performing and understanding quantitative research. Participants will gain an introductory understanding of the theory behind TSCS models and a working understanding of how to estimate, select, and interpret these models.

Overview. After a review of the theory and estimation of linear regression, we will cover the following topics: diagnosing dynamic behavior in time series; modeling time series dynamics using ARIMA models, lagged dependent variables, and distributed lags; cointegration and error correction models; modeling cross-sectional variation using fixed and random effects; appropriate techniques for fixed effects models used on data with many time periods available versus those used with few time periods; coping with panel heteroskedasticity; and presentation and interpretation of TS and TSCS models, especially using dynamic simulation techniques. Time permitting, we will cover advanced topics based on student interest, which in past years have included multiple imputation for missing data in panel datasets, models for binary TSCS data, and linkages between panel data and more general approaches to hierarchical linear models.

Prerequisites. Students should enter the course with a solid understanding of first year statistics as taught in a standard political science doctoral program, an interest in data with either a time series or time series cross-sectional (panel) data structure, and either exposure to, or willingness to learn, the R statistical package.

R. Extensive in-class code examples will use the R statistical package, which is powerful, free, open source, widely used, and rapidly becoming the standard for quantitative work in political science and other fields. You can obtain R at <http://www.r-project.org/>. Throughout the course, I will provide example code in R, and can only promise detailed homework help for R, not other statistics packages.

Course Readings

Selections from the following books and articles will be provided as part of the course materials. Students seeking to get a headstart should focus on the assigned readings for the first week; those without a background in R should especially concentrate on Zuur.

Nathaniel Beck and Jonathan N. Katz. 1995. “What to Do (And Not to Do) With Time Series Cross-Section Data.” *American Political Science Review* 89(3): 634–647.

Nathaniel Beck and Jonathan N. Katz. 2011. “Modeling dynamics in Time-Series–Cross-Section political economy data.” *Annual Review of Political Science* 14:331–52.

Nathaniel Beck, Jonathan N. Katz, and Richard Tucker. 1998. “Taking time seriously: Time-Series–Cross-Section analysis with a binary dependent variable.” *American Journal of Political Science* 42(4) 1260–1288.

Janet M. Box-Steffensmeier, John R. Freeman, Matthew P. Hitt, and Jon C.W. Pevehouse. 2014. *Time Series Analysis for the Social Science*. Cambridge University Press.

Paul S.P. Cowpertwait and Andrew V. Metcalfe. 2009. *Introductory Time Series with R*. Springer-Verlag.

Andrew Gelman and Jennifer Hill. 2007. *Data analysis using regression and multilevel/hierarchical models*. University of Cambridge Press.

James Honaker and Gary King. 2010. “What to do about missing values in Time-Series Cross-Section data.” *American Journal of Political Science* 54(2): 561–581.

Gary King, Michael Tomz, and Jason Wittenberg. 2000. “Making the Most of Statistical Analyses.” *American Journal of Political Science* 44(2): 347–361.

Giovanni Millo. 2014. “Robust standard error estimators for panel models: a unifying approach.” MPRA Paper No. 54954.

Bernhard Pfaff. 2008. *Analysis of Integrated Series with R*. Springer-Verlag.

David Roodman. 2009. “How to do xtabond2: An introduction to difference and system GMM in Stata.” *The Stata Journal*. 9(1): 86–136.

Jeffrey M. Wooldridge. 2010. *Econometric Analysis of Cross-Sectional and Panel Data*. MIT Press. 2nd Edition.

Alain F. Zuur, Elena N. Ieno, and Erik H.W.G. Meesters. 2009. *A Beginner's Guide to R*. Springer-Verlag.

Course outline

Readings assigned for a particular day should be read in advance of class. Students seeking feedback on homework should turn in the listed problems each day. It is strongly recommend that students new to R work through the example code presented in Zuur as they read.

Week 1

Day 1 · 23 July · Introduction, Review of the Linear Model and its Properties; R Basics

Readings: On R: Zuur, Ch. 1, 2, and 3
On linear regression: Woolridge, Ch. 4
On matrix algebra (opt.): Matrix Handout (Kevin Quinn)

Day 2 · 24 July · Estimation and Simulation; Linear Regression in R

Readings: On R: Zuur, Ch. 4 and 5
On simulation: King, Tomz, and Wittenberg 2000
Homework: Problem 1

Day 3 · 25 July · Basic Concepts for Time Series: Trends, Lags, and Cycles

Readings: On time series: Box-Steffensmeier *et al*, Ch. 1, 2
Alternate reading: Cowpertwait & Metcalf, Ch. 1.1, 1.4, 1.6, 2.1–2.5
On R: Zuur, Ch. 6
Homework: Problems 2 and 3

Day 4 · 26 July · Modeling Stationary Time Series

Readings: On time series: Box-Steffensmeier *et al*, Ch. 3
Alternate reading: Cowpertwait & Metcalf, Ch. 4, Ch. 5.1–5.4, 5.9–5.11, Ch. 6
Homework: Problem 4 or Bonus Problem A

Day 5 · 27 July · Modeling Nonstationary Time Series

Readings: On time series: Box-Steffensmeier *et al*, Ch. 5, 6
Alternate reading: Cowpertwait & Metcalf, Ch. 7
On cointegration: Pfaff, Ch. 4
Homework: Problems 5 and 6

Week 2

Day 6 · 30 July · Basic Concepts for Panel Data

Readings: On random effects: Woolridge, Ch. 10.1–10.4
On hierarchical models: Gelman and Hill, Ch. 11, 12, 13

Homework: Problem 7

Day 7 · 31 July · Panel Data Models with Many Time Periods

Readings: On fixed effects: Woolridge, Ch. 10.5–10.7
On panels with large- T : Beck & Katz 2011

Homework: Begin work on Problem 8

Day 8 · 1 August · Panel Data Models with Few Time Periods

Readings: On panels with small- T : Roodman 2009 (skip code examples)

Homework: Continue work on Problem 8

Day 9 · 2 August · Heteroskedasticity in Panel Data / In-Sample Simulation for Panel Data Models

Readings: On panel-corrected standard errors: Beck & Katz 1995
On heteroskedasticity and autocorrelation correction: Millo 2014

Homework: Complete Problem 8

Day 10 · 3 August · Course Wrap-up / Advanced Topics

Possible topics: Multiple imputation; Binary outcomes; Course Q&A;
Student research presentations

Readings: On multiple imputation for TSCS: Honaker and King 2010
On binary outcomes in TSCS: Beck, Katz & Tucker 1998

Homework: No additional homework