CSSS/POLS 512:
Time Series and Panel Data for the Social Sciences

Christopher Adolph · Associate Professor · Political Science and CSSS

University of Washington · Spring Quarter 2018

Class Meets
TTh 4:30–5:50 PM
Johnson Hall 175

Section Meets
F 1:00–2:20 PM
Savery Hall 117

Office
Gowen Hall 145
cadolph@uw.edu

Teaching Assistant
Daniel Yoo
dhyoo@uw.edu

Overview and Class Goals. Time series (TS) data – also called longitudinal data – and time series cross-sectional (TSCS) data – also called panel data – are widely used in the social sciences. In some cases, TS and TSCS data are the ideal format for exploring change over time within several units, which might be individuals, organizations, regions, or other entities observed over time. In other cases, panel data are used to either expand the number of observations, to gain leverage over unobservables, or both. Finally, time series and panel data provide opportunities for causal inference often unavailable in strictly cross-sectional datasets. As useful as these data structures are, they typically derive from data generating processes that violate the usual assumptions of linear regression, requiring a variety of specialized techniques for valid inference. Appropriate methods vary subtly by the nature of the time series or panel dataset and the goals of inference and require training to choose correctly and deploy well.

Learning Goals. This course provides a survey of regression models for time series and time series cross-sectional data and associated analytic techniques. In particular, we focus on methods used in political science and allied fields (including sociology, public health, business, education, and public policy) to study continuous outcomes at the level of organizations or political units, and sometimes individuals (especially in settings where individual growth or development can be neglected). Students will learn how to explore their data to choose appropriate models, and how to understand those models once estimated. In particular, our emphasis lies in modeling dynamics and panel structures with *continuous* outcomes, as well as on interpretation and fitting of models. (Other courses cover the modeling of panel data with discrete outcomes.) Specific topics vary and may include trends and seasonality, ARIMA models, lagged dependent variables, distributed lags, cointegration and error correction models, fixed and random effects, panel heteroskedasticity, missing data imputation, and causal inference using panel data.

Prerequisites. No specific courses are required; however, students should have a solid grounding in linear regression, as provided, for example, by courses in Political Science (POLS 501 and POLS 503), Sociology (SOC 504 and SOC 506), or Statistics (STAT/CSSS 504). Familiarity with (or a willingness to quickly learn) matrix notation for regression models is essential, as is basic proficiency in the R statistical language. Students may use alternative packages when they are able to comprehensively achieve class goals using those packages, but in-class support will be provided only for R.

Course Requirements. Course evaluation will be based on problem sets (50% of course grade based on approximately three or four over the term, due as printed copies in class), a student poster presentation in class (10% of course grade), and a research paper (40% of course grade; due Tuesday, 5 June 2018 at noon *both* as a printed copy in my mailbox in Gowen Hall *and* as an electronic copy sent to cadolph@uw.edu). Students are *strongly* encouraged to collaborate in groups of two or three on the paper, which should apply methods studied in the course (or with instructor approval, related methods of similar sophistication) to a student-chosen time series or panel dataset. Pure methodology papers in time series and/or panel data analysis are also acceptable. Further rules and suggestions for the research paper are provided on the course website.

Office Hours. Christopher Adolph: Thursdays, 3:00 – 4:20 PM and by appointment in Gowen 145. Daniel Yoo: Tuesdays, 3:00 – 4:20 PM in Smith 220.

Course Website. Consult `http://faculty.washington.edu/cadolph/panUW` for problem sets, notes, and announcements.

Course textbooks

Required (sharing encouraged; some selections provided)

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. ISBN-10: 0521691559. ISBN-13: 978-0521691550. Amazon: $35.99.	*Main course text on time series; highly readable and comprehensive coverage of the most relevant methods for social science applications, with a political science focus.*
Paul S.P. Cowpertwait & Andrew V. Metcalfe. 2009. *Introductory Time Series with R*. Springer-Verlag. ISBN-10: 0387886974. ISBN-13: 978-0387886978. Amazon: $32.37.	*Secondary text on time series; straightforward general introduction focused on implementation of common methods in R.*
Andrew Gelman and Jennifer Hill. 2007. *Data analysis using regression and multilevel/hierarchical models*. University of Cambridge Press. ISBN-10: 052168689X. ISBN-13: 978-0521686891. Amazon: $47.55.	*Excellent general introduction to linear regression and hierarchical modeling, of which panel data models are a specialized subset of approaches; good for code and intuition, especially on simulation methods.*
Jeffrey M. Wooldridge. 2010. *Econometric Analysis of Cross-Sectional and Panel Data*. MIT Press. 2nd Edition. ISBN-10: 0262232588. ISBN-13: 978-0262232586. Amazon: $84.84.	*Classic modern text covering linear regression and panel data models from an econometric perspective.*

Optional books for further study

Badi L. Baltagi. 2013. *Econometric Analysis of Panel Data*. Fifth Edition. Wiley.

Stephen L. Morgan and Christopher Winship. 2014. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Second Edition. Cambridge University Press.

Gary King. 1989. *Unifying Political Methodology*. University of Michigan Press.

Norman Matloff. 2011. *The Art of R Programming: A Tour of Statistical Software Design*. No Starch Press.

Will H. Moore and David A. Siegel. 2013. *A Mathematics Course for Political & Social Research*. Princeton University Press.

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

Alain F. Zuur, Elena N. Ieno, and Erik H.W.G. Meesters. 2009. *A beginner's guide to R*. Springer.

Required and optional articles

Nathaniel Beck and Jonathan Katz. 1995. "What to Do (And Not to Do) With Time Series Cross-Section Data." *American Political Science Review*.

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.

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: Interpretation and Presentation" *American Journal of Political Science* 44(2): 341–355.

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

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

Course outline

This outline of topics is a guideline and may be altered to meet course needs. In particular, the pace of the course may vary to make sure we are moving as fast as possible conditional on everyone understanding the material. Students should come to class having read the material for the next topic to be covered. Optional material is marked "(opt.)".

Part I: Review of Fundamentals

Week 1 · *27–29 March* · Course Introduction / Review of Linear Regression & Simulation

Resources: Review POLS/CSSS 510 lectures on MLE, simulation
Woolridge, Ch. 4, review on linear regression
King, Tomz, and Wittenberg, 2000
Math review (opt.): `www.csss.washington.edu/MathCamp/Review/`
Zuur Ch 1–6 (opt.; for R beginners)
Handout on matrix algebra (opt.)

Part II: Analysis of Time Series Data

Week 2 · *3–5 April* · Basic Concepts for Time Series: Trends, Lags, and Cycles

Readings: Box-Steffensmeier *et al*, Ch. 1, 2.
Cowpertwait & Metcalf, Ch. 1.1, 1.4, 1.6, 2.1–2.5.

Week 3 · *10–12 April* · Modeling Stationary Time Series

Readings: Box-Steffensmeier *et al*, Ch. 3.
Cowpertwait & Metcalf, Ch. 4, 5.1–5.4, 5.9–5.11, 6.

Optional: Box-Steffensmeier *et al*, Ch. 4.

PROBLEM SET 1 DUE THURSDAY 12 APRIL IN CLASS

Week 4 · *17–19 April* · Modeling Nonstationary Time Series

Readings: Box-Steffensmeier *et al*, Ch. 5, 6.
Cowpertwait & Metcalf, Ch. 7.

Optional: Pfaff, Ch. 4.

Part III: Analysis of Panel Data

Week 5 · *24–26 April* · Basic Concepts for Panel Data

Readings: Gelman and Hill, Ch. 11, 12, 13.
Woolridge, Ch. 10.1–10.4.

PROBLEM SET 2 DUE THURSDAY 26 APRIL IN CLASS

Week 6 · *1–3 May* · Panel Data Models with Many Time Periods

Readings: Beck & Katz 2011.
Woolridge, Ch. 10.5–10.7.

Week 7 · *8–10 May* · Panel Data Models with Few Time Periods

Readings: Roodman 2009 (skip code examples)

PROBLEM SET 3 DUE THURSDAY 10 MAY IN CLASS

Week 8 · *15–17 May* · Panel Heteroskedasticity / In-Sample Simulation for Panel Data Models

Readings: Beck & Katz 1995
Millo 2014

Part IV: Advanced Topics

Week 9 · *22 May* · Special Topics – TBD

We will consider one or more advanced topics, to be chosen with class input. Possible topics include missing data imputation, and linkages between this course and event history analysis. Additional readings may be provided.

Part V: Student Poster Presentations

Weeks 9–10 · *24–31 May* · Student Poster Presentations

Students will prepare and present a poster on their research projects in progress; this yields valuable feedback prior to final write-ups. Requirements and suggestions for poster construction will be discussed in class. Presentation dates will be assigned to minimize discrepancy with student preferences. Early presentations are assumed to be less complete and evaluated accordingly. Students unable to present on certain days due to schedule conflicts should inform the instructor.

IF NEEDED, PROBLEM SET 4 DUE THURSDAY 24 MAY IN CLASS

FINAL PAPER DUE TUES 5 JUNE AT NOON IN MY GOWEN MAILBOX AND BY EMAIL