

Panel Data and Longitudinal Analysis

ICPSR Summer Program in Quantitative Methods of Social Research

July 20-August 14, 2020

Time:	10:00AM-12:00PM (Eastern Daylight Time, EDT)
Location:	Online instruction
Instructor:	Dr. Andrew Q. Philips
Email:	andrew.philips@colorado.edu
Instructor Office Hours:	12:00-1:00PM EDT (M/W), 5:00-6:00PM EDT (T/TH)
Teaching Assistant:	Christina Kamis
TA Email:	christina.kamis@duke.edu
TA Office Hours:	4:00-5:00PM EDT (M/W), 12:00-1:00PM EDT (T/TH/F)
Teaching Assistant:	Anna Russian
TA Email:	aerussia@iu.edu
TA Office Hours:	2:00-3:00PM EDT (M-F)

COURSE DESCRIPTION: Data collected over both units (e.g., survey respondents, states, countries) and time (e.g., days, months, years)—variants of which are known as time series cross-sectional, longitudinal, or panel data—are common in the social sciences. By gaining leverage across units and over time, these data help us answer important questions that would be difficult if we only looked at a single point in time (e.g., cross section) or single unit (e.g., time series): the relationship between growth and democracy, whether or not the resource curse exists, or how economic perceptions shape support for the government. Despite these advantages, panel data often show forms of heterogeneity as well as temporal and spatial dependence that make standard regression approaches inappropriate.

This course is designed to provide you with a broad understanding of the field of panel data analysis. The first week of the course will be spent familiarizing ourselves with the structure and properties of panel data. We will cover early approaches to modeling out characteristics such as unit heterogeneity and spatio-temporal dependence. In the second week, we move to various approaches to addressing heterogeneity, such as random and fixed effects. We also cover testing for and modeling dynamics. In the third week, we discuss models designed to account for heterogeneity in the effects, especially in regards to dynamic data. In the last week, we will cover approaches for small- T , large- N datasets (e.g., longitudinal surveys). Throughout, we will also discuss several smaller topics in panel data, such as pseudo-panels, missing data, and models for dichotomous dependent variables. We will use both Stata and/or R for many of these topics.

By the end of this course you should be able to:

- Understand a variety of threats to inference when working with panel data
- Understand the prominent approaches to modeling panel, longitudinal, and TSCS data
- Apply what you have learned to your own research

PREREQUISITES: At least one semester long graduate-level econometrics course (e.g., a matrix or scalar regression course). We will use both R/RStudio and Stata in this course. If you

are not familiar with both, you should review Philips' "Introduction to Stata" and "Introduction to R" in the course readings folder.

We will also release course materials on Canvas. This course will be taught virtually, and we will use Zoom, so please have that downloaded before the first day (no need to create a paid account if you don't already have one).

REQUIRED TEXTS: There are no required texts for this course. Course materials will be made available to you on the first day. There are several additional texts you might find helpful (grouped by topic):

- Time Series
 - Pickup, M. 2014. *Introduction to Time Series Analysis*. SAGE Publications. Quantitative Applications in the Social Sciences. 1st Edition.
 - Box-Steffensmeier, J.M., J.R. Freeman, M.P. Hitt, and J.C.W. Pevehouse. 2015. *Time series analysis for the social sciences*. Cambridge University Press.
 - Enders, W. 2010. *Applied Econometric Time Series*. 3rd Edition. John Wiley & Sons.
- Panel Data
 - Asteriou, Dimitrios and Stephen G. Hall. 2016. *Applied Econometrics*, 3rd edition. Palgrave.
 - Hsiao C. 2014. *Analysis of Panel Data*, 3rd Edition. New York, NY: Cambridge University Press.
 - Baltagi, B. 2008. *Econometric analysis of panel data*, 4th Edition. John Wiley & Sons.
 - Soderbom, Mans and Francis Teal (with Markus Eberhardt, Simon Quinn and Andrew Zeitlin). 2015. *Empirical Development Economics*, 1st edition. Routledge.
 - Wooldridge, J.M. 2010. *Econometric analysis of cross section and panel data*. MIT Press.
- Multilevel Modeling/Hierarchical Linear Models
 - Gelman, A., and Jennifer H. 2006. *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.
 - Snijders, T. A. and Bosker, R. J. 2012. *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Sage Publications, 2nd edition.

TENTATIVE SCHEDULE:¹ Below is the tentative schedule. Because some topics may take more than one day, while others less, we will go by "topics" rather than days. We will spend as much time as we need to on a given topic as well.

Readings are roughly listed by priority (i.e., top-most is probably most crucial to understanding the topic; lowest is less important but offers nuance or additional information). Also note that panel data analysis is a huge topic with contributions from all social science fields. There are several topics related to panel data that we simply lack the time to cover in much detail, such as multi-level/hierarchical linear modeling and spatial statistics.

- Topic 1: Regression review and matrix algebra

¹Note: Exact schedule may change. Topics roughly the same.

- Philips, Andrew Q. 2019. “R: A brief introduction.”
- Philips, Andrew Q. 2019. “Stata: A brief introduction.”
- Greene, William H. *Econometric analysis*. 2017. 8th edition. Pearson. *Appendix A*
- Topic 2: Panel data fundamentals, describing and summarizing panel data
 - Philips, Andrew Q. n.d. *Panel Data Analysis for the Social Sciences*. Chapter 2. “Why Pool?”
 - Beck, N. 2001. “Time-series-cross-section data: What have we learned in the past few years?” *Annual Review of Political Science* 4(1):271-293.
- Topic 3: Identifying heterogeneity and spatio-temporal dependence
 - Phillips, P. C. and Sul, D. 2003. “Dynamic panel estimation and homogeneity testing under cross section dependence.” *The Econometrics Journal* 6(1):217-259.
 - Pesaran, M. H. 2004. “General diagnostic tests for cross section dependence in panels.”
- Topic 4: FGLS procedures for standard error corrections
 - 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.
 - King, G. and M.E. Roberts. 2015. “How robust standard errors expose methodological problems they do not fix, and what to do about it.” *Political Analysis* 23: 159-179.
 - Parks, R. W. 1967. “Efficient estimation of a system of regression equations when disturbances are both serially and contemporaneously correlated.” *Journal of the American Statistical Association* 62(318):500-509.
- Topic 5: Approaches to heterogeneity: Fixed and random effects
 - Stimson, James A. 1985. “Regression in space and time: A statistical essay.” *American Journal of Political Science* 914-947.
 - Zhu, L. 2012. “Panel Data Analysis in Public Administration: Substantive and Statistical Considerations.” *Journal of Public Administration Research and Theory* 23:395-428.
 - Arceneaux, K. and Nickerson, D.W. 2009. “Modeling certainty with clustered data: A comparison of methods.” *Political Analysis* 17(2):177-190.
- Topic 6: FE/RE continued: Should I use fixed or random effects?
 - Kittel, B., and H. Winner. 2005. “How reliable is pooled analysis in political economy? The globalization-welfare state nexus revisited.” *European Journal of Political Research* 44(2):269-293.
 - Clark, T.S., and Linzer, D.A., 2015. “Should I use fixed or random effects?” *Political Science Research and Methods* 3(2):399-408.
 - Mummolo, J. and Peterson, E. 2018. “Improving the interpretation of fixed effects regression results.” *Political Science Research and Methods* 6(4):829-835.
 - Jordan, Soren and Andrew Q. Philips. Working Paper. “Improving the interpretation of random effects regression results.”
- Topic 7: Alternative models for unit heterogeneity
 - Bell, A., and Jones, K., 2015. “Explaining fixed effects: Random effects modeling of time-series cross-sectional and panel data.” *Political Science Research and Methods*, 3(1):133-153.
 - Plumper, Thomas, and Vera E. Troeger. 2007. “Efficient estimation of time-invariant and rarely changing variables in finite sample panel analyses with unit fixed effects.” *Political Analysis* 15:124-139.
 - Plumper, T. and Troeger, V. E. 2011. “Fixed-effects vector decomposition: properties, reliability, and instruments.” *Political Analysis* 19(2):147-164.
 - Mundlak, Y. 1978. “On the pooling of time series and cross section data.” *Econometrica: Journal of the Econometric Society* 69-85.
- Topic 8: Modeling and interpretation under dynamic models
 - Williams, L.K., and G.D. Whitten. 2012. “But wait, there’s more! Maximizing substantive inferences from TSCS models” *The Journal of Politics* 74(3):685-693.

- Beck, Nathaniel and Jonathan N. Katz. 2011. “Modeling dynamics in time-series-cross-section political economy data.” *Annual Review of Political Science*: 331-352.
- Williams, L.K., and G.D. Whitten. 2011. “Dynamic simulations of autoregressive relationships.” *The Stata Journal* 11(4):1-12.
- Topic 9: Panel unit root testing
 - Hlouskova, Jaroslava and Martin Wagner. 2006. “The performance of panel unit root and stationarity tests: Results from a large scale simulation study.” *Econometric Reviews* 25(1):85-116.
 - Maddala, G.S. and Shaowen Wu. 1999. “A comparative study of unit root tests with panel data and a new simple test.” *Oxford Bulletin of Economics and Statistics, Special Issue* 61(S1):631-652.
- Topic 10: Panel cointegration and models for cointegrating panel data
 - Westerlund, Joakim. 2005. “New simple tests for panel cointegration.” *Econometric Review* 24(3):297-316.
 - Neal, Timothy. 2014. “Panel cointegration analysis with `xtpedroni`.” *The Stata Journal* 14(3):684-692.
- Topic 11: Modeling dichotomous-CSTS data
 - Beck, Nathaniel, 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* 42:1260-1288.
 - Carter, David B. and Curtis S. Signorino. 2010. “Back to the future: Modeling time dependence in binary data.” *Political Analysis* 18(3):271-292.
 - Philips, Andrew Q. Forthcoming. “`mkduration`: An easy way to create duration variables in binary cross-sectional time series data” *The Stata Journal*.
- Topic 12: Effect heterogeneity: fixed and random slopes
 - Wilson, S.E. and D.M. Butler. 2007. “A Lot More to Do: The Sensitivity of Time-Series Cross-section Analyses to Simple Alternative Specifications.” *Political Analysis* 15:101-123.
 - Zellner, A. 1962. “An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias.” *Journal of the American Statistical Association* 57:348-368.
- Topic 13: Dynamic heterogeneity: Mean-group and pooled mean-group estimators
 - Chapter 22 Dynamic Heterogeneous Panel Data Models of Asteriou, Dimitriou and Stephen G. Hall, *Applied Econometrics*, 3rd edition. Palgrave, 2016.
 - Cross-Section Dependence and Nonstationary Data of Mans Soderbom and Francis Teal with Markus Eberhardt, Simon Quinn and Andrew Zeitlin, *Empirical Development Economics*, 1st edition. Routledge, 2015. Chapter 27.
 - Pesaran, M.H. and Smith, R., 1995. “Estimating long-run relationships from dynamic heterogeneous panels.” *Journal of econometrics* 68(1):79-113.
 - Pesaran, M.H. ,Shin, Y., and Smith, R.P. 1999. “Pooled mean group estimation of dynamic heterogeneous panels.” *Journal of the American Statistical Association* 94(446):621-634.
- Topic 14: Dynamic heterogeneity cont.: Common correlated effects
 - Ditzen, Jan. “Estimating Dynamic Common Correlated Effects in Stata.” *The Stata Journal*, 18:3 (2018): 585-617.
 - Blackburne, Edward F. III and Mark W. Frank. 2007. “Estimation of nonstationary heterogeneous panels.” *The Stata Journal* 7(2):197-208.
- Topic 15: Endogeneity, Nickell bias, and inconsistency, oh my! Instrumental variable approaches
 - Plumper, Thomas and Vera E. Troeger. 2019. “Not so harmless after all: The fixed-effects model.” *Political Analysis* 27:21-45.
 - Anderson, T. W. and Hsiao, C. 1981. “Estimation of dynamic models with error components.” *Journal of the American statistical Association* 76(375):598-606.

- Topic 16: GMM estimators
 - Wawro, G. 2002. “Estimating Dynamic Panel Data Models in Political Science.” *Political Analysis* 10(1):25-48.
 - Roodman, David. 2009. “How to do xtabond2: An Introduction to Difference and System GMM in Stata.” *The Stata Journal* 9(1):86-136.
- Topic 17: GMM estimators continued
 - Estimation of Dynamic Effects with Panel Data of Mans Soderbom and Francis Teal with Markus Eberhardt, Simon Quinn and Andrew Zeitlin, *Empirical Development Economics*, 1st edition. Routledge, 2015. Chapter 24.
 - Roodman, David. “A Note on the Theme of Too Many Instruments.” *Oxford Bulletin of Economics and Statistics* 71.1 (2009), 135-158.
- Topic 18: Transformed-likelihood, quasi- and full-maximum likelihood estimators
 - Pickup, M. and V. Hopkins. 2020. “Transformed-Likelihood Estimators for Dynamic Panel Models with a Very Small T .” *Working Paper*.
 - Pickup, M., P. Gustafson, D. Cubranic, and G. Evans 2017. “OrthoPanels: An R Package for Estimating a Dynamic Panel Model with Fixed Effects Using the Orthogonal Reparameterization Approach.” *The R Journal* 9(1):60-76.
 - Kripfganz, S. 2016. “xtdpqml: Quasi-maximum likelihood estimation of linear dynamic short- T panel data models.” *The Stata Journal*.
 - Williams, R., Allison, P. and Moral-Benito, E., 2018. “Linear dynamic panel-data estimation using maximum likelihood and structural equation modeling.” *The Stata Journal* 18(2): 293-326.
- Topic 19: Pseudo-panels
 - Lebo, M. J. and Weber, C. 2015. “An effective approach to the repeated cross-sectional design.” *American Journal of Political Science* 59(1):242-258.
 - Lau, Richard, David J. Andersen, and David Redlawsk. 2008. “An Exploration of Correct Voting in Recent U.S. Presidential Elections.” *American Journal of Political Science* 52(2): 395-411.
- Topic 20: Missing data, multiple imputation, and the consequences of unbalanced panels
 - Honaker, James, and Gary King. 2010. “What to do about missing values in time-series cross-section data.” *American Journal of Political Science* 54(2):561-81.
 - Kropko, Jonathan, Ben Goodrich, Andrew Gelman, and Jennifer Hill. 2014. “Multiple Imputation for Continuous and Categorical Data: Comparing Joint Multivariate Normal and Conditional Approaches.” 22:497-519.
 - Rubin, Donald B. 1976. “Inference and missing data.” *Biometrika* 63(3):581-592.

Last updated: July 19, 2020