Panel Data and Longitudinal Analysis

ICPSR Summer Program in Quantitative Methods of Social Research

July 15-August 2, 2024 (second session)

Time:	1:30PM-4:30PM (Eastern Daylight Time, EDT)
Location:	Central Campus Classroom Bldg 3460 (CCCB 3460) (and virtual)
Instructor:	Dr. Andrew Q. Philips, University of Colorado Boulder
Email:	andrew.philips@colorado.edu
Instructor Office Hours:	M-F 12:15-1:15PM EDT, 1450 Thompson ISR
Instructor:	Dr. Mark Pickup, Simon Fraser University
Email:	mark.pickup@sfu.ca
Instructor Office Hours:	M-F 12:15-1:15PM EDT, 1450 Thompson ISR
Teaching Assistant:	Ramses Llobet, University of Washington
TA Email:	rllobet@uw.edu
TA Office Hours:	4:45PM-5:45 (M-TH) and 12:15-1:15PM F, 1450 ISR
Teaching Assistant:	Waseem Haider, University of Turku
TA Email:	waseem.w.haider@utu.fi
TA Office Hours:	4:45PM-5:45 (M-TH) and 12:15-1:15PM F, 1450 ISR

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), as well as various other topics. Throughout, we will also discuss several smaller topics in panel data. 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 both in-person and virtually. Virtual attendees 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). Please also have R/RStudio and Stata installed too. TAs can help you during office hours should you need help with installing these programs.

OFFICE HOURS: See above for office hours. Students are encouraged to meet with the instructors and TAs to go over any questions related to the course, including implementing what we've learned in their own projects. No advance notice is required; simply stop by.

Note that all instructors and TAs are also available to meet by appointment via Zoom, schedules permitting. Please email to set up a meeting.

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), but you do not need to purchase these:

- 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/Econometrics

  - Asteriou, Dimitrious 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.
  - Greene, William. 2011. *Econometric analysis.* 7th Edition. Upper Saddle River, NJ: Prentice-Hall.

- 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.

GRADES: While you do not have to take this course for credit, those of you wishing to earn ICPSR credit for this course must let the instructor/TAs know during the first week. To receive a grade you need to complete *all* assignments (there are several throughout the course) *on time*. These will be graded by the TAs. All others not taking the course for credit are still encouraged to complete these assignments too to assess how well they are learning the course material. The following seven assignments will be given (with deadlines subject to change based on the pace of the class):

- Assignment 1: Due July 16

- Assignment 2: Due July 18

- Assignment 3: Due July 21

- Assignment 4: Due July 23

- Assignment 5: Due July 25

- Assignment 6: Due July 29

- Assignment 7: Due July 31

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

  - Philips, Andrew Q. 2023. "`R`: A brief introduction."
  - Philips, Andrew Q. 2023. "`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. *Pooled Data Analysis for the Social Sciences*. Chapter 2.
  - 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.
  - Skim: Croissant Y, Millo G. 2008. "Panel Data Econometrics in R: The plm Package." Journal of Statistical Software, 27(2). URL http://www.jstatsoft.org/v27/i02/.

- Topic 3: Identifying spatio-temporal dependence

  - De Hoyos, R. E. and Sarafidis, V. (2006). "Testing for cross-sectional dependence in panel-data models." *The Stata Journal* 6(4):482-496.
  - Pesaran, M. H. 2021. "General diagnostic tests for cross section dependence in panels." *Empirical Economics* 60: 13-50.
  - Phillips, P.C. and Sul, D. 2003. "Dynamic panel estimation and homogeneity testing under cross section dependence." *The Econometrics Journal* 6(1):217-259.

- Topic 4: Robust estimation and FGLS procedures for standard error corrections

---

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

- Beck, N and J. Katz. 1995. "What To Do (and Not To Do) with Time Series Cross-Section Data." *American Political Science Review* 89:634-47.

- Mansournia, Mohammad Ali, Maryam Nazemipour, Ashley I. Naimi, Gary S. Collins, and Michael J. Campbell. 2021. "Reflection on modern methods: demystifying robust standard errors for epidemiologists." *International Journal of Epidemiology* 50(1): 346-351.

- Cameron, A. Colin and Douglas L. Miller. 2015. "A practitioner's guide to cluster-robust inference." *The Journal of Human Resources* 50(2): 317-372.

- 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.

- Abadie, A., Athey, S., Imbens, G.W. and Wooldridge, J.M., 2023. "When should you adjust standard errors for clustering?." *The Quarterly Journal of Economics*, 138(1): 1-35.

- 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.

  - Kropko, Jonathan and Robert Kubinec. 2020. "Interpretation and identification of within-unit and cross-sectional variation in panel data models" *PLoS ONE*: 1-22.

- 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. 2022. "Improving the interpretation of random effects regression results." *Political Studies Review*: 1-11.

- 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.

  - Bell, Andrew, Malcolm Fairbrother and Kelvyn Jones. 2019. "Fixed and random effects models: Making an informed choice." *Quality and Quantity* 53: 1051-1074.

  - Dieleman, Joseph and Tara Templin. 2014. "Random-Effects, Fixed-Effects and the within-between Specification for Clustered Data in Observational Health Studies: A Simulation Study" *PLoS ONE* 9(10): 1-17.

  - Schunck, Reinhard. 2013. "Within and between estimates in random-effects models: Advantages and drawbacks of correlated random effects and hybrid models." *The Stata Journal* 13(1): 65-76.

  - 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

  - Philips, Andrew Q. n.d. *Pooled Data Analysis for the Social Sciences*. Chapter 3.

  - 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: Review of 'fixed and random effects'

  - Hsiao C. 2014. *Analysis of Panel Data*, 3rd Edition. New York, NY: Cambridge University Press pp. 31-50 & 80-86.
  - Kropko J. and Kubinec R. 2020. "Interpretation and identification of within-unit and cross-sectional variation in panel data models." *PLoS ONE* 15(4):e0231349.

- Topic 12: Endogeneity, Nickell bias, and inconsistency, oh my! Instrumental variable approaches

  - 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.
  - Nickell, S. 1981. "Biases in Dynamic Models with Fixed Effects." *Econometrica* 49(6):1417-1426.
  - 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 13: GMM estimators for small T, large N

  - 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.
  - "xtdpdsys — Arellano–Bover/Blundell–Bond linear dynamic panel-data estimation." *Stata Documentation*
  - Roodman, David. "A Note on the Theme of Too Many Instruments." *Oxford Bulletin of Economics and Statistics* 71.1 (2009), 135-158.

- Topic 14: Transformed-likelihood, quasi- and full-maximum likelihood estimators for small T, large N

  - Pickup, M. and V. Hopkins. 2020. "Transformed-Likelihood Estimators for Dynamic Panel Models with a Very Small T." *Political Science Research and Methods* 10(2):333-352.
  - 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. "xtdpdqml: Quasi-maximum likelihood estimation of linear dynamic short-T panel data models." *The Stata Journal.*

- Topic 15: General-to-Specific: Some guidance on Model/Estimator Choices for small T, large N

  - Campos, J. N.R. Ericsson, and D.F. Hendry 2005. "General-to-specific Modeling: An Overview and Selected Bibliography." *Board of Governors of the Federal Reserve System: International Finance Discussion Papers* No. 838.

- Topic 16: Cross-sectional effect heterogeneity in static models

   – Hsiao C. 2014. *Analysis of Panel Data*, 3rd Edition. New York, NY: Cambridge University Press pp. 167-180.

   – Beck, N. and J.N. Katz. 2007. "Random Coefficient Models for Time-Series–Cross-Section Data: Monte Carlo Experiments." *Political Analysis* 15:182–195.

   – 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 17: Cross-sectional effect heterogeneity in dynamic models: Mean-group and pooled mean-group estimators

   – Chapter 22 Dynamic Heterogeneous Panel Data Models of Dimitrious Asteriou and Stephen G. Hall, *Applied Econometrics*, 3rd edition. Palgrave, 2016.

   – 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 18: Cross-sectional effect heterogeneity in dynamic models cont.: Common correlated effects

   – Ditzen, Jan. 2018. "Estimating Dynamic Common Correlated Effects in Stata." *The Stata Journal* 18(3):585-617.

   – Blackburne, Edward F. III and Mark W. Frank. 2007. "Estimation of nonstationary heterogeneous panels." *The Stata Journal* 7(2):197-208.

   – Advanced: Chudik, Alexander and M. Hashem Pesaran. 2015. "Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors." *Journal of Econometrics* 188(2):393-420,

- Topic 19: Hierarchical longitudinal modelling

   – Immerzeel, Tim and Mark Pickup. 2015. "Populist Radical Right Parties Mobilizing 'the People'? The Role of Populist Radical Right Success in Voter Turnout." *Electoral Studies* 40: 347-360.

   – Rabe-Hesketh, Sophia and Anders Skrondal. 2012. *Multilevel and Longitudinal Modeling Using Stata, Volumes I*. Third Edition. College Station, Texas: Stata Press. Chapters 5 and 9-10.

- 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 11, 2024