Panel Data and Longitudinal Analysis

ICPSR Summer Program in Quantitative Methods of Social Research

July 20-August 14, 2020

Time: 10:00AM-12:00PM
Location: Online instruction
Instructor: Dr. Andrew Q. Philips
Email: andrew.philips@colorado.edu
Teaching Assistant: TBD
TA Email: TBD
TA Office Hours:

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 Dropbox, so please sign up if you do not have an account before the first day of class (a basic account is free). 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**

- **Panel Data**

- **Multilevel Modeling/Hierarchical Linear Models**

**TENTATIVE SCHEDULE:**
Readings are 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.

**Week 1**

- **Day 1: Regression review and matrix algebra**

- **Day 2: Panel data fundamentals, describing and summarizing panel data**

- **Day 3: Identifying heterogeneity and spatio-temporal dependence**

---

1Note: Exact schedule may change. Topics roughly the same.
• Day 4: FGLS procedures for standard error corrections

• Day 5: Approaches to heterogeneity: Fixed and random effects

**Week 2**

• Day 1: FE/RE continued: Should I use fixed or random effects?

• Day 2: Alternative models for unit heterogeneity

• Day 3: Modeling and interpretation under dynamic models

• Day 4: Panel unit root testing

Day 5: Panel cointegration and models for cointegrating panel data


**Week 3**

Day 1: Modeling dichotomous-CSTS data

- Philips, Andrew Q. Working Paper. “mkduration: An easy way to create duration variables in binary cross-sectional time series data”

Day 2: Effect heterogeneity: fixed and random slopes


Day 3: Dynamic heterogeneity: Mean-group and pooled mean-group estimators


Day 4: Dynamic heterogeneity cont.: Common correlated effects


Day 5: Endogeneity, Nickell bias, and inconsistency, oh my! Instrumental variable approaches


**Week 4**

Day 1: GMM estimators


- **Day 2: GMM estimators continued**

- **Day 3: Transformed-likelihood, quasi- and full-maximum likelihood estimators**

- **Day 4: Pseudo-panels**

- **Day 5: Missing data, multiple imputation, and the consequences of unbalanced panels**

_Last updated:_ April 20, 2020