Theory and Hypothesis Testing with Panel Data

Alexander Pommells; Damian Boldt

Department of Political Science

Florida State University, Tallahassee, FL

Abstract

In political science research, researchers are generally concerned with using counterfactuals to determine the effect of a treatment variable. Counterfactuals are relatively easy to understand when data are only measured at one time period. However, in panel data (where multiple units are measured across a given length of time) formulating counterfactuals is much more complicated, because there is a lot more data involved. This added complexity means that there often is a lack of transparency of the difference between the theoretical hypothesis and what the data measures. In this project, we performed a literature review of articles from top international relations journals, hand-coded the kinds of data used for the articles and contrasted that with what would be expected based on the framing of their hypotheses. The goal of this research is to provide context that is important for interpreting political science research to close the gap between theory and hypothesis.
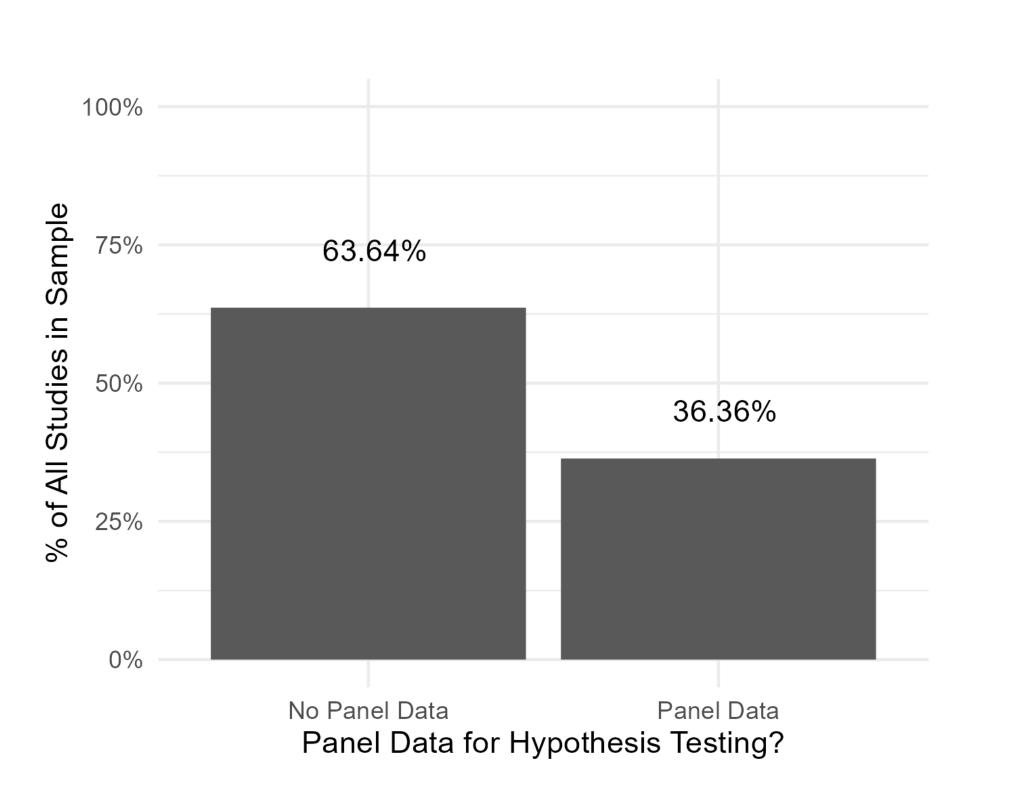
Background/Theory

Background: Panel data is a method of organizing data in which specific units are measured at different points over a set period of time. The goal of collecting data in this way is to be able to compare these units against each other at the same point in time, as well as to be able to compare these units against themselves over time. Within political science research, panel data is used often, both in analyses of domestic and international phenomena. The potential outcomes for a given unit in panel data analysis can be written as
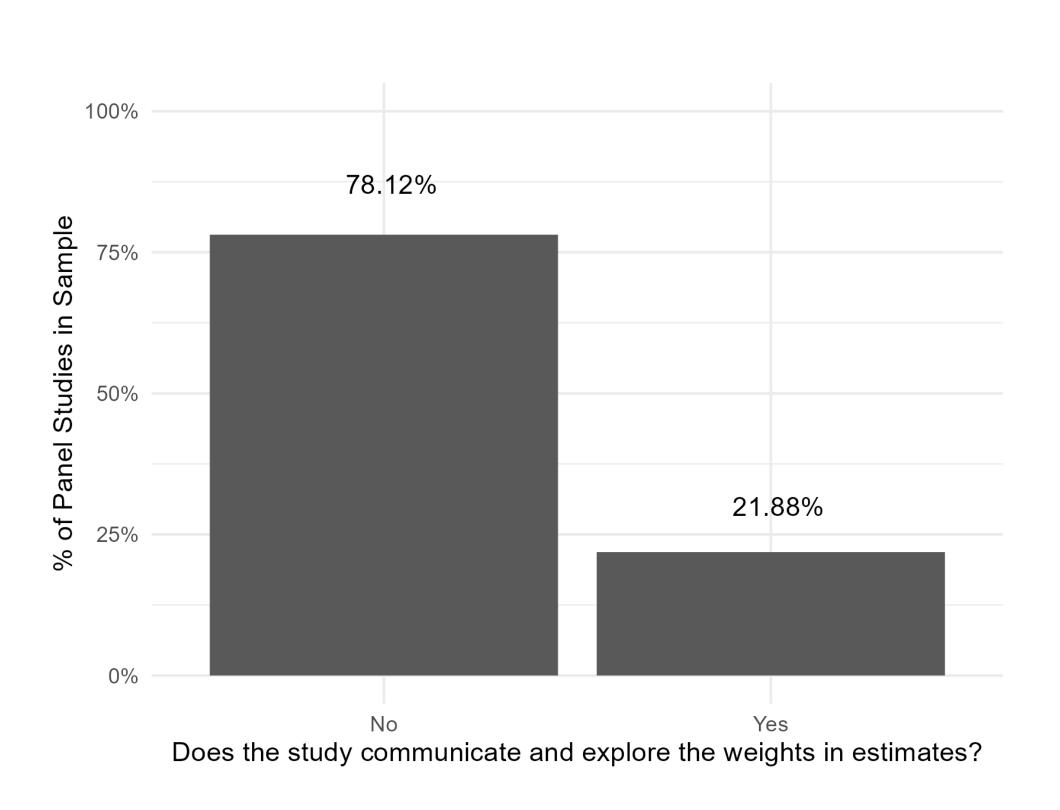
$$Y_{it}(\{v_{t-L}, v_{t-L+1}, \ldots, v_{t-1}\}, v_t, \{v_{t+1}, \ldots, v_{t+F-1}, v_{t+F}\})$$
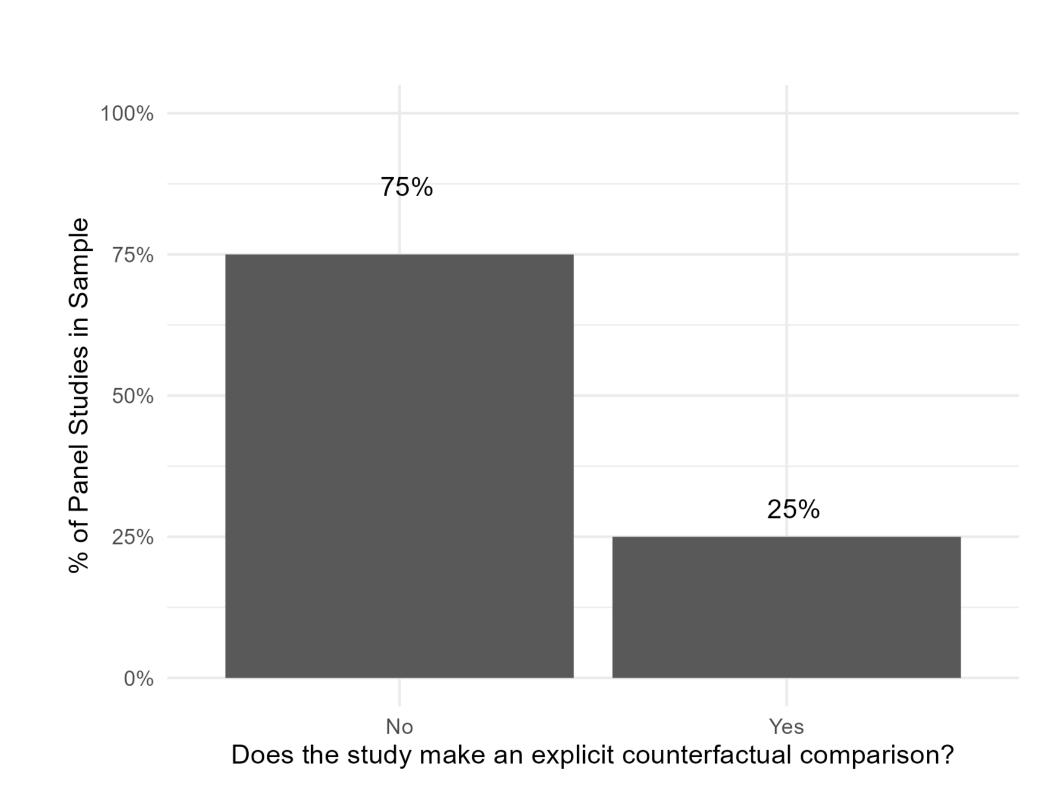
where v_t is a Treatment indicator of unit i at time t.

Theory: Panel data is uniquely complex because of the amount of data being evaluated. As a result, it allows for a wide array of possible comparisons that may be consistent with the researcher's central argument. As such, they often average out a lot of different comparisons in order to develop an empirical variable that best contorts around their hypothesis. However, researchers are often unclear in regard to which comparisons they consider to be important within the context of their work. This lack of clarity makes it more difficult to evaluate how well the comparisons that they are using to measure their results correspond with their theoretical hypothesis.

Figures/Results

Methods

We reviewed articles from two reputable international relations journals published in the year 2022: International Organization and Journal of Conflict Resolution We then coded a variety of variables, and conducted descriptive analysis on the data that was collected.

The variables of particular interest were as follows
- Whether or not the article uses panel data
- Whether or not the article stated an explicit hypothesis
- Whether or not the article stated a reference group
- Whether or not the article acknowledged the weighting of estimates

Conclusions

Throughout this project, we have attempted to establish two key things through the theory behind our work and the empirical examination of the literature

1. Accurate counterfactuals are difficult to develop for panel data, because of the amount of data that needs to be considered, and the length of the time period that that data is measured across

2. This difficulty at times results in a difficulty with connecting the theory behind panel data studies and what they measure empirically

We developed our theory for why this might be the case, before examining recent articles in multiple international relations journals, and found evidence that demonstrates our claims on several grounds. We found that panel data is quite common in international relations, making up over one-third of the studies that we looked at. This indicates that this issue is important to the field of social sciences and specifically international relations. Additionally, we found that most of the studies that we examined do not state relevant counterfactuals in their hypotheses, and furthermore that most of these studies do not specify the role that weights play in their analysis. These factors taken together demonstrate a significant lack of clarity in matching theoretical expectations of the studies examined with what the data gathered actually measures. Through presenting this newfound data, we hope to demonstrate the necessity for greater transparency of relevant counterfactuals in panel data research.

References