Introduction to the Virtual Issue: Panel Data Analysis and Regression Discontinuity

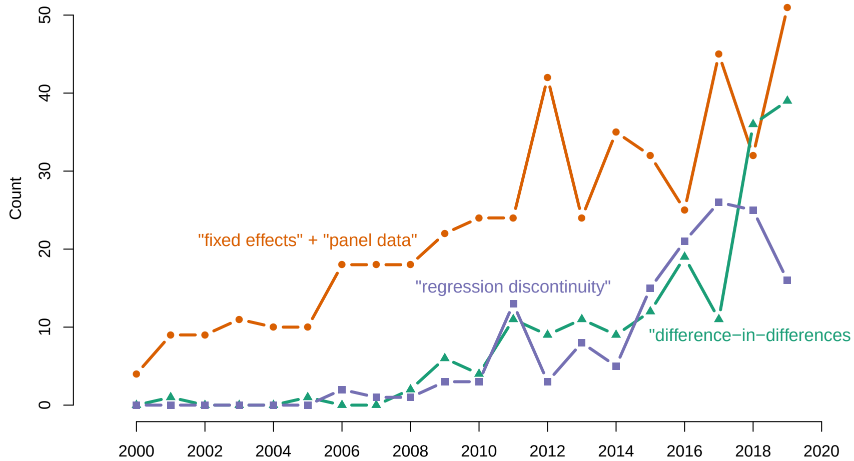
Yiqing Xu
Department of Political Science
Stanford University, Palo Alto CA 94035
email: yiqingxu@stanford.edu

Over the past ten years, the "causal inference revolution" has dramatically changed the landscape of political science and social sciences in general. The so-called (causal) identification problem, i.e., the challenge to identify parameters that have meaningful causal interpretations using observable data, has become part of the daily vocabulary in political science seminars around the world. Moreover, randomized control trials (RCTs), including various types of survey experiments, lab experiments, and field experiments, have entered the standard toolkit of political scientists.

Though RCTs deepen our understanding of politics by improving both precision and clarity of the arguments we make based on them, most political scientists would probably agree that studying politics using observational data remains extremely valuable, for two reasons. First, conclusions drawn from an experiment using a specific sample under a specific context may not be generalizable to the population of interest or to other contexts (i.e. it may not have external validity). Second, and perhaps more importantly, not all political science questions can be studied using experimental tools. In fact, many key variables central to our discipline, such as institutional changes and civil conflicts, are difficult, if not impossible, for researchers to experimentally "manipulate." Hence, researchers often turn to alternative design-based methods applicable to observational data (see, for example, the 2009 *Political Analysis* special issue on "Natural Experiments in Political Science").

Two of the most commonly used groups of methods for causal inference with observational data are (1) methods related to panel data or time-series-cross-section (TSCS) data and (2) regression discontinuity (RD) designs. Figure 1 shows the total numbers of articles mentioning "panel data" plus "fixed effects," "difference-in-differences," and "regression discontinuity" in the five top publications in political science from 2000 to 2019. Our discipline's increasing interests in, and wider acceptance of, applying these methods to address important political science questions is clear.

Figure 1. Number of Papers Published in the Five Top Political Science Journals: by Method



Note: Data are based on Google Scholar advanced searches (accessed on April 27, 2020). The five journals are *American Political Science Review*, *American Journal of Political Science*, *Journal of Politics*, *Comparative Political Studies*, *International Organization*. Papers may appear in more than one of the three categories.

Articles in this virtual issue represent the efforts of political methodologists either to develop more reliable and versatile approaches to panel data analysis and RD designs or to better understand the advantages and disadvantages of existing common practices in recent years. They are as follows:

1. Gaibulloev, Khusrav, Todd Sandler and Donggyu Sul (2014). "Dynamic Panel Analysis under Cross-Sectional Dependence." *Political Analysis* 22(2): 258–273.
2. Keele, Luke J. and Rocío Titiunik (2015). "Geographic Boundaries as Regression Discontinuities." *Political Analysis* 23(1): 127–155.
3. Caughey, Devin and Jasjeet S. Sekhon (2017). "Elections and the Regression Discontinuity Design: Lessons from Close U.S. House Races, 1942–2008." *Political Analysis* 19(4):385– 408.
4. Xu, Yiqing (2017). "Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models." *Political Analysis* 25(1): 57–76.
5. Choi, Jin-young and Myoung-jae Lee (2018). "Regression Discontinuity with Multiple Running Variables Allowing Partial Effects." *Political Analysis* 26(3): 258–274.
6. Plümper, Thomas and Vera E. Troeger (2019). "Not So Harmless After All: The Fixed-effects Model." *Political Analysis* 27(1): 21-45.
7. Ding, Peng and Fan Li (2019). "A Bracketing Relationship between Difference-in-Differences and Lagged-Dependent-Variable Adjustment." *Political Analysis* 27(4): 605–615.

Below I briefly introduce each article, followed by a discussion on future research agenda in the two broadly defined research areas.

Articles in this Virtual Issue

The first group of articles deal with longitudinal data or panel data. In political science, we also refer to longitudinal data in which the number of time periods, T, is big as time-series-cross-section (TSCS) data to highlight the importance of temporal variations within each unit (Beck and Katz 1995). To make causal arguments using observational panel data, political scientists usually rely on the following *workhorse* models: (1) difference-in-differences (DiD) models, (2) fixed effect (FE) models, and (3) models with lagged dependent variables (LDVs). These models rely on distinct, and often untestable, identification assumptions.

Two challenges immediately emerge. The first is that the within estimator is inconsistent when T is small and both FEs and LDVs are included in the specification (Nickell 1981), known as the Nickell bias with dynamic panel analysis. The conventional solution econometricians have proposed is to use an GMM/IV approach, but studies show that an GMM/IV estimator often produces finite sample biases even bigger than the Nickell bias (Beck and Katz 2011). The second challenge is that the error terms may be not only temporally correlated within each unit—this can be addressed by clustering the standard errors at the unit level—but also cross-sectionally dependent on each other. Neglecting the cross-sectional correlations may lead to biases in both the point estimates of treatment effects and estimates of their uncertainties. Four articles in this virtual issue address these challenges from different angles.

Gaibulloev, Sandler and Sul (2014) propose a potential solution for each of the two challenges. To minimize the Nickell bias, they subset data into subgroups, such as defined by regions or continents, and run separate FE regressions controlling for LDVs. The rationale behind this approach is that in each regression, N is smaller than T; as a result, the estimator is asymptotically pivotal (and its limiting distribution is centered around the true parameter), given a correct model specification. To deal with cross-sectionally correlated errors, the authors propose to use *factor-augmented models*, such as common correlated regression (CCE), iterative principal component (IPC) and projected principal component (PPC) estimators, to deal with cross-sectionally correlated errors. Applying these two approaches in their paper to the effect of terrorism on economic growth reveals qualitatively different results from those in the existing literature. This paper demonstrates the importance of taking both temporal and cross-sectional correlations seriously when using observational panel data. One caveat is that factor-augmented models require both N and T to be large. Hence, the proposal to subset data may exacerbate the incidental parameter problem in addition to its effect on power and multiple testing issues.

In **Xu (2017)**, I introduce the factor-augmented approach (specifically, the IPC estimator) to a multi-period DiD setup, in which treatment reversal never occurs, i.e., once a unit receives the treatment, it stays treated in the remaining time periods. A crucial difference between the DiD approach and two-way FE models is that with DiD, the average treatment effect on the treated (ATT) can be non-parametrically identified by averaging the differences in the actual outcomes for the treated units in the post-treatment periods and their predicted counterfactuals (based on additive fixed effects, strictly exogenous covariates, and/or additional factors). In other words, this approach flips the problem of controlling for cross-sectional correlations in the error terms by using these correlations for better counterfactual prediction when the "parallel trends" assumption appears to fail. Hence, the main contribution of this paper is to clarify that the causal inference problem with observational panel data is indeed a missing data problem and can be addressed with weaker modeling assumptions than standard FE models. The main drawbacks of this method are two-fold: (1) much like other factor-augmented approaches, it requires both N and T to be large; and (2) because counterfactual prediction in this context is essentially model-based extrapolation, it may lead to erroneous conclusions when treated and control units lack overlaps (e.g., in factor loadings) or are fundamentally different in other ways.

Plümper and Troeger (2019) caution against using FE models as a generic solution to analyzing observational panel data. A series of Monte Carlo simulations indicate that biases from FE models can be even bigger than those from a pooled OLS estimator when dynamics in the outcome variable Y and a key explanatory variable X are mis-specified. The reason for this pattern is that FE models, by solely relying on within-unit variations, amplify two types of biases: The first are induced by a time-varying confounder Z if its within correlations with both X and Y are bigger than its between correlations (e.g., if Z represents some unit-specific trends). The second type of biases are caused by spurious temporal correlations between X and Y (e.g., when they are independent of each other but both highly autoregressive). Plümper and Troeger do not advocate for a return to the pooled OLS estimator, but they warn researchers against treating FE models as a panacea in panel data analysis and that it is often necessary to model dynamic relationships in key variables of interest.

Ding and Li (2019) extend the famous bracketing relationship between DiD and LDV models to non-parametric settings with two groups (treated and control) and two periods. Angrist and Pischke (2009) show that estimates from DiD and LDV models offer an upper-bound and a lower-bound, respectively, for the true effect. Specifically, if the parallel trends assumption is correct, an LDV model will under-estimate the true effect, while if the (sequential) ignorability assumption is correct, a DiD model will over-estimate the true effect. Note that with a two-period two-group setup, a twoway FE estimator is mathematically equivalent to DiD. Ding and Li (2019) prove that this bracketing relationship holds when one replaces the linear LDV model with some non-parametric adjustment methods, such as inverse propensity score reweighting. This result is important because the choice between these two assumptions is often arbitrary in applied research and new methods of the non-parametric adjustment are more frequently applied to panel/TSCS data (Imai, Kim and Wang 2018; Strezhnev 2018; Hazlett and Xu 2018). Future work is needed to further extend this result to empirical settings with more than two periods and/or two groups.

The second group of articles in this virtual issue concerns regression discontinuity (RD) designs. **Keele and Titiunik (2015)** proposes the Geographic Regression Discontinuity (GRD) design, in which the discontinuity in treatment assignment is geographic. This paper clarifies the GRD's identification assumptions and proposes a method to estimate geographically located treatment effects. Specifically, it shows that GRD is equivalent to a standard RD with *two* running variables and that using a one-dimensional distance metric can lead to "bad matches." It also develops a spatial balance test that compares nearest geographic neighbors and can shed light on the validity of the identification assumptions. Since the publication of this papers and other companion research, the GRD has become an increasingly popular tool among political scientists; 11 papers using GRDs have appeared in the five top political science journals alone.

Relatedly, **Choi and Lee (2018)** study RD designs with multiple running variables. In this setup, a unit is considered fully treated if each running variable passes a pre-specified cutoff. For example, if the treatment is the Republican Party controlling both chambers of the legislature, it switches to 1 when Republicans won elections in both the House and the Senate. Different from previous work, the authors relax the restriction that units that not all fully treated are categorized as untreated; instead, they develop a new estimator, based on a weaker continuity assumption, to allow partial effects due to each running variable passing its cutoff. In the above example, therefore, the number of treatment status increases to four. Using data from the US congress from 1789 to 2004, they find statistically significant partial effects of party control on legislative productivity.

The last paper in this virtual issue is **Caughey and Sekhon (2017)**. This paper points out potential caveats of applying RD designs in election settings. Using data from the US House of Representatives from 1942 to 2008, they show considerable imbalance in pre-treatment variables of candidates who barely win and who barely loose. This paper is included in this issue because it starts an important conversation on the validity of RD designs in the study of politics, of which strategic behavior is a central component. For example, Eggers et al. (2015) report that other legislative elections around the world, and other US elections, do not show such anomalies; Erikson and Rader (2017) argue that the systematic differences between winners and losers in close elections can be explained almost completely by candidates' incumbent party status. **Caughey and Sekhon (2017)** serves as a much-needed reminder of the necessity of routine diagnostic checks, such as the McCrary test and a balance or equivalence test, when an RD design is being executed.

Future Research Agenda

Despite the progress researchers have made in recent years in causal inference using panel data and with RD designs, many questions remain. First and foremost, because most existing approaches to panel data analysis reply heavily on parametric models with relatively stringent assumptions, there is a notable gap between current practices and a truly design-based perspective. Several recent papers attempt to fill this gap with simplified settings (e.g., Imai and Kim 2019, Athey and Imbens 2018), but more work needs to be done to extend their results more general contexts, including treatment reversal and continuous treatments.

Second, inference remains challenging when (1) the number of treated units is small (such as in synthetic control settings) or (2) interference takes places across both time and unit dimensions. In the first scenario, a full Bayesian approach may be a plausible solution. New inferential methods need to be developed to accommodate various cases in the second scenario—one such example is Aronow, Samii and Assenova (2017), in which a cluster-robust variance estimator is developed for dyadic data. Further more, powerful machine learning tools developed in computer science and statistics can be utilized to improve counterfactual prediction (e.g. Athey et al. 2018) or better incorporate propensity scores.

On the study of RD designs, to the best of my knowledge, recent research has been focusing on establishing robust estimation strategies and inferential methods (e.g. Calonico, Cattaneo and Titiunik 2014; Imbens and Wager 2019) and incorporating covariates to improve efficiency (Calonico et al. 2019). More research is needed to address the external validity concerns of RD designs and to link the local estimates to a bigger segment of the population of interest.

Concluding Remarks

Along with the rise of using experimental methods in political science, casual inference with observational data has become increasingly popular among political scientists. In this endeavor, panel data analysis and regression discontinuity designs are two common choices that, if properly applied, can produce credible casual estimates. Articles included in this virtual issue constitute a sample of methodological innovations in these two areas in political science in recent years. Future research is needed to more closely link panel data methods with a design-based perspective and expand the external validity of RD designs.

About the Author

Yiqing Xu is an Assistant Professor of Political Science at Stanford University. He has worked on methods and applications related to causal inference with panel/TSCS data in the social sciences . His published and working papers are available at http://yiqingxu.org.

References

Angrist, Josh D. and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Aronow, Peter M, Cyrus Samii and Valentina A Assenova. 2017. "Cluster–Robust Variance Estimation for Dyadic Data." *Political Analysis* 23(4):564–577.

Athey, Susan and Guido W Imbens. 2018. Design-based analysis in difference-in-differences settings with staggered adoption. Technical report National Bureau of Economic Research.

Athey, Susan, Mohsen Bayati, Nikolay Doudchenko, Guido Imbens and Khashayar Khosravi. 2018. Matrix completion methods for causal panel data models. Technical report National Bureau of Economic Research.

Beck, Nathaniel and Jonathan N. Katz. 1995. "What to do (and not to do) with Time-Series Cross-Section Data." *American Political Science Review* 89(3):634–647.

Beck, Nathaniel and N. Katz, Jonathan. 2011. "Modeling Dynamics in Time-Series-Cross-Section Political Economy Data." *Annual Review of Political Science* 14:331–352.

Calonico, Sebastian, Matias D Cattaneo, Max H Farrell and Rocio Titiunik. 2019. "Regression discontinuity designs using covariates." *Review of Economics and Statistics* 101(3):442–451.

Calonico, Sebastian, Matias D Cattaneo and Rocio Titiunik. 2014. "Robust nonparametric confidence intervals for regression-discontinuity designs." *Econometrica* 82(6):2295–2326.

Caughey, Devin and Jasjeet S Sekhon. 2017. “Elections and the Regression Discontinuity Design: Lessons from Close U.S. House Races, 1942–2008.” *Political Analysis* 19(4):385–408.

Choi, Jin-young and Myoung-jae Lee. 2018. “Regression Discontinuity with Multiple Running Variables Allowing Partial Effects.” *Political Analysis* 26(3):258–274.

Ding, Peng and Fan Li. 2019. “A Bracketing Relationship between Difference-in-Differences and Lagged-Dependent-Variable Adjustment.” *Political Analysis* 27(4):605–615.

Eggers, Andrew, Olle Folke, Anthony Fowler, Jens Hainmueller, Andrew B. Hall and James M. Snyder. 2015. “On the Validity Of the Regression Discontinuity Design for Estimating Electoral Effects: New Evidence from Over 40,000 Close Races.” *American Journal of Political Science* 59(1):259–27.

Erikson, Robert S and Kelly Rader. 2017. “Much Ado About Nothing: RDD and the Incumbency Advantage.” *Political Analysis* 25(2):269–275.

Gaibulloev, Khusrav, Todd Sandler and Donggyu Sul. 2014. “Dynamic Panel Analysis under Cross-Sectional Dependence.” *Political Analysis* 22(2):258–273.

Hazlett, Chad and Yiqing Xu. 2018. “Trajectory Balancing: A General Reweighting Approach to Causal Inference with Time-Series Cross-Sectional Data.” Working Paper, UCLA and UCSD.

Imai, Kosuke and In Song Kim. 2019. “When Should We Use Linear Fixed Effects Regression Models for Causal Inference with Longitudinal Data.” *American Journal of Political Science* 63(2):467–490.

Imai, Kosuke, In Song Kim and Erik Wang. 2018. “Matching Methods for Causal Inference with Time-Series Cross-Section Data.” Working Paper, Princeton University.

Imbens, Guido and Stefan Wager. 2019. “Optimized regression discontinuity designs.” *Review of Economics and Statistics* 101(2):264–278.

Keele, Luke J and Rocío Titiunik. 2015. “Geographic Boundaries as Regression Discontinuities.” *Political Analysis* 23(1):127–155.

Nickell, Stephen. 1981. “Biases in Dynamic Models with Fixed Effects.” *Econometrica* 49(6):1417–1426.

Plümper, Thomas and Vera E. Troeger. 2019. "Not So Harmless After All: The Fixed-effects Model." *Political Analysis* 27(1):21–45.

Strezhnev, Anton. 2018. "Semiparametric weighting estimators for multi-period difference-in-differences designs." Working Paper, Harvard University.

Xu, Yiqing. 2017. "Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models." *Political Analysis* 25(1):57–76.