1 Research

My primary research area is econometrics, with a focus on causal inference. I have also worked on topics in algorithmic fairness and in empirical microeconomics. Broadly, my research in econometrics seeks to help empirical economists draw more credible conclusions from data. My work often highlights hidden assumptions implicitly relied on in empirical work, which may be questionable in many contexts. I then develop new methodology that either avoids the most questionable assumptions, or allows the researcher to assess the sensitivity of their conclusions to violations of these assumptions.

1.1 Econometrics Research

Difference-in-differences. Much of my research has been on improving the credibility of difference-in-differences (DiD) designs, one of the most popular tools for policy analysis in economics. The basic idea of DiD is to estimate the impact of the introduction of a new policy or intervention (typically called a "treatment") by comparing the change in the outcome for treated individuals around the time of the policy introduction to the analogous change for a comparison group of individuals who were not treated. For example, we might look at the change in health outcomes in states that implemented Medicaid expansion, and compare this to the analogous change in health outcomes for states that did not expand Medicaid. The key assumption for the validity of DiD designs is the so-called "parallel trends" assumption, which states that the average change in the outcome for the treated group would have been the same as the average change in the outcome for the comparison group *if not* for the implementation of the policy. In practice, researchers often try to validate the parallel trends assumption by testing whether the two groups' outcomes appeared to have been moving in parallel prior to the implementation of the policy (a "pre-trends" test).

In Roth (2022), I highlighted several important limitations with the common approach for validating the parallel trends assumption using pre-trends tests. The first relates to a lack of statistical power: even if the parallel trends assumption is violated prior to the treatment, owing to noise in the data the researcher may fail to find a statistically significant violation. In a systematic survey of recent papers published in top economics journals, I show that statistical power is often quite low to detect meaningful violations of parallel trends. The second issue raised in the paper is that conditioning the analysis on having "passed" the test for pre-trends introduces a form of selection bias known as pre-testing, which I show in many cases can further exacerbate the bias from a violation of parallel trends.

In related work with Pedro Sant'Anna (Roth and Sant'Anna, 2023), I also provide arguments as to why we should often be skeptical of the validity of the parallel trends assumption *ex ante*. That paper shows that, except in special cases, the parallel trends assumption will be sensitive to the functional form of the outcome—e.g., do we specify the outcome in levels or in logs? More formally, we provide a complete characterization of when parallel trends is invariant to functional form, and show that this occurs if and only if the treatment is as good as randomized, there are no time effects, or some combination thereof. Outside of these special cases, the validity of the parallel trends assumption will necessarily depend on the choice of functional form.

To address these concerns about the validity of the parallel trends assumption, Ashesh Rambachan and I developed methodology for more credibly assessing the sensitivity of conclusions to violations of the parallel trends assumption (Rambachan and Roth, 2023). Instead of assuming that the parallel trends assumption holds exactly, our approach is based on the idea that the violations of parallel trends prior to the treatment are informative about the post-treatment violations. For example, one formulation of the approach is to assume that the post-treatment violations are no more than M times larger than the largest pre-treatment violation, for some number M. We then develop confidence intervals (CIs) that are valid for the causal effect of interest under this restriction. These CIs account both for the potential violations of parallel trends as well as for the statistical uncertainty in measuring the pre-trends. The approach enables a natural form of sensitivity analysis, as it allows researchers to make statements along the lines of, "our result remains statistically significant allowing for violations of the parallel trends assumption up to 2 times larger than that seen in the pre-treatment period." In follow-on work (Kwon and Roth, 2024a), we provide a related approach from the Bayesian (rather than Frequentist) statistical perspective, which allows the researcher to incorporate prior information about what they expect violations of parallel trends to look like.

While often the parallel trends assumption will be questionable, one special case where its validity is guaranteed is when the timing of treatment is randomized. Although researchers often use DiD in such settings, in Roth and Sant'Anna (2023) we show that one can actually obtain substantially more precise estimates by exploiting the randomized rollout directly. We derive the most efficient estimator under random treatment timing in a class of estimators that nests many commonly-used DiD approaches. In an application to the staggered rollout of a police training program, we find that the proposed efficient estimator can reduce the length of confidence intervals by a factor of 8 relative to existing approaches.

The large number of recent papers on DiD (by myself and others) has made it difficult for practitioners to keep up with the state-of-the-art best practices. In Roth, Sant'Anna, Bilinski and Poe (2023), we provide an accessible review of the recent literature, accompanied with practical recommendations for applied researchers. This paper has become a standard reference, with over 1000 Google Scholar citations.

Finally, in a recent working paper (Roth, 2024), I show that the event-study plots produced by software for implementing some of the recent DiD methods may be misleading, as they differ in important ways from conventional event-study plots to which empirical researchers are accustomed. This note has influenced some of the most popular packages for DiD to update their default settings for creating event-study plots to increase interpretability.

Other econometrics work. My other econometrics work has focused on improving empirical practice in a variety of domains.

In Chen and Roth (2024), we identify problems with the common practice of estimating treatment effects for "log-like" transformations of the outcome, and propose some practical solutions. As background, when the outcome Y is strictly positive, the average treatment effect in logs has the useful feature that its units approximate percentage changes. In practice, however, the outcome of interest (e.g. earnings) may equal zero for some observations, in which case $\log(Y)$ is not well-defined. Researchers therefore commonly use "log-like" transformations such as $\log(1 + Y)$ or $\text{arcsinh}(Y)$ that approximate the logarithm for large values of Y but are defined at zero. In Chen and Roth (2024), we show that treatment effects for such transformations should not be interpreted as percentage effects: In contrast with a percentage, which is unit-free, the treatment effects for log-like transformations are arbitrarily sensitive to the units of the outcome when the treatment affects the extensive margin. We further prove a "trilemma", which shows that when the outcome can equal zero, there is no treatment effect parameter that is: (i) an average of individual level effects, (ii) invariant to the units of the outcome, and (iii) point-identified. In light of this result, we enumerate a variety of possible paths forward that relax one of the criteria in the trilemma. We discuss the economic rationale for each of these solutions, and illustrate how they can be used in practice in three empirical applications.

In Andrews, Roth and Pakes (2023), we provide computationally tractable and powerful methods for inference on parameters defined by conditional moment inequalities. As background, many economic models imply that some inequalities should hold on average: for example, if a consumer chooses good A over good B, then their expected utility from good A must have been higher than for good B. Such restrictions are known as moment inequalities, and they arise in a large range of economic settings—including many models of consumer choice, regression with interval-valued outcomes, as well as the bounds on treatment effects in DiD settings discussed earlier (Rambachan and Roth, 2023). There are two practical challenges in conducting inference based on moment inequalities, however: first, existing methods become computationally challenging when there more than a few parameters to estimate (including nuisance parameters, such as coefficients on control variables); second, the power of existing methods is often very sensitive to the inclusion of inequalities that are far from binding. In Andrews et al. (2023), we show that the moment inequalities that arise in economic applications often have a particular structure in which the nuisance parameters enter linearly. We exploit this structure to develop tests that are computationally tractable (by exploiting linear programming) and which are insensitive to the inclusion of slack moments. In Rambachan and Roth (2023), we further show that the proposed tests achieve optimal local asymptotic power in a variety of settings.

In Kwon and Roth (2024b), we develop practical tools for understanding *why* a particular economic treatment affects an outcome of interest. Researchers often conjecture that the effect of the treatment operates through a particular mechanism (or set of mechanisms) M. Although formal tools exist for testing such hypotheses, they typically rely on strong assumptions, such as that M is as good as randomly assigned conditional on observables. In the paper, we show that under much weaker assumptions, one can obtain lower-bounds on the fraction of individuals whose outcome is affected by the treatment despite having no change in M. These lower-bounds allow us to test whether a particular M can fully explain the treatment effect, and if not, to quantify the magnitude of the alternative mechanisms.

Finally, in Rambachan and Roth (2024), we study the problem of conducting statistical inference in settings where the full population is observed (e.g. we have data on all 50 US states). In such settings, the traditional statistical framework that views the data as having been sampled from some super-population may be unnatural. A more natural approach in such settings is to view the uncertainty as *design-based*, i.e. arising from the stochastic assignment of treatment rather than from sampling. Existing design-based frameworks have primarily focused on the setting where the treatment is randomly assigned. In our paper, we extend the design-based framework of uncertainty to cover common quasi-experimental strategies in economics such as DiD and instrumental variables.

Statistical software. I have (co-)developed statistical software to implement the methods in many of the papers described above, including the `HonestDiD`, `pretrends`, and `staggered` Stata and R packages.

1.2 Other work.

Algorithmic fairness. Another strand of my research has studied the fairness of algorithms. There is often concern that if an algorithm is trained on data generated by a biased decision maker, then it will learn to reflect this bias. In Rambachan and Roth (2020), we show that whether this is the case depends crucially on how the algorithm is trained. If the algorithm is trained directly to predict the human decision, then as expected the algorithm will inherit the human's biases. On the other hand, we show that if the algorithm is trained to predict an objective outcome in a sample of people selected by a biased human decisionmaker (e.g. predict productivity among previously hired applicants), then the algorithm will actually *reverse* the human's bias. Much of the algorithmic fairness literature has focused on binary decisions (e.g. grant a loan or not). However, in practice algorithms often produce rankings—e.g. LinkedIn's search algorithm produces an ordered list of candidates. In collaboration with researchers at LinkedIn (Roth, Saint-Jacques and Yu, 2022), I developed statistical tools for testing for algorithmic bias when the algorithm produces a ranked list rather than a binary decision.

Empirical microeconomics. In Katz, Roth, Hendra and Schaberg (2022), we study a class of job training programs called sectoral employment programs, which focus on providing participants with credentials to obtain jobs in high-paying sectors. We show that sectoral training programs have been unusually effective relative to other job training programs, and analyze data from the WorkAdvance randomized trial to provide evidence on the mechanisms. In Roth (2019), I study the impacts of Wisconsin's Act 10, which severely weakened teachers' unions, on the labor supply of teachers. I show that many teachers were induced to retire early as a result of the reform. Finally, many of my econometrics papers contain a substantial empirical component. For example, Roth and Sant'Anna (2023) provides the most precise estimates to date of procedural justice training programs for police officers. Our work also revealed a statistical error in a prior analysis by Wood, Tyler and Papachristos (2020), which led to a correction in *PNAS* and a re-analysis co-authored with the authors of the flawed study (Wood, Papachristos and Tyler, 2021a; Wood, Tyler, Papachristos, Roth and Sant'Anna, 2021b).

2 Teaching and Service

Teaching. At Brown, I have been teaching ECON 1630 (Mathematical Econometrics I), which covers basic probability, regression analysis, and methods for causal inference such as DiD, instrumental variables, and regression discontinuity. In collaboration with Peter Hull, I have developed all of the materials for the course, including detailed slides and problem sets. The course gives a modern and rigorous introduction to econometrics, particularly causal inference, while also teaching students to implement these tools in practice. The lectures alternate between applied examples and the technical methodology, and the assignments for the course walk the students through applications of the methodology to real empirical applications. In the classroom, I emphasize student participation, and am proud of the back-and-forth I have fostered with my students. This has been reflected in the teaching evaluations for the course—which have averaged a 4.67 out of 5 in the five sections I have taught—with one student writing "Professor Roth has done an amazing job this semester in running an engaging class setting" and another "I think Professor Roth has done an excellent job of taking what is a very complicated topic and distilling it down to what we needed to know... Professor Roth also did an excellent job of listening to students and actually answering the question they are asking, noting if something needs to be explained more clearly based on the question".

I have also had the opportunity to give several mini-courses and tutorials on DiD. I have given mini-courses (1-3 days in length) at the University of Exeter and Tel Aviv University, and online through the University of Michigan and the *Mixtape Sessions* online platform run by Professor Scott Cunningham (Baylor). I also gave shorter tutorials for the Chamberlain Online Seminar in Econometrics and One World YoungStatS. The material for these courses is publicly available on my webpage and has received over 45,000 views.

Advising. I enjoy advising student research at both the undergraduate and graduate level. I have been the supervisor of three undergraduate theses in my time at Brown, and have served on the graduate dissertation committee of Patrick Vu (chair), Tommaso Coen, and Vincent Starck. I am also serving on the committee of two current graduate students. In addition, I regularly meet with students about econometric issues they encounter in their research, and provide feedback at both the econometrics and applied micro student workshops.

Service. I have served the department in a variety of roles. I was on the junior hiring committee in 2022-2023, and have been serving as an undergraduate concentration advisor. I have co-organized both the econometrics and applied microeconomics seminars. I have also helped to organize a variety of conferences and events, including an interdisciplinary conference on the economics of algorithms, the Greater New York Econometrics Colloquium, and a Replication Games event in collaboration with the Institute for Replication.

References

Andrews, Isaiah, Jonathan Roth, and Ariel Pakes, "Inference for Linear Conditional Moment Inequalities," *The Review of Economic Studies*, November 2023, *90* (6), 2763–2791.

Chen, Jiafeng and Jonathan Roth, "Logs with Zeros? Some Problems and Solutions," *The Quarterly Journal of Economics*, 2024, *139* (2), 891–936.

Katz, Lawrence F., Jonathan Roth, Richard Hendra, and Kelsey Schaberg, "Why Do Sectoral Employment Programs Work? Lessons from WorkAdvance," *Journal of Labor Economics*, April 2022, *40* (S1), S249–S291. Publisher: The University of Chicago Press.

Kwon, Soonwoo and Jonathan Roth, "(Empirical) Bayes Approaches to Parallel Trends," *AEA Papers and Proceedings*, May 2024, *114*, 606–609.

_ and _ , "Testing Mechanisms," April 2024. arXiv:2404.11739 [econ, stat].

Rambachan, Ashesh and Jonathan Roth, "Bias In, Bias Out? Evaluating the Folk Wisdom," *LIPIcs, Volume 156, FORC 2020*, 2020, *156*, 6:1–6:15.

_ and _ , "A More Credible Approach to Parallel Trends," *The Review of Economic Studies*, October 2023, *90* (5), 2555–2591.

_ and _ , "Design-Based Uncertainty for Quasi-Experiments," February 2024. arXiv:2008.00602 [econ, stat].

Roth, Jonathan, "Union reform and teacher turnover: Evidence from Wisconsin's Act 10," *Working paper*, 2019.

_ , "Pretest with Caution: Event-Study Estimates after Testing for Parallel Trends," *American Economic Review: Insights*, September 2022, *4* (3), 305–322.

_ , "Interpreting Event-Studies from Recent Difference-in-Differences Methods," January 2024. arXiv:2401.12309 [econ, stat].

_ and Pedro H. C. Sant'Anna, "Efficient Estimation for Staggered Rollout Designs," *Journal of Political Economy Microeconomics*, November 2023, *1* (4), 669–709. Publisher: The University of Chicago Press.

_ and Pedro HC Sant'Anna, "When is parallel trends sensitive to functional form?," *Econometrica*, 2023, *91* (2), 737–747.

_ , Guillaume Saint-Jacques, and YinYin Yu, "An Outcome Test of Discrimination for Ranked Lists," in "Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency" FAccT '22 Association for Computing Machinery New York, NY, USA June 2022, pp. 350–356.

_ , Pedro H. C. Sant'Anna, Alyssa Bilinski, and John Poe, "What's trending in difference-in-differences? A synthesis of the recent econometrics literature," *Journal of Econometrics*, August 2023, *235* (2), 2218–2244.

Wood, George, Andrew V. Papachristos, and Tom R. Tyler, "Correction for Wood et al., Procedural justice training reduces police use of force and complaints against officers," *Proceedings of the National Academy of Sciences*, July 2021, *118* (27).

_ , Tom R. Tyler, and Andrew V. Papachristos, "Procedural justice training reduces police use of force and complaints against officers," *Proceedings of the National Academy of Sciences*, May 2020, *117* (18), 9815–9821.

_ , _ , _ , Jonathan Roth, and Pedro HC Sant'Anna, "Revised findings for "Procedural justice training reduces police use of force and complaints against officers"," Technical Report, Center for Open Science 2021.