Yong Cai

Northwestern
Economics

Placement Director:	Professor Alessandro Pavan	(+1) 847-491-8266	alepavan@northwestern.edu
Placement Administrator:	Lola Ittner	(+1) 847-491-5213	econjobmarket@northwestern.edu

Contact Information

Department of Economics
Northwestern University
2211 Campus Drive
Evanston, IL 60208

Mobile: (+1) 773-796-0405
yongcai2023@u.northwestern.edu
www.sites.northwestern.edu/ycz9493/
Citizenship: Singaporean

Fields

Econometrics

Education

Northwestern University
Ph.D. Economics — 2023 (expected)

London School of Economics and Political Science
M.Sc. Econometrics and Mathematical Economics (Distinction) — 2017
B.Sc. Econometrics and Mathematical Economics (First Class Honours) — 2016

Fellowships & Awards

Dissertation University Fellowship — 2022–23
Distinguished Teaching Assistant Award — 2021
Economics Department Prize — 2016
Economics Examiners' Prize — 2014
Singapore LSE Trust Study Award — 2013–16

Teaching Experience

Teaching Assistant, Northwestern University
ECON 201 Introduction to Macroeconomics — 2018–21
ECON 381-1 Econometrics — 2021
ECON 480-3 Econometrics — 2019–20

Course Instructor, Northwestern University
Econometrics Review for Incoming Ph.D. Students — 2019

Teaching Assistant, London School of Economics
EC210 Intermediate Macroeconomics — 2016–17

Research Experience

Research Assistant, Northwestern University
Professor Eric Auerbach — 2020–21
Professor Ivan Canay — 2018–19

Research Assistant, London School of Economics
Professor Shengxing Zhang — 2017
Professor David Baqaee — 2016
Behavioural Research Lab — 2014–17

Refereeing

Journal of Business and Economic Statistics

Job Market Paper

"Linear Regression with Network Centrality Measures"

This paper studies the properties of linear regression on centrality measures when network data is sparse – that is, when there are many more agents than links per agent – and when they are measured with error. We make three contributions in this setting: (1) We show that OLS estimators can become inconsistent under sparsity and characterize the threshold at which this occurs, with and without measurement error. This threshold depends on the centrality measure used. Specifically, regression on eigenvector is less robust to sparsity than on degree and diffusion. (2) We develop distributional theory for OLS estimators under measurement error and sparsity, finding that OLS estimators are subject to asymptotic bias even when they are consistent. Moreover, bias can be large relative to their variances, so that bias correction is necessary for inference. (3) We propose novel bias correction and inference methods for OLS with sparse noisy networks. Simulation evidence suggests that our theory and methods perform well, particularly in settings where the usual OLS estimators and heteroskedasticity-consistent/robust t-tests are deficient. Finally, we demonstrate the utility of our results in an application inspired by De Weerdt and Dercon (2006), in which we study the relationship between consumption smoothing and informal insurance in Nyakatoke, Tanzania.

Publications

"On the Implementation of Approximate Randomization Tests in Linear Models with a Small Number of Clusters" with Ivan Canay, Deborah Kim and Azeem Shaikh.

Accepted, Journal of Econometric Methods.

This paper provides a user's guide to the general theory of approximate randomization tests developed in Canay et al. (2017a) when specialized to linear regressions with clustered data. An important feature of the methodology is that it applies to settings in which the number of clusters is small – even as small as five. We provide a step-by-step algorithmic description of how to implement the test and construct confidence intervals for the parameter of interest. In doing so, we additionally present three novel results concerning the methodology: we show that the method admits an equivalent implementation based on weighted scores; we show the test and confidence intervals are invariant to whether the test statistic is studentized or not; and we prove convexity of the confidence intervals for scalar parameters. We also articulate the main requirements underlying the test, emphasizing in particular common pitfalls that researchers may encounter.

Working Papers

"A Modified Randomization Test for the Level of Clustering"

R&R, Journal of Business and Economic Statistics.

Suppose a researcher observes individuals within a county within a state. Given concerns about correlation across individuals, at which level should they cluster their observations for inference? This paper proposes a modified randomization test as a robustness check for their chosen specification in a linear regression setting. Existing tests require either the number of states or number of counties to be large. Our method is designed for settings with few states and few counties. While the method is conservative, it has competitive power in settings that may be relevant to empirical work.

"On the Performance of the Neyman Allocation with Small Pilots" with Ahnaf Rafi. *Submitted.*

The Neyman Allocation and its conditional counterpart are used in many papers on experiment design, which typically assume that researchers have access to large pilot studies. This may not be realistic. To understand the properties of the Neyman Allocation with small pilots, we study its behavior in a novel asymptotic framework for two-wave experiments in which the pilot size is assumed to be fixed while the main wave sample size grows. Our analysis shows that the Neyman Allocation can lead to estimates of the ATE with higher asymptotic variance than with (non-adaptive) balanced randomization, particularly when the population is relatively homoskedastic. We also provide a series of empirical examples showing that the Neyman Allocation may perform poorly for values of homoskedasticity that are relevant for researchers. Our results suggest caution when employing experiment design methods involving the Neyman Allocation estimated from a small pilot study.

"Heterogeneous Treatment Effects for Networks, Panels, and other Outcome Matrices" with Eric Auerbach. *Submitted.*

We are interested in the distribution of treatment effects for an experiment where units are randomized to treatment but outcomes are measured for pairs of units. For example, we might measure risk sharing links between households enrolled in a microfinance program, employment relationships between workers and firms exposed to a trade shock, or bids from bidders to items assigned to an auction format. Such a double randomized experimental design may be appropriate when there are social interactions, market externalities, or other spillovers across units assigned to the same treatment. Or it may describe a natural or quasi experiment given to the researcher. In this paper, we propose a new empirical strategy based on comparing the eigenvalues of the outcome matrices associated with each treatment. Our proposal is based on a new matrix analog of the FrÃ©chet-Hoeffding bounds that play a key role in the standard theory. We first use this result to bound the distribution of treatment effects. We then propose a new matrix analog of quantile treatment effects based on the difference in the eigenvalues. We call this analog spectral treatment effects.

"It's not always about the money, sometimes it's about sending a message: Evidence of Informational Content in Monetary Policy Announcements" with Santiago Camara and Nicholas Capel.

This paper introduces a transparent framework to identify the informational content of FOMC announcements. We do so by modelling the expectations of the FOMC and private sector agents using state of the art computational linguistic tools on both FOMC statements and New York Times articles. We identify the informational content of FOMC announcements as the projection of high frequency movements in financial assets onto differences in expectations. Our recovered series is intuitively reasonable and shows that information disclosure has a significant impact on the yields of short-term government bonds.

"Panel Data with Unknown Clusters"

Clustered standard errors and approximate randomization tests are popular inference methods that allow for dependence within observations. However, they require researchers to know the cluster structure ex ante. We propose a procedure to help researchers discover clusters in panel data. Our method is based on thresholding an estimated long-run variance-covariance matrix and requires the panel to be large in the time dimension, but imposes no lower bound on the number of units. We show that our procedure recovers the true clusters with high probability with no assumptions on the cluster structure. The estimated clusters are independently of interest, but they can also be used in the approximate randomization tests or with conventional cluster-robust covariance estimators. The resulting procedures control size and have good power.

"Some Finite Sample Properties of the Sign Test"

This paper contains two finite-sample results about the sign test. First, we show that the sign test is unbiased against two-sided alternatives even when observations are not identically distributed. Second, we provide simple theoretical counterexamples to show that correlation that is unaccounted for leads to size distortion and over-rejection. Our results have implication for practitioners, who are increasingly employing randomization tests for inference.

Languages English (native), Mandarin Chinese (fluent)

Programming R, Python, Matlab, Stata

References

Professor Ivan Canay
Department of Economics
Northwestern University
2211 Campus Drive
Evanston, IL 60208
(+1) 847-491-2929
iacanay@northwestern.edu

Professor Joel Horowitz
Department of Economics
Northwestern University
2211 Campus Drive
Evanston, IL 60208
(+1) 847-491-8253
joel-horowitz@northwestern.edu

Professor Eric Auerbach
Department of Economics
Northwestern University
2211 Campus Drive
Evanston, IL 60208
(+1) 847-491-4416
eric.auerbach@northwestern.edu

Updated October 2022.