MENGSI GAO

mengsi.gao@berkeley.edu
https://sites.google.com/view/mengsi-gao

BUSINESS ADDRESS
Department of Economics
530 Evans Hall, #3880
Berkeley, CA 94720-3880

DESIRED RESEARCH AND TEACHING FIELDS
Econometrics

FIELDS OF CONCENTRATION
Primary: Econometrics
Secondary: Industrial Organization

DISSERTATION TITLE: "Essays in Causal Inference and Network Econometrics"

Expected Date of Completion:	May 2025
Principal Advisor:	Bryan Graham and Peng Ding
Other References:	Michael Jansson and Demian Pouzo

PRE-DOCTORAL STUDIES	**DEGREE**	**DATE**	**FIELD**
Duke University	M.A.	2019	Economics
Zhejiang University	B.A.	2017	Economics

WORKING PAPERS
- "Endogenous Interference in Randomized Experiments" (**Job Market Paper**)
- "Causal Inference in Network Experiments: Regression-based Analysis and Design-based Properties," with Peng Ding. Submitted.
- "Identification and Inference on Treatment Effects under Covariate-Adaptive Randomization and Imperfect Compliance," with Federico Bugni, Filip Obradović, and Amilcar Velez. Submitted.
- "On the Power Properties of Inference for Parameters with Interval Identified Sets," with Federico Bugni, Filip Obradović, and Amilcar Velez. Submitted.

PUBLICATIONS
- "Inference under Covariate-Adaptive Randomization with Imperfect Compliance," with Federico Bugni, *Journal of Econometrics*, vol. 237 (1), 2023.
- "Uniform Nonparametric Inference for Time Series using Stata," with Jia Li and Zhipeng Liao, *Stata Journal*, Vol.20(3), 2020, pp. 706–721.

WORK IN PROGRESS
- "Misspecified Regressions with Mixed Regressors," with Peng Ding.

PROFESSIONAL EXPERIENCE

RESEARCH
Research Assistant, Department of Economics, U.C. Berkeley
- Professor Vira Semenova (May 2022 - Aug 2022, May 2023 - Aug 2023)
- Professor Demian Pouzo and Frederico Finan (May 2021 - Aug 2021)
- Professor Giovanni Compiani (May 2020 - Aug 2020)

Research Assistant, Department of Economics, Duke University
- Professor Federico Bugni (Feb 2018 - May 2019)
- Professor Giuseppe Lopomo (June 2018 - May 2019)

TEACHING
Tutor/Teaching Assistant, Department of Economics, U.C. Berkeley (Fall 2019 - Spring 2024)
Graduate Econometrics, Advanced Econometrics, Intermediate Econometrics
Teaching Assistant, Department of Economics, Duke University (Spring 2019)
Graduate Game Theory

SEMINAR AND CONFERENCES

2024	California Econometrics Conference
2024	Guanghua School of Management, Peking University
Fall 2023, Fall 2024	UC Berkeley Econometrics Seminar
Spring 2023, Fall 2023	Guest speaker at Stats 256 (Graduate causal inference, instructed by Peng Ding)

ACADEMIC SERVICE

JOURNAL REFEREE SERVICE

Journal of Econometrics, Journal of Business & Economic Statistics, Journal of the American Statistical Association, Journal of Applied Statistics, Journal of Causal Inference, Biometrics

SEMINAR COORDINATOR

UC Berkeley Econometrics Seminar Series (Spring 2022 - Fall 2023)

FELLOWSHIPS AND AWARDS

2024	U.C. Berkeley Dissertation Completion Fellowship
2019	Department of Economics Master's Program Award for Academic Excellence
2018	Duke University M.A. Merit Scholar
2017 - 2019	Duke University Economics Master's Scholar

OTHER INFORMATION

Languages:	English (fluent), Mandarin (native)
Citizenship:	China

SELECTED PAPER ABSTRACT

- "Endogenous Interference in Randomized Experiments" (**Job Market Paper**)

  Abstract: This paper investigates treatment effect identification and inference in randomized controlled trials with network interference. Two key network features characterize the setting and introduce endogeneity: (1) latent variables may affect both network formation and outcomes, and (2) the intervention may alter network structure, mediating treatment effects. I make three contributions. First, I define parameters within a post-treatment network framework, distinguishing direct effects of treatment from indirect effects mediated through changes in network structure. I provide a causal interpretation of the coefficients in a linear outcome model. For estimation and inference, I focus on a specific form of peer effects, captured by the fraction of treated friends, a special case of the linear-in-means model. Second, I establish the consistency and asymptotic normality of ordinary least squares estimators without unobserved confounders. Third, I address endogeneity using shift-share instrumental variables, demonstrating consistency and asymptotic normality of IV estimators in relatively sparse networks and proposing a modification for denser networks. Finally, I revisit Prina (2015) as an empirical illustration, demonstrating that treatment can influence outcomes both directly as well as through changes in network structure.

- "Causal Inference in Network Experiments: Regression-based Analysis and Design-based Properties," with Peng Ding.

  Abstract: Investigating interference or spillover effects among units is a central task in many social science problems. Network experiments are powerful tools for this task, which avoids endogeneity by randomly assigning treatments to units over networks. However, it is non-trivial to analyze network experiments properly without imposing strong modeling assumptions. Previously, many researchers have proposed sophisticated point estimators and standard errors for causal effects under network experiments. We further show that regression-based point estimators and standard errors can have strong theoretical guarantees if the regression functions and robust standard errors are carefully specified to accommodate the interference patterns under network experiments. We first recall a well-known result that the Hajek estimator is numerically identical to the coefficient from the weighted-least-squares fit based on the inverse probability of the exposure mapping. Moreover, we demonstrate that the regression-based approach offers three notable advantages: its ease of implementation, the ability to derive standard errors through the same weighted-least-squares fit, and the capacity to integrate covariates into the analysis, thereby enhancing estimation efficiency. Furthermore, we analyze the asymptotic bias of the regression-based network-robust standard errors. Recognizing that the covariance estimator can be anti-conservative, we propose an adjusted covariance estimator to improve the empirical coverage rates. Although we focus on regression-based point estimators and standard errors, our theory holds under the design-based framework, which assumes that the randomness comes solely from the design of network experiments and allows for arbitrary misspecification of the regression models.

- "Identification and Inference on Treatment Effects under Covariate-Adaptive Randomization and Imperfect Compliance," with Federico Bugni, Filip Obradović, and Amilcar Velez.

  Abstract: Randomized controlled trials (RCTs) frequently utilize covariate–adaptive randomization (CAR) (e.g., stratified block randomization) and commonly suffer from imperfect compliance. This paper studies the identification and inference for the average treatment effect (ATE) and the average treatment effect on the treated (ATT) in such RCTs with a binary treatment. We first develop characterizations of the identified sets for both estimands. Since data are generally not i.i.d. under CAR, these characterizations do not follow from existing results. We then provide consistent estimators of the identified sets and asymptotically valid confidence intervals for the parameters. Our asymptotic analysis leads to concrete practical recommendations regarding how to estimate the treatment assignment probabilities that enter in estimated bounds. In the case of the ATE, using sample analog assignment frequencies is more efficient than using the true assignment probabilities. On the contrary, using the true assignment probabilities is preferable for the ATT.

- "On the Power Properties of Inference for Parameters with Interval Identified Sets," with Federico Bugni, Filip Obradović, and Amilcar Velez.

  Abstract: This paper studies the power properties of confidence intervals (CIs) for a partially-identified parameter of interest with an interval identified set. We assume the researcher has bounds estimators to construct the CIs proposed by Stoye (2009), referred to as CI_α^1, CI_α^2, and CI_α^3. We also assume that these estimators are "ordered": the lower bound estimator is less than or equal to the upper bound estimator. Under these conditions, we establish two results. First, we show that CI_α^1 and CI_α^2 are equally statistically powerful, and both dominate CI_α^3. Second, we considera favorable situation in which there are two possible bounds estimators to construct these CIs, and one is more efficient than the other. One would expect that the more efficient bounds estimator yields more powerful inference. We prove that this desirable result holds for CI_α^1 and CI_α^2, but not necessarily for CI_α^3.

- "Misspecified Regressions with Mixed Regressors," with Peng Ding.

  Abstract: For analytic convenience, existing statistical theories either assume random or fixed regressors. Consequently, they do not cover the practical case of estimating the average treatment effect in experiments with randomized treatments and non-randomized, fixed pretreatment covariates. To fill the gap, we develop the theory for regressions with mixed regressors that contain both random and non-random, fixed components. Importantly, our theory allows for misspecification of the regression functions. We start with the canonical least-squares regression, discussing the interpretation of the regression coefficients and the estimation of the standard errors. We then develop the theory for estimating equations, which covers the canonical example of the two-stage least-squares estimation. We start with the theory for independent data and also extend the discussion to clustered data.