

YIFEI DING

Department of Economics
 University of California at Riverside
 900 University Avenue
 Riverside CA 92521, USA

Phone: (951) 231-5175
 Email: yding067@ucr.edu
 Website: <https://yifeiding-ucr.github.io/>

Placement Director

Associate Professor Joseph Cummins
 Phone: (951) 827-1582
 Email: joseph.cummins@ucr.edu

Placement Coordinator

Gary Kuzas
 Phone: (951) 827-1474
 Email: gary.kuzas@ucr.edu

Education

Department of Economics, University of California, Riverside	<i>Expected 06/2025</i>
Ph.D. in Economics	
Wang Yanan Institute for Studies in Economics, Xiamen University	<i>06/2018</i>
M.A. in Applied Statistics	
Wang Yanan Institute for Studies in Economics, Xiamen University	<i>06/2015</i>
B.A. in Mathematic Finance	
College of Chemistry and Chemical Engineering, Xiamen University	<i>06/2015</i>
B.S. in Chemistry	

Fields of Interest

Causal Inference, Econometric Theory, Machine Learning (Prediction and Causal Inference), Applied Econometrics

Working Papers

- **Deep Learning for Individual Heterogeneity with Generated Regressors by Adversarial Training** (with Ruoyao Shi)

[**Abstract**] We propose a semiparametric framework that combines machine learning with generated regressors via control function to capture individual heterogeneity while addressing endogeneity and sample selection bias in complex econometric models. This approach models individual heterogeneity through high-dimensional observable characteristics, with generated regressors supporting the control function to manage endogeneity or sample selection bias flexibly across various economic structures. Leveraging a tailored deep learning architecture, our framework integrates control functions and parameter functions seamlessly, enabling its adaptation to diverse econometric models. Using adversarial

training, we achieve sup-norm convergence rates of parameter functions and control function at the optimal min-max rate, which enhances robustness and yields valid inferences for inferential structural parameters in high-dimensional settings. Extending the Double Machine Learning (DML) approach, we incorporate endogenous components and establish a new influence function that directly includes generated regressors, broadening the framework's applicability across econometric models. With automatic differentiation in PyTorch, the influence function applies directly to data, streamlining inference and supporting various structural parameters without additional calculations. This integration makes the framework particularly useful in applied settings where individual heterogeneity and endogeneity are critical, such as personalized policy-making, targeted economic interventions, and customized optimizations in technology. Our simulations demonstrate superior performance, validating this framework's practical use in econometric analysis where heterogeneity and endogeneity are key considerations.

• Estimating Partial Effects Using Machine Learning

[**Abstract**] In this paper, we explore the use of machine learning techniques for estimating partial derivatives, which is a critical step towards understanding causal relationships in econometric analysis. By leveraging modern machine learning methods, such as tree-based models and deep neural networks, we assess their effectiveness in recovering regression functions and estimating partial derivatives. We introduce a novel tree-based model, Boosting Smooth Transition Regression Trees (BooST), and compare its performance with other models, including Boosting of Symmetric Smooth Additive Regression Trees (SMARTboost) and deep neural networks (DNNs). Simulations, based on the Friedman data generating process (DGP), demonstrate the superiority of BooST in estimating partial effects across various signal-to-noise environments and in the presence of redundant variables. The empirical applications, including the study of Engel curves, further highlight the ability of BooST to outperform other machine learning models in accurately estimating partial derivatives. Our findings suggest that BooST provides a powerful tool for nonparametric regression and causal inference, especially in econometric contexts where accurate estimation of marginal effects is crucial.

• A Comparative Study of Machine Learning Models for Prediction: Insights from Tree-Based Models and Deep Neural Networks

[**Abstract**] The growing influence of machine learning (ML) and big data technologies has significantly reshaped many scientific disciplines, including econometrics. This paper conducts a detailed comparative analysis of various tree-based and deep learning models, focusing on their prediction capabilities. The models examined include neural networks (e.g., MLP, ResNet), and several advanced tree-based models (e.g., Boost-Smooth, SMARTboost and Random Forest). Additionally, we explore different prediction combination techniques to evaluate whether combining predictions from multiple models enhances predictive accuracy. Using simulations from the comprehensive data generating processes (DGP), we systematically compare the performance of these models under varying levels of noise and

the presence of irrelevant features. Our findings reveal that tree-based models like SMARTboost and BooST demonstrate robust performance, particularly in low signal-to-noise scenarios, where they often outperform neural networks. Moreover, the inclusion of ensemble methods, such as median and simple average combinations, further improves prediction stability. Two real-world economic applications—Engel curve prediction and stock price crash risk prediction—highlight the practical implications of our analysis, showing the advantages of tree-based methods in capturing both linear and nonlinear data structures, while DNNs struggle in noisy and nonlinear environments. Our study emphasizes the need for careful model selection and the potential benefits of hybridizing prediction models for complex data tasks.

Publications & Conference Papers

1. Comparing Methods for Continuous Treatment (with Meng Xu), CODE@MIT, 2023.

[**Abstract**] This paper presents a comparative study of two advanced methodologies for estimating the effects of continuous treatments on outcome variables in large-scale tech applications. We focus on dose-response curves and marginal effects to address various business scenarios, such as the impact of ad frequency on user conversions, geolocation campaigns on local engagement, and latency on app performance. Our investigation centers around two promising approaches: entropy balancing for continuous treatment and double/debiased machine learning (DML). Using semi-synthetic data based on Snapchat user behavior, we evaluate these methods’ performance in terms of scalability, flexibility, and precision in handling high-dimensional, non-linear relationships between outcome variables, continuous treatments, and confounders. The study finds that tree-based machine learning models, particularly XGBOOST and Boostsmooth, outperform balancing approaches in estimating dose-response curves, while the balancing method performs best for marginal effect estimation. Notably, our findings also challenge the efficacy of the kernel-based selection model in the double machine learning process, prompting a reconsideration of improving its utility in real-world applications.

Presented at CODE@MIT, 2023.

2. Does “Too-Connected” Network of Shareholders Exacerbate Crash Risk? (with Haiqiang Chen, Yang Chen and Muqing Song), *China Economic Quarterly* (经济学季刊), 2023(03): 1070-1087.

[**Abstract**] Using quarterly data from the top 10 largest shareholders of A-share stock markets from 2003 to 2018, we construct a network of influential shareholders. Our findings reveal that firms with more interconnected shareholders face higher crash risk, especially when dominated by financial institutional shareholders or those with higher shareholding ratios. In contrast, state ownership and robust corporate governance significantly mitigate this risk. Mechanism analysis shows that firms with higher network centrality tend to have a higher goodwill-to-market value ratio, a greater proportion of related-party transactions to total assets, and larger M&A premiums, yet exhibit lower corporate governance transparency. These results suggest that overly connected shareholder networks may encourage tunneling behavior, exacerbating the crash risk for listed companies.

Research & Professional Experience

- **Applied Research Scientist Intern, Snap Inc.** *06/2023 – 09/2023*
 - **Project:** Conducted research on causal inference for continuous treatment effects using Double Machine Learning (DML), focusing on dose-response curve estimation and marginal effects for potential applications of studying the impact of business metrics like ad frequency and app latency on user engagement.
 - **Key Methodologies:** Applied entropy balancing and DML techniques to both synthetic and semi-synthetic data derived from Snapchat user data (20 million observations across 100+ variables) to rigorously test model performance.
 - **Findings:** Demonstrated that tree-based models, particularly BooST, outperformed entropy balancing and DNNs in predicting DRC within this large-scale data, with BooST showing superior accuracy and interpretability in high-dimensional, nonlinear data scenarios, surpassing XGBoost.
 - **Outcome:** Presented findings at CODE@MIT 2023, demonstrating BooST's potential practical application for optimal ad exposure and latency thresholds to enhance user engagement.
- **Research Assistant, Xiamen University** *09/2016 – 08/2017*
 - **Project:** *Estimating and Testing Threshold Models with Time-Varying Thresholds* funded by National Science Foundation of China (71571152).
 - Developed research ideas on shareholder networks and stock market crash risk, built project databases using web crawlers (Rcurl and Rvest), and assisted in empirical analysis and paper writing.
- **Research Assistant, Xiamen University** *06/2016 – 08/2016*
 - Employed R and NetLogo to simulate the formation and evolution of shareholder networks and their impact on financial markets.

Journal Referee

Journal of Quantitative Economics

Teaching Experience

- **Instructor, University of California, Riverside** *09/2024*
 - Ph.D. Course: Statistics in Mini-Math Camp
- **Teaching Assistant, University of California, Riverside** *09/2020 – Present*
 - Ph.D. Courses:
 - Econometric Method I - Econ 205A: 6.99/7 (Fall 2021)

- Micro theory (Grader)

- Undergraduate Courses:

- Data Analysis in Economics - Econ 106: 6.8/7 (Fall 2023)
- Introductory Econometrics I - Econ 107: 6.3/7 (Fall 2020), 7/7 (Summer 2022), 6.9/7 (Spring 2023), 6.3/7 (Spring 2024)
- Statistics for Economics - Econ 101: 6.2/7 (Winter 2021), 6.2/7 (Spring 2021), 6.3/7 (Summer 2021), 6.8/7 (Winter 2023)
- Intermediate Microeconomic Theory I - Econ 104A: 6.3/7 (Winter 2022), 6.1/7 (Fall 2022)
- Introduction to Microeconomics - Econ 002: 6.1/7 (Spring 2022)

• **Teaching Assistant, Xiamen University**

09/2016 – 06/2017

- Ph.D. Course: Advanced Econometrics II (Best TA)

- Master Course: Time Series Analysis

Programming Skills

Languages: R (Extensive), Python, GoogleCloud, HPCC, SQL, L^AT_EX, Eviews, Stata

Conference & Seminar Presentations

- The Causal AI Conference, San Francisco *06/2024*
- The Conference on Digital Experimentation (CODE@MIT), MIT *11/2023*
- Econometrics Colloquium, UC Riverside *06/2023, 11/2023, 05/2024, 11/2024*
- Brown Bag Seminars, UC Riverside *05/2023, 11/2023, 11/2024*
- 2018 Workshop of Resource Security and Economic Sciences, Xiamen, China *05/2018*
- 1st Forum of China Econometrics, Xiamen, China *12/2017*

Awards & Honors

- Dissertation Year Fellowship, University of California, Riverside *2024*
- Conference Travel Grant, GSA, University of California, Riverside *2023*
- Dean’s Distinguished Fellowship, University of California, Riverside *09/2019 – Present*
- Graduate Fellowships, University of Xiamen *09/2015, 09/2017*
- Academic Excellence Fellowships, University of Xiamen *05/2015, 09/2013, 09/2012*
- “Yangliangli” Scholarship, University of Xiamen *04/2013*
- “Chongqing Ziguang” Scholarship, University of Xiamen *04/2012*

References

Ruoyao Shi (Advisor)

Assistant Professor

Department of Economics

University of California, Riverside

Email: ruoyao.shi@ucr.edu

Phone: (951) 827-1494

Weixin Yao

Professor and Vice Chair

Department of Statistics

University of California, Riverside

Email: weixin.yao@ucr.edu

Phone: (951) 827-6007

Veronica T Sovero

Teaching Assistant Professor

Department of Economics

University of California, Riverside

Email: veronica.sovero@ucr.edu

Phone: (951) 827-0224

Tae-Hwy Lee

Professor

Department of Economics

University of California, Riverside

Email: taelee@ucr.edu

Phone: (951) 827-1509

Meng Xu

Lead Applied Research Scientist

Snap Inc.

Email: mxu@snap.com

Phone: (310) 854-9802