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 *Expected 03/2025-06/2025*
Ph.D. in Economics

Wang Yanan Institute for Studies in Economics, Xiamen University *06/2018*
Master of Arts in Applied Statistics

Wang Yanan Institute for Studies in Economics, Xiamen University *06/2015*
Bachelor of Arts in Mathematic Finance

College of Chemistry and Chemical Engineering, Xiamen University *06/2015*
Bachelor of Science 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)

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.

- **Key Contributions**: 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.

- **Methodology**: 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.

- **Impact**: This framework bridges advanced machine learning techniques with econometric models, which can be broadly applied to various structure parameters studying their individual heterogeneities.

• Estimating Partial Effects Using Machine Learning

This work investigates causal inference by estimating partial derivative using machine learning models as nonparametric alternatives for recovering regression functions. By introducing BoostSmooth, I leverage its smooth tree structure, which allows for analytic derivation of partial derivatives—providing clear, interpretable insights into the relationships between covariates and outcomes. This capability is particularly valuable in policy contexts where understanding marginal effects, such as the impact of training programs on employment, directly influences resource allocation decisions.

- **Key Contributions**: Through simulations and empirical applications, we showcase the BoostSmooth model's advantages in estimating partial derivatives compared to other machine learning methods across various signal-to-noise ratios and redundant variable settings.

- **Findings**: BoostSmooth effectively recovers nonlinear function structures and performs well in identifying linear function structures, making it versatile in various modeling scenarios.

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

This paper conducts a systematic comparison of prediction models across different machine learning techniques, including tree-based and deep neural network models. The research explores the effectiveness of various prediction combination methods.

- **Key Models**: Tree-based models include Boosting smooth transition regression trees, Boosting symmetric smooth additive regression trees, traditional boosting regression trees, and random forests. Deep neural network models include multilayer perceptrons and deep residual learning.

- **Combination Methods**: We categorize prediction combination methods into three main types: 1. Simple Methods: Simple average, median, Bates and Granger (1964), Newbold and Granger (1974) and inverse rank combination. 2. Regression-based Methods: OLS and constrained least squares regression combinations. 3. Eigenvector Methods: Standard eigenvector combination methods are also evaluated.

- **Findings**: Boosting-based models, especially the newly developed BoostSmooth and SMARTBoost models, perform superiorly in prediction accuracy compared to other tree-based and deep learning models. In terms of prediction combinations, the regression-based prediction combination methods and median combination robustly predominate other prediction combination methods in our simulations and empirical applications.

Research & Professional Experience

• Applied 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 business metrics like ad frequency and app latency.
- **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 practical application for optimal ad exposure and latency thresholds to enhance user engagement

• Master Thesis, Xiamen University *09/2017 –06/2018*

Title: *Network of Shareholders and Stock Price Synchronicity: Empirical Evidence from the A-share Market*

- Developed a social network model for the top 10 shareholders to measure network topology variables (centrality, validity, and length).
- Demonstrated a significant positive causal relationship between shareholder networks and stock price synchronicity using spatial panel data and network models.

• 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.

Publications & Conference Papers

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

Journal Referee

Journal of Quantitative Economics

Teaching Experience

• **Instructor, University of California, Riverside** *09/2024*
Course: Statistics in Mini-Math Camp

• **Teaching Assistant, University of California, Riverside** *09/2020 –Present*
- Graduate Courses: Econometric Method I, Micro theory (Grader)
- Undergraduate Courses: Data Analysis in Economics, Introductory Econometrics I, Statistics for Economics, Intermediate Microeconomic Theory I, Introduction to Microeconomics

• **Teaching Assistant, Xiamen University** *09/2016 –06/2017*
- Graduate Courses: Advanced Econometrics II, Time Series Analysis
- **Award**: Best TA

Programming Skills

Languages: R (Extensive), Python, GoogleCloud, HPCC, SQL, LaTeX, Eviews, Stata

Conference & Seminar Presentations

• Econometrics Colloquium, UC Riverside *11/2024*
• Brown Bag Seminars, UC Riverside *11/2024*
• The Causal AI Conference, San Francisco *06/2024*
• Econometrics Colloquium, UC Riverside *05/2024*
• Econometrics Colloquium, UC Riverside *11/2023*
• Brown Bag Seminars, UC Riverside *11/2023*
• The Conference on Digital Experimentation (CODE@MIT), MIT *11/2023*
• Econometrics Colloquium, UC Riverside *06/2023*
• Brown Bag Seminars, UC Riverside *05/2023*
• 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 of Economics, University of California, Riverside
Email: ruoyao.shi@ucr.edu
Phone: (951) 827-1494

Tae-Hwy Lee
Professor of Economics, University of California, Riverside
Email: taelee@ucr.edu
Phone: (951) 827-1509

Weixin Yao
Professor and Vice Chair of Statistics, University of California, Riverside
Email: weixin.yao@ucr.edu
Phone: (951) 827-6007

Veronica T Sovero
Teaching Assistant Professor of Economics, University of California, Riverside
Email: veronica.sovero@ucr.edu
Phone: (951) 827-0224