Zhouyu Shen

shenzhouyu@gsm.pku.edu.cn
Guanghua School of Management, Peking University, Beijing, China

EMPLOYMENT

Guanghua School of Management, Peking University

- Assistant Professor of Econometrics and Statistics, 2025-

EDUCATION

University of Chicago (UCHICAGO)— *Chicago, IL, US*
Ph.D. in Econometrics and Statistics, Booth School of Business
Advisor: Dacheng Xiu
September 2020 – June 2025
University of Science and Technology of China (USTC) — *Hefei, Anhui, China*
B.Sc. in Statistics, School of the Gifted Young
Advisor: Yu Chen and Zhengjun Zhang
September 2016 – June 2020

RESEARCH INTERESTS

Machine Learning, High Dimensional Statistics, Time-series Econometrics, Financial Econometrics

DISSERTATION COMMITTEE

Dacheng Xiu (Chair)
Booth School of Business, UChicago
Email: dacheng.xiu@chicagobooth.edu
Phone: (773) 834-7191

Christian Hansen
Booth School of Business, UChicago
Email: chansen1@chicagobooth.edu
Phone: (773) 834-1702

Tengyuan Liang
Booth School of Business, UChicago
Email: tengyuan.liang@chicagobooth.edu
Phone: (773) 702-7935

Chao Gao
Department of Statistics, UChicago
Email: chaogao@uchicago.edu
Phone: (773) 834-4178

WORKING PAPERS

Deep Autoencoders for Nonlinear Factor Models: Theory and Applications
Submitted — with Dacheng Xiu.
Abstract: Autoencoders are neural networks widely used in unsupervised learning for dimensionality reduction and feature extraction. This paper provides non-asymptotic guarantees for deep autoencoders within a nonlinear factor model, showing they can effectively extract latent components with errors that diminish with increasing dimensionality and sample size. The extracted factors converge to the true latent factors, up to a functional transformation. We extend these results to supervised autoencoders, supporting their use in factor-augmented prediction and structured matrix completion. Finally, we illustrate the practical value of autoencoders in macroeconomic forecasting, asset return prediction, and noise reduction for causal analysis.

Can Machines Learn Weak Signals?
Submitted — with Dacheng Xiu.
Winner of the 2024 Bates-White Prize for Best Paper at SoFiE Annual Conference.
Abstract: In high-dimensional regression scenarios with low signal-to-noise ratios, we assess the predictive performance of several prevalent machine learning algorithms. Theoretical insights show Ridge regression's superiority in exploiting weak signals, surpassing a zero benchmark. In contrast, Lasso fails to exceed this baseline, indicating its learning limitations. Simulations reveal that Random Forest generally outperforms Gradient Boosted Regression Trees when signals are weak. Moreover, Neural Networks with ℓ_2-regularization excel in capturing nonlinear functions of weak signals. Our empirical analysis across six economic datasets suggests that the weakness of signals, not necessarily the absence of sparsity, may be Lasso's major limitation in economic predictions.

Boosting or Bagging? Navigating Weak Signals
Work in Progress — with Dacheng Xiu and Mingjie Zhang.
Abstract: This study investigates the predictive performance of Bagging and Boosting, two widely used ensemble techniques, in high-dimensional settings characterized by weak signals. Using a novel theoretical framework, we evaluate and compare their effectiveness. Our analysis demonstrates that Bagging excels at capturing weak signals, consistently outperforming the naive benchmark predictor. In contrast, Boosting often underperforms in these environments. Building on these insights, we propose modifications to the traditional random forest algorithm, enhancing its effectiveness for prediction tasks involving weak signals.

Recurrent Neural Networks Meet Time Series: A Theoretical Perspective
Work in Progress — with Xiao Chen, Yu Chen, and Dacheng Xiu.
Abstract: Recurrent Neural Networks (RNNs) are a class of artificial neural networks designed to model sequential data, including text, speech, and time series. This paper investigates the internal mechanisms of RNNs and provides theoretical guarantees within the framework of time series models. Specifically, we analyze a nonlinear autoregressive and moving-average model and derive a statistical error bound for RNN prediction performance. Our analysis demonstrates that RNNs excel at capturing nonlinear dependencies in noise, offering a distinct advantage over traditional nonparametric methods such as feed-forward neural networks.

PUBLICATIONS

Modeling Tail Index with Autoregressive Conditional Pareto Model
Journal of Business & Economic Statistics, Volume 40 (2022) — with Yu Chen and Ruxin Shi.
Abstract: We propose an autoregressive conditional Pareto (AcP) model based on the dynamic peaks over threshold method to model a dynamic tail index in the financial markets. Unlike the score-based approach which is widely used in many articles, we use an exponential function to model the tail index process for its intuitiveness and interpretability. Probabilistic properties of the AcP model and the statistical properties of its parameter estimators of maximum likelihood are studied in this article. Real data are used to show the advantages of AcP, especially, compared to the estimation volatility of GARCH model, the result of AcP is more sensitive to turmoil. The estimated tail index of AcP can accurately reflect the risk of the stock and may even play an early warning role to the turmoil of stock market. We also calculate the tail connectedness based on the estimated tail index of AcP and show that tail connectedness increases during period of turmoil, which is consistent with the result of the score-based approach.

TEACHING EXPERIENCE

Peking University, Guanghua School of Management

- Machine Learning Theory, Ph.D. Course — 2026 Spring

PROFESSIONAL SERVICE

Journal Reviewer for *Journal of Econometrics*, *Journal of the American Statistical Association*, and *Pacific-Basin Finance Journal*.

PRESENTATIONS

- 2023: Statistical Foundations of Data Science and their Applications conference, Princeton University
- 2024: Asian Meeting of the Econometric Society, Zhejiang University; Second Joint Conference on Statistics and Data Science in China, Yunnan University; Peking University; Chinese University of Hongkong
- 2025: City University of Hongkong; PBC School of Finance Tsinghua University; Chinese University of Hongkong; Chinese University of Hongkong-Shenzhen;