Zhouyu Shen
zshen10@chicagobooth.edu
Booth School of Business, University of Chicago, Chicago, US

EDUCATION

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

- **National Scholarship** (2018, top 1%)
- **National Scholarship** (2019, top 1%)
- **First-Class Scholarship**, School of the Gifted Young (2017, top 5%)

RESEARCH INTERESTS

Machine Learning Theory, Factor Analysis, Asset Pricing, High-Dimensional Statistics, Nonparametric Regression, Time Series

RESEARCH

On the Theory of RNNs
Working Paper — Joint with Xiao Chen, Yu Chen, and Dacheng Xiu.
Recurrent Neural Networks (RNNs) represent a class of artificial neural networks specifically designed to model sequential data, such as text, speech, and time series. This paper investigates the internal mechanisms of RNNs by providing theoretical guarantees within the framework of time series models. We analyze a nonlinear autoregressive and moving-average model (NARMA) and establish a statistical error bound for the prediction error of RNNs. The bound comprises two components: the approximation error, influenced by the smoothness of the target function and the dimensionality of the input, and the estimation error, governed by the architecture of the RNN and the mixing properties of the underlying process. Subject to the mixing properties of the underlying model, this bound is capable of approaching the minimax optimal rate for independent data. Accordingly, in comparison to traditional nonparametric regression methods, RNNs are advantageous due to their ability to capture nonlinear dependencies within the noise effectively. The results from our simulation study demonstrate that RNNs outperform both Autoregressive Moving Average (ARMA) models and Deep Neural Networks (DNNs) in predictive performance.

On the Theory of Deep Autoencoders
Working Paper — Joint with Dacheng Xiu.
Autoencoders are pivotal in unsupervised machine learning, widely employed for dimension reduction, feature learning, and signal denoising. This study provides non-asymptotic guarantees for deep autoencoders within a nonlinear factor model framework. We demonstrate that deep autoencoders can effectively retrieve common components from model inputs. The associated error comprises a component that diminishes with increasing dimensionality—akin to the 'blessings of dimensionality' observed in linear factor models—and another component that vanishes with an increasing sample size at the optimal nonparametric regression rate, as if the factors were directly observed. Furthermore, we show that

the extracted factors converge to the true latent factors, albeit through a functional transformation. We conclude by showcasing three economic applications of autoencoders: nowcasting GDP growth, pricing the cross-section of asset returns, and program evaluation.

Can Machines Learn Weak Signals?
Submitted — Joint with Dacheng Xiu.
Winner of the 2024 Bates-White Prize for Best Paper at SoFiE Annual Conference.
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.

Modeling Tail Index with Autoregressive Conditional Pareto Model
Published in Journal of Business & Economic Statistics, Volume 40 (2022) — Joint with Yu Chen and Ruxin Shi.
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

University of Chicago, Booth School of Business
Teaching Assistant

- Statistical Inference, Ph.D. Course — 2024
- Statistical Inference, Ph.D. Course — 2023

University of Science and Technology of China
Teaching Assistant

- Time Series Analysis, Undergraduate Course — 2020
- Stochastic Process, Undergraduate Course — 2019
- Probability Theory, Undergraduate Course — 2019

PROFESSIONAL SERVICE

Journal Reviewer for *Journal of Econometrics* and *Journal of the American Statistical Association.*

Presentations

- Poster presentation at the *Statistical Foundations of Data Science and their Applications* conference, Princeton University, USA, May 2023.
- Presented the paper *Can Machines Learn Weak Signals?* at the *Asian Meeting of the Econometric Society*, Zhejiang University, China, June 2024.
- Presented the paper *Can Machines Learn Weak Signals?* at the *Second JCSDS 2024*, Yunnan University, China, July 2024.