Ruoqing Jiang

Department of Economics
Columbia University
New York, NY 10027

Email: rj2556@columbia.edu
Phone: (+1) 650-475-6528

Placement Committee

Chairs: Sandra Black (sblack@columbia.edu), Mark Dean (mark.dean@columbia.edu)
Administrators: Amy Devine (aed2152@columbia.edu), Rosa Lourentzatos (rcl2109@columbia.edu)

Education

Columbia University — New York, United States
Ph.D. in Economics — 2025 (expected)
M.Phil. in Economics — 2022
M.A. in Economics — 2021

Stanford University — California, United States
M.S. in Statistics — 2018

Tsinghua University — Beijing, China
B.A. in Economics — 2015

Fields of Specialization

Primary: Econometrics
Secondary: Empirical Finance, Empirical Macroeconomics, Statistical Machine Learning

Job Market Paper

Cauchy-Weighted Measure for Nonlinear Conditional Dependencies within the High-Dimensional Factor Model Framework

Conditional independence testing is a fundamental and challenging problem with many important applications. This paper proposes a new approach to test conditional independence within the framework of high-dimensional factor models. A Cauchy-weighted measure is introduced to quantify the dependence between the idiosyncratic components and develop a corresponding conditional independence test. This measure, which ranges from 0 to 1, equals 0 if and only if conditional independence is true. It is robust to extreme values and computationally efficient. The proposed test is asymptotically distribution-free under the null hypothesis and capable of detecting nonlinear dependencies under the alternative hypothesis in high-dimensional scenarios. Furthermore, this paper demonstrates that the approach of first directly estimating the factors using pooled data, followed by performing the test, is invalid when factors are unobserved. Instead, this paper proposes to estimate the factors within the factor-augmented regression model framework and shows that the corresponding test remains valid. Extensive simulation studies and real data analysis are conducted to validate the effectiveness of this method, demonstrating its superior performance in high-dimensional scenarios.

Working Papers

Quasi-Maximum Likelihood Estimation of Weak Factor Models

This paper investigates the quasi-maximum likelihood estimation (QMLE) of high-dimensional weak factor models. Building on prior work that established convergence of maximum likelihood estimates for factors and loadings under strong factor structure assumptions and specific identification conditions, this study extends the theory to cover cases with weaker loadings. Standard inference procedures remain applicable, except for cases involving extremely weak loadings, which yield promising implications for empirical studies. It is shown that the estimates retain consistency and asymptotic normality over a broader spectrum of weak loadings, albeit at slower convergence rates and under stricter sample size requirements. This extended inferential framework addresses not only consistency but also the rate of convergence and limiting distributions. The QMLE approach allows for heteroskedasticity, which is estimated alongside other parameters. Additionally, simulation results compare the efficiency of QMLE with that of the principal components method, providing insights into the relative performance of each approach.

Assessing factor model adequacy in high-dimensional data with cross-fitting

This study evaluates the adequacy of a basic, interpretable factor model that focuses solely on the common factor component to characterize the relationship between the covariate vector and the response variable, in comparison to a more complex factor-augmented sparse regression model. A two-fold cross-fitting method is introduced to enhance test power, eliminating the computational complexities associated with tuning parameters and bootstrapping. Under the null hypothesis, the two statistics derived from disjoint datasets are shown to be asymptotically independent. The combined test statistic follows a standard normal distribution, analogous to the correlation of two uncorrelated variables. Mild conditions are established under which the test achieves an asymptotic power of 1. Under the alternative hypothesis, the combined test statistic exhibits the oracle property, behaving as if the index with the largest absolute correlation value were known, thereby achieving greater power than tests based on maximum-type statistics, which typically require larger critical values. The efficacy of this approach is validated through extensive simulation studies and real data analysis.

Linear Programming Based Pure Exploration in Batched Multi-armed Bandit Problem (with Shengyu Cao, Simai He, Jin Xu and Hongsong Yuan)

Recently multi-armed bandit problem, especially its particular form pure exploration, arises in many real-world scenarios where arms must be sampled in batches due to the limited time the agent can wait for the feedback. Such applications include social media analysis, online marketing and crowdsourcing. The pure exploration problem in batched bandits is further complicated when the number of arms K is large (for instance, it can be exponential in the number of batches R, or even infinite). To deal with the crucial problem of controlling the total sampling cost while maintaining sufficient accuracy of identifying a good arm under these constraints, we introduce a general linear programming framework that can incorporate objectives of different theoretical settings in pure exploration. The linear program leads to a two-stage algorithm that can accelerate the process of selecting arms in its first stage, and can easily incorporate existing pure exploration methods in its second stage. We establish theoretical properties of the two-stage algorithm, and demonstrate that the algorithm also has good numerical performance compared to certain UCB-type or Thompson sampling methods. (https://arxiv.org/abs/2312.13875)

Work in Progress

Instrumental Variable Estimation with a Large Number of Weak Instruments

This paper proposes a novel approach for instrumental variable (IV) estimation using scaled principal component analysis (sPCA). Previous research has demonstrated that latent factors extracted from observed variables closely related to endogenous regressors can effectively construct IVs. However, when a large number of weak IVs are present, traditional principal component analysis (PCA) often underperforms. The sPCA method improves on standard PCA by scaling each observed variable according to its slope with the endogenous variable, assigning greater weight to variables more strongly associated with the endogenous variable. Unlike PCA, which maximizes common variation across variables, sPCA prioritizes those with higher relevance to the endogenous variable. Within a general factor framework, sPCA is shown to outperform PCA in IV estimation given certain data conditions. A real data example further illustrates the improved performance of sPCA for IV estimation.

Teaching Experience

Teaching Assistant, Columbia University

Econometrics I (Graduate)	Fall 2020, 2022
Econometrics II (Graduate)	Spring 2022
Intermediate Macroeconomics (Undergraduate)	Spring, Fall 2021
Introduction to Econometrics (Undergraduate)	Spring 2023

Fellowships and Grants

Dissertation Fellowship, Columbia University	2023 - 2024
Program for Economic Research (PER) Research Summer Fellow	2022
Weatherhead East Asian Institute (WEAI) Research Grant	2021
Dean’s Fellowship, Columbia University	2019 - 2024

Personal

Languages: English (Fluent), Mandarin (Native), Shanghainese (Native)
Software skills: R, Matlab, Stata, Python, C++, Latex
Citizenship: China

References

Jushan Bai
Professor of Economics
Department of Economics
Columbia University
jb3064@columbia.edu

Sokbae Lee
Professor of Economics
Department of Economics
Columbia University
sl3841@columbia.edu

Zhiliang Ying
Professor of Statistics
Department of Statistics
Columbia University
zying@stat.columbia.edu