

Jeff Cai

356 Mendoza College of Business

Department of Information Technology, Analytics, and Operations

University of Notre Dame, Notre Dame, IN 46556

Email: jcai2@nd.edu

Website: <https://jh-cai.com>

RESEARCH INTERESTS Data-driven Decision-making, Network Analysis, Model Selection, Post-selection Inference, Empirical Bayes; Econometrics, Equity Holding Network, Ownership Structure, Healthcare, Revenue Management.

ACADEMIC POSITION *Scharf Family Assistant Professor in Real Estate* 2022 – current
Department of Information Technology, Analytics, and Operations
Mendoza College of Business, University of Notre Dame

EDUCATION **The Wharton School** 2017 – 2022
University of Pennsylvania
Ph.D. in Statistics (Advisor: Linda Zhao)

Thesis: Network, network regression, and equity holding network

Chinese University of Hong Kong 2013 – 2017
Bachelor of Science in Computer Science (Minor in Statistics), First Class Honors

Thesis: Variance-reduced stochastic optimization and acceleration techniques (with Jiacheng Zhuo supervised by James Cheng)

JOURNAL PUBLICATIONS

1. **Cai, J.**, Mandelbaum, A., Nagaraja, C. H., Shen, H., Zhao, L. (2019). “Statistical theory powering data science.” *Statistical Science*, 34(4), 669-691. Available [here](#).
2. Kuchibhotla, A. K., Brown, L. D., Buja, A., **Cai, J.**, George, E. I., Zhao, L. (2020). “Valid post-selection inference in model-free linear regression.” *Annals of Statistics*, 48(5), 2953-2981. Available [here](#).
3. Ogburn, E., **Cai, J.**, Kuchibhotla A., Berk, R., Buja A. (2022). “Practical issues concerning assumption-lean inference for generalized linear models. Discussion on the paper by Vansteelandt and Dukes.” *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 84(3), 715-716. Available [here](#).
4. **Cai, J.**, Yang, D., Zhu, W., Zhao, L. (2023). “Hierarchical vintage sparse PCA. Discussion on the paper by Rohe and Zeng.” *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 85(4), 1076-1080. Available [here](#).

CHAPTERS, ARTICLES, AND COLUMNS

1. Allen, F., **Cai, J.**, Gu, X., Qian, J., Zhao, L., Zhu, W. (2023). “Centralization or decentralization? The evolution of state-ownership in China.” *VoxChina*. Available at [here](#).
2. **Cai, J.**, Gu, X., Zhao, L., Zhu, W. (2023+). “State ownership in China: An equity network perspective.” *The Arc of Chinese Economy*. Brookings Institution Press.

NON-REFEREED
PUBLICATIONS

1. **Cai, J.**, Pashley, N., Zhao, L. (2021).
“2021 IMS Membership Survey: Report.” *IMS Bulletin*, 50(5), 6-9. Available [here](#).
2. **Cai, J.** (2015). “Project on glacier recognition with SIFT and CNN.” *The Arctic Explorers from S.H. Ho College, 2015*, pp. 182-187. Available [here](#).

PREPRINTS AND
WORKING PAPERS

1. **Cai, J.**, Chen, R., Wainwright, M., Zhao, L. (2024).
“Doubly high-dimensional contextual bandits: An interpretable model with applications to assortment and pricing.” *Reject & Resubmit at Management Science*. Available at <https://ssrn.com/abstract=4568525>.
2. **Cai, J.**, Yang, D., Zhu, W., Shen, H., Zhao, L. (2024).
“Network regression and supervised centrality estimation.” *Reject & Resubmit at Journal of the American Statistical Association*. Available at <https://arxiv.org/abs/2111.12921>.
3. Allen, F., **Cai, J.**, Gu, X., Qian, J., Zhao, L., Zhu, W. (2024).
“Ownership network and firm growth: What do forty million companies tell about the Chinese economy?”
AFA 2021, NBER Chinese Economy Meeting 2020, MFA 2020, FMA 2020, NSF 6th Annual Conference for Network Science, The Sixth Network Science in Economics Conference, CFRC 2021, CICF 2021. Under revision at *Management Science*. Available at SSRN: <https://ssrn.com/abstract=3465126>.
China Financial Research Conference (CFRC) 2021 Best Paper Award (3 out of 534 papers).
4. Allen, F., **Cai, J.**, Gu, X., Qian, J., Zhao, L., Zhu, W. (2024).
“Centralization or decentralization? The evolution of state-ownership in China.”
CICF 2021, MFA 2022, FIRS 2022. In submission to *Journal of Political Economy*. Available at SSRN: <https://ssrn.com/abstract=4283197>.
China International Conference in Finance (CICF) 2021 XiYue Best Paper Award (2 out of 2065 papers).
5. **Cai, J.**, Chen, R., Huang, Q., Wainwright, M., Zhao, L., Zhu, W. (2024).
“Optimal assortment and pricing via generalized MNL models with Poisson arrival.”
2024 ESIF Economics and AI+ML Meeting. In submission to *Operations Research*.
6. **Cai, J.**, Chen, R., Wainwright, M., Zhao, L. (2024).
“Personalized reinforcement learning: With applications to sepsis management in ICU.” *Draft available*.
7. Zhang, X., **Cai, J.**, Li, J., Abbasi, A. (2024).
“Towards a holistic representation of online customer journeys: A tensor-based framework.” *INFORMS Workshop on Data Science 2023*. *Draft available*.
8. **Cai, J.**, Gu, X., Zhao, L., Zhu, W. (2022).
“Microscopic dynamics of equity ownership networks in China.” *Draft available*.
9. **Cai, J.**, Han, X., Ritov, Y., Zhao, L. (2021).
“Nonparametric empirical Bayes estimation and testing for sparse and heteroscedastic signals.” Available at <https://arxiv.org/abs/2106.08881>.
10. Kuchibhotla A. K., Brown L. D., Buja A., **Cai, J.** (2019).
“All of linear regression.” Available at <https://arxiv.org/abs/1910.06386>.
11. **Cai, J.**, Kuchibhotla, A. K., Zhao, L. (2020).
“Valid post-selection inference for the average treatment effect with covariate adjustment in randomized experiments.” *Draft available*.

12. **Cai, J.**, Brown, L., Kuchibhotla A. K., Zhao, L. (2020).
“Generalized Cp and a predictive model selection test in assumption-lean framework.”
Draft available.
13. Kuchibhotla A. K., **Cai, J.** (2020).
“Computation of PoSI statistics.” *Draft available.*
14. **Cai, J.**, Cai, E., Harrison, A., Meyer, M., Zhao, L., Zhao, M. (2018).
“Self-reported Chinese company data: Can it be trusted?” *Draft available.*

SOFTWARES

1. **SNP**: R package (C++-based) on Spike-and-Nonparametric Prior (SNP) for Sparse and Heteroscedastic Signals. *Available at <https://github.com/cccfam/SNP>.*
2. **tmax**: R package (C++/FORTRAN-based) on assumption-lean Post-Selection Inference (PoSI). *Available at <https://post-selection-inference.github.io/R/>.*
3. **SuperCENT**: R package (C++-based) on Supervised Centrality Estimation and Network Prediction (SuperCENT). *Available at <https://jh-cai.com/SuperCENT>.*
4. **tmax-GPU**: GPU version of **tmax** (C++/CUDA-based) with Tan, X. Available soon.

INVITED TALKS

- Optimal Assortment and Pricing via Generalized MNL Models with Poisson Arrival.
 - *ESIF Economics and AI+ML Meeting, Ithaca, 2024.*
- Personalized Reinforcement Learning.
 - *Statistics Empowering Data Science (SEEDS) Conference, Los Angeles, 2023.*
 - *IMS International Conference on Statistics and Data Science (ICSIDS), Lisbon, 2023.*
 - *INFORMS Annual Meeting (INFORMS), Phoenix, 2023.*
- Doubly High-Dimensional Contextual Bandits: An Interpretable Model with Applications to Assortment and Pricing.
 - *Production and Operations Management Society Annual Meeting (POMS), Minneapolis, 2024.*
 - *The 15th International Conference of the ERCIM WG on Computational and Methodological Statistics (CMSTAT), Berlin, 2023.*
 - *Joint Statistical Meeting (JSM), Toronto, 2023.*
 - *Manufacturing and Service Operations Management Conference (MSOM), Montreal, 2023.*
- Centralization or Decentralization? The Evolution of State-Ownership in China.
 - *The Arc of Chinese Economy, University of Pennsylvania, Philadelphia, 2022.*
 - *China International Conference in Finance (CICF), Virtual, 2021.*
- Equity Ownership Network and Supervised Centrality Estimation
 - *IMS International Conference on Statistics and Data Science (ICSIDS), Florence, 2022.*
 - *The 14th International Conference of the ERCIM WG on Computational and Methodological Statistics (CMSTAT), London, 2022.*
- Network Regression and Supervised Centrality Estimation.
 - *Young Scholars Conference on Machine Learning in Economics and Finance, Federal Reserve Bank of Philadelphia, 2021.*
 - *Statistics department seminar, George Mason University (joint with Linda Zhao), 2021.*

- *The 14th International Conference of the ERCIM WG on Computational and Methodological Statistics (CMSTAT), London, 2021.*
- Semi-Supervised Centrality Estimation and Network Prediction.
Joint Statistical Meeting (JSM), Virtual, 2021.
- Ownership Network and Firm Growth: What do Forty Million Companies Tell about the Chinese Economy?
The Sixth Network Science in Economics Conference, Virtual, 2021.
- Nonparametric Empirical Bayes Estimation and Testing for Sparse and Heteroscedastic Signals.
The 12th International Conference of the ERCIM WG on Computational and Methodological Statistics (CMSTAT), London, 2019.
- Innovation and Equity Holding Network.
Workshop on Innovation and Intellectual Property: Firm Strategies and Policy Challenges in a Rapidly Changing World, Beijing, 2019.
- Nonparametric Empirical Bayes Methods for Sparse and Noisy Signals.
 - *Joint Statistical Meeting (JSM) Denver, 2019.*
 - *The Fourth Workshop on Higher-Order Asymptotics and Post-Selection Inference (WHOA-PSI-4) (represented Linda Zhao), Washington University of St. Louis, 2019.*
 - *The International Chinese Statistical Association Conference (ICSA), Chicago, 2017.*
- Generalized Cp and a Predictive Model Selection Test in Assumption-Learn Framework.
 - *The International Chinese Statistical Association Conference (ICSA), Naikai University, Tianjin, 2019.*
 - *Joint Statistical Meeting (JSM), Vancouver, 2018.*
 - *The Third Workshop on Higher-Order Asymptotics and Post-Selection Inference (WHOA-PSI-3) (joint with Linda Zhao), Washington University of St. Louis, 2018.*
 - *Joint Statistical Meeting (JSM), Baltimore, 2017.*
 - *Workshop on Chinese Firm-level Data, Beijing, 2017.*

AWARDS

- Murray Teaching Award 2021
- China Financial Research Conference (CFRC) Best Paper Award 2021
“Ownership Network and Firm Growth: What do Forty Million Companies Tell about the Chinese Economy?” (3 out of 534 papers).
- China International Conference in Finance (CICF) XiYue Best Paper Award 2021
“Centralization or Decentralization? The Evolution of State-Ownership in China.” (2 out of 2065 papers).
- Mack Institute Research Fellowship. Co-PI: **Cai, J.**, Zhao, L., Zhu, W. 2020
“Networks, Machine Learning, and Asset Pricing.”
- Penn China Research and Engagement Fund (CREF). 2019
PI: Harrison, A., Meyer, M., Zhao, L., Zhao, M.
- George James Term Fund Award 2019
- Wharton Global Initiative Fund 2017 – Present
- National Science Foundation Grant DMS-1512084. PI: Brown, L., Zhao, L. 2018

SERVICE

- Reviewer for *Annals of Statistics*, *Journal of the American Statistical Association*, *Annals of Applied Statistics*, *Observational Studies*.
- Faculty mentor, Building Bridges Mentoring Program, University of Notre Dame, 2022.
- Organizing committee for Workshop in Honor of Ed George's 70th Birthday: Perspectives in Statistical Modeling and Inference, Philadelphia, 2021. <http://jh-cai.com/edgeorge70>
- Ph.D. Student Mentor, Wharton Doctoral Council, 2020.
- Co-organizer of Workshop on Innovation and Intellectual Property: Firm Strategies and Policy Challenges in a Rapidly Changing World, Beijing, 2019. <http://jh-cai.com/pwcc2019>
- Co-organizer of Lawrence D. Brown Memorial Service and Workshop, Philadelphia, 2018. <http://jh-cai.com/larrybrown2018>
- Organizing committee for Workshop on Chinese Firm-level Data, Beijing, 2017. <http://jh-cai.com/pwcc2017>
- Co-organizer of Data Science Live. Spring 2019, Fall 2019, Spring 2021, Summer 2021. <https://jh-cai.com/modern-data-mining/dsl>

TEACHING

- Co-director for 2022-2024 Wharton Data Science Academy of Wharton Youth Program
- Ph.D. Mentor, Wharton Directed Reading Program, Fall 2021.
- Lead teaching fellow for 2021 Wharton Data Science Academy of Wharton Youth Program
- Co-developed [STAT 471/571/701 Modern Data Mining](#) (cross-listed for undergraduates, graduates and MBAs) and served as head TA Spring 2021 and TA Fall 2017
- Wharton Executive Education OneConnect Program with Two Sigma, Summer 2019

SKILLS

- Programming: C/C++, R, Python, L^AT_EX (proficient); FORTRAN, Matlab, Java (experienced)
- Database management: SQL (PostgreSQL), graph database (Neo4j)
- Web design and development: Javascript (AngularJS), HTML5/CSS, Django; Photoshop
- Server administration: Unix