

Jeff Cai

434B.1 Academic Research Building
265 South 37th Street
Philadelphia, PA 19104

Phone: +1 (484)-888-3894
Email: junhui@wharton.upenn.edu
Website: <https://jh-cai.com>

RESEARCH INTERESTS	Network Analysis, Model Selection, Post-selection Inference, Nonparametric Empirical Bayes; Econometrics, Equity Holding Network, Ownership Structure.
EDUCATION	<p>The Wharton School September 2017 – May 2022 (expected) University of Pennsylvania, Philadelphia, USA <i>Doctoral Candidate in Statistics (Advisor: Linda Zhao)</i></p> <p>Chinese University of Hong Kong September 2013 – May 2017 Hong Kong, PRC <i>Bachelor of Science in Computer Science (Minor in Statistics), First Class Honors</i></p> <p>Thesis: Variance-reduced stochastic optimization and acceleration techniques (with Jiacheng Zhuo supervised by James Cheng)</p>
JOURNAL PUBLICATIONS	<ol style="list-style-type: none">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.” <i>Annals of Statistics</i>, 48(5), 2953-2981. Available here.Cai, J., Mandelbaum, A., Nagaraja, C. H., Shen, H., Zhao, L. (2019). “Statistical theory powering data science.” <i>Statistical Science</i>, 34(4), 669-691. Available here.Ogburn, E., Cai, J., Kuchibhotla A., Berk, R., Buja A. (2021+). “Practical issues concerning assumption-lean inference for generalized linear models. Discussion on the paper by Vansteelandt and Dukes.” <i>Journal of the Royal Statistical Society. Series B: Statistical Methodology (To appear)</i>. Available here.
NON-REFEREED PUBLICATIONS	<ol style="list-style-type: none">Cai, J., Pashley, N., Zhao, L. (2021). “2021 IMS Membership Survey: Report.” <i>IMS Bulletin</i>, 50(5), 6-9. Available here.Cai, J. (2015). “Project on glacier recognition with SIFT and CNN.” <i>The Arctic Explorers from S.H. Ho College, 2015</i>, pp. 182-187. Available here.
PREPRINTS AND WORKING PAPERS	<ol style="list-style-type: none">Cai, J., Yang, D., Zhu, W., Shen, H., Zhao, L. (2021). “Network regression and supervised centrality estimation.” <i>In submission to Review of Economic Studies</i>. Available at https://arxiv.org/abs/2111.12921.Allen, F., Cai, J., Gu, X., Qian, J., Zhao, L., Zhu, W. (2021). “Ownership network and firm growth: What do forty million companies tell about the Chinese economy?” <i>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</i>. In submission to <i>Quarterly Journal of Economics</i>. Available at SSRN: https://ssrn.com/abstract=3465126. China Financial Research Conference (CFRC) 2021 Best Paper Award (3 out of 534 papers).

3. **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>. Under review.
4. Allen, F., **Cai, J.**, Gu, X., Qian, J., Zhao, L., Zhu, W. (2020).
“Centralization or decentralization? The evolution of state-ownership in China.” Available at *CICF 2021*.
China International Conference in Finance (CICF) 2021 XiYue Best Paper Award (2 out of 2065 papers).
5. **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*.
6. **Cai, J.**, Brown, L., Kuchibhotla A. K., Zhao, L. (2020).
“Generalized Cp and a predictive model selection test in assumption-lean framework.” *Draft available*.
7. Kuchibhotla A. K., **Cai, J.** (2020).
“Computation of PoSI statistics.” *Draft available*.
8. **Cai, J.**, Gu, X., Zhao, L., Zhu, W. (2020).
“Microscopic dynamics of equity ownership networks in China.” *Draft available*.
9. Kuchibhotla A. K., Brown L. D., Buja A., **Cai, J.** (2019).
“All of linear regression.” Available at <https://arxiv.org/abs/1910.06386>.
10. **Cai, J.**, Cai, E., Harrison, A., Meyer, M., Zhao, L., Zhao, M. (2018).
“Self-reported Chinese company data: Can it be trusted?” *Draft available*.
11. **Cai, J.**, Zhu, W., Zhao, L. (2021).
“Common versus idiosyncratic risk.” *Preliminary draft*.

SOFTWARES

1. **SNP**: R package (C++-based) on Spike-and-Nonparametric Prior (SNP) for Sparse and Heteroscedastic Signals. Available at <https://github.com/cccfan/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

- 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.
- Centralization or Decentralization? The Evolution of State-Ownership in China.
China International Conference in Finance (CICF), 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-Lean 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

- 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

- Ph.D. Mentor, Wharton Directed Reading Program, Fall 2021.
- Teaching fellow for 2021 Wharton Data Science Academy of Wharton Youth Program
 - Launched the program; co-developed and taught the course
 - Led 5 teaching fellows for 92 students in 18 groups
- 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
 - Syllabus
 - Developed new modules on advanced R tutorials, Exploratory Data Analysis, Principle Component Analysis, Neural Network, and Convolutional Neural Network
 - Restructured lecture notes with up-to-date case studies in R-markdown
 - Lead a project to develop a Python version of all the course materials in Jupyter notebook
- 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