Ying Jin

Email: yjinstat@wharton.upenn.edu Website: https://ying531.github.io
Address: Room 441, Academic Research Building, 265 South 37th Street, Philadelphia, PA 19104

Academic Appointment

Assistant Professor 2025 - now
Department of Statistics and Data Science
Wharton School of Business, *University of Pennsylvania*
Applied Mathematics and Computational Science (AMCS), Affiliated Faculty

Wojcicki-Troper Postdoctoral Fellow 2024 - 2025
Data Science Initiative & Harvard Medical School, *Harvard University*

Education

Ph.D. in Statistics 2019 - 2024
Stanford University
Advisors: Emmanuel Candès, Dominik Rothenhäusler

B.S. in Mathematics 2015 - 2019
B.A. in Economics (Finance)
Tsinghua University

Research[1]

Research interests: Distribution-free inference; Uncertainty quantification; Causal inference; Generalizability; Selective inference; Distributional robustness; Replicability; Data-driven decision making.

Preprints & Under Revision

8. **Jin**, Y.*, Moon, I.*, and Zitnik, M. (2026). Act or Defer: Error-Controlled Decision Policies for Medical Foundation Models. medRxiv
7. Zheng, M. and **Jin**, Y. (2026). Distributionally Robust Conformal Prediction for Reliable AI under Distribution Shift. arXiv:2602.10018
6. Yang, Y. and **Jin**, Y.. Online selective conformal prediction with asymmetric rules: A permutation test approach. arXiv:2601.02998
5. **Jin**, Y. and Zubizarreta, J. R. (2025). Cross-Balancing for Data-Informed Design and Efficient Analysis of Observational Studies. arXiv:2511.15896
4. Gui, Y.*, **Jin**, Y.*, Nair, Y.*, and Ren, Z. (2025). ACS: An Interactive Framework for Conformal Selection. arXiv:2507.15825
3. Nair, Y., **Jin**, Y., Yang, J., and Candès, E. J. (2025). Diversifying Conformal Selections. arXiv:2506.16229
2. Bai, T. and **Jin**, Y. (2024). Optimized Conformal Selection: Powerful Selective Inference After Conformity Score Optimization. arXiv:2411.17983

[1]* = equal contribution or alphabetical ordering; †= co-senior author

1. **Jin**, Y.*, Guo, K.*, and Rothenhäusler, D. (2023). Diagnosing the Role of Observable Distribution Shift in Scientific Replications. arXiv:2309.01056

Journal Publications

12. **Jin**, Y. (2026). Replicability Within One Study: Harnessing Multiplicity for Observational Causal Inference. *Harvard Data Science Review*. HDSR column
11. **Jin**, Y., Egami N., and Rothenhäusler, D. (2025). Beyond Reweighting: On the Predictive Role of Covariate Shift in Effect Generalization. *Proceedings of the National Academy of Sciences*. arXiv:2412.08869
10. **Jin**, Y. and Candès, E. J. (2025). Model-free Selective Inference under Covariate Shift via Weighted Conformal P-values. *Biometrika*. arXiv:2307.09291
9. **Jin**, Y.*, Ren, Z.*, and Zhou Z. (2025). Sensitivity Analysis under the f-Sensitivity Models: A Distributional Robustness Perspective. *Operations Research*. arXiv:2203.04373
8. **Jin**, Y.* and Ren, Z. (2025). Confidence on the Focal: Conformal Prediction with Selection-Conditional Coverage. *Journal of the Royal Statistical Society: Series B*. arXiv:2403.03868
7. **Jin**, Y.*, Ren, Z.*, Yang, Z., and Wang, Z. (2025). Policy Learning 'without' Overlap: Pessimism and Generalized Empirical Bernstein's Inequality. *Annals of Statistics*. arXiv:2212.09900
6. **Jin**, Y. and Rothenhäusler, D. (2023). Modular Regression: Improving Linear Models by Incorporating Auxiliary Data. *Journal of Machine Learning Research (JMLR)*. arXiv:2211.10032
5. **Jin**, Y. and Candès, E. J. (2023). Selection by Prediction with Conformal P-values. *Journal of Machine Learning Research (JMLR)*. arXiv:2210.01408
4. **Jin**, Y. and Rothenhäusler, D. (2023). Tailored Inference for Finite Populations: Conditional Validity and Transfer Across Distributions. *Biometrika*. arXiv:2104.04565
3. **Jin**, Y.*, Ren, Z.*, and Candès, E. J. (2023). Sensitivity Analysis of Individual Treatment Effects: A Robust Conformal Inference Approach. *Proceedings of the National Academy of Sciences*. arXiv:2111.12161
2. **Jin**, Y., and Ba, S. (2022). Towards Optimal Variance Reduction in Online Controlled Experiments. *Technometrics*. arXiv:2110.13406 (Internship project at LinkedIn)
1. **Jin**, Y.*, Yang, Z.*, and Wang, Z.* (2024). Is Pessimism Provably Efficient for Offline RL?. *Mathematics of Operations Research*. Short version appeared at ICML 2021. arXiv:2012.15085

Conference Publications

8. Laghuvarapu, S., **Jin**, Y.†, and Sun, J.† (2026). ConfHit: Conformal Generative Design via Nested Testing. *International Conference on Learning Representations*.
7. Li, M. M., Li, K., Ektefaie Y., **Jin**, Y., Huang, Y., Messica S., Cai, T., and Zitnik, M. (2026). Controllable Sequence Editing for Biological and Clinical Trajectories. *International Conference on Learning Representations*. arXiv:2502.03569
6. Huang, K.*, **Jin**, Y.*, Li, R.*, Li, M., Candès, E. J., and Leskovec, J. (2025). Automated Hypothesis Validation with Agentic Sequential Falsifications. *International Conference on Machine Learning (ICML)*. arXiv:2502.09858
5. Yu Gui*, **Jin**, Y.*, and Ren, Z*. (2024). Conformal Alignment: Knowing When to Trust Foundation Models with Guarantees. *Conference on Neural Information Processing Systems (NeurIPS)*. arXiv:2405.10301
4. Wang, J., Dong, P., **Jin**, Y., Zhan, R., and Zhou, Z. (2024). Adaptively Learning to Select-Rank in Online Platforms. *International Conference on Machine Learning (ICML)*. arXiv:2406.05017

3. Huang, K., **Jin**, **Y.**, Candès, E. J., and Leskovec, J. (2023). Uncertainty Quantification over Graph with Conformalized Graph Neural Networks. *Conference on Neural Information Processing Systems (NeurIPS), Spotlight.* arXiv:2305.14535

2. **Jin**, **Y.** (2023). Upper bounds on the Natarajan dimensions of some function classes. *IEEE International Symposium on Information Theory (ISIT).* arXiv:2209.07015

1. La Cava W., Orzechowski, P., Burlacu, B., de França, F. O., Virgolin, M., **Jin**, **Y.**, Kommenda, M., and Moore, J. H. (2021). Contemporary Symbolic Regression Methods and their Relative Performance. *Neural Information Processing Systems Track on Datasets and Benchmarks (NeurIPS).*

Pre-PhD Work

2. **Jin**, **Y.**, Lu, J., and Wang, Z. (2020). Computational-Statistical Tradeoffs in Inferring Combinatorial Structures of Ising Model. *International Conference on Machine Learning (ICML).* PMLR 119:4901-4910

1. **Jin**, **Y.**, Guo J., Kang, J., and Guo, J. (2020). Bayesian Symbolic Regression. *Proceedings of AAAI Workshop on Statistical Relational Artificial Intelligence (AAAI).* arXiv:1910.08892

Academic Service

- **Seminar/workshop organizing**
  - Organizier, *Online Causal Inference Seminar*, September 2021 - Now.
    Duties include inviting speakers and discussants, moderating and hosting the seminars.
  - Program committee member, *ICML 2021 Workshop on Reinforcement Learning Theory*, July 2021.
  - Co-organizer, *ICLR 2024 Workshop on Machine Learning for Genomics Exploration*, May 2024.
  - Area chair, *NeurIPS 2024 Workshop on AI for New Drug Modalities (AIDrugX)*, December 2024.
  - Session organizer, *Generalizability, transportability, and distribution shift*, American Causal Inference Conference, May 2025.
  - Session organizer, *Conformal inference and statistical testing for reliable deployment of AI/ML models*, Joint Statistical Meetings, August 2025.
  - Co-organizer, *NeurIPS 2025 Workshop on Causality for Science*, December 2025.
- **Invited discussion**
  - For "Distribution-free inference for regression: discrete, continuous, and in between" by Yonghoon Lee, *International Seminar on Selective Inference*, February 2022.
  - For "CAP: A General Algorithm for Online Selective Conformal Prediction with FCR Control" by Changliang Zou, *International Seminar on Selective Inference*, June 2024.
  - For "Exploration, Confirmation, and Replication in the Same Observational Study: A Two Team Cross-Screening Approach to Studying the Effect of Unwanted Pregnancy on Mothers' Later Life Outcomes" by Dylan Small, *Causal Seminar at Harvard Data Science Initiative*, September 2024.
  - For "Exploration, Confirmation, and Replication in the Same Observational Study: A Two Team Cross-Screening Approach to Studying the Effect of Unwanted Pregnancy on Mothers' Later Life Outcomes" by Dylan Small, *Online Causal Inference Seminar*, April 2025.
  - For "Online selective conformal inference: adaptive scores, convergence rate and optimality" by Etienne Roquain, *International Seminar on Selective Inference*, February 2026.

- **Journal referee**: *Journal of the Royal Statistics Society: B; Annals of Statistics; Biometrika; Journal of the American Statistical Association; Annals of Applied Probability; Management Science; Journal of Machine Learning Research; Bernoulli; IEEE Transactions on Information Theory; Nature Machine Intelligence; ACM Computing Surveys; Statistica Sinica; Biometrics; Statistics in Medicine; Machine Learning; AIMS Mathematics; Journal of Causal Inference; IEEE Transactions on Artificial Intelligence; Transactions on Machine Learning Research.*
- **Conference reviewer**: *American Causal Inference Conference (ACIC) 2023; International Conference on Machine Learning (ICML); Neural Information Processing Systems (NeurIPS); NeurIPS Dataset and Benchmarks Track; International Conference on Learning Representations (ICLR); International Conference on Artificial Intelligence and Statistics (AISTATS); AAAI Conference on Artificial Intelligence.*

Invited and Contributed Talks

2026

- Joint Conference on Statistics and Data Science (JCSDS), July 2026.
- International Chinese Statistical Association (ICSA) China Conference, June 2026.
- International Chinese Statistical Association (ICSA) Applied Statistics Symposium, June 2026.
- Department of Statistics Seminar, University of Illinois Urbana-Champaign, April 2026.
- Department of Statistics and Probability Seminar, Michigan State University, April 2026.
- Causality and Machine Learning workshop, Tsinghua Sanya International Mathematics Forum (TSIMF), January, 2026.

2025

- IMS International Conference on Statistics and Data Science (ICSDS), December 2025.
- International Joint Conference CFE-CMStatistics, December 2025.
- Biostatistics/Bioinformatics Seminar, University of Maryland, December 2025.
- International Conference on Econometrics and Statistics (EcoSta 2025), August 2025.
- Join Statistical Meetings (IMS Lawrence D. Brown PhD Student Award session), August 2025.
- Join Statistical Meetings (Uncertainty Quantification for Large Language Models), August 2025.
- International Conference on Multiple Comparison Procedures (MCP), August 2025.
- Institute for Interdisciplinary Information Sciences (IIIS) Seminar, Tsinghua University, July 2025.
- Statistics Seminar, Shanghai Jiao Tong University, July 2025.
- Statistics Seminar, East China Normal University, July 2025.
- Statistics Seminar, Fudan University, July 2025.
- Statistics Seminar, Shanghai University of Finance and Economics, July 2025.
- Joint Conference on Statistics and Data Science (JCSDS), July 2025.
- 2025 WNAR/IMS Annual Meeting, June 2025.
- ASQ Chemical and Process Industries Division Webinar, June 2025.
- International Seminar on Selective Inference, April 2025.
- ENAR Spring Meeting, March 2025.
- IMSI Long Program on Uncertainty Quantification and Artifical Intelligence, March 2025.

2024

- Harvard Applied Statistics Workshop, September 2024.
- IMS International Conference on Statistics and Data Science (ICSDS), December 2024.

- Conference on Statistical Learning and Data Science (SLDS), November 2024.
- Data Science and Engineering Lab, Michigan State University, October 2024.
- INFORMS Annual Meeting, October 2024.
- Professor Junwei Lu's Group Meeting, Harvard University, October 2024.
- Joint Statistical Meetings, August 2024.
- International Conference on Econometrics and Statistics (EcoSta 2024), July 2024.
- Hangzhou International Conference on Frontiers of Data Science, China, July 2024.
- Joint Conference on Statistics and Data Science, China, July 2024.
- Inaugural Berkeley-Stanford Workshop on Veridical Data Science (lightning talk), May 2024.
- Design and Analysis Conference, Virginia Tech, May 2024.
- Clubear Statistics Organization (Virtual Talk), China, April 2024.
- INFORMS Optimization Society Conference (IOS 2024), March 2024.
- Statistics & Data Science Department Seminar, University of Pennsylvania, February 2024.
- Biostatistics Department Seminar, Columbia University, February 2024.
- School of ISyE Seminar, Georgia Tech, February 2024.
- Statistics Department Seminar, University of Wisconsin-Madison, February 2024.
- Statistics Department Seminar, Virginia Tech, February 2024.
- Statistics Department Seminar, Carnegie Mellon University, January 2024.
- Statistics Department Seminar, Columbia University, January 2024.
- ORIE Department Seminar, Cornell University, January 2024.
- Statistics Department Seminar, Harvard University, January 2024.
- Statistics, Operations, and Technology Department Seminar, Stern School of Business, New York University, January 2024.
- OM&BA Department Seminar, Carey School of Business, Johns Hopkins University, January 2024.
- MLBoost Online Seminar, January 2024.

2023

- Applied Mathematics Department Seminar, Brown University, December 2023.
- IMS International Conference on Statistics and Data Science (ICSDS), December 2023.
- Rising Stars in Data Science Workshop, University of Chicago, November 2023.
- Prof. Chiara Sabatti's group meeting, Stanford University, November 2023.
- Workshop on Operations Research and Data Science, Duke University, November 2023.
- Stanford-Berkeley Joint Colloquium Student Seminar, Stanford University, October 2023.
- INFORMS Annual Meeting, October 2023.
- Genentech Incorporation, September 2023.
- Joint Statistical Meetings, August 2023.
- Joint Conference on Statistics and Data Science in China, July 2023.
- ICSA China Conference, June 2023.
- Statistics Seminar, Suzhou University, China, June 2023.
- Causality in Practice Conference (Thematic Quarter for Causality), June 2023.
- INRIA Causal Inference Group, May 2023.
- International Seminar on Selective Inference, May 2023.
- International Conference on Design of Experiments (ICODOE), May 2023.
- Online Reinforcement Learning Theory Seminar, May 2023.
- DoorDash Causal Inference and Experimentation Team, May 2023.
- One World Mathematics of Machine Learning Online Seminar, April 2023.
- Data-Driven Decision-Making Seminar, Stanford Graduate School of Business, January 2023.

2022 and earlier

- Stanford Statistics Student Seminar, December 2022.
- Stanford Statistics Industrial Affiliates Annual Conference, November 2022.
- Data-Driven Decision-Making Seminar, Stanford Graduate School of Business, November 2022.
- Stanford Causal Science Conference, November 2022.
- INFORMS Workshop on Data Mining and Decision Analytics, October 2022.
- ICSA Applied Statistics Symposium (Student Paper Award presentation), June 2022.
- Conference on Digital Experimentation (CODE), November 2021.
- Joint Statistical Meetings, August 2021.
- Online Reinforcement Learning Theory Seminar, April 2021.

Software

- **BSR**, developer, https://github.com/ying531/MCMC-SymReg
  Python package for Bayesian Symbolic Regression method in Jin et al. (2020).
- **SRBench**, contributor, https://github.com/cavalab/srbench
  Large-scale benchmark for symbolic regression methods in La Cava et al. (2021).
- **condinf**, developer, https://github.com/ying531/condinf
  transinf, developer, https://github.com/ying531/transinf
  R packages for conditional inference for finite populations in Jin and Rothenhäusler (2023).
- **cfsensitivity**, contributor, https://github.com/zhimeir/cfsensitivity
  R package for sensitivity analysis and robust conformal inference of individual treatment effects under unmeasured confounding in Jin, Ren, and Candès (2023).
- **ConfSelect**, developer, https://github.com/ying531/conformal-selection
  R package for (Weighted) Conformalized Selection, which conducts calibrated selection of large outcomes with (weighted) conformal p-values in Jin and Candès (2022, 2023).
- **repDiagnosis**, developer, https://github.com/ying531/repDiagnosis
  R package implementing Jin, Guo, and Rothenhäusler (2023) for diagnosing replication studies.
  awesome-replicability-data, developer, https://github.com/ying531/awesome-replicability-data
  Online collection of publicly-available, individual-level datasets of replication study pairs.
  Rshiny app, developer, https://mbzlnj-ying-jin.shinyapps.io/shiny/
  Online live app for the method in Jin, Guo, and Rothenhäusler (2023) for diagnosing replication studies.

Industrial Experience

Data Science Applied Research Intern, *LinkedIn Applied Research Team* June - September 2021

Honors and Awards

IMS Lawrence D. Brown PhD Student Award, *Institute of Mathematical Statistics* 2025
Jack Youden Prize, *American Society for Quality's Chemical and Process Industries Division* 2024
Ingram Olkin Interdisciplinary Dissertation Award, *Department of Statistics, Stanford University* 2024
Rising Star in Data Science, *University of Chicago* 2023
Student Paper Award, *ICSA Applied Statistics Symposium* 2022
Tom Ten Have Award Runner up, *American Causal Inference Conference* 2022

D. E. Shaw Zenith Fellowship	2021
1st Place, Citadel West Coast Regional Datathon	2020
Outstanding Graduate Award (2/91), *Tsinghua University*	2019
Beijing Outstanding Graduate Award	2019
Qualcomm Scholarship for Research	2017
President Scholarship for Comprehensive Excellence (1/91), *Tsinghua University*	2017
Scholarship for Academic Excellence, *Tsinghua University*	2015
Scholarship for Freshmen, *Tsinghua University*	2014
Silver Medal in the 29th China Mathematics Olympics Final	2013