Jian Cao

Department of Economics
Trinity College Dublin
2 College Green
Dublin, Ireland

https://www.jiancao.net
+353 (89) 944 1053
caoj@tcd.ie
https://github.com/jian-frank-cao

Academic Background

Trinity College Dublin
Assistant Professor in Economics — *August 2024 –Present*
Research Fellow — *January 2022 –July 2024*

California Institute of Technology
Visitor — *January 2022 –December 2024*
Postdoctoral Scholar in Data Science and Election Integrity — *July 2019 –December 2021*

Florida State University
Senior Researcher — *August 2018 –June 2019*
Ph.D. Economics — *August 2018*
M.S. Economics — *August 2016*

China University of Mining & Technology
Master of Financial Engineering — *June 2014*

Henan University of Economics & Law
B.A. Economics — *June 2010*

Research Interests

Computational economics, industrial organization, data science
— advanced econometric modeling, Bayesian methods for large-scale economic datasets, market consolidation

Peer-Reviewed Articles

Cao, Jian, Thomas Chadefaux. 2024. "Dynamic Synthetic Controls: Accounting for Varying Speeds in Comparative Case Studies." *Political Analysis*.
doi: 10.1017/pan.2024.14.

Li, Zhuofang, **Jian Cao**, Nicholas Adams-Cohen, and R. Michael Alvarez. 2023. "The Effect of Misinformation Intervention: Evidence from Trump's Tweets and the 2020 Election." *Multidisciplinary International Symposium on Disinformation in Open Online Media 2023 Proceedings*.
doi: 10.1007/978-3-031-47896-3_7.

Cao, Jian, Seo-young Silvia Kim, and R. Michael Alvarez. 2022. "Bayesian Analysis of State Voter Registration Database Integrity." *Statistics, Politics and Policy*.
doi: 10.1515/spp-2021-0016.

Cao, Jian, Christina M. Ramirez, and R. Michael Alvarez. 2021. "The politics of vaccine hesitancy in the United States." *Social Science Quarterly*.
doi: 10.1111/ssqu.13106.

Alvarez, R. Michael, **Jian Cao**, and Yimeng Li. 2021. "Voting Experiences, Perceptions of Fraud, and Voter Confidence." *Social Science Quarterly*.
doi: 10.1111/ssqu.12940.

Srikanth, Maya, Anqi Liu, Nicholas Adams-Cohen, **Jian Cao**, R. Michael Alvarez, and Anima Anandkumar. 2021. "Dynamic Social Media Monitoring for Fast-Evolving Online Discussions." *Knowledge Discovery and Data Mining 2021 Proceedings*.
doi: 10.1145/3447548.3467171.

Cao, Jian, Nicholas Adams-Cohen, and R. Michael Alvarez. 2021. "Reliable and Efficient Long-Term Social Media Monitoring." *Journal of Computer and Communications*.
doi: 10.4236/jcc.2021.910006.

Working Papers

"Multiple Imputation for Large Multi-Scale Data With Linear Constraints."
(with Paul Beaumont)

"Ballot Rejections and Ballot Curing in Washington State."
(with Canyon Foot, Jay Lee, R. Michael Alvarez, Paul Manson, and Paul Gronke)

Work in Progress

"Dynamic Interaction Panel Estimation: Accounting for Complex Interdependence in Panel Data."
(with Thomas Chadefaux)

"Enhancing Regression Analysis through Self-Aligned DTW-Derived Speed Profiles in Time Series Data."
(with Thomas Chadefaux)

"The Parallel Quasi-Monte Carlo Bayesian Multi-Scale Multiple Imputation Method."
(with Paul Beaumont)

"Mailing It In: Voter Confidence in Vote-By-Mail In the 2020 Presidential Election."
(with R. Michael Alvarez and Seo-young Silvia Kim)

Research Experience

TRINITY COLLEGE DUBLIN
Research Fellow   *January 2022 –July 2024*
Project: Patterns of Conflict Emergence (ERC Grant No 101002240, €2 million)

Leading a project on identifying patterns in pre-conflict actions using multi-source data, aimed at enhancing conflict prediction models.

Collaborating with interdisciplinary teams to synthesize findings into actionable insights for policy planning.

CALIFORNIA INSTITUTE OF TECHNOLOGY

Visitor   *January 2022 –Present*
Postdoctoral Scholar in Data Science and Election Integrity   *July 2019 –December 2021*
Project: Election Auditing

Developed probabilistic matching and Bayesian multivariate models using GCP for large election database auditing in California and Florida.

Implemented entity resolution and anomaly detection on daily snapshots of voter registration databases that contain more than 20 million records and detected $10\times$ more true anomalies than the existing methods did.

Collaborated with California State, Los Angeles County, and Orange County election officials to address voter data security issues, resulting in formal recognition from the Secretary of State for contributions to the 2020 general election.

Project: Twitter Monitoring

Developed serverless architectures using GCP, AWS, and Oracle for long-term Twitter monitoring. They ingest, process, and store more than 4.5 billion tweets (30 TB in size) related to COVID-19, primary/general elections, and protests.

Work closely with the Computer Science team and implemented topic, spatial, network, and sentiment analyses on the collected tweets and identified COVID-19 misinformation and voting issues in the 2020 Election cycle.

Shared a 30TB COVID-19 Twitter dataset with NVIDIA and hosted workshops to implement machine learning algorithms for misinformation and hate speech mitigation.

FLORIDA STATE UNIVERSITY

Senior Researcher   *August 2018 –June 2019*
Project: Large Missing Data Multiple Imputation

Developed the fastest and most accurate Bayesian inference method for missing data multiple imputation.

Developed a parallel-sequential imputation method that can impute large multi-scale data sets with 1.5 billion observations (500 GB in size).

Project: Economic Impact Modeling

Analyzed the economic impact of Florida's housing and small business policies.

Developed a NETS-based impact analysis tool that has 1000 times finer resolution than the existing methods.

Teaching

Trinity College Dublin	Assistant Professor, *ECP77101*: Introductory Mathematics (2024);
	ECU33081: Mathematical Economics A (2024).
	ECU44100: Economics Capstone (2024, 2025).

Caltech   Post-doc lecturer, *SS 224*: Social Science Data (2019, 2021).

Florida State University   Teaching assistant, *ECO 3431*: Analysis of Economic Data (2017).

Henan University of Economics & Law   Teaching assistant, Statistics (2008).

Teaching Achievements

2024   Second highest-rated (rated 4.63/5 by students) undergraduate Economics module at Trinity College Dublin (ECU33081: Mathematical Economics A).

2017   1st place in the teaching assistant demonstration at Florida State University.

Software

dsc   A R package for the Dynamic Synthetic Control method, an enhanced synthetic control method that accounts for misalignment in time series caused by inherent speed differences.

spike   A Python package for deploying long-term Twitter monitors and managing files on Google Drive, Google Cloud Storage, and SFTP file systems.

Conference and Seminar Presentations

2024   APSA 2024 (American Political Science Association Annual Conference);
TCD Econ WiP Seminars.

2023   APSA 2023 (American Political Science Association Annual Conference);
MISDOOM 2023 (Multi-disciplinary International Symposium on Disinformation in Open Online Media Annual Conference).

2021   KDD 2021 (Knowledge Discovery and Data Mining);
ESRA 2021 (Election Sciences, Reform, & Administration Conference);
RRoCCET 2021 (Research Running on Cloud Compute & Emerging Technologies).

2020   ESRA 2020 (Election Sciences, Reform, & Administration Conference);
VoteCal (Vote California);
Caltech COVID Dynamic Talk.

2019   Caltech SISL Talks (Social and Information Sciences Laboratory);
FSU GSMC Talks (Graduate Modern Statistics Club).

Peer Review Activities

Political Behavior, American Journal of Political Science, SN Computer Science, Journal of Computer and Communications.

References

Paul Beaumont (primary advisor)
Associate Professor of Economics
Florida State University
beaumont@fsu.edu

R. Michael Alvarez
Flintridge Foundation Professor
of Political and Computational
Social Science
California Institute of Technology
rma@caltech.edu

Thomas Chadefaux
Professor in Political Science
Trinity College Dublin
thomas.chadefaux@tcd.ie

Animashree Anandkumar
Bren Professor of Computing and
Mathematical Sciences
California Institute of Technology
arangelf@caltech.edu