**C.J. Duan**

Email: compute@dulun.com | Website: www.dulun.com

## Professional Summary

Causal Inference Data Scientist with a Ph.D. in Industrial Management and extensive experience applying advanced causal inference methodologies, statistical modeling, and machine learning to real-world problems in healthcare, life sciences, and consumer sectors. Demonstrated expertise in treatment effect estimation, propensity score methods, and causal diagrams (DAGs), with hands-on skills in Python and leading causal inference libraries (DoWhy, EconML, CausalImpact, statsmodels, CausalForest). Adept at collaborating with clinical, product, and commercial teams to deliver data-driven insights for evidence-based business and clinical decisions.

## Technical Skills and Toolsets

- Causal Inference & Statistics: Propensity Score Matching, Instrumental Variables, A/B Testing, Treatment Effect Estimation, Causal Diagrams (DAGs), Survival Analysis, Bayesian Inference  
- Programming & Frameworks: Python (DoWhy, EconML, CausalImpact, statsmodels, CausalForest, PyTorch, SKlearn), R (Shiny, Leaflet), Stan, Pyro, Docker  
- Data & Cloud: Healthcare Datasets (EHR, claims, clinical trials), AWS (SageMaker, Lambda, Bedrock), SQL  
- Software Development: Version Control (Git), Modular Code Design, Testing Frameworks, CI/CD (CircleCI, Docker)  
- Visualization & Analytics: SAS, SPSS, Tableau, Power BI

## Professional Experience

Adjunct Professor of Data Analytics  
Purdue University Global, School of Business and IT  
April 2024 – Present  
- Teach data analytics with an emphasis on causal inference and statistical modeling for healthcare applications.  
- Guide students in applying machine learning and causal methods to real-world medical datasets.

Adjunct Professor of Data Science  
University of Maryland Global Campus  
April 2023 – Present  
- Developed and taught advanced courses on Bayesian inference, statistical modeling, and experimental design for treatment effect analysis.

Contract Research Data Scientist  
PepsiCo (via Insight Global)  
December 2021 – July 2022  
- Designed and implemented Bayesian Media Mix Models using Stan for causal attribution of marketing interventions.  
- Applied time-series causal inference techniques to estimate incremental ROI and optimize media spend.

Assistant Professor of Quantitative Methods  
Troy University (Global and AL Campus)  
March 2009 – May 2017  
- Led research in quantitative modeling, causal inference, and Bayesian statistics; published on bias mitigation and treatment effect estimation.

## Project Highlights

DRC Lab, Healthcare  
Project: MuST Model for Hospital Readmission Prediction (2023–Present)  
- Led development of a multimodal transformer model integrating EHR and medical imaging for causal analysis of readmission risk factors.  
- Designed and interpreted causal diagrams (DAGs) for healthcare data pipelines.

DRC Lab, BioTech  
Project: scGPT for Single-Cell Multi-omics (2023–Present)  
- Replicated state-of-the-art generative models for causal inference on biological datasets.

PepsiCo, CPG/Retail  
Project: Bayesian Media Mix Modeling (2021–2022)  
- Developed state-space models and treatment effect estimations for campaign evaluation.

Troy University, Sports Analytics  
Research: Bayesian Analysis of Home Field Advantage in Soccer (2017–2020)  
- Designed causal frameworks for isolating treatment effects in sports analytics.

## Education

Ph.D. in Industrial Management  
Clemson University

## Selected Publications & Research

- Biases in Machine Learning for Phishing Detection, Journal of Business Analytics, 2021  
- Bayesian Analysis of Home Field Advantage in Soccer, Journal of Business Analytics, 2020  
- Revenue Management Models in CPG: Presented robust demand estimation and causal modeling techniques for marketing optimization

## Contact & References

Available upon request.