Lily Liu

☎ 310-895-5226 ✉ lilyliu8262@gmail.com

EDUCATION

University of California, Los Angeles   2022
PhD in Economics, Department of Economics (Primary Field: Econometrics)   *GPA: 3.9/4.0*

Peking University   2017
M.A. in Economics, National School of Development/CCER (Primary Field: Applied Economics)

EXPERIENCE

University of California, Los Angeles   2019 – 2022
Research Scientist/Data Scientist/Project Manager   *Los Angeles, CA*

Estimation of Average Treatment Effects (ATE) with Network Effects
- Estimated ATE with strategic interaction in networks (e.g. Competitive Pricing Strategy with Strategic Substitutability)
- Identified latent network structures using the K-Means Clustering algorithm, proved $\sqrt{n}$ consistent two-step estimation
- Applied it to heterogeneous treatment effects and strategically formed networks with High Dimensional features

Estimation of Average Treatment Effects (ATE) for Massively Imbalanced Data
- Examined maximum likelihood estimator when events occur with a probability $< 10\%$ using simulated e-commerce data
- Designed random/nonrandom experiments to generate different distributions of propensity scores with Monte Carlo
- Reduced bias in estimators of price elasticity in Logistic Regression by 5%, ATE by 2% in nonrandom experiments

Evaluation of Social Experiments at the California Center for Population Research
- Identified neighborhood effects in housing experiments(Moving to Opportunity) with revealed preference incentive analysis
- Studied short and long term effects of the largest American youth employment and training program on labor markets

Model Selection Tests and Bootstrap for Panel and Network Data with High Dimensional Heterogeneity
- Corrected bias of estimators for irregular data structures (e.g. imbalanced, high-dimensional panel and network data)
- Constructed test statistics with modified likelihood functions to select strictly non-nested, overlapping, and nested models
- Derived the limiting null distribution of the test statistics, conducted power analysis, and proved bootstrap validity

Publications

Hahn, J. and Liu, X., 2022. Jackknife bias reduction for simulated maximum likelihood estimator of discrete choice models. Economics Letters, p.110784.

RAND Corporation   2017 – 2018
Research Scientist/Data Scientist   *Santa Monica, CA*
- Investigated treatments effects of economic policy experiments on wage, education, marriage, and health outcomes
- Engineered informative features with demographic properties, family backgrounds, treatment intensities, social networks
- Combined Regularized Regression algorithms, Random Forest, Gradient Boosted Trees with Causal Inference methods
- Showed large rural-urban disparity and the significant gender gap in many massive government intervention experiments

Peking University   2015 – 2017
Data Analytics Project Manager   *Beijing, China*
- Led field survey teams to collect data by designing social experiments and controlling for selection bias into the treatment
- Improved response rates and external validity, reduced reporting/randomization bias with effective questionnaire designs
- Addressed noncompliance issues by offering vouchers to incentivize respondents to participate in social experiments

PROJECTS

A/B Testing on Email Marketing Campaign (Python, A/B Test, Time Series)
- Designed email multi-arm experiments to improve conversion rate by dividing users ($\approx 500k$) into 24 treatment groups
- Collected data with possible combinations of email types, user segments, delivery schedules; identified the best user group
- Conducted one sample proportional A/B test; recommended increasing sending frequency by 50% with time series analysis

Text Analysis on Amazon Reviews (Python, Machine Learning, Natural Language Processing)
- Automated analysis of user experience reviews ($\approx 570k$) to determine each reviewer's sentiment with listing products
- Built a term-doc incidence matrix with review observations; created cloud word; conducted exploratory data analysis
- Applied N-Gram, TFIDF methods to vectorize text, built Logistic models, Naïve Bayes classifiers with AUC$= 96\%$

Credit Card Fraud Detection (Python, Deep Learning, Imbalanced Data)
- Trained Neural Networks and Decision Trees to detect fraud transactions with high AUC scores ($> 90\%$)
- Performed upsampling minority, downsampling majority, and SMOTE for imbalanced classification datasets

TECHNICAL SKILLS

- Programming Languages: Python (Pandas, NumPy, SciKit-Learn, TensorFlow, PyTorch, BeautifulSoup, Matplotlib, Seaborn), SQL, R, Stata, Matlab
- Statistical Machine Learning Techniques: Linear and Logistic Regression, Naïve Bayes, Decision Tree, Random Forest, Gradient Boosting, K-Nearest Neighbors, K-Means Clustering, Latent Dirichlet Allocations, Principal Component Analysis, Regularization, Ensemble Methods (Boosting and Bagging), Cross-Validation
- Causal Inference Methods: A/B Testing, Instrumental Variables Estimation, Propensity Score Matching, Synthetic Control, Difference-in-differences Estimation, Binary Variable Models, Design of Experiment