Lily Liu

310-895-5226 lilyliu8262@gmail.com https://lilyliu8262.github.io/econ/

EDUCATION

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

Peking University 2017
Master of Finance, National School of Development/CCER (Primary Field: Economics and Finance)

EXPERIENCE

University of California, Los Angeles 2019 – 2022
Financial Econometrician *Los Angeles, CA*

Bias-corrected Model Selection Tests of Contingent-claims Prices with High Dimensional Heterogeneity
- Corrected bias for Black-Scholes model; reduced variance ($\approx 25\%$) for deep-out-of-money options with Monte Carlo
- Applied it to Vasicek model to price discount bonds with simulation-based estimates; showed smaller biases and variances
- 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

Prediction of the Chinese Stock Market with Machine Learning Algorithms
- Investigated predictability in long-only Chinese stock market with retail investors' dominating presence (99% of accounts)
- Explored different machine learning methods' predictive ability; neural networks outperform others (out-of-sample R^2)
- Large stocks and state-owned enterprises are predictable in long run; small stocks have higher short-term predictability

Estimation of Economic Policy Effects for Massively Imbalanced Data
- Examined maximum likelihood estimators when events occur with a probability $< 10\%$ using simulated imbalanced data
- Designed Monte Carlo simulations to generate different distributions of propensity scores for mortgage prepayment
- Reduced bias in estimators of price elasticity in Logistic Regression by 5%, average effects by 2% in simulations

Publications

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

Hahn, J., Ridder, G. and Liu, X., 2022. Estimation of average treatment effects (ATE) for massively imbalanced data. Econometric Reviews (forthcoming).

RAND Corporation 2017 – 2018
Macroeconomist *Santa Monica, CA*
- Investigated macroeconomic policy effects of government interventions on wealth accumulation and economic growth rate
- Engineered informative features with demographic population properties, social-economic status and policy intensities
- 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 macroeconomic policies

Peking University 2015 – 2017
Data Analytics Project Manager *Beijing, China*
- Evaluated general and distributional macroeconomic effects of economic policy by collecting microeconomic data
- Improved policy evaluation accuracy by increasing response rates, enhancing external validity, reducing estimation bias

PROJECTS

Delta-hedging PnL Simulation and Variance Swap Replication
- Investigated the PnL path dependency of delta-hedged option portfolio by simulating various underlying paths
- Showcased the lack of PnL path dependency for variance swap in contrast to vanilla options with replications

Asset Allocation Models
- Developed asset allocation models using MVO (Mean-Variance Optimization) and Risk Parity on different asset classes
- Generalized models by incorporating risk budget constraints through portfolio volatility decomposition with marginal risk
- Built a momentum factor based long-short overlay on equities; improved performance through diversification benefit

Prediction of Housing Prices (Python, Supervised learning, Time Series)
- Predicted housing prices using Regularized Regression algorithms, Random Forest, Gradient Boosted Decision Trees
- Engineered informative features with past prices, housing properties, locations, and demographics of the neighborhood
- Fine-tuned the model with grid search and cross validation, the best model has a root mean square error of less than \$50k

Stock Prediction with Twitter Sentiment Analysis (Python, Machine Learning, Natural Language Processing)
- Predicted stock prices using LSTM and ARIMA by incorporating sentiment features from Twitter as financial lexicons
- Applied N-Gram and TFIDF methods to vectorize text; improved 10% accuracy by involving indicators for sentiment

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, SQL, R, Stata, Matlab
- Statistical/Machine Learning Modeling 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, Synthetic Control, Causal Inference