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 (Primary Field: Financial Econometrics) *GPA: 3.9/4.0*

Peking University 2017
Master of Finance, National School of Development/China Center for Economic Research (Primary Field: Finance)

EXPERIENCE

Goldman Sachs 2023 – present
Quantitative Researcher

- Assessed volatility and PnL of portfolios under different market shock scenarios by monitoring various hedging strategies
- Reviewed performance of portfolio optimization and pricing models; performed sensitivity testing; validated the soundness
- Enhanced asset allocation models by incorporating risk budget constraints on assets via portfolio volatility decomposition

University of California, Los Angeles 2019 – 2022
Financial Econometrician

Asset Pricing Models with Machine Learning Algorithms

Trading Signals

∗ Identified various predictive signals for stock markets: price trends in the US and liquidity in Chinese long-only markets
∗ Developed optimal alpha strategies based on over $1k$ signal features, trained tree models, neural networks on $35k$ stocks
∗ Tracked out-of-sample Sharpe ratio(> 3) with large-scale data (over 60 years) by decomposing variance-covariance matrices
∗ Constructed weighted portfolios based on principal component analysis; improved performance by utilizing cross-predictors

Model Tests

∗ Corrected bias for Black-Scholes model; reduced variance ($\approx 25\%$) for deep-out-of-money options with Monte Carlo
∗ Constructed bias-corrected model selection tests of contingent-claims prices with high dimensional stock heterogeneity
∗ Derived the limiting null distribution of the test statistics, conducted power analysis, and proved bootstrap validity

Stock Price Prediction with Web Scraped Data and Natural Language Processing

∗ Predicted stock prices using features from web-scraped data, including job openings, sales, purchases, news, social media
∗ Built predictive models using LSTM and ARIMA by incorporating sentiment features from Twitter as financial lexicons
∗ Applied N-Gram and TFIDF methods to vectorize text; improved accuracy by 10% by incorporating sentiment indicators

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., 2023. Estimation of average treatment effects (ATE) for massively imbalanced data. Econometric Reviews (forthcoming).

RAND Corporation 2017 – 2018
Financial Economist

Estimation of Macroeconomic 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
· 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

Predicting Housing Prices with Supervised Learning Using Time Series Data

· 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

TECHNICAL SKILLS

· Programming Languages: Python, SQL, R, Matlab
· Econometric/Statistical/Machine Learning Methods: Cross-Validation, Synthetic Control, Causal Inference, Latent Dirichlet Allocations, Principal Component Analysis, Regularization, Ensemble Methods (Boosting and Bagging), Linear and Logistic Regression, Naïve Bayes, Decision Tree, Random Forest, Gradient Boosting, K-Nearest Neighbors, K-Means Clustering