HENGXU (HUGH) LIN

hengxu.lin@outlook.com | (86) 137-1934-4881 | Homepage | LinkedIn

EDUCATION

Columbia University in the City of New York - Columbia Business School　　　　　　　　　New York, US
MS. in Financial Economics (Pre-Ph.D)　　　　　　　　　　　　　　　　　Sept 2021 - **Expected May 2023**
- **Anticipated Coursework(All Ph.D-level):** Optimization, Applied Multivariate Statistics, High Dimensional Statistics, Reliable Statistics Learning, Financial Data Science, Econ & Statistical Inference, Time Series Analysis, Microeconomics Analysis
- **Programming**: Python (Torch/Tensorflow), Linux Server Programming, C++, Java/C#, SQL, 50k+ lines programming experience

Sun Yat-Sen University　　　　　　　　　　　　　　　　　　　　　　　Guangzhou, China
BS. in Mathematics and Applied Mathematics, BBA. in Accounting　　　　　　　　　Aug 2015 - June 2020
- **Coursework:** Neural Networks, Machine Learning, Probability & Statistics, Optimization, Calculus, Algebra, Real Analysis, ODE, Algorithms, Financial Engineering, Investment, Corporate Finance, Econometrics, Micro/Macro-Economics
- **Academic Honors**: Outstanding Graduate (highest distinct,Top 0.5%), First-place Scholarship, Microsoft Stars of Tomorrow

EXPERIENCE

Microsoft Research Asia (MSR Asia)　　　　　　　　　　　　　　　　　　　Beijing, China
Research Intern, Machine Learning Group, Advisor: Dr. Weiqing Liu & Dong Zhou　　　　June 2020 - Present
- **Multi-Domain Learning on Stock Data (Under review of 2021 ACM-SIGKDD)**
  - **Multiple Patterns Adaptor**: Proposed a temporal router adaptor to automatically assign samples into multiple domains based on its memory mapping, applied Sinkhorn algorithms in optimal transport to cluster on the memory bank;
  - **Ablation Study**: Researched on the influence of number of domains, performance of hidden states and memory mapping combination, and improvement on different SOTA models, enhanced top baselines' RankIC by 1%, Annual Return by 3.1%.
- **Self-supervised Learning on Stock Data Distribution**
  - **Domain Identification**: Assumed data's non-i.i.d. via clustering and dimension decomposition (t-SNE/Autoencoder);
  - **Deep Clustering**: Combined reconstruction loss in Autoencoder with K-Means based pseudo labels, projected data into a linear separable hidden space, achieved 3% enhancement to baseline (manifold clustering);
  - **Contrastive Learning**: Applied a contrastive method combining representing learning with optimal transport, which is the best online clustering algorithm, achieved oracle accuracy on synthetic data compared to supervised learning (99%).
- **Qlib: 1st open-source AI platform for Quantitative Finance**
  - **Development**: Contributed to alpha factor-base and baseline testing. (github.com/microsoft/qlib)

Sunshine Quant Investment Consulting (Prop Trading)　　　　　　　　　　　　Shenzhen, China
Research Intern, Quantitative Trading, Advisor: Dr. Guang Yang　　　　　　　　　Feb 2020 - May 2020
- **Portfolio Optimization**: Built Black-Litterman model combined with risk parity strategy and adversarial learning prediction, enhanced index's return by approx. 4%, max drawdown by 30%. (github.com/linhx25/BlackLittermanModel)
- **Adversarial Learning:** Constructed an architecture of GAN with the MLP as the discriminator and the LSTM as the generator for forecasting the return; empowered single LSTM's accuracy by 2%, achieved average accuracy of 58% and MSE of 0.44%.

Morgan Stanley Huaxin Fund Management (Mutual Fund)　　　　　　　　　　　Shenzhen, China
Risk Analyst, Risk Management　　　　　　　　　　　　　　　　　　　　May 2019 - Aug 2019
- **Performance Attribution**: Based on Brinson model, created C# program for funds' return breakdown into security selection and industry allocation, improved attribution algorithm thereby hastening the process by 5 times;
- **Fair Trading Analysis**: Computed profit discrepancies in trading window according to regulations in analyzing unfair trades.

Southern China Center for Statistical Science　　　　　　　　　　　　　　　Guangzhou, China
Financial Engineering Program Leader, Supervisor: Prof. Xueqin Wang　　　　　　　Sept 2017 - Feb 2019
- **Numpy.NET**: Developed data structure emulating Numpy in C#, allowing efficient functions in time series analysis model;
- **Hidden Markov Model**: Utilized the EM algorithm to estimate MSVAR's parameters. (github.com/linhx25/MarkovSwitching)

PROJECTS

Undergraduate Thesis (Best Paper) – "*Readability, Opaqueness and Crash Risk*", Advisor: Prof. Keming Wang
- Developed a package (Python) to process firms' 10-K fillings (217G), applied NLP method for annual report textual analysis;
- Conducted robustness test to eliminate endogeneity: searched alternative variables, tested fixed effect of panel data;
- Proved via research that 10-K of high-risk firms are of low readability and high opaqueness. (github.com/linhx25/FReader)

Kaggle's PUBG Machine Learning Prediction (Ranked Top 1%)
- Feature Engineering: extracted, combined features and tested linear/non-linear models with regularization in pre-experiment;
- Implemented Random Forest, Genetic Algorithm, Particle Swarm and GBDT in Python for prediction, with 1.95% MAE;
- Ranked Top 1% in Kaggle's Project competition. (kaggle.com/c/pubg-finish-placement-prediction/leaderboard , Team SYSU).

OTHER INFORMATION

Languages: Cantonese, English (GRE: 329; TOEFL: 109)　　|　　**Interests**: Hip-hop dance, Climbing (elevation 6,000 m), Cooking