734-968-8570
Ann Arbor
pycai@umich.edu

Peiyao Cai

Personal Website
GitHub: peiyaocai

Education

Master of Mathematics **Jan 2022 - June 2023 (expected)**
University of Michigan, Ann Arbor

GPA: 4.0(4.0)
Main courses: Real Analysis(A), Algebraic Combinatorics(A+), Advanced Linear Algebra, Discrete Stochastic Process.

Master of Statistics **Sep 2021 - June 2023 (expected)**
University of Michigan, Ann Arbor

GPA: 3.98(4.0)
Main courses: Probability Distribution Theory(A), Statistical Learning: Regression(A), Computational Methods & Tools in Statistics(A), Statistical Inference(A).

Bachelor of Finance **Sep 2017 - June 2021**
Peking University

Core GPA: 3.6(4.0)
Main courses: Macroeconomics(A), Data Analysis and Statistical Software(A), Mathematical Methods in Finance(A+), Analysis of Financial Time Series(A).

Exchange Student **Sep 2019 - Dec 2019**
Northeastern University

Main courses: International Finance, Investment Theories.

Publications

Differential Score Matching Graphical Models

Under revision of JMLR, joint work with Kean Ming Tan and Junwei Lu

Abstract: It is of great interest to make estimation and learn the structure of differential graphical models. In this paper, we use the score matching method to give a joint estimation for differential graph between two independent exponential family graphical models. The "Untangle and Chord" method enable us to further make the estimator unbiased and thus make valid inferences on single edges. One particular important question is to test the maximum degree in the differential graph. A node with high degree (i.e. a hub node) in the differential graph means existence of a special variable whose interactions with other variable are significantly different between two groups of samples. Testing on hub nodes has many applications in genetics and neuroscience, such as mutated genes detection. We adapted the "Skip Down" algorithm to test the existence of hub nodes. Numerical study and real data application are provided to illustrate the effectiveness of proposed method.

Research Experience

Differential Score Matching Graphical Models **Apr 2022 - Oct 2022**
Supervisors: Kean Ming Tan(University of Michigan), Junwei Lu(Harvard University) *Ann Arbor, MI*

- Proposed a novel estimator for differential edge parameters between two independent probabilistic graphical models based on score matching method, which applies to general exponential graphical models.
- Designed a joint ADMM algorithm to solve the estimation problem with extra L1 penalties and return solutions under sparsity assumption.

- Further developed a general framework for estimation, getting de-biased estimator and making inferences for differential graph.
- Proved theorems, designed numerical simulations and real data applications, wrote the paper draft independently.

High-Frequency Price Jumps and News Impacts **Apr 2021 - Sep 2022**
Supervisor: Chenxu Li(Guanghua School of Management, Peking University)

- Cleaned the original high-frequency trading data sets for 44 Don Jones Index companies from year 2003 to 2018 with the sample size reaching 100 million level, reorganized the data structure and constructed 19 firm-idiosyncratic features for firms under different trading days with different frequencies. Constructed the well organized panel data for further analysis.
- Combined the high-frequency stock return time series and company-related news from Thomson Reuters News Database together to detect abnormal stock return residuals and their characteristics for each company.
- Set different truncation cutoff to determine high-frequency price jumps for companies, and further ran probit regressions to detect variables' contributions for predicting news related jumps.
- Applied machine learning algorithms including probit-lasso regression and random forest to study the microstructure of high-frequency stock market.

The Tournament Promotion Model in China's Prefecture-Level Cities **Mar 2021 - Jul 2021**
Supervisor: Li-An Zhou *Beijing*

- Worked with Professor Zhou to conduct a DID-based method to evaluate the Tournament Promotion Model in China's prefecture-level cities.
- Established database of government annual report at the level of different prefecture-level city throughout the whole country.

US Mutual Fund Seasonalities **Jun 2021 - Sep 2021**
Supervisors: Yingguang Zhang(Peking University), Jiacui Li(University of Utah) *Beijing*

- Used Python to assist professors to pre-process datasets containing million of observations, including labeling each observation with right signs, selecting useful datasets and constructing important statistical variables.
- Used R to run various tests and regressions with multiple fixed effects to detect and verify seasonalities in US stock return.

Earnings Announcement Return Cycle **Apr 2021 - Jul 2021**
Supervisor: Yingguang Zhang(Peking University) *Beijing*

- Assisted professor to construct various different Change-of -Forecast Variables to examine whether the earnings announcement return cycle can accord with the cycle of analysts change their forecast.

Work Experience

Equity Capital Market Intern **Jun 2020 - Sep 2020**
Northeast Security

- Responsible for updating weekly capital market database and writing weekly reports. Designed a Python Crawler program to help collect important data on website automatically.
- Maintained company's financial system: updated financial data, proofread figures and balance in Balance Sheet etc.
- Participated in a directional private placement project of a domestic investment bank with a volume of 6 billion RMB , and was responsible for writing and checking important compliance documents.

Data Analytics Intern **Jan 2020 - Mar 2020**
PwC China

- Extracted financial data of Chinese enterprises from the data of the past 12 years with the volume of millions. Constructed the corresponding factors according to the Beneish model.
- Applied basic model including logit regression to advanced algorithm including random forest classification and xgb regressor to quantify the probabilities of degrading of Chinese bonds.
- Constructed 27 different factors that could influence the probability of bond degrading. Constructed ROC curve to detect the goodness of fit of the machine learning model and the AUC area came to 0.92.

Industrial Research Intern **Jun 2019 - Sep 2019**
New China Fund Ltd

- According to daily condition of A shares market and fluctuations of specific indexes, composed the daily news report about the industry of wind power and new energy.
- Participated in various of strategy meetings, analysed the potential investment opportunities toward wind power industry.
- Carried out the industrial research in wind power independently. According to the combination between marco politics and the business model of specific corporation, I wrote research report to facilitate the department to choose the investment object.

Skills

Programming	Python, R, Git, LaTeX, Matlab, MarkDown
Communication	Chinese (native), English (business)
Other	Github, Microsoft Office