ZHENZHI (MARK) HE

(773) 219-4881 | zhenzhi.he@kellogg.northwestern.edu

EDUCATION

Kellogg School of Management (Northwestern University), IL   **June 2026 (expected)**
Ph.D. Candidate in Finance, core sequences: Asset Pricing (3.9/4), Macro (3.9/4), Econometrics (3.8/4)
- 2023   Dissertation Research Grant from Dean's Office, Kellogg School of Management
- 2020-2026 Ph.D. Fellowship, Kellogg School of Management

The University of Chicago, Chicago, IL   **March 2020**
Master of Arts in Economics, core courses: Machine Learning (A), Analysis In Rn-3 (A)
- 2018-2019 Tuition Award, UChicago

Peking University, Guanghua Schools of Management and Mathematical Sciences, Beijing   **June 2018**
Bachelor of Management, Minor in Mathematics, Excellent Graduate of Beijing
- 2017-2018 Excellent Graduate Student of Beijing (top 1%), Beijing Municipal Commission of Education
- 2017-2018 National Scholarship of China (top 2% over 267), Ministry of Education of China
- 2016-2017 May 4th Scholarship (top 10%), Peking University

RESEARCH

Earnings Calls and Cross-section Asset Prices   June 2024 - present
Coauthor: Cong Zhang
- Found that the OpenAI text embedding (3,072 dimensions) of earnings calls can predict the future stock returns.
- Applied projected PCA to model α and β, and constructed a portfolio based on the predicted α in the out-of-sample. The portfolio generated an average monthly return of 300 bp with an annual Sharpe ratio of 1.48. The highest month return reached 18.5% while the worst was -16%, with a max drawdown of -27%. After adjusting for various risk factors benchmarks, the portfolio maintained a risk-neutralized return of 229 to 279 bp.
- The predicted α and β explain 1.5% and 10.13% in-sample variations of returns, while 3 bp and 0.2 bp in out-sample.
- Identified that text embeddings provide additional information beyond 61 stock characteristics. Combining text embeddings with these characteristics increases the out-of-sample Sharpe ratio by 0.12.
- Found that predictability is significantly higher for stocks with lower analyst coverage and greater forecast dispersion.

Endogenous Information Acquisition: the Impact of Ex-Ante Beliefs on Analyst Questions   Mar 2024 - present
- Utilized a text-tiling algorithm to segment managers' presentations and linked them to analyst questions.
- Found that more optimistic analysts prior to calls focused on more positively-toned and less risk-laden content.

Forecasting Cryptocurrency Returns Using Neural Networks   June 2023 – Nov 2023
- Constructed a neural network utilizing an encoder architecture in conjunction with a gated recurrent unit (GRU) to effectively capture both time-series and cross-sectional information of crypto perpetual futures returns, resulting in a 11% information coefficient when predicting one minute future returns in the test set.
- Enhanced out-of-sample prediction accuracy by 1% through data segmentation and ensemble modeling.

Technology Clusters Over Centuries   Sep 2022 - present
Coauthor: Bryan Seegmiller, Lawrence Schmidt, Dimitris Papanikolaou
- Collected OpenAI text embeddings for patent summaries to identify evolving technology clusters of patents over time.
- Utilized LLMs (Mixtral, Mistral, and Pegasus) to generate patent summaries and evaluate them using ROUGE metrics.
- Applied a scalable hierarchical cluster with a consistent distance cutoff, revealing fewer clusters relative to the number of patents in later years, suggesting smoother technological progress than implied by patent counts.

EXPERIENCE

Full-Time Research Professional, Kellogg School of Management, Northwestern University   Aug 2019 - July 2020
Supervisor: Efraim Benmelech, Anthony DeFusco, Jacopo Ponticelli
- Contributed to the success of 7 projects, with 4 of them achieving successful publication.

RESEARCH INTERESTS

Machine Learning in Asset Pricing, Cross-section of returns, Large Language Model, Text Mining, Neural Network

SKILLS

Programming: Python (pandas, spacy, gensim, numpy, pytorch, transformers, llama_cpp), R, Stata, MATLAB, Java (disruptor)
Language: Mandarin (native), English (fluent)