MF 703

Programming for Mathematical Finance

Fall 2023

Course Project Proposal

Project Theme	Index tracking portfolio optimization based on deep learning framework ——exemplified on the CSI 300 Index
Team Members and Student ID	Ziyuan Zhang
	Zhitong Ye
	Yingzi Li
	Yang Xia
	Weichen Zhu

1. Research Background and Research Objectives

Index tracking portfolio optimization is popular form of passive investment strategy. It operates by constraining trading costs and management fees, and purchasing either a part or the entirety of the target index's constituent securities to construct a portfolio that replicates the performance of the benchmark or target index. The goal is for the index products to have a high correlation with the target index and achieve a similar return and volatility. Compared to other mutual funds, index funds are a relatively stable and efficient investment tool, playing a significant role in modern portfolio management practices.

Effective index tracking technology not only plays a crucial role in enhancing the tracking accuracy of index products but also provides valuable insights for the design of benchmark portfolios in risk management and risk budgeting. Therefore, exploring ways to replicate an index with fewer stocks, reduce tracking error and management costs, and even achieve performance superior to the index, is of great significance for developing new methodologies and providing decision-making references.

The CSI 300 Index consists of 300 securities that are large in scale, highly liquid, and representative, selected from the Shanghai and Shenzhen stock markets in China. It essentially includes the leading listed companies from various sub-sectors in the Shanghai and Shenzhen markets, offering a comprehensive reflection of the overall performance of large-cap leaders. From

2005 to 2022, the average dividend yield of the CSI 300 Index was approximately 2.11%. In the ten years from August 2013 to July 2023, the annualized return of the CSI 300 Total Return Index reached 8.78%. It effectively mirrors the pace of China's macroeconomic development and the structural dividends in China's high-quality economic growth. It serves as a powerful tool to partake in the developmental achievements of top-notch leading companies, affirming the index's long-term investment value. Therefore, we will construct a portfolio to track the CSI 300 Index, offering the foundation for devising investment strategies and effectively yielding profits[1].

2. Research Methodology

We aim to use a deep learning framework to optimize index tracking portfolio problems. The neural network model we use, which includes fixed noise, is used to optimize the portfolio weight of each stock to achieve accurate tracking of the benchmark index [2].

Specifically, we can use the softmax probability generated by trained models to construct investment portfolios. And use specially designed neural networks to directly optimize portfolio weights. In addition, fixed noise is introduced. In the absence of fixed noise, the portfolio weight will change daily during the training process, which violates the assumption that the portfolio weight should remain unchanged until the rebalancing cycle arrives. By introducing fixed noise, even if the input data changes

during the training process, the portfolio weight of each stock will remain unchanged. This is crucial for solving the optimization problem of index tracking investment portfolios.

In full replication and partial replication, the input of the neural network, Deep NNF, is the cumulative change of each stock, and the output is the cumulative change of the constructed portfolio. The neural network is trained to minimize the difference between the cumulative change of the constructed portfolio and the cumulative change of the benchmark index.

For full replication, the portfolio is constructed from the softmax layer in the neural network. The output of softmax layer is used as the portfolio weight of each stock. When the rebalancing time arrives, the neural network is trained again with more recent data to rebalance the portfolio. For partial replication, the specific number of stocks are selected by the proposed model and the portfolio weights are optimized so that the portfolio can track the index. The given number of stocks are selected in descending order based on the portfolio weights of the neural network. The neural network is trained with the chosen companies again and the final portfolio is constructed[3].

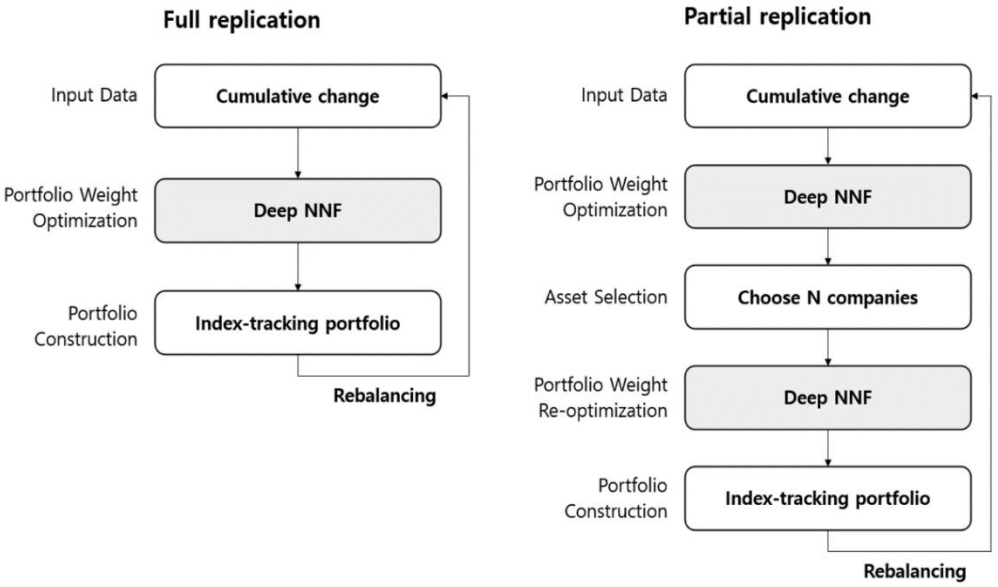

Table 1 The process of full replication and partial replication

We will adopt the following steps to conduct our research:

Data Collection and Preprocessing: Firstly, we will gather daily trading data of the past five years for the CSI 300 Index and all its constituent stocks. This data includes opening price, closing price, highest price, lowest price, and trading volume, among others. Then, we will preprocess these data, including data cleaning, handling missing values, and data normalization.

Stock Selection with Deep Autoencoder: We will use a deep autoencoder to train on the aforementioned data. By analyzing the output of the autoencoder, we will select a subset of stocks from the constituents of the CSI 300 Index, whose behavior patterns most represent the behavior patterns of the whole CSI 300 Index.

Index Replication with Deep Neural Networks: Having selected our representative stocks, we will use deep neural networks to learn these stocks' behavior patterns and attempt to replicate the behavior patterns of the CSI 300 Index[4]. We will train our model using historical data and test it with the most recent year's data.

Dynamic Weight Calculation: We will assign dynamic weights to each selected stock based on its importance in the model. These weights will be used in our index replication strategy.

Performance Evaluation: We will evaluate the performance of our strategy by calculating the tracking error between the replicated index and the actual CSI 300 Index. We will also compare the performance of our strategy with other index replication strategies.

3. Expected Outcomes

Effective Stock Selection: We expect that through the deep autoencoder, we can select a subset of stocks from all constituents of the CSI 300 Index that best represent the behavior patterns of the entire index.

Accurate Index Replication: We expect that through deep neural networks and dynamic weight calculation, our strategy can

accurately replicate the behavior patterns of the CSI 300 Index.

Lower Tracking Error: We expect the tracking error of our strategy to be lower than other index replication strategies, indicating better performance of our strategy.

Practical Investment Strategy: Lastly, we hope our research can provide investors with a practical investment strategy, helping them replicate the performance of the CSI 300 Index with lower costs and risks.

4. Timeline

From November 2nd to December 7th, we will divide our tasks into five weeks:

November 2nd - December 7th (5 weeks)

Week 1 (November 2nd - November 9th)

Data preprocessing and data inspection

Identify potential flaws and omissions in the data

Begin constructing the project dataset

Week 2 (November 10th - November 17th)

Model fitting

Select the appropriate machine learning/deep learning model

Commence model training

Collect model performance data

Week 3 (November 18th - November 25th)

Code adjustments and model optimization

Optimize model performance

Fine-tune the code for efficiency and readability

Document improvements to the code and model

Week 4 (November 26th - December 3rd)

Report writing

Write the formal project report

Include sections on project background, methodology, results, and discussion

Perform editing and proofreading

Week 5 (December 4th - December 7th)

Create a 10-minute presentation (PPT)

Prepare a presentation with slides

Write a presentation script

Prepare all materials required for the final submission

December 7th - Submission Deadline (December 14th)

Complete and submit the project

Finish all tasks before the submission deadline

Conduct a final review and check to ensure everything is in order

References

[1] ETF apartment Huatai-Pinebridge Fund Management Co., Ltd. 2023. "In the middle of 2023, look again at the CSI 300." https://caifuhao.eastmoney.com/news/20230803093203306219160

[2] Kwak, Yuyeong, Junho Song, and Hongchul Lee. 2021. "Neural Network with Fixed Noise for Index-Tracking Portfolio Optimization," Expert systems with applications 183, 183 (November). New York: Elsevier Ltd: 115298. doi:10.1016/j.eswa.2021.115298.

[3] Savitri, Nadia Astria, and Deni Saepudin. 2022. "Deep Learning Implementation for Portfolio Optimization Index Tracking LQ45." IEEE. doi:10.1109/icoseit55604.2022.10029954.

[4] Ouyang, Hongbing, Xiaowei Zhang, and Hongju Yan. 2019. "Index Tracking Based on Deep Neural Network," Cognitive systems research 57, 57 (October). Elsevier B.V: 107—14. doi:10.1016/j.cogsys.2018.10.022.

[5] Yu, Huanhuan, Rongda Chen, and Guoping Zhang. 2014. "A SVM Stock Selection Model within PCA," Procedia Computer Science 31, 31. Elsevier BV: 406—12. doi:10.1016/j.procs.2014.05.284.