Deep Learning final project:
Indexing with deep portfolios

mattia pividori (s284690)

Course of AA 2022-2023 - Data Science and Scientific Computing

1 Introduction

As per a recent research published by the U.S. Fed in May 2020, over the past couple of decades there has been a substantial shift in the global asset management industry from active to passive investment strategies: to provide an idea of the phenomenon through a few figures, Fed's estimates are that -as of March 2020- in the U.S. passive funds made up 48% of the AUM in equity funds and 30% for bond funds, whereas both shares were less than 5% in 1995.

Passive Investing refers to investment strategies wherein investors usually have a long-term goal and invest accordingly in instruments that try to mirror the performances of major stock indexes instead of trying to beat them (which is the aim of *Active Investing*). Mirroring an index means basically the expectation to enjoy the same return and volatility of the target index chosen (i.e. the 'benchmark'), while bearing risk and management fees lower than the active strategies, and a generally good level of liquidity. This is done mainly in three ways: via Index Funds, Exchange-Traded Funds (ETFs) and Direct Equity (i.e. the investor directly buys stocks and makes the necessary changes in the weights of each asset so as to ensure the best replication). All those three options share at their core a strategy known as *Indexing* (also *Index Investing*), a passive form of investment where portfolio's holdings mimic as closely as possible the holdings of the chosen benchmark index and only change when the composition of the underlying index changes (i.e. defined as 'rebalancing').

As a consequence, within *Indexing*, the method used to build a tracking portfolio is vital and, besides that, it is even more important as today index-linked products are ubiquitous in the financial industry: think for instance to futures, swaps, options, annuities, life-insurances, CDs and structured products as well as all hedging or arbitrage strategies which use all -one way or another- indices as their basis.

An *index* (also *market-index*) is a proxy of an underlying market through a market capitalization weighted sum of a selected collection of assets belonging to that market and is created to track alternatively the performance of a broad market (e.g. S&P 500), a market segment (e.g. sector indices such as health-care or technology sector, or Country indices), an investment strategy (e.g. ESG, infrastructure, or Factors' strategies) or an asset class (e.g. Fixed Income indices or Commodity indices).

In particular, the main index-tracking strategies currently employed in theory and practice are: **(i)** to own all the constituent securities of the target in the same proportions of the original index (*full replication*), **(ii)** to hold a representative sample (*partial replication*) of the securities in the index, **(iii)** to synthetically replicate the index via the use of derivatives (replication of index performance by means of linking investment return or strike price to it)[5]. Among the pros of full replication we include its consistency with the benchmark index, low tracking errors and a general transparency and ease of understanding, while many are the cons such as the low liquidity of smaller components, the associated high transaction costs and the difficulty in handling the tracking basket due to its large size and frequent changes in weights. On the other hand, the partial replication method is generally more easy to implement as fewer securities (less costs) are needed to achieve the purpose of indexing, but there is no consensus on how to measure the approximation of the replica portfolio to the market index, as it can vary with varying portfolio's objectives. The most common measures are the standard deviation of the active returns, the square root of the second-order moment of the active returns, Mean Absolute Deviation (MAD), Maximum Absolute Deviation (Max), Mean Absolute Downside Deviation (MADD), Downside Maximum Absolute Deviation (DMax) [8]. Moreover, even if the common goal of index tracking is to obtain the weights of tracker's constituents from the minimization of one (or more) of the above-mentioned measures into an objective function (so to end up with a ex-post return of the replica portfolio as close as possible to the performance of the benchmark index), and all are usually constructed from historical data of benchmark's constituents (e.g. correlation coefficients, market-values, weights, liquidity, etc..), many different techniques are employed like simple ranking, random sampling, stratified sampling, linear and quadratic programming, and more recently machine learning approaches such as genetic algorithms or evolutionary heuristics (literature references in [5][6][8]). This leads to select different subsets of the target indices or different weights for the same assets (in short, different optimal replicas).

More recently, a portfolio construction approach for partial replication based on deep learning has been proposed by Heaton et al. [4] in an attempt to exploit more efficiently historical information about constituents, target indices and correlations between them [7]. The main idea is to select stocks by measuring the Euclidean distance between original and reconstructed returns of benchmark's components using autoencoders, so as to better capture the non-stationary and non-linearity of financial stocks time-series, as well as to reflect the non-linear interactions between stocks with the aim to reduce the overall tracking error [3].

Further variations of this concept have subsequently been developed, experimenting on the S&P500 [1], expanding the framework with a dynamic asset-weight calculation method and applying it to HSI index [6], focusing on dimension estimation and reduction applied on the Russell 3000 index [2], trying various less and more complex autoencoders' structures of recent development on the FTSE 100, S&P 500 and HSI indices [5] and on the CSI 300 Index [8].

2 PROVVISORIO: Abstract

We investigate the use of deep learning (DL) algorithms for tracking stock indexes. These algorithms have the potential to outperform traditional financial models because of their ability to extract valuable hidden information. We replicate the IBB Index by optimizing for the difference between the returns of the tracking portfolio and the target index. We use auto-encoders (AE), successful in other contexts, to identify the complex non-linear relationships between index constituents and select a subset of them to construct a tracking portfolio. Literature results indicate that AE-based strategies are more effective than conventional ones when the tracking portfolio is made up of a small number of stocks. We would like to verify if DL algorithms with hierarchical architectures are really effective in index-tracking problems, identifying and taking advantage of data interactions currently not visible to any current financial theory. Finally, we also attempt to replicate the '1% problem' described in Heaton et al., modifying the target data by replacing all returns below -5% during the calibration phase, in an effort to create an index tracker that exhibits negative correlation during periods of significant declines.

References

[1] Saurabh Aggarwal and Somya Aggarwal. "Deep investment in financial markets using deep learning models". In: *International Journal of Computer Applications* 162.2 (2017), pp. 40–43.

[2] Nitish Bahadur, Randy Paffenroth, and Kelum Gajamannage. "Dimension Estimation of Equity Markets". In: *2019 IEEE International Conference on Big Data (Big Data)*. IEEE. 2019, pp. 5491–5498.

[3] Catherine S Forbes and Worapree Maneesoonthorn. "Discussion on 'deep learning for finance: deep portfolios'". In: *Applied Stochastic Models in Business and Industry* 33.1 (2017), pp. 33–15.

[4] James B. Heaton, Nick G. Polson, and Jan Hendrik Witte. "Deep learning for finance: deep portfolios". In: *Applied Stochastic Models in Business and Industry* 33.1 (2017), pp. 3–12.

[5] Saejoon Kim and Soong Kim. "Index tracking through deep latent representation learning". In: *Quantitative Finance* 20.4 (2020), pp. 639–652.

[6] Hongbing Ouyang, Xiaowei Zhang, and Hongju Yan. "Index tracking based on deep neural network". In: *Cognitive Systems Research* 57 (2019), pp. 107–114.

[7] Vadim Sokolov. "Discussion of 'deep learning for finance: deep portfolios'". In: *Applied Stochastic Models in Business and Industry* 33.1 (2017), pp. 16–18.

[8] Chi Zhang et al. "Stock-index tracking optimization using auto-encoders". In: *Frontiers in Physics* 8 (2020), p. 388.