Methods for Solving Problems in Financial Portfolio Construction, Index Tracking and Enhanced Indexation

A thesis submitted for the degree of

Doctor of Philosophy

by

Hakim Mezali

Supervised by

Prof. John Beasley

Department Mathematical Sciences
School of Information Systems, Computing and Mathematics
Brunel University, London

June 2013

Abstract

The focus of this thesis is on index tracking that aims to replicate the movements of an index of a specific financial market. It is a form of passive portfolio (fund) management that attempts to mirror the performance of a specific index and generate returns that are equal to those of the index, but without purchasing all of the stocks that make up the index.

Additionally, we consider the problem of out-performing the index - Enhanced Indexation. It attempts to generate modest excess returns compared to the index. Enhanced indexation is related to index tracking in that it is a relative return strategy. One seeks a portfolio that will achieve more than the return given by the index (excess return).

In the first approach, we propose two models for the objective function associated with choice of a tracking portfolio, namely; minimise the maximum absolute difference between the tracking portfolio return and index return and minimise the average of the absolute differences between tracking portfolio return and index return. We illustrate and investigate the performance of our models from two perspectives; namely, under the exclusion and inclusion of fixed and variable costs associated with buying or selling each stock.

The second approach studied is that of using Quantile regression for both index tracking and enhanced indexation. We present a mixed-integer linear programming of these problems based on quantile regression.

The third approach considered is on quantifying the level of uncertainty associated with the portfolio selected. The quantification of uncertainty is of importance as this provides investors with an indication of the degree of risk that can be expected as a result of holding the selected portfolio over the holding period. Here a bootstrap approach is employed to quantify the uncertainty of the portfolio selected from our quantile regression model.

Certificate of Originality

"I hereby certify that the work presented in this thesis is my original research and has not been presented for a higher degree at any other university or institute".

....

Hakim Mezali

Acknowledgements

The completion of the this thesis owes sincere gratitude to a number of individuals who have been instrumental towards its completion in one way or another.

I am heavily indebted to my supervisor Prof. John Beasley for his invaluable guidance, support and advise throughout the years of research. I am particularly grateful for his patience and humility in his dealings with me, especially during times in which I experienced difficulties in expressing myself as fluently as would have liked.

I am grateful for having my research financially supported by CARISMA and I would like to acknowledge my gratitude to the Department of Mathematics, for making Brunel such a great working environment and for giving me opportunities to teach throughout my studies.

I would also like to extend sincere gratitude to Dr. Abdallah K. Ally for introducing me to the world of statistics. Special thanks goes to my dear friends in Brunel and outside. I would like to thank Rahim Al-Hamzawi, Ali Al-Kenani, Fatmir Qirezi and Dr. Nafiu Hussaini.

Last, but by no means least, sincere and heartfelt gratitude goes out to my family, my wife Amel and my two lovely kids Younes and Iman for their unwavering and perpetual support during times of ease and hardships.

Author's Publication

1. **Mezali, H**. and Beasley, J. (2011). Index tracking with fixed and variable transaction cost, *Optimization Letters*, 63, 17-19.
2. **Mezali, H**., and Beasley, J. (2012) Quantile regression for index tracking and enhanced indexation, *Journal of Operational Research Society*, 64, 1676-1692.
3. **Mezali, H**., Ally, A. and Beasley, J. (2013). Bootstrap approach to quantifying uncertainty of index tracking and enhanced indexation, *working paper*.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Research definition

In recent years passive fund investment strategies have become very popular, especially amongst mutual fund managers and pension funds. These strategies are often adopted by investors who believe that financial markets are efficient. Such strategies involve building an investment portfolio designed to track a particular benchmark stock/equity index (such as the FTSE 100 in London or the S&P 500 in New York). Index tracking is often referred to as a passive investment strategy and can be contrasted with active management, which typically involves frequent trading in the hope of outperforming a relevant market benchmark.

When the objective is to track the index, a practical alternative for an investor is to invest in an index fund, rather that the investor purchasing all of the individual index stocks. This is a fund that attempts to mirror the performance of a specific index and generate returns that are equal to those of the index. Since portfolio decisions are automatically (algorithmically) made, and transactions are infrequent, expenses tend to be lower than those of actively managed funds. Common criteria imposed on tracking funds such as these are that they should achieve approximately the same returns as a specified market index through investment in an appropriately selected set of stocks from the index.

The simplest case of passive management is the index fund that is designed to exactly replicate a well defined stock index. If a fund invests in all of the stocks in the index in such a way that its investment in each stock mirrors index composition (e.g. if a stock makes up 10% of the index then it makes up 10% of the investment) then the fund is said to be following a full (or complete) replication strategy. Although exact replication is the simplest technique for constructing an index fund, many index funds are not constructed in this way.

Full replication is possible for constructing an index fund, however, as the number of stocks in the index grows it can be an expensive strategy in terms of transaction cost (e.g. see Beasley et al. (2003)). This is because, stocks typically enter and leave the index at regular intervals and as a consequence the entire fund must be rebalanced as this occurs to mirror the index as it changes. Because of these disadvantages, many passively managed funds, especially those that are tracking large indices, hold fewer stocks than are included in the index they are tracking

In this thesis we do not adopt full replication, **in essence we view the index tracking problem as a decision problem, namely to decide the subset of stocks to choose so as to mirror or reproduce the performance of the index over time. We call the subset of stocks we choose a tracking portfolio**.

Enhanced indexation (sometimes referred to as enhanced index tracking) is related to index tracking in that it is a relative return strategy. One seeks a portfolio that will achieve more than the return given by the index. Here the aim is often to achieve returns that are only slightly above what the index itself returns. This is a relative return strategy since if we want a return 2% above that of the index, and the index falls by 10% in the year, then a portfolio that falls by only 8% is appropriate. Note here how in such a relative return strategy we do not attempt to ensure that we achieve a positive return. Constructing an enhanced indexation fund (portfolio), can be accomplished using a mathematical model that is closely related to a model for index tracking.

The objective of this thesis is to contribute to the development of efficient and effective portfolio selection algorithms. We present methods for solving problems in financial portfolio construction, index tracking and enhanced indexation. Our formulations are mixed-integer linear programs for index tracking and enhanced indexation.

The formulations proposed include transaction costs, a constraint limiting the number of stocks that can be in the portfolio and a limit on the total transaction cost that can be incurred. Numerical results are presented for eight test problems drawn from major world markets, where the largest of these test problems involves over 2000 stocks.

1.2 Thesis Outline

The outline of the thesis is as follows. Chapter 2 presents a literature survey relating to index tracking and enhanced indexation.

Chapter 3 presents our two mixed-integer formulations of the index tracking problem. In particular we explicitly consider both fixed and variable transaction costs (the fee associated with trading) and limit the total transaction cost that can be incurred. In this chapter we propose two approaches for the objective function associated with choice of a tracking portfolio, namely; minimise the maximum absolute difference between the tracking portfolio return and index return and minimise the average of the absolute differences between tracking portfolio return and index return. Our formulations are based upon tracking an index by comparing the returns from the index with the returns from the tracking portfolio.

Chapter 4 applies Quantile Regression to two problems in financial portfolio construction, index tracking and enhanced indexation. Quantile regression differs from traditional least-square regression in that one constructs regression lines for the quantiles of the dependent variable in terms of the independent variable. In this approach we apply quantile regression, as first defined by Koenker and Bassett (1978).

Chapter 5 focuses on quantifying the level of uncertainty associated with portfolio selection. In index tracking and enhanced indexation the quantification of uncertainty is of importance as this provides investors with an indication of the degree of risk that can be expected as a result of holding the portfolio selected over the holding period. In this chapter a bootstrap approach is employed to quantify the uncertainty of the portfolio selected from our quantile regression model.

Finally, Chapter 6 summarises the main results of the research, highlighting the contribution to knowledge we have made as well as proposing recommendations for possible future research directions.

Chapter 2

Literature review

Index tracking involves building an investment portfolio designed to track a particular benchmark index. At its simplest, it requires holding all stocks in the index, and weighting each stock-holding so each investment is held in proportion to its contribution to the index being tracked. If this is done, the index fund is said to be following a full replication strategy. Full replication is possible but as the number of stocks in the index grows it can be an expensive strategy in terms of transaction costs.

Then in essence we can view the index tracking as the problem of reproducing the performance of a stock market index over time, but without purchasing all of the stocks that make up the index. It is a decision problem, namely to decide the subset of stocks to choose so as to (hopefully perfectly) mirror the performance of the index over time.

Enhanced indexation deals with the situation where we want to both track the index (so getting the market return), but also want to out-perform the index. For example we might want a stock (equity) portfolio that exceeds the return on a specified index by 2% per year. This can be accomplished using mathematical models that are closely related to index tracking models.

2.1 Introduction

Despite the increasing popularity of passive investment strategies, the attention given in the academic literature to implementation and to algorithmic problems arising in the process of index tracking and enhanced indexation is still relatively small compared to the numerous articles dedicated to the classical problem of portfolio risk and return optimisation. In our literature survey below we discuss papers relating to index tracking and enhanced indexation.

In this chapter we first present investment preliminaries and the historical and practical context behind financial portfolio optimisation. Then we discuss previous studies in the literature relating to constructing portfolios for index tracking and enhanced indexation. In general, algorithms for index tracking can often be extended with only minor modifications to deal with enhanced indexation (or both). However, for simplicity, we survey index tracking and enhanced indexation separately below.

We organize this chapter in the following way. Investment preliminaries are examined in Section 2. Historical and practical context is considered in Section 3. The literature survey for index tracking and enhanced indexation is presented in Sections 4 and 5. The chapter concludes with a summary in Section 6.

2.2 Investment preliminaries

2.2.1 Market index

A market index shows the movement of a particular market as a whole, revealing if the total value (i.e. the market capitalisation) of all companies listed has increased or decreased. Indices are calculated on an entire market as well as being available for a particular sector of a market. The FTSE 100 share index, for example is made up of the 100 largest UK registered companies in terms of their market capitalisation.

Additionally, a market index is a series of pure numbers and is used for making comparison between different index numbers and for following the fortunes of particular sample groups. Index numbers are constructed with a fixed base date and base value.

The key point that we need to grasp is that returns are relative. As an example, if we make an investment in a single stock and that stock goes up (increases in price) by 5% over the year we might at first sight be happy. But suppose we then learn that the market (say as represented by an equity index like the Dow Jones or S&P 500) has risen by 10% that year. Obviously the increase of 5% on the stock does not appear as attractive as it did at first sight.

In a discrete time manner we can calculate return on an investment as:

return(%)=100(change in value)/(original value)

However, in quantitative finance we almost always use a different measure of return calculated as:

return(%)=100* $\log_e$[**(new value)/(original value)**]

This is sometimes referred to as continuous time return. Under continuous time return, if we are earning interest on an initial investment of A at fractional interest rate of r for t years, we will have at the end of the period a sum equal to Ae^{rt}.

Some indices are of fixed cardinality (number of stocks/companies in the index fixed), some are not. For example the S&P500 is a fixed cardinality index with precisely 500 stocks (companies) in the index, the Wilshire 5000 is not. The composition of all indices changes over time (as the underlying companies change, some cease to exist, others grow large enough to warrant inclusion in the index).

Consider the data shown in Table 2.1, the index value quoted, 6229.80 is calculated as: sum over the companies in the index of number of shares issued multiplied by current price = total worth of the company (total market capitalisation, market cap) divided by a large constant value to turn the answer into a meaningful number. Most equity indices are calculated in this way but not all.

To illustrate Table 2.1 index value is useful for investors to track changes in

Table 2.1: Example of the FTSE 100 index values

Index	Values
Index Value	6,229.80
Trade Time	4:36PM
Change	43.20 (0.70%)
Prev Close	6,186.60
Open	6,186.60
Day's Range	6,186.60-6,229.80

market values over period of time. For this example, the FTSE 100 index is made up of the 100 largest UK-based companies and is computed by combining 100 stocks together into one index value. Investors can track changes in the index's value over time and use it as a benchmark against which to compare their own portfolio returns.

From the Table 2.1 we notice that the index opening value is 6,186.60 and is equal to the previous day's closing price. However, the index value at the trade time finished higher than the open value. A measurement of change in the index value over a period of time as in this example is 43.20, which increased by (0.70%) compared to the index opening value. The Day's range is the value range (low - high) in the latest trading day. In other words it is the difference between the highest and the lowest index value of a set time period.

2.2.2 Diversification

Portfolios with only a few assets may be subject to a high degree of risk, represented by a relatively large variance in return. As a general rule, the variance of the return of a portfolio can be reduced by including additional assets in the portfolio, a process referred to as diversification. As an investor, we could invest our entire wealth in one stock (i.e company). If we do so, we are exposed to both company-specific risk and market risk. However, if we expand our portfolio to include other assets or stocks, we are diversifying, and by doing so, we can reduce our exposure to company-specific risk.

For this purpose, when we make an investment in a single stock, we are exposed to market risk, by choosing to invest in the market (such as the FTSE 100 in London or the S&P 500 in New York) where the stock is traded; and stock risk, the individual stock may do better or worse than the average, (e.g. the change in the market) as represented by the change in the index. As a consequence the key concept in stock investment is diversification (not putting all of your stock in one market). This helps to reduce risk (by spreading, and hopefully reducing, your stock risk). However, assuming we invest in just one market, (such as the FTSE 100 in London) we are still exposed to market risk. If we choose to invest in more than one market (such as the FTSE 100 in London or the S&P 500 in New York) we may reduce market risk.

Although we can reduce risk we do not know the future and irrespective of how we choose our stock portfolio we are taking risk. What we hope is that by using past data (for example in relation to stock prices) in a systematic and mathematical fashion we can make better portfolio decisions.

2.2.3 Transaction cost

When we consider investing in stocks, also known as equities and shares, we will have to pay a transaction cost associated with buying or selling stock. Transaction costs are often given in basis points, **one basis point (bp)** is 1/100 of one percent. As an example of a transaction cost we typically need to pay some commission to an intermediary third party if we decide to buy (or sell) one unit of a stock. Such transaction costs vary by stock, typically according to how liquid (easily bought/sold) the stock is and by how much we wish to trade (e.g. number of units of the stock we are buying (or selling)).

The portfolio construction problem has one common feature, the trade-off between gaining a better position by rebalancing the position on one hand and the occurrence of additional transaction costs as a consequence of such an action on the other hand. The principle to decide on trading or not can be formulated as: trade only if the gain from trading pays the transaction costs.

2.3 Historical and practical context

Ever since the pioneering work of Markowitz (1952) optimisation has been at the center of work concerned with decisions relating to deciding the composition of financial portfolios. As such both practitioners and academic researchers have been willing to tradeoff the disadvantages of optimisation (multiple optimal solutions, solution sensitivity) for its advantages (clear modelling framework, computational efficiency, algorithmic decision-making).

2.3.1 Historical context

Indexing was initially made available to institutional investors in 1971, individual investors were able to easily invest in an index tracking fund when the Vanguard 500 index fund made its debut in 1976.

In the last three decades, index funds have gradually increased their share of the overall market; not only for an individual investor's saving but also for institutional funds such as pension and insurance funds. In 2006, more than $120 billion of individual investor savings were invested in indexed mutual funds, and institutional investors contributed several hundred billion more to institutional index funds (Damodaran (2012)).

As a consequence as index funds have grown, the choices have also proliferated. While the first few funds all indexed themselves to the S&P 500, we now see funds indexed to almost every conceivable index. Most of these funds are sampled funds rather than full replication funds and we call the sample of stocks we choose a **tracking portfolio**.

The earliest approach to solving the portfolio problem is the so called mean-variance approach. It was pioneered by H. Markowitz (see Markowitz (1952)) and is only suitable for one-period decision problems. It consists of a one-off decision at the beginning of the period ($t = 0$) and no further actions until end of the period ($t = T$). It still has great importance in real-life applications and is widely applied in risk management.

Prior to Markowitz's work, investors focused on assessing the risks and returns of individual stocks in constructing their portfolios. Markowitz proposed that investors should focus on selecting portfolios based on their overall risk-return characteristics instead of merely constructing portfolios from stocks that each individually have attractive risk-return characteristics, in other words investors should select portfolios not individual stocks.

Markowitz's mean-variance portfolio optimisation model employs variance as the measure of risk and the objective of the model is to find the weighting of the stocks that minimise the variance of a portfolio and give a desired expected return. To proceed with the explanation of Markowitz mean-variance portfolio optimisation we need some notation, let:

N be the number of assets (e.g. stocks/equities) available for an investment

u_i be the expected average return of asset i

ρ_{ij} be the correlation between the return for asset i and j ($-1 \leq \rho_{ij} \leq +1$)

s_i be the standard deviation in return for asset i

R be the desired expected return from the portfolio chosen

Then the decision variable are:

w_i the proportion of the total investment associated with (invested in) asset i ($0 \leq w_i \leq 1$)

Using the standard Markowitz mean-variance approach we have that the portfolio optimisation problem is:

$$\text{minimise} \quad \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j \rho_{ij} s_i s_j \tag{2.1}$$

subject to

$$\sum_{i=1}^{N} w_i u_i = R \tag{2.2}$$

$$\sum_{i=1}^{N} w_i = 1 \tag{2.3}$$

$$(0 \leq w_i \leq 1) \qquad i = 1, \ldots, N \tag{2.4}$$

Equation (2.1) minimises the total variance (risk) associated with the portfolio. This equation is sometimes written as $\sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j \sigma_{ij}$, as it can be expressed in terms of σ_{ij} the covariance between the returns associated with assets (i) and (j) since the covariance $\sigma_{ij} = \rho_{ij} s_i s_j$.

Equation (2.2) is the expected rate of return of the portfolio; it is found by taking the weighted sum of the individual rates of return. Equation (2.3) ensures that the weight proportions add to one. Equation (2.4) is the non-negativity constraint. This formulation (equations (2.1) - (2.4)) is a simple nonlinear programming problem. As the objective is quadratic, computationally effective algorithms exist to calculate the optimal solution for any particular data set. Note here that above we have, for a given return, found the minimum risk portfolio. Logically we could have specified the risk we were prepared to take and found the maximum return portfolio that had this specified risk. Whilst this is a logical equivalent the way presented above is the way we proceed in numeric practice. Numerically finding a minimum risk portfolio that has a specified return is much easier than finding a maximum return portfolio that has a specified risk.

Building on the work of Markowitz and to overcome the limitations and problems raised by modelling the portfolio in a discrete time setting; a continuous-time approach for modelling the stock prices and the actions of the investors was proposed by Merton (1971). It must be regarded as the real starting point of continuous-time portfolio theory. By applying standard methods and results from stochastic control theory to the portfolio problem he was able to obtain explicit solutions for some special examples. However, the crucial point in his approach is that the whole problem reduces to solving the Hamilton-Jacobi-Bellman equation of dynamic programming. This typically leads to the problem of solving a highly non-linear partial differential equation for which even a numerical solution may prove elusive. Despite these limitations, the Merton (1971) approach is still popular in finance.

With the growing application of stochastic approaches to finance in the early 1980's Harrison and Kreps (1979), Harrison and Pliska (1981) and Karatzas (1989), introduce the martingale approach to portfolio optimisation. It is based on results of stochastic calculus and on convex optimisation.

The most significant improvements in continuous-time models have been the introduction of additional constraints and of transaction costs to the portfolio problem. The work on constraints can be divided into work concerning constraints both on the trading strategies and on the wealth of an investor. Typical constraints on the strategies include short selling and leverage constraints, bounds for the wealth held in one asset or incomplete market constraints (see for example (Cox and Huang (1991), Cvitani and Karatzas (1992) and Xu and Shreve (1992)). Moreover, as rebalancing of the holdings is the essential action of an investor solving the portfolio problem, transaction costs and their impact on the form of the optimal strategy cannot be ignored. Magill and Constantinides (1976) was amongst the first papers dealing with the process of rebalancing a portfolio taking into consideration transaction costs.

The risk and return model that has been in use the longest and is still the standard in most real world analysis is the capital asset pricing model (CAPM). The standard form of the equilibrium relationship for asset returns was developed independently by Sharpe (1964) who formalised the capital asset pricing model (CAPM) and parallel work was also performed by Lintner (1965), which follows logically from the Markowitz mean-variance portfolio theory as described above.

The capital asset pricing model (CAPM) was formulated to show how the expected return on an asset could be related to its risk, while at the same time providing a precise definition of the meaning of risk. What would be expected in terms of risk is that a portfolio made up of one asset is likely to be more volatile than a portfolio made up of a range of assets (a diversified portfolio). Investors could therefore lower their risks, in particular company-specific risks, by purchasing a diversified portfolio of assets. This approach may reduce company risk, but the overall equity market risk still exits. Therefore, every asset is made up of two elements of risk, one related to the market and the other related to the company.

As stated by CAPM, the expected return of an asset equals the risk-free rate plus the assets's beta multiplied by the expected excess return of the market portfolio. Specifically, let R_i and R_m be random variables for the simple returns of the stock and the market over some specified period. Let R_f be the known risk-free rate, also expressed as a simple return, and we obtain the capital asset pricing model in the form

$$E(R_i) = R_f + \beta_i(E(R_m) - R_f) \tag{2.5}$$

This is the form in which it is most often written where:

$E(R_i)$ is the expected return on the asset

β_i (the beta) is the sensitivity of the expected excess asset returns to the expected excess market returns, where $\beta_i = Cov(R_i, R_m)/Var(R_m)$

$E(R_m)$ is the expected return of the market

$E(R_m) - R_f$ is known as the market premium (the difference between the expected market rate of return and the risk-free rate of return)

Restated, in terms of risk premium, which states that the individual risk premium equals the market premium times β_i we find that:

$$E(R_i) - R_f = \beta_i(E(R_m) - R_f) \tag{2.6}$$

The CAPM assumes that:

- There are no transaction costs and that everyone has access to the same information. Making these assumptions allows investors to keep diversifying without additional cost.
- An individual investor cannot affect the price of a stock by his buying or selling action.
- Investors are expected to make decisions solely in terms of expected values and standard deviations of the returns on their portfolios.
- One other assumption deals with homogeneity of expectation. First, investors are assumed to be concerned with the mean and variance of returns (or prices over a single period), and all investors are assumed to define the relevant period in exactly the same manner. Second, all investors are assumed to have identical expectations with respect to the necessary inputs to the portfolio decision.

To this end, many assumptions behind the capital asset pricing model may be untenable.

2.3.2 Portfolio Management Strategies

Portfolio fund management strategies refer to the approaches that are applied in order to generate the highest possible returns at lowest possible risks. Portfolio management involves a series of decisions and actions that are made by the investor, whether individual or institution. Portfolios must be managed whether investors follow a passive approach to selecting and holding their financial assets, or an active approach.

Active fund management strategy (also called active investing) refers to a portfolio management strategy that involves making precise investments for outperforming an investment benchmark index. The portfolio manager that follows the active management strategy exploits market inefficiencies by buying undervalued stocks or securities or by short selling overvalued securities. Any of these procedures can be used alone or in combination.

This active approach to portfolio management involves managers observing the market as a whole and deciding about the industries and sectors that are expected to perform well in the ongoing economic cycle. After the decision is made on the sectors, the specific stocks are selected on the basis of companies that are expected to perform well in that particular sector.

Passive fund management (also called passive investing) is a financial strategy in which an investor (or a fund manager) invests in accordance with a pre-determined strategy that doesn't entail any forecasting (e.g. any use of market timing or stock picking would not qualify as passive management). The idea is to invest in an index fund that replicates as closely as possible the performance of a specified index benchmark. By tracking an index, an investment portfolio typically gets good diversification, low turnover (good for keeping down transaction costs), and extremely low management fees.

Active and passive fund management strategies have their respective advantages and disadvantages.

- The primary advantage of active management is that it allows portfolio managers to select a variety of investments rather than investing in the market as a whole, this is usually not the case in passive management.
- Secondly, in order to generate profits, the investors consider that some market segments are less efficient than others and also they manage the volatility or risks of market by investing in less-risky and high-quality companies.
- In addition, investors may take additional risk for achieving higher-than-market

returns and may follow a strategy for avoiding certain industries in comparison to the market as a whole.

The drawback of active management is the chance that bad investment choices are made by the fund manager. The costs related to active management are higher in comparison to passive management. Higher transaction costs due to frequent trading with active fund management strategies reduces the fund's return. In active management an investor is exposed to both market and company risk, whilst in passive management an investor is exposed to market risk.

In recent years passive management has been receiving a higher profile as an investment alternative. The simplest explanation for the difference in returns between actively managed funds and index funds is trading costs. Index tracking funds are inexpensive to create, to run and incur minimal transaction costs and management fees. By contrast, the trading costs and fees of actively managed funds are higher. **In this thesis we focus on passive fund management. It is essentially an algorithmic approach to investment decisions that are made in order to systematically reproduce the performance of an index (i.e. index tracking) or to generate excess return (i.e. enhanced indexation)**.

2.4 Index tracking survey

2.4.1 Index tracking model

Index tracking model involves building an investment portfolio designed to track a particular benchmark index over time. If a fund invests in all of the stocks in the index in such way that its investment in each stock mirrors index composition (e.g. if a stock makes up 10% of the index then it makes up 10% of the investment) then the fund is said to be following a full/complete replication strategy.

Full replication is possible, however as the number of stocks in the index grows it can be an expensive strategy in terms of transaction cost. This is because stock typically enter/leave the index at regular intervals and so the entire fund must be rebalanced as this occurs to mirror the index as it changes and any new money that is invested in (or money taken out) the fund must be spread across all stocks to mirror the index.

Then in essence we can view the index tracking problem as a decision problem, namely to decide the subset of stocks to choose so as to (hopefully perfectly) mirror/reproduce the performance of the index over time. We call the subset of stocks we choose a tracking portfolio.

Suppose that we observe over time 0,1,2,. . . ,T the value of N stocks, as well as the value of the index we want to track. Further suppose that we are interested in deciding the best set of K stocks to hold (where $K < N$), as well as their appropriate quantities. In index tracking we want to answer the question: "what will be the best set of K stocks to hold, as well as their appropriate quantities, so as to best track the index in the future. Our approach in index tracking is a historical look-back approach. To ask the historical question: "what would have been the best set of K stocks to have held, as well as their appropriate quantities, so as to have best tracked the index in the past (i.e. over the time period [0,T])?" and then hold the stocks that answer this question into the future. This idea forms the foundation of the methodology presented in the following chapters as it extends to enhanced indexation model as shall become apparent.

A significant number of papers relating to (equity) index tracking have been discussed both by academics and practitioners. In this section we present our literature review relating to the index tracking problem. Note also here that an extensive discussion as to metaheuristics for the index tracking problem has recently been given by di Tollo and Maringer (2009). Also Metaxiotis and Liagkouras (2012) have recently given a review of multiobjective evolutionary algorithms in portfolio optimisation and some of the papers they reference are concerned with index tracking.

2.4.2 Heuristic algorithms

The term heuristics refers to techniques based on experience for various tasks such as research, problem solving, discovery and learning. Heuristic methods enhance finding a desirable solution in conditions where a comprehensive search is unfeasible.

Heuristic methods are attractive because, while being a robust method for large practical portfolio problems, they are relatively independent of the objective function and offer solutions in a reasonable time. In this section, we give a review of some previous index tracking papers using heuristic methods namely: genetic algorithm, tabu search and simulated annealing.

Beasley et al. (2003), considered the problem of index tracking when transaction costs exist. In their formulation of the problem the total transaction cost and the number of stocks in the tracking portfolio are limited. They presented a population heuristic (PH) for the solution of the index tracking problem and used reduction tests in order to reduce the size of the search space, hence enabling the PH to be more effective. Computational results for the Hang Seng, DAX 100, FTSE 100, S&P 100 and Nikkei 225 indices were presented. Computation times varied between 1.7 and 285.4 minutes.

Maringer and Oyewumi (2007), presented a heuristic algorithm for index tracking based upon differential evolution (see Storn and Price (1997)), where the nonlinear objective relates to minimising the squared differences between tracking portfolio return and index return. They do not consider transaction costs. Computational results are given for tracking the Dow Jones Industrial Average index (which contains 65 assets) over the period March 2000 to November 2006.

Maringer (2008), presented an approach where tracking portfolio deviations above index return are treated differently from tracking portfolio deviations below index return. He does not consider transaction costs and uses a heuristic based on differential evolution (see Storn and Price (1997)). Computational results are given for tracking the Dow Jones Industrial Average index (which contains 65 assets) over the period March 2000 to November 2006.

Krink et al. (2009) presented a model for index tracking where the nonlinear objective relates to minimising the squared differences between tracking portfolio return and index return. Although they do not consider transaction costs they do introduce a constraint on the change in the proportion invested in each asset at a rebalance. Their heuristic uses differential evolution (see Storn and Price (1997)), albeit modified with a number of different algorithmic components. Computational results are presented related to tracking the Nikkei 225 index (which contains 225 assets) over the period November 2005 to January 2007, as well as for the Dow Jones Industrial Average index (which contains 65 assets) over the period April 2002 to December 2003.

Li et al. (2011) presented a multi-objective model, where one objective relates to the minimisation of tracking error and the other objective relates to the maximisation of excess return (return over and above the index). Their model addresses both index tracking and enhanced indexation. Transaction costs are included in their approach and are explicitly limited. They solve their model using an immunity based heuristic, albeit modified to deal with the constraints in their model. Computational results are given for five publicly available test problems involving up to 225 assets.

Guastaroba and Speranza (2012) presented a model that includes fixed and variable transaction costs, as well as a constraint upon the total transaction cost incurred when rebalancing from an existing portfolio. They also constrain the maximum number of assets that can be held in the tracking portfolio. In their model they track an index by reference to the absolute deviation between a scaled index value and the tracking portfolio value, rather than by tracking the return on the index. An heuristic based upon kernel search (Angelelli et al., 2010, 2012) is presented. Computational results are given for eight data sets involving up to 2151 assets.

Murray and Shek (2012) proposed a local relaxation algorithm that explores the inherent structure of the objective function. It solves a sequence of small, local, quadratic-programs by first projecting asset returns onto a reduced metric space, followed by clustering in this space to identify sub-groups of assets that best accentuate a suitable measure of similarity amongst different assets. They used a heuristic method such as the centroid of initial clusters. Computational results, using two data sets consisting of 500 and 3,000 stocks in the US, spanning the period between January 2002 and January 2010 for the first and between May 2005 and April 2010 for the second data set and with different cardinality constraints. They compare the performance of their proposed algorithm against the commonly used heuristic of successive truncation, followed by a more in depth comparison of their algorithm against a leading commercial solver, CPLEX. They indicate that the proposed algorithm can lead to a significant performance gain over popular branch-and-cut methods and also the local relaxation heuristic method proposed is able to obtain a better solution than CPLEX.

2.4.3 Genetic algorithm

A Genetic Algorithm (GA) is a search heuristic that mimics the process of natural evolution. This heuristic is routinely used to generate useful solutions to optimisation and search problems. Genetic algorithms belong to the larger class of evolutionary algorithms (EA), which generate solutions to optimisation problems using techniques inspired by natural evolution, such as inheritance, mutation, selection, and crossover.

Jeurissen and van den Berg (2005) presented an index tracking approach using a hybrid genetic algorithm. They defined tracking error (the variance of the difference between the returns of the tracking portfolio and the index) as a measure of fitness. The weights associated with each stock in the tracking portfolio were decided by using a genetic algorithm. Computational results for 25 stocks from the Dutch AEX index were presented, however no computation times were reported.

Jeurissen and van den Berg (2008) presented an index tracking model based on minimising the variance of the difference between tracking portfolio return and index return. Their model, which is a quadratic program, does not consider transaction costs. They use a hybrid genetic algorithm and present computational results for tracking the Dutch AEX index (which involves 25 assets) using just 10 assets over a one year period from March 2004.

Van Montfort et al. (2008) presented a model for index tracking that has a quadratic objective based on minimising the squared differences between tracking portfolio return and index return. To deal with the computational difficulties that arise when they introduce binary decision variables into their model they present several different heuristics. Transaction costs are included in their approach and are explicitly limited. Computational results are presented for tracking the MSCI Europe index for one year from July 2004.

Ruiz-Torrubiano and Suarez (2009) presented an approach for index tracking based on using a genetic algorithm to decide the set of assets to be included in the tracking portfolio, with quadratic programming being used to decide the proportion invested in each of the chosen assets. They do not consider transaction costs. Computational results are given for five publicly available data sets involving up to 225 assets.

Liu et al. (2012) presented multi-period portfolio selection problems in a fuzzy environment by considering return, transaction cost, risk and skewness of portfolio to provide investors with additional choices. In their models, the return is characterized by the possibilistic mean value and the risk is measured by possibilistic variance. The skewness is quantified by the third order moment about the possibilistic mean value of a return distribution. To solve their models, they first present a TOPSIS-compromised programming approach to convert them into single objective programming models. Then, they design a genetic algorithm with a penalty term to solve their models. Computational results are given for four cases, using Chinese Stock data for the weekly closing prices of four risky assets from January 2001 to

January 2010.

Wang et al. (2012) presented a model for index tracking that minimises the mean absolute difference between tracking portfolio return and index return. They introduce a conditional value at risk constraint to control downside risk. Their model is a mixed-integer linear program which they solve (for one small problem involving 31 assets) using Cplex. For a larger problem with 89 assets they use the genetic algorithm of Ruiz-Torrubiano and Suarez (2009).

2.4.4 Markowitz models

Markowitz (1952) proposed the mean variance methodology for portfolio selection. It has served as a basis for the development of modern financial theory. Konno and Yamazaki (1991) used the mean absolute deviation risk function to replace the risk function in Markowitz's model to formulate a mean absolute deviation portfolio optimization model. Roll (1992) used the sum of the squared deviations of returns on a portfolio from the benchmark as the tracking error and proposed a mean variance index tracking portfolio selection model. So, it is possible to apply the standard Markowitz portfolio model to index tracking.

Rohweder (1998) presented a tracking error optimisation model which includes a term relating to transaction costs in the objective function. He also introduced an alternative technique to control the risk, portfolio segmentation, which does not require the estimation of covariances. His technique controls tracking error risk by dividing the portfolio into an active and a passive subportfolio. He presented simulation results for 200 European stocks.

Yu et al. (2006) presented a Markowitz model for index tracking where their approach assumes that index tracking relates to constraining the probability that the return from the tracking portfolio falls below index return (downside risk); or to higher order moments of downside risk. They assume that stock returns are jointly normally distributed and that short selling is allowed. They presented a small numeric example using stocks from the Hang Seng index.

Garcia et al. (2011) presented a paper arguing for consideration of frontier curvature when deciding an index tracking portfolio, where the frontier is the standard Markowitz frontier based on mean-variance analysis. Their model does not consider transaction costs. Although they present a solution algorithm (based on Tabata and Takeda (1995)) no computational results are given.

2.4.5 Other research papers

Ghandar et al. (2010) presented an evolutionary approach to designing a fuzzy rule based system for deciding tracking portfolio composition. Although transaction costs are considered when rebalancing occurs there is no explicit constraint on the transaction cost incurred. They comment that their rules can be combined with user knowledge if so desired. Computational results are given over the time period 2003 to 2010 for the S&P ASX 200.

Chen and Kwon (2012) presented a model for index tracking, based upon robust optimisation (e.g. see Ben-Tal and Nemirovski (1998)), that creates a tracking portfolio based upon the similarity between the assets in the decided tracking portfolio and the assets in the index. Transaction costs are not considered, although the number of assets held in the tracking portfolio is constrained. Computational results are given for tracking the S&P 100 using daily return data over the period January 2002 to January 2007.

Zhang et al. (2012) presented mean-semivariance-entropy model for multi-period portfolio selection by taking into account four criteria viz; return risk, transaction cost and diversification degree of the portfolio. They propose a bi-objective optimization model for multi-objective portfolio selection. They proposed a hybrid algorithm for solving the multi-period portfolio selection. They express the idea of their model and the effectiveness of the designed algorithm, with two examples for simulating the real transaction. The first example is a multi-period portfolio selection problem with trapezoidal fuzzy returns, while the second one demonstrates a multi-period portfolio decision-making with triangle fuzzy returns. All the assets are from the Shanghai stock exchange which cover the period from January 2001 to January 2010.

Clements et al. (2013) investigated the use of a stochastic approach in forming a stock price index. First, they set out the basics of index-number theory and related it to conventional indexes of stock prices. Second, they applied their stochastic approach to share prices. Finally, they applied their framework to the issue of portfolio tracking and investigated whether it is possible to ignore certain stocks on the basis of their contribution to the index. They used daily data for 20 stocks underlying the S&P/ASX20 index, for the period from January 2003 to December 2008.

2.5 Enhanced indexation survey

Enhanced indexation is concerned with finding portfolios that give additional return with respect to an underlying index. The term enhanced indexation is used to describe any strategy that is used in conjunction with index tracking for the purpose of outperforming a specific index benchmark. Enhanced indexation is a relatively unconsidered area in the scientific literature. All the work considered below was published relatively recently. In general, algorithms developed for index tracking can often be extended with only minor modifications to deal with enhanced indexation (or both).

Ghandar et al. (2010) presented an evolutionary approach to designing a fuzzy rule based system for deciding portfolio composition. Their approach addresses both index tracking and enhanced indexation. Although transaction costs are considered when rebalancing occurs there is no explicit constraint on the transaction cost incurred. They comment that their rules can be combined with user knowledge if so desired. Computational results are given over the time period 2003 to 2010 for the

S&P ASX 200

Lejeune and Samatli-Pac (2010) formulated the enhanced indexation problem as a mixed-integer nonlinear programming problem. They regard the problem as one of constructing a portfolio whose variance is below a given limit with a desired probability. They present two variants of an outer approximation algorithm for the solution of this nonlinear problem. Asset returns are modeled using a factor model. Computational results are given for forming portfolios containing up to 30 assets (from a universe of up to 1000 assets with price data from 1997-2005) where the market benchmark is the S&P 500.

Meade and Beasley (2011) presented a Sortino ratio portfolio selection strategy designed to achieve returns in excess of the market index. Their strategy is designed to identify and exploit momentum (the tendency of either high or poorly performing stocks to continue to exhibit high or poor performance for a long period, i.e. of the order of a year or longer). They use a genetic algorithm and present results for a number of test problems involving up to 1200 assets.

Roman et al. (2011) used second-order stochastic dominance to construct an enhanced indexation portfolio. In their approach, based on Fabian et al. (2011), they construct a portfolio which stochastically dominates the index. Computational results are given for three test problems involving up to 491 assets.

Guastaroba and Speranza (2012) presented a model that includes fixed and variable transaction costs, as well as a constraint upon the total transaction cost incurred when rebalancing from an existing portfolio. They also constrain the maximum number of assets that can be held in the portfolio. In their model they track an index by reference to the absolute deviation between a scaled index value and the tracking portfolio value, rather than by tracking the return on the index. They presented a modification of their model to deal with enhanced indexation. An heuristic based upon kernel search is presented and computational results are given for eight test problems involving up to 2151 assets.

Lejeune (2012) formulated the enhanced indexation problem as a stochastic game theoretic model which is reformulated as a convex second-order cone programming problem. Computational results are given for forming enhanced indexation portfolios (from a universe of 700 assets with price data from 1999-2004) to out-perform the Dow Jones, Russell 2000 and S&P 500 indices over period 2005-2006.

Thomaidis (2012) presented a model for enhanced indexation based on the application of cointegration technique. He extended the previous work of Alexander and Dimitriu (2005) in designing a trading portfolio that outperforms a market benchmark. He proposed a technique that consistently explores the space of feasible portfolio configurations, taking into account constraints on the total number of assets as well as on the trading position. Computational results were presented using data sets involving 65 Dow Jones stocks and the time period spanned by his sample data is from 20 June 2001 to 12 November 2008. He investigated the empirical performance of this strategy taking into account transaction costs and other market frictions.

Thomaidis (2013) proposed an integrated and interactive procedure for designing an enhanced indexation strategy with predetermined investment goals and risk constraints. He considered restrictions on the total number of tradable assets and non-standard investment objectives, focusing on the probability that the enhanced strategy under-performs the market. In dealing with the inherent complexity of the resulting cardinality-constraint formulations, he applies three nature-inspired optimisation techniques: simulated annealing, genetic algorithms and particle swarm optimisation. Computational results were benchmarked against the American Dow Jones index.

2.6 Conclusion

In this chapter we first presented investment preliminaries where we considered a broad range of investment philosophies from market indices, portfolio diversification and the impact of the transaction costs in trading. We then presented the historical and practical context of portfolio optimisation. We also discussed the portfolio management strategies where a portfolio must be managed whether investors follow a passive approach or an active approach to selecting and holding their financial assets.

In addition, we considered recently published studies in the literature relating to index tracking and enhanced indexation. We categorised index tracking literature into four parts where we consider Heuristics models, Markowitz models, Genetic models which is a part of Heuristic model and other research work. Additionally, we discussed a review of the literature relating to enhanced indexation.

Overall we would summarise that the majority of the work reviewed does neglect the impact of the transaction costs when constructing a portfolio. Also, we can conclude that there is no single mathematical perspective for the problem of index tracking and enhanced indexation. Furthermore, most authors adopted their own model and typically use just data sets of their own, not publicly available data used by others. As a consequence it is difficult to perform a systematic data driven comparison of different approaches. Finally, computational results were sometimes not detailed while computational times were missing for some work considered in our review.

Chapter 3

Index Tracking with Fixed and Variable Transaction Costs

3.1 Introduction

The focus of this chapter is on presenting two mixed-integer linear programming formulations of the portfolio construction problem; index tracking. In particular we explicitly consider both fixed and variable transaction costs. The performance of the proposed formulations are investigated through computational studies and the approaches are applied to five data sets.

In this chapter we investigate two different index tracking models that account for fixed and variable transaction costs when constructing and/or rebalancing an index tracking portfolio. Additionally, we consider constraints limiting the stocks that can be bought/sold as well as limiting the total transaction cost that can be incurred.

In the following sections, we first give our notation and define our decisions variables. Then we present and discuss the constraints associated with index tracking problem. Furthermore, we give the objectives considered and clarify our contribution in relationship to earlier work. We finalise the chapter by giving information on our data sets and presenting computational results for our index tracking formulation.

3.2 Formulations

In this section we present our two mixed-integer linear programming formulations. We first present our notation, then the constraints of the problem which are common to both formulations. We then present our objectives which differ between formulations.

3.2.1 Notation

Suppose that we observe over time 0,1,2,...,T the value of N assets, as well as the value of the index we are tracking. Further, suppose that we are interested in deciding the best set of K assets to hold (where $K < N$), as well as their appropriate quantities. Building on the notation of Canakgoz and Beasley (2009) let:

ε_i be the minimum proportion of the tracking portfolio (henceforth TP) that must be held in asset i if any of the asset is held

δ_i be the maximum proportion of the TP that can be held in asset i

X_i be the number of units of asset i in the current TP

V_{it} be the value (price) of one unit of asset i at time t

I_i be the value of the index at time t

R_t be the single period continuous time return for the index at time t, $R_t = \log_e(I_t/I_{t-1})$

r_{it} be the single period continuous time return for asset i at time t, $r_{it} = \log_e(V_{it}/V_{it-1})$

C be the total value (≥ 0) of the current TP $[X_i]$ at time T plus cash change (either new cash to be invested or cash to be taken out) so, $C = \sum_{i=1}^{N} V_{iT} X_i$ + cash change

f_i^b be the fractional transaction cost associated with buying one unit of asset i at time T, so that buying one unit of asset i at time T costs $f_i^b V_{iT}$

f_i^s be the fractional transaction cost associated with selling one unit of asset i at time T, so that selling one unit of asset i at time T costs $f_i^s V_{iT}$

F_i^b be the fixed cost of buying any of asset i at time T

F_i^s be the fixed cost of selling any of asset i at time T

M_i^b be the maximum number of units of asset i that can be bought at time T (assuming we choose to buy some of asset i)

M_i^s be the maximum number of units of asset i that can be sold at time T (assuming we choose to sell some of asset i)

γ be the limit ($0 \leq \gamma \leq 1$) on the proportion of C that can be consumed by transaction cost

Then our decision variables are:

x_i the number of units (≥ 0) of asset i that we choose to hold in the new TP

z_i =1 if any of asset i is held in the new TP, =0 otherwise

α_i^b =1 if any of asset i is bought, =0 otherwise

α_i^s =1 if any of asset i is sold, =0 otherwise

y_i^b the number of units (≥ 0) of asset i that are bought

y_i^s the number of units (≥ 0) of asset i that are sold

Without significant loss of generality (since the sums of money involved are large) we allow $[x_i, y_i^b, y_i^s]$ to take fractional values. Note also that as $x_i \geq 0$ we are excluding short selling (shorting) from our model.

The formulations presented below deal with the general situation of rebalancing of an existing TP $[X_i \; i = 1, \ldots, N]$ to a new portfolio $[x_i \; i = 1, \ldots, N]$. If we are creating a new TP from cash then we simply set $X_i = 0 \; i = 1, \ldots, N$.

3.2.2 Constraints

The constraints of the problem are:

$$\sum_{i=1}^{N} z_i = K \tag{3.1}$$

$$\varepsilon_i z_i \leq x_i V_{iT}/C \leq \delta_i z_i \qquad i = 1, \ldots, N \tag{3.2}$$

$$x_i = X_i + y_i^b - y_i^s \qquad i = 1, \ldots, N \tag{3.3}$$

$$\alpha_i^b + \alpha_i^s \leq 1 \qquad i = 1, \ldots, N \tag{3.4}$$

$$y_i^b \leq M_i^b \alpha_i^b \qquad i = 1, \ldots, N \tag{3.5}$$

$$y_i^s \leq \min[M_i^s, X_i] \alpha_i^s \qquad i = 1, \ldots, N \tag{3.6}$$

$$\sum_{i=1}^{N} V_{iT} x_i = C - \sum_{i=1}^{N} [f_i^b V_{iT} y_i^b + f_i^s V_{iT} y_i^s + F_i^b \alpha_i^b + F_i^s \alpha_i^s] \tag{3.7}$$

$$\sum_{i=1}^{N}[f_i^b V_{iT} y_i^b + f_i^s V_{iT} y_i^s + F_i^b \alpha_i^b + F_i^s \alpha_i^s] \leq \gamma C \tag{3.8}$$

$$y_i^b, y_i^s, x_i \geq 0 \qquad i = 1, \ldots, N \tag{3.9}$$

$$\alpha_i^b, \alpha_i^s, z_i \in [0, 1] \qquad i = 1, \ldots, N \tag{3.10}$$

Equation (3.1) ensures that there are exactly K assets in the new TP. Equation (3.2) ensures that if an asset i is not in the new TP ($z_i = 0$) then x_i is also zero; it also ensures that if the asset is chosen to be in the new TP ($z_i = 1$) then the amount of the asset held satisfies the proportion limits defined. Equation (3.3) defines the number of units of asset i held after rebalancing (we currently hold X_i, we buy y_i^b, we sell y_i^s, so after trading we hold $X_i + y_i^b - y_i^s$ and this must equal x_i).

Equation (3.4) prevents simultaneously buying and selling of asset i; in other words if we trade the asset, and equation (3.4) does not force us to, we can either buy or sell, but not both. Equation (3.5) relates the number of units y_i^b of asset i bought to the zero-one variable α_i^b. This equation forces y_i^b to be zero if α_i^b is zero, whilst if α_i^b is one it ensures that the number of units bought cannot exceed the maximum M_i^b allowed. Equation (3.6) is as equation (3.5) except that it relates to selling the asset, where here we cannot sell more than $\min[M_i^s, X_i]$ units of asset i.

Equation (3.7) is a balance constraint such that the total value of the new TP at time T equals the value of the current TP at time T plus the cash change (i.e. C) minus the total transaction cost. Equation (3.8) limits the total transaction cost incurred appropriately. Equation (3.9) defines the continuous variables to be non-negative and equation (3.10) is the integrality condition for the zero-one variables.

3.2.3 Constraint discussion

We would comment here as to the role played in our constraints by the minimum and maximum proportions (equation (3.2)). Clearly if these factors were not present (equivalently $\varepsilon_i = 0$ and $\delta_i = 1$) then the problem would be less constrained and the value achieved by our objective function (considered below) could be improved. The reason these factors are present, as indeed they are present in previous work (Beasley et al., 2003; Canakgoz and Beasley, 2009; Guastaroba and Speranza, 2012), in that they reflect practical considerations adopted when forming an index tracking portfolio.

In practical applications although one optimises on in-sample data the underlying issue is how the index tracking portfolio chosen performs on (unseen) out-of-sample data. So, for example, if we use a model (of any kind) to decide an index tracking portfolio today, when past asset returns are known, and buy and hold that index tracking portfolio into the future (when asset returns are unknown) how will it perform?

Decision-makers in such situations are concerned to avoid situations where optimising using in-sample data leads to a portfolio with too much (or too little) invested in an asset. If there is too much invested in an asset then the portfolio is not diversified and the decision-maker is exposing themselves to risk associated with returns from a single asset. This situation can be avoided by making use of the maximum proportion factor (δ_i). If there is too little invested in an asset then the issue essentially becomes one of administrative convenience, making a very small investment in an asset when its effect on portfolio performance can hardly be significant seems unnecessary. This situation can be avoided by making use of the minimum proportion factor (ε_i).

Clearly in practice the actual values adopted for the minimum and maximum proportions are a matter of judgment by the decision-maker. What is important though is that any model for index tracking allows such factors to be considered.

This is why they have been included in the formulation we have presented above.

More generically a number of the constraints in our formulation involve factors designed to allow the decision-maker the opportunity to shape the index tracking portfolio produced to reflect their own preferences. Constraints of this type include:

- Equation (3.1), where the factor K relates to the number of assets to hold in the portfolio
- Equation (3.2), where (as discussed above) the factors ε_i and δ_i relate to the minimum and maximum proportion of total portfolio value invested in any asset
- Equations (3.5) and (3.6), where the factors M_i^b and M_i^s relate to the maximum number of units of each asset that can be bought/sold
- Equation (3.8), where the factor γ relates to the maximum proportion of total portfolio value that can be consumed by transaction cost

These factors (and their associated constraints) can essentially be seen as internally derived, coming from the decision-maker. Some constraints though are externally derived, that is they are imposed upon the decision-maker, either as a matter of logic or as a matter of market structure.

An example of a constraint imposed upon the decision-maker as a matter of logic is equation (3.3) which is a balance constraint for an asset relating the number of units held after rebalancing to the number of units held before rebalancing and the number of units bought/sold.

An example of a factor imposed upon the decision-maker as a matter of market structure is the presence of fixed and variable transaction costs (as reflected in equations (3.7) and (3.8). When trading an asset (for example via a third-party broker) transaction costs will be incurred and almost always this transaction cost will have a variable cost component (so the total transaction cost paid depends upon the level of trade, i.e. the number of units bought/sold). For some assets there may also be a fixed cost component. The usual reason for a transaction fixed cost component is that the asset is one that is less commonly traded (so the volume of trading in the market is much less than for other assets). In such situations the fixed cost is imposed by the third-party broker to discourage very small trades in the asset. Echoing a point we made above what is important is that any model for index tracking allows factors such as fixed and variable transaction costs to be considered. This is why they have been considered in the formulation we have presented above.

3.2.4 Objectives

Above we have discussed the constraints we have presented and the reasoning behind them. We now go on to give the objectives we considered.

We will adopt the same weight approximation for TP returns as in Canakgoz and Beasley (2009) where the weight w_i associated with asset i in the TP is given by:

$$w_i = x_i V_{iT}/(C - \gamma C) \qquad i = 1, \ldots, N \tag{3.11}$$

and the return on the TP at time t is given by $\sum_{i=1}^{N} w_i r_{it}$.

On a technical note here equation (3.11) is predicated on the limit (γ) on the proportion of C consumed by transaction cost being small. If we have a problem in which transaction cost is effectively unrestricted (so $\gamma = 1$) then we define w_i using $w_i = x_i V_{iT}/C$.

The objective in index tracking is to minimise the difference between the returns obtained from the chosen portfolio and the index being tracked.

In this chapter we propose two approaches for the objective function associated with choice of a TP, namely:

- minimise the maximum absolute difference between TP return and index return. This is:

$$\text{minimise } \max\{ \mid \sum_{i=1}^{N} w_i r_{it} - R_t \mid t = 1, \ldots, T\} \tag{3.12}$$

- minimise the average of the absolute differences between TP return and index return. This is:

$$\text{minimise } \sum_{t=1}^{T} \mid \sum_{i=1}^{N} w_i r_{it} - R_t \mid /T \tag{3.13}$$

  The objective adopted here equation (3.13) is effectively a goal programming style objective where we are minimising an equally weighted sum of deviations from the target return R_t.

Although both equations (3.12) and (3.13) are nonlinear we can linearise them in a standard way. For equation (3.12) introduce a single variable d (≥ 0) and our formulation, which we denote by ***MINIMAX***, then is:

$$\text{minimise } d \tag{3.14}$$

subject to equations (3.1)-(3.11) and:

$$d \geq \sum_{i=1}^{N} w_i r_{it} - R_t \qquad t = 1, \ldots, T \tag{3.15}$$

$$d \geq R_t - \sum_{i=1}^{N} w_i r_{it} \qquad t = 1, \ldots, T \tag{3.16}$$

For equation (3.13) introduce variables d_t (≥ 0, $t = 1, \ldots, T$) and our formulation, which we denote by ***MINIAVERAGE***, then is:

$$\text{minimise } \sum_{t=1}^{T} d_t/T \tag{3.17}$$

subject to equations (3.1)-(3.11) and:

$$d_t \geq \sum_{i=1}^{N} w_i r_{it} - R_t \qquad t = 1, \ldots, T \tag{3.18}$$

$$d_t \geq R_t - \sum_{i=1}^{N} w_i r_{it} \qquad t = 1, \ldots, T \tag{3.19}$$

Both MINIMAX and MINIAVERAGE are mixed-integer linear programs. In terms of the size of these programs then, before any algebraic manipulation to eliminate variables and/or constraints, MINIMAX involves $7N + 1$ variables and MINIAVERAGE involves $7N + T$ variables. Both MINIMAX and MINIAVERAGE involve $6N + 2T + 3$ constraints.

3.2.5 Contribution and relationship to earlier work

We should clarify here the contribution of this work and the relationship between the formulations seen above and:

- our earlier work as in Beasley et al. (2003) and Canakgoz and Beasley (2009)
- the recently published work of Guastaroba and Speranza (2012), that also deals with fixed and variable transaction costs

We would note that a number of constraints in our formulations are as seen in other work (Beasley et al., 2003; Canakgoz and Beasley, 2009; Guastaroba and Speranza, 2012), as indeed they are seen in other papers by other authors. This is natural since constraints for many optimisation problems are often expressed mathematically exactly as in previous work in the literature. The differences between the formulations presented in this contribution and previous work can be summarised as:

- we include fixed costs related to trade in an asset; these are not included in Canakgoz and Beasley (2009). Fixed costs are included in Guastaroba and Speranza (2012), but they do not distinguish between the type of trade (i.e. in their model a fixed cost is incurred if a trade occurs), whereas in this work we have different fixed costs for buying or selling an asset.

- we model variable transaction costs in a different manner from that adopted in (Beasley et al., 2003; Canakgoz and Beasley, 2009; Guastaroba and Speranza, 2012)

- in our model transaction costs detract from the value of the portfolio held after trading, in the model of Guastaroba and Speranza (2012) transaction costs do not (so they assume that any transaction costs incurred are paid out of a separate fund)

- With respect to the objective adopted:

  - Beasley et al. (2003) adopt a nonlinear objective and use a genetic algorithm heuristic solution approach; we have linear objectives and will adopt (as will become apparent below) optimal solution approaches based on mixed-integer linear programming

  - Canakgoz and Beasley (2009) adopt an approach based upon linear regression; we do not use regression at all within our formulations

  - Guastaroba and Speranza (2012) adopt an approach based upon minimising the total absolute deviation between a scaled index value and the tracking portfolio value; our approach is based upon tracking an index by comparing the returns from the index with the returns from the tracking portfolio

3.3 Computational results

3.3.1 Data

To test our formulation, we used the same data sets as in Beasley et al. (2003) and Canakgoz and Beasley (2009), which are publicly available from OR-Library Beasley

(1990), *http://people.brunel.ac.uk/∼mastjjb/jeb/info.html.* These test problems contain weekly price data (for $T = 290$ weeks) for assets drawn from a number of major world equity indices. Table 3.1 shows the test problems we considered.

Table 3.1: Test problems

Index	Number of stocks N	Number of selected stocks K
Hang Seng	31	10
DAX 100	85	10
FTSE 100	89	10
S&P 100	98	10
Nikkei 225	225	10

The computational results presented below are for our approach as coded in AMPL and solved using ILOG Cplex (version 11.0) IBM ILOG Cplex Solver (2011) as the mixed-integer optimiser. We used Cplex default parameter settings, except that we changed the tolerance parameters so as to find the genuine optimal solution. The reason for this is that Cplex, by default, finds a solution within a specified tolerance of the genuine optimal and since we have real-valued MINIMAX and MINIAVERAGE objective functions (equations (3.14) and (3.17) we wanted to avoid the situation where we missed the genuine optimal solution. We used a Windows 2.4GHz, Core 2 Duo Pentium, pc with 4Gb memory. Unless otherwise stated we:

- used $K = 10$, so a tracking portfolio with ten stocks
- used an initial tracking portfolio of value 10^6 composed of the first $K = 10$ stocks in equal proportions, i.e. $X_i = (10^6/K)/V_{i0}$ $i = 1, ..., K$; $X_i = 0$ $\forall i > K$
- used $\varepsilon_i = 0.01$ and $\delta_i = 1$ $\forall i$
- used $f_i^b = f_i^s = 0.01$ and $F_i^b = F_i^s = 100$ $\forall i$
- used $M_i^b = C/V_{iT}$ and $M_i^s = X_i$ $\forall i$
- imposed a computational time limit of one hour (3600 seconds)

3.3.2 Zero fixed transaction cost

To illustrate how our approaches perform in terms of index tracking we took the in-sample time period [0,145] for each of our data sets and solved our two formulations MINIMAX and MINIAVERAGE, but with fixed transaction costs of zero (in other words only the variable transaction cost was incurred). Table 3.2 gives the results obtained in terms of the optimal (minimal) in-sample objective function value and the computation time (in seconds). We also give the value of the objective function when computed out-of-sample (over the period [146,290]). Note here that the in-sample values are as given in equations (3.14) and (3.17). The out-of-sample values are computed directly from the tracking portfolio held. In other words the out-of-sample value given for MINIMAX is computed using:

$$\max\{\ |\ \log_e(\sum_{i=1}^{N} x_i V_{it} / \sum_{i=1}^{N} x_i V_{it-1})\ - R_t\ |\ t = 146, \ldots, 290\} \tag{3.20}$$

The out-of-sample value given for MINIAVERAGE is computed using:

$$\sum_{t=146}^{290}\ |\ \log_e(\sum_{i=1}^{N} x_i V_{it} / \sum_{i=1}^{N} x_i V_{it-1})\ - R_t\ |\ /145 \tag{3.21}$$

Some problems in Table 3.2 reached the self-imposed computational time limit of 3600 seconds. These problems are indicated by the time being enclosed in brackets. In such cases the solution values reported are those associated with the best mixed-integer feasible solution found before the time limit was reached.

To illustrate Table 3.2 we have that for the S&P 100 with $N = 98$ assets and $\gamma = 0.0075$, so a transaction cost limit of 0.75% of portfolio value, the minimal MINIMAX objective function value is 0.01313 and this is found in 662.1 seconds. Out-of-sample the tracking portfolio associated with this minimal solution has an objective function (equation (3.20)) value of 0.03003.

For MINIAVERAGE the minimal objective function value is 0.00485 and this is found in 1989.3 seconds. Out-of-sample the tracking portfolio associated

Table 3.2: Results with fixed transaction costs of zero

Index	Number of assets	Transaction cost limit (γ)	MINIMAX Objective value In-sample	MINIMAX Objective value Out-of-sample	MINIMAX Time (secs)	MINIAVERAGE Objective value In-sample	MINIAVERAGE Objective value Out-of-sample	MINIAVERAGE Time (secs)
Hang Seng	31	0.0025	0.03129	0.03810	0.1	0.00953	0.00896	0.4
		0.005	0.02220	0.03096	0.4	0.00673	0.00699	0.8
		0.0075	0.01596	0.02588	3.0	0.00516	0.00540	1.9
		0.01	0.01102	0.01854	3.8	0.00360	0.00368	1.4
DAX 100	85	0.0025	0.03054	0.07452	0.2	0.01038	0.01358	0.4
		0.005	0.02171	0.07476	1.0	0.00757	0.01158	1.4
		0.0075	0.01563	0.07365	12.3	0.00584	0.00950	18.7
		0.01	0.01171	0.06234	87.9	0.00431	0.00713	281.4
FTSE 100	89	0.0025	0.02480	0.03361	0.6	0.00789	0.00842	2.3
		0.005	0.01855	0.03567	5.6	0.00646	0.00816	47.4
		0.0075	0.01430	0.02820	46.0	0.00534	0.00708	1978.9
		0.01	0.01212	0.02442	(3600)	0.00459	0.00665	(3600)
S&P 100	98	0.0025	0.03166	0.03259	0.7	0.00751	0.00667	3.0
		0.005	0.01727	0.02836	8.4	0.00607	0.00661	42.0
		0.0075	0.01313	0.03003	662.1	0.00485	0.00604	1989.3
		0.01	0.01096	0.02138	(3600)	0.00410	0.00521	(3600)
Nikkei 225	225	0.0025	0.02018	0.03216	2.9	0.00558	0.00702	28.1
		0.005	0.01415	0.03521	139.4	0.00461	0.00829	(3600)
		0.0075	0.01115	0.03179	(3600)	0.00403	0.00710	(3600)
		0.01	0.01099	0.02432	(3600)	0.00383	0.00666	(3600)

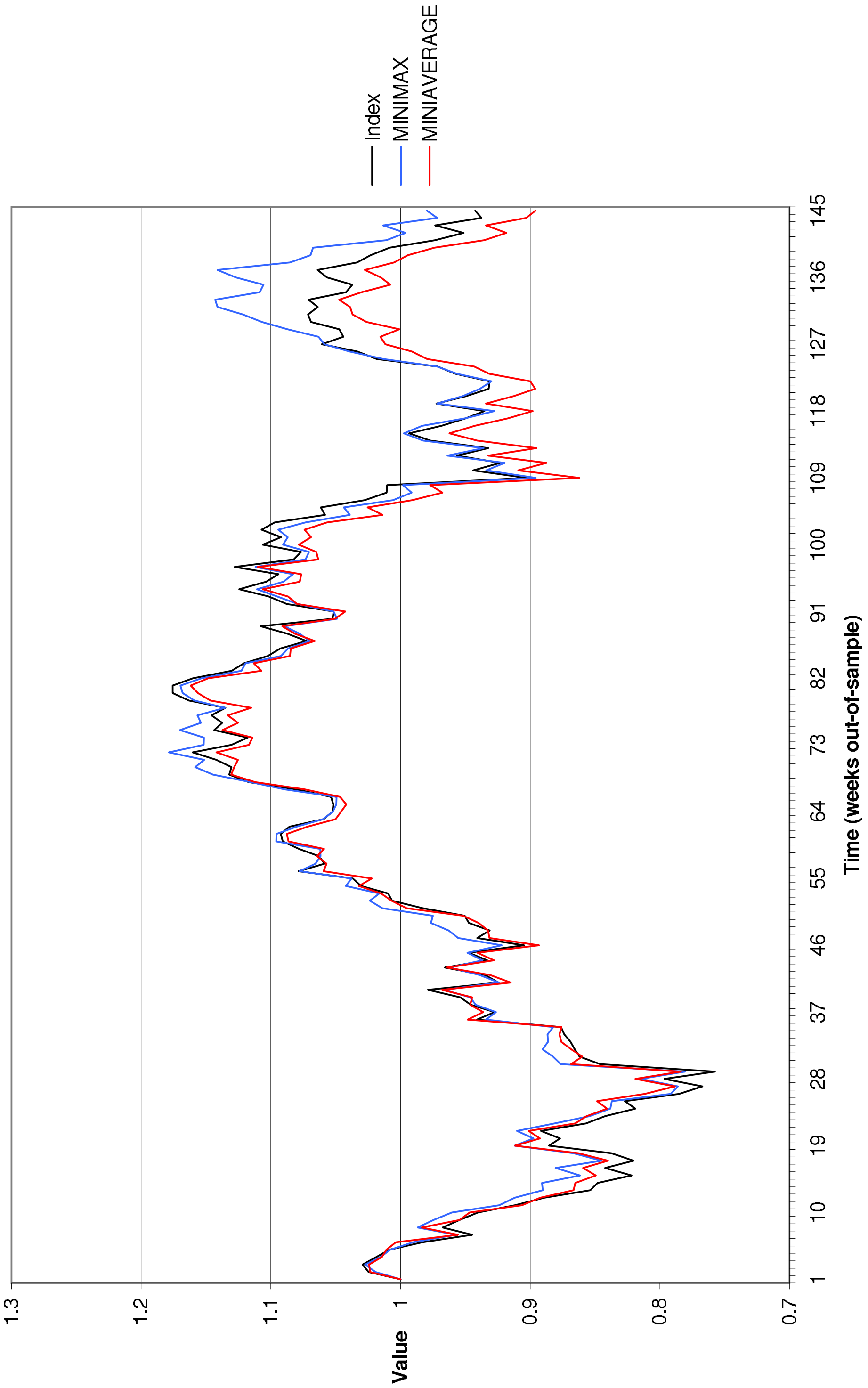

Figure 3.1: Out-of-sample performance for the Nikkei 225 with $\gamma = 0.01$ and no fixed costs

with this minimal solution has an objective function (equation (3.21)) value of 0.00604.

Examining Table 3.2 we can see that, as we might expect, the time required increases as we increase the transaction cost limit (γ) and as the size of the problem (number of assets) increases. Comparing the objective function values we have that the average value of (out-of-sample objective value/in-sample objective value) is 2.26 for MINIMAX and 1.32 for MINIAVERAGE. In other words out-of-sample we do see a degradation in performance as compared to in-sample (as we would expect since in-sample we can directly optimise on known data, but out-of-sample we hold the portfolio unchanged) but not too high a degradation.

In order to provide a graphical illustration of the quality of results Figure 3.1 shows the out-of-sample performance of the index and the tracking portfolios chosen by MINIMAX and MINIAVERAGE for the Nikkei 225 with $\gamma = 0.01$. Note here that we are tracking an index over a 145 week period (nearly 3 years) with a fixed portfolio containing just ten assets (when the index has 225 assets). In such circumstances it is hardly surprising that we fail to perfectly track the index out-of-sample. However it is clear that (visually at least) we do track the index well, sustained large deviations from the index only becoming apparent from weeks 100 onward in Figure 3.1.

Table 3.3 shows the results when, out-of-sample, we perform a linear least squares regression of the return from the tracking portfolio, i.e. $\log_e(\sum_{i=1}^{N} x_i V_{it} / \sum_{i=1}^{N} x_i V_{it-1})$, against the return from the index R_t. In that table we show the value of the regression intercept and slope. We also show the value of the coefficient of determination R^2 which is a measure of how good a fit the regression line is. Whilst ideally we would like an intercept of zero and a slope of one (with a value for R^2 of one) it is clear that, recalling we are

Table 3.3: Out-of-sample regression results with fixed transaction costs of zero

Index	Number of assets	Transaction cost limit (γ)	MINIMAX Intercept	MINIMAX Slope	MINIMAX R^2	MINIAVERAGE Intercept	MINIAVERAGE Slope	MINIAVERAGE R^2
Hang Seng	31	0.0025	-0.00070	0.94711	0.85021	-0.00061	0.94776	0.85605
		0.005	-0.00130	0.96846	0.89803	0.00051	0.94193	0.89705
		0.0075	0.00020	0.94435	0.92668	0.00006	0.94834	0.93169
		0.01	0.00060	0.95779	0.96612	0.00032	0.96838	0.96434
DAX 100	85	0.0025	0.00105	0.84571	0.44464	0.00124	0.85674	0.47552
		0.005	0.00113	0.85524	0.51486	0.00098	0.84640	0.53842
		0.0075	0.00082	0.85303	0.61378	0.00073	0.88872	0.61804
		0.01	-0.00007	0.79018	0.70698	0.00079	0.90249	0.72510
FTSE 100	89	0.0025	0.00135	0.79663	0.60921	0.00126	0.73629	0.53840
		0.005	0.00069	0.82568	0.67544	0.00032	0.73687	0.55158
		0.0075	0.00055	0.80278	0.69971	0.00021	0.85420	0.67476
		0.01	0.00046	0.92108	0.77576	0.00008	0.87649	0.71041
S&P 100	98	0.0025	-0.00011	0.91920	0.72371	-0.00001	0.95227	0.78054
		0.005	-0.00039	0.95087	0.77407	-0.00041	0.95856	0.79951
		0.0075	-0.00080	0.93418	0.78633	-0.00052	0.97518	0.82268
		0.01	-0.00041	0.97133	0.83488	-0.00016	0.95116	0.86539
Nikkei 225	225	0.0025	-0.00077	1.01502	0.88216	-0.00015	1.00488	0.90847
		0.005	-0.00040	1.06403	0.87240	-0.00051	1.04947	0.86016
		0.0075	0.00023	1.02320	0.88623	-0.00045	1.02251	0.90116
		0.01	0.00024	0.93028	0.90308	-0.00036	0.97425	0.91230

choosing a tracking portfolio with just $K = 10$ assets, we will not achieve this ideal. For MINIMAX we can see from Table 3.3 that the intercept values are very small, with a mean value of 0.00012, and that the slope values are close to one, with a mean value of 0.91581.

For MINIAVERAGE the corresponding values are 0.00017 and 0.91964. Both MINIMAX and MINIAVERAGE have similar average R^2 values (of approximately 0.77). For simple linear regression the coefficient of determination is the square of the correlation coefficient so that we have an average correlation (out-of-sample) between tracking portfolio return and index return of $\sqrt{0.77} = 0.88$, which is statistically highly significant given the number of out-of-sample observations.

3.3.3 Non-zero fixed transaction cost

Table 3.4 deals with the same problems as Table 3.2, but where now fixed costs (as well as variable costs) are incurred when we trade an asset. Comparing Table 3.4 and Table 3.2 it is clear that introducing fixed costs makes the problem harder to solve for MINIAVERAGE. For MINIAVERAGE although five test problems in Table 3.2 encounter the computational time limit there are 13 problems that encounter the same time limit in Table 3.4. By contrast MINIMAX has 4 problems that encounter the computational time limit in both Table 3.2 and Table 3.4.

Whereas in Table 3.2 the average value of (out-of-sample objective value/in-sample objective value) is 2.26 for MINIMAX and 1.32 for MINIAVERAGE; in Table 3.4 it is 2.27 for MINIMAX and 1.30 for MINIAVERAGE, so here the presence of fixed costs seems to make little difference.

Figure 3.2 shows the same information as Figure 3.1, but for the case

Table 3.4: Results with non-zero fixed and variable transaction costs

Index	Number of assets	Transaction cost limit (γ)	MINIMAX Objective value In-sample	MINIMAX Objective value Out-of-sample	MINIMAX Time (secs)	MINIAVERAGE Objective value In-sample	MINIAVERAGE Objective value Out-of-sample	MINIAVERAGE Time (secs)
Hang Seng	31	0.0025	0.03206	0.03714	0.5	0.00975	0.00908	0.5
		0.005	0.02301	0.03158	1.8	0.00687	0.00713	42.4
		0.0075	0.01722	0.02355	5.7	0.00526	0.00545	(3600)
		0.01	0.01190	0.02577	9.7	0.00381	0.00380	32.4
DAX 100	85	0.0025	0.03252	0.07462	0.3	0.01054	0.01384	1.5
		0.005	0.02424	0.06946	7.9	0.00775	0.01199	18.0
		0.0075	0.01624	0.07438	20.8	0.00619	0.01011	(3600)
		0.01	0.01316	0.06393	416.1	0.00470	0.00708	(3600)
FTSE 100	89	0.0025	0.02563	0.03598	4.6	0.00803	0.00836	119.5
		0.005	0.01958	0.03379	10.6	0.00662	0.00869	(3600)
		0.0075	0.01510	0.02753	172.3	0.00552	0.00812	(3600)
		0.01	0.01320	0.03002	(3600)	0.00483	0.00731	(3600)
S&P 100	98	0.0025	0.03492	0.03356	0.6	0.00773	0.00696	10.6
		0.005	0.01864	0.03238	29.0	0.00627	0.00719	(3600)
		0.0075	0.01403	0.03650	373.3	0.00508	0.00605	(3600)
		0.01	0.01119	0.03005	(3600)	0.00438	0.00569	(3600)
Nikkei 225	225	0.0025	0.02268	0.02926	18.7	0.00574	0.00708	(3600)
		0.005	0.01520	0.03299	457.4	0.00484	0.00681	(3600)
		0.0075	0.01201	0.03525	(3600)	0.00424	0.00811	(3600)
		0.01	0.01106	0.03381	(3600)	0.00403	0.00661	(3600)

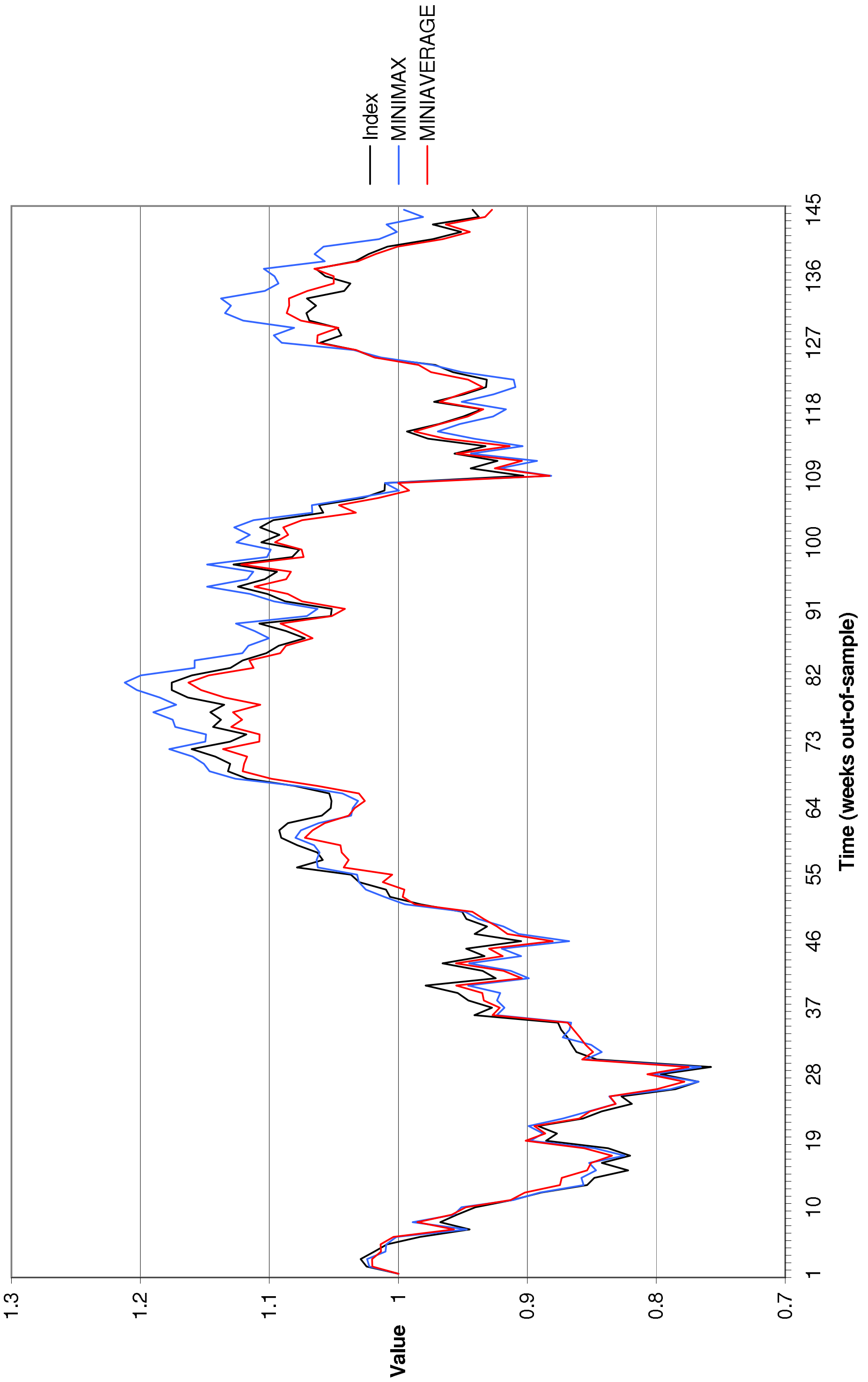

Figure 3.2: Out-of-sample performance for the Nikkei 225 with $\gamma = 0.01$

Table 3.5: Out-of-sample regression results with non-zero fixed and variable transaction costs

Index	Number	Transaction	MINIMAX			MINIAVERAGE		
	of assets	cost limit (γ)	Intercept	Slope	R^2	Intercept	Slope	R^2
Hang Seng	31	0.0025	-0.00089	0.96624	0.85472	-0.00070	0.95321	0.85154
		0.005	-0.00140	0.96608	0.89352	0.00047	0.94269	0.89425
		0.0075	0.00010	0.95464	0.92578	0.00003	0.95029	0.93125
		0.01	0.00049	0.99908	0.94134	0.00030	0.96761	0.96231
DAX 100	85	0.0025	0.00138	0.84617	0.41010	0.00130	0.86518	0.47050
		0.005	0.00063	0.81420	0.49097	0.00108	0.85524	0.52671
		0.0075	0.00090	0.86056	0.60011	0.00084	0.87577	0.58557
		0.01	0.00040	0.85010	0.73706	0.00029	0.83858	0.71310
FTSE 100	89	0.0025	0.00128	0.79481	0.59400	0.00120	0.74295	0.54227
		0.005	0.00091	0.81246	0.66599	0.00006	0.74187	0.51717
		0.0075	0.00055	0.80146	0.70383	0.00033	0.72680	0.56704
		0.01	0.00047	0.91237	0.74073	-0.00051	0.81998	0.64213
S&P 100	98	0.0025	-0.00004	0.92322	0.72006	0.00002	0.94613	0.77053
		0.005	-0.00049	0.91853	0.75274	-0.00073	0.91701	0.74763
		0.0075	-0.00093	0.92976	0.73891	-0.00050	0.97050	0.82036
		0.01	-0.00064	0.94376	0.81426	-0.00066	0.99548	0.83835
Nikkei 225	225	0.0025	-0.00019	1.03559	0.90186	-0.00002	0.99554	0.90679
		0.005	-0.00040	1.05182	0.87515	-0.00020	1.02411	0.91586
		0.0075	-0.00041	1.06774	0.89038	-0.00125	1.01804	0.87646
		0.01	0.00040	1.03393	0.88819	-0.00013	0.94996	0.90743

where fixed costs are incurred. Visually tracking performance seems to deteriorate after about 55 weeks out-of-sample, whereas in Figure 3.1 tracking performance did not seem to deteriorate until about 100 weeks out-of-sample. Note that one effect of fixed costs (which are present in Figure 3.2, unlike Figure 3.1) is to limit flexibility in changing from the current tracking portfolio (given a fixed transaction cost limit of γC, equation (3.8)) and so for that reason alone we would expect tracking to be worse in the presence of fixed costs.

Table 3.5 shows the same information as Table 3.3, but for the case where fixed costs are incurred. For MINIMAX the mean intercept value is 0.00011 and the mean slope value is 0.92413. For MINIAVERAGE the corresponding values are 0.00006 and 0.90485. Both MINIMAX and MINIAVERAGE have similar average R^2 values (of approximately 0.75) giving a correlation coefficient of $\sqrt{0.75} = 0.87$, and this is again statistically highly significant given the number of out-of-sample observations.

We would note here that we have (for computational reasons) considered just one set of values for fixed costs. In practice, as discussed above, the fixed costs associated with trade in an asset would be externally derived as a matter of market structure. What is of importance in any practical application is that any formulation/model adopted can deal with such fixed costs. Clearly our formulations can deal with fixed costs (potentially different for buying/selling and different for each asset). Computationally the results presented in Table 3.2 and Table 3.4 indicate that introducing fixed costs makes the problem harder to solve for MINIAVERAGE, but not for MINIMAX.

3.3.4 Variation with K

To gain insight into how the results vary with the number of assets K chosen to be in the portfolio we took one of our data sets (the DAX 100 with

$N = 85$) and solved it for all values of the transaction cost limit γ; with $K = 10, 15, 20, 25, 50, 75$; both with and without fixed costs. The results can be seen in Table 3.6, which has the same format as Table 3.2 and Table 3.4. The results for $K = 10$ in Table 3.6 are the same as those for $K = 10$ in Table 3.2 and Table 3.4, but are repeated in Table 3.6 for ease of comparison.

Considering the in-sample results it is clear that as K increases from its initial value of 10 the in-sample objective function value falls. This is a reflection of the fact that increasing K provides more flexibility in that we can hold more assets in the portfolio chosen and hence allows a lower objective function value to be achieved.

It is noticeable however that for $K = 75$ we often see an increase in the in-sample objective function value over that seen for $K = 50$. For example for MINIMAX with $\gamma = 0.01$ and fixed transaction costs of zero in Table 3.6 the in-sample objective function value decreases from $K = 10$ to $K = 50$ (decreasing from 0.01171 to 0.00179), but increases to 0.00543 at $K = 75$. This is a reflection of the fact that since we are imposing a lower limit (minimum proportion) of $\varepsilon_i = 0.01$ for each asset then, for $K = 75$, the flexibility to vary the investment in each asset is more limited (since $K\varepsilon_i = 0.75$, i.e. 75% of the total investment is constrained to be in the assets chosen). Obviously if we were to remove the minimum proportion constraint (equivalently set $\varepsilon_i = 0$) then we would have more flexibility. However, since the results in Table 3.2 and Table 3.4 have been produced with this minimum proportion constraint we have retained it here for consistency of comparison.

With regard to out-of-sample objective function values we can see the same effect as in-sample. Typically the objective function value falls as K increases from its initial value of 10, but sometimes rises between $K = 50$ and $K = 75$. Obviously out-of-sample the precise effect seen for any particular

Table 3.6: Results with a varying number of assets K in the portfolio

Transaction cost limit (γ)	K	Fixed transaction costs of zero						Non-zero fixed and variable transaction costs					
		MINIMAX			MINIAVERAGE			MINIMAX			MINIAVERAGE		
		Objective value		Time	Objective value		Time	Objective value		Time	Objective value		Time
		In-sample	Out-sample	(secs)	In-sample	Out-sample	(secs)	In-sample	Out-sample	(secs)	In-sample	Out-sample	(secs)
0.0025	10	0.03054	0.07452	0.2	0.01038	0.01358	0.4	0.03252	0.07462	0.3	0.01054	0.01384	1.5
	15	0.02423	0.06800	0.4	0.00826	0.01109	0.7	0.02624	0.06877	3.0	0.00859	0.01149	4.6
	20	0.01841	0.06613	1.2	0.00660	0.00893	0.6	0.02002	0.06704	4.5	0.00688	0.00901	2.3
	25	0.01431	0.06500	3.3	0.00527	0.00700	0.5	0.01631	0.06548	8.8	0.00556	0.00738	1.4
	50	0.01062	0.07380	3.1	0.00370	0.00593	0.5	0.01275	0.07512	19.2	0.00412	0.00614	5.9
	75	0.00977	0.06939	5.6	0.00387	0.00532	4.2	0.01454	0.07112	9.4	0.00532	0.00605	6.6
0.005	10	0.02171	0.07476	1.0	0.00757	0.01158	1.4	0.02424	0.06946	7.9	0.00775	0.01199	18.0
	15	0.01542	0.06845	5.6	0.00557	0.00828	1.8	0.01692	0.07061	11.9	0.00600	0.00899	13.4
	20	0.01137	0.07098	5.4	0.00434	0.00698	1.7	0.01282	0.07196	17.1	0.00478	0.00707	18.6
	25	0.00820	0.07052	5.9	0.00316	0.00515	1.7	0.01010	0.07164	29.9	0.00370	0.00530	12.6
	50	0.00547	0.06810	15.1	0.00220	0.00449	1.6	0.00777	0.06854	26.3	0.00289	0.00465	10.2
	75	0.00622	0.06491	8.3	0.00252	0.00396	9.0	0.01004	0.06810	36.3	0.00375	0.00514	6.9
0.0075	10	0.01563	0.07365	12.3	0.00584	0.00950	18.7	0.01624	0.07438	20.8	0.00619	0.01011	(3600)
	15	0.01039	0.07025	56.9	0.00389	0.00619	45.4	0.01178	0.06622	152.4	0.00432	0.00612	(3600)
	20	0.00746	0.06512	48.8	0.00281	0.00505	32.7	0.00923	0.06951	1607.3	0.00329	0.00524	102.7
	25	0.00500	0.06914	919.2	0.00206	0.00434	3320.2	0.00634	0.06884	2089.9	0.00246	0.00477	202.6
	50	0.00302	0.06724	34.4	0.00122	0.00327	176.1	0.00525	0.06711	(3600)	0.00201	0.00387	56.0
	75	0.00548	0.06463	25.1	0.00227	0.00408	17.9	0.00754	0.06634	382.3	0.00287	0.00463	37.5
0.01	10	0.01171	0.06234	87.9	0.00431	0.00713	281.4	0.01316	0.06393	416.1	0.00470	0.00708	(3600)
	15	0.00696	0.06899	2307.8	0.00269	0.00506	(3600)	0.00841	0.06558	(3600)	0.00312	0.00506	(3600)
	20	0.00498	0.06382	(3600)	0.00196	0.00414	(3600)	0.00623	0.06571	(3600)	0.00238	0.00452	(3600)
	25	0.00337	0.06694	(3600)	0.00147	0.00335	(3600)	0.00439	0.06725	(3600)	0.00184	0.00422	(3600)
	50	0.00179	0.06484	(3600)	0.00075	0.00298	(3600)	0.00351	0.06625	(3600)	0.00140	0.00340	(3600)
	75	0.00543	0.06367	28.4	0.00226	0.00406	16.3	0.00622	0.06592	3199.7	0.00250	0.00386	(3600)

instance depends upon the portfolio chosen, since we are applying our (in-sample) optimised portfolio to out-of-sample data.

In practice, as discussed above, the factor K relating to the number of assets chosen to be in the portfolio is decided according to the preference of the decision-maker. Utilising the formulations presented in this chapter it is clear that the decision-maker can (by utilising historic asset data) gain numeric insight into the effect on in-sample performance of differing values of K. This enables them to make an informed decision as to the value of K to adopt.

3.4 Conclusions

we presented two mixed-integer linear programming formulations for index tracking. In particular we explicitly considered both fixed and variable transaction costs and limited the total transaction cost that could be incurred. We proposed two approaches for the objective function associated with choice of a tracking portfolio, namely; minimise the maximum absolute difference between the tracking portfolio return and index return and minimise the average of the absolute differences between tracking portfolio return and index return. Our formulations are based upon tracking an index by comparing the returns from the index with the returns from the tracking portfolio. The main results of the chapter can be summarised as follows:

- Computational results indicated that good quality out-of-sample results for tracking the indices considered could be achieved.
- The computational times taken for optimisation all the data sets considered were low.

Chapter 4

Quantile Regression for Index Tracking and Enhanced Indexation

Quantile regression differs from traditional least-squares regression in that one constructs regression lines for the quantiles of the dependent variable in terms of the independent variable. In this Chapter we apply quantile regression to two problems in financial portfolio construction, index tracking and enhanced indexation.

We present a mixed-integer linear programming formulation of these problems based on quantile regression and our formulation includes transaction costs, a constraint limiting the number of stocks that can be in the portfolio and a limit on the total transaction cost that can be incurred. Numerical results are presented for eight test problems drawn from major world markets, where the largest of these test data involves over 2000 stocks.

4.1 Introduction

Any reader of this Chapter will probably be familiar with standard least-squares regression. In graphical form that involves plotting a dependent variable ($y_i,\ i = 1, \ldots, n$) against an independent variable ($x_i,\ i = 1, \ldots, n$) and then fitting a straight line, of the form $y = \alpha + \beta x$, to the data. The regression coefficients (α and β) are calculated so as to minimise the sum of squared differences of the actual values from the estimated values, i.e. minimise $\sum_{i=1}^{n}\ (y_i - (\alpha + \beta x_i))^2$. Finding α and β is computationally simple since there exist closed-form equations for their calculation

More technically a regression of this type assumes a distribution of possible y-values at each x_i (one realisation y_i from the distribution that exists at x_i being observed) and the regression relationship captures the linear relationship between the mean y-values at each x_i. It is clear that one may be interested in discerning the relationship between the quantiles of the distribution of y-values at each x_i.

Quantiles are values which divide the cumulative probability distribution. So for example the 50% quantile corresponds to the median of a distribution. The lower and upper quartiles of a distribution correspond to the 25% and 75% quantiles respectively.

In quantile regression, as first defined by Koenker and Bassett (1978), a linear equation relates how the quantiles of the dependent variable vary with the independent variable. Computationally the coefficients in this linear equation cannot be derived in a closed-form fashion, but instead are found as a result of solving a linear program.

Since its inception quantile regression has been widely used. As an indication of this the seminal paper by Koenker and Bassett (1978) has, at the time of writing, over 1500 citations in the Web of Knowledge (http://wok.mimas.ac.uk) and over 4300 citations in Google Scholar (http://scholar.google.com). However, the potential of quantile regression for use in constructing index tracking and enhanced indexation portfolios seems to have been overlooked. In this Chapter we apply the quantile regression technique to two portfolio construction problems, index tracking and enhanced indexation.

The remainder of this Chapter is organized as follows. In Section 2 we give further insight into the quantile regression technique. Our mixed-integer linear programming formulations for index tracking and enhanced indexation based on quantile regression are examined in Section 3. Computational results are presented in Section 4 and finally in Section 5 we present our conclusions.

4.2 Quantile regression

In order to provide insight into quantile regression we show, for a small example, the quantile regression lines and indicate how they are calculated. Readers interested in greater insight into quantile regression are referred to (Koenker and Bassett, 1978; Koenker and Hallock, 2001; Yu et al., 2003; Hao and Naiman, 2007).

Let the dependent variable be $(y_i,\ i = 1, \ldots, n)$, with the independent variable being $(x_i,\ i = 1, \ldots, n)$ and τ the quantile of interest $(0 \leq \tau \leq 1)$ covering the whole distribution. In other words we are interested in the regression line relating the τ'th quantile of y to x. Suppose that this line is $\alpha_\tau + \beta_\tau x_i$, where the regression intercept and slope are dependent on the quantile of interest.

Unlike the ordinary least square approach, the unknown parameters

α_τ and β_τ are estimated by minimising a non-differentiable loss function as illustrated in Figure 4.1 . In particular one minimises the sum of residuals

$$\sum_{i=1}^{n} \rho_\tau(y_i - \alpha_\tau - \beta_\tau x_i) \tag{4.1}$$

where

$$\rho_\tau(u) = u(\tau - I(u < 0)) = \begin{cases} \tau u, & \text{if } u \geq 0 \\ (\tau - 1)u, & \text{if } u < 0 \end{cases}$$

and for $i = 1, 2, \ldots, n$, y_i is the observed response corresponding to independent variables x_i.

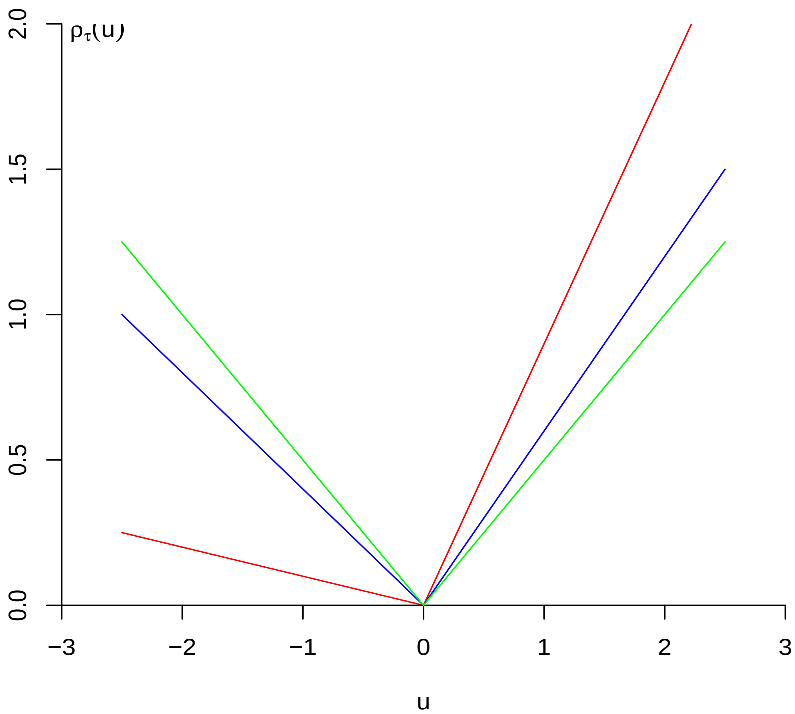

Figure 4.1: Quantile regression check function at τ=0.90 (red line), τ=0.60 (blue line), τ=0.50 (green line)

Define the residual $u_i = y_i - (\alpha_\tau + \beta_\tau x_i)$ then in quantile regression the values of α_τ and β_τ are those that:

$$\text{minimise } \tau[\sum_{i=1,\ u_i \geq 0}^{n} |u_i|] + (1-\tau)[\sum_{i=1,\ u_i<0}^{n} |u_i|] \tag{4.2}$$

In Equation (4.2) the first summation term is the sum of the positive residuals (so the observed value lies above the regression line) and the second summation term is the sum of negative residuals (so the observed value lies below the regression line). Here the positive residuals receive a weight of τ and the negative residuals a weight of $(1-\tau)$. The effect of this is that as τ increases (and we seek to minimise) there will be fewer positive residuals and/or they will be closer to the regression line.

Consider the data shown in Table 4.1. This data is plotted in Figure 4.2 and the quantile regression lines for τ = 0.20, 0.50 and 0.80 are shown in that figure. We also show there the standard least-squares (mean) regression line. The corresponding values for the intercept and slope are shown in Table 4.2. These quantile regression results were produced using the R statistical programming language and the quantreg package (Koenker, R. (2012)).

Table 4.1: Quantile regression example data

Dependent variable (y)	Independent variable (x)
2.21	0.63
3.21	0.88
3.12	0.89
3.21	0.94
3.41	1.13
3.75	1.15
3.20	1.20
5.85	1.23
4.20	1.30

Considering this example we can see that as τ increases the quantile

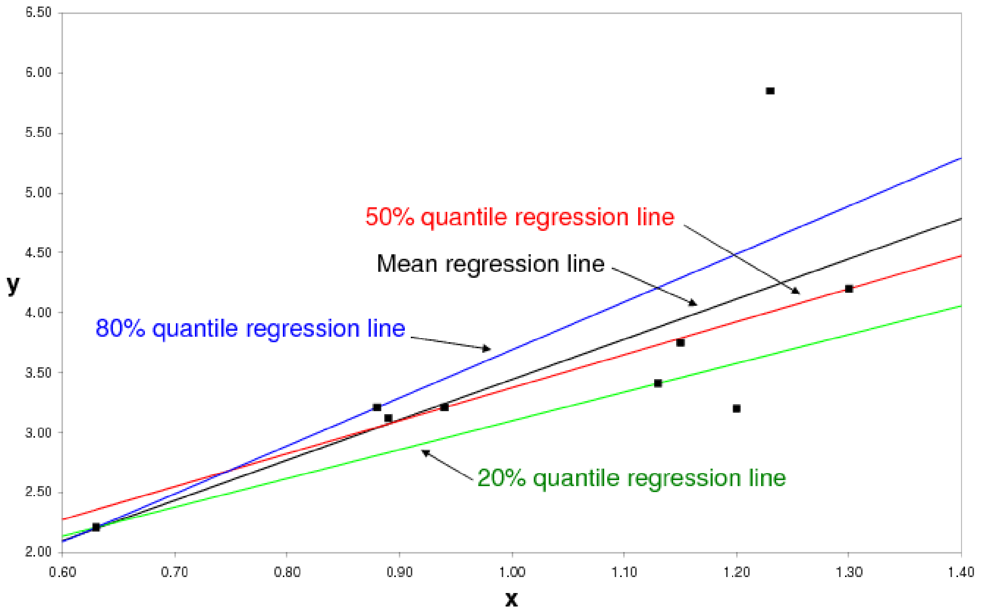

Figure 4.2: Quantile regression example plot

Table 4.2: Regression coefficients

Regression	Intercept	Slope
$\tau = 0.20$	0.69800	2.40000
$\tau = 0.50$	0.62500	2.75000
$\tau = 0.80$	-0.31000	4.00000
mean	0.08372	3.35898

regression lines move upward (so fewer data points lie above the line, more below) as we would expect. Note how the mean regression line and the 50% quantile (median) regression line are different. Note also how both the intercepts and slopes change as τ changes, in particular how the quantile regression lines are not parallel to each other.

Quantile regression has a number of features that the reader should be aware of:

- Firstly, the values for intercept and slope are derived from the solution to a minimisation problem, Equation (4.2). This contrasts with ordinary least-squares (mean) regression where the values for intercept and slope are given by closed-form equations.

- Secondly, as the values for intercept and slope are derived from the solution to a minimisation problem, Equation (4.2), they are not uniquely defined. In other words there may, for a particular value of τ, be two or more sets of values for α_τ and β_τ that achieve the same optimal minimal value for Equation (4.2). This contrasts with ordinary least-squares (mean) regression where the values for intercept and slope are uniquely defined.

- Thirdly, although often the values for intercept and slope are different for different τ values (such as in Table 4.2) this need not be the case. For the data shown in Table 4.1, for example, the quantile regression lines for τ=0.90 and τ=0.95 are identical, each having intercept -1.61200 and slope 6.06667.

In terms of the procedure adopted to solve Equation (4.2) (which is nonlinear) we can linearise it in a standard way. Introduce variables u_i^+ and u_i^-, representing the absolute values of the positive and negative residuals respectively, and then solve:

$$\text{minimise } \tau[\sum_{i=1}^{n} u_i^+] + (1-\tau)[\sum_{i=1}^{n} u_i^-] \tag{4.3}$$

subject to

$$u_i^+ \geq [y_i - (\alpha_\tau + \beta_\tau x_i)] \qquad i = 1, \ldots, n \tag{4.4}$$

$$u_i^- \geq -[y_i - (\alpha_\tau + \beta_\tau x_i)] \qquad i = 1, \ldots, n \tag{4.5}$$

$$u_i^+, u_i^- \geq 0 \qquad i = 1, \ldots, n \tag{4.6}$$

This problem is a linear program and so easily solved computationally to find values for α_τ and β_τ.

4.3 Formulation

In this section we present our mixed-integer linear programming formulation for index tracking and enhanced indexation based upon quantile regression. We first present our notation, then the constraints (note that our notation and constraints are described in the same way as in Section (3.2) with more quantile parameters in this Chapter) and finally the objective. We also highlight what we believe to be the contribution of the work done here.

4.3.1 Notation

Suppose that we observe over time 0,1,2,...,T the value of N assets, as well as the value of the index. In our formulation we have a current portfolio $[X_i,\ i = 1, \ldots, N]$ and we are interested in constructing a new portfolio (containing K assets, where $K < N$) that will (hopefully) perform better than our existing portfolio (either for index tracking or enhanced indexation depending upon our interest). Our formulation can deal with both portfolio creation (where we create a portfolio from cash, equivalently $X_i = 0\ i = 1, \ldots, N$) and portfolio rebalancing (where we change from an existing portfolio to a new portfolio).

Building on the notation of Canakgoz and Beasley (2009) let:

τ be the quantile of interest (e.g. $\tau = 0.50$ for 50% quantile) where $0 \leq \tau \leq 1$

ε_i be the minimum proportion of the portfolio that must be held in asset i if any of the asset is held

δ_i be the maximum proportion of the portfolio that can be held in asset i

V_{it} be the value (price) of one unit of asset i at time t

I_t be the value of the index at time t

R_t be the single period continuous time return for the index at time t, $R_t = \log_e(I_t/I_{t-1})$

r_{it} be the single period continuous time return for asset i at time t, $r_{it} = \log_e(V_{it}/V_{it-1})$

$\hat{\alpha}_{i\tau}$ and $\hat{\beta}_{i\tau}$ be the τ quantile regression intercept and slope for asset i when the returns r_{it} from asset i are regressed against index returns R_t, i.e. the regression equation is that the τ quantile for the return on asset i at time t is given by $\hat{\alpha}_{i\tau} + \hat{\beta}_{i\tau} R_t$

C be the total value (≥ 0) of the current portfolio $[X_i]$ at time T plus cash change (either new cash to be invested or cash to be taken out) so, $C = \sum_{i=1}^{N} V_{iT} X_i$ + cash change

f_i^b be the fractional transaction cost associated with buying one unit of asset i at time T, so that buying one unit of asset i at time T costs $f_i^b V_{iT}$

f_i^s be the fractional transaction cost associated with selling one unit of asset i at time T, so that selling one unit of asset i at time T costs $f_i^s V_{iT}$

γ be the limit ($0 \leq \gamma \leq 1$) on the proportion of C that can be consumed by transaction cost

Then our decision variables are:

x_i the number of units (≥ 0) of asset i that we choose to hold in the new portfolio

G_i the transaction cost (≥ 0) incurred in buying/selling asset i

z_i =1 if any of asset i is held in the new portfolio, =0 otherwise

Without significant loss of generality (since the sums of money involved are large) we allow $[x_i]$ to take fractional values. Note also that as $x_i \geq 0$ we are excluding short selling (shorting) from our model.

4.3.2 Constraints

The constraints of the problem are:

$$\sum_{i=1}^{N} z_i = K \tag{4.7}$$

$$\varepsilon_i z_i \leq x_i V_{iT}/C \leq \delta_i z_i \qquad i = 1, \ldots, N \tag{4.8}$$

$$G_i \geq f_i^s (X_i - x_i) V_{iT} \qquad i = 1, \ldots, N \tag{4.9}$$

$$G_i \geq f_i^b (x_i - X_i) V_{iT} \qquad i = 1, \ldots, N \tag{4.10}$$

$$\sum_{i=1}^{N} G_i \leq \gamma C \tag{4.11}$$

$$\sum_{i=1}^{N} x_i V_{iT} = C - \sum_{i=1}^{N} G_i \tag{4.12}$$

$$x_i, G_i \geq 0 \qquad i = 1, \ldots, N \tag{4.13}$$

$$z_i \in \{0, 1\} \qquad i = 1, \ldots, N \tag{4.14}$$

Equation (4.7) ensures that there are exactly K assets in the portfolio. Equation (4.8) ensures that if an asset i is not in the portfolio ($z_i = 0$) then x_i is also zero; it also ensures that if the asset is chosen to be in the portfolio ($z_i = 1$) then the amount of the asset held satisfies the proportion limits defined. Equations (4.9) and (4.10) define the transaction cost and equation (4.11) limits the total transaction cost incurred. Equation (4.12) is a balance constraint such that the total value of the new portfolio at time T equals the value of the current portfolio at time T plus the cash change (i.e. C) minus the total transaction cost. Equation (4.13) ensures that the continuous variables are non-negative and equation (4.14) is the integrality condition for the zero-one variables.

Although not presented here note that it is a simple matter to extend the formulation given above to represent common situations found in financial portfolio optimisation. These include:

- imposing upper and/or lower limits on the proportion invested in sets of assets (often called class or sector constraints)
- lot size constraints (minimum transaction units) which specify that the holding (x_i) in any asset i must be an integer multiplier of a known constant and/or that the trade in asset i (in moving from X_i to x_i) must be an integer multiplier of a known constant

4.3.3 Objective

Recall that R_t is the single period continuous time return for the index at time t, given by $R_t = \log_e(I_t/I_{t-1})$ for $t = 1, \ldots, T$. Also r_{it} is the single period continuous time return for asset i at time t, given by $r_{it} = \log_e(V_{it}/V_{it-1})$ for

$t = 1, \ldots, T$.

If we quantile regress the returns r_{it} from asset i against the returns R_t from the index we will have a (quantile) regression line with intercept $\hat{\alpha}_{i\tau}$ and slope $\hat{\beta}_{i\tau}$. In other words the regression equation is that the τ quantile for the return on asset i at time t is given by $\hat{\alpha}_{i\tau} + \hat{\beta}_{i\tau} R_t$. As mentioned previously above the values for $\hat{\alpha}_{i\tau}$ and $\hat{\beta}_{i\tau}$ cannot be derived from closed-form equations, but are instead derived from the solution to a linear programming problem. This linear programming problem is solved for each asset i:

$$\text{minimise } \tau[\sum_{t=1}^{T} u_t^+] + (1-\tau)[\sum_{t=1}^{T} u_t^-] \tag{4.15}$$

subject to

$$u_t^+ \geq [r_{it} - (\alpha_\tau + \beta_\tau R_t)] \qquad t = 1, \ldots, T \tag{4.16}$$

$$u_t^- \geq -[r_{it} - (\alpha_\tau + \beta_\tau R_t)] \qquad t = 1, \ldots, T \tag{4.17}$$

$$u_t^+, u_t^- \geq 0 \qquad t = 1, \ldots, T \tag{4.18}$$

So here $\hat{\alpha}_{i\tau}$ and $\hat{\beta}_{i\tau}$ are the optimal values for α_τ and β_τ when this linear program is solved. Equations (4.15)-(4.18) are as Equations (4.3)-(4.6), but particularised to the T observations of return for asset i, $r_{it},\ t = 1, \ldots, T$, and the T observations of index return, $R_t,\ t = 1, \ldots, T$.

Now the weight w_i associated with asset i in the portfolio is given by:

$$w_i = x_i V_{iT}/C \qquad i = 1, \ldots, N \tag{4.19}$$

and the return on the portfolio at time t is given by $\sum_{i=1}^{N} w_i r_{it}$.

If we quantile regress the returns $\sum_{i=1}^{N} w_i r_{it}$ from the portfolio against the returns R_t from the index we will have a (quantile) regression line with a particular intercept and slope. Here we shall assume that we can approximate this quantile regression intercept and slope using the weighted sum of the individual asset quantile regressions, i.e. that:

- the quantile regression intercept for the portfolio can be approximated by $\sum_{i=1}^{N} w_i \hat{\alpha}_{i\tau}$
- the quantile regression slope for the portfolio can be approximated by $\sum_{i=1}^{N} w_i \hat{\beta}_{i\tau}$

In Canakgoz and Beasley (2009) a number of objectives for use in regression-based index tracking were suggested. In that work they used ordinary least-squares (mean) regression. The objective adopted was a three-stage one: where the first stage was to minimise the absolute difference between the regression intercept and zero; the second stage was to minimise the absolute difference between the regression slope and one; the third stage was to minimise total transaction cost. The logic there for the first two stages was that ideally at each time period we would have portfolio return equal to index return. If we could achieve this ideal then the regression line would have intercept zero and slope one. The logic for the third stage was to minimise transaction cost, whilst retaining the values for intercept and slope achieved at the first two stages.

Here we shall adopt the same approach, the difference being that whereas in Canakgoz and Beasley (2009) there was only one set of regression parameters available (as they dealt with just mean regression), here we have many sets of regression parameters (one set for each possible value of τ, where $0 \leq \tau \leq 1$).

A natural choice for index tracking is simply to use $\tau = 0.50$, i.e. to use a 50% (median) regression line. If we can find a portfolio for which the median regression line has an intercept of zero and a slope of one (when portfolio returns are quantile regressed against index returns) then this would seem a good candidate for an index tracking portfolio. Here the logic is that ideally at each time period we would have portfolio return equal to index return. If we could achieve this ideal then the regression line would have intercept zero and slope one.

For enhanced indexation the value of τ to use is less clear cut. However suppose (given a value for τ) we can find a portfolio for which the quantile regression line has an intercept of zero and a slope of one (when portfolio returns are quantile regressed against index returns). This quantile regression line corresponds to the line where the portfolio return is equal to the index return. Above this line the portfolio return exceeds the index return, below this line the portfolio return is less than the index return. In enhanced indexation we seek to out-perform the index, and so we would like to choose a portfolio such that the majority of portfolio returns lie above this quantile regression line (on which the portfolio return is equal to the index return). This implies that the value of τ should be less than 0.50.

Given a value for $\tau < 0.50$, for example $\tau = 0.45$ for the purposes of illustration, then a portfolio for which the quantile regression line has an intercept of zero and a slope of one would have (in quantile terms) 55% of the portfolio returns above the line (so with a return exceeding that of the index), 45% below the line (so with a return less than that of the index), and overall that would seem a reasonable enhanced indexation portfolio.

In the light of the above discussion we can formulate a ***single approach*** which will, depending upon the value of τ adopted, produce an index tracking, or enhanced indexation, portfolio. This approach is:

- First, minimise $\mid \sum_{i=1}^{N} w_i \hat{\alpha}_{i\tau} - 0 \mid$ subject to Equations (4.7)-(4.14),(4.19) to find a portfolio with a quantile regressed intercept as close to zero as possible. This objective is nonlinear, but can be linearised in the same manner as was done above (in going from Equation (4.2) to Equations (4.3)-(4.6)). Introduce a variable $D \geq 0$ and solve:

$$\text{minimise } D \tag{4.20}$$

subject to

$$D \geq [\sum_{i=1}^{N} w_i \hat{\alpha}_{i\tau}] \tag{4.21}$$

$$D \geq -[\sum_{i=1}^{N} w_i \hat{\alpha}_{i\tau}] \tag{4.22}$$

and Equations (4.7)-(4.14),(4.19)

This problem is a mixed-integer linear program and so can be solved using standard software, such as IBM ILOG Cplex IBM ILOG Cplex Solver (2012) which we used. Let the optimal solution value be D^*.

- Second, minimise $| \sum_{i=1}^{N} w_i \hat{\beta}_{i\tau} - 1 |$ subject to Equations (4.7)-(4.14),(4.19)-(4.21),(4.22) and $D = D^*$ to find a portfolio with a quantile regressed slope as close to one as possible, but which retains the minimal value (D^*) achieved previously. Again this is nonlinear, but is easily linearised by introducing a variable $E \geq 0$ and solving

$$\text{minimise } E \tag{4.23}$$

subject to

$$E \geq [\sum_{i=1}^{N} w_i \hat{\beta}_{i\tau} - 1] \tag{4.24}$$

$$E \geq -[\sum_{i=1}^{N} w_i \hat{\beta}_{i\tau} - 1] \tag{4.25}$$

$$D = D^* \tag{4.26}$$

and Equations (4.7)-(4.14),(4.19),(4.21),(4.22)

Let the optimal solution value be E^*.

- Third, minimise $\sum_{i=1}^{N} G_i$ subject to Equations (4.7)-(4.14),(4.19),(4.21),(4.22)-(4.24),(4.26) and $E = E^*$ to find a portfolio with as low a transaction cost as possible but which retains the minimal values (D^* and E^*) achieved previously.

4.3.4 Contribution

A number of the constraints in our formulation are as seen in other work (Beasley et al., 2003; Canakgoz and Beasley, 2009; Guastaroba and Speranza, 2012) , as indeed they are seen in other papers by other authors. This is natural since constraints for many optimisation problems are often expressed mathematically exactly as in previous work in the literature. ***The contribution of this Chapter lies not in the constraints presented, but rather in the application of quantile regression to the problem of constructing financial portfolios for index tracking and enhanced indexation***. To the best of our knowledge this is the first time that quantile regression has been applied to these problems. Moreover the quantile concept means that we can, within the same model/approach, easily capture both index tracking and enhanced indexation objectives.

4.4 Computational results

4.4.1 Test problems

To test our formulation we used the same test problems as in Chapter 3, Section 3.3.1 but added three more data sets for large capital market indices Table 4.3.

Table 4.3: Test problems

Index	Number of stocks N	Number of selected stocks K
Hang Seng	31	10
DAX 100	85	10
FTSE 100	89	10
S&P 100	98	10
Nikkei 225	225	10
S&P 500	457	40
Russell 2000	1318	90
Russell 3000	2151	70

The computational results presented below (Windows dual Xenon 3.06GHz Pentium pc with 2Gb memory) are for our approach as coded in FORTRAN and AMPL using ILOG Cplex (version 12.1) IBM ILOG Cplex Solver (2012) as the mixed-integer optimiser. We used Cplex default parameter settings, except that we changed the tolerance parameters so as to find the genuine optimal solution. The reason for this is that Cplex, by default, finds a solution within a specified tolerance of the genuine optimal and since we have real-valued objective functions we wanted to avoid the situation where we missed the genuine optimal solution. In detail we used the Cplex commands *set mip tol mip 0* and *set mip tol abs 0* to set the tolerance parameters to zero to find the genuine optimal solution. For those readers more familiar with Cplex these correspond to setting both the Cplex parameters EpGap and EpAGap dealing with relative and absolute mixed-integer tolerances to zero. Quantile regression was performed using the R statistical programming language and the quantreg package Koenker, R. (2012). We:

- used an initial portfolio of value 10^6 composed of the first K stocks in equal proportions, i.e. $X_i = (10^6/K)/V_{i0} \;\; i = 1, ..., K$; $X_i = 0 \; \forall i > K$
- used $\varepsilon_i = 0.01$ and $\delta_i = 1 \; \forall i$, i.e. a lower proportion limit for any asset that appears in the decided portfolio of one percent
- used $f_i^b = f_i^s = 0.01 \; \forall i$, i.e. transaction cost was one percent of the value of the assets bought/sold

4.4.2 Index tracking

To illustrate how our approach performs in terms of index tracking we took an in-sample time period [0,145] for each of our test problems and, for a range of values for the transaction cost limit γ, used the approach given above to decide a tracking portfolio. We then calculated the quantile regression coefficients ($\tau = 0.50$) when the returns from this tracking portfolio were regressed against the index out-of sample (in [146, 290]). Table 4.4 gives the results obtained. In that Table we give:

- the optimal values D^* and E^* associated with intercept and slope respectively
- the computation time taken in seconds (this includes both the time taken to calculate the quantile regression coefficients as well as the time taken by the Cplex optimiser to solve our mixed-integer programs to proven optimality)
- the out-of-sample intercept and slope quantile regression coefficients

Table 4.4: In-sample and out-of-sample tracking results

Index	Transaction	In-sample		Time	Out-of-sample	
(N, K)	cost limit γ	D^*	E^*	(secs)	Intercept	Slope
Hang Seng	0.0025	0	0.00858	0.9	-0.00127	0.94676
(31,10)	0.0050	0	0	0.6	-0.00287	0.98818
	0.0075	0	0	0.8	-0.00287	0.98818
	0.01	0	0	0.7	-0.00287	0.98818
DAX 100	0.0025	0	0.26392	1.8	0.00227	0.93844
(85,10)	0.0050	0	0.11761	1.6	0.00213	1.12548
	0.0075	0	0.01081	1.8	0.00017	1.15890
	0.01	0	0	1.5	0.00045	1.15425
FTSE 100	0.0025	0.00067	0.09893	1.9	-0.00006	0.87620
(89,10)	0.0050	0	0	1.8	0.00087	0.81862
	0.0075	0	0	2.0	0.00087	0.81862
	0.01	0	0	2.4	0.00087	0.81862
S&P 100	0.0025	0	0	2.0	0.00079	0.86235
(98,10)	0.0050	0	0	2.0	0.00079	0.86235
	0.0075	0	0	2.3	0.00079	0.86235
	0.01	0	0	2.1	0.00079	0.86235
Nikkei 225	0.0025	0	0.05195	4.7	0.00054	0.98544
(225,10)	0.0050	0	0	5.0	-0.00023	1.03811
	0.0075	0	0	8.6	-0.00023	1.03811
	0.01	0	0	5.1	-0.00023	1.03811
S&P 500	0.0025	0.00014	0.17560	15.0	0.00200	1.21864
(457,40)	0.0050	0	0	10.9	0.00204	1.25729
	0.0075	0	0	14.0	0.00204	1.25729
	0.01	0	0	10.7	0.00204	1.25729
Russell 2000	0.0025			infeasible		
(1318,90)	0.0050	0	0	56.0	0.00188	1.22387
	0.0075	0	0	31.9	0.00188	1.22387
	0.01	0	0	31.1	0.00188	1.22387
Russell 3000	0.0025	0	0	62.1	0.00343	1.10307
(2151,70)	0.0050	0	0	67.3	0.00343	1.10307
	0.0075	0	0	71.4	0.00328	1.10784
	0.01	0	0	75.5	0.00328	1.10784
Average		0.00003	0.02346	16.0	0.00090	1.04044

To illustrate Table 4.4 consider the Nikkei 225 with $N = 225$ assets and $K = 10$ assets in the tracking portfolio. When $\gamma = 0.0025$, so a transaction cost limit of 0.25% of portfolio value, the minimal D^* value is zero and the minimal E^* value is 0.05195 with the total computation time being 4.7 seconds. Out-of-sample the quantile regression intercept is 0.00054 and the quantile regression slope is 0.98544. As we are considering index tracking here we are using $\tau = 0.50$.

One of the problems (Russell 2000 with $\gamma = 0.0025$) in Table 4.4 is infeasible, indicating that it is not possible to trade from the initial portfolio to a portfolio with $K = 90$ assets that satisfies the lower proportion constraint ($\varepsilon_i = 0.01 \ \forall i$) within the transaction cost limit.

One feature of the results in Table 4.4 is that in some cases the out-of-sample results are identical for different transaction cost limit (γ) values. This is a direct consequence of the approach we have used, in that if it is possible to achieve minimal values of $D^* = E^* = 0$ at one transaction cost level then it is possible to achieve the same minimal zero values at higher transaction cost limit values (since the last step in our solution approach minimises transaction cost incurred, whilst preserving the minimal values of D^* and E^* achieved). Hence, once $D^* = E^* = 0$ has been achieved at one transaction cost limit, all higher transaction cost limits can potentially give the same tracking portfolio (any difference being due to multiple optimal solutions, i.e. different portfolios each with $D^* = E^* = 0$ and the same transaction cost associated with trading from the initial portfolio to the decided portfolio). Multiple optimal solutions can occur, for example for the Russell 3000 with $\gamma = 0.0050$ and $\gamma = 0.0075$ we have identical minimal values of zero for both D^* and E^*, but different out-of-sample values, implying the decided portfolios for these two cases must have been different.

Considering the averages at the foot of Table 4.4 we can observe that computation times are very reasonable, an average of 16 seconds with no problem taking more than 76 seconds. In-sample the average value of D^* is very close to zero and the average value of E^* is also close to zero. Out-of-sample (over the time period [146,290], so over nearly three years of weekly observations), the average quantile regression intercept is very close to zero with the average quantile regression slope differing from one by only 0.04.

With regard to the absolute difference between the out-of-sample intercept and zero, and the absolute difference between the out-of-sample slope and one, then for Table 4.4 these values are 0.00159 and 0.12969 respectively (averaged over all cases in Table 4.4). The maximum values for absolute difference between the out-of-sample intercept and zero, and absolute difference between the out-of-sample slope and one, in Table 4.4 are 0.00343 and 0.25729 respectively. Clearly it is a matter of judgment, but given that we have a fixed value for the number of assets (K) in the portfolio much smaller than the number of assets (N) in the index, and given that these values are over nearly three years of weekly observations, they do not appear too unreasonable.

The index tracking test problems shown in Table 4.4 were also considered in Beasley et al. (2003) and Canakgoz and Beasley (2009). With respect to Beasley et al. (2003) a direct comparison is difficult as they adopt a nonlinear objective. With respect to Canakgoz and Beasley (2009) then, as they also used regression (albeit mean, least-squares, regression), a comparison can be made.

In Canakgoz and Beasley (2009) there are 27 zero D^* values and four non-zero D^* values (0.00076, 0.00053, 0.00077 and 0.00209 in Table 2 of Canakgoz and Beasley (2009)). Hence the average D^* value is $(0.00076 + 0.00053 + 0.00077 + 0.00209)/31 = 0.00013$. There are 25 zero E^* values and six E^* values that are different from zero ($|0.94513 - 1|$, $|0.68501 - 1|$, $|0.82413 - 1|$, $|0.96548 - 1|$, $|1.00937 - 1|$ and $|1.15789 - 1|$ in Table 2 of Canakgoz and Beasley (2009)). Hence the average E^* value is $(|0.94513 - 1| + |0.68501 - 1| + |0.82413 - 1| +$

$|0.96548 - 1| + |1.00937 - 1| + |1.15789 - 1|)/31 = 0.02411$.

Both of these values (0.00013 and 0.02411 respectively), associated with mean regression, are higher than the same values seen at the foot of Table 4.4 associated with median quantile regression. Hence with respect to in-sample statistics, at least for the test problems considered, it seems that (on average) median regression is better than mean (least-squares) regression.

Out-of-sample the picture is more mixed. In Canakgoz and Beasley (2009) the average out-of-sample intercept is 0.00106, worse than the corresponding (median) intercept of 0.00090 seen at the foot of Table 4.4. However in Canakgoz and Beasley (2009) the average out-of-sample slope is 0.99725, which is closer to one than the corresponding (median) slope of 1.04044 seen at the foot of Table 4.4.

4.4.3 Enhanced indexation

Tables 4.5 and 4.6 deal with the same problems as Table 4.4, but for enhanced indexation. In Table 4.5 we present results for $\tau = 0.45$ and in Table 4.6 we present results for $\tau = 0.40$. These tables have the same format as Table 4.4, but include an additional column relating to the average excess return (AER). Here AER is defined as the average yearly out-of-sample (percentage) excess return, return over and above index return in the same period, as computed directly from portfolio returns in the out-of-sample period. Here as we are seeking an enhanced indexation portfolio we are seeking excess return, return over and above the index.

Recalling that we are dealing with weekly data, average yearly out-of-sample (percentage) excess return (AER), is defined as:

$$[5200/145] \sum_{t=146}^{290} [log_e(\sum_{i=1}^{N} X_i^{opt} V_{it} / \sum_{i=1}^{N} X_i^{opt} V_{it-1}) \ - R_t] \qquad (4.27)$$

where X_i^{opt} is the number of units of asset i held in the portfolio as given by the optimisation process (after minimisation of transaction cost).

Comparing Tables 4.5 and 4.6 we can see that, whereas in-sample in Table 4.5 we have a significant number of zero values for D^* and/or E^*, in Table 4.6 we have no such values. This indicates that in moving from quantile regression with $\tau = 0.45$ to quantile regression with $\tau = 0.40$ it is much harder to find a regression line with the desired intercept and slope.

Here the desired intercept is zero, equivalently $D^* = 0$, the desired slope is one, equivalently $E^* = 0$.

Considering the averages at the feet of Tables 4.5 and 4.6 we can observe that computation times are very reasonable, across both tables an average of no more than 18 seconds with no problem taking more than 129 seconds. In-sample the average value of D^* is very close to zero in both Tables 4.5 and 4.6, but it is clear that there is a significant difference in the average values of E^* between these tables (0.12825 in Table 4.5 associated with $\tau = 0.45$, but 0.30663 in Table 4.6 associated with $\tau = 0.40$). Out-of-sample (over the time period [145,290], so over nearly three years of weekly observations), in Tables 4.5 and 4.6 the average quantile regression intercept is very close to zero with the average quantile regression slope much closer to one for $\tau = 0.45$ than for $\tau = 0.40$. With regard to the absolute difference between the out-of-sample intercept and zero, and the absolute difference between the out-of-sample slope and one, then for Table 4.5 these values are 0.00135 and 0.12063 respectively (averaged over all cases in Table 4.5), whilst the corresponding values for Table 4.6 are

Table 4.5: In-sample and out-of-sample enhanced indexation results, $\tau = 0.45$

Index	Transaction	In-sample		Time	Out-of-sample		
(N, K)	cost limit γ	D^*	E^*	(secs)	Intercept	Slope	AER
Hang Seng	0.0025	0.00243	0.13725	0.8	-0.00216	0.88354	-5.03
(31,10)	0.0050	0.00137	0.19561	0.7	-0.00312	0.85191	-8.35
	0.0075	0.00039	0.25405	0.7	-0.00299	0.84681	-9.72
	0.01	0	0.20148	0.8	-0.00338	0.84604	-9.79
DAX 100	0.0025	0.00157	0.43969	2.1	0.00041	0.94156	10.85
(85,10)	0.0050	0.00049	0.48290	1.8	0.00130	0.96133	11.01
	0.0075	0	0.36638	1.6	-0.00032	1.00280	11.74
	0.01	0	0.09469	1.7	-0.00163	1.26620	16.13
FTSE 100	0.0025	0	0.15674	1.8	0.00096	0.83021	3.01
(89,10)	0.0050	0	0.04185	1.8	0.00188	0.90412	6.00
	0.0075	0	0	1.8	0.00246	0.87509	7.13
	0.01	0	0	1.6	0.00246	0.87509	7.13
S&P 100	0.0025	0.00050	0.00959	2.0	-0.00052	1.13749	5.32
(98,10)	0.0050	0	0	1.9	-0.00071	1.17571	6.71
	0.0075	0	0	2.0	-0.00071	1.17571	6.71
	0.01	0	0	2.2	-0.00071	1.17571	6.71
Nikkei 225	0.0025	0.00263	0.14958	4.7	-0.00116	0.91166	3.19
(225,10)	0.0050	0.00145	0.19030	4.4	-0.00019	0.81163	7.11
	0.0075	0.00051	0.18358	4.8	-0.00048	0.79613	9.90
	0.01	0	0.01904	4.5	-0.00220	0.98295	1.73
S&P 500	0.0025	0.00122	0.31956	9.3	-0.00242	1.41017	0.63
(457,40)	0.0050	0	0.20342	10.5	-0.00097	1.16534	2.77
	0.0075	0	0	10.5	-0.00219	0.93692	6.38
	0.01	0	0	10.1	-0.00219	0.93692	6.38
Russell 2000	0.0025			infeasible			
(1318,90)	0.0050	0.00403	0.11128	61.2	0.00052	1.07666	7.09
	0.0075	0.00120	0.13308	32.0	-0.00056	1.00869	5.03
	0.01	0	0	30.2	-0.00036	0.98825	1.99
Russell 3000	0.0025	0.00589	0.07351	56.0	-0.00048	1.11735	16.25
(2151,70)	0.0050	0.00261	0.10108	49.9	-0.00080	1.01473	14.31
	0.0075	0.00019	0.11094	47.2	-0.00024	1.01730	12.29
	0.01	0	0	53.4	0.00124	1.17578	10.89
Average		0.00085	0.12825	13.4	-0.00062	1.00322	5.53

0.00318 and 0.22441. The maximum values for absolute difference between the out-of-sample intercept and zero, and absolute difference between the out-of-

Table 4.6: In-sample and out-of-sample enhanced indexation results, $\tau = 0.40$

Index	Transaction	In-sample		Time	Out-sample		
(N, K)	cost limit γ	D^*	E^*	(secs)	Intercept	Slope	AER
Hang Seng	0.0025	0.00693	0.14154	0.7	-0.00406	0.88816	-4.35
(31,10)	0.0050	0.00585	0.20572	0.7	-0.00466	0.84234	-6.38
	0.0075	0.00483	0.26515	0.7	-0.00446	0.80993	-8.60
	0.01	0.00383	0.31817	0.8	-0.00480	0.79307	-10.94
DAX 100	0.0025	0.00461	0.46180	2.1	-0.00214	0.99809	10.15
(85,10)	0.0050	0.00340	0.47403	2.1	-0.00239	0.95701	10.45
	0.0075	0.00243	0.48319	1.7	-0.00181	0.95496	10.50
	0.01	0.00177	0.50297	1.7	-0.00155	0.97142	11.43
FTSE 100	0.0025	0.00248	0.28769	1.9	-0.00248	0.76935	-0.59
(89,10)	0.0050	0.00171	0.38288	1.7	-0.00378	0.76222	-1.13
	0.0075	0.00104	0.46112	1.7	-0.00415	0.70118	-0.35
	0.01	0.00052	0.51713	1.8	-0.00348	0.65650	1.08
S&P 100	0.0025	0.00509	0.05129	2.0	-0.00272	1.10230	5.31
(98,10)	0.0050	0.00401	0.12515	2.1	-0.00246	1.26613	9.46
	0.0075	0.00330	0.16287	2.0	-0.00236	1.25671	12.73
	0.01	0.00272	0.29411	2.1	-0.00286	1.39590	16.93
Nikkei 225	0.0025	0.00507	0.15464	4.5	-0.00234	0.88336	-1.22
(225,10)	0.0050	0.00411	0.21205	4.5	-0.00178	0.81851	-1.17
	0.0075	0.00336	0.21987	4.2	-0.00337	0.77910	-3.65
	0.01	0.00266	0.25214	4.2	-0.00458	0.73670	-7.00
S&P 500	0.0025	0.00630	0.22979	10.1	-0.00349	1.37734	-0.36
(457,40)	0.0050	0.00459	0.25440	9.1	-0.00548	1.33654	-2.40
	0.0075	0.00340	0.28019	8.6	-0.00773	1.43651	-6.94
	0.01	0.00257	0.29523	8.9	-0.01144	1.61619	-9.63
Russell 2000	0.0025			infeasible			
(1318,90)	0.0050	0.01011	0.31803	67.4	0.00063	0.96141	11.53
	0.0075	0.00675	0.42647	73.3	-0.00131	0.80416	9.44
	0.01	0.00480	0.48769	30.7	-0.00132	0.63131	7.45
Russell 3000	0.0025	0.01235	0.13736	128.1	-0.00203	1.02016	19.22
(2151,70)	0.0050	0.00874	0.29088	47.0	-0.00110	0.76174	21.88
	0.0075	0.00609	0.36847	45.7	0.00106	0.70167	20.02
	0.01	0.00405	0.44342	83.0	0.00085	0.66881	18.86
Average		0.00450	0.30663	17.9	-0.00302	0.95673	4.25

sample slope and one, in Table 4.5 are 0.00338 and 0.41017 respectively, whilst the corresponding values for Table 4.6 are 0.01144 and 0.61619.

From the out-of-sample AER columns in Tables 4.5 and 4.6 we can see that the average excess return (return over and above the index) is 5.53% per year for $\tau = 0.45$, but only 4.25% per year for $\tau = 0.40$. Looking in greater detail at the AER columns in Tables 4.5 and 4.6 we have much wider variability in terms of AER as τ decreases from 0.45 to 0.40. In the AER column in Table 4.5 associated with $\tau = 0.45$ we have only 4 values that are negative, the other 27 values are positive. By contrast in Table 4.6 associated with $\tau = 0.40$ we have 15 values that are negative, 16 values that are positive.

Table 4.7 shows the average out-of-sample annual excess returns (AER) values for a number of different τ values. These average values seen are the averages over the four transaction cost limits considered. The values presented in Table 4.7 for $\tau = 0.45$ and $\tau = 0.40$ can be deduced from Tables 4.5 and 4.6 respectively. For the other values of τ considered we have presented detailed results in Appendix A.

In Table 4.7 we have, for the Hang Seng for example, that the average out-of-sample annual excess return is -8.22% for $\tau = 0.45$, implying that out-of-sample the portfolios chosen do worse than the index (recall that this value of -8.22% is an average over four different transaction cost limits γ). For the Hang Seng for $\tau = 0.30$, by contrast, the AER value is 3.52%, implying that out-of-sample the portfolios chosen out-perform the index by an average of 3.52% per year.

The final column in Table 4.7 gives the correlation between the nine values of τ and the associated AER values seen. For the Hang Seng for example this is -0.69, the minus sign here implying that (in general) as τ decreases the value of AER increases. With nine observations then (utilising standard statistical

Table 4.7: Out-of-sample enhanced indexation average AER

Index (N, K)	Value of τ									Correlation
	0.45	0.40	0.35	0.30	0.25	0.20	0.15	0.10	0.05	
Hang Seng (31,10)	-8.22	-7.57	4.26	3.52	3.11	3.32	3.29	2.08	3.84	-0.69
DAX 100 (85,10)	12.43	10.63	12.70	13.85	14.22	14.06	12.29	10.78	3.14	0.50
FTSE 100 (89,10)	5.82	-0.25	-0.15	-0.20	-0.76	-0.71	-0.52	-4.31	2.49	0.45
S&P 100 (98,10)	6.36	11.11	-6.09	-7.76	-6.86	-2.33	-7.55	-8.64	-8.97	0.74
Nikkei 225 (225,10)	5.48	-3.26	-4.42	3.85	3.64	-1.56	4.42	7.62	8.07	-0.55
S&P 500 (457,40)	4.04	-4.83	-0.36	9.54	6.82	9.20	7.15	5.79	5.52	-0.52
Russell 2000	4.70	9.47	10.86	10.46	11.00	12.76	12.20	11.76	38.79	-0.69
Russell 3000	13.44	20.00	20.75	20.66	21.36	35.42	35.29	35.51	34.10	-0.91

tables) any correlation coefficient whose modulus is greater than 0.67 can be said to be significantly different from zero at the 5% significance level, so here four of the eight indices (Hang Seng, S&P 100, Russell 2000, Russell 3000) display significant correlations.

Note here how in Table 4.7 the correlation coefficient splits the indices into two sets. One set, containing the DAX 100, FTSE 100 and S&P 100, where the correlation coefficient is positive, and hence AER decreases as τ decreases. The other set containing the remaining five indices with a negative correlation, so AER increases as τ decreases.

Considering Table 4.7 we can see that for three of the eight test problems (DAX 100, Russell 2000, Russell 3000) we can consistently generate positive AER values irrespective of the value of τ chosen. For the Hang Seng positive AER values are not seen until τ is 0.35 or less. In general these results indicate that the value of τ to adopt will be dependent on the index/market considered and hence computational investigation is needed to decide an appropriate τ value for any particular index/market at any particular time.

Directly comparing the results in this work with the results in Canakgoz and Beasley (2009) for each test problem is difficult as in Canakgoz and Beasley (2009) a number of enhanced indexation results are presented based on desired excess return, whereas in this work we generate enhanced indexation results based on varying τ. However, across all test problems and all relevant results presented, the average out-of-sample AER value in Canakgoz and Beasley (2009) is 6.08%. Averaging the out-of-sample AER values seen in Table 4.7 we obtain 7.05%. In terms of out-of-sample performance therefore we can conclude that the results presented indicate that the quantile regression approach presented in this chapter is (on average) competitive with the approach given in Canakgoz and Beasley (2009).

4.4.4 Alternative approaches

As mentioned above our approach is based on that given in Canakgoz and Beasley (2009). Conceptually we have three factors of interest, each of which we wish to minimise:

- the absolute difference between the regression intercept and zero;
- the absolute difference between the regression slope and one;
- total transaction cost.

Above we have adopted a three-stage approach, where the first stage was to minimise the absolute difference between the regression intercept and zero; the second stage was to minimise the absolute difference between the regression slope and one; the third stage was to minimise total transaction cost. In their paper Canakgoz and Beasley (2009) do discuss how alternative approaches can be formed based on these three factors, but do not give any computational results. In this chapter we do consider alternative approaches to the one discussed above.

Clearly with three factors of interest there are a number of alternative approaches, e.g.

1. first minimise the absolute difference between the regression slope and one (minimise E); then minimise the absolute difference between the regression intercept and zero (minimise D); then minimise total transaction cost (minimise $\sum_{i=1}^{N} G_i$)

2. first minimise a weighted sum of the absolute difference between the regression intercept and zero and the absolute difference between the regression slope and one (minimise $\lambda_1 D + \lambda_2 E$); then minimise total transaction cost (minimise $\sum_{i=1}^{N} G_i$); where λ_1 and λ_2 are the weighting parameters

3. minimise a weighted sum of all three factors, so minimise a weighted sum of the absolute difference between the regression intercept and zero, the absolute difference between the regression slope and one and total transaction cost (minimise $\lambda_1 D + \lambda_2 E + \lambda_3 \sum_{i=1}^{N} G_i$; where λ_1, λ_2 and λ_3 are the weighting parameters)

In all of these minimisations we (in the same manner as presented above) retain the optimal value from the previous minimisation at each stage.

In this section we will present computational results for the first two of these approaches. The difficulty with the third approach is that it is not (in our mind) clear how to set values for the weights ($\lambda_1, \lambda_2,\ \lambda_3$) so as to weigh together two factors which are absolute differences (both small, see the values for D^* and E^* in the tables of computational results, Tables 4.4-4.6, presented above) and transaction cost, which is in monetary units and possibly large. For that reason we have not explored the third approach computationally here.

First alternative approach

In this section we present results for the first alternative approach above, namely first minimise the absolute difference between the regression slope and one (minimise E); then minimise the absolute difference between the regression intercept and zero (minimise D); then minimise total transaction cost (minimise $\sum_{i=1}^{N} G_i$).

Table 4.8 and Table 4.9 have the same format as Table 4.4 and Table 4.7, but are for this first alternative approach. Detailed tables (first alternative approach) of results for individual τ values (such as given in Table 4.5 and Table 4.6) are given in the Appendix B.

Comparing Table 4.4 and Table 4.8, which are the index tracking results, we see that (as we would expect) any cases in Table 4.4 where the values for D^* and E^* are both zero also have both D^* and E^* zero in Table 4.8. It is clear that since for 24 of the 31 cases seen there both D^* and E^* are zero it is hard to draw any conclusions from the 7 cases where there are non-zero values. On the limited evidence available in those tables it seems that either of the two approaches produces good results for index tracking.

Comparing Table 4.7 and Table 4.9 we can see that some of the indices have positive correlations between the value of τ and the average AER, some negative. Except for the Nikkei 225 all of these correlations have the same sign in Table 4.9 as in Table 4.7. These tables each contain average AER values for 8 indices and 9 different values of τ, so 72 cases in total. In only 21 of these 72 cases is the AER value in Table 4.9 better than (or equal to) the value in Table 4.7. Over all 72 cases the average AER in Table 4.9 is 4.24%, as compared with 7.02% in Table 4.7.

Table 4.8: In-sample and out-of-sample tracking results, first alternative

Index	Transaction	In-sample		Time	Out-of-sample	
(N, K)	cost limit γ	D^*	E^*	(secs)	Intercept	Slope
Hang Seng	0.0025	0.00011	0	0.8	-0.00289	0.97444
(31,10)	0.0050	0	0	0.7	-0.00287	0.98818
	0.0075	0	0	0.7	-0.00287	0.98818
	0.01	0	0	0.7	-0.00287	0.98818
DAX 100	0.0025	0.00047	0.25081	2.0	0.00377	1.01353
(85,10)	0.0050	0.00024	0.11533	1.7	0.00170	1.13299
	0.0075	0.00037	0.00834	1.7	0.00088	1.14627
	0.01	0	0	1.8	0.00045	1.15425
FTSE 100	0.0025	0.00193	0.05019	2.1	-0.00041	0.82096
(89,10)	0.0050	0	0	1.8	0.00087	0.81862
	0.0075	0	0	1.8	0.00087	0.81862
	0.01	0	0	2.0	0.00087	0.81862
S&P 100	0.0025	0	0	1.9	0.00079	0.86235
(98,10)	0.0050	0	0	3.0	0.00079	0.86235
	0.0075	0	0	1.9	0.00079	0.86235
	0.01	0	0	1.9	0.00079	0.86235
Nikkei 225	0.0025	0.00077	0	5.4	0.00059	1.00925
(225,10)	0.0050	0	0	5.0	-0.00023	1.03811
	0.0075	0	0	4.7	-0.00023	1.03811
	0.01	0	0	13.9	-0.00023	1.03811
S&P 500	0.0025	0.00155	0	15.4	0.00124	1.24486
(457,40)	0.0050	0	0	9.6	0.00204	1.25729
	0.0075	0	0	10.7	0.00204	1.25729
	0.01	0	0	9.7	0.00204	1.25729
Russell 2000	0.0025			infeasible		
(1318,90)	0.0050	0	0	30.9	0.00188	1.22387
	0.0075	0	0	30.3	0.00188	1.22387
	0.01	0	0	29.3	0.00188	1.22387
Russell 3000	0.0025	0	0	63.7	0.00343	1.10307
(2151,70)	0.0050	0	0	55.4	0.00343	1.10307
	0.0075	0	0	63.4	0.00328	1.10784
	0.01	0	0	71.4	0.00328	1.10784
Average		0.00018	0.01370	14.4	0.00087	1.04342

It seems reasonable to conclude therefore that (on average) this alternative

Table 4.9: Out-of-sample enhanced indexation average AER, first alternative

Index (N, K)	Value of τ									Correlation
	0.45	0.40	0.35	0.30	0.25	0.20	0.15	0.10	0.05	
Hang Seng (31,10)	-5.27	-5.24	0.94	0.67	-0.58	-0.21	-0.68	0.45	3.57	-0.77
DAX 100 (85,10)	5.22	5.16	4.86	5.00	5.56	7.22	5.19	0.67	-0.63	0.62
FTSE 100 (89,10)	6.10	-1.10	1.42	1.73	5.89	3.62	-3.04	2.05	3.16	0.14
S&P 100 (98,10)	6.31	6.56	-2.48	1.66	-5.68	-4.83	-4.40	-5.89	-2.94	0.78
Nikkei 225 (225,10)	0.69	-2.14	-5.05	-6.52	-8.14	-4.08	-5.20	-4.52	-4.31	0.45
S&P 500 (457,40)	6.93	1.72	5.07	5.46	5.05	7.61	7.08	6.99	7.02	-0.56
Russell 2000	4.42	4.57	5.56	7.91	8.71	9.49	10.13	11.17	12.15	-0.99
Russell 3000	12.56	12.38	14.61	14.80	15.72	21.26	27.35	27.68	30.97	-0.95

approach, leading to the out-of-sample results shown in Table 4.9, is worse than the approach given before which leads to the results shown in Table 4.7.

Second alternative approach

In this section we present results for the second alternative approach above, namely first minimise a weighted sum of the absolute difference between the regression intercept and zero and the absolute difference between the regression slope and one (minimise $\lambda_1 D + \lambda_2 E$); then minimise total transaction cost (minimise $\sum_{i=1}^{N} G_i$); where λ_1 and λ_2 are the weighting parameters.

Clearly there are a multiplicity of values that λ_1 and λ_2 can take but here we shall just consider the case $\lambda_1 = \lambda_2 = 1$. This corresponds to first minimising $D + E$, then minimising total transaction cost. The results for this alternative are shown in Table 4.10 and Table 4.11.

Comparing Table 4.8 and Table 4.10 we can see that this second alternative is (for index tracking at least) effectively identical to the first alternative examined above. In-sample the results for D^* and E^* are identical,

Table 4.10: In-sample and out-of-sample tracking results, second alternative

Index	Transaction	In-sample		Time	Out-of-sample	
(N, K)	cost limit γ	D^*	E^*	(secs)	Intercept	Slope
Hang Seng	0.0025	0.00011	0	0.8	-0.00289	0.97445
(31,10)	0.0050	0	0	0.6	-0.00287	0.98818
	0.0075	0	0	0.6	-0.00287	0.98818
	0.01	0	0	0.6	-0.00287	0.98818
DAX 100	0.0025	0.00047	0.25081	1.9	0.00377	1.01353
(85,10)	0.0050	0.00024	0.11533	1.5	0.00170	1.13299
	0.0075	0.00037	0.00834	1.5	0.00088	1.14627
	0.01	0	0	1.5	0.00045	1.15425
FTSE 100	0.0025	0.00193	0.05019	2.2	-0.00041	0.82096
(89,10)	0.0050	0	0	2.0	0.00087	0.81862
	0.0075	0	0	2.0	0.00087	0.81862
	0.01	0	0	1.8	0.00087	0.81862
S&P 100	0.0025	0	0	2.1	0.00079	0.86234
(98,10)	0.0050	0	0	1.9	0.00079	0.86234
	0.0075	0	0	1.9	0.00079	0.86234
	0.01	0	0	2.2	0.00079	0.86234
Nikkei 225	0.0025	0.00077	0	5.3	0.00059	1.00925
(225,10)	0.0050	0	0	5.3	-0.00022	1.03811
	0.0075	0	0	4.8	-0.00022	1.03811
	0.01	0	0	5.2	-0.00022	1.03811
S&P 500	0.0025	0.00155	0	13.4	0.00124	1.24485
(457,40)	0.0050	0	0	9.4	0.00204	1.25729
	0.0075	0	0	11.4	0.00204	1.25729
	0.01	0	0	10.3	0.00204	1.25729
Russell 2000	0.0025			infeasible		
(1318,90)	0.0050	0	0	29.3	0.00188	1.22387
	0.0075	0	0	30.2	0.00188	1.22387
	0.01	0	0	29.7	0.00188	1.22387
Russell 3000	0.0025	0	0	89.7	0.00343	1.10307
(2151,70)	0.0050	0	0	52.1	0.00343	1.10307
	0.0075	0	0	50.9	0.00343	1.10307
	0.01	0	0	65.5	0.00343	1.10307
Average		0.00018	0.0137	14.1	0.00088	1.04311

Table 4.11: Out-of-sample enhanced indexation average AER, second alternative

Index (N, K)	Value of τ									Correlation
	0.45	0.40	0.35	0.30	0.25	0.20	0.15	0.10	0.05	
Hang Seng (31,10)	-5.27	-5.24	0.94	0.67	-0.58	-0.21	-0.68	0.45	3.57	-0.77
DAX 100 (85,10)	6.72	5.16	4.86	5.00	5.56	7.22	5.19	0.67	-0.63	0.69
FTSE 100 (89,10)	6.10	-1.10	1.42	1.73	5.89	3.62	-3.04	2.05	3.16	0.14
S&P 100 (98,10)	6.31	6.56	-2.48	1.67	-5.68	-4.83	-4.40	-5.89	-2.94	0.78
Nikkei 225 (225,10)	0.69	-2.14	-5.05	-6.52	-8.14	-3.68	-5.20	-4.52	-4.31	0.44
S&P 500 (457,40)	6.93	1.72	5.07	5.46	5.05	7.61	7.08	6.99	7.00	-0.55
Russell 2000	4.42	4.57	5.56	7.91	8.66	9.49	10.13	11.17	12.15	-0.99
Russell 3000	12.56	12.38	14.61	14.80	15.72	21.25	27.35	27.68	30.97	-0.95

although we see some slight differences in terms of computation time and out-of-sample intercept/slope values.

Comparing Table 4.9 and Table 4.11 we can again see some slight differences (e.g. for $\tau = 0.45, 0.30, 0.25, 0.20$ and for the correlation) but otherwise the results are identical. Detailed tables (second alternative approach) of results for individual τ values (such as given in Table 4.5 and Table 4.6) are given in the Appendix C.

Hence, at least for the instances examined (and for $\lambda_1 = \lambda_2 = 1$), there appears little to choose between this second and first alternative. However we would note here that (taking index tracking and enhanced indexation together) both of these two alternatives are worse than the approach given before which leads to the results shown in Table 4.4 and Table 4.7.

4.4.5 Robustness

Our approach to deciding an index tracking, or an enhanced indexation, portfolio is based on optimisation. Clearly with any optimisation model there are issues related to robustness, especially:

- Discovering, and choosing between, multiple optimal solutions (if they exist).
- Would a very small change to an input parameter make a large change to the portfolio found as a result of optimisation?
- Would allowing a very small change in the optimised solution value (so allowing solutions whose values lie in a very small neighborhood around the optimal solution value) allow large changes in the portfolios found?

In our quantile regression approach this issue of robustness is especially relevant since some of the input parameters (the quantile regression intercept and slope parameters $\hat{\alpha}_{i\tau}$ and $\hat{\beta}_{i\tau}$) may not be uniquely defined (as mentioned above).

It would be beyond the scope of this chapter to directly address these issues. However we would make the following points with regard to robustness and our quantile regression approach:

- Ever since the pioneering work of Markowitz (1952) optimisation has been at the centre of work concerned with decisions relating to deciding the composition of financial portfolios. As such both practitioners and academic researchers have been willing to tradeoff the disadvantages of optimisation (multiple optimal solutions, solution sensitivity) for its advantages (clear modelling framework, computational efficiency, algorithmic decision-making). Whether practitioners and academic researchers in finance will be willing to forgive the disadvantages of quantile regression (non-uniquely defined quantile regression derived input parameters) when solving portfolio decision problems remains to be seen.

However we would note here that the citation statistics mentioned above for the seminal paper by Koenker and Bassett (1978) indicate that many other areas of science do not appear to regard non-uniquely defined quantile regression parameters as a bar to using the technique.

- The specific optimisation objectives we have adopted (Equation (4.20) and Equation (4.23)) have associated optimal solution values that may be small (e.g. see the values for D^* and E^* in Table 4.5 and Table 4.6). Since, especially for practitioners, a small variation in what is already a relatively small solution value may be of little consequence this does imply that there may be an acceptable neighborhood around the optimal solution value within which different portfolios could exist.

4.5 Conclusions

In this Chapter we considered two problems in financial portfolio construction, index tracking and enhanced indexation. We presented a mixed-integer linear programming formulation of these problems based on quantile regression. Computational results were presented for eight data sets drawn from major world markets which indicated that good quality out-of-sample results for tracking the indices considered could be achieved. With respect to enhanced indexation the computational results presented indicated that excess returns (returns in excess of index return) could be achieved out-of-sample and that the average out-of-sample return was competitive with that associated with previous work presented in the literature.

Chapter 5

Bootstrap Approach to Quantifying Uncertainty in Index Tracking and Enhanced Indexation

The focus of this chapter is on demonstrating the creation of portfolio uncertainty bands using a bootstrapped procedure. We restrict attention to the model presented in Chapter 4 and use re-sampling statistical techniques to build in-sample portfolio uncertainty bands. Further, we propose a number of ways in which the in-sample bootstrapped portfolios, which collectively form an uncertainty band, can be employed to improve out-of-sample portfolio performance for both index tracking and enhanced indexation.

5.1 Introduction

While a great deal of attention has been directed towards formulating models little effort has been invested in quantifying the level of uncertainty associated with the portfolio selected by these models. In index tracking and enhanced indexation the quantification of uncertainty is of importance as this provides investors with an indication of the degree of risk that can be expected as a result of holding the selected portfolio over the holding period.

The employment of past historical data to feed into the optimisation implicitly implies that the past is an accurate representation of the future. While this may be the case over a relatively near future, one would expect that as the holding time period increases the performance of the portfolio will deteriorate as, from a passive investment perspective, there is no mechanism to dynamically update the portfolio without incurring additional transaction costs.

In this chapter, we extend the work of Mezali and Beasley (2012) presented in Chapter 4 and demonstrate how to construct uncertainty bands for portfolios selected by this model. While the focus is on this model which encompasses both index tracking and enhanced indexation the methodology presented here can be extended to any model that is formulated with regression components, such as that of Canakgoz and Beasley (2009).

The objective of the work presented in this chapter is to illustrate how, under some assumptions, one can construct uncertainty bands for portfolios selected at a specific point in time and held for a given period. The uncertainty bands give an indication of the likely outcome that a portfolio may experience and thus provide a quantifiable measure of the monetary values that can be achieved as result of holding portfolios. In order to accomplish this goal we shall combine optimisation and statistical techniques, where for the latter we employ the bootstrapping re-sampling technique.

The remainder of the chapter is organised as follows. In Section 2 we give further insight into the bootstrapping technique. In Section 3 we introduce our formulation for the proposed method. Computational results are presented in Section 4 and finally in Section 5 we summarise and give concluding remarks.

5.2 Bootstrappping Procedure

Bootstrapping is a common statistical tool for generating an approximate sampling distribution of a statistic from one sample, in order to estimate a parameter. The idea was first introduced by the seminal work of Efron (1979) and has since become very popular due to its intuitiveness and the fact that no stringent conditions are attached to its application. More recently the popularity of this computationally intensive approach has increased due to technological advancements and availability of relatively cheap, powerful and efficient computers.

In this section we present the idea of the bootstrap approach from an application perspective by way of an example. Theoretical treatment of the subject can be found in Efron and Tibshirani (1993); Shao and Tu (1995) and Davison and Hinkley (1997).

Suppose that one is interested in the average height of some population of interest. However, due to some constraints, such as financial and time, only a fraction (n) of the entire population (m) is sampled with realisations $\mathbf{x} = \{x_1, x_2, ..., x_n\}$ and where $n < m$. From this realisation one can calculate a sample mean $\overline{x} = (\sum_{i=1}^{n} x_i)/n$. In order to make inference for the whole population, rather than just the realised sample, one would need to obtain a distribution of likely values for the unobserved population mean, μ, and given this information a probabilistic statement can be formulated on the likely range of μ with a given degree of confidence. For instance,

$$\Pr[\hat{\mu}_L < \mu < \hat{\mu}_U] = 0.95 \tag{5.1}$$

to indicate a 95% confidence that the true unknown mean μ would lie between $\hat{\mu}_L$ and $\hat{\mu}_U$, where the subscript L and U denote lower and upper bounds respectively and in the content of the example they represent 2.5 and 97.5 percentiles.

From the theory of statistics, the calculation of the lower and upper bounds depends on the assumption attached to the data. The most common of these assumptions are the data are identically distributed from a normal distribution with unknown mean μ and standard deviation σ, $\mathrm{N}(\mu, \sigma)$. If the sample size n is large ($n \geq 30$) then $\hat{\mu}_L$ and $\hat{\mu}_U$ forming a $100(1-\theta)\%$ confidence interval ($\theta \in (0,1)$) can be respectively calculated as $\overline{x} - z_{1-\theta/2}\hat{\sigma}/\sqrt{n}$ and $\overline{x} + z_{1-\theta/2}\hat{\sigma}/\sqrt{n}$ where z_θ denotes the inverse cumulative distribution of a Normal distribution at level θ and $\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$ is an estimate of the true but unknown σ^2. Where the sample is small ($n < 30$) $z_{1-\theta/2}$ is replaced by student-t distribution with $n-1$ degrees of freedom, $t_{1-\theta/2,n-1}$.

In order to obtain the limits $\hat{\mu}_L$ and $\hat{\mu}_U$ using the re-sampling technique, the idea of a non-parametric bootstrap involves sampling with replacement a large number of times from the original sample, $\mathbf{x}$. More specifically, one samples directly from $\mathbf{x}$ and obtains $\mathbf{x}^* = \{x_1^*, x_2^*, ..., x_n^*\}$ followed by a recalculation of $\overline{x}^*$. By repeating this process a large number of times, say 1000, one is able to map out the distribution of $\overline{x}$ and from this the lower and upper bounds for a given confidence level can be calculated. That is, once the bootstrapped samples $\{\overline{x}^{1*}, \overline{x}^{2*}, ..., \overline{x}^{1000*}\}$ are obtained they are sorted in ascending order and the lower and upper limits $\hat{\mu}_L$ and $\hat{\mu}_U$ are respectively given by the $[n(\theta/2)]$ and $[n(1-\theta/2)]$ data points, where $[a]$ denotes an integer part of a.

This idea forms the foundation of the methodology presented in the following discussion as it extends to constructing confidence intervals for unknown regression parameters as shall become apparent.

5.2.1 Portfolio selection using the bootstrapping technique

Liang et al. (1996) offered bootstrap simulation as a tool for quantifying the uncertainty in the composition of portfolios. They used this bootstrap simulation in an attempt to estimate the amount of real estate investors should hold to achieve optimum portfolio performance. The bootstrap method has shown itself to be useful in situations where the number of available data points is relatively small and the assumptions of parametric techniques do not hold. However, the confidence intervals produced were large.

Hatemi and Roca (2006) examined the simple case of international portfolio diversification involving the three largest stock markets in the world US, UK and Japan. Based on standard portfolio analysis, they first examined whether or not using diversification by US investors into the UK and Japanese markets would have been beneficial. They compared the risk-adjusted returns in the US market and that of the tangency portfolio consisting of the US, UK and Japanese markets. Then they undertook an analysis of the three markets based on a bootstrapping approach. They considered that the results of this study can be expanded to accommodate more markets and can also be done from the point of view of investors from the other markets. It can also be replicated on disaggregated scales.

Bartlmae (2009) introduced a framework for constructing portfolios, addressing two of the major problems of classical mean-variance optimization in practice: low diversification and sensitivity to information ambiguity. In order to address these issues, he used a bootstrapping method to incorporate the effects of input parameter variation. He investigated these methods by the use of Monte Carlo sampling. Firstly in order to overcome the problem of non-intuitive and undiversified portfolios, he introduced a method to construct portfolios that show a higher degree of diversification. He did this by introducing a diversification on the portfolio weights. In a second step, he applied bootstrapping to assess the input parameter ambiguity. By this method, more robust portfolios can be found. He incorporated these methods into a portfolio construction procedure.

Chen et al. (2012) applied the bootstrapping technique proposed by Kosowski et al. (2006) to examine whether the performance of enhanced-return index funds are based on luck or superior enhancing skills. They showed the advantages of using the bootstrap to rank fund performance. Their results show evidence of enhanced-return index funds with positive and significant alphas after controlling for luck and sampling variability.

Kopa (2012) considered robustness and bootstrap techniques in portfolio efficiency testing with respect to second-order stochastic dominance (SSD). He applied a computational method to test whether a US market portfolio, is (SSD) efficient with respect to 48 US industry representative portfolios. Moreover, he presented a robust version of a (SSD) portfolio efficiency test that allows for small errors in data and he analysed their impact on the market portfolio (SSD) efficiency. He improved his results, by applying the bootstrap technique to estimate the p-value of market portfolio (SSD) efficiency.

5.3 Bootstrap Uncertainty Bands

In this section we demonstrate how one can construct bootstrapped uncertainty bands for a portfolio selected through optimising a quantile regression-based model. More specifically, once we have optimised the mixed-integer program presented in chapter 4 and have obtained values for our decision variables G_i, x_i and z_i the question we wish to address is: what is the range of values we can expect the portfolio to realise over the in-sample period. In the first instance the aim is to construct uncertainty bands for the in-sample period and thereafter we present practical ways in which these bands will be employed in real life application.

In order to quantify a degree of uncertainty associated with a portfolio we exploit the fact that in formulating the objective functions for index tracking ($\tau = 0.50$) and enhanced indexation ($\tau = 0.45$) it is assumed that, for a given quantile of interest τ, the quantile regression intercept and slope for the portfolio can be approximated by the weighted sum of individual asset regressions. That is,

$$\hat{\alpha}_\tau = \sum_{i=1}^{N} w_i \hat{\alpha}_{i\tau} \tag{5.2}$$

and

$$\hat{\beta}_\tau = \sum_{i=1}^{N} w_i \hat{\beta}_{i\tau} \tag{5.3}$$

where

$$r_{it} = \alpha_{i\tau} + \beta_{i\tau} R_t \tag{5.4}$$

with $\hat{\alpha}_{i\tau}$ and $\hat{\beta}_{i\tau}$ denoting the respective quantile-specific estimates of intercept and slope obtained from quantile regressing the returns of asset i against the returns of the index R_t. Since, according to the theory of statistics, the individual intercepts $\hat{\alpha}_{i\tau}$ and slopes $\hat{\beta}_{i\tau}$ are estimates of true yet unknown parameters $\alpha_{i\tau}$ and $\beta_{i\tau}$ and thus, the theory implies, there is uncertainty associated with these coefficients. In order to quantify the uncertainty on an estimated parameter a probabilistic statement with an associated confidence level is attached to it, with the most common of these being the 95% confidence interval. For instance, to calculate 95% confidence interval for the slope parameter from equation (5.4) one would calculate the lower and upper bounds, $\beta_{i\tau}^L$ and $\beta_{i\tau}^U$ such that

$$\Pr[\beta_{i\tau}^L < \beta_{i\tau} < \beta_{i\tau}^U] = 0.95 \tag{5.5}$$

Equation (5.5) states that we are 95% confident that the true unknown parameter $\beta_{i\tau}$ will be enclosed within $\beta_{i\tau}^L$ and $\beta_{i\tau}^U$. In order to be able to calculate the lower and upper bounds one needs to approximate the distribution $\hat{\beta}_{i\tau}$. In the following discussion we demonstrate how these limits are calculated using bootstrap.

5.3.1 Bootstrapping quantile regression parameters

In chapter 4 it was mentioned that in quantile regressing asset returns against the returns of the index, as shown in equation (5.4), the slope and intercept parameters are obtained from solving a linear program by minimising the following objective function

$$\min \sum_{t=1}^{T} \rho_\tau(r_{it} - (\alpha_{i\tau} + \beta_{i\tau} R_t)), \tag{5.6}$$

where

$$\rho_\tau(u) = u(\tau - I(u < 0)) = \begin{cases} \tau u, & \text{if } u \geq 0 \\ (\tau - 1)u, & \text{if } u < 0 \end{cases}$$

In order to facilitate bootstrapping we make use of an equivalent definition represented by Yu and Moyeed (2001). That is, through defining the quantile regression model (5.4) as

$$r_{it} = \alpha_{i\tau} + \beta_{i\tau} R_t + \sigma u_t \tag{5.7}$$

where the error terms u_t follow an Asymmetric Laplace Distribution (ALD) with probability density function

$$f(u;\mu,\sigma,\tau) = \frac{\tau(1-\tau)}{\sigma} \begin{cases} \exp\left(\frac{(u-\mu)(1-\tau)}{\sigma}\right), & \text{if } u \leq \mu \\ \\ \exp\left(\frac{-(u-\mu)\tau}{\sigma}\right), & \text{if } u > \mu, \end{cases}$$

with $\mu \in (-\infty, \infty)$ and $\sigma > 0$ as the location and scale parameters respectively.

Defining the quantile regression in the format of equation (5.7) implies that the minimisation of (5.6) can equivalently be viewed as maximisation of the likelihood function (see (Yu and Zhang (2005))

$$\max \left(\frac{\tau(1-\tau)}{\sigma}\right)^T \exp\left\{-\sum_{t=1}^{T} \rho_\tau \left(\frac{r_{it} - (\alpha_{i\tau} + \beta_{i\tau} R_t)}{\sigma}\right)\right\}. \tag{5.8}$$

Yu and Moyeed (2001) noted that for a specific quantile of interest τ the ALD errors, u_t, can be represented by a combination of two exponential random variables (numbers) such that

$$u_t = \left(\frac{e_1}{\tau} - \frac{e_2}{1-\tau}\right) \tag{5.9}$$

where e_1 and e_2 are independent standard exponentially distributed random variables with mean 1, $\lambda(1)$. The scale parameter σ is estimated by

$$\hat{\sigma} = \frac{1}{T}\sum_{t=1}^{T}\rho_\tau(r_{it} - (\hat{\alpha}_{i\tau} + \hat{\beta}_{i\tau}R_t)). \tag{5.10}$$

This formulation of quantile regression presentation is very useful as it facilitates an easy to implement bootstrap procedure and, as a consequence, confidence intervals to reflect the uncertainty associated with parameters. The procedure to build uncertainty bands for quantile regression parameters $\alpha_{i\tau}$ and $\beta_{i\tau}$ is as follows:

1. Minimise equation (5.6) through linear programming and obtain estimated parameters $\hat{\alpha}_{i\tau}$ and $\hat{\beta}_{i\tau}$
2. Calculate the scale parameter $\hat{\sigma}$ from equation (5.10)
3. Generate T independent standard exponential random numbers $\{e_1\}_{t=1}^{T}$ and $\{e_2\}_{t=1}^{T}$ and calculate u_t from equation (5.9)
4. Define $r^*_{it} = r_{it} - \hat{\sigma}u_t$
5. Re-fit model $r^*_{it} = \alpha_{i\tau} + \beta_{i\tau}R_t$ and obtain new parameters $\hat{\alpha}^*_{i\tau}$ and $\hat{\beta}^*_{i\tau}$
6. Repeat steps 3 to 5 a large number of times M

After carrying out the procedures outlined above one would obtain M bootstrapped values for the intercept, $\{\hat{\alpha}^{*1}_{i\tau}, \hat{\alpha}^{*2}_{i\tau}, ..., \hat{\alpha}^{*M}_{i\tau}\}$, and slope, $\{\hat{\beta}^{*1}_{i\tau}, \hat{\beta}^{*2}_{i\tau}, ..., \hat{\beta}^{*M}_{i\tau}\}$. Let $[a]$ be an integer value of a and $\theta \in (0, 1)$ then to calculate the lower and upper bounds of the $100(1-\theta)\%$ confidence interval one will order the bootstrapped values in ascending order and from these ordered values the $[M\theta/2]$-th and $[M(1-\theta/2)]$-th data are the lower and upper limits forming the interval for a parameter of interest at a given quantile τ.

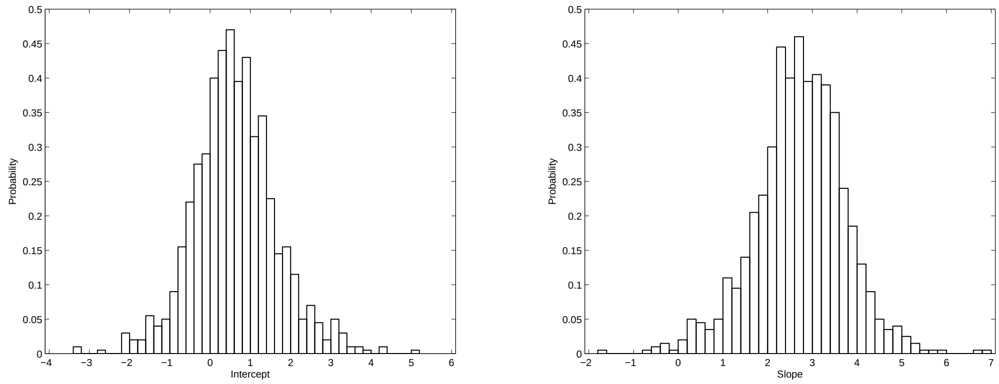

Figure 5.1: Bootstrapped confidence interval for the median ($\tau = 0.50$) intercept and slope parameters based on the data presented in Table 4.1 in chapter 4.

To illustrate we go back to an example presented in chapter 4 with fitted coefficients presented in Table 4.2. In applying the bootstrap procedure outlined previously, for the median ($\tau = 0.50$) and $M = 1000$, one obtains the distributions of the intercept and slope as showed in Figure 5.1. In this particular example the fitted intercept and slope are 0.625 and 2.75 respectively and as can be observed from Figure 5.1 these values are enveloped within their respective distributions. 95% confidence intervals for the intercept and slope parameters are [-1.433, 2.919] and [0.5045, 4.682] respectively. It takes only 3.4 seconds to implement 1000 bootstrap replications of quantile regression.

5.3.2 Bootstrapping portfolio value

Our next step is to link the bootstrapped quantile regression coefficients and the construction of uncertainty bands of a selected portfolio. It is noteworthy to emphasise that the creation of bootstrapped uncertainty bands is carried out post-optimisation. That is, for a given τ (eg, 0.45 or 0.50) the original model presented in chapter 4 is optimised and K assets to make up the portfolio are selected, $\{k_1, k_2, ..., k_K\} \subset \{1, 2, ..., N\}$ together with their associated quantity of units X_i^{opt} and weights w_i^{opt} for the bootstrap application, and thus the remainder of this chapter, we restrict our attention only on these and will henceforth use a k index rather than i. Further, since our attention in this chapter is on describing our proposed method (for convenience) we ignore uncertainty in the intercept parameters $\alpha_{k\tau}$ and focus attention purely on the slope $\beta_{k\tau}$.

For a given τ, $\tau = 0.50$ for indexation and $\tau = 0.45$ for enhanced indexation, let Z be a matrix whose columns are populated with M bootstrapped slope parameters for each of the K selected assets from a particular index. Explicitly,

$$Z = \begin{pmatrix} \hat{\beta}^{*1}_{k_1\tau} & \hat{\beta}^{*1}_{k_2\tau} & \cdots & \hat{\beta}^{*1}_{k_K\tau} \\ \hat{\beta}^{*2}_{k_1\tau} & \hat{\beta}^{*2}_{k_2\tau} & \cdots & \hat{\beta}^{*2}_{k_K\tau} \\ \vdots & \vdots & \cdots & \vdots \\ \hat{\beta}^{*M}_{k_1\tau} & \hat{\beta}^{*M}_{k_2\tau} & \cdots & \hat{\beta}^{*M}_{k_K\tau} \end{pmatrix}$$

where, for instance, the first column contains M bootstrapped slope coefficients for the first of the K selected assets. Given the matrix Z we proceed by optimising the following objective function

$$\text{minimise} \mid \hat{\beta}_\tau - \sum_{k=k_1}^{k_K} w_k^* \hat{\beta}^*_{k\tau} \mid \tag{5.11}$$

subject to

$$\sum_{k=k_1}^{k_K} w_k^* = 1 \tag{5.12}$$

$$\varepsilon_k \leq w_k^* \leq \delta_k \qquad k = k_1, \ldots, k_K \tag{5.13}$$

where $\hat{\beta}_\tau = \sum_{k=1}^{K} w_k^{opt} \hat{\beta}_{k\tau}$ with the weights, w_k^{opt}, are as obtained from the initial optimisation. Note that, $\hat{\beta}^*_{k\tau}$ correspond to each row of the matrix Z and thus the objective function given by Equation (5.11) is optimised M times with each optimisation producing different weights, $\{\boldsymbol{w}_m^* = \{w_{k_1}^*, w_{k_2}^*, ..., w_{k_K}^*\}\}_{m=1}^M$. The optimisation are conducted using a variant of Genetic Algorithm available in MATLAB in the form of `simulannealbnd` function.

Equipped with bootstrapped weights w_k^* one can obtain the associated number of units of assets from equation (4.19)

$$X_k^* = w_k^* C / V_{kT} \qquad k = k_1, \ldots, k_K \tag{5.14}$$

In summary, the step by step procedure taken to obtain the bootstrap uncertainty bands are as follows:

1. Optimisation of the objective function given by (5.11) to obtain w_k^*
2. Calculate number of units of each selected asset from 5.14
3. Calculate the value of bootstrapped portfolio at each time step $t = 1, 2, ..., T$ by $\sum_{t=1}^{T} X_k^* V_{kt}$ where V_{kt} is the price of asset k at time t
4. Repeat steps 1 to 3 $M = 1000$ times

5.4 Computational results and discussion

To provide a graphical illustration of the quality of results, figures 5.2 and 5.3 respectively show the evolution of the in-sample value of portfolios (shown in red) for the median and quantile ($\tau = 0.45$) regression-based models together with bootstrapped uncertainty bands (blue region). Note that the value of the original portfolio is calculated by $\sum_{t=1}^{T=145} X_k^{opt} V_{kt}$ and value of each M bootstrapped portfolios are given by $\sum_{t=1}^{T=145} X_k^* V_{kt}$. Figures 5.4 and 5.5 show out-of-sample value of portfolios for all 8 test problems described in chapter 4

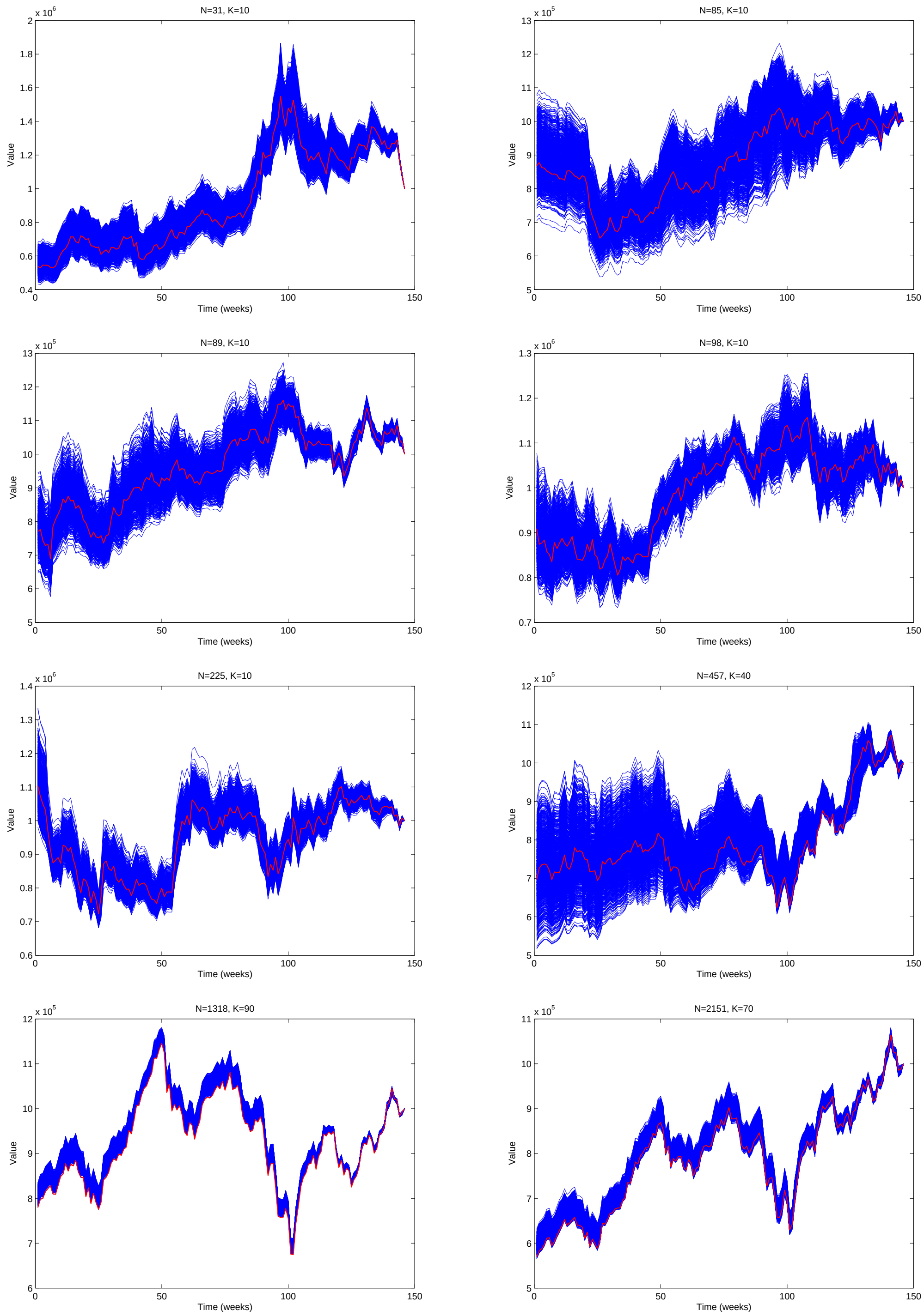

Figure 5.2: Quantile regression model (Chapter 4) in-sample uncertainty bands for $\tau = 0.5$. Test problems 1, 3, 5, 7 and Test problems 2, 4, 6, 8 are on the left and right panels respectively.

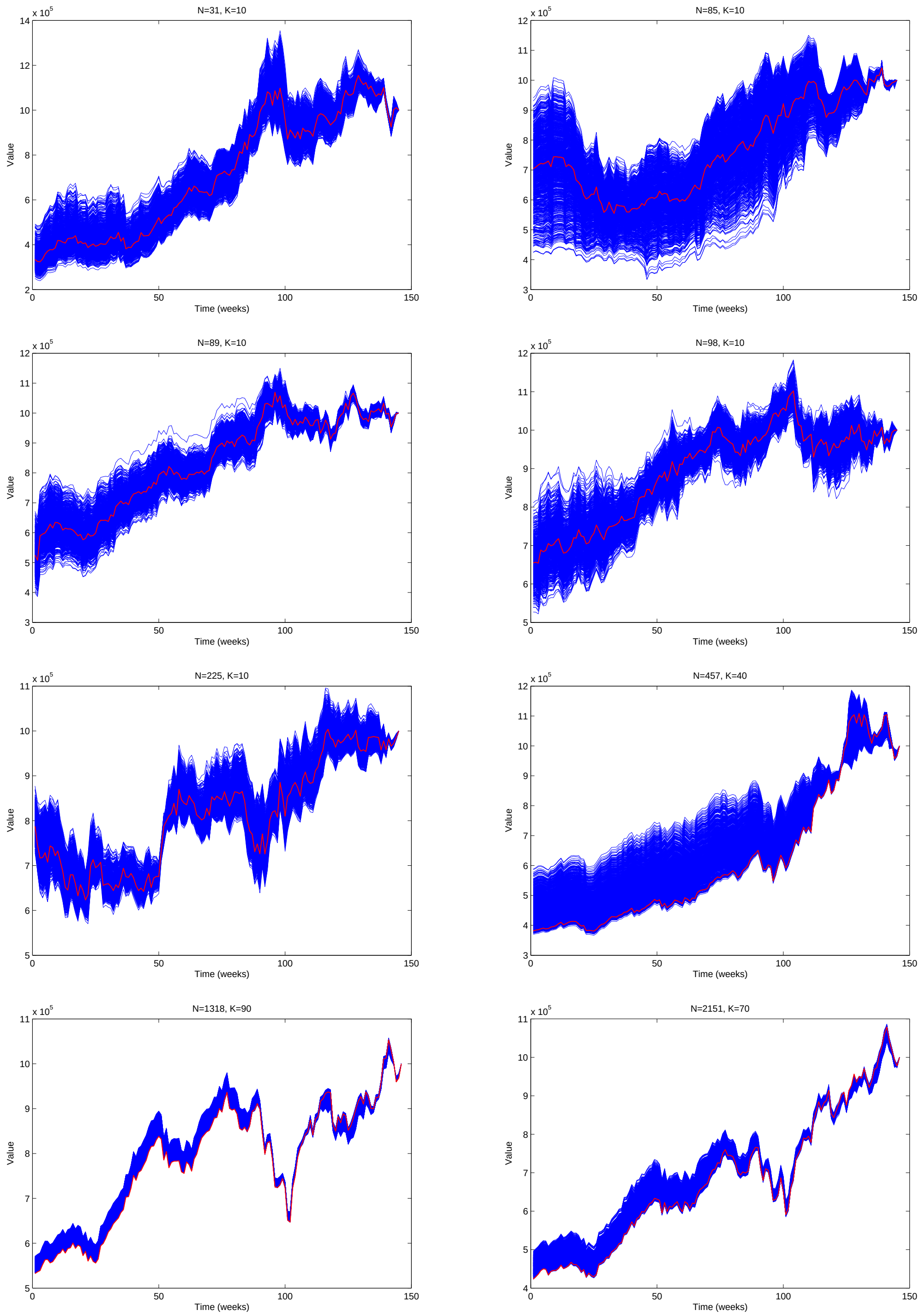

Figure 5.3: Quantile regression model (Chapter 4) in-sample uncertainty bands for $\tau = 0.45$. Test problems 1, 3, 5, 7 and Test problems 2, 4, 6, 8 are on the left and right panels respectively.

table 4.3. The values of the portfolios are calculated as was done for the in-sample results with time t ranging from $t = 146$ to $T = 290$ with associated prices within the out-of-sample period.

From a statistical point of view, an important property of uncertainty bands is to be able to fully encapsulate the variable of interest, in this case the value of portfolios. From Figures 5.2 and 5.3 it can be seen that the portfolios, for both quantiles, clearly envelope the original portfolios shown in red. Further, if the model is optimised at time T one does not have visibility of future prices (from $t > T$) however Figures 5.4 and 5.5 are included for illustration. As it can be observed from 5.4 and 5.5 for both indexation $\tau = 0.5$ and enhanced indexation $\tau = 0.45$ in applying the bootstrapped weights w_k^*, or equivalently, number of unit of assets x_k^*, obtained from the in-sample period one is in a sense also capturing the evolution of the value of the out-of-sample portfolio for all test problems.

For both indexation and enhanced indexation models it appears that for those test problems such as Russell 2000 and Russell 3000, in which the number of assets selected (K) is relatively large the value of the original portfolio is shown to be on the boundary of the uncertainty bands. This could be a result of the fact for K relatively large the weights of the original portfolio w_k^{opt} are approaching the proportion lower limit of $\varepsilon \geq 0.01$ and thus in optimising to calculate the boostrapped weights the optimiser is restricted within a narrower search space which becomes visible in the plots in the form of narrower uncertainty bands.

In what follows we present a number of ways in which the uncertainty bands constructed for the in-sample period can be interpreted. The objective and emphasis here is to demonstrate how these uncertainty bands can be implemented in real applications.

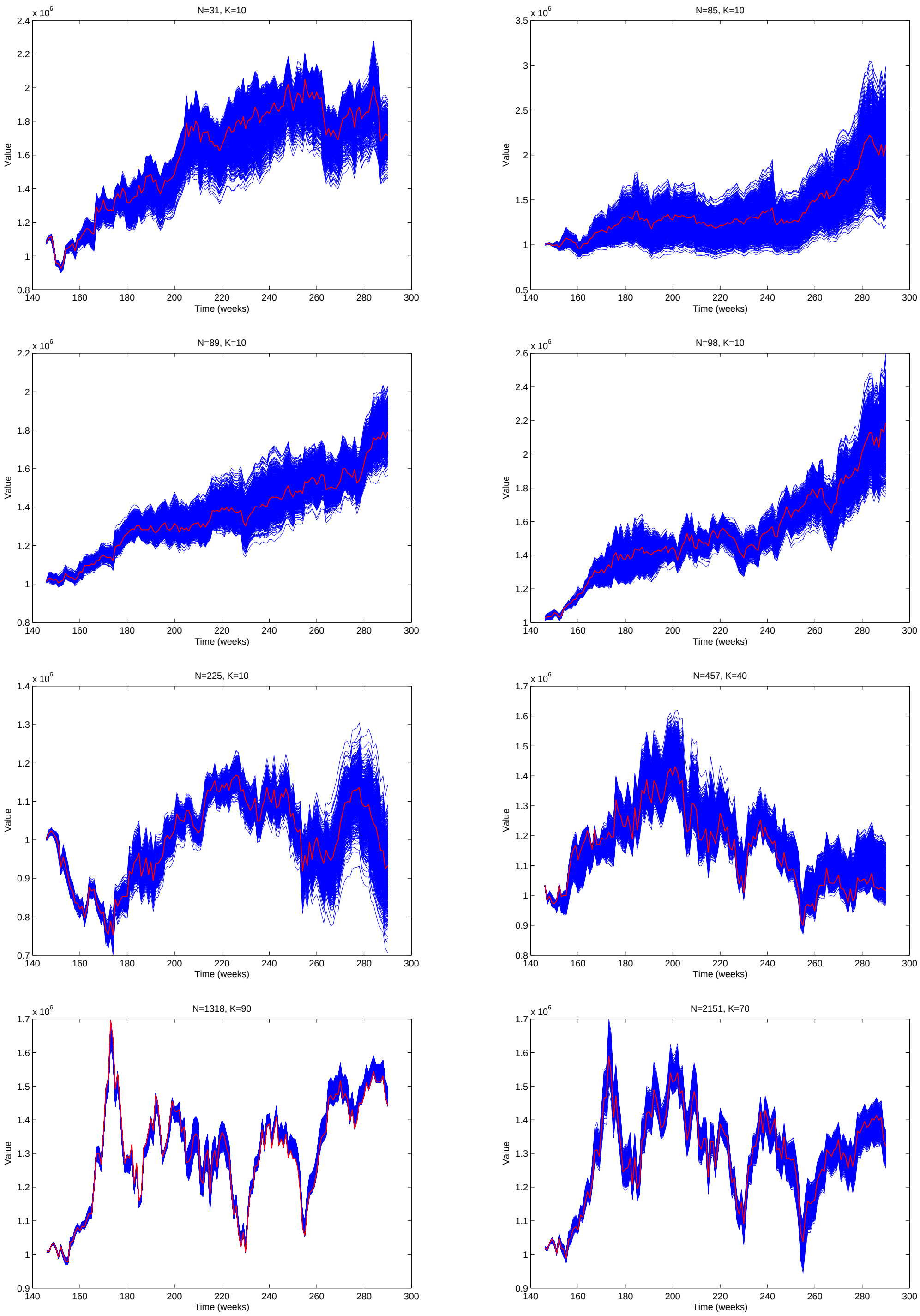

Figure 5.4: Quantile regression model (Chapter 4) out-of-sample uncertainty bands for $\tau = 0.5$. Test problems 1, 3, 5, 7 and Test problems 2, 4, 6, 8 are on the left and right panels respectively.

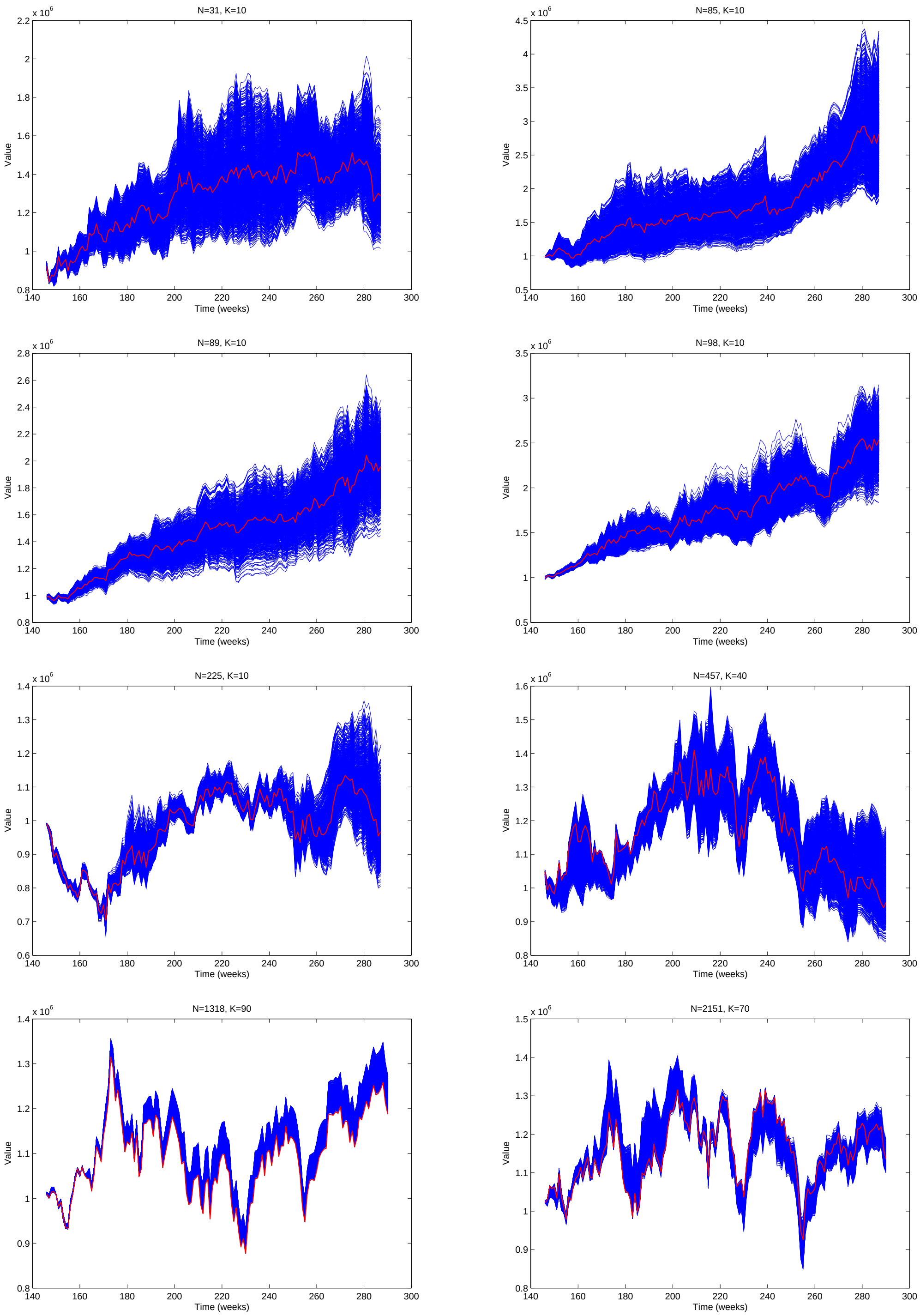

Figure 5.5: Quantile regression model (Chapter 4) out-of-sample uncertainty bands for $\tau = 0.45$. Test problems 1, 3, 5, 7 and Test problems 2, 4, 6, 8 are on the left and right panels respectively.

5.4.1 Improving portfolio returns for enhanced indexation

Once the in-sample uncertainty bands are constructed for a specific portfolio the M individual bootstrapped portfolios, which collectively form the uncertainty bands, can be used in conjunction with a user-defined measure of performance to provide an indication of the distribution or variability of the chosen measure of performance. The resultant distribution calculated over the entire in-sample period can be used to quantify in-sample confidence intervals in respect of the chosen measure of performance.

To be consistent with chapter 4 here we use the AER to assess the performance of selected portfolio:

$$\text{AER}^{in} = \frac{5200}{145} \sum_{t=1}^{145} \left[log_e \left(\frac{\sum_{k=1}^{N} X_k^{opt} V_{kt}}{\sum_{k=1}^{K} X_k^{opt} V_{kt-1}} \right) - R_t \right] \tag{5.15}$$

where, as before, X_k^{opt} denote the optimal number of units of asset k to be held in the portfolio with k identifying only those assets that have been selected from the original optimisation. To differentiate from the bootstrapped portfolios, when we refer to the original portfolio we mean the one from which the original number of units X^{opt} was obtained. In the following discussion whenever an asterisk is used it implies that a bootstrapped portfolio is referred to, otherwise it is the original portfolio.

Recall that each replication of the bootstrap procedure produces different weights w_k^* and corresponding X_k^*. By substituting X_k^* in equation (5.15) in place of X_k^{opt} one obtains M values of the in-sample AER, $\{AER_j^{in*}\}_{j=1}^{M}$, with each j corresponding to a bootstrap replication.

While the magnitude of the AER provides a measure of performance of a portfolio over a given time period in comparing alternative portfolios it is also instructive to incorporate the frequency with which a given portfolio produces positive or negative returns. To this end, for each bootstrapped portfolio observed through time, $X_k^{j*}V_{kt}$, we calculate

$$\mathfrak{J}_j^* = \sum_{t=1}^{145} \mathfrak{L}\left[\log_e\left(\frac{\sum_{k=1}^{K} X_k^{j*}V_{kt}}{\sum_{k=1}^{K} X_k^{j*}V_{kt-1}}\right) < 0\right] \qquad j = 1, \ldots, M \tag{5.16}$$

where $\mathfrak{L}(.)$ is an indication function taking a value of 1 if $\mathfrak{L}(.)$ is true and zero otherwise. Equation (5.16) calculates the number of negative returns over the entire in-sample period for each of the M bootstrapped portfolios. By combining the AER together with the frequency with which portfolio values decrease (5.16) we define

$$\text{AER}_j^{fre*} = \frac{\text{AER}_j^{in*}}{\mathfrak{J}_j^*} \qquad j = 1, \ldots, M \tag{5.17}$$

The equation (5.17) centres on comparing alternative bootstrapped portfolios, with different weights and number of units of assets, in addition to the original portfolio. For instance, suppose on analysing the in-sample performance of bootstrapped portfolios one finds that there are two portfolios with the same AER^{in*}. Then the better of these two portfolios to be held into the future would be the one with the highest AER_j^{fre*} value and thus greater number of positive returns.

Figures 5.6 shows the distribution of AER_j^{fre*} for $\tau = 0.45$. Note that, this distribution is constructed based on all possible values of AER_j^{fre*} with the x-axis showing the ranges for AER^{fre*} values .

An implicit assumption in forming the original portfolio centres on the idea that the weight w_k^{opt} and corresponding number of units X_k^{opt} selected from the original optimisation is the optimal selection given the constraints imposed. If this assumption is tenable, then one would expect that the performance of this portfolio would always be superior then any other portfolio as the objective functions of the original portfolio was explicitly designed to out-perform and

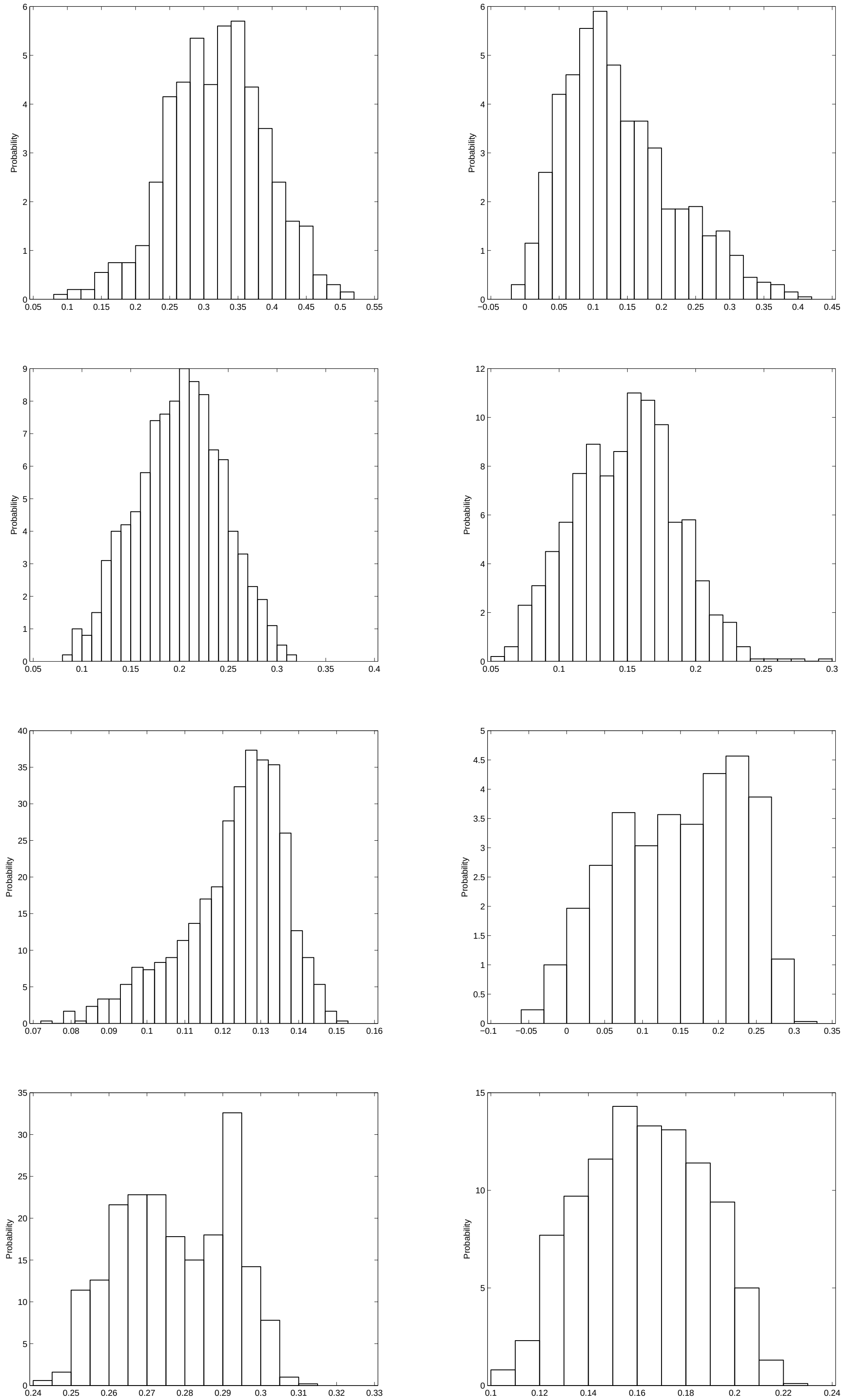

Figure 5.6: Distribution of bootstrapped in-sample AER^{fre*} for enhanced indexation $\tau = 0.45$. Test problems 1, 3, 5, 7 and Test problems 2, 4, 6, 8 are on the left and right panels respectively.

track the index in the case of $\tau = 0.45$ and $\tau = 0.5$ respectively.

The availability of bootstrapped portfolios provide a platform on which this assumption of optimality of the original portfolio can be tested. Since the AER^{fre} of the original portfolio is captured within the distribution of the bootstrapped AER_j^{fre*} this implies that there is at least one bootstrapped portfolio, with identifiable number of units X_k^*, that outperforms the optimal portfolio, at least in the in-sample period. Further, by extrapolation, one would also expect that the bootstrapped portfolio outperforming the original portfolio in the in-sample period will also out-perform it in the out-of-sample period. Since investors are interested in obtaining the highest returns we define the maximum of AER_j^{fre*}

$$\text{AER}^{max*} = \max\left(\frac{\text{AER}_j^{in*}}{\mathfrak{I}_j^*}\right) \qquad j = 1, \ldots, M \tag{5.18}$$

with X_k^{max*} as the number of units of asset k in a portfolio that is expected to produce the greatest return in the out-of-sample period. To test this assumption we evaluate the AER

$$\text{AER}^{out*} = \frac{5200}{h} \sum_{t=146}^{146+h} \left[log_e\left(\frac{\sum_{k=1}^{K} X_k^{max*} V_{kt}}{\sum_{k=1}^{K} X_k^{max*} V_{kt-1}}\right) - R_t \right] \tag{5.19}$$

for different duration $h = 52, 104$ and 145 weeks and compare the outcome with that of the original portfolio with number of units X_k^{opt}.

Table 5.1 displays out-of-sample AER results based on equation (5.19) together with the original portfolio (X_k^{opt}) for enhanced indexation ($\tau = 0.45$). Table 5.1 shows that for three cases of holding time of 52, 104 and 145 weeks the bootstrapped selected portfolio outperforms the original portfolio in most of the eight test problems. Note that, in those cases where both original and bootstrapped portfolios AER are negative the lesser of the two losses is

Table 5.1: Out-of-sample enhanced indexation results, $\tau = 0.45$

Index	AER (h=52 weeks)		AER (h=104 weeks)		AER (h=145 weeks)	
(N, K)	Quantile regression	Bootstrapping	Quantile regression	Bootstrapping	Quantile regression	Bootstrapping
Hang Seng (31,10)	-2.135	0.814	-6.675	-6.586	-9.79	-12.38
DAX 100 (85,10)	37.48	62.17	14.51	23.48	16.13	25.28
FTSE 100 (89,10)	13.29	24.75	9.44	17.16	7.13	12.92
S&P 100 (98,10)	10.19	17.93	10.03	18.49	6.71	13.69
Nikkei 225 (225,10)	-1.19	-1.85	1.59	2.39	1.73	1.92
S&P 500 (457,40)	11.69	11.27	14.37	13.87	6.38	5.48
Russell 2000 (1318,90)	-3.38	-2.47	1.52	1.87	1.99	2.36
Russell 3000 (2151,70)	6.26	6.91	12.72	12.80	10.89	11.45
Average	9.02	14.94	7.19	10.43	5.15	7.59
Average difference	5.92		3.24		2.44	

considered to be superior. Table 5.1 also displays the average AER across all eight indices. The average difference ($\mathrm{AD} = 0.125\sum_{\ell=1}^{8}\mathrm{AER}_{\ell}^{out*} - \mathrm{AER}_{\ell}^{out}$) for all holding time period h is positive and decreasing with h with values ranging from approximately 6% for $h = 52$, 3% for h=104 and above 2% for $h = 145$ weeks.

5.4.2 Improving tracking performance

Building on the previous subsection in which an approach is presented for improving out-of-sample returns based on in-sample bootstrapped portfolios here we extend this idea for tracking portfolios. The objective is to use the in-sample bootstrapped portfolios for $\tau = 0.5$ to improve out-of-sample tracking performance through examining regression parameters when the in-sample bootstrapped portfolio returns are regressed against that of in-sample index returns.

A selected portfolio will perfectly track the index if on quantile regressing $\tau = 0.50$ out-of-sample index returns against the selected portfolio one obtains an intercept of zero and a slope of unity. That is, performing the following regression

$$log_e\left(\frac{X_k^{opt}V_{kt}}{X_k^{opt}V_{kt-1}}\right) = \alpha_{\tau=0.5}^{out} + \beta_{\tau=0.5}^{out}R_t, \qquad t = 146, \ldots, T = 290 \qquad (5.20)$$

should ideally produce $\hat{\alpha}_{\tau=0.5}^{out} = 0$ and $\hat{\beta}_{\tau=0.5}^{out} = 1$. However, since at time T when the original model is optimised an analyst has no visibility of future prices and thus, as before, we employ the additional information we have obtained from the in-sample bootstrapped portfolios to compare and select a bootstrapped portfolio that is expected to be superior than the original portfolio. The criterion of selection is now $\hat{\alpha}_{\tau=0.5}^{in} = 0$ and $\hat{\beta}_{\tau=0.5}^{in} = 1$ and thus for each of the 1000 bootstrapped replications we calculate the intercept from quantile regressing ($\tau = 0.50$) the bootstrapped portfolio returns against the index

$$log_e\left(\frac{X_k^{*j}V_{kt}}{X_k^{*}V_{kt-1}}\right) = \alpha_{\tau=0.5}^{in*j} + \beta_{\tau=0.5}^{in*j}R_t, \qquad t = 1, \ldots, T = 145 \qquad j = 1, ..., M \tag{5.21}$$

Once the corresponding parameters $\{\hat{\alpha}_{\tau=0.5}^{in*j}\}_{j=1}^{M}$ and $\{\hat{\beta}_{\tau=0.5}^{in*j}\}_{j=1}^{M}$ are obtained we calculate the errors associated with these over the entire in-sample period

$$\xi^{*j} = \sum_{t=1}^{T}\left|log_e\left(\frac{X_k^{*j}V_{kt}}{X_k^{*}V_{kt-1}}\right) - \hat{\alpha}_{\tau=0.5}^{in*j} - \hat{\beta}_{\tau=0.5}^{in*j}R_t\right|, \quad j = 1, ..., M \tag{5.22}$$

Finally, the bootstrapped portfolio with the lowest error $\min\{\xi^{*j}\}_{j=1}^{M}$ (with associated number of units X^{min*}) is then compared with the original portfolio. The intuition is that, since the bootstrapped portfolio is associated with the lowest error it closely tracks the in-sample index and thus as a by-product it is also expected that when this portfolio is held into the future it will also closely mimic the index and thus produce better fits compared to the original portfolio. To examine this assumption we quantile regress ($\tau = 0.50$) the out-of-sample returns of the selected bootstrapped portfolio against those of the index (as shown in 5.20) by replacing X_k^{opt} with X^{min*}. Further, we also compare the intercept and slope coefficients of the selected bootstrapped portfolio with those obtained from the original portfolio.

Table 5.2 displays the coefficients for the slope and intercept for the quantile regression and the bootstrapping models, for three different holding time periods using transaction cost limit (γ)=0.01. As we are considering index tracking here we are using $\tau = 0.50$. It it can be observed that for the 52 weeks holding period, the out-of-sample average quantile regression intercept and slope are [0.0021 and 1.1292]; and the average bootstrapping regression intercept and slope are [0.0018 and 1.1072].

Table 5.2: Out-of-sample index tracking results for holding time period of 52, 104, and 145 weeks, $\tau = 0.50$

Index	QR 52 weeks		Bootstrap 52 weeks		QR 104 weeks		Bootstrap 104 weeks		QR 145 weeks		Bootstrap 145 weeks	
(N, K)	Intercept	Slope	Intercept	Slope	Intercept	Slope	Intercept	Slope	Intercept	Slope	Intercept	Slope
Hang Seng (31,10)	0.0017	1.0956	0.0016	1.0894	0.0006	1.0642	0.0001	1.0372	-0.0028	0.9881	-0.0004	1.0195
DAX 100 (85,10)	0.0037	1.3274	0.0028	1.2156	-0.0002	1.0020	0.0003	1.0097	0.0004	1.1542	0.0001	1.07790
FTSE 100 (89,10)	0.0009	0.9582	0.0007	0.9977	0.0009	0.9456	0.0001	0.9490	0.0008	0.8186	0.0004	0.8482
S&P 100 (98,10)	-0.0003	1.0732	-0.0001	1.0527	-0.0005	1.0320	-0.0003	1.0193	0.0007	0.8623	0.0005	0.9644
Nikkei 225 (225,10)	-0.0001	1.0284	-0.0002	0.9996	0.0000	1.0355	0.0000	1.0261	-0.0002	1.0381	-0.0002	1.0252
S&P 500 (457,40)	0.0039	1.2203	0.0037	1.2217	0.0024	1.2123	0.0025	1.2054	0.002	1.2572	0.0002	1.2131
Russell 2000 (1318,90)	0.0023	1.2474	0.0021	1.1837	0.0012	1.3123	0.0015	1.2344	0.0018	1.2238	0.0007	1.0875
Russell 3000 (2151,70)	0.0045	1.0829	0.0042	1.0969	0.0037	1.1768	0.0037	1.1825	0.0032	1.1078	0.0003	1.0902
Average	0.0021	1.1292	0.0018	1.1072	0.001	1.0976	0.001	1.0830	0.0007	1.0563	0.0002	1.0293

Out-of-sample (over the time periods [146,250] and [146,290] of weekly observations, the average bootstrapping model intercept is close to zero with the average bootstrapping model slope differing from one by only 0.08 and 0.02 respectively.

With regard to the absolute difference between the out-of-sample intercept and zero, and the absolute difference between the out-of-sample slope and one, then for Table 5.2 these values are 0.0013 and 0.0943 for quantile regression and 0.0010 and 0.0731 for the bootstrapping model respectively (averaged over all cases in Table 5.2). Clearly it is a matter of judgment, that in most cases the bootstrapping model increases the robustness of results in terms of intercept close to zero and slope close to one.

5.4.3 Projecting prediction intervals

An implicit assumption employed in constructing an index tracking model lies in the usage of historical data as a representative benchmark of future market fluctuations such that a portfolio selected at a given point in time using historical data would, in the optimised model under consideration, be the optimal portfolio to be held into the future, at least in the not very distant future.

Using this implicit assumption we propose that the bootstrapped uncertainty bands constructed from a given model should also have the characteristic of identifying a spectrum of possible realisations of the selected portfolio extrapolated into the future. It is argued that this assumption of extrapolating the interpretation of the confidence intervals into the future is robust to the length of the holding horizon from the selection time point with a shorter period such as 13 weeks producing more representative results as the behaviour of the market more closely reflects the period for which the model was optimised. By extrapolative interpretation it is meant that the uncertainty bands using the in-sample prices are used to characterise future uncertainty. In terms of application this will involve an analyst choosing a cut-off time of the most recent period for which the uncertainty bands will be used for extrapolation. For instance, if the in-sample uncertainty bands of a portfolio is built using 145 weeks then an analyst can choose, say, the recent 13 weeks (weeks 132 to 145) to provide an indication of future uncertainty.

From a statistical point of view the assumption of extrapolative interpretation of the portfolio uncertainty implies the distribution of the portfolio returns does not change through the holding period. While in practical applications, when one optimises a model in order to select a portfolio at a specific point in time, one does not have visibility of future prices, however, a proposition put forward here is that if the holding horizon is not very distant from the selection time the distributions of in-sample and out-of-sample returns should not be significantly different and thus make this interpretation tenable. In order to test such an assumption on our data we apply the Kolmogorov-Smirnov test for equality of in-sample and out-of-sample portfolio returns distributions for different holding periods. To this end, we proceed by presenting the mechanism of the Kolmogorov-Smirnov.

Let $\boldsymbol{r}_{in} = \{r_1, r_2, \ldots, r_T\}$ and $\boldsymbol{r}_{out} = \{r_{T+1}, r_{T+2}, \ldots, r_{T+h}\}$ respectively denote the in-sample and out-of-sample returns obtained from a portfolio selected by a specific model, where h is the holding horizon. Further, let F_{in} and F_{out} denote the cumulative distributions of the in-sample and out-of-sample returns respectively. The two sample Kolmogorov-Smirnov tests for equality of the two return distributions with the hypothesis

$$H_0 : F_{in} = F_{out} \qquad\qquad H_1 : F_{in} \neq F_{out} \tag{5.23}$$

uses the test statistics $KS = \max \mid F_{in} - F_{out} \mid$. For a given significance level θ, where we will use 0.05, the test will detect if there is statistically significant difference between the two distributions. For our implementation we employ the *kstest2* function in *MATLAB* and test four different horizons, $h = 13$, $h = 26$, $h = 52$ and $h = 145$ weeks.

Table 5.3: Kolmogorov Smirnov test results of equality of in-sample and out-of-sample distributions for QR model $\tau = 0.50$ and $\tau = 0.45$.

Index	Holding time period (QR $\tau = 0.50$)				Holding time period (QR $\tau = 0.45$)			
	13	26	52	145	13	26	52	145
Hang Seng	0.934	0.788	0.659	0.159	0.861	0.996	0.842	0.004*
DAX 100	0.850	0.684	0.255	0.021*	0.702	0.111	0.060	0.012*
FTSE 100	0.839	0.169	0.108	0.049*	0.928	0.613	0.278	0.175
S&P 100	0.151	0.019*	0.210	0.294	0.268	0.053	0.254	0.045*
Nikkei 225	0.380	0.241	0.790	0.428	0.004*	0.018*	0.818	0.996
S&P 500	0.817	0.940	0.550	0.074	0.237	0.328	0.397	0.004*
Russell 2000	0.868	0.444	0.092	0.059	0.408	0.064	0.012*	0.014*
Russell 3000	0.943	0.100	0.025*	0.043*	0.299	0.281	0.069	0.013*

Table 5.3 shows the statistic p values obtained from testing for the equality of tracking portfolios in-sample and out-of-sample return distributions using a two sample Kolmogorov-Sminorv test for different horizons. All tests are conducted with a 5% significance level and thus a p value less than $p < 0.05$ (marked with * in the table) indicates that the test of equality of distribution is rejected and if $p \geq 0.05$ there is no evidence to suggest that the distributions of the in-sample and out-of-sample returns are statistically different. From the table one can observe, in the case of quantile regression with $\tau = 0.50$ model, that when all the out-of-sample period are compared with the in-sample period the test rejects the hypothesis of equality for index DAX 100, FTSE 100 and

Russell 3000. However, when the out-of-sample is reduced to 13, 26 and 52 weeks, as is customary for real life applications, all but the S&P 100 and the Russell 3000 distributions are classified to be statistically the same. Some minor differences are apparent for results obtained from the quantile regression with $\tau = 0.45$ model where for shorter holding periods the distribution of the in-sample and out-of-sample are not statistically different. These results indicate that over short holding horizons the assumption of equality of in-sample and out-of-sample portfolio return distributions is plausible and thus confirming the validity of the methodology.

5.5 Conclusion

In this chapter we concentrated on increasing the robustness of results given in association with chapter 4. We introduced a new approach that demonstrated how one can construct bootstrapped uncertainty bands for a portfolio selected through optimising a quantile regression-based model. We focused in quantifying the level of the uncertainty associated with portfolio selection in index tracking and enhanced indexation. We first showed how to capture the uncertainty visually for the out-of-sample value of the portfolio for the 8 test problems described in chapter 4 for index tracking and enhanced indexation.

Moreover, we presented a number of numerical ways in which the uncertainty bands can be implemented in real life applications. We first presented how to improve the portfolio returns for enhanced indexation by providing a measure of performance of a portfolio over a given time period. This approach for improving out-of-sample returns was based on in-sample bootstrapped portfolios. Secondly, we extended this idea for tracking portfolios and the objective was to use the in-sample bootstrapped portfolios to improve out-of-sample tracking performance through examining regression parameters (i.e. slope and intercept). Finally, we used the assumption of extrapolating the interpretation of the confidence intervals into the future. We used the Kolmogorov Smirnov test to extrapolate interpretation of the uncertainty bands using the in-sample prices that are used to characterise future uncertainty. We showed that from a statistical point of view the assumption of extrapolative interpretation of portfolio uncertainty implies the distribution of the portfolio returns does not change through the holding time period.

Chapter 6

Concluding Remarks

The objective of this thesis was to contribute to the development of efficient and effective portfolio selection algorithms. We presented methods for solving problems in financial portfolio construction, index tracking and enhanced indexation. Our formulations were mixed-integer linear programs for index tracking and enhanced indexation. In contrast to the majority of previous work, our formulations for both index tracking and enhanced indexation presented in this thesis include transaction costs, constrain the number of stocks that can be held, and also constrain the total transaction cost that can be incurred.

6.1 Main Contributions

In chapter 3 we presented two mixed-integer linear programming formulations for index tracking. In particular we explicitly considered both fixed and variable transaction costs and limited the total transaction cost that could be incurred. We proposed two approaches for the objective function associated with choice of a tracking portfolio, namely; minimise the maximum absolute difference between the tracking portfolio return and index return and minimise the average of the absolute differences between tracking portfolio return and index return. Our formulations are based upon tracking an index by comparing the returns from the index with the returns from the tracking portfolio. The main results indicated that good quality out-of-sample results for tracking the indices considered could be achieved. The computational times for all the data sets considered were low. The work presented in this chapter has been published in the Springer Optimization Letters journal (see Mezali and Beasley (2011)).

In chapter 4 we applied Quantile Regression to two problems in financial portfolio construction, index tracking and enhanced indexation. The contribution of this Chapter lies in the application of quantile regression to the problem of constructing financial portfolios for index tracking and enhanced indexation. According to our knowledge this is the first time that quantile regression has been applied to these problems. Moreover by using the quantile regression concept, we managed to capture within the same model/approach, both index tracking and enhanced indexation objectives.

Computational results were presented for eight data sets drawn from major world markets which indicated that good quality out-of-sample results for tracking the indices considered could be achieved. With respect to enhanced indexation the computational results presented indicated that excess returns (returns in excess of index return) could be achieved out-of-sample and that the average out-of-sample return was competitive with that associated with previous work presented in the literature. The work presented in this chapter has been published in the Journal of the Operational Research Society (JORS) (see Mezali and Beasley (2012)).

In chapter 5 we focused on quantifying the level of uncertainty associated with portfolio selection. In index tracking and enhanced indexation the quantification of uncertainty is of importance as this provides investors with an indication of the degree of risk that can be expected as a result of holding the selected portfolio over the holding period. We presented a bootstrap approach to quantify the uncertainty of portfolio selected from regression models. We proposed a number of ways in which the in-sample bootstrapped portfolios, which collectively form an uncertainty band, can be employed to improve out-of-sample portfolio performance for both index tracking and enhanced indexation.

6.2 Recommendations for Future Research

In this thesis we have presented and evaluated new methods for index tracking and enhanced indexation. However, there are a number of extensions that could be explored.

6.2.1 Using the concept of rebalancing over time

Given the evolution of prices of the stocks comprising a particular index the goal of the models presented above is to select a number of stocks and their appropriate quantities, which, when held over a period of time in the future, will closely track the returns on the index (or exceed index return). However, regardless of the immediate accuracy of the model in selecting the portfolio of stocks tracking accuracy deteriorates over time. That is, when the same portfolio is held for a very long time the difference between the returns of the portfolio and that of index widen over time. In order to maintain the accuracy of the model in tracking the index the chosen portfolio needs to be rebalanced after some appropriately chosen time such that a balance is achieved between the transaction costs and accuracy of the tracking model.

Using the in-sample information from time $t = 0$ to T the inputs are fed into the model and we obtain a portfolio ($[x_i]$ $[i = 1, \ldots, N]$ in terms of the number of units of stock i included in the portfolio) that will be held out of sample into the future for a period of 13 weeks for example . At time $T+13$ the value of the portfolio is $C = \sum_{i=1}^{N} V_{i(T+13)} x_i$ which is then rebalanced so that our existing portfolio changes from X_i to x_i. The rebalancing process is performed sequentially in a moving window fashion such that at the first rebalance time, $T + 13$, the in-sample period is $t = 13$ to $T + 13$ and for the k-th rebalance ($k \in \mathbb{Z}$) the in-sample period is from $t = 13k$ to $T + 13k$. It is a rolling forward approach to validating the proposed models or other exiting models through out of sample testings.

6.2.2 Forecasting stocks prices

The employment of past historical data to feed into the optimisation implicitly implies that the past is an accurate representation of the future. While this may be the case over a relatively near future, one would expect that as the holding time period increases the performance of the portfolio will deteriorate as, from a passive investment perspective, there is no mechanism to dynamically update the portfolio without incurring additional transaction costs.

We suggest that we forecast prices of stocks over a holding period and using the forecasted prices with the in-sample data in fitting the existing models. In this respect, this approach should incorporate more information about the data and hopefully we can increase stability of the tracking performance out-of-sample. This would involve using a time series model such as Autoregresive moving average (arima) to forecast the prices of each stock making up a particular index and incorporating these forecasts as if they were actual observations. Testing this approach will be based on graphical judgment and comparison with the results that have been already achieved in this thesis.

Appendix A

Remaining detailed tables of results for individual τ (Quantile regression for enhanced indexation)

A.1 Enhanced indexation

This chapter provides the remaining details of tables of results for individual τ values (such as given in Tables 4.5 and 4.6) quantile regression for enhanced indexation approach, Chapter 4 Section 4.4.3

Table A.1: In-sample and out-of-sample enhanced indexation results, $\tau = 0.35$

Index	Transaction	In-sample		Time	Out-of-sample		
(N, K)	cost limit γ	D^*	E^*	(secs)	Intercept	Slope	AER
Hang Seng	0.0025	0.01041	0.09347	0.7	-0.00565	0.89724	-0.46
(31,10)	0.0050	0.00928	0.11083	0.7	-0.00382	0.87776	3.44
	0.0075	0.00834	0.12779	0.7	-0.00321	0.93421	5.83
	0.01	0.00745	0.14375	0.7	-0.00122	0.88565	8.23
DAX 100	0.0025	0.00741	0.4203	1.8	-0.00513	1.04072	11.22
(85,10)	0.0050	0.00598	0.39927	1.6	-0.00546	1.06007	12.01
	0.0075	0.00487	0.39112	1.8	-0.00641	1.09891	13.17
	0.01	0.00408	0.36427	2.4	-0.00554	1.10631	14.41
FTSE 100	0.0025	0.00618	0.3153	1.7	-0.00366	0.74588	-0.54
(89,10)	0.0050	0.00505	0.41192	2.0	-0.00549	0.82603	-0.90
	0.0075	0.00408	0.49533	2.5	-0.00472	0.69927	0.21
	0.01	0.00322	0.56961	2.1	-0.00536	0.64974	0.64
S&P 100	0.0025	0.00915	0.03503	2.0	-0.00260	0.79653	-1.16
(98,10)	0.0050	0.00797	0.02004	2.2	-0.00360	0.90712	-2.66
	0.0075	0.00701	0.07209	2.2	-0.00414	0.87139	-7.23
	0.01	0.00615	0.16284	2.1	-0.00494	0.82720	-13.31
Nikkei 225	0.0025	0.00866	0.18093	5.1	-0.00424	0.85363	-1.03
(225,10)	0.0050	0.00744	0.21585	4.8	-0.00459	0.83245	-3.23
	0.0075	0.0063	0.24763	5.9	-0.00554	0.74315	-5.80
	0.01	0.0053	0.29356	5.4	-0.00503	0.71532	-7.62
S&P 500	0.0025	0.01517	0.30641	10.6	-0.00578	1.22262	1.45
(457,40)	0.0050	0.01286	0.18618	11.0	-0.00638	1.25761	-0.21
	0.0075	0.01056	0.05993	10.2	-0.00848	1.37167	-0.78
	0.01	0.00834	0.0151	10.1	-0.01009	1.35275	-1.90
Russell 2000	0.0025			infeasible			
(1318,90)	0.0050	0.01617	0.28073	50.2	-0.00157	0.88481	11.61
	0.0075	0.01196	0.39745	81.5	-0.00255	0.72718	10.00
	0.01	0.00928	0.5239	33.7	-0.00214	0.50074	10.98
Russell 3000	0.0025	0.01899	0.13972	153.9	-0.00457	1.07307	20.11
(2151,70)	0.0050	0.01459	0.27936	57.7	-0.00335	0.77638	21.06
	0.0075	0.01124	0.41178	53.7	-0.00120	0.54531	21.55
	0.01	0.00851	0.54258	51.4	-0.00148	0.49701	20.29
Average		0.00877	0.26497	18.5	-0.00445	0.88960	4.50

Table A.2: In-sample and out-of-sample enhanced indexation results, $\tau = 0.30$

Index	Transaction	In-sample		Time	Out-of-sample		
(N, K)	cost limit γ	D^*	E^*	(secs)	Intercept	Slope	AER
Hang Seng	0.0025	0.01469	0.10762	0.7	-0.00777	0.88037	-0.91
(31,10)	0.0050	0.01304	0.1146	0.7	-0.00570	0.92557	2.10
	0.0075	0.01142	0.11729	0.7	-0.00296	0.92192	4.75
	0.01	0.01002	0.13452	0.7	-0.00203	0.90525	8.15
DAX 100	0.0025	0.00998	0.41643	1.8	-0.00751	0.99196	11.49
(85,10)	0.0050	0.00838	0.40433	1.8	-0.00778	1.12381	13.11
	0.0075	0.00714	0.38759	1.9	-0.00673	1.10229	14.78
	0.01	0.00625	0.35765	1.8	-0.00588	1.08499	16.00
FTSE 100	0.0025	0.00926	0.30908	1.6	-0.00558	0.74435	-0.39
(89,10)	0.0050	0.00794	0.3968	1.7	-0.00737	0.81160	-0.92
	0.0075	0.00681	0.47504	1.6	-0.00857	0.80856	-0.20
	0.01	0.00591	0.5576	1.8	-0.00775	0.76398	0.72
S&P 100	0.0025	0.01271	0.02863	2.1	-0.00565	0.83769	-1.32
(98,10)	0.0050	0.01109	0.15429	1.9	-0.00434	0.85246	-6.82
	0.0075	0.00967	0.19348	1.9	-0.00511	0.90300	-10.29
	0.01	0.00842	0.21732	1.8	-0.00665	0.91144	-12.61
Nikkei 225	0.0025	0.01229	0.2222	5.1	-0.00551	0.89761	0.60
(225,10)	0.0050	0.0109	0.2432	4.8	-0.00585	0.84723	1.10
	0.0075	0.00968	0.291	5.0	-0.00503	0.80809	5.35
	0.01	0.00859	0.33681	5.0	-0.00532	0.73568	8.33
S&P 500	0.0025	0.01971	0.20327	10.8	-0.00661	1.07569	3.42
(457,40)	0.0050	0.01709	0.00521	10.4	-0.00605	0.88827	7.59
	0.0075	0.01454	0.24245	10.3	-0.00668	0.67079	11.92
	0.01	0.01203	0.4508	10.3	-0.00766	0.51652	15.24
Russell 2000	0.0025			infeasible			
(1318,90)	0.0050	0.02217	0.28588	66.7	-0.00420	0.87312	10.48
	0.0075	0.01707	0.41799	59.8	-0.00214	0.70915	10.29
	0.01	0.0136	0.53934	32.4	-0.00355	0.46032	10.62
Russell 3000	0.0025	0.02568	0.13521	230.1	-0.00791	1.00496	19.79
(2151,70)	0.0050	0.02008	0.29753	50.2	-0.00479	0.73536	21.08
	0.0075	0.0158	0.45088	62.3	-0.00246	0.52566	21.04
	0.01	0.01204	0.58557	52.1	-0.00356	0.50952	20.72
Average		0.01239	0.29289	20.6	-0.00564	0.83314	6.62

Table A.3: In-sample and out-of-sample enhanced indexation results, $\tau = 0.25$

Index	Transaction	In-sample		Time	Out-of-sample		
(N, K)	cost limit γ	D^*	E^*	(secs)	Intercept	Slope	AER
Hang Seng	0.0025	0.0188	0.11887	0.7	-0.00795	0.88749	-1.26
(31,10)	0.0050	0.01644	0.11631	0.7	-0.00432	0.92869	1.62
	0.0075	0.01419	0.12219	0.7	-0.00326	0.91400	4.65
	0.01	0.0125	0.1216	0.7	-0.00309	0.91192	7.42
DAX 100	0.0025	0.0136	0.36268	1.6	-0.00966	1.04803	11.68
(85,10)	0.0050	0.01172	0.35407	1.6	-0.00804	1.12293	13.36
	0.0075	0.01016	0.32693	1.6	-0.00733	1.08639	15.24
	0.01	0.00896	0.2977	1.6	-0.00803	1.00600	16.61
FTSE 100	0.0025	0.01373	0.29718	1.5	-0.00671	0.71399	-0.82
(89,10)	0.0050	0.012	0.39956	1.8	-0.00778	0.72764	-1.02
	0.0075	0.0105	0.48078	2.1	-0.00907	0.69944	-0.70
	0.01	0.00913	0.56245	1.7	-0.01029	0.61600	-0.48
S&P 100	0.0025	0.01687	0.05458	1.9	-0.00608	0.85335	-1.07
(98,10)	0.0050	0.01478	0.01045	1.8	-0.00553	0.91534	-4.30
	0.0075	0.01275	0.09541	2.1	-0.00707	0.92442	-9.42
	0.01	0.01112	0.15001	1.9	-0.00824	0.89611	-12.63
Nikkei 225	0.0025	0.015	0.22734	5.1	-0.00644	0.88893	0.60
(225,10)	0.0050	0.01333	0.23961	4.2	-0.00640	0.84063	0.89
	0.0075	0.01202	0.27642	4.8	-0.00689	0.80477	5.00
	0.01	0.01092	0.32689	5.0	-0.00790	0.75876	8.08
S&P 500	0.0025	0.02608	0.16616	10.7	-0.00974	1.04439	3.00
(457,40)	0.0050	0.02268	0.04555	11.8	-0.00817	0.96508	5.57
	0.0075	0.01928	0.29968	10.3	-0.00730	0.76735	8.49
	0.01	0.01596	0.5006	11.4	-0.00934	0.56857	10.21
Russell 2000	0.0025			infeasible			
(1318,90)	0.0050	0.02894	0.2716	112.0	-0.00685	0.93855	12.20
	0.0075	0.02263	0.4063	55.7	-0.00474	0.68038	10.74
	0.01	0.01826	0.52005	31.5	-0.00417	0.46606	10.06
Russell 3000	0.0025	0.03294	0.11045	196.4	-0.00823	1.02174	21.09
(2151,70)	0.0050	0.02624	0.28261	52.8	-0.00565	0.75395	21.36
	0.0075	0.02061	0.43949	54.5	-0.00444	0.56574	21.97
	0.01	0.01605	0.55879	52.0	-0.00455	0.54109	21.02
Average		0.01639	0.27556	20.7	-0.00688	0.83412	6.42

Table A.4: In-sample and out-of-sample enhanced indexation results, $\tau = 0.20$

Index	Transaction	In-sample		Time	Out-of-sample		
(N, K)	cost limit γ	D^*	E^*	(secs)	Intercept	Slope	AER
Hang Seng	0.0025	0.02406	0.14055	0.6	-0.00930	0.89916	-1.26
(31,10)	0.0050	0.0214	0.14456	0.8	-0.00573	0.91736	1.62
	0.0075	0.01893	0.1574	0.6	-0.00523	0.94197	4.65
	0.01	0.01704	0.17898	0.6	-0.00432	0.89167	8.28
DAX 100	0.0025	0.01725	0.36213	1.8	-0.01397	1.08786	11.49
(85,10)	0.0050	0.01501	0.35427	1.7	-0.01182	1.06317	13.60
	0.0075	0.01304	0.34215	1.7	-0.01086	1.02762	14.99
	0.01	0.01158	0.31368	1.5	-0.01033	1.00397	16.17
FTSE 100	0.0025	0.01789	0.2532	1.8	-0.00833	0.63605	-0.83
(89,10)	0.0050	0.01583	0.35213	1.7	-0.00995	0.71242	-0.89
	0.0075	0.01393	0.4496	1.8	-0.01184	0.58604	-0.68
	0.01	0.0122	0.52236	1.6	-0.01289	0.63900	-0.45
S&P 100	0.0025	0.02127	0.0231	2.1	-0.00818	0.95350	0.78
(98,10)	0.0050	0.01849	0.15319	1.8	-0.00726	0.95845	-2.76
	0.0075	0.01578	0.20008	1.9	-0.00805	0.95283	-3.63
	0.01	0.01395	0.27239	1.7	-0.00986	0.99327	-3.71
Nikkei 225	0.0025	0.01854	0.21677	4.5	-0.00896	0.90353	-1.99
(225,10)	0.0050	0.01689	0.2253	4.1	-0.00872	0.88815	-3.35
	0.0075	0.0154	0.25492	4.8	-0.00886	0.87689	-1.16
	0.01	0.01408	0.29596	4.7	-0.00869	0.81664	0.27
S&P 500	0.0025	0.03346	0.21766	10.8	-0.01229	1.14165	3.71
(457,40)	0.0050	0.02897	0.04757	10.8	-0.00795	0.95528	7.93
	0.0075	0.02448	0.3153	9.5	-0.00831	0.80542	11.69
	0.01	0.0202	0.48135	9.2	-0.01013	0.58225	13.46
Russell 2000	0.0025			infeasible			
(1318,90)	0.0050	0.03722	0.26444	200.0	-0.00889	0.95563	12.36
	0.0075	0.02948	0.42221	70.5	-0.00595	0.62955	10.64
	0.01	0.02395	0.51641	31.7	-0.00786	0.50872	15.27
Russell 3000	0.0025	0.0414	0.10084	184.6	-0.01420	1.03464	20.70
(2151,70)	0.0050	0.03308	0.27864	55.7	-0.02044	0.73906	35.00
	0.0075	0.02623	0.41874	50.8	-0.02162	0.56231	40.18
	0.01	0.02054	0.55888	49.3	-0.02627	0.47262	45.79
Average		0.02102	0.28499	23.4	-0.01055	0.84312	8.64

Table A.5: In-sample and out-of-sample enhanced indexation results, $\tau = 0.15$

Index	Transaction	In-sample		Time	Out-of-sample		
(N, K)	cost limit γ	D^*	E^*	(secs)	Intercept	Slope	AER
Hang Seng	0.0025	0.02937	0.14869	0.7	-0.01203	0.93749	-1.26
(31,10)	0.0050	0.02626	0.14895	0.7	-0.00780	0.92982	1.62
	0.0075	0.02333	0.15696	0.7	-0.00695	0.92908	4.65
	0.01	0.02101	0.17669	0.8	-0.00661	0.90979	8.15
DAX 100	0.0025	0.02282	0.37496	1.9	-0.01573	1.16835	10.98
(85,10)	0.0050	0.02016	0.36449	2.1	-0.01508	1.14648	11.81
	0.0075	0.01783	0.3248	1.9	-0.01511	1.19110	13.46
	0.01	0.01601	0.27921	1.9	-0.01560	1.14943	12.90
FTSE 100	0.0025	0.02325	0.2549	1.9	-0.01000	0.71053	-0.68
(89,10)	0.0050	0.02111	0.32934	2.1	-0.01249	0.62642	-0.86
	0.0075	0.01914	0.40548	2.0	-0.01421	0.54633	-0.55
	0.01	0.01728	0.48137	2.0	-0.01786	0.69021	0.00
S&P 100	0.0025	0.02572	0.02382	2.4	-0.01146	0.95307	-1.26
(98,10)	0.0050	0.02253	0.07792	2.3	-0.00907	0.95414	-6.98
	0.0075	0.01949	0.13272	2.3	-0.00981	0.92296	-9.63
	0.01	0.01702	0.18451	2.3	-0.01363	0.94480	-12.33
Nikkei 225	0.0025	0.0233	0.25024	5.7	-0.00994	0.85763	0.60
(225,10)	0.0050	0.02173	0.29531	5.6	-0.01066	0.85111	3.46
	0.0075	0.02038	0.32247	5.3	-0.01200	0.85580	4.99
	0.01	0.01926	0.37949	5.0	-0.01318	0.80628	8.64
S&P 500	0.0025	0.04016	0.23238	9.9	-0.01435	1.08306	2.85
(457,40)	0.0050	0.035	0.03509	10.1	-0.01317	0.94769	5.94
	0.0075	0.02983	0.30636	10.6	-0.01553	0.77764	8.83
	0.01	0.02486	0.5045	10.6	-0.01825	0.60135	10.97
Russell 2000	0.0025			infeasible			
(1318,90)	0.0050	0.04659	0.26817	49.0	-0.01132	0.96118	10.68
	0.0075	0.03736	0.42891	60.3	-0.00710	0.74760	11.16
	0.01	0.03069	0.53383	40.1	-0.00967	0.51662	14.77
Russell 3000	0.0025	0.05121	0.10166	103.3	-0.01729	1.16101	20.53
(2151,70)	0.0050	0.04119	0.26426	57.8	-0.02358	0.79503	35.02
	0.0075	0.03278	0.39708	50.3	-0.02818	0.74285	40.72
	0.01	0.02592	0.54967	46.4	-0.03438	0.60756	44.89
Average		0.02654	0.28175	16.1	-0.01394	0.87169	8.20

Table A.6: In-sample and out-of-sample enhanced indexation results, $\tau = 0.10$

Index	Transaction	In-sample		Time	Out-of-sample		
(N, K)	cost limit γ	D^*	E^*	(secs)	Intercept	Slope	AER
Hang Seng	0.0025	0.03598	0.10896	0.9	-0.01764	1.07625	-0.97
(31,10)	0.0050	0.03322	0.09906	0.8	-0.01409	0.99128	1.10
	0.0075	0.03075	0.10112	0.8	-0.01421	0.95519	2.74
	0.01	0.02907	0.11457	0.7	-0.01483	0.96885	5.46
DAX 100	0.0025	0.0287	0.34458	2.1	-0.02255	1.30118	11.97
(85,10)	0.0050	0.02541	0.34724	1.8	-0.02086	1.20970	13.26
	0.0075	0.02284	0.33351	1.8	-0.01799	1.11728	12.12
	0.01	0.02049	0.26189	1.9	-0.01569	0.88788	5.76
FTSE 100	0.0025	0.03174	0.24222	2.1	-0.01106	0.71158	-2.14
(89,10)	0.0050	0.02941	0.22159	2.0	-0.01168	0.77554	-3.34
	0.0075	0.02765	0.20787	1.9	-0.01501	0.92157	-5.19
	0.01	0.02629	0.22139	2.0	-0.01701	0.94304	-6.56
S&P 100	0.0025	0.03296	0.05437	2.1	-0.01080	0.97610	-3.90
(98,10)	0.0050	0.02865	0.14476	2.0	-0.01269	0.90645	-8.32
	0.0075	0.02498	0.17132	2.1	-0.01536	0.93954	-9.79
	0.01	0.02211	0.23391	2.2	-0.01663	0.92838	-12.54
Nikkei 225	0.0025	0.02906	0.26047	5.3	-0.01196	0.82883	2.16
(225,10)	0.0050	0.02701	0.32146	4.8	-0.01268	0.81515	6.19
	0.0075	0.02533	0.37012	4.9	-0.01496	0.79549	8.98
	0.01	0.02389	0.44739	4.8	-0.01773	0.74275	13.15
S&P 500	0.0025	0.05076	0.37049	10.3	-0.02090	1.21667	2.95
(457,40)	0.0050	0.04454	0.20621	9.1	-0.01740	1.10538	4.65
	0.0075	0.03837	0.00884	9.0	-0.01591	1.02629	6.93
	0.01	0.03237	0.15909	9.6	-0.02091	0.96395	8.61
Russell 2000	0.0025			infeasible			
(1318,90)	0.0050	0.06052	0.2388	72.6	-0.01472	0.95083	10.53
	0.0075	0.04866	0.39709	40.3	-0.01187	0.81610	10.83
	0.01	0.03999	0.49362	30.5	-0.01019	0.49481	13.91
Russell 3000	0.0025	0.06452	0.06201	100.2	-0.02337	1.20847	20.45
(2151,70)	0.0050	0.05218	0.23544	46.7	-0.03557	0.86162	35.15
	0.0075	0.04189	0.38368	53.7	-0.04317	0.64566	41.12
	0.01	0.0335	0.51042	49.6	-0.04816	0.52726	45.33
Average		0.03429	0.24753	15.4	-0.01831	0.92287	7.44

Table A.7: In-sample and out-of-sample enhanced indexation results, $\tau = 0.05$

Index	Transaction	In-sample		Time	Out-of-sample		
(N, K)	cost limit γ	D^*	E^*	(secs)	Intercept	Slope	AER
Hang Seng	0.0025	0.04777	0.10489	0.7	-0.01932	0.98130	-0.14
(31,10)	0.0050	0.04521	0.10423	0.5	-0.01350	1.04456	2.65
	0.0075	0.0427	0.07396	0.6	-0.01172	0.99555	5.37
	0.01	0.04042	0.0853	0.6	-0.01402	0.99050	7.46
DAX 100	0.0025	0.04443	0.50662	1.7	-0.03161	1.35978	10.05
(85,10)	0.0050	0.03945	0.37343	1.7	-0.03101	0.96267	5.39
	0.0075	0.03447	0.22039	1.6	-0.03106	0.48182	-2.16
	0.01	0.02982	0.16664	1.7	-0.03056	0.48761	-0.73
FTSE 100	0.0025	0.04682	0.38367	1.7	-0.01412	0.87375	0.18
(89,10)	0.0050	0.04322	0.40395	1.6	-0.01288	0.73962	2.19
	0.0075	0.03984	0.40331	1.7	-0.01364	0.72089	3.21
	0.01	0.03683	0.39611	1.8	-0.01660	0.73428	4.37
S&P 100	0.0025	0.04581	0.15214	1.9	-0.01535	0.96305	-3.89
(98,10)	0.0050	0.03939	0.25464	1.7	-0.01803	0.99059	-8.08
	0.0075	0.0343	0.27909	1.8	-0.01867	0.95502	-10.96
	0.01	0.03	0.28862	1.8	-0.02030	0.84883	-12.96
Nikkei 225	0.0025	0.03884	0.31762	4.7	-0.01632	0.85139	2.18
(225,10)	0.0050	0.03584	0.37177	4.3	-0.02086	0.76476	5.80
	0.0075	0.03305	0.44248	4.1	-0.01984	0.79125	10.94
	0.01	0.03059	0.46666	4.4	-0.02287	0.70548	13.37
S&P 500	0.0025	0.06462	0.34865	8.8	-0.03026	1.13273	3.15
(457,40)	0.0050	0.05648	0.19173	9.9	-0.02777	1.03724	3.97
	0.0075	0.04896	0.01303	9.9	-0.02211	1.10574	6.29
	0.01	0.04145	0.15306	8.9	-0.03080	1.00246	8.67
Russell 2000	0.0025			infeasible			
(1318,90)	0.0050	0.07132	0.16588	48.6	-0.03611	1.07651	29.84
	0.0075	0.05716	0.36319	47.7	-0.05077	0.69956	41.62
	0.01	0.04553	0.47967	47.9	-0.06147	0.53086	44.92
Russell 3000	0.0025	0.08816	0.04412	109.4	-0.03213	0.99695	20.00
(2151,70)	0.0050	0.07132	0.16588	47.0	-0.03611	1.07651	29.84
	0.0075	0.05716	0.36319	46.3	-0.05077	0.69956	41.62
	0.01	0.04553	0.47967	44.4	-0.06147	0.53086	44.92
Average		0.04602	0.27624	15.1	-0.02684	0.87522	9.97

Appendix B

Detailed tables of results for individual τ (first alternative approach quantile regression for index tracking and enhanced indexation)

B.1 First alternative approach tables details

This chpater provides details of tables of results for individual τ values (such as given in Tables 4.5 and 4.6) first alternative approach quantile regression for index tracking and enhanced indexation, Chapter 4 Section 4.4.4.

Table B.1: In-sample and out-of-sample enhanced indexation results, $\tau = 0.45$

Index	Transaction	In-sample		Time	Out-of-sample		
(N, K)	cost limit γ	D^*	E^*	(secs)	Intercept	Slope	AER
Hang Seng	0.0025	0.00339	0	0.6	-0.00532	0.99664	-2.90
(31,10)	0.0050	0.00235	0	0.6	-0.00381	0.99220	-6.59
	0.0075	0.00167	0	0.6	-0.00331	0.99258	-5.65
	0.01	0.00111	0	0.6	-0.00279	0.99405	-5.95
DAX 100	0.0025	0.00287	0.2658	1.7	-0.00054	1.02493	9.84
(85,10)	0.0050	0.00318	0.12738	1.7	-0.00016	1.07664	3.02
	0.0075	0.0032	0.00547	1.6	-0.00209	0.91740	-5.58
	0.01	0.00074	0	1.6	0.00054	1.23131	13.59
FTSE 100	0.0025	0.00083	0.06543	1.8	0.00035	0.86437	4.46
(89,10)	0.0050	0.00032	0	2.0	0.00102	0.88119	5.69
	0.0075	0	0	1.7	0.00246	0.87509	7.13
	0.01	0	0	1.7	0.00246	0.87509	7.13
S&P 100	0.0025	0.00074	0	2.0	-0.00112	1.16614	5.10
(98,10)	0.0050	0	0	1.9	-0.00071	1.17571	6.71
	0.0075	0	0	1.9	-0.00071	1.17571	6.71
	0.01	0	0	2.3	-0.00071	1.17571	6.71
Nikkei 225	0.0025	0.00379	0	5.4	-0.00145	1.01651	0.26
(225,10)	0.0050	0.00211	0	4.1	-0.00151	1.04496	0.10
	0.0075	0.00088	0	4.1	-0.00096	1.05242	1.71
	0.01	0.00004	0	4.1	-0.00267	1.00974	0.68
S&P 500	0.0025	0.0039	0.03298	9.0	0.00174	1.08745	8.46
(457,40)	0.0050	0.00101	0	8.8	-0.00217	0.93213	6.51
	0.0075	0	0	8.9	-0.00219	0.93692	6.38
	0.01	0	0	8.6	-0.00219	0.93692	6.38
Russell 2000	0.0025			infeasible			
(1318,90)	0.0050	0.0045	0	64.2	0.00087	1.20617	8.08
	0.0075	0.00139	0	28.8	-0.00021	1.12816	3.19
	0.01	0	0	30.1	-0.00036	0.98825	1.99
Russell 3000	0.0025	0.00619	0	52.4	-0.00071	1.16259	15.49
(2151,70)	0.0050	0.00278	0	47.9	-0.00104	1.17817	12.81
	0.0075	0.00038	0	47.4	0.00037	1.15754	11.03
	0.01	0	0	66.9	0.00124	1.17578	10.89
Average		0.00153	0.01603	13.4	-0.00083	1.04608	4.63

Table B.2: In-sample and out-of-sample enhanced indexation results, $\tau = 0.40$

Index	Transaction	In-sample		Time	Out-of-sample		
(N, K)	cost limit γ	D^*	E^*	(secs)	Intercept	Slope	AER
Hang Seng	0.0025	0.00801	0	0.7	-0.00550	0.98053	-3.32
(31,10)	0.0050	0.00714	0	0.7	-0.00523	1.00211	-6.14
	0.0075	0.00645	0	0.6	-0.00550	1.02867	-5.68
	0.01	0.00579	0	0.5	-0.00528	1.06906	-5.82
DAX 100	0.0025	0.0058	0.29239	1.9	-0.00266	1.02849	9.84
(85,10)	0.0050	0.00639	0.14265	1.4	-0.00193	1.08721	3.00
	0.0075	0.00644	0.01207	1.6	-0.00260	0.92899	-5.61
	0.01	0.00364	0	1.6	-0.00138	1.21784	13.42
FTSE 100	0.0025	0.00432	0.06727	2.4	-0.00261	0.87471	1.24
(89,10)	0.0050	0.0034	0	2.5	-0.00107	0.87355	0.71
	0.0075	0.00255	0	2.0	-0.00272	0.91935	-1.49
	0.01	0.00177	0	1.9	-0.00467	0.85553	-4.87
S&P 100	0.0025	0.00517	0	2.2	-0.00254	1.01841	3.81
(98,10)	0.0050	0.00415	0	2.0	-0.00218	1.08597	5.21
	0.0075	0.00348	0	2.1	-0.00162	1.17905	7.77
	0.01	0.00303	0	2.1	-0.00115	1.12206	9.45
Nikkei 225	0.0025	0.00622	0	5.5	-0.00276	1.03135	-2.99
(225,10)	0.0050	0.005	0	5.8	-0.00291	1.04529	-2.28
	0.0075	0.00428	0	5.0	-0.00256	1.08354	-0.72
	0.01	0.00364	0	5.4	-0.00419	1.06781	-2.58
S&P 500	0.0025	0.00858	0	10.8	-0.00078	1.21788	4.67
(457,40)	0.0050	0.00525	0	11.0	-0.00204	0.99535	3.00
	0.0075	0.00416	0	10.9	-0.00251	1.12352	0.80
	0.01	0.00322	0	10.5	-0.00532	1.29344	-1.60
Russell 2000	0.0025			infeasible			
(1318,90)	0.0050	0.01132	0	33.5	-0.00198	1.19462	7.47
	0.0075	0.00799	0	46.8	-0.00145	1.04575	4.61
	0.01	0.00632	0	46.0	-0.00327	1.04211	1.63
Russell 3000	0.0025	0.01273	0	50.7	-0.00203	1.10342	15.82
(2151,70)	0.0050	0.00933	0	61.4	-0.00201	1.02432	14.24
	0.0075	0.0069	0	53.2	-0.00165	0.98493	10.56
	0.01	0.00508	0	63.5	-0.00219	1.05542	8.91
Average		0.00573	0.01659	14.4	-0.00278	1.05098	2.68

Table B.3: In-sample and out-of-sample enhanced indexation results, $\tau = 0.35$

Index	Transaction	In-sample		Time	Out-of-sample		
(N, K)	cost limit γ	D^*	E^*	(secs)	Intercept	Slope	AER
Hang Seng	0.0025	0.01115	0	1.1	-0.00667	0.96085	-2.18
(31,10)	0.0050	0.01014	0	0.6	-0.00510	1.01051	-0.48
	0.0075	0.00921	0	0.6	-0.00427	1.02722	2.33
	0.01	0.00846	0	0.5	-0.00294	1.04302	4.08
DAX 100	0.0025	0.00863	0.28718	1.6	-0.00474	1.09767	9.84
(85,10)	0.0050	0.00902	0.13751	1.6	-0.00309	1.02789	3.00
	0.0075	0.00899	0.00078	1.7	-0.00362	0.93321	-5.61
	0.01	0.00577	0	1.6	-0.00317	1.23197	12.21
FTSE 100	0.0025	0.00884	0.07571	1.9	-0.00316	0.81863	1.47
(89,10)	0.0050	0.00759	0	1.7	-0.00319	0.89505	0.95
	0.0075	0.0063	0	1.7	-0.00374	0.99899	1.57
	0.01	0.0053	0	1.7	-0.00466	1.05635	1.67
S&P 100	0.0025	0.00933	0	2.3	-0.00352	0.92274	-0.98
(98,10)	0.0050	0.00803	0	1.8	-0.00295	0.86322	-1.26
	0.0075	0.00717	0	1.8	-0.00337	0.88473	-3.88
	0.01	0.0065	0	1.7	-0.00413	1.01475	-3.79
Nikkei 225	0.0025	0.01014	0	4.5	-0.00443	1.00957	-3.09
(225,10)	0.0050	0.00885	0	4.1	-0.00459	1.06873	-5.83
	0.0075	0.00768	0	4.0	-0.00410	1.10284	-5.08
	0.01	0.00662	0	4.1	-0.00514	1.14875	-6.20
S&P 500	0.0025	0.01662	0.06329	16.7	-0.00308	1.15695	10.01
(457,40)	0.0050	0.01299	0	8.7	-0.00261	0.90276	7.64
	0.0075	0.01058	0	8.7	-0.00235	1.15198	4.71
	0.01	0.00835	0	10.2	-0.01076	1.34091	-2.08
Russell 2000	0.0025			infeasible			
(1318,90)	0.0050	0.01824	0	105.7	-0.00380	1.14312	5.02
	0.0075	0.01438	0	31.0	-0.00433	1.02247	5.30
	0.01	0.01217	0	27.4	-0.00330	0.92831	6.37
Russell 3000	0.0025	0.01959	0	151.3	-0.00535	1.11189	16.47
(2151,70)	0.0050	0.01554	0	69.0	-0.00239	1.00475	17.46
	0.0075	0.01258	0	76.1	-0.00393	1.01429	13.88
	0.01	0.01018	0	53.9	-0.00354	1.08363	10.62
Average		0.01016	0.01821	19.3	-0.00407	1.03154	3.04

Table B.4: In-sample and out-of-sample enhanced indexation results, $\tau = 0.30$

Index	Transaction	In-sample		Time	Out-of-sample		
(N, K)	cost limit γ	D^*	E^*	(secs)	Intercept	Slope	AER
Hang Seng	0.0025	0.01577	0	0.7	-0.00765	0.97765	-3.01
(31,10)	0.0050	0.01413	0	0.6	-0.00540	1.02418	-0.23
	0.0075	0.01251	0	0.6	-0.00278	1.00043	2.55
	0.01	0.01114	0	0.7	-0.00228	1.02826	3.36
DAX 100	0.0025	0.01183	0.28878	2.0	-0.00541	1.06561	10.99
(85,10)	0.0050	0.01216	0.13757	1.7	-0.00474	1.03534	2.84
	0.0075	0.0121	0.00147	1.7	-0.00623	1.01366	-5.91
	0.01	0.00828	0	1.7	-0.00585	1.27648	12.06
FTSE 100	0.0025	0.01244	0.07722	2.1	-0.00398	0.83968	1.47
(89,10)	0.0050	0.01127	0	1.8	-0.00435	0.88922	2.27
	0.0075	0.00962	0	1.7	-0.00547	0.97103	1.41
	0.01	0.00842	0	2.1	-0.00880	1.13569	1.76
S&P 100	0.0025	0.01281	0	2.4	-0.00487	0.88009	0.63
(98,10)	0.0050	0.01142	0	2.0	-0.00529	0.97864	-1.55
	0.0075	0.01038	0	1.9	-0.00532	1.12618	2.98
	0.01	0.00944	0	1.6	-0.00445	1.11890	4.59
Nikkei 225	0.0025	0.01525	0.018	4.4	-0.00455	1.01275	-0.85
(225,10)	0.0050	0.01321	0	4.6	-0.00554	1.11033	-5.17
	0.0075	0.01176	0	4.0	-0.00636	1.13923	-9.61
	0.01	0.01034	0	4.4	-0.00704	1.12395	-10.46
S&P 500	0.0025	0.02148	0.02945	10.6	-0.00407	1.16601	10.01
(457,40)	0.0050	0.0171	0	10.2	-0.00636	0.89036	7.54
	0.0075	0.01484	0	10.2	-0.00665	1.06477	4.64
	0.01	0.01264	0	9.3	-0.01356	1.32129	-0.35
Russell 2000	0.0025			infeasible			
(1318,90)	0.0050	0.02519	0	53.5	-0.00564	1.11580	7.68
	0.0075	0.02084	0	53.1	-0.00616	1.02492	7.99
	0.01	0.01825	0	31.0	-0.00530	0.91046	8.06
Russell 3000	0.0025	0.02646	0	104.2	-0.00712	1.11805	16.66
(2151,70)	0.0050	0.02167	0	54.6	-0.00441	0.99363	17.38
	0.0075	0.01814	0	59.5	-0.00524	0.98520	14.51
	0.01	0.01518	0	51.6	-0.00448	1.07915	10.63
Average		0.01439	0.01782	15.8	-0.00566	1.04571	3.71

Table B.5: In-sample and out-of-sample enhanced indexation results, $\tau = 0.25$

Index	Transaction	In-sample		Time	Out-of-sample		
(N, K)	cost limit γ	D^*	E^*	(secs)	Intercept	Slope	AER
Hang Seng	0.0025	0.02132	0	0.7	-0.01012	0.96928	-5.31
(31,10)	0.0050	0.01819	0	0.7	-0.00700	1.03320	-2.04
	0.0075	0.01573	0	0.6	-0.00354	1.01322	2.01
	0.01	0.01426	0	0.6	-0.00339	1.02416	3.04
DAX 100	0.0025	0.01529	0.25945	1.7	-0.01034	1.14765	9.84
(85,10)	0.0050	0.01552	0.11466	1.4	-0.00763	1.11945	3.00
	0.0075	0.01461	0	1.9	-0.00715	0.98518	-2.18
	0.01	0.01099	0	1.6	-0.00829	1.23031	11.57
FTSE 100	0.0025	0.01738	0.07411	1.8	-0.00483	0.83656	1.47
(89,10)	0.0050	0.01661	0	1.8	-0.00738	0.95412	5.25
	0.0075	0.01503	0	1.6	-0.00582	1.05963	7.83
	0.01	0.01383	0	1.5	-0.00758	1.10040	9.00
S&P 100	0.0025	0.01693	0	2.1	-0.00561	0.89756	-3.75
(98,10)	0.0050	0.01479	0	1.8	-0.00550	0.95445	-4.62
	0.0075	0.01318	0	1.9	-0.00690	0.94036	-5.97
	0.01	0.01199	0	2.0	-0.00748	0.96940	-8.38
Nikkei 225	0.0025	0.01809	0.02526	4.5	-0.00595	1.02407	-1.79
(225,10)	0.0050	0.01576	0	4.1	-0.00904	1.06141	-11.79
	0.0075	0.01408	0	4.3	-0.00905	1.07110	-9.72
	0.01	0.0125	0	4.2	-0.00862	1.13227	-9.26
S&P 500	0.0025	0.02831	0.01265	15.6	-0.01166	1.33331	4.81
(457,40)	0.0050	0.02269	0	8.7	-0.00725	0.96894	6.23
	0.0075	0.01965	0	8.7	-0.00690	0.97748	5.77
	0.01	0.01692	0	8.8	-0.00875	0.99284	3.37
Russell 2000	0.0025			infeasible			
(1318,90)	0.0050	0.03255	0	107.2	-0.00650	1.10071	7.55
	0.0075	0.02755	0	29.9	-0.00663	0.96905	9.04
	0.01	0.02453	0	39.5	-0.00636	0.86104	9.54
Russell 3000	0.0025	0.03376	0	110.3	-0.00943	1.11317	16.51
(2151,70)	0.0050	0.02792	0	47.6	-0.00798	0.94822	17.99
	0.0075	0.02367	0	85.8	-0.00647	0.94424	16.33
	0.01	0.02017	0	59.1	-0.00652	0.83142	12.06
Average		0.01883	0.01568	18.1	-0.00728	1.01820	3.14

Table B.6: In-sample and out-of-sample enhanced indexation results, $\tau = 0.20$

Index	Transaction	In-sample		Time	Out-of-sample		
(N, K)	cost limit γ	D^*	E^*	(secs)	Intercept	Slope	AER
Hang Seng	0.0025	0.02698	0.0008	0.6	-0.01304	0.94377	-5.47
(31,10)	0.0050	0.02333	0	0.8	-0.01053	1.01067	-5.75
	0.0075	0.02096	0	0.6	-0.00589	1.04628	3.76
	0.01	0.01943	0	0.6	-0.00691	1.06389	6.61
DAX 100	0.0025	0.0195	0.26665	1.6	-0.01286	1.10790	9.84
(85,10)	0.0050	0.01976	0.10209	1.7	-0.01146	1.14849	3.00
	0.0075	0.0173	0	1.7	-0.01211	1.23227	3.27
	0.01	0.01437	0	1.6	-0.01060	1.29995	12.78
FTSE 100	0.0025	0.02203	0.02556	1.8	-0.00609	0.84322	1.47
(89,10)	0.0050	0.01999	0	1.9	-0.00756	0.93974	2.34
	0.0075	0.01881	0	1.8	-0.00910	0.90083	5.41
	0.01	0.01793	0	1.6	-0.00811	0.85576	5.24
S&P 100	0.0025	0.02128	0	2.0	-0.00821	0.90168	-1.76
(98,10)	0.0050	0.0186	0	2.0	-0.00848	1.03058	-3.00
	0.0075	0.01688	0	1.7	-0.00808	0.97843	-4.96
	0.01	0.01535	0	1.9	-0.00764	0.90459	-9.59
Nikkei 225	0.0025	0.02386	0.02929	6.1	-0.00783	1.03690	-3.42
(225,10)	0.0050	0.02053	0	4.2	-0.00769	1.06147	-3.97
	0.0075	0.01901	0	4.1	-0.00946	1.08858	-5.42
	0.01	0.01763	0	4.3	-0.01269	0.96228	-3.52
S&P 500	0.0025	0.03488	0.061	10.2	-0.00754	1.17151	9.69
(457,40)	0.0050	0.029	0	10.3	-0.00913	0.96977	6.94
	0.0075	0.02501	0	9.8	-0.01015	1.04751	6.95
	0.01	0.02147	0	9.9	-0.01163	1.10464	6.87
Russell 2000	0.0025			infeasible			
(1318,90)	0.0050	0.04164	0	58.3	-0.00886	1.08910	8.01
	0.0075	0.03551	0	229.5	-0.00773	0.84193	10.60
	0.01	0.03141	0	100.0	-0.00636	0.77080	9.85
Russell 3000	0.0025	0.04213	0	112.9	-0.01215	1.13534	19.30
(2151,70)	0.0050	0.03542	0	49.8	-0.00996	0.96585	23.81
	0.0075	0.03022	0	74.0	-0.00841	0.92803	23.25
	0.01	0.02585	0	55.8	-0.00739	0.88077	18.66
Average		0.02407	0.01566	24.6	-0.00915	1.00847	4.86

Table B.7: In-sample and out-of-sample enhanced indexation results, $\tau = 0.15$

Index	Transaction	In-sample		Time	Out-of-sample		
(N, K)	cost limit γ	D^*	E^*	(secs)	Intercept	Slope	AER
Hang Seng	0.0025	0.03283	0.00774	0.6	-0.01682	0.99701	-5.47
(31,10)	0.0050	0.02857	0	0.7	-0.00903	1.03040	-1.56
	0.0075	0.02569	0	0.7	-0.00652	1.01514	0.63
	0.01	0.02362	0	0.6	-0.00574	1.01532	3.69
DAX 100	0.0025	0.02873	0.25253	1.9	-0.01491	1.06956	11.00
(85,10)	0.0050	0.03212	0.06549	2.2	-0.01705	0.92861	2.87
	0.0075	0.02134	0	1.9	-0.01456	1.19519	5.40
	0.01	0.01828	0	1.7	-0.01179	1.19243	1.48
FTSE 100	0.0025	0.02854	0.03306	1.9	-0.00989	0.75842	0.35
(89,10)	0.0050	0.02507	0	1.7	-0.00912	0.91866	-0.39
	0.0075	0.0231	0	1.7	-0.00955	0.71907	-4.68
	0.01	0.02161	0	1.5	-0.01283	0.76487	-7.44
S&P 100	0.0025	0.02574	0	2.0	-0.00974	0.89670	-2.23
(98,10)	0.0050	0.02283	0	2.0	-0.01212	0.92175	-2.72
	0.0075	0.02053	0	2.3	-0.01012	0.98313	-5.29
	0.01	0.01858	0	2.1	-0.01133	0.96798	-7.35
Nikkei 225	0.0025	0.02841	0.03566	5.9	-0.01108	0.99765	-4.08
(225,10)	0.0050	0.02496	0	4.2	-0.01067	1.03675	-3.16
	0.0075	0.02352	0	4.2	-0.01074	1.01159	-4.38
	0.01	0.02216	0	4.1	-0.01494	1.03467	-9.19
S&P 500	0.0025	0.04223	0.0444	9.4	-0.01166	1.19068	8.16
(457,40)	0.0050	0.03502	0	9.8	-0.01267	0.98856	5.70
	0.0075	0.03082	0	9.4	-0.01329	1.04395	6.50
	0.01	0.02701	0	10.0	-0.01362	1.07905	7.95
Russell 2000	0.0025			infeasible			
(1318,90)	0.0050	0.05206	0	33.2	-0.01482	1.09715	8.36
	0.0075	0.04462	0	27.9	-0.01033	0.89752	11.40
	0.01	0.03962	0	28.5	-0.00990	0.78782	10.64
Russell 3000	0.0025	0.05194	0	170.0	-0.01759	1.25463	19.58
(2151,70)	0.0050	0.04363	0	52.0	-0.01614	1.10317	25.76
	0.0075	0.03727	0	52.6	-0.01846	0.91581	31.50
	0.01	0.03196	0	52.3	-0.02008	0.85628	32.54
Average		0.03008	0.01416	16.1	-0.01249	0.98934	4.37

Table B.8: In-sample and out-of-sample enhanced indexation results, $\tau = 0.10$

Index	Transaction	In-sample		Time	Out-of-sample		
(N, K)	cost limit γ	D^*	E^*	(secs)	Intercept	Slope	AER
Hang Seng	0.0025	0.03773	0	0.7	-0.01690	1.03068	-3.55
(31,10)	0.0050	0.0345	0	0.8	-0.01155	1.07424	-0.40
	0.0075	0.03164	0	0.6	-0.00972	0.97837	1.12
	0.01	0.0299	0	0.6	-0.01213	0.95739	4.61
DAX 100	0.0025	0.03774	0.20376	1.7	-0.02164	1.15915	6.86
(85,10)	0.0050	0.04435	0.01669	2.3	-0.02223	1.09512	-1.58
	0.0075	0.02676	0	1.7	-0.01440	0.92357	0.74
	0.01	0.02328	0	1.5	-0.01257	0.81882	-3.33
FTSE 100	0.0025	0.03668	0.05752	1.8	-0.01233	0.74833	0.79
(89,10)	0.0050	0.032	0	1.9	-0.01269	0.86932	4.12
	0.0075	0.03025	0	1.6	-0.01146	0.85928	2.68
	0.01	0.02872	0	1.7	-0.01252	0.89100	0.62
S&P 100	0.0025	0.03347	0	1.9	-0.01321	0.99955	-1.43
(98,10)	0.0050	0.03012	0	1.9	-0.01259	0.96345	-3.59
	0.0075	0.0269	0	1.9	-0.01367	1.00850	-8.54
	0.01	0.02443	0	1.8	-0.01344	1.00777	-10.00
Nikkei 225	0.0025	0.03545	0	5.8	-0.01338	0.97724	-3.83
(225,10)	0.0050	0.03031	0	4.6	-0.01427	1.04915	-2.82
	0.0075	0.02865	0	4.1	-0.01230	1.00604	-3.97
	0.01	0.02711	0	3.9	-0.01555	0.98084	-7.45
S&P 500	0.0025	0.05381	0.12091	9.3	-0.01559	1.21103	7.85
(457,40)	0.0050	0.0446	0	10.4	-0.01929	0.90635	6.18
	0.0075	0.03838	0	9.1	-0.01550	1.02232	6.98
	0.01	0.03349	0	9.3	-0.02221	1.03848	6.96
Russell 2000	0.0025			infeasible			
(1318,90)	0.0050	0.06615	0	46.2	-0.01839	1.00370	10.99
	0.0075	0.05612	0	39.1	-0.01133	0.83181	11.95
	0.01	0.04964	0	35.4	-0.01031	0.76752	10.56
Russell 3000	0.0025	0.065	0	100.0	-0.02395	1.20231	19.46
(2151,70)	0.0050	0.05443	0	50.4	-0.01993	1.06797	26.16
	0.0075	0.04639	0	46.5	-0.02673	0.80276	30.64
	0.01	0.03995	0	51.2	-0.03409	0.99098	34.47
Average		0.038	0.01287	14.5	-0.01600	0.97558	4.62

Table B.9: In-sample and out-of-sample enhanced indexation results, $\tau = 0.05$

Index	Transaction	In-sample		Time	Out-of-sample		
(N, K)	cost limit γ	D^*	E^*	(secs)	Intercept	Slope	AER
Hang Seng	0.0025	0.0484	0	0.9	-0.01978	1.06866	-1.12
(31,10)	0.0050	0.04558	0	0.7	-0.01786	1.06947	2.61
	0.0075	0.04301	0	0.7	-0.01155	0.98244	5.22
	0.01	0.04081	0	0.8	-0.01091	0.94767	7.58
DAX 100	0.0025	0.05732	0.32524	2.5	-0.03562	1.11813	4.94
(85,10)	0.0050	0.06552	0.05636	2.4	-0.02883	1.04838	-1.58
	0.0075	0.03917	0	2.1	-0.02717	0.94891	-1.65
	0.01	0.03252	0	1.8	-0.01160	0.75638	-4.22
FTSE 100	0.0025	0.05126	0.11593	1.9	-0.01223	0.74775	4.07
(89,10)	0.0050	0.04923	0	1.7	-0.01348	0.86698	3.86
	0.0075	0.04527	0	2.0	-0.01197	0.83102	3.11
	0.01	0.04176	0	2.0	-0.01474	0.82007	1.59
S&P 100	0.0025	0.04745	0	2.7	-0.01888	1.02513	-0.68
(98,10)	0.0050	0.04232	0	1.9	-0.01766	0.99195	-1.90
	0.0075	0.03782	0	1.8	-0.01952	0.90826	-3.91
	0.01	0.03426	0	2.1	-0.02077	0.84411	-5.26
Nikkei 225	0.0025	0.0468	0	5.7	-0.02036	0.97802	-6.20
(225,10)	0.0050	0.04154	0	4.4	-0.02291	1.08786	-3.53
	0.0075	0.03881	0	4.6	-0.01924	0.95357	-4.19
	0.01	0.03657	0	4.7	-0.01764	0.90908	-3.33
S&P 500	0.0025	0.06965	0.10813	8.8	-0.02231	1.11203	7.78
(457,40)	0.0050	0.05739	0	9.8	-0.02944	0.71525	6.19
	0.0075	0.04902	0	8.8	-0.02240	1.08548	6.32
	0.01	0.04269	0	11.4	-0.02524	1.12116	7.80
Russell 2000	0.0025			infeasible			
(1318,90)	0.0050	0.08657	0	199.9	-0.02542	1.12434	11.27
	0.0075	0.07206	0	32.1	-0.01997	1.08103	11.19
	0.01	0.06273	0	38.0	-0.01672	0.89035	14.00
Russell 3000	0.0025	0.08866	0	127.0	-0.03471	1.06009	20.28
(2151,70)	0.0050	0.07268	0	83.6	-0.03892	1.19041	26.44
	0.0075	0.06128	0	58.0	-0.03774	0.85580	33.28
	0.01	0.05224	0	50.5	-0.06015	0.70276	43.89
Average		0.05163	0.01954	21.8	-0.02277	0.96266	5.93

Appendix C

Detailed tables of results for individual τ (second alternative approach quantile regression for index tracking and enhanced indexation)

C.1 Second alternative approach tables details

This chapter provides details of tables of results for individual τ values (such as given in Tables 4.5 and 4.6) for the second alternative approach quantile regression for index tracking and enhanced indexation, Chapter 4 Section 4.4.4.

Table C.1: In-sample and out-of-sample enhanced indexation results, $\tau = 0.45$

Index	Transaction	In-sample		Time	Out-of-sample		
(N, K)	cost limit γ	D^*	E^*	(secs)	Intercept	Slope	AER
Hang Seng	0.0025	0.00339	0	0.6	-0.00532	0.99665	-2.90
(31,10)	0.0050	0.00235	0	0.5	-0.00381	0.99220	-6.59
	0.0075	0.00167	0	0.6	-0.00331	0.99258	-5.65
	0.01	0.00111	0	0.6	-0.00279	0.99404	-5.95
DAX 100	0.0025	0.00287	0.2658	1.7	-0.00054	1.02493	9.84
(85,10)	0.0050	0.00304	0.12746	1.5	0.00033	1.10103	5.67
	0.0075	0.00306	0.00555	1.6	-0.00150	0.96082	-2.24
	0.01	0.00074	0	1.5	0.00054	1.23131	13.59
FTSE 100	0.0025	0.00083	0.06543	1.9	0.00035	0.86437	4.46
(89,10)	0.0050	0.00032	0	1.7	0.00102	0.88118	5.69
	0.0075	0	0	1.6	0.00246	0.87509	7.13
	0.01	0	0	1.8	0.00246	0.87509	7.13
S&P 100	0.0025	0.00074	0	2.0	-0.00112	1.16615	5.10
(98,10)	0.0050	0	0	1.8	-0.00071	1.17569	6.71
	0.0075	0	0	2.1	-0.00071	1.17569	6.71
	0.01	0	0	1.9	-0.00071	1.17569	6.71
Nikkei 225	0.0025	0.00379	0	4.5	-0.00145	1.01651	0.26
(225,10)	0.0050	0.00211	0	4.3	-0.00151	1.04497	0.10
	0.0075	0.00088	0	4.6	-0.00096	1.05243	1.71
	0.01	0.00004	0	4.1	-0.00267	1.00974	0.68
S&P 500	0.0025	0.0039	0.03298	9.4	0.00174	1.08745	8.46
(457,40)	0.0050	0.00101	0	8.6	-0.00217	0.93213	6.51
	0.0075	0	0	8.9	-0.00219	0.93691	6.38
	0.01	0	0	9.3	-0.00219	0.93691	6.38
Russell 2000	0.0025			infeasible			
(1318,90)	0.0050	0.0045	0	70.6	0.00087	1.20618	8.08
	0.0075	0.00139	0	28.2	-0.00021	1.12817	3.19
	0.01	0	0	27.9	-0.00036	0.98825	1.99
Russell 3000	0.0025	0.00619	0	53.7	-0.00071	1.16259	15.49
(2151,70)	0.0050	0.00278	0	49.0	-0.00104	1.17818	12.81
	0.0075	0.00038	0	45.2	0.00037	1.15753	11.03
	0.01	0	0	46.3	0.00124	1.17578	10.89
Average		0.00152	0.01604	12.8	-0.00079	1.04827	4.82

Table C.2: In-sample and out-of-sample enhanced indexation results, $\tau = 0.40$

Index	Transaction	In-sample		Time	Out-of-sample		
(N, K)	cost limit γ	D^*	E^*	(secs)	Intercept	Slope	AER
Hang Seng	0.0025	0.00801	0	0.7	-0.00550	0.98053	-3.32
(31,10)	0.0050	0.00714	0	0.5	-0.00523	1.00211	-6.14
	0.0075	0.00645	0	0.6	-0.00550	1.02867	-5.68
	0.01	0.00579	0	0.6	-0.00528	1.06906	-5.82
DAX 100	0.0025	0.0058	0.29239	1.6	-0.00266	1.02849	9.84
(85,10)	0.0050	0.00639	0.14265	1.9	-0.00193	1.08721	3.00
	0.0075	0.00644	0.01207	1.5	-0.00260	0.92899	-5.61
	0.01	0.00364	0	1.9	-0.00138	1.21784	13.42
FTSE 100	0.0025	0.00432	0.06727	1.8	-0.00261	0.87471	1.24
(89,10)	0.0050	0.0034	0	1.6	-0.00107	0.87355	0.71
	0.0075	0.00255	0	1.6	-0.00272	0.91936	-1.49
	0.01	0.00177	0	1.6	-0.00467	0.85553	-4.87
S&P 100	0.0025	0.00517	0	1.9	-0.00254	1.01841	3.81
(98,10)	0.0050	0.00415	0	1.7	-0.00218	1.08597	5.21
	0.0075	0.00348	0	1.9	-0.00162	1.17905	7.77
	0.01	0.00303	0	1.8	-0.00115	1.12206	9.45
Nikkei 225	0.0025	0.00622	0	4.2	-0.00276	1.03135	-2.99
(225,10)	0.0050	0.005	0	4.0	-0.00291	1.04529	-2.28
	0.0075	0.00428	0	4.1	-0.00256	1.08354	-0.72
	0.01	0.00364	0	4.2	-0.00419	1.06781	-2.58
S&P 500	0.0025	0.00858	0	8.4	-0.00078	1.21789	4.67
(457,40)	0.0050	0.00525	0	8.3	-0.00204	0.99535	3.00
	0.0075	0.00416	0	8.6	-0.00251	1.12351	0.80
	0.01	0.00322	0	8.3	-0.00532	1.29344	-1.60
Russell 2000	0.0025			infeasible			
(1318,90)	0.0050	0.01132	0	28.8	-0.00198	1.19462	7.47
	0.0075	0.00799	0	193.6	-0.00145	1.04575	4.61
	0.01	0.00632	0	41.7	-0.00327	1.04211	1.63
Russell 3000	0.0025	0.01273	0	52.8	-0.00203	1.10343	15.82
(2151,70)	0.0050	0.00933	0	58.6	-0.00201	1.02432	14.24
	0.0075	0.0069	0	43.4	-0.00165	0.98492	10.56
	0.01	0.00508	0	44.0	-0.00219	1.05542	8.91
Average		0.00573	0.01659	17.3	-0.00278	1.05098	2.68

Table C.3: In-sample and out-of-sample enhanced indexation results, $\tau = 0.35$

Index	Transaction	In-sample		Time	Out-of-sample		
(N, K)	cost limit γ	D^*	E^*	(secs)	Intercept	Slope	AER
Hang Seng	0.0025	0.01115	0	0.6	-0.00667	0.96085	-2.18
(31,10)	0.0050	0.01014	0	0.7	-0.00510	1.01054	-0.48
	0.0075	0.00921	0	0.5	-0.00427	1.02722	2.33
	0.01	0.00846	0	0.6	-0.00294	1.04302	4.08
DAX 100	0.0025	0.00863	0.28718	1.6	-0.00474	1.09767	9.84
(85,10)	0.0050	0.00902	0.13751	1.5	-0.00309	1.02789	3.00
	0.0075	0.00899	0.00078	1.6	-0.00362	0.93321	-5.61
	0.01	0.00577	0	1.7	-0.00317	1.23197	12.21
FTSE 100	0.0025	0.00884	0.07571	1.9	-0.00316	0.81863	1.47
(89,10)	0.0050	0.00759	0	1.8	-0.00319	0.89505	0.95
	0.0075	0.0063	0	1.6	-0.00374	0.99899	1.57
	0.01	0.0053	0	1.6	-0.00466	1.05636	1.67
S&P 100	0.0025	0.00933	0	2.2	-0.00352	0.92273	-0.98
(98,10)	0.0050	0.00803	0	1.8	-0.00295	0.86323	-1.26
	0.0075	0.00717	0	2.0	-0.00337	0.88473	-3.88
	0.01	0.0065	0	2.1	-0.00413	1.01475	-3.79
Nikkei 225	0.0025	0.01014	0	4.9	-0.00443	1.00957	-3.09
(225,10)	0.0050	0.00885	0	4.2	-0.00459	1.06873	-5.83
	0.0075	0.00768	0	4.3	-0.00410	1.10284	-5.08
	0.01	0.00662	0	4.1	-0.00514	1.14875	-6.20
S&P 500	0.0025	0.01662	0.06329	12.3	-0.00308	1.15695	10.01
(457,40)	0.0050	0.01299	0	8.8	-0.00261	0.90276	7.64
	0.0075	0.01058	0	9.1	-0.00235	1.15198	4.71
	0.01	0.00835	0	9.7	-0.01076	1.34091	-2.08
Russell 2000	0.0025			infeasible			
(1318,90)	0.0050	0.01824	0	60.9	-0.00380	1.14312	5.02
	0.0075	0.01438	0	34.6	-0.00433	1.02247	5.30
	0.01	0.01217	0	30.3	-0.00330	0.92831	6.37
Russell 3000	0.0025	0.01959	0	132.7	-0.00535	1.11190	16.47
(2151,70)	0.0050	0.01554	0	71.6	-0.00239	1.00475	17.46
	0.0075	0.01258	0	66.6	-0.00393	1.01429	13.88
	0.01	0.01018	0	42.7	-0.00354	1.08363	10.62
Average		0.01016	0.01821	16.8	-0.00407	1.03154	3.04

Table C.4: In-sample and out-of-sample enhanced indexation results, $\tau = 0.30$

Index	Transaction	In-sample		Time	Out-of-sample		
(N, K)	cost limit γ	D^*	E^*	(secs)	Intercept	Slope	AER
Hang Seng	0.0025	0.01577	0	0.6	-0.00765	0.97765	-3.01
(31,10)	0.0050	0.01413	0	0.7	-0.00540	1.02418	-0.23
	0.0075	0.01251	0	0.6	-0.00278	1.00043	2.55
	0.01	0.01114	0	0.6	-0.00228	1.02826	3.36
DAX 100	0.0025	0.01183	0.28878	1.7	-0.00541	1.06561	10.99
(85,10)	0.0050	0.01216	0.13757	1.6	-0.00474	1.03534	2.84
	0.0075	0.01211	0.00147	1.6	-0.00623	1.01366	-5.91
	0.01	0.00828	0	1.6	-0.00585	1.27648	12.06
FTSE 100	0.0025	0.01244	0.07722	1.7	-0.00398	0.83968	1.47
(89,10)	0.0050	0.01127	0	1.8	-0.00435	0.88925	2.27
	0.0075	0.00962	0	1.7	-0.00547	0.97103	1.41
	0.01	0.00842	0	1.9	-0.00880	1.13570	1.76
S&P 100	0.0025	0.01281	0	2.2	-0.00487	0.88009	0.63
(98,10)	0.0050	0.01142	0	1.9	-0.00529	0.97864	-1.54
	0.0075	0.01038	0	1.7	-0.00532	1.12619	2.98
	0.01	0.00944	0	1.8	-0.00445	1.11890	4.59
Nikkei 225	0.0025	0.01525	0.018	4.2	-0.00455	1.01275	-0.85
(225,10)	0.0050	0.01321	0	4.0	-0.00554	1.11033	-5.17
	0.0075	0.01176	0	4.1	-0.00636	1.13923	-9.61
	0.01	0.01034	0	4.0	-0.00704	1.12395	-10.46
S&P 500	0.0025	0.02148	0.02945	9.3	-0.00407	1.16601	10.01
(457,40)	0.0050	0.0171	0	8.7	-0.00636	0.89036	7.54
	0.0075	0.01484	0	8.8	-0.00665	1.06477	4.64
	0.01	0.01264	0	8.5	-0.01356	1.32129	-0.35
Russell 2000	0.0025			infeasible			
(1318,90)	0.0050	0.02519	0	42.5	-0.00564	1.11580	7.68
	0.0075	0.02084	0	65.3	-0.00616	1.02493	7.99
	0.01	0.01825	0	28.7	-0.00530	0.91046	8.06
Russell 3000	0.0025	0.02646	0	95.1	-0.00712	1.11806	16.66
(2151,70)	0.0050	0.02167	0	42.9	-0.00441	0.99362	17.38
	0.0075	0.01814	0	43.7	-0.00524	0.98521	14.51
	0.01	0.01518	0	42.4	-0.00448	1.07915	10.63
Average		0.01439	0.01782	14.1	-0.00566	1.04571	3.71

Table C.5: In-sample and out-of-sample enhanced indexation results, $\tau = 0.25$

Index	Transaction	In-sample		Time	Out-of-sample		
(N, K)	cost limit γ	D^*	E^*	(secs)	Intercept	Slope	AER
Hang Seng	0.0025	0.02132	0	0.6	-0.01012	0.96928	-5.31
(31,10)	0.0050	0.01819	0	0.6	-0.00700	1.03321	-2.04
	0.0075	0.01573	0	0.5	-0.00354	1.01322	2.01
	0.01	0.01426	0	0.6	-0.00339	1.02417	3.04
DAX 100	0.0025	0.01529	0.25945	1.7	-0.01034	1.14765	9.84
(85,10)	0.0050	0.01552	0.11466	1.6	-0.00763	1.11945	3.00
	0.0075	0.01461	0	1.6	-0.00715	0.98518	-2.18
	0.01	0.01099	0	1.5	-0.00829	1.23030	11.57
FTSE 100	0.0025	0.01738	0.07411	1.7	-0.00483	0.83656	1.47
(89,10)	0.0050	0.01661	0	1.7	-0.00738	0.95413	5.25
	0.0075	0.01503	0	1.7	-0.00582	1.05962	7.83
	0.01	0.01383	0	1.6	-0.00758	1.10041	9.00
S&P 100	0.0025	0.01693	0	2.0	-0.00561	0.89756	-3.75
(98,10)	0.0050	0.01479	0	1.8	-0.00550	0.95445	-4.62
	0.0075	0.01318	0	1.8	-0.00690	0.94036	-5.97
	0.01	0.01199	0	1.9	-0.00748	0.96940	-8.38
Nikkei 225	0.0025	0.01809	0.02526	4.6	-0.00595	1.02407	-1.79
(225,10)	0.0050	0.01576	0	4.0	-0.00904	1.06141	-11.79
	0.0075	0.01408	0	4.1	-0.00905	1.07110	-9.72
	0.01	0.0125	0	4.1	-0.00862	1.13227	-9.26
S&P 500	0.0025	0.02831	0.01265	11.8	-0.01166	1.33331	4.81
(457,40)	0.0050	0.02269	0	8.4	-0.00725	0.96894	6.23
	0.0075	0.01965	0	9.0	-0.00690	0.97747	5.77
	0.01	0.01692	0	8.6	-0.00875	0.99284	3.37
Russell 2000	0.0025			infeasible			
(1318,90)	0.0050	0.03255	0	64.2	-0.00662	1.09111	7.41
	0.0075	0.02755	0	28.4	-0.00662	0.96906	9.04
	0.01	0.02453	0	35.9	-0.00636	0.86104	9.54
Russell 3000	0.0025	0.03376	0	131.3	-0.00943	1.11318	16.51
(2151,70)	0.0050	0.02792	0	43.2	-0.00798	0.94822	17.99
	0.0075	0.02367	0	43.7	-0.00647	0.94424	16.33
	0.01	0.02017	0	40.7	-0.00652	0.83142	12.06
Average		0.01883	0.01568	15.0	-0.00728	1.01789	3.14

Table C.6: In-sample and out-of-sample enhanced indexation results, $\tau = 0.20$

Index	Transaction	In-sample		Time	Out-of-sample		
(N, K)	cost limit γ	D^*	E^*	(secs)	Intercept	Slope	AER
Hang Seng	0.0025	0.02698	0.0008	0.6	-0.01304	0.94377	-5.47
(31,10)	0.0050	0.02333	0	0.8	-0.01053	1.01067	-5.75
	0.0075	0.02096	0	0.7	-0.00589	1.04628	3.76
	0.01	0.01943	0	0.6	-0.00691	1.06389	6.61
DAX 100	0.0025	0.0195	0.26665	2.2	-0.01286	1.10790	9.84
(85,10)	0.0050	0.01976	0.10209	2.0	-0.01146	1.14849	3.00
	0.0075	0.0173	0	1.9	-0.01211	1.23226	3.27
	0.01	0.01437	0	1.6	-0.01060	1.29997	12.78
FTSE 100	0.0025	0.02203	0.02556	1.8	-0.00609	0.84322	1.47
(89,10)	0.0050	0.01999	0	2.0	-0.00756	0.93975	2.34
	0.0075	0.01881	0	2.0	-0.00910	0.90083	5.41
	0.01	0.01793	0	1.8	-0.00811	0.85577	5.24
S&P 100	0.0025	0.02128	0	2.0	-0.00821	0.90168	-1.76
(98,10)	0.0050	0.0186	0	2.0	-0.00848	1.03059	-3.00
	0.0075	0.01688	0	1.7	-0.00808	0.97843	-4.96
	0.01	0.01535	0	1.7	-0.00764	0.90459	-9.59
Nikkei 225	0.0025	0.0222	0.02936	5.1	-0.00703	0.99546	-1.79
(225,10)	0.0050	0.02053	0	4.4	-0.00769	1.06147	-3.97
	0.0075	0.01901	0	4.2	-0.00946	1.08858	-5.42
	0.01	0.01763	0	4.5	-0.01269	0.96228	-3.52
S&P 500	0.0025	0.03488	0.061	14.4	-0.00754	1.17151	9.69
(457,40)	0.0050	0.029	0	10.3	-0.00913	0.96977	6.94
	0.0075	0.02501	0	10.0	-0.01015	1.04751	6.95
	0.01	0.02147	0	10.5	-0.01163	1.10464	6.87
Russell 2000	0.0025			infeasible			
(1318,90)	0.0050	0.04164	0	69.0	-0.00886	1.08912	8.01
	0.0075	0.03551	0	72.7	-0.00773	0.84193	10.60
	0.01	0.03141	0	39.1	-0.00636	0.77080	9.85
Russell 3000	0.0025	0.04213	0	111.3	-0.01215	1.13534	19.29
(2151,70)	0.0050	0.03542	0	45.0	-0.00996	0.96581	23.81
	0.0075	0.03022	0	44.0	-0.00841	0.92803	23.25
	0.01	0.02585	0	44.8	-0.00739	0.88077	18.66
Average		0.02401	0.01566	16.6	-0.00912	1.00713	4.92

Table C.7: In-sample and out-of-sample enhanced indexation results, $\tau = 0.15$

Index	Transaction	In-sample		Time	Out-of-sample		
(N, K)	cost limit γ	D^*	E^*	(secs)	Intercept	Slope	AER
Hang Seng	0.0025	0.03284	0.00774	0.6	-0.01682	0.99701	-5.47
(31,10)	0.0050	0.02857	0	0.7	-0.00903	1.03040	-1.56
	0.0075	0.02569	0	0.6	-0.00652	1.01514	0.63
	0.01	0.02362	0	0.6	-0.00574	1.01532	3.69
DAX 100	0.0025	0.02873	0.25253	2.2	-0.01491	1.06956	11.00
(85,10)	0.0050	0.03212	0.06549	1.7	-0.01705	0.92861	2.87
	0.0075	0.02134	0	1.6	-0.01456	1.19519	5.40
	0.01	0.01828	0	1.6	-0.01179	1.19244	1.48
FTSE 100	0.0025	0.02854	0.03306	1.9	-0.00989	0.75842	0.35
(89,10)	0.0050	0.02507	0	1.8	-0.00912	0.91866	-0.39
	0.0075	0.0231	0	1.7	-0.00955	0.71907	-4.68
	0.01	0.02161	0	1.8	-0.01283	0.76487	-7.44
S&P 100	0.0025	0.02574	0	1.8	-0.00974	0.89669	-2.23
(98,10)	0.0050	0.02283	0	2.0	-0.01212	0.92175	-2.72
	0.0075	0.02053	0	1.9	-0.01012	0.98312	-5.29
	0.01	0.01858	0	2.0	-0.01133	0.96797	-7.35
Nikkei 225	0.0025	0.02841	0.03566	4.9	-0.01108	0.99765	-4.08
(225,10)	0.0050	0.02496	0	4.3	-0.01067	1.03675	-3.16
	0.0075	0.02352	0	4.4	-0.01074	1.01159	-4.38
	0.01	0.02216	0	4.5	-0.01494	1.03467	-9.19
S&P 500	0.0025	0.04223	0.0444	9.0	-0.01166	1.19068	8.16
(457,40)	0.0050	0.03502	0	8.9	-0.01267	0.98856	5.70
	0.0075	0.03082	0	9.0	-0.01329	1.04395	6.50
	0.01	0.02701	0	8.9	-0.01362	1.07905	7.95
Russell 2000	0.0025			infeasible			
(1318,90)	0.0050	0.05206	0	39.3	-0.01482	1.09715	8.36
	0.0075	0.04462	0	29.7	-0.01033	0.89752	11.40
	0.01	0.03962	0	27.1	-0.00990	0.78780	10.64
Russell 3000	0.0025	0.05194	0	184.4	-0.01759	1.25463	19.58
(2151,70)	0.0050	0.04363	0	43.8	-0.01614	1.10317	25.76
	0.0075	0.03727	0	44.3	-0.01846	0.91581	31.50
	0.01	0.03196	0	42.4	-0.02008	0.85628	32.54
Average		0.03008	0.01416	15.8	-0.01249	0.98934	4.37

Table C.8: In-sample and out-of-sample enhanced indexation results, $\tau = 0.10$

Index	Transaction	In-sample		Time	Out-of-sample		
(N, K)	cost limit γ	D^*	E^*	(secs)	Intercept	Slope	AER
Hang Seng	0.0025	0.03773	0	0.7	-0.01690	1.03069	-3.55
(31,10)	0.0050	0.0345	0	0.7	-0.01155	1.07424	-0.40
	0.0075	0.03164	0	0.6	-0.00972	0.97837	1.12
	0.01	0.0299	0	0.6	-0.01213	0.95739	4.61
DAX 100	0.0025	0.03774	0.20376	2.0	-0.02164	1.15915	6.86
(85,10)	0.0050	0.04435	0.01669	1.9	-0.02223	1.09512	-1.58
	0.0075	0.02676	0	1.5	-0.01440	0.92357	0.74
	0.01	0.02328	0	1.5	-0.01257	0.81881	-3.33
FTSE 100	0.0025	0.03668	0.05752	1.7	-0.01233	0.74833	0.79
(89,10)	0.0050	0.032	0	2.3	-0.01269	0.86933	4.12
	0.0075	0.03025	0	1.6	-0.01146	0.85928	2.68
	0.01	0.02872	0	1.6	-0.01252	0.89100	0.62
S&P 100	0.0025	0.03347	0	1.7	-0.01321	0.99955	-1.43
(98,10)	0.0050	0.03012	0	1.8	-0.01259	0.96345	-3.59
	0.0075	0.0269	0	1.9	-0.01367	1.00849	-8.54
	0.01	0.02443	0	1.8	-0.01344	1.00777	-10.00
Nikkei 225	0.0025	0.03545	0	5.3	-0.01338	0.97724	-3.83
(225,10)	0.0050	0.03031	0	4.2	-0.01427	1.04915	-2.82
	0.0075	0.02865	0	4.1	-0.01230	1.00604	-3.97
	0.01	0.02711	0	4.0	-0.01555	0.98084	-7.45
S&P 500	0.0025	0.05381	0.12091	12.3	-0.01559	1.21103	7.85
(457,40)	0.0050	0.0446	0	9.1	-0.01929	0.90635	6.18
	0.0075	0.03838	0	8.4	-0.01550	1.02232	6.98
	0.01	0.03349	0	8.3	-0.02221	1.03848	6.96
Russell 2000	0.0025			infeasible			
(1318,90)	0.0050	0.06615	0	44.7	-0.01839	1.00371	10.99
	0.0075	0.05612	0	43.7	-0.01133	0.83181	11.95
	0.01	0.04964	0	39.3	-0.01031	0.76752	10.56
Russell 3000	0.0025	0.065	0	87.8	-0.02395	1.20231	19.46
(2151,70)	0.0050	0.05443	0	43.8	-0.01993	1.06797	26.16
	0.0075	0.04639	0	41.0	-0.02673	0.80276	30.64
	0.01	0.03995	0	45.3	-0.03409	0.99097	34.47
Average		0.038	0.01287	13.7	-0.01600	0.97558	4.62

Table C.9: In-sample and out-of-sample enhanced indexation results, $\tau = 0.05$

Index	Transaction	In-sample		Time	Out-of-sample		
(N, K)	cost limit γ	D^*	E^*	(secs)	Intercept	Slope	AER
Hang Seng	0.0025	0.0484	0	0.8	-0.01978	1.06865	-1.12
(31,10)	0.0050	0.04558	0	0.5	-0.01786	1.06948	2.61
	0.0075	0.04301	0	0.6	-0.01155	0.98244	5.22
	0.01	0.04081	0	0.6	-0.01091	0.94767	7.58
DAX 100	0.0025	0.05732	0.32524	1.8	-0.03562	1.11813	4.94
(85,10)	0.0050	0.06552	0.05636	1.8	-0.02883	1.04838	-1.58
	0.0075	0.03917	0	1.8	-0.02717	0.94890	-1.65
	0.01	0.03252	0	1.6	-0.01160	0.75638	-4.22
FTSE 100	0.0025	0.05126	0.11593	1.8	-0.01223	0.74775	4.07
(89,10)	0.0050	0.04923	0	1.7	-0.01348	0.86697	3.86
	0.0075	0.04527	0	1.7	-0.01197	0.83102	3.11
	0.01	0.04176	0	1.6	-0.01474	0.82007	1.59
S&P 100	0.0025	0.04745	0	2.2	-0.01888	1.02513	-0.68
(98,10)	0.0050	0.04232	0	1.9	-0.01766	0.99195	-1.90
	0.0075	0.03782	0	2.2	-0.01952	0.90826	-3.91
	0.01	0.03426	0	1.8	-0.02077	0.84410	-5.26
Nikkei 225	0.0025	0.0468	0	4.5	-0.02036	0.97802	-6.20
(225,10)	0.0050	0.04154	0	4.3	-0.02291	1.08787	-3.53
	0.0075	0.03881	0	4.1	-0.01924	0.95357	-4.19
	0.01	0.03657	0	4.3	-0.01764	0.90908	-3.33
S&P 500	0.0025	0.06957	0.10817	9.4	-0.02241	1.11371	7.69
(457,40)	0.0050	0.05739	0	9.5	-0.02944	0.71525	6.19
	0.0075	0.04902	0	9.2	-0.02240	1.08548	6.32
	0.01	0.04269	0	9.2	-0.02524	1.12116	7.80
Russell 2000	0.0025			infeasible			
(1318,90)	0.0050	0.08657	0	76.8	-0.02542	1.12434	11.27
	0.0075	0.07206	0	33.9	-0.01997	1.08104	11.19
	0.01	0.06273	0	36.4	-0.01672	0.89035	14.00
Russell 3000	0.0025	0.08866	0	145.3	-0.03471	1.06008	20.28
(2151,70)	0.0050	0.07268	0	62.1	-0.03892	1.19041	26.44
	0.0075	0.06128	0	45.9	-0.03774	0.85580	33.28
	0.01	0.05224	0	43.4	-0.06015	0.70276	43.89
Average		0.05162	0.01954	16.9	-0.02277	0.96272	5.93

Bibliography

Alexander, C. and A. Dimitriu (2005). On the application of cointegration analysis in enhanced indexing. *Economics Letters*, 391–396.

Angelelli, E., R. Mansini, and M. Speranza (2010). Kernel search: A general heuristic for the multi-dimensional knapsack problem. *Computers & Operations Research 37*, 2017–2026.

Angelelli, E., R. Mansini, and M. Speranza (2012). Kernel search: A new heuristic framework for portfolio selection. *Computational Optimization and Applications 51*, 345–361.

Bartlmae, K. (2009). Portfolio construction: Using bootstrapping and portfolio weight resampling for construction of diversified portfolios. *Business Intelligence and Financial Engineering*, 261–265.

Beasley, J., N. Meade, and T.-J. Chang (2003). An evolutionary heuristic for the index tracking problem. *European Journal of Operational Research 148*, 621–643.

Beasley, J. E. (1990). Or-library: distributing test problems by electronic mail. *Journal of the Operational Research Society 41*, 1069–1072ma.

Ben-Tal, A. and A. Nemirovski (1998). Robust convex optimization. *Mathematics of Operations Research 23*, 769–805.

Canakgoz, N. and J. Beasley (2009). Mixed-integer programming approaches for index tracking and enhanced indexation. *European Journal of Operational Research 196*, 384–399.

Chen, A. S., Y. C. Chu, and M. T. Leung (2012). The performance of enhanced-return index funds: evidence from bootstrap analysis. *Quantitative Finance 12*(3), 383–395.

Chen, C. and R. Kwon (2012). Robust portfolio selection for index tracking. *Computers & Operations Research 39*, 829–837.

Clements, K., H. Izan, and Y. Lan (2013). Volatility and stock price indexes. *Applied Economics 45*, 3255–3262.

Cox, J. C. and C. F. Huang (1991). A variational problem arising in financial economics. *Journal of Mathematical Economics 20*(5), 465–487.

Cvitani, J. and I. Karatzas (1992). Convex duality in constrained portfolio optimization. *The Annals of Applied Probability*, 767–818.

Damodaran, A. (2012). *Investment Philosophies.* Wiley Finance. New York.

Davison, A. C. and D. V. Hinkley (1997). *Bootstrap methods and their application*, Volume 1. Cambridge university press.

di Tollo, G. and D. Maringer (2009). Metaheuristics for the index tracking problem. *Metaheuristics in the Service Industry, Lecture Notes in Economics and Mathematical Systems 624*, 127–154.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The annals of Statistics 7*, 1–26.

Efron, B. and R. Tibshirani (1993). *An introduction to the bootstrap*, Volume 57. Chapman & Hall/CRC.

Fabian, C. I., G. Mitra, D. Roman, and V. Zverovich (2011). An enhanced model for portfolio choice with ssd criteria: a constructive approach. *Quantitative Finance 11*, 1525–1534.

Garcia, F., F. Guijarro, and I. Moya (2011). The curvature of the tracking frontier: A new criterion for the partial index tracking problem. *Mathematical and Computer Modelling 54*, 1781–1784.

Ghandar, A., Z. Michalewicz, R. Zurbruegg, and C. Cheong (2010). Index tracking fund enhancement using evolving multi-criteria fuzzy decision models. *IEEE Congress On Evolutionary Computation 345*, 1–8.

Guastaroba, G. and M. Speranza (2012). Kernel search: An application to the index tracking problem. *European Journal of Operational Research 217*, 54–68.

Hao, J. and D. Q. Naiman (2007). Quantile regression. *Sage Publications, London, UK*.

Harrison, J. M. and D. M. Kreps (1979). Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory 20*(3), 381–408.

Harrison, J. M. and S. R. Pliska (1981). Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and their Applications 11*(3), 215–260.

Hatemi, J. A. and E. Roca (2006). A re-examination of international portfolio diversification based on evidence from leveraged bootstrap methods. *Economic Modelling 23*(6), 993–1007.

IBM ILOG Cplex Solver (2011). Technical Report 12.0, Last accessed April 21th 2011.

IBM ILOG Cplex Solver (2012). Technical Report 12.0, Last accessed September 29th 2012.

Jeurissen, R. and J. van den Berg (2005). Index tracking using a hybrid genetic algorithm. pp. 6–pp.

Jeurissen, R. and J. van den Berg (2008). Optimized index tracking using a hybrid genetic algorithm. *IEEE Congress on Evolutionary Computation 345*, 2327–2334.

Karatzas, I. (1989). Optimization problems in the theory of continuous trading. *SIAM Journal on Control and Optimization 27*(6), 1221–1259.

Koenker, R. and G. Bassett (1978). Regression quantile. *Econometrica 46*, 33–50.

Koenker, R. and K. F. Hallock (2001). Quantile regression. *Journal of Economic Perspectives 15*, 143–156.

Koenker, R. (2012). Package 'quantreg'. Technical Report Last accessed April 21th 2012.

Konno, H. and H. Yamazaki (1991). Mean absolute portfolio optimization model and its application to tokyo stock market. *Management Science 37*(5), 519–531.

Kopa, M. (2012). Robustness and bootstrap approaches to second-order stochastic dominance (ssd) portfolio efficiency testing. pp. 443–448.

Kosowski, R., A. Timmermann, R. Wermers, and H. White (2006). Can mutual fund stars really pick stocks? new evidence from a bootstrap analysis. *The Journal of Finance 61*(6), 2551–2595.

Krink, T., S. Mittnik, and S. Paterlini (2009). Differential evolution and combinatorial search for constrained index-tracking. *Annals of Operations Research 172*, 153–176.

Lejeune, M. A. (2012). Game theoretical approach for reliable enhanced indexation. *Decision Analysis 9*, 146–155.

Lejeune, M. A. and G. Samatli-Pac (2010). Construction of risk-averse enhanced index funds. *Stochastic Programming E-Print Series, http://edoc.hu-berlin.de/docviews/abstract.php?id=37366*.

Li, Q., L. Sun, and L. Bao (2011). Enhanced index tracking based on multi-objective immune algorithm. *Expert Systems with Applications 38*, 6101–6106.

Liang, Y., F. C. Myer, and J. R. Webb (1996). The bootstrap efficient frontier for mixed-asset portfolios. *Real Estate Economics 24*(2), 247–256.

Lintner, J. (1965). The valuation of risk assests and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics 47*, 13–37.

Liu, Y. J., W. G. Zhang, and W. J. Xu (2012). Fuzzy multi-period portfolio selection optimization models using multiple criteria. *European Journal of Operational Research 48*, 3042–3053.

Magill, M. J. and G. M. Constantinides (1976). Portfolio selection with transaction costs. *Journal of Economic Theory 13*(2), 245–263.

Maringer, D. (2008). Constrained index tracking under loss aversion using differential evolution. *Studies in Computational Intelligence (SCI) 100*, 7–24.

Maringer, D. and O. Oyewumi (2007). Index tracking with constrained portfolios. *Intelligent Systems in Accounting, Finance and Management 15*, 57–71.

Markowitz, H. (1952). Portfolio selection. *Journal of Finance 7*, 77–91.

Meade, N. and J. E. Beasley (2011). Detection of momentum effects using an index out-performance strategy. *Quantitative Finance 11*, 313–326.

Merton, R. C. (1971). *Optimum consumption and portfolio rules in a continuous-time model.*

Metaxiotis, K. and K. Liagkouras (2012). Multiobjective evolutionary algorithms for portfolio management: A comprehensive literature review. *Expert Systems with Applications 39*, 11685–11698.

Mezali, H. and J. E. Beasley (2011). Index tracking with fixed and variable transaction costs. *Optimization Letters 4*.

Mezali, H. and J. E. Beasley (2012). Quantile regression for index tracking and enhanced indexation. *Journal of Operational Research Society 64*, 1–17.

Murray, W. and H. Shek (2012). A local relaxation method for the cardinality constrained portfolio optimization problem. *Computational Optimization and Applications 53*, 681–709.

Rohweder, H. C. (1998). Implementing stock selection ideas: does tracking error optimization do any good? *The Journal of Portfolio Management 24*(3), 49–59.

Roll, R. (1992). A mean variance analysis of tracking error. *Journal of Portfolio Management 18*, 13–22.

Roman, D., G. Mitra, and V. Zviarovich (2011). Enhanced indexation based on second-order stochastic dominance. *Social Science Research Network*.

Ruiz-Torrubiano, R. and A. Suarez (2009). A hybrid optimization approach to index tracking. *Annals of Operations Research 166*, 57–71.

Shao, J. and D. Tu (1995). The jackknife and bootstrap.

Sharpe, W. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance 19*, 425–442.

Storn, R. and K. Price (1997). Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization 11*, 341–359.

Tabata, Y. and E. Takeda (1995). Bicriteria optimization problem of designing an index fund. *Journal of the Operational Research Society 46*, 1023–1032.

Thomaidis, N. (2012). A soft computing approach to enhanced indexation. *Natural Computing in Computational Finance 380*, 61–77.

Thomaidis, N. (2013). On the application of cointegration analysis in enhanced indexing. *Applied Economics Letters 20:4*, 391–396.

Van Montfort, K., E. Visser, and L. F. van Draat (2008). Index tracking by means of optimized sampling. *The Journal of Portfolio Management 34*, 143–151.

Wang, M., C. Xu, F. Xu, and H. Xue (2012). A mixed 0–1 lp for index tracking problem with cvar risk constraints. *Annals of Operations Research 196*, 591–609.

Xu, G. L. and S. E. Shreve (1992). A duality method for optimal consumption and investment under short-selling prohibition. *The Annals of Applied Probability 2*(1), 87–112.

Yu, K., Z. Lu, and J. Stander (2003). Quantile regression: applications and current research areas. *The Statistician 52*, 331–350.

Yu, K. and R. A. Moyeed (2001). Bayesian quantile regression. *Statistics & Probability Letters 54*(4), 437–447.

Yu, K. and J. Zhang (2005). A three-parameter asymmetric laplace distribution and its extension. *Communications in StatisticsTheory and Methods 34*(9-10), 1867–1879.

Yu, L., S. Zhang, and X. Y. Zhou (2006). A downside risk analysis based on financial index tracking models. pp. 213–236.

Zhang, W. G., Y. J. Liu, and W. J. Xu (2012). A possibility mean-semivariance-entropy model for multi-period portfolio selection with transaction costs. *European Journal of Operational Research 222*, 341–349.