Portfolio management under transaction costs:
Model development and Swedish evidence
Portfolio management under transaction costs:  
Model development and Swedish evidence

Rickard Olsson  
Master of Science

Umeå Studies in Business Administration No. 56  
Umeå School of Business  
Umeå University
Abstract
Portfolio performance evaluations indicate that managed stock portfolios on average underperform relevant benchmarks. Transaction costs arise inevitably when stocks are bought and sold, but the majority of the research on portfolio management does not consider such costs, let alone transaction costs including price impact costs. The conjecture of the thesis is that transaction cost control improves portfolio performance. The research questions addressed are: Do transaction costs matter in portfolio management? and Could transaction cost control improve portfolio performance? The questions are studied within the context of mean-variance (MV) and index fund management. The treatment of transaction costs includes price impact costs and is throughout based on the premises that the trading is uninformed, immediate, and conducted in an open electronic limit order book system. These premises characterize a considerable amount of all trading in stocks.

First, cross-sectional models of price impact costs for Swedish stocks are developed using limit order book information in a novel fashion. Theoretical analysis shows that the price impact cost function of order volume in a limit order book with discrete prices is increasing and piecewise concave. The estimated price impact cost functions are negatively related to market capitalization and historical trading activity, while positively related to order size and stock return volatility. Total transaction costs are obtained by adding the relevant commission rate to the price impact cost.

Second, the importance of transaction costs and transaction cost control is examined within MV portfolio management. I extend the standard MV model by formulating a quadratic program for MV portfolio revisions under transaction costs including price impact costs. The extended portfolio model is integrated with the empirical transaction cost models developed. The integrated model is applied to revise portfolios with different net asset values and across a wide range of risk attitudes. The initial (unrevised) portfolios are capitalization-weighted and contain all Swedish stocks with sufficient data. The standard MV model, which neglects transaction costs, realizes non-trivial certainty equivalent losses relative to the extended model, which, in addition, exhibits lower turnover, higher diversification, and lower transaction costs incurred. The evidence suggests that transaction cost control improves performance in MV revisions, and that price impact costs are worthwhile to consider.

Third, the research questions are studied within index fund management. I formulate two index fund revision models under transaction costs including price impact costs. Each model is integrated with the empirical transaction cost models. Transaction costs including price impact costs, cash flows, and corporate actions are incorporated in the empirical tests, which use ten years of daily data. In the tests, the two index fund revision models and several alternative approaches, including full replication, are applied to track a Swedish capitalization-weighted stock index. Instead of using an extant index, an index is independently calculated according to a consistent methodology, mimicking that of the most used index in the Nordic region, the OMX(S30). The alternative approaches are tested under a number of variations including different tracking error measures and different types and degrees of transaction cost control. Index funds implemented by the index fund revision models under transaction cost control dominate, in all dimensions of tracking performance considered, their counterparts implemented without transaction cost control as well as the funds implemented by full replication. Price impact costs constitute the majority of the transaction costs incurred. Additional results indicate that some common tracking error measures perform similar and that the technique to control transaction cost by constructing an index fund from a pre-defined subset of the most liquid index stocks is not efficient.

The overall conclusion of the thesis is that transaction costs matter, that transaction cost control improves portfolio performance, and that price impact costs are important to consider.
To the memory of my father
Acknowledgments

I would like to thank without implicating in any way the following persons and organizations for aid of various kinds:

- Professor Lars Hassel - main supervisor
- Doctor Claes-Göran Larsson - supervisor
- Professor Johan Knif - reviewer of the manuscript at the pre-defense seminar
- Doctor Jörgen Hellström - reviewer of the manuscript at the pre-defense seminar
- Associate Professor Xavier de Luna
- Professor Kurt Brännäs
- Associate Professor Barbara Cornelius
- Professor Maria Bengtsson
- Professor Emeritus Nils Hakansson
- Doctor Mattias Hamberg
- Doctor Henrik Nilsson, Anders Isaksson, Anneli Linde, and Tobias Svanström
- Christian Fagerlund, Gabriel Fogelsjöö, Mattias Lundgren, Mattias Olofsson, and Joakim Oscarsson
- Jan Wallanders och Tom Hedelius Stiftelse, Sparbanksstiftelsen Norrland, J.C. Kempes Minnes Stipendiefond och J.C. Kempes Minnes Akademiska Fond, and Borgerskapets i Umeå Forskningsstiftelse

And to family and friends, to whom I at times during this work have not paid the attention they deserve, I sincerely apologize.

Umeå, October 2005
List of figures

Figure 2.1 Average price as a function of order quantity, $\bar{p}(q)$ ...................................34
Figure 2.2 Price impact cost as a function of order quantity, $\pi(q)$ ..................................35
Figure 4.1 The OMX index, the return index (RI) and the price index (PI) ..................149
Figure 4.2 Price index level as a proportion of the OMX ............................................150
Figure 4.3 Tracking error of RI and PI, and the ratio of their levels .........................152
Figure 4.4. Index funds with and without transaction cost control as fractions of the index level ..............................................................................................................169

List of tables

Table 2.1 Tick sizes, stock price intervals and price impacts .........................................28
Table 2.2 A limit order book snapshot, average price and price impact .......................34
Table 2.3 Summary statistics across order books .........................................................56
Table 2.4 Price impact costs by market capitalization and order size ......................60
Table 2.5 Price impact regressions .............................................................................61
Table 2.6 Summary statistics for estimated regressions in the simulation analysis ....64
Table 2.7 Statistics on forecasted percentage price impact cost based on the holdout sample ..............................................................................................................65
Table 2.8 Forecast errors of the percentage price impact cost in the holdout sample ..66
Table 3.1 Descriptive statistics for the capitalization-weighted initial portfolio ..........87
Table 3.2 Larger-NAV portfolio revisions without maximum weight constraints .......91
Table 3.3 Larger-NAV portfolio revisions with maximum weight constraints ...........92
Table 3.4 Smaller-NAV portfolio revisions without maximum weight constraints....93
Table 3.5 Smaller-NAV portfolio revisions with maximum weight constraints .........94
Table 4.1 Tracking performance and transaction cost statistics .................................148
Table 4.2 Descriptive statistics based on daily returns of the price index (PI), the total return index (RI), and the OMX .................................................................150
Table 4.3 Tracking of the OMX with the price index and tracking of the price index with the return index .....................................................................................151
Table 4.4 Full replication with no dividends, dividends paid out at close after 1 and 8 days, respectively .........................................................................................153
Table 4.5 Full replication under conditional updating with transaction costs and dividends ..............................................................................................................157
Table 4.6 Opt1: Minimizing a tracking error measure and transaction costs using a trade-off parameter under transaction costs and dividends ..........................161
Table 4.7 Opt2: Minimizing transaction costs while keeping a tracking error measure less than a pre-specified limit ..............................................................164
Table 4.8 Opt2: Meeting a tracking error limit without transaction cost control .......168
Table 4.9 Opt1: Outdated versus updated historical index return series ...............173
Table 4.10 Opt2: SDTE versus MATE meeting a TElimit...........................................175
Chapter 1

Introduction

1.1 Background and motivation

The problem of managing a portfolio of stocks and other securities is an important one. Most individuals, and many organizations, are faced with the problem. The ultimate result of the portfolio management process is often far-reaching. For an individual, the choice of a portfolio of assets for retirement savings is one example. The issue is taking on additional force as many Western welfare states move towards an increasing reliance on individualized pension solutions such as defined contribution systems, in which pension benefits are directly linked to contributions and investment returns. This development is driven by the global demographic change that is about to produce the largest group of pensioners in history (Mitchell and Smetters 2003). In the thriving fund management industry, realized portfolio performance is probably the primary means of competition. Institutions, such as foundations, derive substantial portions of their income from portfolio returns.

A matter in debate here in Sweden is the recently introduced national system for supplementary retirement savings, the Premium Pension system. In this system, every Swedish wage-earner is allotted a personal savings account. Each year, an amount equal to 2.5% of the annual salary two years ago is deposited in the account. It is for the account-holder to decide on how the capital in the account should be allocated. From nearly 700 mutual funds, the individual can choose a minimum of one and a maximum of five funds. Most fund categories, from government bond funds to international hedge funds, are represented. The account-holder determines the proportion of total capital to be allocated to each one of the chosen funds. All the capital in the account has to be invested in funds, that is, uninvested cash is not
Introduction
Chapter 1

allowed. Investment decisions can be effectuated electronically through the internet whenever the individual so desires. Finally, premium pension payments can start from the age of 61.

The attention of the media and the public has primarily concerned the poor performance of the premium pension accounts. By year-end 2004, more than four years after the system was launched, the value of the average premium pension account was 8% lower than the sum of all contributions. Only 31% of the accounts had had a positive return on the capital invested. As of July 2005, the average account had, however, increased 6% relative to total contributions, and almost 80% of the accounts were in surplus, thanks to strong stock markets and favorable currency movements. Although the amounts contributed each year to the premium pension account may appear small, it has been estimated by the Premium Pension Authority that the premium pension could constitute as much as 1/3 of an individual’s total national pension. The management of the premium pension account, a portfolio problem, is thus a matter of considerable importance for most Swedes.

What determines the performance of the premium pension accounts? The portfolio decisions made by the individuals are of course one factor. Another factor is the performance of the mutual funds among which the pension savers can choose. The performance of the mutual funds, in turn, is a function of the portfolio decisions of their managers. That the mutual fund portfolios are efficiently managed is hence essential for future premium pensions. Past performance is probably the strongest sales argument for a fund, so, for a fund management company, efficient portfolio management is also vital.

Empirical evidence on fund performance

What is the empirical evidence on fund management and fund performance?

Ever since the seminal work of Jensen (1968), mutual fund performance evaluations continue to evidence that most mutual funds underperform relevant benchmarks on a net return, risk-adjusted basis (see, e.g., Elton et al. 1993; Malkiel 1995; Gruber 1996; Carhart 1997; Chalmers et al. 2001). Regarding Swedish mutual

---

1 Generally speaking, this portfolio problem is technically quite demanding to solve optimally. One reason is the restriction on the number of funds to choose among.
funds, Dahlquist et al. (2000, p. 410) find that performances are level with or below benchmarks:

“This conditional evaluation of the funds suggests that the performance of regular equity funds has been neutral. The equity funds in the public savings program (with certain tax advantages) appear to have had a negative relative performance (before tax).”

As a result of the research on fund performance,

“a conventional belief has developed in the academic community that mutual fund managers as a group have no special ability to identify and profit from mispriced securities.” Edelen (1999, p. 440)

The above studies focus on net returns, that is, the actual bottom-line returns that mutual funds deliver to investors. These are net of transaction costs, fees, taxes, and fund expenses. The major components of transaction costs are brokerage commissions and price impact costs (Keim and Madhavan 1997). The price impact cost of a given transaction may be defined as any temporary adverse price movement induced by the transaction (see, e.g., Kraus and Stoll 1972).

Another line of research, pursued by, e.g., Grinblatt and Titman (1989), Daniel et al. (1998), Chen et al. (2000), Wermers (2000), and Pinnuck (2003), evaluates performance in terms of the performance of the individual stocks held and traded in mutual fund portfolios. The empirical performance as measured on this effectively gross return basis - i.e., before transaction costs, taxes, and other expenses, and adjusted for risk - is summarized by Chen et al. (2000, p. 367) as “suggestive of the funds possessing superior stock selection skills”. That fund managers may realize abnormal returns on their holdings or trades does not necessarily imply that superior net returns are delivered to investors. Wermers (2000, p. 1690) for instance conclude that:

“Thus, considering only their stock holdings, mutual fund managers hold stocks that beat the market portfolio by almost enough to cover their expenses and transactions costs.”

Likewise, Chalmers et al. (2001, p. 16) conclude, “our evidence suggests that fund managers fail to recover any of their trading expenses.” Moreover, the empirical findings of Edelen (1999) indicate that a part of the underperformance of mutual funds is explained by the costs of their uninformed, liquidity-motivated trading induced by investor flows. Edelen (1999) and Wermers (2000) point out that the observed
relations between gross returns, expenses, transaction costs, and uninformed trading are consistent with the notion of market efficiency laid out in Grossman and Stiglitz (1980).

To summarize, the majority of mutual funds underperform relevant benchmarks on a net return, risk-adjusted basis. Mutual fund managers seem to have some ability to generate superior gross returns due to stock selection skills, but this is of little comfort to mutual fund investors as these superior gross returns are dwarfed by transaction costs, expenses, and fees.

**Implications for investors, fund managers and further research**

What are the implications of this research for investors, fund managers, and further research?

With regard to investors, the general conclusion in the literature is that they should invest in index funds rather than in actively managed funds:

> “These studies do not provide a promising picture of active mutual fund management—instead, the studies conclude that investors are better off, on average, buying a low-expense index fund.” Wermers (2000, p. 1656)

Most funds are actively managed. They aim actively at beating pre-specified benchmark indexes. A large part of the remainder of funds are index funds. These funds are said to be passively managed, because they aim at tracking or replicating, rather than surpassing, the performance of their benchmarks.

The studies referred to above concerned the performance of mutual funds relative to costless benchmark indexes rather than to index funds. Gruber (1996), Wermers (2000), and Frino and Gallagher (2001), among others, study the performance of actively managed funds and index funds, and find that the majority of actively managed funds are outperformed by index funds on a risk-adjusted basis; Kyhlberg and Wallander (1987) observe that two Swedish index funds performed at least as well as other Swedish equity funds during the period 1978-1987.

Performance evaluations of index funds, however, indicate that funds tracking the same index exhibit varying tracking performance (Sinquefield 1993; Gruber 1996, Frino and Gallagher 2001, 2002; Frino et al. 2004; Blume and Edelen 2004; Elton et al. 2004). One explanation for the varying tracking performances is that different approaches (or methods) are employed to select and revise index funds. Respects in
which the approaches may differ include the type and degree of transaction cost control used. Sinquefield (1993) and Keim (1999) provide indirect evidence that transaction cost control improves performance for index funds. That the management of index funds is as efficient as possible is of obvious importance to both fund managers and investors.

The amount of research investigating the importance of transaction costs and transaction cost control in index fund management does, however, appear to be limited:

“The majority of the work relating to index tracking presented in the literature does not consider transaction costs and only considers the problem of creating an initial tracking portfolio from cash.” Beasley et al. (2003, p. 623)

In fact, it seems as if more research on index funds in general is warranted, considering the following quotations:

“The method of selecting and revising passive portfolios [index funds] has received virtually no attention in the literature,” Rudd (1980, p. 57)

“Index funds have grown significantly over the past decade, however empirical research concerning these passive investment offerings is surprisingly scarce in the literature.” Frino and Gallagher (2001, p. 44)

“Finally we would comment that index tracking is an important problem that, in our view, has received insufficient attention in the literature.” Beasley et al. (2003, p. 641)

With regard to fund managers as a collective, it seems that they may have a certain ability to generate profitable investment ideas, but as Arnott and Wagner (1990, p. 79) point out:

“An investment idea, no matter how rich or insightful, is merely an idea until it is implemented. Effective implementation is not automatic. A poor implementation process can overwhelm the value of the idea.”

Trading is the implementation of investment ideas, and trading or transaction costs are the costs of implementation. So, in addition to cutting expenses and improving security valuation, a possible means for funds to enhance returns is to improve implementation, that is, to exercise better control of transaction costs as stocks are bought and sold. Quite radically, Keim and Madhavan (1998, p. 65) even suggest that:

“Given the voluminous evidence on the unprofitability of active portfolio management, we suggest that resources would be put to better use in attempting to understand and reduce trade costs than trying to exploit scarce market inefficiencies.”
All else equal, reduced transaction costs imply increased gross returns. This would allow the fund company to inflate expenses and charge higher fees, and if less than the total cost reduction were used for such purposes, the fund would also improve its net return performance. This would benefit investors and probably also make the fund more attractive to new customers. Reducing transaction costs can thus be beneficial to both the fund company and current investors.

The probably most well known model for portfolio management is the mean-variance (MV) model of Markowitz (1952, 1959, 1987). Transaction cost control is not integrated in this model. The MV model is a cornerstone of financial theory, but its use among investment practitioners has been reported as limited (see, e.g., Michaud 1989, 1998; Konno and Yamazaki 1991; Black and Litterman 1992; Jorion 1992; Fisher and Statman 1997). Reasons offered for this “enigma” (Michaud 1989) include that the MV model generates high portfolio turnover rates, which may result in large transaction costs, and that obtained portfolios often are counter-intuitive in that they exhibit lack of diversification with extreme allocations into just a few securities (see, e.g., Michaud 1989, 1998; Best and Grauer 1991a, 1991b; Green and Hollifield 1992; Fisher and Statman 1997; Jagannathan and Ma 2003). It seems interesting to examine how transaction cost control in the mean-variance model affects portfolio turnover, transaction costs, and lack of diversification. Insomuch that excessive transaction costs are caused by high portfolio turnover when transaction costs are not controlled, it appears likely that transaction costs as well as turnover will be lower when transaction costs are controlled. Extreme portfolio allocations can be produced by extreme weight changes. To the extent that such extreme weight changes would induce large transaction costs, it is probable that portfolio allocations will be less extreme with transaction cost control than without.

1.2 Research questions

It appears that transaction cost control may be a key to improved performance in index fund management as well as in mean-variance portfolio management. Based on this, the following overall research questions are formulated: Do transaction costs matter in
portfolio management? and Could transaction cost control improve portfolio performance?

An investigation of these questions within the context of a particular portfolio (management) problem would typically require the following elements: models of individual stocks’ transaction costs, a portfolio decision model that incorporates the transaction costs, empirical test procedures, and relevant data. Unavailable or non-existent elements would have to be developed.

1.3 Initial literature review

At the beginning of this research project, a literature search was conducted for models that integrate equity index fund and mean-variance portfolio management, respectively, with transaction costs including price impact. It was desired that any such model should be possible to implement in a standard quadratic programming solver, and that larger-scale problems should be possible to solve in reasonable time using the quite limited computational equipment available to the project initially. A handful of studies were found including Pogue (1970), Chen et al. (1971), Rosenberg and Rudd (1979), Rudd (1980), Perold (1984), and Adcock and Meade (1994). Of these studies, Rudd (1980) and Adcock and Meade (1994) implement and test stock index fund management models under transaction costs that are exclusive of price impact and invariant across stocks and time. The other studies develop mean-variance models under transaction costs. Many models in the above studies did appear difficult to implement, requiring non-standard solution algorithms. In addition, restrictions on the return generation process are imposed in several of the models, where factor structures are used in place of the variance-covariance return matrix. Few of the models consider explicitly price impact costs. The ones that do, Pogue (1970) and Perold (1984), use an increasing stepwise function to model price impact; the percentage price impact cost is constant in a given volume interval but increases between intervals. More elaborate models of transaction costs require several steps, and this complicates modeling and computation. In view of these difficulties, it was decided that an independent attempt should be made to formulate mean-variance and index fund models under transaction costs including price impact costs.
A parallel literature search for empirical models of transaction costs including price impact was undertaken. It is important that such a model allows the percentage transaction cost to vary with the size of the transaction as this allows explicit modeling of price impact costs. Models that failed to satisfy this criterion were discarded. Only a few models were found, among them Loeb (1981, 1991), Chan and Lakonishok (1995), and Keim and Madhavan (1997). These cross-sectional models are based on data relating to institutional trading in U.S. stocks. No transaction cost models where the percentage cost is a function of order size could be identified for Swedish stocks. Therefore, a decision was made to develop such models directly for stocks trading on the Stockholm Stock Exchange.

Besides the research questions stated above, the literature review highlighted other issues related to index fund management worthy of study. Several alternative approaches, or methods, for equity indexing may be distinguished in the literature, including full replication, optimization, regression, and stratified sampling (see, e.g., Rudd 1980; Andrews et al. 1986). It has been suggested that some approaches are better suited than others for tracking particular kinds of indexes (see, e.g., Rudd 1980; Andrews et al. 1986). The empirical evidence on the alternative approaches ability to track different kind indexes is, however, limited and does not allow straightforward comparison of the tracking performances of the alternative approaches. For instance, full replication is reported to be the most frequently used approach in practice (see, e.g., Liesching and Manchanda 1990; Frino and Gallagher 2001; Blume and Edelen 2004), but it has not been implemented and tested empirically in research.

A heuristic technique to control transaction costs in index fund management is to restrict the number of stocks in the index fund relative to the number of stocks in the index (see, e.g, Rudd 1980; Connor and Leland 1995; Larsen and Resnick 1998; Bamberg and Wagner 2000). This can be done by constructing the index fund from a pre-defined subset of the stocks in the index, where the subset contains the stocks that are deemed to have the lowest transaction costs. No empirical evidence is, however, presented on whether this use of subsets fulfills its purpose.

In most approaches to indexing, desired and realized tracking performance are expressed in terms of a so-called tracking error measure, which is a mathematical
function that measures the how closely the index fund tracks the index. Different tracking error measures are used in the literature (see, e.g., Rudd 1980; Sharpe 1992; Roll 1992; Larsen and Resnick 1998; Rudolf et al. 1999). A priori arguments for and against different tracking error measures are offered, but empirical evidence allowing direct comparison of the performances of different measures in a relevant test setting is typically not provided.

The literature search also revealed that studies that implement and test different approaches to index fund management, e.g., Rudd (1980), Meade and Salkin (1989, 1990), Adcock and Meade (1994), Larsen and Resnick (1998), and Walsh et al. (1998), do not consider many of the factors that actually present the challenges to equity indexing. Examples of such factors are investor and dividend cash flows, corporate actions (e.g., new issues and mergers and acquisitions), and transaction costs including price impact. In addition, these factors typically occur on a daily time-scale, whereas most of the studies employ data of weekly or monthly frequency. In response to this, I decided that the empirical work on index fund management preferably should be based on daily data and include cash flows, corporate actions, and transaction costs including price impact costs.

To summarize, integration of transaction cost control in portfolio management appears potentially rewarding, but the majority of the work on portfolio management problems does not consider transaction costs, let alone transaction costs including price impact (see, e.g., Leinweber 2002, 1995). There are several possible reasons for this. It may be that transaction costs are unimportant in portfolio problems in the respect that portfolios selected with and without regard to transaction costs do not differ much, and, in particular, that portfolios in the former category do not exhibit any economically significant improvements over portfolios in the latter. Moreover, incorporating transaction costs in portfolio problems generally increases modeling complexity and computational burden (see, e.g., Rudd and Rosenberg 1979; Konno and Yamazaki 1991).

Though conceptually clear, the cost of a transaction, and particularly the price impact part, is in general difficult to measure unambiguously (Arnott and Wagner 1990). Moreover, theory and empirical evidence suggest that the cost of a transaction
may depend on a number of factors including, but not limited to, order complexity, stock-specific characteristics, investment strategy, and market architecture (see, e.g., Keim and Madhavan 1998). Furthermore, the access to data necessary for accurate measurement and modeling of transaction costs is in general restrictive in that such data often are proprietary, apply to a limited number of institutional customers and stocks (Bessembinder 2003), or are prohibitively expensive to acquire (Chalmers et al. 2000).

1.4 Premises regarding trading and transaction costs
In response to the aforementioned issues related to transaction costs, the treatment of transaction costs in this thesis will be based on the premises that the trading is uninformed, demands immediacy of execution, and takes place in an electronic open limit order book market, the Stockholm Stock Exchange (SSE). This is motivated as follows.

Uninformed trading derives from risk tolerance changes and portfolio cash imbalances (Bagehot 1971; Stoll 2003), from traders “acting on information they believe has not yet been fully discounted in the market price but which in fact has” (Bagehot 1971, p. 13), and from those trading “on noise as if it were information.“ Black (1986, p. 531). The motive for informed trading is to exploit superior information about asset values or price changes (Bagehot 1971, p. 143).

Since 1989, trading in Swedish stocks take place in an electronic limit order book system. In this order-driven (or auction) market, public investors can trade directly with each other. The SSE is a pure limit order market in the sense that all traders trade on basically equal terms, and in that there are no designated intermediaries such as market markers responsible for maintaining a liquid market. Investors electronically submit orders to buy or to sell. A limit order is an order to trade a fixed quantity at a certain limit price or better. A market order is an order to trade a fixed quantity at the best available price. The limit order book for a given stock at a given time is the set of all submitted and unexecuted buy and sell orders, with volumes consolidated at each price level. The prices and the consolidated volumes in the limit order book are publicly visible. A trade executes automatically when the price of a new order matches
the price of an order in the limit order book. Orders execute in priority of price and time of submission. As these automated order-driven systems are relatively recent phenomena, they are generally less well understood than intermediated order-driven systems and so-called quote-driven systems.² Ahn et al. (2001, p. 767), for instance, establish that:

“Although many stock exchanges around the world are based on pure limit order books, very few empirical papers investigate the role of limit-order traders in an order-driven market without any designated market maker.”

Virtually all new trading systems that are being implemented are electronic order-driven systems (Brockman and Chung 2000). Their practical relevance is thus increasing.

According to market microstructure theory, the whole adverse price movement or price impact induced by an uninformed trade should be temporary (see, e.g., Kraus and Stoll 1972) and, therefore, fully recognizable as a cost. If trading is uninformed, without any permanent price effects, then the prices and volumes displayed in the limit order book provide a means for measuring in a rather uncontroversial manner the price impact cost of trades that are executed immediately at the best prices available. Specifically, the whole price impact cost function of order volume is observable for the volume displayed in the order book. That access was freely available (through Ecovision, a real-time financial information system) to the Stockholm Stock Exchange’s limit order book market system, influenced the decision to perform the empirical analysis of the importance of transaction cost control in portfolio problems using stocks traded on the Stockholm Stock Exchange, and with the stocks’ price impact costs derived from their limit order books.

The commonness of uninformed immediate trading

Uninformed trading that seeks immediacy of execution does not represent an idiosyncrasy in the markets. On the contrary, it is possible that a substantial amount of all trading is of this kind. Empirical market microstructure studies indicate that

² A quote-driven market – perhaps the antipode to an order-driven - is one in which a dealer (or market maker) takes the opposite side of every transaction. Dealers quote ask and bid prices and the number of shares they are willing to trade at these prices. Public investors cannot trade directly with each other, but must buy at the dealers ask and sell at the dealers bid. Trades are typically conducted by telephone.
uninformed trades outnumber by an order of a magnitude informed trades in both quote-driven (see, e.g., Easley et al. 1996; Nyholm 2002) and order-driven markets (see, e.g., Brown et al. 1999; Brockman and Chung 2000; Ma and Yang 2004). That immediacy is frequently demanded in the marketplace is supported by empirical evidence. Of the $83 billions worth of institutional equity transactions analyzed by Keim and Madhavan (1995, 1997), market orders, which demand immediate execution at the best prices available, constitute approximately 90% of all orders in number and value. For an order-driven market, Griffiths et al. (2000) document that of all orders submitted, around 80% in number, and 50% in terms of number of shares, sought immediate execution. Biais et al. (1995) also analyze an order-driven market and find that almost 50% of all orders are such that immediate execution is obtained.

Index fund management in view of the premises

One part of the trading that is uninformed and immediate represents implementations of solutions to the index fund problem. The trading - due to investor flows, index reconstitutions, and dividends - of index funds is per definition uninformed and their demand for immediacy appears strong. Keim and Madhavan (1995, 1997), for instance, report that in their sample market orders constitute 90% of all orders submitted by index funds. Despite that index funds generally exhibit lower portfolio turnover ratios than actively managed funds (see, e.g., Wermers 2000), the trading of index funds should still represent an important part of total trading as index fund investments have been estimated to represent around 1/3 of global assets under management (Cerulli Associates 2001 cited Asset growth grinds to a halt 2002). The index fund problem has been stated as how to replicate the performance of an index while minimizing transaction cost (see, e.g., Rudd 1980; Frino and Gallagher 2002; Konno and Wijayanayake 2001; Beasley et al. 2003). As transaction cost control is made explicit in the index fund problem, it should provide a rather ideal context in which to address the research questions of this thesis.

Mean-variance portfolio management in view of the premises

The remaining part of the uninformed immediate trading represents implementations of solutions to portfolio problems other than index fund problems. However, rather than considering the importance of transaction cost control in many different portfolio
problems - besides the index fund problem - the analysis is confined to the mean-variance problem of Markowitz (1952, 1959, 1987). One reason is the central position of the mean-variance framework in the literature. Another reason, probably related to the preceding one, is that the single-period mean-variance portfolio model can provide good approximate solutions to various other portfolio problems, featuring for example different utility functions and multi-period considerations (see, e.g., Pulley 1981, 1983; Ziemba and Kallberg 1983; Amilon 2001). The mean-variance problem is more closely linked to utility theory than is the index fund problem. The mean-variance problem hence better allows for analyzing in utility-theoretic terms the importance of transaction costs and transaction cost control in portfolio choice. This is important, for utility theory is a cornerstone of portfolio choice. Moreover, the mean-variance model is a natural starting point for developing index fund models (see, e.g., Roll 1992). Also, studying the importance of transaction cost control in the context of two different portfolio problems rather than one, broadens the scope of the thesis.

1.5 Purpose

The main purpose of this thesis is to examine the importance of transaction costs and transaction cost control in mean-variance and index fund management involving stocks traded on the Stockholm Stock Exchange. The transaction costs shall include price impact and be incurred by trading that is uninformed, immediate and executed in the electronic open limit order book system of the said exchange. Empirically, consideration will be given to portfolios of sizes intended to be representative of typical Swedish mutual funds.

A secondary purpose, relating to index fund management, is to examine the efficiency of pre-defined subsets as a means to control transaction costs, and how different tracking error measures and alternative indexing approaches affect tracking performance. In consequence, the amount of research work that will be done on index fund problem in this thesis will be greater than on the mean-variance problem. In addition to personal preference, this state of things is consistent with the premises in that index fund trading is exclusively uninformed, whereas the trading incurred in conjunction with mean-variance portfolio management is not exclusively uninformed.
1.6 Research tasks

In view of the initial literature review and the decisions presented therein, for a given portfolio problem, pursuing the main purpose would reasonably amount to completing the following research tasks:

(i) develop models of individual stocks’ transaction costs including price impact costs,

(ii) formulate (at least) one portfolio decision model that incorporates these transaction costs, and

(iii) design and conduct empirical tests that, based on (i) and (ii), seek to assess the importance of transaction costs including price impact and transaction cost control.

A fourth research task, relating to index fund management, is to

(iv) design and conduct empirical tests that address the issues of the secondary purpose.

The adopted premises regarding trading and transaction costs effectively means that research task (i) will be common to the two portfolio problems, and, consequently, that the same models of transaction costs can be used in the work on both problems. Research tasks (ii) and (iii) will both be dealt with within each of the two portfolio problems. An important implication of the decision to independently develop transaction cost models is that these have to be finalized before any real empirical work on the mean-variance and index fund problems can be done.

Research tasks (i), (iii), and (iv) require relevant data. An overview of the methods and the data that have been used in the work on these research tasks follow next. The last section of this introductory chapter outlines the remainder of the thesis.

1.7 Methods and data

The thesis work comprises theoretical/analytical model development as well as empirical research and several techniques and tools have been utilized.

Gauss, a mathematical and statistical system that features a matrix-based programming language, has been an important tool in the work on this thesis. All studies, computations, and statistical analyses have been programmed in GAUSS.
Certain steps of the data processing have been carried out in Microsoft Excel and in some special applications. Although mathematical programming is used frequently throughout the thesis, I have not programmed any solution algorithms myself, but instead relied on the algorithms that come with Gauss.

Research task (i) involves deriving the price impact cost function in a limit order book. The derivation uses calculus. Empirical price impact cost functions are estimated using ordinary least squares regression analysis (OLS). The robustness of the estimated models are examined by simulation.

Mathematical programs are developed for each portfolio problem through the course of research task (ii) and this involves matrix algebra. A quadratic program for mean-variance portfolio revision under transaction costs, which allows for price impact, is derived. The importance of transaction cost control in mean-variance portfolio revision is examined in a utility-theoretic setting by means of an empirical test based on Swedish stock and fund data. This comprises solving quadratic programs, and is done with QProg, the quadratic programming solver in Gauss.

I derive two index fund revision formulations under transaction costs including price impact. These are both related to the mean-variance portfolio revision formulation under transaction cost. One is a QP and the other a quadratically constrained QP (QCQP). In the empirical tests, linear and quadratic programs are solved using QProg. QCQPs are solved using a sequential quadratic programming (SQP) solver available in Gauss. Moreover, a mean absolute deviation model for index fund management is formulated and implemented in the form of a linear program (LP), and solved with QProg. A heuristic index fund implementation based on Satchell and Hwang (2001) is solved using the SQP solver. A flexible index calculation methodology is implemented in Gauss. The index study uses a computerized “portfolio research system” that comprises a securities inventory, optimization, and trading in the presence of transaction costs, corporate actions, and cash flows, all on a daily basis. The system was also programmed and implemented in Gauss.

Offered below is a condensed description of the data employed in this thesis; fuller descriptions are provided in the respective chapters where the data are used. A variety of data sources have been utilized.
The cross-sectional price impact cost models are estimated on limit order book information for stocks trading on the SSE’s A-list, O-list, New Market and Aktietorget. In the fall of 2000, limit order book information was retrieved electronically from Ecovision, a real-time financial information system. Additional data, market capitalizations and stock quotes, were collected from the SIX Trust database and the online archive of stock prices of Affärsvärdiden (a business publication), respectively.

As mentioned, I had decided that the empirical work on index fund management preferably should be based on daily data and include cash flows, corporate actions, and transaction costs. During the initial phase of the project, a data vendor was asked about whether they could contribute index composition and related data to this project. Their response was interpreted as being positive, but no data were delivered. The requested data therefore had to be compiled independently. The collection process was semi-manual. Significant toil was required to compile, structure, and error check the data. A comprehensive set of daily data was compiled for the period 1987-2000. The empirical tests of the importance of transaction cost control in both mean-variance and index fund problem use this data base. Data have been collected from various public sources, both printed and electronic, including SIX Trust, the Stockholm Stock Exchange, Datastream, and Ecovision. The data set include virtually all common stocks traded on the Stockholm Stock Exchange between 1987 and 2000. The data comprise 13 years (3275 days) of daily closing stock prices, market capitalizations based on all shares outstanding, traded volumes, and corporate action adjustment factors. The dividends data include dividend amount, ex-dividend date and payout date. For the period 1990 to 2000, the data include corporate actions such as mergers and acquisitions, conversions, spin-offs, and new issues. This data have been collected from the Swedish Tax Agency, the Stockholm Stock Exchange, and various other sources.

1.8 Disposition

The remainder of the thesis is organized as follows.

Chapter 2 Transaction costs concerns research task (i). Cross-sectional models of price impact costs for Swedish stocks are developed using limit order book
information in a new way. The models developed allow for explicit consideration of price impact cost as they are functions of order size. Theoretical analysis shows that the price impact cost function of order volume for market buy and sell orders in a limit order book with discrete prices is increasing and piecewise concave. The way in which the price impact cost functions are defined and measured in the modeling is consistent with uninformed, immediate trading. Based on limit order book data for a large cross-section of Swedish equities, empirical price impact cost functions are estimated separately for buy and sell orders. The robustness and predictive performance of the estimated models are evaluated on a holdout sample. Total transaction costs are obtained by adding the relevant commission rate to the price impact cost.

In Chapter 3 Mean-variance portfolio management under transaction costs, research tasks (ii) and (iii) are treated within the mean-variance problem. I extend the standard mean-variance model by formulating a quadratic program for mean-variance portfolio revisions under transaction costs including price impact costs. The transaction cost specification used, and that allows for explicit modeling of price impact cost, is a quadratic function (without a constant term) of portfolio weight change. This specification is consistent with recent empirical models including those developed in Chapter 2. In lieu of possibly more elaborate transaction cost functions, the approach taken is believed to provide a good balance between realism, computational cost, and ease of implementation. An empirical test is carried out to examine the importance of transaction costs and transaction cost control in mean-variance portfolio revision. The extended mean-variance model is integrated with the empirical transaction cost models developed in Chapter 2. The integrated model is applied to revise portfolios of different sizes in terms of net asset value (NAV), and across a broad range of risk attitudes. The initial (unrevised) portfolios are capitalization-weighted and contain all Swedish stocks with sufficient data. The test estimates for each revision the transaction costs and portfolio turnover incurred as well as the expected utility and diversification of the resulting portfolio. The integrated model’s performance is compared to that of the standard mean-variance model as well as to the performances of some portfolio revision models involving less elaborate transaction cost specifications.
Chapter 4 Index fund management under transaction costs deals with research tasks (ii), (iii), and (iv) within the context of index fund management. Material contained in the previous chapters are further expanded and integrated. I develop two different models for index fund revision under transaction costs including price impact costs. Both of them build on the extended mean-variance portfolio revision model developed in Chapter 3. The models are integrated with the transaction cost models of Chapter 2. The importance of transaction costs and transaction cost control in index fund revision is examined by a number of large-scale empirical tests based on daily data. Transaction costs including price impact, cash flows, and corporate actions are incorporated in the test setting. In the tests, the two index fund revision models and several alternative approaches, including full replication, are applied to track a Swedish capitalization-weighted stock index. The different approaches are implemented and tested under a number of variations including different tracking error measures and different types and degrees of transaction cost control. Comprehensive results on tracking performance and transaction costs are provided.

The thesis is summarized and concluded in Chapter 5. Theoretical, methodological, and empirical contributions as well as directions for future research are presented in this chapter.
Chapter 2

Transaction costs

In this chapter, models of individual Swedish stocks’ transaction costs including price impact are developed. The premises for the modeling are that the trading is uninformed, demands immediacy of execution, and takes place in a computerized public limit order book system. The empirical analysis considers stocks trading in the limit order book system of the Stockholm Stock Exchange. The transaction cost models are needed in Chapters 3 and 4, where the importance of transaction cost control will be examined within the context of mean-variance and index fund management, respectively.

2.1 Introduction

“FINANCIAL SCHOLARS AND PRACTITIONERS are interested in transaction costs for obvious reasons: the net gains to investments are affected by such costs and market equilibrium returns are likely to be influenced by cross-sectional differences in costs.” Roll (1984, p. 1127)

“Obtaining accurate measures of trade execution costs and assessing the reasons for their systematic variation is important to individual investors, portfolio managers, those evaluating brokerage firm or financial market performance, and corporate managers considering where to list their shares. Interest in measuring trading costs appears to have increased in recent years.” Bessembinder (2003, p. 233)

Models of trading or transaction costs are potentially useful in many instances including the evaluation of investment strategies (see, e.g., Wermers 2000; Chalmers et al. 2001), portfolio construction (see, e.g., Chapter 3 and 4 here, and Keim and Madhavan 1998), design of market indexes (Shah and Thomas 1998) and policymaking (Keim and Madhavan 1998).

A trading cost model has two components: one explicit, consisting primarily of commission costs, and one implicit, consisting of price impact and opportunity cost (Keim and Madhavan 1997). Commissions are in Sweden often pre-negotiated and

---

3 Precursors to this chapter have been presented at Umeå Workshop in Accounting and Finance 11 June 2001, and at Research Institute USBE Finance Workshop 14 April 2003.
expressed as a fixed percentage of order value. Opportunity costs are subsequently argued to be less relevant in the context of this study and will therefore be omitted. The price impact cost is often defined as the volume-weighted average price degradation relative to the quote midpoint prevailing at the time the decision to trade was made (Keim and Madhavan 1998). Note that price impact cost defined this way subsumes the so-called bid-ask half-spread. As the explicit costs usually are known for Swedish stocks, this chapter focuses on price impact costs.

In the absence of any existing trading cost models for Swedish stocks, the choice was either to adopt an existing model for foreign equities, or to develop one specifically for Swedish equities. Transaction cost models, in which the percentage cost of a transaction is a function of trade (or order) size and thus allow for explicit modeling of price impact costs, are found in, e.g., Loeb (1983, 1991), Grossman and Sharpe (1986), Chan and Lakonishok (1997), and Keim and Madhavan (1997). These models are cross-sectional and concern U.S. stocks. For a given stock, the percentage volume-weighted average trading cost is typically modeled as a function of trade size, market capitalization, stock price level, stock return volatility, investment strategy, and exchange listing. In general, the models express trading costs as decreasing in market capitalization and stock price, while increasing in trade size and volatility.

Albeit possible, it seems quite unsatisfactory to use a cross-sectional model estimated on, say, U.S. data for the prediction of transaction costs for Swedish stocks. One reason is the different architectures of the marketplaces. The Stockholm Stock Exchange (SSE) is an automated open limit order book market, while the two major U.S. markets are not; New York Stock Exchange (NYSE) is an auction market with specialists, and NASDAQ (National Association of Securities Dealers Automated Quotation System) operates largely as a dealer market, and these facts may imply systematic differences in transaction costs. Moreover, translating in a meaningful way

---

models, where some variables are expressed in U.S. dollars, to Swedish kronor may be
difficult, considering, among other things, the large variation in exchange rates over
time. In addition, some of the existing models involve variables irrelevant (e.g., U.S.
exchange affiliation) for Swedish stocks. Hence, the choice fell on developing models
specific to Swedish stocks, based on Swedish data. This resulted in the following
theoretical and empirical analyses of price impact costs in a limit order book setting:

The theoretical analysis reveals that the (quoted) price impact cost (the absolute
volume-weighted average price degradation relative to the quote midpoint) for market
buy and sell orders in a limit order book with discrete prices, is an increasing
piecewise concave function of order volume. This detail appears to have been
unnoticed in the literature. Moreover, this definition of price impact cost is consistent
with trading that is uninformed and immediate.

Instead of using detailed institutional order and trade data as Chan and Lakonishok
(1997) and Keim and Madhavan (1997) do, I use data in the form of limit order book-
“snapshots”, which are simultaneous records of the five best ask and bid price levels
and corresponding volumes. The orderbooks for virtually all listed Swedish stocks are
sampled at a dozen arbitrary times during a period of 16 consecutive trading days in
2000. For any security, at any given time, this information allows the the whole quoted
price impact cost function to observed. Those price impact observations are pooled
and related cross-sectionally by means of regression analysis to a set of explanatory
variables identified in the literature. The estimated models are functions of order size,
thus allowing for explicit modeling of price impact costs. The price impact costs of
market buy and sell orders are analyzed separately, and this permits identification of
possible asymmetries between them.

To shed light on the robustness and possible performance out-of-sample of the
estimated models, a simulation analysis is performed. An estimation sample and a
holdout sample is randomly selected from the full data set. The identified models are
re-estimated in the estimation sample and then used to predict price impact costs in the
holdout sample. This is repeated one hundred times. I introduce an error metric
designed to measure predictive performance in a meaningful way.
The main contributions of this chapter are as follows. I derive the price impact cost as a function of order size in a limit order book with discrete prices. Utilizing limit order book data in a novel way, cross-sectional models of price impact costs for Swedish equities are estimated and tested for robustness and performance out-of-sample.

According to microstructure theory for quote-driven markets, the cost of transacting a given quantity of a stock is a function of the dealer’s or the liquidity provider’s order, inventory, option and information costs, and his or hers beliefs, preferences and wealth (see, e.g., Stoll 2003). The transaction cost is, partially, represented by the dealer’s quoted bid and ask and corresponding volumes. In an order-driven market, submitted limit orders provide liquidity. Hence, the quoted price schedule or price impact cost function should reflect the aggregated order, inventory, option and information costs of the traders having submitted the limit orders, as well as their beliefs, preferences and wealth. Limit orders also represent the implementation of investment strategies, of which liquidity provision may be one. Submitted limit orders are binding commitments (until cancellation), which dealer quotes need not be (Lee et al. 1993), and the order book contains visible information about the cost of transacting beyond the best prices and volumes quoted. Empirical analysis of price impact costs based on limit order book data may thus provide information about transaction costs and market microstructure that is complementary to what analysis of quote-driven systems can reveal. Moreover, researchers indicate a need for more research on electronic order-driven markets. Ahn et al. (2001, p. 767), for instance, establish that,

“Although many stock exchanges around the world are based on pure limit order books, very few empirical papers investigate the role of limit-order traders in an order-driven market without any designated market maker.”

The next section contains a description of the Swedish stock market and the electronic limit order book trading system. A theoretical analysis of the price impact cost function in a limit order book is carried out in Section 2.3. Certain features in the background of the present study have significantly influenced its design. These and important research design issues including estimation approaches and possible approaches to data collection are described in Section 2.4, which also contains a
discussion aimed at identifying possible cross-sectional determinants of price impact costs in the vast literature on trading costs and bid-ask spreads. The data are described in Section 2.5. The design and the results of the empirical analysis are presented in section 2.6. In Section 2.7, the chapter is concluded.

2.2 The Stockholm Stock Exchange trading system

Since 1989 most of the trading in Swedish equities is carried out in a continuous, computerized open limit order market system operated by the Stockholm Stock Exchange. In 1999, an improved version of the original system was launched.

Two types of orders can be submitted in the system: limit orders and market orders. A market order is an order to buy or sell a fixed quantity at the best available prices. A limit order is an order to buy or sell a fixed quantity only at a certain limit price or better. New limit orders are entered into the electronic limit order book where they are stored until executed or cancelled. A trade executes automatically when the price of a new order matches the price of an order in the limit order book. Orders in the limit order book are executed giving priority firstly according to price and secondly according to time of submission, i.e., higher bid prices and lower ask prices have priority, and if the price is the same for several orders, they are executed in the order they were submitted. A market order is executed according to the rules of priority against limit orders in the order book until the order is filled (given that the order book is deep enough to accommodate the volume of the order). Any unfilled portion of a market order is converted into a limit order.

Trading is done in whole lots. Normally there are 100 or 200 shares per lot. Odd lots are routed to a "small order system" where they are executed at the same price as the price applicable for whole trading lots at any given time.

Order prices are required to be stated in multiples of a fixed minimum price unit or minimum tick size. Tick sizes are delineated in Table 2.1.7

---

5 In the year 2000, trading was continuous except at the open when there was a single-price auction procedure.
6 The value of one lot should approximate a certain statutory amount, equal to approximately SEK 18000.
7 This is the tick size regime at the time of the study. The tick size regime has however varied over time.
Table 2.1 Tick sizes, stock price intervals and price impacts

This table shows the price intervals, the corresponding tick sizes, and the resulting price impacts bounds or halfspreads, which are defined as: Halfspread = |(Bid or Ask) – Midpoint|/Midpoint, where Midpoint is (Ask+Bid)/2; Min. halfspread = (Tick size/2)/(High price in interval - Tick size/2); Max. halfspread = (Tick size/2)/(Low price in interval + Tick size/2).

<table>
<thead>
<tr>
<th>Price interval (SEK)</th>
<th>Price impact bounds from minimum tick size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Start</td>
</tr>
<tr>
<td>0.00 - 4.99</td>
<td>0.00</td>
</tr>
<tr>
<td>5.00 - 9.95</td>
<td>5.00</td>
</tr>
<tr>
<td>10.00 - 49.90</td>
<td>10.00</td>
</tr>
<tr>
<td>50.00 - 499.50</td>
<td>50.00</td>
</tr>
<tr>
<td>500.00 -</td>
<td>500.00</td>
</tr>
</tbody>
</table>

In the SSE limit order book trading system, there are no designated market makers or specialists; liquidity provision is thus purely order-driven. Members of the exchange (investment banks, brokerages etc.) act as broker-dealers, i.e., they both trade in their own account and submit orders on behalf of their customers. Since 1 Jan. 2000, there is no entry fee for members, only a monthly fee based on transaction volume (amounting to a minimum of SEK 100000 per year). In addition, all members pay fees for system connections (Factbook 1999, p. 24).

The SSE limit order market is highly transparent. All members have access in real-time to the aggregated quantities offered and demanded at every price level in the limit order book for any given security.8 Also revealed to members are dealer identification codes, which appear in order of priority at each level. Information available to the public - and for this study - is limited to the five best price levels and the corresponding volumes at each side of the market. Furthermore, data on completed transactions in the trading system, i.e., prices, quantities and counterparts, are continuously transmitted to the public via various financial information providers. The system is thus not (completely) anonymous. In particular, submitters of limit orders may be identified by the dealer codes, but, obviously, the identities of those having entered orders that execute directly against submitted limit orders are not known by the limit order submitters until after trade completion.

---

8Members, however, have the opportunity to use hidden orders. These are not visible in the limit order book, and they have lower priority than all visible orders with the same price.
It is possible to transact outside the electronic trading system, by so-called off-exchange registration (as opposed to automatic matching). During regular trading hours, smaller trades (less than 500 or 250 whole lots depending on the issues’ liquidity category) must be executed within the bid-ask spread, while larger trades may be executed outside the spread. All such transactions have to be reported to the trading system within five minutes after completion. Transactions completed after the regular trading session may be executed outside the spread and must be reported to the system 15 minutes before the market opens the following day.9

The limit order book data were retrieved from Ecovision which is a financial information system that provides such data in real-time for all shares traded in SSE's electronic trading system. The data for this study emanate from the period 22 Aug. 2000 to 12 Sept. 2000, comprising 16 trading days. During that time shares on the following lists were traded in the system: SSE’s A-list, O-list, and the New Market; and Aktietorget.

Actively traded shares of the largest firms that meet certain requirements regarding market capitalization (300 MSEK minimum), widespread ownership, ability to provide information to the public, and verifiable economic history, could be found on the Stockholm Stock Exchange's A-list. Firms that do not possess verifiable economic histories, but fulfill the other requirements (to a certain degree) of the A-list, may list their shares on the O-list of the SSE. Moreover, at the beginning of 1998, the SSE launched a list called the New Market or Nya Marknaden intended for small capitalization stocks. In contrast to the companies listed on the A-list and the O-list, the SSE is not responsible for firms listed on the New Market. Instead, each company has one of the exchange's members as a sponsor, which, pursuant to an agreement with the Exchange, is responsible for ensuring that the company fulfills the requirements for trading on the New Market.

Aktietorget is a market place operated by an organization distinct from the SSE. This market is intended primarily for trading in development stage companies. There are requirements for listing in terms of book value of equity, widespread ownership

---

9 It may be worthwhile to note that some of the securities that traded in SSE’s system, also traded on other markets, e.g., SEAQ International, Nasdaq, and NYSE, during the period of this study.
and dissemination of information to the market. All trading in shares on Aktietorget, however, takes place in the electronic limit order system of the SSE.\textsuperscript{10}

### 2.3 Price impact cost as a function of order volume in the limit order book

What is price impact cost? The price impact cost can be defined as any temporary adverse price movement or price impact\textsuperscript{11} induced by a transaction. In a world with perfect liquidity, one should be able to instantaneously buy and sell large quantities of a security at the same price. The price impact cost would then be zero. In reality, however, there is a spread between the best ask and bid, and, in general, the larger quantities one buys or sells, the worse the prices get.

It may be instructive to think about this in the following way. Consider an imaginary market order to transact (buy or sell) a certain number of shares out of the order book, and compute the average price per share for that transaction. This yields the average price for that particular order volume. For a given order book, the whole average price function of order volume is obtained if the average price is calculated per every possible order volume.

To derive the price impact cost for a given volume, compare the (volume-weighted) average price for the given order volume to a price representing a situation of perfect liquidity. A natural candidate for such an ideal price is the quote midpoint, i.e., the average of the best ask and bid, as of the time the trade was decided to be executed.\textsuperscript{12} The percentage average price impact cost for a given volume is then defined as the absolute percentage degradation of the average price for the given volume relative to the midpoint.\textsuperscript{13} For a given order book, the whole average price impact cost function

---

\textsuperscript{10} By the time of the study, there was also a market place called the SBI-list. The SBI-list featured its own trading system and is therefore not included in our data. However, the number of companies was quite limited and their nature was pretty similar to the companies listed on Aktietorget and the New Market.

\textsuperscript{11} Note that price impact and price impact cost are sometimes used interchangeably here as well as in the literature.

\textsuperscript{12} Admittedly, the quote midpoint is somewhat arbitrary, although standard in the literature.

\textsuperscript{13} To define price impact cost as the average price change relative to the quote midpoint at, or to the last trade price before, trade decision time, is common in the literature, see, e.g., Chan and Lakonsihok (1995), Keim and Madhavan (1995, 1997), and Griffiths et al. (2000).
of order volume is obtained if the absolute percentage degradation of the average price relative to the midpoint is calculated for every possible order volume.

Note that when price impact cost is defined in this way it is actually expressed in the form of a (rate of) return. Moreover, price impact cost defined like this implies that the whole adverse price movement induced by the transaction is temporary, and thus recognizable as a cost. Temporary price impacts are consistent with uninformed trading. That the price impact cost measurements are based on market buy and sell orders is consistent with trading that demands immediacy of execution. Below, an analysis of the (quoted) price impact cost function in a limit order book is provided.

For either the bid or the ask side of the order book, let \( P_k \) and \( Q_k \) represent the quoted price and order quantity, respectively, at level \( k \), where \( k=1, 2, \ldots, 5 \). The average price impact cost as a function of order size \( q \), and where \( k \) represents the level in the order book where the \( q \)th share would be executed, could then be expressed as

\[
\pi(q) = \frac{\bar{p}(q) - \text{MIDP}}{\text{MIDP}},
\]

(2.1)

where \( \bar{p}(q) \), the average price per share for the volume \( q \), equals

\[
\bar{p}(q) = \frac{\sum_{i=1}^{k-1} P_i Q_i + P_k \left( q - \sum_{i=1}^{k-1} Q_i \right)}{q} = \frac{\bar{p}_{k-1} \sum_{i=1}^{k-1} Q_i + P_k \left( q - \sum_{i=1}^{k-1} Q_i \right)}{q}, \quad \text{for } k = 1, \ldots, 5,
\]

and with \( k \) selected so that

\[
\sum_{i=1}^{k-1} Q_i < q \leq \sum_{i=1}^{k} Q_i.
\]

Note the convention that \( \bar{p}_k = \bar{p}\left(\sum_{i=1}^{k} Q_i\right) = \sum_{i=1}^{k} P_i Q_i / \sum_{i=1}^{k} Q_i \); \( \bar{p}_k \) is the volume-weighted average price per share over levels 1 to \( k \), or, put differently, the average price per share that would result if all shares at levels 1 to \( k \) were transacted. The intuition behind the expression \( \bar{p}(q) = \frac{\sum_{i=1}^{k-1} P_i Q_i + P_k \left( q - \sum_{i=1}^{k-1} Q_i \right)}{q} \) is analogous; the left term in the numerator on the right hand side represents the total value of transacting all shares at levels 1 to \( k-1 \), and the right term represents the value of the shares at level \( k \) that is
included in an order of size \( q \). Dividing the numerator, which thus expresses the value of transacting \( q \) shares, by \( q \) then yields the volume-weighted average price \( \bar{p}(q) \).

The price impact cost as a function of order volume in the limit order book has a specific form. It is an increasing\(^{14}\) piecewise concave function, which is demonstrated below. This follows from that price in the order book is related to accumulated order volume and order book level in stepwise fashion, i.e., the price is constant for the volume at any given level, but it increases (decreases) for asks (bids) between levels. Also note that the volume or number of shares is a discrete variable. However, later on, order volume will be represented as a proportion of outstanding shares rather than the number of shares; this revised quantity is not discrete. The breakpoints of this piecewise function appear at the accumulated volumes and average prices that correspond to the transitions from one level to another in the order book. The price impact cost function is not differentiable in the breakpoints. With only one share at each level, the function would be non-differentiable over its whole domain. In reality, however, the number of shares at any level is greater than one, because trading is done in whole lots.

Since the price impact cost function \( \pi(q) = \frac{[\bar{p}(q) - MIDP]}{MIDP} \) involves an absolute value, write it as a piecewise function, separating the intervals that give positive values for the inside of the absolute value from those that give negative values. For asks \( \bar{p}(q) - MIDP > 0 \), therefore differentiate \( \pi(q) = \frac{\bar{p}(q) - MIDP}{MIDP} \) with respect to the order quantity \( q \) to obtain

\[
\frac{d\pi}{dq} = \frac{(-\bar{p}_{k-1} + P_k)\sum_{t=1}^{k-1} Q_t}{q^2 MIDP}.
\]

For asks, prices increase with the level in the order book, \( k \), therefore \( P_k \geq \bar{p}_{k-1} \), i.e. the volume-weighted average price per share \( \bar{p}_{k-1} \) for levels 2 to \( k-1 \) is lower than \( P_k \), except for \( k = 1 \) where it is constant and equal to \( P_k \). The first derivative is thus zero in

\(^{14}\) Note: A function \( f \) is increasing, if, for all \( x \) and \( y \) in the domain, \( x \leq y \) implies that \( f(x) \leq f(y) \).
the segment corresponding to the first level \((k = 1)\) in the order book, and positive for segments corresponding to \(k > 1\). Moreover, the rate of change is inversely related to \(q^2\), so in a given segment \(k > 1\), \(\pi\) increases at a decreasing rate. Taking the second derivative of \(\pi(q)\) yields

\[
\frac{d^2 \pi}{dq^2} = \frac{2(\bar{p}_{k-1} - p_k) \sum_{i=1}^{k-1} Q_i}{q^3 \text{MIDP}}.
\]

The second derivative is non-positive since \(p_k \geq \bar{p}_{k-1}\). Hence, for the ask side in the order book, the price impact function \(\pi(q)\) is constant for level \(k = 1\), and for levels \(k > 1\) it is an increasing, piecewise concave function.

For bids, \(\bar{p}(q) - \text{MIDP} < 0\). Therefore, because of the absolute value, analyze the negative of the price impact function, i.e., \(\pi(q) = -\frac{(\bar{p}(q) - \text{MIDP})}{\text{MIDP}}\). A parallel analysis to that for asks reveals that the price impact function \(\pi(q)\) for bids is constant for segment \(k = 1\), and for all segments \(k > 1\) it is an increasing, piecewise concave function.

Table 2.2, Panel A, shows an example of a limit order book snapshot. Panel B exhibits, for the bid side, the price impact cost and average price calculated on the accumulated volumes at the end of each order book level. A plot of the average price as a function of order quantity on the bid side is presented in Figure 2.1.\(^{15}\) The corresponding average price impact cost function is depicted in Figure 2.2.

\(^{15}\) Observe that the quantity dimension of the average price and the price impact functions is discrete, but that the plots in Figure 2.1 and Figure 2.2 for convenience are continuous.
Table 2.2 A limit order book snapshot, average price and price impact
Panel A shows an example of a limit order book snapshot for the stock of ABB as of 22 Aug. 2000, 11.30. MIDP is calculated as the average of the best bid and best ask quote. Panel B shows components of the calculation of the price impact cost at accumulated volumes at each level in the order book. (Level 0 corresponds to the first share).

Panel A

<table>
<thead>
<tr>
<th>Level (k)</th>
<th>Orders to buy</th>
<th>MIDP</th>
<th>Orders to sell</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Q</td>
<td>Bid</td>
<td>Ask</td>
</tr>
<tr>
<td>1</td>
<td>2000</td>
<td>1087</td>
<td>1090</td>
</tr>
<tr>
<td>2</td>
<td>2200</td>
<td>1085</td>
<td>1093</td>
</tr>
<tr>
<td>3</td>
<td>1000</td>
<td>1082</td>
<td>1095</td>
</tr>
<tr>
<td>4</td>
<td>200</td>
<td>1075</td>
<td>1096</td>
</tr>
<tr>
<td>5</td>
<td>25</td>
<td>1074</td>
<td>1097</td>
</tr>
</tbody>
</table>

Panel B

<table>
<thead>
<tr>
<th>Level (k)</th>
<th>$\sum_{i=1}^{k} Q_i$</th>
<th>$\sum_{i=1}^{k} P_i Q_i$</th>
<th>Average price $\bar{p}_k$</th>
<th>Price impact cost $\pi(q)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 0</td>
<td>0</td>
<td>0</td>
<td>1087</td>
<td>0.138%</td>
</tr>
<tr>
<td>1</td>
<td>2000</td>
<td>2174000</td>
<td>1087</td>
<td>0.138%</td>
</tr>
<tr>
<td>2</td>
<td>4200</td>
<td>4561000</td>
<td>1085.95</td>
<td>0.234%</td>
</tr>
<tr>
<td>3</td>
<td>5200</td>
<td>5643000</td>
<td>1085.19</td>
<td>0.304%</td>
</tr>
<tr>
<td>4</td>
<td>5400</td>
<td>5885000</td>
<td>1084.81</td>
<td>0.339%</td>
</tr>
<tr>
<td>5</td>
<td>5425</td>
<td>5884850</td>
<td>1084.76</td>
<td>0.343%</td>
</tr>
</tbody>
</table>

Figure 2.1 Average price as a function of order quantity, $\bar{p}(q)$
Note that it is possible for the price impact cost function to take quite different shapes depending on the values of its parameters (this is corroborated by the data). For the sake of argument, assume that there is only one share per level. Then, for instance, with only two levels populated, and with only one share at the first level, the form of the function is concave. And with only one share at each level and with prices that increase faster than linearly between levels, the price impact function becomes convex.

It appears unclear whether the form of the volume-weighted average price (and price impact cost) function has been recognized in the literature. For example, Figure 2 in Glosten (1994) of average price looks as if only the average prices given the accumulated volumes at the breakpoints are considered, and that it is assumed that the rate of change of the average price is constant between the breakpoints. Also see Figure 8 in Niemeyer and Sandås (1993).


2.4 Price impact cost modeling using limit order book data

2.4.1 Background

The requirements of Chapter 3 and Chapter 4 have significantly influenced the design of the research in this chapter. These and certain other research design considerations are described in this subsection.

Trading costs are often measured using the implementation shortfall approach (Perold 1988). In this approach, the performance of the actual portfolio is compared to the performance of an imaginary paper portfolio. The performance of the paper portfolio is based on the assumption that every order executes in its entirety at an ideal price (usually the quote midprice) prevailing at the time the decision to trade was made. It is also assumed that no commissions or other fees are incurred. Total trading costs then, are defined as the difference in return performance between the ideal portfolio and the actual portfolio. Total trading costs can be decomposed into explicit costs, mainly commissions, and implicit costs, consisting of price impact and opportunity costs (Keim and Madhavan 1998). If there is a difference in performance between the paper portfolio and the actual portfolio after explicit costs and price impact costs are accounted for, then that difference is attributable to opportunity costs. The opportunity costs (or gains) are thus due to the parts of the desired orders that were not filled or were executed with a delay. The total difference between the actual portfolio and the paper portfolio is called the implementation shortfall and it equals the sum of explicit, price impact, and opportunity costs.

As noted previously, this chapter focuses on price impact costs, that is transaction costs excluding explicit costs and opportunity costs. There are essentially two reasons for this. First, in the Swedish marketplace explicit costs are generally negotiated in advance and expressed as a constant percentage, that, when added to the price impact function, yield the full transaction cost function. This implies that commissions and price impact costs are independent. Keim and Madhavan (1997) indicate a need to jointly analyze explicit costs and price impact costs, because they may be systematically related in that higher commissions could encourage a broker to execute the order in such a manner that it results in a lower price impact cost than otherwise. This concern is without merit here, since I will proceed as if all trades are executed
electronically without any brokers intermediating. There are and have been several ways to trade electronically. Only exchange members are, and have been, able to submit orders directly to the Stockholm Stock Exchange, but it is and has been possible for non-members to trade directly. Non-members achieve this by renting access through a member, who provide them with a connection and relevant equipment such as trading terminals and applications. One way in which such an arrangement is paid for is through commissions. In addition, since the mid 1990s it has been possible to trade electronically through various internet brokerages.

Second, the transaction cost models will be used in the implementation and evaluation of investment strategies or portfolios where trading is uninformed (see Chapter 3 and Chapter 4). In particular, Chapter 4 examines alternative equity indexing approaches, for which according to Keim and Madhavan (1998, p. 54) opportunity costs are zero. Opportunity costs are associated with missed trading opportunities. These costs emanate from the idea that there is information that has value, but as that value will decay over time, the information has to be exploited in a timely manner (Keim and Madhavan 1998). However, while passive strategies are uninformed with regard to asset values and valuations, failure to timely and adequately update, e.g., an index fund in the event of an index reconstitution may lead to tracking error, which is associated with some kind of cost or loss.

Furthermore, apart from lacking the data required for their estimation, it is, I argue, possible to identify situations under which opportunity costs should not receive any explicit consideration in portfolio optimization applications with transaction costs at all. In the context of portfolio optimization with transaction costs, the optimizer weighs, in terms of expected utility, changes in portfolio expected return and risk against the transaction costs incurred from changes in portfolio holdings.16 The modeling should therefore also reflect, for the investor and trading horizon etc. in question, the risk of non-execution, that is, the risk that trades of sizes beyond some

---

16 The optimizer outputs the optimal changes in the fractions of portfolio wealth allocated to individual holdings. For a given stock, such a changed fraction or amount of money includes transaction costs. In principle, with transaction costs as compared to without, a sale (buy) transaction thus implies selling (buying) more (fewer) shares to comply with the desired amounts of money. Opportunity costs can thus arise due to either wrongly predicted transaction costs or that the optimizer suggested trades that were impossible to fill (or were overfilled).
levels sometimes cannot be absorbed in the market at all. Thus, with perfect prediction of transaction costs and with the optimizer consciously accounting for non-execution, only trades that are possible to fill in a timely manner are targeted.\textsuperscript{17} Under these circumstances, there is little room for opportunity costs to assert themselves. In reality, predictions are less than perfect, and the actual, resulting portfolio may differ from the portfolio suggested by the optimizer, thereby inducing opportunity costs. However, for short enough horizons and small enough volumes, in open limit order markets and sometimes also in other market structures, the price impact and the depth of every security are known. The differences between actual and desired trades will then be minimal.

For the transaction costs model to comply with Chapter 3 and Chapter 4, a few restrictions and requirements apply. They are described as follows. The modeling should reflect that trading in Swedish stocks is done in an open electronic limit order book system and that the trades are assumed to be completed immediately. The indexing study in Chapter 4 covers the period 1990-2000 and the trading cost model should be possible to apply to that whole period and to stocks that have ceased to exist. The tests of Chapter 3 include virtually all listed Swedish equities, including the very smallest issues in terms of capitalization. The data set gathered contains daily closing stock prices, traded volumes, and market capitalizations. Consequently, the models should be expressed in terms of the variables in the database and cover not only larger capitalization stocks traded on the Stockholm Stock Exchange's lists for most actively traded stocks, but also less liquid, small-capitalization issues.

Given these restrictions and the absence of existing trading cost models for Swedish stocks, I choose to develop a model specifically for Swedish equities instead of adopting a model estimated on, for example, U.S. data. It can be noted that the most important explanatory variables in existing models are identical to, or derivable from, those available in this data set. Actually, the data set is rich enough to allow for the introduction of two variables previously not used in the literature where transaction costs are measured using trading record data.

\textsuperscript{17} Perhaps indicative of such considerations, Keim and Madhavan (1997), for instance, find completion rates of around 95\% among their sample of institutional traders.
In addition, the portfolio and index fund revision programs to be used in Chapter 3 and Chapter 4 require $t_i$, the total (percentage volume-weighted average) trading cost function for a given stock $i$, to be specified as a linear function of order size: $t_i = c_i + \pi_i$, where $c_i$ is the constant percentage explicit cost representing commissions etc. for stock $i$, and $\pi_i$ is the price impact cost function for stock $i$, which, in turn, is $\pi_i = \alpha_i + \beta_i \text{ORDERSIZE}_i$, where $\text{ORDERSIZE}_i$ is the number of shares in the trade relative to the total number of outstanding shares for stock $i$, and $\alpha_i$ and $\beta_i$ coefficients uniquely determined for the given stock. In view of existing models of transaction costs the required specification does not appear very severe, since it is practically identical to the ones employed in for example Keim and Madhavan (1997) and Chan and Lakonishok (1997). Nevertheless, I will test other functional forms as well to see whether they fit the data better, and if that is the case they will have to be approximated by the required linear specification. In any case, Olsson (1999) shows that several existing and essentially non-linear trading costs models, e.g., Loeb (1991) and Grinold and Kahn (1995), may be successfully approximated by functions linear in order size.

Given the above prerequisites, I proceed to discuss how to estimate a model that expresses price impact cost as a function of the available explanatory variables and what data to use for the price impact cost measurements. After that, the explanatory variables are discussed in the light of market microstructure theory.

### 2.4.2 Modeling of transaction costs: Estimation approaches and data

Transaction cost models can in principle be estimated for one stock at a time or cross-sectionally. One can start from a theoretical structural model of price formation in a certain market architecture or use a more data driven approach. Which blend of approaches to choose depends on factors such as the data available and the purposes of the modeling.

Most theoretical models of price formation consider only one share at a time and describe the spread or price impact in terms of the intraday price dynamics and the equilibria of games and activities of agents in the market conceptualized in the form of order processing costs, adverse selection and inventory control considerations. The
settings are idealized and stylized, but the resulting models are complex and the references to directly observable or easily measurable variables are sometimes limited, and the data requirements associated with the estimation of these models can be quite formidable. Moreover, relatively little is known about these models’ descriptive validity with regard to price impact as a function of order size (see de Jong, Nijman, and Röell 1996; Sandås 2001). Given the purpose of this study and the data available, it thus seems impractical to start from a (structural) model of price formation in an automated limit order market like that of Glosten (1994).

Many studies employ publicly available (intra-daily) trade and quote data to analyze trading costs and liquidity (see for example Bessembinder 2003 and the references therein). However, issues arise with trade and quote data when it comes to measuring trading cost, including the matching of transactions to quotes and the classification of transactions as buyer or seller initiated (Bessembinder 2002). Order flow data, including information about order type (market or limit order), time of order submission, and multiple-fill orders (packages), are not provided. Estimating price impact costs on trades and quote data may underestimate the price impacts of larger trades, because such trades may have been broken up in smaller parts; and symmetrically, observed trade size ranges may be limited in magnitude.\(^\text{18}\) Finally, no readily available database of this kind exists for Swedish stocks.

It is perhaps obvious, but nevertheless important to realize that unambiguous comparisons of liquidity in terms of quoted spreads or price impacts are possible only for orders of the same size. Much of the microstructure literature focuses on the size of the spread between the best ask and bid, and pays less attention to the order size dimension; usually referred to as the depth (i.e., the number of shares at the best levels).\(^\text{19}\) This is to some extent natural, since in quote-driven markets only the prices and quantities at the best levels are visible, and the quoted bid and ask price impacts are thus constant over these quantities; public trade and quote databases such as TAQ and TORQ, only contain the best quoted prices and volumes. Unambiguous

\(^{18}\) A recent study by Breen et al. (2002) uses this approach.

\(^{19}\) See Lee et al. (1993, p. 346), who note “Since market liquidity has both a price dimension (the spread) and a quantity dimension (the depth), it is surprising that much of the literature focuses only on the spread.”
comparisons of liquidity in terms of quoted bid-ask spreads (across securities and or over time) are possible, but only over the minimum quoted volume (among all observations).

The preferred type of data is probably (institutional) “trading records”\textsuperscript{20}, i.e., data that includes detailed information on trader identity, trade objectives, order characteristics and trade completion (Chan and Lakonishok 1997; Keim and Madhavan 1997). For instance, such data make it possible to realistically assess the price impacts of large orders that are split and executed over several days. However, as Bessembinder (2003) points out, access to this type of data is restricted and the data often applies to a limited number of institutional customers, and or to a limited number of stocks. In consequence, Bessembinder emphasizes the importance of using publicly available data in research. Moreover, trading record data might be self-censored in the respect that the observed, executed trades probably represent trades that were possible to do without too severe price concessions, while the most expensive trades never were submitted to the market, and could therefore not be observed; the high fill rates of around 95% reported in several recent studies (see Keim and Madhavan 1998) may be indicative of this phenomenon.\textsuperscript{21}

Void of trading record data, I opted to use publicly available information in the form of “snapshots” or instantaneous records of individual stocks’ limit order books. In a real-world application, the most efficient approach to predicting transaction costs under full immediacy in an automated limit order book market seems to be to use the information in the order book in real-time. Given that an order book does not change for some time frame, the price impact cost prediction will be more or less perfect over the volumes in the order book and, in addition, opportunity costs may be more or less eliminated, since it is possible to target only trades that are possible to fill. This approach will not be possible to use in, e.g., Chapter 4, because I do not have any historical record of limit order book observations covering the whole period of that

\textsuperscript{20} Stoll (2003) uses this term. Bessembinder (2003) uses the term “customer order data”.

\textsuperscript{21} This means that cross-sectional analysis must be approached with care, since the cost observations for large trades are likely to be concentrated to more liquid securities, and the cost estimates for large trades may be underestimated.
study, i.e., from 1990 to 2000. Instead, the order books were sampled at a dozen arbitrary times for virtually all listed Swedish stocks during a period of 16 consecutive trading days in 2000. For any security, at any given time, this information allows observing the whole quoted price impact cost function, separately for the ask and bid side. Those price impact observations are pooled and related cross-sectionally to the explanatory variables by means of regression analysis.

The estimated models can then be applied backwards in time in the optimization applications. It is thus assumed that the estimated coefficients are constant and the estimated models descriptively valid for the whole period between 1990 and 2000. It appears reasonable to believe that the Swedish market was less liquid with higher transaction costs in the beginning of the period. Given constant coefficients and that the explanatory variables were the same then as now, that is also what is likely to be reflected by this modeling. The presumed lower liquidity in the earlier years would be reflected by the values of the variables prevailing at that time. Specifically, the in the earlier years generally smaller market capitalizations and the lower trading activity, which are two important explanatory variables in the estimated models, imply lower liquidity and higher transaction costs.

That the liquidity of Swedish stocks increased between 1990 and 2000 is supported by empirical evidence. For trades in Swedish stocks, Chiyachantana et al. (2004) report an average price impact cost of 0.32% for 1997-1998, and of 0.35% for 2001. They employ proprietary data from the Plexus Group on global trades conducted by some 200 institutional traders. Similar but less comprehensive data from State Street Global Advisors are used by Perold and Sirri (1998), whose estimate of the average price impact cost is 1.59% for the period 1987-1991.

To summarize, the approach taken here measures price impact under full immediacy; every trade is assumed to be executed as a market order immediately and the whole price effect will be regarded as temporary, which is consistent with uninformed trading.

It is probably possible to achieve lower price impacts by using more patient trading, i.e., by using limit orders instead of market orders. A problem with non-aggressive limit orders is the risk of non-execution and the associated opportunity
costs. Griffiths et al. (2000) compute the total costs for aggressive respective patient trading strategies in a limit order system. Although aggressive orders have higher price impacts costs than patient orders, which often earn negative price impacts costs, the low fill rates of patient orders incur opportunity costs, which when accounted for reduce, and in some cases eliminate the differences. That index funds trade relatively aggressively is observed by Keim and Madhavan (1997).

The price impact costs estimates are thus likely to be rather pessimistic (i.e., high), and probably higher than those a more patient trading strategy could achieve. Also, note that in a limit order book, price and quantities represent binding commitments to trade, while quotes on NYSE for instance are not binding (Lee et al. 1993, p. 357). Indeed, several studies document that trades in quote-driven structures receive improvement relative to quotes, that is, are executed at prices better than those quoted (see, e.g., Werner 2003). The conservatism of the price impact estimates is further accentuated by the fact that hidden cannot be observed. There are only data on the five best levels in any order book, so the estimates will as a consequence not take account of orders outside the five best levels. However, (according to private communication with traders,) levels outside the fifth best are rarely populated and hidden orders are not so common on the SSE (to the extent this can be measured).22 Moreover, the price impact cost observations are for orders of quite limited sizes, as one can only observe what is in the order book at a given time.

This study appears to be the first one that uses all public prices and quantities in the consolidated limit order book to estimate cross-sectional models of buy and sell price impact costs, where the percentage cost is a function of order size.

2.4.3 Cross-sectional determinants of trading costs

In the previous section, I established, for individual stocks, the form of the quoted price impact cost as a function of order volume in a limit order book. In the preceding subsections, the prerequisites for the study were outlined, and, in the light of these, different modeling approaches were discussed in terms including data availability and estimation procedures. The focus of this subsection is to identify and discuss potential

---

22 Degryse (1999, p. 1343) report that on Brussels CATS market, invisible (hidden) orders constitute about 16% of the entire limit order book (the unit of measurement is unfortunately not specified by Degryse).
determinants of transaction costs and price impact costs. How these factors could be captured by the explanatory variables available is a question that will be answered here. There is a vast theoretical and empirical literature on market microstructure and transaction costs (which is reviewed in, e.g., Cohen et al. 1979; Keim and Madhavan 1998; Coghenour and Shastri 1999; Biais, Glosten and Spatt 2002; Madhavan 2002; Stoll 2003).

I want to find a set of cross-sectional determinants that can be used to predict the whole buy and sell price impact functions for stocks traded in an order-driven open limit order market. In an ideal situation, there would be a general, empirically supported, dominant theory for order-driven markets from which relevant variables could be selected. There is a body of theoretical literature focusing on various aspects of order-driven markets including Cohen et al. (1981), Glosten (1994), Handa and Schwartz (1996), Parlour (1998), Foucault (1999), and Handa, Schwartz and Tiwari (2003). However, none of these studies provides such a general, dominant framework. For some recent empirical evidence relating to the theories put forth in some of these studies, see, e.g, Ranaldo (2004).

Most microstructure theory concerns quote-driven market structures with designated market makers or dealers acting under varying degrees of competition, and bid-ask spreads rather than whole bid and ask price impact functions. In these theories, the bid-ask spread is the mechanism through which the dealer gets compensation for the costs of the services provided, primarily predictable immediacy, and the risks assumed. Microstructure models consequently explain the size and various other aspects of the spread in terms of certain costs faced by a dealer, that is, from the supply side. Numerous empirical studies have investigated different aspects of these theories. In resemblance with the situation for pure limit order markets, there is, however, no single, dominant theory. In selecting and motivating the explanatory variables, I will as a compromise consider theories and models concerning both order and quote-driven markets.

Microstructure models explain the size and various other aspects of the spread in terms of the following four categories of costs faced by a dealer (see, e.g., Stoll 2003). *Order processing costs* are the direct costs for the dealer’s operations and handling of
transactions and include capital and labor costs as well as possible profits in terms of monopoly or cartel rents. These costs include both variable and fixed parts.

*Inventory holding costs* arise because dealers are mandated to post two-sided quotes continuously, and in doing this, they will sometimes have to take on risky inventory, which might make them deviate from their optimal portfolio allocations.

In quoting bids and asks, a dealer risks trading with investors who possess superior information about the value of the security. Such informed investors, unidentifiable for the market maker, will only trade when the quoted prices are in their favor. The dealer will thus lose to informed investors and thereby incur *adverse selection* or *information costs*.

A dealer who places a limit order bid (ask) has written a put (call) *option* to the market. If new public information arrives that justifies a higher (lower) price, the call (put) will be exercised by someone in the market unless the ask (bid) is timely adjusted by the market maker. However, the options are not exactly free, because the bid-ask spread will reflect the values or premiums of the options, where the exercise prices are the bid and the ask.

Taken together, in order to survive, a dealer must quote wide enough spreads so that over time the compensation earned is sufficient to cover the costs incurred.

The foregoing exposition was largely based on theory for quote-driven markets and developed from the perspective of a dealer. Few real-world trading systems are pure quote- or order-driven. NASDAQ is basically a dealer market, but has features of an auction market in that (since 1997) customer orders are displayed and may execute directly against each other. NYSE is a mixed auction/dealer market, because customer orders are centralized and executed if matched; however, the NYSE specialists trade for their own account to maintain liquidity in their assigned stocks (Stoll 2003).

As noted, most microstructure theory concerns quote-driven markets structures with designated market makers or dealers acting under varying degrees of competition. The extent to which these theories are descriptively valid for order-driven automated limit order markets is an open question. Agents in order-driven markets certainly do, however, share, to some extent, features and considerations with market makers in the quote-driven markets. One important difference may be that on the Stockholm Stock
Exchange there are no agents contracted to provide immediacy by continuously quoting two-sides. Instead, immediacy is provided by submitted limit orders. The resulting quoted bid and ask price impact functions (the consolidated quoted prices and quantities) should reflect the aggregated adverse selection risks, order processing, inventory holding, and option costs of the agents who submitted the limit orders. How that aggregation works dynamically for agents with possibly heterogeneous beliefs, preferences and wealth is, of course, a complex matter.

Empirical evidence from studies that decompose the bid-ask spread, using quote and trade data, into order processing, adverse selection, and inventory holding costs, corroborates the existence of at least the two former costs even in limit order markets. de Jong et al. (1996) implement an empirical variant of Glosten (1994), extended to include order processing costs, and find for ten stocks on the Paris Bourse (CAC) statistically significant adverse selection and order cost components. For small transactions, order costs constitute 70% and adverse selection 30% of the small volume quoted spread, while for large transactions the figures are 55% and 45% of the large volume spread, respectively. In addition, they find that price revisions (or permanent price impacts) after large trades are substantially bigger than after small trades (60% versus 25% of the respective quoted spread). Complementary results suggest that the permanent price impacts may be larger, ranging from 41% to 115%, and that inventory holding costs are negligible.

Based on the model (which neglects inventory costs) in Lin et al. (1995), Brockman and Chung (1999) estimate order processing and adverse selection components to 45% and 33%, respectively, of the quoted bid-ask spread for stocks traded on the Stock Exchange of Hong Kong (SEHK). The order cost component as a proportion of the spread is invariant across mean daily dollar trade volume deciles, while the adverse selection component is inversely related to volume. The magnitude of the relative spread increases with decreasing volume, and this implies that the order cost component fraction of the relative spread increases as volume decreases, and that the adverse selection component fraction of the relative spread increases even faster as volume decreases.
Chan (2000) uses an adaptation of a model in Glosten and Harris (1988), and finds empirically for stocks on SEHK that the adverse information component dominates the inventory holding component controlling for trade size, with both components being statistically significant.\footnote{In the empirical analysis, Glosten and Harris (1988, p. 135) actually assume away inventory costs, maintaining order costs.}

For a sample of stocks listed on the Tokyo Stock Exchange, Ahn et al. (2002) by means of the approach of Madhavan et al. (1997), find that the adverse selection component of the implied bid-ask spread increases with trade size, while the combined order processing and inventory holding cost component falls with trade size (suggesting indirectly that order processing costs dominate inventory holding costs). The adverse selection spread component as a fraction of stock price is larger for low priced, small cap stocks than for higher price, large cap stocks. The same holds true for the order cost component. However, the order component is larger than the adverse selection component for low price, small cap stocks, but for high price, large cap stocks the opposite is true, for the adverse selection component declines faster as stock price increases.

Overall, empirical results of spread decomposition studies seem to support the existence of order processing and adverse selection costs increasing in trade size, while the evidence for inventory holding costs is less convincing.\footnote{The situation appear to be similar for quote-driven markets, with the possible exception of Huang and Stoll (1997) and George et al. (1991) who find substantial inventory costs and small information costs, see, e.g., Coughenour and Shastri (1999).}

\textit{Variable selection}

As mentioned, \textit{price impact cost} is measured as in, e.g., Chan and Lakonishok (1997) and Keim and Madhavan (1997), as the absolute volume-weighted average price degradation relative to the quote midpoint (at the time the order book is observed). In computing the price impact, full immediacy is assumed, i.e., every transaction is treated as if it were a market order. The whole price impact is regarded as temporary, that is, recognizable as a cost. This is consistent with uninformed trading for which there should be no permanent price effects. Price impact cost, measured this way,
however, includes adverse information as well as other costs as defined by the prices and quantities in the order book, and as assessed by the market in the form of the submitted limit orders. It is probably not possible for limit order submitters to know exactly when uninformed trading will take place, because trading is largely anonymous and follows strict priority rules. Limit order traders may however weigh in the probability that such trading is higher in association with index reconstitutions and similar events, and adjust their limit orders accordingly; but submitted limit orders are still likely to contain some compensation for informed trading risk.

Order size. Transaction costs are for several reasons expected to increase with order or transaction size. In a limit order book, the quoted (volume-weighted average) price impact cost is constant for the best levels and a strictly increasing function of order size beyond the best levels by construction. As was mentioned earlier, the quoted price impact function in a limit order book is a piecewise concave function, and whose shape can vary markedly depending on its parameters.

Stoll (1978a, p. 1141) shows theoretically that a dealer’s inventory holding cost in dollars (in percentage per share in order) rises quadratically (linearly) with transaction size in dollars.

Theoretical adverse information-based models that offer an integrated treatment of price impact and order size, where order size is a continuous variable, include Kyle (1985) and Glosten (1994). In Glosten (1994), the quoted (bid and ask) price impact functions are inherent in the liquidity supply schedule, and by construction increasing in order size. Kyle (1985) models price change (or impact) in monetary units, due to adverse information, as a positive linear function of transaction size $q$ with slope $\lambda$ and zero intercept.\(^{25}\)

According to microstructure theory, the order cost component (including possibly both fixed and variable parts) of the price impact cost as a function of order size could either increase, be constant, or decrease. There is thus no clear relation, except for very large orders, which are likely to be increasingly expensive.

\(^{25}\) The inverse of the slope, $1/\lambda$, is termed depth in Kyle (1985). The volume-weighted price impact cost is linear in $q$ and equal to $q\lambda/(2 \text{MIDP})$. 
In Sweden, commissions are often charged as a fixed percentage of order value. There is probably a certain amount of fixed costs, independent of order size, associated with each order. For institutional portfolios it seems reasonable to assume that the fixed costs per order are small relative to the value of typical order sizes. Then, given linear percentage inventory and information cost functions, the percentage price impact cost would be a (weakly) concave function of order size, which asymptotically becomes linear, as the fixed costs vanish relative to the monetary order value. Under these circumstances, a linear function should provide a good approximation to true percentage price impact costs.

Relatively few empirical studies explicitly consider quoted or effective price impacts (or spreads) as functions of order or transaction size measured as a continuous variable. (As noted, this modeling is prevented in markets where only the best bid and ask and accompanying volumes are publicly displayed, because the quoted price impact cost at a given time does not change with order volume.) Based on data from actual institutional trading, Chan and Lakonishok (1997) and Keim and Madhavan (1997) find that temporary price impact costs increase with trade size. Loeb’s (1983, 1991) empirical analysis of block trading finds permanent price impact costs positively related to order size. A carefully designed study by Griffiths et al. (2000) establishes in an order-driven setting a positive relation between price impact and order size for non-passive orders using a highly detailed database that allows them, among other things, to reconstruct orders with multiple fills. Studies based on trade and quote data (as opposed to trading record data or data with extensive order flow information), using a variety of approaches, find empirical evidence of various measures of price impact being positively related to different measures of transaction size (see, in addition to the above-mentioned studies by de Jong et al. 1996 and Ahn et al. 2002, e.g., Bessembinder and Kaufman 1997, who find in regressions that average temporary price impact increases with average trade size, Glosten and Harris 1988, Brennan and Subramanyam 1995, and Hasbrouck 1991).

It is clear that the quoted bid and ask price impact functions for a given stock, in the limit order book are increasing functions of order size. Interesting empirical questions concerning the relation between price impact and order size, are whether bid
and ask price impacts for a given stock are equally sensitive to order size and whether the price impact-order size relation varies across stocks.

Following Keim and Madhavan (1997) and Chan and Lakonishok (1997), I will use, as the unit of measurement for order size, the number of shares in the order relative to all shares outstanding. In Chan and Lakonishok (1997), where buyer and seller-initiated trades are analyzed together, the coefficient $\beta$ on order size is positive, and constant across stocks. Chan and Lakonishok (1995), however, document purchases as being more sensitive to order size than sales. Keim and Madhavan (1997) analyze separately buyer and seller-initiated trades and find that the coefficient on order size is positive and constant across stocks, but that it is higher for seller-initiated than for buyer-initiated trades. For a given stock, a buy order of a particular size is, however, usually more expensive than a sell order of the same size. Keim and Madhavan argue that buy orders are more likely than sells to be associated with information, and that liquidity providers in quoting asks therefore include a larger compensation for adverse selection risk than for bids (also see Griffiths et al. 2000, p. 81; Burdett and O’Hara 1987). In contrast to Keim and Madhavan (1997) and Chan and Lakonishok (1995), who only study bullish markets, Chiyachantana et al. (2004) study both rising and falling markets. They find empirical evidence that buy price impact costs are larger than sell impact costs in a rising market, but that the relation is reversed in falling markets.

A priori, the constancy of the coefficient on order size across stocks seems somewhat unmotivated; maybe it is a result of an attempt to keep the modeling simple. Note, however, that this constancy implies, all else equal, that if order size were restated in monetary units the corresponding coefficient would be proportionally decreasing in market value, that is, a stock with a $g$ times larger market value than another stock, would have a $g$ times smaller coefficient. Furthermore, accumulated order quantity in the order book, perhaps best expressed as a proportion of shares outstanding, could also, in itself, be considered a measure of liquidity. This might, as will be discussed later, introduce a selection bias when price impact costs for large volumes are estimated cross-sectionally, since there is risk that such observations exclusively represent stocks whose liquidity is above average.
Market capitalization (the stock price times the number of all outstanding shares in issue) is taken to be a relative measure of liquidity, i.e., how expensive it is to trade the share. The larger a firm is, the cheaper their shares are to trade. The reason for this may be that large firms are more actively traded, they are followed by a higher number of market participants and more information is generated about them (see, e.g., Harris 1994; Keim and Madhavan 1998). Traders thus know that the likelihood of finding a counterpart is relatively high. It is also probable that the competition in liquidity provision is greater the greater a stock’s market capitalization is. Adverse selection and inventory holding costs as well as economic rents are thus likely to be inversely related to market capitalization, and so are also price impact costs. (Market capitalization is in addition empirically related to other potentially important factors including monetary trading activity, stock price level, and stock return volatility.) Usually, in the typical OLS setting, the variable undergoes a logarithmic transformation. According to Keim and Madhavan (1998), many studies identify market capitalization (and order size) as the most important factor(s) in explaining trading costs. Price impact costs are anticipated to be decreasing in market capitalization.

The level of trading activity in a stock is for several reasons expected to be inversely related to the price impact cost of the stock. Demsetz (1968, p. 41) argue that since waiting and inventory holding costs are expected to decrease as trading activity increases, “The fundamental force working to reduce the spread is the time rate of transactions”, and he uses empirically the number of transactions per day to explain the ask-bid spread. In the operationalization of the Stoll (1978a) model of inventory holding costs, Stoll (1978b) models spread size as decreasing in historical total dollar-trading volume (over 64 days).

In addition to the trading volume dimension of trading activity, Tinic and West (1974) document spreads to be inversely related to “trading continuity” (p. 733) measured as the number of days with trading divided by the total number of days in their sample. In an order-driven auction market setting, Cohen et al. (1981, p. 300) derive a model under which “Thinner securities will, ceteris paribus, have larger equilibrium market spreads.” For a limit order market, Foucault (1999, p. 117-118)
establishes theoretically “a negative relationship between the spread and transaction frequency.” Copeland and Galai (1983, p. 1463) argue that the probability of informed trading (and thus adverse information risk) is likely to be higher for thinly traded stocks, presumably closely held, and, therefore, there should be a negative correlation between spread and trading volume, holding the transaction size constant. Easley et al. (1996) find empirically that stocks with higher trading volumes, have higher arrival rates of both informed and uninformed traders, as well as higher probability of information events occurring than have stocks with lower trading volumes. However, the interaction of arrival rates of informed respective uninformed traders and information event probabilities is (in terms of their theoretical modeling) such that higher volume stocks actually are associated with less risk of information-based trading or, equivalently, have lower adverse selection costs than do stocks with lower trading volumes. (Using a variant of the Easley et al. 1996 framework, Grammig et al. 2001 document that informed traders select to trade on an anonymous limit order market rather than on a non-anonymous, floor-based, specialist-operated exchange; that estimated probabilities of informed trading are positively related to the bid–ask spread and, particularly, to the adverse selection component of the spread; and that cross-sectional differences in the probability of informed trading on the respective markets are directly related to differences in the adverse selection component.)

In light of the above notions, I introduce three explanatory variables intended to reflect various aspects of (historical) trading activity. The first variable, $TOTTRVOL$, measures total trading volume in monetary units over the 22 preceding trading days (approximately one month). The second, $PROPTR$, is a trading-continuity variable measuring the proportion of days with non-zero trading over the preceding 22 trading days.

The third variable, $TRVOL$, is computed as the average of the daily turnover rate the preceding 22 trading days, where the daily turnover is equal to the traded volume in number of shares as a proportion of all outstanding shares. Similarly defined turnover rates have been used as measures of adverse selection costs, hypothesized and empirically documented as positively related to spreads (Stoll 1978b; Tripathy and Peterson 1991). The argument for this is that trading will be large relative to the
number of shares outstanding in stocks where certain investors believe they have information that others do not (Stoll 1978b, p. 1170). In contrast, Chan (2002) in empirical analysis of an open order-driven market, found a negative relation between the turnover rate and percentage price impact cost.26

In addition to reflecting the risk of adverse information, the historical turnover rate may proxy several other dimensions that should be positively related to liquidity. For instance, a higher turnover rate may indicate higher floatation or more owners. The former represent a larger supply of stocks, which should, all else equal, make the stock more liquid and attractive for various investors. Demsetz (1968) argues (and observes empirically) that the transaction rate should be (is) approximately proportional to the number of owners.

Whether observed price impact costs increase or decrease with the turnover rate appears to be an empirical question. In addition, variables defined similarly are often used by stock exchanges and organizations that manage stock market indexes to rank stocks after liquidity. It should be expected that price impact costs are negatively related to the first two measures of historical trading activity.

The stock price level, often measured as the quote midpoint, is potentially related to liquidity in several ways (Keim and Madhavan 1998). One argument is arbitrage-based: “Were all factors other than price per share equal, traders would use limit orders to equalize the spread per dollar regardless of the price per share. Consequently, spreads would be proportional to the per share price.” (Benston and Hagerman 1974, p. 355; also see Demsetz 1968).

Another argument for a relation between stock price and transaction costs concerns the minimum tick size (Branch and Freed 1977, also see Stoll 1978b and Harris 1994). The minimum tick size is the smallest increment for which stock prices can be quoted. If the minimum tick size is constant for all stock price levels, the higher the stock price level, the smaller the minimal relative price changes will be, implying gentler price impacts for shares with higher prices. The minimum tick size in the SSE limit order trading-system is, however, not constant (see Table 2.1), and a monotone effect on

---

26 Chan (2000, p. 7) does not offer any motivation for this relation, apart from two seemingly invalid references, Glosten and Harris (1998) and Brennan and Subramanyam (1995), which neither uses turnover rate.
price impacts from the tick size via the stock price level is therefore expected only for stock prices in excess of SEK 500.\textsuperscript{27} In addition, price has been documented as related to other relevant explanatory variables. At most, a weak negative relation is expected between price and price impact costs.

Stock return volatility has in the literature been documented as positively related to transaction costs. Holding other factors constant, Stoll (1978a), among others, argues and shows empirically that higher asset risk measured as stock return volatility increases inventory management costs and as a consequence price impact costs. I adopt the arguments which Stoll (1978a) and Stoll and Ho (1980, p. 261, fn. 1) provide for using - as a measure of the risk relevant to holding inventory - total risk, i.e., return volatility, rather than systematic risk or unsystematic risk. Empirical support for this view is provided by Stoll (1978b) and Menya and Paudyal (1996).

Adverse information models and option theory-based arguments also imply that spreads are positively related to return volatility (e.g., Easley and O’Hara 1987, p. 81; Copeland and Galai 1983). On intuitive grounds, it is reasonable to expect that riskier stocks, whose fundamental values are more uncertain, to a higher degree are subject to adverse information risk. Moreover, Foucault’s (1999, p. 101) theoretical limit order book model suggests “that posted spreads are positively related to asset volatility.” Price impact costs are thus expected to increase with stock return volatility.

### 2.5 Data

During the period of this study, Ecovision provided limit order book data for shares trading on the SSE’s A-list, O-list, New Market and Aktietorget. Market capitalizations were retrieved from the SIX Trust system. The source for stock quote data is Affärsvärdens archive of stock prices.

At twelve arbitrary points in time during 22 Aug. 2000 to 12 Sept. 2000, between 10.30 and 16.30, that is one hour after the opening and one hour before the close, respectively, the limit orderbooks for all Swedish equities in Ecovision were sampled. This generated almost 5000 order book "snapshots". There are several reasons why I

\textsuperscript{27} The regressions were estimated separately for each price interval with constant tick size, but the coefficient on price did not achieve statistical significance.
preferred not to use data from the first and last hour. The order books of many less
frequently traded shares do not become populated until after the first hour and for
other shares there might be effects remaining from the batch auctions at the opening.

In essence, each order book snapshot has a firm identifier, a timestamp, and the five
best asks and bids and the corresponding volumes. The order book snapshots were
augmented with market capitalizations, computed as closing best bid times total
number of outstanding shares, and closing best bid prices.

The order book snapshots were scrutinized for anomalies such as non-decreasing
bid price levels and non-increasing ask price levels, unpopulated order books, equal
best bid and best ask, price levels without corresponding volumes and vice versa,
missing market capitalizations and closing best bid prices (the different databases do
not cover exactly the same firms). In total, some 450 snapshots out of approximately
5000 were deleted as a result of the scrutiny.

For each side of every orderbook, separate price impact functions were computed.
Each price impact cost function was augmented with the following explanatory
variables: \( \text{MIDP} \), the quote midpoint, computed as the average of the best bid and ask;
\( \text{MC} \), the market capitalization equal to \( \text{MIDP} \) times the number of outstanding shares;
\( \text{TRVOL} \) is the average of the daily traded volume as a proportion of outstanding shares
over the preceding 22 days; \( \text{PROPTR} \) is the proportion of days with non-zero traded
volume during the previous 22 days; \( \text{VOLR} \), the annualized standard deviation in daily
logarithmic (closing bid) price relatives estimated according to Campell et al. (1997,
Eq. 9.3.28), assuming a continuous-time price process using all available observations
Table 2.3 Summary statistics across order books
This table shows summary statistics across order books for the following variables: Inside half-spread = Min. price impact cost = |Best ask (bid) - MIDP| / MIDP; Max. price impact cost = Avg. price impact cost at max ORDERSIZE; ORDERSIZE = Quoted order size (1/1000 of the total no. of outstanding shares); MC = Market capitalization equal to MIDP times no. of outstanding shares; MIDP = quote midpoint, i.e., the average of the best ask price and the best bid price; TRVOL = the average, over the previous 22 days, of the daily traded volume as a proportion of all outstanding shares; PROPTR = proportion of days with trading the previous 22 days; and VOLR = annualized return volatility using all available observations during 21 Jul. 2000-12 Sep. 2000.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std</th>
<th>Min</th>
<th>25%</th>
<th>Median</th>
<th>75%</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A. Asks (n = 3920)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inside half-spread (%)</td>
<td>1.02</td>
<td>1.23</td>
<td>0.05</td>
<td>0.32</td>
<td>0.60</td>
<td>1.20</td>
<td>10.71</td>
</tr>
<tr>
<td>Max. price impact cost (%)</td>
<td>3.19</td>
<td>2.31</td>
<td>0.14</td>
<td>1.44</td>
<td>2.48</td>
<td>4.38</td>
<td>11.00</td>
</tr>
<tr>
<td>Max. ORDERSIZE</td>
<td>0.64</td>
<td>0.68</td>
<td>0.004</td>
<td>0.18</td>
<td>0.42</td>
<td>0.87</td>
<td>5.00</td>
</tr>
<tr>
<td>MC (billion SEK)</td>
<td>13.34</td>
<td>80.17</td>
<td>0.01</td>
<td>0.36</td>
<td>1.12</td>
<td>5.11</td>
<td>1449</td>
</tr>
<tr>
<td>MIDP (SEK)</td>
<td>108.3</td>
<td>120.5</td>
<td>0.385</td>
<td>35.2</td>
<td>76.375</td>
<td>136.75</td>
<td>1092</td>
</tr>
<tr>
<td>TRVOL *10000</td>
<td>25.37</td>
<td>32.92</td>
<td>0.07</td>
<td>6.17</td>
<td>15.56</td>
<td>29.90</td>
<td>294.21</td>
</tr>
<tr>
<td>PROPTR</td>
<td>0.94</td>
<td>0.14</td>
<td>0.05</td>
<td>0.95</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>VOLR</td>
<td>0.58</td>
<td>0.39</td>
<td>0.05</td>
<td>0.28</td>
<td>0.48</td>
<td>0.79</td>
<td>2.83</td>
</tr>
<tr>
<td>Panel B. Bids (n = 3843)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inside half-spread (%)</td>
<td>1.03</td>
<td>1.18</td>
<td>0.05</td>
<td>0.33</td>
<td>0.61</td>
<td>1.27</td>
<td>9.27</td>
</tr>
<tr>
<td>Max. price impact cost (%)</td>
<td>2.88</td>
<td>2.00</td>
<td>0.17</td>
<td>1.42</td>
<td>2.27</td>
<td>3.91</td>
<td>9.34</td>
</tr>
<tr>
<td>Max. ORDERSIZE</td>
<td>0.57</td>
<td>0.66</td>
<td>0.003</td>
<td>0.14</td>
<td>0.33</td>
<td>0.74</td>
<td>5.00</td>
</tr>
<tr>
<td>MC (billion SEK)</td>
<td>13.36</td>
<td>80.79</td>
<td>0.01</td>
<td>0.36</td>
<td>1.16</td>
<td>5.11</td>
<td>1449</td>
</tr>
<tr>
<td>MIDP (SEK)</td>
<td>109.2</td>
<td>123.2</td>
<td>0.385</td>
<td>35.5</td>
<td>76.5</td>
<td>136.5</td>
<td>1092</td>
</tr>
<tr>
<td>TRVOL *10000</td>
<td>24.97</td>
<td>32.58</td>
<td>0.04</td>
<td>5.76</td>
<td>15.09</td>
<td>29.88</td>
<td>294.21</td>
</tr>
<tr>
<td>PROPTR</td>
<td>0.94</td>
<td>0.15</td>
<td>0.05</td>
<td>0.95</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>VOLR</td>
<td>0.58</td>
<td>0.39</td>
<td>0.05</td>
<td>0.28</td>
<td>0.48</td>
<td>0.79</td>
<td>2.83</td>
</tr>
</tbody>
</table>

To mitigate the effects from unrepresentative or erroneous order book observations, ask snapshots (bid snapshots), where the maximum price impact cost is at or above the 90th percentile (90th percentile), or where the maximum order size is greater than 5/1000 of all shares outstanding, are deleted. This reduced the number of ask (bid) price impact snapshots to 3920 (3843). Table 2.3 contains summary statistics of the data on the order book-level. Securities with extremely large volumes in their order books are at one hand desirable to include in the analysis, since they are informative about price impact costs for large volumes, but their inclusion poses, however, at the same time a risk in that they might be unrepresentatively liquid; and basing price impact estimates for large order sizes on such securities could introduce downward bias.
It is clear from Table 2.3 that the data exhibit large variation in most dimensions. The average inside half-spread, for instance, is around 1% for both asks and bids, while the respective maxima are 10.7% and 9.3%.

For the purpose of parsimony, and to reduce potential multicollinearity problems in the regressions, the trading activity variable \( TOTTRVOL \) is dropped. The correlation between the logarithm of that variable and the logarithm of \( MC \) is 0.84, which is high. Moreover, the correlations between \( TOTTRVOL \) and the other variables (not reported) are similar to the correlations between \( MC \) and the other variables. In choosing between the two variables, \( MC \) was kept, because it is easier to measure and does not require time-series data.

### 2.6 Empirical analysis

It is probably a reasonable approximation to consider the price impact cost function as defined by the limit order book, as being continuous in order size.\(^{28}\) This implies that the function has (approximately) infinitely many points along the order size dimension. For each such price impact function there is a set of explanatory variables, i.e., market capitalization, return volatility, and others. The objective here is to find a model that will allow us to predict price impact cost as a function of order size and the other explanatory variables. There are several possible ways to do this. One is to approximate each price impact function by fitting a simpler (polynomial) function (of reasonable order) with order size as independent variable to it. The next step would be to model and estimate the coefficients of the approximating function in terms of the explanatory variables. Since the price impact functions have order size domains of widely differing numerical ranges this will probably be problematic.

Another option is to select a limited number of points along the price impact function. One may for instance select one random point from each price impact function (representing the price impact cost of a hypothetical trade of the random size), and append the explanatory variables. This is repeated for all price impact functions. Then it is possible to fit a regression model to the pooled data, separately for asks and bids.

---

\(^{28}\) Note that order size, strictly taken, is a discrete variable.
Using only one point from each price impact function is a simple approach, but the information in the price impact cost functions could probably be utilized more efficiently by using more points. Therefore, I decided to use 10 points from each price impact function. The 10 pairwise observations of price impact cost and order size were selected equidistantly along the order size dimension. Selecting 10 points from each price impact function gives every order book observation the same weight. Price impact functions whose maximum order size exceeds \(3/1000\) of all outstanding shares were truncated at \(3/1000\); and the 10 observations were selected equidistantly within this range. The reason for this is twofold. First, it was deemed unlikely that a need to predict price impacts outside this order size would ever arise (in the portfolio optimization applications). Second, the few observations outside this boundary come from a very limited number of order books, and those few observations might affect the estimated regressions unjustly. The price impact costs were expressed in percentages and the order sizes in \(1/1000\) of the outstanding number of shares in the issue. Appended to each such pairwise observation were the corresponding explanatory variables.

When a regression model is fit, separately for asks and bids, by OLS to all the observations obtained when each order book generates 10 observations, the residuals will exhibit autocorrelation up to a lag of roughly the same number as the number of points selected from each price impact function, i.e., 10. The reason is that the price impact cost for a given stock is an increasing function of order size. In the presence of autocorrelation, the usual OLS standard error estimates will be biased, and this generally makes the customary significance tests inappropriate.

To account for the non-zero autocorrelations, and for possible heteroscedasticity due to the heterogeneous sample of securities, the Newey and West (1987) heteroscedasticity and autocorrelation consistent covariance estimator is applied. This estimator requires a lag truncation parameter that specifies the number of autocorrelations to consider. West and Newey (1994) suggest, given a sample size \(n\), that the lag truncation parameter be computed as \(4(n/100)^{2/9}\), rounded downwards to the nearest integer. Given the sample sizes for asks and bids of around 40000 observations each based on 10 points from each side of the order book, the value of the
lag truncation parameter is 15.\textsuperscript{29} This means that it is likely that all non-zero autocorrelations are accounted for by the estimator; selecting for instance 20 points from each price impact function decreases that likelihood, since the lag truncation parameter then becomes 17.

In Table 2.4 average price impact costs are tabulated by market capitalization and order size quintiles. In computing a cell mean, no more than a single randomly selected observation from any orderbook (eligible for the cell) is used. This is believed to provide a more representative weighting than if all observations of a given orderbook, eligible for the cell, are used. In particular, the cells representing the largest order sizes risk being dominated by observations from the few order books whose order size domains stretch out far enough, and these order books probably represent securities that are very liquid (in the respect that large quantities are quoted and possible to transact). The cell average price impact costs might therefore be biased downwards, if they are based on an unjustifiably high proportion of observations from the most liquid stocks.

Table 2.4 should provide a glimpse of the magnitude and variation of price impacts over two important dimensions. The quintile breakpoints for the market capitalization dimension are virtually the same for asks and bids. For the order size dimension, the breakpoints differ between Panel A and B. This reflects varying patterns of total depths for asks and bids. In both Panel A and B, price impacts are almost monotonically increasing in trade size and decreasing in market capitalization. They also appear significant both economically and statistically. Inventory and information cost theories are both consistent with price impact costs increasing in order size. Asymmetric information problems are likely to increase as market capitalization decreases. With some exceptions, for a given market capitalization quintile and order size quintile, price impact costs for the ask side appear larger than for the bid side, and the difference seems to increase with order size and decrease with market capitalization. The comparison is rough, because it controls for just two factors, and the quintile breakpoints are not identical for asks and bids.

\textsuperscript{29}Campbell, Lo and MacKinlay (1997, p. 535) actually recommend using a lag truncation parameter that exceeds the number of non-zero autocorrelations.
Table 2.4 Price impact costs by market capitalization and order size
Percentage price impact costs by market capitalization (billion SEK) and order size (1/1000 of outstanding shares) quintiles. From each side of every order book price impact functions were computed. From every such price impact function, 10 pairwise observations of price impact and order size are selected equidistantly along order size; price impact functions whose order size domain exceeded 3 were cut-off at 3. Quintiles are determined separately for asks and bids. In each cell, first row is mean price impact and second row standard error. In computing a cell mean, no more than a single randomly selected observation from any order book is used.

Panel A. Asks

<table>
<thead>
<tr>
<th>Order size quintiles</th>
<th>0.01-0.29</th>
<th>0.29-0.64</th>
<th>0.64-2.37</th>
<th>2.37-6.71</th>
<th>6.71-1449</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.000-0.032</td>
<td>2.12</td>
<td>1.05</td>
<td>0.92</td>
<td>0.97</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td>0.066</td>
<td>0.026</td>
<td>0.036</td>
<td>0.043</td>
<td>0.032</td>
</tr>
<tr>
<td>0.032-0.103</td>
<td>2.81</td>
<td>1.40</td>
<td>1.44</td>
<td>1.69</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td>0.118</td>
<td>0.039</td>
<td>0.060</td>
<td>0.062</td>
<td>0.039</td>
</tr>
<tr>
<td>0.103-0.227</td>
<td>2.61</td>
<td>1.87</td>
<td>1.71</td>
<td>1.83</td>
<td>1.08</td>
</tr>
<tr>
<td></td>
<td>0.075</td>
<td>0.050</td>
<td>0.057</td>
<td>0.061</td>
<td>0.038</td>
</tr>
<tr>
<td>0.227-0.500</td>
<td>3.25</td>
<td>2.70</td>
<td>2.08</td>
<td>1.90</td>
<td>1.21</td>
</tr>
<tr>
<td></td>
<td>0.079</td>
<td>0.064</td>
<td>0.064</td>
<td>0.072</td>
<td>0.050</td>
</tr>
<tr>
<td>0.500-3.000</td>
<td>5.29</td>
<td>3.56</td>
<td>2.39</td>
<td>1.99</td>
<td>1.28</td>
</tr>
<tr>
<td></td>
<td>0.091</td>
<td>0.082</td>
<td>0.077</td>
<td>0.137</td>
<td>0.075</td>
</tr>
</tbody>
</table>

Panel B. Bids

<table>
<thead>
<tr>
<th>Order size quintiles</th>
<th>0.01-0.28</th>
<th>0.28-0.65</th>
<th>0.65-2.42</th>
<th>2.42-6.76</th>
<th>6.76-1449</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.000-0.026</td>
<td>2.04</td>
<td>1.06</td>
<td>0.96</td>
<td>0.98</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td>0.06</td>
<td>0.03</td>
<td>0.04</td>
<td>0.04</td>
<td>0.03</td>
</tr>
<tr>
<td>0.026-0.084</td>
<td>2.35</td>
<td>1.40</td>
<td>1.36</td>
<td>1.50</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>0.12</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.03</td>
</tr>
<tr>
<td>0.084-0.190</td>
<td>2.34</td>
<td>1.67</td>
<td>1.61</td>
<td>1.70</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td>0.07</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.03</td>
</tr>
<tr>
<td>0.190-0.433</td>
<td>2.84</td>
<td>2.42</td>
<td>2.08</td>
<td>1.78</td>
<td>1.07</td>
</tr>
<tr>
<td></td>
<td>0.07</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
<td>0.05</td>
</tr>
<tr>
<td>0.433-3.000</td>
<td>4.56</td>
<td>3.16</td>
<td>2.46</td>
<td>1.70</td>
<td>1.14</td>
</tr>
<tr>
<td></td>
<td>0.08</td>
<td>0.07</td>
<td>0.08</td>
<td>0.12</td>
<td>0.07</td>
</tr>
</tbody>
</table>
2.6.1 Regression results

To assess how price impact costs are jointly affected by market capitalization and order size and other explanatory variables, regression analysis is performed. The results are presented in Table 2.5. Other specifications and transformations were tested but the ones in Table 2.5 provided good balance between parsimony, explanatory power, and out-of-sample performance.30

Table 2.5 Price impact regressions

Price impact regressions by OLS for asks and bids separately over all observations. \( \pi = \) average percentage price impact; \( LNMC = \) natural log. of market capitalization in SEK; \( RTRVOL = \) square root of the average daily traded volume as a proportion of outstanding shares over the previous 22 days; \( PROPTR = \) proportion of days with trading previous 22 days; \( LNVOLR = \) log. of annualized return volatility over the 22 prev. days; \( LNMIDP = \) log. of quote midpoint (SEK); \( ORDERSIZE = \) no. of shares in order as 1/1000 of outstanding shares; \( \text{Prob. of F-test} = \) the p-value for the F-test31 of the overall significance of the regression (based on the usual OLS estimate of residual variance). Adjusted \( R^2 \) are also reported (based on the usual OLS estimate of residual variance). t-values are computed using Newey-West (1987) heteroscedastic autocorrelation consistent standard errors, with lag truncation parameter = \( 4(n/100)^{2/9} \) rounded downwards to the nearest integer.

Panel A. Asks \((n = 39200)\)

<table>
<thead>
<tr>
<th>( \gamma_1 )</th>
<th>( \gamma_2 )</th>
<th>( \gamma_3 )</th>
<th>( \gamma_4 )</th>
<th>( \gamma_5 )</th>
<th>( \beta )</th>
<th>Adj. ( R^2 )</th>
<th>Prob. of F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
<td>8.82</td>
<td>-0.128</td>
<td>-19.4</td>
<td>-3.25</td>
<td>0.796</td>
<td>1.20</td>
<td>0.42</td>
</tr>
<tr>
<td>t-value</td>
<td>34.2</td>
<td>-11.7</td>
<td>-25.3</td>
<td>-13.8</td>
<td>-22.2</td>
<td>23.0</td>
<td></td>
</tr>
</tbody>
</table>

Panel B. Bids \((n = 38430)\)

<table>
<thead>
<tr>
<th>( \gamma_1 )</th>
<th>( \gamma_2 )</th>
<th>( \gamma_3 )</th>
<th>( \gamma_4 )</th>
<th>( \gamma_5 )</th>
<th>( \gamma_6 )</th>
<th>( \beta )</th>
<th>Adj. ( R^2 )</th>
<th>Prob. of F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
<td>7.47</td>
<td>-0.111</td>
<td>-17.2</td>
<td>-2.14</td>
<td>0.720</td>
<td>-0.0727</td>
<td>1.02</td>
<td>0.41</td>
</tr>
<tr>
<td>t-value</td>
<td>33.6</td>
<td>-9.60</td>
<td>-23.8</td>
<td>-11.2</td>
<td>20.9</td>
<td>-3.67</td>
<td>18.9</td>
<td></td>
</tr>
</tbody>
</table>

Overall, the regression results are consistent with microstructure theory. The regressions exhibit highly significant F-values and most coefficients have expected signs and highly significant autocorrelation-heteroscedasticity adjusted t-values. The adjusted coefficients of determination for the regressions are slightly above 40%. The ones obtained by Keim and Madhavan (1997) were below 10%. The findings common to both sides of the orderbook are that price impact costs decrease with market

---

30 For instance, a specification was tested, where order size was given in monetary units, but to no avail.

31 This F-statistic is the test statistic of the F-test that all the slope coefficients are zero, in a linear regression model with an intercept.
capitalization and the two measures of historical trading activity, and increase with order size and stock return volatility. The negative coefficient on \( RTRVOL \) is consistent with the empirical findings of Chan (2000) in a limit order market, but inconsistent with, e.g., Stoll (1978b). The coefficient on quote midpoint is negative and statistically significant for bids but not for asks. That these two variables would behave in this fashion was to some extent expected as discussed above.

Table 2.4 indicates that there is a tendency for price impact costs to be greater for the ask side than for the bid side, and that the difference increases with order size and decreases with market capitalization. This pattern is consistent with the estimated regressions in Table 2.5,\(^{32}\) in that the coefficient on order size is substantially larger for the ask side than for the bid side, and that the coefficient on market capitalization is more negative for asks than for bids. The notion that price impact costs for the ask side, i.e., for market buy orders, are higher and increase faster with order size than for bids, i.e., for market sell orders, is supported by the buy and sell price impact cost functions for the “average”\(^{33}\) stock, which are \( \pi = 1.81 + 1.20 \text{ORDERSIZE} \) and \( \pi = 1.28 + 1.02 \text{ORDERSIZE} \), respectively. As noted, this is consistent with, e.g., Keim and Madhavan (1997) and Chan and Lakonishok (1995).\(^{34}\) One explanation would be that submitters of limit orders consider it more risky to provide liquidity on the ask side than on the bid side, and that they therefore require higher compensation for exposing their limit orders to incoming buy orders than to incoming sell orders.

The results of Chiyachantana et al. (2004) are, however, suggestive of that the price impact asymmetry could be explained by the direction of the market. No attempt was made to control for market direction in the price impact regressions. One reason is that the order book snapshot data did not include any information on market index performance.

\(^{32}\) Considering the similar magnitudes of the other coefficients for bids and asks, respectively.

\(^{33}\) These were obtained by taking the average (of the means) of the explanatory variables across Panel A and B in Table 2.3.

\(^{34}\) In Keim and Madhavan (1997), the coefficient on order size is actually greater for sells than for buys (disregarding the value of other variables and coefficients).
A rough examination of the role of the market condition can, however, be made on basis of whole day market returns. During the period of 16 days from which order book data were sampled, the Affärsvärdens General Index (AFGX), a broad Swedish stock market index, realized a compound return of 2.6%, but during the days from which order book snapshots actually were taken, the compound return was -2.2%. Based on these market returns, it seems difficult to determine whether the market was bullish, bearish, or neutral during the period studied. It is, in consequence, difficult to draw any conclusion as to the cause of the observed price impact asymmetry. An disadvantage of using whole day market returns as a measure of market condition is that the whole day market condition needs not be consistent with the market condition at the time from which an order book observation originates.

The coefficient on ORDERSIZE is here modeled as constant across the price impact functions of securities with widely different characteristics. This is consistent with previous studies, but nevertheless somewhat unappealing. It is, however, important to keep in mind that ORDERSIZE is defined as the value of the trade relative to the market capitalization of a given firm. As noted, this means that an order of a given monetary value induces a greater price impact cost for a smaller capitalization stock than for a larger capitalization stock. In view of the adverse selection argument this relation appears reasonable. In quoting or accommodating a trade of a given monetary value, a liquidity provider would require a higher compensation in the firm with the smaller market capitalization, since there is probably a higher risk that there is private information about that entity.

In an attempt to improve on the invariant coefficients, I estimated regressions (not reported) where the coefficient on ORDERSIZE was allowed to vary as a function of LNMC, PROPTR, and the other variables. This was accomplished by using z-scored interaction terms. The coefficients on the interactions, however, failed to achieve statistical significance.

### 2.6.2 Simulation analysis

In order to shed light on the robustness and possible out-of-sample predictive performance for the estimated models, a simulation analysis is conducted. A holdout sample is randomly selected containing 20% of all order book observations.
Observations from the remaining 80% of the order books are used to fit the same regression models as above. The estimated models are then used to predict price impact costs in the holdout sample. This procedure is repeated or replicated 100 times, and done separately for each side of the order book.

In Table 2.6, summary statistics over the estimated regressions are presented.

**Table 2.6 Summary statistics for estimated regressions in the simulation analysis**

A holdout sample was randomly selected containing 20% of all order book observations. The remaining 80% of the order books were used to fit the same regression models as in Table 4. This was repeated 100 times, and done separately for asks and bids. The table reports summary statistics for these 100 repetitions. Price impacts regressions for asks and bids separately over all observations. \( \pi \) = average percentage price impact cost; \( LNMC \) = natural log. of market capitalization in SEK; \( RTRVOL \) = square root of the average daily traded volume as a proportion of outstanding shares over the previous 22 days; \( PROPTR \) = proportion of days with trading previous 22 days; \( LNVOLR \) = log of return volatility over the 22 prev. days; \( LNMDP \) = log. of quote midpoint(SEK); \( ORDERSIZE \) = no. of shares in order as 1/1000 of outstanding shares; Adjusted R\(^2\) are also reported.

**Panel A. Asks (n = 31360)**

Model: \( \pi = \gamma_1 + \gamma_2 LNMC_i + \gamma_3 RTRVOL_i + \gamma_4 PROPTR_i + \gamma_5 LNVOLR_i + \beta ORDERSIZE_i + \epsilon_i \)

<table>
<thead>
<tr>
<th>( \gamma_1 )</th>
<th>( \gamma_2 )</th>
<th>( \gamma_3 )</th>
<th>( \gamma_4 )</th>
<th>( \gamma_5 )</th>
<th>( \beta )</th>
<th>Adj. R(^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean 8.82</td>
<td>-0.13</td>
<td>-19.33</td>
<td>-3.25</td>
<td>0.80</td>
<td>1.20</td>
<td>0.42</td>
</tr>
<tr>
<td>Std 0.14</td>
<td>0.01</td>
<td>0.38</td>
<td>0.12</td>
<td>0.02</td>
<td>0.03</td>
<td>0.01</td>
</tr>
<tr>
<td>Min 8.28</td>
<td>-0.14</td>
<td>-20.16</td>
<td>-3.53</td>
<td>0.74</td>
<td>1.13</td>
<td>0.40</td>
</tr>
<tr>
<td>25% 8.73</td>
<td>-0.13</td>
<td>-19.57</td>
<td>-3.33</td>
<td>0.79</td>
<td>1.18</td>
<td>0.41</td>
</tr>
<tr>
<td>Median 8.83</td>
<td>-0.13</td>
<td>-19.34</td>
<td>-3.25</td>
<td>0.80</td>
<td>1.20</td>
<td>0.42</td>
</tr>
<tr>
<td>75% 8.93</td>
<td>-0.12</td>
<td>-19.07</td>
<td>-3.18</td>
<td>0.81</td>
<td>1.22</td>
<td>0.42</td>
</tr>
<tr>
<td>Max 9.07</td>
<td>-0.11</td>
<td>-18.53</td>
<td>-2.83</td>
<td>0.85</td>
<td>1.27</td>
<td>0.43</td>
</tr>
</tbody>
</table>

**Panel B. Bids (n = 30744)**

Model: \( \pi = \gamma_1 + \gamma_2 LNMC_i + \gamma_3 RTRVOL_i + \gamma_4 PROPTR_i + \gamma_5 LNVOLR_i + \gamma_6 LNMDP_i + \beta ORDERSIZE_i + \epsilon_i \)

<table>
<thead>
<tr>
<th>( \gamma_1 )</th>
<th>( \gamma_2 )</th>
<th>( \gamma_3 )</th>
<th>( \gamma_4 )</th>
<th>( \gamma_5 )</th>
<th>( \gamma_6 )</th>
<th>( \beta )</th>
<th>Adj. R(^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean 7.48</td>
<td>-0.11</td>
<td>-17.2</td>
<td>-2.14</td>
<td>0.72</td>
<td>-0.07</td>
<td>1.02</td>
<td>0.41</td>
</tr>
<tr>
<td>Std 0.13</td>
<td>0.01</td>
<td>0.39</td>
<td>0.11</td>
<td>0.02</td>
<td>0.01</td>
<td>0.03</td>
<td>0.01</td>
</tr>
<tr>
<td>Min 7.10</td>
<td>-0.12</td>
<td>-18.3</td>
<td>-2.37</td>
<td>0.67</td>
<td>-0.09</td>
<td>0.96</td>
<td>0.39</td>
</tr>
<tr>
<td>25% 7.38</td>
<td>-0.12</td>
<td>-17.5</td>
<td>-2.24</td>
<td>0.71</td>
<td>-0.08</td>
<td>1.00</td>
<td>0.41</td>
</tr>
<tr>
<td>Median 7.49</td>
<td>-0.11</td>
<td>-17.2</td>
<td>-2.14</td>
<td>0.72</td>
<td>-0.07</td>
<td>1.02</td>
<td>0.41</td>
</tr>
<tr>
<td>75% 7.56</td>
<td>-0.11</td>
<td>-17.0</td>
<td>-2.06</td>
<td>0.73</td>
<td>-0.07</td>
<td>1.04</td>
<td>0.41</td>
</tr>
<tr>
<td>Max 7.82</td>
<td>-0.09</td>
<td>-16.1</td>
<td>-1.89</td>
<td>0.76</td>
<td>-0.05</td>
<td>1.11</td>
<td>0.43</td>
</tr>
</tbody>
</table>

For both asks and bids, the regression coefficients estimated in the reduced (80%) random subsamples seem quite stable across the 100 replications, and the average coefficient values are close to the ones in Table 2.6 that were obtained when all observations were used for estimation.

For a given replication and a given observation in the holdout sample, a forecasted price impact is generated by inserting the values of the observation’s explanatory
variables into the regression model with coefficients estimated on the remaining 80% of observations. Statistics over the predicted values can be found in Table 2.7.

**Table 2.7 Statistics on forecasted percentage price impact cost based on the holdout sample**

In each of the 100 replications the Average, Minimum, Median, and Maximum forecasted price impact cost were computed. The row labeled Mean is the average over the 100 average, minimum, median and maximum forecasted price impact costs obtained in a single replication. The number of observations in any holdout sample equals 20% of the number of observations in the full sample.

<table>
<thead>
<tr>
<th></th>
<th>Average</th>
<th>Minimum</th>
<th>Median</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A. Asks</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>1.99</td>
<td>-0.65</td>
<td>1.82</td>
<td>7.79</td>
</tr>
<tr>
<td>Std</td>
<td>0.04</td>
<td>0.09</td>
<td>0.04</td>
<td>0.65</td>
</tr>
<tr>
<td>Min</td>
<td>1.90</td>
<td>-0.81</td>
<td>1.73</td>
<td>6.43</td>
</tr>
<tr>
<td>25%</td>
<td>1.96</td>
<td>-0.72</td>
<td>1.79</td>
<td>7.31</td>
</tr>
<tr>
<td>Median</td>
<td>1.98</td>
<td>-0.68</td>
<td>1.82</td>
<td>7.71</td>
</tr>
<tr>
<td>75%</td>
<td>2.01</td>
<td>-0.56</td>
<td>1.85</td>
<td>8.60</td>
</tr>
<tr>
<td>Max</td>
<td>2.10</td>
<td>-0.43</td>
<td>1.93</td>
<td>8.94</td>
</tr>
<tr>
<td><strong>Panel B. Bids</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>1.84</td>
<td>-0.54</td>
<td>1.71</td>
<td>7.41</td>
</tr>
<tr>
<td>Std</td>
<td>0.03</td>
<td>0.10</td>
<td>0.03</td>
<td>0.45</td>
</tr>
<tr>
<td>Min</td>
<td>1.76</td>
<td>-0.70</td>
<td>1.61</td>
<td>6.23</td>
</tr>
<tr>
<td>25%</td>
<td>1.81</td>
<td>-0.62</td>
<td>1.69</td>
<td>7.04</td>
</tr>
<tr>
<td>Median</td>
<td>1.84</td>
<td>-0.58</td>
<td>1.71</td>
<td>7.50</td>
</tr>
<tr>
<td>75%</td>
<td>1.86</td>
<td>-0.43</td>
<td>1.74</td>
<td>7.78</td>
</tr>
<tr>
<td>Max</td>
<td>1.92</td>
<td>-0.32</td>
<td>1.79</td>
<td>8.00</td>
</tr>
</tbody>
</table>

Table 2.7 allows for checking whether the numerical range of the forecasted price impact costs is reasonable. This seems to be case, although some negative forecasted price impact costs are observed. These observations refer to Ericsson and are caused by the very large market capitalization of that firm during the period of the study. In the empirical analyses in Chapter 3 and Chapter 4, negative price impact cost forecasts are set to zero.

In evaluating the out-of-sample forecast performance, I use as error metric the actual price impact cost minus the forecasted one; by multiplying this forecast error with the money value of the order quantity, the money value of the forecast error is obtained. In addition to having a meaningful interpretation, this error metric avoids the problems with close-to-zero denominators that metrics, which express the forecast error as a proportion of the forecast or the actual, might have. Table 2.8 displays the forecast performance.
Table 2.8 Forecast errors of the percentage price impact cost in the holdout sample

The forecast error (FE) metric employed is the actual percentage price impact cost minus the forecasted one. For each of the 100 replications, the average, min, median, maximum and t-value of the forecast errors are computed. In the row labeled Mean, the average of the 100 average, min, median, max and t-values enter. In the row labeled standard deviation, the standard deviation across the 100 replications of the average, min, etc., enters, and so on.

<table>
<thead>
<tr>
<th></th>
<th>Signed FE</th>
<th>Absolute FE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Avg</td>
<td>Min</td>
</tr>
<tr>
<td>Panel A. Asks</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.00</td>
<td>-4.78</td>
</tr>
<tr>
<td>Std</td>
<td>0.04</td>
<td>0.43</td>
</tr>
<tr>
<td>Stderr</td>
<td>0.00</td>
<td>0.04</td>
</tr>
<tr>
<td>Min</td>
<td>-0.10</td>
<td>-5.51</td>
</tr>
<tr>
<td>25%</td>
<td>-0.02</td>
<td>-5.20</td>
</tr>
<tr>
<td>Median</td>
<td>0.00</td>
<td>-4.66</td>
</tr>
<tr>
<td>75%</td>
<td>0.03</td>
<td>-4.46</td>
</tr>
<tr>
<td>Max</td>
<td>0.13</td>
<td>-3.87</td>
</tr>
</tbody>
</table>

|                  |           |             |             |             |            |     |         |
| Panel B. Bids    |           |             |             |             |            |     |         |
| Mean             | 0.00      | -4.06       | -0.19       | 7.12        | -0.34      | 0.90 | 88.85   |
| Std              | 0.05      | 0.17        | 0.03        | 0.55        | 3.19       | 0.02 | 2.16    |
| Stderr           | 0.00      | 0.02        | 0.00        | 0.06        | 0.32       | 0.00 | 0.22    |
| Min              | -0.10     | -4.55       | -0.27       | 5.86        | -7.52      | 0.84 | 83.18   |
| 25%              | -0.03     | -4.14       | -0.22       | 6.64        | -2.39      | 0.89 | 87.30   |
| Median           | 0.00      | -4.03       | -0.19       | 7.07        | 0.01       | 0.90 | 88.90   |
| 75%              | 0.03      | -3.97       | -0.17       | 7.57        | 1.89       | 0.91 | 90.48   |
| Max              | 0.10      | -3.51       | -0.11       | 7.91        | 6.72       | 0.96 | 94.74   |

Overall, for both sides of the market, the out-of-sample performance is quite satisfactory: the bias or the average signed forecast error is on average close to zero percent, while the accuracy or the average absolute forecast error is on average below one percent; the maximum absolute forecast error for asks (bids) was 10.6% (7.9%). The mean median signed forecast error is for asks –0.20% and for bids –0.19%. Caution should be exercised in interpreting the t-values, because as with the regression residuals, the forecast errors are not uncorrelated. In particular, both the signed and absolute forecast errors exhibit positive autocorrelations up to an order of 10. Thus, autocorrelation consistent standard errors would be larger than the unadjusted ones, and the t-values lower. The actual autocorrelation consistent standard errors (not reported) are usually more than two times greater than the unadjusted ones. Hence, it cannot be rejected that the mean signed forecast errors are zero; neither can it safely be

---

35 These were computed using eq. 2.4.41 in Campbell et al. (1997).
rejected that the quartiles, 25% and 75% quantiles, of the forecast error distributions are zero. This applies to both the ask and the bid side.

2.6.3 Improvements and extensions
The efficiency of the estimation approach may be improved. One option would be to use generalized least squares (GLS) where the error variance process be specified as an autoregression of order roughly equal to the number of points selected from each price impact function. The error process would be estimated in a first step. With the greater efficiency of GLS maybe the intuitively attractive specification with z-scored interactions allowing for different slopes of the price impact functions of individual securities would achieve significant coefficients. Further non-linear forms of the explanatory variables should be tested.

It may be worthwhile to estimate a simultaneous system as some variables are likely to be interrelated (see, e.g., Glosten and Harris 1988; Brennan and Subramanyam 1993).

More limit order book snapshots could be collected, preferably at regular intra-daily points in time. It would then be possible to analyze possible systematic variation over time and the effects from market-wide conditions (commonalities) on individual securities.

2.7 Conclusions
I showed analytically that the quoted price impact cost (the absolute volume-weighted average price degradation relative to the quote midpoint) of market buy and sell orders in a limit order book with discrete prices is an increasing piecewise concave function of order volume. This detail seems to have to been unnoticed in the literature (see, e.g., Glosten 1994). This particular functional form implies, for one thing, that quoted price impact cost functions can have widely different shapes.

Based on electronic limit order book data for a large cross-section of Swedish equities, regression analysis was used to estimate empirical price impact cost functions separately for buy and sell orders. For both the ask and bid side, the estimated price impact cost function was found to be decreasing in market capitalization and (two different measures of) historical trading activity; and increasing in order size
(measured as the number of shares in the order relative to all outstanding shares), stock return volatility and, for bids only, quote midpoint. These results are largely consistent with microstructure theory and empirical results obtained for other markets. The separate analysis of bids and asks allowed the generation of evidence indicating that the price impact cost for market buy orders, on average, is higher and increases faster in order size than it does for market sell orders.

To shed light on the robustness and possible out-of-sample predictive performance of the estimated models, a simulation analysis was carried out. I introduced as error metric the actual percentage price impact cost minus the forecasted one. By multiplying this percentage forecast error with the monetary order value, the money value of the forecast error is obtained. The robustness and out-of-sample forecast performance of the estimated models were satisfactory. For both asks and bids, the mean signed forecast error was close to 0%, while the mean absolute forecast error was below 1%. The maximum absolute forecast errors for the bid side and the ask side were 7.9% and 10.6%, respectively.

Bessembinder (2003) and Chalmers et al. (2001) emphasize the costs and difficulties associated with obtaining estimates of transaction costs for individual stocks, e.g.:

“However, from a practical standpoint, trading expenses are difficult to estimate and not readily available from any data source.” Chalmers et al. (2001, p. 15)

The approach presented here resolves some of the issues involved in price impact cost modeling in electronic open limit order book markets, thereby providing a solution to a recognized need.
Chapter 3

Mean-variance portfolio management under transaction costs

In this chapter, I extend the standard mean-variance portfolio model by formulating a model for mean-variance portfolio revision under transaction costs including price impact. The transaction cost models estimated in Chapter 2 are integrated with the extended model, and the importance of transaction cost control is assessed empirically within the integrated framework.

3.1 Introduction

“...there has been a strong tendency to overlook the transactions costs in favor of a simplified treatment of the revision problem. In our view, all of these simplifications are less than satisfactory.” Rosenberg and Rudd (1979, p. 27)

The mean-variance model of Markowitz (1952, 1959, 1987) is an established analytical framework for portfolio problems. The original formulation concerns the problem of creating a portfolio from cash. In the absence of transaction costs, this problem is identical to the more general problem of revising an existing portfolio, but with transaction costs present, the portfolio revision problem becomes distinct from the creation problem. Papers addressing portfolio revisions with transaction costs in a mean-variance quadratic programming (QP) framework include Pogue (1970), Chen et al. (1971), Rudd and Rosenberg (1979), Perold (1984), and Adcock and Meade (1994).

The mean-variance approach requires the optimization of a quadratic function with possibly many variables and restrictions and is often computationally demanding. Historically, computing power has been limited and available algorithms relatively inefficient. To achieve computational feasibility, researchers have had certain features of the particular problems addressed aptly exploited, and models of transaction costs

---

36 The mean-variance portfolio revision model under transaction costs including price impact developed here was first presented at Umeå Workshop in Finance and Accounting in May 1999. A presentation of a paper related to the empirical part of this chapter was given at the Research Institute USBE Finance Workshop April 2002.
simplified. This has sometimes resulted in loss of generality and realism. Adcock and Meade (1994), for instance, present an approach to transaction costs designed rather exclusively for a particular quadratic programming algorithm. In addition, they use - as do also Chen et al. (1971) and Rudd and Rosenberg (1979) – models of individual stocks’ transaction costs in which the percentage cost of a transaction is independent of the size of the transaction. This modeling is limited in that the percentage transaction cost cannot vary with trade size. Price impact costs can thus not be explicitly considered. Pogue (1970) uses an increasing stepwise function to model variable percentage transaction costs; the percentage cost is constant in a given volume interval, but increases between intervals. Realistic models of transaction costs require several steps. This complicates modeling and computation.

Other, more heuristic approaches for controlling transaction costs in connection with portfolio revisions have been suggested by, e.g., Smith (1967) and Schreiner (1980). The latter writer presents a method that straightforwardly extends the portfolio creation model by constraining turnover of the existing portfolio to a designated maximum rate. This will limit total transaction costs, but not necessarily in an optimal fashion. Incorporating transaction costs explicitly in the portfolio revision model as will be done here, will generally (indirectly) limit turnover, and for the right reasons.

The use of MV optimization among investors has been reported as limited (see, e.g., Michaud 1989, 1998; Konno and Yamazaki 1991; Black and Litterman 1992; Jorion 1992; Fisher and Statman 1997). One of the reasons given for this, apart from the one that the standard MV model disregards transaction costs, is that mean-variance efficient portfolios often lack diversification, and this make them intuitively unappealing to investors (see, e.g. Michaud 1989, 1998; Best and Grauer 1991a, 1991b; Green and Hollifield 1992; Fisher and Statman 1997; Jagannathan and Ma 2003). It appears possible that the inclusion of transaction costs in the portfolio optimization model could mitigate the problem of lack of diversification. Lack of diversification, or an extreme portfolio allocation, is often the result of extreme portfolio alterations. As extreme portfolio alterations in general incur large transaction costs, it seems reasonable to expect that extreme alterations – and extreme allocations -
would be less likely to occur if transaction costs, including volume-dependent price impact costs, are controlled.

I formulate a quadratic program for mean-variance portfolio revisions under variable percentage transaction costs. The volume-weighted average percentage transaction cost in a given security is modeled as linearly increasing in the volume transacted. This modeling of transaction costs is consistent with recent empirical research, e.g., Chan and Lakonishok (1997) and Keim and Madhavan (1997, also see 1998), and with the price impact cost models developed in Chapter 2, and, thus, allows for explicit consideration of price impact costs. The negative contribution to a portfolio’s return from changing the portfolio weight for a given security will be a quadratic function, without constant term, of the weight change. In lieu of possibly more realistic and complex models of transaction costs, the approach taken here is believed to provide a good balance between computational cost, realism, and ease of implementation.

A revised portfolio obtained without full consideration of transaction costs is optimal only in the true absence of the transaction costs neglected, and could, net of the neglected transaction costs, be quite far from optimal. An empirical test is carried out to assess the importance of transaction costs and transaction cost control in portfolio revisions. The models of price impact costs estimated in Chapter 2 are combined with the commission rates estimated in Dahlquist et al. (2000) to form (full) transaction cost models. These transaction cost models are integrated with the portfolio revision program formulated here. The integrated model is used to analyze revisions of typical Swedish mutual fund portfolios across a broad range of risk attitudes. The performance of the integrated model, whose transaction cost specification includes price impact costs, is compared to that of the standard MV model, which disregards transaction costs, as well as to the performances of some MV portfolio revision models involving less elaborate transaction cost specifications. The tests may thus be said to consider mean-variance portfolio revisions conducted under different degrees of consideration of transaction costs. Performance is evaluated in terms of portfolio utility, portfolio diversification, turnover rates, and incurred levels of transaction costs. Certainty equivalent losses - relative to a revised portfolio obtained with full
consideration of transaction costs - of revised portfolios obtained with less than full consideration of transaction costs, are computed. Additional results concern the effects on portfolio utility and diversification from a maximum weight constraint on individual holdings as imposed by Swedish law.

The price impact costs models of Chapter 2 were developed for trading that seeks immediacy of execution, and where the whole price effect is regarded as temporary. Opportunity costs are omitted, because they were argued to be less relevant for trades executed with immediacy in an open limit order book, when the transaction costs are accounted for by the optimizer. The empirical results obtained in this chapter thus apply to portfolio revisions where the trading is uninformed and immediate.

In the next section the standard MV portfolio selection model is presented and some of its issues discussed. In Section 3.3, transaction costs are introduced and the standard MV model is extended to handle portfolio revisions under transaction costs including price impact. Section 3.4 presents the test design and data. The empirical analysis follows in Section 3.5. This chapter is concluded in Section 3.6.

3.2 Background

3.2.1 Mean-variance portfolio selection

The mean-variance model was introduced in the 1950’s by Markowitz (1952, 1959, 1987) and others as a normative model for portfolio selection. Investors are assumed to be risk-averse, single-period, expected-utility maximizers, and utility functions and probability distributions of returns such that expected utility can be maximized or close to maximized by choosing portfolios from the mean-variance efficient set, that is, the subset of feasible portfolios that for every given variance of return offer maximum mean return.

Let \( x \) be a \( n \times 1 \) column vector, whose elements \( x_1, \ldots, x_n \) denote the weight or proportion of the investor’s wealth allocated to the \( i \)th security in the portfolio. There is a budget constraint, that is, the weights sum to unity: \( \sum_{i=1}^{n} x_i = 1 \), or in matrix notation \( \mathbf{1}'x = 1 \), where the “\( \mathbf{1} \)” represents a \( n \times 1 \) column vector of ones, and “\( \mathbf{1}' \)” the transpose. \( \mathbf{r} \) is a \( n \times 1 \) column vector of mean returns \( R_1, \ldots, R_n \). \( \mathbf{V} \) is the \( n \times n \) variance-
covariance matrix of returns. For a given portfolio $p$, the mean return is $R_p = r'x$, and the variance is $\sigma_p = x'Vx$. For computational reasons, $V$ is assumed positive semi-definite, which means that $x'Vx \geq 0$ for all $x$. (This is expected to hold true since (portfolio) variance is by definition a non-negative quantity.) Let $\lambda, \lambda > 0$, be the investor’s risk aversion parameter.

The mean-variance (MV) problem could then be stated as:

$$\max_{\{x\}} r'x - \lambda x'Vx$$

subject to $1'x = 1$, and

$$0 \leq x_i, \quad i = 1, 2, ..., n.$$ (3.1)

For a fixed $\lambda$ and returns $r$ and variances-covariances $V$, the standard MV portfolio problem is to choose the weights $x$ such that (3.1) is being maximized, here subject to (i) the budget restriction, i.e., weights must sum to unity, and (ii) non-negative weights, i.e., short selling is prohibited. The above is a quadratic optimization problem, as quadratic terms of the decision variables, $x_i$, appear in the goal function, with linear restrictions. The solution can be obtained by quadratic programming. Let $x_s$ be the solution vector to the above problem for a given $\lambda$, then the portfolio $\{R_s = r'x_s, \sigma_s = x_s'Vx_s\}$ is a mean-variance efficient portfolio.

Formulating the MV problem as in (3.1) allows for several meaningful interpretations. First, the mean-variance efficient set is obtained as $\lambda$ varies from 0 to infinity. Second, the optimal portfolio solution to (3.1) exactly maximizes expected utility for an investor who has negative exponential preferences and makes normal probability assessments and whose Arrow-Pratt (constant) absolute risk aversion coefficient is $2\lambda$; in addition, the numerical value of the objective is equal to the portfolio’s certainty equivalent. Third, the goal function in (3.1) is equivalent to the “general” goal function derived by Pulley (1981). In that context, the optimal portfolio provides approximately maximum expected utility for investors with Arrow-Pratt relative risk aversion coefficients equal to $2\lambda$ more or less regardless of the specific

---

37 A check for positive semi-definiteness of a matrix could be done by inspection of its eigenvalues, which should be greater than or equal to zero, with at least one eigenvalue equal to zero.

38 To obtain the numerical value of the expected utility, the analysis must be done in gross returns (returns plus unities).
form of utility functions and return distributions given that holding periods are short and or returns close to zero. Fourth, Tobin (1958, 1969) shows that if either the utility function is quadratic or if returns are multivariate normal, then the MV efficient set will contain the portfolios with maximum expected utility.

Additional motivations for the MV approach include the following ones. Under fairly general conditions on preferences and beliefs, Samuelson (1970) and Ohlson (1975) show that as the holding period approaches zero, expected utility converges to a function of mean and variance. Tsiang (1972) demonstrates that MV approximations to expected utility improve as the proportion of an individual’s total wealth invested in risky assets decreases. Therefore, if the proportion of wealth at risk is small, MV approximations should be expected to perform well.

The use of (3.1) may thus be justified in at least two ways. One could by varying $\lambda$ from 0 to infinity, compute the efficient set and pick from it the portfolio with maximum expected utility given one’s preferences, or one could solve (3.1) for a given value of $\lambda$ that corresponds to either one’s relative or absolute Arrow-Pratt risk aversion coefficient depending on what assumptions one is willing to make. And, a priori, one would expect (3.1) to perform well if the holding period is short and or returns small, the proportion of wealth at risk small, and returns close to multivariate normal.

There is considerable empirical evidence, see, e.g., Pulley (1981, 1983), Kroll, Levy and Markowitz (1984), Hlawitschka (1994), and Amilon (2001) that optimal portfolios selected on the basis of functions of means and variances closely approximate portfolios selected by expected utility maximization. Amilon (2001) examines portfolios of Swedish stocks and derivatives containing around 120 securities, while prior research considers portfolios of only 10-20 securities. Moreover, Pulley (1981) and Kallberg and Ziemba (1983) find that investors with differing utility functions but similar risk aversion hold similar optimal portfolios, while Grauer (1986) and Grauer and Hakansson (1993) for longer (yearly) holding periods and for higher levels of risk aversion, document differences in optimal portfolio composition between power utility functions and MV approximations.
As noted, the MV portfolio model is, in spite of its appealing features, reported to be largely ignored by investment practitioners. Lack of diversification of MV optimal portfolios is suggested as one reason for this. Obtained solutions may also be unstable in that they can be very sensitive to small perturbations in input parameters; a small change in the expected return of a given security might cause large changes in optimal portfolio composition (Best and Grauer 1991a, 1991b; Meade and Salkin 1990). Another reason concerns estimation risk (see, e.g., Kalymon 1971; Barry 1974; Klein and Bawa 1976; Jobson and Korkie 1980; Jorion 1992; Chopra and Ziemba 1993; Grinold and Kahn 1995). The inputs to the MV portfolio selection model are measured with error, but treated as error-free in the optimization process, and in solutions. An optimal portfolio is therefore only an estimate, or approximation, of unknown accuracy, to a “true” optimal portfolio. This is the estimation risk. The combined effects of estimation risk and unstable solutions are complicating and might lead to “error maximization” (Jobson and Korkie 1981; Michaud 1989, 1998; Fisher and Statman 1997). This refers to the phenomenon that estimation errors do not tend to “average out” in the optimization process, but rather are magnified.

Here, the aforementioned problems will not be directly dealt with. The empirical analysis will, however, shed light on how the incorporation of transaction cost control in the MV model affect portfolio diversification. Moreover, on a somewhat speculative note, volatile securities might be more prone to estimation error and thereby more likely to contribute to error maximization. They are, however, also documented to have higher transaction costs (see Chapter 2 and, e.g., Keim and Madhavan 1998), and this will make large transactions in them expensive, which in turn may help to alleviate error maximization. In passing, one may note that a suggested remedy for the problem of error maximization is to constrain portfolio weights (see, e.g., Frost and Savarino 1988), and this is in effect tantamount to restricting portfolio turnover, which is proposed as a means for controlling transaction costs (Schreiner 1980, also see Markowitz 1987). As mentioned, restricting turnover will usually limit the transaction

---

39 Michaud (1998), for instance, suggests a method for improving MV optimization with regard to error maximization.
costs but not necessarily in an optimal fashion. Below, the effects of a maximum weight constraint as imposed by Swedish law will be investigated.

If one uses (3.1) to select an optimal portfolio, but without consideration of transaction costs, the resulting portfolio after transaction costs could be quite far from optimal. Hence, in order to achieve or at least approach the true optimal portfolio, transaction costs must be accounted for. Moreover, (3.1) represents a portfolio creation problem, as opposed to a portfolio revision problem. In the first instance, you create a portfolio by buying assets from a given amount of money while in the second instance you revise an existing portfolio through buying and selling. In the absence of transaction costs (3.1) is perfectly capable of handling portfolio revisions, because the existing portfolio can be translated into cash without cost, making the problem one of portfolio creation. With transaction costs, however, the revision problem becomes distinct from the creation problem. Below transaction costs are introduced and (3.1) is reformulated to handle mean-variance portfolio revisions under transaction costs including price impact. Finally, in that formulation no attempt is made to account for the tax consequences that some transactions may lead to. Under, for instance, Swedish legislation, a fund is not taxed for capital gains, only for net asset value, and there is thus no need to consider such tax consequences. (For an individual investor owning shares in the fund there might nevertheless be significant tax effects.)

3.3 Model development

3.3.1 Transaction costs

In the MV portfolio optimization literature, e.g., Chen et al. (1971), Rudd and Rosenberg (1979), and Adcock and Meade (1994), transaction costs have, when included, usually been represented by constant percentages probably intended to be representative of the volume-weighted average percentage cost of a trade of typical size in a given security. In other words, price impact cost is not explicitly modeled as a function of order size. Moreover, the average percentage cost of a trade of typical size
for a given security depends on many factors including investment strategy, portfolio size, and composition, and is as a consequence difficult to estimate.\footnote{In a series of recent papers, Konno and different co-authors elaborate on the mean-absolute deviation (MAD) portfolio model of Konno and Yamazaki (1991). In particular, the basic model is extended to accomodate minimization of non-convex transaction costs as a function of weight change, see, e.g., Konno (2003). Solutions are obtained by the application of a branch and bound algorithm (“the hypercube subdivision algorithm”).}

Let $c_i$ be the constant percentage transaction cost, and $x_i$ the weight change for the $i$th security, then the cost at the portfolio level or return charge from investment in security $i$ is $x_i \times c_i$, i.e. proportional to, or a linear function of, the weight. A problem in modeling constant percentage transaction costs arises when weights might have both negative and positive signs, as in portfolio revision problems and problems where negative positions in some assets are allowed. For example, a negative weight times a cost coefficient ($-x_i \times c_i$) produces erroneously a negative cost, i.e., a positive contribution to return. Here this will be resolved by using one variable to represent a decrease in a given holding and another one to represent an increase; such a mathematical program is called a semi-linear program (Fourer 1989).\footnote{Corresponding to the semi-linear program there is a semi-linear function, which belongs to the class of piecewise linear functions.}

One approach, originally suggested by Pogue (1970), to increase the realism and flexibility in the modeling of transaction costs beyond the constant percentage cost model, is to use a separable convex piecewise linear function for the transaction costs. Such a function is specified by a number of breakpoints and a sequence of increasing slopes between the breakpoints. A piecewise linear function could be represented by a piecewise linear program, which could be transformed into an equivalent linear program (see Fourer 1989 and Ho 1985), which could be inserted into a quadratic program and be solved by means of quadratic programming.

More elaborate models of transaction costs require many breakpoints and slopes, and this makes modeling and computation more demanding.\footnote{For instance, breakpoints and slopes must be determined such that they provide an adequate representation of each security’s transaction costs function. Hybrid approaches involving P-L and other functions are also possible.} I therefore suggest an extension of the constant percentage cost model in the form of a model where the
volume-weighted average percentage transaction cost for a given security is linearly increasing in the size of the transaction; this allows price impact to be taken into account. For security \( i \), the total volume-weighted average percentage trading cost for a given weight change is represented as \( T_i^* = c_i^* + d_i^* x_i^* \), where \( \cdot = \text{"+"} \) denotes a purchase or a weight increase, and \( \cdot = \text{"−"} \) denotes a sale or a weight decrease, \( c_i^* \) and \( d_i^* \) are coefficients uniquely determined for the security, and \( x_i^* \) is the change in portfolio weight. The cost at the portfolio level or return charge from a transaction of size \( x_i^* \) in security \( i \) is equal to \( x_i^* (c_i^* + d_i^* x_i^*) \), i.e., a quadratic function, without constant term, of weight change. The increase in computational expense for the innovation relative to the standard MV portfolio selection model derives chiefly from a doubled number of variables, and, relative to the portfolio revision model with linear transaction costs, the increase is nil.\(^{43}\)

**3.3.2 Portfolio revision under transaction costs including price impact**

In the portfolio revision problem, the basic idea is to represent the resulting, revised portfolio (vector) \( \mathbf{x} \) as the sum of the initial, unrevised portfolio, which is denoted \( \mathbf{x}_0 \), and the weight changes – which can be either increases or decreases – needed to achieve the optimal revised portfolio.

Let \( x_i^+ \) and \( x_i^- \), represent the increase and decrease, respectively, in security \( i \). Furthermore, it is required that \( x_i^+, x_i^- \geq 0 \). Let \( x_{i0} \) be the initial weight in security \( i \), then the resulting weight is \( x_i = x_{i0} + x_i^+ - x_i^- \). The revised portfolio \( \mathbf{x} \) can then be expressed as \( \mathbf{x} = \mathbf{x}_0 + \mathbf{x}^+ - \mathbf{x}^- \). If the constant percentage transaction cost is \( c_i \), then the cost for a purchase transaction is \( x_i^+ \times c_i \), and for a sale \( x_i^- \times c_i \). Because \( x_i^+, x_i^- \geq 0 \), a consistent expression is obtained for the transaction cost of both weight increases and decreases. It also allows for having different cost coefficients for buying and selling, i.e., \( c_i^+ \) and \( c_i^- \). Also note the important fact that \( x_i^+ \) and \( x_i^- \) are mutually exclusive in the sense that either one is zero. On the portfolio level this cost component for revising the portfolio can be expressed as \( \mathbf{x}^+ \mathbf{c}^+ + \mathbf{x}^- \mathbf{c}^- \).

\(^{43}\) Konno and Yamazaki (1991) mention computational cost as a reason for the limited use of the standard MV approach.
Let $d^+$ and $d^-$ be diagonal matrices with the $n$ securities’ quadratic and non-negative cost terms, $d_i^+$ and $d_i^-$, along their respective main diagonals. The contribution to transaction costs from these terms when moving the portfolio from $x_0$ to $x$ equals $x'^+d^+x^+ + x'^-d^-x^-$. In the goal function of (3.1), substitute $x$ with $x = x_0 + x^+ - x^-$ and insert the linear and quadratic transaction cost components:

$$
\text{max}_{(x',x)} \left( r'(x_0 + x^+ - x^-) - \lambda (x_0 + x^+ - x^-)'V (x_0 + x^+ - x^-) \right) - x^+'c^+ - x^-'c^- - x^'+d^+x^+ - x^-'d^-x^-.
$$

(3.2)

Adjust bounds and constraints for the substitution, drop constant terms and rearrange to obtain the program:

$$
\text{max}_{(x',x)} \left( r' - 2\lambda x_0'V \right)(x^+ - x^-) - x^+'c^+ - x^-'c^- - \lambda (x^+ - x^-)'V (x^+ - x^-) - x^'+d^+x^+ - x^-'d^-x^- \quad (3.3)
$$

subject to

- $1'(x_0 + x^+ - x^-) = 1$ (budget restriction)
- $0 \leq x_{i0} + x_i^+ - x_i^-$, $i = 1, 2, ..., n$ (short selling constraint)
- $0 \leq x_i^+, x_i^-$, $i = 1, 2, ..., n$.

The above form is unapt for quadratic programming. Before the program is further transformed, it is convenient to have the following notation introduced:

$$
X = \begin{pmatrix} x^+ \\ x^- \end{pmatrix}, \quad W = \begin{pmatrix} V & -V \\ -V & V \end{pmatrix}, \quad D = \begin{pmatrix} d^+ & 0 \\ 0 & d^- \end{pmatrix}.
$$

For a symmetric matrix $V$, the following identity applies:

$$
(x^+ - x^-)'V(x^+ - x^-) = x'^+Vx^+ + x'^-Vx^- - 2x'^+Vx^-
$$

$$
= \begin{pmatrix} x^+ \\ x^- \end{pmatrix}' \begin{pmatrix} V & -V \\ -V & V \end{pmatrix} \begin{pmatrix} x^+ \\ x^- \end{pmatrix} = X'WX \quad (3.4)
$$

Similarly, for symmetric (and diagonal) matrices $d^+$ and $d^-$, the following applies:

$$
x'^+d^+x^+ + x'^-d^-x^- = \begin{pmatrix} x^+ \\ x^- \end{pmatrix}' \begin{pmatrix} d^+ & 0 \\ 0 & d^- \end{pmatrix} \begin{pmatrix} x^+ \\ x^- \end{pmatrix} = X'DX. \quad (3.5)
$$

In view of (3.4) and (3.5), the goal function in (3.3) may now be expressed as

---

44 Observe that after having dropped these terms, the value of goal function can no longer be interpreted as an expected utility value (in the cases where that had been admissible before).
\[
X' \begin{pmatrix}
  r - 2\lambda Vx_0 \\
  -(r - 2\lambda Vx_0)
\end{pmatrix} - X' \begin{pmatrix}
  c^+ \\
  -c^-
\end{pmatrix} - \lambda X'WX - X'DX \\
= X' \begin{pmatrix}
  r - 2\lambda Vx_0 - c^+ \\
  -r + 2\lambda Vx_0 - c^-
\end{pmatrix} - \lambda X'(W + D/\lambda)X.
\]

The budget restriction \(1'(x_0 + x^+ - x^-) = 1\) is transformed and becomes in block matrix notation \(\begin{pmatrix} 1 \\ -1 \end{pmatrix}' \begin{pmatrix} x^+ \\ x^-
\end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}' X = 0\), and implies that the sum of weight increases and decreases must be zero.

For a variety of reasons, including legal requirements and fund policy statements, it may be desirable to restrict the minimum and maximum weight for a given security by specifying bounds on its weight. In general, the bounds \(l_i \leq x_i \leq u_i\), where \(x_i = x_{i0} + x_i^+ - x_i^-\) and where \(l_i\) and \(u_i\) represent the lower and upper limits, respectively, could be split and expressed in terms of \(x_i^+\) and \(x_i^-\) in the following manner:

\[
l_i - x_{i0} \leq x_i^+ \leq u_i - x_{i0}, \quad i = 1, 2, ..., n, \text{ and}
\]

\[
u_i - x_{i0} \leq x_i^- \leq l_i - x_{i0}, \quad i = 1, 2, ..., n.
\]

Some of these bounds could become inactivated by the non-negativity condition imposed on \(x_i^+\) and \(x_i^-\). For instance, if the initial weight for stock \(i, x_{i0}\), is below its lower bound \(l_i\), the lower bound for the weight increase \(x_i^+\) will be set to a value such that it forces a minimum weight increase of \(l_i - x_{i0}\), so that the resulting weight \(x_i\) will obey the lower bound \(l_i\); the lower and upper bounds for the weight decrease will then both equal zero, and so on. Finally, combining this with the objective and the budget constraint above, the following program is arrived at:

\[
\max_{\{X\}} \quad X' \begin{pmatrix}
  r - 2\lambda Vx_0 - c^+ \\
  -(r + 2\lambda Vx_0 - c^-)
\end{pmatrix} - \lambda X'(W + D/\lambda)X' \\
\text{subject to} \quad \begin{pmatrix} 1 \\ -1 \end{pmatrix}' X = 0,
\]

\[
max(l_i - x_{i0}, 0) \leq x_i^+ \leq max(u_i - x_{i0}, 0), \quad i = 1, 2, ..., n,
\]

\[
max(x_{i0} - u_i, 0) \leq x_i^- \leq max(x_{i0} - l_i, 0), \quad i = 1, 2, ..., n.
\]
This is a concave maximization problem (or equivalently, a convex minimization problem). The objective function is concave, for its second derivative, \(-\lambda(W + D)\), is negative semidefinite. To see this, consider the form \(\lambda X'(W + D/\lambda)X\), which is positive semidefinite, because the variance-covariance matrix \(V\) is positive semidefinite (by construction), and \(D/\lambda\) is diagonal with non-negative elements along the main diagonal. This can also be observed if the form is expanded as follows:

\[
\lambda X'(W + D/\lambda)X = \\
\lambda X'WX + X'DX = \\
\lambda(x^+ - x^-)'V(x^+ - x^-) + x^+d^+x^+ + x^-d^-x^-.
\]

The importance of transaction cost control in mean-variance portfolio revision will be examined next using the above program in conjunction with the empirical cross-section models of transaction costs for Swedish stocks estimated in Chapter 2. In addition, (3.6) largely forms the basis for the index fund revision models to be derived and tested in Chapter 4.

### 3.4 Test design and data

An empirical test is conducted to assess the importance of transaction costs and transaction cost control in mean-variance portfolio revision.

I assume that the true model of transaction costs includes price impact, and that it is a quadratic function, without constant term, of weight change, with coefficients given by the empirical models of price impact cost estimated in Chapter 2, and by the commission rates reported in Dahlquist et al. (2000). This transaction cost model is integrated with the mean-variance revision model under variable percentage transaction costs formulated in this chapter. The integrated model is applied to revise two typical Swedish mutual fund portfolios of different sizes in terms of net asset value (NAV) across a wide range of risk attitudes. The initial (unrevised) portfolios are capitalization-weighted and contain all Swedish stocks (with sufficient data) as of 30 Dec. 1999. In view of the existing literature, this is a quite large portfolio revision problem, as the initial portfolio contains 180 stocks. The performance of the integrated model is compared to that of the standard MV model, which neglects transaction costs altogether, as well as to the performances of some revision models involving less
elaborate transaction cost specifications. Performance is evaluated in terms of portfolio utility, portfolio diversification, turnover rates, and incurred levels of transaction costs.

The magnitude of transaction costs for a given portfolio is related to its size, or net asset value (NAV), and turnover rate. Everything else equal, the transaction cost for a trade of a given size measured as portfolio weight change, increases with the NAV of the portfolio. For portfolios differing only with respect to turnover rate, trading costs increase with increasing turnover given that one amortizes expected trading costs as suggested by Rudd and Clasing (1982) and Grinold and Kahn (1995), which will be done here.

Dahlquist et al. (2000) report statistics on Swedish mutual funds. They estimate the average NAVs, portfolio turnover rates, and commission rates for two categories of funds, called Equity I and Equity II. Equity I are “regular equity funds”, while Equity II are “Allemansfonder”, which “are part of a public savings program and offer tax benefits”. For Equity II funds, the average NAV is 1862 MSEK, the average turnover rate 0.47, and the average commission rate 0.2%. For Equity II, the corresponding numbers are 568 MSEK, 0.75, and 0.4%. On average, Equity I funds are thus smaller, have higher turnover rates, and pay higher commission rates than Equity II funds do. The differences appear reasonable in that a larger fund may negotiate lower commissions than a smaller fund. And, all else equal, the greater the NAV, the more price impact costs a given turnover incurs, and perhaps thereof the lower average turnover for Equity II funds. Even though objections can be raised, let the average Equity I fund represent a typical somewhat smaller Swedish fund in terms of NAV, turnover rate, and commission rate, and the average Equity II fund represent a somewhat larger Swedish fund in the same terms. So, to provide evidence relevant to a somewhat typical smaller and a somewhat typical larger fund in terms of NAV, the arbitrary initial portfolio is revised using their respective turnover rates, net asset values, and commissions. Additional revisions are, as a robustness check, performed of portfolios with other NAVs, turnover rates, and commissions.

The quasi-experimental nature of the empirical test conducted here should be emphasized. A typical Swedish stock fund does not hold exactly 180 stocks, and is not exactly capitalization-weighted, but lacking relevant and comprehensive data on
Swedish funds, these are reasonable choices. In addition, there is for example no risk free asset to invest in, which could possibly reflect the common policy requirement of stock funds of being (virtually) fully invested in stocks at all times. Moreover, the empirical test considers portfolio revisions at a single point in time. A natural extension would be to consider revisions over many periods. Such an analysis is provided in Chapter 4 but for various index fund revision models. The focus in this chapter is, however, on how different degrees of consideration of transaction costs relate to portfolio utility, turnover, diversification, and incurred transaction costs. If this analysis were to be performed over time, other factors, such as the correspondence between parameter forecasts and their realizations, would come into play, whereas the quasi-experimental setting used here prevents this from happen.

Law applicable to Swedish investment companies and investment funds impose upper bounds on the portfolio weights of individual holdings. For regular equity funds, individual holdings are not allowed to exceed 10% of a fund’s NAV (given that the total value of such holdings is less than 40% of the fund’s NAV; otherwise the maximum weight is 5%). In addition, the value of an individual holding must not exceed 5% of the total value of all shares outstanding in the holding (in terms of voting power). Restrictions like these are probably intended ensure a minimum degree of diversification for funds. On the other hand, however, it can be argued, that they may force funds to hold suboptimal portfolios. To shed light on the issue, the portfolio revisions are performed with and without such constraints. As an approximation to the restrictions just discussed, a maximum weight constraint, $x_{max,i}$, is imposed on any individual holding $i$, as

$$x_{max,i} = \min(0.1, 0.05 \frac{Mcap_i}{Pval})$$

where $Pval$ is the value of the portfolio in MSEK and $Mcap_i$ the market capitalization in MSEK of all outstanding shares of stock $i$.

The next subsection describes the empirical transaction costs model. Subsection 3.4.2 discusses amortization of transaction costs. In Subsection 3.4.3 the data are presented. The empirical analysis is contained in Section 3.5, and the chapter is summarized in Section 3.6.
3.4.1 Empirical model of transaction costs

Empirical models of ask and bid price impact costs were estimated in Chapter 2. The ask price impact cost, which in the context of a portfolio is the price impact cost of a buy transaction or a weight increase, is

\[
\pi^+_i = 8.82 - 0.128 \ln{MC}_i - 19.4 \ln{RTRVOL}_i - 3.25 \ln{PROPTR}_i + 0.796 \ln{VOLR}_i + 1.20 \text{ORDERSIZE}_i
\]  
(3.7)

The bid price impact cost, which in a portfolio context is the price impact cost associated with a sell or a weight decrease, is

\[
\pi^-_i = 7.47 - 0.111 \ln{MC}_i - 17.2 \ln{RTRVOL}_i - 2.14 \ln{PROPTR}_i + 0.720 \ln{VOLR}_i - 0.0727 \ln{MIDP}_i + 1.02 \text{ORDERSIZE}_i
\]  
(3.8)

In the portfolio revision program (3.6), transaction costs are expressed as \( T^*_i = c^*_i + d^*_i x^*_i \), that is, in terms of portfolio weight changes \( x^*_i \), but the models (3.7) and (3.8) of ask and bid price impact cost are expressed in \( \text{ORDERSIZE} \), defined as the number of shares in the trade divided by the total number of outstanding shares in thousands. By proper scaling of the coefficient on \( \text{ORDERSIZE} \), the equations (3.7) and (3.8) can be re-expressed in terms of portfolio weight changes \( x^*_i \).\(^{45}\) Let \( P_{val} \) be the value of the portfolio in MSEK and \( M_{cap,i} \) the market capitalization in MSEK of all outstanding shares of stock \( i \), then based on (3.7), the percentage price impact cost, \( \pi^+_i \) for a purchase of size \( x^+_i \) in security \( i \), can be represented as

\[
\pi^+_i = c^+_i + d^+_i x^+_i,
\]

where

\[
c^+_i = 8.82 - 0.128 \ln{MC}_i - 19.4 \ln{RTRVOL}_i - 3.25 \ln{PROPTR}_i + 0.796 \ln{VOLR}_i,
\]

\[
d^+_i = 1.20 \times 1000 \times P_{val} / M_{cap,i}.
\]

Similarly, for a sale transaction in security \( i \), the percentage price impact cost becomes:

\[
\pi^-_i = c^-_i + d^-_i x^-_i,
\]

where

\(^{45}\) See 3.7 Appendix B for details.
$$c_i^- = 7.47 - 0.111 \text{LNMC}_i - 17.2 RTRVOL_i - 2.14 \text{PROPTR}_i$$
$$+ 0.720 \text{LNVOLR}_i - 0.0727 \text{LNMINP}_i,$$

$$d_i^- = 1.02 \times 1000 \frac{P\text{val}}{M\text{cap}_i}.$$  

Moreover, the price impact equations do not include commissions, \(\text{comm}\), which in Sweden usually are fixed percentages of order value, identical for buys and sells. Total percentage transaction costs are for buys then obtained as \(T_i^+ = \pi_i^+ + \text{comm}\), and for sells as \(T_i^- = \pi_i^- + \text{comm}\).

According to Dahlquist et al. (2000), average commission rates for the period 1992-1997 are 0.4% and 0.2% for smaller and larger funds NAV-wise, respectively, and these are the commission rates that will be used here. In view of the findings of other studies, these numbers seem reasonable. For trades in Swedish stocks during the period September 1996 to December 1999, Domowitz et al. (2001) find that explicit costs, mainly commissions, constitute 0.26% of order value. They use proprietary data from Elkins/McSherry Co., Inc. on global trades conducted by 136 institutional traders, of whom 105 are pension funds, 27 are investment managers, and 4 are brokers. Chiyachantana et al. (2004) employ comparable data from the Plexus Group and find that average commissions are 0.13% in 2001. Perold and Sirri (1998) use similar, but less comprehensive data from State Street Global Advisors, and observe for the period 1987-1991 average commissions and taxes of 0.59%.

The forecast total transaction costs generated by the above models express the immediate cost of a transaction of given size. These immediate costs are amortized over the stocks’ expected holding period as described below, using the average turnover rate for each fund category.

### 3.4.2 Amortization of transaction costs

Models of transaction costs express the immediate cost of buying or selling a stock. In the literature, it is customary to amortize such forecasted immediate costs over the stock’s expected holding period and express them on the same (holding period) basis as the stock returns from which they are to be subtracted; usually annual returns are employed (Rudd and Clasing 1982; Grinold and Kahn 1995). The expected transaction cost per year for a single holding is then the product of (i) the roundtrip cost, i.e., the
cost of buying plus selling the holding, and (ii) the annual turnover rate which is the
inverse of the holding period in years. For instance, imagine a portfolio containing one
stock that (without price impact) costs 1% to buy and 2% to sell and that the holding
period is 2 years, then the turnover rate is 0.5, i.e. half of the portfolio is turned over
per year, and the amortized annual cost becomes 1.5%. This roundtrip cost figure is
then used for both purchase and sale transactions.

Note that accurate forecasting of the turnover rates for individual securities in a
given portfolio is important, since every such forecast has the potential to largely
influence a stock’s amortized transaction costs. Moreover, using observed turnover
rates for portfolios perhaps managed with inadequate control of transaction costs as
predictors of future turnover rates is hardly optimal. The costs for the transactions that
are not immediate are probably difficult to forecast, and are, in the context of a given
portfolio, partially endogenously determined; for example, one common predictor of a
stock’s non-immediate transaction cost is the future market value and that is a function
of the stock’s expected return. Undoubtedly, more research is needed in this area.

Furthermore, in view of the single-period utility-maximization framework, the
adequacy of the approach just described remains to be rigorously motivated. In the
empirical analysis, the amortization method just described will nevertheless be used.
As a robustness check, two other amortization schemes, whereof one was no
amortization at all were tested. The results were qualitatively similar to those
presented here.

3.4.3 Data

Average turnover, fund net asset values and commission rates emanate as mentioned
from Dahlquist et al. (2000). Included in the initial portfolio are all stocks traded in
Sweden that, as of last Dec. 1999, have market capitalization, based on all shares
outstanding, above 10 MSEK, and 5 years of price data. Arithmetically annualized
returns were computed from (splits adjusted) price data of monthly frequency. The
empirical work in this chapter and in Chapter 4 are based on the same data base. As

46 For instance, when using amortized transaction costs one is not strictly maximizing end-of-a-single-period
portfolio value. However, using unamortized transaction costs means that one maximizes end-of-a-single-period
portfolio value but before the transaction costs needed to liquidate the portfolio.
only a small subset of that data base are used here, the reader is referred to Section 4.3.10 for a complete description of the data base. The variable $MIDP$ in the price impact costs models will be represented by the closing price.

Table 3.1 reveals that there is considerable variation in all variables. The maximum portfolio weight of 35% refers to Ericsson.

Table 3.1 Descriptive statistics for the capitalization-weighted initial portfolio
All figures are as of the close of the market 29 Dec. 1999. Annual mean return and annual standard deviation are the arithmetically annualized mean return and standard deviation of the stocks in a given portfolio based on 5 years of monthly data. Market cap. is the market capitalization in million SEK based on all shares outstanding. Price is the closing price in SEK.

<table>
<thead>
<tr>
<th></th>
<th>Annual mean return (%)</th>
<th>Annual standard deviation (%)</th>
<th>Market cap. (MSEK)</th>
<th>Price (SEK)</th>
<th>Weights of initial portfolio (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(n=180)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>17.8</td>
<td>34.6</td>
<td>15381.7</td>
<td>151.4</td>
<td>0.6</td>
</tr>
<tr>
<td>Std</td>
<td>18.5</td>
<td>13.7</td>
<td>76353.1</td>
<td>222.2</td>
<td>2.8</td>
</tr>
<tr>
<td>Min</td>
<td>-27.0</td>
<td>19.1</td>
<td>14.8</td>
<td>0.47</td>
<td>0.00</td>
</tr>
<tr>
<td>25%</td>
<td>8.5</td>
<td>27.0</td>
<td>382.3</td>
<td>48.5</td>
<td>0.01</td>
</tr>
<tr>
<td>Median</td>
<td>16.4</td>
<td>31.2</td>
<td>1464.6</td>
<td>99</td>
<td>0.05</td>
</tr>
<tr>
<td>75%</td>
<td>24.0</td>
<td>39.7</td>
<td>6512.1</td>
<td>177</td>
<td>0.24</td>
</tr>
<tr>
<td>Max</td>
<td>80.2</td>
<td>105.9</td>
<td>980587.6</td>
<td>1730</td>
<td>35.42</td>
</tr>
</tbody>
</table>

3.5 Empirical analysis

The details of the procedure employed for comparing portfolio revisions performed with varying degrees of regard to transaction costs are as follows. An initial, capitalization-weighted portfolio is formed based on all stocks meeting the data requirements. The portfolio’s mean return vector and covariance matrix required as inputs to the portfolio revision program are estimated by the corresponding sample moments using 5 years of annualized data. Portfolio weights are restricted to be non-negative, i.e., short selling is prohibited, reflecting either a common institutional constraint, or a means to reduce error maximization or portfolio turnover, or all three.

The initial portfolio is then revised for four different degrees of consideration of transaction costs. The resulting portfolios are labeled accordingly as follows: $P_1$ is obtained using the full model of transaction costs including price impact, i.e., $x_i^*(c_i^* + comm + d_i^*x_i^*)$; $P_2$ neglects the quadratic part and uses $x_i^*(c_i^* + comm)$, effectively assuming constant percentage transaction costs, that is, price impact costs are not explicitly modeled as functions of order size or weight change; $P_3$ neglects the
linear part and uses $x_i^t dx_i^t dx_i^t$, which allows the percentage cost to vary as a function of trade size; and $P_4$ is obtained using the standard MV model and neglects transaction costs altogether.

For a given capitalization segment this is done for seven different levels of absolute risk aversion ($ARA$) from 2 to 8 covering a broad spectrum of attitudes towards risk. According to Chopra and Ziemba (1993) an $ARA$ of 2 represents a very aggressive investor, an $ARA$ of 4 a typical institutional investor, and an $ARA$ of 8 a very conservative investor.

The revised portfolios $P_2$, $P_3$, and $P_4$, obtained with less than full consideration of transaction costs, are optimal only in the true absence of the (part of the) transaction costs neglected, and will, net of the neglected transaction costs, be less than optimal. Two returns will be reported for each revised portfolio: $R_p$ is a revised portfolio’s return before any disregarded transaction costs are deducted, and $R_p^*$ is the portfolio return net of all (regarded and disregarded) transaction costs. $R_p^*$ is calculated as $R_p$ minus any neglected transaction costs incurred in the course of a revision (it is supposed that other portfolio features are unaffected by the ignored transaction costs, e.g. portfolio variance). The true optimal revised portfolio, $P_1$, is assumed obtained when the revision is performed with the full model of transaction costs and with the transaction costs amortized according to Rudd and Clasing (1982) and Grinold and Kahn (1995). I thus want to compare the optimal portfolio $P_1$ to the suboptimal portfolios $P_2$, $P_3$, and $P_4$, respectively, and assess the utility losses that are due to the neglected transaction costs.

Meaningful comparisons of the expected utility scores of different portfolios may be done in terms of the portfolios’ cash or certainty equivalents as suggested by Dexter et al. (1980). The certainty equivalent ($CE$) of a risky portfolio is the certain, or risk-free, gross return (i.e., one plus the rate of return) that yields the same utility as the risky portfolio does. The $CE$ is an appropriate measure to base comparisons on, since it takes into account the investor’s risk aversion and the inherent uncertainty in returns, and it is independent of utility units (Chopra and Ziemba 1993). Assume that returns are jointly normally distributed, and that the investor has a negative exponential utility function, then the certainty equivalent is
\[ CE = (1+r)'x - ARA/2 \ x'V \ x, \]
where \( x \) is the vector of portfolio weights, \( r \) the vector of mean returns, \( V \) the covariance matrix, and \( ARA \) is the investor’s (constant) absolute risk aversion parameter (expressed in gross returns). See 3.7 Appendix A for derivation and details.

The percentage certainty equivalent loss (CEL) from holding the suboptimal portfolio \( P_i \), where \( i=2, 3, 4 \), instead of the optimal portfolio \( P_1 \), is calculated as

\[ CEL_i = \frac{CE_{R_i} - CE_{P_i}}{CE_{R_i}}, \]

where \( CE_{R_i} \) and \( CE_{P_i} \) are the respective certainty equivalents of portfolios \( P_1 \) and \( P_i \).

Caution must be exercised in comparing to each other the results obtained for revisions of portfolios which differ with regard to capitalization, risk aversion and constraints, since the result for a given revised portfolio is sensitive to the relative efficiency, or optimality, of the initial portfolio. For instance, for a given level of risk aversion, the smaller-NAV initial portfolio may be close to optimal and relatively less transactions are therefore probably needed to achieve the optimal revised portfolio, while the initial larger-NAV portfolio may be relatively inefficient necessitating high turnover to reach the optimal portfolio. The improvement or disimprovement realized by moving from the initial portfolio \( P_0 \) to the optimal revised portfolio \( P_1 \) obtained with full regard to transaction costs for a given \( ARA \), is measured in analogy with the certainty equivalent loss. This measure of portfolio proximity is defined as the signed percentage certainty equivalent difference or distance:

\[ CED = \frac{CE_{P_1} - CE_{P_0}}{CE_{P_1}}, \]

where \( CE_{P_1} \) and \( CE_{P_0} \) are, for a given \( ARA \), the certainty equivalents of the optimal revised portfolio and the initial portfolio, respectively. The measure will be non-negative, since the initial portfolio will be unchanged if no improvement is possible. One may, however, expect some improvement, as the initial portfolio is (arbitrarily) capitalization-weighted and therefore, most likely, suboptimal to the optimal portfolio revised with full regard to transaction costs.

First, some important aspects of the portfolio revisions performed with full and no regard to transaction costs, respectively, are described. Second, results are presented concerning utility losses experienced due to neglect of transaction costs relative to full
consideration of those costs. These two steps are then repeated but with focus on the relative importance of the quadratic and linear part of the transaction costs specification. Up to this point, the analysis will have concerned the results in Table 3.2 and Table 3.4 of portfolio revisions performed without a maximum weight constraint imposed. Lastly, the results in Table 3.3 and Table 3.5 of portfolio revisions under maximum weight constraints are reviewed and compared to the results obtained without such constraints.

3.5.1 Transaction costs incurred with the full transaction cost model

With the full quadratic model of transaction costs, for all the levels of risk aversion considered and for both the smaller and the larger NAV portfolios, the revisions generate substantial - compared to typical levels of portfolio returns - transaction costs. According to Table 3.2, revision of the larger-NAV portfolio results in total expected amortized transaction costs ($TC$) varying from 5.30% to 5.94%.\(^{47}\) For the smaller-NAV portfolio in Table 3.4, $TC$ varies from 5.36% to 5.86%.

Incidentally, for both smaller and larger-NAV portfolio revisions with full consideration of transaction costs, $TC$ decreases almost monotonically with $ARA$, the level of absolute risk aversion. For instance, moving from the initial portfolio to the optimal portfolio for $ARA=2$ generates the largest $TC$. However, inspection of the percentage certainty equivalent distances $CED$s between the initial portfolio and the optimal ones ($P_1$) in Table 3.2, reveals that $TC$ is positively correlated with those distances, and that the optimal large-cap portfolio for $ARA=2$, with the highest $TC$, is in fact farthest from the initial portfolio as measured by $CED$. The observation that $TC$ increases in $CED$ applies to Table 3.4 as well, and suggests that the certainty equivalent distance $CED$ may be used to explain the amount of transaction cost different portfolio revisions generate. Overall, the results in Table 3.2 and Table 3.4 for given $ARA$ and degree of account of transaction costs are similar for smaller and larger portfolios NAV-wise.

\(^{47}\) Note that these are expected amortized transaction costs, and not the transaction costs incurred immediately in association with the portfolio revision.
Table 3.2 Larger-NAV portfolio revisions without maximum weight constraints

Revision with Full ($P_1$), (L)inear part ($P_2$), (Q)uadratic part ($P_3$) and Nil ($P_4$) regard to transaction costs for negative exponential investors with different constant absolute risk aversion coefficients ($ARA$). Net asset value of unrevised portfolio is 1862 MSEK and its turnover rate 0.47. The mean vector and covariance matrix of returns were estimated by the sample moments using 5 years of monthly returns annualized arithmetically. The mean and standard deviation of the unrevised portfolio’s return are 41.5%, and 24.4%, respectively.

$$Rp$$ is portfolio mean annual return before any disregarded transaction costs are deducted, $$Sp$$ is portfolio standard deviation, $$TC$$ is total expected amortized transaction costs, $$TCL$$ is the linear part of $$TC$$, $$TCQ$$ is the quadratic part of $$TC$$, $$TO$$ is turnover, $$Rp^*$$ is portfolio return net of all (regarded and disregarded) transaction costs, $$CE$$ is the certainty equivalent gross return for negative exponential preferences and normal returns based on $$Rp^*$$, $$CED$$ is the certainty equivalent distance of a given portfolio based on $$Rp^*$$ relative to the initial portfolio, and $$CEL$$ is the certainty equivalent loss relative to the portfolio obtained under full consideration of transaction costs. $N_{>0}$ is the number of stocks (with weights>1e-9). $N_{>0.1}$ is the number of stocks whose weights exceed 0.10. $Maxx$ is the weight of the largest holding.

<table>
<thead>
<tr>
<th>$ARA$</th>
<th>$Rp$ (%)</th>
<th>$Sp$ (%)</th>
<th>$TC$ (%)</th>
<th>$TCL$ (%)</th>
<th>$TCQ$ (%)</th>
<th>$TO$ (%)</th>
<th>$Rp^*$ (%)</th>
<th>$CE$ (%)</th>
<th>$CED$ (%)</th>
<th>$CEL$ (%)</th>
<th>$N_{&gt;0}$</th>
<th>$N_{&gt;0.1}$</th>
<th>$N_{Law}$</th>
<th>$Maxx$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>$P_1$ Full</td>
<td>66.9</td>
<td>25.0</td>
<td>5.94</td>
<td>1.36</td>
<td>4.59</td>
<td>82</td>
<td>66.9</td>
<td>1.607</td>
<td>39.1</td>
<td>11</td>
<td>4</td>
<td>4</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>$P_1$ L</td>
<td>71.3</td>
<td>27.3</td>
<td>23.78</td>
<td>1.58</td>
<td>22.20</td>
<td>93</td>
<td>49.1</td>
<td>1.417</td>
<td>11.9</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>$P_1$ Q</td>
<td>67.0</td>
<td>24.9</td>
<td>6.04</td>
<td>1.39</td>
<td>4.65</td>
<td>83</td>
<td>65.6</td>
<td>1.594</td>
<td>0.8</td>
<td>10</td>
<td>4</td>
<td>4</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>$P_1$ Nil</td>
<td>71.2</td>
<td>27.1</td>
<td>23.02</td>
<td>1.61</td>
<td>21.41</td>
<td>94</td>
<td>48.2</td>
<td>1.409</td>
<td>12.3</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>0.39</td>
</tr>
<tr>
<td>3</td>
<td>$P_1$ Full</td>
<td>65.9</td>
<td>23.6</td>
<td>5.75</td>
<td>1.36</td>
<td>4.39</td>
<td>80</td>
<td>65.9</td>
<td>1.576</td>
<td>37.7</td>
<td>15</td>
<td>4</td>
<td>4</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>$P_1$ L</td>
<td>69.3</td>
<td>24.4</td>
<td>18.33</td>
<td>1.44</td>
<td>16.89</td>
<td>85</td>
<td>52.5</td>
<td>1.435</td>
<td>8.9</td>
<td>8</td>
<td>3</td>
<td>4</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>$P_1$ Q</td>
<td>66.1</td>
<td>23.7</td>
<td>5.95</td>
<td>1.41</td>
<td>4.54</td>
<td>82</td>
<td>64.7</td>
<td>1.563</td>
<td>0.8</td>
<td>12</td>
<td>4</td>
<td>4</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>$P_1$ Nil</td>
<td>70.0</td>
<td>25.1</td>
<td>21.67</td>
<td>1.61</td>
<td>20.06</td>
<td>93</td>
<td>48.3</td>
<td>1.388</td>
<td>11.9</td>
<td>7</td>
<td>3</td>
<td>4</td>
<td>0.38</td>
</tr>
<tr>
<td>4</td>
<td>$P_1$ Full</td>
<td>64.8</td>
<td>22.5</td>
<td>5.50</td>
<td>1.33</td>
<td>4.21</td>
<td>77</td>
<td>64.8</td>
<td>1.547</td>
<td>36.5</td>
<td>17</td>
<td>4</td>
<td>4</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>$P_1$ L</td>
<td>68.6</td>
<td>23.5</td>
<td>18.73</td>
<td>1.44</td>
<td>17.30</td>
<td>84</td>
<td>51.3</td>
<td>1.403</td>
<td>9.3</td>
<td>7</td>
<td>3</td>
<td>3</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>$P_1$ Q</td>
<td>65.2</td>
<td>22.7</td>
<td>5.80</td>
<td>1.40</td>
<td>4.40</td>
<td>80</td>
<td>63.8</td>
<td>1.535</td>
<td>0.7</td>
<td>15</td>
<td>4</td>
<td>4</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>$P_1$ Nil</td>
<td>68.5</td>
<td>23.4</td>
<td>18.24</td>
<td>1.47</td>
<td>16.77</td>
<td>84</td>
<td>50.3</td>
<td>1.394</td>
<td>9.9</td>
<td>7</td>
<td>3</td>
<td>3</td>
<td>0.43</td>
</tr>
<tr>
<td>5</td>
<td>$P_1$ Full</td>
<td>64.2</td>
<td>21.9</td>
<td>5.36</td>
<td>1.35</td>
<td>4.01</td>
<td>76</td>
<td>63.5</td>
<td>1.499</td>
<td>34.9</td>
<td>26</td>
<td>4</td>
<td>4</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>$P_1$ L</td>
<td>67.8</td>
<td>22.7</td>
<td>17.62</td>
<td>1.47</td>
<td>16.16</td>
<td>84</td>
<td>51.7</td>
<td>1.388</td>
<td>8.8</td>
<td>7</td>
<td>3</td>
<td>3</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>$P_1$ Q</td>
<td>64.3</td>
<td>21.9</td>
<td>5.64</td>
<td>1.37</td>
<td>4.27</td>
<td>77</td>
<td>62.9</td>
<td>1.509</td>
<td>0.9</td>
<td>17</td>
<td>4</td>
<td>4</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>$P_1$ Nil</td>
<td>67.7</td>
<td>22.6</td>
<td>16.86</td>
<td>1.49</td>
<td>15.37</td>
<td>84</td>
<td>50.8</td>
<td>1.381</td>
<td>9.3</td>
<td>7</td>
<td>4</td>
<td>4</td>
<td>0.41</td>
</tr>
<tr>
<td>6</td>
<td>$P_1$ Full</td>
<td>63.5</td>
<td>21.3</td>
<td>5.54</td>
<td>1.36</td>
<td>4.18</td>
<td>76</td>
<td>63.5</td>
<td>1.499</td>
<td>34.9</td>
<td>26</td>
<td>4</td>
<td>4</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>$P_1$ L</td>
<td>67.2</td>
<td>22.2</td>
<td>16.78</td>
<td>1.49</td>
<td>15.28</td>
<td>84</td>
<td>52.0</td>
<td>1.372</td>
<td>8.5</td>
<td>8</td>
<td>3</td>
<td>3</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>$P_1$ Q</td>
<td>63.5</td>
<td>21.2</td>
<td>5.59</td>
<td>1.38</td>
<td>4.22</td>
<td>77</td>
<td>62.1</td>
<td>1.486</td>
<td>0.9</td>
<td>18</td>
<td>4</td>
<td>4</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>$P_1$ Nil</td>
<td>67.1</td>
<td>22.1</td>
<td>16.13</td>
<td>1.51</td>
<td>14.62</td>
<td>84</td>
<td>51.0</td>
<td>1.363</td>
<td>9.0</td>
<td>8</td>
<td>3</td>
<td>3</td>
<td>0.41</td>
</tr>
<tr>
<td>7</td>
<td>$P_1$ Full</td>
<td>62.6</td>
<td>20.7</td>
<td>5.45</td>
<td>1.35</td>
<td>4.10</td>
<td>76</td>
<td>62.6</td>
<td>1.476</td>
<td>34.2</td>
<td>33</td>
<td>4</td>
<td>4</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>$P_1$ L</td>
<td>65.9</td>
<td>21.2</td>
<td>18.03</td>
<td>1.52</td>
<td>16.51</td>
<td>83</td>
<td>49.4</td>
<td>1.336</td>
<td>9.5</td>
<td>11</td>
<td>3</td>
<td>4</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>$P_1$ Q</td>
<td>62.7</td>
<td>20.7</td>
<td>5.56</td>
<td>1.38</td>
<td>4.18</td>
<td>76</td>
<td>61.3</td>
<td>1.463</td>
<td>0.9</td>
<td>25</td>
<td>4</td>
<td>4</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>$P_1$ Nil</td>
<td>65.7</td>
<td>21.1</td>
<td>19.94</td>
<td>1.57</td>
<td>18.38</td>
<td>84</td>
<td>45.8</td>
<td>1.302</td>
<td>11.8</td>
<td>9</td>
<td>4</td>
<td>5</td>
<td>0.38</td>
</tr>
<tr>
<td>8</td>
<td>$P_1$ Full</td>
<td>61.4</td>
<td>20.0</td>
<td>5.30</td>
<td>1.33</td>
<td>3.97</td>
<td>74</td>
<td>61.4</td>
<td>1.454</td>
<td>33.6</td>
<td>37</td>
<td>4</td>
<td>4</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>$P_1$ L</td>
<td>64.0</td>
<td>20.0</td>
<td>24.77</td>
<td>1.51</td>
<td>23.26</td>
<td>80</td>
<td>40.8</td>
<td>1.247</td>
<td>14.2</td>
<td>12</td>
<td>3</td>
<td>5</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>$P_1$ Q</td>
<td>61.8</td>
<td>20.1</td>
<td>5.51</td>
<td>1.38</td>
<td>4.13</td>
<td>76</td>
<td>60.4</td>
<td>1.442</td>
<td>0.8</td>
<td>30</td>
<td>4</td>
<td>4</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>$P_1$ Nil</td>
<td>63.9</td>
<td>19.9</td>
<td>13.40</td>
<td>1.61</td>
<td>29.78</td>
<td>83</td>
<td>32.5</td>
<td>1.166</td>
<td>19.8</td>
<td>11</td>
<td>4</td>
<td>6</td>
<td>0.35</td>
</tr>
</tbody>
</table>
Table 3.3: Larger-NAV portfolio revisions with maximum weight constraints
Revision with Full ($P_1$), (L)inear part ($P_2$), (Q)uadratic part ($P_3$) and Nil ($P_4$) regard to transaction costs for negative exponential investors with different constant absolute risk aversion coefficients (ARA). Net asset value of unrevised portfolio is 1862 MSEK and its turnover rate 0.47. The mean vector and covariance matrix of returns were estimated by the sample moments using 5 years of monthly returns annualized arithmetically. The mean and standard deviation of the unrevised portfolio’s return are 41.5%, and 24.4%, respectively. $R_p$ is portfolio mean annual return before any disregarded transaction costs are deducted, $S_p$ is portfolio standard deviation, $TC$ is total expected amortized transaction costs, $TC_L$ is the linear part of $TC$, $TC_Q$ is the quadratic part of $TC$, $TO$ is turnover, $R_p^*$ is portfolio return net of all (regarded and disregarded) transaction costs, $CE$ is the certainty equivalent gross return for negative exponential preferences and normal returns based on $R_p^*$, $CED$ is the certainty equivalent distance of a given portfolio based on $R_p^*$ relative to the initial portfolio, and $CEL$ is the certainty equivalent loss relative to the portfolio obtained under full consideration of transaction costs. $N_{c=0}$ is the number of stocks (with weights > 1e-9). $N_{c=0.1}$ is the number of stocks whose weights exceed 0.10. $Max$ is the weight of the largest holding. $CEwL$ is percentage certainty equivalent loss due to weight constraints.

<table>
<thead>
<tr>
<th>ARA</th>
<th>$R_p$ (%)</th>
<th>$S_p$ (%)</th>
<th>$TC$ (%)</th>
<th>$TC_L$ (%)</th>
<th>$TC_Q$ (%)</th>
<th>$TO$ (%)</th>
<th>$R_p^*$ (%)</th>
<th>$CE$ (%)</th>
<th>$CED$ (%)</th>
<th>$CEL$ (%)</th>
<th>$N_{c=0}$</th>
<th>$N_{c=0.1}$</th>
<th>$CEwL$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>$P_1$ Full</td>
<td>60.2</td>
<td>23.4</td>
<td>4.13</td>
<td>1.15</td>
<td>2.98</td>
<td>63</td>
<td>60.2</td>
<td>1.547</td>
<td>33.9</td>
<td>27</td>
<td>9</td>
<td>3.7</td>
</tr>
<tr>
<td></td>
<td>$P_2$ L</td>
<td>65.0</td>
<td>26.2</td>
<td>17.40</td>
<td>1.19</td>
<td>16.22</td>
<td>71</td>
<td>48.8</td>
<td>1.419</td>
<td>8.3</td>
<td>16</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$P_3$ Q</td>
<td>60.2</td>
<td>23.4</td>
<td>4.12</td>
<td>1.16</td>
<td>2.96</td>
<td>63</td>
<td>59.0</td>
<td>1.536</td>
<td>0.8</td>
<td>24</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$P_4$ Nil</td>
<td>64.9</td>
<td>26.0</td>
<td>17.64</td>
<td>1.27</td>
<td>16.37</td>
<td>74</td>
<td>47.3</td>
<td>1.405</td>
<td>9.2</td>
<td>14</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>$P_1$ Full</td>
<td>59.6</td>
<td>22.8</td>
<td>3.83</td>
<td>1.10</td>
<td>2.73</td>
<td>61</td>
<td>59.6</td>
<td>1.517</td>
<td>32.6</td>
<td>34</td>
<td>8</td>
<td>3.7</td>
</tr>
<tr>
<td></td>
<td>$P_2$ L</td>
<td>64.6</td>
<td>25.5</td>
<td>17.43</td>
<td>1.23</td>
<td>16.20</td>
<td>74</td>
<td>48.4</td>
<td>1.386</td>
<td>8.7</td>
<td>16</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$P_3$ Q</td>
<td>59.9</td>
<td>23.2</td>
<td>3.94</td>
<td>1.15</td>
<td>2.79</td>
<td>63</td>
<td>58.7</td>
<td>1.507</td>
<td>0.7</td>
<td>29</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$P_4$ Nil</td>
<td>64.9</td>
<td>25.4</td>
<td>18.39</td>
<td>1.33</td>
<td>17.06</td>
<td>77</td>
<td>46.2</td>
<td>1.365</td>
<td>10.1</td>
<td>14</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>$P_1$ Full</td>
<td>58.9</td>
<td>22.2</td>
<td>3.66</td>
<td>1.06</td>
<td>2.60</td>
<td>59</td>
<td>58.9</td>
<td>1.490</td>
<td>31.5</td>
<td>40</td>
<td>8</td>
<td>3.6</td>
</tr>
<tr>
<td></td>
<td>$P_2$ L</td>
<td>63.1</td>
<td>23.7</td>
<td>14.86</td>
<td>1.26</td>
<td>13.60</td>
<td>74</td>
<td>49.5</td>
<td>1.382</td>
<td>7.3</td>
<td>16</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$P_3$ Q</td>
<td>59.0</td>
<td>22.3</td>
<td>3.67</td>
<td>1.08</td>
<td>2.59</td>
<td>60</td>
<td>57.9</td>
<td>1.480</td>
<td>0.7</td>
<td>35</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$P_4$ Nil</td>
<td>63.3</td>
<td>24.0</td>
<td>16.29</td>
<td>1.33</td>
<td>14.96</td>
<td>77</td>
<td>47.1</td>
<td>1.356</td>
<td>9.0</td>
<td>16</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>$P_1$ Full</td>
<td>58.1</td>
<td>21.5</td>
<td>3.55</td>
<td>1.04</td>
<td>2.51</td>
<td>59</td>
<td>58.1</td>
<td>1.466</td>
<td>30.6</td>
<td>47</td>
<td>8</td>
<td>3.7</td>
</tr>
<tr>
<td></td>
<td>$P_2$ L</td>
<td>62.4</td>
<td>23.1</td>
<td>15.00</td>
<td>1.24</td>
<td>13.76</td>
<td>73</td>
<td>48.6</td>
<td>1.353</td>
<td>7.7</td>
<td>21</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$P_3$ Q</td>
<td>58.1</td>
<td>21.5</td>
<td>3.57</td>
<td>1.06</td>
<td>2.51</td>
<td>60</td>
<td>57.1</td>
<td>1.455</td>
<td>0.7</td>
<td>39</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$P_4$ Nil</td>
<td>62.5</td>
<td>23.2</td>
<td>15.58</td>
<td>1.27</td>
<td>14.31</td>
<td>74</td>
<td>46.9</td>
<td>1.335</td>
<td>8.9</td>
<td>17</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>$P_1$ Full</td>
<td>57.3</td>
<td>20.8</td>
<td>3.53</td>
<td>1.06</td>
<td>2.47</td>
<td>60</td>
<td>57.3</td>
<td>1.443</td>
<td>29.9</td>
<td>51</td>
<td>7</td>
<td>3.7</td>
</tr>
<tr>
<td></td>
<td>$P_2$ L</td>
<td>60.6</td>
<td>21.6</td>
<td>13.30</td>
<td>1.24</td>
<td>12.06</td>
<td>73</td>
<td>48.5</td>
<td>1.345</td>
<td>6.8</td>
<td>27</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$P_3$ Q</td>
<td>57.2</td>
<td>20.7</td>
<td>3.68</td>
<td>1.11</td>
<td>2.57</td>
<td>63</td>
<td>56.1</td>
<td>1.433</td>
<td>0.7</td>
<td>44</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$P_4$ Nil</td>
<td>60.2</td>
<td>21.3</td>
<td>14.05</td>
<td>1.37</td>
<td>12.68</td>
<td>77</td>
<td>46.2</td>
<td>1.326</td>
<td>8.1</td>
<td>18</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>$P_1$ Full</td>
<td>56.2</td>
<td>20.0</td>
<td>3.56</td>
<td>1.07</td>
<td>2.49</td>
<td>62</td>
<td>56.2</td>
<td>1.422</td>
<td>29.3</td>
<td>56</td>
<td>7</td>
<td>3.7</td>
</tr>
<tr>
<td></td>
<td>$P_2$ L</td>
<td>59.0</td>
<td>20.4</td>
<td>13.26</td>
<td>1.29</td>
<td>11.97</td>
<td>75</td>
<td>47.0</td>
<td>1.324</td>
<td>6.9</td>
<td>30</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$P_3$ Q</td>
<td>56.3</td>
<td>19.9</td>
<td>3.73</td>
<td>1.13</td>
<td>2.60</td>
<td>64</td>
<td>55.1</td>
<td>1.412</td>
<td>0.7</td>
<td>50</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$P_4$ Nil</td>
<td>58.7</td>
<td>20.2</td>
<td>13.36</td>
<td>1.39</td>
<td>11.97</td>
<td>77</td>
<td>45.3</td>
<td>1.311</td>
<td>7.8</td>
<td>21</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>$P_1$ Full</td>
<td>55.3</td>
<td>19.3</td>
<td>3.61</td>
<td>1.09</td>
<td>2.52</td>
<td>64</td>
<td>55.3</td>
<td>1.403</td>
<td>28.9</td>
<td>62</td>
<td>7</td>
<td>3.5</td>
</tr>
<tr>
<td></td>
<td>$P_2$ L</td>
<td>57.5</td>
<td>19.4</td>
<td>13.72</td>
<td>1.32</td>
<td>12.39</td>
<td>76</td>
<td>45.1</td>
<td>1.301</td>
<td>7.3</td>
<td>29</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$P_3$ Q</td>
<td>55.3</td>
<td>19.3</td>
<td>3.78</td>
<td>1.15</td>
<td>2.63</td>
<td>66</td>
<td>54.2</td>
<td>1.393</td>
<td>0.7</td>
<td>54</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$P_4$ Nil</td>
<td>57.4</td>
<td>19.3</td>
<td>13.65</td>
<td>1.42</td>
<td>12.23</td>
<td>77</td>
<td>43.7</td>
<td>1.289</td>
<td>8.2</td>
<td>21</td>
<td>12</td>
<td></td>
</tr>
</tbody>
</table>
Table 3.4 Smaller-NAV portfolio revisions without maximum weight constraints
Revision with Full ($P_1$), (L)inear part ($P_2$), (Q)uadratic part ($P_3$) and Nil ($P_4$) regard to transaction costs for negative exponential investors with different constant absolute risk aversion coefficients ($ARA$). Net asset value of unrevised portfolio is 568 MSEK and its turnover rate 0.75. The mean vector and covariance matrix of returns were estimated by the sample moments using 5 years of monthly returns annualized arithmetically. The mean and standard deviation of the unrevised portfolio’s return are 41.5%, and 24.4%, respectively. $R_p$ is portfolio mean annual return before any disregarded transaction costs are deducted, $S_p$ is portfolio standard deviation, $TC$ is total expected amortized transaction costs, $TC_L$ is the linear part of $TC$, $TC_Q$ is the quadratic part of $TC$, $TO$ is turnover, $R_p^*$ is portfolio return net of all (regarded and disregarded) transaction costs, $CE$ is the certainty equivalent gross return for negative exponential preferences and normal returns based on $R_p^*$, $CED$ is the certainty equivalent distance of a given portfolio based on $R_p^*$ relative to the initial portfolio, and $CEL$ is the certainty equivalent loss relative to the portfolio obtained under full consideration of transaction costs. $N_{x>0}$ is the number of stocks (with weights $>1e-9$). $N_{x>0.1}$ is the number of stocks whose weights exceed 0.10. Maxx is the weight of the largest holding.

<table>
<thead>
<tr>
<th>$ARA$</th>
<th>$R_p$</th>
<th>$S_p$</th>
<th>$TC$</th>
<th>$TC_L$</th>
<th>$TC_Q$</th>
<th>$TO$</th>
<th>$R_p^*$</th>
<th>$CE$</th>
<th>$CED$</th>
<th>$CEL$</th>
<th>$N_{x&gt;0}$</th>
<th>$N_{x&gt;0.1}$</th>
<th>$N_{x=Law}$</th>
<th>Maxx</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>$P_1$ Full</td>
<td>68.6</td>
<td>25.6</td>
<td>5.86</td>
<td>2.69</td>
<td>3.17</td>
<td>84</td>
<td>68.6</td>
<td>1.621</td>
<td>40.2</td>
<td>9</td>
<td>4</td>
<td>4</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td>$P_2$ L</td>
<td>70.9</td>
<td>26.8</td>
<td>13.01</td>
<td>2.87</td>
<td>10.14</td>
<td>88</td>
<td>60.7</td>
<td>1.536</td>
<td>5.2</td>
<td>8</td>
<td>4</td>
<td>4</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>$P_3$ Q</td>
<td>68.3</td>
<td>25.1</td>
<td>5.88</td>
<td>2.74</td>
<td>3.14</td>
<td>84</td>
<td>65.6</td>
<td>1.593</td>
<td>1.7</td>
<td>9</td>
<td>4</td>
<td>4</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>$P_4$ Nil</td>
<td>71.2</td>
<td>27.1</td>
<td>13.55</td>
<td>3.13</td>
<td>10.42</td>
<td>94</td>
<td>57.7</td>
<td>1.503</td>
<td>7.2</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>0.39</td>
</tr>
<tr>
<td>3</td>
<td>$P_1$ Full</td>
<td>67.4</td>
<td>23.9</td>
<td>5.70</td>
<td>2.71</td>
<td>2.99</td>
<td>82</td>
<td>67.4</td>
<td>1.589</td>
<td>38.8</td>
<td>11</td>
<td>4</td>
<td>4</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>$P_2$ L</td>
<td>69.2</td>
<td>24.3</td>
<td>10.81</td>
<td>2.75</td>
<td>8.06</td>
<td>84</td>
<td>61.2</td>
<td>1.523</td>
<td>4.1</td>
<td>8</td>
<td>3</td>
<td>3</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>$P_3$ Q</td>
<td>67.5</td>
<td>23.8</td>
<td>5.93</td>
<td>2.83</td>
<td>3.10</td>
<td>84</td>
<td>64.7</td>
<td>1.562</td>
<td>1.7</td>
<td>9</td>
<td>4</td>
<td>4</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>$P_4$ Nil</td>
<td>70.0</td>
<td>25.1</td>
<td>12.89</td>
<td>3.13</td>
<td>9.77</td>
<td>93</td>
<td>57.1</td>
<td>1.476</td>
<td>7.1</td>
<td>7</td>
<td>3</td>
<td>3</td>
<td>0.43</td>
</tr>
<tr>
<td>4</td>
<td>$P_1$ Full</td>
<td>65.4</td>
<td>21.8</td>
<td>5.44</td>
<td>2.62</td>
<td>2.82</td>
<td>77</td>
<td>65.4</td>
<td>1.534</td>
<td>36.7</td>
<td>17</td>
<td>4</td>
<td>4</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>$P_2$ L</td>
<td>67.9</td>
<td>22.6</td>
<td>10.93</td>
<td>2.82</td>
<td>8.11</td>
<td>84</td>
<td>60.2</td>
<td>1.491</td>
<td>4.5</td>
<td>8</td>
<td>3</td>
<td>3</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>$P_3$ Q</td>
<td>66.8</td>
<td>23.0</td>
<td>5.95</td>
<td>2.85</td>
<td>3.10</td>
<td>83</td>
<td>64.0</td>
<td>1.535</td>
<td>1.7</td>
<td>11</td>
<td>4</td>
<td>4</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>$P_4$ Nil</td>
<td>68.5</td>
<td>23.4</td>
<td>11.01</td>
<td>2.84</td>
<td>8.16</td>
<td>84</td>
<td>57.5</td>
<td>1.466</td>
<td>6.1</td>
<td>7</td>
<td>3</td>
<td>3</td>
<td>0.42</td>
</tr>
<tr>
<td>5</td>
<td>$P_1$ Full</td>
<td>65.4</td>
<td>21.9</td>
<td>5.44</td>
<td>2.62</td>
<td>2.82</td>
<td>77</td>
<td>65.4</td>
<td>1.534</td>
<td>36.7</td>
<td>17</td>
<td>4</td>
<td>4</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>$P_2$ L</td>
<td>67.9</td>
<td>22.8</td>
<td>10.93</td>
<td>2.82</td>
<td>8.11</td>
<td>84</td>
<td>60.2</td>
<td>1.491</td>
<td>4.3</td>
<td>8</td>
<td>3</td>
<td>3</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>$P_3$ Q</td>
<td>66.8</td>
<td>23.0</td>
<td>5.95</td>
<td>2.85</td>
<td>3.10</td>
<td>83</td>
<td>64.0</td>
<td>1.535</td>
<td>1.7</td>
<td>11</td>
<td>4</td>
<td>4</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>$P_4$ Nil</td>
<td>67.7</td>
<td>22.6</td>
<td>10.36</td>
<td>2.88</td>
<td>7.48</td>
<td>84</td>
<td>57.3</td>
<td>1.446</td>
<td>5.7</td>
<td>7</td>
<td>4</td>
<td>4</td>
<td>0.41</td>
</tr>
<tr>
<td>6</td>
<td>$P_1$ Full</td>
<td>64.7</td>
<td>21.3</td>
<td>5.44</td>
<td>2.65</td>
<td>2.79</td>
<td>77</td>
<td>64.7</td>
<td>1.510</td>
<td>36.0</td>
<td>19</td>
<td>4</td>
<td>4</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>$P_2$ L</td>
<td>67.2</td>
<td>22.2</td>
<td>10.34</td>
<td>2.83</td>
<td>5.71</td>
<td>83</td>
<td>59.3</td>
<td>1.449</td>
<td>4.0</td>
<td>10</td>
<td>3</td>
<td>3</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>$P_3$ Q</td>
<td>65.1</td>
<td>21.5</td>
<td>5.75</td>
<td>2.81</td>
<td>2.94</td>
<td>80</td>
<td>62.2</td>
<td>1.484</td>
<td>1.7</td>
<td>16</td>
<td>4</td>
<td>4</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>$P_4$ Nil</td>
<td>67.1</td>
<td>22.1</td>
<td>10.03</td>
<td>2.91</td>
<td>7.12</td>
<td>84</td>
<td>57.1</td>
<td>1.424</td>
<td>5.7</td>
<td>8</td>
<td>3</td>
<td>3</td>
<td>0.41</td>
</tr>
<tr>
<td>7</td>
<td>$P_1$ Full</td>
<td>64.0</td>
<td>20.8</td>
<td>5.42</td>
<td>2.65</td>
<td>2.76</td>
<td>76</td>
<td>64.0</td>
<td>1.488</td>
<td>35.3</td>
<td>24</td>
<td>4</td>
<td>4</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>$P_2$ L</td>
<td>65.5</td>
<td>21.0</td>
<td>9.78</td>
<td>2.72</td>
<td>7.06</td>
<td>79</td>
<td>58.4</td>
<td>1.429</td>
<td>3.9</td>
<td>12</td>
<td>3</td>
<td>3</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>$P_3$ Q</td>
<td>64.1</td>
<td>20.8</td>
<td>5.59</td>
<td>2.74</td>
<td>2.85</td>
<td>78</td>
<td>61.4</td>
<td>1.461</td>
<td>1.8</td>
<td>16</td>
<td>4</td>
<td>4</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>$P_4$ Nil</td>
<td>65.7</td>
<td>21.1</td>
<td>11.95</td>
<td>3.00</td>
<td>8.95</td>
<td>84</td>
<td>53.8</td>
<td>1.382</td>
<td>7.1</td>
<td>9</td>
<td>4</td>
<td>4</td>
<td>0.38</td>
</tr>
<tr>
<td>8</td>
<td>$P_1$ Full</td>
<td>63.2</td>
<td>20.4</td>
<td>5.36</td>
<td>2.64</td>
<td>2.72</td>
<td>76</td>
<td>63.2</td>
<td>1.466</td>
<td>34.7</td>
<td>31</td>
<td>4</td>
<td>4</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>$P_2$ L</td>
<td>64.0</td>
<td>20.1</td>
<td>12.08</td>
<td>2.73</td>
<td>9.35</td>
<td>77</td>
<td>54.7</td>
<td>1.385</td>
<td>5.5</td>
<td>13</td>
<td>3</td>
<td>4</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>$P_3$ Q</td>
<td>63.3</td>
<td>20.4</td>
<td>5.52</td>
<td>2.72</td>
<td>2.79</td>
<td>77</td>
<td>60.6</td>
<td>1.440</td>
<td>1.8</td>
<td>18</td>
<td>4</td>
<td>4</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>$P_4$ Nil</td>
<td>63.9</td>
<td>19.9</td>
<td>17.57</td>
<td>3.07</td>
<td>14.50</td>
<td>83</td>
<td>46.3</td>
<td>1.305</td>
<td>11.0</td>
<td>11</td>
<td>4</td>
<td>6</td>
<td>0.35</td>
</tr>
</tbody>
</table>
Table 3.5 Smaller-NAV portfolio revisions with maximum weight constraints

Revision with Full ($P_1$), (L)inear part ($P_2$), (Q)uadratic part ($P_3$) and Nil ($P_4$) regard to transaction costs for negative exponential investors with different constant absolute risk aversion coefficients ($ARA$). Net asset value of unrevised portfolio is 568 MSEK and its turnover rate 0.75. The mean and standard deviation of the unrevised portfolio’s return are 41.5%, and 24.4%, respectively. $R_p$ is portfolio mean annual return before any disregarded transaction costs are deducted, $S_p$ is portfolio standard deviation, $TC$ is total expected amortized transaction costs, $TC_L$ is the linear part of $TC$, $TC_Q$ is the quadratic part of $TC$, $TO$ is turnover, $R_p^*$ is portfolio return net of all (regarded and disregarded) transaction costs, $CE$ is the certainty equivalent gross return for negative exponential preferences and normal returns based on $R_p^*$, $CED$ is the certainty equivalent distance of a given portfolio based on $R_p^*$ relative to the initial portfolio, and $CEL$ is the certainty equivalent loss relative to the portfolio obtained under full consideration of transaction costs. $N_{c>0}$ is the number of stocks (with weights $>1e{-9}$). $N_{x=x_{max}}$ is the number of stocks whose weights exceed 0.1. $Maxx$ is the weight of the largest holding. $CEwL$ is percentage certainty equivalent loss due to weight constraints.

<table>
<thead>
<tr>
<th>$ARA$</th>
<th>$P_1$ Full</th>
<th>$P_2$ L</th>
<th>$P_3$ Q</th>
<th>$P_4$ Nil</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>61.4 23.7</td>
<td>65.8 27.1</td>
<td>61.4 23.8</td>
<td>65.9 27.1</td>
</tr>
<tr>
<td></td>
<td>4.38 21.5</td>
<td>12.98 22.8</td>
<td>4.35 22.0</td>
<td>15.87 26.3</td>
</tr>
<tr>
<td></td>
<td>63.1 22.3</td>
<td>10.70 21.4</td>
<td>64. 21.4</td>
<td>78. 13.24</td>
</tr>
<tr>
<td></td>
<td>61.4 1.557</td>
<td>55.1 1.477</td>
<td>59.2 1.535</td>
<td>50.1 1.427</td>
</tr>
<tr>
<td></td>
<td>34.8 26.0</td>
<td>5.2 17.0</td>
<td>1.4 16.0</td>
<td>8.4 11.0</td>
</tr>
<tr>
<td></td>
<td>26.8 9.0</td>
<td>3.9</td>
<td>11.0</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>60.9 23.3</td>
<td>65.0 25.9</td>
<td>61.1 23.5</td>
<td>64.8 25.6</td>
</tr>
<tr>
<td></td>
<td>4.19 2.11</td>
<td>10.63 2.32</td>
<td>4.24 2.20</td>
<td>12.62 2.63</td>
</tr>
<tr>
<td></td>
<td>60.9 1.528</td>
<td>56.7 1.466</td>
<td>58.9 1.506</td>
<td>50.1 1.427</td>
</tr>
<tr>
<td></td>
<td>33.5 17.0</td>
<td>4.0 17.0</td>
<td>2.0 17.0</td>
<td>8.4 14.0</td>
</tr>
<tr>
<td></td>
<td>32.8 13.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>60.0 22.5</td>
<td>65.2 25.4</td>
<td>60.4 22.8</td>
<td>64.8 25.2</td>
</tr>
<tr>
<td></td>
<td>3.86 2.01</td>
<td>10.63 2.32</td>
<td>4.24 2.20</td>
<td>12.62 2.63</td>
</tr>
<tr>
<td></td>
<td>60.0 1.498</td>
<td>53.1 1.420</td>
<td>58.2 1.478</td>
<td>52.2 1.424</td>
</tr>
<tr>
<td></td>
<td>32.2 14.0</td>
<td>4.0 16.0</td>
<td>2.0 14.0</td>
<td>8.4 13.0</td>
</tr>
<tr>
<td></td>
<td>38.0 13.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>59.3 22.0</td>
<td>61.8 22.3</td>
<td>59.0 21.6</td>
<td>62.1 22.4</td>
</tr>
<tr>
<td></td>
<td>3.70 2.00</td>
<td>11.68 2.44</td>
<td>4.01 2.20</td>
<td>13.48 2.69</td>
</tr>
<tr>
<td></td>
<td>59.3 1.472</td>
<td>52.6 1.402</td>
<td>56.8 1.452</td>
<td>49.8 1.387</td>
</tr>
<tr>
<td></td>
<td>31.2 15.0</td>
<td>4.8 19.0</td>
<td>2.0 15.0</td>
<td>7.4 14.0</td>
</tr>
<tr>
<td></td>
<td>42.7 10.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>58.2 21.0</td>
<td>60.7 21.4</td>
<td>57.8 20.6</td>
<td>60.4 21.1</td>
</tr>
<tr>
<td></td>
<td>3.74 2.08</td>
<td>11.22 2.44</td>
<td>4.06 2.30</td>
<td>13.62 2.72</td>
</tr>
<tr>
<td></td>
<td>58.2 1.449</td>
<td>51.8 1.381</td>
<td>56.8 1.452</td>
<td>48.6 1.360</td>
</tr>
<tr>
<td></td>
<td>30.5 14.0</td>
<td>4.7 30.0</td>
<td>2.0 14.0</td>
<td>7.6 15.0</td>
</tr>
<tr>
<td></td>
<td>48.7 10.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>56.8 20.0</td>
<td>59.3 20.4</td>
<td>56.9 19.9</td>
<td>59.1 20.1</td>
</tr>
<tr>
<td></td>
<td>3.69 2.10</td>
<td>10.88 2.41</td>
<td>4.10 2.35</td>
<td>12.53 2.74</td>
</tr>
<tr>
<td></td>
<td>56.8 1.428</td>
<td>50.8 1.363</td>
<td>54.5 1.407</td>
<td>46.6 1.324</td>
</tr>
<tr>
<td></td>
<td>29.8 9.0</td>
<td>4.5 32.0</td>
<td>1.4 41.0</td>
<td>7.2 17.0</td>
</tr>
<tr>
<td></td>
<td>53.7 14.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>55.9 19.3</td>
<td>57.8 19.3</td>
<td>56.0 19.3</td>
<td>57.9 19.3</td>
</tr>
<tr>
<td></td>
<td>3.76 2.15</td>
<td>11.06 2.44</td>
<td>4.10 2.35</td>
<td>11.93 2.75</td>
</tr>
<tr>
<td></td>
<td>55.9 1.409</td>
<td>49.2 1.342</td>
<td>53.7 1.388</td>
<td>46.0 1.311</td>
</tr>
<tr>
<td></td>
<td>29.5 7.0</td>
<td>4.7 35.0</td>
<td>1.5 46.0</td>
<td>7.0 21.0</td>
</tr>
<tr>
<td></td>
<td>58.7 10.0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$RP$ is portfolio mean annual return before any disregarded transaction costs are deducted, $SP$ is portfolio standard deviation, $TC$ is total expected amortized transaction costs, $TCL$ is the linear part of $TC$, $TCQ$ is the quadratic part of $TC$, $TO$ is turnover, $RP^*$ is portfolio return net of all (regarded and disregarded) transaction costs, $CE$ is the certainty equivalent gross return for negative exponential preferences and normal returns based on $RP^*$, $CED$ is the certainty equivalent distance of a given portfolio based on $RP^*$ relative to the initial portfolio, and $CEL$ is the certainty equivalent loss relative to the portfolio obtained under full consideration of transaction costs. $N_{c>0}$ is the number of stocks (with weights $>1e{-9}$), $N_{x=x_{max}}$ is the number of stocks whose weights exceed 0.1. $Maxx$ is the weight of the largest holding. $CEwL$ is percentage certainty equivalent loss due to weight constraints.
3.5.2 Portfolio turnover rates
The turnover rates (TO), computed as the sum of weight increases (or decreases), experienced with full consideration of transaction costs appear quite high on an absolute basis ranging from 74% to 84% across Table 3.2 and Table 3.4. They are also smaller than the turnover rates obtained with nil regard to transaction costs; the ratios of the former to the latter turnover rates vary from 86% to 95%. Thus, the importance of explicitly restricting turnover as suggested by Schreiner (1980) appears somewhat reduced when transaction costs are given consideration.

3.5.3 Total neglect of transaction costs
With transaction costs neglected in the portfolio revisions, the total (regarded plus disregarded) transaction costs TC suffered are quite large; for larger-NAV revisions (Table 3.2) between 16.13% and 31.40%, for smaller-NAV (Table 3.4) between 10.03% and 17.57%. Thus, without any (explicit) consideration of transaction costs in the portfolio revision program at all, there seems to be a strong need for an alternative means to limit transaction costs, and particularly so for the larger-NAV portfolio, for which TC in several cases exceed 15%. Restricting turnover probably helps serving that purpose.

3.5.4 Distribution of portfolio weights
Four measures of portfolio concentration or allocative extremeness are used: the number of stocks (with positive weights) in a given revised portfolio (N_{x>0}); the weight of the largest holding in any revised portfolio (Maxx); the number of individual stocks that have weights greater than 0.10 (N_{x>0.1}); and the number of individual stocks whose weights exceed the constraint xmaxi imposed by Swedish law (N_{x>Law}). Note that these measures are not independent.

With full regard to transaction costs, the revised smaller and larger-NAV portfolios contain 9 to 31 stocks and 11 to 37 stocks, respectively. With respect to the maximum weight of a single holding in any revised portfolio obtained with full regard of trading costs, the respective maximum weight ranges are for smaller-NAV 0.25-0.31 and for larger-NAV portfolios 0.26-0.29.

With transaction costs neglected, the portfolio allocations are extremer and concentrated to in some cases as few as five holdings. Moreover, the maximum
weights $\maxx$ are likewise substantially larger ranging from 0.35 to 0.43. The number of weights that exceed 0.10 and $\maxx$, respectively, in individual portfolios is virtually identical to when transaction costs are accounted for. Portfolios revised under transaction costs thus demonstrated themselves to be less extreme in terms of concentration than portfolios revised without regard to transaction costs, and therefore probably more appealing to investor intuition.

The results presented this far indicate that the magnitude of the transaction costs incurred in the portfolio revisions is non-trivial when compared to expected portfolio returns, and that neglect of transaction costs lead to high portfolio turnover, extremer allocations, and large transaction costs.

### 3.5.5 Certainty equivalent losses

Although transaction costs appear to have major nominal consequences, it remains to be documented if less than full consideration of those costs in portfolio revisions lead to significant negative effects on investor welfare. Certainty equivalent losses (CELs) for neglect of such costs relative to full consideration of them are therefore computed. Total neglect of transaction costs results in substantial CELs ranging from 9.0% to 19.8% for the larger-NAV portfolio revisions, and from 5.7% to 11.0% for the smaller-NAV portfolio revisions.

### 3.5.6 Linear versus quadratic part

With full consideration of transaction costs, $TC_Q$, the quadratic part of the transaction costs function, contributed more to total transaction costs $TC$ than did $TC_L$, the linear part. This holds true for all the levels of risk aversion considered. For the smaller-NAV portfolio revisions, the ratio $TC_Q/TC_L$ ranges from 1.0 to 1.2, and for the larger-NAV revisions, from 3.0 to 3.4.

To shed light on whether it is worthwhile to use a transaction cost specification that allows for price impact costs to be explicitly modeled as functions of order size, revision are performed with alternating pieces of the transaction cost function excluded. As noted, with just the linear part included, percentage transaction costs in

---

48 The test was also conducted for a 50 MSEK and a 10 billion SEK portfolio. Total neglect of transaction costs in revisions of the 50 MSEK (10 GSEK) portfolio rendered non-trivial (huge) CELs; neglect of the linear (quadratic) transaction cost part was more serious than neglect of the quadratic (linear) part.
individual stocks are effectively constant, and no explicit consideration is given to price impact costs. Revisions performed with the linear part of the transaction cost function excluded exhibit smaller turnover rates ($TO$) than the revisions performed with the quadratic part excluded. Moreover, neglect of the quadratic part results in more extreme portfolio allocations than neglect of the linear part. Furthermore, in terms of utility losses, it seems that it is less dangerous to disregard the linear part of the transaction cost specification than it is to disregard the quadratic part. Across the smaller and larger-NAV portfolio revisions, neglect of the linear part results in only moderate $CEL$s, ranging from 0.8% to 1.7%. Disregarding the quadratic part, however, leads to considerably larger welfare losses; for the smaller-NAV portfolio revisions, the $CEL$ varies from 3.9% to 5.5%, and for the larger-NAV revisions from 8.5% to 14.2%.

Taken together, the results suggest that it is more serious to neglect the quadratic part, which enables price impact costs to be considered explicitly, than the linear part, which does not. In addition, revisions performed with either the quadratic or the linear term of the transaction cost specification excluded, exhibit marked differences with respect to portfolio allocations, turnover rates, and total (regarded and disregarded) transaction costs incurred. In general, the specification with just the quadratic cost term generates results similar to, albeit less good than, the results obtained with the full specification, while the results for the linear specification are more similar to, albeit less bad than, the results obtained with transaction costs neglected.

Note that $c^*_i$, the coefficient of the linear part of the price impact cost function, is not based on an order size of typical volume but was estimated jointly with the coefficient $d^*_i$ of the quadratic part. To put the comparison on an (more) equal footing one should ideally use, for the coefficient of the linear part, the average percentage cost of a transaction of typical size. As mentioned, this quantity is, however, difficult to estimate.

### 3.5.7 Effects of the maximum weight constraint as imposed by law

Table 3.3 and Table 3.5 outline the results for the revision of a larger and a smaller portfolio in terms of NAV when the maximum weight constraint $x_{max_i}$ in the form of...
upper bounds on individual holdings are imposed; the non-negativity weight constraint is still in place.

The observation that $TC$ increase in $CED$ carries over to Table 3.3 and Table 3.5. A comparison between the larger NAV-portfolios revised with the complete model of transaction costs in Table 3.3 and Table 3.2, holding the risk attitude constant, reveals that the portfolios revised under a maximum weight constraint exhibit (i) more diversification in terms of $N_{x>0}$ (2-3 times), (ii) lower $TC$ and $TO$, (iii) lower expected return and standard deviation, and (iv) lower utility ($CE$). The results for smaller-NAV portfolios in Table 3.5 and Table 3.4 relate to each other in virtually the same manner as the results for the larger NAV-portfolios.

The higher diversification, in terms of $N_{x>0}$, induced by the maximum weight constraint, comes at the cost of utility. The quantity $CEwL$ expresses, for a given $ARA$ and under full consideration of transaction costs, the percentage certainty equivalent loss for a weight constrained portfolio relative to an unconstrained one.\footnote{$CEwL = (CE\text{unconstrained} - CE\text{constrained}) / CE\text{unconstrained}$} The rightmost columns of Table 3.3 and Table 3.5 show that these constraints incur, with the full transaction cost model, cash equivalent losses of around 4\% for the smaller-NAV portfolio revisions, and somewhat less for the larger-NAV portfolio revisions.

If a portfolio is revised under less than full control of transaction costs, there might be a chance that a maximum weight constraint helps to bring down turnover and thereby transaction costs to a point where a utility gain could be realized vis-à-vis the unconstrained case. This occurs for three of the larger NAV-portfolio revisions, for $ARA$ equal to 7 and 8 under total neglect of transaction costs, and for $ARA=8$ under neglect of quadratic transaction costs. It also occurs for the smaller NAV-portfolio revision for $ARA=8$ under complete neglect of transaction costs. This again demonstrates that restrictions on turnover, weights, or weight changes could be productive, in case transaction costs are ignored.

3.5.8 Extensions and limitations

True transaction costs were assumed to be a quadratic function of weight change, with coefficients given by the models estimated in Chapter 2 and with commission rates
taken from Dahlquist et al. (2000), and amortized as suggested by Rudd and Clasing (1982) and Grinold and Kahn (1995). The findings with regard to the magnitude of transaction costs apply of course only directly to the portfolios revised, and - to the extent that the revised portfolios are representative in terms of size, turnover, and composition - of other portfolios. The price impact costs were based on uninformed immediate execution against displayed limit orders in the order book. Therefore, it is possible that the estimated transaction costs are high if compared to those that a less aggressive trading strategy would have incurred. However, Griffiths et al. (2000) show that implementation shortfall reduce, and sometimes even eliminate, the cost differences of patient and aggressive trading.

For convenience, the tests assumed an investor who has negative exponential preferences and who believes that returns are multivariate normal. Hence, results and conclusions are valid only under these conditions, but previous research, by, e.g., Kallberg and Ziemba (1983), shows that under normal returns, different utility functions result in virtually identical optimal portfolios being selected for similar levels of absolute risk aversion. Similar levels of transaction costs and turnover rates would thus be experienced regardless of the form of the utility function. In addition, it should be noted that the magnitudes of the mean returns in the sample were quite large, and that the goodness of the MV approximation is expected to deteriorate as the level of return increase.

Furthermore, input parameters, that are mean returns, variances, covariances, and transaction costs, were treated as free of measurement error, i.e., no consideration was given to estimation risk. A possible extension of this study would be to incorporate estimation risk and other utility functions. Finally, the empirical exercise was quasi-experimental. It considered, for instance, portfolio revisions at a single point in time. A natural extension would be to investigate the performance over time for portfolios revised with varying levels of transaction cost control.

---

50 I have implemented the transaction costs specification used here for, e.g., logarithmitic utility and so-called empirical probability assessment approach (see Grauer and Hakansson 1993).
3.6 Summary

I extended the standard mean-variance portfolio selection model by formulating a quadratic program for mean-variance portfolio revisions with transaction costs, including price impact costs. The true and full model of transaction costs were assumed to be a quadratic function (without a constant term) of weight change, with coefficients given by the empirical price impact cost models of Chapter 2 and with commission rates taken from Dahlquist et al. (2000). Empirical tests were performed to analyze how different degrees of consideration of transaction costs in portfolio revisions are related to investor welfare, portfolio allocations, turnover rates, and incurred levels of transaction costs. The initial portfolio was capitalization-weighted and included all listed Swedish stocks with sufficient data. To provide evidence relevant to a typically smaller and a typically larger Swedish mutual fund in terms of net asset value (NAV), the initial portfolio was revised using turnover rates, net asset values, and commissions typical for each fund-size.

With full consideration of transaction costs, that is, for revisions performed with the full transaction cost specification in place, for a wide range of attitudes towards risk, and for smaller and larger portfolios in terms of NAV, the portfolio revisions generated total expected amortized transaction costs of 5.3% to 5.9%. Transaction costs of these magnitudes are non-trivial compared to typical levels of expected portfolio returns. The certainty equivalent distance between a given initial portfolio and a given revised optimal portfolio was observed to have explanatory power with regard to the magnitude of transaction costs incurred in the transition. With complete neglect of transaction costs, the total transaction costs incurred were greater, ranging from 10% to 31%.

With full consideration of transaction costs, the revised portfolios exhibited allocations that were less extreme and probably more intuitively appealing to investors than the allocations obtained with transaction costs disregarded. The turnover rates experienced for the portfolio revisions performed with full consideration of transaction cost were smaller than the ones observed with transaction costs neglected. Transaction costs thus demonstrated themselves to have significant consequences in nominal terms for the revised portfolios.
The effects on investor welfare from varying degrees of neglect of transaction costs relative to full consideration of such costs in mean-variance portfolio revisions were measured in terms of certainty equivalent losses. The portfolios revised without control of transaction costs incurred substantial certainty equivalent losses in the range 6-20% relative to the portfolios revised under transaction cost control. Given the net asset values of the portfolios analyzed, these certainty equivalent losses correspond to cash losses in the range 34-370 MSEK. In general, the effects of neglecting transaction costs were more serious for the larger portfolios NAV-wise than for the smaller ones.

A maximum weight constraint on individual holdings, as imposed by Swedish law, led to higher diversification in terms of the number of (non-zero) portfolio holdings, but it induced cash equivalent losses of around 4%, compared to when no such constraint was in place. In case transaction costs are ignored, then are restrictions on turnover, weights, or weight changes probably useful.

Overall, the evidence suggests that transaction costs are important to take account of in mean-variance portfolio revisions, and that it is worthwhile to use a transaction cost specification that includes price impact costs.

3.7 Appendix

Appendix A

Called for by the empirical analysis, a discussion of portfolio certainty equivalents in the special case of negative exponential utility and joint normal returns is undertaken.

In the empirical tests, I want to measure the proximity of different portfolios. Comparing the expected utility scores of two different portfolios by calculating the difference or ratio of the scores does, however, not produce an absolute standard of comparison. Expected utility functions (of the von Neumann-Morgenstern type) are invariant to linear transformations, i.e., give identical rankings of portfolios under linear transformations. Thus, by suitable transformation of the utility function, one can make a difference of scores arbitrarily close to zero, or a ratio arbitrarily close to one, without affecting the ranking of the portfolios.

Meaningful comparisons of the expected utility scores of different portfolios, may, however, be done in terms of the portfolios’ cash or certainty equivalents, as suggested
by Dexter et al. (1980). The certainty equivalent (CE) of a risky portfolio is the certain, or risk-free, gross return (which is equal to one plus the rate of return) that yields the same utility as the risky portfolio does. The CE is an appropriate measure to base comparisons on, since it takes into account the investor’s risk aversion and the inherent uncertainty in returns, and it is independent of utility units (Chopra and Ziemba 1993). Below the calculation of the certainty equivalent for a given portfolio is shown; in the empirical analysis, such certainty equivalents are used in two different measures of relative portfolio proximity. Assume that returns have a joint normal distribution, and that the investor has a negative exponential utility function 

\[ U(W) = -\exp(-aW), \]

where \( W \) is wealth or portfolio value at the end of the single period, and \( a \) the investor’s constant absolute risk aversion parameter expressed in units of terminal wealth (i.e., money). Under these assumptions, Freund (1956) shows that expected utility \( EU(W) \) is a function of expected value, \( E(W) \), and variance, \( V(W) \), of terminal portfolio value:

\[
EU(W) = E(-\exp(-aW)) = -\exp[-a E(W) + a^2/2 V(W)].
\]

Often, it is preferable to work with returns instead of terminal wealth. Define \( ARA \), the investor’s absolute risk aversion parameter expressed in gross returns, as \( ARA = a W_0 \).

Then, since \( W = W_0(1+r) \), where \( W_0 \) is the initial value of the portfolio and \( r \) the portfolio rate of return, rewrite \( EU(W) \) in terms of returns and \( ARA \):

\[
EU(W) = -\exp[-a E(W_0(1+r)) - a^2/2 V(W_0(1+r))] = -\exp[-ARA ( E(1+r) - ARA/2 V(1+r) )]
\]

Obviously, maximizing the so-called mean-variance objectives \( E(1+r) - ARA/2 V(1+r) \) and \( E(r) - ARA/2 V(r) \)\(^5\) is equivalent to maximizing \( EU(W) = E(-\exp(-ARA(1+r))) \); hence, fulfilling either objective results in the same optimal portfolio being selected.

Note that in terms of the vector of portfolio weights, \( x \), the vector of mean returns, \( r \), and the covariance matrix \( V \), the expected portfolio gross return and variance equal

\[ E(1+r) = (1+r)^T x \text{ and } V(1+r) = x^T V x, \]

respectively.

Recall that the certainty equivalent of a risky portfolio is the risk-free return that yields the same utility as the risky portfolio. This is the gross return \( CE \) that solves

\(^5\) \( E(r) - ARA/2 V(r) \) is equivalent as an objective, because the unities of the gross returns are constants with regard to the maximization.
\[ U(W_0 CE) = EU(W_0 (1+r)), \]

which can be expanded to

\[ -\exp(-ARA CE) = -\exp[-ARA (E(1+r) + ARA/2 V(1+r))]. \]

Solving for the certainty equivalent return, yields:

\[ CE = E(1+r) - ARA/2 V(1+r). \]

Appendix B

The equivalence of 1.20 ORDERSIZE and \(d^+_i x^+_i\), where \(d^+_i = 1.2 \times 1000 \times Pval / Mcapi\):

\[
1.20 \text{ORDERSIZE} = \\
1.2 \times \text{No. of shares in trade} / (\text{Tot no. of shares outst.}/1000) = \\
1.2 \times 1000 \times \text{No. of shares in trade} \times \text{Price} / (\text{Tot no. of shares outst.} \times \text{Price}) = \\
1.2 \times 1000 \times x^+_i \times Pval / Mcapi = \\
1.2 \times 1000 \times Pval / Mcapi \times x^+_i = \\
d^+_i x^+_i.
\]
Chapter 4

Index fund management under transaction costs

The index fund problem has been stated as how to achieve a desired tracking performance while controlling transaction costs (see, e.g., Rudd 1980, Frino and Gallagher 2002, and Beasley et al. 2003). The trading of index funds is per definition uninformed and index funds have been observed as frequent demanders of immediacy. The index fund problem thus seems to be a rather ideal context in which to address the research questions of this thesis. I develop two index fund revision models under transaction costs including price impact costs. The two models are integrated with the empirical models of price impact costs of Chapter 2. The importance of transaction costs and transaction cost control in index fund management is examined by a number of large-scale empirical tests, where the integrated models and some alternative approaches, including full replication, are applied to track a Swedish capitalization-weighted stock index. Other issues in index fund management that are examined are the efficiency of pre-defined subsets as means to control transaction costs and how different tracking error measures affect tracking performance.

4.1 Introduction

“The method of selecting and revising passive portfolios [index funds] has received virtually no attention in the literature,” Rudd (1980, p. 57)

“Index funds have grown significantly over the past decade, however empirical research concerning these passive investment offerings is surprisingly scarce in the literature.” Frino and Gallagher (2001, p. 44)

“Finally we would comment that index tracking is an important problem that, in our view, has received insufficient attention in the literature.” Beasley et al. (2003, p. 641)

Most funds are actively managed and aim at beating a benchmark. In contrast, the aim of an index fund, is to replicate, or track, the performance of a pre-specified benchmark, most commonly a capitalization-weighted stock market index. The first equity index fund was launched on the U.S. market in 1971 by Wells Fargo Bank (Bernstein 1992). The index fund concept was a rather natural spin-off from the, at that time, emerging modern portfolio theory, with its fundamental ideas of market
efficiency and market portfolio optimality, as implied by the Capital Asset Pricing Model (CAPM) (see, e.g., Black and Scholes 1974). A common argument for indexing and against active management goes as follows (see, e.g., Rosenberg 1981 and Sharpe 1991; Etzioni 1992 discusses the validity of the argument). The average value-weighted return, before costs, for all portfolios invested in a given market must equal the return of the market (index), as these portfolios are the market. Before costs, adequately constructed index funds will deliver the same average return as the market. After costs, the average index fund will underperform the market index, but to a lesser degree than the average actively managed portfolio, because of lower costs.

Moreover, compared to investments in closed-end funds and actively managed funds, index fund investments probably incur lower agency costs in that the performance relative to an index is easily monitored (Edelen 1999). The eventual success of indexing is perhaps best reflected by the large amounts of money invested in index funds; recently, it was estimated that of global assets under management around 1/3 - the equivalent of USD 10 trillion or SEK 75 trillion - was indexed (Cerulli Associates 2001, 2003 cited Asset growth grinds to a halt 2002 and Global asset management industry making a comeback 2003, respectively).

Since the seminal work of Jensen (1968), mutual fund performance evaluations continue to indicate that most actively managed funds perform worse than their benchmark indexes (see, e.g., Elton et al. 1993; Malkiel 1995; Carhart 1997; Chalmers et al. 2001; Pinnuck 2003). In addition, Gruber (1996) and Frino and Gallagher (2001), among others, find that actively managed funds are outperformed by index funds on a risk-adjusted basis. Also, indexing has had profound effects on fund management thinking and practices. Concepts such as benchmark portfolios, active and passive strategies, and tracking error, all relate to indexing in some way.

On the surface, it may appear trivial for a fund to replicate the performance of a capitalization-weighted stock market index. However, performance evaluations of index funds find that funds tracking the same index exhibit varying tracking performance (Sinquefield 1993; Gruber 1996; Frino and Gallagher 2001, 2002, 2004; Blume and Edelen 2004; Elton et al. 2004).
Work by, e.g., Black and Scholes (1974), Rudd (1980, 1986), Sinquefield (1991), Gruber (1996), Keim (1999), Frino and Gallagher (2001, 2002, 2004), Blume and Edelen (2004), and a priori theorizing suggest that the tracking performance of an equity index fund is related to several factors including the following ones: (i) the approach (or method) including possible criteria - such as the desired tracking performance, usually expressed in the form of a tracking error measure, which is a function that measures how closely the fund tracks the index - according to which the fund is selected and revised; (ii) the index methodology, i.e., the rules according to which the index is calculated; (iii) the transaction costs, the dividend behavior, the corporate action activity, and other idiosyncratic features, such as the risk-return characteristics, of the index constituents and other securities in the investment universe; (vi) the cash flows to and from the fund; (v) the size of the fund; and (vi) the quality of the data used in implementing the fund.52

Several studies including Rudd (1980), Meade and Salkin (1989, 1990), Adcock and Meade (1994), Larsen and Resnick (1998), Walsh et al. (1998), Bamberg and Wagner (2000), Rey and Seiler (2001), Konno and Wijayanayake (2001), and Beasley et al. (2003) implement and test empirically the tracking performance of different approaches to equity indexing. Corporate actions, dividends, and other cash flows are, however, not considered in any of those studies. Rudd (1980), Adcock and Meade (1994), and Beasley et al. (2003) consider explicitly transaction costs when the index fund is selected and revised, but use transaction cost functions that are exclusive of price impact costs and invariant across stocks and time. In a rarely referenced article, Konno and Wijayanayake (2001) develop and test an index fund selection model based on the mean-absolute deviation framework of Konno and Yamazaki (1991). The model allows for non-convex transaction costs including price impact, but their empirical test does not consider price impact costs and the transaction costs are invariant across stocks.

52 Expenses, including advisory, administrative, distribution and other fees, are of course also important, but are not given explicit consideration here, as the focus is on the costs related directly to a fund’s portfolio decisions and resulting trading.
The index to track is a central element in a study of alternative indexing approaches. The obvious choice would be to use an extant index like the Swedish OMX(S30) index, which is the most well-known and used index in Sweden as well as in the Nordic region. In essence, the OMX index is capitalization-weighted and comprises the 30 largest stocks listed on the Stockholm Stock Exchange in terms of monetary trading volume. The trading in standardized derivatives based on the OMX amounts to SEK 800 billion annually. In addition, there are warrants, equity index-linked bonds, exchange traded funds (ETFs), index funds, and a variety of non-standardized derivatives based on the OMX (Stockholm Stock Exchange 2003).

Over time, the OMX index methodology has, however, been changed several times. Therefore, because replication performance is expected to be related to index methodology, I decided to independently calculate an index, using a consistent, fully computerized rule-based methodology, mimicking that according to which OMX is constructed and maintained; the methodology is quite generic to those used for most of the indexes of Standard & Poors, Dow Jones, FTSE, MSCI, and other index providers. This approach also solves several issues related to data availability and quality. For instance, complete daily data on the exact composition of the index is generated, whereas such data have not been readily available on the OMX. Moreover, several studies, e.g., Walsh et al. (1998), Rudolf et al. (1999), and Bamberg and Wagner (2001) implement index tracking using index return series data that are based on the historical, and possibly outdated, index composition. I examine whether index funds constructed using such data perform differently than index funds constructed using complete and current index composition data.

OMX is a so-called price index, that is, it is not calculated with dividends reinvested. Index funds do, however, receive dividends. As dividends increase fund returns, it seems unfair to use a price index as benchmark. Because of this, the different indexing methods will be evaluated against a so-called total return index, where dividends are reinvested, rather than against a price index. The detailed index composition data allow precise calculation of the total return index.

---

53 A cursory review shows that many Swedish index funds have chosen price indexes as benchmarks.
In the literature, several approaches to equity indexing may be distinguished including full replication, optimization, regression, and stratified sampling. It has been suggested, based on rather tentative empirical evidence, however, that some approaches are better suited than others for tracking particular kinds of indexes; full replication, for instance, is argued to be relatively more suitable, the fewer and the more liquid stocks an index contains (Rudd 1980; Andrews et al. 1986).

Given the nature of the index to be tracked and other considerations, I decided to implement and test more thoroughly three approaches, full replication and two optimization-based ones, each with quite a few variations. Full replication in its basic form involves holding, at all times, all index constituents in the same proportions as in the index, and without transaction cost control. Optimization, essentially, involves selecting a portfolio so that some criterion of goodness, usually based on a measure of tracking error, is optimized or satisfied. Optimization-based approaches also permit explicit transaction cost control.

The motive for involving full replication is that the approach is expected to produce good tracking performance relative to the index that will be used here, as this index contains relatively few and liquid stocks. The performance of full replication would thereby serve as a qualified reference point for the competing optimization-based approaches. Moreover, no study has actually implemented and tested the performance of full replication, although it is reported to be the approach most commonly used in practice (see, e.g., Frino and Gallagher 2001; Blume and Edelen 2004). Probably, it is not expressed explicitly, the performance of the index is taken as a surrogate for the performance of full replication. This is correct only if the tracking errors arising from transaction costs and delayed dividend payments are minimal. Whether this is the case or not is analyzed empirically here.

To be able to address the question about the importance of transaction cost control in index fund management more directly, I formulate two optimization-based approaches for index fund revision under transaction costs including price impact. The first, Opt1, is a mathematical program that features a goal function where transaction costs are balanced against a tracking error measure utilizing a trade-off parameter. The
second, \( \text{Opt2} \), is a program where the objective is to minimize transaction costs while keeping a tracking error measure less than or equal to a pre-specified limit.

A technique to control transaction costs is to restrict the number of stocks in the index fund relative to the number of stocks in the index (see, e.g., Rudd 1980; Connor and Leland 1995; Larsen and Resnick 1998; Bamberg and Wagner 2000; Rey and Seiler 2001). One way of implementing the technique is to construct the index fund from a pre-defined subset of the stocks in the index, where the subset contains the stocks that are expected to have the lowest transaction costs. There is, however, no empirical evidence on whether this use of subset constraints fulfills its purpose.

The question of what tracking error measure to use is stressed in the literature (see, e.g., Rudolf et al. 1999; Bamberg and Wagner 2000; Rey and Seiler 2001; Satchell and Hwang 2001; Beasley et al. 2003). Empirical evidence allowing direct comparison of the performances of different measures in a relevant test setting is typically not provided.

The main purpose of this chapter, then, is to investigate, with regard to a capitalization-weighted Swedish stock index with relatively few and liquid stocks, the tracking performance of three alternative equity indexing approaches, which, among other things, feature different types and degrees of transaction cost control. Additional purposes of the chapter are to examine the efficiency of pre-defined subsets as means to control transaction costs and how different tracking error measures affect tracking performance.

Previous research is extended by the development of a research design that incorporates transaction costs that are inclusive of price impact costs and variable across stocks and time; cash flows in the form of dividends; and corporate actions, including mergers and acquisitions, new issues, and other capitalization changes. Though important for retail index funds, investor cash flows are not considered here, because such data could not be accessed. The research design required the (largely manual) collection and structuring of a comprehensive database comprising more than 10 years of daily data.

The three approaches are tested using a number of variations including (where applicable): with and without transaction costs, with conditional and unconditional
Chapter 4  Index fund management under transaction costs

updating; with different tracking error measures, with different levels of desired/accepted tracking error, with and without tracking error measure autocorrelation adjustment, with pre-defined subsets of different sizes, and with dividends included/excluded. Some issues are best investigated isolated from transaction costs and dividends, including the usefulness of linear tracking error measures in the form of a mean absolute tracking error portfolio selection program, and a heuristic implementation based on the observations made by Satchell and Hwang (2001).

I believe that advances in knowledge about the empirical performance and, hence, the efficiency of different approaches to equity indexing are important, considering, for one thing, the enormous amounts of money invested in index funds.

Important pieces of the research design have been developed in the foregoing chapters. Empirical cross-sectional price impact costs models for Swedish stocks trading in the Stockholm Stock Exchange’s automated limit order book system were developed in Chapter 2. That the trading of index funds is uninformed and often demands immediacy is recognized in the transaction cost models, where costs are measured and expressed so as to reflect the costs of trading that is uninformed and immediate. In Chapter 3, I formulated a mean-variance portfolio revision model with transaction costs inclusive of price impact, to which the two index fund revision models under transaction costs derived here are related. The empirical analysis in Chapter 3 highlighted in a utility-theoretic setting the importance of controlling transaction costs inclusive of price impact in portfolio revision problems. The empirical performance of the two index fund revision formulations under transaction costs derived in this chapter will offer insights into whether it is worthwhile to control transaction costs when index funds are selected.

On a different note, the approach developed here constitutes a computerized approach to portfolio management. Barely any human involvement would be necessary. The study features a computerized “portfolio research system”. Apart from historical returns, all the input needed is a list on tomorrow’s index composition and rudimentary information about corporate actions and cash flows. Trades would
possibly be generated based on real-time information on the public limit order book and according to the indexing method currently preferred.

The chapter proceeds as follows. In Section 4.2, operational problems of index fund management are discussed including implications of transaction costs. Based largely on the material covered in the foregoing section, the research design is developed in Section 4.3. This section includes subsections on the choice of what indexing tracking approaches to test, development of the two index fund revision programs under variable transaction costs, choice of tracking error measures, index methodology, calculation rules and their implementation, and data collection and processing. Section 4.4 contains the empirical analysis, and Section 4.5 a concluding summary.

4.2 Operational problems of index fund management

The operation of an index fund aimed at replicating a capitalization-weighted index would be trivial in the absence of transaction costs, cash flows to/from the fund and if the stocks in the index were non-changing and never subjected to corporate actions. Perfect replication would then be achieved by investing the initial contribution in all stocks in the index and in the same proportions as in the index. However, an index is a mathematical abstraction and perfect replication may in practice not be possible, even in the absence of transaction costs.

Dividends, for instance, are assumed reinvested in the index at the close of the day prior to the ex-dividend day, while in the reality dividends are paid out to and received by the fund some weeks later. Similar, although subtler, difficulties arise in connection with rights issues (where the index calculation assumes that the pay in and the issuance of new shares are in effect from the beginning of the ex-rights day, while these events in practice usually occur some weeks later).

Moreover, perfect replication requires immediate fund rebalancing in response to additions and deletions of stocks from the index as well as to corporate actions that alter the capitalizations and weights of the index constituents. Index calculations suppose that all rebalancing transactions occur at prices prevailing at the close of the market and without any transaction costs incurred. In reality, rebalancing incurs transaction costs, which contribute to prevent perfect replication.
With the passage of time, the composition of the fund is thus likely to diverge from that of the index, and this induces tracking error. This suggests that the fund should be rebalanced so as to minimize tracking error. Rebalancing, however, incurs transaction costs, which lower returns and in turn produces tracking error. The benefit of rebalancing in terms of reduced tracking error must therefore be weighed against the detriment in terms of transaction costs. The problem central to the operation of an index fund may thus be stated as how to maintain a desired/accepted level of tracking error (volatility) while minimizing transaction costs.

4.2.1 Implications of transaction costs for tracking error

The transaction cost of an order may be broken up in a direct part including commissions and fees, which are to be paid at once, and an indirect part in the form of the price impact cost, that represents the temporary, adverse price movement induced by the order. Also assume that the price impact is fully temporary, which it should be for uninformed trading according to theory. Suppose that a full replication index fund has decided to revise the holding of stock Y by a proportion X that amounts to Z in monetary units, or U number of shares, valued at the decision-time mid-point stock price P. Two implications of transaction costs for index tracking are then the following ones.

First, regarding the transaction (the buying or selling) of a value of Z monetary units (dollars) of stock Y, one needs to realize that the transacted value net of transaction costs, call it $Z'$, will be lower than the amount $Z$. If one buys for $Z$, price impact will cause the average transaction price to be above the post-transaction price, implying that, $Z'$, the value of the acquired shares taken at the post-transaction price $P$, will be lower than $Z$. In addition, part of $Z$ will have to cover commissions and other direct costs. The number of shares afforded $U'$ will thus be lower than $U$. An order to sell $Z$ monetary units of stock Y, or $U$ number of stocks, incurs direct costs and price impact that cause the average transaction price to be below the decision-time price $P$. Hence, $Z'$, the money received for the $U$ stocks sold, will be lower than $Z$. Thus, for a purchase as well as a sell transaction, there is a loss of value equal to the difference between $Z$ and $Z'$, and this loss will induce tracking error.
Second, tracking error will also arise to the extent that the transaction of Z’ instead of Z does not cause the weight change for stock Y to be exactly X, the desired portfolio weight change of stock Y. Issues related to this second implication are treated analytically elsewhere by the author.

4.3 Research design

4.3.1 Guidelines for index fund management
It is convenient here, and probably close to actual practice, to view the guidelines for the management of an index fund as an exogenously given policy that establishes the index to track; how the tracking performance is to be measured, which usually involves a tracking error (dispersion) measure; and the desired tracking performance. The policy may include the set of investable securities, often referred to as the universe, and various constraints, such as bounds on the portfolio weights of individual holdings or a maximum number of holdings in the fund. It is then for the index fund manager to choose and implement an index tracking approach so as to achieve, in the best manner possible, the policy guidelines. The study in this chapter may thus be viewed as an investigation of how different tracking approaches perform under different sets of policy guidelines.

The research design developed below is largely compatible with this structure. The index to track has been presented. Details about the index methodology are given below in Subsection 4.3.9. On basis of prior research and a priori considerations, several index tracking approaches and measures of tracking error will be devised and applied to track the given index under different desired tracking performances and constraints.

4.3.2 Choice of index tracking approaches to test
In addition to full replication and optimization there are, in broad terms, at least two (types of) approaches to equity indexing.

Stratified sampling means that stocks are selected from a pre-defined subset of the firms in the index so as to achieve the same sector and capitalization weightings as the index. Contrary to full replication, stratified sampling is argued to be relatively less
suitable for tracking indexes with few and liquid stocks (Rudd 1980; Andrews et al. 1986). Stratified sampling will therefore not be considered in the present study.

Regression applied to index tracking, in its basic form, estimates the weights as the regression coefficients in a restricted multiple linear regression of index returns on individual stock returns. For the coefficients to be interpretable as portfolio weights, the restriction that weights sum to unity is imposed.

The approaches, except full replication, are sometimes constrained to select the index fund from a subset of the stocks in the index, and may in such instances be referred to as sampling or approximative replication (see, e.g., Rudd 1980; Andrews et al. 1986; Larsen and Resnick 1998; Wagner 1998). As was discussed previously, allowing only a subset of the index stocks to be considered in the formation of an index fund may be a severe restriction and should therefore be used with care. Moreover, optimization, where the objective is to minimize tracking error volatility, without restrictions on the number of assets, will select the (full replication) index portfolio.

As mentioned, three approaches, each with quite a few variations, will be implemented and tested more thoroughly. The first approach is full replication, which should provide high tracking performance if holding and transaction costs and indivisibility problems are small enough. Also implemented is a variation of full replication in which the index fund is updated conditional on a given forecasted tracking error measure exceeding a pre-defined limit. This is done for several different tracking error measures.

The two optimization-based approaches that incorporate transaction cost control are derived below. Both approaches are implemented and tested using different tracking error measures and different subset constraints, among other things.

Regression indexing

Alderson and Zivney (1989) pioneered the use of OLS to estimate weights for an index tracking portfolio. I do not consider regression-based techniques for index portfolio selection here for reasons including the following ones: they do not allow for adaptive handling of transaction costs, restrictions are not always easily imposed, and it is not straightforward to optimize towards a particular value for tracking error, i.e.,
stop minimizing when a particular value of the goal function is reached. For a clear exposition of equity indexing based on regression analysis, including robust regression techniques, see Bamberg and Wagner (2000).

Other approaches and variations
In addition to the three base approaches and variations thereof, a few other approaches are examined. I implement and test an index portfolio selection program based on a mean absolute tracking error measure, without transaction costs, however. Its performance in the absence of transaction costs is examined and compared to those of formulations involving other tracking error measures. Satchell and Hwang (2001) observe that in tracking error calculations portfolio weights are treated as fixed or non-stochastic ex ante, but ex post calculations use actual weights which usually are variable or stochastic, and this, they show, will on average make ex post tracking error volatility larger than its ex ante counterpart. I implement and test a heuristic index fund selection formulation based on Satchell and Hwang (2001) that attempts to take into account stochasticity in weights.

4.3.3 Index tracking - preliminaries

Preliminaries
Perfect replication is generally not possible. It may not even be desirable as a direct goal, because it is possible that the costs incurred in trying to achieve such an exacting goal actually would deteriorate tracking performance; whether or to which extent this is the case, is however an empirical matter and will as such receive attention in the empirical analysis here. It thus seems potentially beneficial that some kind of operationalization of the acceptable degree of lack of tracking performance is done. In the literature, this is typically done in terms of a measure of the expected dispersion of (the distribution of) the tracking error, which is the return difference between the fund and the index.54 Such measures may thus also be referred to as measures of tracking risk.

---

54 Note that the term tracking error is not consistently used in the literature (or here). For example, tracking error sometimes refer to mistracking in general, to the return difference between the fund and the index, and to (explicitly defined) measures of the dispersion of the return differences.
Common measures of tracking error dispersion in the literature are tracking error standard deviation or variance (e.g., Rudd 1980; Meade and Salkin 1990; Roll 1992; Pope and Yadav 1994; Bamberg and Wagner 2000), root mean squared tracking error (e.g., Rudolf et al. 1999; Rey and Seiler 2001; Beasley et al. 2003), and mean absolute tracking error (e.g., Larsen and Resnick 1998; Rudolf et al. 1999; Frino and Gallagher 2001).

Other measures that have been employed include the median absolute (deviation about the median) tracking error (Bamberg and Wagner 2000) and the standard error of a regression of index fund returns on index returns (Frino and Gallagher 2001, 2002, 2004; Rey and Seiler 2001); Larsen and Resnick (1998) use the coefficient of determination for such a regression as a measure of tracking performance.

Once the policy guidelines have been stated, they need to be operationalized. For an optimization-based approach this amounts to formulating a mathematical program, which involves finding a goal function and a set of constraints consistent with the policy. In the literature, the typical policy has been to minimize a particular tracking error measure, subject to standard portfolio constraints. As an example, if the tracking error measure is standard deviation and minimal tracking error standard deviation is desired, an appropriate goal function in a quadratic program formulation is to minimize the tracking error variance.

Two exceptions from tracking error measure minimization are Jorion (2004), who uses a target value for tracking error variance in the form of an equality constraint, and Meade and Salkin (1990), who, conditional on a forecasted tracking error measure exceeding a limit, update an index fund by minimizing the tracking error measure.

A variation of these themes, and one that will be implemented here, is to use a formulation, where the particular tracking error measure is constrained not to exceed a specified limit. The tracking error measures and corresponding goal functions and constraints that will be examined empirically are presented below. Before that, however, a discussion is undertaken of the sets of securities involved in the formulation of the index fund revision problem.

---

55 Not all studies observe this principle. Larsen and Resnick (1998), for instance, minimize tracking error standard deviation, but measure realized performance in terms of mean absolute tracking error.
The set of investable securities

The set of securities, which shall be considered when the index fund is to be revised, needs to be specified. The investable set, often called the universe, may be viewed as a part of the policy guidelines. For instance, this set can be restricted to include only the stocks in index or a subset of the index members. Another option would be to allow for inclusion all the stocks in the particular market. Still another option is to allow the fund to select from the set that is the union of the index stocks and the fund’s current holdings, which may include cash.

Let $B$ be the (non-empty) set of stocks in the benchmark, or index. Let $U$ denote the universe (or the set) of investable securities. Let $X_0$ be the union (containing all unique members) of the fund’s current holdings (including cash, if any) and the universe $U$. Now form the set $S$ as the union of $B$ and $X_0$, i.e., $S = B \cup X_0$, and let $n$ be the cardinality (the number of members) of $S$, i.e., $n = \text{Card}(S)$. The set $S$ is the set of securities that is considered in the index fund revisions; it may contain both investable and not investable securities. (Unless $U$ is specified, $S$ is quite general in that $S$ depends on how the universe of investable stocks is defined.) In the empirical analysis, the investable set $U$ is the union of the current fund holdings (that are not delisted the next day), the current index stocks (that are not delisted the next day), and cash.

Notation

Let $x$ be the $n \times 1$ column vector corresponding to $S$ of portfolio weights for the index fund, $x_0$ the $n \times 1$ column vector (corresponding to $S$) of current index fund weights, and $b$ the $n \times 1$ column vector (corresponding to $S$) of index weights that will be in effect at, or immediately after, the portfolio construction. Given an objective and constraints, the decision problem is to determine $x$.

If the index fund is to be constructed from a pre-defined subset of the index stocks, then this can be implemented by setting the appropriate bounds constraints on $x$, so that only stocks in the subset are allowed to have non-zero weights.

Let $\mathbf{r}$ be the $n \times 1$ column vector of the investable securities’ returns (for the upcoming period), which are considered jointly distributed random variables. Let
\( \mathbf{r} = \mathbf{E}(\tilde{\mathbf{r}}) \) be the \( n \times 1 \) column vector of expected returns on the investable securities, and \( \text{Var}(\tilde{\mathbf{r}}) = \mathbf{E}[(\tilde{\mathbf{r}} - \mathbf{r})(\tilde{\mathbf{r}} - \mathbf{r})'] \) the \( n \times n \) covariance matrix of the same securities’ returns.

**Illustrative problem formulation**

If the fund policy, for instance, is to minimize tracking error standard deviation, then the index fund problem (without regard to transaction costs) could be stated as:

\[
\text{minimize } \text{Var}(x'\tilde{\mathbf{r}} - \mathbf{b'}\tilde{\mathbf{r}}) = (x - \mathbf{b})'\text{Var}(\tilde{\mathbf{r}})(x - \mathbf{b})
\]

with respect to \( x \), and subject to the standard portfolio constraint that weights sum to unity. Then, if there are no subset constraints or other complications, the trivial solution is \( x = \mathbf{b} \).

Moreover, let \( \mathbf{R} \) be the \( T \times n \) matrix of historical returns over \( T \) prior periods for the \( n \) securities in \( \mathbf{S} \), \( \mathbf{r}_b \) the \( T \times 1 \) column vector with actual realized benchmark returns, and \( \mathbf{r}_p \) the \( T \times 1 \) column vector of realized returns of the replicating portfolio. Note that some securities in \( \mathbf{S} \) may have less than \( T \) periods of historical returns in \( \mathbf{R} \), and that the “missing” returns in some situations therefore have to be imputed.

**4.3.4 Forecasting and estimation issues**

Models of portfolio and index fund choice involve \( \mathbf{E}(\tilde{\mathbf{r}}) \) and \( \text{Var}(\tilde{\mathbf{r}}) \), which are expectations of random variables, and whose future realizations have to be forecasted.

As was discussed in Chapter 2, optimized portfolios or solution vectors are sensitive to forecast errors in input parameters, so accurate forecasting is important. The approach taken here will be a simple one, namely to use (historical) sample realizations. The covariance matrix \( \text{Var}(\tilde{\mathbf{r}}) \) will be forecasted as \( \mathbf{V} = \text{Var}(\mathbf{R}) \). In view of this, the above objective becomes to minimize \( (x - \mathbf{b})'\mathbf{V}(x - \mathbf{b}) \) with respect to \( x \).

In essence, the optimization approach may be viewed as selecting a fund, which over a historical period would have had an optimal tracking performance (according to the specified measure and given the current composition of the index). This implies that a potentially important source of tracking error is related to the correspondence between the optimal performance over the historical period and the actual future tracking performance. Recent empirical evidence indicates, however, that little if nothing is to be gained by using, in tracking error optimization based on daily data under no-short-sale restrictions, other estimators than the sample covariance matrix:
"When the no-short-sale restriction is already in place, minimum variance and minimum tracking error portfolios constructed using the sample covariance matrix perform as well as those constructed using covariance matrices estimated using factor models and shrinkage methods. The use of daily return data instead of monthly return data helps when constructing minimum tracking error portfolios but not in constructing minimum variance portfolios." Jagannathan and Ma (2003, p. 1677)

Mean returns are subject to high estimation risk and difficult to forecast, and optimized portfolios are generally extremely sensitive to changes in the means (Best and Grauer 1991a, 1991b), so small forecast errors of $r$ might induce large errors in solution vectors. Chopra and Ziemba (1993) demonstrate that errors in mean returns are an order of magnitude more serious than errors in the covariance matrix. Because of this, mean returns $r$ are set to zero. This may be viewed as an extreme shrinkage operation; so-called shrinkage estimators have been suggested as a means for reducing estimation errors (for an overview, see Michaud 1998, Ch. 8). In addition, Rudd and Clasing (1982) and Rosenberg and Rudd (1979) argue that mean returns are too judgmental and should therefore be excluded from index tracking formulations.

**Forward-looking and backward-looking assessments of tracking error**

It is important to distinguish between forward-looking and backward-looking assessments of tracking performance. Backward-looking assessment is intended to measure realized performance, that is, what has happened, while forward-looking assessment is concerned with future performance, that is, what will happen.

It seems reasonable that measurements of realized tracking performance are based on the actual realized returns $r_p$ and $r_b$. Let $\hat{e} = r_p - r_b$ be the realized tracking error vector. An often used measure is the (sample) standard deviation of the difference between $r_p$ and $r_b$, that is, $s(\hat{e})$. It is possible to use measures based on $r_p$ and $r_b$, such as $s(\hat{e})$, to forecast the (future) tracking performance of the fund (see Meade and Salkin 1990 in the context of conditional updating).

Objectives used for revising by means of optimization or regression the current fund so as to achieve an desirable level of future tracking performance, have also been based on $r_b$ (Walsh et al. 1998; Bamberg and Wagner 2000). I argue that one should if possible avoid using measures that involves $r_b$ and $r_p$ in forward-looking situations, because the index composition today, i.e., the composition that is going into the future, may be different from what it was historically, that is, when the index’ returns were $r_b$.
- and likewise for \( r_p \). One should, instead, compute what the historical return of the fund and the index would have been, given their current compositions, and use these return series - instead of \( r_p \) and \( r_b \) - to obtain a (historical) tracking error vector \( e \) more relevant for forward-looking purposes. The vector \( e \) is computed in the following manner:

\[
e = R_x - R_b = R(x-b) = (x-b)'R,
\]

where \( x \), \( b \), and \( R \) are based on the set \( S \).

As was emphasized by Bamberg and Wagner (2000) and Satchell and Hwang (2001), this computation effectively treats the weight vectors \( b \) and \( x \) as constants. Thus, to really obtain the return series corresponding to the vectors \( R_x \) and \( R_b \), rebalancing so as to keep the weights constant would be required, as portfolio weights change over time, unless all portfolio constituents generate identical returns over any given period.

The fund policy is assumed to indicate the desired tracking performance and how the tracking performance is to be measured. So, for a given index fund, in the backward-looking situation or performance evaluation, the particular measure (of dispersion) to apply to \( \hat{e} \) is given. It seems reasonable that the measure that will be used in forward-looking situations is consistent with the measure used for performance evaluation, and that, in case optimization is to be used to select the fund, a consistent mathematical program is devised. In forward-looking situations, the vector \( e \) will be used here, and hence, in effect, weights will be treated as constant, which is in conformity with the majority of prior research. As noted, however, an approach with variable weights based on Satchell and Hwang (2001) is tested. Moreover, I examine empirically if tracking performance is affected by whether an updated index return vector \( R_b \) or a possibly outdated one, \( r_b \), is used.

Note on discrete vs. continuous compounding of returns
Several studies, e.g., Pope and Yadav (1994), Walsh et al. (1998), Rudolf et al. (1999), Bamberg and Wagner (2000), use continuously compounded returns, or log returns (the natural logarithm of price relatives), instead of discretely compounded returns, or simple returns. A problem using log returns is that there is no exact relation between the return on individual stocks and the portfolio over a given period. The weighted
sum of the constituents’ log returns is not equal to the log return of the portfolio. Neither is the weighted sum of the constituents’ log returns equal to the portfolio’s simple return. With simple returns there is an exact portfolio relation in that the return of a portfolio over a given period is equal to the weighted sum of the returns of the constituents over the same period. Therefore, when returns are computed for individual stocks, e.g., the return matrix $\mathbf{R}$, and for portfolios, simple returns will be used and not log returns.

Simple returns in time-series are disadvantageous in statistical analysis for they have a lower bound of $-1$ and this makes them incompatible with the normal distribution. Moreover, for simple returns, cumulative subperiod returns do not equal the (compound) return for the full period. Tracking error is the return deviation between the fund and the index. The cumulative deviation between a fund and an index over a given period, is not equal to the sum of single (subperiod) simple return deviations (tracking errors). Log returns possess the convenient property that the total compound return (the log price relative) for the full period equals the sum of subperiod log returns. This applies, however, to a single return series. So, if the fund returns and the index returns are computed using log returns, it is not the case that the sum of the differences of their log returns is equal to the cumulative deviation of the fund and the index.

In the following, simple returns are used in the analysis of return and tracking error return series. Almost all computations performed in this chapter, have, however, also been done with log returns (instead of simple returns). The results were virtually unchanged. Moreover, using simple returns is relevant to the extent that tracking error measurements in practice are based on such returns.

### 4.3.5 Tracking error measures

Measures of tracking error (dispersion) having appeared in the literature include the ones below.

#### Tracking error volatility measures

The traditional measure of tracking error dispersion is tracking error standard deviation, that is, the square root of the central second moment or of the variance (see Roll 1992, also Rudd 1980), based on realized tracking error $\hat{e}$:
\[
SDTE = \sqrt{\text{Var}(\hat{e})} = \sqrt{\text{Var}(r_p - r_b)}, \quad (4.1)
\]

and based on the historical tracking error \(e\):\(^{56}\)
\[
SDTE = \sqrt{\text{Var}(e)} = \sqrt{\text{Var}(R_x - R_b)} = (x-b)'\text{Var}(R)(x-b) = (x-b)'V(x-b). \quad (4.2)
\]

According to Frino and Gallagher (2001, p. 47), \(SDTE\) is “the standard methodology used in industry.” Let \(V=\text{Var}(R)\) represent the sample variance-covariance matrix of \(R\), then a consistent objective for quadratic optimization is (see, e.g., Roll 1992):\(^{57}\)
\[
\text{min } (x-b)'V(x-b).
\]

Another tracking error measure is the square root of the non-central second moment, or the root mean square, of the realized tracking error \(\hat{e}\):
\[
RMSTE = \sqrt{1/(T-1) \hat{e}'\hat{e}} = \sqrt{1/(T-1)(r_p - r_b)'(r_p - r_b)},
\]

and based on \(e\):
\[
RMSTE = \sqrt{1/(T-1) \ e'\ e}
\]
\[
= \sqrt{1/(T-1)(R_x - R_b)'(R_x - R_b)}
\]
\[
= \sqrt{1/(T-1)(x-b)'R'R(x-b)}
\]

In contrast to Frino and Gallagher (2001), Rey and Seiler (2001) report that \(RMSTE\) rather than \(SDTE\) is the measure most frequently used in practice. Using as, e.g., Rudolf et al. (1999) and Rey and Seiler (2001), \(\text{min } (R_x-r_b)'(R_x-r_b)\) for an objective could be dangerous, for \(r_b\) may be based on securities not in \(b\) at the time of optimization. A more appropriate objective\(^{58}\) should be
\[
\text{min } (x-b)'R'R(x-b).
\]

\(RMSTE\) and \(SDTE\) are sometimes referred to as tracking error volatility measures. Rey and Seiler (2001) and Beasley et al. (2003) report that they prefer \(RMSTE\) over \(SDTE\), because, as was pointed out by Roll (1992), the latter measure implies that a tracking portfolio that constantly over- or underperforms the index by a fixed amount has a tracking error volatility of zero, while the former does not. In expectation, the

\(^{56}\) I will not use separate notations depending on whether \(\hat{e}\) or \(e\) is involved, as this will be obvious.

\(^{57}\) Rudolf et al. (1999, p. 87) wrongly claim that \(e'e\) is equivalent to \(\text{Var}(e)\) referring to Roll (1992).

\(^{58}\) If one needs to optimize towards a particular value (e.g., a limit), then that value should be scaled so that it becomes consistent with the numerical value of a given objective.
root mean squared error is per definition equal to the variance plus the mean squared, so \( \text{RMSTE} \) and \( \text{SDTE} \) relate as

\[
\text{RMSTE}^2 = \text{SDTE}^2 + \overline{TE}^2,
\]

where \( \overline{TE}^2 \) denotes mean tracking error squared. Thus, if \( \overline{TE}^2 \) is close to zero, \( \text{RMSTE} \) and \( \text{SDTE} \) will also be close. How close is an empirical matter, but neither Rey and Seiler (2001) nor Beasley et al. (2003) present any evidence allowing a direct comparison of the empirical performances of the measures. Such evidence will, however, be produced here, and since the measures can be used as objectives as well as ex-post measures, the evidence will cover both aspects. Because the performance of \( \text{SDTE} \) and \( \text{RMSTE} \) as ex-post measures were identical, except for the worst performing subset funds, for which small differences were observed, the results for \( \text{RMSTE} \) as an ex-post measure are not reported throughout.

**Autocorrelation adjustment of tracking error standard deviation**

Estimates of tracking error standard deviation are usually computed from short-interval returns, typically daily or weekly returns. The short-interval estimates are then converted to a common basis. This is done to facilitate comparisons and to achieve compatibility with reporting standards, among other things. The common basis is usually annual, and the operation is then referred to as annualizing. Without autocorrelation in the shorter-interval returns, the conversion is straightforward, as is demonstrated below. However, if there are positive (negative) autocorrelation in the shorter-interval returns and if that is not accounted for, annualized standard deviations will be biased downwards (upwards).

The problem of comparing tracking error standard deviations based on returns measured over intervals of different length in the presence of autocorrelation was pointed out by Meade and Salkin (1989) and Pope and Yadav (1994). In annualizing standard deviations of short-interval tracking errors in the presence of autocorrelation, the latter pair of writers applies the results of Campbell et al. (1997) to obtain annualized autocorrelation corrected tracking error standard deviations. In particular, Campbell et al. (1997, Eq. 2.4.41) show that in the presence of autocorrelation, the ratio of variances computed from returns of different frequencies relate as
\[
\frac{\text{Var}(r_t[q])}{q \text{ Var}(r_t)} = 1 + 2 \sum_{k=1}^{q-1} \left(1 - \frac{k}{q}\right) \hat{\rho}(k),
\]  

(4.3)

where \(\text{Var}(r_t[q])\) is the variance based on returns measured over \(q\) 1-periods (longer-interval), \(\text{Var}(r_t)\) is the variance based on 1-period returns (shorter-interval), and \(\hat{\rho}(k)\) is the estimated \(k\)th order autocorrelation coefficient of \(\{r_t\}\). Denote by

\[
A[q] = 1 + 2 \sum_{k=1}^{q-1} \left(1 - \frac{k}{q}\right) \hat{\rho}(k)
\]

the autocorrelation adjustment factor. Then

\[
\text{Var}(r_t[q]) = q \text{ Var}(r_t) \left(1 + 2 \sum_{k=1}^{q-1} \left(1 - \frac{k}{q}\right) \hat{\rho}(k)\right) = q \text{ Var}(r_t) A[q].
\]

(4.4)

The annualized autocorrelation corrected return variance, assuming, for instance, 252 days in a year, can then be computed from daily return variance as:

\[
\text{Var}(r_t[252]) = 252 \text{ Var}(r_t) A[252].
\]

Absent autocorrelation in the short-interval returns, that is, with \(A[q]\) equal to one, the annualized (autocorrelation corrected) variance is simply equal to \(q\) times the short-interval return variance. However, if short-interval returns exhibit autocorrelation, that is, with \(A[q]\) different from one, the simple relation no longer holds, and the annualized autocorrelation corrected variance is given by (4.4).

Usually, evaluations of tracking error volatility are based on annualized estimates of daily or monthly returns. Therefore, in the empirical analysis autocorrelation-corrected annualized tracking error standard deviations are calculated with (4.4) using, for monthly data, \(q=12\) and, for daily data, \(q=252\), taking into account all possible non-zero autocorrelations up to the annual level, and with returns replaced by tracking errors (return differences). Regardless of whether individual autocorrelation coefficients are statistically significant or not, they do affect the longer interval returns, and should thus be accounted for. No tracking error volatilities are computed from yearly returns, for with ten years of price data there will be only nine or ten return observations, and with so few observations sampling error is likely to be an issue.

The relation (4.3) is developed for variances and applies, therefore, directly to \(SDTE\), which is based on the variance, while the application to others measures, such as \(RMSTE\), which is based on mean squared error, would be ad hoc in the absence of
relevant theory. Autocorrelation adjustments are therefore not applied to other measures than SDTE.

For all the index funds implemented and tested, except the subset funds, the realized annualized daily and monthly tracking error standard deviations never deviated much (the maximum absolute ratio of the daily and the monthly estimate minus one equaled 17%). Moreover, the deviations of the unadjusted monthly estimates from the adjusted daily estimates were similar to the deviations of the unadjusted daily estimates from the adjusted daily estimates, except for the subset funds, for which the unadjusted monthly estimates were closer to the adjusted daily estimates. More importantly, the realized autocorrelation-corrected annualized daily and monthly tracking errors never deviated much for any index fund implemented and tested (the maximum absolute ratio of the daily and the monthly estimate minus one was 10%). Because of this, the monthly estimates were not reported.

Other measures of tracking performance
Several studies measure tracking error by **Sigma**, the standard error of a regression of index fund returns \( r_{pt} \) on index returns \( r_{bt} \) (see, e.g., Frino and Gallagher 2001, 2002; Frino et al. 2004; Rey and Seiler 2001):

\[
    r_{pt} = \alpha_p + \beta_p r_{bt} + \varepsilon_{pt}.
\]

This also yields the so-called risk-adjusted performance of the index fund in terms of alpha and beta estimates (see, e.g., Frino and Gallagher 2001, p. 50). The standard error of the regression, or residual volatility, is calculated as the square root of the residual variance divided by the number of observations minus two (the degrees of freedom):

\[
    \hat{\sigma}_p = \sqrt{\frac{1}{T - 2} \sum_{t=1}^{T} \hat{\varepsilon}_{pt}^2}
\]

In case beta is equal to one and the regression is well-specified otherwise, then the regression standard error \( \hat{\sigma}_p \) will be close to SDTE. It would be possible to use the coefficient of determination, \( R^2 \), as a gauge of tracking performance as Larsen and Resnick (1998) do. Blume and Edelen (2004), however, argue convincingly that it is difficult to discriminate between reasonably performing index funds on basis of \( R^2 \).
Larsen and Resnick (1998), Gallagher and Frino (2001, 2002), and Frino et al. (2004) use the following mean absolute tracking error (MATE) measure:

\[
MATE = \frac{1}{T} \sum_{t=1}^{T} |e_t|, \tag{4.5}
\]

where \(e_t\) is the tracking error of period \(t\).

Rudolf et al. (1999) express concern about the interpretability of quadratic tracking error measures, e.g., root mean square or standard deviation, and suggest several alternatives, including one defined as MATE, but which they call “mean absolute deviation (MAD)”. It may, however, be worthwhile to point out that the above definition of “mean absolute deviation” is different from that prevailing in other contexts, where it refers to the mean absolute deviation from the mean or the mean absolute deviation from the median\(^{59}\); Satchell and Hwang (2001, p. 243) actually declare that Rudolf et al. (1999) use the mean absolute deviation from the mean. The mean absolute tracking error definition (4.5) is not (generally) equivalent with the other two measures, whose interpretations, as measures of the dispersion of a distribution, are not entirely straightforward or familiar. For instance, Herrey (1965, p. 258) writes that “…the properties of the mean absolute deviation are more complex and less well known than those of the standard deviation…” In view of this, it seems less clear why a measure defined as MATE should be easier to interpret than, for example, the standard deviation tracking error.

Herrey (1965) reports that for a normal distribution, the mean absolute deviation from the mean, MAD, and the standard deviation, \(\sigma\), relate as \(MAD = \left(\sqrt{2/\pi}\right)\sigma \approx 0.8\,\sigma\). If the mean of the normal distribution is zero, then MATE and MAD are equivalent and the relation applies to MATE as well. Under these conditions, the numerical value of MATE will be approximately equal to 0.8 times the standard deviation. In the empirical analysis, an index fund selection formulation based on MATE is implemented and tested.

\(^{59}\) See, e.g., the portfolio formulation of Konno and Yamashiki (1991) or statistics-oriented literature, e.g., Bonett and Seier (2003).
A possible advantage with a measure that is based on absolute tracking errors is its robustness in that it should be “less sensitive against outliers than the mean square models.” (Rudolf et al. 1999, p. 88).

Bamberg and Wagner (2000) estimate the tracking error standard deviation by the square root of the sample variance, and by a robust alternative, the standardized median absolute deviation from the median. A robust estimator of the standard deviation could be useful in conjunction with conditional updating of an index fund. If there is a single extreme tracking error rather than a systematic error in the recent history, or sample, on which the update is conditioned, then the ordinary sample standard deviation might signal a need to update, whereas the robust measure might not. However, if the tracking performance is to be measured in terms of the ordinary standard deviation of the tracking error, then an update in response to an extreme observation may help to achieve the desired tracking performance, given that the update improves the tracking performance and that the outlier negatively affects overall realized tracking performance. In addition, Bamberg and Wagner (2000) find empirical evidence that when there are extreme observations in-sample, then robust regression estimators generates better out-of-sample performance than do least squares.

I devise a heuristic procedure for forecasting a measure of tracking error dispersion that is in the spirit of Satchell and Hwang (2001), Satchell and Hwang measure (SHTE).

“Such strategy-based estimates could be calculated and would be a useful research contribution.” Satchell and Hwang (2001, p. 244)

The procedure takes the current (or effective next day) weights of the index and the fund, and computes what, based on $\mathbf{R}$, the historical returns on the fund and the index, respectively, would have been with varying weights starting from the first day in $\mathbf{R}$. A measure of tracking error could then be computed from these return series and be used as a forecast. (Note that this procedure only calculates one possible realization of the information in $\mathbf{R}$ and that the return series are dependent on the order of the returns in $\mathbf{R}$; starting, e.g., from the last day in $\mathbf{R}$ and computing the returns backwards in time, would lead to different results.)
I implement and test an index fund selection model based on the just described procedure. It selects the index fund that produces a minimal (value of the selected) tracking error measure based on historical index and fund return series computed with varying weights. Transaction costs are not considered in the approach.

4.3.6 **Formulation of index fund revision models under transaction costs**

“The majority of the work relating to index tracking presented in the literature does not consider transaction costs and only considers the problem of creating an initial tracking portfolio from cash.” Beasley et al. (2003, p. 623)

Below, I formulate two index fund revision programs under transaction costs including price impact costs. These are related to the mean-variance portfolio revision model under transaction costs, in particular the first one. As was mentioned above, in the index fund revision problems, no explicit forecast of mean returns will be included.

To derive the first program, therefore, in the goal function of (3.2) in Chapter 3, delete $R$, replace $x$ with $x - b = x_0 + x^+ - x^- - b$, replace $\lambda$ (or Lambda) with $\tau$ (or Tau), and insert the components of the transaction cost functions to obtain:

$$
\max_{\{x^+, x^-, \lambda\}} -\tau (x_0 + x^+ - x^- - b)'V (x_0 + x^+ - x^- - b) - x^+ c^+ - x^- c^- - x^+ d^+ x^- - x^- d^- x^-
$$

which as a goal function is equivalent to following minimization problem:

$$
\min_{\{x^+, x^-\}} \tau (x_0 + x^+ - x^- - b)'V(x_0 + x^+ - x^- - b) + x^+ c^+ + x^- c^- + x^+ d^+ x^- + x^- d^- x^-.
$$

This objective minimizes the sum of tracking error variance, effectively $SDTE$, and transaction costs, where the trade-off parameter $\tau$ determines the emphasis that should be put on minimizing tracking error variance relative to transaction cost. In relative terms, emphasis on tracking error and transaction costs are $\tau/(1 + \tau)$ and $1/(1 + \tau)$, respectively. An issue with this formulation is the value of $\tau$; it has to be determined empirically, as there are no suggestions about its value a priori. Rearrange (4.6), state it in the same form as equation (3.6), drop constant terms, and include constraints to obtain the complete formulation:

$$
\min_{\{X\}} X' \left( \begin{array}{c} 2\tau V(x_0 - b) + c^+ \\ -2\tau V(x_0 - b) + c^- \end{array} \right) + \tau X'(W + D/\tau)X
$$

(Opt1)
subject to \[
\begin{bmatrix}
1 \\
-1
\end{bmatrix} X = 0, \\
\max(l_i - x_{i0}, 0) \leq x_i^+ \leq \max(u_i - x_{i0}, 0), \quad i = 1, 2, ..., n, \\
\max(x_{i0} - u_i, 0) \leq x_i^- \leq \max(x_{i0} - l_i, 0), \quad i = 1, 2, ..., n.
\]

I also suggest the following intuitively appealing index fund revision formulation, where the objective is to minimize transaction costs while keeping tracking error volatility less than or equal to a pre-defined limit (leaving out the some of the constraints):\(^{60}\)

\[
\begin{aligned}
\min_{\{X\}} & \quad X' \begin{bmatrix} c^- \\ c^+ \end{bmatrix} + X'DX \\
\text{subject to} & \quad (x - b)'V(x - b) \leq TE\text{limit}
\end{aligned}
\]

which, expanded with the complete set of constraints, becomes:

\[
\begin{aligned}
\min_{\{X\}} & \quad X' \begin{bmatrix} c^- \\ c^+ \end{bmatrix} + X'DX \\
\text{subject to} & \quad X'WX + 2 X' \begin{bmatrix} V(x_0 - b) \\ -V(x_0 - b) \end{bmatrix} -(x_0 - b)'V(x_0 - b) \leq TE\text{limit}, \\
& \quad \begin{bmatrix} 1 \\
-1
\end{bmatrix} X = 0, \\
& \quad \max(l_i - x_{i0}, 0) \leq x_i^+ \leq \max(u_i - x_{i0}, 0), \quad i = 1, 2, ..., n, \\
& \quad \max(x_{i0} - u_i, 0) \leq x_i^- \leq \max(x_{i0} - l_i, 0), \quad i = 1, 2, ..., n.
\end{aligned}
\]

As noted, the first approach is perhaps more heuristic, but has a few advantages vis-à-vis the second alternative. It is implemented as a quadratic program and is solved in significantly shorter time than the second formulation, which is a quadratically constrained quadratic program (QCQP). The second approach might in some cases fail to satisfy the tracking error constraint. Note that it is straightforward in both formulations above to change the preferred tracking error measure from SDTE to RMSTE by replacing \(V\) with \(R'R\).

---

\(^{60}\) A selection formulation - rather than a revision formulation - based on the same logic, but within the mean-absolute deviation framework of Konno (1991), is suggested by Konno and Wijayanayake (2001).
4.3.7 **Conditional and unconditional updating**

Revising a full replication index fund when a forecasted tracking error measure exceeds a pre-defined limit seems like a natural and potentially successful variation of full replication, which is why I choose to implement it. It augments full replication by allowing some tracking error to exist without enforcing full revision of the fund, thereby potentially saving transaction costs and reducing tracking error. This is conceptually similar to the second index fund revision approach where the objective is to minimize transaction costs while keeping tracking error volatility less than or equal to a pre-defined limit. However, when the condition is triggered the fund will be revised into a fully replicating one, and not one that just meets the tracking error limit.

*Statistically based conditional updating*

A variation of conditional updating would be to update only when the forecasted tracking error measure exceeds the limit by an amount that is statistically significant at a desired level. This would typically entail a chi-square test of variance, which is based on a normal distributed population. The test is, however, not robust to deviations from normality, which are expected to be present. Therefore, I choose not to implement this kind of conditional updating.

4.3.8 **Pre-defined subsets as means to control transaction costs**

Many studies construct an index fund from a pre-defined subset of the stocks in the index (e.g., Larsen and Resnick 1998; Bamberg and Wagner 2000); Beasley et al. (2003) is an exception in that they develop a method that has the distinctive strength of being able to select among all stocks in the universe, the \( m \) stocks that yield the best projected tracking performance according to the preferred measure. Using a pre-defined subset is to intended to reduce transaction, handling, and computational costs by excluding a priori from consideration in the index fund the most illiquid, or otherwise unwanted stocks in the index.

Constructing an index fund from a subset to reduce transaction costs, however, appears reasonable if no explicit consideration is given to transaction costs in the selection process of the index fund (or if one has low faith in such an approach). Using a subset instead of the full set of index stocks will generally lead to increased tracking error. In addition, restricting the number of stocks in an index fund implies that the
average weight will be larger than if no such restrictions were imposed. Larger weights seem to imply larger trades and perhaps more frequent updates in order to maintain a desirable level of tracking error. If there are high fixed costs (back office etc.) associated with holding certain stocks - which, at least for developed markets in the recent history, appears rather unlikely - and if these costs are not reflected in the transaction cost model, then excluding these stocks may be sound practice.

Rudd (1980) offers another reason for excluding stocks and that is that certain stocks may be ineligible on basis of fund policy considerations. For indexes with very many constituents, as the Wilshire 5000, which contains over 5000 stocks, computational expense might motivate the use of a pre-defined subset.

To examine empirically the effects of using subsets on tracking performance and incurred transaction costs, index funds are constructed from at most $m$ stocks, where those stocks are the $m$ largest ones in terms of market capitalization at the time of the construction. Ranking on capitalization is a simple and probably efficient way of implementing indexing based on a subset, for the index tracked here is itself capitalization-weighted, and, in terms of the empirical findings of Chapter 2, transaction costs decrease with capitalization, all else equal. In addition, it appears reasonable to assume that handling and similar costs also decrease with capitalization. In prior studies, the size of the subset as a proportion of the full set of index stocks have varied; for instance, Rey and Seiler (2001) use (in one test setting) 1 stock out of 29, or 3.4%, and Bamberg (1998) use 5 out 30, or 17%, whereas Bamberg and Wagner (2000), use 20 out of 30, or 67%. Subsets of 10 and 20 stocks, respectively, are considered here. The results for the index funds constructed using the smaller subset are, however, not reported, because their tracking performance was less than satisfactory.

**Cash position**

Whether a cash holding should be permitted in an index fund, and whether it can help to bring down transaction costs are questions related to the preceding discussion about subsets. Connor and Leland (1995) point out that standard equity indexes have a zero weight in cash, while index funds generally maintain a positive cash holding, and that this induces tracking error. The advantage with a positive cash holding is that it can
accommodate temporal cash inflows and outflows, which, if they were to be unconditionally allocated to stocks, could cause unnecessarily large transaction costs.

A positive cash position is generally permitted in the index funds implemented in the empirical tests. For full replication funds, however, cash is not allowed, as the fund must hold the index stocks in the exact proportions they have in the index. For subset funds, a cash holding is generally not allowed. When a subset fund is revised, only the stocks in the pre-defined subset are considered. In an attempt to increase the flexibility of subset index funds, cash is however allowed to enter the fund under certain circumstances. A subset fund may contain fewer stocks than the number of stocks in the subset; it is not necessary that holding all the stocks in the subset is optimal considering tracking error and transaction costs. If the fund holds fewer stocks than the cardinality of the subset, and there is a positive cash inflow to the fund, then a positive cash position is permitted to be established. This requires that the fund meets its desired tracking performance, otherwise the fund needs to be revised, and, as noted, in the revisions only stocks in the subset are considered.

The behavior of the cash position is captured in the empirical analysis by a variable measuring the average portfolio weight for cash.

4.3.9 Index methodology: calculation rules and their implementation

The methodology used in computing the index is described below. The methodology is largely based on the OMX rules circa 2000, which since then have remained virtually unchanged. The index methodology is implemented in the form of a computer program that calculates the index. The approach is flexible in the respect that by changing rules and parameters in the program it is possible to calculate indexes with other characteristics.

Periodic review

The OMX index is subject to periodic review two times per year. Any changes to the index due to the reviews is made effective as of the beginning of the first trading day

---

61 The OMX index methodology is described in the internet publication Addendum OMX, which is updated from time to time. The two versions of Addendum OMX that have been used here, were downloaded as of 14 Apr. 1999 and 17 Nov. 2000 (see Stockholm Stock Exchange 1999, 2000).

in January and July, respectively. (This implies that a full replication index fund should be updated at the close of the foregoing day.) The first premise for the OMX index is that it shall comprise the 30 stocks (or Swedish depository receipts), which, among all stocks on the Stockholm Stock Exchange, have the largest trading volume measured in SEK during a certain control period. Stocks may be included in, or excluded from, the index depending on their relative trading volume in the preceding control period, which constitutes the 6 months beginning 7 months before the review date. (A first construction of the index to be effective from the opening of the market 2 Jan. 1990 is made on the basis of a review conducted at the close of the last trading day in 1989, 29 Dec. 1989.) If a stock is taken over or converted during the control period, or between the end of the control period and the first day of the upcoming calendar half-year, then this is kept track of and the traded volumes are accumulated accordingly.

At the semi-annual reviews, the number of shares outstanding is updated for each of the index constituents. The number of shares shall apply for the whole period until the next review, unless it changes significantly (in the opinion of the Index Calculator) - and in which case it is updated. Here, however, the number of shares outstanding for an individual index constituent is updated if it changes by more than 5% between the periodic reviews. This is consistent with the index methodologies of, e.g., STOXX, FTSE, and MSCI.

For a period of time, OMXCap, a so-called capped version of the OMX index was computed. In the capped index, constituents’ portfolio weights were limited to a maximum of 20% by applying a cap factor. Capping is intended to protect the index from lack of diversification. Moreover, law applicable to investment companies and investment funds usually impose upper bounds on the portfolio weights of individual holdings. Capping the index in a manner consistent with such bounds should make it easier for index funds to replicate the index. Several international indexes, e.g., those provided by STOXX, are capped at 10%. In this study, a 20% weight cap is used for consistency with the OMXCap and with the OMX, but note that Swedish law basically prescribes maximum weights of 10%; the 10% constraint can, however, be circumvented by registering the fund as a “Nationell fond” for which less restrictive
rules regarding portfolio weights apply. In association with the periodic reviews, weight cap factors are computed and applied to weights so as to prevent weights of individual constituents from exceeding 0.2. Between periodic reviews, the cap factors are kept unchanged, implying that weights are allowed to vary freely along with movements in market prices and changes in the number of shares outstanding (smaller than 5%), and the weights may hence come to exceed 0.2.

To limit index composition changes and increase index “stability” the following “inertial rule” (see Indices account at www.stockholmsborsen.se) based on buffer zones (see Stoxx Ltd., 2003) is applied in the OMX methodology, and here. If, during the control period, an index stock is not among the 45 most traded stocks on the exchange, it shall be replaced by the non-index stock with the highest traded volume during the control period. If a stock is listed on the exchange, but is not an index stock, and is among the 15 most traded stocks on the exchange during the control period, that stock shall replace the index stock which has the lowest traded volume.

**Ongoing review**

Between periodic reviews, the composition of the index may also change in response to corporate actions including mergers and acquisitions, IPOs, new issues, conversions, voluntary delistings, share repurchases, spin-offs, splits, and bonus issues. If the index is computed with dividends reinvested, dividend reinvestment affects index composition.

In addition to the aforementioned events, index composition changes might on rare occasions be effectuated at the discretion of the Index Calculator; this could happen when a constituent, in the opinion of the Index Calculator, is illiquid or otherwise unsuitable for inclusion in the index and therefore has to be deleted. Due to their ad hoc nature, this last type of index composition changes is not implemented in this study.

Certain aspects of the implementation of the index methodology are described below. How these issues are treated in the index calculations, in a full replication index fund, and, when applicable, in other indexing approaches, is also related.
The index number and the index divisor - formulas

The OMX index, like many other stock market indexes, is essentially calculated as a chain-linked Laspeyres price index number. Basically, this means that the OMX index represents, at different points in time, the development of the total market value of all the stocks in the index based on the number of shares outstanding at the first point. This provides a meaningful way of measuring development over time for the value of the aggregate of stocks in the index. At some points in time (and for a variety of reasons), however, stocks in the index and or their number of shares outstanding may change. In such instances, to uphold the continuity of the index, a so-called index divisor is applied (see, e.g., Dow Jones & Company, Inc., 2002, p. 8).

The procedure for calculating the index number at time $t$, $I_t$, is as follows. The capped market capitalization of the stocks in the index at time $t$ is:

$$M_t = \sum_{i=1}^{n} p_{it} q_{it} f_{it},$$

where $p_{it} =$ price of stock $i$ at $t$, $q_{it} =$ the number of shares outstanding of stock $i$ at $t$, and $f_{it} =$ the cap factor for stock $i$ at $t$.

An initial index divisor is computed as $D_0 = \frac{M_0}{I_0}$, where $I_0$ is the (arbitrary) base index value, usually a multiple of 10. Here a base index value of 1000 will be used; $D_0$ will then equal $1/1000$ of $M_0$. The index at the close day $t$ is then computed as

$$I_t = \frac{M_t}{D_t}. \quad (4.7)$$

If there are index composition changes, or corporate actions, of $CA$ monetary units after the close of day $t$, effective day $t+1$, the index divisor for $t+1$, $D_{t+1}$, is updated so as to ensure that the continuity of the index is maintained:

$$D_{t+1} = \frac{M_t + CA}{I_t}.$$

Rewriting $D_{t+1}$ in terms of $D_t$ by using (4.7) shows how the new divisor relates to the old:

$$D_{t+1} = \frac{M_t + CA}{M_t / D_t} = D_t \left( 1 + \frac{CA}{M_t} \right).$$
For instance, if a stock with market cap of $CA$ at close day $t$, is to be added to the index effective at $t+1$, then the total market cap of the stocks in the index going into the next day, $t+1$, will be an amount $CA$ larger than $M_t$. As a consequence, the index going into the next day will also be larger, and would, unless adjusted for, destroy the continuity of the index. By increasing the next day’s divisor by the same proportion that the market cap of the index increases due to the added stock, the index will maintain its continuity; this is referred to as “rebasing the index”.

The foregoing discussion applied to a price index, which does not take account of dividend payments. In a total return index, dividends are assumed to be received and reinvested in perfect proportions across stocks in the index after the close the day prior to the ex-dividend day. This is achieved through adjustment (decreasing) of the index divisor.

*A central implementation construct: the weightlist (wlist)*

A central construct for the implementation of the index methodology as well as for the index tracking approaches, is the index weightlist ($wlist$). At the close day $t$, the index $wlist$ contains the index components and their weights effective the next day, $t+1$.

If, just prior to the close of the market at time $t$, the fund components and their weights deviate from the $wlist$ effective day $t+1$, a full replication fund needs to be updated according to the $wlist$. All the trading of the fund is thus supposed to take place at the close each day. In the absence of transaction costs and dividends, a fund that is revised at the close each day so as to match the $wlist$ will achieve perfect replication of the index.

*Treatment of dividends and corporate actions*

*Dividends*

The way dividends distort the performance of the fund relative to the index is somewhat intricate. In the index, dividends are assumed to be received and reinvested in perfect proportions across stocks in the index after the close the day prior to the ex-dividend day. (Note that this implies that the $wlist$ for the index with and without dividends reinvested are identical.) This means that the dividends will not increase the index level at the close of day $t$. However, through the adjustment made to the index
divisor of day $t+1$, the index level at the close of $t+1$ will include the dividends amount as if the dividends had been invested in the index after the close at time $t$, but before the opening day $t+1$.

Unfortunately, it is not possible for a fund to replicate this performance. One reason is that in reality dividends are paid out 8 trading days (that is the median, while the average is 8.47) after the day before the ex-dividend day.\textsuperscript{63} And if one assumes that the fund receives the dividends just prior to the close (possibly by borrowing the dividends amount from some willing counterpart at a negligible interest rate)\textsuperscript{64} and reinvests them immediately in perfect proportions, the market value of the fund will be inconsistent with the level of the index at the close day $t$.

The effects of different dividend payment delays on tracking performance are analyzed empirically. Based on that analysis, it will be assumed that dividends are paid out and available for trading at the close on the ex-dividend day. Given that the fund is perfectly replicated and that there are no transaction costs, this implies that there will be a return difference at the ex-dividend date, because the opening value of the index includes the dividends, whereas the fund NAV reflects the received dividends at the close. If the benchmark return is positive on the ex-dividend date, the index will exhibit a superior performance relative to the fund, because the dividends allocated to the index will grow at the benchmark rate of return, whereas the fund receives the dividends not until the end of the day.

Conversions and mergers and acquisitions (M&A)

Conversions and M&A are accounted for in this study. They are incorporated in the index calculations as follows. Consider an issue in the index at the close day $t$, and that this issue is subject to a conversion to a surviving issue (a new or an existing one), or an M&A where the payment is made in shares of the acquiring firm, effective the next day; the acquiring or surviving firm’s stock must be in the universe, otherwise the issue is deleted from the index effective from day $t+1$. The market value of the index

\textsuperscript{63} Source: Own calculations based on the full sample from 1987 up to and including 1999; the corresponding numbers including non-trading days are 13 for the median, 12.64 for the mean.

\textsuperscript{64} Stock index futures - if available on the relevant index and if the index fund is allowed to use such instruments - can also be used to handle cash and cash flows.
at the close day $t$ is computed based on the closing price and number of shares in the unconverted issue. The issue is then replaced by the surviving issue in the index. The market cap as of day $t$ of the non-surviving issue is subtracted and the market cap of the survivor as of day $t$ is added to the divisor of day $t+1$, to maintain the chainlinking of the index. The $wlist$ is updated consequently, to reflect the components and weights effective day $t+1$. When the buyer pays in cash, the acquired stock is deleted from the index (and the buying firm is not added to the index), and the $wlist$ is updated accordingly.

For a fund the following occurs in association with a conversion or an M&A. The fund’s NAV is computed using the to be converted issue’s closing price and number of outstanding shares at the close day $t$. Then, to be in effect for day $t+1$, the conversion takes place and the issue is replaced by the surviving issue using the same number of shares, while in the case of a M&A the number of shares is updated according to the conditions of the M&A. The $wlist$ thus contains the weight for the merged line effective the next day. Since the new shares in the merged line are not available for trading at the end of day $t$, the trading at $t$ will take place in the shares that are to be taken over in the following way. First, the holdings at $t$ are converted according to the terms of the M&A; this yields the holdings effective for day $t+1$, given no trading. These holdings are then compared to the $wlist$ that is to be in effect for the next day. Any differences are recorded, i.e., required weight changes. The recorded differences are translated back to what weight changes the fund must undertake in the holdings at the close of day $t$. If several lines are merged into a single one, no trading is done in any line but the single line with the largest capitalization. The fund’s NAV is computed based on the holdings and prices at close of day $t$.

**Splits and bonus issues**

No divisor adjustments are necessary for stock splits and bonus issues, since market capitalization does not change and the share number and share price are adjusted (automatically) prior to the opening of trading on the split's ex-date.

**Spin offs**

Data on spin-offs have been collected, but are not included in the study. In addition, the OMX index methodology does not consider spin-offs. One can for instance assume
that the proceeds from selling stock spun off are used to cover the operational costs for the fund.

Rights issues

In a rights issue, i.e., an issuance of new equity capital where current stockholders have priority to participate, the following happens in the real world.

Registered holders of the stock, at the end of a preannounced day, receive, based on the number of shares they hold, a certain number of rights. One right usually allows its holder to purchase one newly issued share at the subscription price that - to attract buyers - is a price below the stock’s recent price range. Holders of stock at the close, will, effective the next day, the ex-rights day, also own rights. Since a right allows the holder to buy newly issued stock at a reduction, it has an economic value. The rights will trade for some weeks, and after that the rights are (expected to be) exercised, the subscription price paid in and the new shares received (in the index the pay in and the issuance of new shares are assumed to be in effect from the beginning of the ex-rights day). Although the market value of the firm will increase as the newly issued stock is paid in, the market value per share or ex-rights stock price will decrease, because there will be more shares outstanding.

As mentioned, in the index calculation the issuance of, and the payment for, the new shares are effective the ex-rights day. The weight of the issue in the index will (in general) increase as a consequence of the rights issue. The fund must accommodate this increase in order to maintain tracking implying that shares must be sold in other holdings and invested in the issue with the larger weight. It appears, however, that the most straightforward way of doing this would be to decrease other holdings so as to raise enough money to buy the required amount of shares needed to achieve the target weight. This approach, however, is unattractive for several reasons. First, it implies that, effective the next day, the fund will be the owner of a number of rights that are not included in the index. The rights, therefore, need to be sold at their theoretical value and the proceeds reinvested in the stock at the ex-rights price at a point in time after the close, but before the opening, the ex-rights day, so that perfect weighting is attained; the value of the rights plus stocks at their ex-rights price will (of course)
equal the value of the stocks cum-rights, so the weighting will be perfect given the assumption.

In a more realistic setting, several problems and inefficiencies arise. Trading may have to be done in both the stock and the rights, but price and transaction cost data on rights are not available. A way to assure exact tracking bar transaction costs is to assume participation in the new issue, since then the desired weighting of the fund can be in effect at the right time. The amount of selling in other securities is the same, but the index fund would only try to hold so many shares that the value of those shares plus the value of the payment for the newly issued shares, obtained by the exercise of rights, results in the desired weighting of the fund. Assuming that the fund is perfectly weighted just prior the new issue, this scheme might sometimes include some selling of the stock, so that the value of the shares held plus the payment for the newly issued shares meets the target value.

Consider an index fund that is far from perfectly weighted just prior to the new issue. In such cases either selling or buying of the stock subject to the issue may occur. Due to time constraints, this more complex trading scheme surrounding new issues was not implemented in software.

Instead, the fund is assumed to transact at the close the number of shares required to achieve the target weighting for the next day, the ex-rights day; this means that the fund will be the holder of rights effective the next day. These rights should be sold and the proceeds reinvested in shares. For simplicity, however, it is presumed that the rights are converted to shares costlessly at their respective theoretical values and in time for being in effect the ex-rights day. It appears likely that a fund that is run by a bank or a large institution could perform such a transaction.

The ratio of the value of the rights distributed per dollar held of the stock is one determinant of the amount of trading that has to be done in rights and stock. Others are the weight of the stock in the index, the size of the new issue, and the size of the fund. Even for a fund of large size, a quite large new issue in a large holding only results in limited weight changes, implying limited total transaction costs on the day before the ex-rights day; this in turn implies (most likely) even more limited transaction costs for the trading in the rights and the stocks. One cannot, however, rule out the possibility
that the chosen approach, in at least some cases, underestimates transaction costs, and
overimproves the tracking ability of the fund vis-à-vis a more realistic and complex
scheme. A deeper analysis of optimal index fund trading around new issues is left for
future research.

4.3.10 Data
Data have been collected from various public sources, both electronic and printed. The
data set includes virtually all (around 900) common stocks traded on the Stockholm
Stock Exchange between 1987 and 2000. The data comprise 13 years (3275 days) of
daily closing stock prices, market capitalizations based on all shares outstanding,
traded volumes, and corporate action adjustment factors. The dividend data include
dividend amount, ex-dividend date, and payout date.

For 1990-2000, the data include a record on mergers and acquisitions, conversions,
and spin-offs; the record is intended to be complete with regard to the firms eligible
for inclusion in the OMX. These were collected from the Swedish Tax Agency, the
Stockholm Stock Exchange, the Aktiespararens deklarationsbilaga, among others. The
M&A data contain information on the identities of the buyer (surviving firm) and
seller (non-surviving firm), terms of payment, i.e., cash and/or stock, and how much
that is received per each share transacted, and date of transaction. (Transactions settled
in cash will not affect index calculations, because the non-surviving stock will not be
included in the index.) In total, there are 38 non-cash M&A which are included in the
study (in addition, there are 10 cash and stocks and 129 cash-only transactions). There
are 80 conversions, and each record have a date for the conversion and an identity of
the converted stock (non-surviving) and of the surviving stock. The dates for the
conversions, and the dates and the terms for the M&As have been scrutinized and
checked against actual market capitalization data and other relevant data, and, if
necessary, been changed so as to achieve consistency with the actual data.

Computerized error checking has been performed to detect abnormalities in returns,
dividends, turnovers, market capitalizations etc. Suspected errors have been checked
against other sources including micro-fiche copies of newspaper’s stock market
notations. However, errors might still be present in the data. Unless such remaining
errors are extreme, or lead to extreme consequences, one should not worry too much
about them, because the empirical study is not intended to reflect what would have been exactly the best historical indexing strategy, but rather to provide an idea about the performance of the different approaches under typical and varying conditions that are representative of the market environment in general.

4.4 Empirical analysis
The empirical analysis proceeds as follows. Section 4.4.1 begins with some notes related to the implementation of the empirical analysis. Section 4.4.2 discusses some issues related to the gauging of tracking performance. The relations between the OMX index and the total return index and the price index constructed here are examined in Section 4.4.3.

Section 4.4.4 concerns, in the absence of transaction costs, the empirical tracking performance of full replication under different dividend-payout and reinvestment schemes. Provided in Section 4.4.5 is a brief discussion of Swedish mutual index funds based on the OMX index and their tracking performance. Results regarding the empirical performance of full replication in the presence of transaction costs and with conditional and unconditional updating are presented in Section 4.4.6.

Empirical results are found in Section 4.4.7 for the index fund revision formulation in which a trade-off parameter balances transaction costs and a tracking error measure. The analysis is performed for different values of the trade-off parameter and for a subset constraint.

Section 4.4.8 contains an empirical analysis of the index fund formulation, where the objective is to minimize transaction costs subject to a tracking error measure being kept at or below a limit. The analysis is performed for several different limits and for a subset constraint.

Section 4.4.9 focuses on how tracking performance is affected when the two optimization-based index fund formulations are used without transaction cost control. Section 4.4.10 discusses the efficiency of pre-defined subsets as means to reduce transaction costs.

In Section 4.4.11, the results are presented of an empirical analysis of tracking error minimization based on different tracking error measures and whether different
measures lead to different performance. The different tracking error measures do not take into account transaction costs or dividends. To prevent transaction costs and dividends from affecting the analysis, the tests utilize the developed “portfolio research system” in its capacity as a research laboratory by switching off transaction costs and dividends. First considered are the two measures of tracking error volatility, $SDTE$ and $RMSTE$. Also studied are the performance, in terms of $SDTE$ and $RMSTE$, of index funds constructed using tracking error measures based on returns inferred from the current index composition are compared to the performance of index funds constructed using tracking error measures based on returns inferred from historical index composition.

The implementation based on Satchell and Hwang (2001) is also tested in the absence of transaction costs and dividends. Finally, a formulation featuring a mean absolute tracking error measure constrained to meet a limit is implemented and tested. Its performance is compared to that of an equivalent formulation based on $SDTE$ instead of $MATE$.

### 4.4.1 Implementation notes

**Study period.** The test period begins at the opening of the market of 1 Jan. 1990 and ends at the close of 29 Dec. 1999, which is one day before the last day of 1999. If the last day were used, the index funds would have been updated according to the new $wlist$ effective as of the first trading day of 2000. In the presence of transaction costs or restrictions on the number of stocks to include in the fund, this would introduce unnecessary tracking error. It would be possible to extract the NAV of a given index fund right before the update the last day and use that in the calculations. For simplicity, however, the second last day of 1999 is used instead. The first index values and fund values will emanate from the close of the last date of 1989, 29 Dec., and the last values come from 29 Dec. 1999, which makes a total of 2509 days. The number of daily returns and, consequently, the number of observations in tracking error measure and other calculations that involve returns, will thus be 2508.

**Annualization.** The 13-year period, 1 Jan. 1987 to 30 Dec. 1999, which the data cover, contains 3275 trading days. There is thus, on average, 251.9 ($=3275/13$) trading days per year, and 20.99 such days per month. For the 10-year period 2 Jan. 1990 to 30
Dec. 1999, the corresponding numbers are 250.9 and 20.9, respectively. Based on this, I decided to use, when annualizing or converting measurements based on data of different frequencies to a common basis, the convention of 21 trading days per month and, consequently, 252 days per year.

*Initial portfolio – composition and size.* Each index fund, excepting subset funds, is initially, at the opening of 1 Jan. 1990, an exact replica of the index. For subset index funds, the subset fund with minimum tracking error is used as initial portfolio. For all index funds, the initial NAV is 1000 MSEK. This number roughly agrees with the NAV of the largest, at the time, Swedish index mutual fund. Moreover, funds whose NAVs are of this magnitude are probably large enough to be economically attractive for mutual fund companies, which charge annual fees as a fixed percentage of NAV.

*Transaction costs.* Price impact costs are given by the models estimated in Chapter 2. The commission rate is set to 0.2% for the whole study period. The figure is reasonable and consistent with the average commission reported by Dahlquist et al. (2000) for larger Swedish mutual funds during 1992-1997. The price impact costs and commissions are incorporated in the index fund revision models in the same manner as they were incorporated in the mean-variance revision model in Chapter 3, Section 3.4.1, except that no amortization is done.

It is assumed that the transaction costs are predicted without error, that is, for a given transaction, the same coefficients are used for predicted and incurred costs. In a limit order book market such a presumption is fairly realistic, since one can observe the order book the very moment before a trade is executed. The closing price represents the price at the time the decision to trade was made.

*Cash security.* A security with constant price and, thus, zero, return will represent cash. Numerical problems are avoided in the optimizations if very small, but non-zero, transaction cost coefficients are used for transactions in the cash security.

*Minimum transactions.* If a required weight change is such that it corresponds to trading less than one share of a stock then no trading will be done. Stock holdings whose value net of transaction costs are less than the value of one share are automatically deleted.
**Forecasting and estimation.** The matrix $\mathbf{R}$ contains, for each security in the set $S$, a historical return series consisting of 60 daily returns. These returns are exclusive of dividends. Swedish stocks pay dividends once a year. The forecasted tracking error measures will be based on the in-sample returns $\mathbf{R}$. Increasing in-sample returns with paid dividends, is, however, likely to introduce a bias, because no dividends will be paid in the relatively near future for which the forecast is intended to apply to. A missing return observation for a stock is replaced by the median of the returns of the stocks in the set $S$ that date.

### 4.4.2 Gauging tracking performance

Even though the ultimate goal of indexing may be to have zero tracking error all the time, other goals are used to achieve operational efficiency.

As a first principle, in gauging the performance of a given approach, it is important that performance is measured in terms of the tracking error measure employed, and in view of whether the corresponding objective and constraints were achieved and satisfied, respectively. This verifies the internal consistency of the approach. Next, it can be checked what tracking performance, in a wider sense, the fulfillment of the operationalized objective, led to. Tracking performance in that wider sense relates to, e.g., the final relative deviation between the net asset value of the fund and the index level (scaled appropriately). Also important in this respect are the maximum negative and positive relative deviations of the same quantity during the test period. These measures reflect the compound effects of transaction costs and other effects that cause tracking error. The epitome of tracking performance measures is probably the standard deviation tracking error. This will be recognized here in that the annualized autocorrelation adjusted tracking error standard deviation will be used as a main gauge of tracking performance. The total amount of transaction costs incurred ($\sum TC$) is a convenient gauge of the efficiency of a given approach’s transaction cost control.

As a part of the analysis of the latter stage, statistical tests can be used to complement the evaluation of realized tracking performance. It is important to keep in mind that there is a difference between statistical and economical significance.
For instance, it could be tested whether the average tracking error is different from zero by a paired t-test, where the average tracking error is divided by tracking error standard deviation divided by the square root of the number of observations.

A chi-square test of variance may be used to test whether the realized tracking error variance is larger than a desired level. The chi-square variance test applies under a normal distribution and is very sensitive to deviations from normality. Therefore, a distribution-free Kolmogorov-Smirnov (K-S) test is used to test if a particular tracking error comes from a normal distribution. The test examines the maximum deviation between two distributions. Because the mean and variance of the empirical (tracking error) distribution are estimated, the results of Lilliefors (1967) are used for the critical values. For all tracking error series tested, however, the null hypothesis that the tracking errors are normally distributed could be rejected at very high levels of significance.

Comprehensive statistics will be generated on the tracking performance and transaction costs for each index fund implemented. The statistics are organized under four different headings, Ex-post tracking performance, Updating, Fund transaction costs, and Stock transaction costs. Table 4.1 outlines these statistics and their definitions.

---

65 The K-S test was used out of convenience; there are, according to D’Agostino et al (1990), better tests, such as the ones of D’Agostino and Pearson or Anderson-Darling. The K-S test used is based on a Gauss program by Paul Söderlind.
Table 4.1 Tracking performance and transaction cost statistics

<table>
<thead>
<tr>
<th>Subset</th>
<th>The size of the pre-defined subset (0 indicates that no subset is imposed).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tau</td>
<td>The trade-off between transaction costs and tracking error measure minimization in Opt1</td>
</tr>
<tr>
<td>TELimit</td>
<td>The value of the tracking error measure limit</td>
</tr>
</tbody>
</table>

Panel A. Ex-post tracking performance

| AvgTE (%) | Average daily tracking error (difference between fund and index return) |
| MinTE (%) | Minimum daily tracking error |
| MaxTE (%) | Maximum daily tracking error |
| SDTE (%) | Annualized standard deviation of daily tracking error (using 252 days) |
| αSDTE (%) | Autocorrelation-adjusted annualized standard deviation of daily tracking error (using 252 days) |
| RMSTE (%) | Annualized root mean square of daily tracking error |
| MATE (%) | Annualized mean absolute daily tracking error |
| MinNAV | Minimum ratio of fund net asset value and index level |
| MaxNAV | Maximum ratio of fund net asset value and index level |
| EndNAV | Ending ratio of fund net asset value and index level |
| αSDRp (%) | Autocorrelation-adjusted annualized standard deviation of daily fund return (using 252 days) |
| Alpha (%) | Intercept (times 100) of OLS regression of fund returns on index returns |
| Beta | Slope of OLS regression of fund returns on index returns |
| Sigma (%) | Annualized standard error of OLS regression of fund returns on index returns |
| AvgCash (%) | Average portfolio weight for cash holding |
| AvgTO (%) | Annual average portfolio turnover computed as the average daily turnover (taken as the minimum of the sum of weight increases and the sum of weight decreases each day) times 252 |
| Avgx+ (%) | Average weight increase across all transactions |
| Avgx– (%) | Average weight decrease across all transactions |
| MinN | Minimum number of fund holdings including cash (right after close but before any corporate actions effective next day) |
| AvgN | Average number of fund holdings including cash (-“-”) |
| MaxN | Maximum number of fund holdings including cash (-“-”) |

Panel B. Updating

| AvgFcast (%) | Average forecast tracking error measure at the end of each day but before any trading |
| AvgObj (%) | Average forecast tracking error measure at the end of each day based on desired weights, but before trading and transaction costs. |
| AvgObjTC (%) | Average forecast tracking error measure based on the fund weighting after trading and transaction costs. |
| TECnd | No. of times the tracking error measure update condition has been triggered. |
| DieMCnd | No. of times the fund is being revised because a holding ceases to exist or, in the case of full replication, is not in the index (anymore). |
| SubsetCnd | No. of times the the fund update was induced by violation of the subset constraint. |
| Updates | No. of actual Updates of the fund (trades<1 share in any given stock are not effectuated). |

Panel C. Fund transaction costs

| ΔTC (MSEK) | Total monetary value of transaction costs (not compounded). |
| BuyΔTC | Buy transaction costs as a proportion of ΔTC |
| BuyQΔTC | The quadratic, volume-dependent, part of buy transaction costs as a proportion of BuyΔTC |
| BuyCΔTC | Commissions on buy transactions as a proportion of BuyΔTC |
| SellQΔTC | The quadratic, volume-dependent, part of sell transaction costs as a proportion of SellΔTC |
| SellCΔTC | Commissions on sell transactions as a proportion of SellΔTC |

Panel D. Stock transaction costs

| AvgTC+ (%) | Average buy transaction cost excluding cash transactions |
| MaxTC+ (%) | Maximum total buy transaction cost excluding cash transactions |
| Avgoz+ (%) | Average buy ordersize excepting cash transactions |
| Maxx+ (%) | Maximum buy transaction size in any stock in thousandths of shares outstanding |
| AvgTC– (%) | Average sell transaction cost excluding cash transactions |
| MaxTC– (%) | Maximum sell transaction cost in percentage excluding cash transactions |
| Avgoz– (%) | Average sell ordersize excepting cash transactions |
| Maxx– (%) | Maximum sell transaction size in any stock in thousandths of shares outstanding |
4.4.3 The indexes of the study and the OMX

A price index (PI) and a total return index (RI) are calculated using a consistent methodology, mimicking that according to which the OMX index is calculated. It should thus be interesting to compare the behavior of these indexes to that of the OMX.

Figure 4.1 The OMX index, the return index (RI) and the price index (PI).

In Figure 4.1, the behavior of the three indexes during the 1990s are depicted. Table 4.2 outlines key statistics based on daily returns of the three indexes. Each index is scaled to have a base index value of 1000. The total return index increased around seven times during the decade, while the price index and the OMX increased almost six times. Moreover, according to the figure, the development of the price index and the OMX was fairly similar.
Table 4.2 Descriptive statistics based on daily returns of the price index (PI), the total return index (RI), and the OMX

$SD$ is the annualized standard deviation of the return. $aSD$ is the autocorrelation adjusted return standard deviation. $EndLevel$ is the value of the index at the end of the study period, that is, at the close of 29 Jan. 1999. The number of observations is 2508.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>PI</th>
<th>RI</th>
<th>OMX</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.00080</td>
<td>0.00088</td>
<td>0.00078</td>
</tr>
<tr>
<td>Std</td>
<td>0.01329</td>
<td>0.01329</td>
<td>0.01344</td>
</tr>
<tr>
<td>Stderr</td>
<td>0.00027</td>
<td>0.00027</td>
<td>0.00027</td>
</tr>
<tr>
<td>Min</td>
<td>-0.06709</td>
<td>-0.06709</td>
<td>-0.06653</td>
</tr>
<tr>
<td>25%</td>
<td>-0.00690</td>
<td>-0.00676</td>
<td>-0.00693</td>
</tr>
<tr>
<td>Median</td>
<td>0.00074</td>
<td>0.00084</td>
<td>0.00072</td>
</tr>
<tr>
<td>75%</td>
<td>0.00832</td>
<td>0.00837</td>
<td>0.00822</td>
</tr>
<tr>
<td>Max</td>
<td>0.11544</td>
<td>0.11544</td>
<td>0.11653</td>
</tr>
</tbody>
</table>

$SD$ (%)  21.09   21.10   21.33  
$aSD$ (%) 22.55   22.41   23.32  
EndLevel  5974.75 7201.34 5674.88

Figure 4.2 and Table 4.3, the second column, provide more detailed information about the tracking performance of the price index relative to the OMX.

Figure 4.2 Price index level as a proportion of the OMX
Table 4.3 Tracking of the OMX with the price index and tracking of the price index with the return index
Tracking errors are computed as the daily return difference of PI and OMX and of RI and PI, respectively.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>PI-OMX</th>
<th>RI-PI</th>
</tr>
</thead>
<tbody>
<tr>
<td>AvgTE (%)</td>
<td>0.0019</td>
<td>0.0075</td>
</tr>
<tr>
<td>SDTE (%)</td>
<td>1.91</td>
<td>0.62</td>
</tr>
<tr>
<td>aSDTE (%)</td>
<td>1.53</td>
<td>0.45</td>
</tr>
<tr>
<td>MinNAV</td>
<td>0.993</td>
<td>1.000</td>
</tr>
<tr>
<td>MaxNAV</td>
<td>1.111</td>
<td>1.205</td>
</tr>
<tr>
<td>EndNAV</td>
<td>1.053</td>
<td>1.205</td>
</tr>
<tr>
<td>aSDRp (%)</td>
<td>22.55</td>
<td>22.41</td>
</tr>
<tr>
<td>Alpha (%)</td>
<td>0.0030</td>
<td>0.0075</td>
</tr>
<tr>
<td>Beta</td>
<td>0.985</td>
<td>1.000</td>
</tr>
<tr>
<td>Sigma (%)</td>
<td>1.88</td>
<td>0.62</td>
</tr>
</tbody>
</table>

In the second half of 1999, the steeply increasing market value of Ericsson pushes its weight in the (uncapped) OMX above 40%. In the price index, this effect was dampened by the 20% weight cap, and therefore the price index lost ground to the OMX, as indicated by the curve in Figure 4.2. (The maximum weight of Ericsson in the (capped) price index is 0.307.) During the period of the study, the maximum (minimum) ratio between the price index and the OMX is 1.111 (0.993). At the end of the decade, the price index is 5.3% above the OMX. The average daily return difference $\text{AvgTE}$ is 0.0019% and the annualized standard deviation of the daily return difference, $\text{SDTE}$, equals 1.91%. (A paired t-test does not reject at conventional levels of significance the null hypothesis of zero tracking error.) By and large, it seems fair to conclude that the price index is representative of the OMX.

The so-called risk-adjusted performance of the index fund in terms of estimates of alpha and beta of a regression of the price index returns on the OMX returns are computed. $\text{Alpha}$ is 0.003%, which is quite close to $\text{AvgTE}$, and $\text{Beta}$ is close to one. The annualized standard deviation of the regression residuals, $\text{Sigma}$, equals 1.88%, which thus, as expected, is close to $\text{SDTE}$.

The total return index differs from the price index in that dividends are reinvested, effective at the beginning of the ex-dividends date. Figure 4.3 displays the total return index divided by the price index as well as the tracking error.
Figure 4.3 shows that the greater part of dividends are paid out in the second quarter and that they give rise to a yield (the dividend amount relative to the sum market value of the index constituents) of around 2%, which when reinvested and compounded over 10 years results in the total return index being around 20% greater than the price index.

Table 4.3, the third column, gives the details about tracking error for the total return index vis-à-vis the price index. The average tracking error is statistically significant based on a t-test.

4.4.4 Full replication without transaction costs: effects of dividends

First, full replication is applied to track the price index without transaction costs. Second, the effect on tracking performance of delayed dividend payments in a setting without transaction costs is examined. In one scheme, dividends are assumed received at the end of the ex-dividend date, and in the other at the close 8 trading days after the day before the ex-dividend date. The latter scheme corresponds to the average actual payout time, while the former scheme should be achievable for a bank or institution.
### Table 4.4 Full replication with no dividends, dividends paid out at close after 1 and 8 days, respectively

No transaction costs.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>No dividends</th>
<th>Delay of dividends (in days)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Panel A. Ex-post tracking performance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$AvgTE$ (%)</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>$SDTE$ (%)</td>
<td>0.00</td>
<td>0.02</td>
</tr>
<tr>
<td>$aSDTE$ (%)</td>
<td>0.00</td>
<td>0.02</td>
</tr>
<tr>
<td>$MinNAV$</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$MaxNAV$</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$EndNAV$</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$aSDRp$ (%)</td>
<td>22.55</td>
<td>22.40</td>
</tr>
<tr>
<td>$Alpha$ (%)</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>$Beta$</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$Sigma$ (%)</td>
<td>0.00</td>
<td>0.02</td>
</tr>
<tr>
<td>$AvgCash$ (%)</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>$AvgTO$ (%)</td>
<td>19.3</td>
<td>21.0</td>
</tr>
<tr>
<td>$Avgx^+$ (%)</td>
<td>0.45</td>
<td>0.23</td>
</tr>
<tr>
<td>$Avgx^-$ (%)</td>
<td>0.39</td>
<td>0.24</td>
</tr>
<tr>
<td>$MinN$</td>
<td>28</td>
<td>28</td>
</tr>
<tr>
<td>$AvgN$</td>
<td>29.6</td>
<td>29.6</td>
</tr>
<tr>
<td>$MaxN$</td>
<td>31</td>
<td>31</td>
</tr>
<tr>
<td>Panel B. Updating</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$AvgFcast$ (%)</td>
<td>0.017</td>
<td>0.019</td>
</tr>
<tr>
<td>$AvgObj$ (%)</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>$TECnd$</td>
<td>2508</td>
<td>2508</td>
</tr>
<tr>
<td>$DieMCnd$</td>
<td>26</td>
<td>185</td>
</tr>
<tr>
<td>$Updates$</td>
<td>146</td>
<td>287</td>
</tr>
</tbody>
</table>

The figures in the first column, Panel A, Table 4.4, show that without dividends and transaction costs, full replication (exactly) replicates the price index. In full replication funds, cash is not allowed, which is reflected by $AvgCash$ being zero. The average annual portfolio turnover, $AvgTO$, defined as the average daily turnover (taken as the minimum of the sum of weight increases and the sum of weight decreases each day) times 252, is increased due to the dividends. $MinN$, $AvgN$, and $MaxN$ describes (right after the close but before any corporate actions effective the next day) the minimum, mean, and maximum number of holdings in the fund, including cash. The index never contains more than 30 stocks, while here the fund holds at some date 31. This peculiarity is because the number of stocks is measured before any corporate actions effective the next day, and here the 31st stock is converted to another line (already in the index), thereby reducing the number of stocks to 30, effective the next
day. The reason that the index and the funds contain fewer holdings than 30 is because of delistings.

In Panel B, $TEC_{nd}$ equals 2508 and this is the number of times the tracking error measure update condition is triggered. For full replication with zero tracking error, any tiny weight deviation from the index triggers an update. Because no tracking error is accepted, the condition is triggered 2508 times or *every* date. However, in this case this often leads to no actual update of the fund, as the implied changes will be miniscule, i.e., less than one share would have to be traded. $DieMC_{nd}$ is the number of times the fund is being revised because a holding ceases to exist or, in the case of full replication, is not in the index. So when dividends are included, this variable reflects fund revisions induced by changes in the $w_{list}$ as well as by cash receipts from dividend payments. $Updates$ is the number of times the fund is actually updated. $AvgFcast$, the mean of the forecasted tracking error measure at the end of each date, but before any trading, is 0.02% which is small but non-zero. $AvgObj$ is the mean of the tracking error measure/objective computed using updated or desired weights, but before trading and transaction costs. $AvgObj$ shows that the tracking error measure objective is effectively reduced to zero after updating.

Columns 2 and 3 of Table 4.4 depict the tracking performance of full replication without transaction costs for dividend payments delayed 1 and 8 trading days, respectively, relative to the dividend reinvestment in the index. The 1-day delay results in virtually no tracking error gauged by, e.g., $aSDTE$, which equals 0.02%. The 8-day delay results in an annualized daily tracking error standard deviation ($SDTE$) of 0.8%, which is quite large considering, for instance, the annualized tracking error standard deviations of 0.18-0.72% for the S&P 500 index funds analyzed in Frino and Gallagher (2001). Note, however, that the autocorrelation tracking error standard deviation, $aSDTE$, for the 8-day delay is a substantially lower 0.14%. Recall that the tracking errors of these S&P 500 index funds are achieved in the presence of dividends, trading costs, and investor flows. In view of such low tracking errors it appears likely that S&P 500 index funds somehow reduce the tracking errors caused by dividend payment delays, possibly by techniques similar to those discussed earlier.
In the tests that follow, when dividends are included, the 1-day delay convention is used.

4.4.5 Tracking performance of Swedish index mutual funds

It would be interesting to compare the performances of the index funds implemented here to those of actual Swedish index funds. Today, there are several mutual index funds that track the OMX price index. Direct comparisons are, however, difficult to make. One reason is that the OMX index and the indexes calculated here are based on similar, but not identical, methodologies. Another reason is that the performance of the mutual funds are affected by other costs than transaction costs, such as administrative and marketing costs, which the funds constructed here are not affected by. Moreover, none of the index mutual funds based on the OMX have been in existence for the full period of this study. It is even difficult to conduct a relevant tracking performance evaluation of the index mutual funds, as there exists no total return version of the OMX index, which would be the appropriate benchmark considering the fact that the funds receive dividends.

Despite the said obstacles, a brief analysis is attempted of the tracking performance of the first OMX index mutual fund, which was launched on 24 June 1996 by Erik Penser Fonder AB. Its $SDTE$ and $AvgTE$ relative to the OMX from inception to the end of 1999 are 1.94% and 0.0037%, respectively. The positive $AvgTE$ is partly explained by the fact that the OMX index is a price index, whereas the Penser fund benefits from dividends. In an effort to adjust the OMX index for dividends, an approximative dividend component, derived from the total return index and the price index calculated here, was applied to the OMX index. Relative to this approximate total return version of the OMX index, the Penser index fund exhibits a $SDTE$ of 1.82% and an $AvgTE$ of -0.0024%. The tracking error standard deviation of the Penser index fund appears large compared to those of the U.S. and Australian index funds analyzed in Frino and Gallagher (2001) and Frino and Gallagher (2002), respectively. The tracking error standard deviations of the Australian funds are below 1%.
4.4.6 **Full replication under conditional updating with transaction costs and dividends**

A straightforward extension of full replication is to revise the index fund using the full replication method only when the projected tracking error measure exceeds a limit. This would be appropriate for a fund whose objective is to keep tracking error below a certain limit. A potential problem with the full replication method given such an objective is that it will (generally) overdo the revisions, i.e., the fund could have met its tracking error target without turning itself into a fully replicating fund at each update, thereby incurring less transaction costs. On the other hand, this may lead to fewer updates than if the fund is subjected to revisions just large enough to meet the tracking error limit. In any case, there is no explicit transaction cost control.

Table 4.5 presents the results for full replication with conditional updating based on tracking error standard deviation in the presence of transaction cost and dividends with a one-day delay.

It is satisfying that the realized tracking error standard deviation, \(SDTE\), increases monotonically with \(TELimit\). Between the autocorrelation adjusted tracking error standard deviations, \(aSDTE\), and the \(TELimit\) there is, however, no monotonic relation. Actually, both \(aSDTE\) and \(SDTE\) for the index funds with the three lowest tracking error limits violate the limits. The ratios between the ending fund net asset values and the index, \(EndNAV\), varies between 0.904 and 0.912. The corresponding NAV’s are, thus, 6510 MSEK and 6568 MSEK, respectively, so in monetary terms, the difference is 58 MSEK.

\(Alpha\) is close to \(AvgTE\) and \(Beta\) is near one for all the index funds. \(Sigma\) equals \(SDTE\).
Table 4.5 Full replication under conditional updating with transaction costs and dividends

<table>
<thead>
<tr>
<th>Statistic</th>
<th>0</th>
<th>0.2</th>
<th>0.5</th>
<th>1</th>
<th>2</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A. Ex-post tracking performance</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AvgTE (%)</td>
<td>-0.0037</td>
<td>-0.0037</td>
<td>-0.0037</td>
<td>-0.0041</td>
<td>-0.0040</td>
<td>-0.0040</td>
</tr>
<tr>
<td>SDTE (%)</td>
<td>0.60</td>
<td>0.61</td>
<td>0.63</td>
<td>0.67</td>
<td>0.69</td>
<td>0.75</td>
</tr>
<tr>
<td>aSDTE (%)</td>
<td>0.61</td>
<td>0.60</td>
<td>0.56</td>
<td>0.62</td>
<td>0.62</td>
<td>0.61</td>
</tr>
<tr>
<td>MinNAV</td>
<td>0.910</td>
<td>0.912</td>
<td>0.911</td>
<td>0.903</td>
<td>0.904</td>
<td>0.905</td>
</tr>
<tr>
<td>MaxNAV</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>EndNAV</td>
<td>0.910</td>
<td>0.912</td>
<td>0.912</td>
<td>0.904</td>
<td>0.905</td>
<td>0.905</td>
</tr>
<tr>
<td>aSDRp (%)</td>
<td>22.51</td>
<td>22.50</td>
<td>22.47</td>
<td>22.41</td>
<td>22.40</td>
<td>22.40</td>
</tr>
<tr>
<td>Alpha (%)</td>
<td>-0.0037</td>
<td>-0.0036</td>
<td>-0.0036</td>
<td>-0.0039</td>
<td>-0.0039</td>
<td>-0.0039</td>
</tr>
<tr>
<td>Beta</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.999</td>
<td>0.998</td>
<td>0.998</td>
</tr>
<tr>
<td>Sigma (%)</td>
<td>0.60</td>
<td>0.61</td>
<td>0.63</td>
<td>0.67</td>
<td>0.69</td>
<td>0.75</td>
</tr>
<tr>
<td>AvgCash (%)</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>AvgTO (%)</td>
<td>21.1</td>
<td>20.1</td>
<td>19.7</td>
<td>19.5</td>
<td>18.7</td>
<td>18.7</td>
</tr>
<tr>
<td>Avgx+ (%)</td>
<td>0.13</td>
<td>0.23</td>
<td>0.24</td>
<td>0.24</td>
<td>0.21</td>
<td>0.21</td>
</tr>
<tr>
<td>Avgx- (%)</td>
<td>0.15</td>
<td>0.31</td>
<td>0.31</td>
<td>0.31</td>
<td>0.30</td>
<td>0.30</td>
</tr>
<tr>
<td>MinN</td>
<td>28</td>
<td>28</td>
<td>28</td>
<td>28</td>
<td>28</td>
<td>28</td>
</tr>
<tr>
<td>AvgN</td>
<td>29.6</td>
<td>29.6</td>
<td>29.6</td>
<td>29.6</td>
<td>29.5</td>
<td>29.5</td>
</tr>
<tr>
<td>MaxN</td>
<td>31</td>
<td>31</td>
<td>31</td>
<td>31</td>
<td>31</td>
<td>31</td>
</tr>
<tr>
<td><strong>Panel B. Updating</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AvgFcast (%)</td>
<td>0.019</td>
<td>0.060</td>
<td>0.095</td>
<td>0.136</td>
<td>0.154</td>
<td>0.164</td>
</tr>
<tr>
<td>AvgObj (%)</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>AvgObjTC (%)</td>
<td>0.003</td>
<td>0.007</td>
<td>0.007</td>
<td>0.007</td>
<td>0.007</td>
<td>0.007</td>
</tr>
<tr>
<td>TEcnd</td>
<td>2508</td>
<td>45</td>
<td>30</td>
<td>18</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>DieMCnd</td>
<td>185</td>
<td>185</td>
<td>185</td>
<td>185</td>
<td>185</td>
<td>185</td>
</tr>
<tr>
<td>Updates</td>
<td>480</td>
<td>205</td>
<td>193</td>
<td>189</td>
<td>186</td>
<td>185</td>
</tr>
<tr>
<td><strong>Panel C. Fund transaction costs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \Sigma TC ) (MSEK)</td>
<td>191</td>
<td>188</td>
<td>186</td>
<td>184</td>
<td>167</td>
<td>154</td>
</tr>
<tr>
<td>Buy( \Sigma TC )</td>
<td>0.59</td>
<td>0.59</td>
<td>0.59</td>
<td>0.59</td>
<td>0.60</td>
<td>0.56</td>
</tr>
<tr>
<td>BuyQ( \Sigma TC )</td>
<td>0.83</td>
<td>0.83</td>
<td>0.83</td>
<td>0.83</td>
<td>0.83</td>
<td>0.81</td>
</tr>
<tr>
<td>BuyC( \Sigma TC )</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>SellQ( \Sigma TC )</td>
<td>0.84</td>
<td>0.85</td>
<td>0.85</td>
<td>0.85</td>
<td>0.84</td>
<td>0.83</td>
</tr>
<tr>
<td>SellC( \Sigma TC )</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td><strong>Panel D. Stock transaction costs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AvgTC+ (%)</td>
<td>1.02</td>
<td>1.06</td>
<td>1.06</td>
<td>1.04</td>
<td>1.02</td>
<td>1.01</td>
</tr>
<tr>
<td>MaxTC+ (%)</td>
<td>7.34</td>
<td>7.34</td>
<td>7.36</td>
<td>7.36</td>
<td>7.36</td>
<td>7.36</td>
</tr>
<tr>
<td>Avgoz+(‰)</td>
<td>0.14</td>
<td>0.21</td>
<td>0.21</td>
<td>0.19</td>
<td>0.18</td>
<td>0.17</td>
</tr>
<tr>
<td>Maxx+ (‰)</td>
<td>4.91</td>
<td>4.92</td>
<td>4.93</td>
<td>4.93</td>
<td>4.93</td>
<td>4.93</td>
</tr>
<tr>
<td>AvgTC- (%)</td>
<td>0.89</td>
<td>1.44</td>
<td>1.46</td>
<td>1.44</td>
<td>1.47</td>
<td>1.49</td>
</tr>
<tr>
<td>MaxTC- (%)</td>
<td>6.09</td>
<td>6.12</td>
<td>6.06</td>
<td>6.06</td>
<td>6.06</td>
<td>6.06</td>
</tr>
<tr>
<td>Avgoz- (‰)</td>
<td>0.16</td>
<td>0.63</td>
<td>0.63</td>
<td>0.64</td>
<td>0.66</td>
<td>0.67</td>
</tr>
</tbody>
</table>

Not unexpectedly, the number of fund updates, \( \text{Updates} \), decreases monotonically with \( \text{TElimit} \). \( \text{AvgObj} \), the tracking error measure based on updated or desired weights,
but before trading and transaction costs, is zero, whereas $\text{AvgObjTC}$, the average forecast tracking error measure based on the fund weighting after trading and transaction costs, is small and positive. The reason is that transaction costs cause the achieved weights to deviate from the desired weights, but the effect is minor.

The total monetary amount (not compounded) lost in transaction costs decreases with $T\text{E}_{\text{limit}}$, from 191 to 154 MSEK. The sum of monetary buy transaction costs incurred as a proportion of $\Sigma TC$, $Buy\Sigma TC$, is just below 60%, and is thus greater than sell transaction costs which constitute slightly more than 40% of $\Sigma TC$. $BuyQ\Sigma TC$, the monetary transaction costs from the quadratic, volume-dependent part of the transaction cost specification as fraction of $Buy\Sigma TC$, is above 80%. Commissions on buy transactions as a proportion of $Buy\Sigma TC$, $BuyC\Sigma TC$, are 0.05. $SellQ\Sigma TC$ is for all funds greater than 80%. Commissions on sales make up 5% of $Sell\Sigma TC$.

Panel D of Table 4.5 provides evidence on the transaction costs incurred for individual stocks. The average total percentage buy transaction cost, $AvgTC^+$, is in the range 1.01-1.06%. The maximum percentage total buy transaction cost, $MaxTC^+$, is around 7.35%.

The maximum weight increase for any stock expressed in thousandths of the number of outstanding shares, $Maxx^+$, is near 4.9 thousandths. The transaction cost model was estimated for order sizes in the range 0 to 3 thousandths of the number shares outstanding. For buys, this then means that the transaction cost model is extrapolated outside the domain for which it was estimated.

Across the full replication index strategies, the mean percentage transaction cost for sales, $AvgTC^-$, decreases with target tracking error standard deviation and ranges from 0.89% to 1.49% and are thus larger than for buy transactions. This is natural as the average sell order sizes, $Avgoz^-$, are larger. The maximum transaction cost $MaxTC^-$ varies from 6.06% to 6.12%. The maximum weight decrease for any stock expressed in thousandths of the number of outstanding shares, $Maxx^-$, varies from 3.3 to 3.6 thousandths.

As expected, conditional updating can be combined with full replication to reduce the number of times a fund is updated. The funds implemented with larger $T\text{E}_{\text{limits}}$
are updated fewer times, and incur lower total transaction costs. The overall tracking performance of the different funds is, however, quite similar.

4.4.7 Minimizing a tracking error measure and transaction costs using a trade-off parameter (Opt1)

This approach involves a trade-off parameter between transaction costs and the tracking error measure. The trade-off parameter is in a way similar to, but must not be confused with the risk aversion parameter of the mean-variance portfolio revision model. The lack of a priori knowledge about what is a reasonable value of the parameter is an issue that is addressed by testing a wide range of values for the parameter. Since the parameter is multiplied with the tracking error measure term in the objective, sufficiently large values of the parameter implies that all emphasis will be on tracking error minimization. In effect, this turns the approach into full replication if the full set of index stocks is available for selection; however, if, e.g., cash is allowed in the fund, full replication and this approach do not necessarily coincide. Conversely, small enough values of the parameter put all emphasis on minimizing transaction cost.

After some experimentation, five different $\tau$ values were selected, ranging from 1 to $10^7$, where the emphasis on tracking error measure minimization increases with $\tau$. The approach is also varied in that index funds are created from a pre-defined subset of 20 stocks. This gives a total of 10 different index funds for which results are presented in Table 4.6.

The index funds constructed without a subset restriction, exhibit for $\tau$ values of 1 and 100 $aSDTE$s of 3.5% and 2.3%, respectively; and such large tracking errors are probably not acceptable for ordinary index funds. As the emphasis on tracking error minimization is low, or inversely, high on transaction cost minimization, total monetary transaction costs $\Sigma TC$ are small, amounting to 12 and 23 MSEK, respectively. The fund with the smallest $\tau$ never bought a stock. Because of the large tracking error deviations for these two funds, there is a comparatively sizeable spread between their respective $MinNAV$ and $MaxNAV$. A high tracking error standard deviation, indicates a greater likelihood of outperforming the index, but also of
underperformance, which is evidenced by the fund obtained with \( \text{Tau}=1 \), whose \( \text{EndNAV} \) is 0.78.

For the funds with \( \text{Tau} \) ranging from 1000 to \( 10^7 \), most of the different realized tracking error measures are below 1%, and probably low enough to be considered as successful indexing. There seems to be a minimum of sorts with regard to tracking error for \( \text{Tau} \) equal to 10000, in that smaller and larger \( \text{Tau} \)s result in funds with higher \( \text{aSDTE} \). Despite that tracking error minimization is given greater weight for the fund with a \( \text{Tau} \) of \( 10^7 \), tracking error for that fund is increased relative to the fund whose \( \text{Tau} \) is \( 10^4 \), and the reason appears to be transaction costs. The monetary transaction costs, \( \Sigma TC \), increase monotonically with \( \text{Tau} \).

As expected, the fund with a \( \text{Tau} \) of \( 10^7 \) performs very similar to the full replication funds in Table 4.5, but, in addition to having a different maximum number of holdings \( \text{MaxN} \) than the full replication funds, the \( \text{Tau} \) \( 10^7 \) fund has a non-zero average cash position, as indicated by \( \text{AvgCash} \).

For funds constructed with \( \text{Tau} \) ranging from 100 to \( 10^7 \), the ending NAV to ending index value, \( \text{EndNAV} \), decreases monotonically with \( \text{Tau} \). In particular, the funds with \( \text{Tau} \) from 1000 to \( 10^7 \), illustrate the benefits of controlling transaction costs. These four funds have autocorrelation adjusted tracking error standard deviations \( \text{aSDTE} \) in the range 0.45-0.72%. These \( \text{aSTDEs} \) are comparable to the \( \text{aSTDEs} \) for full replication, presented in Table 4.5, of 0.56-0.62%. However, the ending NAVs to ending index value, \( \text{EndNAV} \), are greater than, or equal to, the ones achieved by the full replication funds. If the fund with a \( \text{Tau} \) of \( 10^7 \) fund is excluded, then the difference in terms of \( \text{EndNAV} \) is at least 1 percentage point, which, considering that the ending index value is approximately 7200 MSEK, amounts to 72 MSEK. For the \( \text{Tau} \) 1000 fund with a \( \text{aSDTE} \) of 0.69%, the \( \text{EndNAV} \) is 0.954 and the difference 4.2 percentage points, or 302 MSEK, to the full replication fund with the highest \( \text{EndNAV} \) ratio.
Table 4.6 **Opt1**: Minimizing a tracking error measure and transaction costs using a trade-off parameter under transaction costs and dividends

The tracking error measure is \( SDTE \).

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Subset/TElimit</th>
<th>0/1</th>
<th>0/100</th>
<th>0/10^2</th>
<th>0/10^4</th>
<th>0/10^6</th>
<th>20/1</th>
<th>20/100</th>
<th>20/10^3</th>
<th>20/10^4</th>
<th>20/10^6</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A. Ex-post tracking performance</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AvgTE (%)</td>
<td></td>
<td>-0.011</td>
<td>-0.002</td>
<td>-0.002</td>
<td>-0.003</td>
<td>-0.004</td>
<td>0.006</td>
<td>0.000</td>
<td>-0.008</td>
<td>-0.021</td>
<td>-0.065</td>
</tr>
<tr>
<td>SDTE (%)</td>
<td></td>
<td>4.44</td>
<td>3.24</td>
<td>0.96</td>
<td>0.41</td>
<td>0.59</td>
<td>7.23</td>
<td>3.77</td>
<td>1.77</td>
<td>1.74</td>
<td>3.02</td>
</tr>
<tr>
<td>aSDTE (%)</td>
<td></td>
<td>3.48</td>
<td>2.27</td>
<td>0.72</td>
<td>0.45</td>
<td>0.61</td>
<td>5.09</td>
<td>2.14</td>
<td>2.36</td>
<td>4.00</td>
<td>12.74</td>
</tr>
<tr>
<td>MinNAV</td>
<td></td>
<td>0.746</td>
<td>0.941</td>
<td>0.953</td>
<td>0.932</td>
<td>0.912</td>
<td>0.914</td>
<td>0.933</td>
<td>0.821</td>
<td>0.586</td>
<td>0.196</td>
</tr>
<tr>
<td>MaxNAV</td>
<td></td>
<td>1.100</td>
<td>1.079</td>
<td>1.010</td>
<td>1.000</td>
<td>1.000</td>
<td>1.107</td>
<td>1.025</td>
<td>1.004</td>
<td>1.003</td>
<td>1.000</td>
</tr>
<tr>
<td>EndNAV</td>
<td></td>
<td>0.782</td>
<td>0.958</td>
<td>0.954</td>
<td>0.932</td>
<td>0.912</td>
<td>1.093</td>
<td>0.978</td>
<td>0.823</td>
<td>0.587</td>
<td>0.196</td>
</tr>
<tr>
<td>aSDRp (%)</td>
<td></td>
<td>20.36</td>
<td>21.44</td>
<td>22.13</td>
<td>22.47</td>
<td>22.48</td>
<td>23.77</td>
<td>22.59</td>
<td>22.81</td>
<td>23.80</td>
<td>27.92</td>
</tr>
<tr>
<td>Alpha (%)</td>
<td></td>
<td>-0.002</td>
<td>0.002</td>
<td>-0.001</td>
<td>-0.003</td>
<td>-0.004</td>
<td>-0.002</td>
<td>-0.003</td>
<td>-0.008</td>
<td>-0.021</td>
<td>-0.065</td>
</tr>
<tr>
<td>Beta</td>
<td></td>
<td>0.897</td>
<td>0.945</td>
<td>0.993</td>
<td>0.999</td>
<td>1.000</td>
<td>1.088</td>
<td>1.037</td>
<td>1.006</td>
<td>1.002</td>
<td>1.001</td>
</tr>
<tr>
<td>Sigma (%)</td>
<td></td>
<td>3.87</td>
<td>3.03</td>
<td>0.95</td>
<td>0.41</td>
<td>0.59</td>
<td>6.99</td>
<td>3.69</td>
<td>1.76</td>
<td>1.74</td>
<td>3.02</td>
</tr>
<tr>
<td>AvgCash (%)</td>
<td></td>
<td>13.93</td>
<td>8.72</td>
<td>0.60</td>
<td>0.13</td>
<td>0.08</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AvgTO (%)</td>
<td></td>
<td>2.2</td>
<td>8.9</td>
<td>16.4</td>
<td>20.4</td>
<td>21.2</td>
<td>6.6</td>
<td>13.2</td>
<td>40.4</td>
<td>151.6</td>
<td>588.6</td>
</tr>
<tr>
<td>MinN</td>
<td></td>
<td>17</td>
<td>24</td>
<td>30</td>
<td>29</td>
<td>28</td>
<td>12</td>
<td>16</td>
<td>18</td>
<td>19</td>
<td>16</td>
</tr>
<tr>
<td>AvgN</td>
<td></td>
<td>22.2</td>
<td>29.7</td>
<td>36.2</td>
<td>33.0</td>
<td>30.2</td>
<td>15.5</td>
<td>18.4</td>
<td>19.8</td>
<td>20.0</td>
<td>19.8</td>
</tr>
<tr>
<td>MaxN</td>
<td></td>
<td>31</td>
<td>36</td>
<td>42</td>
<td>36</td>
<td>33</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
</tr>
</tbody>
</table>

All the index funds, but the two funds with the largest \( aSDTE \), have Betas near one, Sigmas near \( SDTE \), and Alphas near AvgTE. AvgCash is positively related to the
emphasis on transaction cost minimization, indicating that a positive cash balance is potentially helpful in minimizing transaction costs.

\[ \text{Buy} \Sigma \text{TC} \] increases from zero to 0.59 as the emphasis on transaction cost minimization decreases. This is reasonable, because when the emphasis on minimizing transaction costs is high, only forced trades are executed, and these are never purchases. For forced trades, induced by, e.g., delistings, transaction costs are not controlled. This explains the high \( \text{AvgTC}^- \) of 4.81%. When the emphasis is shifted, however, sales become more common, and the costs for them as a proportion of total transaction costs increase. For both purchases and sales, the quadratic parts, \( \text{BuyQ} \Sigma \text{TC} \) and \( \text{SellQ} \Sigma \text{TC} \), which allow for explicit consideration of price impact costs by letting the percentage costs be volume-dependent, again contribute more to monetary transaction costs than do the linear parts, which represent constant percentage costs. For both buys and sells, commissions are the least important part of the monetary transaction costs.

For buy transactions, the average and maximum percentage transaction costs in individual stocks, \( \text{AvgTC}^+ \) and \( \text{MaxTC}^+ \), respectively, appear reasonable. Moreover, it is only for the fund with a \( \text{Tau} \) of \( 10^7 \) that the transaction cost model is extrapolated.

For sales, the average percentage transaction costs in individual stocks, \( \text{AvgTC}^- \), also seem reasonable, except for the fund discussed above with a \( \text{Tau} \) of 1. For the fund constructed using \( \text{Tau}=100 \) there are very high values on \( \text{MaxTC}^- \) and \( \text{Maxx}^- \). These values refers to BGB, a real estate firm that eventually went bankrupt. Similar observations can be made in Table 4.7 and in Table 4.8. The monetary amounts corresponding to the large percentage transaction costs in the stock are, however, small.\(^{66}\)

The subset index funds show high \( aSDTEs \). The two funds with the smallest \( \text{Tau} \) do, however, exhibit a favourable performance in terms of NAV, whereas the tracking

\(^{66}\) The large values regarding individual stocks’ transaction costs in Table 4.7 for \( TElimit=2\% \), and in Table 4.8 for \( TElimit=5\% \) also refers to the stock of BGB. At the close before the delisting, the total market cap for BGB is 0.288 MSEK and the stock price 0.01 SEK. In the most extreme instance, Table 4.8 for \( TElimit=5\% \), the index fund is forced to sell, due to the delisting, its holding of 74.89\% of the BGB’s market cap, or 0.0216 MSEK, which was bought just prior, probably because of the extreme return pattern of BGB. The sale results in a transaction cost of 81.67\%, or 0.0176 MSEK.
performance of the other subset funds are weak. Moreover, there is no evidence that the use of the pre-defined subset decreases transaction costs.

### 4.4.8 Minimizing transaction costs while keeping a tracking error measure less than a pre-specified limit (Opt2)

In this approach, whenever the forecasted tracking error measure is above a desired limit, the fund is revised so as to satisfy the tracking error limit while minimizing transaction costs. For the root mean square and standard deviation tracking error measures, this amounts to solving a QCQP. The approach is straightforward and easy to understand, though solving a relatively sizeable QCQP is computationally demanding. However, if it is implemented with a subset constraint, it is not certain that the $TE_{\text{limit}}$ is attainable, that is, the possible minimum tracking error achievable by the fund might not be low enough to meet the constraint. When this happens (as a simple workaround), the tracking error measure is minimized using quadratic programming without regard to transaction costs. (A possible improvement is to use a limit somewhat narrower than the desired tracking error. This may limit the number of revisions and the transaction costs incurred.)

Index funds are constructed for five different limits of tracking error standard deviation, namely 0.2, 0.5, 1, 2, and 5%. The approach is also implemented using a subset of 20. Hence, 10 index funds are implemented, and the results are presented in Table 4.7. It is probably reasonable to consider the first three limits as typical or acceptable levels of tracking error for standard index funds, whereas the two largest values are not. The desired/accepted level of tracking error for an index fund is, however, a matter of subjective preference, so the inclusion of the larger values may be viewed as an attempt to offer additional coverage in that respect.

Without the subset restriction, this approach performs well in terms of realized tracking error standard deviations, ending fund values relative to ending index value, and transaction costs incurred.

The realized $SDTE$ increases monotonically with the limit, and violate moderately the three lowest limits. The autocorrelation adjusted tracking error standard deviations are lower than the unadjusted ones and it is only the 0.2% limit that is violated in terms of this measure.
Table 4.7 Opt2: Minimizing transaction costs while keeping a tracking error measure less than a pre-specified limit

*TELIMIT* is based on \(SDTE\).

### Panel A. Ex-post tracking performance

<table>
<thead>
<tr>
<th>Statistic</th>
<th>0/0.2</th>
<th>0/0.5</th>
<th>0/1</th>
<th>0/2</th>
<th>0/5</th>
<th>20/0.2</th>
<th>20/0.5</th>
<th>20/1</th>
<th>20/2</th>
<th>20/5</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\text{AvgTE} (%))</td>
<td>-0.003</td>
<td>-0.002</td>
<td>-0.003</td>
<td>-0.004</td>
<td>0.001</td>
<td>-0.065</td>
<td>-0.064</td>
<td>-0.036</td>
<td>-0.008</td>
<td>0.001</td>
</tr>
<tr>
<td>(\text{SDTE} (%))</td>
<td>0.50</td>
<td>0.64</td>
<td>1.04</td>
<td>1.84</td>
<td>3.84</td>
<td>3.02</td>
<td>3.02</td>
<td>3.11</td>
<td>2.16</td>
<td>3.82</td>
</tr>
<tr>
<td>(\text{aSDTE} (%))</td>
<td>0.47</td>
<td>0.48</td>
<td>0.68</td>
<td>1.36</td>
<td>2.53</td>
<td>12.81</td>
<td>12.85</td>
<td>13.27</td>
<td>2.42</td>
<td>2.58</td>
</tr>
<tr>
<td>(\text{MinNAV})</td>
<td>0.925</td>
<td>0.938</td>
<td>0.937</td>
<td>0.920</td>
<td>0.971</td>
<td>0.195</td>
<td>0.199</td>
<td>0.402</td>
<td>0.816</td>
<td>0.942</td>
</tr>
<tr>
<td>(\text{MaxNAV})</td>
<td>1.000</td>
<td>1.002</td>
<td>1.008</td>
<td>1.031</td>
<td>1.080</td>
<td>1.000</td>
<td>1.000</td>
<td>1.004</td>
<td>1.004</td>
<td>1.044</td>
</tr>
<tr>
<td>(\text{EndNAV})</td>
<td>0.927</td>
<td>0.941</td>
<td>0.938</td>
<td>0.924</td>
<td>0.971</td>
<td>0.195</td>
<td>0.199</td>
<td>0.404</td>
<td>0.824</td>
<td>0.989</td>
</tr>
</tbody>
</table>

### Panel B. Updating

<table>
<thead>
<tr>
<th>Statistic</th>
<th>0/0.2</th>
<th>0/0.5</th>
<th>0/1</th>
<th>0/2</th>
<th>0/5</th>
<th>20/0.2</th>
<th>20/0.5</th>
<th>20/1</th>
<th>20/2</th>
<th>20/5</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\text{AvgFcast} (%))</td>
<td>0.206</td>
<td>0.472</td>
<td>0.890</td>
<td>1.659</td>
<td>3.364</td>
<td>0.832</td>
<td>0.834</td>
<td>1.029</td>
<td>1.712</td>
<td>3.577</td>
</tr>
<tr>
<td>(\text{AvgObj} (%))</td>
<td>0.200</td>
<td>0.500</td>
<td>0.999</td>
<td>1.993</td>
<td>4.799</td>
<td>0.811</td>
<td>0.823</td>
<td>1.082</td>
<td>1.927</td>
<td>4.448</td>
</tr>
<tr>
<td>(\text{AvgObjTC} (%))</td>
<td>0.201</td>
<td>0.500</td>
<td>1.000</td>
<td>1.994</td>
<td>4.799</td>
<td>0.811</td>
<td>0.823</td>
<td>1.083</td>
<td>1.929</td>
<td>4.451</td>
</tr>
<tr>
<td>(\text{TECnd})</td>
<td>963</td>
<td>643</td>
<td>500</td>
<td>280</td>
<td>80</td>
<td>2508</td>
<td>2432</td>
<td>1133</td>
<td>321</td>
<td>77</td>
</tr>
<tr>
<td>(\text{DieMCnd})</td>
<td>13</td>
<td>15</td>
<td>15</td>
<td>13</td>
<td>10</td>
<td>8</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>(\text{SubsetCnd})</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>100</td>
<td>98</td>
<td>61</td>
<td>31</td>
</tr>
<tr>
<td>(\text{Updates})</td>
<td>962</td>
<td>644</td>
<td>503</td>
<td>286</td>
<td>90</td>
<td>2508</td>
<td>2439</td>
<td>1209</td>
<td>387</td>
<td>109</td>
</tr>
</tbody>
</table>

### Panel C. Fund transaction costs

| \(\Sigma TC\) (MSEK) | 135 | 99 | 63 | 34 | 22 | 950 | 936 | 675 | 181 | 65 |
| \(\text{Buy}\Sigma TC\) | 0.58 | 0.54 | 0.47 | 0.35 | 0.23 | 0.51 | 0.51 | 0.50 | 0.35 | 0.25 |
| \(\text{BuyQ}\Sigma TC\) | 0.80 | 0.77 | 0.74 | 0.72 | 0.80 | 0.68 | 0.68 | 0.75 | 0.80 | 0.83 |
| \(\text{BuyC}\Sigma TC\) | 0.06 | 0.07 | 0.08 | 0.09 | 0.05 | 0.08 | 0.08 | 0.06 | 0.05 | 0.05 |
| \(\text{SellQ}\Sigma TC\) | 0.82 | 0.80 | 0.78 | 0.76 | 0.70 | 0.69 | 0.69 | 0.75 | 0.81 | 0.81 |
| \(\text{SellC}\Sigma TC\) | 0.06 | 0.07 | 0.07 | 0.08 | 0.10 | 0.08 | 0.08 | 0.06 | 0.05 | 0.05 |

### Panel D. Stock transaction costs

| \(\text{AvgTC}^+\) (%) | 0.82 | 0.93 | 0.93 | 1.09 | 1.21 | 0.97 | 0.98 | 1.03 | 1.05 | 0.82 |
| \(\text{MaxTC}^+\) (%) | 6.01 | 5.59 | 4.60 | 8.28 | 7.23 | 8.88 | 8.88 | 10.86 | 5.70 | 3.72 |
| \(\text{AvgTC}^-\) (%) | 0.04 | 0.05 | 0.06 | 0.07 | 0.07 | 0.09 | 0.09 | 0.13 | 0.13 | 0.08 |
| \(\text{MaxTC}^-\) (%) | 3.67 | 2.10 | 1.49 | 2.52 | 1.74 | 6.28 | 6.30 | 7.92 | 3.67 | 1.09 |
| \(\text{AvgTC}^-\) (%) | 0.76 | 0.83 | 0.93 | 1.22 | 1.39 | 0.92 | 0.93 | 1.00 | 1.09 | 1.26 |
| \(\text{MaxTC}^-\) (%) | 6.11 | 6.35 | 6.37 | 27.80 | 12.61 | 9.16 | 9.20 | 12.38 | 8.75 | 6.97 |
| \(\text{AvgTC}^-\) (%) | 0.06 | 0.09 | 0.17 | 0.45 | 0.64 | 0.09 | 0.10 | 0.18 | 0.35 | 0.52 |
| \(\text{MaxTC}^-\) (%) | 3.58 | 4.52 | 4.12 | 22.07 | 7.19 | 7.78 | 7.81 | 10.92 | 7.85 | 5.98 |
The ending NAVs as a proportion of the ending index value, range from 0.92 to 1.04.

For the three index funds with the lowest \( TElimit \), Alpha is near AvgTE, Beta is close to one, and Sigma close to SDTE.

Total incurred transaction costs in monetary units, \( \Sigma TC \) decreases with the value of the tracking error limit, from 135 to 22 MSEK, as do the average annual portfolio turnover, AvgTO, from 18 to 6%. Not surprisingly, AvgCash increases with \( TElimit \), because a fund with a less demanding \( TElimit \), does not have to invest cash immediately to achieve its \( TElimit \).

Buy transaction costs, \( \text{Buy} \Sigma TC \), constitute a larger proportion of \( \Sigma TC \) than sell transaction costs for the two lowest \( TElimit \)s, and vice versa for the larger limits. A possible explanation for this is that for large limits, few sell trades are done and a relatively large proportion of these will involve forcing stocks that have ceased to exist out of the fund. For such forced trades transaction costs are not controlled. This reasoning is supported by the observation that the average portfolio weight change for sell transactions, \( \text{Avg}x^- \), is larger than its buy transaction counterpart, \( \text{Avg}x^+ \), for all but the smallest limit and increasingly so. \( \text{Max}x^- \) and \( \text{Max}x^+ \) also support this.

\( \text{Buy}Q \Sigma TC \) and \( \text{Sell}Q \Sigma TC \) are never smaller than 0.69. Commissions, \( \text{Buy}C \Sigma TC \) and \( \text{Sell}C \Sigma TC \), are in the range 5-10%.

The transaction costs and and trade sizes for individual stocks seem reasonable, excepting the transactions in the stock of BGB, to which the large values refer.

For the five funds created from a subset of 20 stocks, the lower \( TElimit \)s are practically unachievable. The mean tracking error measure for the fund immediately after revision, but before trading, \( \text{AvgObj} \), for the three lowest limits is above the limits. (In addition, not reported explicitly, the 0.2% limit was never achieved and the 0.5% limit was achieved a few times.) The number of holdings in the subset funds is restricted not to exceed 20, but is allowed to be lower as indicated by MinN. The two subset funds with the largest \( TElimit \)s also have a positive cash position. This is permitted as long as the number of holdings, including cash, does not exceed 20. It also required that the forecasted tracking error measure meets the \( TElimit \), otherwise
the fund is updated, and then the investable set is constrained to 20 largest index stocks capitalization-wise.

Because of the difficulty in achieving high and stable tracking performance, the subset funds had to be frequently updated as indicated by the variable \textit{Updates}. For all but the two largest limit values, this approach based on a subset of 20 stocks results in extremely large portfolio turnovers, transaction costs, realized tracking errors, and low ending fund net asset values to ending index values.

On a different note, Jorion (2003) raises the concern that, without restrictions on short-selling, focusing on meeting a tracking error standard deviation equality constraint could produce portfolios with undesirable absolute risk-return characteristics, and, in particular, that no attention is paid to absolute risk, i.e., standard deviation, which could result in the portfolio having a markedly larger total standard deviation than the index. For this problem to arise, it is however necessary that no weight restrictions are imposed. Moreover, it appears, according to the results in Jorion (2003), that quite large values of the tracking error standard deviation constraint are required for the problem to be material. Though the prerequisites for the problem are not fulfilled here, it is reassuring to observe that the absolute risk measured as the autocorrelation-adjusted standard deviation, \textit{aSDRp}, of the index funds constructed without a subset constraint, never deviates much from that of the index of 22.41%.

\subsection*{4.4.9 Impact of neglect of transaction costs on tracking performance}

For \textit{Opt1}, in Table 4.6, the cases with \textit{Tau} of $10^7$ are close to neglect of transaction costs. The performance of the index fund constructed (without a subset restriction) and with a \textit{Tau} of $10^7$ is probably acceptable (on an absolute basis). Its performance is, however, unambiguously worse than the performances of the \textit{Opt1} funds with \textit{Taus} of $10^5$ and 10000, which are constructed with more emphasis on transaction cost minimization than the fund whose \textit{Tau} is $10^7$. These two funds have lower tracking error as well as lower transaction costs $\Sigma TC$ and higher ending fund values \textit{EndNAV}. For the subset index fund with a \textit{Tau} of $10^7$, the performance is simply miserable.

It is possible to conduct a more direct comparison of index funds implemented with and without regard to transaction costs for \textit{Opt2} than for \textit{Opt1}. For each index fund obtained with \textit{Opt2} (without a subset restriction) in the left panel of Table 4.7, where
transaction costs are minimized while the tracking error measure is kept below a limit, there is a fund tabulated in the left panel of Table 4.8 obtained (without a subset restriction and) without regard to transaction costs, i.e., the tracking error measure is kept below a limit but without minimizing transaction costs.

As expected, for each parameter setting or fund, the performance is equal or better in Table 4.7 than in Table 4.8. It is, however, important to remember that “better” is in relation to the particular objective and constraints of Opt2. For instance, the fund in Table 4.8 obtained with a $TE_{limit}$ of 5%, has lower tracking error $aSDTE$ than the corresponding fund in Table 4.7, where transaction costs were controlled for, 1.84% versus 2.53%. The fund constructed under transaction cost control, however, has lower monetary transaction costs $\Sigma TC$, 22 MSEK vs. 28 MSEK, higher ending value $EndNAV$, 1.04 vs. 0.88, and a more attractive range of maximum and minimum deviations, $MinNAV$ and $MaxNAV$, as a proportion of the index, 0.97-1.08 vs. 0.85-1.02. Since both funds satisfy the tracking error constraint of 5%, the performance of the fund obtained under transaction costs control has to be considered as superior despite its greater tracking error dispersion.

$Alpha$ equals zero and $Sigma$ is near $SDTE$ for the five index funds constructed without the subset constraint. $Beta$ decreases with $TE_{limit}$ from close to one to 0.97.

$Buy\Sigma TC$ behaves in similar fashion to when transaction costs were explicitly controlled. $BuyQ\Sigma TC$ and $SellQ\Sigma TC$ are larger than 0.68, and commissions, $BuyC\Sigma TC$ and $SellC\Sigma TC$, are in the range 5-8%, which, on the whole, is similar to what was observed under transaction cost control.
Table 4.8 *Opt2*: Meeting a tracking error limit without transaction cost control

**TElimit** is based on **SDTE**.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>0/0.2</th>
<th>0/0.5</th>
<th>0/1</th>
<th>0/2</th>
<th>0/5</th>
<th>20/0.2</th>
<th>20/0.5</th>
<th>20/1</th>
<th>20/2</th>
<th>20/5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A. Ex-post tracking performance</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>AvgTE (%)</strong></td>
<td>-0.004</td>
<td>-0.005</td>
<td>-0.005</td>
<td>-0.005</td>
<td>-0.005</td>
<td>-0.065</td>
<td>-0.064</td>
<td>-0.037</td>
<td>-0.008</td>
<td>-0.003</td>
</tr>
<tr>
<td><strong>SDTE (%)</strong></td>
<td>0.60</td>
<td>0.75</td>
<td>1.14</td>
<td>1.86</td>
<td>3.06</td>
<td>3.02</td>
<td>3.02</td>
<td>3.15</td>
<td>2.17</td>
<td>3.73</td>
</tr>
<tr>
<td><strong>aSDTE (%)</strong></td>
<td>0.62</td>
<td>0.73</td>
<td>0.91</td>
<td>1.29</td>
<td>1.84</td>
<td>12.81</td>
<td>12.84</td>
<td>13.36</td>
<td>2.88</td>
<td>2.54</td>
</tr>
<tr>
<td><strong>MinNAV</strong></td>
<td>1.00</td>
<td>1.001</td>
<td>1.009</td>
<td>1.013</td>
<td>1.016</td>
<td>1.000</td>
<td>1.000</td>
<td>1.004</td>
<td>1.004</td>
<td>1.005</td>
</tr>
<tr>
<td><strong>MaxNAV</strong></td>
<td>0.902</td>
<td>0.897</td>
<td>0.897</td>
<td>0.883</td>
<td>0.195</td>
<td>0.199</td>
<td>0.392</td>
<td>0.806</td>
<td>0.908</td>
<td></td>
</tr>
<tr>
<td><strong>aSDRp (%)</strong></td>
<td>22.38</td>
<td>22.13</td>
<td>21.85</td>
<td>21.79</td>
<td>22.26</td>
<td>27.97</td>
<td>28.01</td>
<td>26.49</td>
<td>22.35</td>
<td>22.68</td>
</tr>
<tr>
<td><strong>Panel B. Updating</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>AvgFcast (%)</strong></td>
<td>0.209</td>
<td>0.481</td>
<td>0.909</td>
<td>1.637</td>
<td>2.728</td>
<td>0.832</td>
<td>0.834</td>
<td>1.029</td>
<td>1.718</td>
<td>3.502</td>
</tr>
<tr>
<td><strong>AvgObj (%)</strong></td>
<td>0.200</td>
<td>0.500</td>
<td>0.999</td>
<td>1.984</td>
<td>4.26</td>
<td>0.811</td>
<td>0.823</td>
<td>1.081</td>
<td>1.898</td>
<td>3.824</td>
</tr>
<tr>
<td><strong>AvgObjTC</strong></td>
<td>0.200</td>
<td>0.500</td>
<td>1.000</td>
<td>1.985</td>
<td>4.26</td>
<td>0.811</td>
<td>0.823</td>
<td>1.082</td>
<td>1.900</td>
<td>3.827</td>
</tr>
<tr>
<td><strong>TECnd</strong></td>
<td>1180</td>
<td>801</td>
<td>556</td>
<td>246</td>
<td>34</td>
<td>2508</td>
<td>2431</td>
<td>1137</td>
<td>329</td>
<td>61</td>
</tr>
<tr>
<td><strong>DieMCond</strong></td>
<td>12</td>
<td>11</td>
<td>11</td>
<td>13</td>
<td>15</td>
<td>8</td>
<td>8</td>
<td>7</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td><strong>SubsetCond</strong></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>100</td>
<td>109</td>
<td>111</td>
<td>101</td>
</tr>
<tr>
<td><strong>Updates</strong></td>
<td>1164</td>
<td>798</td>
<td>555</td>
<td>255</td>
<td>49</td>
<td>2508</td>
<td>2438</td>
<td>1209</td>
<td>440</td>
<td>164</td>
</tr>
<tr>
<td><strong>Panel C. Fund transaction costs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ΣTC (MSEK)</strong></td>
<td>181</td>
<td>153</td>
<td>105</td>
<td>54</td>
<td>28</td>
<td>950</td>
<td>936</td>
<td>677</td>
<td>192</td>
<td>78</td>
</tr>
<tr>
<td><strong>BnqΣTC</strong></td>
<td>0.57</td>
<td>0.54</td>
<td>0.51</td>
<td>0.41</td>
<td>0.37</td>
<td>0.51</td>
<td>0.51</td>
<td>0.50</td>
<td>0.37</td>
<td>0.27</td>
</tr>
<tr>
<td><strong>BnqΣTC</strong></td>
<td>0.77</td>
<td>0.74</td>
<td>0.71</td>
<td>0.68</td>
<td>0.71</td>
<td>0.68</td>
<td>0.68</td>
<td>0.74</td>
<td>0.77</td>
<td>0.79</td>
</tr>
<tr>
<td><strong>SellQΣTC</strong></td>
<td>0.06</td>
<td>0.07</td>
<td>0.07</td>
<td>0.08</td>
<td>0.06</td>
<td>0.08</td>
<td>0.08</td>
<td>0.06</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td><strong>Panel D. Stock transaction costs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>AvgTC+ (%)</strong></td>
<td>0.95</td>
<td>0.98</td>
<td>1.03</td>
<td>1.00</td>
<td>1.18</td>
<td>0.97</td>
<td>0.98</td>
<td>1.05</td>
<td>1.01</td>
<td>0.86</td>
</tr>
<tr>
<td><strong>MaxTC+ (%)</strong></td>
<td>7.23</td>
<td>6.42</td>
<td>9.08</td>
<td>5.12</td>
<td>68.05</td>
<td>8.88</td>
<td>8.88</td>
<td>10.78</td>
<td>6.69</td>
<td>3.42</td>
</tr>
<tr>
<td><strong>Avggoz+ (%)</strong></td>
<td>0.02</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.16</td>
<td>0.09</td>
<td>0.09</td>
<td>0.11</td>
<td>0.07</td>
<td>0.03</td>
</tr>
<tr>
<td><strong>Maxx+ (%)</strong></td>
<td>4.82</td>
<td>5.06</td>
<td>7.27</td>
<td>3.01</td>
<td>52.02</td>
<td>6.28</td>
<td>6.30</td>
<td>7.85</td>
<td>4.55</td>
<td>1.73</td>
</tr>
<tr>
<td><strong>AvgTC- (%)</strong></td>
<td>0.90</td>
<td>0.92</td>
<td>0.99</td>
<td>1.11</td>
<td>3.45</td>
<td>0.92</td>
<td>0.93</td>
<td>1.04</td>
<td>1.12</td>
<td>1.43</td>
</tr>
<tr>
<td><strong>MaxTC- (%)</strong></td>
<td>6.22</td>
<td>5.82</td>
<td>6.86</td>
<td>9.78</td>
<td>81.67</td>
<td>9.16</td>
<td>9.20</td>
<td>12.29</td>
<td>9.84</td>
<td>6.88</td>
</tr>
<tr>
<td><strong>Avggoz- (%)</strong></td>
<td>0.03</td>
<td>0.04</td>
<td>0.07</td>
<td>0.16</td>
<td>2.27</td>
<td>0.09</td>
<td>0.10</td>
<td>0.15</td>
<td>0.22</td>
<td>0.58</td>
</tr>
<tr>
<td><strong>Maxx- (%)</strong></td>
<td>4.40</td>
<td>4.28</td>
<td>4.88</td>
<td>7.89</td>
<td>74.89</td>
<td>7.78</td>
<td>7.81</td>
<td>10.83</td>
<td>8.51</td>
<td>5.82</td>
</tr>
</tbody>
</table>
Figure 4.4 illustrates the performance relative to the index level of the two index funds implemented using Opt2 with and without transaction control, respectively. The fund constructed without transaction cost control is outperformed. Their respective EndNAVs are 0.941 and 0.897, a difference of 4.4%, which translates into a monetary difference of 317 MSEK. A paired t-test rejects that the average return difference between the funds is zero.

Moreover, control of transaction costs reduces incurred transaction costs by 54 MSEK (=153-99) over a ten-year period non-compounded. Roughly, this means that on average 5.4 MSEK per year could be used to increase fund expenses, e.g., compensation to the fund managers. Alternatively, the cost savings could be transferred to investors who would realize improved performance in terms of lower tracking error standard deviation, higher EndNAV, MinNAV, and MaxNAV.

Table 4.7 and Table 4.8, the rightmost panels, display the tracking performances of subset index funds implemented using Opt2 with and without transaction cost minimization, respectively. The performance of the funds constructed under transaction cost control is similar or better than that of the funds constructed without transaction cost control.
4.4.10 Efficiency of pre-defined subsets as means to control transaction costs

As noted, the results for Opt1 presented in Table 4.6 do indicate that the use of pre-defined subsets reduce transaction costs. Likewise, the results for subset index funds implemented with Opt2 under explicit transaction cost control in the right panel of Table 4.7 indicate that use of pre-defined subsets could impair tracking performance significantly.

The right panel of Table 4.8 displays the tracking performances of subset index funds constructed using Opt2 but without transaction cost minimization in the mathematical program. Each one of these funds, except the fund with a Telimit of 5%, underperforms significantly and unambiguously the corresponding fund that is constructed without a subset restriction and whose tracking performance is displayed in the left panel of Table 4.8. There is again no evidence that the use of pre-defined subsets reduces transaction costs.

The results obtained here, for an index with relatively few and liquid stocks, demonstrate that subset restrictions can be severe. However, if there are many index constituents, or if there are high fixed costs per stock, then subset constraints may be productive.

4.4.11 Impact of different tracking error measures on tracking performance

In several studies, e.g., Rudolf et al. (1999), Rey and Seiler (2001), Satchell and Hwang (2001), and Beasley et al. (2003) the question of what tracking error measure to use is emphasized. The empirical performances of a number of different tracking error measures are presented below.

**SDTE and RMSTE**

The standard deviation tracking error SDTE has been used as the forward-looking tracking error measure hitherto. Most analyzes, however, have also been performed with the root mean square tracking error measure, RMSTE, in place of SDTE. The results of these analyzes are as follows.

The results for full replication under conditional updating based on RMSTE were more or less identical to the results in Table 4.5 which were based on SDTE.
In *Opt1*, under transaction costs and dividends with a one day delay, using *SDTE* or *RMSTE* in the objective does not seem to matter, as they each generate results that, in the dimensions of Table 4.6, are indistinguishable from one another.

For *Opt2*, under transaction costs and dividends excepting subset funds, *SDTE* and *RMSTE* generate virtually identical results in terms of Table 4.7.

The formulas for *SDTE* and *RMSTE* do not involve transaction costs and dividends, but the above results were, in fact, affected by those factors. Additional results, unaffected by transaction costs and dividends, are presented below. This should illustrate the usefulness of the portfolio system developed in its capacity as a research laboratory.

*Outdated vs. updated historical index return vectors in tracking error measures*  
It was argued previously that the possibly outdated index return vector \( \mathbf{r}_b \), based on the historical index composition, should be avoided in tracking error measures used for forward-looking purposes and optimization. It is possible to use \( \mathbf{r}_b \) with *Opt1* and *Opt2* under transaction costs and dividends. The resulting tracking performance could be compared to the tracking performances achieved when the index vector \( \mathbf{R}_b \) based on the current index composition was used. This, however, would require that *Opt1* and *Opt2*, which use \( \mathbf{R}_b \), are reformulated to incorporate \( \mathbf{r}_b \).

Moreover, the results could be affected by transaction costs and dividends, although these factors are not directly related to whether \( \mathbf{R}_b \) or \( \mathbf{r}_b \) is used in the tracking error measure. Therefore, to isolate transaction costs and dividends from affecting the results, the impact on tracking performance of using a benchmark return vector based on historical (outdated) index composition instead of a benchmark return vector based on current index composition is analyzed in the absence of transaction costs and dividends.

Table 4.9 presents the results for minimization of *SDTE* and *RMSTE* with no transaction costs or dividends using an index return vector based on (i) the current index composition \( \mathbf{R}_b \), and (ii) the historical index composition \( \mathbf{r}_b \). Thus, the following four tracking error minimization objectives are implemented and tested:

\[
SDTE (\mathbf{R}_b): \text{Var}(\mathbf{R}_x - \mathbf{R}_b) = (\mathbf{x} - \mathbf{b})'\text{Var}(\mathbf{R})(\mathbf{x} - \mathbf{b}),
\]

\[
RMSTE (\mathbf{R}_b): (\mathbf{R}_x - \mathbf{R}_b)'(\mathbf{R}_x - \mathbf{R}_b) = \mathbf{b}'(\mathbf{R}'\mathbf{R})\mathbf{b} + \mathbf{x}'(\mathbf{R}'\mathbf{R})\mathbf{x} - 2 \mathbf{x}'(\mathbf{R}'\mathbf{R})\mathbf{b},
\]
\[
SDTE(r_b): \quad \text{Var}(Rx-r_b) = \text{Var}(Rx) + \text{Var}(r_b) - 2\text{Cov}(Rx,r_b)
\]
\[
= x'\text{Var}(R)x + \text{Var}(r_b) - 2x'[E(r_b'R) - E(r_b)E(R)],
\]

\[
RMSTE(r_b): \quad (Rx-r_b)'(Rx-r_b) = Rx'Rx + r_b'r_b - 2 r_b'Rx = xR'Rx - 2x'(r_b'R).
\]

The objectives corresponding to SDTE(Rb) and SDTE(rb) were implemented in the form of their sample counterparts, and the objectives corresponding to RMSTE(Rb) and RMSTE(rb) were scaled in consistency with this.

In this streamlined setting with neither dividends nor transaction costs, the results in Table 4.9 demonstrate that the objectives based on returns inferred from the current index composition outperformed the ones based on returns inferred from historical index composition.

For the objectives consistent with the tracking error measure SDTE, and no subset restriction, the realized tracking error \(aSDTE\) is zero for SDTE(Rb), and a hefty 1.50% for SDTE(rb), the objective based on historical index composition. For the corresponding subset funds, the differences are however smaller.

For RMSTE(Rb), the index fund without subset constraint and the fund constrained to a subset of 20 stocks perform close to identical to the corresponding funds obtained under SDTE. The high degree of similarity between funds constructed on basis of RMSTE and SDTE, respectively, suggests that the choice between the two tracking error measures is of less importance.

Again, subset index funds underperform but this time in a setting without dividends and transaction costs. The earlier impression that pre-defined subsets are severe restrictions is reinforced.
### Table 4.9 Opt1: Outdated versus updated historical index return series

Only tracking error minimization. No transaction costs. No dividends.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>SDTE (Rb)</th>
<th>SDTE (rb)</th>
<th>RMSTE (Rb)</th>
<th>RMSTE(rb)</th>
<th>SHTE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Subset</td>
<td>Subset</td>
<td>Subset</td>
<td>Subset</td>
<td>Subset</td>
</tr>
<tr>
<td>AvgTE (%)</td>
<td>0.000</td>
<td>-0.002</td>
<td>-0.001</td>
<td>-0.003</td>
<td>0.000</td>
</tr>
<tr>
<td>MinTE (%)</td>
<td>0.00</td>
<td>-0.66</td>
<td>-1.25</td>
<td>-1.21</td>
<td>0.00</td>
</tr>
<tr>
<td>MaxTE (%)</td>
<td>0.38</td>
<td>0.72</td>
<td>0.58</td>
<td>0.38</td>
<td>0.78</td>
</tr>
<tr>
<td>SDTE (%)</td>
<td>1.37</td>
<td>1.40</td>
<td>1.86</td>
<td>1.35</td>
<td>1.47</td>
</tr>
<tr>
<td>aSDTE (%)</td>
<td>1.18</td>
<td>1.50</td>
<td>2.10</td>
<td>1.16</td>
<td>1.65</td>
</tr>
<tr>
<td>RMSTE (%)</td>
<td>1.37</td>
<td>1.40</td>
<td>1.86</td>
<td>1.35</td>
<td>1.47</td>
</tr>
<tr>
<td>MinNAV</td>
<td>1.000</td>
<td>0.954</td>
<td>0.954</td>
<td>0.908</td>
<td>0.959</td>
</tr>
<tr>
<td>MaxNAV</td>
<td>1.000</td>
<td>1.004</td>
<td>1.017</td>
<td>1.003</td>
<td>1.042</td>
</tr>
<tr>
<td>EndNAV</td>
<td>1.000</td>
<td>0.958</td>
<td>0.969</td>
<td>0.929</td>
<td>0.991</td>
</tr>
<tr>
<td>aSDRp (%)</td>
<td>22.55</td>
<td>22.85</td>
<td>22.36</td>
<td>22.82</td>
<td>22.41</td>
</tr>
<tr>
<td>Alpha (%)</td>
<td>0.000</td>
<td>-0.002</td>
<td>-0.001</td>
<td>-0.003</td>
<td>0.000</td>
</tr>
<tr>
<td>Beta</td>
<td>1.000</td>
<td>1.002</td>
<td>0.998</td>
<td>1.002</td>
<td>0.998</td>
</tr>
<tr>
<td>Sigma (%)</td>
<td>1.37</td>
<td>1.40</td>
<td>1.86</td>
<td>1.35</td>
<td>1.47</td>
</tr>
<tr>
<td>AvgTO (%)</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Avgx+ (%)</td>
<td>0.18</td>
<td>0.24</td>
<td>0.17</td>
<td>0.31</td>
<td>0.18</td>
</tr>
<tr>
<td>Avgx- (%)</td>
<td>0.15</td>
<td>0.12</td>
<td>0.15</td>
<td>0.16</td>
<td>0.15</td>
</tr>
<tr>
<td>MinN</td>
<td>28</td>
<td>16</td>
<td>20</td>
<td>15</td>
<td>28</td>
</tr>
<tr>
<td>AvgN</td>
<td>29.6</td>
<td>19.8</td>
<td>29.0</td>
<td>19.5</td>
<td>29.6</td>
</tr>
<tr>
<td>MaxN</td>
<td>32</td>
<td>20</td>
<td>33</td>
<td>20</td>
<td>32</td>
</tr>
</tbody>
</table>

### Panel A. Ex-post tracking performance

### Panel B. Updating

Satchell and Hwang tracking error measure

In the rightmost column of Table 4.9, the results are presented for the index fund selection model and a tracking error measure based on variable weights as suggested by Satchell and Hwang (2001). With no subset constraint imposed, the results are practically identical to those obtained for \( \text{RMSTE}(\text{Rb}) \) and \( \text{SDTE}(\text{Rb}) \). This is in line with expectations, as Satchell and Hwang (2001) state that a premise, for a bias to exist between measures based on constant and variable weights, is that the fund and...
Index are not identically weighted, whereas the results here indicate that the fund and the index are close to identically weighted for most of the time.

**Mean absolute tracking error**

Suggested advantages of using a tracking error measure based on absolute tracking errors, such as $MATE$, rather than a measure based on squared tracking errors, such as $SDTE$, include interpretability and robustness. As noted, for normally distributed tracking errors with zero mean, the numerical value of $MATE$ should be around 80% of the numerical value of $SDTE$.

Results for $SDTE$ and $MATE$ meeting a $TElimit$ without transaction costs and dividends, are presented in Table 4.10. The $MATE$ measured on daily tracking errors was (arbitrarily, though consistent with the annualization of daily tracking error standard deviation) annualized by multiplication with the square root of 252.

The difference between realized tracking error and the corresponding tracking error limits is smaller for $MATE$ than for $SDTE$. The index funds implemented using $MATE$ violate by smaller amounts their limits than do funds constructed on basis of $SDTE$.

For the three lowest values of the tracking error limits, the index funds constructed with $MATE$ and $STDE$, respectively, perform close to equal in terms of $aSTDE$ as well as $MinNAV$, $MaxNAV$, and $EndNAV$.

The observed ratio between $MATE$ and $SDTE$ for a given index fund is near 0.8, which is close to what was hypothesized. This applies regardless of whether the index funds were constructed on basis of $SDTE$ or $MATE$. The ratio, however, decreases as $TElimit$ increases. One explanation for this could be that the mean tracking error $AvgTE$ deviates more and more from zero as $TElimit$ increases, and as this happens $MATE$ starts to deviate more and more from $MAD$, for which the relation actually applies. Another explanation is that the empirical tracking error distribution gets less normal as $TElimit$ increases.
Table 4.10 Opt2: SDTE versus MATE meeting a TLimit
No transaction costs. No dividends.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>SDTE</th>
<th>MATE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>AvgTE (%)</td>
<td>0.00</td>
<td>0.28</td>
</tr>
<tr>
<td>aSDTE (%)</td>
<td>0.00</td>
<td>0.21</td>
</tr>
<tr>
<td>MATE (%)</td>
<td>0.00</td>
<td>0.20</td>
</tr>
<tr>
<td>MinNAV</td>
<td>1.00</td>
<td>0.990</td>
</tr>
<tr>
<td>MaxNAV</td>
<td>1.00</td>
<td>1.002</td>
</tr>
<tr>
<td>EndNAV</td>
<td>1.00</td>
<td>0.992</td>
</tr>
<tr>
<td>aSDRp (%)</td>
<td>22.55</td>
<td>22.56</td>
</tr>
<tr>
<td>Alpha (%)</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Beta</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Sigma (%)</td>
<td>0.00</td>
<td>0.28</td>
</tr>
<tr>
<td>AvgCash (%)</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>AvgTO (%)</td>
<td>19.6</td>
<td>22.8</td>
</tr>
<tr>
<td>Avgx+ (%)</td>
<td>0.18</td>
<td>0.02</td>
</tr>
<tr>
<td>Avgx- (%)</td>
<td>0.15</td>
<td>0.04</td>
</tr>
<tr>
<td>MinN</td>
<td>28</td>
<td>29</td>
</tr>
<tr>
<td>AvgN</td>
<td>29.6</td>
<td>32.2</td>
</tr>
<tr>
<td>MaxN</td>
<td>32</td>
<td>35</td>
</tr>
</tbody>
</table>

Panel A. Ex-post tracking performance

As a tracking error measure, MATE appears to more robust than SDTE in the sense that the TLimit condition, TECnd, is triggered fewer times for MATE than for SDTE. This is also reflected in the number of actual fund revisions, Updates. The average portfolio turnover rate is also lower for MATE except for when the TLimit is zero. However, when the funds managed under MATE are rebalanced, the average transaction sizes, Avgx+ and Avgx-, are much larger than for the SDTE funds. In the presence of transaction costs, it may still be the case that MATE funds, despite fewer revisions, incur larger tracking errors and transaction costs than SDTE funds. Whether this is the case or not is left for future research to find out.
Minimizing a serial correlation adjusted tracking error objective

In the results above, the annualized autocorrelation corrected tracking errors generally differ from the uncorrected ones. The tracking error goal functions and constraints are not corrected for autocorrelation. Given that autocorrelated-corrected annualized tracking error measures are the standard for comparisons, it appears that one possible way of increasing the degree to which desired tracking performance is achieved, would be to use objectives, constraints, and conditions that are corrected for autocorrelation. Provisional experiments with formulations adjusted for autocorrelation (in-sample) exhibited limited success. In particular, they demonstrated overfitting with regard to the autocorrelation adjustment.

4.5 Summary of results and conclusions

In this chapter, I implemented and tested more thoroughly the tracking performance of three alternative equity indexing approaches, full replication and two optimization-based ones, which feature different types and degrees of transaction cost control. Though full replication is reported to be the most frequently used method in practice according to, e.g., Blume and Edelen (2004), it has not been tested explicitly in research. The standard unconditional full replication approach is complemented by conditional variants. To address the question about the importance of transaction cost control more directly, I formulated two programs for index fund revision under transaction costs including price impact costs: (i) a program that features a goal function in which transaction costs are balanced against a tracking error measure utilizing a trade-off parameter ($Opt1$), and (ii) a program where the objective is to minimize transaction costs while keeping a tracking error measure less than or equal to a pre-specified limit ($Opt2$).

The three approaches were employed to track a capitalization-weighted Swedish stock index with relatively few and liquid stocks. Previous research was extended by the development of a research design that incorporates transaction costs that are inclusive of price impact and variable across stocks and time, cash flows in the form of dividends, and corporate actions including mergers and acquisitions, new issues and other capitalization changes. The empirical investigation uses more than ten years of
daily data and it is the first one to provide large-scale evidence in a setting that incorporates explicitly many of the factors that present the challenges to index fund management.

Another distinctive feature of the study in this chapter is that, instead of using an existing index, an index was calculated independently according to a consistent, computerized methodology, mimicking that according to which the OMX index is calculated; the OMX index comprises the 30 most traded stocks on the Stockholm Stock Exchange and it is the most well known and used index in Sweden and the Nordic region. Having an index that is calculated according to a consistent methodology is desirable because index methodology is a determinant of tracking performance. Also, the approach solves several issues related to data availability and quality, including the calculation of a total return index with dividends reinvested.

The empirical relation between the OMX and the independently calculated price index was examined. The performance of the indexes were quite similar, despite that the index methodology for the OMX have been changed several times. In this respect, the index calculated here was successful.

Full replication was applied to track both the price index and the return index in a laboratory setting without transaction costs. The impact of dividends and different dividend receipt schemes on tracking performance was examined. Without dividends, the tracking performance of the index fund was identical to that of the index. A 1-day dividend receipt delay, relative to the ex-dividend day, resulted in a very low tracking error standard deviation. An 8-day delay resulted in larger tracking error standard deviation. Based on those results, the first convention was adopted in the tests that followed.

Full replication was dominated, in all dimensions of tracking performance considered, by the two optimization-based approaches with explicit control of transaction costs. The optimization-based approaches were implemented and tested without transaction cost control. The resulting index funds were dominated across all dimensions of tracking performance by their counterparts obtained with control of transaction costs.
As in Chapter 3, the results suggest that price impact costs are important to consider. Across all the index funds implemented, for both purchase and sale transactions, the quadratic part of the transaction cost specification, which allow for explicit consideration of price impact costs by letting the percentage costs be volume-dependent, constitutes more than 70% of the monetary transaction costs. For both buys and sells, commissions were usually the least important part, and they never represented more than 10% of the total monetary transaction costs. In view of this, current reporting practices of mutual funds appear imperfect, because only commissions, which appear to constitute a small part of total transaction costs, are disclosed.

The realized transaction costs and trade sizes for trades in individual stocks were documented and found reasonable. Occasionally, the transaction sizes were such that the transaction cost models needed to be extrapolated outside the domain for which they were estimated.

Connor and Leland (1995) show that a cash position can help to reduce transaction costs for an index fund. That finding is corroborated here in that index funds constructed with increasing emphasis on transaction cost minimization also exhibit increasingly large average cash holdings.

Many studies that implement and test different approaches to index fund management, e.g., Rudd (1980), Larsen and Resnick (1998), Wagner (1998), Rudolf et al. (1999), Bamberg and Wagner (2000), and Rey and Seiler (2001), construct index funds from a pre-defined subset of the index stocks. It was investigated whether transaction costs could be reduced by constructing index funds from a pre-defined subset of the 20 most liquid index stocks, where increasing market capitalization was used as a proxy for increasing liquidity. The usefulness of subsets as means for controlling transaction costs in the tracking of an index with relatively few and liquid stocks was zero, and the use of subsets rather amplified tracking error and transaction costs incurred.

The performance of some alternative tracking error measures was tested. Objectives based on tracking error standard deviation performed in a manner indistinguishable from that of root mean square when they were evaluated in terms of
autocorrelation adjusted tracking error standard deviation, in a laboratory setting that excludes transaction costs and dividends. Rey and Seiler (2001) and Beasley et al. (2003) argue that the root mean square tracking error measure is to prefer over tracking error standard deviation. The analysis here, however, suggest that the two are quite similar in that they produce funds with similar tracking performance.

Meade and Salkin (1989) and Pope and Yadav (1994) indicate that there is a risk that estimates of the tracking error standard deviation based on short-interval returns could be biased due to autocorrelation in the tracking errors calculated from such returns. I complemented their research by calculating autocorrelation adjusted tracking error standard deviations for all index funds implemented. For reasonably performing index funds, the autocorrelation-adjusted tracking error standard deviation was, in several instances, particularly for larger tracking errors, lower than the unadjusted standard deviation, thus indicating a potential need for adjustment.

Rudolf et al. (1999) advocate, as an objective for indexing, a mean absolute tracking error measure over quadratic tracking error measures, such as root mean square and standard deviation, on grounds that it provides a more immediate interpretation of the optimized value of the objective function. I argued that this claim was difficult to understand. Actually, Herrey (1965) reports results which suggest that, under certain, quite reasonable conditions, the numerical value of $MATE$ is approximately equal to 0.8 times the $SDTE$; the empirical findings support this relation. An index fund selection formulation featuring a mean absolute tracking error measure constrained to meet a limit was implemented and tested. Its performance was compared to that of an equivalent formulation based on $SDTE$ instead of $MATE$. The results obtained in a setting without transaction costs and dividends, indicated that the formulations produce index funds whose tracking performances are quite similar.

A heuristic implementation based on the observations made by Satchell and Hwang (2001) generated, in the absence of transaction costs and dividends, tracking performance very similar to what was obtained for $SDTE$.

That accurate data are important for index fund management was observed. Empirical evidence was found that index tracking based on a historical time-series of index returns rather than on complete data based on index composition at the time of
the optimization could lead to inferior tracking performance. This finding has bearing on the results obtained in the studies by, e.g., Walsh et al. (1998), Rudolf et al. (1999), Bamberg and Wagner (2001), and Beasley et al. (2003) that implement and test different indexing approaches based on historical time-series data instead of on current index composition data.

The research design was embodied in a computer program or “portfolio research system” which features a securities inventory, optimization, and trading in the presence of transaction costs, corporate actions, and cash flows, all on a daily basis. The system demonstrated itself useful as a research laboratory in which idealized conditions can be furnished so as to meet different research demands.

Overall, the results indicate that transaction costs in general are significant and that transaction cost control can enhance performance. For example, an index fund, with an initial net asset value of 1000 MSEK, managed with Opt2 to meet a $T_{E_{\text{limit}}}$ of 0.5%, and under control of transaction costs, had after ten years a net asset value that was 317 MSEK, or 5%, larger than that of an index fund constructed to meet the same $T_{E_{\text{limit}}}$ but without transaction cost control. The cost savings could be used to increase fund expenses, e.g., the compensation to the fund managers. Alternatively, the savings could be passed on to investors, who would realize improved performance in terms of lower tracking error standard deviation and higher returns. To reflect the performance for a fund investor, an annual management fee and other expenses need to be subtracted, but this has not been done here.

A tentative analysis was conducted of the tracking performance of an existing OMX index mutual fund. Its annualized tracking error standard deviation of around 1.8% appears high compared to those of the U.S. and Australian index funds analyzed in Frino and Gallagher (2001) and Frino and Gallagher (2002), respectively. This impression is retained when a comparison is made with the index funds implemented here. Recall, however, that the funds constructed in this study were unaffected by investor cash flows and other costs than transaction costs. Without knowledge of the index mutual fund’s targeted tracking performance, it is, however, impossible to judge how successful the fund has been.
A factor that was not accounted for in the study in this chapter, is investor cash flows, that is, withdrawals from and deposits to a fund by investors. The reason was lack of data. Investor flows will generally force an index fund to update more frequently so as to maintain the desired tracking performance after infusions, and to raise cash to meet withdrawals. In addition, Edelen (1999) finds evidence that the transaction costs induced by investor flows have explanatory power with regard to the underperformance of mutual funds. It appears likely that transaction cost control should be increasingly important in the presence of investor flows. An interesting opportunity for future research should hence be to investigate the effects of investor flows on index fund management and performance.
Chapter 5

Concluding summary

The first section in this chapter contains a summary of the thesis. The departure points for the thesis as well as the contents of Chapters 2, 3, and 4 are recapitulated. After that, overall conclusions are drawn with regard to the research questions and the main purpose initially stated. The theoretical, methodological, and empirical contributions of this work are outlined in Section 5.3. The final section contains a discussion of possible directions for future research.

5.1 Thesis summary

At the outset of this thesis it was conjectured that transaction cost control could be a key to improved performance in portfolio management. Based on this, two research questions were formulated: Do transaction costs matter in portfolio management? and Could transaction cost control improve portfolio performance?

I decided to study these questions within the context of two different portfolio management problems, the index fund problem and the mean-variance problem. In response to a number of issues related to transaction costs, and particularly to price impact costs, it was decided that the treatment of transaction costs throughout the thesis should include price impact cost and be based on the premises that trading is uninformed, immediate, and executed in the computerized public limit order book market of the Stockholm Stock Exchange. As a considerable amount of all trading is uninformed and immediate, the outcomes of this research may be relevant to many parties. Moreover, almost all trading systems that are being implemented are electronic limit order book systems. Their practical relevance is thus increasing.

The index fund problem was selected for several reasons. A general conclusion of the fund performance evaluation literature is the recommendation of index funds as the preferred investment vehicle (see, e.g., Wermers 2000). Since as much as 1/3 of all money under management globally may be invested in index funds, improvements to the efficiency of index fund management could result in aggregate welfare gains that
are sizeable. Several researchers, e.g., Frino and Gallagher (2001) and Beasley et al. (2003), indicate that the index fund problem is less well-understood and that additional studies are needed. Transaction cost control is made explicit in the problem formulation (Rudd 1980; Frino and Gallagher 2001; Beasley et al. 2003). As the trading of index funds per definition is uninformed and often conducted under immediacy, the index fund problem allows for a rather unambiguous treatment of transaction costs.

The part of uninformed immediate trading that is not conducted by index funds, represent implementations of solutions to portfolio problems other than index fund problems. Rather than studying the research questions within the context of many different portfolio problems, in addition to the index fund problem, I decided to confine the analysis to one particular portfolio problem, the mean-variance portfolio management problem of Markowitz (1952, 1959, 1987). The mean-variance approach was selected because of its central position in the literature. In addition, it has been demonstrated to being able to approximate various other portfolio problems (Ziemba and Kallberg 1983; Amilon 2001). Because the mean-variance problem has a closer connection to utility theory than the index fund problem, it better allows for analyzing in utility-theoretic terms the importance of transaction costs and transaction cost control in portfolio choice.

For each portfolio problem, I attempt to answer the research questions posed by completing the following three research tasks:

(i) develop models of individual stocks’ transaction costs,

(ii) formulate (at least) one portfolio decision model that incorporates these transation costs, and

(iii) design and conduct empirical tests, based on (i) and (ii), that seek to assess the importance of transaction costs and transaction cost control.

The premises employed regarding trading and transaction costs allowed for a common treatment of research task (i) across the portfolio problems. I decided to structure the thesis according to the research tasks and the two portfolio problems as follows. Chapter 2 deals with research task (i). In Chapter 3, research task (ii) and (iii) are treated within the context of the mean-variance problem, and in Chapter 4, research
tasks (ii) and (iii) are treated within the context of index fund management. The contents of Chapters 2, 3, and 4 are summarized below.

Chapter 2 Transaction costs

The development of models of Swedish stocks’ transaction costs including price impact costs involved the following steps.

I showed analytically that the price impact cost - the absolute volume-weighted average price degradation relative to the quote midpoint - for market buy and sell orders in a limit order book with discrete prices is an increasing piecewise concave function of order volume. This detail seems to have been unnoticed in the literature (see, e.g., Glosten 1994 and Niemeyer and Sandás 1993).

Based on electronic limit order book data for a large cross-section of Swedish equities, I estimated empirical price impact cost functions separately for market buy and sell orders. The estimated price impact cost functions are decreasing in market capitalization and historical trading activity; and increasing in order size (measured as the number of shares in the order relative to all outstanding shares), stock return volatility, and, for sell orders only, quote midpoint. These results are largely consistent with microstructure theory and empirical results obtained for other markets (see, e.g., Keim and Madhavan 1997). The separate analysis of bids and asks, enabled the generation of evidence indicating that the price impact costs for market buy orders, on average, are higher and increase faster in order size than for market sell orders. This suggests that submitters of limit orders require higher compensation for providing liquidity to buyers than to sellers.

Simulation analysis were employed to examine the robustness and out-of-sample performance of the estimated models. The error metric used was the actual percentage price impact cost minus the forecast. By multiplying this percentage forecast error with the monetary order value, the monetary value of the forecast error is obtained. The robustness and the out-of-sample performance of the estimated models were satisfactory. For both buy and sell orders, the mean signed forecast error was close to 0%, while the mean absolute forecast error was below 1%. The maximum absolute price impact cost forecast errors for sell and buy orders were 7.9% and 10.6%,
respectively. Total transaction costs are obtained by adding the relevant commission rate to the price impact cost.

Chapter 3 Mean-variance portfolio management under transaction costs

I extended the standard mean-variance portfolio selection model by formulating a quadratic program for mean-variance portfolio revisions under transaction costs including price impact costs. The transaction cost specification used, and that allows for explicit modeling of price impact costs, is a quadratic function (without constant term) of portfolio weight change, and is thus consistent with the empirical transaction cost models developed in Chapter 2. In lieu of more elaborate transaction cost specifications, the approach taken is believed to provide a good balance between realism, computational cost, and ease of implementation.

The empirical transaction cost models of Chapter 2 were integrated with the extended portfolio model. An empirical test were performed to analyze how varying degrees of consideration of transaction costs in portfolio revisions are related to investor welfare, portfolio allocations, turnover rates, and incurred levels of transaction costs. The integrated model was applied to revise portfolios of different sizes in terms of net asset value and across a broad range of risk attitudes. The sizes of the portfolios were intended to be typical for Swedish mutual funds. The initial (unrevised) portfolios are capitalization-weighted and contain all Swedish stocks with sufficient data. The test estimated for each revision the transaction costs and portfolio turnover incurred as well as the expected utility and diversification of the resulting portfolio. Additional results concerned the effects on portfolio utility and diversification of a maximum weight constraint on individual holdings as imposed by Swedish law.

The integrated model’s performance were compared to that of the standard mean-variance model as well as to the performances of some portfolio revision models involving less elaborate transaction cost specifications. The portfolio revisions conducted with the integrated model incurred transaction costs of 5-6%, whereas revisions performed with standard mean-variance model, which neglects transaction costs, experienced transaction costs that were greater, ranging from 10% to 31%. Compared to typical levels of portfolio returns, the incurred transaction costs are non-
trivial. The standard mean-variance model realized certainty equivalent losses of 6-20% relative to the extended model, which, in addition, exhibited lower turnover, higher diversification, and lower transaction costs incurred.

The maximum weight constraint on individual holdings, led to higher diversification in terms of the number of (non-zero) portfolio holdings, but induced certainty equivalent losses of around 4% compared to when no maximum weight constraint was in place.

In general, the effects of neglecting transaction costs were more serious for the larger portfolios NAV-wise than for the smaller ones.

Overall, the evidence suggests that transaction cost control are important in mean-variance portfolio revisions, and that it is worthwhile to use a transaction cost specification that includes price impact costs.

Chapter 4 Index fund management under transaction costs

To being able to address directly the question of the importance of transaction cost control in index fund management, I formulated two index fund revision models under transaction costs including price impact costs. The first model features an objective where transaction costs are balanced against a tracking error measure utilizing a trade-off parameter. In the second model, a tracking error measure is kept below a pre-specified limit, while transaction costs are minimized. Each model is integrated with the empirical transaction cost models developed in Chapter 2.

The importance of transaction costs and transaction cost control in index fund management were examined by a number of large-scale empirical tests using daily data. Previous research was extended by the development of a research design that incorporates transaction costs that are inclusive of price impact and variable across stocks and time, cash flows in the form of dividends, and corporate actions including mergers and acquisitions, new issues and other capitalization changes. The empirical investigation covers more than ten years of daily data and it is the first one to provide large-scale evidence in a setting that incorporates explicitly many of the factors that present the challenges to index fund management.

In the tests, the two index fund revision models and several alternative approaches, including full replication, are applied to track a Swedish capitalization-weighted stock
index. The approaches are implemented and tested under a number of variations including different tracking error measures and different types and degrees of transaction cost control.

Instead of using an extant index, an index is independently calculated according to a consistent methodology, mimicking that of the most well-known and used index in the Nordic region, the OMX index. This way, issues related to data availability and quality are resolved. The empirical findings indicate that accurate data indeed are important. That the index is based on a consistent methodology is desirable, for tracking performance is expected to be related to index methodology.

Comprehensive results are provided on tracking performance and transaction costs. Index funds implemented by the index fund revision models under transaction cost control dominate, in all dimensions of tracking performance considered, their counterparts implemented without transaction cost control as well as the funds implemented by full replication.

The results are overall indicative of that transaction costs are significant and that transaction cost control can substantially improve performance. For example, an index fund of typical size managed under control of transaction costs, had after a decade a net asset value that was 317 MSEK, or 5%, larger than that of an index fund constructed to meet the same tracking error standard deviation limit but without transaction cost control. Although adequate comparisons are difficult do, the tracking performance of the index funds implemented here does not appear to be inferior to that of an existing OMX index mutual fund.

As in Chapter 3, the results suggest that price impact costs are important to consider. Across all the index funds implemented, price impact costs constitute the major part of incurred transaction costs. Brokerage commissions were usually the least important part, and they never represented more than 10% of the total transaction costs. The realized transaction costs and trade sizes for trades in individual stocks were found reasonable.

Additional analyses highlighted other issues raised in the literature. The question of what tracking error measure to use is pointed up in prior research (see, e.g., Rudolf et al. 1999; Bamberg and Wagner 2000; Rey and Seiler 2001; Satchell and Hwang 2001;
Beasley et al. 2003). The results obtained this study indicate that several common tracking error measures perform similar. The concern expressed by Meade and Salkin (1989) and Pope and Yadav (1994) that estimates of the tracking error standard deviation based on short-interval returns could be biased, is supported by the results found here. There is thus a potential need for adjusting such estimates, which can be done by the methods in Campbell et al. (1997).

It has been common in research to construct index funds from a pre-defined subset of the index stocks (e.g., Rudd 1980; Larsen and Resnick 1998; Wagner 1998; Rudolf et al. 1999; Bamberg and Wagner 2000; Rey and Seiler 2001). The technique of constructing an index fund from a pre-defined subset of the most liquid index stocks was not found as an efficient means for transaction cost control. Connor and Leland (1995) demonstrate that a positive cash holding can reduce transaction costs for an index fund. That finding was corroborated here.

5.2 Overall conclusion
The answers to the two research questions raised intially are both in the affirmative. Yes, transaction costs matter, and yes, control of transaction costs improves portfolio performance. The results also suggest the importance of considering price impact costs in portfolio management. The research questions were studied within mean-variance and index fund revision problems, where the treatment of transaction costs were based on the premises that trading is uninformed, immediate, and executed in the electronic open limit order book system of the Stockholm Stock Exchange. Price impact cost models were estimated on data from one period. In the empirical tests, these price impact cost models were applied backwards in time. Reasonable commission rates were added to the price impact costs to obtain full transaction costs. In addition, transaction costs are related to fund size, that is, net asset value. The funds considered in the empirical tests were of sizes intended to be representative of typical Swedish mutual funds. The empirical results thus only apply under those premises and to the fund sizes considered.

5.3 Contributions
The following are considered the major contributions of the thesis.
5.3.1 Theoretical contributions

• Identification of the functional form of the price impact cost function in a limit order book. The form of this function appears not to have been recognized in the literature (see Glosten 1994 and Niemeyer and Sandás 1993).

• Extension of the standard mean-variance model by formulating a mean-variance revision model under a transaction cost including price impact costs.

• Formulation of two different index fund revision models under transaction costs including price impact costs.

Compared to the majority of prior studies, the mean-variance and index fund models developed here allow for more elaborate specifications of transaction costs without compromising significantly computational efficiency or ease of implementation.

5.3.2 Methodological contributions

• Development of a novel approach for cross-sectional modeling of price impact costs using public limit order book information. The usefulness of transaction cost estimates for individual stocks is emphasized in the literature (see, e.g., Keim and Madhavan 1998; Chalmers et al. 2000; Leinweber 2002; Bessembinder 2003). Also emphasized are the difficulties and costs associated with acquiring such estimates. For open limit order book markets, the approach presented here thus provides a solution to a recognized need.

• Based on the measure of portfolio proximity proposed by Dexter et al. (1980), a procedure was developed for analyses of how different transaction cost specifications and weight constraints, among other things, in mean-variance portfolio revision, impact central “portfolio analytic” quantities, such as expected utility, turnover, diversification, and transaction costs.

• Development of a research design and a “portfolio research system” that enable comprehensive analyses of the performances of index fund and other portfolio strategies in a setting that incorporates transaction costs, cash flows, and corporate actions, all on a daily basis. A research system of this kind is believed to represent a step towards more reliable and detailed analyses of financial processes that evolve over time, such as investment strategies. The higher daily data resolution is probable to better represent the actual window of opportunity for decision making.
Also, important economic factors, such as news events and dividend payments, to name a few, are likely to manifest themselves fully within single days. The interrelation between such effects and other economic factors, such as prices, will probably be easier to detect and study on a daily time-scale than on coarser time-scales when, e.g., price changes are likely to reflect the aggregate influence of many factors.

- Development of a program for calculation of benchmark indexes. This enables calculation of benchmark indexes according to consequent methodologies, which is potentially important, since the performance of a given tracking approach is expected to be related to index methodology. The approach also solves several issues related to data availability and quality. The empirical tests indicated that such issues, unless resolved, could have detrimental effects on index fund performance. The finding that accurate data is important has implications for the results obtained in the studies by, e.g., Walsh et al. (1998), Rudolf et al. (1999), Bamberg and Wagner (2001), and Beasley et al. (2003).

5.3.3 Empirical contributions

- Empirical models of price impact costs for Swedish stocks were developed and validated. Separate models were estimated for buy and sell price impact costs. The fairly parsimonious models performed satisfactory both in-sample and out-of-sample. This research contributes to the market microstructure literature by providing evidence on the magnitudes and determinants of price impact costs for a large cross-section of stocks trading in an electronic open limit order book market, namely the Stockholm Stock Exchange.

- Empirical evidence were provided on the importance of transaction costs and transaction cost control in mean-variance portfolio revision involving Swedish stocks. The evidence indicated that incurred transaction costs are non-trivial, and that control of transaction costs enhances performance in terms of portfolio utility, diversification, turnover, and the magnitude of transaction costs incurred. The results are suggestive of the significance of the price impact cost component of transaction costs. The mean-variance portfolio revision model formulated here and
that explicitly considers price impact costs, empirically outperformed revision models with less elaborate transaction cost specifications.

- Large-scale empirical evidence was generated on the tracking performance of a number of alternative approaches to index fund management. Detailed results were obtained on the importance of transaction costs and transaction cost control. The research design considered transaction costs including price impact, cash flows in the form of dividends and various corporate actions, all on a daily time-scale. Additional issues in the literature that were shed light on include how different tracking error measures affect tracking performance (Rudolf et al. 1999; Bamberg and Wagner 2000; Rey and Seiler 2001; Satchell and Hwang 2001; Beasley et al. 2003), the tracking performance of unconditional and conditional full replication, and the efficiency of pre-defined subsets as means to control transaction costs (Rudd 1980; Connor and Leland 1995; Larsen and Resnick 1998; Bamberg and Wagner 2000; Rey and Seiler 2001).

### 5.3.4 Integral contribution

The research in this thesis represents an effort to integrate portfolio management, trading, and transaction costs, much for the reasons, and much in the manner, described by Leinweber (2002):

“Reliable transaction cost forecasts can be applied ... in the often discussed but seldom-observed integration of portfolio management and trading. Most portfolio construction tools and optimizers use overly simple assumptions about transaction costs, e.g., they are the same for all stocks and are expressed as so many cents per share or a fixed percentage of the order size. These assumptions are clearly at variance with the real world. The portfolios that would emerge from these systems if they incorporated more realistic estimates of transaction costs are very different from the portfolios produced under the naïve assumptions.” Leinweber (2002, p. 5, also see 1995)

The elements researched and developed in this thesis constitute when integrated an almost fully computerized approach to portfolio and index fund management. With regard to index fund management, apart from historical returns, all the input needed is a list on tomorrow’s index composition and rudimentary information about corporate actions and cash flows. It would be possible to generate trades on basis of real-time information in a public limit order book system and according to the indexing approach and tracking performance currently preferred. For backtesting of investment
strategies, the estimated cross-sectional price impact cost models should be useful. If real-time information on transaction costs are unavailable, they could also be useful in the implementation of strategies.

The indexing study features a computerized portfolio research system that can be employed in research on trading strategies and other issues, where it is potentially important to consider daily portfolio alterations in the presence of transaction costs, dividends, and corporate actions. As an example, the system enables the calculation of so-called factor portfolios that are net of transaction costs. Because the magnitude of transaction costs incurred is related to the scale (the net asset value) of the fund, these factor portfolios need to be derived specifically for any given fund size. Constraints, such as restrictions on short-selling, could also be imposed on the factor portfolios. I believe that such scale-dependent, net-of-transaction costs factor portfolios can be potentially interesting in risk-adjusted performance measurement. Finally, the usefulness was demonstrated of the “portfolio research system” as a laboratory in which idealized conditions can be furnished to satisfy different research needs.

5.4 Practical implications

The empirical tests considered portfolios of sizes typical for, or deemed to be large enough to be economically interesting to, fund management companies. The results indicated that transaction costs in general are important and that transaction cost control can considerably improve performance, both in mean-variance and index fund management. In mean-variance portfolio management, portfolios revised with no control of transaction costs suffered certainty equivalent losses of 6-20% relative to portfolios revised under transaction cost control. Given the net asset values of the portfolios analyzed, these certainty equivalent losses correspond to cash losses in the range 34-370 MSEK.

For a typical index fund in terms of net asset value and tracking performance, control of transaction cost reduced transaction costs incurred by 54 MSEK, or 1/3, over a ten-year period non-compounded. A measure that takes compounding into account is the ending net asset value. After ten years, the net asset value of the fund managed under transaction cost control was 317 MSEK, or 5%, larger than that of a
comparable fund managed without transaction cost control. The cost savings could be used for the benefit of the fund management company or the fund investors, or both.

The public reporting by mutual fund of their trading costs is confined to brokerage commissions. Commissions were generally found as the least important part of the transaction costs incurred. There is thus room for improvement in the transaction cost reporting practices of mutual funds.

5.5 Future research

Throughout the thesis suggestions for further research have been made. In this section, the most important ones will be reviewed and, in some cases, expanded upon. Some new material will also be introduced.

As to the modeling of price impact costs in Chapter 2, it seems worthwhile to collect more limit order book snapshots and over a longer period. It would then be possible to analyze possible systematic variation over time and the effects from market-wide conditions (commonalities) on individual stocks’ price impact costs. In fact, work to retrieve more limit order book data has begun.

The empirical analysis of the importance and transaction costs and transaction cost control in mean-variance portfolio management performed in Chapter 3, could be extended in several ways. A natural extension would be to investigate the performance over time for mean-variance portfolios revised with varying levels of transaction cost control. The adequacy of the transaction cost amortization scheme used, and that was adopted from Rudd and Clasing (1982) and Grinold and Kahn (1995), should be examined. Other utility functions, possibly from the class of isoelastic (or power) functions, such as the logarithmic, should be considered. This class of utility functions exhibits - at least in the absence of transaction costs - several noteworthy properties, including that they are myopic, i.e., the investor needs only consider the coming period’s return distributions, as the return distributions for periods beyond the coming do not affect the current period’s optimal decision (Mossin 1968). I have for instance implemented the transaction costs specification used here for logarithmitic utility and so-called empirical probability assessment approach, an approach used by, e.g., Grauer and Hakansson (1993).
Inherent in all these extensions is the question of how well the mean-variance model (which strictly is a single-period model, or an approximation to portfolio models based on exact utility functions) and models based on exact utility functions, perform in a multi-period setting under transaction costs. Investigation of these issues would seem to call for the formulation and empirical testing of multi-period investment models that consider different utility functions, transaction costs, and cash flows. However, solving such multi-period models for more than a few securities and periods and under realistic return distributions, is computationally very hard. Moreover, the number of periods to consider is not always obvious.

The empirical investigation of index fund management in Chapter 4 did not consider investor flows, which, as Edelen (1999) demonstrates, is a factor that has explanatory power with regard to the underperformance of actively managed funds. To extend the analyses in Chapter 4 to include effects of investor flows on index fund management and performance, should be an interesting opportunity for future research. This, of course, requires that such data could be accessed. Moreover, the tests in Chapter 4 could be extended to reflect the performance available to an index fund investor by deducting a management fee and other expenses from the assets of the fund.


References


References


