Revisiting Index Methodology for Thinly Traded Stock Market

Case: Helsinki Stock Exchange

Mika Vaihekoski*

November 10, 2020

Abstract

Stock market indices play a central role in portfolio management and academic research. This paper reviews and discusses the main issues in index construction, especially on thinly traded stock markets and in a historical setting with deficiency of information. The main methods to deal with missing price observations are studied. As a case in point, a newly collected historical database for the Finnish stock market that covers the period from the establishment of the Helsinki Stock Exchange (HSE) in October 1912 forward is used. The HSE suffered from severe thin trading with only approximately 20% of the stocks having a daily transaction in the early part of the sample. Overall, the results show that index construction methodology have a major impact on the index as well as its statistical properties. The results also highlight the impact of corporate capital actions, often the hardest information to obtain, on the market index performance.

Keywords: index methodology, stock market index, thin trading, financial history, Helsinki Stock Exchange, cliometrics

JEL classification: F3, G12, G15.

*Mika Vaihekoski is with the Turku School of Economics, University of Turku. Address: Turku School of Economics, Department of Accounting and Finance, 20014 University of Turku, Finland. E-mail: mika.vaihekoski@utu.fi. This research has been supported by OP Bank Research Foundation, the Foundation for Advancement of Finnish Securities Market (Suomen Arvopaperimarkkinoiden Edistämissäätiö), and Nasdaq Nordic Foundation.

1. INTRODUCTION

Stock market indices play a central role in portfolio management and academic research. For portfolio managers, the indices play a central role as benchmarks and key measures of market development. For risk analysts, it is important to have long enough market indices to get more accurate measures of the true distribution of the returns. For the wider public, the market indices aggregate information in a convenient way. For the academic research, the indices are a central ingredient in many asset pricing studies. Although modern index methodologies are typically documented, especially those calculated by professional organizations (c.f., e.g., CRSP (2020), MSCI (2020)), the discussion on their foundations leaves room for improvement. As such, surprisingly little have been written on how minor changes to the index methodology can have a major impact on the outcome.

Moreover, there is also a growing interest in the historical development of financial markets and stock markets in particular. As a result, a number researchers[1] and organizations[2] have worked to create historical stock market indices for different markets. Creating long-term historical stock market indices does not go without problems, however. The availability, breath, and accuracy of information typically decreases the further back in history one goes. In addition, stock markets typically become increasingly thinly traded. Studies on index construction for thinly traded stock markets with missing price observations are scarce. Most of the studies have focused on how to deal with the outcome – an index with serial correlation induced by thin trading (c.f., e.g., Jokivuolle, 1995). The aim of this study is to discuss the main issues in index construction with a focus on methods to deal missing price observations due to infrequent (thin) trading.

The Helsinki Stock Exchange (HSE) in Finland and the time period from its opening in October 1912 forward to December 1981 is used as an example. The HSE is a stock market

[1] E.g., Dimson, Marsh, and Staunton (2002), Annaert, Buelens, and De Ceuster (2012); in Nordic countries, see, e.g. Frennberg and Hansson (1992), Nielsen and Risager (2001), Nyberg and Vaihekoski (2010), Rydqvist (2019), and Waldenström (2007).

[2] E.g., the work done over the years at the Center for Research in Securities Prices LLC ("CRSP") and EUROFIDAI just to mention only a few.

well-known to suffer from thin trading during this sample period (Berglund and Liljeblom (1988)). In the early part of the sample, thin trading was severe as approximately only 20% of the stocks had a daily transaction. Towards the end of the sample period, the situation improved a bit as more than 70% of the stocks traded daily. This paper also continues the work set forth in Nyberg and Vaihekoski (2010). They collected a database on Finnish stock market and used it to create a monthly all-share value-weighted total return stock market index for Finland that, when combined with available similar index series from 1970 to the present day, covered the entire history of the Finnish stock market.

However, the data collected by Nyberg and Vaihekoski (2010) and thus their index had some shortcomings that they acknowledge in their study. First, they only had access to bid offers from a secondary source (newspapers), not transaction prices and not from a primary source. Furthermore, the ex dividend month was unknown – and thus they assumed it to be April. For this study, a new database is collected using the official price quotation lists by the Stock Exchange. This allows one to avoid potential mistakes in the secondary source, and to collect, not only the bid offers, but also the ask offers as well as the actual realized prices which can be used to deal with the thin trading in index creation. In addition, it turned out that the Stock Exchange has also kept track of dividend payments as well as of shares' nominal values together with their minumum trading lots. This is a huge improvement in data quality that allows one also to consider alternative ways to construct the stock index.

Overall, the results show that the compound monthly growth rate (geometric mean) for the sample period is 1.03 per cent (13.16% p.a.) using the baseline index. The result is similar to that of Nyberg and Vaihekoski for 1912–1970 and Berglund et al. (1983) for the 1970–1981 periods. It is also evident that it is of major importance to collect information also on corporate actions even if it is often the most difficult information to acquire from historical sources. In Finland, the bonus and rights issues were frequently used by the companies during the sample period, and as such they make a significant contribution to investors' return. Compounded monthly percentage return would have been 0.37 per cent

(4.71 % p.a.) lower had one ignored them in the index construction.

The remainder of the paper is as follows. Section 2 provides first a review of methods to calculate the stock market index after which the issues on thinly traded stock market are discussed. Section 3 presents the market used as an example in this study, Finland. A short review of the trading in Helsinki Stock Exchange as well as review the data collection procedure is provided. Section 4 discusses the characteristics of thin trading in the HSE and finally demonstrates how different aspects of index methodology impacts the outcome. Section 5 concludes and offers suggestions for future research.

2. STOCK MARKET INDEX METHODOLOGY

2.1. Basic methodology

One can calculate stock market indices in a number of different ways. The main differences in index construction are related to the choice of companies to include in the index, handling of dividends and corporate actions as well as in the weights given to companies in the index. Historically, the main interest has been to use the index to measure the return of a passive market portfolio, but more recently a number of benchmark indices has been created to measure the return of an active investment strategy. Although certain recommendation for the index calculation methodology can be given, in practice there are still certain room for subjective judgement which makes the issue all the more interesting.

Typically the goal of a stock market index is to measure the return of a buy-and-hold investor. As such, we can develop some criteria for a good index. The index should measure the accumulation of investor's wealth if one assumes that there are no transaction costs of taxes involved in rebalancing the index (e.g., adding a new listed security into the index implies selling a certain portfion of the original securities and investing the proceeds into the new one). Moreover, the buy-and-hold index should reflect the returns of a passive investor. Thus, the tracking accuracy of the index of the wealth of a buy-and-hold investor,

can used as a benchmark for validity of the index methodology. Moreover, the index should be investable, all money should be invested in the stocks all the time (no side-portfoliso), and the index should not suffer from different biases, such as lookahead bias or survivorship bias.

Probably the most generic formula to calculate a stock market index is as follows

$$I_{m,t} = I_{m,t-1} \times \frac{\sum_{i=1}^{N_{t-1}} w_{i,t-1}(P_{i,t} + D_{i,t} + f^a_{i,t}) \times f^m_{i,t}}{\sum_{i=1}^{N_{t-1}} w_{i,t-1} P_{i,t-1}}, \tag{1}$$

where $I_{m,t}$ is the index value at time t, N_{t-1} is the number of different stock series included in the index at time $t-1$, $w_{i,t-1}$ is the weight for each stock i in the index at time $t-1$, $P_{i,t-1}$ is the price of the stock i at time $t-1$, $P_{i,t}$ is the ex-dividend and corporate action price of the stock i at time t, $f^a_{i,t}$ and $f^m_{i,t}$ are additive and multiplicative adjustment factors for corporate actions for stock i respectively, and $D_{i,t}$ is the (cash) dividend paid for stock i at time t. Note that in theory index constituents should be selected at the beginning of the investment period (time $t-1$). This removes survivorship as well as ex post selection bias, and it makes the index replicable – hence N_{t-1} is used instead of N_t. In practice, however, one often makes an exception with the delisted companies and removes them one period before actual delisting.

Defining a new relative weight variable $w^s_{i,t-1} = (w_{i,t-1}P_{i,t-1}) / \sum_{i=1}^{N_{t-1}} (w_{i,t-1}P_{i,t-1})$, we can rewrite equation (1) using the following return notation

$$I_{m,t} = I_{m,t-1} \times (1 + \sum_{i=1}^{N_{t-1}} w^s_{i,t-1} R_{i,t}), \tag{2}$$

where $R_{i,t}$ is the familiar *percentage* return for asset i from time $t-1$ to time t, adjusted for corporate actions.

A number of different ways to weight companies (or stock series) in the index exist. Both equations (1) and (2) can easily accommodate most of them although equation (2) is often more convenient to work with. If we set all $w_{i,t-1} = 1$, the end results is a price weighted

index such as the well-known Dow Jones Industrial Average index. Setting $w^s_{i,t-1} = 1/N_{t-1}$ gives us the equally weighted index.[3] Note that the equal weighting implicitly entails frequent trading – one needs to sell stocks with above average returns to buy those with returns below average to keeps the weights equal.

If we define the weights in equation (1) in terms of number of stocks, we get the equation used in CRSP (2020) and MSCI (2020) although with some minor differences. Setting $w_{i,t-1} = q_{i,t-1}$, where $q_{i,t-1}$ is the total number of shares available for stock series i at time $t-1$, we can define a new term $MCAP_{i,t-1} = w_{i,t-1}P_{i,t-1}$ which represents market capitalization value of the series. The end result represents the commonly used market capitalization weighted index (market value-weighted index).[4]

The value-weighted index is probably the most commonly used index in the academic literature as it is convenient to calculate and it measures the buy and hold return of the stock market portfolio. As such, it is also widely used as a benchmark and followed by the public. The use of market capitalization weights has also certain implications regarding investors' behavior. Namely, it implicitly implies that the dividends are reinvested in the index constituents in proportion to their respective weights – not on the dividend paying stock. This is easy to see with an example.

Assume that we have two companies, A and B, in our index. For convenience, both stocks are originally at time $t = 1$ valued at 100 and both companies have the same number of shares issued. During the first period, company A pays a dividend of 50 and the ex-dividend price closes at 75. The second period has A's price going up 10 per cent. Company B's stock trades put at all times. Using equation (2), the index value at time $t = 2$ is $100 \times (1 + 50\% \cdot 25\% + 50\% \cdot 0\%) = 112.50$. Similarly, at time $t = 3$, the index value is $112.50 \times (1 + 75/(75 + 100) \cdot 10\% + 100/(75 + 100) \cdot 0\%) = 117.32$. Now, these values should correspond to the wealth of an investor who owns one of both stocks and who invests the

[3]Note that naming convention for index weighting is based on selected relative weights. Equal absolute weights, $w_{i,t-1} = 1$, imply price-based relative weights $w^s_{i,t-1} = P_{i,t-1}/\sum_{i=1}^{N_{t-1}} P_{i,t-1}$.

[4]A volume-weighted index can be created similarly by replacing number of shares with trading volume.

dividend back to the market, i.e. A and B stocks, after receiving the dividend during the first period. Starting with the original wealth of 200, at time $t = 2$, before dividend reinvestment, investor's wealth is 75 (A share) + 100 (B share) + 50 (cash), but after using 75/175 parts of the dividend to buy A shares and the rest to buy B shares, her wealth becomes 96.43 (1.29 A shares) + 128.57 (1.29 B shares) = 225 which equals that shown by the index. Similarly, at time $t = 3$, her wealth becomes $1.29 \times 75 \cdot (1 + 10\%) + 1.29 \times 100 = 234.64$ which equals $225 \times (117.32/112.50)$. The assumption that the dividend is invested back to the market is mathematically convenient. If we, on the other hand, assume that the dividend is invested back to the dividend paying stock, the weights would depend on the historical dividend payments in an increasing complex manner.[5]

We can also utilize a number of other weighting schemas and/or firm filtering approaches. For example, we can set an upper limit on all weights to create a capped index. This leads to cascading weight calculation. On the other hand, we can filter away companies that do not meet certain criteria or restrict the market coverage in some other way by by using an upper limit on N_{t-1}. We may, for example, want to focus on the largest 100 companies, or to filter away companies which do not meet certain requirements on trading volume, free float, and other characteristics of investability. We can also contruct an index to track certain active investment strategies. The constituents of the index and the weight of the stocks can also be rebalanced less frequently (e.g. quarterly) than the index values.

NASDAQ (2013) uses the following equation to calculate their all-share total return indexes for the Nordic Stock markets, including the Helsinki Stock Exchange (modified to reflect the notation used here)

$$I_{m,t} = \frac{\sum_{i=1}^{N_t} q_{i,t} P_{i,t}}{\sum_{i=1}^{N_t} q_{i,t}(P_{i,t-1} - d_{i,t}) \times f_{i,t}} \times I_{m,t-1}, \tag{3}$$

[5] In the example, if the investor had invested the dividend back to the dividend paying asset, her wealth at time $t = 2$ would be 125 (1.67 A shares) + 100 (1 B share). At time $t = 3$, her wealth would be $1.67 \cdot 75 \cdot (1+10\%)+100 = 237.50$. Similarly, the index at time $t = 3$ would equal $112.50 \times (1+125/225 \cdot 10\%+ 100/225 \cdot 0\%) = 118.75$ which indicates the correct wealth development i.e. $225 \times (122.14/112.50) = 237.50$.

where $q_{i,t}$ is the number of shares of company i applied in the index at time t.

The main difference between equation (3) and (1) is in the handling of the dividends. Equation (3) adjusts the previous closing prices – a method sometimes known as a start-of-day price adjustment – whereas equation (1) adjusts the end-of-day prices. The somewhat counter-intuitive naming becomes clear with the following example, which also shows that, although the difference might seem like a minor one, it is not always the case. Assuming again that we have a company with a stock priced at 100. The next day, the company pays a dividend of 50 and the ex-dividend price ends up at 75. Now, assuming that this is the only security included in the index, the index value at time $t+1$ would be $100 \times 75/(100-50) = 150$ using equation (3) and $100 \times (75 + 50)/100 = 125$ using using equation (1). In fact, the former method will always imply higher return for as long as the dividend is positive.

The difference comes down to the assuption made regarding the timing of the reinvestment back into the market. Equation (1) assumes that the dividend is reinvested in the market at the end of the ex-dividend period using available closing prices. This is totally executable transaction provided that the stocks trade.[6] Equation (3), on the other hand, assumes that one is able to reinvest the dividend at the beginning of the period using the dividend adjusted closing prices from the previous period. Alternatively put, equation (3) assumes that the closing prices from the previous period, corrected for dividend, equal opening prices and the stocks trade at that price, but in reality this is usually not the case. In fact, this feature can make index tracking quite difficult.

On the other hand, if we are creating a time-aggregated index, such as a monthly index, assuming that dividend are always paid at the end of the period can bias the index, if the dividends are paid, on average, earlier in the month. Namely, had the dividends been invested on the market when paid, the investor would have earned the market return on the dividend. Without adjustment, the index excludes the effect of this intra-month return. A smaller issue is the time discrepancy between the dividend payment and its receipting.

[6]Obviously, if the stock exchange is closed, one cannot trade on the market. Now, the dividend and other adjustments have to be postponed until the stock exchange is open again.

Historically, the owners did not receive the dividend exactly the ex-dividend day, but few days later. In practice, however, researchers have often assumed that the effects of both on the return are negligible and thus they can be ignored.

Another source of variation in the index methodologies are the adjustments due to other corporate actions, such as (reverse) splits, company consolidations, stock dividends (bonus issues), capital repayments, and cash rights issues, just to name a few. On a broad scale, there are two types of adjustment factors. The adjustment factor can be expressed in multiplication form as in equation (3) or in can be broken down for convenience to multiplicative and addivite terms as in (1). In most cases, this is simply a matter of taste as $P_{i,t-1} + f_{i,t} = P_{i,t-1} f'_{i,t}$ where $f'_{i,t} = 1 + f_{i,t}/P_{i,t-1}$.

The multiplicative adjustment is often convenient when the number of shares is affected by the action. Typical examples include (reverse) splits and bonus issues. The additive adjustment is convenient when the action provides some additional value for the holder of a stock in question. This is the case typically in cash rights issue or when shares of carved-off subsidiare are distributed to the old owners. In the latter case, one needs to calculate the cash value of the corporate action.

The calculation of the actual adjustment factor is rather straight forward in most cases and one can review the methods in the literature (c.f., e.g. NASDAQ (2013)). For (reverse) splits and some other actions, the adjustment is rather simple to derive. However, for some actions, the adjustment factor might be harder to derive, especially if the action can be considered to have option like features.[7] As the number of different actions that can take place is vast, and sometimes they take place simultaneously, it is not possible to review different adjustment factors here. Instead, one is referred e.g. to NASDAQ (2013) for a list of possible actions and adherent adjustments. Note that one can avoid calculating the adjustment factor if one can observe its value on the market. For example, if a company

[7] A typical example is a rights issue. The rights should be considered as options and they should be valued correspondingly. The situation become more complicated with "dirty issues" where the rights are written to an unlisted stock series or some other security without a proper market price.

organizes a rights offering, rights can be traded on the market. However, this is not always the case on all markets and the longer back in history one goes.

However, there is one issue that one needs to consider, especially with cash rights offerings, which are commonly used method for companies to raise new equity capital in many markets. The issue is related to the consequences of the investor's action on the index weights. It would be convenient, if we could follow the same principle as with the dividends. Namely, simply assume that the right(s) are sold and the proceeds are invested with the market weights. However, this is not the case as the next example shows.

Assume again that we have two stocks, A and B, in our index. Both stocks are valued at 100 at time $t = 1$ and both companies have one share issued. The next period, company B organizes a cash rights offering giving the old owners of the stock a right to buy one new share for 50 for each share they own. Assume for simplicity that there are no other news regarding stock B, and the new value of the company is 100 (old value) + 50 (newly raised capital), which is now divided between two stocks, priced at 75. Assuming that the investor can sell the right for its exercise value 25, the adjusted return for stock B is easy to calculate (0%). However, the main question remains: what happens to the cash received from the sell? Assuming that the price of stock A stays put, the post-issue total market capitalization value is 250 which is divided 40%/60% to stocks A and B. Since the investor's total wealth remains at 200 (100+75+25), she must invest 80 to A and 120 to B to have her portfolio reflect the market weights at the end of period 2. Now, if she invests the cash (25) received from selling the right to A and B according to the market weights, she cannot attain the correct weighting. Even investing all to stock B is not enough – she has to trade on the stocks (sell part of stock A and invest the proceedings and 25 to stock B). The end result is doable, but one cannot say that the cash received from selling the rights is invested with the market weights (as we did with the dividends).

What happens if the investor participates in the offering and invests 50 to buy a new stock B? To finance the purchase, she can sell part of stock B (50/75 raises the required

50), and end up with 1.33 shares of stock B. Her portfolio's weights would remain at 50-50% which again requires her to trade to reach the 40/60% weighting. If she, on the other hand, decides to sell part of her portfolio with market weights to finance the purchase, she will sell 0.2 parts of stock A (raising 20 = 40% times 50) and 0.4 parts of stock B (raising 30 = 60% times 50). As a result, her portfolio will consist of 0.8 shares of stock A and 1.6 shares of stock B, valued at 80 and 120 which reflect the market weights. So, ultimately, all three approaches will yield the same result indexwise, but perhaps the last approach (the investors are assumed to "participate in rights offerings, sell the market") is most consistent with the "raise the dividends, invest them in the market" assumption.

Adjustments for the corporate actions face the same issues as the handling of dividends, i.e., the adjustment either on the post-action (end-of-period) price observation in the numerator as in equation (1) or on the previous pre-action (beginning-of-period) price observation in the denominator as in equation (3). The latter approach is used e.g. by Nasdaq (2013) and CRSP (2020). For (reverse) splits issues, the choice does not make a difference given that in the (3) one uses the number of shares post-split ($q_{i,t}$) in the equation. However, for some corporate actions the choice makes a difference. The difference corresponds to the timing of the dividends discussed earlier.[8] As a final comment, the discussion concerning the intra-period timing of the dividends and its effect on the return ensues for the corporate actions as well as since the investor can utilize the right as soon as it becomes available and not at the end of the aggregation period.

In practice, any index methodology always leaves some room for subjective judgement, especially on the following five issues. First, what companies and series should be included in the index and with what market cap? For example, one may want to exclude some securities

[8]To see why, assume a company with two stocks, each priced at 100, that issues new stocks to the old owners of the company so that for every two owned, they can buy a new (post-split) stock for 40. At the same time the company splits the stock with a two-for-one ratio. The value of the company post-issue is 240. If we divide it with the value with the number of post-split stocks (4+1), the price of each stock should be 48. Obviously the value of a right to buy shares for 40 is at least four. Now, imagine an investor who owns these two stocks. She can sell the rights for 8 (or whatever price is observed in the market) and invest the money back to the stock in question or back to the index. Thus the case is similar to that of dividends and the earlier discussion on the timing of the adjustment enteils.

from the index (e.g. certificates, depositary receipts, ETFs, mutual fund type of holding companies, non-corporations, or companies whose majority is owned by the government). Historically, it is not always easy to find enough information on the nature of the assets to make the decision. Another issue are caused by the dual-listed (or higher) companies – dual-listing can happen in different contries and/or regional stock exchances. Should they be included on both market indices or only on the one where to company's domicile is? And should we use full market capitalization for both indices or divide it somehow? If a company is dual-listed on a second market and for that purpose the company has issued certain proportion of stocks to the second market, one can utilize them to calculate the market cap for the second market, especially if the cross-market transactions are made difficult (c.f., e.g. ADRs).

Second, some companies might have multiple stock series/classes some of which may not be listed. There are different ways to proceed. In a company centric approach, we focus on companies and use only one entry for the company in the index. In the stock series centric approach, all series are treated separately in the index. If some of the series are unlisted, one can simply ignore them altogether or take them into account, typically as if it has the same price as the listed series.[9] In practice, this simply adds on to the market values of the company. If two (or more) stock classes/series are listed, one can treat them as separate items, or one can combine them using some kind of criteria to choose the series to represent the company. The latter alternative is often justifiable when the other series, perhaps newly issued, temporarily differs from the other series e.g. due to lower dividend right.

Third, one has to decide how to deal with changes in the company or series name/code and the reason behind them. The main issue is whether one should consider a new stock series a new entity without history (and lose one observation) or a continuation of the old series. It is very common for the companies to change their names throughout their history for a variety of reasons. Now, if the company simply changes its name to a new one, it is

[9]Typically the unlisted stock series has more voting rights (c.f., e.g. Facebook) and hence its values should reflect this with some voting premia.

often best to consider it as one entity (series). However, if the firm name is changed because of a combination merger of two companies, the choice is often to take it as a new entity (say, merger of Exxon and Mobile into ExxonMobile). The gray area is often with mergers where the name of the either company is kept or with acquisitions where the acquired company is merged with the acquirer, but the end results might have a totally new name, sometimes even that of the acquired company. Again, the longer into history one goes, the harder it is to find information on the true nature of transactions.

Fourth, one may be forced to consider including stocks into the index whose prices are not for the same trading day. Historically it was not uncommon that some stocks were traded less frequently (say once a week) than other stocks. In addition, at times, certain (groups of) stocks could even trade on alternate days. If one is interested in creating, say, a monthly index with month-end values, one has to deal with this trade-off – either to collect last available month-end values for each series and accept observations that are slightly off month-end or to collect only observations on the last day of the month when all (or most) of stocks are traded.

Finally, the handling of delisting returns. There are various reasons for companies being removed from the stock exchange listing. In some cases, calculation of the delisting returns are quite easy. For example, if there is a public puchase offer, one can utilize the purchase price to calculate the return. However, there are still many cases, where one cannot calculate the return at the time of the delisting. For example, in a bankruptcy, the return might initially seem like -100%, but the investors might later recover some cents on the dollar. Historically, yet more difficult issue to deal with, are delistings, where the company simply disappears from the official quotation list. This might happen at the request of the company (the company might have been listed against its own will in the first place) or the stock exchange may have decided to stop quoting the stock if there is not enough trading interest. In the past, there wasn't necessariyl a requirement to purchase remaining shareholders out.

There are different ways to deal with the delisting returns. One approach is to delete the

delisting period returns altogether, even for those for which they are known. In practice this is simple to do, as it reduces the information requirement and it is easy to do in the index calculation. Acharya and Pedersen (2005), on the other hand, use either the last return available on CRSP, or the delisting return, if available. If not, they assign a return of -30% if the deletion reason is unavailable or bankruptcy, among others, since Shumway (1997) obtains that -30% is the average delisting return.

2.2. Index calculation on thinly traded stock markets

There are different ways to deal with thin trading when calculating a stock market index. The main issue with thin trading is the occasional lack of price observations due to the stocks trading infrequently. Obvisously the issue is more pronounced when one is creating a historical market index and especially for markets where the trading activity is lower to begin with. Having said this, one should not forget that even on present day and larger markets some stocks may trade infrequently. A related issue is the lack of information when one is creating historical indices – we might have only access to parts of the information that one would in an ideal world use to create the index. The question comes down to what should we do in these situations. To some degree the answer depends on the goal we are trying to achieve. For example, if we want to see the overall development of the market, we can settle for imputed price observations, but if we want to focus on the temporal behavior of the investors' returns, then we should focus on realized trade prices.

Creating a value-weighted index is as such a step towards dealing with the issue as it gives lower weight on smaller companies which are typically the ones that suffer the most from thin trading. One could argue that a volume-weighted index could be even better way to deal with thin trading as it would give more weight on those stocks that trade. However, it would not measure the returns of a passive buy-and-hold investors as it would require frequent trading to keep the weights in balance. The value-weighted index, on the other hand, closely tracks the returns on a passive strategy – the amount of trading is minimized and as such

it is well received idea on thinly traded stock markets. The downside of a value-weighted index is that it sets a new requirement in terms of data – one needs to collect information on the number of shares issued. This information is not typically as easily available as the trading data, especially from historical sources. Often one needs to calculate the values using accounting information on book equity and nominal values. Moreover, even if the required information is available, it is often available on an end-of-year basis. For the highest quality data, one needs to track down the intra-year effects of corporate capital operations (splits, issues) on it as well. At times this might be impossible and one has to settle for the lower quality information.

If one decides to create a value-weighted index and one has access to information on the number of shares issued, there are two main questions one need to answer. The first question is related to whether or not to exclude certain companies from the index. As said, we are often interested in creating an all-share survivorship and lookahead bias free index. Thus, including all shares at all times, as soon as they are listed and priced on the market, into our index. In practice, this index is simple to create, and it maximizes the number of assets in the index as well as the use of information. It also has a number of natural uses in practice and in academic research. An all-share index is also a natural choice if the number of listed stocks is limited – often the case on smaller stock markets – and as such one often aims to avoid throwing away data if not absolutely necessary.[10]

The downside of the all-share index is higher impact coming from thin trading. If thin trading is severe, one may choose to focus on more frequently traded stocks instead. It is somewhat of a trade-off – if the selected criteria for index inclusion is too tight, the end result may not represent the true market portfolio anymore. In the end, if we do decide to exclude thinly traded stocks from the index, we need to decide on a criteria to select stock series into the index and to remove them when they have become illiquid. A good criterion should not cause look-ahead bias into the index, i.e., we should not select (deselect) series

[10] A related issue is the question of how often to update index constituents. Namely, less frequently updating (say, once a year) leads to wasting some observations.

that we *ex post* know to be liquid (or illiquid). A natural and investable criteria is to include (exclude) stocks that have been found to be liquid (illiquid) within certain evaluation period in the past (e.g., "select stocks that trade more than half of the days with the past year"). In addition, we need to decide whether the index constituents are updated on a rolling basis or periodically which is often the most convenient way in practice. Moreover, the rules could be set to vary over the time to avoid e.g. situation where it is too stringent (i.e., none of the stocks meet the set criteria) or the opposite (all stocks meet the criteria) if the overall market liquidity changes.

Obviously, focusing only on the most liquid stocks has some advantages but drawbacks as well. On the plus side, if the index is based mostly on realized prices, it is hard to argue that the index does not measure the return collected by a buy-and-hold investor. Of course, at times the prices can reflect smaller trades which does not represent truly executable (larger) trades or which can be subject to market manipulation, but these issues are most likely more serious for thinly traded stocks. On the negative side, if the aim is to produce an all-share index reflecting the overall development of small and large companies, it is questionable if a liquidity filtered index fulfills this goal.

The second question is related to what should we do about missing trade prices. An intuitively appealing method would be to simply exclude those assets from the index for the time period when we do not have returns for them due to lack of price observations. Thus, the number of stocks included in the index and the wealth covered by the index would vary from one period to the next. However, this seemingly innocent decision implicitly equals to selling those stocks that are not included in the index, investing the proceeds to other stocks (market) at the same time, and buying them back as soon as they are again traded. Not only will it lead to unrealistic trading, but it also assumes that the investor is somehow able to foresee the (il)liquidity one period before it happens creating huge amounts of lookahead bias. This method also excludes the realized return over the non-trading period on assets that ultimately trade from the index.

As the observation omission strategy clearly has major drawbacks, better methods have been developed to deal with the missing price observations. The three main methods can be labelled as the search back method, the price imputation method, and the return imputation method, all with their own advantages and disadvantages. As a researcher, one is not always free to choose the best method as it depends on the availability of the data. If we are constructing, say, a monthly index and we have only access to month-end observations, the search back method is not available. On the other hand, if we do have access to daily transaction prices, but not to bid and ask offer prices, the use of the imputation method is limited. If enough information is available, one can also apply both methods. The choice of the best method(s) depends on the characteristics of the market data in question. There is no single right solution, and one has to analyze the data before making the decision.

The search back method is based on substituting earlier values for the missing values. If one is creating, say, a monthly index, the used price observation is the one observed on the selected cut-off date (typically the end of month). If the value on that particular day is unavailable, the search back method uses the observation from the previous day, and if it is also missing, the one from the day before. The obvious question is how far back in time should one search for an observation. If the found observation is from the beginning of the month, the return does not measure the true return for the month.

We can also choose to impute missing price observations. Several different imputation methods exist. The first approach, and probably the most commonly used method, is to impute the prices series with the mid-point of the bid and ask offers, if they are both available (see, e.g., Dimson et al., 2016). This approach has the drawback that the bid offers are often more realistic and they reflect market situations quicker. It is not uncommon for the brokers to keep the asking price the same for weeks on end even if overall price level has dropped. Another drawback is the fact that for most stock markets the ask offer is missing more often and it limits the number of imputations one can make.

If either of the offers is missing, one can choose from three different alternatives. First,

one can settle on using only the mid-points, and ignore all other information. This will severely limit the number of observations. Second, one can use either offer price all the time. Typically one uses bid offers to impute the price obsevations as the bid offers are most often available. The second alternative leads to considerable higher number of observations that the first alternative. However, the returns suffer from the bid-ask bounce. As a result, sometimes researchers have decided to avoid mixing realized prices and bid offers, and they have decided to use bid offers only.

However, there are reasons to consider also using the ask price offers at times to augment the bid prices.[11] This leads to the question of the nonlinear usege of ask offers in the imputation. Say, we observe a bid offer of 100 at time t and the next period we observe only an ask offer of 90 (110). Should we utilize the fact that in some sense the market value of the stock has decreased (increased) by 10 percent even for a buy-and-hold investors, or not? The answer seems to be different for a decrease and for an increase. For the decrease, one is more likely to accept that the value has really dropped, whereas the increase is less likely to represent an acceptable increase in the value.

The third alternative is to mix both mid-points and bid (or ask) offers when they are available. If one decides to mix mid-point of offers with bid (ask) offer observations to maximize the number of price observations, a related question is whether one should adjust the bid (ask) offer by (half of) some historical average bid-ask spread to reduce the bid-ask bounce. To demonstrate this, assume that the bid and ask offers at time t are 90 and 110. The mid-point of offers is thus 100. The next period, there are still no trades, and not even an ask offer, only a bid offer of 95. Without adjustment, mixing the observations as such, the measured return would be -5%, but assuming a historical bid-ask spread of 20%, the adjusted bid offer would be $(1 + 0.2/2) \cdot 95 = 104.50$. This would lead to a return of +4.5%. This method will maximize the number of observations, but to use it, one needs to solve a practical question – over what period should one estimate the historical bid-ask spread?

[11] We can also consider using ask offers in return imputation. If we observe only ask offers in consequtive periods, we can use them to calculate imputed returns using their values.

Now the question is which replacement value should we use? In some sence, if we have access to real trade prices and the number of missing observations is not high, it is advicable to use them, potentially utilizing the search back method if the daily data is available. As such, the tradition is to use the closing prices or the last transaction prices recorded for the day and, if they are not available, trade-weighted average price or the average of highest and lowest prices. If we do not have access to trade prices, the common approach is to use the bid offers as they are typically more often available than the ask offers and they typically behave less erratically. Of course, one can combine both methods – impute missing prices first with bid/ask offers, but if both are missing, then search back one day and repeat the procedure until an observation is found. It is a practical question which one of the methods should be applied first.

Now, it is possible that we do not want to utilize any of the methods above or that even after applying them, we may still have missing observations. The next question is what to do in this situation. For some purposes, a viable solution could be to use average of prices during the given measurement interval. In the case of a monthly index, it means using the average of daily prices that are available within a month. As a result, there is no need for search back or imputations as long as there is at least one observation in the given interval. However, the downside of this method is that it will typically create serious smoothness into the index and thus autocorrelation. In addition, the observed prices can be from the beginning or from the end of the month which also have an impact on the index.

The third method is the *return* imputation method. Obviously imputed returns can be converted into price imputations, but since this method has some of its own unique features, it is covered here separately. Three main return imputation approaches have been proposed in the literature. Note that if there is at least one missing observation after potential use of price imputations, one is forced to use return imputations unless one begins excluding thinly traded assets from the index.

The first approach assumes that the return is zero when we do not have a price observa-

tion. This corresponds to using the last known price until a new one is observed. Intuitively this assumption makes sense, as one can imagine the value to stay the same until new information appears. This approach is probably the most commonly used and the reason is its mathematical conveniency in index calculation. The downside is that this approach produces serial correlation into the index (c.f., e.g., Berglund and Liljeblom (1988), Jokivuolle (1995)). Moreover, the returns show somewhat problematic behavior. Namely, when a new price is eventually observed, potentially after several periods since the previous one, it accumulates the price development over multiple periods, and as a result, the return represents the sum of multiple periods creating returns that can be in either tail of the true return distribution. As a variant of this method, one could mitigate this problem by calculating a constant growth rate across all nontrading periods and using it in the imputation.

Another method assumes that the return is random, drawn from a return distribution that reflects the asset in question. This approach can be justified by the fact that since we do not know the real return, any reasonable value will do. This method keeps that mathematical conveniency of the previous method, but reduces the serial autocorrelation in the index. The drawback of this method is that the random values have to be saved once drawn. Otherwise, recalculating the index would give a different series. In addition, the derivation of the random returns requires information of the past distribution which induces new practical issues into the index calculation. Finally, intuitively it may be diffifult to accept the use of a random return in the index analysis as it implies that all derived indices will differ from each other.

The third approach assumes that the stock gains the same return as the (rest of the) market. This assumption is intituitively appealing as it is often the expectation that the stock prices move together with the market. This approach actually comes quite close to the approach where we exclude non-trading stocks from the index. The difference is in the fact that here part of the money stays invested in the non-trading asset, whereas in the omission method, all money is invested in the market.

Now, to demonstrate the difference between these methods, we use the following example. Let us again assume a market with only two companies A and B. They have both issued one stock, both valued initially at 100. At time $t+1$, the price of A goes up by 20 % whereas for stock B there is no price observation. At time $t+2$, the price of A stays put at 120 and stock B is priced at 108. Assuming a buy-and-hold investor who owns both of these stocks, her wealth at the beginning is 200 and at the end 228, an increase of 14 %. An index that tracks the return on the buy-and-strategy should reflect this at time $t+2$.

Table 1 compares how these three return imputation methods (I-III) behave. For the random return imputation approach, the return at time $t+1$ is assumed to be 4 per cent. The results show how all three return imputation strategies match the return of the buy-and-hold investor at time $t+2$. In fact, for as long as we get the end values right, any intervening values (e.g., due to a typo) will give us the same end value for the index. The choice of return imputation method has an impact on the index only on those periods when the imputation happens, but it does influence the overall return distribution.

The (naïve) method to exclude non-trading stocks from the index and invest the proceeds to the market has also been included in Table 1 for comparison (method IVa). It is clear from the result, that excluding assets from the index does not give the same results as the other methods. Here, excluding asset B from time t to $t+2$ from the index removes its total return of 8 %, but on the other hand, as rest of the market provides a return of 20%, the end value for the index becomes 120. We can go even one step further, and assume that the non-trading asset are removed from the index and the proceeds are kept "on the table" with 0% return (or perhaps with the risk-free rate). This method (IVb) goes against the basic assumption of an all-share index, but for the sake of interest, we can analyze it. Unfortunately, one can see from Table 1 that it produces biased result as well.

Now we can use the collected information to provide interesting information of the index that we have created. Remembering that we are always forced to use some kind of return imputation method, the situation can be written as follows using using equation (2) as the

Table 1: **Comparison of the return imputation methods.** Four different return imputation methods are compared in calculating a value-weighted index. Asset B does not trade at time $t+1$. Method I assumes a return of zero for asset B. Method II assumes that the return equals that of the rest of the market instead. Method III assumes the return is random (here assumed to be 4%). Method IVa excludes non-trading assets from the index during the periods for which we do not have returns. Method IVb assumes that money from non-trading assets can be kept "on the table", earning 0%. Note that the returns are reflected in weights and subsequent returns, if applicable.

	Time t	Time $t+1$	Time $t+2$
Panel A: Observed prices for assets A and B			
Stock A	100	120	120
Stock B	100	-	108
Panel B: Index values			
Method I	100	$100 \cdot (1 + 20\% \frac{100}{200} + 0\% \frac{100}{200}) = 110$	$110 \cdot (1 + 0\% \frac{120}{220} + 8\% \frac{100}{220}) = 114$
Method II	100	$100 \cdot (1 + 20\% \frac{100}{200} + 20\% \frac{100}{200}) = 120$	$120 \cdot (1 + 0\% \frac{120}{240} - 10\% \frac{120}{240}) = 114$
Method III	100	$100 \cdot (1 + 20\% \frac{100}{200} + 4\% \frac{100}{200}) = 112$	$112 \cdot (1 + 0\% \frac{120}{224} + 3.85\% \frac{104}{224}) = 114$
Method IVa	100	$100 \cdot (1 + 20\% \frac{200}{200}) = 120$	$120 \cdot (1 + 0\% \frac{240}{240}) = 120$
Method IVb	100	$100 \cdot (1 + 20\% \frac{100}{200} + 0\% \frac{100}{200}) = 110$	$110 \cdot (1 + 0\% \frac{120}{220} + 0\% \frac{100}{220}) = 110$

starting point

$$I_{m,t} = I_{m,t-1} \times \left(1 + \sum_{i=1}^{N_{t-1}-M_{t-1}} w^{s}_{i,t-1} R_{i,t} + \sum_{i=1}^{M_{t-1}} w^{s*}_{i,t-1} R^{*}_{i,t}\right), \tag{4}$$

where M_{t-1} is the number of missing return observations, $R^{s*}_{i,t}$ is imputed return observation, and $w^{s*}_{i,t-1}$ is the imputed weight. In principle, we could also break down the latter term to observations derived from the search back and from price imputations.[12] Now, this equation implies an interesting ratio, M_{t-1}/N_{t-1} which can be used to evaluate the severity of prevailing thin trading on the market and its impact on the market index. At one end, if there are no observations, the ratio is 100 % and on the other end, if there are no missing values, the ratio is 0 %.[13]

Finally, there are few more issues that we need to deal with. First, we need to decide necessary entry conditions for the stocks to the index. Do we require a trade price before

[12] In fact, we could also include a summation term with zero returns for those $N^{max}_{t-1} - N_{t-1}$ series that are excluded from the index. This can be used to calculate an index coverage ratio.

[13] We can also easily calculate similar ratio using market weights.

a stock can be included in the index or is it enough if a bid, or even ask, offer is available. At extreme, we can have stocks listed on the stock exchange, without a single trade (price observations) during its listing period. If we do not impute price observations in any way, the stock series in question is automatically excluded from the index. Now, if we utilize imputed price series, and there is at least one bid offer price available, we can choose to include the series into the index. It is also possible to require an ask offer before one can invest in the stock. The same question can be asked regarding the criteria for excluding stocks from the index.

Second, we need to consider when and how to take into account the effect of those corporate actions that take place during time periods when there are no price observations. The first question is: Should we take the action into account immediately, even when we do not observe a price (imputed or not), or should we delay taking it into account until we observe a price? If we choose to take them into account immediately, one needs to be careful with the search back method so that the price observation used as replacement is also post action. With the return imputation method, we can impute the return the same way as before and, at least in theory, we can even first apply the imputation as if there were no action, after which we take the action in question into account. However, for most purposes, it is often more convenient and more in line with the theory to assume that imputed return includes the corporate action as well.

To demonstrate the issue, let us assume that asset B in the previous example paid a dividend of 2 at time $t+1$. Now we have two alternatives: take the dividend into account when paid or delay until we have a price observation. In both cases, regardless of the return imputation method, B's return stays the same as before at time $t+1$. The difference can be seen at time $t+2$. If we assume that B's imputed return was zero at $t+1$, and we delay taking the dividend into account until $t+2$, B's return at $t+2$ would be $(108+2)/100-1=10\%$. However, if we decide to take the dividend into account immediately, we first note that the adjusted ex-dividend price at $t+1$ would be 98 and hence B's return at $t+2$ would be

$108/98 - 1 \approx 10.2\%$. At the index level, the former would give us an index value of 115 and the latter an index value of 115.05. The latter approach is more in line with the buy-and-hold investor's wealth development, but in practice it is a bit harder to implement in the index calculation.

The final issue is: How should we derive adjustment factor for corporate actions on thinly traded stock markets? As noted before, in some cases we can choose between using the market value based and theory-based adjustments. For example, if a company organizes a rights offering, these rights can be traded on the market, and as such we have a market price for them. The market prices are not necessarily the same as their theoretical value, as the rights can also suffer from thin trading or there might be some discount on them for practical reasons. Obviously, if the rights are not traded, then we have to settle for the their theoretical value. For the issue rights (which in principle are options) in particular, calculating the theoretical value can be challenging as we do not always have all necessary information, e.g., regarding the time to exercise or variance, to name a few.

3. CASE: HELSINKI STOCK EXCHANGE

3.1. Trading at the Helsinki Stock Exchange

The Helsinki Stock Exchange was officially opened for trading in October 7, 1912. At the end of the month, trading took place on 34 stock series from 30 different companies. The difference, four series, was due to the fact that four companies had recently issued new stocks that traded separate from the old shares because they had right only for half the dividend for the ongoing year.[14]

As Nyberg and Vaihekoski (2010) shows, the trading at the HSE was quite weak during the first four years. For example, the total number of stocks trades in 1913 was only 7709 in

[14]For a more general review of the history of the Helsinki Stock Exchange, see Nyberg and Vaihekoski (2010 and 2014).

896 trades with a value of FIM 2.54 million (corresponding to approximately 10.03 million euros in 2019). Although the trading began to increase markebly towards the end of the 1910s, the HSE decided quote certain stocks less frequently. For a while in 1916 and 1918, industry stocks and all other stocks were quoted on alternative days. From November 1918 onwards, the HSE divided stocks into two categories based on the their trading activity. Stocks considered as less liquid were quoted only on Tuesdays and Fridays whereas the other series were quoted on every day. For example, all banks traded every day whereas all insurance companies belonged on the less frequently traded cateogry.

Originally, the HSE was open for trading on stocks from Monday to Saturday. Beginning from Fall 1945, the HSE decided to open the stock exchange only during the weekdays (the same happened, for example, in the NYSE in September 1952). Of course, the stock exchange was closed on official holidays that were established at the time. On the other hand, during summer months, June, July and August, the stock exchange was closed on Mondays. This tradition lasted surprisingly late, until the end of year 1986. In 1914, trading took place only on Wednesdays from June 2nd until the end of July.

The Helsinki Stock Exchange has been complete closed few times during its history, mostly due to ensuing wars. This happened the first time during the First World War from August 1914 to March 1915 and from July to August 1916. The stock exchange was again closed between January 26th and May 20th in 1918 during the Finnish Civil War that ensued after Finland claimed independence from Russia in December 1917. In 1931, the HSE was closed shortly from September 21st to October 8th. In 1939, the HSE was closed from September 1st to 4th, October 11th to November 22nd, and finally from November 30th to March 31st in 1940.

Trading at the HSE took place in the Stock Exchange Trading Hall downtown Helsinki. In front of the room was a bench for the Chairperson and two secretaries. The brokers sat behind two-person desks organized opposite to the Chair in four columns and several rows. The Chair opened the trading day at at 10 o'clock (9 o'clock Central European

Time) by reading all official messages sent to the Exchange by the companies.[15] The trading began after the messages had been read. At the beginning of the day, the chairman asked for offers on the securities, one at the time. The price offers were given using an outcry system. Trading continued until no trades were recorded. Once all securities had been gone through, beginning from mid-November 1921, the brokers were allowed to trade freely within the price range achieved earlier until the trading ended approximately at 4 pm. This free trading session became important part of the trading activity as it typically accounted for more than one third of the trades.

A mechanical trading system – arguable one the most advanced in the world at the time and similar to the one used e.g. in Stockholm – was taken into use in 1935 and the trading changed somewhat as the brokers used the buttoms at their desk to issue their bid and ask offers. Once the offers met, the broker issuing the bid said out loud how many shares (s)he was willing to buy and the selling broker how many shares (s)he was willing to sell. In 1987 the decision was made to transfer trading gradually to all electronic trading system. The last open outcry was made in March 1990.[16]

The stock exchange was originally organized as a joint operation, a co-operation of its members until 1995. As such, only the members of the Exchange – mostly banks – were allowed to act as brokers and trade at the Exchange. In 1915, the HSE allowed even non-members to trade at the Exchange if they were given a permission to do so. By the year 1918, the number of members had already increased to the point that the Exchange decided to set an upper limit of 66 members. However, gradually the number of members started to decrease and it finally settled for little less than twenty members for the decades to gome.

The brokers trading at the Stock Exchange had to pay an annual fee to the Exchange. The stock exchange was originally organized as a joint operation, a co-operation of the brokers until 1995, and as such annual fees were set to cover HSE's operating expenses.

[15] Originally the days began at 1:30 pm, but it was two weeks later advanced to 11:50 am. Few years later, the days were opened at 11:30 a.m. but since the trading often lasted longer 4 pm, the beginning of the day was moved an hour earlier.

[16] For more information and for a video of the last open outcry, visit the following link.

These fees were originally set to FIM 200–800 depending the banks size of the equity capital (corresponding to 764-3055 euros in 2019) but they were increased gradually over the years with the inflation.

The broker's commission was set by the Exchange. On stocks it was originally set to 1/2 % for buyers and sellers, but is was lowered in October 1913 to 1/4 %. It was later in 1915 increased back to 1/2 per cent on top of which buyers and sellers had to pay 1/20 % (double if the broker was a non-member) fee to the Exchange for each transaction. In 1921, commission was again lowered back to 1/4 %, but at the same time the Exchange increased its own fee to 0.45 % all the while setting a minimum fee of FIM 25. In September 1986 the minimum level was removed.

Tick size was originally set to FIM 0.25. However, in 1916, the tick size was increased to FIM 1 for stocks priced less than one thousand and FIM 5 for those priced more than that. The exception was bank stocks, for which the lower tick size of FIM 0.50 was applied until end of 1929. After the introduction of the new markka (equal to 100 old markkas) in the beginning of 1963, the tick sizes were lowered accordingly.

In addition to the commission, investors had to pay so-called stamp duty to the government on all trades. Originally this duty was 0.8 per cent, but it was lowered to 0.5 % at the end of 1929. In 1941, the duty was increased first to 1.8 % and a bit later to 4 %. The Finnish specialty, Holding Concern company, was an exception – a temporary relief of the stamp duty was given to all trades on its certificates until the end of 1948.[17] In March 1949, the duty was lowered to 1 per cent until the end of the sample period on all stocks traded on the stock exchance. For stocks traded outside the stock exchange, the duty was slightly higher.

[17]For a more detailed explanation of this company, see Nyberg and Vaihekoski (2010).

3.2. Structure of the official trading list

In the beginning, all listed stock series were organized on the official list titled as the 'A' list under different industry categories. Bonds were placed on a separate 'B' list which traded less frequently, typically once a week. For the stocks, five industry categories were initially used (Banks, Insurance, Transport, Industry, and Miscellaneous companies). Companies were placed on the categories in mixed order of interest, not in an alphabetical order. In October 1918, the Stock Exchange began to quote less frequently stocks for which there was less demand. These series were marked on the official quote list with a star (asterisk) before the name of the security until May 1951. Trading took place on Tuesdays and Fridays towards the end of 1936 after which trading took place on all stocks every day.

Beginning in January 1951, the Stock Exchange reorganized its listing structure. The titles A and B lists were now used for stocks, and bond quotation list was kept as totally separate. All stocks were now placed either on A and B list which were below each other on the official list, and the former industrial categorization was dismissed. Stocks whose trading volumes were lower were placed on the B list although the earlier asterisk indicator remained in use for few months and even some of the *-indicated stocks ended up on the A list. Typically, stocks stayed in either list throughout the period they were listed, but occasionally some stocks were transferred from the B list to the A list and vice versa. The use of the A and B list ended in March 1970 after which all companies were listed below each other, but still not in an alphabetical order. In January 1976, the HSE again categorized companies into industries on the official list, this time alphabetically. Four industry categories were used: Banks, Insurance companies, Trade and Transport, and Industry.

There were also a number of other securities listed on the A list. One bond (issued by the City of Helsinki) was added on the A list on October 1913. In May 1914, a separate section was added for bonds ("Obligatiooneja, Obligationer"). In September 1916 the HSE made a decision to move this bond to the B list, but the bonds section returned in May 1921 after a decision was made to transfer three government bonds from the B list to the A list.

More bonds were added few months later. The first participating debenture was listed in February 1929 and it was given its own section heading. In January 1937, a separate section for stock issue rights, due to cash and bonus stock issues, was added for the first time as the trading on them started to take a more formal role. All of these additional sections were removed from the A list in 1942. In October 1946, a section for co-operatives with one company was added and the section for bonds was reintroduced, but with only one bond. After the 1951 the whole quotation list restructurated, bonds were placed without their own section heading at end of the A list. In 1961 new sections for bonds and debentures were added. Issue rights were given their own section heading in 1969.

From 1912 to 1970, the HSE recorded all the information above on a single page per day, i.e. all prices – for stocks, bonds, depentures, warrants – were written on this page. The only exception took place in 1926 when a separate sheet was introduced for the bonds. However, this did not last long, and from 1927 one page was again enough. In July 1978 the number of listed bonds had again grown to the degree that the HSE began to record their prices on a second sheet. From 1979 onwards these official bond quotation lists have been bound as separate volumes.

3.3. DATA

3.3.1. Official quotation list

Throughout the trading day, the secretaries wrote down information of each trade on a special trading ledger book. Similarly, at the end of each trading day, they wrote the quoted prices on a special sheet titled the Official Quotation List (in Finnish and Swedish "Julkinen kurssilista, Officiell kurslista"). These sheets had been prepared especially for the HSE by a printing company.

The sheets had the names of the quoted securites printed on them. Since new securities were introduced to the market throughout the year, the secretaries had to pencil down these new securities until the printing company delivered new updated sheets. As the HSE ordered

new sheets several times in a year, it usually did not take too long for the new securities to be added on the printed sheets. The only exception were those securities that did not trade long – for example newly issued shares with lower right on dividend – so they were not always added on the printed sheets.

At the end of the trading day, these sheets were dated, numbered and the Chair signed them "In fidem". At the beginning of the next year, the sheets for the previous year were collected together and bound into hard-covered books. Since the sheets were rather larger to begin with, these books were quite large in size, weighting typically several kilograms.

The sheets for the official quotation list had certain named columns for the recorded information. The number of columns varied a bit over the years, but for the most part, there were columns for the nominal value of each share, dividend, minimum trading lot, exact name of the security, as well as for the trading information. For some columns, such as security names and their nominal values, the values were pre-printed as they did not change that often, but the main columns, price information, were left empty for hand-written entries.

The number of columns for the official price quotations varied a bit over the years, but the columns for the highest bid and lowest ask offers were always there. In addition, there were either one or two columns for the realized trading prices. Namely, from January 1922 until the end of 1930, the realized prices were recorded separately for open outcry and free trading sessions on two columns. The entry showed the highest and lowest paid prices if they differed. The total rading volume was also recorded on a separate column (for bonds, the volume was the nominal value of the bonds sold). If no transactions took place, or if there were not bid and/or ask offers available, the price columns were left empty or hyphens were inserted.

The nominal values of each stock was also provided on the official lists with one exception February 1973 – January 1974. For the insurance companies, in addition to the full nominal value, there was a separate column from 1912 to 1930 for the paid-in capital. For all companies, if the nominal value changed, e.g., due to split or due to some other corporate

action, the old value was typically manually overwritten and the new value was pencilled down until the printed sheets began to reflect the new value. In January 1963, when the old FIM was replaced with the new FIM in Finland, the nominal values were not corrected until a new set of printed sheets were taken into use in March 1963.

Three columns were typically reserved for the information on dividends. The first two told the previous and the current dividend, percent of the nominal value.[18] Typically at the beginning of the year, the current dividend column was empty until companies made decisions on their dividends and the secretary dutifully pencilled them down on this column. Until February 1961 the third column titled "coupon" indicated the coupon used to collect the dividend. These coupons refer to the fact that in Finland before 1980s, stock certificates had a separate sheet with cuttable coupons for dividends and stock offerings typically for the next ten years. When all coupons had been used, it was replaced with a new one, but not the share certificate itself.

The dividend coupon used to collect the dividend was for the year when the dividend was paid, i.e. if the dividend was paid in Spring 1923 for the financial year 1922, the coupon to be used was 1923 and a similar marking was made on the HSE official quotation list. Although for the most part, the timing of the dividend is obvious from the official list, at times extracting the size of the dividend and the exact month of the dividend payment is not a straightforward process. Especially in 1968 the officials did not use the last dividend columns similarly for all securities – at times they did not write the new dividend if it was the same as the one already pre-printed on the price list.

Since stock issue rights (warrants) and newly issued stocks with lower dividend rights were listed only for a short period, all information for them had to be pencilled down on the official sheets. As a result, for most of the additional details (e.g. nominal values and trading lots) are missing for these securities, but, for example with the newly issued stocks one, can assume them to be the same as for the matching stock series. If these newly issued

[18]The column for previous dividend was added in September 1915.

stocks had a lower right to the dividend for the ongoing year, it was typically written down and often as fraction of the full dividend (e.g., "coupon 1928 3/4").

If a company's financial year deviated from the calendar year, it was marked down with a footnote in the dividend column beginning from year 1955. The same year, the official list includes dividends written down in the form of a+b%. This marking seems to indicate the need of some companies to signal that certain part of the dividend should be considered as one time special dividend, i.e., the a% part indicates the normal dividend and b% the special dividend, paid e.g. when the company had an anniversary year. There were also two occurances (Fennia Insurance and Enso-Gutzeit in 1971) where the special dividends were negative suggesting a one time reduction due to special circumstances.

The HSE officials wrote down information on the dividends until the end of 1972. After January 1973 there was no column reserved for the information on dividends. However, even after this, for some companies the HSE officials have pencilled down the dividend beside the name of the security – in some cases until the end of the sample period.

Minimum trading amounts (lots) for each security had also their own column on the official price quotation lists until January 1973. These trading lots varied originally from one stock to twenty, but later lots of fifty and hundred shares were also used. The lots were defined so that the minimum trade sizes were approximately equal in monetary terms for all stocks. Occasionally the secretaries seem to have been slow to make appropriate adjustments to the trading lot information. For example, KOP Bank organized a reverse split in December 1965 and the post-split prices were more than ten-fold, but the trading lot information was not updated until a year later in January 1967. It could also be that changes to trading lots had to be officially decided and hence the change did not take place until the next month.

3.3.2. Digitalization of the official quotation list

As one of the interests in this paper is to update the Nyberg-Vaihekoski the stock market index for the HSE from the establishment of the Helsinki Stock Exchange in October 1912 until 1970 after which an all-share value-weighted total return WI index by Berglund et al. (1990) is available.[19] This WI-index is widely used as the baseline index for stock market research on Finnish market. When the index re-created here is combined with the WI-index and the total return index that was made available by the HSE from 1991 forward, one can study the whole history of Finnish stock market.

On the other hand, it is known that the pre-existing research databases for Finland do not include detailed price information for the stocks prior to the last day of 1981. Thus, the decision was made to collect and digitalize month-end official quotation lists from 1912 to the end of 1981. The end result, a database that covers month-end information of the official quotation list in its entirety, collected mostly from the primary source, allows researchers to conduct long-term studies on Finnish stock market. In addition, it allows one to create various stock market indices for 1912–1981 and to compare the end result with the WI-index for the overlapping years.

The official quotation lists for the Helsinki Stock Exchange are stored as bound volumes at the Central Archives for Finnish Business Records (ELKA) in Mikkeli. The books for 1913–1979 are available – at present, it is not known whether similar bound books for year 1912 and post-1980 have been created, and if they are, where they are stored. To cover years 1912, 1980 and 1981, the information was augmented with the information from Helsingin Sanomat newspaper – the leading daily newspaper in Finland.[20] Helsingin Sanomat listed the official price list on a daily basis with the same structure as the official price list kept by the HSE together with some additional information (e.g. highest and lowest prices for the

[19]The index has been updated until the end of 1990 by the Department of the Finance at Hanken School of Economics. Unfortunately, the information used to create the index before 1982 has for the large part disappeared.

[20]The only exception is November 1980 when the newspaper did not appear due to journalists' strike. This information is taken from the official quotation lists provided by the HSE to me.

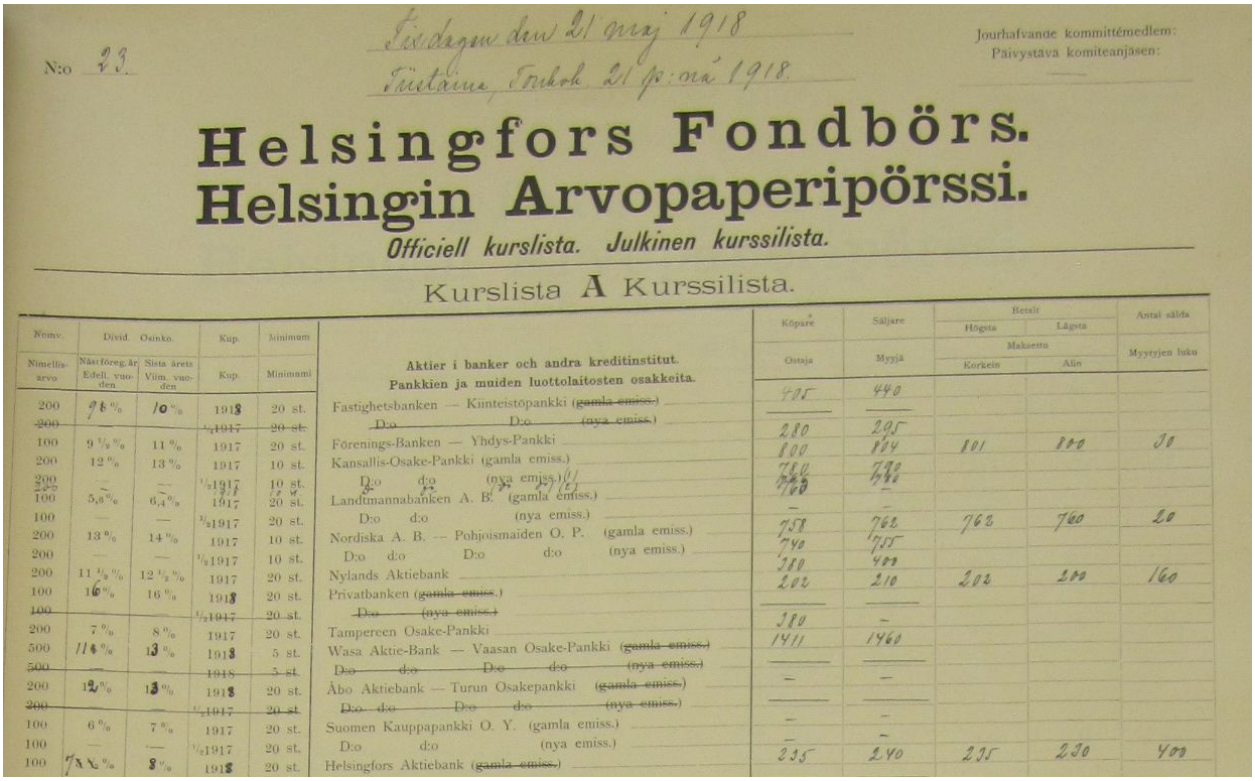
year as well as dividend yields).

In practice, month-end pages from the bound books and newspapers were photographed. Since the less frequently quoted series stocks were not always traded on the last day of the month, I also photographed few previous pages and made certain that I always had the last available official list for all stocks, including those that were quoted less frequently.

The next step was to digitalize these pages. Using an OCR software to do the extraction of the data proved to be impossible due to the fact that most of the OCR software do not work well with data structured as tables although some advanced has been made with this respect (c.f. Transkript software). In addition, the information on the sheets was always a combination of printed security names with manually entered trading information. Finally, many securities were pencilled down, sometimes between the rows, making it almost impossible for the OCR to extract the information accurately enough, if at all. Figure 1 shows part of the price quotation list demonstrating the difficulty in digitalization. Thus, in the end the information was manually digitalized (Internet Appendix 1 provides more details of the process).

All information on the official information sheets (and newspaper pages) was collected and digitalized. In the end product, a database, the data was organized first in the order of the securities, then on dates. The date of the last trading data was stored for securities that did not trade exactly at the end of the month. With the digitalized data, the last trading days for each month is searched for and a list was created (available upon request). For approximately twenty months, the end of month day differs from that in the Nyberg and Vaihekoski (2010) database. In addition, a separate list of all traded securities was created. The latter list includes information of the first month when the trading began as well the first month when the security was delisted.

After this I conducted a number of checks on the reliability of the database. First, the bid and ask offers were compared. If the former were higher that the latter, it would indicate an erranous entry in the database. Similarly, the realized prices could not be higher or lower



N:o 23.

Fredagen den 21 maj 1918
Tisdagen, Tisdag 21 p:nä 1918

Jourhafvande kommittémedlem:
Päivystävä komiteanjäsen:

Helsingfors Fondbörs.
Helsingin Arvopaperipörssi.

Officiell kurslista. Julkinen kurssilista.

Kurslista A Kurssilista.

Nomv. Nimellisarvo	Divid. Osinko. Näst föreg. år Edell. vuoden	Divid. Osinko. Sista årets Viim. vuoden	Kup. Kup.	Minimum Minimumi	Aktier i banker och andra kreditinstitut. Pankkien ja muiden luottolaitosten osakkeita.	Köpare Ostaja	Säljare Myyjä	Betalt Högsta Maksettu Korkein	Betalt Lägsta Maksettu Alin	Antal sålda Myytyjen luku
200	9 %	10 %	1918	20 st.	Fastighetsbanken — Kiinteistöpankki (gamla emiss.)	405	440			
~~200~~	—	—	~~1917~~	~~20 st.~~	~~D:o D:o (nya emiss.)~~					
100	9 ½ %	11 %	1917	20 st.	Förenings-Banken — Yhdys-Pankki	280	295			
200	12 %	13 %	1917	10 st.	Kansallis-Osake-Pankki (gamla emiss.)	800	804	801	800	30
200	—	—	1917	10 st.	D:o d:o (nya emiss.)	780	790			
100	5,8 %	6,4 %	1917	20 st.	Landtmannabanken A. B. (gamla emiss.)	760	—			
100	—	—	1917	20 st.	D:o d:o (nya emiss.)	—	—			
200	13 %	14 %	1917	10 st.	Nordiska A. B. — Pohjoismaiden O. P. (gamla emiss.)	758	762	762	760	20
200	—	—	1917	10 st.	D:o d:o D:o d:o (nya emiss.)	740	755			
200	11 ½ %	12 ½ %	1917	20 st.	Nylands Aktiebank	380	400			
100	16 %	16 %	1918	20 st.	Privatbanken (~~gamla emiss.~~)	202	210	202	200	160
~~100~~	—	—	~~1917~~	~~20 st.~~	~~D:o (nya emiss.)~~					
200	7 %	8 %	1917	20 st.	Tampereen Osake-Pankki	280	—			
500	14 %	13 %	1918	5 st.	Wasa Aktie-Bank — Vaasan Osake-Pankki (~~gamla emiss.~~)	1411	1460			
~~500~~	—	—	~~1918~~	~~5 st.~~	~~D:o d:o D:o d:o (nya emiss.)~~					
200	12 %	13 %	1918	20 st.	Åbo Aktiebank — Turun Osakepankki (~~gamla emiss.~~)	—	—			
~~200~~	—	—	~~1917~~	~~20 st.~~	~~D:o d:o D:o d:o (nya emiss.)~~					
100	6 %	7 %	1917	20 st.	Suomen Kauppapankki O. Y. (gamla emiss.)	—	—			
100	—	—	1917	20 st.	D:o d:o (nya emiss.)	—	—			
100	7 %	8 %	1918	20 st.	Helsingfors Aktiebank (~~gamla emiss.~~)	235	240	235	230	400

Fig. 1. **Official price quotation list for Tuesday May 21, 1918.** The first five columns show the nominal value, dividend percentage for previous and ongoing years, dividend coupon required for the dividend payment, and the minimum trading lot. The middle column shows details of the listed asset – typically the name of the company. The five columns on the right show information of the price quotations. First two show bid and ask offers, the next two the highest and lowest prices for the transactions while the last column shows the number of shares traded.

than the offers although it was possible that there were realized prices but no official bid or ask offers. To remove any potential wrongly placed decimal dots, I also compare the bid offers with those in the old database. If they differed for more than ten per cent, a likely typo exists in either database. Furthermore, I compare consequtive prices. If they differed more than 50 per cent, it warranted a closer scrutiny. Finally, if trading volume was given, price had to be given as well.

3.3.3. Bonus and cash rights issues

During the sample period, Finnish publicly listed companies were quite active in their equity capital operations. Although a variety of different actions can be found, the most notable

and frequently used capital actions were bonus and cash rights issues (equity offerings). Both types of issues give the current shareholders the preemptive right to receive/buy additional shares of the company without diluting their ownership. Although a number of directed issues were also used – typically in a merger to the sellers of the merging company – it was actually quite rare during the sample period for the companies to issues new stocks to the general public. The only exception were banks who often gave their clients a chance to subscribe to new stocks as part of the cash rights issue, sometimes at a higher subscription price than that offered for the current owners.

In these rights issues, a company issues additional shares of stock and gives all share owners a pre-emptive right to subscribe them free of charge (bonus issue) or at a discounted price (cash issue). Early on, companies seems to have required that the shareholders sign a list to participate in the issue. If some holders did not sign the list or contact the company in time, the company sold these rights through the stock exchance. Later in the 1920s, the shareholders were asked to cut a rights coupon from the stock certificate to exercise their right. Often these coupons were traded in the HSE, but not in all cases, especially if the stock price had fallen below the subscription price.

The main motive for companies to organize cash rights issues is to raise new capital. As such, it is an important action for any company. Originally, the government did not tax issues, but in 1916 a 2 per cent tax was levied on all issues to cover initially for the WWI related expences. This stamp duty stayed and it was slowly increased to 4 per cent (6 per cent for bonus issues). However, the duty was temporarily lowered for every now and then, and for cash issues, it was even set down to zero at times. From May 1966 onwards, the duty was removed on all issues. In the 1970s, dividends payed on the newly issues stocks was made tax deductible which further increased motive to issue new stocks. The bonus issues were, on the other hand, motivated by two other reasons. Anecdotal evidence suggests that bonus issues were frequently used by the companies as a means to increase their dividend payments all the while they were often unwilling to increase their dividend rates as such. In

addition, the bonus issues were used as a way to sweeten the cash rights issues and hence it is quite common to see mixed issues during the sample period.

Since the issues can potentially have a major impact on the market index, one needs to collect the relevant information. For the index construction, detailed information on the rules of the issue are crucial. This included, for example, how many coupons were required to receive/purchase a new stock (i.e. subscription ratio), subscription price and payer of the stamp duty (company or investor), whether the newly issued stocks have right for the dividend for the ongoing year, and the payment as well as the timing details of the issue. In cash issues, the payment often took place in two installments and at times the company offered interest on early exercise of the rights and charged interest on late payments. The timing and the time until the rights had to be exercised are important to calculate the value of a right. In addition, it is important to to know the date when the stock traded sans right ("rights off") and whether the rights were traded at the market to calculate and to apply the appropriate adjustment for the price.

Unfortunately, the HSE has not stored detailed information on the issues. Luckily, Kock has collected detailed information on most of the issues, especially from the beginning of 1960 forward. For the period before 1960 the information is less detailed; main missing piece of information is the timing of the issue but in most cases one can deduct the timing from the market reaction or from the rights becoming quoted on the stock exchange. The information provided by him was used to create the Nyberg and Vaihekoski database on splits, bonus issues, and cash issues. Here, the database is augmented with corporate capital actions taking place between 1970–1981 using Kock (1975 and 1982). In addition, HSE's annual statement for 1928, which contains a detailed list of issues and increases in nominal values during 1927 and 1928, is used. Although the list contains only two issues that were not previously included in the database, it does contain detailed information on the timing of the issues. Finally, a few additional updates and correction are also made to the database.

All in all, during 1912-1981 (1970-1981), a total of 20 (8) splits and 12 (2) reverse splits

are found for the listed stock series in the database. Similarly 276 (89) cases of bonus and 462 (120) cash rights issues are found, some of which are mixed combining both bonus and rights issues at the same time. The collected database is used to calculate the adjustment factors for the index calculation as well as to adjust the number of stocks mid-year.

3.3.4. Other information sources

As noted before, the official quotation sheets do not include information on dividends after 1972 except for a few companies. The dividends and their ex-dividend months during 1980 and 1981 are collected from the Helsingin Sanomat newspaper. For the period from 1975 to 1979, I collect missing information on the dividends from annual stock market books by Gunhard Kock. He provides the exact date for the dividend payment. Unfortunately, the books are not available prior to 1976, so the timing of the dividend could not be determined for 1972-74, and in those cases, they are assumed to be paid in April similar as in Nyberg and Vaihekoski (2010).

As the number of stocks is needed to calculate market capitalization values of the series, information on the book equity capital for each series is taken from the year-end database collected by Nyberg and Vaihekoski (2010; updated for the 2014 study). Here the database is augmented with new information collected for 1970-1981 using again the books by Kock. The equity capital book value is used to calculate the number of shares with the help of the nominal values. Since the nominal values are not written on the official quotation sheets between February 1973 and January 1974, I compare the values before and after this period and find that during this period there were no changes in the nominal values.

4. EMPIRICAL RESULTS

4.1. Descriptive analysis

Ultimately, 849 different stocks, stock warrants, and newly issues stocks were listed in the Helsinki Stock Exchange during 1912-1981.[21] Of these securities, 206 were stock series as such, but in addition there were 252 newly issued stock series and 391 different bonus or cash issue rights listed and traded at the HSE.[22] On average, the stock series are listed for 226 months, i.e., just a little less than 20 years, when calculated towards the end of year 1981.[23] The newly issues stock series are typically listed for less than half a year, i.e., until the next dividend payment after which there are no differences between to old and new series and they are merged to the "main series" on the List. The number of different series listed each month can be seen from Figure 2.

We can see that the number of listed stock series has varied quite dramatically during the sample period. After the modest beginning, the HSE started to attract more and more companies towards the end of 1910s all the while the companies started to raise new capital from the market via cash offerings which is clearly seen from the Figure. However, in the 1920s the number of listed companies started to decline, arguable mostly due to active merger activity, especially in the banking sector.

Figure 2 also shows the number of quoted issue rights. The stock exchange began trading rights for the first time in March 1927, but since these rights are typically listed for a month or two, the impact on the total number of listed securities is quite marginal. The number of

[21] It is also interesting to compare the newly created database with the one collected by Nyberg and Vaihekoski (2010). For the 1912-1969 period we found 595 stock (old and new) and stock issue rights series, which is 33 more than Nyberg and Vaihekoski (2010) database had.

[22] These series represent 167 different companies. Companies whose names changed only slightly are considered as one. Namely, in Finland the names changed more or less similarly as a process of going from the original company name in Swedish language (or combination of Swedish and Finnish names) gradually to a more universal name. Some changes were due to changes in the overall development of Finnish grammar/style. For example, 'Pargas Kalkbergs Aktie Bolag' changed its name to 'Paraisten Kalkkivuori O. Y. – Pargas Kalkbergs A. B.' from 1936 forward. From 1970 onwards, its name was changed to 'Paraisten Kalkki Oy - Pargas Kalk Ab'. Later the company name changed to present day Partek Oyj.

[23] The average is 136 months if calculated across stocks series that had been delisted by the end of 1981.

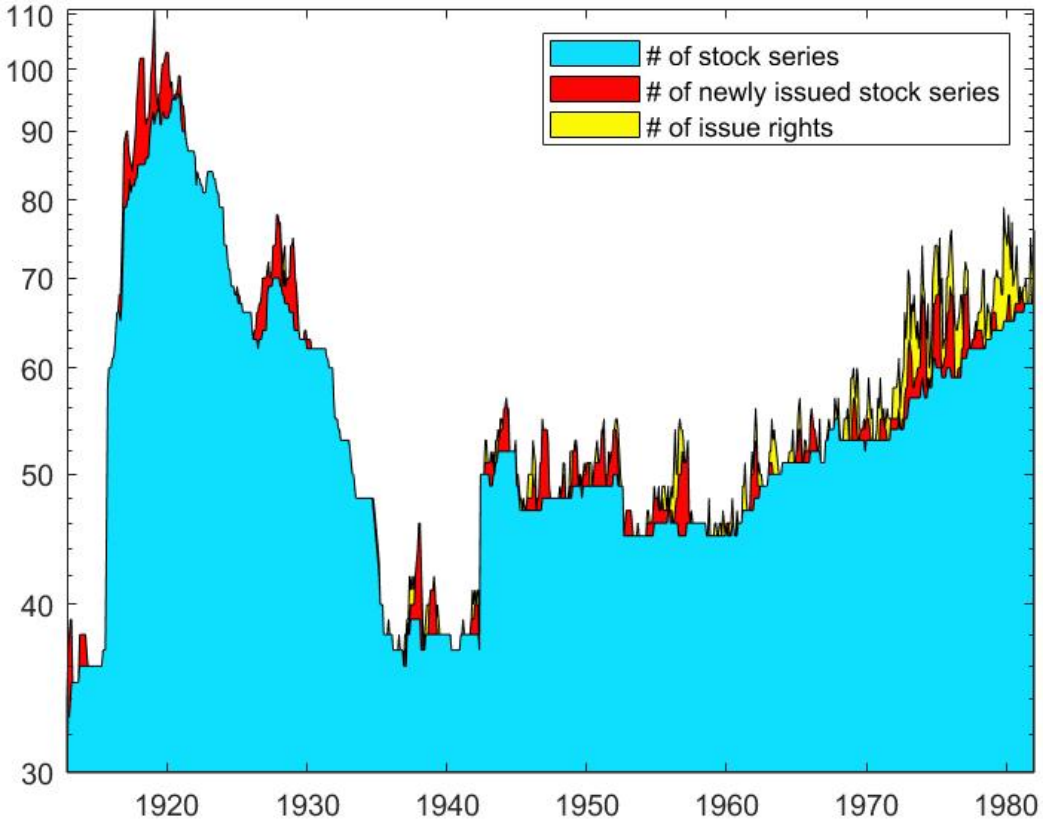


Fig. 2. **Number of listed securities.** Number of stock series, newly issues stock series with temporary lower dividend rights, and cash and bonus issue rights listed in the HSE (month-end). Log-scale used on y-axis for clarity.

listed stock series started again to increase in the 1970s. At the same time companies started to raise more capital from the market. This can be partly attributed to the government's decision to lower tax on issues in 1969. This all contributed to the increase in the number of issue rights (warrants) being quoted on the market. Their number could be even higher since not all rights were listed at the Exchange.

Analysing the rights issues reveal a number of interesting issues. First, typically the rights issues do not seem to be an indication of the company being cash-strapped or in desperate need of capital as such. More likely, they seem to be commonly used by the most successful companies to finance their future growth. There are, however, few issues where investors

are asked successfully to buy new shares at a premium to prevailing market price. Although the initial thought would be to forego such issues, it seems that the investors have seen the offering as a chance to help the company (e.g. to avoid bankruptcy) and thus investing makes sense in the long run especially if the company has appealed on certain duty of ownership.

The second most typical use for the rights issues especially in the 1970s is the introduction and ultimately listing of a new stock class. Namely, towards the end of the sample period companies started to issue stocks that are commonly labelled as preference class stocks. These stocks typically had a lower amount of voting rights, but they often had first right on dividend to make them attractive to investors. The motivation of issuing these stocks is the fact they do not dilute the voting power of the old owners the way as first class stocks would.

It is also interesting to see that Finnish companies have organized quite many bonus issues during the sample period. There are probably many reasons for this but, as always, tax reasons are probably on the top of the list. For example, these bonus shares can be seen as an alternative to dividend payout and as such, they they deferred the taxation until sold whereas the dividend income was taxed immediately. On the other hand, one could interpret bonus issues as a signal of the company's future. Finally, many companies also used bonus issues with the rights issue – in fact, more than half of the right issues are mixed issues. In some sense, bonus issues could have been used as a way to soften the company's request for shareholders' money.

One of the main interests in this study is the nature of the thin trading in the HSE. There are several ways to describe thin trading. Analyzing the stock series and the time they were listed, we can calculate the maximum number of stock month-end observations. In the 1912–1981 sample period, it is 46,709. Of these potential monthly observations, we can observe a transaction price for 16,444 months (representing 35.2 per cent of 46,709). Bid and ask offers are available for 38,587 (82.6%) and 29,031 (62.1%) months respectively. It is obvious that bid offers are available most often, followed ask offers, and then finally by trade

prices. This is arguably typical for most stock markets – here, ratio of bid and ask offers to price observation is 2.35:1.77:1. Thus, one observes more than two times as many bid offers than actual trade prices.

Thin trading does not stay the same and it can vary for a number of reasons. To analyze this, I calculate the percentage of stocks with a month-end bid, ask, or price observation from October 1912 to December 1981. The result can be seen from Figure 3. As expected, thin trading was quite severe at the beginning of the sample period but it starts to ease towards the end of the sample period. For example, in the 1920s only 20 to 30 per cent of the stocks traded, whereas in the 1970s almost 70 per cent of the stocks traded. At the same time, from the 1970 forward almost all stocks have at least a bid offer made in the stock exchange. The development has not been unidirectional, though. Since we are using information from only one day for each month, there are also some outlier looking observations as well. For example, in April 1945 only three stocks had an ask offer whereas there were no major changes in the number of bid offers.

We can also analyze how many price observations are associated with bid and ask offers as well. The results show that bid and ask offer prices are typically posted when stocks are traded. There are only 751 observation months when this is not the case. Similarly, there are 142 (636) traded price observations without bid (ask) offer at the same time. On the other hand, both bid and ask offers are available for 27,161 days which indicates that if we choose to impute prices with the midpoint of the bid and ask offers, we could increase the number of transaction price observations by more than 65%.

It is also interesting to study the distribution of the dividend payments across calendar year. Nyberg and Vaihekoski (2010) assumed that the dividends were always paid in April. Here the results show that during 1912-1972, a total of 3,165 non-zero dividends were paid. April was indeed the most popular month with 38.60% of all dividend payments, followed by March (24.24%), May (14.34%), and February (12.33%).

Finally, I study how many changes were made to the stocks' nominal values. There can

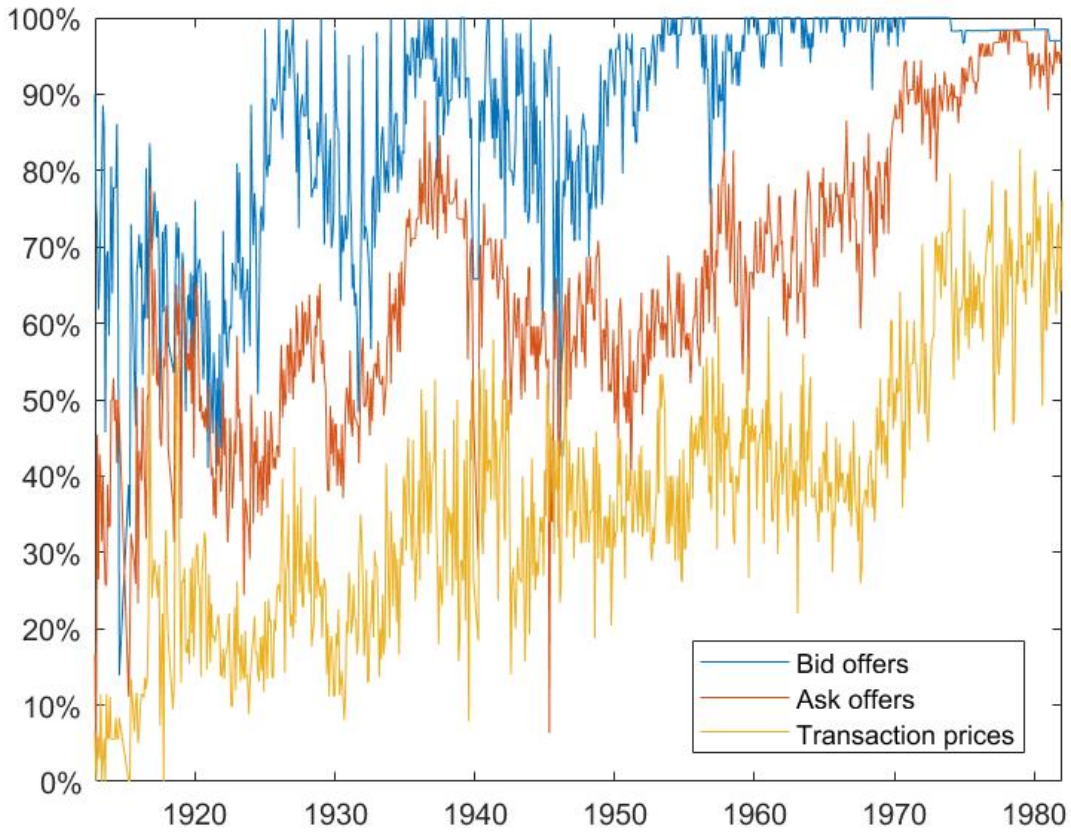


Fig. 3. **Availability of price observations.** For each month-end the number of recorded bid offers, ask offers, and transaction prices is dividend by the number of listed stocks. Months when the stock exchange was closed have been removed from the graph.

be several reasons for increasing the nominal value. For example, as the dividend payout rate was a function of the nominal value, an increase in the nominal value could signal management's improved forecast for the company's future especially if the payout rate was kept the same. In Sweden, the nominal values were quite frequently adjusted for taxation purposes (c.f., Rydqvist, 2019). Here we find 258 changes by the end of 1981. Since 30 of them are related to splits, the rest are due to companies increasing (with few exceptions) the nominal value of their shares using reserve equity capital. Analyzing their occurance frequency distribution over the sample period, one can see that the are quite evenly distributed over time, typically one or two companies per year but there are years without any changes and

years when the changes are more commonplace (especially in the early 1950s).

4.2. All-share baseline index

The baseline index is an all-share, value-weighted, total return monthly index based on bid offers. Zero return imputation method has been used for the missing values. The index is calculated for the most part similar to the Nyberg and Vaihekoski (2010) index (henceforth labelled as the NV index), but without the search back method. The data for the capital operations (splits and issues) is basically the same as for the NV index with minor updates for 1912–1969 and extension for 1970–1981, but otherwise the newly collected database with bid offers, dividends, and nominal values from the primary source as well as information on timing of dividends have been utilized. Theoretical month-end exercise values for bonus and cash issue rights have been used to calculate their adjustment factors.[24] The end result can be seen from Figure 4. The index is calculated until the end of year 1981. The old NV index has also been included for comparison until March 1970. The base value of 100 is set for both in October 1912. In addition, the Berglund et al. (1983) WI index has been included for comparison from March 1970 forward.

We can see that the indices track each other as expected. Some occasional differences are to be expected, as the used bid offers may differ every now and them, but the end results should be close to each other.[25] Table 2 provides descriptive statistics for the percentage returns of the indices. In Panel A, the sample period is from October 1912 to March 1970.

The geometric mean returns, which corresponds to the compound monthly growth rate (CMGR), are 1.030 and 1.027 per cent (13.08% vs. 13.04%), for the new and the old (NV)

[24]The adjustment factor for cash issues is $f^a_{i,t} = Max(0, (S_{i,t} - X_{i,t})/R_{i,t}))$, where $R_{i,t}$ is the ratio of one over fraction of rights given to each stock, and $S_{i,t} - X_{i,t}$ is the difference between the ex rights price for the stock on the market and the issue price. If newly issued stocks have lower right on the dividend for the ongoing year, one should deduct the expected dividend from the market price. Since this information is not available, this latter adjustment is not done. As a result, the adjustment for the cash issues is occasionally slightly too high, but in contrast it does not include rights' time value which makes the adjustment too low.

[25]In fact, differences in the interim observations have only a minor effect on the end-value of a stock specific index ceteris paribus. Similarly, for the aggregate index, the effect is quite minimal. As the geometric mean return utilizes only the first and the last observations of the index, it should also return quite similar values. Arithmetic averages, on the other hand, may differ due to differences in the price observations.

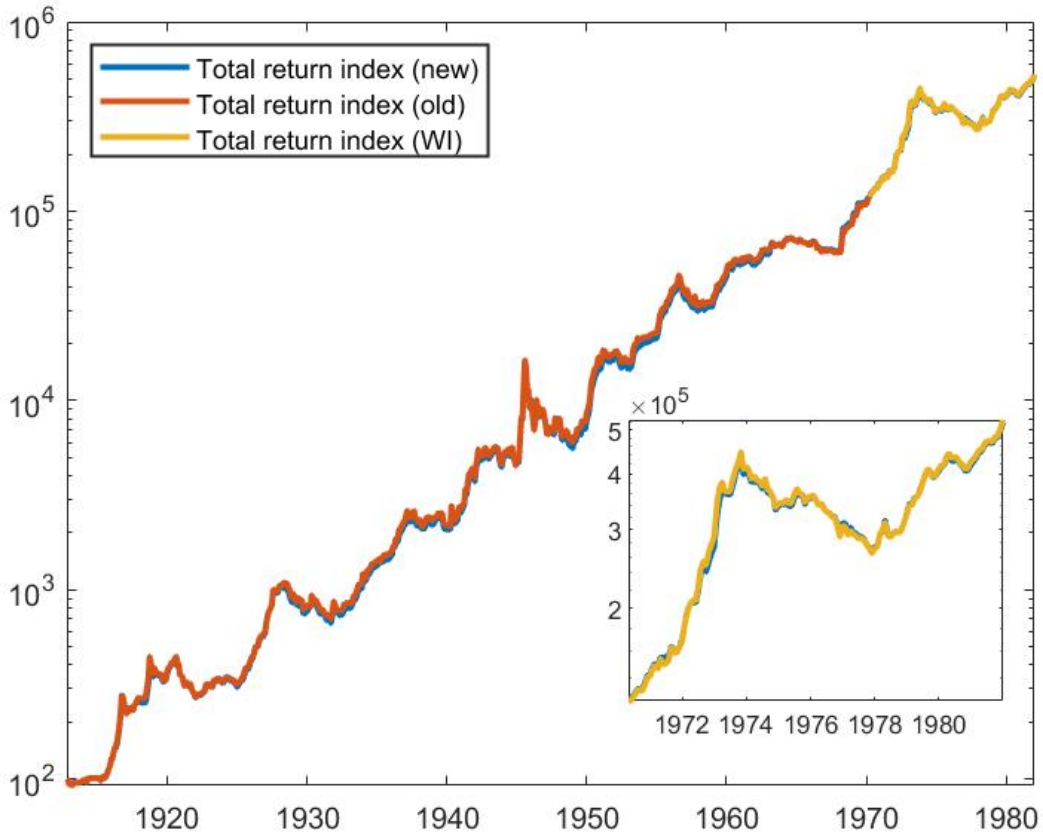


Fig. 4. **Old vs. new indices.** Month-end values for the new value-weighted total return index (1912M10-1981M12), for the old index (1912M10-1970M03) by Nyberg and Vaihekoski (2010), and for the WI index (1970M03-1981M12) by Berglund et al. (1983). The WI index has been scaled to start with the end value of the new index (closer comparison is provided in the small inset on the lower right corner).

index, respectively. Observing a slightly higher CMGR for the new index is to be expected as close to 35 per cent of the dividends were paid before April and only 20 per cent after it. Thus, the assuption of using April as the ex dividend month for all companies in the old index had a minor decreasing effect on the end result.

On the other hand, arithmetic mean returns for the new and the old indices are 1.166 and 1.179 per cent per month (14.9% vs. 15.1% per annum) which is due to the old index's higher volatility. Namely, volatilities are 5.39 and 5.69 per cent per month for the new and the old index, respectively. This is likely to be caused by higher frequency of erraneous

observations in the latter index which typically cause higher swings in returns. Both series show signs of nonnormality with positive skewness and excess kurtosis. The first three autocorrelation coefficients are also shown in the table. The new index shows higher first order autocorrelation (0.25 vs. 0.20) and to some degree the second order as well. We also test for the significance of the arithmetic mean with one sample t-test statistics. It is significantly different from zero (t-value is 5.66 for the new index). Since the test is based on the assumption that the observations are independent from each other, we re-estimate the test statistic using standard errors adjusted for AR(1) process (c.f., Campbell, Lo, and MacKinlay, 1997). The result remain, the mean return is statistically different from zero.

Panel B reports similar descriptive statistics for the March 1970 to December 1981 period. Since the WI-index by Berglund et al. (1983) is available for this period, it is included for comparison. The geometric mean return is 1.014 and 1.064 per cent for the new and the WI index, respectively. The arithmetic mean return for the new index and the WI-index are 1.07 and 1.13 per cent per month (13.7% vs. 14.4% per annum). The difference is most likely due to the missing issue adjustments. Again both return series show positive skewness and excess kurtosis, although to much lower degree than before 1970. The biggest change is in the kurtosis which is likely to be driven with lower thin trading (fewer zero return). Both series also show fairly high first order autocorrelation, although slightly lower than in Panel A. Somewhat surprisingly, the second order autocorrelation coefficients are much higher than before.

4.3. The effect of thin trading adjustments

Next, I analyze the effect of different methods to deal with thin trading on the index. Note that these different methods do not necessarily have a major impact on the geometric mean return, or even the arithmetic mean if the indices close at the near similar values. The difference is expected to be with the other moments of the return distribution and in its serial correlation pattern.

As a starting point, I re-calculate the index using realized price observations instead of bid offers. Obviously, the number of observations reduces dramatically, but one can argue that there prices measure truly tradable returns. The results for the 1912 to 1981 sample are reported in Panel C. The compound monthly growth rate average return is 0.991 per cent (12.57% p.a.) which is surprisingly clearly different than for the baseline index. In fact, it is close to half a per cent lower per annum. Volatility is also slightly lower, 4.99 per cent, but skewness and kurtosis are higher. The autocorrelation patterns are approximately similar.

The difference in results can be explained by analyzing the thin trading characteristics of the index using equation (4) vis-a-vis the baseline index. On average, 55.18 stock series were listed each month during the sample period, and only 20.15 (on average 33.04% of all series) stock series had price observation each month for our based index on average, whereas for the baseline index we have bid observations for 47.29 stock series (87.38%), on average. Of course with the help of zero return imputations, the number of return observations for the price-based index rises to 49.25 (89.80%) stock series, but for the baseline index, the coverage goes up to 99.72%. Now, the lower number of stock series in the price-based index is due to the fact that stock series are excluded until they have their first price observation. This leads to a great deal of companies to be excluded from the index, at least for the first few months of their listing period whereas they are included in the baseline index as soon as there is a bid offer becomes available. As such, the price based index excludes relatively high returns after the listing.

To see how price imputation method may affect the index and returns, I re-calculate the index with the price observations imputed with the mid-point of bid-ask offers. If even the mid-point is missing, bid offers are used. Again no search back is used. The results are again reported in Panel C. As expected, the CMGR is now clearly higher than for the index based on non-imputed prices, and almost at par with the baseline index, 1.023% (vs. 1.027%). Volatility is also slightly higher, 5.34 per cent, but it is likely to be due to bid-trade bounce. Skewness and kurtosis are at par with the baseline index. The first order autocorrelation

is clearly lower (0.185 vs. 0.234). Now the number of usable observations rises to 47.46 (87.68%), on average. Together with the the zero return imputation method, the number of return observation is basically the same as for the baseline index (99.72% of listed series, on average).

As the final analysis, the baseline index is re-calculated using the search-back method to augment missing bid offer observation with earlier values within each month. The lack of daily observations limits its use here, but luckily, we can utilize the old bid offer database collected in Nyberg and Vaihekoski (2010) as they collected the last available bid offer within each month for each asset.[26] In practice, we use month-end bid offers that have been collected for this study and impute missing values with the bid offers from the old database. As a result, the number of observations is maximized, and on average it is 51.84 (94.88% of all). With the zero return imputation, we basically have full coverage of all listed stock series in the index. Panel C shows again the descriptive statistics for the 1912 to 1981 period.

As expected, there are no major differences when compared to the baseline index. The CMGR is slightly higher, 1.032 per cent (13.11% per annum), but the difference is not large. Volatility is slightly higher, 5.33 per cent, as is kurtosis. Autocorrelation coefficients are all lower than that for the baseline index. The differences are most likely due to intra-month non-synchroneity in used prices that is induced by the use of search-back. Overall, the results show that the method to deal with thin trading can have an important and meaningful impact on the index. Here, the difference between the lowest and the highest CMGR is 0.54% per annum which can be considered quite substantial. The return series is also influenced of the choice.

4.4. Performance analysis

We can also analyze the impact of dividends and different capital operations on the index. To do this, I calculate the baseline index without the dividends, but taking into account

[26] Their old database was slightly updated for this study.

capital operations. The end results corresponds to the price index. Second, I re-calculate the index even without adjustments for the issues. This index can be labelled as the raw price index. Splits are taken into account in all cases.

The geometric mean return for the total return index is 1.03% per month (13.05% per annum), but if one exclude the dividends, the mean geometric return of the price index decreases to 0.59% (7.37% p.a.). For the raw price index, the mean return is even lower, 0.24% (2.65% p.a.). This result highlights the fact that capital operations, and especially the rights issues, play a major role when creating a return index for the Finnish stock market. The impact on annual returns (0.35%, 4.72% p.a.) is almost as high as that of the dividends (5.68% p.a.).

Next, a number of different variations of the baseline index are created with different weighting structure. A natural choice is the well-known equally weighted version of the all-share indexes. Figure 5 shows the development of both value-weighted and equally weighted indices. It is clear from the figure that the equally weighted index has faired much better than the value-weighted index during the sample period. This is in line with the expectation as the former index gives higher weight on smaller companies with respect to the latter index. In addition, I calculate the book-equity weighted version of the index (available upon request). Returnwise it falls in between the value-weighted and equally weighted indices.

Finally, it is interesting to analyze the stock market performance of different industries. This can be done with the help of industry indices. For this purpose the listed companies have to be classified into different industries based on their main line of business. Obviously, defining the main line of business can be tricky at times, as it may require information on companies that no longer exists. Luckily the stock exchanges often provide some kind of information with their list classification and since we are dealing with the publicly listed companies, there are typically historical material available (often even as books on corporate history). For companies with multiple lines of business, the choice can also be somewhat subjective. Similarly, the main line of business can change over the years and thus the

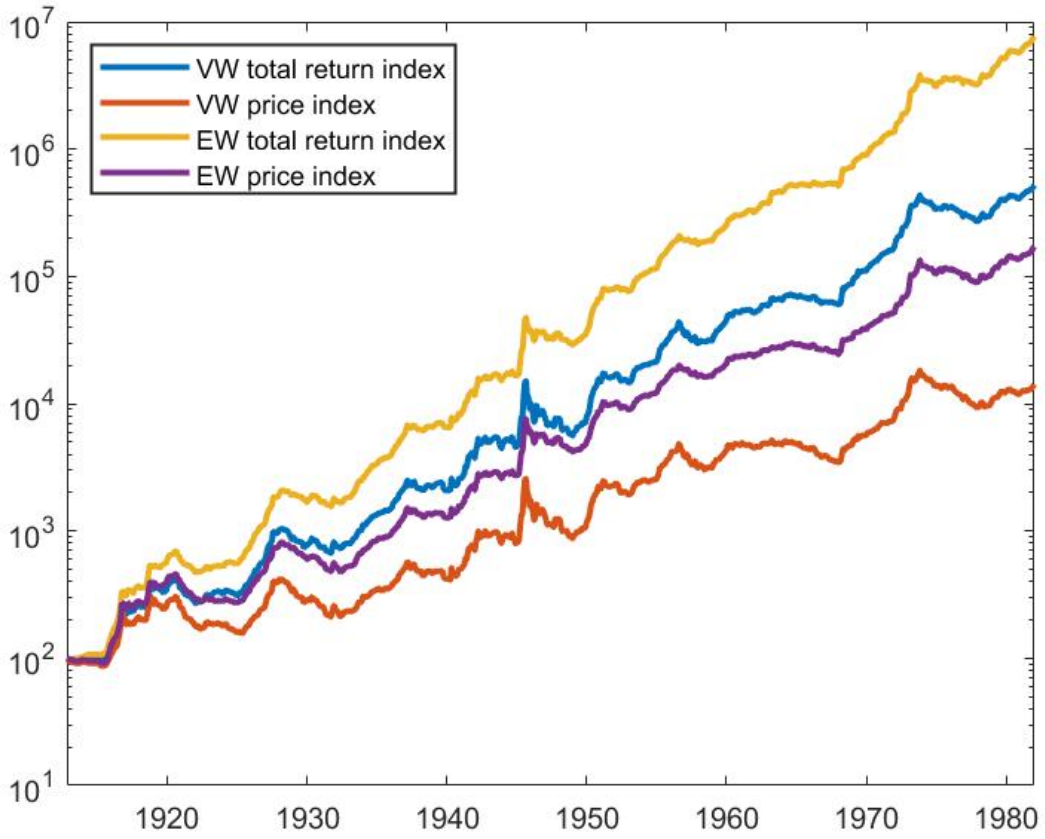


Fig. 5. **Baseline indices.** Monthly values for value-weighted and equally weighted total return and price indices.

classification has to be redone at times.

When creating industry indices, the number of different different companies and industries limits the choices of the indices. Similarly, the longer time period one tries to cover with the index, the more difficult it becomes. A very detailed classification can lead to having only few stocks in the industry category at question and, at extreme, having no stocks in the industry. Although one can create industry indices for shorter periods, here we select few industries which allow us to create industry indices from the beginning of the sample period. Analyzing the 30 companies listed at the beginning of the sample (October 1912), we can see that they represent mostly banking (11 companies), insurance (8 companies), transport/shipping (4 companies) and several other lines of industry (e.g., pawn shops, cot-

ton industry, telephone connections). For the analysis here, two industry indices have been created. Figure 6 shows the development of the value-weighted indices for banks and for insurance companies against the market index.

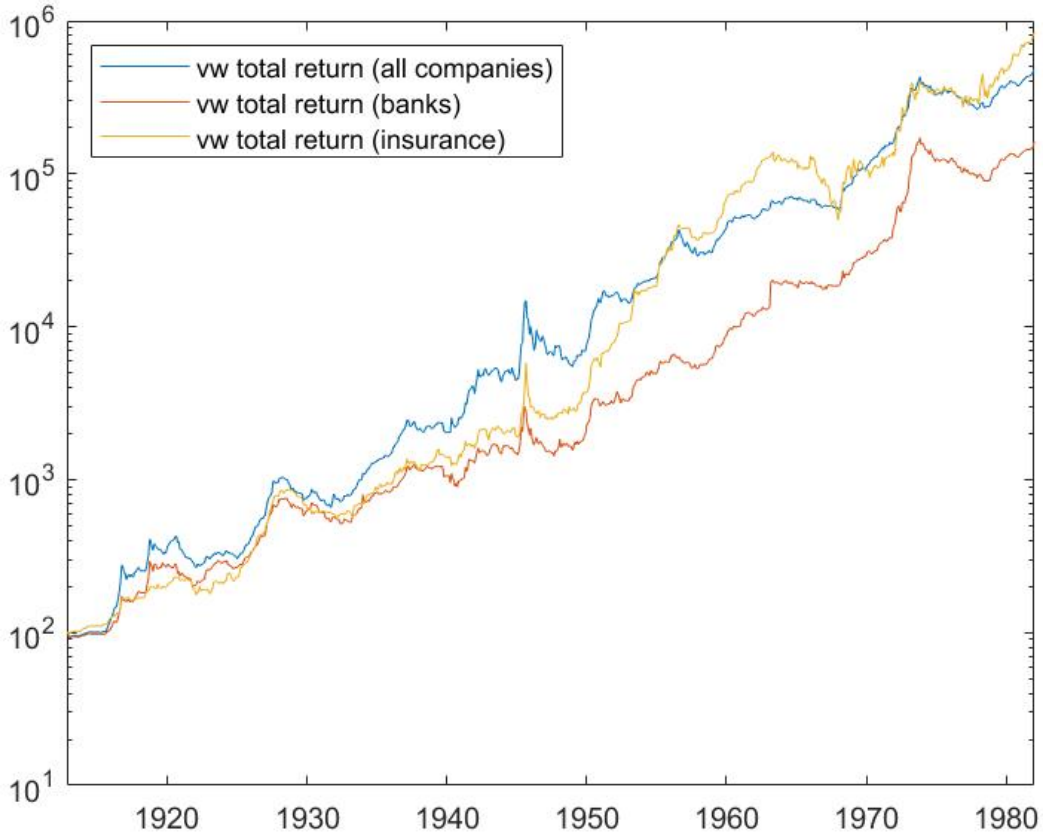


Fig. 6. **Industry indices.** Monthly values for value-weighted industry indices. General all share index has been included for comparison.

Somewhat surprisingly, banks' performance has been inferior to the general market index during the sample period whereas the insurance companies have fared better especially after the WWII. This could be due to heavy regulation on the banking industry during the sample period.

4.5. Additional considerations

As noted earlier, theoretical exercise values for the issue rights have been used in the analysis. To see how close these values are to the market prices, I compare the theoretical value of the rights with the observed market price. Evidence from the 1970s show that the deviation is not large.

One can also discuss if there are still some additional information or information usage that could be used to improve the index. Three pieces of information can be identified for this purpose. First, more detailed information on dividend rights for the newly issues stocks. If the issued stocks had a lower right for the dividend for the ongoing year, the value of the right should be adjusted to reflect investors' expectation for the dividend. As a result, the adjustment for the issues would be lower than the one used. However, this adjustment would apply to only part of the issues, and the adjustment for the issues excludes the time value of the right, so one can argue that the net effect of these omissions is quite small.

Second, information on directed issues. Namely, book equity values are not updated for the directed issues as the information is not available. However, its impact on the index would most likely be minor as the book equity is nonetheless updated at the end of the year. Thus, the only improvement would be the exact intra-year impact on the number of stocks.

Third, information is that of the delisting returns. Following most of the earlier studies, the delisting months have been excluded from the index. The main reason for this is the difficulty of finding information regarding the investors' returns in delistings. For those companies that suffered a bankruptcy, this approach clearly biases the returns upwards. However, in most cases the weight of these companies in the index is small before the actual bankruptcy. For the companies that were delisted as a result of the merger, the delisting return could actually be mildly positive.

For this study, I have analyzed the delistings. Altogether 136 stock series were delisted by the end of 1981. The most typical reason for delisting was a merger which happened in 42 (30.9%) cases. It was followed by company's own application for delisting in 34 (25.0%)

cases, lack of enough trading in 13 (9.6%), and bankruptcy in 9 (6.6%) cases. For 18 cases, the clear reason for delisting could not be found. Of these delistings, it is impossible to calculate delisting returns in 64 cases (46.4%) as the company continued its operations, but as a private firm. For 22 (15.9%) cases, I could not retrieve the information to calculate the return. For the nine backruptcies, the delisting return was assumed to be -100%. For the remaining 40 cases (29.0%), the delisting return could be calculated. Typically, the old owners of stocks were given news stocks in another company. Using the market price of this another company, we can derive the return. On average, the return is 38.7% when bankcruptcies are excluded, and 12.7% when they are included. Overall, I consider that there is some room for improving the accuracy of the constructed index, but impact on the returns is likely to be minor and not a decreasing one.

5. SUMMARY AND CONCLUSIONS

In this paper, the basic methodology to create stock market indices is reviewed and discussed. The main focus is on the methods one can use to create stock market indices for thinly traded stock markets. This is the case on all markets when one goes long enough back in time. It is evident that creating a research quality stock market index, especially one going back long into history, is not for the faint of heart – it requires a lot of time to collect data, and knowledge of index construction methods. A minor mistake or poorly justified approach to deal with thin trading can create serious bias into the index. There are still a number of countries in need of a high quality stock market index.

The Finnish stock market, Helsinki Stock Exchange, is used as the case in point. A new stock market index is developed for the market and compared with the old index of Nyberg and Vaihekoski (2010). The new index is based on a new database collected from the information on the official quotation lists. The ending of the sample period is also extended from 1969 to 1981. A number of industry indices are also created for the sample

period.

The choices made in the index calculation can have a major effect on the outcome. The results in this paper highlight the fact that perhaps even a bigger role is played by the information used to create the indices. For example, stock price data, even when accompanied with dividend data, is not enough to provide good estimates for the stock returns. Corporate capital actions can have a major impact on the investors' returns.

The new database can be combined with the existing stock market databases that are available to researchers from the beginning of 1982 forward. A combined database allows one to conduct historical economic research on a much longer time-span. A number of interesting research questions that require long sample period can be created but they are left for future research.

REFERENCES

Acharya, V. V. – Pedersen, L. H. (2005) Asset pricing with liquidity risk. Journal of Financial Economics, vol. 77, 375-410.

Andersen, S. E. (2011): The Evolution of Nordic Finance. UK: Palgrave Macmillan.

Annaert, J. – Buelens F. – de Ceuster M.J.K. (2012) New Belgian Stock Market Returns: 1832-1914. Explorations in Economic History, vol. 49 (2), 189-204.

Berglund, T. – Liljeblom, E. (1988) Market Serial Correlation on a Small Security Market: A Note. The Journal of Finance, vol. 43 (5), 1265-1274.

Berglund, T. – Wahlroos, B. – Grandell, L. (1983) The KOP and the UNITAS indexes for the Helsinki Stock Exchange in the light of a New Value Weighted Index. Finnish Journal of Business Economics, vol. 32, 30-41.

CRSP (2020) CRSP Index Methodology Guide. January 2020. Available at https://www.crsp.org/etsi.pdf.

Dimson, E. – Marsh, P. – Staunton, M. (2002) Triumph of the Optimists: 101 Years of Global Investment Returns. New Jersey: Princeton University Press.

Frennberg, Per – Hansson, Björn (1992) Computation of a monthly return index for the Swedish stock market 1919-1989. Scandinavian Economic History Review, vol. 40, 3-27.

Jokivuolle, Esa (1995) Measuring true stock index value in the presence of infrequent trading. Journal of Financial and Quantitative Analysis, vol. 30, 455-464.

Kock, Gunhard (1972): Osakesäästäjän käsikirja. Helsinki: Tietoteos Ky. Kock, Gunhard (1975): Pörssitieto 1975/6a. Helsinki: Pörssitieto Ky. Kock, Gunhard (1981): Pörssitieto

1980. Helsinki: Tietoteos Ky. Kock, Gunhard (1987): Pörssitieto 1986. Espoo: Kustannusliike Tietoteos Ky.

MSCI (2020) MSCI Index calculation Methodology. April 2020. Available at https://www.msci.com/eqb/methodology/meth_docs/MSCI_IndexCalcMethodology_Apr2020.pdf.

NASDAQ (2013) Rules for the Construction and Maintenance of the NASDAQ OMX All-Share, Benchmark and Sector Indexes. Version 2.1. Available at https://indexes.nasdaqomx.com/docs/NASDAQ_OMX_Equity_Indexes.pdf.

Nielsen, S. and Risager, O. (2001): Stock Returns and Bond Yields in Denmark, 1922-1998. Scandinavian Economic History Review, vol. 49, 63-82.

Nyberg, Peter – Vaihekoski, Mika (2010) A new value-weighted index for the Finnish stock market. Research in International Business and Finance, vol. 24 (3), 267-283.

Nyberg, Peter – Vaihekoski, Mika (2014): Equity Premium in Finland and Long-Term Performance of the Finnish Equity and Money Markets. Cliometrica, vol. 8 (2), 241-269.

Rydqvist, Kristian (2019) Performance and Development of a Thin Stock Market. Unpublished manuscript, available at SSRN: http://dx.doi.org/10.2139/ssrn.3477972.

Shumway, T. (1997), The delisting bias in CRSP data. Journal of Finance, vol. 52, 327-340.

Waldenström, Daniel (2007): Swedish stock prices and returns and bond yields 1856-2006. Unpublished manuscript. Available at http://www.riksbank.se/Upload/Dokument_riksbank/Monetar_hist/FinancialData_DW.pdf

Table 2: **Descriptive statistics.** This table presents descriptive statistics for the monthly stock market index returns. Percentage returns are used in the analysis with the exception of skewness and kurtosis which have been calculated using continuously compounded returns. Zero returns have been used when the stock market was closed. In Panel A, the new and the old indices are calculated using bid offers, the new index without search-back. The old index is from Nyberg and Vaihekoski (2010). In Panel B, the WI index was created in Berglund et al. (1983). In Panel C, all indices are created in this paper. The first one is the baseline index which uses bid offers but no search-back. The second index uses only observed prices. The third index imputes observed prices with the midpoint of bid-ask offers and, if not available, with the bid offers. The fourth index uses bid offers with intra-month search-back. All indices are value-weighted, all-share total return indices. Autocorrelation coefficients significantly (5%) different from zero are marked with an asterisk.

Variable	**Mean** Geometric	Arithmetic	**Std. dev.**	**Skew-ness**	**Excess Kurtosis**	**Autocorrelation** ρ_1	ρ_2	ρ_3
Panel A: October 1912 – March 1970								
New baseline index	1.030	1.17	5.39	1.88	11.22	0.246∗	0.080∗	0.026
Old index	1.027	1.18	5.69	1.83	10.98	0.195∗	0.035	0.056
Panel B: April 1970 – December 1981								
New baseline index	1.014	1.07	3.48	0.51	2.48	0.234∗	0.193∗	0.103∗
WI index	1.064	1.13	3.61	0.77	3.28	0.235∗	0.171∗	0.115∗
Panel C: October 1912 – December 1981								
New baseline index	1.027	1.15	5.12	1.85	11.75	0.245∗	0.089∗	0.033
Using prices	0.991	1.11	4.99	1.93	13.08	0.245∗	0.082∗	0.057
Using imputed prices	1.023	1.16	5.34	1.62	11.72	0.185∗	0.054∗	0.051
Using bids + search-back	1.032	1.17	5.33	1.83	12.35	0.235∗	0.051∗	0.019