PORTFOLIO CONSTRUCTION
FOR TESTS OF ASSET PRICING MODELS

Mika Vaihekoski[1*]

October 2003

PREPRINT DRAFT

Comments are welcome.

ABSTRACT

Portfolios are commonly used in finance literature to study asset-pricing models. In business practice portfolios are used to detect abnormal performance in certain asset groups or to construct reference assets. However, analyses on practical issues related to portfolio construction are surprisingly few. This paper presents and discusses issues related to portfolio return calculation from theoretical and practical perspectives. Special attention is given both to smaller and emerging stock markets. These stock markets often share common features like low liquidity, multiple stock series, and changes in foreign ownership restrictions that greatly affect portfolio construction.

Keywords: portfolio, asset, pricing models, tests, thin-trading, emerging markets, ownership restriction, multiple stock series, bias, Nokia, Finland, HEX, Matlab

JEL-classification: G11, G12

---

* LTT Research Ltd. and Department of Accounting and Finance, Helsinki School of Economics (HSE). E-mail: mika.vaihekoski@hkkk.fi, tel. +358 9 4313 8694, fax +358 9 408 417, Internet: http://hkkk.fi/~vaihekos/. I would like to thank Eva Liljeblom, Anders Löflund, and seminar participants at the Joint Finance Seminar of HSE and Hanken for their helpful comments. I also wish to thank Karl Felixson and Department of Finance and Statistics, Swedish School of Economics and Business Adm. for providing the Finnish return and trading volume data.

CONTENTS

1 INTRODUCTION

Portfolios of stocks are commonly used by financial researchers and practitioners in their analyses. Researchers use portfolios, e.g., to test asset pricing models or to measure risk premiums related to certain characteristics of companies (e.g., Fama-French type value premiums) or markets (e.g., market-wide bid-ask spreads, P/E-ratio, etc.).[2] Practitioners build portfolios to achieve their investment goals (e.g., minimize risk, maximize return). Besides investments, practitioners may use them to measure the performance of the target or a control group of companies or to study the ability of investment strategies (styles) to produce abnormal profits. Regardless of their purpose, portfolios can provide useful information about the stock market in a concise way that is helpful especially when the asset space is large.

However, even though there are several applications for portfolios, there are hardly any studies on the issues related to their construction from researchers point of view, as most researchers seem to have followed the approach adopted by their peers and practitioners have concentrated on the return/risk optimization. Especially issues related to portfolio construction on small and/or emerging stock markets have been given negligible attention (Vaihekoski, 2000, is one exception). These markets share several similar characteristics and features that affect portfolio construction and force us to consider different methods that can be employed in larger markets. The biggest differences that characterize the small or emerging markets are e.g. the thin trading of many companies, the small number of companies, and the high number of companies with multiple share classes.[3] Many small and emerging markets also have restrictions on the foreign ownership that has often led to the quotation of both restricted and unrestricted shares (see, e.g., Hietala, 1989, and Domowitz, Glen, and Madhavan, 1997, for a summary of the situations in Finland and Mexico, respectively).

We have two aims in this paper. First, we consider the issues related to portfolio construction. This paper extends and generalizes the analysis in Vaihekoski (2000). Second, we analyze in more detail some of the issues that are encountered in non-US

---

[2] Typically Fama-French type risk premiums are estimated as the difference on returns of asset portfolios formed on the basis of one firm characteristic (e.g., size, book-to-market ratio). Portfolios have been used in research also to estimate the risk premium for a non-traded risk factor by constructing mimicking portfolios.

[3] For example, in many smaller markets more than half of the companies have at least two (listed) stock classes whereas e.g. in the New York Stock Exchange only a small number of companies have more than one class of shares (excluding preferred shares who resemble more fixed income securities).

markets, especially on smaller and/or emerging markets. As an empirical application of the analysis, we provide documentation of portfolio construction using the Finnish stock market (i.e., the Helsinki Stock Exchange) as an example.[4] Finland is a good example for our purposes since it brings up most of the typical problems seen in portfolio construction.

This paper proceeds as follows. The problems and issues associated with the stock return calculation are briefly introduced and discussed in section 2. The main emphasis is on issues related to portfolio construction and they are discussed in section 3. In section 4, we present a case study of portfolio construction using Finnish data. The final section offers conclusions and some suggestions for further research.

[4] For example at the end of September 1997, the Finnish stock market was the fifth smallest market included in the MSCI world stock market index right after Norway and before Denmark. Finland represented less than one percent of the total market capitalization of the MSCI world index with its capitalization value of USD 87 billion. Although the Finnish stock market is small, it has traditionally been included among the developed stock markets (see, e.g., MSCI, 1998).

2 PORTFOLIO CONSTRUCTION

2.1 Portfolios vs. Individual Assets in Tests of Asset Pricing Models

Portfolios are commonly used in empirical research, especially to test asset pricing models (APM), since there are many advantages to using portfolios. First, portfolios reduce the noise in the individual asset return series due to non-synchronous trading and other measurement errors. This is especially the case when betas (risk sensitivities) are estimated as independent variables in a two-step cross-sectional estimation.[5] Second, portfolios make the estimation of the covariance matrix of the residual term easier when a time-series approach is used. This is the case especially when the number of assets, N, is "too close" to the number of time series observations, T, or when the number of assets is really large.[6] At the same time the portfolios' residual terms are often lower than for individual assets. Third, many models can be estimated only with a relatively low number of test assets (e.g., MGARCH-type models). Fourth, portfolios allows one to use longer time series than with individual stocks that come and go over time. Fifth, portfolios may have fairly constant beta even though individual stock betas change over time, if the portfolios are formed using a suitable criteria.[7] Portfolios also make the results much compact and easier to report.

Portfolios provide a convenient compromise between market/country indices and individual assets. Aggregate market indices may miss important aspects of the individual markets. Contrary to the U.S. studies most studies on non-U.S. markets have mostly used aggregated market or aggregate industry indices to study asset pricing models and something interesting may have gone undetected. Grouping test assets into portfolios on the basis of an interesting characteristic provide a chance for researcher to look beyond the aggregate market and gain interesting insights into the dataset. On the other hand, portfolios provide a chance to find out and highlight major facts that are hidden in the “noise” of individual assets. For example value-weighted portfolios – the most common type of portfolio used – highlight the most important

5 It is a well-known fact that the results of the beta-pricing and factor models are severely affected by thin trading (cf., Dimson, 1979).

6 Portfolios also make tests of factor models easier (cf., Chen, 1983).

7 See Shanken (1996).

aspects of assets which cannot be observed if individual assets are tested at the same time, since standard tests implicitly give equal weights to all assets.

However, there are also few disadvantages associated with the use of portfolios. First, the power of the tests against the null hypothesis decreases as the number of assets is reduced (Campbell, Lo, and MacKinlay, 1997). This is because the use of portfolios leads to the reduction in heterogeneity in the sample and to the conciliation of possible relevant return characteristics within portfolio averages (cf., Roll, 1977; Gibbons, Ross, and Shanken, 1986). In addition, researchers forming portfolios on the basis of characteristics, which prior evidence has found to be relevant, will be inclined to reject the null hypothesis too often due to the so-called 'data-snooping' bias (see Brennan, Chordia, and Subrahmanuam, 1998; Lo and MacKinlay, 1990).

Second, information is lost when assets are grouped together. For example, cross-correlation and other dependencies between assets are removed to a large extent from the data. Furthermore, the characteristics of the asset return distribution are altered. Finally, multivariate asset pricing tests are sensitive to both the choice of and number of test assets (Stambaugh, 1982). Ultimately, the decision of whether to test asset pricing models using a group of portfolios or a full set of individual assets depends on our research objectives. There is no definitive answer, and one has to weight the pros and cons of using portfolios instead of individual assets every time.

2.2 Criteria for Portfolio Construction

There are almost a limitless number of ways to construct portfolios, as we shall see later. The relevant question is: Are there any criteria or rules that can be used to guide us with the choices we inherently have to make during the process of building portfolios for tests of asset pricing models? Unfortunately, there have not been many studies of these issues.[8] However, we can draw some good practices from the earlier literature and provide some criteria of our own. The three most relevant criteria in order of importance are: investability, feasibility, and the efficient use of available data.

First, since we are interested in studying financial models that provide estimates for the expected returns, portfolios should be investable. The portfolios construction process should minimize the forward-looking bias, i.e. one should not use information

8 Carhart et al. (1996) is one exception.

that is available ex post. Otherwise, the estimates provided by the model are not valid since the investors did not have the information available at the time to form their expectations and make their investment decisions. This is an important criterion with many implications. For example, this criterion implicitly clearly states the need for survivorship-free data.

Second, other things being equal, the portfolios should be feasible; in other words, effects not relevant to the study should be minimized. This rule states that, other things equal, we should choose portfolios that, e.g., minimize transaction costs (e.g., fees, tax etc.) unless they are the focus of the study. This rule also relates the construction process to practitioners' needs to form portfolios economically in a feasible fashion avoiding transactions that are unrealistic in practice, e.g., due to the high costs involved.

Third, other things being equal, portfolios should be constructed to maximize the utilization of all available data. Simply stated, all data should be used as efficiently as possible. It is an important rule for small and/or emerging markets where the small number of stock series and the shortness of the sample time series is a common problem for researchers.

These three criteria are, of course, partially idealistic and, in practice, one has to make trade-off decisions. Therefore, one should not always at all costs choose portfolios that minimize, e.g., the forward-looking bias if this leads to wasting a lot of data. In some cases, the availability of the data can also set limits to our choices. For instance, if we do not have market values available, we cannot calculate value-weighted portfolio returns. In most cases, we have to consider the pros and cons of each choice and make case-by-case decisions as long as they can be justified. Naturally, we have to document our decisions as it can be argued that the results can be greatly affected by the way we have constructed our portfolios.

3 ISSUES IN PORTFOLIO CONSTRUCTION

3.1 General Issues

Once a decision to construct portfolios is made, one has to acknowledge that there are still several issues ahead that need careful consideration. In this sub-section, we present and discuss the most common issues and problems in portfolio construction that are common to all markets. In the next sub-section, we will discuss issues that are typically encountered in non-U.S. markets, especially in small and/or emerging stock markets.

3.1.1 Ex ante vs. ex post approach

The first question relates to our interest in studying ex ante returns (expected returns) or ex post returns (realized returns). Realized returns are frequently analyzed by technical analysts. Their interest is mainly in the predictability of the returns. From a practitioners' point of view, realized returns also reveal information of the strategies that could have been used. Most academic research, however, have studied models for the expected returns, even though they are for practical purposes in most cases forced to test them using realized returns. This clearly presents serious problems for the analysis, but it is not studies here (for a discussion of these issues, see, e.g., Elton, 1999). Thus, the portfolios should be formed on the basis of information that was available at the time of forming the expectations. In other words, studying expected returns forces us to utilize portfolio construction processes that maximized the investability of the portfolios.

Requirement for investability and the ex ante approach go hand in hand. Investability also implicitly includes the requirement for the minimization of survival bias. Survival bias can be a serious issue in portfolio construction. Survival bias comes usually through the use of a database that is backfilled. This can happen, e.g., when historical information is added to the database only for those companies that are listed today. Another source of survival bias is the selection of the companies for the portfolios on the basis of some external information that is available only for companies that have survived through the period. This is very common: most publications provide information only for those companies that are listed at the time of the publication. The good thing about survival bias is that it can be avoided with careful checking of the data once one is aware of it.

3.1.2 Market coverage

The second question is naturally what kind of coverage we want in our portfolios. Should we include stocks only from one country or from several countries? And even if we choose only one country, we may still need to select between multiple stock exchanges (e.g., NYSE vs. NASDAQ etc.). Furthermore, most stock exchanges have several lists (e.g., Main List and OTC-List). Sometimes we are forced to select even between the lists as they may differ so much (e.g., the other one may be far more liquid than the other) or the data is not available for all lists.

3.1.3 Portfolio types

The next question is what kind of portfolios should we construct i.e., what information or variable(s) should we use to group N assets into P ($P < N$) portfolios? In asset pricing tests, the test assets (portfolios) should according to Carhart et al. (1996): a) contain a spread in mean return, b) contain a spread in loadings on the true factors, c) have low cross-correlation, and d) be economically interesting and investable. In particular, to test factor pricing models (with one or more factors), we should construct the portfolios so that they show strong diversification in relation to the risk factors. In addition, if conditional pricing models are tested, we have to consider whether our interest is in the time-series and cross-sectional behavior of the stocks, since the classification information can often also be used as conditioning (time-series) information.

Portfolios can be formed on the basis of market information (e.g., size, trading volume, spread, or beta) or company specific information (e.g., industry classification or accounting measures). Of course a combination of the market and company specific information can also be used (e.g., book-to-market ratio). Sorting criteria based on market information allows for more frequent resembling of the assets into the portfolios while company specific information typically allows only for resembling of the portfolios once a year. Market information is often easier to use since it is usually readily available. Portfolios can also be formed using two-way ranking. In this case, companies are first ranked into groups on the basis of the first criteria and then within each group again on the basis of the second criteria.

3.1.4 Number of portfolios

Fourth question is how many portfolios should we construct. In asset pricing tests the question is related to the time series observations, T, as it sets the upper limit for the number of portfolios. In particular, when N is larger than T the estimated residual matrix is singular (Korajczyk and Viallet, 1990). The number of portfolios depends partly on researcher's interests, measurement problems in the return data[9], and it is also a matter of convenience. Usually the number of portfolios is not more than twenty. On the other hand, the number of portfolios can be lower if the number of available assets is small. The number of portfolio can sometimes be even as low as two, depending on our interest.[10] Usually, however, if the number of assets is limited, the number of portfolios should be kept as high as possible while keeping the number of assets in a portfolio above a certain minimum level.

If the number of portfolios is decided, we have to keep an eye on the number of assets in the portfolios over time so that it does not drop below a certain limit between the revisions. A related problem is caused if the number of companies quoted on the stock market varies remarkably over the sample period. If this happens, the number of assets in portfolios might be satisfactory for a part of the sample period but not all the time. However, sometimes having enough portfolios is more important. So one is faced with a trade-off and the decision has to be made on the basis of the availability of companies over time and requirements of the tests. As a rule of thumb, one should have at least five or more assets in a portfolio.

3.1.5 Weighting scheme

The next issue deals with what kind of weighting scheme should we use for the included in the portfolios. Following earlier research practices, we typically choose to calculate the portfolio return either as value-weighted or equally weighted average of the returns on assets included in the portfolio. There is nothing, however, stopping us from choosing any other weighting scheme as well. In some cases, it might be worthwhile to

---

9 This is especially important in small and/or emerging markets, where the return series show various degrees of autocorrelation and heteroskedasticity as one of the main ideas behind portfolio construction is to reduce these problems.

10 Two portfolios can be used to find out certain spreads in the returns of the assets, for example, in relation to size or market-to-book ratio (cf., Fama-French-type studies).

construct portfolios matching the weighting-scheme of the benchmark index. A common practice in many indices is, e.g., to include an upper and/or lower limit to the weight given to one company. Some these limit structures can require relatively complex programming.[11]

Value-weighted portfolios highlight the role of size (and investability), whereas equal-weighting highlight the cross-sectional variety of all available assets. However, the value-weighted approach has generally been the predominant choice for the research purposes. It also has the advantage of being equivalent to practitioners' return calculation. Furthermore, value-weighted portfolios imply lower transaction costs than the (contrarian) equally weighted portfolio construction process. Empirically, the value-weighted returns usually exhibit lower autocorrelation due to thin trading, especially on higher frequencies (see, e.g., Campbell, Lo, and MacKinlay, 1997). Value-weighted portfolios, however, have the disadvantage of requiring more information (i.e., the value of the equity capital) and work before they can be constructed.

If we choose to weight the stocks using some information, the weights should be available at time of weighting t-1 to make sure that the investability criterion is followed. The next question is when and how often the weights should be updated. In practice, the question is whether we should update weights at the end of the year or some another date and whether we should update weights once a year or more frequently. Typically, the weights are updated yearly to minimize the transaction costs or using the same frequency as the return data (i.e., daily, weekly, or monthly) to reflect the fact that asset pricing models are usually one period models. If researchers have chosen to update the weights only once a year, they typically have updated the weights either in the end of the year or in the middle of the year.[12]

3.1.6 Resembling process

[11] For example, the Finnish HEX Portfolio index is constructed with an upper weight limit of 10 percent per company. This index was developed partly because Nokia's weight in the Finnish all share HEX index was at times more than 50 percent when the mutual funds could not invest more than ten percent in one company.

[12] Updating weights in the end of a year can highlight the January-effect in returns, but on the other hand requires less information that the latter alternative. In practice, we need half a year longer time period to use middle year weights and if we have only a short sample, spending half a year's worth of observations is not worth the benefits.

Once we have the return and the weight matrices available, we need to construct an indicator matrix that indicates appropriate portfolio for each stock. This brings up a few additional questions. First, how stocks should be added into portfolios? Second, how often portfolios' content should be updated i.e., how often the assets should be resembled (re-ranked) into portfolios? Moreover, when should the stocks be added into portfolios?

Following the investability criterion, the stocks should be ranked into portfolios using information that was available before the return period started, i.e., if we measure portfolios return from time t-1 to t, we should use information Z_{t-1} available at time t-1. As a consequence, the return of the ith portfolio can be expressed as follows

$$r_{it}^{p} = \frac{diag(\mathbf{m}_{t-1}'\mathbf{I}_{i,t-1})\mathbf{r}_t}{\mathbf{m}_{t-1}\mathbf{I}_{i,t-1}'} \tag{1}$$

where r_t is a 1×N vector of asset returns, m_{t-1} is a 1×N vector of weights for available stock series at time t-1, $I_{i,t-1}$ is 1×N vector of portfolio indicator values at time t-1. Vector I gets value a value one, if stock n belongs to portfolio i. Weights are set according to the weighting scheme; typically, they are market values or ones (for equally-weighted portfolios).

From the previous equation, it can be immediately seen, that the indicator variable can be updated as frequently as the returns are calculated (e.g., daily, weekly, monthly). Company market value is one example of such information that allows us to update indicator variable as often as needed, but another question is whether we want to do so. However, in most cases we do not have any instrument variable available to update the indicator variable so frequently, and the suitable re-ranking frequency has to be set to be lower (usually annually).

The construction of the indicator variable is one possible source for the look-ahead bias if the information was not available in reality at the time of the re-ranking. Hence, if want to use accounting information, we cannot set the resembling to take place at the end of the year, because the information was typically not available until early next year. In principle, we should find out the actual release date for the company's financial statements, but this information is usually not available. Moreover, it is often hard to pin down the exact date, when the information became publicly available. Therefore, most US researchers have decided to resemble portfolios in the middle of the year (typically at the end of June) to minimize this problem and to reduce any possible effects caused by year-end re-ranking. However, this produces a problem if we have only a small number of companies or short time series available, because it wastes up to six months of data.[13] Therefore, the small look-ahead bias, caused by year-end (or close) re-ranking when the accounting information is used to define the I_{t-1} matrix, can be justifiable.

If we have the possibility to choose how often we want to update the I_{t-1} matrix (as with size portfolios), we have to consider the pros and cons of the alternatives. More frequent updating (resembling) naturally utilizes the data more efficiently since stocks are not excluded from the portfolios until necessary and stocks are included into the portfolios as soon as possible without a long waiting period.[14] Moreover, frequent updating reduces the possible survivorship bias in the dataset. On the other hand, more frequent updating requires more data and increases trading costs in practice, as the owner of a portfolio have to sell and buy stocks to keep the portfolio content in line with the included stocks. Moreover, for the size portfolios, frequent updating leads to a situation where successful companies tend to move up in the size order, i.e., they move towards the largest portfolios as they gain in their market value. Conversely the unsuccessful companies tend to move towards the smallest portfolios. This is bound to have an effect on the portfolio return characteristics.

Finally, we have to deal with new listed stocks and also with de-listing of stocks. The list of publicly listed stocks is evolving all the time – new companies (and stock classes) are listed and old companies disappear. A new company presents a problem when it should be included in the portfolios. Usually, new companies should be included as soon as we have their market values are available, i.e., with one period lag. De-listing of a company presents a bigger problem. As stated earlier it is very burdensome and sometimes even impossible to calculate the return for the final period. Therefore, researchers in practice usually exclude the last period return from the portfolios.[15]

---

13 Typically researchers have data available from the start of the year. Similarly, if a company is listed from January forward, it will be left out of the portfolios until July.

14 For example, if we use mid-year (end of June) resembling procedure, a company listed in mid-July has to wait several months to be included into the portfolios, whereas monthly resembling procedure would waste only half a month's worth of data. Similarly, a company de-listed in mid-May could not be included into the portfolios for the last eleven months using a mid-year resembling.

15 See Shumway and Warther (1999) for more information.

As we have seen in this subsection, there are a number of choices to be made with the portfolio resembling procedure. The alternatives available to a researcher are partly affected by the availability of data. If the amount of data is not the primary concern, researcher has more degrees of freedom to design her portfolios. However, if the data is of concern, one has to consider different trade-offs against one and another.

3.1.7 Company restructuring

We are often also faced with several minor issues when we are constructing portfolios. A very common problem is related to a situation where one private or publicly held company is merged into a publicly listed company, and they continue under the name of either one of the companies. Basically, we have to decide on a case-by-case basis whether the new entity forms a new company or not. If the new entity's business clearly differs from the old company, we should start new data series for the company (and loose two periods of observations) even though it may not be the case in our database. In many cases we have to make manual adjustments into our computer program to take care of these special cases. If the merger does not warrant for a new entity (say, IBM buys a small company), then we can simply continue with the old series.[16] A similar but opposite judgment call takes place when a listed company decides to spin off part of its operation (c.f., e.g., Palm spinning of its operating system as a separately listed company).

3.2 Market Specific Issues

3.2.1 Small number of companies

Small and/or emerging stock markets share several unique features that are of concern in portfolio construction. First, the number of listed companies is small. Whereas the New York Stock Exchange has thousands of listed companies, many small stock markets have only one hundred or so listed companies or less. Having only a few

[16] Looking at the stock exchange codes can be helpful in some cases, but in some cases they might lead us astray since the company might apply for a new listing code even if it is not necessary from our point of view. This leads to an interruption of the company's data series in most databases. In these cases, one has to re-combine the series in the portfolio construction process (in practice: computer program). Similar situation may arise when a company changes its name.

companies restricts the number of portfolios that can be constructed. Moreover, extra steps have to be taken to ensure that the sample data is utilized to the fullest.

The first step is to make every effort to get all available data into the database. Many databases, namely, collect only information of the largest companies covering, for example, 70 percent of the total market capitalization value. Once all data is available, one has to use it as efficiently as possible. For example, a good practice is to remove companies from the portfolios no sooner than one period before their market capitalization values hit zero (i.e., they are removed from the exchange list) if the last period return is not available. Similarly companies should be added to the portfolios one period after their market value becomes available.[17]

3.2.2 Multiple classes of stock

In many countries[18] companies can have dual- or even multiple classes of stock (e.g., common and preference shares[19]), whereas in the USA firms typically have only one class of stock.[20] Some of these classes can be unlisted (typically, controlling shares). Furthermore, there may be restrictions on (foreign) ownership.[21] If this is the case, unrestricted and restricted stock series may be listed separately. A good example of the multiple stock series can be found from the Finnish stock market. Prior to 1993, many companies had four listed share series (ordinary and preference classes of stocks, both

---

17 This also reduces the survival bias since alternatively we could select companies every year and typically only companies that have survived through the year are selected in order to avoid the problems of calculating last period returns.

18 E.g., in Europe the EU has tried to set some limitations to the use of multiple stock classes, but two or more stock classes are still frequently used in many companies. See Nenova (2003) for a recent survey of the countries with dual-class stocks.

19 Here common (or ordinary) stocks refer to the series with premium voting rights, whereas the preference stocks refer to stocks that do have voting rights albeit less than common stocks. Preference shares, on the other hand, have often the first right to the dividend, though often up to a certain limit (e.g., ten percent dividend on the nominal value). In some companies preference stocks may always have higher dividend right than common stocks.

20 Of course, we can have old and newly issued stocks listed separately if they have e.g. different dividend rights (usually newly issued stocks are entitled to the dividend starting next accounting period). The discussion that follows is applicable to some degree even to this kind of situations (in the USA).

21 Many stock markets have set restrictions on foreign ownership, especially emerging markets. These restrictions can be e.g. sector restrictions, limits on single or aggregate foreign ownership (e.g., in Finland foreign ownership was restricted to 40 percent of the equity of the company and 20 percent of the votes prior to 1993), and limits on foreign investments. To decide whether these restrictions affect the portfolio construction, we have to know whether we take the foreign or domestic investor's point of view. If these restrictions are binding, we may have to set limits on the company weights in our portfolios.

available as restricted and non-restricted series).[22] Beginning 1993 all stocks were declared free (i.e., unrestricted), and the maximum number of different series of stock came down to two (ordinary and preference).

The problem with multiple stock series is twofold. Typically small stock markets have only a few companies that are actively traded and having multiple stock series listed separately makes the thin trading effect on the return measurement even more evident as the available liquidity is divided between company's multiple stock series. Thus, selecting all series leads to serious thin trading problem with some series since the trading hardly ever is divided equally between the series. In many cases only one of the series is relative actively traded whereas the other series are typically very illiquid.

On the other hand, company's multiple stock series measure the same thing more or less and thus, the multicollinearity could be a problem in the econometric analysis. To illustrate this point, let us assume that we want to construct industry portfolios. As a result, one of our portfolios contains, say, eight stocks during certain time period. Closer examination may reveal that these eight stocks actually represent only two companies because both companies had four different stocks listed. In a similar fashion, we may find one company to be represented in four of our portfolios if we were to construct size portfolios.

To mitigate these problems typically only one stock series is selected to represent the company. If we choose to use this representative stock series approach, we are faced with the following problems: i) which one of the series to select as the representative series for the company, ii) how to measure the weight for the representative stock series if value-weighted returns are calculated, and iii) what to do with the unlisted series?

If the number of available companies is small, one should try to use of data as efficiently as possible and always select the representative series to be the one that has been available longest during the sample period. This guarantees the maximum number of stock series in the sample at all times. Second, to maximize the feasibility of the portfolios, one should aim to reduce the thin trading effect on the portfolio return measurement. Thus, if there are multiple series that have been available equally long

[22] See Hietala (1989) for more information.

during the sample period, we should select the one that minimizes the thin trading effect as suggested by the feasibility criterion. To do this we have to try to select the most liquid one of the series. Liquidity is very hard to define and the easiest choice is to simply pick the most valuable series measured with market capitalization since in many cases it is also the most liquid one. However, if we have the trading volume information available, we can use that information to improve our selection. A good method is to select series that minimizes the number of non-trading days during the sample period.[23]

Following these two rules can force us to make minor trade-offs, since selecting always the series with the longest data series can lead to selecting series that might end up being illiquid compared to the other series.[24] Moreover, the liquidity criteria can cause a small forward-looking bias if liquidity is defined over the whole sample period. This contradicts with the investability criteria, since liquidity criteria requires ex post information on liquidity. We could mitigate this problem to some degree by making the liquidity decision on a monthly (or yearly) basis and switch the representative series according to their past liquidity. However, switching back and forth between different series would change the distribution of the returns unpredictably since there are various reasons how the liquidity between the share classes may change and is therefore not recommended.[25]

Even though the representative stock series approach reduces the multicollinearity problem, it has at least two drawbacks. First, selecting a representative series (unless based on market value) requires extra work and additional information. For example, if the liquidity is used to select between series, we need time series of the trading volumes. Second, selected representative stock series could change when new information becomes available. For example, if series A is only slightly more liquid than series B in the original sample period, adding new data later may make series B more liquid. This forces us to re-construct the portfolios and makes

---

[23] If the trading is so active that the number of days with no trading is small, a turnover limit (>0) could be used instead. Alternatively, we could select the series with the highest volume accumulated over the whole sample period. However, this alternative could pick series that are thinly traded during most of the time with few large trades (e.g., due to the change of control in a company).

[24] If the listing periods are not overlapping, one can apply these rules only to the overlapping period, and select the available series for the non-overlapping periods.

[25] During, e.g., a corporate takeover battle, the series with the voting premium (common/ordinary shares) is often more traded whereas the preference series is often more traded at other times.

the historical comparison more difficult. Reconstruction does not have to be a big problem if we have build computer programs to do the portfolio construction. Moreover, this approach takes care that we have the most liquid series representing the company.

In countries where there are multiple classes of stocks listed, we should first use the above-mentioned guidelines to choose between different classes of stock (e.g., between common and preference stocks). If restricted and unrestricted stock series are listed separately, we first have to choose which one of them to use as the representative series for that particular class of stock. After selecting between the representing series from the unrestricted and restricted series within each class, we proceed normally and choose between common and preference classes of stock. However, in many countries restrictions on foreign ownership have disappeared during the last few decades. If constricted series are suddenly converted to unrestricted, we have to make further choices. The problem here is that in many cases it is difficult to say whether the unrestricted stock series is a continuum of the old unrestricted stock or (often) more liquid restricted stock price series. A good practice is to let the liquidity up to the conversion date decide, i.e., trading after conversion date does not affect the choice between unrestricted and restricted series, only the choice between ordinary and preference shares.

A good example of this situation is Finland's Nokia. Nokia's ordinary (common) shares became publicly listed in the Helsinki Stock Exchange in early 20th century. In 1981, Nokia issued a new series with lower amount of voting rights (ratio was 1:20) but with higher dividend rights. The ordinary series was denoted with a letter K (Kanta) and the new preference series with a letter E (Etu). On top of this, from 1984 to the end of 1992, both series were traded and quoted separately according to who could own them. Those shares that were available only for Finnish investors were known as restricted shares, and those that were available for both Finnish and foreign investors were known as unrestricted shares. As a result, during 1984–92 Nokia had four different stock series publicly listed: K restricted, K unrestricted, E restricted, and E unrestricted. To make the situation even more complex from our point of view, the Finnish legislation was changed in 1993 and all stocks became freely available to foreign investors. Thus, restricted shares were suddenly merged with unrestricted series (or vice versa, as we shall see later). And finally in 1999, both K and E series were combined together into one new A series (see Figure 1).

Since owners of any one of the four original series could end up with the A series without any action on her part, the representative series for Nokia can be built by combining these four different series. Let us start by assuming that we only have data from 1993 forward. Now, our problem is to choose between K and E free series. Officially K series were merged into E series (which was renamed as A series) and this would lead us to believe that the E (A) series has the longest time series and should therefore picked as the representative series.[26] However, the new A series can be interpreted as a continuum for either one of the K or E series.[27] Since both series now have equally long time series in our sample, we should select the most liquid one of them. As a consequence, we should compare the liquidity of K and E series only up to 1999 as A series' liquidity can be added into either one of the old series. If the accumulated liquidity measure favors, say, K series, our representative series is a combination of K series' returns up to 1999 and A series' thereafter.

If we have longer time series available, say, from 1984 onwards, the situation is basically similar, but now we have to compare four different series. Keeping in mind that A series can be considered as continuum to either one of the K or E series, we are faced with a similar situation in 1993 where the new free series can be interpreted as continuum of either unrestricted or restricted series, since the only change that took place was the change of the name of the series and trading code.[28]

As a researcher, we can calculate certain liquidity measures for free K and E series from 1993 onwards (remember to add A into both of the series or equivalently leave it out of the analysis). Now, we add K free series' liquidity into both unrestricted and restricted K-series' liquidity. Similarly for the E series. Finally, we have four different liquidity measures for four different series (see Figure 2). If the situation is as depicted in Figure 2, the E series is more liquid than K series measured over the whole

[26] The stock exchange merges two series often by simply removing the other series' trading code and continuing with the other one or by creating a new trading code altogether. As a researcher, we are not bound by the trading codes.

[27] Conversation day's return has to excluded, if any. In Nokia's case, the series were merged in 1:1 fashion, so no special conversation return was born.

[28] Naturally, one could argue that the free series cannot continue restricted series, but in many cases the restricted series were by far the most valuable and liquid series and thus, the difference in the true pricing process for the company's stocks is not that dramatic.

sample period. As a result, our representative series for Nokia is build from E restricted from 1984–92, E free from 1993 to 1999, and A from 1999 forward.

Multiple available stock series brings up also the problem of how to measure company's weight in the value-weighted portfolios. Usually, the market values are first calculated separately for each series after which they are added up to represent the company's weight in the portfolios. A problematic issue here is the unlisted stock series.[29] Basically there are two alternatives, either ignore them or value them somehow and add to company's market value. Using the latter alternative brings the valuation problem i.e. what price to use for the unlisted stocks. Using the market price of the listed stock series could be problematic since the unlisted stocks usually have premium voting rights as they are used to control the company. Moreover, in some cases the company might have had two listed series available. As a result selecting a price for the unlisted third series is problematic. Typically, one should pick the most similar one of publicly listed series or use their average price. On the other hand, there are several examples where the controlling stocks are later combined with the preference shares on a 1:1 basis.[30]

Regardless of the method used to calculate the company market value, we have to replace the m in the equation (1) with a new m^c-matrix. Adding series together can be done conveniently using a small loop in a computer (see Appendix C for an example).

3.2.3 Restrictions to foreign ownership

Restriction on foreign ownership can also affect the portfolio construction through weights. Namely, if we test asset pricing models from foreign investors point of view, it seems reasonable that the weight cannot exceed the restrictions set on the foreign ownership. In some cases, some companies may not even be available for foreign investors. A quite similar problem is caused by the government ownership in publicly listed companies. This reduces the so-called free-float of the stocks and hence,

---

29 A good example is for example Sweden's Ericsson whose B shares are listed publicly whereas the A series with superior voting rights is not publicly listed.

30 Recent examples from Finland include, e.g., Nokia and Huhtamäki.

comparing full weights versus benchmarks that take into consideration only the free-float portion of the company can be problematic.

3.2.4 Market domination

Small stock exchanges can also have situation where only one or few companies clearly dominate the market. E.g. in Finland Nokia's market value represents at times more than half of the total market capitalization. This raises two kinds of issues. First, in asset pricing tests we might want to consider removing this company from the portfolios especially if the number of portfolios is small. Second, if a company grows rapidly in size then using lagged market values as weights in the portfolio construction biases the return on that particular portfolio downwards. This can seriously affect tests of asset pricing models when for example monthly returns are used if the market index is calculated using weights lagged only by one day.

3.2.5 Market specific characteristics

Finally, there are bound to be market specific occasions or characteristics that must also be dealt with. Changes in the legislation can have huge impacts on the stock market. Governments are typically concerned with the foreign ownership of the companies and thus they can set additional restrictions on foreign ownership to protect national interests.[31] Companies can also change the characteristics of their stocks themselves. For example, a company could set up new series, change the ownership restriction status of its outstanding series, or combine two or more series.[32] Common to all these problems is that there are no clear rules how to proceed. Thus, the decisions have to be made on a case-by-case basis.

3.3 Data issues

3.3.1 Price series

31 For example, in Finland all stocks of the insurance companies were converted to restricted stocks unless owned by foreigners at the time of the conversion.

32 In Finland, see, e.g., Asko and Unitas.

Portfolio construction requires naturally that we have available return time series for the stocks selected to be in our portfolios. However, in many cases we do not have the return series readily available but the price information and we have to calculate return series ourselves.[33] Unfortunately this is not a trivial exercise – there are a number of issues to consider.

First, we have to consider what kind of price information should be used as the basis for the calculation. Typically we have several different price observations available for each stock like, e.g., best bid and ask offers at the end of the day, and first (opening), last (closing), highest and lowest transaction prices. If intraday data is not available, researchers have usually preferred to use average transaction price measured as the arithmetic average of the highest and lowest transaction prices[34] or the closing transaction price. These so-called realization methods have the advantage that they both reflect real trading opportunities (cf. Roll, 1984). The latter one also has the advantage that it could reduce the non-synchronous trading bias when compared to a market (index) returns.[35]

So far no problems. However, small and emerging markets frequently exhibit thin trading where stocks have one or more of the price observations mentioned above missing. E.g., a stock could have no transactions during a day (closing price is either zero or equal to the closing price during the previous trading day depending on the database) and sometimes even the bid or the ask price is also missing if the illiquidity gets really severe.

If we have bid and ask offers available but no transaction price, our alternatives differ depending whether we have chosen to use the average transaction price or closing price approach. If we have chosen to use the average of highest and lowest transaction prices, we typically use analogously average of the bid and ask offers to calculate the missing price observation. However, if we have chosen to use closing prices, we can either use the last available transaction price as the closing price until the next transaction takes place, or we can use the average of the bid and ask offers as

---

33 Exceptions are research databases such as the CRSP database in the United States. In many cases, however, the data is collected directly from the stock exchange or data vendors such as Reuters or Datastream and we have to calculate returns ourselves.

34 In some cases, we may even have enough data to calculate average (volume-weighted) transaction price using all transactions within a day.

35 Index observations are often taken from the end of trading days whereas average transaction prices often reflect the midday observations.

earlier. The first alternative has the disadvantage that the non-synchronous trading bias is greater compared to the latter method as the last transaction price observation can be from several days back. On the other hand, using only transaction prices reflects more accurately real trading opportunities.

The situation is more complicated if even bid or ask offer (or both) observation is missing. In these situations, one is usually forced to use previous price observation i.e. previous closing or average trading price, or previous average bid-ask offer. Ultimately, the choice between average price method, closing price method, or a combination of them has to be made after their benefits and disadvantages are considered. The way market index is constructed also affects the choice since totally different methods to construct market index and portfolios can induce measurement problems into the analysis.

Once we have constructed the price series we have to consider what kind of corrections we should make to the price series since price series are typically not applicable for return calculation as such. Usually we have to correct them at least for cash dividends, stock dividends, issues, and splits.[36] Sometimes it can be really hard work, not least because we may have to track down all the correct dates for these events if they are not provided in the database.

3.3.2 Return measurement

Finally we have to consider how we measure returns. Traditionally there have been two main alternatives: percentage returns and logarithmic returns (i.e. continuously compounded returns). Percentage returns have the advantage of being frequently used in practice. However, log returns (i.e., logarithmic relative of the price) has become de facto -standard in the financial research. There are several reasons for this. First, multi-period returns are just sums of period returns. Second, log returns reduce the bias in returns induced by bid/ask spread and price discreteness if compared with the proportional percentage returns (see Mucklow, 1994). Third, compound returns often exhibit higher degree of normality than percentage returns due to its symmetry. Fourth, log returns reduce the heteroscedasticity found in most of the stock return

36 The problems related to the return calculation (e.g., how returns are adjusted to splits, dividends, and issues) are not discussed here. There are several good articles discussing return calculation problems (see, e.g., Roll, 1984). For more information how these corrections are done in the Finnish market, see e.g. Berglund, Wahlroos, and Grandell (1983) and Hernesniemi (1990).

series. However, the use of log returns does not go totally without problems: log return on portfolio is not exactly the value-weighted sum of log returns on assets in the portfolio (cf., Campbell, Lo, and MacKinlay, 1997), but the bias is typically small.

Besides choosing the way we want to measure returns, we have to select whether we want to study nominal net returns or perhaps something else. In some cases, we might want to use gross returns, that is, one plus net return. In many asset pricing tests, we analyze returns in excess of risk-free rate. In some cases it is also of interest to study real returns instead of nominal returns. Both excess and real returns require additional information from us.[37] Furthermore, we have to consider whose point of view are we measuring the returns. If we measure returns from foreigner's point of view, we have to consider changes in the currency exchange rates. Normally, this is not a problem but in some emerging countries the currency is not freely traded or the information not otherwise available. In some countries, there might even be restrictions on the capital outflows. Whether or not we want to consider the currency effects, we have to know whether the returns are adjusted for the taxes (cf., e.g., MSCI's indices) and how the dividends are handled (e.g., are they just used to adjust the index or are they re-invested in the market or in the dividend paying stock).

Finally, we have several practical problems. Calculating returns for the last listing period (so-called delisting returns) is always problematic. Namely, there can be several different reasons for delisting (e.g. bankruptcy, merger, buy-out, acquisition, liquidation, or migration to another exchange) and it makes the return calculation almost impossible in practice when we have a large number of assets in our sample and the return are not readily available. Typically researchers have simply removed delisting returns from the sample, but this can possible have large implications for the empirical results.

In most cases we also want to aggregate daily returns over longer periods (e.g., weekly or monthly). If we are using continuously compounded returns, we can just accumulate returns over the desired period. In particular, to get monthly returns we just accumulate daily returns over the whole month. Almost all researchers have chosen to calculate monthly returns by accumulating daily returns from start of the

[37] For example, real returns require time series for the inflation. This series is not always easily available (e.g., for emerging markets) and/or reliably calculated. One also has to consider the choice between preliminary inflation figures against final inflation figures, often available months later. If one chooses to use the announced preliminary figures, they might be hard to find as most databases adjust their figures to reflect the final figures.

month to the end (i.e., 1.–28./30./31.). As far as we know, there are no studies that would have studied how this affects the returns. On the other hand, if we want to calculate weekly returns, we usually accumulate returns from Wednesday to next Wednesday to avoid the day of the week effect found in many markets.[38]

3.3.3 Number of stocks

We need the number of stocks to calculate value-weighted portfolio returns. In some databases the market values are given, but in many cases we have to calculate company's market values ourselves by multiplying the number of stocks with the market price. There are a number of problems related to the number of stocks information. Unfortunately, there are no clear rules how to handle these problems, but in case one should be aware of them and try to deal with them objectively.

First, the dual-listing of the companies causes a problematic situation. The problem here is choosing the right number of shares. In most cases we cannot use the total number of shares to calculate the market value if only small portion of shares where initially sold to a new market and if there exists some impediments for free flow of shares from one market to the other. Typically one can use either the initial number of shares issued to a new market or the market's proportion of total trading as the basis for market value calculation.[39] A related question is: how to account for the ADR-programs? Should the number of stocks represented by the ADRs be deducted from the number of stocks available for a particular market?

Second, we have to find ways to deal with those, often minor, changes in the number of stocks that are not caused by splits or issues. These kinds of changes are often due, e.g., to exercised stock compensation options or convertible bonds. This is a problem especially in practice. The data might not be readily available or it is not properly updated in the databases. For example, the databases tend to update the minor changes only once a year or so.

---

[38] In practice, weekly return calculation can be a little bit problematic since a Wednesday can be a holiday etc. This problem can be overcome e.g. by constructing an indicator variable (for example in a spreadsheet) to show when the week changes. After this it is easy to calculate the weekly returns.

[39] A case in point is the dual-listing of several Swedish companies on the Helsinki Stock Exchange during late 1980s and early 1990s. For example, if ASEA's (one of Europe's largest companies) market value were calculated on the basis of total number of shares (millions) instead of the number of shares sold to Finland (80,000), it would have had a huge impact on the portfolios.

Third and more fundamental question is how company share repurchases should be considered. If a company has X stocks outstanding and it buys Y stocks back with an intention, e.g., to cancel them, the question is whether or not we should use (X–Y) as the number of stocks to calculate the market value. Most often, however, a good practice is to use X for as long as these stocks really exist. We cannot be sure of the company's intentions and it may decide to use the stocks as a currency in a business deal.

Finally, we have to consider the free float issue. A number of indices have started to include into their market capitalization weight calculation only those stocks that are not owned by the government (i.e., the so-called free float). However, this approach is partly questionably, since in many countries the government is free to sell part of their holdings (or all) if the price is right. Moreover, it raises the question what organizations should be considered as part of the government. For example, should we account into the free float the stocks owned, e.g., by the government-led organization that takes care of citizens' mandatory pension savings or not? What about organizations that are mostly funded by the government? In many cases, it is simpler to account for all the stocks that have been issued.

4 CASE: FINLAND

To demonstrate the discussion in the previous section, we use Finland as an example of a portfolio construction in a small stock market. Finland is a good example for our purposes since it brings up most of the typical problems in portfolio construction. Two sorting criteria are used to construct portfolios: size and industry. Monthly value-weighted returns are calculated for both groups of portfolios with monthly updated weights. Stocks are resembled monthly for the size portfolios and annually for the industry portfolios. Sample period is from 1987 to 2000.[40] Thus, we have 168 monthly observations.

4.1 Data

In order to calculate portfolio returns, we need first of all daily return series adjusted for all necessary events (dividends, splits etc.). To make the selection between different classes of stocks, we also need the daily trading volumes (number of shares traded). Since we want to calculate value-weighted portfolio returns, we need the stock price series as well as the number of stocks outstanding for all listed share series.[41] In addition, we need the industry classification data on the company basis to calculate the industry portfolio returns.

4.1.1 Stock prices and returns

Stock returns are calculated for all stocks quoted on the Main List[42] of the Helsinki Stock Exchange (HEX). Stock returns are measured as the continuously compounded return on stock price indices, i.e., the logarithmic relative of the price. This price index is based on last available closing price[43] provided by the HEX and it is corrected for

---

[40] This period is chosen because the open and transparent market for certificates of deposit began 1987 in Finland. The birth of the money market provided the market participants a fair and equal chance for (virtually) risk-free investments and thus, gives the researcher a change to calculate excess returns.

[41] Note that if we had price and number of stocks series from the last trading day of year 1986 we would not loose the first month of 1987 in the portfolio building process. This is because the weighting information has to be available prior to the return calculation period. Note also that since we are calculating monthly returns, it is enough if we collect the price and number of stocks series for the last day of the month. This reduces our job considerably if the number of stocks series is not readily available as it hardly ever is.

[42] The Main List is chosen because the data is readily available and the difference to the OTC and Brokers' List is large in terms of liquidity, size, and transaction costs.

[43] Due to the reporting system of the Helsinki Stock Exchange, prior to 1989 the price for a particular day is calculated as the average trading price. After 1989 the last available closing price is used. If no trades occurred, the bid quote

cash dividends, stock dividends, issues and splits (cf., Berglund, Wahlroos and Grandell, 1983).[44] We select the last closing price (i.e., so-called realization method) as a basis for return calculations, because it was readily available and it has the advantage that it reflects real trading opportunities.

Equity market portfolio returns are calculated combining two different return indexes. Since the HEX total return index is not available prior to 1991, the WI-index is used for the period 1987 through 1990, and the HEX return index thereafter. They both measure the general price movement in the Helsinki Stock Exchange and they are both value-weighted and corrected for cash dividends, splits, stock dividends and new issues.[45] Contrary to the price indices used to calculate stocks' returns, the market index is calculated using the average of bid and ask quotas if no trades take place during the day (if either bid or ask quota is missing, a substitute value is used). All companies are included in the index with their full market value. Market values (weights) are updated daily in the HEX index as far as the price movements goes, but the number of stocks for a company are updated quarterly with some exeptions.[46]

4.1.2 Trading volume

As pointed out earlier, if multiple stock series are available, we have to select a representative series for the company from the available series. Appendix A lists all companies in the Finnish stock market that have had multiple listed stock series during our sample period. To make decision easier, we have grouped the stocks on the basis whether we have to choose only between unrestricted or restricted series, preference or common series, or a combination of the these two.

---

is used or the previous price, if even the bid quote is missing. See Hedvall (1994) for more details on the trading system in the HEX.

[44] The return data was provided by the Department of Finance and Statistics, Swedish School of Economics and Business Administration. We compared this data with the only other academic data source in Finland, namely the University of Oulu, and found the returns to be identical. However, comparing the data with that of Datastream revealed a number of errors in their data series.

[45] The main difference between the WI-index and the HEX-index is how dividends are handled. In the WI-index the dividends are reinvested to the particular stock, whereas in the HEX-index the dividends are reinvested in the market. Other smaller differences include, e.g., what price is used when no transaction price is available. The WI-index is described in more detail in Berglund, Wahlroos and Grandell (1983) and the HEX-index in Hernesniemi (1990) and HEX (2000).

[46] For more information how the index is calculated see Hernesniemi (1990) and HEX (2000).

In some cases we can select the representative series using the principle of selecting the series with the highest number of observations, i.e., the series that has been listed longest during the sample period. But there are still surprisingly many companies with multiple stock series to choose from once we have used our first principle.

When we have an equal number of observations in our sample period, we aim to minimize the thin trading effect in our sample and select the most traded series. To do this we need a measure of liquidity. Here we use the trading volume information to calculate the number of days without trading activity. Namely, the daily number of shares traded acts as an indicator variable where a value of zero indicates that no trades have occurred during a day whereas a positive number indicates that trades have occurred. Adding together the number of days of without trading throughout our sample period gives us our measure of liquidity that is used to select the most liquid series of the available stock series.

Using the liquidity measure and other rules outlined in the previous sections, we can select the representative series for all companies. Results can also be seen from Appendix A. Since all stocks became unrestricted (free) in Finland after 1993, we have to keep in mind what was said in the previous section about a situation where we have to choose between preference and ordinary share classes and both unrestricted and restricted series are available for either one of the series or both.

4.1.3 Market values

To calculate the value-weighted average of the return on the representative stock series we need market values for all companies. Since we already have stock price series[47], the only missing information needed to calculate the market values, is the number of stocks outstanding for each stock series. The market value of a particular stock series is the stock price times the number of shares outstanding. Once we have the market values calculated for each series, we simply add the market values of multiple series – if listed – together to get the market value for the company.

---

[47] Note that since we are using closing prices which, in turn, are based on last available transaction prices, we have to make sure that the closing prices in our database are adjusted for splits and stock issues even if no trades have taken place. In other words, if the last "real" closing price was from January 30th and a split happens on 31st but no trades took place, the closing price for January – if not adjusted – could still reflect the un-split value from January 30th.

Unfortunately there is no good source to find the number of stocks outstanding for each series, especially for the early part of the sample period. HEX provides this information for parts of the sample period. Therefore, we have collected the missing data from books by Kock, KOP, and OKO. The number of stocks outstanding is corrected for splits, subscription and bonus stock issues, tender issues and directed issues.[48] However, some minor adjustments, like stock repurchases and stocks issued due to convertible debt and options, may not be taken into account until later when the accounting year has ended and more detailed information became available.

Weights are then calculated as the total market value of a company's equity capital at the end of previous weight revision period divided by the sum of the market values of all companies included in the portfolios at that time. Weights are updated in our case monthly. This means that portfolio return for a time t to t+1 is the value-weighted average using weights calculated as of t.[49] This is done in order to guarantee the investability of the portfolios and that no forward-looking bias is induced through the weights.

4.1.4 Portfolio sorting information

Size portfolios are constructed by ranking companies into P portfolios on the basis of their total market values as of time t. Since the number of companies available each month varies between 50–106, we cannot construct many portfolios. In this case, we select P to be seven. Ranking (resembling) is done every month at the same time as the weights are updated. All companies with a market value available are always assigned into portfolios, unless their market value is zero next period (i.e., last period returns are excluded). The number of companies in the portfolios has been on average more than ten for all size portfolios. Smaller portfolios have higher number of stocks on average due to the way the size portfolios were constructed (see Table 1).

Panel A in Table 1 also shows the minimum, maximum, and average relative market weights for each one of the size portfolios. Relative market weight is the

[48] The number of stocks outstanding does not, however, take into account the number of stock currently held by the company or its subsidiaries. In addition, the amount of new stocks due to the convertibles is added at the end of the year. Occasionally, some directed issues may go unnoticed, but they are updated at the end of the accounting year. Note also, that the number of stocks does not include unlisted series.

[49] To be exact, the weights are calculated at the end of period t-1 i.e. at the end of the previous month.

portfolios market capitalization valued divided by the total market capitalization value. Similar minimum, maximum, and average statistics are also given for the number of companies in each portfolio. Averages are calculated as the time series mean of the month-end values.

The largest size portfolio accounts, on average, for nearly 60 percent of the total market capitalization value whereas the smallest size portfolio only accounts for less than one percent. If we study its percentage share on a monthly basis (results not reported), we can see that there have been huge variations in its relative weight. Namely, it has varied from 48 percent to over 80 percent.

Industry portfolios are created by sorting firms at the end of each calendar year[50] using the industry classification given by Talouselämä.[51] Talouselämä classifies a company by the business that accounts for more than 60 percent of its net sales during the previous year. If none of the businesses exceed the limit, the company is regarded as a multi-branch corporation. If no other information is available, Kock's book and to some degree also subjective consideration are used.[52] Appendix B shows the company classification used in this study.

Panel B in Table 1 shows the relative values for the different industry portfolios. The results reflect the industry structure in the Helsinki Stock Exchange. Again, many industries show surprisingly large variations in their relative values. Especially the Metal and Electronics industry, and the Multi-business industry portfolios show huge differences in their minimum and maximum relative values. This is mainly a result of the reorientation of the Nokia's business focus from multi-business to pure telecommunication business during the sample period.

Since we have only a limited number of companies per industry available, we are forced to combine subjectively some closely related industries to form bigger portfolios. Unfortunately, this leaves certain companies out of the portfolios since their

---

[50] An exception to this rule are those companies became publicly listed during the calendar year. To maximize the market coverage, they are included in the portfolios as soon as possible using the industry information available at the time of listing. Similarly companies are kept in the portfolios until they are de-listed.

[51] Talouselämä is the leading weekly business magazine in Finland. They publish yearly a roundup review of the biggest 500 Finnish companies.

[52] However, one should notice that although the objectivity has been the goal in every level of the portfolio construction, the industry ranking remains partly subjective, since no SIC-codes are available for the Finnish stock market and all industry classification is partly subjective.

industries cannot be justifiably combined with the other ones. In the end we have seven different industry portfolios. As a rule, we have tried to build portfolios that have on average more than four companies calculated as the monthly average over the whole research period and at no time the number of companies in a portfolio could be less than three (with one exception).

The results are shown in Panel B of Table 1. Typically industry portfolios hold five to ten companies. Naturally, the number of companies in industry portfolios varies a lot reflecting the number of companies listed in each industry sector. The Housing and Construction industry has a short period when only one company was included in the portfolio. This is unfortunate, but to construct meaningful portfolios in the Finnish stock market we have to accept this.[53]

4.2 Construction process

Once we have all necessary data available, we can proceed to the portfolio construction process. Appendix C shows the computer program code needed to construct the portfolios. The language is based on Matlab's, but we have made modified to make it more generic and readable.

The program has four different stages.[54] The first stage calculates the company market values from the asset class specific market values. To do that, we need a $1 \times N$ vector that indicates the representative series for a company. Say, the representative series is four (i.e., column four in matrix marketvalue) and series two through three are other classes of shares for that particular company, then the second, third, and fourth values of the indicator vector are four.

The second stage calculates the weights to be used in the portfolios. As noted earlier, the last day observations are not available, and thus the last day market values are dropped from the matrix.

The third stage calculates the monthly value-weighted returns for size portfolios. We start by constructing a matrix with values indicating the appropriate size portfolio (zero indicates that the asset is not to included in the portfolios). While

---

53 This is quite short period: January 1987-February 1988. On average the number of assets is more than four, which is quite acceptable for a small stock market.

54 Naturally, these stages are preceded with a program that does all the necessary adjustments to the data. In case of Finland, there are a number of adjustments to be made. For example there are a number of companies that need special attention (e.g., Unitas/Merita/Nordea, HPY/Helsingin Puhelin).

constructing this matrix, we have to keep in mind that in a rare case, two companies can have similar market values, and therefore the program has to be built to handle cases like this.

The fourth stage simple calculates the industry portfolio returns. The only difficulty here is to adjust the company specific Y × C matrix, where Y is the number of years and C is the number of companies, into T × N matrix using company vector. Each value of the vector company simply indicates the appropriate representative asset series (column) for the company information. We could have constructed a Y × N company matrix in the first place into our original database, but since we might need company specific information also in other places, it is nice to have a routine that quickly assigns company information into representative stock series.

4.3 Descriptive statistics

Table 2 presents the descriptive statistics for the portfolio and market index return series. Mean and standard deviation are annualized in the Table. In panel A we can see that contrary to the US findings, the mean realized return is generally lower for the smaller companies. This is probably due to the recession in the Finnish economy in the early 1990s, which seems to have hit the small companies more severely than larger companies since smaller firms are typically more dependent on the success of the domestic markets.

Surprisingly, we find that none of the size portfolio can beat the market index. The largest size portfolio comes close, but it is still below the market index. This is somewhat puzzling since we know that the market index and the largest size portfolio are mostly driven by Nokia, who accounts for sixty percent of the capitalization value of the largest size portfolio and nearly half of the total market capitalization value at the end of 1998 (see Figure 3). There can be several explanations for this, but the most likely reason is how the weights are updated in the portfolio and in the market index return calculation. Size portfolios have weights that are updated only once a month, whereas the weights for the market index are updated daily. This can have a huge impact on the returns if there is one company that dominates the market and exhibits constant growth in its relative market value. And it turn out that in Finland all other companies included in the largest size portfolio exhibit below market returns as a group (see Figure 4), which further emphasizes the difference.[55]

If we study the industry portfolios, we can see that the Metal and Electronics industry together with the Multi-Business industry have achieved the highest returns to investors. Recent economic and banking crises are reflected in the Banking and Finance as well as in the Housing and Construction industry portfolios.

Most asset returns also show evidence of non-normality. The Bera-Jarque test for the normality rejects the null hypothesis of normal distribution for eight out of the thirteen portfolios. Furthermore, there is evidence of significant first and in one case even third-order serial correlation in the portfolio returns.

Panel B shows the cross-correlation matrix for the market and portfolio returns. As expected, the largest size portfolio exhibits the highest cross-correlation with the market index (0.969). Size portfolios exhibit in general higher cross-correlations with each other than industry portfolios making them less interesting from diversification's point of view.

55 The difference is caused by the fact that Nokia's weights always lag behind the market index and thus the gains are not fully accumulated into the portfolio's gains. On the other hand, all other stocks in the largest size portfolio have "too high" weights when compared to the market index, which further lowers the return on the largest portfolio.

5 CONCLUSIONS

In this study, we have presented and analyzed the issues that are associated with portfolio construction. There are surprisingly many issues to consider when one constructs portfolio for tests of asset pricing models. Especially this is true for small and/or emerging stock markets. Unfortunately, there are not clear rules to guide portfolio construction. This study, however, have attempted to provide common criteria that can be used to evaluate and solve the issues. As an example, we provide a detailed case of portfolio construction in a small stock market, namely Finland.

This study has shown also that any kind of portfolio building profess leaves a lot of choices to the researcher. This raises questions how much are the empirical results affected by the researchers choices. The effect may be bigger than anticipated, but the answer is left to further studies.

REFERENCES

Berglund, T. – Wahlroos, B. – Grandell, L.: 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, 1983, 32, 30–41.

Brennan, M. J. – Chordia, T. – Subrahmanyam, A.: Alternative factor specifications, security characteristics, and the cross-section of expected asset returns. Journal of Financial Economics, 1998, 49, 345–373.

Campbell, J. Y. – Lo, A. W. – MacKinlay, A. C.: The Econometrics of Financial Markets. USA: Princeton, 1997.

Carhart, M. M. – Krail, R. J. – Stevens, R. L. – Welch, K. D.: Testing the Conditional CAPM. An unpublished manuscript, 1996.

Chen, Nai-Fu: Some Empirical Tests of the Theory of APT. Journal of Finance, 1983, 38, 5, 1393–1414.

Dimson, E.: Risk Measurement when Shares are Subject to Infrequent Trading. Journal of Financial Economics, 1979, 7, 197–226.

Domowitz, I. – Glen, J. – Madhavan, A.: Market Segmentation and Stock Prices: Evidence from an Emerging Market. Journal of Finance, 1997, 52, 3, 1059–1085.

Gibbons, M. R. – Ross, S. A. – Shanken, J.: A Test of the Efficiency of a Given Portfolio. Econometrica, 1989, 57, 1121–1152.

Hedvall, K.: Essays on Market Microstructure of the Helsinki Stock Exchange. Helsinki: Publications of the Swedish School of Economics and Business Administration, 62, 1994.

Hernesniemi, H.: HEX-indeksi [HEX-index]. Helsinki: Elinkeinoelämän tutkimuslaitos, Sarja B:68, 1990.

HEX: Calculation of HEX indices. Helsinki: Helsinki Exchanges Plc., 2000. Available from http://www.hex.fi/.

Hietala, P. T.: Asset Pricing in Partially Segmented Markets: Evidence from the Finnish Market. Journal of Finance, 1989, 44, 3, 697–718.

Kock, G.: Pörssitieto. [Stock Guide.] Finland: Tietoteos ky, 1988–1998.

Korajczyk, R. A. – Viallet, G. J.: An Empirical Investigation of International Asset Pricing. Review of Financial Studies, 1990, 2, 4, 553–585.

KOP: Pörssiyhtiöt [Listed Companies in Finland]. Helsinki: Lakimiesliiton Kustannus and Kansallis-Osake-Pankki, 1990–1997.

Lo, A. W. – MacKinlay, A. C.: Data-snooping biases in tests of financial asset pricing models. Review of Financial Studies, 1990, 3, 3–28.

MSCI: Methodology & Index Policy. USA: Morgan Stanley Capital International, 1998.

Mucklow, B.: Market Microstructure: An Examination of the Effects on Intraday Event Studies. Contemporary Accounting Research, 1994, 10, 2, 355–382.

Nenova, Tatiana: The value of corporate voting rights and control: A cross-country analysis. Journal of Financial Economics, 2003, 68, 3, 325–351.

OKO: Julkisesti Noteeratut Arvopaperit [Publicly Listed Companies in Finland]. Helsinki: Osuuspankkien Keskusosakepankki and Opstock, 1996–1998.

Roll, R.: A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market. The Journal of Finance, 1984, 39, 4, 1127-1139.

Roll, R.: A Critique of the Asset Pricing Theory's Tests: Part I. Journal of Financial Economics, 1977, 4, 129-176.

Shanken, J.: Statistical Methods in Tests of Portfolio Efficiency: A Synthesis. In Maddala, G. S. – Rao, C. R. (eds.): Handbook of Statistics, Vol. 14. USA: Elsevier Science, 1996.

Shumway, Tyler – Warther, Vincent A.: The Delisting Bias in CRSP's Nasdaq Data and Its Implications for the Size Effect. Journal of Finance, 1999, 54, 6, 2361-2379.

Stambaugh, Robert: On the exclusion of assets from tests of the two-parameter model: A sensitivity analysis. Journal of Financial Economics, 1982, 10, 237-268.

Talouselämä: Talouselämän raportti suuryrityksistä [Report on the largest companies in Finland]. Helsinki: Talentum, 1987-1997.

Vaihekoski, Mika: Portfolio Construction in Emerging Markets. Emerging Markets Quarterly, 2000, Fall.

Table 1. Descriptive Statistics of the Portfolio Contents

Monthly resembled value-weighted size portfolios and yearly resembled value-weighted industry portfolios are constructed. Minimum, maximum, and time-series average of the percentage of the total equity market capitalization value represented by portfolios are given in the table. Minimum, maximum, and time-series average number of companies in portfolios are also given in the table. Sample size is 168 monthly observations from January 1987 to December 2000.

PORTFOLIO	Relative Value			Number of Assets		
	Min	Mean[a]	Max	Min	Mean[b]	Max
Panel A: Size portfolios						
1 – Largest	47.96	63.80	92.57	8	11.58	17
2	3.92	19.01	27.71	8	11.72	17
3	1.78	9.84	15.35	8	11.95	18
4	1.05	4.46	7.77	8	12.15	18
5	0.42	2.17	4.15	9	12.29	18
6– Smallest	0.11	0.71	1.41	9	12.40	18
Panel B: Industry portfolios						
Banking & Finance	0.74	11.34	22.72	5	9.32	16
Forestry	4.27	14.77	25.39	4	5.85	8
Trade & Transport	0.99	6.06	11.15	6	9.07	11
Metal & Electronics	5.97	29.11	87.21	6	13.27	34
Food	0.63	5.57	11.19	2	4.76	7
Housing & Construction	0.44	3.34	9.18	1	8.32	13
Multi-Business	0.55	17.99	49.55	6	8.13	10

[a] Time series average of the ratio of the portfolio’s market value to the total market capitalization value at the end of every month.

[b] Time series average of the number of assets in a portfolio at the end of every month.

Table 2. Descriptive Statistics for Portfolio Returns

Descriptive statistics of the portfolio returns are given in the table. The first four sample central moments are given for the market index, and size and industry portfolio returns. The minimum and maximum returns are also provided together with the date when they took place. The normality of the return series is tested using Bera-Jarque test (p-value provided in the table). Mean and standard deviation of the returns are annualized. Sample size is 168 monthly observations from January 1987 to December 2000.

PORTFOLIO	Return Mean	Min	Month[a]	Max	Month[a]	Std. Dev.	Skew-ness	Excess Kurtosis	Bera-Jarque	Autocorrelation[b] r_1	r_2	r_3
Panel A: Sample statistics												
Market portfolio												
HEX index	19.9 %	-21.3 %	Aug. 98	25.8 %	Dec. 99	0.263	0.028	0.584	0.300	0.246*	0.001	0.054
Size portfolios												
1 – Largest	17.7 %	-21.1 %	Aug. 97	25.0 %	Dec. 99	0.280	-0.040	0.245	0.793	0.207*	0.056	0.063
2	5.5 %	-21.9 %	Sep. 97	25.1 %	Oct. 92	0.247	-0.125	1.000	0.024	0.277*	-0.021	0.095
3	4.6 %	-28.2 %	Aug. 92	22.7 %	Nov. 92	0.259	-0.256	1.570	<0.001	0.291*	-0.094	0.005
4	7.9 %	-22.4 %	Aug. 92	24.6 %	Jan. 94	0.244	0.350	1.961	<0.001	0.311*	0.041	0.139
5	3.3 %	-21.0 %	Aug. 92	18.7 %	Dec. 99	0.221	-0.099	1.102	0.012	0.270*	-0.007	0.063
6 – Smallest	3.6 %	-24.4 %	Aug. 92	29.8 %	Oct. 93	0.239	0.670	2.949	<0.001	0.174*	0.046	0.096
Industry portfolios												
Banking & Finance	-1.4 %	-26.9 %	Jun. 92	36.2 %	Oct. 92	0.318	0.343	2.095	<0.001	0.165*	0.039	0.162
Forestry	8.6 %	-20.7 %	Sep. 90	22.3 %	Jan. 90	0.241	-0.084	0.702	0.161	0.126	0.075	-0.155
Trade & Transport	1.9 %	-23.0 %	Aug. 98	18.8 %	Jan. 94	0.225	-0.176	0.673	0.133	0.233*	-0.001	0.201*
Metal & Electronics	23.4 %	-21.5 %	Sep. 90	28.1 %	Dec. 99	0.304	0.339	0.165	0.182	0.174*	-0.115	0.131
Food	5.4 %	-27.7 %	Aug. 98	26.6 %	Mar. 99	0.255	-0.035	2.191	<0.001	0.105	0.104	0.028
Housing & Construction	-4.9 %	-38.1 %	Aug. 92	28.4 %	Jan. 94	0.300	-0.490	2.842	<0.001	0.218*	0.002	0.129
Multi-Business	12.8 %	-24.7 %	Dec. 96	26.8 %	Oct. 92	0.300	-0.264	1.085	0.006	0.268*	0.048	0.096

[a] Months when minimum or maximum returns took place.

[b] Significant autocorrelation coefficients are marked with an asterisk (*).

Table 2 Continued

Panel B: Cross-correlation coefficients

	HEX-Index	1	2	3	4	5	6	Banking	Forestry	Trade	Metal	Food	Housing	Multi-Business
HEX Market Index	1,000													
1 – Largest	0.969	1.000												
2	0.803	0.705	1.000											
3	0.713	0.623	0.778	1.000										
4	0.649	0.532	0.710	0.750	1.000									
5	0.671	0.572	0.714	0.799	0.782	1.000								
6 – Smallest	0.584	0.487	0.671	0.635	0.685	0.709	1.000							
Banking & Finance	0.671	0.625	0.661	0.592	0.559	0.487	0.527	1.000						
Forestry	0.691	0.691	0.701	0.593	0.387	0.499	0.439	0.470	1.000					
Trade & Transport	0.666	0.576	0.754	0.800	0.684	0.740	0.585	0.551	0.605	1.000				
Metal & Electronics	0.895	0.879	0.684	0.627	0.569	0.587	0.504	0.526	0.584	0.557	1.000			
Food	0.399	0.342	0.589	0.467	0.413	0.419	0.435	0.346	0.381	0.469	0.257	1.000		
Housing & Construction	0.578	0.495	0.669	0.723	0.665	0.718	0.663	0.534	0.479	0.632	0.489	0.391	1.000	
Multi-Business	0.761	0.748	0.790	0.692	0.539	0.637	0.525	0.534	0.644	0.599	0.589	0.428	0.529	1,000

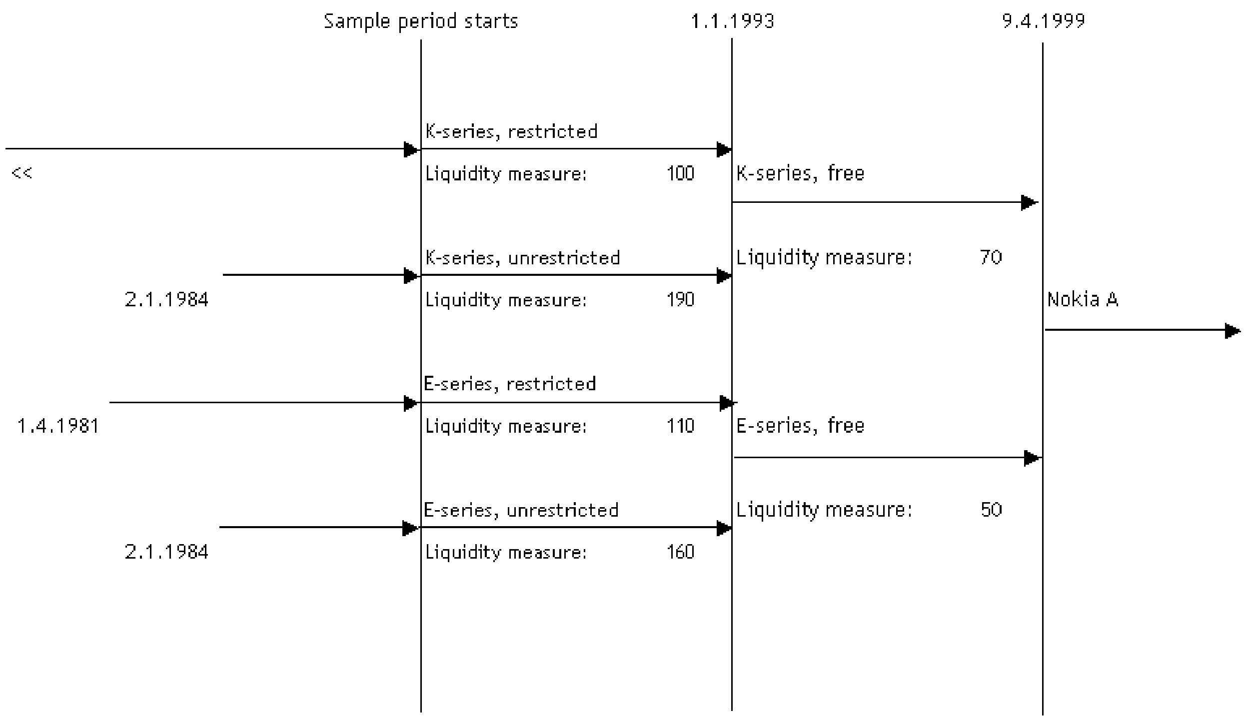

Figure 1. History of Nokia's listed stock series. Nokia has had up to four different stock series listed at the same time in Helsinki Stock Exchange. Liquidity measures in the table are not real. They are just used to show how the most liquid stock can be selected.

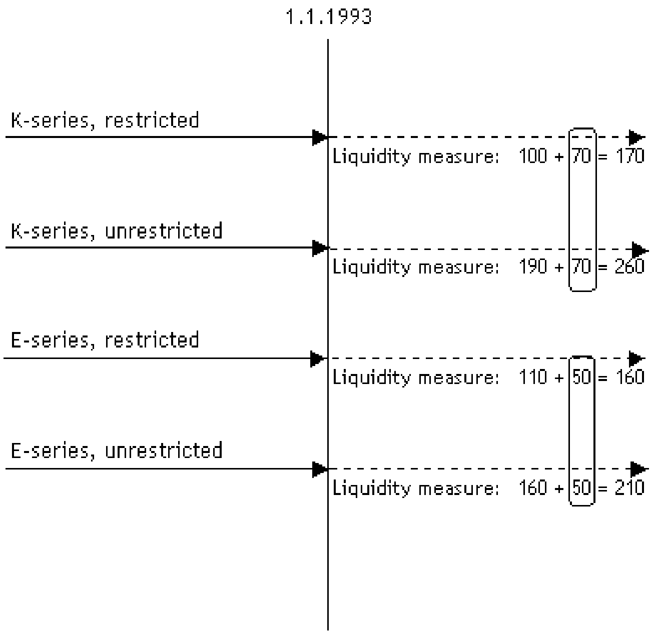

Figure 2. An example of selecting the representative stock series after unrestricted and restricted series are combined. In Finland, all restrictions on foreign ownership were abolished starting 1993 and, thus, all shares became free (unrestricted). New "free" series can be seen as continuum of either restricted or unrestricted series, since owners of either series could continue trading with their shares without any interruptions. Therefore, if we use liquidity of the series to select between K- and E-series, we should add the liquidity measure of the free series 1.1.1993- to both restricted and unrestricted series' liquidity measures and make the choice between K and E based on their whole history.

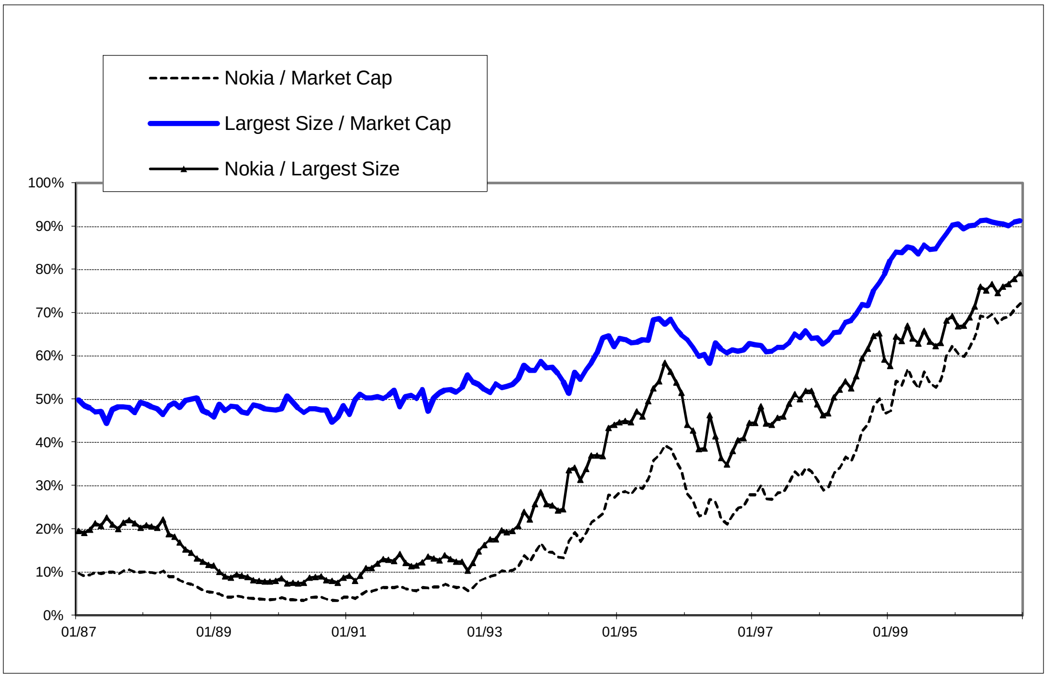

Figure 3. Relative market capitalization values. Nokia's market capitalization value divided by the total market capitalization value, largest size portfolio vs. the total market value, and Nokia vs. the largest size portfolio's market values. Monthly ratios from 1987 to 2000.

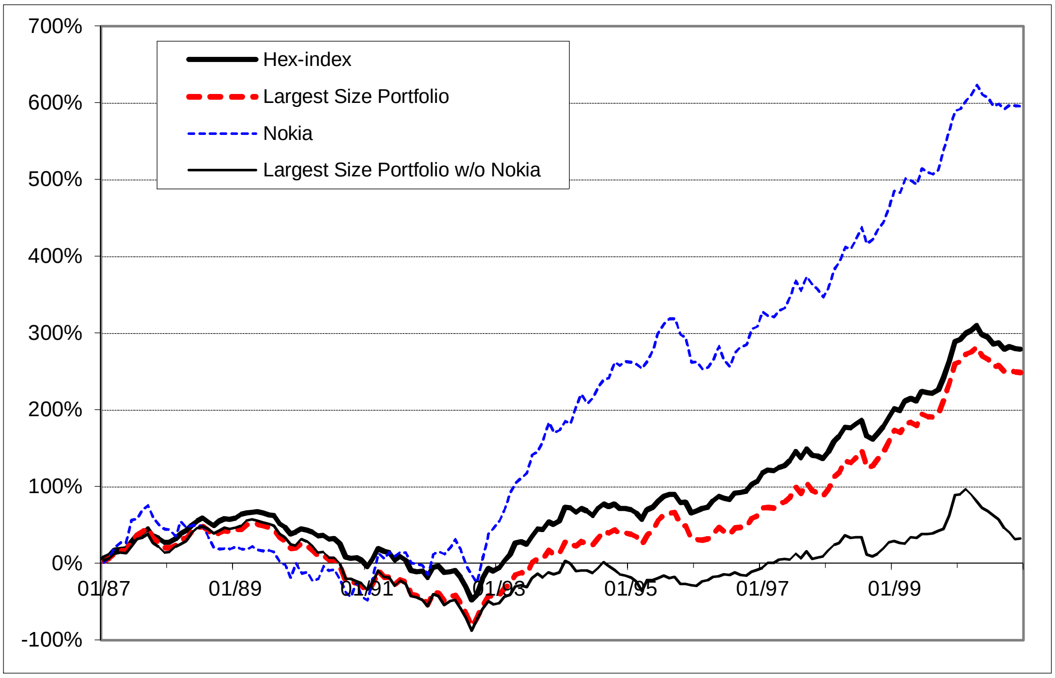

Figure 4. Cumulative returns. Cumulative monthly returns for the market return index (HEX), largest size portfolio, Nokia, and largest size portfolio without Nokia from 1987 to 2000.

Appendix A. Representative Stock Series Selection for the companies in the Helsinki Stock Exchange.

The number of days without trading during a year is given in the table for companies with multiple stock series quoted on the Helsinki Stock Exchange. Stock series marked with an asterisk (*), have been listed longer than other series classes for that company during the sample period and therefore being chosen without further comparison. The stock series symbol is given after the company name if several series available. A, K and I are used symbols to mark the common (ordinary) stock series, B, C, E, R, and II are names for preferred stock series. Note-column specifies whether the series is restricted and unrestricted. In Panel A we have to choose only between restricted and unrestricted series. In Panel B we have to choose between two different classes and restricted/unrestricted series. A minus sign (-) in the year columns marks that the stock is not listed during that year. n/a marks that this information is not applicable or relevant during that year. After 1993 all shares are unrestricted. Sum is the sum of days when no transactions took place during the sample period. Sample period is from 1987 to 2000.

Company	Note	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	Sum	Selected ?
PANEL A																	
Amer A	restr.	0	1	2	16	5	4	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	28	yes
	unrestr.	1	4	11	58	57	55	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	186	
Kone	restr.	3	6	3	3	8	8	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	31	yes
	unrestr.	20	36	49	81	111	66	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	363	
PANEL B																	
Nokia E	restr.	0	1	1	14	5	2	0	0	0	0	0	1			24	yes
	unrestr.	4	1	7	55	56	22									146	
Nokia K	restr.	1	0	0	12	9	6	0	0	0	0	0	1	-	-	29	
	unrestr.	7	10	20	59	48	19									164	

Etc. (full list available from author's web-page)

Appendix B. Industry Portfolio Classification for the Helsinki Stock Exchange.

Company industry classification is given for companies listed in the Helsinki Stock Exchange. Classification is based on the information available at the end of the year. If a company has not been listed during a year, zero is given in the table. Other numbers are as follows: 1 = Banking Industry, 2 = Forestry, 3 = Trade, 4 = Metal Industry, 5 = Technology Industry, 6 = Food Industry, 7 = Other Financial (not Insurance), 8 = Housing & Construction Industry, 9 = Transport & Forwarding, 10 = Clothing Industry, 11 = Multi-Business, 12 = Chemistry & Plastics, 13 = Printing & Publication Industry, and 99 = other industries.

		Industry code													
#	Company	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
1.	AAB	1	1	1	1	1	1	1	1	1	0	0	0	0	0
2.	Aamulehti	0	0	0	0	0	0	0	13	13	13	13	13	0	0
3.	Aga	12	12	12	12	12	0	0	0	0	0	0	0	0	0
4.	Amer	11	11	11	11	11	11	11	11	11	11	11	11	11	11
5.	Arctos	0	0	0	0	0	0	0	0	7	7	7	7	7	0

Etc. (full listing available later on the Internet)

APPENDIX C. Code Listing

```
% ********************************************************************************
% 1 calculate company market values by adding the values of all firm's stock series
    together
% ********************************************************************************

load rep_indicator;              % 1 x N vector, value (1,j) points company
                                   reprentative column; if (1,j) = j, then is
                                   reprentative
load marketvalue;                % (T+1) x N matrix; +1 because 31.12.1986

[r,s] = size(marketvalue);

for i=1:s,
  if rep_indicator(1,i) > 0,
     marketvalue(:,rep_indicator (1,i))
       = marketvalue(:,rep_indicator (1,i))+ marketvalue(:,i);
  endif
end

company_value = marketvalue;

save company_value;

disp('1 - daily company market values - stage completed!');

% ********************************************************************************
% 2 calculate market cap weights (remove last observation, if de-listed next month)
% ********************************************************************************

load company_value;              % (T+1) x N matrix

lagged_company_value = company_value(1:(rows(company_value)-1),:);

company_value = company_value(2:(rows(company_value)),:);

                                 % if company de-listed next day, exclude

lagged_company_value = lagged_company_value .* (company_value > 0);

weights = lagged_company_value;

save weights;

disp('2 - daily lagged company market values - stage completed!');

% ********************************************************************************
% 3 form nofsp value-weighted size portfolios with monthly resembling,
    weight and content revisions
% ********************************************************************************

load weights;                    % T x N matrix
load returns;                    % T x N matrix

[r,s] = size(weights);
nofsp = 7;                       % uset set; number of portfolios to be constructed
mv_s  = zeros(r,s);              % size portfolio indicator matrix;
                                   values are 0 or 1-nofsp

for i=1:r,                       % loop for each period
 limit = 0;
 firms = sum(weights(i,:)>0);    % = number of listed firms at period i
 s1    = fix(firms/nofsp);       % number of companies/portfolio
 s2    = firms-s1*nofsp;         % s2 portfolios have 1 extra company
 for j = 1:nofsp,                % loop for NOFSP portfolios
```

```
   if s2>0
        one = 1;
        s2 = s2-1;
   else
        one = 0;
   end
   for n = 1:(s1+one),           % loop to find S1 firms (+ 1 if needed)

                                 % Find the smallest market value above limit

        minimum = min(weights(i,(weights(i,:)>limit)));

                                 % Check whether more than one company with same
                                   market value

        sara = find(weights(i,:)==minimum);
        s3   = size(sara,2);
                                 % if this is the case (very rare) induce little
                                   difference into market values
        while s3>1
            for k=1:cols(s3),
             weights(i,sara(1,k))=weights(i,sara(1,k))+k;
            end
                                 % Recalculate # of minimum market value companies

            sara = find(weights(i,:)==minimum);
            s3   = size(sara,2);
        end

                                 % The smallest companies get value of nofsp etc.

        mv_s(i,sara(1,1)) = nofsp+1-j;
        limit = minimum;
   end
 end
end

for i=1:nofsp
    monthly_size_portfolio_returns(:,i)
       = sum((returns .* weights .* (mv_s == i))')'
         ./ sum((weights .* (mv_s == i))')';
end

save monthly_size_portfolio_returns;

disp('3 - size portfolio returns calculated - stage completed!');

% *********************************************************************************
% 4 form value-weighted industry portfolios with monthly weight and annual content
    revisions
% *********************************************************************************

load weights;                    % T x N matrix
load returns;                    % T x N matrix
load industry_indicator;         % Y x C matrix; Y = # of years; C = # of companies
load company;                    % 1 x C vector; indicates

[r1,s1] = size(weights);
mv_i  = zeros(r1,s1);            % industry portfolio indicator matrix;
                                   values are 0 or industry class
[r2,s2] = size(industry);
d_indu  = zeros(r1,s1);          % init daily industry matrix
y_indu  = zeros(r2,s1);          % init yearly industry matrix

for i=1:14,                      % loop for each year; 14 years
  for j=1:s2,                    % “spread” company matrix into asset matrix
    mv_i(((i-1)+1):((i-1)+12),company(1,j))=industry(:,j);
  end
```

```
end

monthly_industry_portfolio_returns(1:r1,1)
      = sum((returns .* weights .* (mv_i==1|mv_i==7))')'
        ./ sum((weights .* (mv_i==1|mv_i==7))')';
      % Banking & other finance

monthly_industry_portfolio_returns(1:r1,2)
      = sum((returns .* weights .* (mv_i==2))')'
        ./ sum((weights .* (mv_i==2))')';
      % Forestry

monthly_industry_portfolio_returns(1:r1,3)
      = sum((returns .* weights .* (mv_i==3|mv_i==9))')'
        ./ sum((weights .* (mv_i==3|mv_i==9))')';
      % Trade & transport

monthly_industry_portfolio_returns(1:r1,4)
      = sum((returns .* weights .* (mv_i==4|mv_i==5))')'
        ./ sum((weights .* (mv_i==4|mv_i==5))')';
      % Metal & electronics

monthly_industry_portfolio_returns(1:r1,5)
      = sum((returns .* weights .* (mv_i==6))')'
        ./ sum((weights .* (mv_i==6))')';
      % Food

monthly_industry_portfolio_returns(1:r1,6)
      = sum((returns .* weights .* (mv_i==8))')'
        ./ sum((weights .* (mv_i==8))')';
      % Housing & Construction

monthly_industry_portfolio_returns(1:r1,7)
      = sum((returns .* weights .* (mv_i==11))')'
        ./ sum((weights .* (mv_i==11))')';
      % Multi-business

save monthly_portfolio_return;   % save industry portfolio returns matrix

disp('4 - industry portfolio returns calculated - stage completed!');
```