Journal of Applied Finance & Banking, vol.1, no.1, 2011, 143-162
ISSN: 1792-6580 (print version), 1792-6599 (online)
International Scientific Press, 2011

The stock selection problem: Is the stock selection approach more important than the optimization method? Evidence from the Danish stock market

Klaus Grobys[1]

Abstract

Passive investment strategies basically aim to replicate an underlying benchmark. Thereby, the management usually selects a subset of stocks being employed in the optimization procedure. Apart from the optimization procedure, the stock selection approach determines the stock portfolios' out-of-sample performance. The empirical study here takes into account the Danish stock market from 2000-2010 and gives evidence that stock portfolios including small companies' stocks being estimated via cointegration optimization methods are most beneficial. Only the stock portfolios exhibiting the lowest initial market capitalization corresponding to 29.51% showed a Sharpe ratio of 0.4545 and 0.4824, respectively, being higher than the stock market's Sharpe ratio of 0.4451 concerning the out-of-sample period running from 2003-2010.

JEL classification numbers: C13, C22, G11, G12

Keywords: Stock selection, Optimization procedure, Cointegration, Quasi-Maximum-Likelihood Estimation, Index-Tracking.

---

[1] E-mail: Klaus.grobys@klarna.com

Article Info: *Revised*: April 6, 2011. *Published online* : May 31, 2011

1 Introduction

Active trading strategies are basically targeted on beating the underlying benchmark. Van Montfort, Visser and Finjn van Draat (2008) figure out several drawbacks being associated with active portfolio management. Apart from high transaction costs and high risks, active investment strategies require a remarkable knowledge about the companies being listed at the stock exchange as well as correlative risk factors that the underlying stock market involves. Grobys (2010a) shows though that even passive investment strategies involving only historical information may exhibit better Sharpe-Ratios than the underlying stock market.

Passive investment strategies basically aim to track an underlying benchmark. Thereby, the management usually selects a subset of stocks being employed in the optimization procedure. Selecting n stocks of an index including N stocks would involve running $N!/\left(N-n\right)!n!$ models for backtesting. As N increases, the mechanical selection procedure becomes more and more time consuming. Alexander and Dimitriu (2005) estimate tracking portfolios by selecting 20, 25 and 30 stocks of companies being listed at the S&P 500 in accordance to their price ranking, starting with the highest-priced stocks. This approach rests upon the market capitalization method assuming a company’s market value to be the present value of its expected future earnings.

The stratified index portfolio as applied by Larsen, Bruce, and Resnick (1998), Focardi and Fabozzi (2004) divides the stocks of an index into a number of different categories such as market sectors. The tracking portfolio then accounts for each sector such that each sector is represented in the index portfolio to the same extent like in the actual stock index. Then, one stock exhibiting the highest market capitalization, for instance, can be chosen which will be assigned the weights of all the stocks from this category. Van Montfort, Visser and Finjn van Draat (2008) argue that this approach implicitly involves the assumption that price movements of stocks within the same category are highly correlated, while price movements of stocks of different sectors may deviate from each other.

Fama and French (1996) construct a model resting upon the empirical evidence that firm characteristics such as the relationship between large and small firms and the relationship between firms exhibiting high, respectively, low book-to-market ratios seem to be predictive of average stock returns. In order to figure out the relationship between large and small companies, Fama and French (1996) construct two portfolios where the first portfolio contains a subset of stocks of an index exhibiting the highest market capitalization, while the second portfolio contains a subset of stocks of an index exhibiting the lowest market capitalization. Going long on second and short on the first one results in a risk factor that seems to be significant in the studies of Davis, Fama and French (2000) who analyze data from 1929-1997.

Grobys (2010a) employs stocks being preselected due to data set limitations and considers the stock portfolios' performance when applying different optimization methods. He figures out that portfolios being based on cointegration analysis clearly dominate their traditional counterparts which apply correlation analysis of asset returns based on the seminal work of Markowitz (1952). Even if the stock portfolio is not rebalanced, being often referred to as buy-and-hold strategy, the cointegration based portfolio as shown by Grobys (2010a) beats the index by 79.08% within the overall 10-years out-of-sample period (i.e. from 2000-2010), whereas the annual volatility on average was 1.10 base points lower. But is the optimization procedure the more determining factor as the stock selection approach?

Alexander and Dimitriu (2005) mention that the stock selection criterion has a significant impact on the positive (or negative) alpha being generated by the tracking portfolio. In the following, the Danish leading stock market index OMX 20 is considered. Thereby, five different stock selection criteria are considered that are applied to statistical models being based on cointegration as well as correlation methods. The empirical analysis reveals on the one hand that given the stock selection criteria, cointegration based models exhibit better Sharpe-Ratios compared to correlation based models, basically. On the other hand, the stock selection is more important and under the time of consideration (i.e. from 2003 to 2010), the portfolio containing the stocks involving the smallest market capitalization outperform all other portfolios, irrespective of the optimization procedure being employed.

2 Background

Roll (1992) suggests an optimization model that employs the asset returns and minimizes the tracking error. Then, the parameter estimates correspond to the optimal asset allocation. The weights are held constant until the stock portfolio is rebalanced. The optimization procedure can be considered as passive strategy because only historical data is necessary to estimate the optimal weights. Alexander (1999) argues though that correlation is basically a short run measure and the tracking error of stock portfolios which rest upon correlation analysis can exhibit out of sample random walk behavior.

Alexander and Dimitriu (2005) compare index tracking portfolios being based on cointegration analysis with stock portfolios being based on correlation. They consider different subsets of stocks of the S&P 500 and use three years of daily data in order to run the optimization procedure. They argue that three years of daily data are necessary for ensuring cointegration. Furthermore, they analyze the out-of-sample period from 1993-2003, while constructing portfolios differing in the number of stocks being included as well as the rebalancing moments (i.e. 2-week rebalancing, monthly rebalancing, 3-months rebalancing and 6-months rebalancing). They figure out that the correlation based optimization procedure generates marginally lower transaction costs, while having slightly better Sharpe ratios, too. They conclude that no significant advantages or limitations of a cointegration relationship with the benchmark are empirically evident as long as no weight constraints are taken into account.

However, Grobys (2010a) considers the Swedish stock market index OMX 30 and compares index tracking portfolios being based on correlation and cointegration methods. Thereby, 17 stocks are employed being preselected out of the index containing 30 stocks. The stocks corresponding to a market capitalization of 56.67% are selected due to data set limitations since there is only data of those 17 stocks available that range to 31.12.1996. The out-of-sample period runs from 31.12.1999 to 31.12.2009. Even though the cointegration optimal portfolios do not exhibit a strong cointegration relationship with the benchmark out-of-sample, they dominate their traditional counterparts as they exhibit higher Sharpe ratios as the latter. Cointegration, as defined and developed by Granger (1981) and Engle and Granger (1987), is a feature of some nonstationary time series. If two or more nonstationary time series are cointegrated, a linear combination relationship exhibiting stationarity is said to exist. In the asset allocation framework, whether the value series of a fixed weight portfolio of assets with nonstationary prices is stationary, the assets will involve a cointegrated set. The set of asset weights producing such a portfolio is referred to as the cointegrating vector. In contrast to correlation, cointegration ensures a so called mean reversion of the tracking error. Even though correlation and cointegration based index tracking models do according to Alexander and Dimitriu (2005) not significantly differ from each other with respect to their out-of-sample performance (i.e. at least as long as no restrictions are taken into account), the correlation based models' tracking error is more volatile out-of-sample as the corresponding tracking error of cointegration based models.

In contrast to Alexander and Dimitriu (2005), Phengis and Swanson (2011) apply cointegration analysis in order to select the assets which are employed to construct international portfolios. Their implication is that cointegrated assets involve significant long-run comovements, whereby the diversification gains are lowered. In their studies 21 different stock markets are taken into account, whereas the US-stock represents to domestic market. They compare four different portfolios whereby portfolio 1 is based on maximizing the Sharpe Ratio, portfolio 2 accounts for shrinkage estimators while maximizing the Sharpe ratio, portfolio 3 is an equal weighted portfolio taking into account all markets to the same extent, whereas portfolio 4 selects only those markets which are either weakly exogenous within the cointegration relationships or not a part of the cointegration relationships. Their findings give evidence that the cointegration based asset selection process clearly outperforms all other portfolios even if the assets being involved are only equally weighted.

In the following empirical analysis, both optimization methods are applied to different stock selection approaches, optimizing based on correlation and cointegration as introduced by Alexander (1999). In order to make the models comparable, the quantity of stocks is held constant within the out-of-sample period and the portfolio weights are held constant over time being often referred to as "buy-and-hold strategy".

3 Econometric Methodology

In line with Roll (1992) the optimization problem is given by minimizing the tracking error variance of the following model:

$$r_{OMX,t} = a_1 r_{1,t} + \dots a_N r_{N,t} + \varepsilon_t \tag{1}$$

where $r_{OMX,t}$ denotes the log index returns and $r_{i,t}$ denotes the log returns of stocks $i = 1,\dots,N$. The optimization method being employed here is in line with Grobys (2010a) Quasi-Maximum Likelihood-Estimation (QMLE). Therefore, the mean-variance optimal portfolio may be estimated by maximizing the log-likelihood being given by

$$\log L(\theta,t) = -\frac{T}{2}\log(2\cdot\pi) - \frac{T}{2}\log\sigma^2 - \frac{1}{2}\sum_{t\in T}\left(\frac{\varepsilon_t^2}{\sigma^2}\right), \tag{2}$$

where $\varepsilon_t = r_{OMX,t} - \sum_{i=1}^{N} a_i r_{i,t}$. Furthermore, the models account for two restrictions. On the one hand, the weights should sum up to one, as given by Equation (3). On the other hand, the weights should be positive which may be given by Equation (4) being also in line with Grobys (2010a) as well as van Montefort, Visser and and Fijn van Draat (2008):

$$\sum_{i=1}^{N} a_i = 1 \tag{3}$$

$$a_i > 0 \quad \text{for} \quad i = 1, ..., N\,. \tag{4}$$

Constructing cointegration optimal portfolios though involves in line with Alexander and Dimitriu (2005) first of all running the optimization procedure (see Equation 2) by employing the logarithm of the stock prices $p_{i,t}$ instead of the log-returns $r_{i,t}$ (see Equation 1). The second additional step is in line with Alexander and Dimitriu (2005) testing for cointegration. In contrast to Alexander and Dimitriu (2005) who suggest employing the augmented Dickey fuller test (ADF) in the following, the trace test is employed. In line with Johansen (1988, 1991), Johansen and Juselius (1990, 1991) the multivariate trace test for cointegration tests whether the stock index and the estimated portfolios have a cointegration relationship. The test is performed with respect to the in-sample as well the out-of-sample period in order to evaluate if the cointegration relationship is stable even out-of-sample. The Johansen procedure employs the maximum likelihood estimates of a fully specified error correction model which is given by

$$\Delta Y_t = \mu + \sum_{i=1}^{k} \Gamma_i \Delta Y_{t-i} + \Pi Y_{t-1} + \varepsilon_t$$

where ΔY_t exhibits the vector of the stock market's and portfolios price changes in logarithms at time t, μ is a constant vector, Γ represent the short-run impact, and Π denotes the long-run impact matrix having reduced rank under cointegration. If the rank of Π is equal to one, the stock index and the portfolio under consideration will be cointegrated. To determine the rank r of the estimated long-run matrix $\hat{\Pi}$, the eigenvalues $\tilde{\lambda}_i$ have to be calculated. Thereby, the number of significantly nonzero eigenvalues shows the rank of $\hat{\Pi}$, and can be evaluated by the trace test. Then, the trace test statistic is the result of testing the restriction $r \le q\left(q < n\right)$ against the completely unrestricted model $r \le n$:

$$\lambda_{trace} = -T\sum\nolimits_{i=q+1}^{n} \ln\left(1-\tilde{\lambda}_i\right)$$

where T is the sample size and $\tilde{\lambda}_{r+i},...,\tilde{\lambda}_n$ are the $n-r$ smallest squared canonical correlations. The trace test statistic will be calculated while accounting for both, a constant term only, as well as a constant and trend term.

The Stock selection criteria selecting N out of an index are as follows: The first approach takes N stocks into account that join at the initial allocation day the highest market capitalizations. The second approach though employs N stocks of the smallest companies listed at the stock exchange. The first approach takes into account stocks of the largest company of each business sector that the underlying index may account for. In order to ensure comparability, it will be assumed that the number of different sectors equals the number of stocks (i.e. N) being employed in the optimization procedure. In contrast to the third approach, the fourth one employs stocks of the smallest company concerning each business sector. The fifth approach selects stocks randomly with probability 0.5 for both large and small stocks. Table 1 provides an overview of the stocks, the corresponding sector as well as the initial market capitalization.[2]

4 Limitations of the data set

Stock market data can be downloaded for free on the index provider’s homepage nasdaqomx.com. However, the data availability is limited. For 19 of 20 stocks, data is available from 17.11.2000 onwards. Consequently, the company *Tryg* (security identification number A0HF12) is excluded from stock selection

[2] The stocks which every portfolio involves can be provided by the author upon request.

processes. The in-sample period being used for estimate the models runs from 17.11.2000-14.11.2003 including 745 daily observations. The out-of-sample period runs from 17.11.2003-14.07.2010 accounting for 1665 daily observations. Table 1 show that the stocks of the OMX 20 can be divided into 12 different sectors.

5 Results

In order to account for the number of stocks employed in the optimization procedure to be equal to the number of business sectors, every stock portfolio contains 12 stocks. The range of the corresponding market capitalization runs from 29.51%-82.11%. The weights being estimated for each portfolio are holding constant for the out-of-sample period (i.e. from 17.11.2003 to 14.07.2010) which is often referred to as buy-and-hold strategy. Exhibits 1 and 2 show that the cointegration based portfolios have slightly better Sharpe ratios compared to the correlation based models, except for stock selection approach 2 where the stocks of the 12 smallest companies are employed. Only the stock selection approach 2 exhibits higher Sharpe ratios than the underlying stock market exhibiting a Sharpe ratio of 0.4451 and a Treynor ratio of 10.01 within the out-of-sample period. Apart from that only the latter stock selection approach exhibits cointegration relationships with the benchmark with respect to both periods in-sample and out-of-sample, irrespective of the employed optimization procedure: The p-value of the trace test statistics including a linear deterministic trend is in-sample as well as out of-sample clearly below 5% (i.e. p-value=0.03 and p-value*=0.01 concerning the cointegration based model and p-value=0.05 and p-value*=0.01 regarding the correlation based model). The correlation between in-sample

Table 1a: The initial weight allocation of the OMX 20

Company	Weight	Orderbook	Sector	Sector number
Bang & Olufsen	0.98%	BO B	Consumer Electronics	1
Carlsberg B	2.88%	CARL B	Brewers	2
Coloplast B	2.74%	COLO B	Health Care Supplies	3
Danske Bank	20.70%	DANSKE	Diversified Banks	4
Danisco	2.88%	DCO	Agricultural Products	5
DSV	2.48%	DSV	Trucking	6
FLSmidth & Co.	3.28%	FLS	Construction & Engineering	7
GN Store Nord	2.09%	GN	Health Care Equipment	3
Jyske Bank	2.95%	JYSK	Diversified Banks	4
Lundbeck	3.88%	LUN	Pharmaceuticals	8
A.P. Møller - Mærsk B	13.93%	MAERSK B	Industry/Marine	9
Nordea Bank	3.43%	NDA DKK	Diversified Banks	4

Note: Table 1a,b show the index’s weight allocation on the 18.Dec 2006.

Table 1b: The initial weight allocation of the OMX 20

Company	Weight	Orderbook	Sector	Sector number
Nordea Bank	3.43%	NDA DKK	Diversified Banks	4
Novo Nordisk B	16.29%	NOVO B	Pharmaceuticals	8
Novozymes B	3.14%	NZYM B	Specialty Chemicals	10
Sydbank	2.16%	SYDB	Diversified Banks	4
Topdanmark	2.31%	TOP	Multi-line Insurance	4
TORM	1.62%	TORM	Oil & Gas & Transportation	11
Tryg	3.51%	TRYG	-	-
Vestas Wind Systems	5.19%	VWS	Heavy Electrical Equipment	12
William Demant Holding	3.56%	WDH	Health Care Equipment	3

Note: Table 1a,b show the index's weight allocation on the 18.Dec 2006.

tracking-error and out-of-sample tracking error is 0.21 concerning the cointegration based models (i.e. portfolios 1-5) and 0.82 with respect to correlation based models (i.e. portfolios 6-10). However, the correlation between market capitalization and Sharpe ratio is less ambiguous for both models, as it is -0.69 for cointegration based models and -0.73, suggesting that portfolios that include small stocks perform on average better compared to portfolios accounting for stocks of large companies. Running the trace test while accounting for portfolios 1 and 2 (with the time series in logarithms) shows that they have a cointegration relationship (p-value =0.0486) even though portfolio 1 is not cointegrated with the benchmark (i.e. p-value=0.5791 and p-value*=0.2321 concerning the out-of-sample period). A cointegration relationship offers in accordance to Grobys (2010b) arbitrage opportunities, irrespective of the individual portfolio's beta. Exhibit 3 shows the graphs of both portfolios with respect to the out-of-sample period. Consequently, going short on portfolio 1 and long on portfolio 2 would have generated annual abnormal returns of 4.30% on average with respect to the out-of-sample period. Moreover, all tracking portfolios except for portfolios 2 and 6 exhibited a lower tracking error out-of-sample compared to the in-sample tracking error. A further empirical outcome of this study here is that taking into account different sectors does not involve any benefits, irrespective of the optimization procedure being applied (see portfolios 3 and 7 on exhibits 2 and 3 for instance). The same is true concerning portfolios of randomly selected stocks (i.e. portfolios 5 and 10).

Furthermore, the constructed portfolios are compared with a well-diversified European stock portfolio. The EuroStoxx 50 is a stock index accounting for 50 European companies exhibiting the highest market capitalization. During the out-of-sample period, running from November 17, 2003 until July 14, 2010 the EuroStoxx 50 exhibited a return of 1.78% p.a. and a volatility of 22.93% p.a. which results in a Sharpe ratio of 0.0776. Interestingly, all constructed portfolios, irrespective of the optimization procedure or stock selection procedure outperformed the EuroStoxx 50.

Table 2: Statistical properties of cointegration based stock portfolios

Optimization Method	Cointegration-Analyis				
Portfolio	1	2	3	4	5
Stock selection	1	2	3	4	5
Market capitalization at initial allocation*	82.11%	29.51%	76.93%	45.31%	60.05%
Annual return (out-of-sample)	8.92%	13.22%	9.02%	8.47%	9.38%
Annual Volatility (out-of-sample)	33.14%	29.09%	32.33%	33.62%	29.61%
Treynor ratio (out-of-sample)	8.3364	15.3721	8.5094	7.7706	9.2871
Rank Treynor ratio (out-of-sample)	5	2	4	8	3
Sharpe ratio (out-of-sample)	0.2692	0.4545	0.2790	0.2520	0.3169
Rank Sharpe ratio (out-of-sample)	5	2	4	6	3
In-sample Tracking-Error (TE)	28.03%	11.64%	27.84%	28.80%	23.38%
Out-of-sample Tracking-Erorr (TE)	22.86%	21.79%	21.74%	23.04%	19.03%
p-value Trace-test in sample	0.8661 (0.6520)*	0.6506 (0.0314)*	0.8463 (0.6215)*	0.8694 (0.6305)*	0.8279 (0.5943)*
p-value Trace-test out of sample	0.5204 (0.2087)*	0.8206 (0.0095)*	0.7133 (0.0877)*	0.7456 (0.0581)*	0.7006 (0.0285)*

Note: The OMX 20 exhibited within the out-of-sample period under consideration an annual Mean of 10.01% and a annual volatility of 22.49% corresponding to a Sharpe-Ratio of 0.4451.
* See table 1.

6 Discussion

Davis, Fama and French (2000) figured out that selling stock portfolios containing small companies' stocks and buying stock portfolios containing large companies' stocks resulted at least from 1929 to 1997 in significant positive abnormal returns.

However, Schwert (2002) reports that effects, such as the size effect or the book-to-market effect seem to have weakened over time or simply disappeared after the research article that highlighted those empirical observations had been published. In this study it could be shown that even though portfolio 1 does not exhibit a cointegration relationship with the benchmark, cointegration could be asserted between the latter and portfolio 2 investing in small companies' stocks only.

Table 3: Statistical properties of cointegration based stock portfolios

Optimization Method	Correlation-Analysis				
Portfolio	6	7	8	9	10
Stock selection	1	2	3	4	5
Market capitalization at initial allocation**	82.11%	29.51%	76.93%	45.31%	60.05%
Annual return (out-of-sample)	8.84%	13.33%	8.84%	8.65%	8.80%
Annual Volatility (out-of-sample)	35.45%	27.63%	35.12%	35.72%	35.01%
Treynor ratio (out-of-sample)	7.9640	15.8690	8.0364	7.7232	8.0000
Rank Treynor ratio (out-of-sample)	9	1	6	10	7
Sharpe ratio (out-of-sample)	0.2494	0.4824	0.2517	0.2421	0.2514
Rank Sharpe ratio (out-of-sample)	9	1	7	10	8
In-sample Tracking-Error (TE)	31.85%	11.48%	31.92%	32.30%	31.80%
Out-of-sample Tracking-Erorr (TE)	25.26%	20.07%	24.83%	25.43%	24.73%
p-value Trace-test in sample	0.8648 (0.6622)*	0.8183 (0.0458)*	0.8598 (0.6546)*	0.8670 (0.6558)*	0.8597 (0.6561)*
p-value Trace-test out of sample	0.5791 (0.2321)*	0.6993 (0.0090)*	0.6183 (0.1972)*	0.6183 (0.1972)*	0.6177 (0.1958)*

* See table 1.

Consequently, Fama and French's (1996) empirical fact can be supported by this study since a cointegration relationship ensures the tracking error of a potential statistical arbitrage model to mean revert. Alexander and Dimitriu (2005) analyze the difference between cointegration and tracking error variance allocation methods, given an identical set of stocks in the two portfolios where they apply a naive approach of stock selection, only. Thereby, stocks are selected according to their price ranking, starting with the highest-priced stocks. Comparing the Sharpe ratios of stock portfolios estimated by applying cointegration analysis on the one hand and correlation analysis on the other hand, they figure out that correlation based models generate marginally better Sharpe ratios. However, the studies here suggest rather the opposite. All stock portfolios being based on cointegration clearly dominate their traditional counterparts. The stock selection approach 2 may consequently be considered as exception, only. Moreover, Alexander and Dimitriu (2005) report that the tracking errors from cointegration based models are in accordance to their studies well above those of the correlation based models. They argue that this may be an outcome of the optimization procedure since correlation based portfolios are specifically constructed to minimize the variance of the tracking error. Comparing the out-of-sample tracking errors of cointegration and correlation based models (see exhibits 1 and 2) shows, however, that the opposite is the case concerning this study here. Only the correlation based portfolio regarding stock selection approach 2 generates out-of-sample a tracking error being 1.72 percent points lower than the corresponding model being based on cointegration (see portfolios 2 and 6 on exhibits 1 and 2). Alexander and Dimitriu's (2005) approach corresponds to stock selection approach 1 as suggested here. Even though they conclude that their analysis taking into account no weight constraints exhibits no significant advantages or limitations of a cointegration relationship with the benchmark, it could be shown here that the cointegration based portfolio (i.e. portfolio 1) generated out-of-sample an annual return being 0.08 percent points higher and a volatility being 2.31 base points lower in comparison to the correlation based counterpart (i.e. portfolio 6) resulting in a Sharpe and Treynor ratios which are 7.94%, respectively, 4.68% higher.

Grobys (2010a) considers the Swedish stock market and selects 17 stocks out of an index containing 30 stocks. Thereby, the stock selection approach is related to the limitations to the data set being involved. Consequently, stock portfolios being analyzed contain the same stocks corresponding to a market capitalization of 56.67%. All cointegation based models being estimated exhibit higher Sharpe ratios in comparison to their correlation based counterparts which can be also supported in this study here. Furthermore, Grobys (2010a) mentions that even if optimization procedures based on cointegration analysis are employed, the portfolios do not necessarily show a cointegration relationship with the benchmark out-of-sample (i.e the out-of sample p-values in Grobys (2010a) studies are between 0.10-0.11). This outcome which is not in line with Alexander and Dimitriu (2005) can be supported here, too. Only portfolios 1 and 2 and 7 are cointegrated with the benchmark within the out-of-sample period exhibiting Trace-test p-values of 0.0095, 0.0285 and 0.0090. As those test statistics include a linear trend parameter, portfolios 1 and 2 may involve statistical arbitrage opportunities as mentioned by Grobys (2010b) (see figure 1).

In contrast to Phengis and Swanson (2011), the stock selection process is rather based on market capitalization, whereas the optimization procedure is based on cointegration as the assets' log-prices are employed in the optimization processes. The out-of-sample period is quite similar as Phengis and Swanson's (2011) studies who account for only six years out-of-sample data due to data set limitations. However, figuring out the potential benefits of a cointegration based stock selection processes within the domestic stock portfolio framework is left for future research as well as investigating the performance of both, different out-of-sample time windows and frequent rebalancing of the portfolios.

7 Conclusion

The study here attempts to throw light on the benefits of different stock selection approaches and two portfolio optimization procedures. Both factors are essential regarding portfolio management. Given a subset of stocks, the portfolio manger wants to find an asset allocation maximizing the return with the lowest possible volatility. On the one hand it could be shown that given a certain set of stocks, cointegration models dominate their traditional counterparts, basically. Consequently, earlier studies suggesting this finding could be supported. On the other hand the stock selection problem has a strong impact on the portfolio's performance. Investing in small stocks would have been a beneficial strategy when considering the Danish stock market under the last 10 years.

Figure 1: Statistical Arbitrage opportunities of cointegrated assets

Considering portfolios that account for a high market capitalization at the initial allocation time, it seems to be ambiguous that it was not possible to estimate a cointegration optimal asset allocation. Stock selection approaches 1 and 3 which correspond to initial market capitalizations of 82.11%, respectively, 76.93% do neither generate stock portfolios exhibiting a cointegration relationship with the benchmark in the-sample nor out-of-sample. However, stock selection approach 2 corresponding to the lowest initial market capitalization of 29.51% is the only stock selection approach that generates a portfolio being in- and out-of-sample cointegrated with the benchmark.

Moreover, there may be further need of research concerning the optimal stock selection approaches. The approaches being employed here lean on stock selection procedures which are taken into account in the academic literature. Apart from that it may be possible that the ascertained high negative correlation between initial market capitalization and Sharpe ratio may be an outcome purely by chance. Further studies may take into account stock selection approaches being based on randomly selections and estimate the corresponding statistics to evaluate possible advantages of investing in small stocks, or give further evidence for the ascertained negative correlation. Moreover, the stock portfolios being considered are not rebalanced. Taking into account frequent rebalancing as well as transaction costs may give further insights.

Future research can also take into account other countries while distinguishing between large and small economies. Studies may analyze the performance difference between larger and smaller economies. Another avenue may be that future research takes into account different time windows concerning the out-of-sample performance as well as different in-sample time windows while accounting for different initial states - such as bull and bear markets – at the time the assets are allocated.

ACKNOWLEDGEMENTS: I am grateful for the referee asking me for investigating the risk-return relationship of well diversified other European portfolios.

References

[1] C Alexander, Optimal Hedging Using Cointegration, *Philosophical Transactions of the Royal Society* Series **A**, **357**, (1999), 2039-2058.

[2] C. Alexander and A. Dimitriu, Indexing and Statistical Arbitrage, *The Journal if Portfolio Management*, (**31**), (2005), 50-63.

[3] J.L. Davis, E.F. Fama and K.R French, Characteristics, Covariances and Average Returns, 1929 to 1997, *Journal of Finance*, (**55**), (2000), 389-406.

[4] R. Engle and C.W.J. Granger, Co-integration and error correction, representation, estimation and testing, *Econometrica*, (**55**), (1987), 251–276.

[5] E.F. Fama and K.R. French, Multifactor Explanations of Asset Pricing Anomalies, *Journal of Finance*, (**51**), (1996), 55-84.

[6] S.M. Focardi and FJ. Fabozzi, A Methodology for Index Tracking Based on Time Series Clustering, *Quantitative Finance,* (**4**), (2004), 417-425.

[7] C.W.J. Granger, Some properties of time series data and their use in econometric model specification, *Journal of Econonometr*ics, (**16**), (1981), 121–130.

[8] K. Grobys, Correlation versus Cointegration. Do Cointegration based Index-Tracking Portfolios perform better? Evidence from the Swedish Stock-Market, *German Journal for Young Researchers*, (**2**), (2010), 72-78.

[9] K. Grobys, Do Business Cycles exhibit beneficial Information for Portfolio Management? An Empirical Application of Statistical Arbitrage, *The Review of Finance and Banking*, (**2**), (2010), 43-59.

[10] S. Johansen, Statistical analysis of cointegrating vectors, *Journal of Economic Dynamics and Control*, (**12**), (1988), 231-254.

[11] S. Johansen, Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models, *Econometrica*, (**59**), (1991), 1551-1581.

[12] S. Johansen and K. Juselius, Maximum likelihood estimation and inference on cointegration with applications to the demand for money, *Oxford Bulletin*

of Economics and Statistics, (**52**), (1990), 169–210.

[13] S. Johansen and K. Juselius, Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models, *Econometrica*, (**59**), (1991), 1551–1580.

[14] G.A. Larsen, G. Bruce and G. Resnick, Empirical Insights on Indexing, *The Journal of Portfolio Management*, (**25**), (1998), 51-60.

[15] H.M Markowitz, Portfolio Selection, *Journal of Finance*, (**7**), (1952), 77-91.

[16] K. Montefort, E. Visser and L.F. van Draat, Index Tracking by Means of Optimized Sampling, *The Journal of Portfolio Management*, (**34**), (2008), 143-151.

[17] C. Phengis and P.E. Swanson, Optimization, cointegration and diversification gains from international portfolios: an out-of-sample analysis, *Review of Quantitative Finance and Accounting*, (**36**), (2011), 269-286.

[18] R. Roll, A Mean Variance Analysis of Tracking Error, *The Journal of Portfolio Management*, (**18**), (1992), 13-22.

[19] G.W Schwert, *Anomalies and Market Efficiency,* Handbook of the Economics of Finance, G. Constantinides, R.M. Stulz, and Milton Harris (Eds.), North Holland, Amsterdam, 939-974, 2002.