HOW RANDOM SAMPLING SPOILS PERFORMANCE IN QUADRATIC INDEX TRACKING

MARTIN BOĎA, MÁRIA KANDEROVÁ
Matej Bel University in Banská Bystrica, Faculty of Economics
Department of Quantitative Methods and Information Systems
Tajovského 10, Banská Bystrica, Slovakia
e-mail: maria.kanderova@umb.sk, martin.boda@umb.sk

Abstract

In the paper, market capitalization and random sampling are compared as criterions of asset pre-selection for quadratic index tracking. Both approaches are considered in a simple and industry-stratified variant, and their contribution to performance is assessed against the deteriorating effect of transaction costs that becomes to manifest itself with an increasing number of assets in the tracking portfolio. In a case study for the S&P 500 Index, it is found that market capitalization works remarkably well as against random sampling, and that small portfolios are recommendable since they are not plagued by high transaction costs.

Key words: *asset pre-selection, market capitalization, random sampling, transaction costs, quadratic tracking.*

1. Introduction

In a recent paper of theirs, Boďa and Kanderová (2016) investigated and compared four approaches to asset pre-selection in quadratic portfolio tracking. Whereas partially accounting for the role of transaction costs, the object of their comparison was two market capitalization approaches and two random sampling approaches, in which both market capitalization and random sampling were considered in a simple as well as stratified variant. Transaction costs were introduced into their study through five scenarios differentiated according to the amount of investment, fixed transaction costs, variable transaction costs applied to the number of assets acquired, the investment amount and the number of holdings. Their study resulted in a recommendation of the market capitalization approach over the random sampling approach; however, the number of scenarios did not give an exhaustive insight into the impact of transaction costs on the investment outcome. Following the outline of the aforesaid study, the present paper aims to provide a more thorough view into the reasons why market capitalization should be preferred over random sampling other than those purely economic. In addition, the design of the investigations is helpful in assessing the trade-off between the size of transaction costs and the number of assets in a tracking portfolio and in evaluating their combined effect upon performance.

Before proceeding further, several short comments must be made and they especially concern the methods usually applied for asset pre-selection and the methods available for portfolio selection. Any task of portfolio selection requires that some assets must be chosen as suitable candidates for the resulting portfolio. A universe of assets is frequently reduced using the criterion of market capitalization (e.g. only large cap assets are chosen, or – conversely – the choice is made in favour of small cap assets in some cases) or by the screening method (according to which a suitable criterion of historical performance is made use of such as

Jensen's alpha or the P/E ratio). Admittedly, random sampling is seldom (or rather never) chosen as a criterion of asset pre-selection, possibly because it fails to have any reasonably connection with the economic rationale that assets pre-selected are picked in such a manner that they are likely to help in attaining higher performance. Randomness is void of economic reason and is thus avoided. Nevertheless, although many an objection may be easily raised against asset pre-selection based on randomness, it seems that the properties of random sampling designs in portfolio selections have not been studied hitherto, and at least they are not known to the authors. Having a set of assets pre-selected, it is usually a passive investment strategy that is utilized in allocating funds across the assets and in the specification of weights determining capital allocation. For good reasons explained in sufficient detail e.g. in Prigent (2007, p. 104), the intermediate goal is to mimic the performance of a suitably chosen financial index, and the task of index tracking is posited most frequently in the spirit of quadratic tracking error minimization (see e.g. Rudolf et al., 1999, p. 86). Quadratic tracking error is therefore opted for here in this paper as the method of portfolio choice from amongst assets pre-selected by a suitable algorithm.

The paper considers a number of portfolio selection situations, in which the Standard & Poor's (S&P) 500 Index is tracked using a portfolio of its constituents. As announced, a total of four of asset pre-selection algorithms are included into comparison: simple and stratified market capitalization, simple and stratified random drawing. Stratification is industrial and respects the GICS (Global Industry Classification Standard) taxonomy that was developed in 1999 by MSCI for use by the global financial community, and is applied also to the S&P 500 Index constituents. The four asset pre-selection algorithms serve to produce a set of assets for quadratic tracking, and the investment is carried out in the presence of transaction costs that lessen the merits of tracking portfolios. Transaction costs are recognized in varying scale in order to better capture the severity with which the performance of investment reacts to the size of portfolio (too large portfolios are associated with high transaction costs and their appeal is thus diminished). In addition, their effect is evaluated and confronted against the four algorithms of asset pre-selection.

In what follows, the paper is organized into four more sections. Section 2 comments briefly on asset pre-selection strategies implemented in portfolio selection and spotlights the role of transaction costs in portfolio selection. Section 3 explains the methodology of the paper, whilst Section 4 present the results Section 5 then concludes.

2. Asset Pre-selection and Transaction Costs

Frequently, several criteria are implemented in asset pre-selection such as market capitalization, turnover and relative turnover (see Rey and Seiler, 2001; Alexander and Dimitriu, 2005). Whilst investigating the impact of different pre-selection strategies, Rey and Seiler (2001) demonstrated that tracking error decreases as the number of assets rises, though not in a systematic way, and that including more and more assets does hardly result in significant lower tracking errors. Furthermore, they observed that market capitalization is a superior criterion of asset pre-selection (see Rey and Seiler, 2001, pp. 23-24). Also the study by Alexander and Dimitriu (2005) relied upon market capitalization and found that portfolios should contain at least 20 assets in order to deliver adequate performance with an ability to outperform the benchmark index. Another possibility is to classify assets into several disparate categories that reflect performance characteristics, industrial characteristics or country affiliation of issuers (see Rudd, 1980, p. 60; Fabozzi, 1998, pp. 58-62). After having assets stratified according to categories (industries, countries, and the like), the criterion of market capitalization qualifies assets for pre-selection into the tracking portfolio. The approach of stratified market capitalization was employed e.g. by Larsen and Resnick (1998), Focardi and Fabozzi (2004) and Frino et al. (2004). These studies find that that tracking portfolios composed of large-capitalization assets are preferable to those made up of small-capitalization assets in terms of tracking error. In addition, Larsen and Resnick (1998, p. 59) found that stratification across industry groups contributes only to portfolios created of large-capitalization assets.

Transaction costs encompass trading costs both explicit (direct) and implicit (indirect). The former group of transaction costs represents those that are disclosed prior to the trade and includes commissions, markups, and other fees. Representing the costs that are not determined until after the execution of a trade or set of trades is completed, implicit costs stem from the bid-ask spread and present themselves as illiquidity costs following from wide bid-ask spreads. Several studies examined the influence of transactions costs upon performance, such as Mezali and Beasley (2014), and Domowitz et al. (2001). The reader is recommended to Boďa and Kanderová (2016) for a short summary or to the original papers for further details.

3. Methodology

The paper follows the conventional model of portfolio selection based on quadratic index tracking and espouses the model of transaction costs charging that was adopted by Boďa and Kanderová (2016) in an earlier study of the sort. Therefore, the presentation here is shortened to a traceable minimum and the interested reader is referred to the cited paper. In this, the optimization model does not take into account transaction costs that arise with a particularly chosen portfolio. Transaction costs are treated as exogenous and apply once the optimal solution for weights has been found.

Assume that a history of T historical observations of logarithmic returns is available and that the tracking portfolio is to be composed of k assets. Let $\mathbf{Y} = (Y_1, \ldots, Y_T)'$ denote a $(T \times 1)$ vector of benchmark returns, and $\mathbf{X} = (\mathbf{x}_1 | \ldots | \mathbf{x}_T)'$ denote a $(T \times k)$ matrix of returns of the k assets that are to be represented in the tracking portfolio. In the ensuing exposition, elements of vectors are denoted by left-hand subscripts in conjunction with curly brackets, e.g. an i-th element of a vector $\mathbf{z}$ is written as $\{\mathbf{z}\}_i$. The symbol $\boldsymbol{\omega}$ stands for a $(k \times 1)$ vector of unknown portfolio weights that result from the quadratic optimization problem

$$\min_{\boldsymbol{\omega} \in \Re^k} (\mathbf{Y} - \mathbf{X}\boldsymbol{\omega})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\omega}) \quad \text{subject to} \quad \boldsymbol{\omega}'\mathbf{1} = 1 \quad \text{and} \quad \text{possibly some other constraints,} \tag{1}$$

in which $\mathbf{1}$ is a vector of k ones. This general formulation of the optimization task allows an extension and can be complemented by the constraint banning short sales requiring that $\{\boldsymbol{\omega}\}_i \geq 0$ for $\forall i \in \{1,\ldots,k\}$. This constraint is utilized in the practical investigation and long-only positions are sought, although also other constraints are encountered in practice.

Suppose that a budget B is available for the investment and the investor faces two variants of transaction costs, both of a variable nature. Some costs are charged to the number of assets traded and some to the number of holdings acquired (or sold providing that also short positions are allowed). The lump charges of variable costs are denoted as χ_A per asset traded and χ_H per holding of an asset purchased or sold. Suppose that there are $k^{\#}$ non-zero weights in the optimal solution $\boldsymbol{\omega}^{\#}$ to (1) extended by the requirement of short sales exclusion and assume further that the prices of assets at the instance of portfolio construction are organized in an $(k \times 1)$ vector $\mathbf{P}$. The variable costs then amount to $k^{\#} \cdot \chi_A + \chi_H \cdot \sum |B^{\S} \cdot \{\boldsymbol{\omega}^{\#}\}_i / \{\mathbf{P}\}_i|$, where $B^{\S} \cdot \{\boldsymbol{\omega}^{\#}\}_i / \{\mathbf{P}\}_i$ denotes the holdings of an i-th asset derived from the actual sum that can be allocated to the investment after satisfying all the transactions costs applicable. This suggest the equation $B^{\S} = B - k^{\#}\cdot\chi_A - \chi_H\cdot\sum|B^{\S}\cdot\{\boldsymbol{\omega}^{\#}\}_i/\{\mathbf{P}\}_i|$ that can be straightforward solved for $B^{\S}$. The considerations can be extended also for fixed costs and other transactions costs (see Boďa and Kanderová, 2016), but it transpires that these two forms are sufficient to control for transaction costs.

The analysis revolves around tracking the S&P 500 Index for an investor who is eager to select his portfolio as of 30 March 2015. To this end, he uses historical returns data observed on a weekly basis spanning the period from 9 January 2012 to 30 March 2015 (the in-sample period) and his investment period is about one year from 30 March 2015 until 28 March 2016 (the out-of-sample period). The in-sample period counts 169 effective weekly returns and they are used for portfolio tracking, whilst the out-of-sample period comprises 53 effective weekly returns and prices that serve the purpose of performance evaluation. The portfolio is to be composed out of 502 equities forming the S&P 500 Index as of 30 March 2015 that are classified into 10 GICS sectors. That being said, the practical implementation of the task permits only a lower number of 463 equities owing to the fact that some equities are only fresh constituents with an insufficient history or ceased to be traded during the out-of-sample period.[1] Under random sampling, equities for index tracking are pre-selected as a simple random draw from the basket of 463 equities (for the simple variant) or as a stratified random draw using the stratas represented by the 10 GICS sectors (for the stratified variant). Under market capitalization, only equities with the greatest market capitalization regardless of their sectorial classification (for the simple variant) or across the GICS stratas (for the stratified variant) are chosen. Some rounding errors emerge with the stratification variants of both methods with a negligible effect.

The usefulness of the four asset pre-selection strategies is evaluated in the background of several configurations for transaction costs (graded from almost zero costs to considerable costs) that are designed to help in isolating the role of transaction costs upon the performance of the resultant tracking portfolio. It is assumed in these configurations that the budget available for investment is B = US$ 1000 and this budget is offset by the lump amount of χ_H = US$ 0.001 per one holding of assets acquired and by the amount of variable transaction costs per one unit of asset χ_A varying from US$ 0 to US$ 1 by US$ 0.05 (i.e. there are a total of 21 ranging values of χ_A).[2]

On the basis of the observations in the in-sample period, the weights of long-only tracking portfolios are determined using optimization task (1) for each scenario and for 10, 30, 50, 75, 100, 125 and 150 equities pre-selected with one of the four methods: simple sampling, simple market capitalization, stratified sampling and stratified market capitalization. In contrast to the sampling methods, the pre-selection conducted by means of either market capitalization method is for each number of equities deterministic and unique. There is only one possibility to select a certain number of constituents from the S&P 500 Index so that their market capitalization tops the ranked list. For each random draw the pre-selection result of simple sampling and stratified sampling from the basket of S&P 500 Index components should be in general different. With this in mind, for each scenario and for each number of pre-selected equities under consideration a total of 100 independent draws (without replacement) are

[1] In consequence of a loss in the effective number of equities, the stratification of equities was as follows: Consumer Discretionary (78 equities), Consumer Staples (34 equities), Energy (39 equities), Financials (81 equities), Health Care (49 equities), Industrials (63 equities), Information Technology (62 equities), Materials (26 equities), Telecommunications Services (6 equities), Utilities (25 equities).

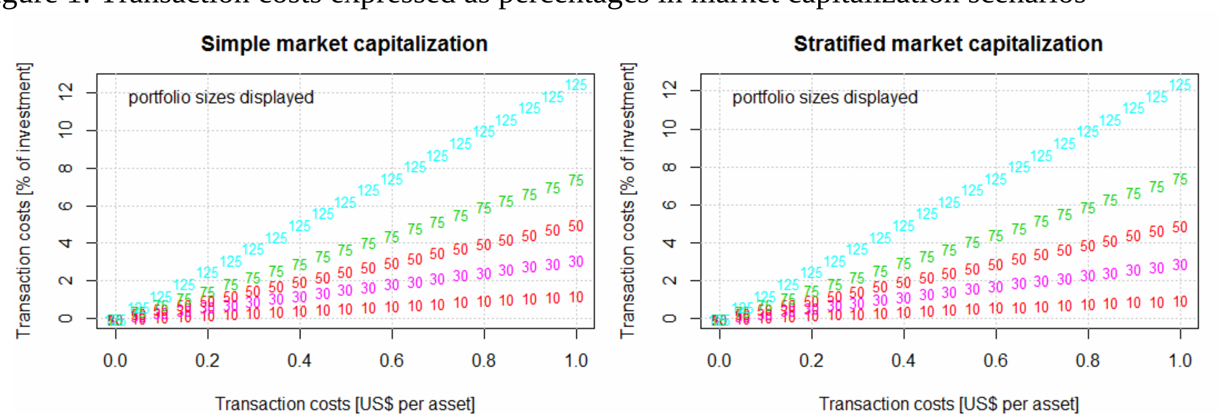
[2] There should be some trade-off between the former and the latter because a highly populated portfolio (composed of many assets) suggest lower holdings, and if there are only few assets in a portfolio, the holdings must be greater. This is the way by which ideas of fixed transaction costs and economies of scale can be incorporated and inserted into the assessment, as if in passing. That being said, this trade-off does not prove itself in this present study.

performed and the performance results are averaged or summarized in a conveniently informative way. The results are shown in the next section.

4. Results

The importance of transactions costs is underlined by a visual representation in Figure 1. Albeit in sampling scenarios the ultimate amount of transactions costs that are applicable at the moment of making an investment depends on a particular drawing of assets, a rough guess regarding the magnitude of transaction costs in individual settings can be inferred from the market capitalization scenarios differentiated according to the number of assets pre-selected k and the unit costs per asset χ_A. In this respect, Figure 1 displays transaction costs recalculated as percentages relative to the intended amount of investment. These percentages are computed as $(1 - B^{\S}/B)\cdot 100$ [%]. Smaller portfolios are associated with lower transactions costs well below 2 % of initial investment, whereas transaction costs for portfolios to be created out of 125 assets in this configuration may climb up to about 12 % of the invested sum.

Figure 1: Transaction costs expressed as percentages in market capitalization scenarios



Source: the authors.

Figures 2 to 4 show values of the tracking portfolio as of 28 Mar 2016 at the end of the out-of-sample period for all considered scenarios. These graphical displays report how the final portfolio value is affected by the amount of transaction costs and the number of assets pre-selected. By the agency of Figure 1, transaction costs charged per unit of an asset that displayed on the horizontal axis can be directly mapped for different sizes of the tracking portfolio and their impact can be translated in terms of percentages. The horizontal line at US$ 1000 reminds of the initial amount of investment. Sampling scenarios are juxtaposed and compared to market capitalization scenarios, which confirms superiority of the market capitalization criterion. In few cases of random sampling final portfolio values are higher than those for market capitalization, but they are merely a lucky chance than a systematic pattern. The market capitalization criterion may be recommended for any size of the tracking portfolio and any magnitude of transaction costs.

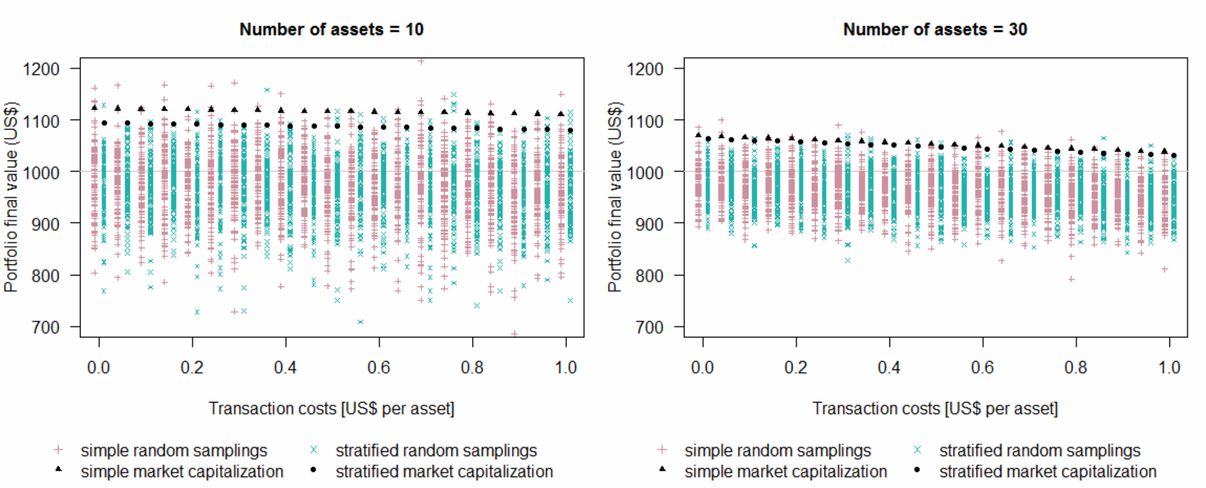
Additionally, three more rules of behaviour are discernible from Figures 2 to 4.

- First, transaction costs have a devastating effect upon performance and the performance of the tracking portfolios (perceived in terms of final portfolio values) more or less monotonously decreases alongside increasing transaction costs, which is especially readable for highly populated scenarios.
- Second, smaller portfolios are preferable over highly populated portfolios. Although tracking portfolios of 10 assets present high variability in terms of portfolio values with

high likelihood of an unsatisfactory result, tracking portfolios of 30 assets are more acceptable. From 50 assets, transaction costs begin to offset the performance they would otherwise show.

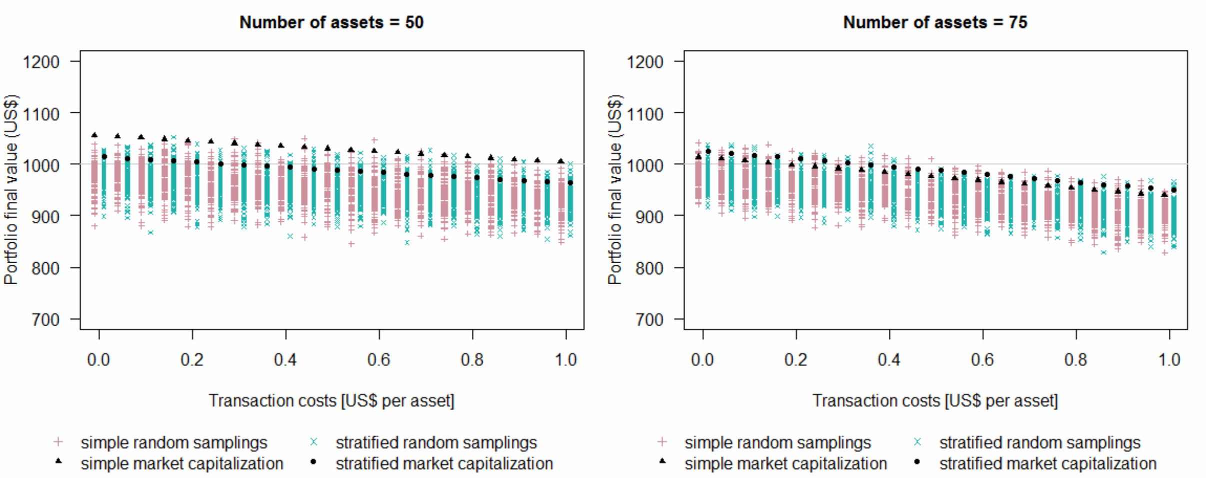
- Third, it cannot be universally said whether simple or market capitalization is recommendable. For 10 and 50 assets simple market capitalization tracking portfolios outperform those formed under stratified market capitalization. For 30 and 100 assets the merits are almost balanced (in slight favour of simple market capitalization). Nonetheless, for 75 and 125 assets, stratified market capitalization criterion seems better than simple one. Still, for smaller portfolios simple market capitalization appears to work remarkably well.

Figure 2: Final portfolio values for 10 and 30 assets pre-selected



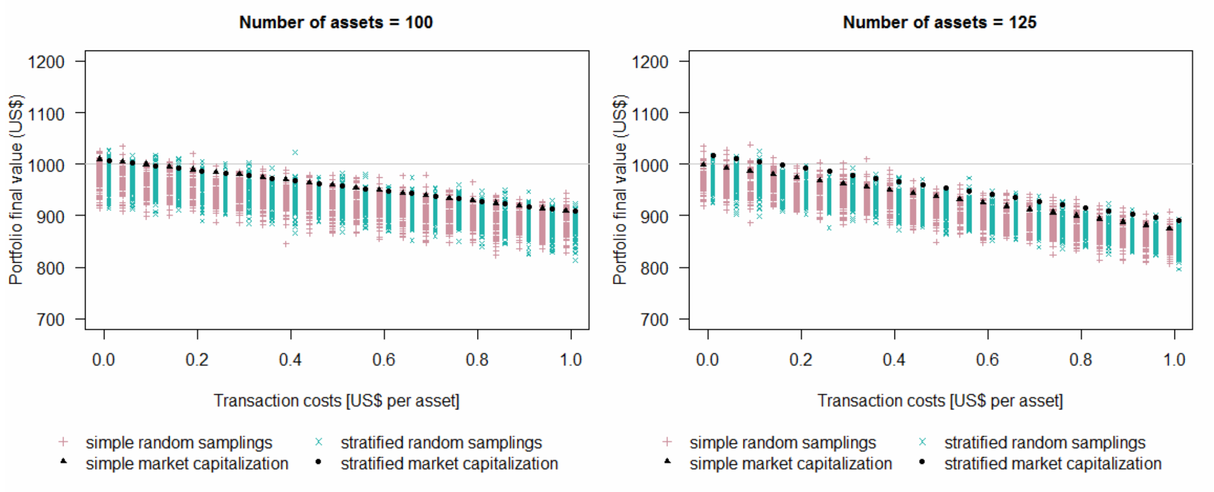
Source: the authors.

Figure 3: Final portfolio values for 50 and 75 assets pre-selected



Source: the authors.

Figure 4: Final portfolio values for 100 and 125 assets pre-selected

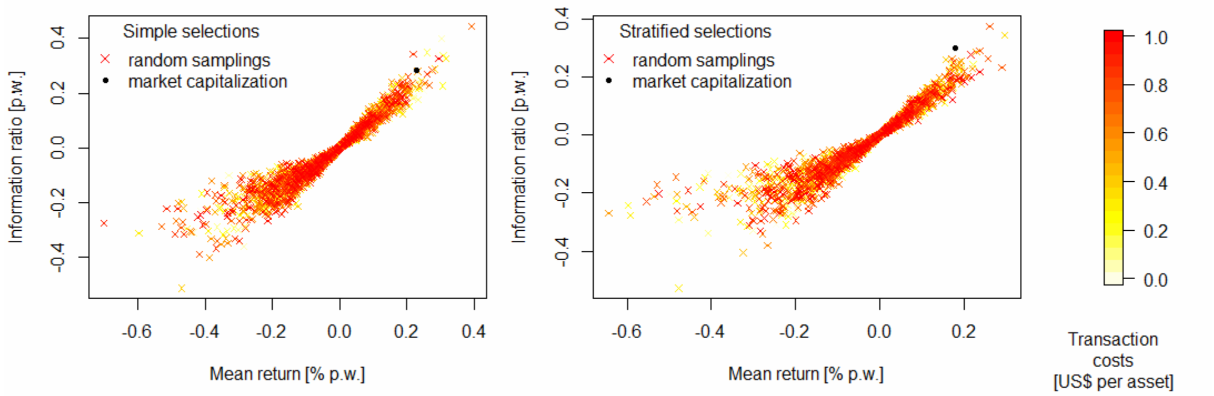


Source: the authors.

Another confirmation of the superiority of the market capitalization criterion is Figures 5 to 10 that plot for each size of the tracking portfolio and for either variant of pre-selection mean weekly returns and weekly information ratios for the out-of-sample period. Combinations of mean returns and information ratios for sampling scenarios are coloured according to the magnitude of transaction costs, but this colouring seem of no avail. It is seen that market capitalization tracking portfolios are capable of yielding comparatively high mean returns at low risk as seen in high values of the information ratio. It is withal apparent that – as might be anticipated on account of the very nature of random sampling – tracking portfolios that arise from sampling are prone to produce negative mean returns somewhat too frequently. Although these statistical descriptives, being in relative terms, do not reflect transaction costs and are nohow influenced by their size, it is it is obvious that with an increasing size of the tracking portfolio both mean returns and information ratios tilt more toward zero values, which is just another point against portfolios of large size.

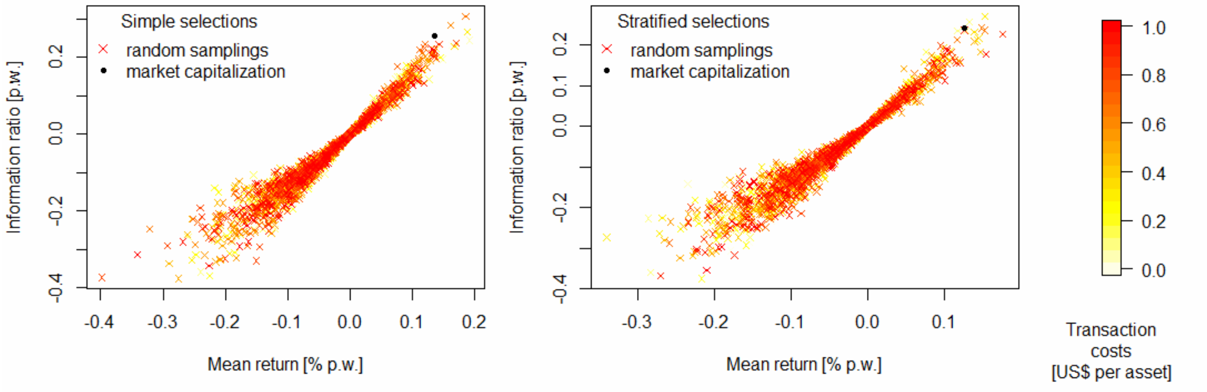
In computations and preparing graphical presentations, the software R version 3.0.1 (R Core Team, 2013) was employed with several of its libraries, `fPortfolio`, `quadprog` and `timeSeries`.

Figure 5: Performance of tracking portfolios for 10 assets pre-selected



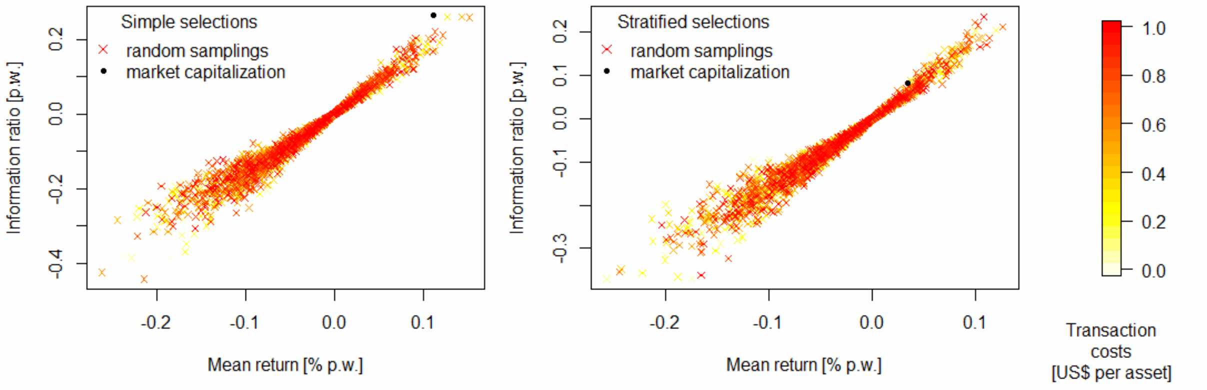
Source: the authors.

Figure 6: Performance of tracking portfolios for 30 assets pre-selected



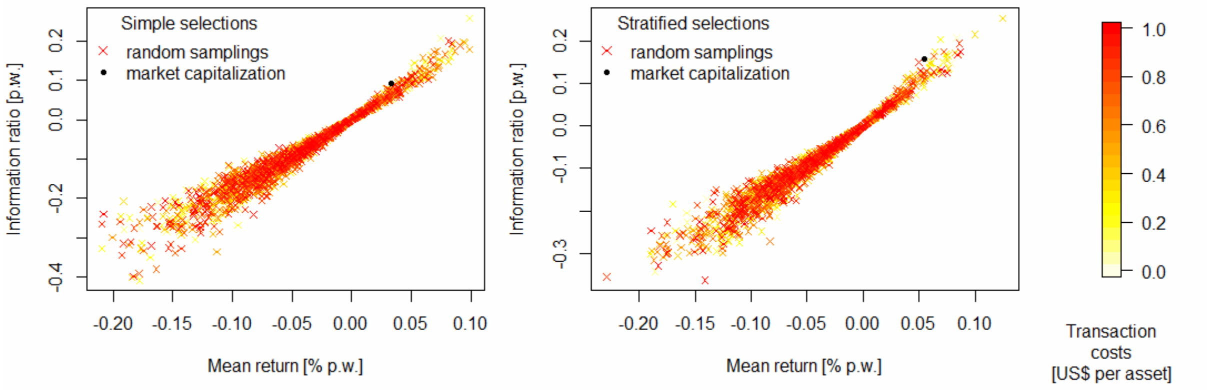
Source: the authors.

Figure 7: Performance of tracking portfolios for 50 assets pre-selected



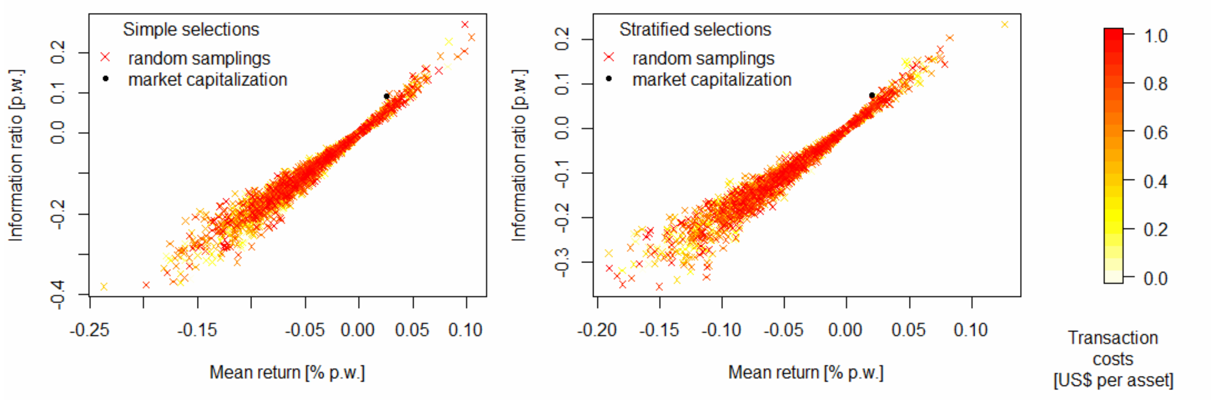
Source: the authors.

Figure 8: Performance of tracking portfolios for 75 assets pre-selected



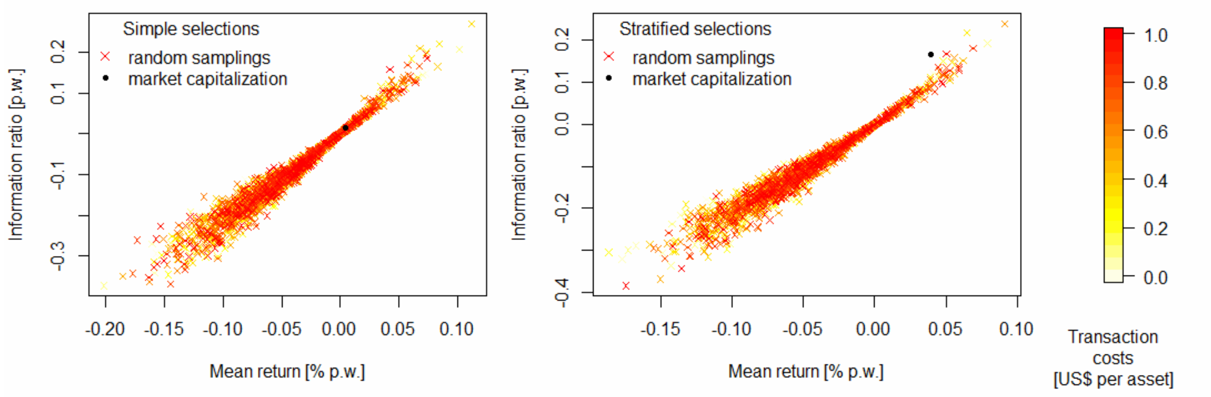
Source: the authors.

Figure 9: Performance of tracking portfolios for 100 assets pre-selected



Source: the authors.

Figure 10: Performance of tracking portfolios for 125 assets pre-selected



Source: the authors.

5. Conclusion

The results here are in full conformity with those in Boďa and Kanderová (2016) and give a more systematic outlook upon the performance of both market capitalization and random sampling when they are used for asset pre-selection in quadratic index tracking. Although Boďa and Kanderová (2016) considered a wider spectrum of transaction costs, they evaluated only 5 configurations of transaction costs by balancing the investment sum and the size of transaction costs. Here, the situation is somewhat simplified, yet at no loss of informational value, as only two forms of transaction costs are taken into consideration that penalize the investment in a monotonous way in relation to the size of the tracking portfolio. The findings say in favour of smaller portfolios – certainly of no higher size than 50 assets and these portfolios may be pre-selected on the basis of market capitalization rather than as thoughtless mechanical draws from a benchmark financial index. Simple market capitalization complies outstandingly well and conforms to this purpose.

Of course, one might object to using the formulation of quadratic tracking error as the method of selecting the tracking portfolio and might wonder if a different method of portfolio optimization would not establish different findings. The reason being, there are other formulations of index tracking (basing e.g. on linear programs and thinking in terms of absolute deviations), and there are a number of active approaches as well. That being said, quadratic index tracking is a standard approach and its economic rationale is indisputable. In addition, this approach also highlights the statistical qualities of square errors. The authors somehow do not believe that a different approach replacing quadratic index tracking would change the conclusion and jeopardize the superiority of small portfolios and market capitalization. Possibly, an active portfolio optimization approach might suggest a different conclusion, but the investigation of this surmise is just a topic for future research.

Acknowledgements

The support of the grant scheme VEGA # 1/0554/16 "Alternatives to replicating portfolio selection in the conditions of frequent rebalancing" is acknowledged.

References

[1] ALEXANDER, C., DIMITRIU, A. 2005. Indexing and statistical arbitrage. In Journal of Portfolio Management, 2005, vol. 31, iss. 2, pp. 50-63.

[2] BOĎA, M., KANDEROVÁ, M. 2016. Comparison of approaches for asset pre-selection in portfolio tracking. In BOĎA, M., MENDELOVÁ, V. (eds.) Applications of Mathematics and Statistics 2016. Conference proceedings. Banská Bystrica : Občianske združenie Financ, 2016. ISBN 978-80-89438-04-4. Pp. 44-57.

[3] DOMOWITZ, I., GLEN, J., MADHAVAN, A. 2001. Liquidity, volatility, and equity trading costs across countries and over time. In International Finance, 2001, vol. 4, iss. 1, pp. 221-255.

[4] ELTON, E. J. et al. 2009. Modern portfolio theory and investment analysis. New York : Wiley. 2009. ISBN 978-0470388327.

[5] FABOZZI, F. J. 1998. Overview of equity style management. In FABOZZI, F. J. (ed.) Active equity portfolio management. New Hope : Frank J. Fabozzi Associates, 1998. ISBN 978-122324930-9. Pp. 57-70.

[6] FOCARDI, S. M., FABOZZI, F. J. 2004. A methodology for index tracking based on time-series clustering. In Quantitative Finance, 2004. vol. 4, iss. 4, pp. 417-425.

[7] FRINO, A. et al. 2004. Index design and implications for index tracking : Evidence from S&P 500 Index funds. In The Journal of Portfolio Management, 2004, vol. 30, iss. 2, pp. 89-96.

[8] LARSEN, G. A, RESNICK, B. G. 1998. Empirical insights on indexing : How capitalization, stratification and weighting can affect tracking error. In Journal of Portfolio Management, 1998, vol. 25, iss. 1, pp. 51-60.

[9] MEZALI, H., BEASLEY, J. E. 2014. Index tracking with fixed and variable transaction costs. In Optimization Letters, 2014, vol. 8, iss. 1, pp. 61-80.

[10] PRIGENT, J.-L. 2007. Portfolio optimization and performance analysis. Boca Raton : Chapman & Hall/CRC. ISBN 1-58488-578-5.

[11] R CORE TEAM 2013. R : A language and environment for statistical computing. Vienna : R Foundation for Statistical Computing, 2013, http://www.r-project.org/.

[12] REY, D. M., SEILER, D. 2001. Indexation and tracking errors. University of Basel, WWZ Department of Finance working paper no. 2/01. [cit. 16-04-2016] http://www.econbiz.de/ archiv1/2009/95490_indexation_trackingerrors.pdf.

[13] RUDD, A. 1980. Optimal selection of passive portfolios. In Financial Management, 1980, vol. 9, iss. 1, pp. 57-66.

[14] RUDOLF, M., WOLTER, H.-J., ZIMMERMANN, H. 1999. A linear model for tracking error minimization. In Journal of Banking & Finance, 1999, vol. 23, iss. 1, pp. 85-103.