

Optimal portfolio selection for the small investor considering risk and transaction costs

Rainer Baule

Published online: 11 November 2008

© The Author(s) 2008. This article is published with open access at Springerlink.com

Abstract A direct application of classical portfolio selection theory is problematic for the small investor because of transaction costs in the form of bank and broker fees. In particular, minimum fees force the investor to choose a comparatively rather small selection of assets. The existence of transaction costs leads to an optimization problem that juxtaposes those costs against the risk costs that arise with portfolios consisting of only a few assets. Despite the non-convex and, thus, complex optimization, an algorithmic solution turns out to be very fast and precise. An empirical study shows that, for smaller investment volumes, transaction costs dominate risk costs so that optimal portfolios contain only a very small number of assets. Based upon these results, the cost-effectiveness of direct investments is compared to alternative vehicles, particularly index certificates and exchange-traded funds, depending on the level of invested wealth.

Keywords Portfolio selection · Transaction costs · Non-convex optimization

1 Introduction

Classical portfolio selection theory calls for investors to diversify their portfolios and divide their wealth among a large number of securities. Applying this theory, private investors with limited investment volume are forced to spread their purchasing power

The author thanks Olaf Korn, Thorsten Poddig, Walter Zucchini, participants of the 18th Australasian Finance and Banking Conference (Sydney), and two anonymous referees for valuable comments and suggestions.

R. Baule (✉)

Chair of Finance, University of Göttingen, Platz der Göttinger Sieben 3, 37073 Göttingen, Germany
e-mail: rbaule@uni-goettingen.de

so thin that most of the assets enter the portfolio with only a very small proportion. While optimal in a theoretical sense, in practice this procedure can be very expensive because of the transaction costs associated with the purchase of each security; hence, the optimal portfolio selection in the presence of transaction costs might differ from the classical solution. Particularly for the small investor, transaction costs in the form of minimum fees play a decisive role, since they appear with each transaction, independent of its size. These minimum fees make the transaction cost function nonlinear, producing a comparably complex optimization problem.

However, the finance industry offers a broad range of products that promise the benefits of diversification without the curse of high transaction costs for small investment amounts. A very popular alternative to a direct investment in single stocks are mutual funds. However, although mutual funds keep direct transaction costs to a minimum, premia and periodic management fees are common. While they are smaller for passively managed index tracking funds, they are not negligible.

A second alternative are index certificates, which are securities issued by a growing number of banks, entitling the holder to participate directly in the performance of the underlying index, that is, to receive an amount of money equivalent to the value of the index at maturity. As issuers usually sell index certificates at a price equal to the current index value, these investments seem to be free of bank fees. However, many indices, such as the EUROSTOXX 50, are price indices, which are calculated without a correction for paid dividends. This involves implicit issuer fees of some 2–3% annually, which makes the certificates comparably less attractive. A second type of hidden costs, even for return indices such as the German DAX, arises because, unlike mutual funds, index certificates are unsecured debt of the issuer and are subject to its credit risk.¹ Although the majority of issuers have high investment-grade credit ratings, the negative value of credit risk can amount to some dozens of base points.² In light of the recent crisis at the financial markets, issuer default risk has even increased.

Exchange-traded funds, a new financial innovation, combine the advantages of passively managed mutual funds and index certificates. These funds have lean management and, thus, lean cost structures and are listed at a special market segment of a stock exchange. In the case of index-tracking funds, they have very competitive management fees. In addition, the funds are separated from the assets of the investment company, so they face no risk of default. Nonetheless, transaction costs are not zero, and the investor is restricted to the passive replication of the index.

There are two situations in which a closer look at the portfolio selection problem of the small investor who intends to invest in single stocks, rather than one of the mentioned alternatives, is worthwhile. The first is when the investor has special knowledge about the (joint) distribution of the single stock returns that is different from common beliefs of the market. In that case, even if the investor was able to purchase low-cost index products, his information might lead him to follow a different strategy than replicating the index. The second situation is when, even

¹ See [Stoimenov and Wilkens \(2005\)](#).

² See [Baule et al. \(2008\)](#).

without private information, a direct investment might be superior to other investment vehicles for certain levels of invested wealth because of the mentioned types of costs. This paper focuses on the second situation by calculating the characteristics of the optimal portfolio of single stocks in the presence of transaction costs and comparing that portfolio to the two major alternatives, i.e., index certificates and exchange-traded funds, for different levels of invested wealth. The primary objective of the paper is to analyze for which investment volumes a direct investment is superior.

Based upon the seminal works of [Markowitz \(1952\)](#) and [Tobin \(1958\)](#), a number of contributions in the 1970s provide some insight into the portfolio selection problem when the number of securities in the portfolio is limited. [Mao \(1970\)](#) analyzes homogenous portfolios with different numbers of securities and finds that relatively few securities are needed to gain most of the benefits of diversification. [Jacob \(1974\)](#) formulates a portfolio selection problem with a cardinality constraint, i.e., a restriction in the number of securities. [Brennan \(1975\)](#) explicitly accounts for transaction costs and derives the optimal number of securities in the portfolio for the [Sharpe \(1963\)](#) index model.

The major part of the literature treating portfolio optimization in the presence of transaction costs focuses on linear, or at least continuous, cost functions. In the case of fixed transaction costs, [Patel and Subrahmanyam \(1982\)](#) and [Blog et al. \(1983\)](#) present algorithms for special forms of the covariance matrix of returns. In the past decade, some contributions address discrete constraints in the form of transaction costs or lots mainly under computational issues. [Sankaran and Patil \(1999\)](#) derive an algorithm for the cardinality-constraint portfolio selection with homogenous correlations. [Bienstock \(1996\)](#) applies branch-and-bound methods for this problem in the general case. [Maringer and Kellerer \(2003\)](#) and [Maringer \(2005\)](#) propose simulated annealing. Several other heuristics—including taboo search, threshold accepting, and genetic programming—are applied to similar problems by [Mansini and Speranza \(1999\)](#), [Bertsimas et al. \(1999\)](#), [Gilli and Këllezli \(2002\)](#), [Chang et al. \(2000\)](#) and [Jobst et al. \(2001\)](#). Recently, [Lobo et al. \(2007\)](#) address the problem of fixed transaction costs and propose an efficient heuristic algorithm for maximizing the expected portfolio return, subject to a number of various constraints.

However, little has been said about the particular optimization problem for small investors who seek to maximize their expected utility. In light of transaction costs, the optimal portfolio will contain only a selection of all available assets, compared to classical Markowitz theory. As a result, these portfolios harbor a measure of unsystematic risk which is not rewarded. In this paper we formulate the optimization problem in terms of costs arising from bearing, on one hand, unsystematic risk and, on the other, transaction costs (Sect. 2.1). We demonstrate how this problem can be solved algorithmically (Sect. 2.2). The empirical part of the paper (Sect. 3) analyzes the trade-off between risk costs and transaction costs for exemplary investment universes covering (i) the EUROSTOXX 50 and (ii) the Dow Jones industrial average 30 stocks, with a view to the relationship between investment volume and transaction cost size. The paper's major contribution is the empirical assessment of a direct investment compared to index certificates and exchange-traded funds for different levels of invested wealth.

2 Theoretical model

2.1 Risk costs versus transaction costs

We consider portfolio optimization in an extended Markowitz framework. Particularly, we reflect on individual investors who maximize their single-period expected utility. The corresponding preference function is expressed solely in terms of expected return and variance, which can be justified by either a quadratic utility function or normally distributed asset returns.³ In the case of normal returns and exponential utility, the preference function can be stated as

$$u(\mu, \sigma^2) = \mu - \gamma\sigma^2, \quad (1)$$

where $\gamma \geq 0$ is the risk aversion parameter.⁴

There are n risky assets with random return X_i with expectation μ_i and standard deviation σ_i , $i = 1, \dots, n$.⁵ Furthermore, there is a riskless asset with return r . The investor wishes to invest a fixed amount of money x in these assets—that is, to choose a portfolio with (random) terminal wealth

$$P = x_0(1 + r) + \sum_{i=1}^n x_i(1 + X_i), \quad (2)$$

subject to the budget constraint

$$\sum_{i=0}^n x_i = x. \quad (3)$$

Furthermore, we introduce short-selling restrictions, an assumption which is reasonable for the private investor. Hence, the single investments must be non-negative, that is,

$$x_i \geq 0 \quad \text{for } i = 1, \dots, n. \quad (4)$$

Transaction costs such as bank or broker fees are associated with any investment $x_i > 0$. The simplest type of transaction costs are linear costs. However, it is very common in banking practice for minimum fees to apply to each transaction, an aspect that is particularly relevant for the small investor. Hence, a more realistic modelling of transaction costs should take this nonlinearity into account. In this paper we allow for transaction costs with minimum fees of the following type:

$$c(x_i) = \max\{a, b x_i\} 1_{\{x_i > 0\}} \quad \text{with } a, b > 0. \quad (5)$$

³ See, e.g., [Haugen \(2001\)](#), p 201–204.

⁴ This is consistent with a utility function $U(z) = -e^{-2\gamma z}$, see, e.g., [Sargent \(1987\)](#), p 154 f.

⁵ The setup is similar to the optimization model of [Maringer \(2005\)](#), p 78 ff.

Within this type of function, costs are constant at the minimum level up to the amount $x_i = a/b$. Above this amount, costs increase linearly. The indicator function $1_{\{x_i > 0\}}$ models the discrete jump from 0 to a when the investment volume increases from zero to greater than zero, i.e., when a security enters into the portfolio.

Transaction costs lower the initial investment volume, or equivalently, lower the expected portfolio return. We assume that the riskless asset is free of transaction costs. This assumption is reasonable since, for instance, German government bills can be purchased, even in small amounts, without costs. Thus, (2) becomes

$$P = x_0(1 + r) + \sum_{i=1}^n (x_i - c(x_i))(1 + X_i). \tag{6}$$

The investor seeks to maximize the preference function of the portfolio’s expected return and variance, $u(E[P]/x - 1, \text{Var}[P]/x^2)$.

Since Markowitz and Tobin, it has been well known that in the absence of transaction costs, the set of efficient portfolios consists of linear combinations of a tangential portfolio T and the riskless asset. That is, there is a parameter λ_C^T ,⁶ the Sharpe ratio of the tangential portfolio, so that, for a given σ , the optimal expected return $\mu_C^T(\sigma) = r + \lambda_C^T \sigma$ can be achieved. Let σ_C^* be the optimal choice regarding the investor’s preference function in absence of transaction costs, i.e.,

$$(\mu_C^*, \sigma_C^*) = \arg \max u(\mu_C^T(\sigma), \sigma). \tag{7}$$

If transaction costs of the form (5) with a discrete component have to be taken into account, it will generally no longer be optimal to buy this portfolio; instead, portfolios with a smaller number of assets and lower transaction costs might be more advantageous.

The crucial point of the transaction cost function is the discrete component: transaction costs jump by a finite amount if an asset is taken into the portfolio, even at infinitesimally small volumes. This component is responsible for the problem’s being no longer convex, so that classical methods cannot be applied.

However, if the set of assets with strictly positive weights in the portfolio is fixed, the transaction cost function is convex; thus, classical portfolio selection on this subset of assets can be performed. Hence, a theoretical approach to solving the optimization problem would be to conduct a classical portfolio selection for each of the 2^n possible subsets.⁷ Accordingly, the set of efficient portfolios is, again, the set of linear combinations of the riskless asset and a new tangential portfolio, which is calculated by the classical method but on a subset of all risky assets. The situation is demonstrated graphically in Fig. 1.⁸

⁶ The subscript C indicates classical Markowitz-Tobin portfolios.

⁷ See Lobo et al. (2007), p 353. Strictly speaking, this is true only for transaction costs that are the sum of a constant and a linear component. However, the following analysis is independent of the shape of the cost function.

⁸ See also Levy (1978), p 645, for a similar illustration.

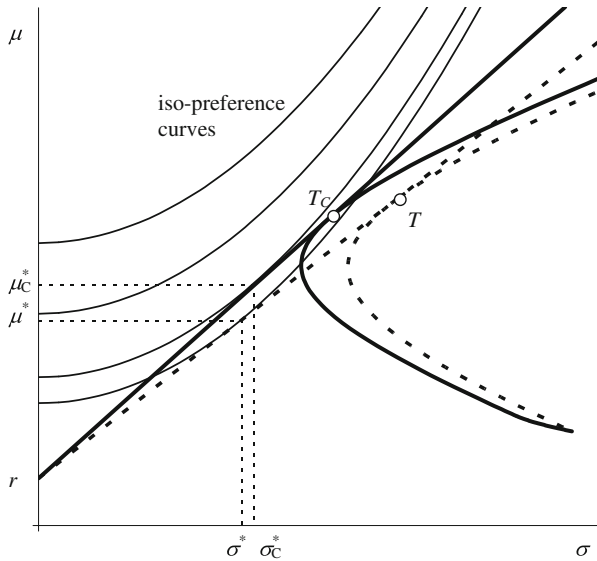


Fig. 1 Portfolio optimization with and without transaction costs: without transaction costs, the solid efficient frontier and the tangential portfolio T_C can be achieved. In the presence of transaction costs, only a selection of assets enters into the portfolio, which leads to the dotted efficient frontier and the tangential portfolio T . For a given preference function, the parameters of the optimal portfolio change from μ_C^* to μ^* and from σ_C^* to σ^* , respectively

The new tangential portfolio is suboptimal in the sense that its Sharpe ratio λ^T is lower than λ_C^T . Furthermore, the remaining transaction costs lower expected portfolio returns. Together, the expected return of a portfolio on the new efficient line becomes

$$\mu = \mu^T(\sigma) = r + \lambda^T \sigma - \phi, \tag{8}$$

where

$$\phi = \sum_{i=1}^n \frac{c(x_i)}{x} (1 + r + \lambda^T \sigma) \tag{9}$$

are the relative transaction costs with respect to the total investment volume, compounded to the end of the period.

From the iso-preference curves $\mu - \gamma \sigma^2 = \text{const.}$, the first-order condition for the optimal portfolio is easily derived as

$$\frac{\partial \mu}{\partial \sigma} = 2\gamma \sigma \stackrel{!}{=} \lambda^T \tag{10}$$

$$\Rightarrow \sigma^* = \frac{\lambda^T}{2\gamma}. \tag{11}$$

Hence, transaction costs reduce the optimal preference value by the amount

$$\begin{aligned}
 u(\mu_C^T(\sigma_C^*), \sigma_C^*) - u(\mu^T(\sigma^*), \sigma^*) &= \lambda_C^T \sigma_C^* - \gamma(\sigma_C^*)^2 - \lambda^T \sigma^* + \gamma(\sigma^*)^2 + \phi \\
 &= \lambda_C^T \frac{\lambda_C^T}{2\gamma} - \gamma \left(\frac{\lambda_C^T}{2\gamma} \right)^2 - \lambda^T \frac{\lambda^T}{2\gamma} + \gamma \left(\frac{\lambda^T}{2\gamma} \right)^2 + \phi \\
 &= \frac{(\lambda_C^T)^2}{4\gamma} - \frac{(\lambda^T)^2}{4\gamma} + \phi \\
 &= (\lambda_C^T - \lambda^T) \frac{\lambda_C^T + \lambda^T}{4\gamma} + \phi \\
 &= (\lambda_C^T - \lambda^T) \frac{\sigma_C^* + \sigma^*}{2} + \phi. \tag{12}
 \end{aligned}$$

While the second component clearly comprises the transaction costs, the first component is a second type of costs in that it describes the decrease in return for bearing unsystematic, unrewarded risk. Hence, we interpret this type of opportunity costs as the risk costs associated with the portfolio.

These two types of costs depend negatively on each other: Higher transaction costs correspond to a larger number of securities in the portfolio, resulting in a (possibly) higher Sharpe ratio of the tangential portfolio and lower risk costs. On the other hand, low transaction costs allow only for a small number of securities and, hence, a larger deviation of the tangential portfolio from the classically optimal one in terms of the Sharpe ratio.

2.2 Portfolio optimization

As the previous section has made clear, portfolio optimization in the presence of transaction costs breaks down to minimizing the sum of transaction costs and risk costs. Once a subset of securities is fixed, the corresponding tangential portfolio and the associated costs are easily calculated. The crucial task is to determine the subset of portfolios which minimizes these costs.

A combinatorial search must be performed in order to find the optimal subset. Each single security can either be in the portfolio or not, so there are 2^n possible combinations. If n is not very small, a combinatorial search is not realistic.⁹ The problem is NP-hard, so there is no algorithm which solves the problem in polynomial time.¹⁰ However, often, algorithms or heuristics do not find the optimum but find an approximate solution that can be very close to the optimum.¹¹ In this paper we adapt

⁹ The numerical effort could be reduced by techniques like branch-and-bound, see [Lobo et al. \(2007\)](#), p 353.

¹⁰ See [Bienstock \(1996\)](#).

¹¹ The good quality of solutions for cardinality constrained portfolios computed by heuristic methods has been extensively demonstrated, see, e.g., [Maringer \(2005\)](#), [Gilli et al. \(2006\)](#). Besides the deterministic algorithm which is described in the following, probabilistic algorithms like genetic programming promise some success, see [Chang et al. \(2000\)](#).

a heuristic proposed by Lobo et al. (2007) for a similar portfolio selection problem. Their basic idea is to solve a sequence of convex problems related to the original non-convex problem. The solutions of this problems converge to a solution which is near the optimum of the original problem.

Transferred to our setting, the algorithm reduces to solving classical portfolio selection problems with adjusted expected returns, where the adjustments in returns reflect the transaction costs. Let μ_i^k denote the expected return for the i -th security within the k -th iteration step, and let $w^k = (w_i^k)_{i=1, \dots, n}$ denote the vector of weights of the tangential portfolio for the classical portfolio selection problem with expected returns μ_i^k in the absence of transaction costs. The weights correspond to the respective investment volumes via $w_i^k = x_i^k/x$. The algorithm is then defined as:

```

define  $\mu_i^0 = \mu_i - b$  and calculate  $w^0$ 
for  $k = 0, 1, 2, \dots$  do
    let  $\mu_i^{k+1} = \mu_i - \max \left\{ b, \frac{a}{(w_i^k + \delta)x} \right\}$ 
    calculate  $w^{k+1}$ 
    if  $\|w^{k+1} - w^k\| < \epsilon$  : STOP
end for
 $w^{k+1}$  is the solution
  
```

Here, δ and ϵ are small, non-negative constants, and $\|\cdot\|$ is a suitable vector norm, e.g., the Euclidean norm. Lobo et al. (2007) show that the stopping criterion is always met, so there is no risk of an endless loop.

Within each step, the transaction costs are linearized to

$$c(x_i^{k+1}) = \max \left\{ b, \frac{a}{(w_i^k + \delta)x} \right\} x_i^{k+1} = \max \left\{ bx_i^{k+1}, \frac{ax_i^{k+1}}{(w_i^k + \delta)x} \right\}. \quad (13)$$

The rationale behind this approach is that, if the procedure has converged, $w_i^{k+1} = x_i^{k+1}/x$ will approximately equal w_i^k , so that, if $w_i^{k+1} \gg \delta$,

$$c(x_i^{k+1}) \approx \max\{a, bx_i^{k+1}\} \quad (14)$$

as desired. The performance of the algorithm is fairly good; usually it converges after 10–20 iterations. The error with respect to the true solution in terms of the preference function is lower than 0.01%.¹²

¹² As will become clear in the empirical part, in many cases buying a very small number of assets is optimal. With a given bound on the number of assets, a complete combinatorial search to find the exact solution becomes achievable again. Calculating the exact solution in every case where the optimal number of assets is bounded by 4 (out of 50) yields the quoted figure.

3 Empirical investigation

3.1 Data

This section applies the theoretical results empirically to a representative optimization problem. In particular, we consider an investor whose universe consists of the 50 blue chips of the EUROSTOXX 50 and analyze how the number of stocks in the optimal portfolio depends on the investment volume. Given the proportion of transaction costs and risk costs and their absolute values, we then compare a direct investment strategy with the aforementioned alternatives, i.e., exchange-traded funds and index certificates.

The parameters are estimated from historical market prices. Therefore, we use a 2-year history (from 2 January, 2006 to 28 December, 2007) of closing quotes for the 50 stocks at the Frankfurt stock exchange, electronic trading (XETRA).¹³ We estimate the variance-covariance-matrix directly using a standard approach. The vector of expected returns is estimated implicitly by applying the capital asset pricing model.¹⁴ Therefore, we use the 2-year history to estimate the beta factors β_i for each stock with respect to the index and calculate the implied expected return via

$$\mu_i = r + \beta_i(\mu_M - r). \quad (15)$$

The market premium $\mu_M - r$ is assumed to be equal to 6.5%, which represents an average estimate of leading academic financial economists (see Welch 2000). The risk-free rate is $r = 4.07\%$ as of 28/12/2007.

Table 1 provides a brief overview of the fees of six major banks and discount brokers in Germany that apply to internet transactions. Fees for personal transactions are usually higher, but it is reasonable to assume in our analysis that the investor seeks to minimize costs and chooses the cheapest order channel.

Table 1 Transaction cost structures for six banks and brokers in Germany as of December 2007

	Minimum fee a (Euro)	Linear component b (%)
Citibank	9.99	0.00
Comdirect bank	9.90	0.25
Commerzbank	19.92	0.80
Cortal Consors	9.95	0.25
Deutsche Bank	20.00	1.00
Maxblue	7.90	0.25

These figures are valid for standard internet orders. Depending on peculiarities such as active-trader discounts, custodian account type, student discounts, etc., the fees may vary

¹³ We used XETRA as a large and liquid trading place, instead of using the respective home exchanges of the stocks in order to avoid problems with asynchronous trading hours.

¹⁴ Jorion (1991) reports evidence that this kind of estimation leads to comparably good results. The same approach is followed by Maringer (2005).

It is evident that there are significant differences among the banks and brokers selected. In addition to the minimum fee and a linear component, there are several other variations—particularly maximum fees, degressive shapes, and discrete jumps—for these and other offerers. However, for our purposes, the discrete minimum fee is of main interest, so the assumed function type (5) is a good model for real transaction costs. For our empirical study, we apply a minimum fee of 10 Euro and a linear component of 0.25%, which are representative values according to Table 1.

3.2 Characteristics of optimal portfolios

Figure 2 shows the risk-return profile of the 50 stocks and the corresponding efficient frontier. It becomes evident that the straight line of efficient portfolios is close to the curved boundary without the riskless asset over a large range. This observation recalls the assumption made to calculate the expected returns: Since we assume the capital asset pricing model holds, all single stocks lie perfectly along the security market line. As, furthermore, the stocks are fairly positively correlated, a major part of stock variances is explained by the market. Hence, their positions in the μ - σ -diagram are densest within an area parallel to the efficient line (which is identical to the capital market line in this case), which explains the shape of the curved boundary. The tangential portfolio has an expected return of 10.3% and a standard deviation of 14.5%.

Based on this data set, we calculate optimal portfolios with respect to transaction costs and total investment volume. Optimal tangential portfolios and, thus, optimal subsets of assets are independent of individual preferences. These preferences however influence the absolute size of transaction and risk costs because they determine the amount invested in the riskless asset and, hence, the position of the optimal portfolio on the efficient line. The following analysis is based upon an average investor whose preference function suggests that the tangential portfolio is optimal in the absence

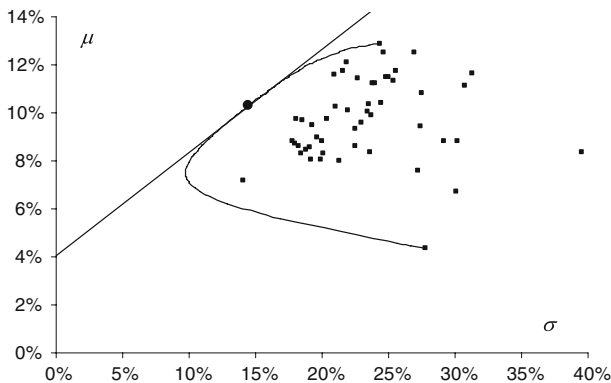


Fig. 2 μ - σ -diagram for the EUROSTOXX 50: the small squares represent the single stocks. Due to the CAPM assumption the curved efficient frontier is close to the efficient line. The parameters of the tangential portfolio (*circle*) are $\mu = 10.3\%$ and $\sigma = 14.5\%$

Table 2 Properties of optimal portfolios for the EUROSTOXX 50 universe: investment volume x , optimal number of assets n^* , average transaction size \bar{x}_i , relative transaction costs ϕ , relative risk costs ψ , transaction costs relative to total costs $\phi/(\phi + \psi)$

x	n^*	\bar{x}_i	ϕ (%)	ψ (%)	$\frac{\phi}{\phi + \psi}$ (%)
1,000	1	1,000	0.66	0.89	43
2,000	1	2,000	0.33	0.89	27
5,000	2	2,500	0.28	0.63	31
10,000	3	3,300	0.23	0.44	35
20,000	6	3,300	0.26	0.20	56
50,000	10	5,000	0.22	0.09	71
100,000	14	7,100	0.22	0.06	79
200,000	24	8,300	0.23	0.02	93
500,000	30	16,700	0.23	0.01	96

of transaction costs,¹⁵ so we calculate portfolios with maximum return, given the standard deviation $\sigma_C^* = 14.5\%$. The results are summarized in Table 2.

It becomes evident that, for very small total investment volumes (up to 2,000 Euro), the optimal number of assets is as small as one. If only one single stock has to be chosen, it is optimal to invest in the stock with the largest ratio $(\mu_i - r)/\sigma_i$,¹⁶ since this stock is the best substitute for the market portfolio. In doing so, the risk costs remain below one percent. With three stocks, the risk costs decrease to 0.44%, which is optimal for an investment volume of 5,000 Euro. The optimal average transaction size increases relatively slowly with the investment amount.¹⁷

With the exception of the smallest volumes, relative transaction costs are fairly independent of the investment volume at values between 0.2 and 0.3% (they can fall below the assumed linear transaction cost component of 0.25% because the optimal portfolio may involve investing in the riskless asset that is considered to be free of transaction costs). Naturally, the risk costs decrease with the number of assets, so for larger investment volumes (above 20,000 Euro), the transaction costs become the major part of the total costs.

3.3 Direct investment versus tailor-made products

This section compares a direct investment against popular alternative investment vehicles offered by the finance industry. Based on the assumption that the investor’s information imply to replicate the index, the cost structures of (classical) mutual index funds, exchange-traded index funds, and index certificates can directly be compared with the sum of the risk costs and the transaction costs of a direct investment. As the risk costs can be bounded by 0.7%, even with only one single stock in our example, and transaction costs decrease sharply with the investment volume, this figure represents the upper limit for the annual fee of a fund if it is to have any chance of being advantageous.

¹⁵ This is equivalent to an absolute risk aversion of $2\gamma = (\mu_M - r)/\sigma_M^2 = 2.98$.

¹⁶ Provided the absence of estimation errors, this is the stock with the highest Sharpe ratio.

¹⁷ The observation that only very few securities can be optimal is in line with results of the index model of Brennan (1975).

Many (even passively managed) mutual funds exceed this threshold. Therefore, we concentrate on exchange-traded funds as alternative vehicles, since these products exhibit more competitive cost structures. As of December 2007, nine exchange-traded funds at the market replicate the EUROSTOXX 50 index. Their annual fees (total expense ratios) range from 0.15 to 0.60%, with the median at 0.30%. The total costs associated with an investment in an exchange-traded fund are the sum of these fees and the transaction costs for a single purchase transaction.

Figure 3 compares the total costs of a direct investment with exchange-traded funds, depending on the investment volume, on a log scale. Not surprisingly, for small investment volumes, exchange-traded funds are superior because of the non-existent risk costs, whereas for larger investment volumes the fund expenses dominate the risk cost, which makes a direct investment favorable. The cut-off point is at 16,000 Euro for the median and at 32,000 Euro for the most cost-attractive fund.

For index certificates, the second alternative, the default risk of the issuer has to be considered. To construct a comparable default-free portfolio, an index certificate can be combined with a credit default swap of the issuer, which generates a compensation payment in the event of the issuer's default. Hence, the credit spreads of the issuers can be applied to quantify the negative value of credit risk as a type of costs associated with index certificates. As of year-end 2007, credit default swaps were liquidly traded for 17 issuers of EUROSTOXX 50 index certificates. The spreads range from 0.10 to 1.43% with the median at 0.28%, so the costs arising from index certificates are comparable to those of exchange-traded funds. However, they are much more unstable—in early 2007, the median was below 0.1%, but in light of the financial market crisis, spreads rose to a maximum in early 2008 with a median above 0.5%.

Thus, for exchange-traded funds and index certificates, the annual costs are crucial for the comparison to a direct investment. Whereas fund fees are rather stable, the costs of default risk exhibit larger variations over the course of time, so there are times where funds are more attractive with respect to index certificates and vice versa. As the major criterion for the choice of a direct investment, Table 3 lists the cut-off points for

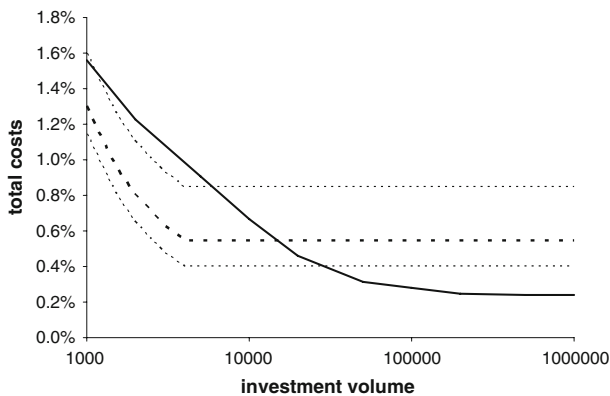


Fig. 3 Total costs for a direct investment (*solid line*) and exchange-traded funds (*dashed lines*), depending on the investment volume (*log scale*), for the EUROSTOXX 50. The three *dashed lines* represent the maximum, minimum, and median of available exchange-traded funds as of December 2007

Table 3 Cut-off points for a direct investment for different annual cost ratios of exchange-traded funds (annual fees) or index certificates (issuer credit spread)

	Cost ratio (%)	Cut-off point
	0.10	42,000
	0.15	32,000
	0.20	22,000
	0.25	18,000
	0.30	16,000
	0.35	13,000
	0.40	11,000
	0.45	9,000
Above the cut-off point, a direct investment is favorable. Data are based on the EUROSTOXX 50 universe	0.50	8,000
	0.55	7,000
	0.60	6,000

several values of annual costs of funds or certificates above which a direct investment is favorable.

3.4 Stability for other investment universes

The results presented so far are based on the EUROSTOXX 50 as one representative example. In this section, we evaluate whether other investment universes lead to substantially different results, using the US market and the stocks of the Dow Jones industrial average 30 index. Table 4 summarizes the optimal portfolios for a direct investment in this market, showing that the results are quite similar to the EUROSTOXX portfolios. For very low investment volumes—only one stock in the optimal portfolio—risk costs are higher, indicating that there is no single stock which is a similarly good approximation of the index, as in the case for the EUROSTOXX. Accordingly, the number of stocks in the optimal portfolio rises slightly faster with the investment volume, as more stocks are needed to reduce unsystematic risk. For larger volumes, risk costs fall below their counterparts for the EUROSTOXX and reach values of only 0.002%.

Figure 4 displays the total costs for both direct investments and exchange-traded funds, depending on the investment volume. As of December 2007, there are only three exchange-traded funds which replicate the Dow Jones industrial average 30 index: one listed at the American Stock Exchange with a total expense ratio of 0.18%, one at Euronext Paris (0.50%), and one at XETRA (0.50%). For the low-cost fund, an investment is favorable up to a volume of 34,000 Euro, which is very similar to the EUROSTOXX case.

Thus, the results are consistent over these two investment universes. The findings concerning the optimality of only few securities can be drawn back to the shape of the covariance matrix, which exhibits moderate to high levels of correlations among the single stocks and leads to limited potential for diversification; in other words, a major part of the total diversification is already achieved by a small number of stocks. This observation relaxes if the covariance matrix is more diagonal-dominant, in which case,

Table 4 Properties of optimal portfolios for the Dow Jones industrial average 30 index

x	n^*	\bar{x}_i	ϕ (%)	ψ (%)	$\frac{\phi}{\phi+\psi}$ (%)
1,000	1	1,000	0.66	1.45	31
2,000	1	2,000	0.33	1.45	19
5,000	3	1,700	0.50	0.62	45
10,000	4	2,500	0.35	0.40	47
20,000	7	2,900	0.33	0.16	68
50,000	9	5,600	0.24	0.07	77
100,000	13	7,700	0.25	0.02	93
200,000	17	11,800	0.25	0.002	99
500,000	18	27,800	0.25	0.002	99

See Table 2 for a legend

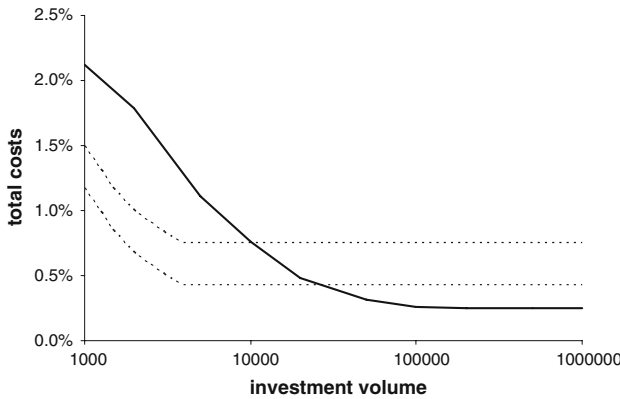


Fig. 4 Total costs for a direct investment (*solid line*) and exchange-traded funds (*dashed lines*), depending on the investment volume (*log scale*), for the Dow Jones industrial average 30. The two *dashed lines* represent the maximum and minimum of available exchange-traded funds as of December 2007

because of lower correlations, a higher level of diversification would be achievable and a larger number of securities would be optimal in the presence of transaction costs. However, the covariance matrix estimated for the EUROSTOXX 50 and the Dow Jones industrial average 30 can be seen as typical, so the results presented here will be similar also for other markets.

4 Conclusion

Transaction costs force small investors to concentrate their money on a comparatively small selection of securities. We analyzed the resulting optimization problem by reformulating it as the minimization of the sum of two types of costs: the transaction costs and the risk costs. Risk costs arise when bearing unsystematic risk, which is unavoidable if the portfolio differs from the optimal portfolio in classical Markowitz

theory. Because of transaction costs, however, the small investor seeks to reduce the number of assets to a level below that of the classically optimal portfolio.

The optimization problem is NP-hard. We propose an efficient heuristic, which proves to be powerful in terms of computational speed and accuracy of the solution. However the algorithm requires input data, particularly the covariance matrix and the vector of expected returns, which may be difficult for the small investor to estimate. Providing this estimation, together with the algorithm, could be a useful service by financial institutions or information providers. Such a service is possible because of the guarantee of convergence and the stability of the results.

The empirical investigation shows that transaction costs can lead to a very small optimum for the number of stocks in the portfolio. Our analysis also provides guidelines for choosing between direct investments in stocks or in alternative vehicles. Direct investment is advantageous if the sum of associated risk and transaction costs is smaller than the annual costs of the alternative investment products. Furthermore, an alternative is reasonable only if it is based on the same information. For instance, if the investor has private information believed to make it possible to outperform the market, a passive index investment is not appropriate.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Baule R, Entrop O, Wilkens M (2008) Credit risk and bank margins in structured financial products: evidence from the German secondary market for discount certificates. *J Futures Mark* 28:376–397
- Bertsimas D, Darnell C, Soucy R (1999) Portfolio construction through mixed-integer programming at Grantham, Mayo, Van Otterloo and Company. *Interfaces* 29:49–66
- Bienstock D (1996) Computational study of a family of mixed-integer quadratic programming problems. *Math Program* 74:121–140
- Blog B, van der Hoek G, Rinnooy Kan AHG, Timmer GT (1983) The optimal selection of small portfolios. *Manage Sci* 29:792–798
- Brennan MJ (1975) The optimal number of securities in a risky asset portfolio when there are fixed costs of transacting: theory and some empirical results. *J Financ Quant Anal* 10:483–496
- Chang TJ, Meade N, Beasley JE, Sharaiha YM (2000) Heuristics for cardinality constrained portfolio optimisation. *Comput Oper Res* 27:1271–1302
- Gilli M, Kellezi E (2002) Portfolio optimization with VaR and expected shortfall. In: Kontoghiorghes EJ, Rustem B, Siokos S (eds) *Computational methods in decision-making, economics and finance*. Kluwer, Dordrecht, pp 167–183
- Gilli M, Kellezi E, Hysi H (2006) A data-driven optimization heuristic for downside risk minimization. *J Risk* 8(4):1–18
- Haugen RA (2001) *Modern investment theory*, 5th edn. Prentice Hall, Upper Saddle River
- Jacob NL (1974) A limited-diversification portfolio selection model for the small investor. *J Finance* 29:847–856
- Jobst NJ, Horniman MD, Lucas C, Mitra G (2001) Computational aspects of alternative portfolio selection models in the presence of discrete asset choice constraints. *Quant Finance* 1:489–501
- Jorion P (1991) Bayesian and CAPM estimators of the means: implications for portfolio selection. *J Bank Finance* 15:717–727
- Levy H (1978) Equilibrium in an imperfect market: a constraint on the number of securities in the portfolio. *Am Econ Rev* 68:643–658
- Lobo MS, Fazel M, Boyd S (2007) Portfolio optimisation with linear and fixed transaction costs. *Ann Oper Res* 152:341–365

- Mansini R, Speranza MG (1999) Heuristic algorithms for the portfolio selection problem with minimum transaction lots. *Eur J Oper Res* 114:219–233
- Mao JCT (1970) Essentials of portfolio diversification strategy. *J Finance* 25:1109–1121
- Maringer D, Kellerer H (2003) Optimization of cardinality constrained portfolios with a hybrid local search algorithm. *OR Spectr* 25:481–495
- Maringer D (2005) Portfolio management with heuristic optimization. In: *Advances in computational management science*, vol 8. Springer, Berlin
- Markowitz HM (1952) Portfolio selection. *J Finance* 7:77–91
- Patel NR, Subrahmanyam MG (1982) A simple algorithm for optimal portfolio selection with fixed transaction costs. *Manage Sci* 28:303–314
- Sankaran JK, Patil AA (1999) On the optimal selection of portfolios under limited diversification. *J Bank Finance* 23:1655–1666
- Sargent TJ (1987) *Macroeconomic Theory*, 2nd edn. Academic Press, San Diego
- Sharpe WF (1963) A simplified model for portfolio analysis. *Manage Sci* 9:277–293
- Stoimenov PA, Wilkens S (2005) Are structured products ‘fairly’ priced? An analysis of the German market for equity-linked instruments. *J Bank Finance* 29:2971–2993
- Tobin J (1958) Liquidity preference as behavior towards risk. *Rev Econ Studies* 25:65–86
- Welch I (2000) Views of financial economists on the equity premium and on professional controversies. *J Bus* 73:501–537