Centre for Computational Finance and Economic Agents

WP005-06

Working Paper Series

Dietmar Maringer

Small is beautiful. Diversification with a limited number of assets

March 2006

University of Essex

CCFEA

www.essex.ac.uk/ccfea

Small Is Beautiful.
Diversification With a Limited Number of Assets

Dietmar Maringer*

March 6, 2006

Abstract

Empirical studies show that investors prefer to hold only a very small number of different assets. This implies that their portfolios are inferior as including more assets would reduce their (unsystematic) risk and would their increase their return / risk premium. At the same time, however, it is common knowledge in the finance industry and in the literature that "small portfolio" with just a few different assets can be very well diversified – provided the "optimal" assets have been chosen. Finding such an optimized portfolio with a constraint on the number of different assets in it, i.e., a cardinality constraint, is a challenge that can hardly be approached with traditional optimization techniques. Hence, proofs for such "small portfolios" had often to resort to simplifying assumptions or anecdotal evidence.

A new approach in numerical optimization are heuristic methods. These methods differ from their traditional counterparts in several important aspects: (i) They incorporate stochastic elements in their search process and are therefore less prone to get stuck (and eventually report) inferior solutions. For many of these methods, there exist convergence proofs stating that they actually are able to identify

* Centre for Computational Finance and Economic Agents (CCFEA), University of Essex, Wivenhoe Park, Colchester CO4 3SQ, United Kingdom.

the true global optimum. (ii) Most of these heuristic methods are extremely flexible and can therefore be applied to all sort of different problems and constraints that would not be approachable otherwise.

This paper addresses the issue of finding an optimal portfolio structure when there is a limit on the number of different assets that may be included. Using Ant Systems, empirical studies are performed for NYSE, FTSE and DAX data. The results confirm that small portfolios can indeed be very well diversified – provided the asset and weight selection has been done with a suitable method. We also find that the heuristic method used in this paper is such a suitable method and is able to outperform standard approaches.

JEL Classification: G11, C14, C61

1 Introduction

One of the central results of Modern Portfolio Theory is that, in perfect markets with no constraints on short selling and frictionless trading without transactions costs, investors will want to hold as many different assets as possible: Any additional security that is not a linear combination of an already included asset will contribute to the portfolio's diversification of risk and could therefore increase the investor's utility.

In practice, however, this situation is rather impractical, since the amount of transactions costs which has to be paid for many different small stocks, would raise the total cost considerably as has been shown in Maringer (2002a) and Maringer (2005). Moreover, the administration of such portfolios with a large number of different assets may become very tedious. Hence, investors seem to prefer portfolios with a rather small number of different assets (see, e.g., Blume and Friend (1975), Börsch-Supan and Eymann (2000), Guiso, Jappelli, and Terlizzese (1996) or Jansen and van Dijk (2002)).

Another important aspect in portfolio selection is that in practice, most of the risk diversification in a portfolio can be achieved with a rather small, yet well chosen set of assets.[1] Hence, in practice, the crucial question of finding the right weight for an asset is linked to the problem whether or not to include this asset in the first place.

Building on Maringer (2001), Maringer (2002b), Keber and Maringer (2001) and Maringer and Kellerer (2003), this paper is concerned with the portfolio optimization problem under cardinality constraints, i.e., when there is an explicit constraint on the number of different assets in the portfolio. In particular, the case of an investor is considered who wants to optimize her Sharpe Ratio in a modified Tobin framework. The remainder of this paper is organized as follows. In section 2, the optimization problem and the optimization method will be presented. Section 3 summarizes the main results from an empirical study of this issue, section 4 concludes.

2 The Model

2.1 Optimization Problem

From a theoretical point of view, the portfolio selection problem with a cardinality constraint can be regarded as *Knapsack Problem* (*KP*).[2] The KP in its simplest version deals with selecting some of the available goods by maximizing the overall value of the resulting combination (objective function) without exceeding the capacity of the knapsack (constraint(s)). The investor's problem, however, demands two significant modifications of the classical KP:

- In the classical KP, each good has a given value which does not depend on what other goods are or are not in the knapsack. For portfolios, however, the "value" of any included asset depends on the overall structure of the portfolio and the other assets in the knapsack because of the diversification effects.

[1] See, e.g., Elton, Gruber, Brown, and Goetzmann (2003, chapter 4).

[2] See Maringer and Kellerer (2003). For a general discussion, see Kellerer, Pferschy, and Pisinger (2004).

- In the classical KP, the goods have fixed weights, and one has to decide whether to take the good or not ("0/1 KP"). The investor has to jointly decide (i) whether to include an asset or not *and* (ii) what amount of her endowment to invest in this asset.

In this paper we assume that investors can choose among one risk-free asset and up to N risky securities and want to maximize their (standardized) risk premium.

Tobin (1958) showed that any portfolio consisting of one risk-free asset and one ore many risky assets will result in a linear relationship between expected return r and the risk associated with this portfolio, σ,

$$r = r_{\mathrm{s}} + \theta_{\mathcal{P}} \cdot \sigma$$

where r_{s} is the return of the safe asset. $\theta_{\mathcal{P}}$ is the risk premium per unit risk and is often referred to as *(ex ante) Sharpe Ratio*, $SR_{\mathcal{P}}$. Given the standard assumptions on capital markets with many risky and one risk-free asset, a rational risk averse investor will therefore split her endowment and invest a portion of α in the safe asset and $(1-\alpha)$ in some portfolio of risky assets, $\mathcal{P}$, where the structure of $\mathcal{P}$ determines $\theta_{\mathcal{P}}$. The investor will therefore choose the weights, x_i, for assets i within the portfolio of risky assets in order to maximize the investment's risk premium per unit risk, $\theta_{\mathcal{P}}$. In passing note that the investor's level of risk aversion is reflected in her α and that the x_i's ought to be the same for any investor. Thus, the portfolio $\mathcal{P}$ (usually called *tangency portfolio*) can be determined without considering the investor's attitude towards risk and regardless of her utility function.

If there exists a market $\mathcal{M} = \{1, .., N\}$ with N assets k of which shall be included in the portfolio $\mathcal{P}$, the investor's optimization problem can be written as follows:

$$\max_{\mathcal{P}} \; \theta_{\mathcal{P}} = SR_{\mathcal{P}} = \frac{r_{\mathcal{P}} - r_{\mathrm{s}}}{\sigma_{\mathcal{P}}}$$

subject to

$$r_{\mathcal{P}} = \sum_{i=1}^{N} x_i \cdot r_i$$

$$\sigma_{\mathcal{P}} = \sqrt{\sum_{i=1}^{N} \sum_{j=1}^{N} x_i \cdot x_j \cdot \sigma_{ij}}$$

$$\sum_{i=1}^{N} x_i = 1 \quad \text{and} \quad \begin{cases} x_i \geq 0 & \forall i \in \mathcal{P} \\ x_i = 0 & \forall i \notin \mathcal{P} \end{cases}$$

$$\mathcal{P} \subset \mathcal{M}$$

$$|\mathcal{P}| = k$$

where σ_{ij} is the covariance between the returns of assets i and j and r_i is the return of asset i.

Like the (actually simpler) "0/1 KP" this optimization problem is NP–hard. It is usually approached in practice by rules of the thumb (based on certain characteristics of the individual assets[3]) or by reducing the problem space by making *a priori* selections (e.g., by dividing the market into several segments and choosing "the best" asset of each segment[4]). As neither of these methods reliably excludes only "irrelevant" combinations, they tend to result in sub-optimal solutions. An alternative way to solve the problem is the use of meta-heuristics which are not based on *a priori* neglecting the majority of the problem space. The method suggested here has its origin in biology, namely ant systems.

2.2 Ant Systems

Evolution has provided ants with a simple, yet enormously efficient method of finding shortest paths.[5] While traveling, ants lay a pheromone trail which helps themselves and their followers to orientate.

---

[3] One popular rule, which will also be applied in this study, states to prefer assets that have high Sharpe Ratios SR_i.

[4] See, e.g., Farrell, Jr. (1997, chapter 4) on Asset Classes. Lacking appropriate information for our data sets, this approach could not be applied.

[5] See Goss, Aron, Deneubourg, and Pasteels (1989).

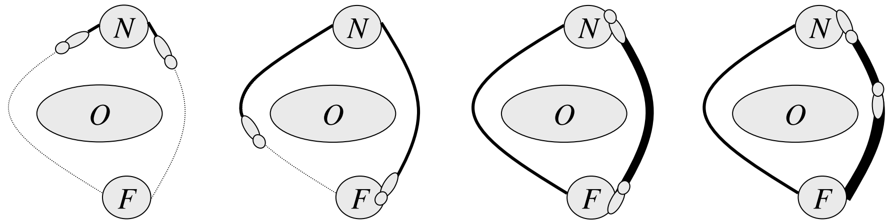


Figure 1: Simple foraging example for a colony with two ants

To illustrate the underlying principle, we assume a nest N and a food source F are separated by an obstacle O (Figure 1) and that there are two alternative routes leaving N both leading to F, yet different in length. Since the colony has no information which of the two routes to choose, the population (here consisting of two ants) is likely to split up and each ant selects a different trail. Since the route on the right is shorter, the ant on it reaches F while the other ant is still on its way. Supplied with food, the ant wants to return to the nest and finds a pheromone trail (namely its own) on one of the two possible ways back and will therefore select this alternative with a higher probability. If it actually chooses this route, it lays a second pheromone trail while returning to the nest. Meanwhile the second ant has reached F and wants to bring the food to the nest. Again, F can be left on two routes: the left one (=long) has now one trail on it, the right one (=short) has already two trails. As the ant prefers routes with more pheromone in it, it is likely to return on the right path – which is the shorter one and will then have a third trail on it (versus one on the left path). The next time the ants leave the nest, they already consider the right route to be more attractive and are likely to select it over the left one. In real live, this self-reinforcing principle is enhanced by two further effects: shorter routes get more pheromone trails as ants can travel on them more often within the same time span then they could on longer routes; and old pheromone trails tend to evaporate making routes without new trails less attractive.

Based on this reinforcing mechanism, the tendency towards the shorter route will increase and due to the new pheromone trails will be preferred. At the same time, there remains a certain probability that routes with less scent will be chosen; this assures that new, yet unexplored alternatives can be considered. If these new alternatives turn out to be shorter (e.g., because to a closer food source), the ant principle will enforce it, and – on the long run – it will become the new most attractive route; if it is longer, the detour is unlikely to have a lasting impression on the colony's behavior.

As discussed in Maringer (2002b, 2005), Dorigo, Maniezzo, and Colorni (1991), Colorni, Dorigo, and Maniezzo (1992a), Colorni, Dorigo, and Maniezzo (1992b) and Dorigo (1992) first introduced this principle to routing problems (such as the Traveling Salesman Problem) by simulating real routes and distances between the cities in artificial nets and implementing an artificial ant colony where the ants repeatedly travel through these nets. Meanwhile, this concept resulted in the closely related meta-heuristics *Ant Systems* (*AS*) and *Ant Colony Optimization* (*ACO*) which have been applied successfully to a wide range of logistic problems and ordering tasks.[6] In particular, the introduction of *elitists* turned out to be very effective. In this concept the best solution found so far is reinforced each iteration in addition to the ants of the colony a certain number of elitist ants are traveling along the best solution found so far and by doing so reinforce this path.[7] In addition Bullnheimer, Hartl, and Strauss (1999) suggest a ranking system where ants with better solution spread more pheromone than the not so good ones and where paths of bad solutions receive no additional scent.

The concept of ant colony optimization and how it can be implemented will be presented in the next section. We will also demonstrate that this approach can be adopted for the Knapsack Problem in general and the investor's problem in particular.

2.3 The Algorithm

2.3.1 Approaching the Knapsack Problem

Applied to the portfolio selection problem, we implement an iterative search strategy where each iteration consists of three stages. In the first stage, artificial ants are to travel within a net consisting of N nodes which represent

[6] For a comprehensive survey on applications as well as the methodical variants, see, e.g., Dorigo and Di Caro (1999) or Bonabeau, Dorigo, and Theraulaz (1999).

[7] See Dorigo, Maniezzo, and Colorni (1996).

the available assets. An arc connecting any two nodes i and j where $i, j \in \mathcal{M}$ and $i \neq j$ shall capture whether a combination of these two is considered favorable or not. This can be achieved by introducing a matrix $\left[\tau_{ij}\right]_{N \times N}$ where τ_{ij} represents the amount of pheromone. The trail information will then be used to calculate the probabilities with which the following ants will select assets.

Let $\mathcal{P}'_a$ be the incomplete portfolio of ant a with $|\mathcal{P}'_a| < k$. If i is some asset already included in this portfolio, i.e., $i \in \mathcal{P}'_a$, whereas j is not, i.e., $j \notin \mathcal{P}'_a$, then the probability of choosing asset j shall be

$$p_{aj} = \begin{cases} \frac{\sum_{i \in \mathcal{P}'_a} \left(\tau_{ij}\right)^{\gamma} \cdot \left(\eta_{ij}\right)^{\beta}}{\sum_{i \in \mathcal{P}'_a} \sum_{h \notin \mathcal{P}'_a} (\tau_{ih})^{\gamma} \cdot (\eta_{ih})^{\beta}} & \forall j \notin \mathcal{P}'_a \\ 0 & \forall j \in \mathcal{P}'_a \end{cases}. \tag{1}$$

The probability p_{aj} is mainly influenced by the amount of pheromone τ_{ij} that is on the paths from nodes $i \in \mathcal{P}'_a$ to node j. γ is a parameter for tuning that influence. In line with other implementations of ant based strategies, a matrix $[\eta_{ij}]_{N \times N}$ is introduced which represents the visibility of j from i. In routing problems, this information (which unlike the pheromone trails remains unchanged during the optimization) provides sort of a map thus providing the ants with guidelines or *a priori* information on preferred combinations. In the asset selection problem, $\eta_{ij} \geq 0$ might be used to indicate whether the investor regards combination i and j as desirable or not by transferring some general rule of the thumb onto $[\eta_{ij}]$. Also, the visibility could be employed to reinforce constraints. E.g., if i and j represent common and preferred stocks, respectively, of the same company and the investor does not want to hold both in her portfolio, she will set $\eta_{ij} = 0$, and the probability, asset j is added to a portfolio $\mathcal{P}'_a$ already containing i will become zero. If, on the other hand, she has a strong preference for this combination, a high value for η_{ij} will increase the probability that both i and j get included in the portfolio.

Results for the Traveling Salesman Problem suggest that in addition to elitists and ranking systems, the ants ought to be provided with some sort of a "road map" which usually is based on some *a priori* heuristics and is captured in the visibility matrix $\left[\eta_{ij}\right]$.[8] In ordering problems such as the Traveling Salesman Problem, the number of updated trails is rather small because the sequence in which the nodes are selected is of central importance. Thus, an ant visiting k nodes will update just $k-1$ arcs and the chance of not updated arcs and evaporation on "good" arcs must not be neglected.

In our problem, however, it is the combination that matters, thus an ant selecting k securities will update $k \cdot (k-1)$ arcs in a symmetric matrix. Having experimented with general rules and incorporated them into $\left[\eta_{ij}\right]$,[9] we found that they have a rather limited effect on the overall result: favorable parameters have been found to have far more influence on the reliability of the results and the speed at which the algorithm converges. We therefore do not introduce heuristics and "save" the visibility matrix for enhanced optimization problems, e.g., with possible individual preferences. In this study we assume that there are no such preferences and that the investor is interested only in maximizing the portfolio's risk premium per unit risk, $\theta_{\mathcal{P}}$. Thus, we set the visibility matrix $\left[\eta_{ij}\right] = \mathbf{1}$ and the parameter for tuning its influence $\beta = 1$. By also setting the parameter $\gamma = 1$, the probability from equation (1) melts down to

$$p_{aj} = \begin{cases} \dfrac{\sum\limits_{i \in \mathcal{P}'_a} \tau_{ij}}{\sum\limits_{i \in \mathcal{P}'_a} \sum\limits_{h \notin \mathcal{P}'_a} \tau_{ih}} & \forall j \notin \mathcal{P}'_a \\ 0 & \forall j \in \mathcal{P}'_a \end{cases} . \tag{1*}$$

Once any ant has chosen their k assets, stage two of the model can be entered and the optimal portfolio weights are determined by some standard solution: When short sales are permitted, any ant could determine the maximum risk premium $\theta_{\mathcal{P}_a}$ that could be achieved with the securities in its

[8] See Bonabeau, Dorigo, and Theraulaz (1999).

[9] E.g., such rules could make use of the general result that diversification will be the higher the lower the correlation between the included assets. Hence, the visibility matrix could be derived from the correlation matrix or the covariance matrix by increasing (decreasing) the visibility between i and j when the correlation or covariance is low (high).

knapsack by the exact closed-form solution [10]

$$\theta_{\mathcal{P}_a} = \frac{r_{\mathcal{P}} - r_{\mathsf{S}}}{\sigma_{\mathcal{P}}}$$

where

$$r_{\mathcal{P}} = \frac{a - r_{\mathsf{S}} \cdot b}{b - r_{\mathsf{S}} \cdot c}$$

$$\sigma^2_{\mathcal{P}} = \frac{a - 2 \cdot b \cdot r_{\mathsf{S}} + c \cdot r^2_{\mathsf{S}}}{(b - c \cdot r_{\mathsf{S}})^2}$$

and

$$\begin{bmatrix} a & b \\ b & c \end{bmatrix} = \begin{bmatrix} \boldsymbol{r}' \\ \boldsymbol{I}' \end{bmatrix} \boldsymbol{\Sigma}^{-1}_{\mathcal{P}} \begin{bmatrix} \boldsymbol{r} & \boldsymbol{I} \end{bmatrix}$$

and $\boldsymbol{\Sigma}^{-1}_{\mathcal{P}}$ the covariance matrix of the assets included in the portfolio. However, since our optimization model disallows negative asset weights, determining the $\theta_{\mathcal{P}_a}$ is regarded as a quadratic programming problem.[11]

In the third stage, when all ants have packed their knapsack and know their $\theta_{\mathcal{P}}$'s, the trail information can be updated which comprises three steps:

- As time goes by pheromone evaporates. Thus, when a period of time has elapsed, only $\rho \cdot \tau_{ij}$ of the original trail is left where $\rho \in [0, 1]$.
- New pheromone trails are laid. In real life, ants tend to be permanently on the run and are permanently leaving new pheromone trails without bothering whether their colleagues have already returned to the nest. In artificial ant systems, however, each ant chooses one path through the net and waits for the other ants to complete their journey before starting the next trip. Thus, artificial ant systems usually assume that any ant a spreads a fixed quantity Q of pheromone on its path which has a length of L_a and by doing so updates the trail by

[10] For a more detailed discussion, see Maringer (2005).

[11] For a more detailed discussion, see Maringer (2005), alternative approaches are discussed, e.g., in Elton, Gruber, Brown, and Goetzmann (2003) and Brandimarte (2002). Maringer and Kellerer (2003) present a heuristic that simultaneously selects asset and optimization the weights.

$\Delta_a \tau_{ij} = {}^{Q}\!/_{L_a}$ for any arc (i, j) along a's path. This implies that the shorter the path the higher the additional trail. Since both in real life and in ant systems L_a is to be minimized whereas here $\theta_{\mathcal{P}_a}$ is to be maximized, we adopt this concept and use ${}^{1}\!/_{\theta_{\mathcal{P}_a}}$ for a substitute of L_a. Thus, the trail update for ant a would be $\Delta_a \tau_{ij} = Q \cdot \theta_{\mathcal{P}_a}$ for all securities $i \neq j$ and $i, j \in \mathcal{P}_a$.

Bullnheimer, Hartl, and Strauss (1999) suggest a ranking system that reinforces the solutions of the best ants of the current population (here called "prodigies"). In our application, this concept allows only the ω best ants to update $[\tau_{ij}]$ where the rank $\mu = 1, .., \omega$ determines the quantity of pheromone Q_μ a prodigy can spread. Assuming a simple linear ranking system where the quantity of pheromone depends on the ant's rank, prodigy μ updates arc (i, j) by

$$\Delta \tau_{ij,\mu} = \begin{cases} Q_\mu \cdot \theta_{\mathcal{P}^\mu} & \forall i, j \in \mathcal{P}^\mu, i \neq j \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where

$$Q_\mu = \begin{cases} ((\omega - \mu) + 1) \cdot Q & \mu \leq \omega \\ 0 & \mu > \omega \end{cases}.$$

- Assuming that, in addition to the ants of the current colony, ε elitist ants are choosing the best portfolio found so far, $\mathcal{P}^*$, and each of them spread Q pheromone, then each elitist updates the matrix by

$$\Delta \tau^*_{ij} = \begin{cases} Q \cdot \theta_{\mathcal{P}^*} & \forall i, j \in \mathcal{P}^*, i \neq j \\ 0 & \text{otherwise} \end{cases}.$$

Combining evaporation and new trails, the pheromone matrix is to be updated according to

$$\tau_{ij} := \rho \cdot \tau_{ij} + \sum_{\mu=1}^{\omega} \Delta \tau_{ij,\mu} + \varepsilon \cdot \Delta \tau^*_{ij} \quad \forall i \neq j. \tag{3}$$

Due to this updates, the next troop of ants can apply their predecessors' experiences: In the next iteration, the probabilities according to (1*) will be

```
initialize pheromone matrix τ_ij = τ^0 ∀i ≠ j and τ_ii = 0;
population size := N;
REPEAT
    FOR a := 1 TO Population size DO
        P_a := {a};
        WHILE |P_a| < k DO
            determine selection probabilities p_aj ∀j ∉ P_a according to
                definition (1*);
            use probabilities p_aj to randomly draw one additional asset j;
            add asset j to the portfolio, P_a := P_a ∪ {j};
        END;
        determine optimal asset weights such that {SR|P_a} → max!;
    END;
    rank ants according to their portfolios' Sharpe Ratios;
    IF max SR_{P_a} > SR_{P*}
        new elitist is found, replace previous elitist P* with new one;
    END;
    update pheromone matrix;
UNTIL convergence criterion met;
REPORT elitist;
```

Listing 1: Pseudo-code for the main Ant System routine

influenced, and the ants' preferences will be shifted towards combinations of securities that have proven successful. Listing 1 summarizes the main steps of the algorithm. The computationally most complex parts of the algorithm are the computation of a portfolio's volatility, $\mathcal{O}(k^2)$, having to sort the population, $\mathcal{O}(A \cdot \ln(A))$ where A is the number of ants in the colony, and the update of the pheromone matrix which is quadratic in k and linear in the number of prodigies plus the elitist, $\mathcal{O}((k^2 - k) \cdot (\omega + 1))$, and quadratic in N due to the evaporation, $\mathcal{O}(N^2)$. The overall complexity of the algorithm is determined by these individual complexities times the number of iterations.

2.3.2 First Applications of the Algorithm

In order to determine the essential parameters for the algorithm we ran a series of tests with random numbers for elitists, ε, prodigies, ω, and factor of evaporation, ρ. We then tried to find correlations between these values and the "effectiveness" (i.e., speed and level of improvement during the iterations) of the respective ant colony. According to these results we chose the colony size to equal the number of available assets, i.e., 100, $\varepsilon = 100$ (i.e., equal to the number of securities and ants per iteration), $\omega = 15$ (i.e., only the best 15 per cent of ants where allowed to update according to equation (2)), and $\rho = 0.5$ (i.e., half of last round's trail information evaporates).

The version of the ant algorithm as just presented was first applied to the portfolio selection problem in Maringer (2001) where the *ex ante* Sharpe Ratio (SR) (which, as argued, is equal to $\theta_{\mathcal{P}}$) is to be maximized for a DAX data set. In Maringer (2002b) it is applied to finding cardinality constrained portfolios under a Markowitz/Black framework, and the computational study therein is based on subsets of a FTSE data set with $N = 25, \ldots, 90$ where the expected return is equal to the market's expected return and the risk is to be minimized, and the algorithm is found to be superior Simulated Annealing and a Monte Carlo approach. In addition, for a number of problems, the heuristically obtained results were compared to those from complete enumeration, and it was found that the Ant algorithm reported the optimal solution in the majority of runs (often all the runs) even when the alternative method Simulated Annealing was unable to identify the optimum even once. In Keber and Maringer (2001), the ant algorithm is compared to Simulated Annealing and Genetic Algorithm based on the SR maximization problem with the same FTSE data set. Again, it was found that the results of population based heuristics are superior to Simulated Annealing, that, however, the population based heuristics also take more time to find appropriate parameter values.

For the sake of simplicity neither of these studies had a non-negativity constraint on the weights which is included in this study. In addition, in all three previous studies, the ant algorithm did exhibit a typical property of this method: Ant algorithms perform best when applied to large problems, whereas it is the rather "small" problems sometimes that cause slightly more problems to the ant algorithm. When k, the number of different asset included in the portfolio, is rather small, the algorithm converges quite fast, and the colony might get stuck in a local optimum which they cannot escape. In Maringer (2002b), e.g., selecting $k = 3$ assets appears more demanding than selecting $k = 6$ assets for all markets with $N \leq 78$. Though even for these cases, the chances of identifying the actual global optimum in an independent run is for the ant algorithms by magnitude higher than the results found by Simulated Annealing, it is desirable to have equally high reliability for small problems.

2.3.3 Refining the Algorithm

In heuristic search strategies, a common stopping criterion is the number of iterations the elitist hasn't changed, i.e., for how many iterations the algorithm has produced no further improvement. The search is then stopped and the current elitist is reported. If there is a chance that the result is only a local one, a new independent run is started and knowledge acquired in the previous run is lost. In Artificial Intelligence, another common way of overcoming a potential local optimum is to introduce a random shock:[12] When an agent hasn't achieved an improvement over a given number of iterations, it is randomly "positioned" at a different location will continue the search from there, yet without necessarily starting a new, perfectly independent run. Based on this idea, we suggest a similar concept to the ant algorithm.

In ant algorithms (as well as in real life), the ants tend to get stuck in a local optimum when the pheromone trails for a good, yet not globally optimal solution are so strong that chances of finding a route aside these tracks are very low – and are even lowered in due course as the (probably) suboptimal routes are reinforced. Enforcing alternative routes therefore demands lowering the pheromone level on these (probably) suboptimal tracks. We suggest a simple means that might do exactly this trick: With a certain probability the initial pheromone matrix (or a weighted combination of the current and initial matrices) is reinstated yet the current elitist is kept. This implies that knowledge and experience acquired in previous runs is kept while the ants have a higher chance of selecting alternative routes. Metaphorically speaking, this corresponds to "rain" where current pheromone trails are washed away or at least blurred. We therefore introduce a reset parameter $\nu \in [0, 1]$ where $\nu = 1$ corresponds to "heavy rain" where all pheromone trails are swept away and the original pheromone matrix is restored; the closer ν is to

[12] See, e.g., Russell and Norvig (2003).

zero, the more of the current trail information endures. This variant has a similar effect as the local updates of the pheromone trail in *Ant Colony Optimization* (*ACO*),[13] a variant of ant algorithms which is to foster diversity within the colony's solutions.

The updating rule (3) for the off-diagonal elements of the pheromone matrix, $[\tau_{ij}]$ with $i \neq j$, can then be enhanced with an option where a part of the old and newly added trails are washed away and the initials trails are restored:

$$\tau_{ij} := \begin{cases} \left(\rho \cdot \tau_{ij} + \sum_{\mu} \Delta\tau_{ij,\mu} + \varepsilon \cdot \Delta\tau^{*}_{ij}\right) \cdot (1-\nu) + \tau^{0} \cdot \nu & \text{“rain”} \\ \rho \cdot \tau_{ij} + \sum_{\mu=1}^{\omega} \Delta\tau_{ij,\mu} + \varepsilon \cdot \Delta\tau^{*}_{ij} & \text{“sunshine”} \end{cases} \tag{3*}$$

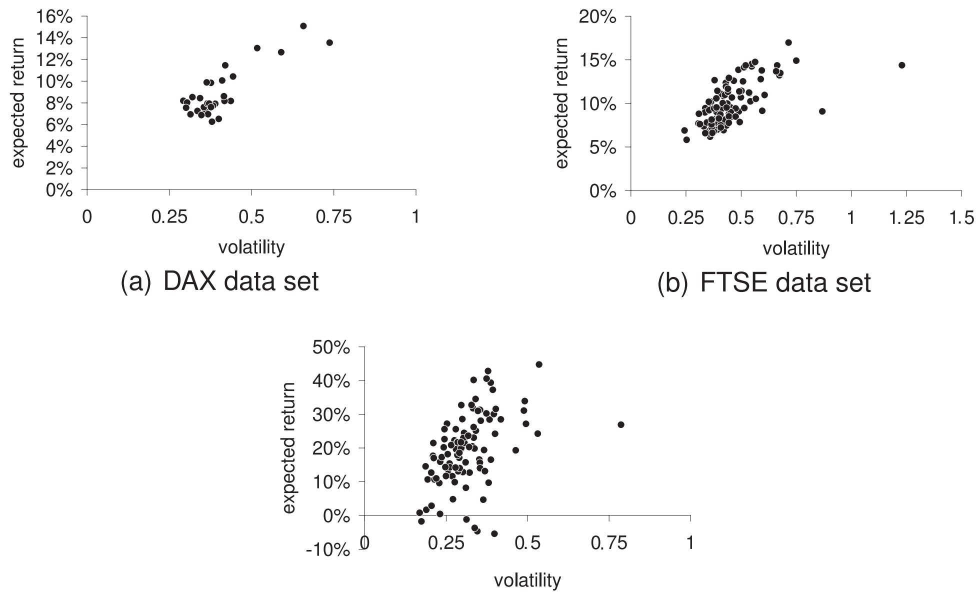
where the option “rain” is chosen with a probability of p_{rain} and the alternatively chosen option “sunshine” corresponds to the original updating rule (3). Whether this concept is advantageous or not, was tested in a computational study; the main results will be discussed in section 3.2.

3 Empirical Study

3.1 Data

The empirical study in this paper is based on data sets for the DAX, the FTSE, and the S&P 100. The *DAX data set* contains the 30 stocks represented in the German stock Index DAX30. The *FTSE data set* is based on the 100 stocks contained in the London FTSE100 stock index; four of the stocks, however, had to be excluded due to missing data. In both cases we used daily quotes over the period July 1998 – December 2000. Based on the corresponding historic returns we calculated the covariances σ_{ij} which are used for estimators of future risk. The expected returns, r_i, were generated with a standard *Capital Asset Pricing Model* (*CAPM*) approach according to $r_i = r_s + (r_{\mathcal{M}} - r_s) \cdot \beta_i$ with an expected safe return of r_s = 5%, expected

[13] See Bonabeau, Dorigo, and Theraulaz (1999, p. 49) and the literature quoted therein.



(a) DAX data set

(b) FTSE data set

(c) S&P 100 data set

Figure 2: Estimated return and risk for the data sets

market risk premia of $r_{\text{DAX}} - r_{\text{S}} = 5.5\%$ and $r_{\text{FTSE}} - r_{\text{S}} = 6\%$, respectively,[14] and with beta coefficients, β_i, coming from the historic returns. The distributions of the assets in the return-volatility space are depicted in Figures 2(a) and 2(b), respectively. In the light of recent developments in the capital markets, we want to point out that we focus exclusively on the selection problem and that in this optimization problem the mean and variance of returns are regarded as exogenously determined.

[14] The values for the safe interest rate and the markets' risk premia were chosen to represent what then would have made reasonable guesses. With the focus on the optimization where the estimates for risk and return can be considered exogenously determined, the actual values proofed to have little influence on the conclusions drawn in the computational study.

Estimating the assets' returns via the CAPM implies that their (estimated) Sharpe Ratio differ in their correlation with the market:

$$
\begin{aligned}
SR_i &= \frac{\overbrace{(r_{\mathrm{s}} + (r_{\mathcal{M}} - r_{\mathrm{s}}) \cdot \beta_i)}^{=r_i} - r_{\mathrm{s}}}{\sigma_i} \\
&= \frac{(r_{\mathcal{M}} - r_{\mathrm{s}}) \cdot \frac{\sigma_i \cdot \rho_{i\mathcal{M}}}{\sigma_{\mathcal{M}}}}{\sigma_i} \\
&= \frac{r_{\mathcal{M}} - r_{\mathrm{s}}}{\sigma_{\mathcal{M}}} \cdot \rho_{i\mathcal{M}} \\
&= SR_{\mathcal{M}} \cdot \rho_{i\mathcal{M}}
\end{aligned}
$$

where M is the respective market index. Though this does not affect the covariances of any two assets and therefore does not have an immediate effect on a portfolio's volatility and Sharpe Ratio, the values for the third data set, the *S&P 100 data set*, are estimated differently. Based on daily returns for the stocks in the S&P 100 stock index from November 1995 through November 2000 and for 23 country, 42 regional and 38 sector indices, the expected returns for the stocks where estimated from the first 1 000 days with an combined APT[15] and GARCH[16] approach: First, for any asset the bundle of five indices was determined that explains most of the asset's return in sample. Next, the expected returns and volatility for the indices where estimated with a GARCH model and the assets' expected out of sample returns based on the individual APT models. The assets' volatilities where estimated with a GARCH model, the covariances where determined with the assets' volatilities and their historic correlation coefficients. The factor selection process is presented in Maringer (2004) which also offers a more detailed presentation of the underlying method. The results from this estimation procedure appear quite reliable: Only for eight of the 100 assets, the actual out of sample returns differ statistically significant from their expected values.[17] The

[15] See Ross (1976). A detailed presentation of the data set and the factor selection problem can be found in Maringer (2004).

[16] See Engle (1982), Bollerslev (1986) and Engle (1995).

[17] Significance test at the usual 5% level of significance; corresponding tests for the DAX and FTSE data sets had to be omitted in lack of out of sample data. The following presentation will therefore focus strictly on the selection problem given a certain market situation which can be considered realistic.

volatilities and expected returns for this data set are depicted in Figure 2(c). Though comparable to the DAX data set in the range of the assets' volatilities and the FTSE data set in the number of assets, the S&P data set differs from the others in the range of the expected returns since it also contains assets with negative expected returns.[18]

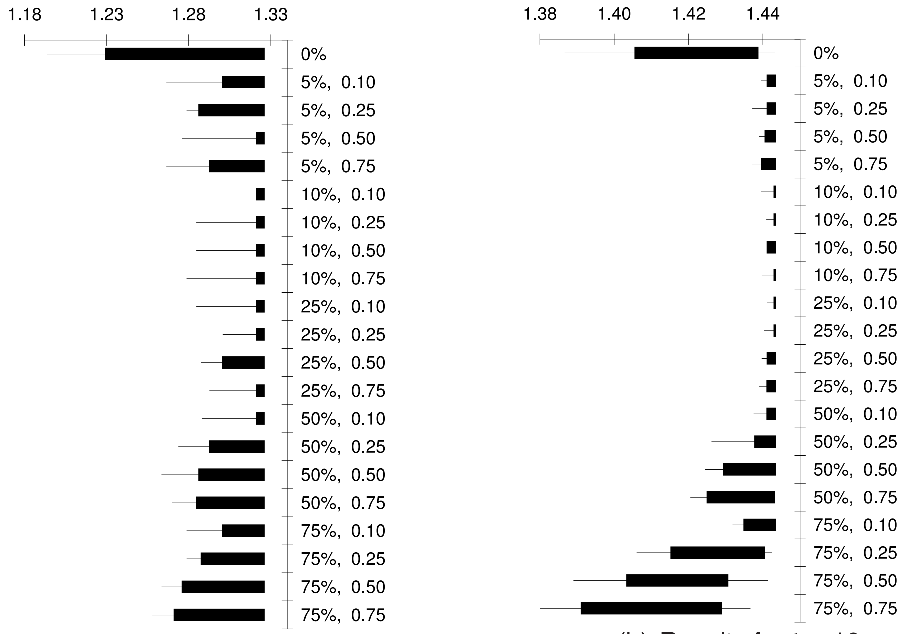
3.2 Computational Study for the Modified Update Rule

In order to test whether the concept of casually resetting the pheromone matrix has a favorable effect on the algorithm's performance or not, we ran a series of Monte Carlo experiments where in the initialization step the value of the reset parameter ν was randomly chosen from $(0.1, 0.25, 0.50, 0.75)$ and the probability for "rain" was chosen to be $p_{rain} \in \{5\%, 10\%, 25\%, 50\%, 75\%\}$, i.e., with a probability of p_{rain} of the update steps, the modified update rule (labeled "rain" in (3*) was applied and with a probability of $(1 - p_{rain})$, the original update rule (3) (labeled "sunshine" in (3*)) was applied. For each combination of parameters and different value of assets in the portfolio, k, approximately 120 independent runs were performed. For comparison, we also ran the algorithm in its previous version with update rule (3) without the "rain" modification by simply setting $p_{rain} = 0\%$; here 1 000 independent runs were performed. For any parameter setting, the number of iterations per run was limited to 200, the colony size equaled the number of included assets, k.

Based on the S&P 100 data set, the two cases $k = 3$ and $k = 10$ are considered. These two problems differ considerable in the number of candidate solutions: the former comes with just 161 700 alternatives,[19] the latter with 1.73×10^{13} alternatives.

[18] When assets are negatively correlated with the index, both the APT and the CAPM predict negative risk premia in equilibrium which, when exceeding the safe interest rate, can also result in negative expected returns. The economic argument behind is that an investor is willing to pay an ("insurance") premium for assets that react opposite to the market trend and are therefore well suited for diversification and hedging.

[19] From a practical point of view, for $k = 3$, the number of alternatives would be small enough for an exhaustive search. Nonetheless, a heuristic optimization method ought to work well with small problems; hence this case is considered here, too.



(a) Results for $k = 3$ (b) Results for $k = 10$

Figure 3: Range from best to worst reported result (lines) and range covering 90% of reported solutions (black area) with the traditional update rule (3) (with $p_{rain} = 0\%$) and with rain according to rule (3*) for different values of p_{rain} (in %) and ν (as decimals)

Figure 3 depicts the range for the reported solutions depending on the different values for ν and p_{rain}. As can be seen for either value of k, the version without rain reports quite diverse solutions. Though the global optimum is found eventually, a high number of runs is necessary to reduce the likelihood that just a local optimum is reported: for $k = 3$, in just 16% of all runs, the global optimum was found, and for $k = 10$, in just 2 of the 1 000 independent runs the global optimum was reported.

With appropriate values for p_{rain} and ν, on the other hand, the algorithm performs significantly better: For the case with $k = 3$, in two thirds of the runs the global optimum was identified by any of the tested version with $p_{rain} \in (5\%, 10\%, 25\%)$ (and arbitrary positive value for ν) or with $\nu = 0.10$ (and arbitrary positive values for p_{rain}). For the case where $k = 10$, in two thirds of the runs the global optimum was reported with the parameter combinations $(p, \nu) = (5\%, 0.10)$ and $(10\%, 0.25)$, and for any positive value of ν with $p_{rain} \in (5\%, 10\%, 25\%)$, half of the runs reported the global optimum.

Additional experiments showed that the performance of the traditional version without rain could be improved by increasing the number of iterations, yet never reached the modified version's high ratio of runs in the global optimum was identified.

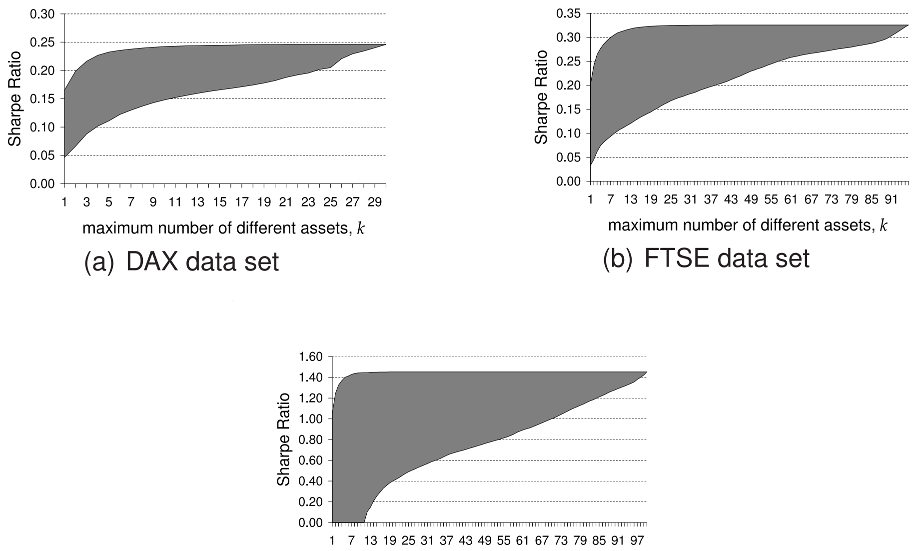
3.3 Financial Results

According to the theory, increasing the number of assets k in the portfolio $\mathcal{P}$ causes an increase in the risk premium $SR_{\mathcal{P}}$ provided the right assets are chosen and assigned the optimal weights. In addition, the marginal contribution of any additional security to the portfolio's diversification is decreasing. Both effects can be found in the results for either of the markets: The graphs in Figure 4 depict the bandwidth within which the Sharpe Ratio will lie when the asset weights are optimized for the best and the worst possible combination of k assets.[20] As can easily be seen, in all three markets, a relatively small yet well chosen set of assets can achieve a higher $SR_{\mathcal{P}}$ than a large yet badly chosen set of assets. E.g., in the FTSE data set, the optimal combination of $k = 5$ assets might outperform a poor combination of $k = 84$ assets: $SR^{\max}_{\mathcal{P}(k=5)} = 0.2863 > SR^{\min}_{\mathcal{P}(k=84)} = 0.2860$. Note that this is all due to good or bad selection of the assets and not due to suboptimal asset weights.[21]

The bandwidth for the Sharpe ratio will be the larger the smaller the portfolio and the larger and more diverse the market is: selecting any $k = 15$ assets from the DAX data set, e.g., the achievable SRwill be in the range from 0.1658 to 0.2447; is the same number of different assets is selected from the FTSE and the very diverse S&P data set, the Sharpe Ratios will range from 0.1294 to 0.3202 and from 0.2545 to 1.4497, respectively.

[20] Finding the "worst" combinations, too, represents an optimization problem where the sign in the objective function is changed, yet not the way the weights $x_i \forall i \in \mathcal{P}$ are determined. This assures that low values for the Sharpe Ratios actually do come from selecting the "wrong" assets and not by an inappropriate loading of the weights.

[21] See Solnik (1973).



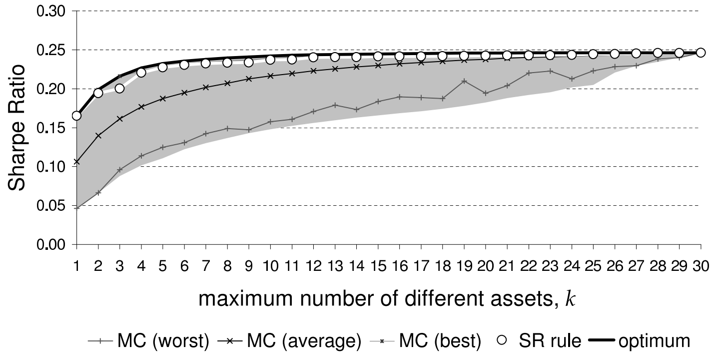
(a) DAX data set

(b) FTSE data set

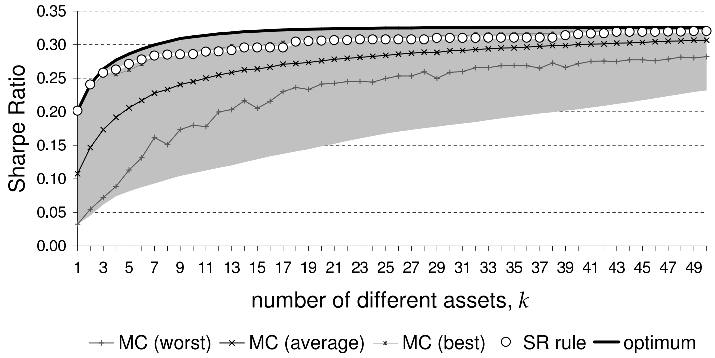
(c) S & P 100 data set

Figure 4: Range for Sharpe Ratios for portfolios under cardinality constraint with optimized weights

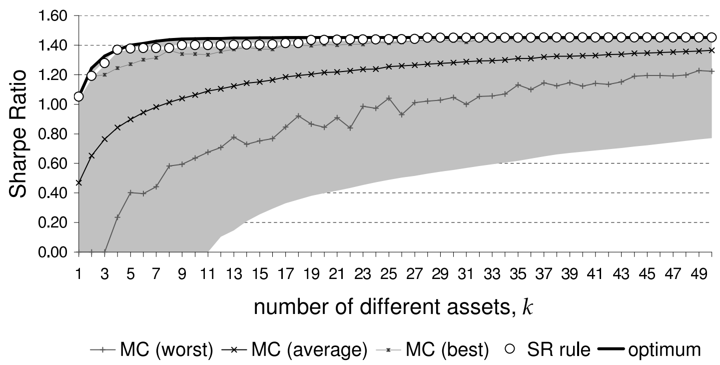
A simple rule of the thumb suggests to prefer assets which themselves have a high Sharpe Ratio. According to this rule, the available assets are sorted by their *SR*in descending order, and the first k assets are selected for the portfolio. A downside of this rule is that it does not consider the correlation or covariance between the assets which largely affects the portfolio's volatility. Hence, this rule will not necessarily find the optimal solution, particularly when k is rather small. Having a method that has a higher chance of identifying the actually best combination, one can also evaluate how large the gap between the SR rule based portfolios and the optima is. As can be seen from Figure 5 for the DAX data, the Sharpe ratios of portfolios selected with this popular rule could mostly be achieved with one or more assets less. For the FTSE data set, this is even more apparent: For the optimal portfolio with $k = 10$, the SR is higher than for a portfolio with $k = 38$ when selected with the SR rule. The consequences of this gap become even more severe when the investor faces transactions costs that contain a fixed fee as can be seen from the results Maringer (2005, chapter 3). Other rule-based selection methods



(a) DAX data set



(b) FTSE data set



(c) S & P 100 data set

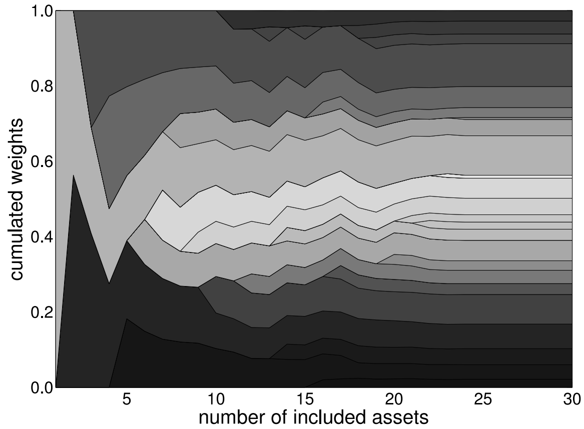
Figure 5: Sharpe Ratios for portfolios under cardinality constraint with different selection processes

such as selections based on the companies' industry, size, or geographic aspects exhibit equal shortcomings.

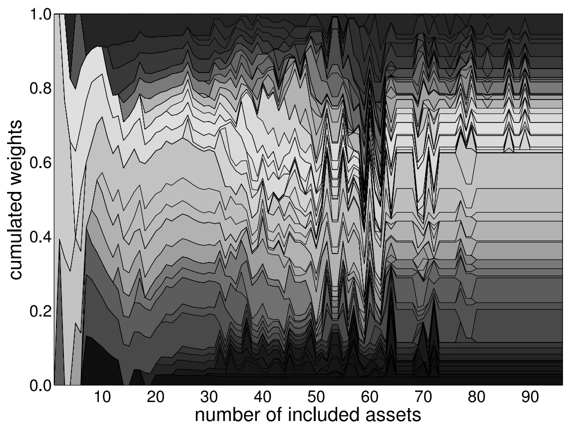
For the Monte Carlo approach, k assets are drawn randomly and their weights are optimized. This selection process has been replicated a 1 000 times, and the best, the worst, and the average *SR* for any k and data set are plotted on the bandwidth for the possible outcomes. As can be seen for the larger markets, it is very unlikely to randomly draw the worst possible solution – yet it is also unlikely that the optimal solution is chosen: In the best of the 1 000 replications, a solution close to that from the SR rule is found, on average, however, a random selection is significantly below what could be achieved with a superior selection method: the upper limit, indicating the optima are the results found with the heuristic search method.

As neither the SR rule nor the MC approach includes the correlations and covariances between the assets into the selection process, a main aspect from portfolio selection might be lost. A closer look at what assets actually are selected and what weights they are given also confirms that the decision of whether to include a certain asset or not depends on what other assets are included. In Figure 6 the cumulated asset weights are depicted for the different values of k. In particular the results for the FTSE data set illustrate that the optimal selection with k assets cannot be determined by simply searching the asset that fits best to the solution with $k - 1$ assets: in smaller portfolios one asset might serve as a substitute for a bundle of other assets which, however, cannot be included because of the constraints (be it transactions costs, be it cardinality). Also, what makes a good choice in a portfolio with few different assets might or might not be a good choice for large portfolios.

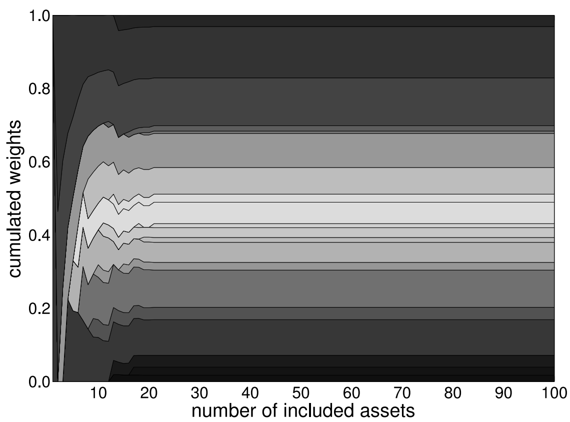
The results for the S&P 100 data set (Figure 6(c)) also exhibits a particularity of this data set: Given the estimates for the assets' returns and covariances, only a limited number of assets are actually assigned positive weights, i.e., even for large k only a small number of different assets is included in the portfolio, and the cardinality constraint is no longer binding. The selection of these assets depends to some extent on the choice of the safe interest rate, r_S, as (geometrically speaking) the tangency line from the Tobin efficiency line crosses the y-axis of the mean-variance-diagram at a different point, yet



(a) DAX data set



(b) FTSE data set



(c) S & P 100 data set

Figure 6: Cumulated weights for optimal portfolios under cardinality constraint

the basic results are unchanged. The SR rule "filters" most of the assets that ought not to be included in any of the optimal portfolios but again still ignores some better combinations in the lack of considering the covariances. In this type of market situation, however, a Monte Carlo approach will be likely to also pick one or several of these undesirable assets – and will therefore be clearly inferior to a heuristic search strategy.

4 Conclusion

For various reasons, investors tend to hold a rather small number of assets. In this paper, a method has been presented to approach the associated NP hard optimization problem of selecting the optimal set of assets under a given market situation and expectations. The main results from this empirical study are twofold: (i) the well known fact of decreasing marginal contribution to diversification is not only confirmed, but can be exploited by identifying those assets that, in combination, offer the highest risk premium; (ii) it has been shown that alternative rules, frequently found in practice, are likely to underperform as they offer solutions with risk premia lower than would be possible under the same constraints and market situations.

References

BLUME, M. E., AND I. FRIEND (1975): "The Asset Structure of Individual Portfolios and Some Implications for Utility Functions," *The Journal of Finance*, 30(2), 585–603.

BOLLERSLEV, T. (1986): "Generalized Autoregressive Conditional Heteroscedasticity," *Journal of Econometrics*, 31(3), 307–327.

BONABEAU, E., M. DORIGO, AND G. THERAULAZ (1999): *Swarm Intelligence. From Natural to Artificial Systems*. Oxford University Press, New York and Oxford.

BÖRSCH-SUPAN, A., AND A. EYMANN (2000): "Household Portfolios in Germany," Working paper, University of Mannheim.

BRANDIMARTE, P. (2002): *Numerical Methods in Finance. A MATLAB-Based Introduction*, Probability and Statistics. Wiley, New York.

BULLNHEIMER, B., R. HARTL, AND C. STRAUSS (1999): "A New Rank Based Version of the Ant System — A Computational Study," *Central European Journal of Operations Research*, 7, 25–38.

COLORNI, A., M. DORIGO, AND V. MANIEZZO (1992a): "Distributed Optimization by Ant Colonies," in *Proceedings of the First European Conference on Artificial Life ECAL '91*, ed. by F. Varela, and P. Bourgine, pp. 134–142. Elsevier, Paris.

——— (1992b): "Investigation of Some Properties of an Ant Algorithm," in *Proceedings of the Parallel Problem Solving from Nature*, ed. by R. Männer, and B. Manderick, pp. 509–520. Elsevier, Brussels.

DORIGO, M. (1992): "Optimization, Learning and Natural Algorithms," Dissertation, Politecnico di Milano, Milano.

DORIGO, M., AND G. DI CARO (1999): "The Ant Colony Optimization Meta-Heuristic," in *New Ideas in Optimization*, ed. by D. Corne, M. Dorigo, and F. Glover. McGraw-Hill, London.

DORIGO, M., V. MANIEZZO, AND A. COLORNI (1991): "Positive Feedback as a Search Strategy," Technical Report 91-016, Politecnico di Milano, Milano, Italy.

——— (1996): "The Ant System: Optimization by a Colony of Cooperating Agents," *IEEE Transactions on Systems, Man, and Cybernetics - Part B*, 26(1), 29–42.

ELTON, E. J., M. J. GRUBER, S. J. BROWN, AND W. N. GOETZMANN (2003): *Modern Portfolio Theory and Investment Analysis*. Wiley & Sons, Inc., Hoboken, NJ, 6th edn.

ENGLE, R. (1982): "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50(1), 987–1007.

ENGLE, R. F. (ed.) (1995): *ARCH: Selected Readings*. Oxford University Press, Oxford, UK.

FARRELL, JR., J. (1997): *Portfolio Management: Theory and Application*. McGraw-Hill, New York et al., 2nd edn.

GOSS, S., S. ARON, J. DENEUBOURG, AND J. PASTEELS (1989): "Self-Organized Shortcuts in the Argentine ant," *Naturwissenschaften*, 76, 579–581.

GUISO, L., T. JAPPELLI, AND D. TERLIZZESE (1996): "Income Risk, Borrowing Constraints and Portfolio Choice," *American Economic Review*, 86(1), 158–172.

JANSEN, R., AND R. VAN DIJK (2002): "Optimal Benchmark Tracking with Small Portfolios," *The Journal of Portfolio Management*, 28(2), 9–22.

KEBER, C., AND D. MARINGER (2001): "On Genes, Insects and Crystals: Determing Marginal Diversification Effects with Nature Based Methods," presentation, Annual Meeting of the Society of Computational Economics at Yale University.

KELLERER, H., U. PFERSCHY, AND D. PISINGER (2004): *Knapsack Problems*. Springer, New York.

MARINGER, D. (2001): "Optimizing Portfolios with Ant Systems," in *International ICSC Congress on Computational Intelligence: Methods and Applications (CIMA 2001)*, pp. 288–294. ICSC.

——— (2002a): "Portfolioselektion bei Transaktionskosten und Ganzzahligkeitsbeschränkungen," *Zeitschrift für Betriebswirtschaft*, 72(11), 1155–1176.

——— (2002b): "Wertpapierselektion mittels Ant Systems," *Zeitschrift für Betriebswirtschaft*, 72(12), 1221–1240.

——— (2004): "Finding the Relevant Risk Factors for Asset Pricing," *Computational Statistics and Data Analysis*, forthcoming.

——— (2005): *Portfolio Management with Heuristic Optimization*. Springer, New York et al.

Maringer, D., and H. Kellerer (2003): "Optimization of Cardinality Constrained Portfolios with a Hybrid Local Search Algorithm," *OR Spectrum*, 25(4), 481–495.

Ross, S. A. (1976): "The Arbitrage Theory of Capital Asset Pricing," *The Journal of Economic Theory*, 13(3), 341–360.

Russell, S., and P. Norvig (2003): *Artificial Intelligence. A Modern Approach*. Prentice Hall, Englewood Cliffs, NJ, 2nd edn.

Solnik, B. (1973): "The Advantages of Domestic and International Diversification," in *International Capital Markets*, ed. by E. J. Elton, and M. J. Gruber. North-Holland Publishing Company, Amsterdam.

Tobin, J. (1958): "Liquidity Preference as Behavior Towards Risk," *Review of Economic Studies*, 26(1), 65–86.