

Sparse, mean reverting portfolio selection using simulated annealing

Norbert Fogarasi^a and Janos Levendovszky^b

Department of Networked Systems and Services, Budapest University of Technology and Economics, Budapest, Hungary

^a*E-mail: fogarasi@hit.bme.hu*

^b*E-mail: levendov@hit.bme.hu*

Abstract. We study the problem of finding sparse, mean reverting portfolios based on multivariate historical time series. After mapping the optimal portfolio selection problem into a generalized eigenvalue problem, we propose a new optimization approach based on the use of simulated annealing. This new method ensures that the cardinality constraint is automatically satisfied in each step of the optimization by embedding the constraint into the iterative neighbor selection function. We empirically demonstrate that the method produces better mean reversion coefficients than other heuristic methods, but also show that this does not necessarily result in higher profits during convergence trading. This implies that more complex objective functions should be developed for the problem, which can also be optimized under cardinality constraints using the proposed approach.

Keywords: mean reversion, convergence trading, parameter estimation, stochastic optimization, simulated annealing.

1. Introduction

Ever since publication of the seminal paper of Markowitz (1952), the selection of portfolios which are optimal in terms of risk-adjusted returns has been an active area of research both by academics and financial practitioners. A key finding of Markowitz is that diversification, as a means of reducing the variance and therefore increasing the predictability of an investment portfolio, is of paramount importance even at the cost of reducing expected return. The resulting, so-called “mean-variance” approach to portfolio selection involves tracing out the surface of efficient portfolios, which can be found via quadratic programming and successfully applied by practitioners (Konno and Yamazaki, 1991; Feiring et al., 1994). Recent improvements to the model have considered including more complex measures of risk (Elton et al., 2007), transaction costs (Adcock and Meade, 1994; Yoshomito, 1996) and sparsity constraints (Streichert et al., 2003; Anagnostopoulos and Mamanis, 2011) to this basic model. These additional constraints have made the portfolio selection problem

significantly more complex, justifying the use of various heuristic methods, such as genetic algorithms (Loraschi et al., 1995; Arnone et al., 1993), neural networks (Fernández and Gómez, 2007) and even simulated annealing (Chang et al., 2000; Armananzas and Lozano, 2005), which is also used in this paper. A comprehensive study of the metaheuristics in portfolio selection can be found in Di Tollo and Rolli (2008).

At the same time, mean reversion, as a classic indicator of predictability, has also received a great deal of attention over the last few decades. It has been shown that equity excess returns over long horizons are mean reverting and therefore contain an element of predictability (Poterba and Summers, 1988; Fama and French, 1988; Manzan, 2007).

While there exist simple and reliable methods to identify mean reversion in univariate time series, selecting portfolios from multivariate data which exhibit this property is a much more difficult problem. This can be approached by the Box-Tiao procedure (Box and Tiao, 1977) to extract cointegrated vectors by solving a generalized eigenvalue problem.

In his recently published article, d'Aspremont (2011) posed the problem of finding sparse portfolios which maximize mean reversion. The impetus is that by finding mean reverting portfolios, one can develop conversion trading strategies to produce profits, as opposed to the buy and hold strategy associated with portfolios which maximize excess returns or minimize risk. Sparseness, he argues, is desirable for reducing transaction costs associated with convergence trading as well as for increasing the interpretability of the resulting portfolio. He developed a new approach to solve the problem by using semidefinite relaxation and compared the efficiency of this solution to the simple greedy algorithm in a number of markets. Inspired by this work, we recently suggested some further heuristics for portfolio selection, compared these to the theoretically optimal exhaustive solution and outlined a simplified method to estimate the parameters of the underlying stationary first order vector autoregressive VAR(1) model, which also provides a model fit measure (Fogarasi and Levendovszky, 2011). Subsequently, we suggested improvements to estimating the underlying Ornstein-Uhlenbeck process and outlined a simple convergence trading algorithm based on decision theory (Fogarasi and Levendovszky, 2012).

In this paper, we summarize the above results and combine them into a comprehensive framework (see Fig. 1). We also outline a new method for portfolio selection using simulated annealing. This new method is compared to the existing methods via extensive simulations on both generated and historical time series.

The structure of the paper is as follows:

- In section 2, after giving a formal presentation of the problem, we review the existing heuristics and present a novel approach using simulated annealing.
- In section 3, we review the methods for estimating the parameters of a VAR(1) model and explain a novel approach to trading based on decision theory.

- Finally, in section 4, the methodology on generated VAR(1) data is validated and significant trading gains are demonstrated. We also analyze the performance on historical time series of real data: the daily close prices of 8 different maturity U.S. swap rates and the daily close prices of stocks comprising the S&P 500 index.

2. Sparse mean reverting portfolio selection

In this section, the model is described together with the foundations of mean reverting portfolio identification. Our approach follows the one published by d'Aspremont (2011), and in Section 2.3 we present the methods published by Fogarasi and Levendovszky (2011).

2.1. Mean reverting portfolios

The problem we examine in this paper is the selection of sparse, mean reverting portfolios which follow the so-called Ornstein-Uhlenbeck process (Ornstein and Uhlenbeck, 1930). Let $s_{i,t}$ denote the price of asset i at time instant t , where $i = 1, \dots, n$ and $t = 1, \dots, m$ are positive integers. We form portfolios, of value p_t of these assets with coefficients x_i , and assume they follow an Ornstein-Uhlenbeck process given by:

$$dp_t = \lambda(\mu - p_t) dt + \sigma dW_t, \quad (1)$$

where W_t is a Wiener process and $\lambda > 0$ (mean reversion coefficient), μ (long-term mean), and $\sigma > 0$ (portfolio volatility) are constants.

Considering the continuous version of this process and using the Ito-Doebelin formula (Ito, 1944) to solve it, we get:

$$p(t) = p(0) e^{-\lambda t} + \mu (1 - e^{-\lambda t}) + \int_0^t \sigma e^{-\lambda(t-s)} dW(s) \quad (2)$$

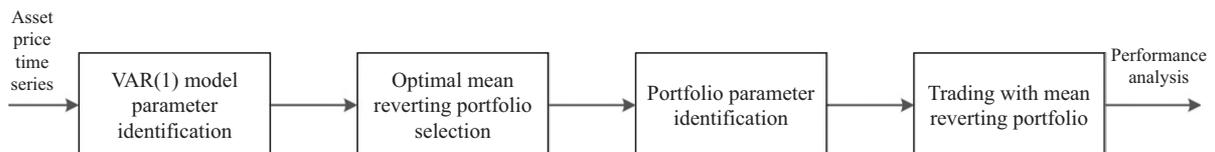


Fig. 1. Overall framework for identifying and trading sparse mean reverting portfolios.

which implies that

$$E[p(t)] = p(0) e^{-\lambda t} + \mu (1 - e^{-\lambda t}) \quad (3)$$

and asymptotically

$$\lim_{t \rightarrow \infty} p(t) \sim N \left(\mu, \sqrt{\frac{\sigma^2}{2\lambda}} \right) \quad (4)$$

For trading, λ is a key parameter, as it determines how fast the process returns to the mean, as well as inversely indicating the level of uncertainty around the mean (via the standard deviation of the asymptotic Gaussian distribution). Hence, the larger the λ , the more suitable is the mean reverting portfolio for convergence trading, as it quickly returns to the mean and contains a minimum amount of uncertainty around the mean. Therefore, we shall be concerned with finding sparse portfolios which are optimal in the sense that they maximize λ . Later on, we show that this approach may be overly simplistic if our purpose is to maximize profits achieved via convergence trading. In this case, the distance from the long-term mean, the model fit and other factors should also be taken into account.

2.2. Mean reverting portfolio as a generalized eigenvalue problem

In this section, we view asset prices as a first order, vector autoregressive VAR(1) process. Let $s_{i,t}$ denote the price of asset i at time instant t , where $i = 1, \dots, n$ and $t = 1, \dots, m$ are positive integers, and assume that $\mathbf{s}_t^T = (s_{1,t}, \dots, s_{n,t})$ is subject to a first order vector autoregressive process, VAR(1), defined as follows:

$$\mathbf{s}_t = \mathbf{A}\mathbf{s}_{t-1} + \mathbf{W}_t, \quad (5)$$

where \mathbf{A} is an $n \times n$ matrix and $\mathbf{W}_t \sim N(0, \sigma I)$ are i.i.d. noise terms for some $\sigma > 0$.

One can introduce a portfolio vector $\mathbf{x}^T = (x_1, \dots, x_n)$, where component x_i denotes the amount of asset i held. In practice, assets are traded in discrete units, so $x_i \in \{0, 1, 2, \dots\}$ but for the purposes of our analysis we allow x_i to be any real number, including negative ones, which denote the ability to short sell assets. Multiplying both sides by vector \mathbf{x} (in the inner product sense), we obtain

$$\mathbf{x}^T \mathbf{s}_t = \mathbf{x}^T \mathbf{s}_{t-1} \mathbf{A} + \mathbf{x}^T \mathbf{W}_t \quad (6)$$

Following the treatment in d'Aspremont (2011), we define the *predictability* of the portfolio as

$$\nu(\mathbf{x}) := \frac{\text{var}(\mathbf{x}^T \mathbf{A}\mathbf{s}_{t-1})}{\text{var}(\mathbf{x}^T \mathbf{s}_t)} = \frac{E(\mathbf{x}^T \mathbf{A}\mathbf{s}_{t-1} \mathbf{s}_{t-1}^T \mathbf{A}^T \mathbf{x})}{E(\mathbf{x}^T \mathbf{s}_t \mathbf{s}_t^T \mathbf{x})}, \quad (7)$$

provided that $E(\mathbf{s}_t) = 0$, so the asset prices are normalized for each time step. The intuition behind this portfolio predictability is that the greater this ratio, the more \mathbf{s}_{t-1} dominates the noise, and therefore the more predictable \mathbf{s}_t becomes. Therefore, we will use this measure as a proxy for the portfolio's mean reversion parameter λ in (1). Maximizing this expression will yield the following optimization problem for finding the best portfolio vector \mathbf{x}_{opt} :

$$\mathbf{x}_{\text{opt}} = \arg \max_{\mathbf{x}} \nu(\mathbf{x}) = \arg \max_{\mathbf{x}} \frac{\mathbf{x}^T \mathbf{A} \mathbf{G} \mathbf{A}^T \mathbf{x}}{\mathbf{x}^T \mathbf{G} \mathbf{x}} \quad (8)$$

where \mathbf{G} is the stationary covariance matrix of process \mathbf{s}_t . Based on (8), we observe that the problem is equivalent to finding the eigenvector corresponding to the maximum eigenvalue in the following generalized eigenvalue problem (d'Aspremont, 2011):

$$\mathbf{A} \mathbf{G} \mathbf{A}^T \mathbf{x} = \lambda \mathbf{G} \mathbf{x} \quad (9)$$

λ can be obtained by solving the following equation:

$$\det(\mathbf{A} \mathbf{G} \mathbf{A}^T - \lambda \mathbf{G}) = 0 \quad (10)$$

2.3. Sparse portfolio selection

In the previous section, we outlined the process of selecting a portfolio which maximizes predictability by solving a generalized eigenvalue problem. However, we are seeking the optimal portfolio vector that exhibits a mean reverting property under sparseness constraint. Adding this to (8), we can formulate our constrained optimization problem as follows:

$$\mathbf{x}_{\text{opt}} = \arg \max_{\mathbf{x} \in \mathbf{R}^n, \text{card}(\mathbf{x}) \leq L} \frac{\mathbf{x}^T \mathbf{A} \mathbf{G} \mathbf{A}^T \mathbf{x}}{\mathbf{x}^T \mathbf{G} \mathbf{x}} \quad (11)$$

where *card* denotes the number of non-zero components, and L is a given integer $1 \leq L \leq n$.

The cardinality constraint poses a serious computational challenge, as the number of subspaces in which optimality must be checked grows exponentially.

In fact, Natarjan (1955) shows that this problem is equivalent to the subset selection problem, which is proven to be NP-hard. However, as a benchmark metric, we can compute the theoretically optimal solution which, depending on the level of sparsity and the total number of assets, could be computationally feasible (Fogarasi and Levendovszky, 2011). We also describe three polynomial time, heuristic algorithms for an approximate solution of this problem.

Exhaustive search method

The brute force approach of constructing all $\binom{n!}{L!(n-L)!}$ L -dimensional submatrices of \mathbf{G} and \mathbf{AGA}^T , and then solving all the corresponding eigenvalue problems to find the theoretical optimum is, in general, computationally infeasible. However, for relatively small values of n and L , or as an off-line computed benchmark, this method can provide a very useful basis of comparison. Indeed, for the practical applications considered by d'Aspremont (2011) (selecting sparse portfolios of $n = 8$ U.S. swap rates and $n = 14$ FX rates), this method is fully applicable and could be used to examine the level of sub-optimality of other proposed methods.

Greedy method

A reasonably fast heuristic algorithm is the so-called greedy method, first presented by d'Aspremont (2011). Let I_k be the set of indices belonging to the k non-zero components of \mathbf{x} . One can then develop the following recursion for constructing I_k with respect to k .

$$\text{When } k = 1, \text{ we set } i_1 = \arg \max_{j \in [1, n]} \frac{(\mathbf{AGA}^T)_{jj}}{\mathbf{G}_{jj}}.$$

Suppose now that we have a reasonable approximate solution with support set I_k given by $(\mathbf{x})_k = \arg \max_{\{\mathbf{x} \in \mathbb{R}^n : x_{J_k^C} = 0\}} \frac{\mathbf{x}^T \mathbf{AGA}^T \mathbf{x}}{\mathbf{x}^T \mathbf{G} \mathbf{x}}$, where I_k^C is the

complement of the set I_k . This can be solved as a generalized eigenvalue problem of size k . We seek to add one variable with index i_{k+1} to the set I_k to produce the largest increase in predictability by scanning each of the remaining indices in I_k^C . The index i_{k+1} is then given by

$$i_{k+1} = \arg \max_{i \in I_k^C} \max_{\{\mathbf{x} \in \mathbb{R}^n : x_{J_i} = 0\}} \frac{\mathbf{x}^T \mathbf{AGA}^T \mathbf{x}}{\mathbf{x}^T \mathbf{G} \mathbf{x}}, \quad (12)$$

where $J_i = I_k^C \setminus \{i\}$

which amounts to solving $(n - k)$ generalized eigenvalue problems of size $k + 1$. We then define $I_{k+1} = I_k \cup \{i_{k+1}\}$, and repeat the procedure until

$k = n$. Naturally, the optimal solutions of the problem might not have increasing support sets $I_k \subset I_{k+1}$, hence the solutions generated by this recursive algorithm are potentially far from optimal. However, the cost of this method is relatively low: with each iteration costing $O(k^2(n - k))$, the complexity of computing solutions for all target cardinalities k is $O(n^4)$. This recursive procedure can also be repeated forward and backward to improve the quality of the solution.

Truncation method

A simple and very fast heuristic that we can apply is the following. First, compute \mathbf{x}_{opt} , the unconstrained, n -dimensional solution of the optimization problem in (8) by solving the generalized eigenvalue problem in (9). Next, consider the L largest values of \mathbf{x}_{opt} and construct $L \times L$ dimensional submatrices \mathbf{G}' and $(\mathbf{AGA}^T)'$ corresponding to these dimensions. Solving the generalized eigenvalue problem in this reduced space and padding the resulting \mathbf{x}'_{opt} with 0's will yield a feasible solution $\mathbf{x}_{\text{opt}}^{\text{trunc}}$ to the original constrained optimization problem. The big advantage of this method is that with just two maximum eigenvector computations, we can determine an estimate for the optimal solution. The intuition behind this heuristic is that the most significant dimensions in the solution of the unconstrained optimization problem could provide, in most cases, a reasonable guess for the dimensions of the constrained problem. This is clearly not the case in general, but nonetheless, the truncation method has proven to be a very quick and useful benchmark for evaluating other methods.

Simulated annealing with constraint satisfaction over a discretized grid

Another approach to solving (9) is to restrict the portfolio vector \mathbf{x} to contain only integer values. Indeed, when it comes to trading methods utilizing the optimal portfolio vector, only an integer number of assets can be purchased in most markets, so this restriction more closely resembles the truth. If the per unit price of the asset is relatively large, this can give rise to a material difference to considering all real-valued portfolio vectors.

As such, the equation in (11) becomes a combinatorial optimization problem over an n -dimensional discrete grid where the subspace of allowable solutions is limited to dimensionality L . The main difficulty in combinatorial optimization problems is that simpler algorithms such as greedy methods and local search methods tend to become trapped in local minima.

A large number and variety of stochastic optimization methods have been developed to address this problem, one of which is simulated annealing as proposed by Kirkpatrick et al. (1983).

At each step of the algorithm, we consider a neighboring state \mathbf{w}' of the current state \mathbf{w}_n and decide between moving the system to state \mathbf{w}' , or staying in \mathbf{w}_n . Metropolis et al. (1953) suggested the use of the Boltzmann-Gibbs distribution as the probability function to make this decision as follows:

$$P(\mathbf{w}_{n+1} = \mathbf{w}') = \begin{cases} 1 & \text{if } E(\mathbf{w}_n) > E(\mathbf{w}') \\ e^{-\frac{E(\mathbf{w}') - E(\mathbf{w}_n)}{T}} & \text{if } E(\mathbf{w}_n) \leq E(\mathbf{w}') \end{cases} \quad (13)$$

where $E(\cdot)$ is the function to be minimized and T is the temperature of the system. If T is fixed and the *neighbor function* which generates the random neighboring state at each step is *non-periodic* and *ergodic* (any state is reachable from any other state by some sequence of moves), then this system corresponds to a regular time-homogenous finite Markov chain. For the Metropolis transition probability function, the stationary probability distribution of finding the system at a given state \mathbf{w} is given by the Boltzmann distribution:

$$\pi_{\mathbf{w}} = e^{-\frac{E(\mathbf{w})}{TK_B}} \quad (14)$$

where K_B is the Boltzmann constant, equal to $1.38 \cdot 10^{-23}$ J/K. The stationary distribution given by (14) indicates that the maximum of function $E(\mathbf{w})$ will be reached with maximum probability

(Salamon et al., 2002). Furthermore, when the temperature parameter T is decreased during the procedure (also known as *cooling*), convergence in distribution to the uniform measure over globally optimal solutions has been proven by Geman and Geman (1984), provided that the cooling schedule is slow enough to be at least inversely proportional to the logarithm of time. In practice, the time performance of inverse logarithmic cooling is often untenable, so faster heuristic algorithms have been developed as a result, such as exponential, geometric and adaptive cooling and applied successfully to a wide range of combinatorial optimization problems (Chen and Aihara, 1995; Ingber, 1993; Johnson et al., 1989; Johnson et al., 1991).

In our application, based on the formulation in (8), the energy function to be minimized is defined as

$$E(\mathbf{w}) = -\frac{\mathbf{w}^T \mathbf{A}^T \mathbf{G} \mathbf{A} \mathbf{w}}{\mathbf{w}^T \mathbf{G} \mathbf{w}} \quad (15)$$

A very important feature of the simulated annealing method is that the cardinality constraint can be easily built into the search, by mapping the unconstrained \mathbf{w} vector into a randomly selected L dimensional subspace on each step. The idea is that on each step, we randomly select k of the n indices $(i_1, \dots, i_k); i_j \in \{1, \dots, n\}$ and we set the corresponding elements of \mathbf{w} to be 0: $\mathbf{w}_{i_j} := 0, j = 1, \dots, k$. Full pseudocode for the neighbor function and further details of the whole algorithm are presented in the Appendix.

The high-level pseudocode of the general simulated annealing algorithm used is as follows:

```

s ← Greedy_sol; e ← E(s)
sbest ← s; ebest ← e
k ← 0
reject ← 0
while ~stop(reject, k, e, ebest)

    snew ← neighbor(s)

    enew ← E(snew)
    if P(e, enew, temp(k/kmax)) > random() then
        s ← snew; e ← enew;
        reject ← 0
    else reject ← reject + 1
    if enew > ebest then
        sbest ← snew; ebest ← enew
    k ← k + 1
return sbest

// Start from Greedy sol.
// Initial "best" solution
// Energy evaluation count
// consecutive rejections
// While stop conditions
// (see Appendix) not met
// Pick some neighbour
// See Appendix for details
// Compute its energy.
// Should we move to it?
// Yes, change state.
// Reset reject counter
// Increment reject counter
// Is this a new best?
// Save to 'best found'.
// One more evaluation done
// Return the best found.

```

The parameters of the simulated annealing algorithm (function `temp`, which determines the cooling schedule, and function `stop`, determining the halting condition) need to be selected for acceptable speed of convergence. Whilst the logarithmic cooling schedule, shown to converge to the optimal solution by Geman and Geman (1984), has proven to be too slow in the application to this particular problem, in the Appendix we show the details of a modified geometric cooling algorithm with appropriate stopping conditions. This has free parameters to adapt to the size of the specific problem at hand.

3. Model parameter estimation and convergence trading

In this section, we summarize our methods proposed in earlier studies (Fogarasi and Levendovszky, 2011; Fogarasi and Levendovszky, 2012) for estimating the model parameters and performing convergence trading.

3.1. Estimation of VAR(1) parameters

As explained in the preceding sections, in the knowledge of the parameters \mathbf{G} and \mathbf{A} , we can apply various heuristics to approximate the L -dimensional optimal sparse mean reverting portfolio. However, these matrices must be estimated from historical observations of the random process \mathbf{s}_t . While there is a vast amount of literature on the topic of parameter estimation of VAR(1) processes, recent research has focused on sparse and regularized covariance estimation (Banerjee et al., 2008; d'Aspremont et al., 2008; Rothman et al., 2008). As a new approach, Fogarasi and Levendovszky (2011) have suggested finding a dense estimate for \mathbf{G} which best describes the observed historical data and dealing with dimensionality reduction with the apparatus outlined in section 2.

We start by estimating the recursion matrix \mathbf{A} . If the number of assets n is greater than or equal to the length of the observed time series m , then \mathbf{A} can be estimated by simply solving the linear system of equations:

$$\hat{\mathbf{A}}\mathbf{s}_{t-1} = \mathbf{s}_t \quad (16)$$

This gives a perfect VAR(1) fit for our time series in cases where we have a large portfolio of potential assets (e.g. considering all 500 stocks which make up the S&P 500 index), from which a sparse mean reverting subportfolio is to be chosen.

In most practical applications, however, the length of the available historical time series is greater than the number of assets considered, so \mathbf{A} is estimated using, for example, least squares estimation techniques, as derived by Fogarasi and Levendovszky (2012). The resulting estimate is

$$\hat{\mathbf{A}} = \left(\sum_{t=2}^m \mathbf{s}_t \mathbf{s}_{t-1}^T \right) \left(\sum_{t=2}^m \mathbf{s}_{t-1} \mathbf{s}_{t-1}^T \right)^+, \quad (17)$$

where \mathbf{M}^+ denotes the Moore-Penrose pseudoinverse of matrix \mathbf{M} . Note that the Moore-Penrose pseudoinverse is preferred to regular matrix inversion, in order to avoid problems which may arise due to the potential singularity of $\mathbf{s}_{t-1}^T \mathbf{s}_{t-1}$.

Assuming that the noise terms in equation (5) are i.i.d. with $\mathbf{W}_t \sim N(0, \sigma_W I)$ for some $\sigma_W > 0$, we obtain the following estimate for σ_W using $\hat{\mathbf{A}}$ from (17):

$$\hat{\sigma}_W = \sqrt{\frac{1}{n(m-1)} \sum_{t=2}^m \left\| \mathbf{s}_t - \hat{\mathbf{A}}\mathbf{s}_{t-1} \right\|^2} \quad (18)$$

In the case that the terms of \mathbf{W}_t are correlated, we can estimate the covariance matrix \mathbf{K} of the noise as follows:

$$\hat{\mathbf{K}} = \frac{1}{m-1} \sum_{t=2}^m (\mathbf{s}_t - \hat{\mathbf{A}}\mathbf{s}_{t-1})(\mathbf{s}_t - \hat{\mathbf{A}}\mathbf{s}_{t-1})^T \quad (19)$$

Using the results of (17) and (19), Fogarasi and Levendovszky (2012) have derived a numerical method for computing the covariance matrix, based on the following iteration:

$$\mathbf{G}(k+1) = \mathbf{G}(k) - \delta(\mathbf{G}(k) - \hat{\mathbf{A}}\mathbf{G}(k)\hat{\mathbf{A}}^T - \hat{\mathbf{K}}) \quad (20)$$

where δ is a constant between 0 and 1, and $\mathbf{G}(i)$ is the covariance matrix estimate on iteration i . Provided that the starting point for the numerical method, $\mathbf{G}(0)$, is positive definite (eg. the sample covariance matrix), and since our estimate of \mathbf{K} is positive definite, by construction, this iterative method will produce an estimate which will be positive definite. We have also shown that comparing this estimate to the sample covariance estimate gives rise to a model fit parameter which can later be used in determining our level of confidence in a mean reverting portfolio.

3.2. Estimation of Ornstein-Uhlenbeck parameters

Having identified the portfolio with maximal mean reversion that satisfies the cardinality constraint, our task now is to develop a profitable convergence trading strategy. The immediate decision that we face is whether the current value of the portfolio is below the mean and is therefore likely to rise so buying is advisable, above the mean and therefore likely to fall so selling is advisable, or close to the mean in an already stationary state, in which case no action will likely result in a profit. In order to formulate this as a problem of decision theory, we first need to estimate the mean value of the portfolio.

As an alternative to the classical methods of mean estimation (sample mean estimate, least

squares linear regression estimate, maximum likelihood estimate), Fogarasi and Leventovszky (2012) have presented a novel mean estimation technique based on pattern matching. The idea is to consider $\mu(t) = \mu + (\mu(0) - \mu)e^{-\lambda t}$, the continuous process of the expected value of the portfolio at time step t with $\mu(0) = p(0)$. Intuitively, this describes the value of the portfolio, without noise, in the knowledge of the long term mean and the initial portfolio value. In the knowledge of observations $\mathbf{p}_t = (p(t-1), p(t-2), \dots, p(0))$ we can use maximum likelihood estimation techniques to determine which of the convergence patterns $\mu(t)$ matches best. The result is the following mean estimate:

$$\hat{\mu} := \frac{\sum_{i=1}^t \sum_{j=1}^t (\mathbf{U}^{-1})_{i,j} [\mu(0) (2e^{-\lambda(i+j)} - e^{-\lambda i} - e^{-\lambda j}) - 2\mathbf{p}_j (e^{-\lambda i} - 1)]}{\sum_{i=1}^t \sum_{j=1}^t 2(\mathbf{U}^{-1})_{i,j} (e^{-\lambda(i+j)} - e^{-\lambda i} - e^{-\lambda j} + 1)}, \quad (21)$$

where $\mathbf{U}_{ij} := \text{cov}(p(t-i), p(t-j)) = \frac{\sigma^2}{2\lambda} (e^{-\lambda(j-i)} - e^{-\lambda(2t-i-j)})$ is the time-covariance matrix of \mathbf{p}_t for $j \geq i$ and $\boldsymbol{\mu}_t = (\mu(t-1), \dots, \mu(0))$.

3.3. A simple convergence trading strategy

In this section, we review a simple model of trading in which we restrict ourselves to holding only one portfolio at a time. This can be perceived as a walk in a binary state-space, depicted in Fig. 2.

As a result, the main task after identifying the mean reverting portfolio and obtaining an estimate for its long-term mean is to verify whether $\mu(t) < \mu$

or $\mu(t) \geq \mu$ based on observing samples $\{p(t) = \mathbf{x}^T \mathbf{s}(t), t = 1, \dots, T\}$. This verification can be perceived as a problem of decision theory, since direct observations on $\mu(t)$ are not available.

If process $p(t)$ is in a stationary state, then the samples $\{p(t), t = 1, \dots, T\}$ are generated by Gaussian distribution $N\left(\mu, \sqrt{\frac{\sigma^2}{2\lambda}}\right)$. As a result, for a given rate of acceptable error ε we can select an α for which

$$\int_{\mu-\alpha}^{\mu+\alpha} \frac{1}{\sqrt{2\pi\sigma^2/2\lambda}} e^{-\frac{(u-\mu)^2}{\sigma^2/2\lambda}} du = 1 - \varepsilon \quad (22)$$

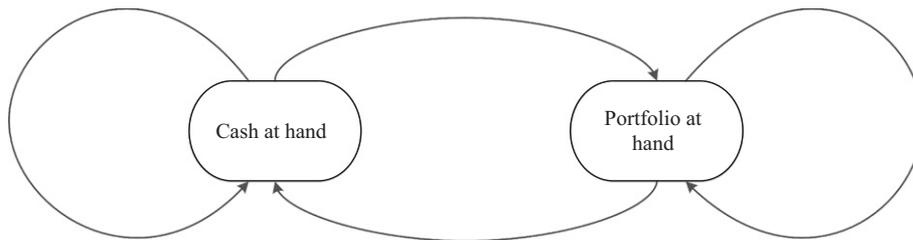


Fig. 2. Trading as a walk in a binary state-space.

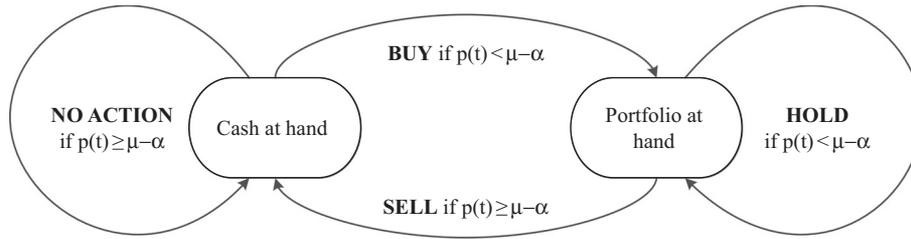


Fig. 3. Flowchart for simple convergence trading of mean reverting portfolios.

As such, having observed the sample $p(t) \in [\mu - \alpha, \mu + \alpha]$, it can be said that we accept the stationary hypothesis which holds with probability $1 - \varepsilon$. Thus, we can augment Fig. 2 with precise conditions as depicted in Fig. 3.

4. Performance analysis

All of the algorithms outlined in this paper have been implemented in MATLAB and extensive numerical simulations were run on both generated and real historical data. The results of these simulations are presented in this section.

4.1. Performance of portfolio selection and trading on generated data

In order to compare the portfolio selection methods outlined in section 2, we generated synthetic VAR(1) data as follows. We first generated an $n \times n$ random matrix \mathbf{A} , ensuring that all of its eigenvalues were smaller than 1. We then generated random i.i.d. noise $\mathbf{W}_t \sim N(0, \sigma I)$ for an arbitrary selection of σ . Finally, we used equation (5) to generate the random sequence s_t , ensuring that all of its values were positive. We then used the methods of section 3 to compute the estimates $\hat{\mathbf{A}}$, $\hat{\mathbf{K}}$ and $\hat{\mathbf{G}}$ and computed optimal sparse portfolios, maximizing the mean-reversion coefficient λ .

Having run 3,000 independent simulations for selecting sparse portfolios of five assets out of a universe of ten, we found that the greedy method generates the theoretically best result produced by an exhaustive search in 70% of the cases. Of the remaining 30% where an improvement over the greedy method is possible, simulated annealing managed to find an improvement in slightly over one-third of the cases, namely in 11% of all simulations. The impact of the improvement produced by the simulated annealing method can be significant, as illustrated in Fig. 4 by one specific generated example

where simulated annealing substantially outperforms the greedy method. We also note that the truncation method performs poorly in this analysis providing mean reversion coefficients lower than other methods in over 99% of the generated cases.

In the following simulation, we increased the asset population to 20 and restricted cardinality of the optimal portfolio to ten. Intuitively, this causes the greedy method to go astray more frequently and thus would provide more room for improvement. The simulations indeed confirm this intuition, as in a simulation of 1,000 independent cases, the greedy method was able to find the theoretical optimum in only 1% of the cases. Under these settings, simulated annealing outperformed the greedy method in 25% of all cases. This indicates that for larger problem sizes, simulated annealing becomes more attractive compared to other simple heuristic methods. This finding is also confirmed when analyzing the runtimes of the different algorithms. As expected, truncation and greedy methods are clearly the fastest of the four examined methods. The speed of simulated annealing depends largely on the settings of the parameters (see Appendix for more details), but it is generally slower than even an exhaustive search for smaller values of n . However, over $n = 21$, simulated annealing is faster than exhaustive search. At the other end of the spectrum, the greedy and truncation methods, although clearly less optimal than exhaustive and simulated annealing, are very fast and therefore can be used for real-time algorithmic trading applications for most reasonably-sized problems. For a total asset size of 100, computing all sparse portfolios of cardinalities 1 to 100 took only two seconds with the truncation method and only 31 seconds with the greedy algorithm on a Pentium 4, 3.80 GHz machine. Simulated annealing can be used for lower frequency (daily or infrequent intraday trading) for most problem sizes. The optimal exhaustive method is only practical for daily or intraday trading if the total number of assets does not exceed 25. Fig. 5 shows more details of the runtimes of the different algorithms.

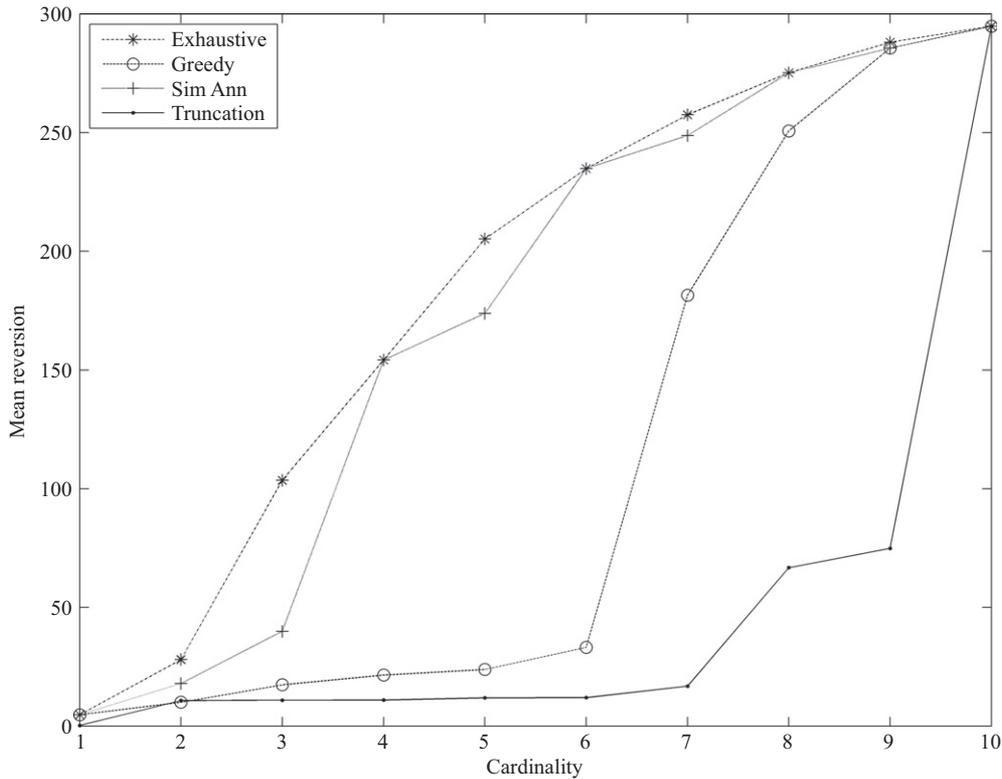


Fig. 4. Comparison of portfolio selection methods of various cardinalities on $n = 10$ -dimensional generated VAR(1) data.

In order to prove the economical viability of our methodology, we implemented the simple convergence trading strategy, as outlined in section 4.3. We generated $n = 10$ -dimensional VAR(1) sequences of length 270 of which the first 20 time steps were used to estimate the parameters of the model and find the optimal portfolio of size $L = 5$ using the different methods. The following 250 (approximate number of trading days in a year) were used to trade the portfolio, using the simple linear regression estimate of μ . Running each of the algorithms on 2,000 different time series, we found that all methods generated a profit in over 97% of the cases. The size of the profit, starting from \$100, using the risky strategy of betting all of our cash on each opportunity, increased monotonically in most simulations, reaching as high as 400% in some cases. The biggest loss across the 2,000 runs was 37% of our initial wealth, showing the robustness of the method on generated data. Fig. 6 shows a typical pattern of successful convergence trading on mean reverting portfolios selected from generated VAR(1) data. We can observe that the more frequent the movement around the estimated long-term mean, the more trading opportunities arise, hence the larger

the profitability of the methodology. Fig. 7 shows the histogram of trading gains achieved by the simulated annealing method. All four methods produced average profits in the same order of magnitude, with the distribution of trading gains very similar. This is despite the fact that the exhaustive method produced mean reversion coefficients on average 15 times those produced by the truncation method, and three times those produced by the greedy method and simulated annealing. This implies that the profits reached by this simple convergence trading strategy are not directly proportional to the lambda produced by the portfolio selection method. In order to maximize trading gains, other factors (such as the model fit, the amount of diversion from the long-term mean, etc) would need to be taken into account. This topic is the subject of further research.

4.2. Performance of portfolio selection and trading on historical data

We consider daily close prices of the 500 stocks which make up the S&P 500 index from July 2009 to July 2010. We use the methods of section 3 to

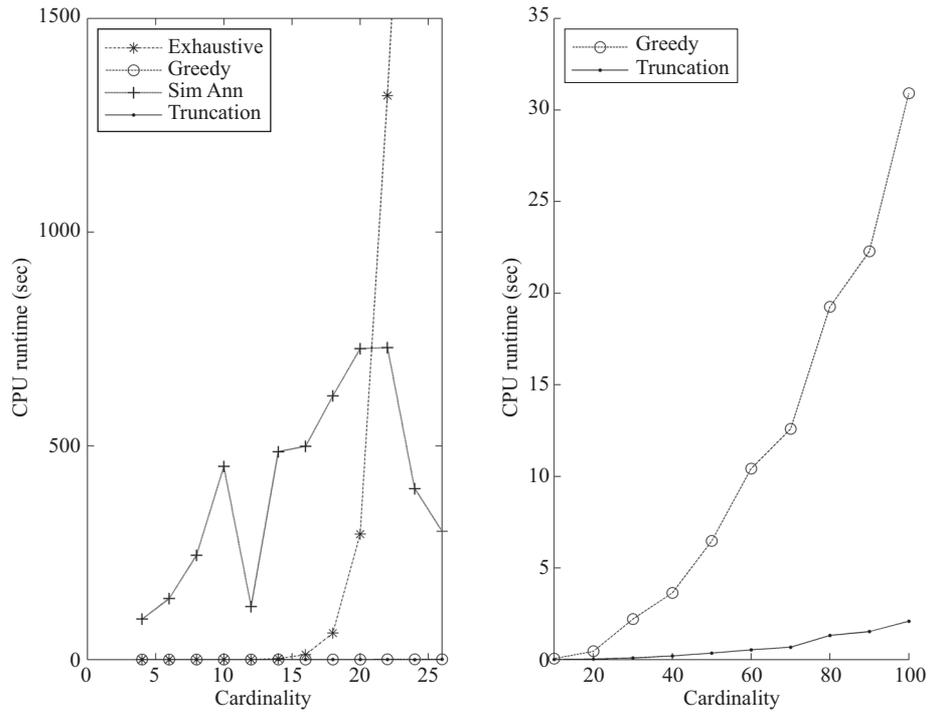


Fig. 5. CPU runtime (in seconds) versus total number of assets n , to compute a full set of sparse portfolios, with cardinality ranging from 1 to n , using the different algorithms.

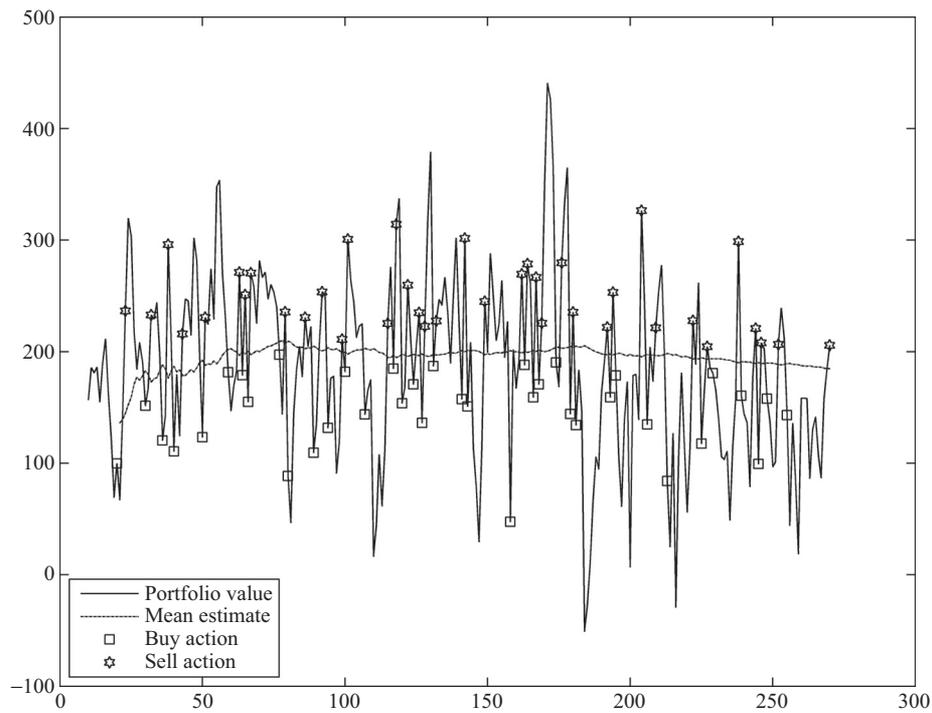


Fig. 6. Typical example of convergence trading over 250 time steps on a sparse mean reverting portfolio. Weighted mix of $L = 5$ assets were selected from $n = 10$ -dimensional generated VAR(1) process by simulated annealing during the first 20 time steps. A profit of \$1,440 was achieved with an initial cash of \$100 after 85 transactions.

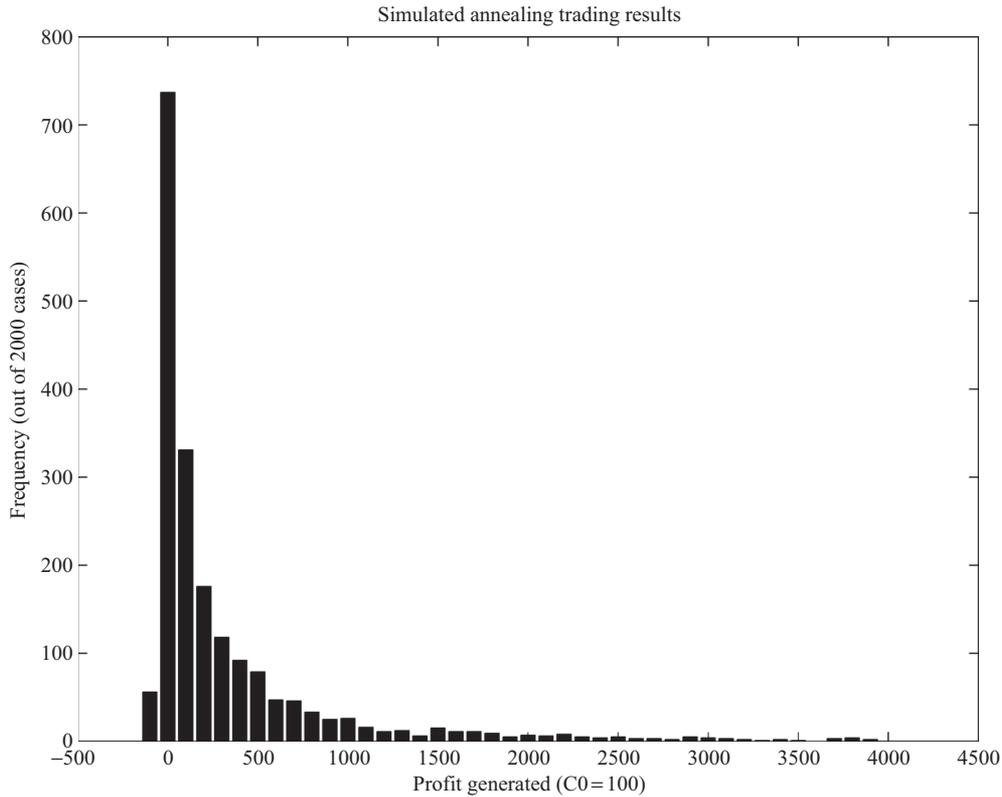


Fig. 7. Histogram of profits achieved over 2,000 different generated VAR(1) series by simulated annealing.

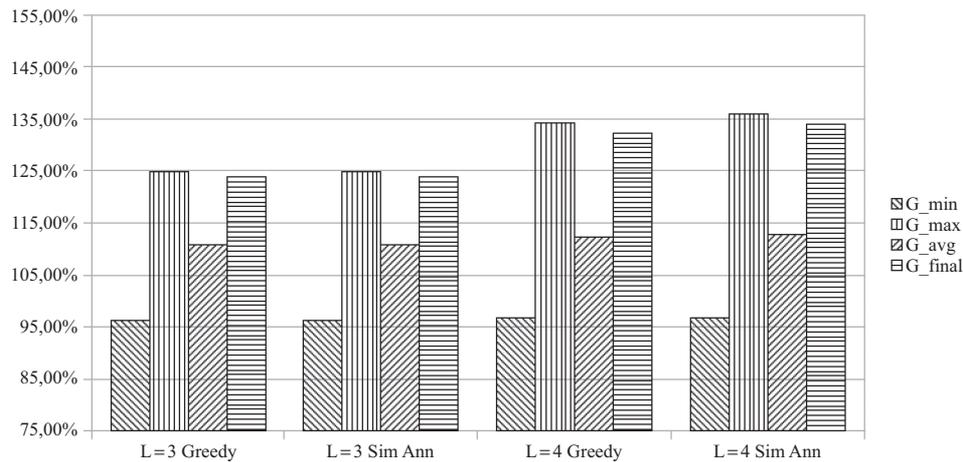


Fig. 8. Comparison of minimum, maximum, average and final return on S&P500 historical data of the greedy and simulated annealing methods for sparse mean reverting portfolios of size $L = 3$ and $L = 4$.

estimate the model parameters on a sliding window of eight observations and select sparse, mean reverting portfolios using the algorithms of section 2. Using the simple convergence trading methodology in section 4 for portfolios of sparseness $L = 3$ and 4, the methods produced annual returns in the range

of 23-34% (note that the return on the S&P 500 index was 11.6% for this reference period). Simulated annealing performed similarly to the greedy method for most cardinalities, but produced better mean reversion coefficients and better profit return for the case $L = 4$ (see Fig. 8).

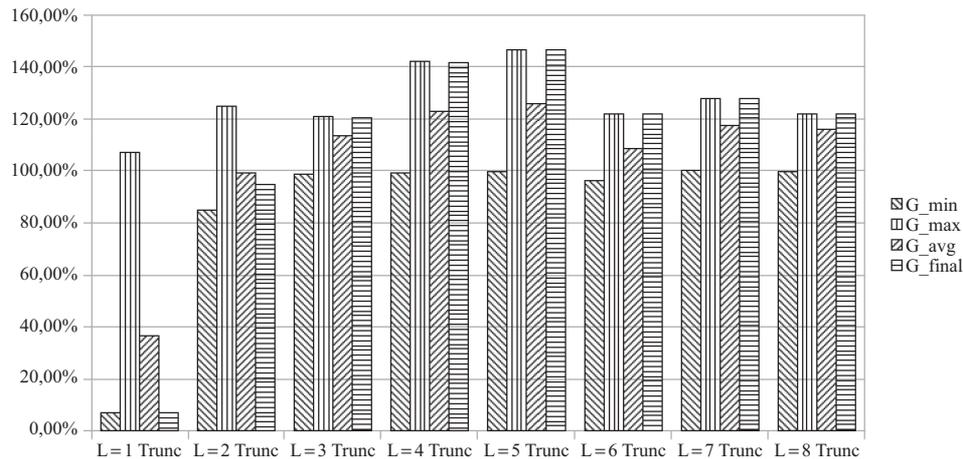


Fig. 9. Comparison of minimum, maximum, average and final return on historical US swap data of the truncation method for sparse mean reverting portfolios of cardinality $L = 1-8$.

Next, we studied U.S. swap rate data for maturities 1Y, 2Y, 3Y, 4Y, 5Y, 7Y, 10Y, and 30Y from 1995 to 2010. Similar to the S&P 500 analysis, we used sliding time windows of eight observations to estimate model parameters, and applied the simple convergence trading methodology on heuristically selected mean reverting portfolios. Due to a poorer model fit, we observed more moderate returns of 15-45% over the 15 years for portfolio cardinalities of $L = 4-6$. The methods generated best results for cardinalities $L = 4$ and 5 (see Fig. 9).

5. Conclusions and directions for future research

We have examined the problem of selecting optimal sparse mean reverting portfolios based on observed and generated time series. After reviewing the novel parameter estimation techniques suggested in earlier publications, we have adapted simulated annealing to the problem of optimal portfolio selection. We have shown that the exhaustive method can be a viable practical alternative for problems with an asset universe size up to 25, and examined the relative performance of simulated annealing versus the greedy method whilst considering the theoretically best solution. We concluded that even for small problem sizes (ten assets and portfolios restricted to cardinality of five), simulated annealing can outperform the greedy method in 10% of the cases, very significantly in some cases. This ratio improved to 25% with the doubling of the problem size

and cardinality restriction, suggesting that simulated annealing becomes more appealing for larger problem sizes, whilst having the asymptotic runtime of simpler heuristics. Furthermore, the viability of the entire framework has been demonstrated by introducing a simple convergence trading strategy which proved to be very profitable on generated VAR(1) sequences and also viable on real historical time series of S&P 500 and US swap rate data. However, the fact that all four methods produced very similar profit profiles implies that the relationship between the mean reversion coefficient produced by the selected sparse portfolio and the profits achieved by trading is nontrivial. Our conclusion is that more factors (such as distance from the mean, model fit, etc.) need to be taken into account in the selection of the objective function. However, our method can be directly applied to these more complex objective functions, making our suggested numerical approach viable. Finally, more sophisticated trading methods, involving multiple portfolios and better cash management, could be developed to further enhance trading performance, and they could be analyzed with the introduction of transaction costs, short-selling costs or an order book.

Acknowledgments

This work was partially supported by the European Union and the European Social Fund through project FuturICT.hu (grant no.: TAMOP-4.2.2.C-11/1/KONV-2012-0013).

The authors also wish to thank the anonymous referee and the editor, Philip Maymin for their valuable comments and suggestions to improve the quality of the paper.

Appendix – Simulated Annealing technical details

In this Appendix, we discuss some of the technical details of our implementation of the simulated annealing algorithm for solving the cardinality constrained maximum eigenvalue problem required for sparse mean reverting portfolio selection.

As explained in section 2.6, optimality in distribution of the solution of simulated annealing can be proven, as long as the *cooling schedule* is slow enough to be at least inversely proportional to the logarithm of time (Geman and Geman, 1984). However, for concrete applications this schedule is “utterly impractical [...] and amounts to a random search in state space.” (Nourani and Andersen, 1998) As such, we implemented the exponential cooling schedule of the form

$$T(t) = T_0 \alpha^t \quad (23)$$

In our implementation, we used $T_0 = 0$, $\alpha = 0.8$ and a maximum of 3,000 repeats at each temperature level, but moving to the next temperature when 100 successful moves have been made. This technique is used to ensure there are a sufficient number of trials at each temperature level, but it also enables the algorithm to proceed if there is reasonable progress.

The same principles must guide our selection of *stopping conditions*. Since our specific application contains no *a priori* target or lower limit for the minimization problem, we set a stopping temperature

$T_{stop} = 10^{-8}$. However, if the algorithm finds no improvement over 10,000 consecutive trials, we also stop the algorithm. Another optimization heuristic we found successful for difficult surfaces is to revert to the best solution thus far after 500 unsuccessful attempts.

As indicated in section 2.6, it is practical to select a competitive *starting point* for the algorithm because, by our construction, the final solution is guaranteed to be at least as good as the starting point. Given the linear scalability of eigenvectors, we can give any scalar multiple of the solution of the greedy algorithm as the starting point for the algorithm. Given that it only operates over discrete whole values, it has been found advisable to give a large scalar multiple of the normalized greedy solution as a starting point – for the purposes of our numerical runs, we used 100 times the normalized greedy solution.

As for the *neighbor function*, as explained in section 2.6, we need a two-step approach to first randomly decide the up to L new non-zero values, and then to randomly decide the up to L new dimensions, given the starting values and dimensions.

For the selection of non-zero values, two different approaches would be to either generate them completely anew, without regard to the starting point, or to perturb the starting non-zero values by uniformly generated random values between 0 and a constant k_1 . A similar choice can be made about the non-zero dimensions as well; they can either be generated completely randomly, independent of the starting point, or we can select $k_2 \leq L$ random new dimensions to change within the dimensions of our starting point. Having implemented various combinations of these choices and comparing the speed of convergence and quality of solution produced on generated data, we found that the following adaptive algorithm works best in practice.

```

function neighbor (s0, T)
for i←1 upto L
    deltas[i]=rand(0,max(1,floor(100*T))
P ← random permutation of nonzero indices in s0
Q ← random permutation of zero indices in s0
n ← rand(0,floor(L*T, (n-L)*T))
for i←1 upto n
    swap P[i] and Q[i]
snew=vector(P,s0[P]+deltas)
return snew

```

```

// s0 is starting point,
// Generate L random
// perturbations to values
// # dimensions to change
// insert perturbed values
// into perturbed P dims

```

The intuition behind this algorithm is that at higher temperatures more dimensions are allowed to change and by larger amounts, but as the system is cooled, only smaller changes to s_0 are allowed.

References

- Adcock, C.J., Meade, N., 1994. A simple algorithm to incorporate transaction costs in quadratic optimization. *Eur. J. Oper. Res.* 79, 85–94.
- Anagnostopoulos, K.P., Mamanis, G., 2011. The mean-variance cardinality constrained portfolio optimization problem: An experimental evaluation of five multiobjective evolutionary algorithms. *Exp. Sys. App.* 38, 14208–14217.
- Armananzas, R., Lozano, J.A., 2005. A multiobjective approach to the portfolio optimization problem. *IEEE congress on evolutionary computation 1*, 1388–1395.
- Arnone, S., Loraschi, A., Tettamanzi, A., 1993. A genetic approach to portfolio selection. *Neur. Netw. W.* 6, 597–604.
- Banerjee, O., El Ghaoui, L., d'Aspremont A., 2008. Model selection through sparse maximum likelihood estimation. *J. Mach. Learn. Res.* 9, 485–516.
- Box, G.E., Tiao, G.C., 1977. A canonical analysis of multiple time series. *Biometrika* 64(2), 355–365.
- Chang, T.-J., Meade, N., Beasley, J.E., Sharaiha, Y.M., 2000. Heuristics for cardinality constrained portfolio optimization. *Comp. Oper. Res.*, 27, 1271–1302.
- Chen, L., Aihara, K., 1995. Chaotic simulated annealing by a neural network model with transient chaos. *Neur. Netw.* 8(6), 915–930.
- d'Aspremont, A., Banerjee, O., El Ghaoui, L., 2008. First-Order Methods for Sparse Covariance Selection. *SIAM J. Matr. Anal. Appl.* 30(1), 56–66.
- d'Aspremont, A., 2011. Identifying small mean reverting portfolios. *Quant. Financ.* 11(3), 351–364.
- Di Tollo, G., Rolli, A., 2008. Metaheuristics for the portfolio selection problem. *Int. J. Operations Res.* 5, 13–35.
- Elton, E.J., Gruber, M.J., Brown, S.J., Goetzmann W.N., 2007. *Modern Portfolio Theory and Investment Analysis*. Wiley, New York.
- Fama, E., French, K., 1988. Permanent and Temporary Components of Stock Prices. *The J. Pol. Econ.* 96(2), 246–273.
- Feiring, BR., Wong, WL, Poon, M., Chan Y.C., 1994. Portfolio selection in downside risk optimization approach – application to the Hong Kong stock market. *Int. J. Sys. Sci.* 25:1921–1929.
- Fernández, A., Gómez, S., 2007. Portfolio selection using neural networks. *Comp. Oper. Res.* 34, 1177–1191.
- Fogarasi, N., Levendovszky, J., 2011. A simplified approach to parameter estimation and selection of sparse, mean reverting portfolios. To appear in *Periodica Polytechnica Electrical Engineering and Computer Science*.
- Fogarasi, N., Levendovszky, J., 2012. Improved parameter estimation and simple trading algorithm for sparse, mean reverting portfolios. *Annales Univ. Sci. Budapest., Sect. Comp.* 37, 121–144.
- Geman, S., Geman, D., 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Patt. Anal. Mach. Intellig.* 6, 721–741.
- Ingber, L., 1993. Simulated annealing: Practice versus theory. *Math. Comput. Modelling* 18, 29–57.
- Ito, K., 1944. Stochastic integral. *Proc. Imperial Acad. Tokyo* 20, 519–524.
- Johnson, D.S., Aragon, C.R., McGeoch, L.A., Schevon, C., 1989. Optimization by simulated annealing: an experimental evaluation. Part I, graph partitioning. *Oper. Res.* 37, 865–892.
- Johnson, D.S., Aragon, C.R., McGeoch, L.A., Schevon, C., 1991. Optimization by simulated annealing: an experimental evaluation; Part II, graph coloring and number partitioning. *Oper. Res.* 39, 378–406.
- Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P., 1983. Optimization by simulated annealing. *Science* 220, 671–680.
- Konno, H., Yamazaki, H., 1991. Mean/absolute deviation portfolio optimization model and its applications to Tokyo Stock Market. *Manag. Sci.* 37, 519–531.
- Loraschi, A., Tettamanzi, A., Tomassini, M., Verda, P., 1995. Distributed genetic algorithms with an application to portfolio selection problems. In: Pearson, D.W., Steele, N.C., Albrecht, R.F. (Eds.), *Artificial Neural Nets and Genetic Algorithms*. Springer-Verlag, Vienna, pp. 384–387.
- Manzan, S. 2007. Nonlinear mean reversion in stock prices. *Quant. Qual. Anal. Soc. Sci.* 1(3), 1–20.

- Markowitz, H., 1952. Portfolio Selection. *J. Financ.* 7(1), 77–91.
- Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equations of state calculations by fast computing machines. *J. Chem. Phys.* 21(6), 1087–1092.
- Natarjan, B.K., 1995. Sparse approximate solutions to linear systems. *SIAM J. Comput.* 24(2), 227–234.
- Nourani, Y., Andersen, B., 1998. A comparison of simulated annealing cooling strategies. *J. Phys.* 31, 8373–8385.
- Ornstein, L.S., Uhlenbeck, G.E., 1930. On the theory of the Brownian motion. *Phys. Rev.* 36(5), 823.
- Poterba, J.M., Summers, L.H., 1988. Mean reversion in stock prices: Evidence and implications. *J. Financ. Econ.* 22(1), 27–59.
- Rothman, A., Bickel, P., Levina, E., Zhu, J., 2008. Sparse permutation invariant covariance estimation. *Electron. J. Stat.* 2, 494–515.
- Salamon, P., Sibani, P., Frost, R., 2002. Facts, Conjectures, and Improvements for Simulated Annealing. *SIAM Monographs on Mathematical Modeling and Computation*. Society for Industrial and Applied Mathematics, Philadelphia, PA.
- Streichert, F., Ulmer, H., Zell, A., 2003. Evolutionary algorithms and the cardinality constrained portfolio optimization problem. In: *Selected Papers of the International Conference on Operations Research*, Heidelberg, Germany, pp. 253–260.
- Yoshimoto, A., 1996. The mean-variance approach to portfolio optimization subject to transaction costs. *J. Oper. Res. Soc. Japan* 39, 99–117.