Annales Univ. Sci. Budapest., Sect. Comp. 37 (2012) 121–144

IMPROVED PARAMETER ESTIMATION AND SIMPLE TRADING ALGORITHM FOR SPARSE, MEAN REVERTING PORTFOLIOS

Norbert Fogarasi and János Levendovszky

(Budapest, Hungary)

Communicated by Imre Kátai

(Received January 15, 2012; revised March 13, 2012;
accepted March 19, 2012)

Abstract. We examine the problem of finding sparse, mean reverting portfolios based on multivariate historical time series. After mapping optimal portfolio selection into a generalized eigenvalue problem, two different heuristic algorithms are referenced for finding the solution in a subspace which satisfies the cardinality constraint. Having identified the optimal portfolio, we outline the known methods for finding the long-term mean and introduce a novel approach based on pattern matching. Furthermore, we present a simple convergence trading algorithm with a decision theoretic approach, which can be used to compare the economic viability of the different methods and test the effectiveness of our end-to-end process by extensive simulations on generated and historical real market data.

1. Introduction

Mean reversion, as a classic indicator of predictability in financial markets, has received a lot of attention over the last few decades. It has been shown

Key words and phrases: Mean reversion, sparse estimation, convergence trading, parameter estimation, VAR(1) model, covariance selection, financial time series.
https://doi.org/10.71352/ac.37.121

that equity excess returns over long horizons are mean-reverting and therefore contain an element of predictability [9, 13, 17]. Convergence trading, by estimating the parameters of mean reverting portfolios has also been proposed and studied in a number of previous research publications [3, 8].

In a recently published article, d'Aspremont posed the problem of finding mean-reverting portfolios which are sparse [6]. The Box–Tiao procedure [4] is used to extract cointegrated vectors by solving a generalized eigenvalue problem. Imposing a cardinality constraint is desirable for reducing transaction costs associated with convergence trading and for increasing the interpretability of the resulting portfolio. The simple greedy algorithm is introduced which is used as a benchmark for the more complex semidefinite relaxation method. In [10] we followed the same approach and proposed a simplified approach to model identification based on historical data as well as introducing two new methods for portfolio selection.

In this paper, we outline the above in some more detail and make two important, new contributions. Firstly, in Section 4.3, we propose a novel method, based on pattern matching, to estimate the long-term mean of an Ornstein–Uhlenbeck mean reverting process based on historical observations. We show that this estimate is more reliable than the mean estimates used in [6] and other related literature. Secondly, in Section 4.4, we introduce a simple convergence trading algorithm based on a decision theoretic approach, which allows the comparison, in terms of economic viability, of the different model identification, parameter estimation and portfolio selection algorithms. The construction of this allows an end-to-end process to be developed as shown in Figure 1.

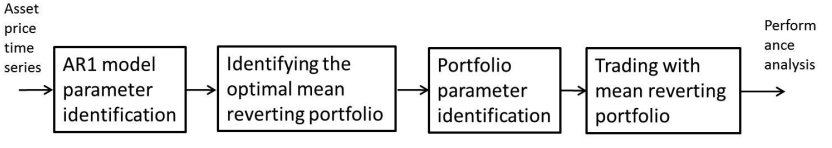

Figure 1. Overall framework for identifying and trading sparse mean reverting portfolios

The structure of the paper is as follows.

- In Section 2, after giving a formal presentation of the problem, we outline some previously suggested methods for portfolio selection, following the treatment of [6] and [10].

- In Section 3, we outline how the VAR(1) model parameters can be identified based on historical data, following the methods outlined in [10] and making some minor corrections to them.

- In Section 4, we present a novel, decision theoretic approach to trading where an estimate is first obtained for the long-term mean of the process and the agent makes hypotheses about the state of the process based on this estimate.
- In Section 5, 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 eight different maturity U.S. swap rates, and the daily close prices of stocks comprising the S&P 500 index.
- Finally, in Section 6 some conclusions are drawn and directions for future research are outlined.

2. Sparse, mean reverting portfolio selection

In this section, the model is described together with the foundations of identifying mean reverting portfolios. Our approach follows the one published in [6], and in Section 2.3 we present the methods published in [10].

2.1. Mean reverting portfolios

The problem we examine in this paper is the selection of sparse, mean reverting portfolios which follow the Ornstein–Uhlenbeck process [15]. Let $s_{i,t}$ denote the price of asset i at time instant t, where $i = 1, \ldots, n$ and $t = 1, \ldots, m$ are positive integers. We form a portfolio, of value p_t of these assets with coefficients x_i, and assume that asymptotically, as the size of the time increment tends to zero, it follows an Ornstein–Uhlenbeck process given by:

$$dp_t = \lambda\left(\mu - p_t\right) dt + \sigma dW_t, \tag{2.1}$$

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

Using the Ito-Doeblin formula [12] to solve this, we get:

$$p(t) = p\left(0\right) e^{-\lambda t} + \mu\left(1 - e^{-\lambda t}\right) + \int_0^t \sigma e^{-\lambda(t-s)} dW\left(s\right), \tag{2.2}$$

which implies that

$$E[p(t)] = p\left(0\right) e^{-\lambda t} + \mu\left(1 - e^{-\lambda t}\right), \tag{2.3}$$

and asymptotically

$$\lim_{t\to\infty} p(t) \sim N\left(\mu, \sqrt{\frac{\sigma^2}{2\lambda}}\right). \tag{2.4}$$

For trading, the mean reverting coefficient (λ) is a key parameter, as it determines how fast the process gets back 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 it contains a minimum amount of uncertainty around the mean. Therefore, we will be concerned with finding sparse portfolios which are optimal in the sense that they maximize λ.

2.2. Mean reverting portfolio as a generalized eigenvalue problem

In this section we view the 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 $i = 1, \ldots, n$ and $t = 1, \ldots, m$ are positive integers and assume that $\mathbf{s}_t^T = (s_{1,t}, \ldots, s_{n,t})$ is subject to a first order vector autoregressive process, VAR(1), defined as follows:

$$\mathbf{s}_t = \mathbf{A}\mathbf{s}_{t-1} + \mathbf{W}_t, \tag{2.5}$$

where $\mathbf{A}$ is an $n \times n$ matrix and $\mathbf{W}_t \sim N(0, \sigma_W I)$ are i.i.d. noise terms, independent of $\mathbf{s}_{t-1}$, for some $\sigma_W > 0$.

One can introduce a portfolio vector $\mathbf{x}^T = (x_1, \ldots, x_n)$, where component x_i denotes the amount of asset i held. In practice, assets are traded in discrete units, so $x_i \in \{0, 1, 2, \ldots\}$, 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

$$\mathbf{x}^T\mathbf{s}_t = \mathbf{x}^T\mathbf{A}\mathbf{s}_{t-1} + \mathbf{x}^T\mathbf{W}_t. \tag{2.6}$$

Following the treatment in [6] and [4], we define the *predictability* of the portfolio as

$$\nu(\mathbf{x}) := \frac{\mathrm{var}(\mathbf{x}^T\mathbf{A}\mathbf{s}_{t-1})}{\mathrm{var}(\mathbf{x}^T\mathbf{s}_t)} = \frac{E(\mathbf{x}^T\mathbf{A}\mathbf{s}_{t-1}\mathbf{s}_{t-1}^T\mathbf{A}^T\mathbf{x})}{E\left(\mathbf{x}^T\mathbf{s}_t\mathbf{s}_t^T\mathbf{x}\right)}, \tag{2.7}$$

provided that $E(\mathbf{s_t}) = 0$, so the asset prices are normalized. The intuition behind this portfolio predictability is that the greater this ratio is, 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 portfolios mean reversion parameter λ in (2.1). Maximizing this expression will yield the following optimization problem for finding the best portfolio vector $\mathbf{x_{opt}}$:

$$\mathbf{x_{opt}} = \arg\max_{\mathbf{x}} \nu(\mathbf{x}) = \arg\max_{\mathbf{x}} \frac{\mathbf{x}^T\mathbf{A}\mathbf{G}\mathbf{A}^T\mathbf{x}}{\mathbf{x}^T\mathbf{G}\mathbf{x}}, \tag{2.8}$$

where $\mathbf{G}$ is the covariance matrix of the stationary process $\mathbf{s_t}$. Based on (2.8) we can see that the problem is equivalent to finding the eigenvector corresponding to the maximum eigenvalue in the following generalized eigenvalue problem [6]:

$$\mathbf{A}\mathbf{G}\mathbf{A}^T\mathbf{x} = \alpha\mathbf{G}\mathbf{x}. \tag{2.9}$$

α can be obtained by solving the following equation:

$$\det(\mathbf{A}\mathbf{G}\mathbf{A}^T - \alpha\mathbf{G}) = 0. \tag{2.10}$$

Note that this can be transformed into a traditional eigenvalue problem by introducing the variable $\mathbf{u} := \mathbf{G}^{1/2}\mathbf{x}$ so that we have

$$\mathbf{G}^{-1/2}\mathbf{A}\mathbf{G}\mathbf{A}^T\mathbf{G}^{-1/2}\mathbf{u} = \alpha\mathbf{u}. \tag{2.11}$$

2.3. Sparse portfolio selection

In the previous section we outlined how to select a portfolio which maximizes predictability by solving a generalized eigenvalue problem. However, we seek the optimal portfolio vector exhibiting mean reverting property under a sparseness constraint. Adding this to (2.8), we can formulate our constrained optimization problem as follows:

$$\mathbf{x_{opt}} = \arg\max_{\mathbf{x}\in\mathbf{R}^n, card(\mathbf{x})\leq L} \frac{\mathbf{x}^T\mathbf{A}\mathbf{G}\mathbf{A}^T\mathbf{x}}{\mathbf{x}^T\mathbf{G}\mathbf{x}} \tag{2.12}$$

where *card* denotes the number of non-zero components, and L is a given positive 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 shows that this problem is equivalent to the subset selection problem, which is proven to be NP-hard [14]. 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 [10]. We also describe two polynomial time heuristic algorithms for an approximate solution of this problem.

2.3.1. Exhaustive search method

The brute force approach of constructing all $\left(\frac{n!}{L!(n-L)!}\right)$ 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 a benchmark computed off-line, this method can provide a very useful basis of comparison. Indeed, for the practical applications considered in [6] (selecting sparse portfolios of $n = 8$ U.S. swap rates and $n = 14$ FX rates), this method is fully applicable and can be used to see the level of sub-optimality of other proposed methods.

2.3.2. Greedy method

A reasonably fast heuristic algorithm, first presented by d'Aspremont in [6] is the so-called greedy method which we will briefly explain. 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. When $k = 1$, we set $i_1 = \underset{j\in[1,n]}{\arg\max} \frac{\left(\mathbf{AGA}^T\right)_{jj}}{\mathbf{G}_{jj}}$. Suppose now that we have a good approximate solution with support set I_k given by $(\mathbf{x})_k = \underset{\left\{\mathbf{x}\in R^n : x_{I_k^C}=0\right\}}{\arg\max} \frac{\mathbf{x}^T\mathbf{AGA}^T\mathbf{x}}{\mathbf{x}^T\mathbf{Gx}}$, 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

$$i_{k+1} = \underset{i\in I_k^C}{\arg\max} \max_{\left\{\mathbf{x}\in R^n : x_{J_i}=0\right\}} \frac{\mathbf{x}^T\mathbf{AGA}^T\mathbf{x}}{\mathbf{x}^T\mathbf{Gx}}, \text{where } J_i = I_k^C \setminus \{i\}, \tag{2.13}$$

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

2.3.3. 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 (2.8) by solving the generalized eigenvalue problem in (2.9). 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 $\mathbf{x}_{\mathbf{opt}}^{trunc}$ 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 heaviest 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.

3. Estimation of model 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 the historical observations of the random process. There is vast literature on the topic of parameter estimation of VAR(1) processes; recent research has focused on sparse and regularized covariance estimation [1, 5, 18]. However, our approach will be to gain a dense estimate for $\mathbf{G}$ which best describes the observed historical data and to deal with dimensionality reduction by the apparatus outlined in Section 2. Another important objective that we pose for the parameter fitting is to provide a measure of goodness of fit of the real time series to the VAR(1) model, which we can use in the portfolio selection and trading parts of our overall algorithm.

3.1. Estimation of matrix A

We recall from our earlier discussion that we assume that $\mathbf{s_t}$ follows a first order autoregressive process as in equation (2.5). We first observe that if the number of assets n is greater than or equal to m, the length of the observed time

series, then $\mathbf{A}$ can be estimated by simply solving exactly the linear system of equations:

$$\hat{\mathbf{A}}\mathbf{s}_{t-1} = \mathbf{s}_t. \tag{3.1}$$

Note that if $n > m$, this system is underdetermined, so there are infinitely many solutions. In this case, we considered the subsystem consisting of the first m assets to obtain a unique solution. This gives a perfect VAR(1) fit for our time series for cases where we have a large portfolio of potential assets in relation to the amount of data observed (e.g. considering daily close prices over a 1-month period of all 500 stocks which make up the S&P 500 index), from which a sparse mean-reverting portfolio is to be chosen.

In most of the applications, however, the length of the available historical time series is greater than the number of assets considered, so typically $m > n$ and thus (3.1) is overdetermined, and $\mathbf{A}$ is estimated using, for example, least squares estimation techniques, as in

$$\hat{\mathbf{A}} = \arg\min_{\mathbf{A}} \sum_{t=2}^{m} \|\mathbf{s}_t - \mathbf{A}\mathbf{s}_{t-1}\|^2 \tag{3.2}$$

where $\|\cdot\|$ denotes the Euclidian norm.

Equating to zero the partial derivatives of the above expression with respect to each element of the matrix $\mathbf{A}$, we obtain the following system of equations:

$$\sum_{k=1}^{n} \hat{\mathbf{A}}_{i,k} \sum_{t=2}^{m} s_{k,t-1} s_{j,t-1} = \sum_{t=2}^{m} s_{i,t} s_{j,t-1} \ \forall i, j = 1, \ldots, n. \tag{3.3}$$

Solving for $\hat{\mathbf{A}}$ and switching back to matrix-vector notation, we obtain

$$\hat{\mathbf{A}} = \left(\sum_{t=2}^{m} \mathbf{s}_t \mathbf{s}_{t-1}^T \right) \left(\sum_{t=2}^{m} \mathbf{s}_{t-1} \mathbf{s}_{t-1}^T \right)^{+}, \tag{3.4}$$

where $\mathbf{M}^{+}$ denotes the Moore–Penrose pseudoinverse of matrix $\mathbf{M}$. Note that the Moore-Penrose pseudoinverse is used rather than regular matrix inversion, due to possible singularity of the sum of outer products in the above expression.

3.2. Estimation of the covariance matrix of W

Assuming that the noise terms in equation (2.5) are i.i.d. with $\mathbf{W}_t \sim$ $\sim N(0, \sigma_W I)$ for some $\sigma_W > 0$, we obtain the following estimate for σ_W using $\hat{\mathbf{A}}$ from (3.4):

$$\hat{\sigma}_W = \sqrt{\frac{1}{n\,(m-1)} \sum_{t=2}^{m} \left\| \mathbf{s}_t - \hat{\mathbf{A}}\mathbf{s}_{t-1} \right\|^2}. \tag{3.5}$$

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

$$\hat{\mathbf{K}} = \frac{1}{m-1} \sum_{t=2}^{m} (\mathbf{s}_t - \hat{\mathbf{A}}\mathbf{s}_{t-1})(\mathbf{s}_t - \hat{\mathbf{A}}\mathbf{s}_{t-1})^T. \tag{3.6}$$

This noise covariance estimate will be used below in the estimation of the covariance matrix.

3.3. Estimation of covariance matrix G

There are two independent approaches to estimating the covariance of a VAR(1) process based on a sample time series. On the one hand, the sample covariance and various maximum likelihood-based regularizations thereof can provide a simple estimate and have been studied extensively for the more general case of multivariate Gaussian distributions [1, 5, 7, 18]. In our treatment, we take the approach of using the sample covariance matrix directly without any reguralization or finding structure via maximum likelihood, as sparsifying and structure finding will be left for the apparatus of the sparse portfolio selection, explained in Section 2. As such, we will define $\hat{\mathbf{G}}_\mathbf{1}$ as the sample covariance defined as

$$\hat{\mathbf{G}}_1 := \frac{1}{m-1} \sum_{t=1}^{m} (\mathbf{s}_t - \bar{\mathbf{s}})\,(\mathbf{s}_t - \bar{\mathbf{s}})^T, \tag{3.7}$$

where $\bar{\mathbf{s}}$ is the sample mean vector of the assets defined as

$$\bar{\mathbf{s}} := \frac{1}{m} \sum_{t=1}^{m} \mathbf{s}_t. \tag{3.8}$$

On the other hand, starting from the definition of the VAR(1) process in (2.5) and assuming the more general case that the terms of $\mathbf{W}_t$ are correlated with covariance matrix $\mathbf{K}$, we must have

$$\mathbf{G}_t = \mathbf{A}\mathbf{G}_{t-1}\mathbf{A}^T + \mathbf{K}, \tag{3.9}$$

which implies that in the stationary case

$$\mathbf{G} = \mathbf{A}\mathbf{G}\mathbf{A}^T + \mathbf{K}. \tag{3.10}$$

Having estimated $\mathbf{A}$ and $\mathbf{K}$, as in the previous sections, this is a Lyapunov equation with unknown $\mathbf{G}$ and can be solved analytically to obtain another covariance estimate $\hat{\mathbf{G}}_\mathbf{2}$. This estimate exists, and is unique, in case no two

eigenvalues of $\mathbf{A}$ have a product equal to 1. In this case, we note that since $\mathbf{K}$ is symmetric, $\hat{\mathbf{G}}_\mathbf{2}$ is also guaranteed to be symmetric. Furthermore, studies pertaining to the solution of Lyapunov equations show that if $\mathbf{K}$ is positive definite and $\mathbf{A}$ has all its eigenvalues inside the unit disk, the solution of the equation is also positive definite [2, 16]. However, it is possible that our estimate of $\mathbf{A}$ does not satisfy this condition, in which case $\hat{\mathbf{G}}_\mathbf{2}$ is not necessarily positive definite and, as such, it may not be a permissible covariance estimate. In order to overcome this issue, in case the solution of the Lyapunov equation is non-positive-definite, the following iterative numerical method can be used to obtain a permissible covariance estimate $\hat{\mathbf{G}}_\mathbf{2}$:

$$\mathbf{G}(k+1) = \mathbf{G}(k) - \delta(\mathbf{G}(k) - \mathbf{A}\mathbf{G}(k)\mathbf{A}^T - \mathbf{K}), \tag{3.11}$$

where δ is a constant between 0 and 1, $\mathbf{G}(i)$ is the covariance matrix estimate on iteration i. Provided that the starting point for the numerical method, $\mathbf{G}(0)$, is positive definite (e.g. 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. It can also be seen that if we select δ to be sufficiently small, this numerical estimate will converge to the solution of the Lyapunov equation in (3.10), in case that is positive definite. In Section 4, some numerical results are presented, which show that for generated VAR(1) data, these two covariance estimates are equivalent, provided that appropriately sized sample data is available for the given level of noise. However, for historical financial data, the two estimates can differ significantly. A large difference between the two estimates indicates a poor fit of the data to the VAR(1) model, hence we can define the following measure of model fit:

$$\beta := \left\| \hat{\mathbf{G}}_1 - \hat{\mathbf{G}}_2 \right\|, \tag{3.12}$$

where $\|\mathbf{M}\|$ denotes the largest singular value of the matrix $\mathbf{M}$. This goodness of model fit parameter will be used during trading to set our level of confidence that the resulting portfolio indeed follows a mean reverting process and therefore our convergence trading strategy will be profitable.

4. Trading as a decision theoretic problem

Having identified the portfolio with maximal mean reversion, satisfying 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 we ought to

buy, above the mean and therefore likely to fall so we ought to short sell, or close to the mean in an already stationary state, in which case no profitable action is available. In order to formulate this as a decision theoretic problem, we first need to estimate the mean value of the portfolio. We present three different methods of identifying μ from the observations of the optimal portfolios time series, $\{\mathbf{p}_t, t = 1, \ldots, m\}$.

4.1. Sample mean estimator

The simplest estimate of the mean of the process based on the time series observation of the historical values of the portfolio is the one given by the sample mean, defined as follows:

$$\hat{\mu}_1 := \frac{1}{m} \sum_{t=1}^{m} \mathbf{p}_t \tag{4.1}$$

This estimate is akin to a measure used by technical traders and gives a poor estimate of the long-term mean of the mean-reverting process, in case it is in its transient state, trending towards the true long-term mean. However, it has served as a very useful benchmark for evaluating other methods.

4.2. Mean estimation via least squares linear regression

Following the treatment of [19], we can perform a linear regression on pairwise consecutive samples of $\mathbf{p}$ as follows:

$$\mathbf{p}_{t+1} = a\mathbf{p}_t + b + \varepsilon \tag{4.2}$$

for $t = 1, \ldots, m-1$. Then, an estimate of μ can be obtained from the coefficients of the regression as follows:

$$\hat{\mu}_2 := \frac{b}{1-a} \tag{4.3}$$

Note that there is an instability in this estimate for $a = 1$, in which case $\hat{\lambda} = -\ln a = 0$ and hence the process is deemed not to be mean reverting based on the observed sample.

4.3. Mean estimation via pattern matching

In our description of this novel mean estimation technique, we start from the definition of the discrete Orstein–Uhlenbeck process in equation (2.1) and

consider its continuous solution given in equation (2.2). Disregarding the noise to consider only the expected value of the process, we can rewrite equation (2.3) as follows:

$$\mu(t) = \mu + (\mu(0) - \mu)\, e^{-\lambda t} \tag{4.4}$$

where $\mu(t) = E[p(t)]$ is a 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.

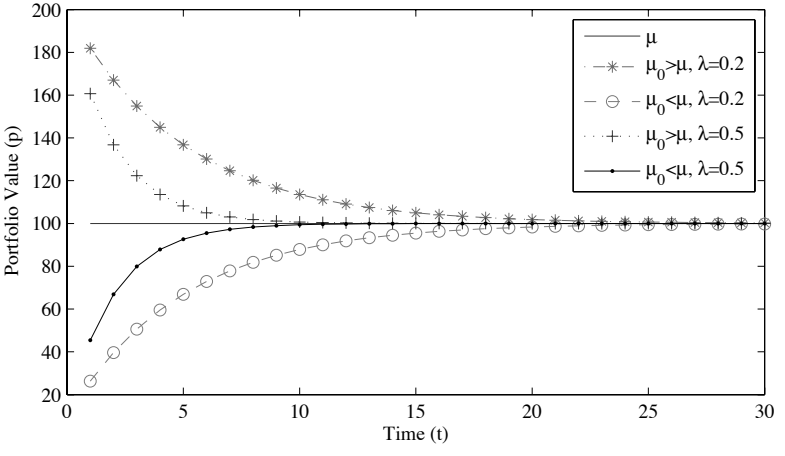

Figure 2. Some sample patterns of $\mu(t)$

In Figure 2, we show some typical tendencies of $\mu(t)$ for various relative values of $\mu(0)$ and μ. The idea behind pattern matching is to observe historical time series values of $\mathbf{p}_t = (p(t-1), p(t-2), \ldots, p(0))$ and use maximum likelihood estimation techniques to determine which of the patterns $\mu(t)$ matches best. We first observe that $\mathbf{p}_t = (p(t-1), p(t-2), \ldots, p(0))$ is subject to a multivariate Gaussian distribution, the density function of which is given by

$$D(\mathbf{p}_t) = \frac{1}{\sqrt{(2\pi)^t \det(\mathbf{U})}} e^{-\frac{1}{2}(\mathbf{p}_t - \mu_t)^T \mathbf{U}^{-1} (\mathbf{p}_t - \mu_t)} \tag{4.5}$$

where $\mathbf{U}_{ij} := \operatorname{cov}(p(t-i), p(t-j)) = \frac{\sigma^2}{2\lambda}\left(e^{-\lambda(j-i)} - e^{-\lambda(2t-i-j)}\right)$ is the time-covariance matrix of $\mathbf{p}_t$ for $j \geq i$ and $\mu_t = (\mu(t-1), \ldots, \mu(0))$. The idea is now to consider class Φ of all t-length vectors, μ_t which consist of sequences satisfying (4.4). The maximum likelihood pattern matching estimate can now

be formulated as follows:

$$\hat{\mu}_{\mathbf{t}} = \arg\max_{\mu_t \in \Phi} \frac{1}{\sqrt{(2\pi)^t \det(\mathbf{U})}} e^{-\frac{1}{2}(\mathbf{p}_t - \mu_t)^T \mathbf{U}^{-1} (\mathbf{p}_t - \mu_t)}, \tag{4.6}$$

which is equivalent to

$$\hat{\mu}_{\mathbf{t}} = \arg\min_{\mu_t \in \Phi} \left\{ \mu_t^T \mathbf{U}^{-1} \mu_t - 2\mu_t^T \mathbf{U}^{-1} \mathbf{p}_t \right\}. \tag{4.7}$$

Using the definition of μ_t and equation (4.4) to expand (4.7), then taking the derivative of this quadratic expression with respect to μ, equating to zero and solving, we obtain the following closed-form estimate for the long term mean:

(4.8)
$$\hat{\mu}_3 := \frac{\sum_{i=1}^{t} \sum_{j=1}^{t} \left(\mathbf{U}^{-1}\right)_{i,j} \left[\mu(0) \left(2e^{-\lambda(i+j)} - e^{-\lambda i} - e^{-\lambda j}\right) - 2\mathbf{p}_j \left(e^{-\lambda i} - 1\right)\right]}{\sum_{i=1}^{t} \sum_{j=1}^{t} 2 \left(\mathbf{U}^{-1}\right)_{i,j} \left(e^{-\lambda(i+j)} - e^{-\lambda i} - e^{-\lambda j} + 1\right)}.$$

Having observed the historical portfolio values in $\mathbf{p}_t$ and $\mu(0) = \mathbf{p}_0$, we can substitute λ computed via linear regression [19] and $\sigma = \mathbf{x}^T \hat{\mathbf{K}} \mathbf{x}$ and use this equation to obtain an estimate for the long term mean.

4.4. A simple convergence trading strategy

Our task is now to define an algorithm which results in a trading strategy based on buying sparse mean reverting portfolios below the estimated mean, and selling above the estimated mean. A simple model of trading in which we restrict ourselves to holding only one portfolio at a time can be perceived as a walk in a binary state-space, depicted in Figure 3.

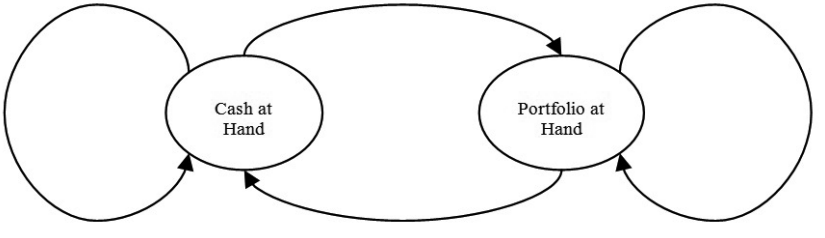

Figure 3. Trading as a walk in a binary state-space

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

or $\mu(t) \geq \mu$ based on observing the samples $\{p(t) = \mathbf{x}^T\mathbf{s}(t), t = 1, \ldots, m\}$. This verification can be perceived as a decision theoretic problem, since direct observations of $\mu(t)$ are not available. If process $p(t)$ is in stationary state, then the samples $\{p(t), t = 1, \ldots, m\}$ are generated by a 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

$$\int_{\mu-\alpha}^{\mu+\alpha} \frac{1}{\sqrt{2\pi\sigma^2/2\lambda}} e^{-\frac{(u-\mu)^2}{\sigma^2/\lambda}} du = 1 - \varepsilon \tag{4.9}$$

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 our trading strategy can be summarized as follows:

- If the observed sample $p(t) < \mu - \alpha$, then we accept the hypothesis that $\mu(t) < \mu$. The probability of error associated with this hypothesis is $\int_{-\infty}^{\mu-\alpha} \frac{1}{\sqrt{2\pi\sigma^2/2\lambda}} e^{-\frac{(u-\mu)^2}{\sigma^2/\lambda}} du = \varepsilon/2$. In this case, we buy the portfolio if we have cash at hand, and we hold the portfolio if we already have one.
- If the observed sample $p(t) > \mu + \alpha$, then we accept the hypothesis that $\mu(t) > \mu$. The probability of error associated with this hypothesis is $\int_{\mu+\alpha}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2/2\lambda}} e^{-\frac{(u-\mu)^2}{\sigma^2/\lambda}} du = \varepsilon/2$. In this case, if we have a portfolio, we sell it, otherwise we perform no action.
- If the observed sample $p(t) \in [\mu - \alpha, \mu + \alpha]$, then we accept the hypothesis that $\mu(t) = \mu$. The probability of error associated with this hypothesis is $1 - \int_{\mu-\alpha}^{\mu+\alpha} \frac{1}{\sqrt{2\pi\sigma^2/2\lambda}} e^{-\frac{(u-\mu)^2}{\sigma^2/\lambda}} du = \varepsilon$. Then sell if a portfolio is held, or perform no action if only cash is held.

We can now extend Figure 3 to present a complete flowchart for our proposed simple trading strategy in Figure 4.

4.5. Technical issues of trading

Having worked out the mathematical and algorithmic foundations for mean reverting trading, in this section some of the technical questions are examined which arise during trading. One important question is whether to spend all our available cash at the time we identify a mean-reverting portfolio which is

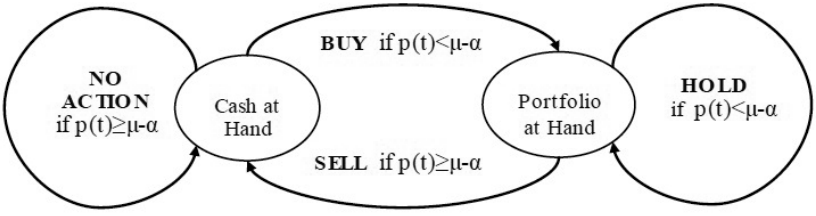

Figure 4. Flowchart for simple convergence trading of mean reverting portfolios

below its long-term mean level or to use some *cash management* strategy to only spend part of our current holdings. Indeed, we have two parameters, β and λ which we can use to help establish our level of confidence in the profitability of the portfolio. For simplicity, we have used the approach of spending all of our available cash each time we identify an appropriate portfolio. Furthermore, the simple convergence trading strategy we described in the previous section allows holding only one portfolio at a time. We could enhance this to have the ability to hold a number of portfolios at once, continually estimating the remaining profitability of each and comparing this to the best available portfolio in the market at each time step. Finally, in our numerical results, we have assumed that we have the ability to buy and sell assets without any transaction costs and we also have the ability to short sell. In order to make our results more realistic, we could introduce a bid-ask spread or a more sophisticated order book model to estimate the true profitability of these methods in the presence of market friction.

5. Performance analysis

In this section, we will review some results of the numerical tests which were produced for validating the methods outlined earlier. We first tested the model parameter estimation methods on generated data to show their viability and observe their limitations. We then produced a simple convergence trading engine which has proven to perform extremely well on the generated VAR(1) data, validating the theoretical value of our methodology. Finally, we ran our parameter estimation, portfolio optimization and convergence trading strategies on various historical financial time series.

5.1. Performance of parameter estimation

In order to test the relative accuracy of the three different estimates of μ presented in Section 4, we used simulated Ornstein–Uhlenbeck processes. Note that for simulation purposes, as shown in [11], instead of equation (2.1), it is more appropriate to use the discrete version of the differential equation, given by

$$p(t) = e^{-\lambda\Delta t}p(t-1) + \left(1 - e^{-\lambda\Delta t}\right)\mu + \sigma\sqrt{\frac{(1-e^{-2\lambda\Delta t})}{2\lambda}}dW(t) \tag{5.1}$$

As we can see in Figure 5, for a reasonably sized and sufficiently large sample, the linear regression method produces a smaller Mean Squared Error (MSE) than the other two estimates.

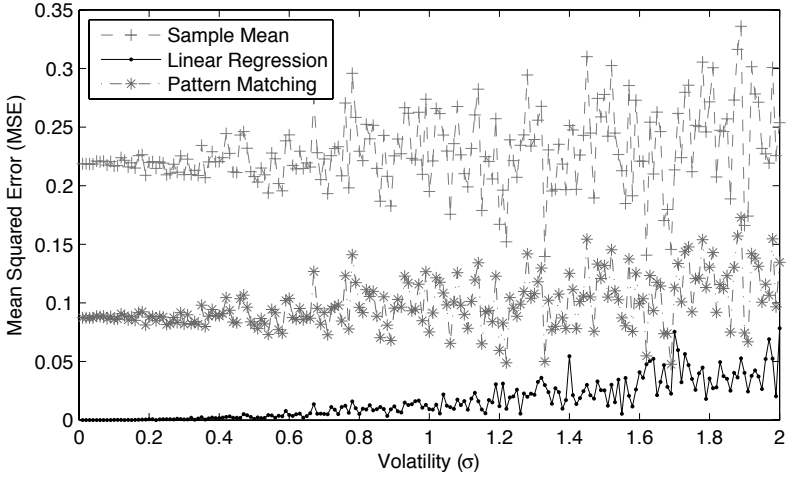

Figure 5. MSE of each μ estimate as a function of σ. Fixing $\lambda = 1$, $\mu_0 = 0$ and sample size of 100, we generated noisy sequences as per (5.1) with $\mu = 1, 2, \ldots, 50$ and $\sigma = 0.01, 0.02, \ldots, 2.00$

On the other hand, Figure 6 shows that the linear regression method is unstable for smaller sample size (in these experiments, we used 20) and for small values of λ. Indeed, we noted this instability in the description of this estimator and for smaller sample sizes, and for smaller values of λ we are more likely to encounter this. Note also that as we increase λ, the pattern matching estimate produces increasing MSE with the limit of producing identical MSE to the sample mean method.

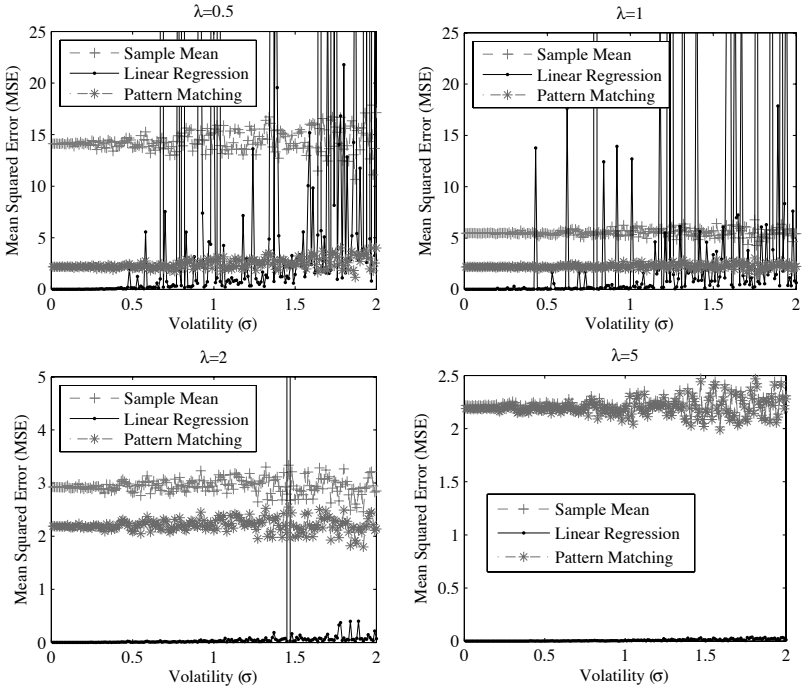

Figure 6. MSE of each μ estimate as a function of σ. $\lambda = 0.5, 1, 2, 5$, $\mu_0 = 0$ and sample size of 20

5.2. Performance of portfolio selection and trading on generated data

In order to compare the portfolio selection methods outlined in Section 2, we again generated VAR(1) data with random $\mathbf{A}$ matrix and noise with covariance matrix $\mathbf{K}$. 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. We found that in a large number of cases, the greedy method yields portfolios whose mean reversion coefficient is close to the theoretical best, produced by the Exhaustive method. Having run 1000 simulations on independently generated VAR(1) data, we observed that the exhaustive method produced better mean reversion coefficients than the greedy method in 59.3% of the cases and outperformed the truncation method in 99.8% of the cases. The greedy method produced lambdas which were, on average, 2.26% worse than the optimal lambda found by the exhaustive search while truncation produced lambdas which were 7.34% worse on average. We also

found a number of examples where the greedy method yielded portfolios whose mean reversion coefficients were significantly below the theoretical optimum and where other polynomial time heuristic methods could be found to improve upon this (for an example, see Figure 7).

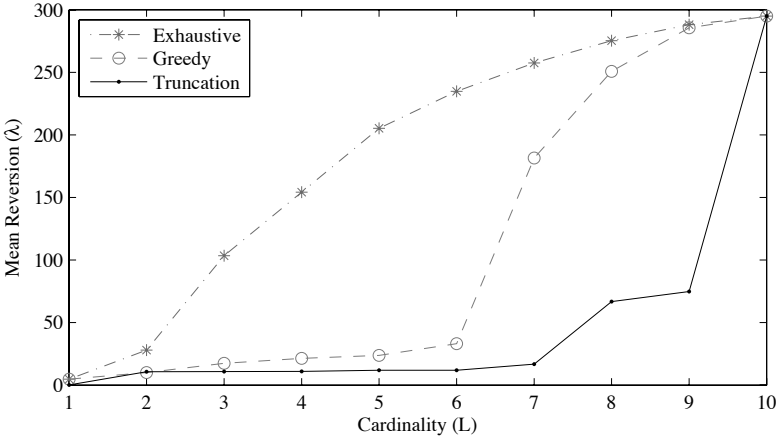

Figure 7. Comparison of portfolio selection methods of various cardinality on n=10-dimensional generated VAR(1) data

In order to examine the runtime of the portfolio selection algorithms, we ran repeated simulations of selecting sparse portfolios from n assets for all cardinalities from 1 to n and plotting the total CPU time taken against n for each of the proposed methods (Figure 8). We observe that the truncation method is the fastest, taking less than 3 seconds on a Pentium 4, 3.80 GHz machine to select all 100 subportfolios of 100 assets. The same took over 30 seconds for the greedy method, which suggests that while the truncation method is well suited for real-time algorithmic trading where sub-second algorithms are required for a given cardinality, the greedy method could also be used for intraday electronic trading. The exhaustive search could also be viable for intraday trading for asset populations of 20 or under, but the run times become hours on our test hardware beyond 22 assets.

In order to prove the economical viability of our methodology, we implemented the simple convergence trading strategy, outlined in Section 4.4. We then generated $n = 10$-dimensional VAR(1) sequences of length 270 of which the first 20 time steps were used to estimate the model parameters and find the optimal portfolio of size $L = 5$ using the different methods, and the following 250 (approximate number of trading days in a year) were used to trade the portfolio. Running each of the algorithms on 2000 different time series and

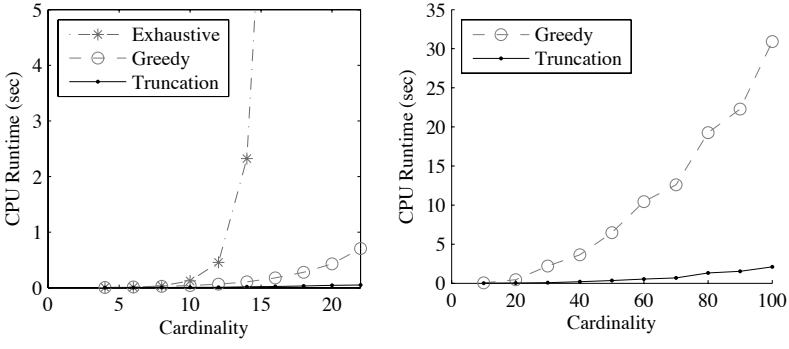

Figure 8. CPU runtime (in seconds) versus total number of assets n, to compute a full set of sparse portfolios, with cardinality ranging from 1 to n, using the different algorithms

fixing $\varepsilon = 0.02$ as the error level, 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, grew exponentially over time, in some cases hitting as much as \$14 million over the 250 trading days. Figure 9 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. Figure 10 shows the histogram of trading gains achieved by the truncation method. All three methods produced average profits of the same order of magnitude and the distribution of trading gains was 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 3 times those produced by the greedy method. 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 goodness of 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.

5.3. 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 until July 2010. We use the methods of Section 3

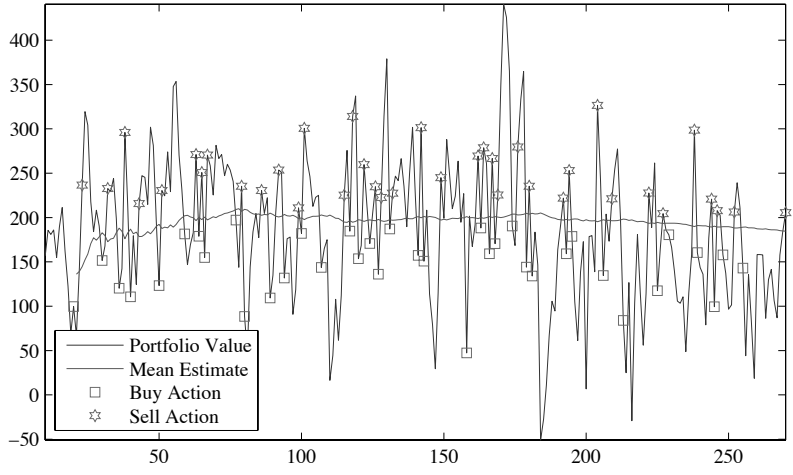

Figure 9. Typical example of convergence trading over 250 time steps on a sparse mean-reverting portfolio. Weighted mix of L=5 assets were selected from n=10-dimensional generated VAR(1) process by simulated annealing during the first 20 time steps. A profit of \$1440 was achieved with an initial cash of \$100 having made 85 transactions

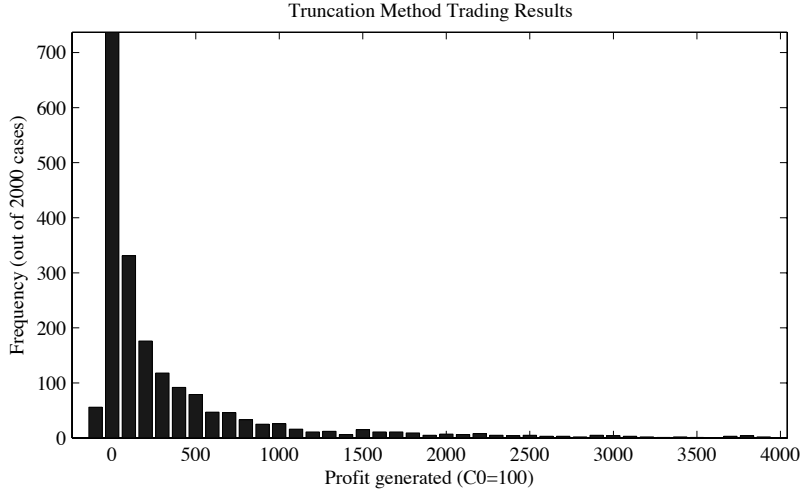

Figure 10. Histogram of profits achieved over 2000 different generated VAR(1) series. Note that in 30 cases the profit exceeded \$4000 (not shown on the histogram) with maximum profit of \$12.6 million reached

to estimate the model parameters on a sliding window of 8 observations and select sparse, mean reverting portfolios using the algorithms of Section 2. Using the simple convergence trading methodology of 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). Detailed results are presented in Figure 11.

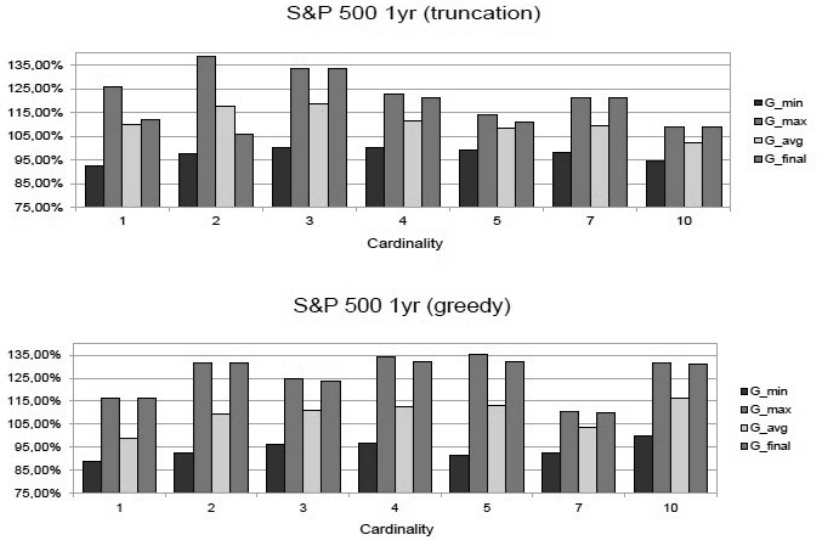

Figure 11. Comparison of minimum return, maximum return, average return, and final return over a trading horizon of 1 year, on S&P 500 historical data of the truncation and greedy methods for sparse mean-reverting portfolios of size 1, 2, 3, 4, 5, 7, and 10

Next, we studied U.S. swap rate data for maturities 1Y, 2Y, 3Y, 4Y, 5Y, 7Y, 10Y, and 30Y from 1995 to 2010. Similarly to the S&P 500 analysis, we used sliding time windows of 8 observations to estimate model parameters and applied the simple convergence trading methodology on heuristically selected mean reverting portfolios. Due to poorer model fit, we observed more moderate returns of 15-45% over the 15 years for portfolio cardinalities of 4 to 6. The methods generated best results for cardinalities 4 and 5 (see Figure 12).

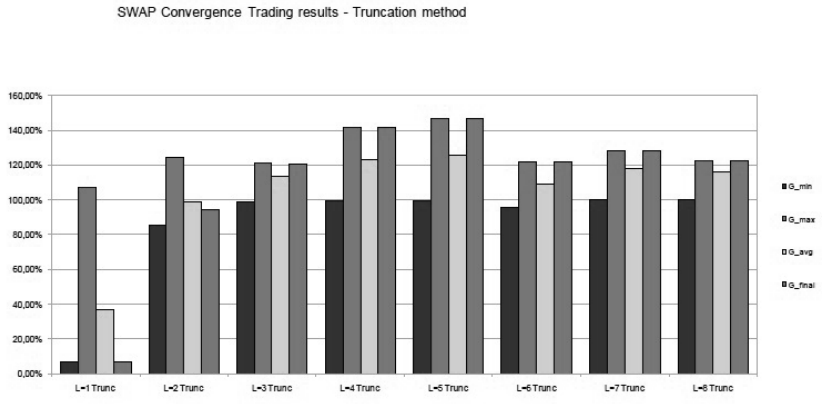

Figure 12. Comparison of minimum, maximum, average and final return on historical U.S. swap data of the truncation method for sparse, mean reverting portfolios of cardinality 1 to 8

6. Conclusions and directions for future research

In this paper we have examined the problem of selecting optimal sparse mean reverting portfolios based on observed and generated historical time series. Having identified the portfolio, we have proposed a novel way to estimate its long term mean, based on pattern matching. We have empirically shown the relative stability and accuracy of this mean estimate, in comparison to the other methods used in the literature on a variety of generated data. Furthermore, we have introduced a simple trading algorithm based on a decision theoretic approach and have shown its economic viability on both generated and real historical data. Having examined the relationship between λ and the profits generated by the simple trading method, we observe that the relationship is not trivial. Further study of this will provide insight into what other parameters should be considered during the optimal portfolio selection. It appears that whilst maximizing λ gives rise to an elegant formulation due to the mapping to the generalized eigenvalue problem, this method does not maximize the subsequent profits produced. Finally, more sophisticated trading methods, involving multiple portfolios and better cash management as indicated in Section 4.4 could be developed to further enhance trading performance and they could be analyzed in the presence of transaction costs, short-selling costs and order book.

Acknowledgements. The authors would like to acknowledge Róbert Sipos for useful discussions and his help in running many of the simulations. We also thank the referee for many valuable comments.

References

[1] **Banerjee, O., L. El Ghaoui and A. d'Aspermont,** Model selection through sparse maximum likelihood estimation, *Journal of Machine Learning Research*, **9** (2008), 485–516.

[2] **Barraud, A.Y.,** A numerical algorithm to solve A XA - X = Q, *IEEE Trans. Auto. Contr.*, **22** (1977) , 883–885.

[3] **Boguslavsky, M. and E. Boguslavskaya,** Arbitrage under power, *Risk Magazine*, **6** (2004), 69–73.

[4] **Box, G.E. and G.C. Tiao,** A canonical analysis of multiple time series, *Biometrika*, **64(2)** (1977), 355–365.

[5] **D'Aspremont, A, O. Banerjee and L. El Ghaoui,** First-order methods for sparce covariance selection, *SIAM Journal on Matrix Analysis and its Applications*, **30(1)** (2008), 56–66.

[6] **D'Aspremont, A,** Identifying small mean-reverting portfolios, *Quantitative Finance*, **11:3** (2011), 351–364.

[7] **Dempster, A,** Covariance selection, *Biometrics*, **28** (1972), 157–175.

[8] **Dixit, A.K. and R.S. Pindyck,** *Investment Under Uncertainty*, Princeton University Press, Princeton, NJ., 1994.

[9] **Fama, E. and K. French,** Permanent and temporary components of stock prices, *The Journal of Political Economy*, **96(2)** (1988), 246–273.

[10] **Fogarasi, N. and J. Levendovszky,** A simplified approach to parameter estimation and selection of sparse, mean reverting portfolios, *Periodica Polytechnica* (to appear).

[11] **Gillespie, D.T.,** Exact numerical simulationof the Ornstein-Uhlenbeck process and its integral, *Physical Review E*, **54:2** (1996), 2084–2091.

[12] **Ito, K.,** Stochastic integral, *Proc. Imperial Acad. Tokyo*, **20** (1944), 519–524.

[13] **Manzan, S.,** Nonlinear mean reversion in stock prices, *Quantitative and Qualitative Analysis in Social Sciences*, **1(3)** (2007), 1–20.

[14] **Natarjan, B.K.,** Sparse approximate solutions to linear systems, *SIAM J. Comput.*, **24(2)** (1995), 227–234.

[15] **Ornstein, L. S. and G.E. Uhlenbeck,** On the theory of the Brownian motion, *Physical Review*, **36(5)** (1930), 823.

[16] **Penzl, T.,** Numerical solution of generalized Lyapunov equations, *Advances in Comp. Math.*, **8** (1998), 33–48.

[17] **Poterba, J.M. and L.H. Summers,** Mean reversion in stock prices: Evidence and implications, *Journal of Financial Economics*, **22(1)** (1988), 27–59.

[18] **Rothman, A., P. Bickel, E. Levina and J. Zhu,** Sparse permutation invariant covariance estimation, *Electronic Journal of Statistics*, **2** (2008), 494–515.

[19] **Smith, W.,** On the simulation and estimation of mean-reverting Ornstein-Uhlenbeck process, *Technical Report, February 2010*, `http://commoditymodels.com/2010/02/24/`

N. Fogarasi and J. Levendovszky
Department of Telecommunications
Budapest University of Technology and Economics
Budapest, Hungary
`fogarasi@hit.bme.hu`
`levendov@hit.bme.hu`