

ALGORITHMIC FINANCE

Optimizing sparse mean reverting portfolios

I. Róbert Sipos; János Levendovszky

Algorithmic Finance (2013), 2:2, 127-139

DOI: 10.3233/AF-13021

Abstract, HTML, and PDF:

<http://algorithmicfinance.org/2-2/pp127-139>

Aims and Scope *Algorithmic Finance is a high-quality academic research journal that seeks to bridge computer science and finance, including high frequency and algorithmic trading, statistical arbitrage, momentum and other algorithmic portfolio management strategies, machine learning and computational financial intelligence, agent-based finance, complexity and market efficiency, algorithmic analysis on derivatives, behavioral finance and investor heuristics, and news analytics.*

Managing Editor

Philip Maymin, NYU-Polytechnic Institute

Deputy Managing Editor

Jayaram Muthuswamy, Kent State University

Advisory Board

Kenneth J. Arrow, Stanford University
Herman Chernoff, Harvard University
David S. Johnson, AT&T Labs Research
Leonid Levin, Boston University
Myron Scholes, Stanford University
Michael Sipser, Massachusetts Institute of Technology
Richard Thaler, University of Chicago
Stephen Wolfram, Wolfram Research

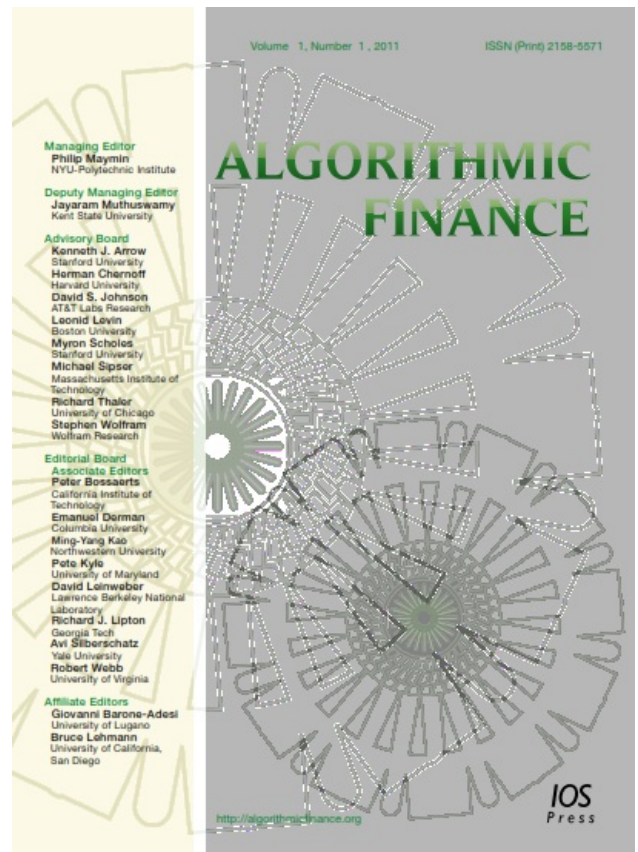
Editorial Board

Associate Editors

Peter Bossaerts, California Institute of Technology
Emanuel Derman, Columbia University
Ming-Yang Kao, Northwestern University
Pete Kyle, University of Maryland
David Leinweber, Lawrence Berkeley National Laboratory
Richard J. Lipton, Georgia Tech
Avi Silberschatz, Yale University
Robert Webb, University of Virginia

Affiliate Editors

Giovanni Barone-Adesi, University of Lugano
Bruce Lehmann, University of California, San Diego



Subscription, submission, and other info:
www.iospress.nl/journal/algorithmic-finance

Unique Features of the Journal *Open access:* Online articles are freely available to all. *No submission fees:* There is no cost to submit articles for review. There will also be no publication or author fee for at least the first two volumes. *Authors retain copyright:* Authors may repost their versions of the papers on preprint archives, or anywhere else, at any time. *Enhanced content:* Enhanced, interactive, computable content will accompany papers whenever possible. Possibilities include code, datasets, videos, and live calculations. *Comments:* Algorithmic Finance is the first journal in the Financial Economics Network of SSRN to allow comments. *Archives:* The journal is published by [IOS Press](http://www.iospress.nl). In addition, the journal maintains an [archive on SSRN.com](http://www.ssrn.com) *Legal:* While the journal does reserve the right to change these features at any time without notice, the intent will always be to provide the world's most freely and quickly available research on algorithmic finance. *ISSN:* Online ISSN: 2157-6203. Print ISSN: 2158-5571.

Optimizing sparse mean reverting portfolios

I. Róbert Sipos* and János Levendovszky†

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

*E-mail: siposr@hit.bme.hu

†E-mail: levendov@hit.bme.hu

Abstract. In this paper we investigate trading with optimal mean reverting portfolios subject to cardinality constraints. First, we identify the parameters of the underlying VAR(1) model of asset prices and then the quantities of the corresponding Ornstein-Uhlenbeck (OU) process are estimated by pattern matching techniques. Portfolio optimization is performed according to two approaches: (i) maximizing the predictability by solving the generalized eigenvalue problem or (ii) maximizing the mean return. The optimization itself is carried out by stochastic search algorithms and Feed Forward Neural Networks (FFNNs). The presented solutions satisfy the cardinality constraint thus providing sparse portfolios to minimize the transaction costs and to maximize interpretability of the results. The performance has been tested on historical data (SWAP rates, SP 500, and FOREX). The proposed trading algorithms have achieved 29.57% yearly return on average, on the examined data sets. The algorithms prove to be suitable for high frequency, intraday trading as they can handle financial data up to the arrival rate of every second.

Keywords: mean reversion, convergence trading, parameter estimation, VAR(1) model, financial time series.

1. Introduction

Portfolio optimization was first investigated by Markowitz (1952) in the context of diversification to minimize the associated risk and maximize predictability. Mean reversion is a good indicator of predictability, as a result, identifying mean reverting portfolios has become a key research area (d’Aspremont, 2011; Fogarasi & Levendovszky, 2011). Assuming that the asset price vector follows a VAR(1) process, portfolio optimization can be reduced to solving a generalized eigenvalue problem (Box & Tiao, 1977). d’Aspremont (2011) analyzed the problem of finding mean-reverting portfolios with cardinality constraints, resulting in sparse portfolios to minimize the transaction costs. With this constraint, the optimal portfolio selection becomes NP hard (Natarjan, 1995), but he proposed a heuristic algorithm providing the optimal portfolio in polynomial time. In this paper, we further advance the optimal portfolio selection by using FFNNs. Furthermore, as an alternative objective function we introduce the maximum average return and optimize it by using the underlying marginal probability density

functions (PDFs) of the OU processes. In order to estimate the parameters of OU processes we use pattern matching methods.

Trading is then perceived as a walk in the “buy/sell” action space. The methods have been tested numerically on SWAP, SP500 and FOREX rates, and the results exhibit profits.

The structure of the paper is as follows.

- in section 2, the model and the notations are introduced;
- in section 3, we optimize the portfolio by two approaches: (i) maximizing the predictability by solving the generalized eigenvalue problem; and (ii) maximizing the average return;
- in section 4, the model identification is treated;
- in section 5, portfolio optimization is carried out with cardinality constraints by using stochastic search algorithms and FFNNs;
- in section 6, the trading algorithm is described;
- in section 7, a detailed performance analysis is given based on historical data;
- in section 8, some conclusions are drawn.

2. Theoretical foundations

In this section we describe the model and the concept of mean reverting portfolios. Our approach follows the one published in (d'Aspremont, 2011), however, in section 3.2 we develop novel heuristic approaches for portfolio optimization.

2.1. The model

The time series describing the prices of assets is denoted by $\mathbf{s}_t^T = (s_{1,t}, \dots, s_{n,t})$ where $s_{i,t}$ is the price of asset i at time instant t . The portfolio vector is denoted by $\mathbf{x}^T = (x_1, \dots, x_n)$ where x_i gives the number of possessed quantity from asset i . The value of the portfolio at time t is denoted by $p(t)$ and defined as $p(t) = \mathbf{x}^T \mathbf{s}_t = \sum_{i=1}^n x_i s_{i,t}$. Our objective is to find the optimal portfolio \mathbf{x}_{opt} which maximizes a pre-defined objective function subject to cardinality constraint which specifies that the number of non-zero components in \mathbf{x}_{opt} must not exceed a given number l . The optimal portfolio is sought according to different objective functions, such as maximizing the predictability or the average return. These objective functions are developed under the assumption that the portfolio value $p(t)$ exhibits mean reverting properties and follows an Ornstein-Uhlenbeck (OU) process (Ornstein & Uhlenbeck, 1930). This is a frequent assumption in trading (Fama & French, 1988; Manzan, 2007; Ornstein & Uhlenbeck, 1930; Poterba & Summers, 1988) which follows from the VAR(1) nature of the underlying asset process.

2.2. Portfolios as OU processes and the related VAR(1) model

The OU process is characterized by the following stochastic differential equation

$$dp(t) = \vartheta(\mu - p(t))dt + \sigma dW(t), \quad (1)$$

where $W(t)$ is a Wiener process and $\vartheta > 0$ (mean reversion coefficient), μ (long-term mean) and $\sigma > 0$ (volatility) are constants. By using the Itô-Doeblin formula (Itô, 1944), one can obtain the following solution:

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

which implies that

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

and asymptotically

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

Parameter ϑ determines the convergence speed of the process towards the mean, and inversely indicating the level of uncertainty (via the standard deviation of the asymptotic Gaussian distribution (4)). Hence, for convergence trading, larger ϑ implies a better portfolio, as it quickly returns to the mean with a minimum amount of uncertainty.

Since, in real trading environments, the time is treated as a discrete quantity, in the following discussion we view the asset prices as a first order, vector autoregressive VAR(1) process, which is considered to be the discrete representative of the continuous OU process (d'Aspremont, 2011). 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 a matrix of type $n \times n$ and $\mathbf{W}_t \sim N(0, \sigma I)$ are i.i.d.r.v.-s for some $\sigma > 0$. In practice, assets are traded in discrete units, thus $x_i \in \{0, 1, 2, \dots\}$ 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 with vector \mathbf{x} (in the inner product sense), we obtain the portfolio value as follows:

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

The variance of the portfolio is

$$\sigma^2(t) = E(p_t^2) = E(\mathbf{x}^T \mathbf{s}_t \mathbf{s}_t^T \mathbf{x}) = \mathbf{x}^T \mathbf{G}\mathbf{x}, \quad (7)$$

due to independence $\sigma^2(t) = \sigma^2(t-1) + \sigma_{noise}^2$, where

$$\sigma^2(t-1) = E(\mathbf{x}^T \mathbf{A}\mathbf{s}_t \mathbf{s}_t^T \mathbf{A}^T \mathbf{x}) = \mathbf{x}^T \mathbf{A}\mathbf{G}\mathbf{A}^T \mathbf{x}, \quad (8)$$

and

$$\sigma_{noise}^2 = \mathbf{x}^T E(\mathbf{W}_t \mathbf{W}_t^T) \mathbf{x} = \mathbf{x}^T \mathbf{K}\mathbf{x}. \quad (9)$$

\mathbf{G} is the covariance matrix of the VAR(1) process and \mathbf{K} is the covariance matrix of the noise. Defining the predictability of any selected mean-reverting \mathbf{x} portfolio can be formulated as (Box & Tiao, 1977):

$$\lambda = \frac{\sigma^2(t-1)}{\sigma^2(t)} = \frac{\sigma^2(t-1)}{\sigma^2(t-1) + \sigma_{noise}^2}. \quad (10)$$

It is clear that if λ is high then the contribution of the noise is small, as a result the portfolio is more predictable.

3. Portfolio optimization

In this section we discuss the optimal portfolio selection subject to two different objective functions:

1. Optimizing the portfolio subject to the traditional λ maximization (d'Aspremont, 2011);
2. Optimizing the portfolio subject to maximizing the average return as a novel approach.

3.1. Mean reverting portfolio as a generalized eigenvalue problem

As mentioned before, λ is a key parameter for trading. The traditional way to identify the optimal sparse mean-reverting portfolio is to find a portfolio vector subject to maximizing its predictability. One may note that

$$\begin{aligned} \mathbf{x}_{\text{opt}} &= \arg \max_{\mathbf{x}} \lambda = \arg \max_{\mathbf{x}} \frac{\sigma^2(t-1)}{\sigma^2(t)} \\ &= \arg \max_{\mathbf{x}} \frac{\mathbf{x}^T \mathbf{A} \mathbf{G} \mathbf{A}^T \mathbf{x}}{\mathbf{x}^T \mathbf{G} \mathbf{x}} \end{aligned} \quad (11)$$

is equivalent to finding the eigenvector corresponding to the maximum eigenvalue in the following generalized eigenvalue problem (d'Aspremont, 2011):

$$\mathbf{A} \mathbf{G} \mathbf{A}^T \mathbf{x} = \lambda \mathbf{G} \mathbf{x} \quad (12)$$

which can then be solved as

$$\det(\mathbf{A} \mathbf{G} \mathbf{A}^T - \lambda \mathbf{G}) = 0 \quad (13)$$

under the cardinality constraint. As a result, portfolio optimization is then cast as the following constrained optimization problem:

$$\mathbf{x}_{\text{opt}} : \max_{\mathbf{x}} \frac{\mathbf{x}^T \mathbf{A} \mathbf{G} \mathbf{A}^T \mathbf{x}}{\mathbf{x}^T \mathbf{G} \mathbf{x}}, \text{card}(\mathbf{x}) \leq l \quad (14)$$

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

$$(\mathbf{G}^+)^{1/2} \mathbf{A} \mathbf{G} \mathbf{A}^T (\mathbf{G}^+)^{1/2} \mathbf{u} = \lambda \mathbf{u}, \quad (15)$$

where \mathbf{G}^+ denotes the Moore-Penrose pseudoinverse of matrix \mathbf{G} , and the cardinality constraint is now placed upon $(\mathbf{G}^+)^{1/2} \mathbf{u}$.

3.2. Maximizing the average return

In this section a new objective function is developed which may have a more direct impact on the trading profit, namely maximizing the average return. This approach yields another optimization function which, however, does not have known analytical solutions. This objective function can be related to any portfolio selection. However, in the paper we identify mean-reverting processes but now instead of optimizing their predictability, we optimize the average return. In this specific case, starting from the observed initial value of $p(0)$, based on (3), the objective function can be expressed as follows:

$$\begin{aligned} \Psi(\mathbf{x}) &= \max_{0 \leq t} E(p(t)) - p(0) = \max_{0 \leq t} (\mu(t) - p_0) \\ &= \max_{0 \leq t} ((\mu_0 - \mu) e^{-\vartheta t} + \mu - \mathbf{x}^T \mathbf{s}_0). \end{aligned} \quad (16)$$

Section 4.2 describes methods for estimating ϑ , σ , μ and μ_0 , respectively. Another objective function can be obtained if we take into account that a risk free interest is available with interest rate r_f , which allows discounting the expected future portfolio value over time. We obtain the following optimization by replacing the future value with its net present value:

$$\begin{aligned} \Psi(\mathbf{x}) &= \max_{0 \leq t} \frac{E(p(t))}{(1+r_f)^t} - p_0 = \max_{0 \leq t} \left(\frac{\mu(t)}{(1+r_f)^t} - p_0 \right) \\ &= \max_{0 \leq t} \left(\frac{(\mu_0 - \mu) e^{-\vartheta t} + \mu - \mathbf{x}^T \mathbf{s}_0}{(1+r_f)^t} \right), \end{aligned} \quad (17)$$

where equating the partial derivative to zero with respect to time yields the optimal solution, which is given as (the proof is given in the appendix):

$$t = \frac{1}{\vartheta} \ln \left(\frac{(\mu - \mu_0) (\vartheta + \ln(1+r_f))}{\mu \ln(1+r_f)} \right). \quad (18)$$

In this way, portfolio optimization can again be reduced to a constrained optimization problem:

$$\mathbf{x}_{\text{opt}} = \arg \max_{\mathbf{x}} \Psi(\mathbf{x}, t_{\text{opt}}), \text{card}(\mathbf{x}) \leq l. \quad (19)$$

where function $\Psi(\mathbf{x})$ can be either (16) or (17).

4. The computational approach

In this section we describe how to implement the mathematical model described in section 3 to trading on real time series. The proposed computational approach is depicted by the block diagram shown on Fig. 1.

As a first step, the VAR(1) model parameters (matrices \mathbf{A} , \mathbf{G} , \mathbf{K}) have to be identified based on the available observations, as will be described in section 4.1. In the next block, different portfolio optimization methods (detailed in section 4.2) will produce an optimal sparse portfolio vector: $\mathbf{x}_{\text{opt}}^T = \{x_1, \dots, x_N\}$ under the constraint $\text{card}(\mathbf{x}) \leq l$. Based on the identified portfolio, the selected trading strategy (section 6) should decide on which trading action is to be launched. For the sake of comparison of the profitability of different methods we made performance analysis the results of which are described in section 7.

4.1. VAR(1) model parameter identification

As explained in the preceding sections, with the knowledge of the parameters \mathbf{A} , \mathbf{G} and \mathbf{K} , we can apply various heuristics to approximate the l -dimensional, optimal, sparse mean-reverting portfolio (Fogarasi & Levendovszky, 2011). However, these matrices must be estimated from the historical observations of the random process \mathbf{s}_t .

We recall from our earlier discussion that we assume \mathbf{s}_t follows a stationary, first order autoregressive process. In most cases the linear system of equations is over-determined, hence \mathbf{A} is estimated using least squares estimation techniques, as

$$\hat{\mathbf{A}} : \min_{\mathbf{A}} \sum_{t=2}^T \|\mathbf{s}_t - \mathbf{A}\mathbf{s}_{t-1}\|^2 \quad (20)$$

where $\|\cdot\|^2$ denotes the Euclidian norm.

This gives a VAR(1) fit for our time series for 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.

Solving the minimization problem above, by equating the partial derivatives to zero with respect to each element of the matrix \mathbf{A} , we obtain the following system of linear equations:

$$\begin{aligned} \sum_{k=1}^n \hat{\mathbf{A}}_{i,k} \sum_{t=2}^T s_{t-1,k} s_{t-1,j} \\ = \sum_{t=2}^T s_{t,i} s_{t-1,j} \forall i, j = 1, \dots, n. \end{aligned} \quad (21)$$

Solving (21) for $\hat{\mathbf{A}}$ and switching back to vector notation for \mathbf{s} , we obtain

$$\hat{\mathbf{A}} = \sum_{t=2}^T (\mathbf{s}_{t-1}^T \mathbf{s}_{t-1})^+ (\mathbf{s}_{t-1}^T \mathbf{s}_t), \quad (22)$$

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 because of the potential singularity of $\mathbf{s}_{t-1}^T \mathbf{s}_{t-1}$.

The covariance matrix \mathbf{K} of random variable \mathbf{W}_t can be estimated based on the assumption that the noise terms in equation are i.i.d.r.v.-s with $\mathbf{W}_t \sim N(0, \sigma_W \mathbf{I})$ for some $\sigma_W > 0$, where \mathbf{I} denotes the identity matrix. Then we obtain the following estimate for σ_W using:

$$\hat{\sigma}_W = \sqrt{\frac{1}{n(T-1)} \sum_{t=2}^T \|\mathbf{s}_t - \hat{\mathbf{A}}\mathbf{s}_{t-1}\|^2}. \quad (23)$$

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

$$\hat{\mathbf{K}} = \frac{1}{(T-1)} \sum_{t=2}^T (\mathbf{s}_t - \hat{\mathbf{A}}\mathbf{s}_{t-1})^T (\mathbf{s}_t - \hat{\mathbf{A}}\mathbf{s}_{t-1}). \quad (24)$$

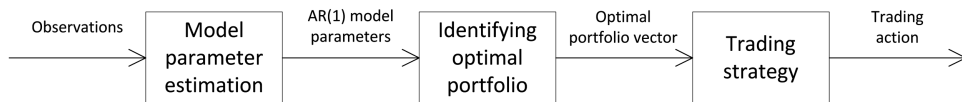


Fig. 1. Computational model.

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

We estimate \mathbf{G} with the sample covariance matrix obtained as:

$$\hat{\mathbf{G}}_1 := \frac{1}{T-1} \sum_{t=1}^T (\mathbf{s}_t - \bar{\mathbf{s}})^T (\mathbf{s}_t - \bar{\mathbf{s}}), \quad (25)$$

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

$$\bar{\mathbf{s}} := \frac{1}{T} \sum_{t=1}^T \mathbf{s}_t. \quad (26)$$

4.2. OU parameter estimation

The objective of this section is to estimate parameter μ and ϑ . Estimating the long term mean (μ) of the process of portfolio valuations (6) is instrumental for mean reverting trading. Having an OU process at hand, this estimation can be performed in multiple ways (Fogarasi & Levendovszky, 2012a). Besides the long term mean the mean at the initial starting point of the process (μ_0) is also a subject of interest from the point of trading. Besides the estimation of μ and μ_0 , the least squares regression can estimate ϑ and σ parameters for any selected portfolio.

Sample mean estimation

As a benchmark for other methods, the long term mean can be estimated by simply taking the empirical average of the observed data (27), while the first value can be assigned to the μ_0 parameter (28).

$$\hat{\mu} := \frac{1}{T} \sum_{t=1}^T \mathbf{x}^T \mathbf{s}_t \quad (27)$$

$$\hat{\mu}_0 := p_0 \quad (28)$$

Least squares regression (van Gelder, 2009)

Rewriting the Ornstein-Uhlenbeck equation stochastic differential equation (1) to the following way

$$\begin{aligned} s_t - s_{t-1} &= \vartheta(\mu - s_{t-1})\Delta t + \sigma(W_t - W_{t-1}) \\ &= \vartheta\mu\Delta t - \vartheta s_{t-1}\Delta t + \sigma(W_t - W_{t-1}) \end{aligned} \quad (29)$$

results in a linear regression in the form of

$$y = a + bx + \varepsilon_t \quad (30)$$

from which the estimation of the OU process parameters can be formulated as

$$\hat{\vartheta} = -\frac{\ln(b)}{\Delta t}, \hat{\sigma} = sd(\varepsilon_t) \sqrt{\frac{2\hat{\vartheta}}{1-b^2}} \text{ and } \hat{\mu} := \frac{a}{1-b}, \quad (31)$$

where $sd(\varepsilon_t)$ denotes the standard deviation of ε_t over t . Note that in the lack of the mean-reverting property, when $\hat{\vartheta}$ is close to zero, the estimation of $\hat{\mu}$ could suffer from numerical instability.

Recursive solution

As the selected linear combination of VAR(1) processes is also VAR(1), then we can use the following recursive function:

$$p_0 = \mathbf{x}^T \mathbf{s}_0 \quad (32)$$

$$\hat{\mathbf{s}}_{t+1} = \mathbf{A} \mathbf{s}_t \quad (33)$$

$$p_{t+1} = \mathbf{x}^T \mathbf{A} \hat{\mathbf{s}}_t \quad (34)$$

This recursion should be performed until the difference between two steps is smaller than a given threshold, when the process is sufficiently close to the mean.

$$\hat{\mu} := p_{t+k}, |p_{t+k} - p_{t+k-1}| < \varepsilon \quad (35)$$

If the recursion is divergent that means that the process has not got the mean reversion property.

Minimizing the mean-square error

Among all possible Ornstein-Uhlenbeck processes described by equation

$$\begin{aligned} \mathbf{m}(\mu_0, \mu, \vartheta, t) &= (\mu_0 - \mu)e^{-\vartheta t} + \mu \\ \mathbf{M} &: \left\{ \mathbf{m}(\mu_0, \mu, \vartheta, t), \vartheta = \hat{\vartheta}, t = 1, \dots, T \right\} \end{aligned} \quad (36)$$

we would like to select the one with least squared error against the observed portfolio valuation vector:

$$\hat{\mu} = \min_{\mu \in M} \|\mathbf{p} - \mathbf{m}\|^2 \quad (37)$$

Solving the $\frac{\partial}{\partial \mu} \|\mathbf{p} - \mathbf{m}\|^2 = 0$ equation to find the optimal mean gives the following estimation:

$$\hat{\mu} = \frac{\sum_{t=1}^T (p_t - \mu_0 e^{-\vartheta t})}{\sum_{t=1}^T (1 - e^{-\vartheta t})}. \quad (38)$$

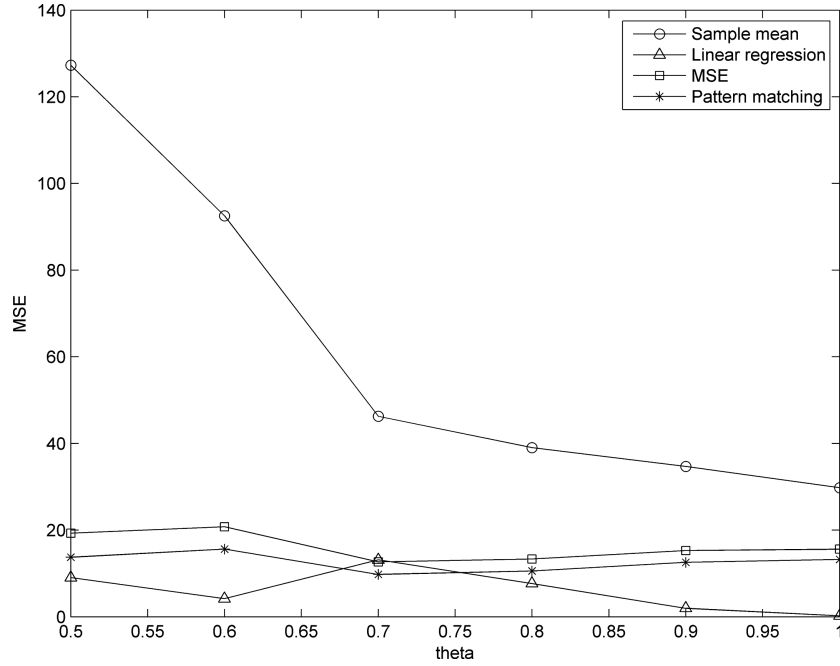


Fig. 2. Comparison of different mean estimation techniques.

Pattern matching (Fogarasi & Levendovszky, 2012a)

Similarly to the previous method, pattern matching aims to find the maximum-likelihood estimation for the observed portfolio valuations among all possible processes (36). Taking into account that the samples follow Gaussian distribution, this leads to the optimization problem

$$\max_{\boldsymbol{\mu} \in \mathcal{M}} \frac{1}{\sqrt{2\pi \det(\mathbf{U})}} e^{-\frac{1}{2}(\mathbf{p}_t - \boldsymbol{\mu}_t)^T \mathbf{U}^{-1}(\mathbf{p}_t - \boldsymbol{\mu}_t)} \quad (39)$$

being equivalent with

$$\min_{\boldsymbol{\mu} \in \mathcal{M}} \boldsymbol{\mu}_t^T \mathbf{U}^{-1} \boldsymbol{\mu}_t - 2\boldsymbol{\mu}_t^T \mathbf{U}^{-1} \mathbf{p}_t, \quad (40)$$

where \mathbf{U} denotes the covariance matrix of \mathbf{p}_t : $\mathbf{U}_{i,j} = \frac{\sigma^2}{2\vartheta} (e^{-\vartheta|i-j|} - e^{-\vartheta(i+j)})$

As the optimization problem is quadratic, it can be solved analytically

$$\frac{\partial}{\partial \boldsymbol{\mu}} \boldsymbol{\mu}_t^T \mathbf{U}^{-1} \boldsymbol{\mu}_t - 2\boldsymbol{\mu}_t^T \mathbf{U}^{-1} \mathbf{p}_t = 0 \quad (41)$$

resulting in the following closed form solution:

$$\hat{\boldsymbol{\mu}} = \frac{(\mathbf{p}^T - \mu_0 \mathbf{v}_2^T) \mathbf{U}^{-1} \mathbf{v}_1}{\mathbf{v}_1^T \mathbf{U}^{-1} \mathbf{v}_1} \quad (42)$$

where $\mathbf{v}_1 = \{(1 - e^{-\vartheta t}), t = 1, \dots, T\}$ and $\mathbf{v}_2 = \{e^{-\vartheta t}, t = 1, \dots, T\}$ respectively.

For the sake of finding the best estimation procedure, the different methods were tested on artificially generated data. Each observation was with the following parameters $\sigma = 1.5$, $T = 8$ and $\mu, \mu_0 \in [0; 100]$, respectively. The comparison was done independently for mean reversion coefficients in $\vartheta \in [0.5; 1]$, and in each case the mean squared error was taken into account for 100 generated processes. The results are shown by Fig. 2.

5. Optimization and dealing with the cardinality constraint

Having identified the model parameters, we can now focus on the optimization. In the absence of proper analytical solutions for the constrained optimization problems posed in section 3.2, we use simulated annealing and FFNNs for obtaining good quality heuristic solutions.

5.1. Portfolio optimization by simulated annealing

Simulated annealing (Kirkpatrick et al., 1983) is a stochastic search method for finding the global optimum in a large search space. In this context

the energy function $J(\mathbf{x})$ is either the first objective function maximizing the eigenvalue (11) (Fogarasi & Levendovszky, 2012b) or the second objective function maximizing the average return (12) for the selected portfolio:

$$\begin{aligned} \text{a) } J(\mathbf{x}) &= \frac{\mathbf{x}^T \mathbf{A} \mathbf{G} \mathbf{A}^T \mathbf{x}}{\mathbf{x}^T \mathbf{G} \mathbf{x}}, \\ \text{b) } J(\mathbf{x}) &= \max_{0 \leq t} \left(\frac{(\mu_0 - \mu) e^{-\theta t} + \mu}{(1+r)^t} - \mathbf{x}^T \mathbf{s}_0 \right). \end{aligned}$$

With an appropriate neighbor function the cardinality constraint $\text{card}(\mathbf{x}) \leq l$ is automatically fulfilled at each step of the algorithm. The neighbor function on each iteration makes two steps: first, changes the non-zero values by ± 1 , and in the second step inserts the changed elements into a randomly chosen l dimensional subspace. Let \mathbf{x} be an arbitrary initialization vector, and then by calling a random number generation a vector \mathbf{x}' is generated subject to uniform distribution over the subspace in which the optimal portfolio vector is sought. Accept the new vector if $J(\mathbf{x}') > J(\mathbf{x})$, or otherwise with $e^{-\frac{J(\mathbf{x}) - J(\mathbf{x}')}{T}}$ probability. Continue the sampling while decreasing the T until zero. The last state vector is now the identified optimal sparse portfolio vector.

Algorithm

Set $\mathbf{x}(0) = \mathbf{0}$ (or use any other arbitrary initialization). Assume we are at Step l .

1. By calling the RNG function generate a vector \mathbf{x}' subject to uniform distribution over bounded the state space in which the optimal portfolio vector is sought.
2. Generate a binary random number subject to the following distribution:

$$P(\xi = 1) = \begin{cases} 1 & \text{if } J(\mathbf{x}') > J(\mathbf{x}(l)) \\ e^{-\frac{J(\mathbf{x}(l)) - J(\mathbf{x}')}{T(l)}} & \text{if } J(\mathbf{x}') < J(\mathbf{x}(l)) \end{cases}$$

3. Accept \mathbf{x}' if $\xi = 1$.
4. Generate k times a number subject to uniform distribution $1/n$ in the interval $(1, \dots, n)$ (if two or more generated numbers coincide then refuse them and carry on with the generation until all of them are different). Let us denote the output of this generation with $(i_1, \dots, i_k); i_j \in \{1, \dots, n\}$.
5. Set the corresponding components of vector \mathbf{x}' to be zero, i.e. $x'_{i_j} := 0, j = 1, \dots, k$ to satisfy the cardinality constraint and accept $\mathbf{x}(l+1) = \mathbf{x}'$.

6. Decrease $T(l)$ according to $T(l) = \frac{A}{\log(l)}$ (or by using $T(l+1) = \alpha^* T(l)$).
7. Otherwise go back to Step 1.

5.2. Portfolio optimization by FFNNs

We can use a universal approximator, like FFNN to estimate the optimal portfolio vector from the identified matrices. FFNNs are described by the following mapping

$$\begin{aligned} y &= \text{Net}(\mathbf{z}, \mathbf{w}) \\ &= \varphi \left(\sum_j w_j^{(L)} \varphi \left(\sum_l w_{jl}^{(L-1)} \dots \varphi \left(w_{nm}^{(1)} z_m \right) \right) \right) \end{aligned} \quad (43)$$

where z is the input vector and \mathbf{w} vector denotes the free parameters. They have universal representation capabilities (Cybenko, 1989) in L^2 in terms of

$$\forall \varepsilon > 0, f(\mathbf{z}) \in L^2 \rightarrow \exists \mathbf{w} : \|f(\mathbf{z}) - \text{Net}(\mathbf{z}, \mathbf{w})\| < \varepsilon \quad (44)$$

Furthermore they can learn and generalize from a finite set of examples $\tau^{(K)} = \{(\mathbf{z}_k, \mathbf{d}_k), k = 1, \dots, K\}$ by using the Back Propagation algorithm (Hagan & Menhaj, 1994). As a result, one may look upon the optimal portfolio selection problem as a mapping from the identified, $\mathbf{A}, \mathbf{G}, \mathbf{K}$ matrices of the underlying VAR(1) process to the optimal sparse portfolio vector \mathbf{x} . In this case, the input vector of FFNN is $\mathbf{z} = (\mathbf{A}, \mathbf{G}, \mathbf{K})$ constructed by matrix flattening and the output is vector \mathbf{x} with $\text{card}(\mathbf{x}) \leq l$ to fulfill the sparsity constraint. One can construct a training set $\tau^{(K)} = \{(\mathbf{z}_k, \mathbf{d}_k), k = 1, \dots, K\}$, by finding the optimal sparse portfolio vectors for some input matrices by exhaustive search. Unfortunately as the input layer has a size of $3N^2$, this solution can be used only for lower dimensional spaces. The construction of the training set is done according to the computational model shown on Fig. 3.

Once the training set has been constructed, the BP algorithm (Hagan & Menhaj, 1994) can be used to optimize the weights (the free parameters) of the corresponding FFNN by minimizing the following objective function:

$$\mathbf{w}_{opt} : \min_{\mathbf{w}} \frac{1}{K} \sum_{k=1}^K \|\mathbf{z}_k - \text{Net}(\mathbf{z}_k, \mathbf{w})\|^2 \quad (45)$$

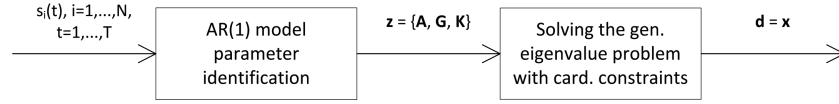


Fig. 3. Construction of the training set for FFNNs.

These FFNNs can be combined with the SA method described in section 5.1 to generate a starting vector for the search, in order to increase the convergence speed.

6. The proposed trading algorithms

In this section we describe the trading algorithms which are used to trade with the optimized portfolio. In the proposed algorithms, trading is described as a walk in a binary state space in which either we already have a portfolio at hand or cash at hand, while the transitions are only affected by the state of the portfolio valuation described in (6) or the evaluations of the potentially owned and the newly identified portfolios by the using a given objective function (see section 3). We introduce two alternatives of the trading strategy which differs in whether is it allowed to sell an owned portfolio before it reached its expected mean and invest to a new portfolio with higher expectations instead. Each trading strategy is formalized by a state chart (Fig. 4 and Fig. 5).

6.1. Trading by the λ of OU process

After the optimal portfolio selection based on maximizing the predictability parameter λ , the trading algorithm is responsible for interpreting the selected portfolios as a sequence of trading actions. In the first place it has to be decided whether the portfolio is below or above its long term mean, or in its stationary state. This can be perceived as a decision theoretic problem. In stationary state the portfolio has a normal distribution,

$$p_t \sim N\left(\mu, \frac{\sigma}{\sqrt{2\vartheta}}\right) \quad (46)$$

then we can formulate hypotheses based on the probability of the observed portfolio value is in the stationary state as follows (Fogarasi & Levendovszky, 2012a):

- Hypothesis 1, the portfolio is under the mean ($p(t) < \mu - \alpha \rightarrow p(t) < \mu$) with the

probability of

$$\int_{-\infty}^{\mu-\alpha} \frac{1}{\sqrt{2\pi}\sigma/\sqrt{2\vartheta}} e^{-\frac{(u-\mu)^2}{\sigma^2/\vartheta}} du = \varepsilon/2. \quad (47)$$

- Hypothesis 2, the portfolio is above the mean ($p(t) > \mu + \alpha \rightarrow p(t) > \mu$) with the probability of

$$\int_{\mu+\alpha}^{\infty} \frac{1}{\sqrt{2\pi}\sigma/\sqrt{2\vartheta}} e^{-\frac{(u-\mu)^2}{\sigma^2/\vartheta}} du = \varepsilon/2. \quad (48)$$

- Hypothesis 3, the portfolio is in stationary state ($p(t) \in [\mu - \alpha; \mu + \alpha] \rightarrow p(t) = \mu$) with the probability of

$$\int_{\mu-\alpha}^{\mu+\alpha} \frac{1}{\sqrt{2\pi}\sigma/\sqrt{2\vartheta}} e^{-\frac{(u-\mu)^2}{\sigma^2/\vartheta}} du = \varepsilon. \quad (49)$$

By adjusting the parameter ε one can set the probability of making the wrong choice if the portfolio has already reached its stationary state. Adjusting this parameter affects how cautious the trading will be.

In this case the agent buys the portfolio only if the identified portfolio has not reached its stationary state. Note that if the portfolio is above its estimated long term mean, then the agent buys the inverse portfolio instead in which every long and short position is switched to its opposite. And after having a portfolio at hand the agent holds it until it reaches its stationary state (estimating the long term mean and making the decision is recalculated in every time instance according to the sliding window used) then sells it.

6.2. Trading by maximizing the average return

In the second case, the portfolio is evaluated by calculating the corresponding objective function. Positive evaluation indicates a profitable portfolio, while negative evaluation indicates that the portfolio may produce a loss. Higher value in each objective function implies a better portfolio.

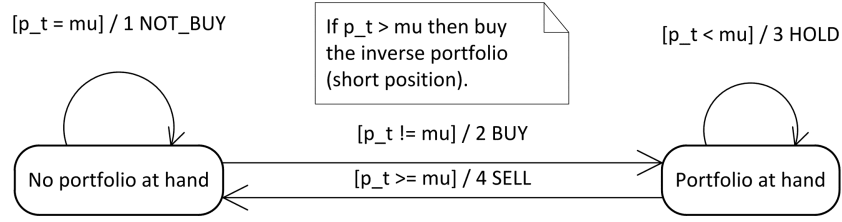


Fig. 4. State chart of “nochange” trading strategy.

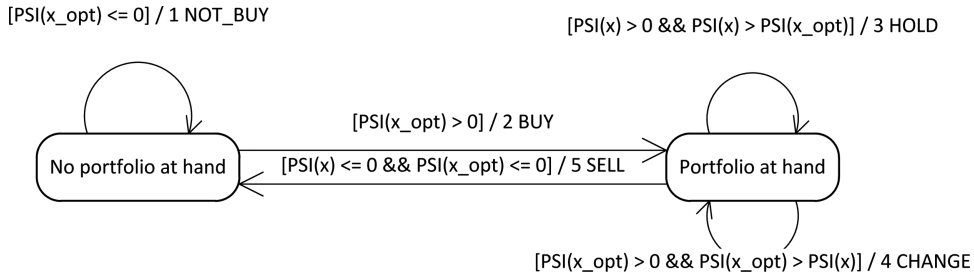


Fig. 5. State chart of “change” trading strategy.

Based on this, the agent buys a portfolio only if it has a positive evaluation. A new trading action is taken if a newly identified portfolio has higher and also positive evaluation than the present one. In this case one can sell the owned portfolio and buy the new one regardless of whether the original portfolio reached its mean or not. This approach treats the present portfolio as a sunk cost, thus only the future expectations are taken into consideration. Hence, we do not have to give up the best available portfolio in favor of a presently unfavorable portfolio. In the case that neither the owned nor the currently identified portfolio has positive evaluation then the agent closes the positions (Fig. 5).

7. Simulation results and performance analysis

An extensive back-testing framework was created to handle trading actions on various input data sets and provide numerical results for comparison of different methods on different financial data series.

7.1. Data sets

For performance analysis we used the following data sets:

- Daily close prices of 500 stocks from the S&P 500 (between July 2010 and July 2011) (Yahoo, 2011);
- U.S. SWAP rates (from the year of 1998) (Morgan Stanley Marketone, 2010);

- FOREX rates (EUR/USD, GBP/USD, AUD/USD, NZD/USD, USD/CHF, USD/CAD between October 2005 and September 2006 in daily resolution).

In the implementation, we used sliding windows for the model parameter estimation, where an optimal portfolio was identified after each step as an input for the trading strategy described in section 6.

For real data, a pre-processing step is required in order to ensure $E(s_t) = 0$ for fitting the VAR(1) model:

$$s_t = z_t - \bar{z}, t = 1, \dots, T, \tag{50}$$

where $\bar{z} := \frac{1}{T} \sum_{t=1}^T z_t$ (z denotes the original time series). This normalization have done in each time window independently, therefore this approach is applicable for forward tests as well.

7.2. Performance measures

For the sake of comparison the following performance measures were calculated for each simulation, where c_t denotes the sum of owned cash and the market value of the owned portfolio at time instance t , while c_0 denotes the initial cash (in each case the agent started with \$10,000).

- Minimal value $G_{\min} = \frac{\min_{0 \leq t \leq T} c_t}{c_0}$

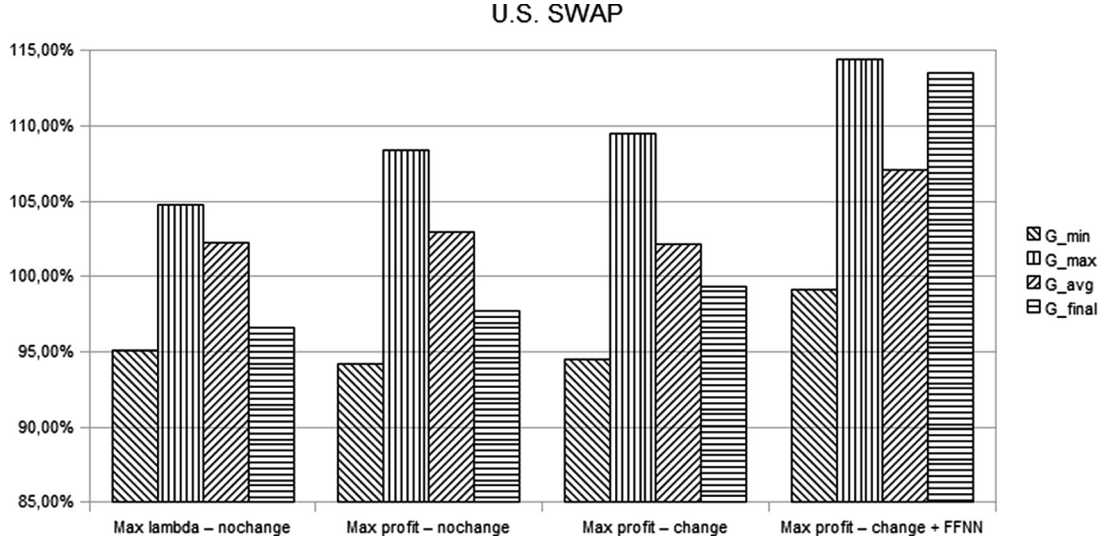


Fig. 6. Trading results on U.S. SWAP data.

- Final value $G_{final} = \frac{cT}{c_0}$
- Maximal value $G_{max} = \frac{\max_{0 \leq t \leq T} c_t}{c_0}$
- Average value $G_{avg} = \frac{\frac{1}{T} \sum_{t=1}^T c_t}{c_0}$
- Trading count N
- Sharpe-ratio $S = \frac{\frac{1}{T} \sum_{t=1}^T \left(\frac{c_t}{c_{t-1}} - 1 \right) - r_f}{\sqrt{\text{var}\left(\left\{ \frac{c_t}{c_{t-1}}, t=1, \dots, T \right\}\right)}}$,

where r_f denotes the available daily risk free interest rate and $\text{var}\left(\left\{ \frac{c_t}{c_{t-1}}, t=1, \dots, T \right\}\right)$ denotes the variance of the daily returns (for the numerical results we used yearly 1% as risk free return).

7.3. Trading results

In this section we show the numerical results obtained on S&P500, SWAP and FOREX mid-prices. The performance of different portfolio identification methods are compared. The effect of deploying a FFNN (section 5.2) is also measured (due to the limitations of FFNNs regarding higher dimensional spaces in this case only the SWAP data was analyzed). Regarding the sparsity constraint, 3 assets were selected in each transaction.

In case of the SWAP and S&P 500 data 8 days were used as an identification time window, while 14 days for FOREX. The usage of relatively short windows let us quickly react to new market situations. Buying portfolios much earlier than they converged

close to their long term mean compensates the rough estimation of model parameters. The long term mean was estimated by linear regression (section 4.2) in each simulation.

In this period, the U.S. SWAP rates had a decreasing tendency, at the end of the year they are worth only 87.89% of their initial prices on average. The bar chart (Fig. 6) shows that all of the introduced methods beat this tendency, and in the scenario when a FFNN was deployed the trading was profitable with a 13.49% yearly profit (with 0.065 daily Sharpe-ratio).

While the S&P 500 asset prices in the studied period rose only with 12.03%, maximizing the λ results in 33.40% and maximizing the average profit with the first trading strategy results in 53.55% profit (with 0.100 daily Sharpe-ratio). On the other hand, changing the portfolios before they could reach the long term mean proved to be a bad strategy (Fig. 7).

The FOREX rates in this period showed (Fig. 8) only a slight increase during the year (1.52%), our methods achieved up to 21.66% profit (with 0.177 daily Sharpe-ratio).

As one can see, the novel portfolio optimization methods outperform the traditional eigenvalue maximization in most scenarios, and also deploying a FFNN increases the performance.

Fig. 9 shows the profit reached in FOREX data using different mean estimation methods. The novel objective function with the corresponding trading strategy described in 6.2 were used. It is shown that using the linear regression results in the best profit.

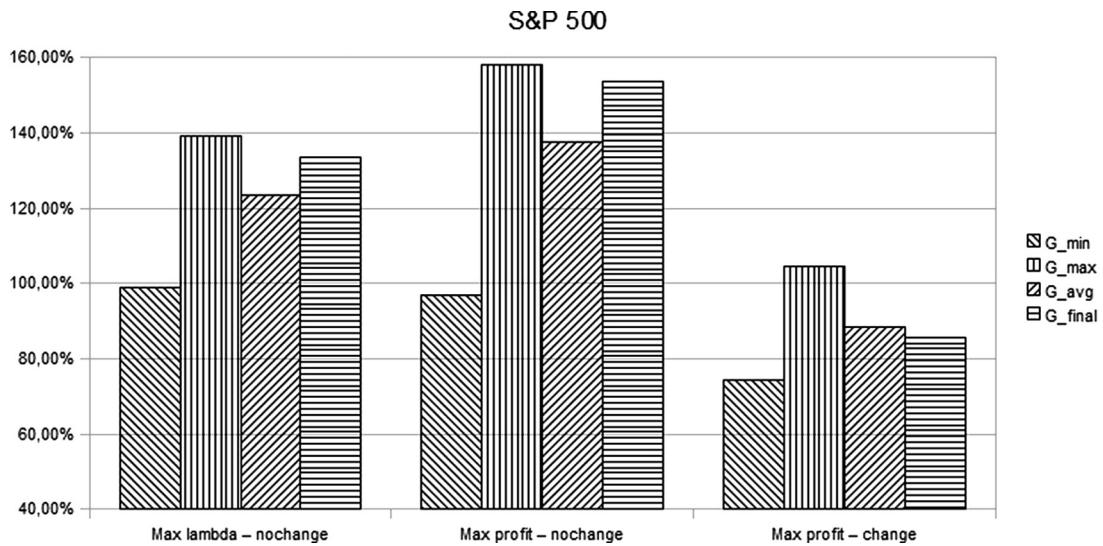


Fig. 7. Trading results on S&P 500 data.

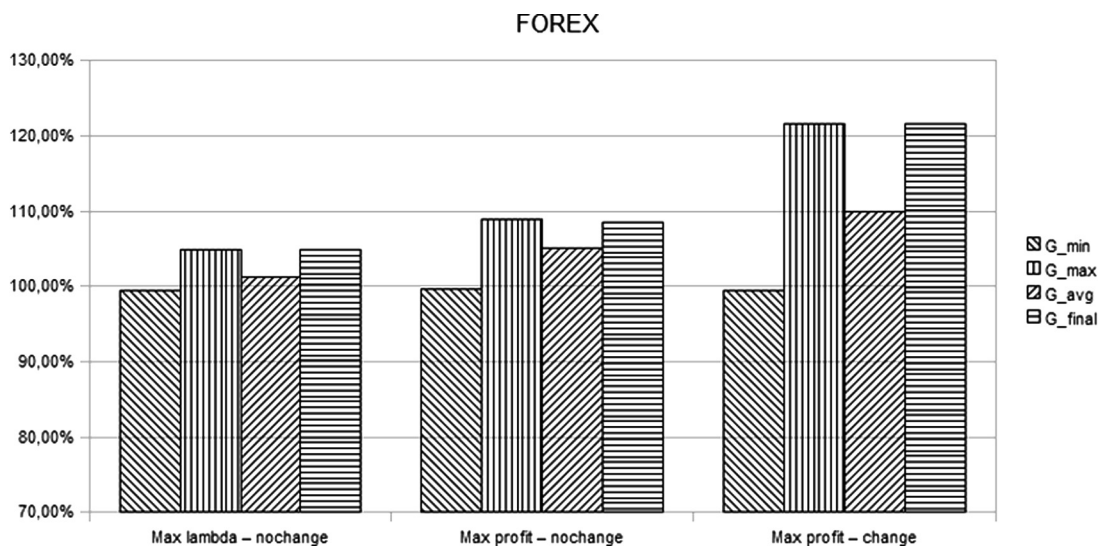


Fig. 8. Trading results on FOREX data.

7.4. Response time analysis

Besides the profit achieved by the introduced methods, the computational time is also an important aspect of usability in real environments. As described in Fig. 1, each trading iteration consists of a parameter identification step, portfolio optimization and making a trading decision. The average, minimal and maximal response times are given for each data set. The measurements were made on a desktop computer with an Intel Core 2 Quad Q9550@2.83GHz CPU on a single thread without any optimization

(e.g. caching). Table 1 contains the response times with maximizing the predictability (section 3.1). In this case, the response time is roughly constant and depends mostly on the number of assets.

The computational time for maximizing the average return (section 3.2) is summarized in Table 2. This method is considerably slower because of the stochastic portfolio optimization, but for large number of assets, the running time is comparable to the generalized eigenvalue problem. Note that the maximum response time can be scaled in an arbitrary way by setting the number of steps in the simulated annealing algorithm,

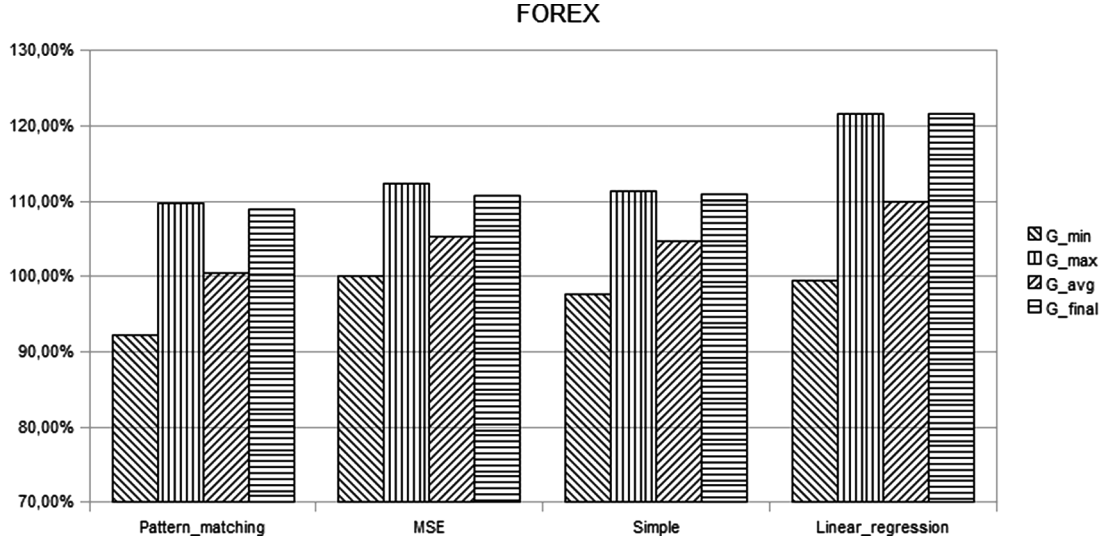


Fig. 9. Comparison of trading results using different mean estimation techniques on FOREX data.

Table 1

Response times (with maximizing the predictability)			
	Minimum	Maximum	Average
SWAP	1 ms	297 ms	9 ms
S&P 500	2138 ms	3176 ms	2406 ms
FOREX	1 ms	19 ms	8 ms

Table 2

Response times (with maximizing the average return)			
	Minimum	Maximum	Average
SWAP	551 ms	16752 ms	1574 ms
S&P 500	2220 ms	2785 ms	2496 ms
FOREX	650 ms	15641 ms	1727 ms

and massively parallel architectures can be used as well to decrease the running time.

8. Conclusions and future work

In this paper we have proposed novel algorithms for optimal trading on mean reverting portfolios with cardinality constraints. We used two objective functions: (i) maximizing the predictability parameter; and (ii) maximizing the mean return. The portfolio optimizations have been carried out by stochastic search and FFNNs. The proposed trading algorithms can reach profit on real financial time series. The performance analysis demonstrated that portfolio selection based on the novel objective functions (maximizing the average profit), could increase trading efficiency and profit compared to the traditional λ

optimization strategy. Also introducing FFNNs can slightly improve the performance. However, there is a room for improvement on the trading strategies in order to take other factors (e.g. introducing stop loss) into consideration, as well.

Appendix

In order to solve the maximization problem stated in (17), we need to equate the partial derivative of the function to zero with respect to time (the function has only one extremum, namely the global maximum we are looking for):

$$\frac{\partial}{\partial t} \left(\frac{(\mu_0 - \mu) e^{-\vartheta t} + \mu}{(1 + r_f)^t} - \mathbf{x}^T \mathbf{s}_0 \right) = 0. \quad (51)$$

Partially differentiate this expression yields the following function:

$$\begin{aligned} & \frac{\partial}{\partial t} \left(\frac{(\mu_0 - \mu) e^{-\vartheta t} + \mu}{(1 + r_f)^t} - \mathbf{x}^T \mathbf{s}_0 \right) \\ &= \frac{e^{-\vartheta t} (\mu - \mu_0) (\vartheta + \ln(1 + r_f)) - \mu \ln(1 + r_f)}{(1 + r_f)^t}. \end{aligned} \quad (52)$$

Then by solving the equation, we obtain the optimal solution in (18):

$$\frac{e^{-\vartheta t} (\mu - \mu_0) (\vartheta + \ln(1 + r_f)) - \mu \ln(1 + r_f)}{(1 + r_f)^t} = 0$$

$$e^{-\vartheta t} = \frac{\mu \ln(1 + r_f)}{(\mu - \mu_0)(\vartheta + \ln(1 + r_f))}$$

$$t = \frac{1}{-\vartheta} \ln \left(\frac{\mu \ln(1 + r_f)}{(\mu - \mu_0)(\vartheta + \ln(1 + r_f))} \right)$$

$$= \frac{1}{\vartheta} \ln \left(\frac{(\mu - \mu_0)(\vartheta + \ln(1 + r_f))}{\mu \ln(1 + r_f)} \right).$$

Acknowledgement

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).

References

- Box, G.E., Tiao, G.C., 1977. A canonical analysis of multiple time series. *Biometrika* 64 (2), 355–365.
- Cybenko, G., 1989. Approximations by superpositions of sigmoidal functions. *Math. controls signals syst.* 2 (4), 303–314.
- d'Aspremont, A., 2011. Identifying small mean-reverting portfolios. *Quant. Financ.* 11 (3), 351–364.
- Fama, E., French, K., 1988. Permanent and temporary components of stock prices. *J. Polit. Econ.* 96 (2), 246–273.
- Fogarasi, N., Levendovszky, J., 2011. A simplified approach to parameter estimation and selection of sparse, mean reverting portfolios. To appear in *Periodica Polytechnica*.
- Fogarasi, N., Levendovszky, J., 2012a. Improved parameter estimation and simple trading algorithm for sparse, mean-reverting portfolios. *Annales Univ. Sci. Budapest. Sect. Comp.* 37, 121–144.
- Fogarasi, N., Levendovszky, J., 2012b. Sparse, mean reverting portfolio selection using simulated annealing. Working Paper.
- Hagan, M.T., Menhaj, M.B., 1994. Training feed-forward networks with the Marquardt algorithm. *IEEE Trans. Neural. Netw.* 5 (6), 989–993.
- Ito, K., 1944. Stochastic integral, vol. 20. Proc. Imperial Acad, Tokyo, pp. 519–524.
- Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P., 1983. Optimization by simulated annealing. *Science* 220, 671–680.
- 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.
- Morgan Stanley Marketone., 2010. SWAP series [DATABASE].
- Natarjan, B.K., 1995. Sparse approximate solutions to linear systems. *SIAM J. Comput.* 24 (2), 227–234.
- 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.
- van Gelder, P., 2009. Statistical modeling of financial markets. In: TU Delft, UFS Workshop, Bloemfontein, 4–5 March 2009.
- Yahoo., 2011. S&P 500 asset series [WWW]. <http://finance.yahoo.com> [Accessed July 2011].