OPTIMIZATION OF A FUTURES PORTFOLIO UTILIZING NUMERICAL MARKET PHASE-DETECTION

L. PUTZIG*, D. BECHERER†, AND I. HORENKO‡

Abstract. This paper presents an application of the recently developed method for simultaneous dimension reduction and meta-stability analysis of high-dimensional time series in context of computational finance. Further extensions are included to combine state-specific principal component analysis (PCA) and state-specific regressive trend models to handle the high-dimensional, non-stationary data. The identification of market phases allows to control the involved phase-specific risk for futures portfolios. The numerical optimization strategy for futures portfolios based on the Tykhonov-type regularization is presented. The application of proposed strategies to online detection of the market phases is exemplified first on the simulated data, then on historical futures prices for oil and wheat between 2005-2008. Numerical tests demonstrate the comparison of the presented methods with existent approaches.

Key words. economic time series analysis, portfolio optimization, investment, regression, market phases, clustering

AMS subject classifications. 62M10, 91B28, 91B84

1. Introduction.

In today's financial industry, computational finance is used for a wide field of applications. Throughout this paper, we will concentrate on the portfolio optimization problem, where, for return maximization and risk minimization problems, the evolution of security returns is estimated with a variety of different methods, in order to find an allocation that optimizes a suitable performance criterion, while limiting the risk [26, 30]. For practical reasons, computational methods have to address the following points:

1. **High-dimensional Optimization:** Financial data are *high dimensional* as thousands of securities are available from different asset classes, (stocks, futures etc.), and data are available often at high frequencies with non-equidistant time stepping (due to weekends, market closures, holidays, . . .). E.g. the Wilshire 5000 consists of 5000 stocks just from U.S. companies. Taking stocks from all over the world and including derivatives like different kinds of options and certificates, commodities, bonds and futures to this pool of assets, one might end up with millions of possible investments, leading to all kinds of numerical problems.
2. **Time Series Analysis:** Market dynamics are most likely not reflected by a stationary process, but involve (hidden) market phases, as economic cycles, harvest cycles (for grain, fruits and other kind of plants), yearly holidays and many more aspects influencing the markets. This poses fundamental problems for estimation: To obtain sensible estimates for mean returns even in stationary linear models, hundreds of years of data would be needed. In shorter intervals, estimation risk becomes significant.
3. **Sensitivity Analysis:** Approaches should be robust with respect to model parameter uncertainty. This is a challenge when estimating time-varying parameters.

Aspects (1,2) cause a high uncertainty of estimates whereas non-stationarity (2) (in observable dimensions, at least) means that estimates should involve time-dependent functions, which could not be estimated in a non-trivial way, even so, it is a well-known problem. I.e. in 1989 J. Hamilton [12] suggested a numerical method for identification of piecewise linear functions, called hidden market phases, assuming stationarity and Gaussianity of returns. While Gaussianity is still a generally accepted assumption, many modern models, especially when build for practical purpose, soften this by using, e.g. heavy tails [28]. Later this was extended to multidimensional data using linear vector autoregressive (VAR) models [22]. Still, standard phase-identification techniques based on filtering approach (like wavelets-based spectral methods, see [1]) have in

*__corresponding author__ Free University Berlin, Institute of Mathematics, Arnimallee 6, 14195 Berlin, putzig@mi.fu-berlin.de

†Humboldt-University Berlin, Institute of Mathematics, Rudower Chaussee 25, 12489 Berlin, becherer@mathematik.hu-berlin.de

‡Free University Berlin, Institute of Mathematics, Arnimallee 6, 14195 Berlin, horenko@mi.fu-berlin.de

general infeasible numerical complexity in high dimensions; other methods (e.g., Kalman-filter [20], VAR [22], (G)ARCH [6, 2]) require stronger assumptions about the underlying dynamics or the methods (as in [3]) are based on perfect observability of the Markov process, which can not be taken for granted in reality. As demonstrated by the actual financial crises, non-robust reliance on simplified assumptions and ignorance of model risk can (among other factors) lead to misperception of risk and thus not less than to a total loss of the investment. Therefore, there is need of a wider class of methods, that are computationally capable to deal with realistic amounts of data, to monitor the evolution of risks and optimize the risk-return profile of investments. While recent publications concentrate on robustness by using worst-case scenarios (e.g. [10, 7, 11]), penalty functions (e.g. [34]) or simply assuming vanishing mean returns, this paper aims at an estimation of non-stationary trends and volatilities directly from prices of futures contracts. Futures markets constitute broad and liquid markets for several asset classes [5]. Leverage can be quite large for futures, if the amplitude of futures price variations is measured relative to the size of the margin requirement. But this also includes a high risk, the minimization of this risk is one of the central points of the paper. We present the recently developed numerical framework work for clustering and analysis of non-stationary data [14, 15, 13, 16] and explain it in the context of computational finance. An extension of the framework utilizing linear regression is introduced, allowing for simultaneous market phase identification, dimension reduction and estimation of the phase-specific trends in financial context. The ability of the extended framework to identify change points online is demonstrated, both on simulated and real market data. Finally, the numerical optimization strategy for phase-specific futures portfolio selection is presented and exemplified on a real 8-dimensional data set of futures data for wheat and oil.

This paper is organized as follows: In Section 2, the FEM-based clustering framework is introduced and its extension towards financial data context is presented. In Section 3, we have a short look at the definition of return and adapt this to *futures markets*, as the classical definition does not fit in this context. The numerical futures portfolio optimization approach is presented. Section 4 demonstrates the application of the proposed numerical framework to simulated and real data, comparison to standard approaches (like the classical methods proposed by Markowitz [26], the flexible GARCH(1,1) model [23] and the simple $\frac{1}{n}$-portfolio). In the end, a practical side constraint is considered, the market neutrality.

2. Finding Market Phases.

In the following, we will briefly described the clustering framework first in general introduced in [14, 15, 13, 16]. Special emphasis will be given to the numerical and interpretational aspects of the framework in context of high dimensional financial applications.

2.1. The Clustering Problem.

Let $S_t : [0, T] \to [0, \infty)^d$ be the observed d-dimensional series of assets prices. We look for K *models* characterized by K distinct sets of a priori unknown *model parameters* $\theta_1, \ldots, \theta_K \in \Omega$, where Ω is the parameter space. Let $g(S_t, \theta_i) : \mathbb{R}^d \times \Omega \to [0, \infty)$ be some *distance function* of the observation S_t and the set of parameters θ_i. For the work done throughout this paper, the θ_i correspond to the changing *market phases*. Then the vector $\Gamma_t = (\gamma_1(t), \ldots, \gamma_K(t))$, called the *affiliation vector*, gives the probability to be in the ith market phase at time t. For a given observation S_t we want to solve the *clustering problem*

$$\sum_{i=1}^{K} \gamma_i(t) g(S_t, \theta_i) \to \min_{\Gamma_t, \Theta}, \tag{2.1}$$

for $\Theta = (\theta_1, \ldots, \theta_K)$. Moreover, as the $\gamma_i(t)$ shall be probabilities, they are subject to the constrains

$$\sum_{i=1}^{K} \gamma_i(t) = 1, \quad \forall t \in [0, T] \tag{2.2}$$

$$\gamma_i(t) \geq 0, \quad \forall t \in [0, T],\ i = 1, \ldots, K. \tag{2.3}$$

As we do not want to solve this problem for a specific time t only, but for the whole time interval $[0, T]$, we now introduce the *averaged clustering functional* L:

$$L(\Theta, \Gamma) = \int_0^T \sum_{i=1}^{K} \gamma_i(t) g(S_t, \theta_i) \to \min_{\Gamma, \Theta}, \tag{2.4}$$

subject to the constrains (2.2-2.3).[1] As was demonstrated in [14, 15, 13, 16] there are practical difficulties in solving problem (2.4) numerically as:

1. the problem is infinite dimensional since $\gamma_i(t)$ is an element of some (not yet specified) function space.
2. the problem is *ill-posed* since the number of unknowns can be higher then the number of known parameters.
3. g is, in general, non-convex and thus only locally optimal solutions might be found.

Problems 1 and 2 are addressed within section (2.2 and 2.3), how to evade problem 3 is shown in section (2.4).

2.2. Persistent Market Phases. As was shown in [14, 15, 13, 16], one of the possibilities to overcome the above mentioned problem is to impose some additional assumptions on Γ. As we look for hidden market phases, we might assume the process switching between phases is slow: When the market switches to a new state, it will stay there for some time. This leads to a new constraint on Γ:

$$\int_0^T (\partial_t \gamma_i(t))^2 \, dt \le \gamma_i^\varepsilon, \quad i = 1, \dots, K, \gamma_i^\varepsilon \in (0, \infty). \tag{2.5}$$

While this reduces the transitions between different states, it is hard to handle this inside the optimization, thus we expand our problem to the *regularized form* as suggested by `Tikhonov` for linear least-square problems [31]:

$$L^\varepsilon(\Theta, \Gamma) = L(\Theta, \Gamma) + \varepsilon^2 \sum_{i=1}^{K} \int_0^T (\partial_t \gamma_i(t))^2 \, dt. \tag{2.6}$$

As was shown in [14], ε^2 can be used to control the persistents of the cluster states, thus preventing a short-term switch into another market phase.

2.3. Finite Elements and discretization. Discretization is, from a data point of view, not really an additional assumption as market data, even if given intraday, is ticked data, thus it is discrete by nature. Continuity of the data is a modeling idealization. Additionally, this limits the function space from which Γ is chosen, addressing problem 1 and 2. Thus let $\mathcal{T}$ be a countable set of finite intervals on $[0, T]$:

$$\mathcal{T} = \{0 = t_1, t_2, \dots, t_N = T\}, \quad t_i < t_{i+1} \, \forall i = 1, \dots, N-1. \tag{2.7}$$

Then we can choose a *finite element basis* $\{v_1(t), \dots, v_N(t)\}$ with

$$v_1(t) \neq 0 \text{ for } t \in (t_1, t_2), v_1(t) = 0 \text{ for } t \notin (t_1, t_2) \tag{2.8}$$

$$v_k(t) \neq 0 \text{ for } t \in (t_{k-1}, t_{k+1}), v_k(t) = 0 \text{ for } t \notin (t_{k-1}, t_{k+1}), \quad k = 2, \dots, N-1 \tag{2.9}$$

$$v_N(t) \neq 0 \text{ for } t \in (t_{N-1}, t_N), v_N(t) = 0 \text{ for } t \notin (t_{N-1}, t_N) \tag{2.10}$$

[1] If the Θ_i are known or fixed to some value, the minimization problem (2.4) can be solved analytically w.r.t. Γ.

$$\gamma_i(t) = \begin{cases} 1, & i = \arg\min g(S_t, \Theta_i) \\ 0, & \text{else} \end{cases}$$

This allows us to write γ_i as

$$\gamma_i = \sum_{k=1}^{N} \hat{\gamma}_i^k v_k + \delta_i^N \tag{2.11}$$

where $\delta^N = \max\{\delta_i^N, i = 1, \dots, K\}$ is the discretization error. Now equation (2.6) could be written as

$$L^\varepsilon(\Theta, \Gamma) = \sum_{i=1}^{K} \sum_{k=1}^{N} \hat{\gamma}_i^k \int_0^T v_k(t) g(S_t, \theta_i)\, dt + \varepsilon^2 \sum_{i=1}^{K} \int_0^T \left(\sum_{k=1}^{N} \hat{\gamma}_i^k \partial_t v_k(t) \right)^2 dt + O(\delta^N) \tag{2.12}$$

Using the local support of v_k, we can define

$$\alpha(\theta_i) = \left(\int_{t_1}^{t_2} v_1(t) g(S_t, \theta_i)\, dt, \dots, \int_{t_{N-1}}^{t_N} v_N(t) g(S_t, \theta_i)\, dt \right), \tag{2.13}$$

and

$$H = \begin{pmatrix} \int_{t_1}^{t_2} v_1^2(t)\, dt & \int_{t_1}^{t_2} v_1(t) v_2(t)\, dt & 0 & \dots \\ \int_{t_1}^{t_2} v_1(t) v_2(t)\, dt & \int_{t_2}^{t_3} v_2^2(t)\, dt & \int_{t_2}^{t_3} v_2(t) v_3(t)\, dt & \dots \\ 0 & \ddots & \ddots & \ddots \end{pmatrix}, \tag{2.14}$$

to create the *discrete clustering problem*, see for example [14]

$$\widetilde{L^\varepsilon}(\Theta, \Gamma) = \sum_{i=1}^{K} \left(\alpha(\theta_i)^T \hat{\gamma}_i + \varepsilon^2 \hat{\gamma}_i^T H \hat{\gamma}_i \right) \to \min_{\Gamma, \Theta}. \tag{2.15}$$

Please note, that up to now, all assumptions made are solely data-concerned and we did not make any probabilistic assumptions on the distribution of the data.

2.4. The distance function. When working with financial data, one might end up with thousands of assets, as, for example, the Wilshire 5000 index consists of 5.000 stocks. With respect to this high dimensionality, one has to think about an easy way to handle these data. The popular idea is to use the *principle component analysis* (short: PCA, see [32, 19]), where only a linear low-dimensional manifold of the data space is used to represent the original. The drawback here is the linearity of the manifold, as a globally optimal linear subspace might not exist. Thus let $T_i \in \mathbb{R}^{d \times n}$ for $i = 1, \dots, K$ and $n < d$ with $T_i^T T_i = \mathtt{Id}$ be the *projection matrices*, and $x_t \in \mathbb{R}^d$ be some time series deduced from S_t, then we can define the distance function by

$$g(x_t, \theta_i) = \|(x_t - \mu_i) - T_i T_i^T (x_t - \mu_i)\|_2^2, \quad \theta_i = (\mu_i, T_i). \tag{2.16}$$

This distance function measures the *reduction error* resulting from the projection of the d-dimensional data on the n-dimensional linear manifold i. x_t could be the returns and T_i could be interpreted as the direction of maximal volatility of x, while μ_i describes the mean behavior [17]. We can now try to get better results by using the time-dependent function $\mu_i(t)$ instead of a constant one, e.g. approximating $\mu_i(t)$ with some regression model as in [15]. Therefore let $\varphi_i(t)$ be some basis of the functional space (e.g. monomials), then the distance function changes to

$$g(x_t, \theta_i) = \|(x_t - \sum_{j=1}^{\omega} \mu_i^j \varphi_j(t)) - T_i T_i^T (x_t - \sum_{j=1}^{\omega} \mu_i^j \varphi_j(t))\|_2^2, \quad \theta_i = ((\mu_i^1, \dots, \mu_i^k), T_i), \tag{2.17}$$

where ω is the *order of the trend*. In fact, we can see $\sum_{j=1}^{\omega} \mu_i^j \varphi_j(t)$ as the smoothed trend so this analysis can be interpreted as a de-trending. Additionally we can estimate the *covariance matrix* of a clustered set (x_t) of the time series for some i by

$$\Sigma_i^{\bar{T}} = \frac{1}{\sum_{t \le \bar{T}} \gamma_i(t)} \sum_{t \le \bar{T}} \gamma_i(t) \left(x_t - \sum_{j=1}^{\omega} \mu_i^j \varphi_j(t) \right) \left(x_t - \sum_{j=1}^{\omega} \mu_i^j \varphi_j(t) \right)^T \tag{2.18}$$

Please note: The combination of PCA, trend estimation and clustering is not just a direct chaining of both algorithms. The estimation of trends and covariance matrices is *embedded* in the minimization problem (2.15).

2.5. Calculating the parameters. Given the analyzed time series x_t and a cluster affiliation Γ, we can now compute the cluster parameters μ_i^j and T_i. Therefore we have to solve for each $i = 1, \dots, K$ the problem

$$\alpha(\theta_i)^T \hat{\gamma}_i + \varepsilon^2 \hat{\gamma}_i^T H \hat{\gamma}_i \rightarrow \min_{\theta_i}, \tag{2.19}$$

$$\Leftrightarrow \qquad \alpha(\theta_i)^T \hat{\gamma}_i \qquad \rightarrow \min_{\theta_i}. \tag{2.20}$$

Using the definition of α and g we get

$$\sum_{k=1}^{N} \gamma_i^k \int_{t_{k-1}}^{t_{k+1}} v_k(t) \|(x_t - \sum_{j=1}^{\omega} \mu_i^j \varphi_j(t)) - T_i T_i^T (x_t - \sum_{j=1}^{\omega} \mu_i^j \varphi_j(t))\|_2^2 \, dt \rightarrow \min_{\mu_i, T_i}, \tag{2.21}$$

with $t_0 = t_1$ and $t_{N+1} = t_N$. First order conditions of optimality imply that partial derivatives w.r.t. μ_i^j and T_i vanish, hence for μ we obtain

$$\begin{aligned} 0 = & \sum_{k=1}^{N} \gamma_i^k \int_{t_{k-1}}^{t_{k+1}} v_k(t) (T_i T_i^T \varphi_l(t) - \varphi_l(t))^T (x_t - T_i T_i^T x_t) \, dt \\ & - \sum_{j=1}^{\omega} \mu_i^j \sum_{k=1}^{N} \gamma_i^k \int_{t_{k-1}}^{t_{k+1}} v_k(t) (T_i T_i^T \varphi_l(t) - \varphi_l(t))^T (\varphi_j(t) - T_i T_i^T \varphi_j(t)) \, dt \end{aligned} \tag{2.22}$$

holds and an analogous equation is obtained for T_i. The latter yields that T_i consists of the eigenvectors corresponding to the n largest eigenvalues of the empirical covariance matrix of the values associated to state i. Now the problem (2.15) could be solved by subspace iteration, thus starting with some arbitrary Γ_0, we solve iteratively solve

$$\Theta_k = \arg\min_{\Theta} \widetilde{L^{\varepsilon}}(\Theta, \Gamma_k) \tag{2.23}$$

and

$$\Gamma_{k+1} = \arg\min_{\Gamma} \widetilde{L^{\varepsilon}}(\Theta_k, \Gamma) \tag{2.24}$$

for $k = 0, 1, \dots$. Although this algorithm is not guaranteed to converge to the global optimum, at least a local minimum will be found. By performing this algorithm several times with different starting values Γ_0, we can approximate the global optimal parameters.

3. Adapting portfolio optimization theory to futures. The usual definition of returns $R_t = \frac{S_{t+\Delta t} - S_t}{S_t}$ can not be naively applied to *futures data*, as the price for a contract is not S_t. Instead, no price is charged for entering a futures contract, but an initial amount is paid into the *margin account* that will hold the accrued gains and losses. So given an *initial margin* of $I \in \mathbb{R}^d$ and a *portfolio* π *in proportion of wealth*, we define returns for *futures* over the period t to $t + \Delta t$ as

$$R_t = \operatorname{sgn}(\pi_t) \frac{S_{t+\Delta t} - S_t}{I} \quad \text{(component-wise)}. \tag{3.1}$$

In contrast to portfolio optimization approaches for the stocks market, it is an important feature of futures that long and short positions can be held equally well. But since the *initial margin* is depending only on the absolute contract size, our portfolio constraints will change, the absolute weights of the components should sum up to one

$$\sum_{i=1}^{d} |\pi_i(t)| = 1, \quad \forall t \in [0, T]. \tag{3.2}$$

Given suitable estimates Σ, μ for covariances and means of returns, the portfolio problem is posed as

$$\pi^T \Sigma \pi \to \min_{\pi}, \quad \text{with } \|\pi\|_1 = 1,\, \pi^T \mu \geq C, \tag{3.3}$$

for a chosen target mean return C. Since this is not yet (as in classical mean variance optimization) a quadratic minimization problem with linear (or quadratic) constrains, we use a splitting into long positions and short positions, similar to the idea explained in [9] in context of image processing, to get an *extended parameter set*

$$\hat{\pi} = (\pi^{\text{long}}, \pi^{\text{short}})^T, \quad \widehat{\Sigma} = \begin{pmatrix} \Sigma & -\Sigma \\ -\Sigma & \Sigma \end{pmatrix}, \quad \hat{\mu} = (\mu, -\mu)^T, \tag{3.4}$$

and apply a penalty term. We define the *extended problem* by

$$\hat{\pi}^T \left(\widehat{\Sigma} + \lambda \begin{pmatrix} 0 & \mathbb{I}_d \\ \mathbb{I}_d & 0 \end{pmatrix} \right) \hat{\pi} \to \min_{\hat{\pi} \in \mathbb{R}^{2d}}, \quad \text{with } \hat{\pi}^T \hat{\mu} \geq C,\, \sum_{i=1}^{2d} \hat{\pi}_i = 1,\, \hat{\pi}_i \geq 0\, \forall i \tag{3.5}$$

with the regularization parameter λ. Please note that for sufficiently large λ, the part

$$\lambda \hat{\pi}^T \begin{pmatrix} 0 & \mathbb{I}_d \\ \mathbb{I}_d & 0 \end{pmatrix} \hat{\pi} = 2\lambda (\pi^{\text{long}})^T \pi^{\text{short}}, \tag{3.6}$$

and the condition $\hat{\pi}_i \geq 0\, \forall i$ guarantee, that for each asset the position is either long or short, but not both at the same time. Simultaneously, the remainder of the extended problem is just

$$\hat{\pi}^T \widehat{\Sigma} \hat{\pi} = (\pi^{\text{long}} - \pi^{\text{short}})^T \Sigma \pi^{\text{long}} - (\pi^{\text{long}} - \pi^{\text{short}})^T \Sigma \pi^{\text{short}} = \pi \Sigma \pi, \tag{3.7}$$

and

$$\hat{\mu}^T \hat{\pi} = \mu^T \pi^{\text{long}} - \mu^T \pi^{\text{short}} = \mu^T \pi. \tag{3.8}$$

4. Numerical examples. Throughout the first part of this section, we will make use of the simulated time series. These time series have a length of 1000 with a change point at time $t_C = 500$. All of the data points are distributed independently according to $\mathcal{N}(0, \Sigma)$ where

$$\Sigma(t) = \alpha^T(t) \begin{bmatrix} 0.85 & 0.4 \\ 0.4 & 0.2 \end{bmatrix} \alpha(t), \quad \alpha(t) = \begin{cases} \mathbb{I}_2, & t \leq 500 \\ \begin{bmatrix} \cos(\rho) & -\sin(\rho) \\ \sin(\rho) & \cos(\rho) \end{bmatrix}, & t > 500 \end{cases}. \tag{4.1}$$

As one can see, after the change point at t_C, the covariance matrix is just rotated by some ρ.

4.1. Covariance for small samples. As known from the literature, for small sample sizes, the maximum likelihood estimator $\frac{1}{k} \sum_{t=t_1}^{t_k} x_t x_t^T$ is not robust enough, e.g. [29, 4]. This will lead to numerical biassing when it comes to portfolio optimization [27, 24, 25]. We will use the Tyler-M-Estimator (e.g. [33, 8]) instead, that is defined by

$$\hat{\Sigma}^k = \frac{d}{k} \sum_{t=t_1}^{t_k} \frac{x_t x_t^T}{x_t^T (\hat{\Sigma}^k)^{-1} x_t}. \tag{4.2}$$

Although the solution $\hat{\Sigma}^k$ to equation (4.2) is only unique up to a multiplicative constant, this is not a problem for the clustering method (2.15) as neither the result of the minimization nor the eigenvectors (for PCA) are dependent on this constant. For figure (4.1), the 100 first points of 100 time series were used (using the distribution stated above) and for every time step $t \in 2, \ldots, 100$ the maximum likelihood estimator and the Tyler-M-Estimator were calculated. Then the largest singular value of the difference $(\hat{\Sigma} - \Sigma)$ was used as a distance measure and the average and standard-deviation were calculated and plotted. As one can see from this example, the standard-deviation of the Tyler-M-Estimator is much smaller then the one for the maximum likelihood estimator (also both converge to the same point).

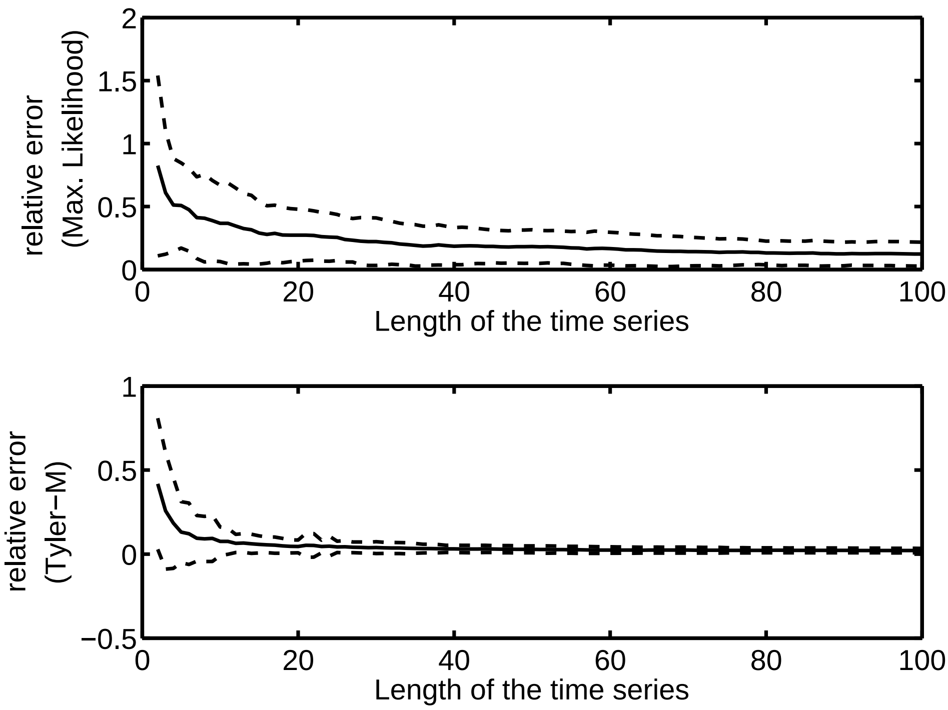

FIGURE 4.1. *Robustness of the MLE and TME for short normally distributed time series. Parameters as in equation (4.1).*

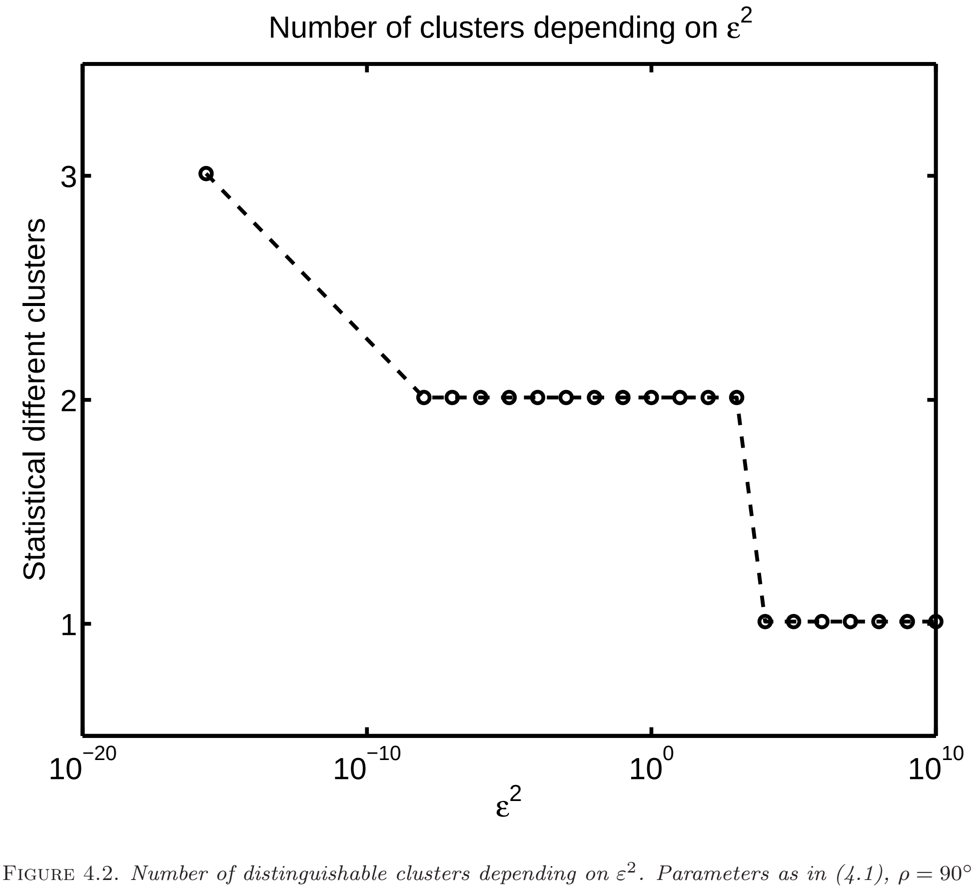

FIGURE 4.2. *Number of distinguishable clusters depending on ε^2. Parameters as in (4.1), $\rho = 90°$.*

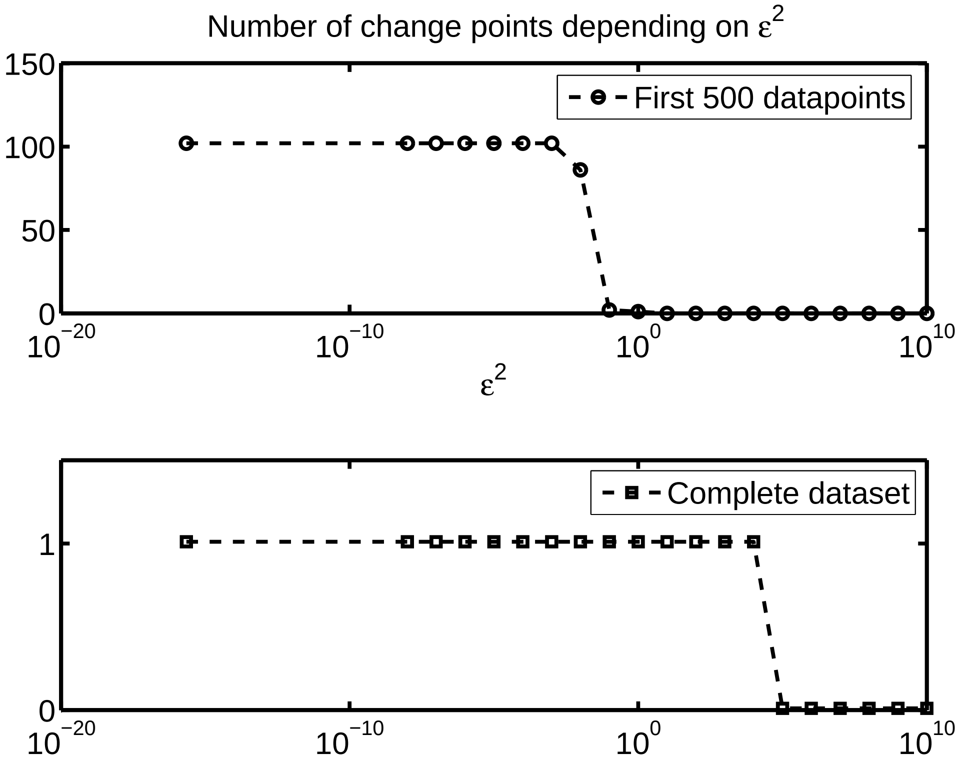

FIGURE 4.3. *Number of detected change points depending on ε^2 for half of the (artificial) data and the complete data set. Parameters as in (4.1), $K = 2, \rho = 90°$.*

4.2. Number of Clusters. In realistic applications, the optimal number of meta-stable clusters is usually a priori unknown. Nevertheless, one can find out the number of statistically distinguishable clusters (see [14]). Therefore we start the clustering process with a rather high number of clusters, estimate the cluster parameters and get the error bounds of for these parameters by bootstrapping of the elements of the cluster. If the parameter sets of different clusters overlap, the statistical difference between those clusters is not high enough thus the number of clusters should be reduced and the process should be started over again [21]. Using this technique for different ε leads to a figure like (4.2). The decrease in the number of cluster is due to the fact that for the increasing ε the algorithm favors meta-stable clusters and merges the rapidly mixing ones together. Thus, while increasing ε, similar clusters will get merged into one. The plateau results from the high difference between the last two remaining clusters. For a real data set we will observe similar behavior when we look for the number of clusters.

4.3. Detecting a change point. Another important problem is to find the actual position of the change points. The number of detected change points depends on ε as one can see in figure (4.3). Although, a high number of change points usually indicates some rapidly mixing clusters that could be united to a meta-stable cluster (e.g. using half the data in figure (4.3)). Decreasing the number of clusters and increasing ε might help to solve this problem. On the other hand, a high value for ε will lead to a longer delay between occurrence and online detection of the change point, this can be seen in figure (4.4). The same problem appears, when the rotation angel becomes to small, see figure (4.5).

4.4. Looking at real data. For the calculations done in this paper, we used the futures prices of the years 2005-2008 for wheat and oil futures (four time to maturity each, quarterly for wheat and monthly for oil, daily closings), as reported at
`http://www.kcbt.com/historical_data.asp` (Kansas City Board of Trade) and

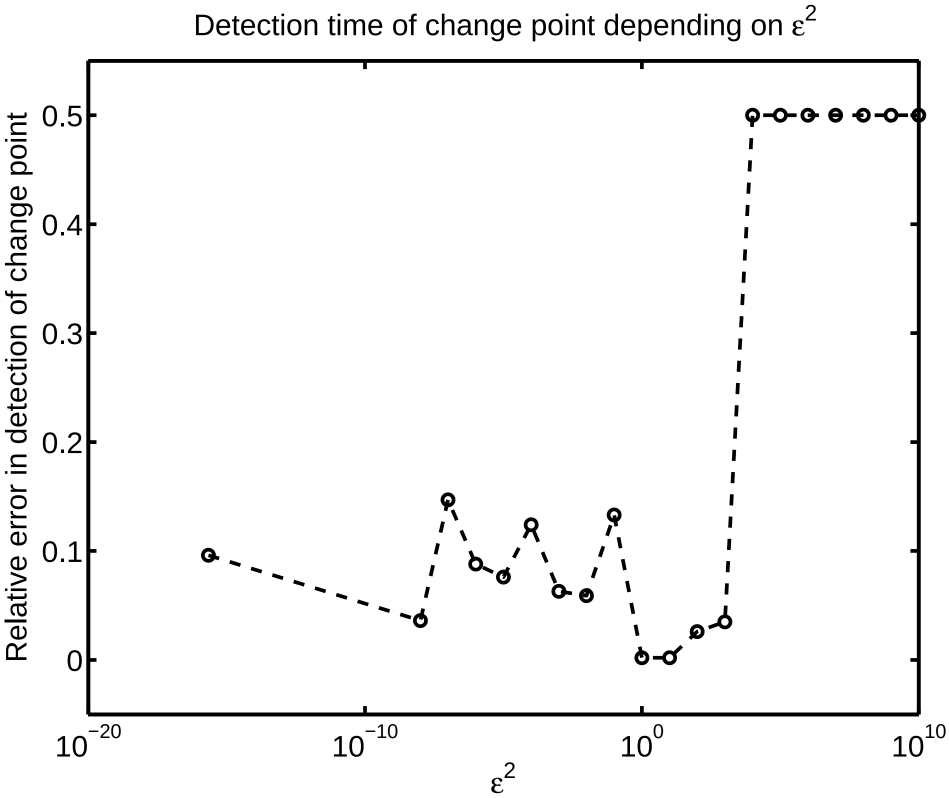

FIGURE 4.4. *Relative error in detection time of the single change point depending on* ε^2. *Parameters as in (4.1),* $K = 2, \rho = 90°$.

http://tonto.eia.doe.gov/dnav/pet/pet_pri_fut_s1_d.htm (U.S. Department of Energy). For both commodities the four futures with the lowest time to maturity were taken into account. Applying the introduced clustering algorithm to these data gives the result shown in figure (4.6). Implied by this picture, an optimal number of meta-stable clusters to use should be two. Than the cluster allocation for the whole time series, see figure (4.7), produces a kind of seasonal cycle that could be due to the harvest cycle of the wheat. The phases differ in the following points:

- The phase shown in figure (4.7), thus phase 1, is characterized by lesser overall trend in the price than phase 2.
- The general direction of the maximum volatility (described by the projection operator T_i from equation (2.15)) in both phases is similar, the first eigenvectors of the covariance matrices are nearly equal, the second eigenvectors span an angle of approximately 5 degrees.
- In phase 1, the optimal portfolio consists of a long position in the wheat futures with longest maturity and short positions in all other wheat futures. Approximately 20% of the wealth are invested in wheat, while the remainder is invested in oil, with a short position in the oil futures with the all but one shortest maturity. In phase 2, the wheat futures with longest and shortest maturity are sold short, with approximately 22% in wheat futures. And for the oil futures, only one long position exists, bought is the one with the all but one longest maturity.

4.5. Comparing portfolios. When looking at the interest rate for U.S. treasury bonds, an investment in January 2005 would have given a yearly interest rate of approx 3.4%. Thus after four years we have a gain of 14.31% (thats a daily rate of 1.415 *bps*). We use this rate as a target rate C (as in equation (3.5)). As stated earlier, we compare our algorithm with the Markowitz approach,

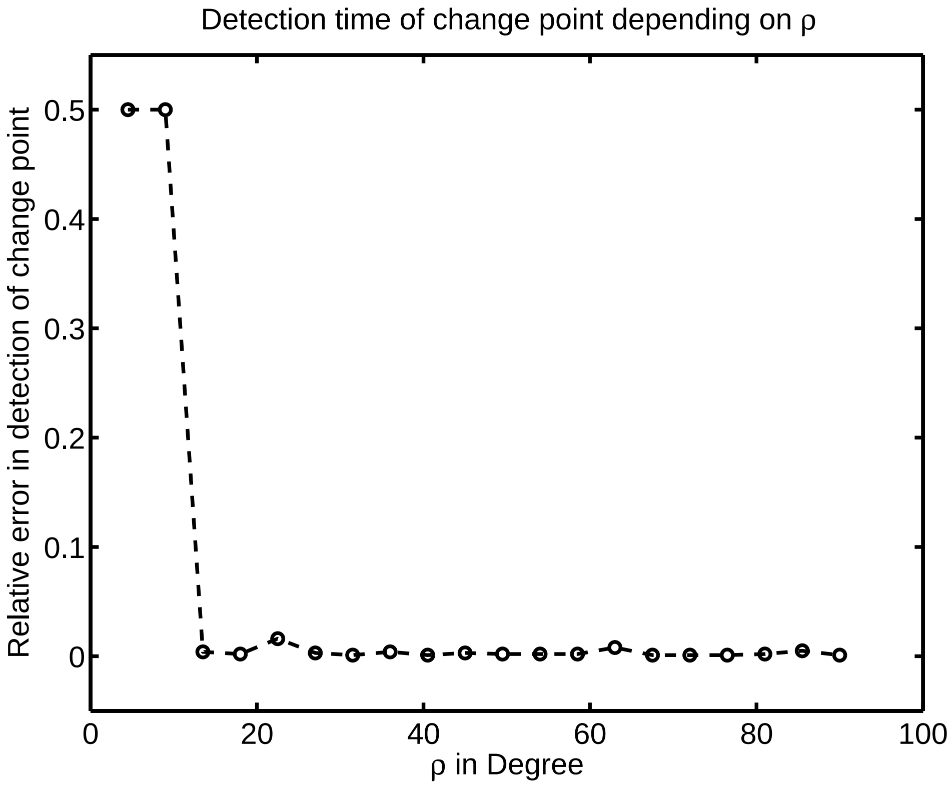

FIGURE 4.5. *Relative error in detection time of the single change point depending on* ρ*. Parameters as in (4.1),* $K = 2, \varepsilon^2 = 1$*.*

the FlexM algorithm from [23] and a simple '$\frac{1}{n}$'-portfolio, where half the assets are bought, the other half sold. For the cluster-based approach, the basis φ with polynomials up to order two was taken. The clustering was done with respect to μ and T. The FlexM algorithm provided by the authors of [23] was used utilizing the parameter `demean`, thus the estimation and subtraction of the mean from the data prior to the portfolio optimization is done by the FlexM-function. The $\frac{1}{n}$-portfolio was build by buying for $\frac{1}{8}$ of the wealth the long maturity futures of each commodity and selling the remainder equally over the short-time futures. Additionally, the Markowitz approach was used by estimating mean and covariance from the whole time series up to the actual point. All algorithm were only supplied with information avaible at time t, so no future knowledge is included. In each case, the portfolio was recalculated and/or re-balanced every 10 business days starting with day 10 of the time series. As one can see in figure (4.8), all algorithms fall short of the target 14.31% mean return. While $\frac{1}{n}$-portfolio is doing better (in terms of return-rate) than the other algorithms, this comes with a higher risk, as one can see in figure (4.9). We define the 'intrinsic risk estimate' for portfolio returns from t to $t + \Delta t$ by

$$\mathcal{R}_t = \sqrt{\pi_t^T \mathrm{Cov}\ (R_t)\pi_t}, \tag{4.3}$$

where the empirical returns R_t are given by equation (3.1), the Cov (R_t) is an estimate for the covariance matrix of the data (based upon data only up to t), which is given by:

- a linear combination of the empirical cluster covariance matrices given by equation (2.18) for the cluster-based algorithm, thus

$$\mathrm{Cov}\ (R_t) = \sum_{i=1}^{K} \gamma_i(t)\Sigma_i^t. \tag{4.4}$$

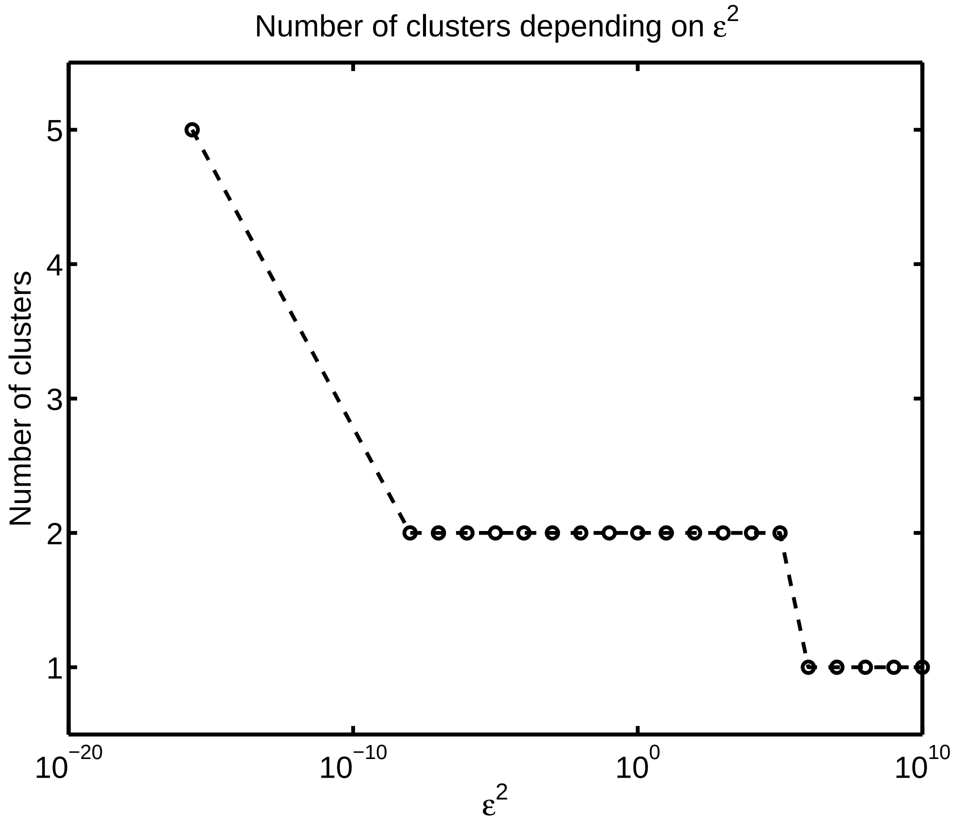

FIGURE 4.6. *Number of distinguishable clusters depending on ε^2, Data set: Prices of oil and wheat futures 2005-2008.*

- the GARCH(1,1) estimate for the FlexM algorithm (see [23]),
- the standard sample covariance matrix for the Markowitz approach and the $\frac{1}{n}$-Portfolio

Although those intrinsic risks shown in figure (4.9) are estimates of the instantaneous variance of the portfolio returns, they are not directly comparable, as the measures differ from each other and the underlying assumptions are different, so in figure (4.10) risk is meant in the sense of standard deviation of the daily returns of the portfolios from the empirical mean, still the cluster-based approach produces the strategy with the lowest risk. And even if we use the deviation from the target mean return C

$$\sqrt{\frac{1}{k-1}\sum_{t=t_1}^{t_{k-1}}(\pi_t^T R_t - C)^2} \tag{4.5}$$

as a risk measure, see results in figure (4.11), the cluster-based approach reaches the best result. The similarity of this risk measure and the standard deviation is easy to see: If the variance minimal portfolio does not satisfy the condition $\hat{\pi}_{\min}^T \hat{\mu} \geq C$, i.e. the target mean return is not reached, due to the convexity of the problem the solution of (3.5 will satisfy $\hat{\pi}^T \hat{\mu} = C$. This guarantees, that for sufficiently large C the expected mean is C and thus equation (4.5) would give the standard deviation. The advantage of equation (4.5) over the empirical standard deviation is the fact that the deviation from the target mean return does not depend on the empirical expectation of the portfolio returns. Indeed: As stated in table (4.1), the empirical standard deviation of the portfolio returns and the empircial deviation from the target mean return are quite near to each other. The cluster-based approach reaches 69.85% of the performance of the $\frac{1}{n}$-portfolio, while it only takes 40.33% of the risk (in sense of empirical standard deviation). Thereby rate of return is the total

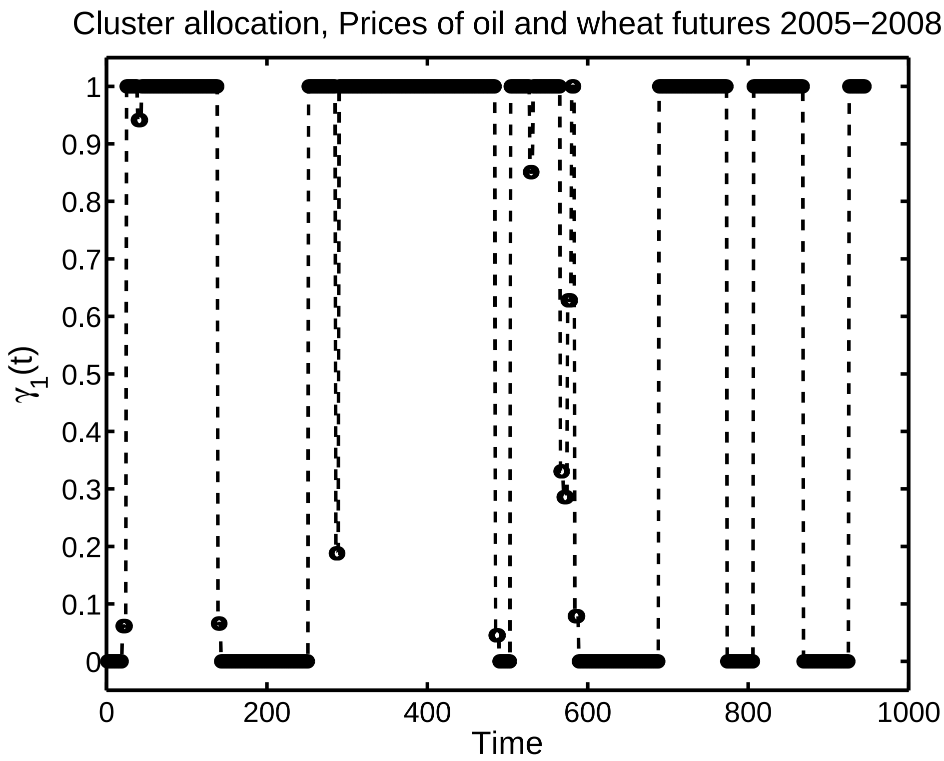

FIGURE 4.7. *Cluster allocation for time series of futures. $K = 2, \varepsilon^2 = 1$, Data set: Prices of oil and wheat futures 2005-2008.*

return minus one and average intrinsic is defined by

$$\frac{1}{N-1}\sum_{t=t_1}^{t_{N-1}} \mathcal{R}_t \tag{4.6}$$

For different target returns the results are similar, also the cluster-based portfolio is less sensible to the target return rate (see figures (4.12,4.13)).

Algorithm	rate of return	averaged intrinsic Risk	standard deviation of returns	deviation from target mean return
cluster-based	4.01%	$1.4164\ 10^{-3}$	$2.4613\ 10^{-3}$	$2.4619\ 10^{-3}$
FlexM	−0.55%	$7.5113\ 10^{-3}$	$7.8968\ 10^{-3}$	$7.8935\ 10^{-3}$
Markowitz	0.30%	$1.5180\ 10^{-3}$	$3.9324\ 10^{-3}$	$3.9325\ 10^{-3}$
$\frac{1}{n}$-Portfolio	5.74%	$3.7022\ 10^{-3}$	$6.1033\ 10^{-3}$	$6.1003\ 10^{-3}$

TABLE 4.1
Rate of return and different risk measures

4.6. Adaption for market neutral investment. A common relevant strategy of portfolio managers is so-called *market neutral investment*. The aim is to bet only from views on relative changes between asset prices, but remain in neutral w.r.t. overall market moves. This is done by an additional constraint to ensure that the accrued notional of contracts within each class u (here wheat and oil) is equal to zero.

$$\sum_{i\in I_u} \pi_i = 0, \quad \forall u, \tag{4.7}$$

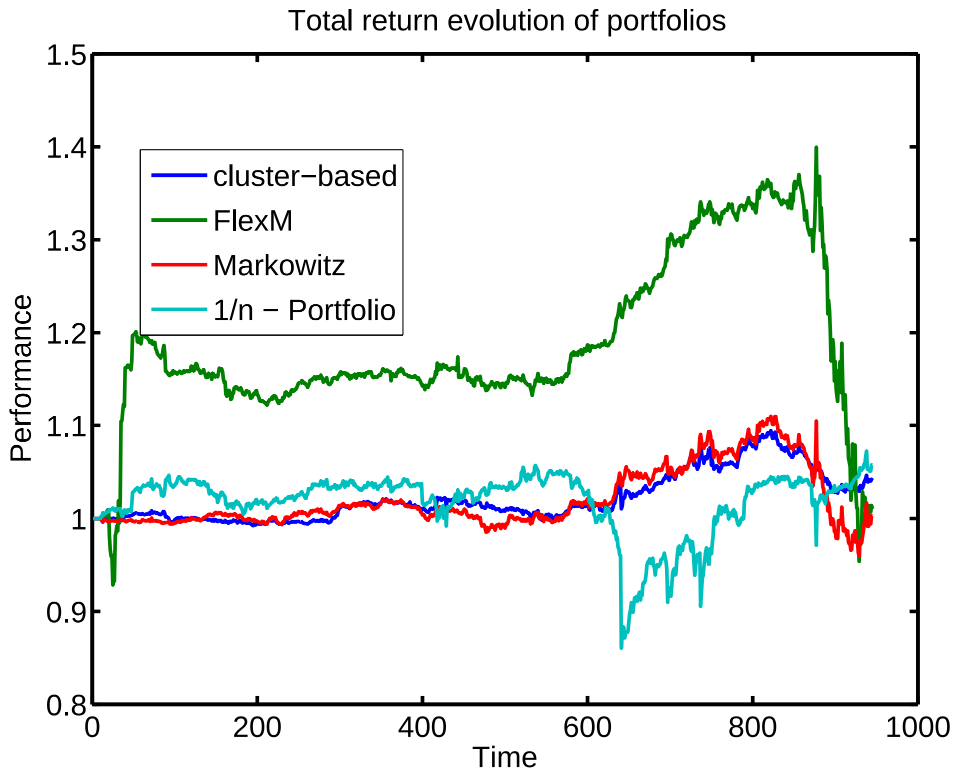

Figure 4.8. *Value evolution of different portfolio-optimization algorithms.*

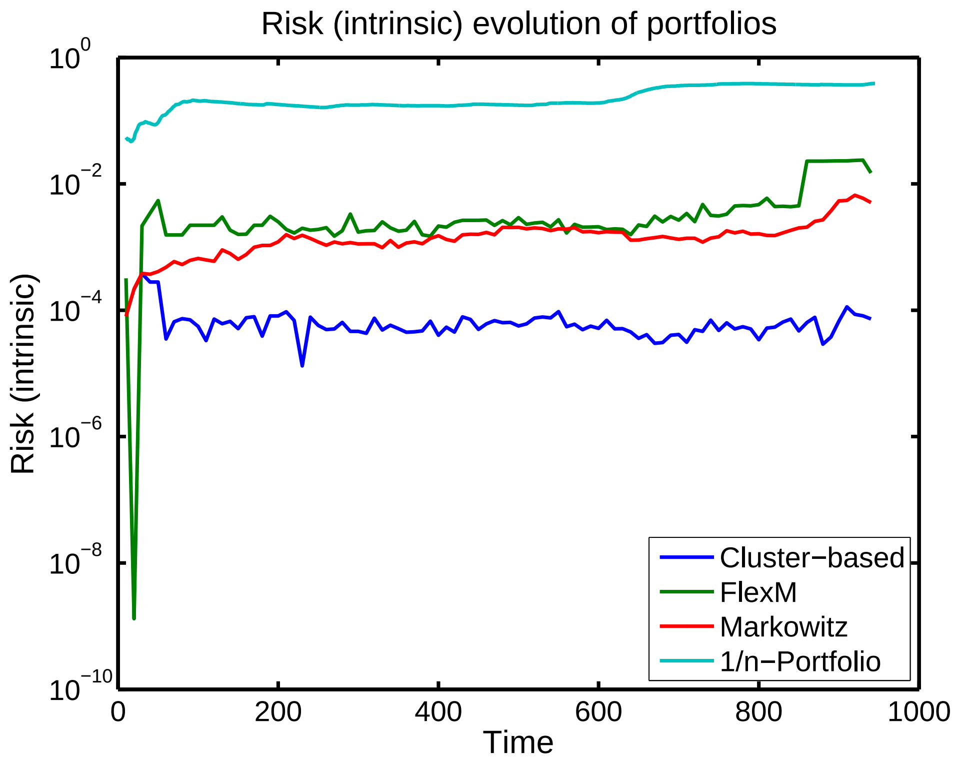

Figure 4.9. *Intrinsic risk (see equation 4.3) evolution of different portfolio-optimization algorithms.*

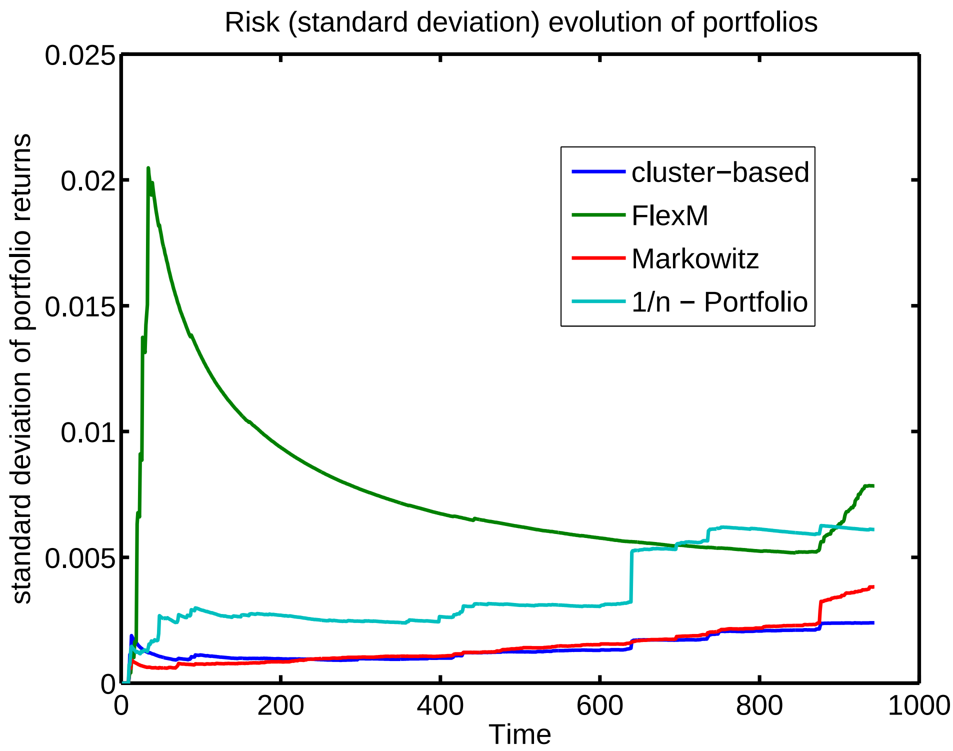

FIGURE 4.10. *Risk (empirical standard deviation) evolution of different portfolio-optimization algorithms.*

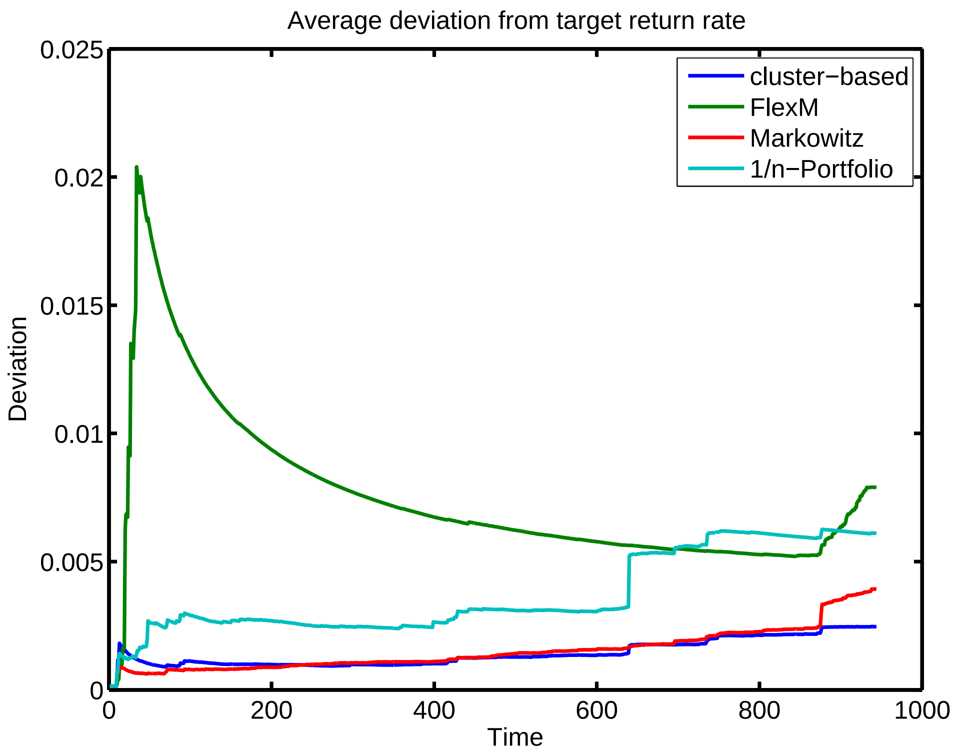

FIGURE 4.11. *Risk (deviation from target mean return) evolution of different portfolio-optimization algorithms.*

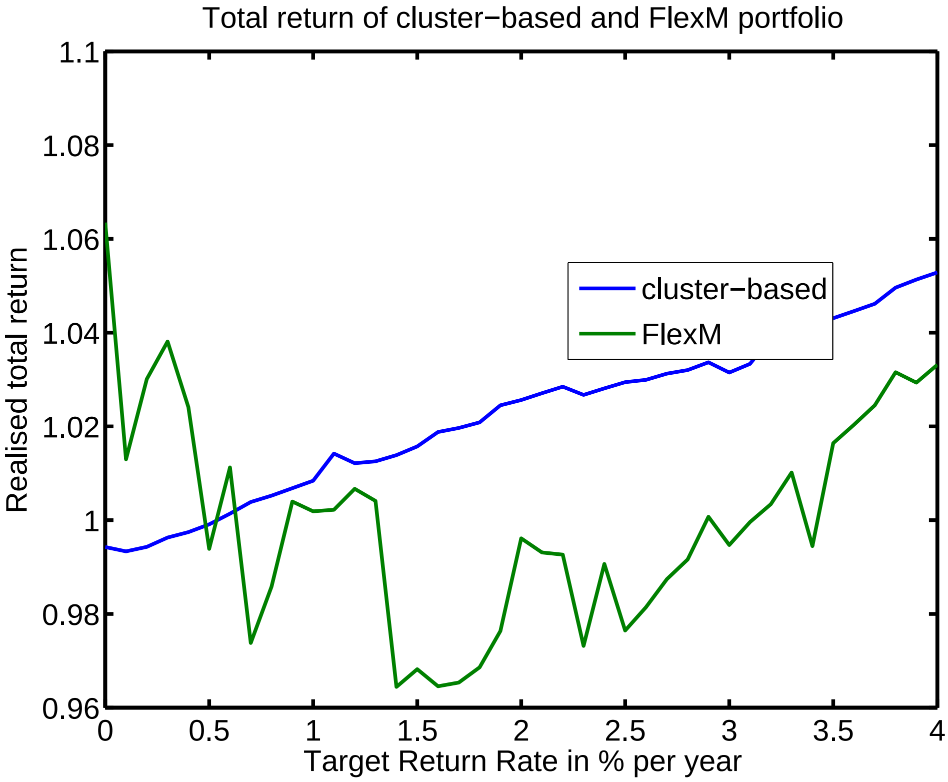

FIGURE 4.12. *Total return depending on the yearly target return rate C in (3.5).*

where I_u is an index-set combining all assets for each selected class of futures within which market-neutrality is requested. For the extended problem, this transforms to

$$\sum_{i \in I_u} (\hat{\pi}_i - \hat{\pi}_{i+d}) = 0, \quad \forall u. \tag{4.8}$$

Thus the market neutral version of the problem is

$$\hat{\pi}^T \left(\widehat{\Sigma} + \lambda \begin{pmatrix} 0 & \mathbb{I}_d \\ \mathbb{I}_d & 0 \end{pmatrix} \right) \hat{\pi} \to \min_{\hat{\pi} \in \mathbb{R}^{2d}}, \quad \text{with } \hat{\pi}^T \hat{\mu} \geq C, \sum_{i=1}^{2d} \hat{\pi}_i = 1, \, \hat{\pi}_i \geq 0 \, \forall i, \sum_{i \in I_u} (\hat{\pi}_i - \hat{\pi}_{i+d}) = 0 \tag{4.9}$$

The additional constraint was used for all algorithms (except the $\frac{1}{n}$-Portfolio). As before, the realized return for each algorithm falls short of the mean target return, see figure (4.14), but the presented cluster-based method produces the result with the least risk, see figure (4.16), and, for this data, even the best return rate, next to the $\frac{1}{n}$-Portfolio where no risk-minimization is done.

5. Conclusion. We adapted and expanded a method for simultaneous clustering, dimension reduction and meta-stability analysis of high-dimensional time series to financial data. A main advantage of the method (compared to standard HMM/GMM-methods of phase identification [12, 22]) is that no a priory probabilistic assumptions about the hidden and observed market processes are necessary in context of the presented method. Furthermore, section 4 indicates that the method is suitable for finding a low-dimensional representation of the data and estimating mean and variance while detecting market phases, delivering information that can be further used for portfolio optimization. The presented framework allows to make use of fast and numerically robust FEM-solvers (developed for numerical solution of partial differential equations) in a new context of computational time series analysis.

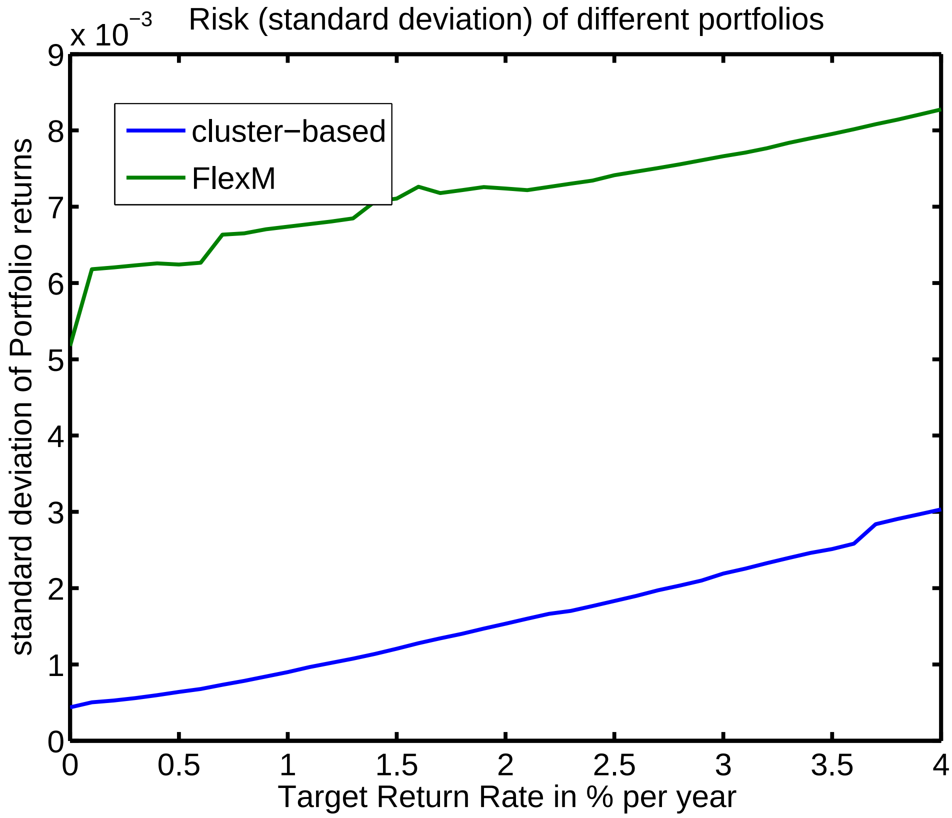

FIGURE 4.13. *Risk (empirical standard deviation) depending on the yearly target return rate C in (3.5).*

The general feasibility of identification of market phases was demonstrated and it was shown, that an expansion of classical portfolio theory utilizing market phase detection can be used to further decrease the intrinsic risk. Additionally, the behavior of the change point detection when applied online to a data stream was demonstrated to be robust enough to qualify for practical purposes. As a standard stationary risk measure, also the standard deviation of portfolio returns was used to compare different portfolio optimization strategies on real financial data. It was demonstrated that application of the presented numerical scheme based on the identification of market phases can decrease the risk of resulting portfolios in terms of this risk measure. Strictly speaking, the standard deviation of portfolio returns is not an equitably risk measure, as it implies the stationarity of the underlying data. If we assume the underlying market process to be non-stationary (as it is done in context of the presented numerical approach), the evaluation and comparison of risks should be done in a measure that reflects this. Numerical studies of robust portfolio optimization methods and empirical quantification of risk by methods that do not rely on stationarity assumptions, are a matter of future research.

Having demonstrated the applicability of the presented computational scheme to a realistic financial data example, a more extensive online data study is clearly needed for validation. Also, transaction costs should be included and different risk measures should be taken into account. In contrast to climatological [13, 15, 16, 17] or biophysical applications [18], repeating phases seem to be seldom in financial time series, an additional time weighting (e.g. exponential decay) should be further investigated.

Acknowledgments. The work was supported by the DFG reseach center Matheon "Mathematics for key technologies" in Berlin.

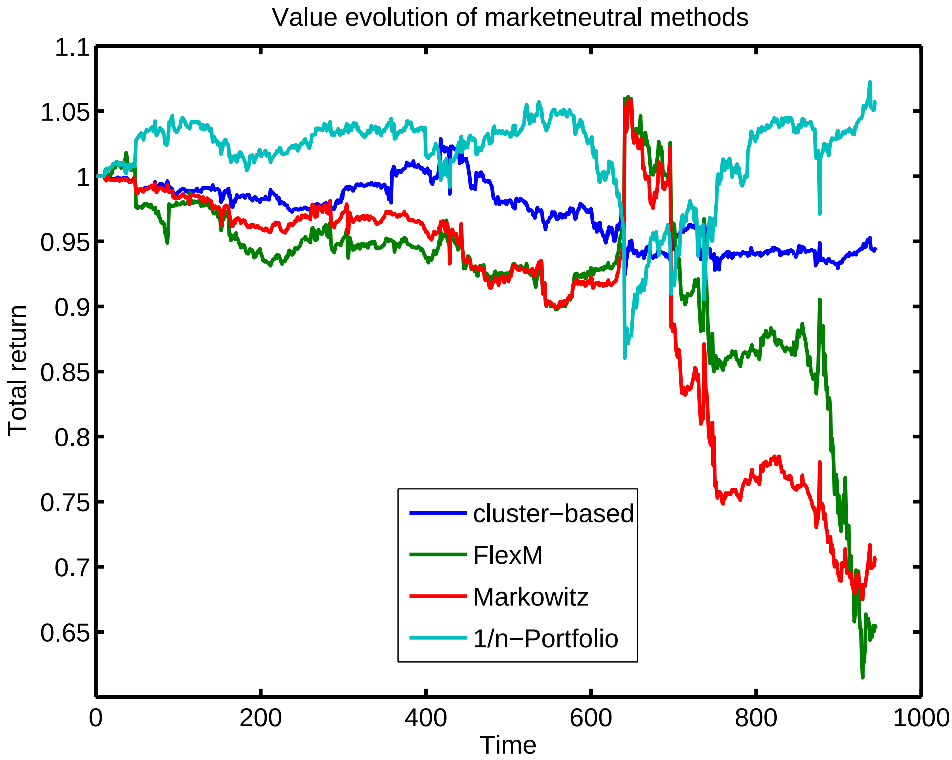

FIGURE 4.14. *Value evolution of different market neutral portfolio-optimization algorithms.*

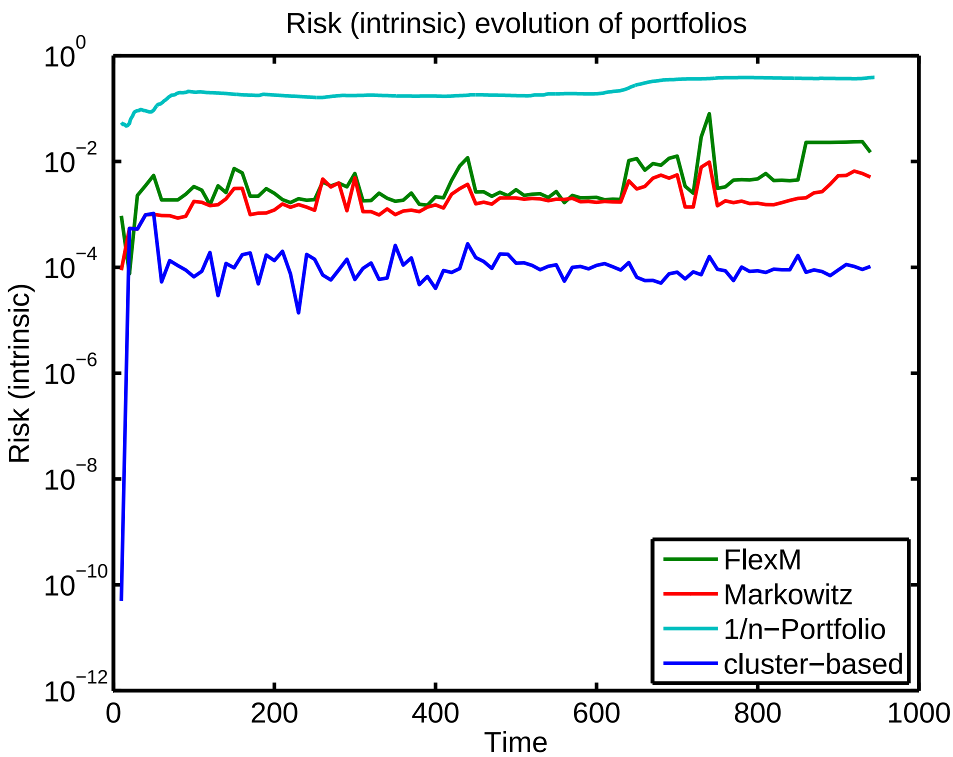

FIGURE 4.15. *Intrinsic risk (4.3) evolution of different market neutral portfolio-optimization algorithms.*

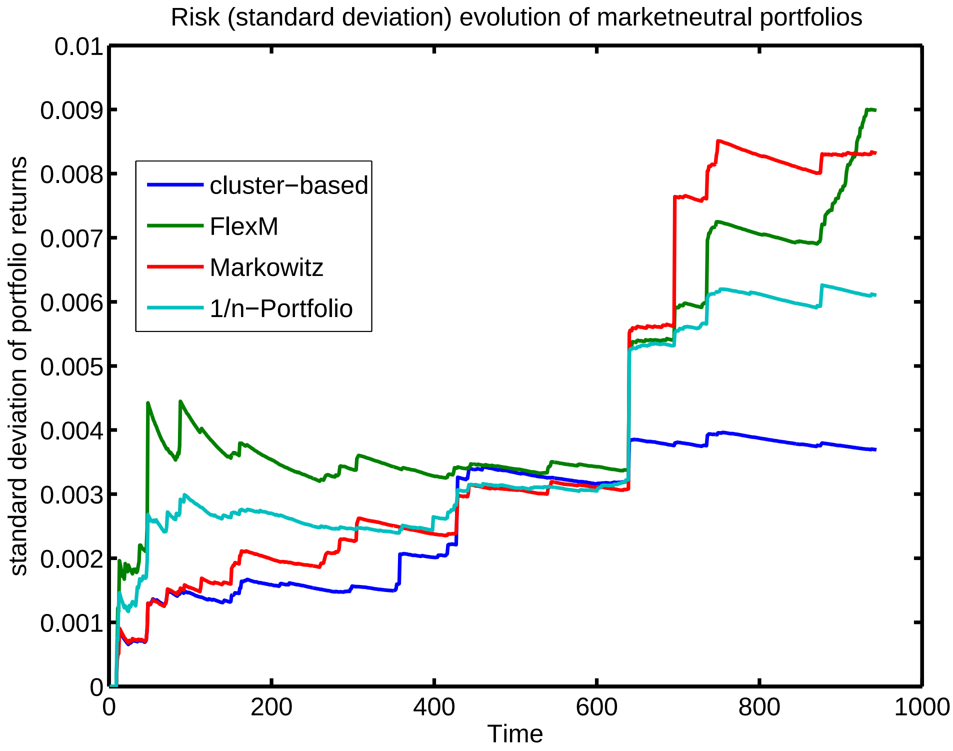

Figure 4.16. *Risk (empirical standard deviation) evolution of different market neutral portfolio-optimization algorithms.*

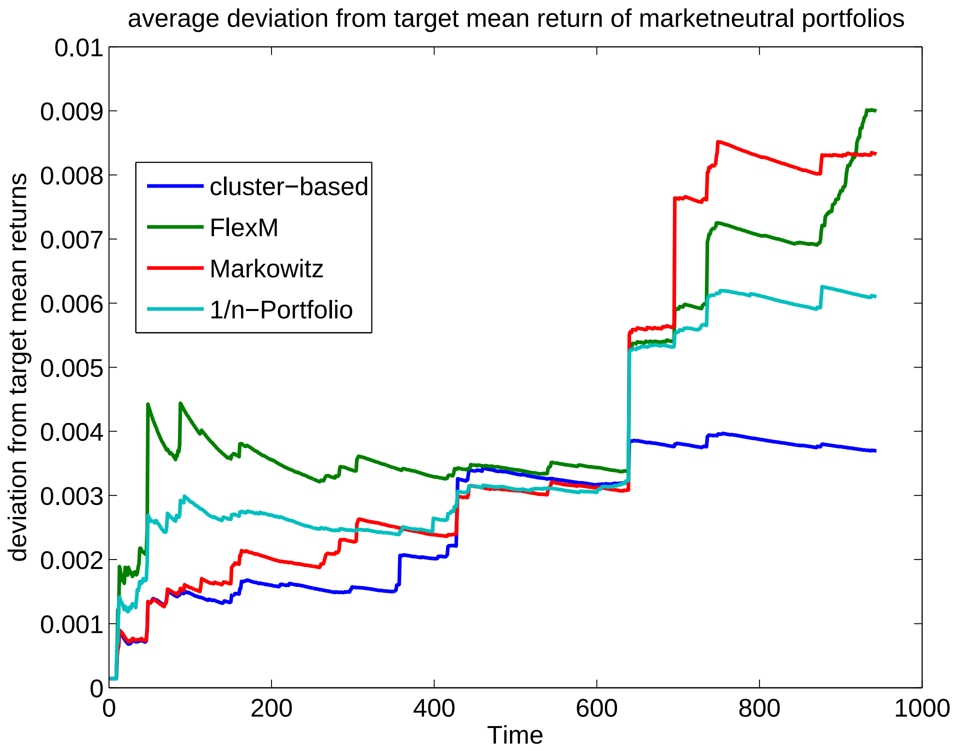

Figure 4.17. *Risk (deviation from target mean return) evolution of different market neutral portfolio-optimization algorithms.*

REFERENCES

[1] A. Akansu and R. Haddad, *Multiresolution Signal Decomposition: Transforms, Subbands, Wavelets*, Academic Press, 1992.

[2] T. Bollerslev, *Generalized autoregressive conditional heteroskedasticity*, Journal of Econometrics, 31 (1986), pp. 307–327.

[3] U. Çelikyurt and S. Özekici, *Multiperiod portfolio optimization models in stochastic markets using the mean-variance approach*, European Journal of Operational Research, 179 (2007), pp. 186–202.

[4] V. DeMiguel and F. Nogales, *Portfolio selection with robust estimation*, Forthcoming in Operations Research, (2007).

[5] D. Duffie, *Futures Markets*, Prentice Hall College Div, 1989.

[6] R. Engle, *Autoregressive conditional heteroscedasticity with estimates of variance of united kingdom inflation*, Econometrica, 50 (1982), pp. 987–1008.

[7] F. Fabozzi, P. Kolm, D. Pachamanova, and S. Focardi, *Robust Portfolio Optimization and Management*, John Wiley & Sons, 2007.

[8] G. Frahm and U. Jaekel, *Tyler's m-estimator, random matrix theory, and generalized elliptical distributions with applications to finance*, submitted to Physica A, (2008).

[9] H. Fu, M. NG, M. Nikolova, and J. Barlow, *Efficient minimization methods of mixed $l2-l1$ and $l1-l1$ norms for image restoration*, SIAM Journal of Sci. Comput., 27(6) (2006), pp. 1881–1902.

[10] L. Garlappi, R. Uppal, and T. Wang, *Portfolio selection with parameter and model uncertainty: A multi-prior approach*, Review of Financial Studies, 20(1) (2007), pp. 41–81.

[11] D. Goldfarb and G. Iyengar, *Corc technical report tr-2002-03 robust portfolio selection problems*, Mathematics of Operations Research, 28 (2002), pp. 1–38.

[12] J. Hamilton, *A new approach to the economic analysis of nonstationary time series and the business cycle*, Econometrica, 57 (1989), pp. 357–384.

[13] I. Horenko, *On simultaneous data-based dimension reduction and hidden phase identification*, J. Atm. Sci, 65 (2008), pp. 1941–1954.

[14] ———, *Finite element approach to clustering of multidimensional time series*, To appear in SIAM J. Sci. Comp., (2009).

[15] ———, *On clustering of non-stationary meteorological time series*, Dyn. of Atmos. and Oc., accepted, (2009).

[16] ———, *On robust estimation of low-frequency variability trends in discrete markovian sequences of atmospheric circulation patterns*, To appear in J. of Atmospheric Sciences, (2009).

[17] I. Horenko, R. Klein, S. Dolaptchiev, and Ch. Schütte, *Automated generation of reduced stochastic weather models i: Simultaneous dimension and model reduction for time series analysis*, Mult. Mod. Sim., 6 (2008), pp. 1125–1145.

[18] I. Horenko and Ch. Schütte, *Likelihood-based estimation of multidimensional langevin models and its application to biomolecular dynamics*, SIAM Mult. Mod. Sim., 7(2) (2008), pp. 731–773.

[19] I. Jolliffe, *Principle Component Analysis*, Springer, 2002.

[20] R. Kalman, *A new approach to linear filtering and prediction problems*, Journal of Basic Engenieering, 82(1) (1960), pp. 35–45.

[21] B. Kedem and K. Fokianos, *Regression models for time series analysis*, Wiley Series in Probability and Statistics, 2002.

[22] H. Krolzig, *Predicting Markov-switching vector autoregressive processes*, Nuffield College Oxford, 2000.

[23] O. Ledoit, P. Santa-Clara, and M. Wolf, *Flexible multivariate garch modeling with an application to international stock markets*, Review of Economics and Statistics, 3(07) (2003), pp. 735–747.

[24] O. Ledoit and M. Wolf, *Improved estimation of the covariance matrix of stock returns with an application to portfolio selection*, Journal of Empirical Finance, 10(5) (2003), pp. 603–621.

[25] ———, *Honey, i shrunk the sample covariance matrix*, Journal of Portfolio Management, 32(1) (2004).

[26] H. Markowitz, *Portfolio selection*, J. of Finance, 7 (1952), pp. 77–91.

[27] R. Michaud, *The markowitz optimization enigma: Is optimized optimal?*, Financial Analysts Journal, (1989).

[28] T. Rolski, H. Schmidli, V. Schmidt, and Jozef Teugels, *Stochastic Processes for Insurance and Finance*, Wiley Series in Probability and Statistics, 1999.

[29] J. Schäfer and K. Strimmer, *A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics*, Statistical Applications in Genetics and Molecular Biology, 4(1) (2005).

[30] W. Sharpe, *Capital asset prices: a theory of market equilibrium under conditions of risk*, J. of Finance, 9 (1964), pp. 425–442.

[31] A. Tikhonov, *On the stability of inverse problems*, Dokl. Akad. Nauk SSSR, 39(5) (1943), pp. 195–198.

[32] R.S. Tsay, *Analysis of financial time series*, Wiley-Interscience, 2005.

[33] D.E. Tyler, *A distribution-free m-estimator of multivariate scatter*, Ann. Statist., 15 (1987), pp. 223–251.

[34] R. Welsch and X. Zhou, *Application of robust statistics to asset allocation models*, REVSTAT - Statistical Journal, 5(1) (2007), pp. 97–114.