Statistics & Decisions 26, 109–143 (2008) / **DOI** 10.1524/stnd.2008.0916

Mean and covariance matrix adaptive estimation for a weakly stationary process. Application in stochastic optimization

Vincent Guigues

Received: April 4, 2007; Accepted: October 19, 2008

Summary: We introduce an adaptive algorithm to estimate the uncertain parameter of a stochastic optimization problem. The procedure estimates the one-step-ahead means, variances and covariances of a random process in a distribution-free and multidimensional framework when these means, variances and covariances are slowly varying on a given past interval. The quality of the approximate problem obtained when employing our estimation of the uncertain parameter is controlled in function of the number of components of the process and of the length of the largest past interval where the means, variances and covariances slowly vary. The procedure is finally applied to a portfolio selection model.

1 Introduction

We consider stochastic optimization problems where the underlying stochastic process $r_t \in \mathbb{R}^n$ ($n \geq 2$) is generated by the model:

$$r_t = \rho_t + \zeta_t, \quad \text{with } \mathbb{E}r_t = \rho_t \text{ and } \mathbb{E}\zeta_t\zeta_t^\top = Q_t, \quad t = 1, \ldots, N, \tag{1.1}$$

where ζ_t are independent random vectors in $\mathbb{R}^n$ with zero mean and N is the number of available observations. The constants ρ_t and Q_t respectively represent the mean and covariance matrix at time step t. If svec(Q) is the symmetric vectorization of the symmetric matrix Q, we focus on stochastic optimization problems that can be expressed as:

$$\mathcal{P}(\theta) \quad = \quad \begin{cases} \min \ f_0(x, \theta), \\ x \in X \subset \mathbb{R}^p; \end{cases} \tag{1.2}$$

where the unknown parameter $\theta = (\rho_{N+1}^\top, \text{svec}(Q_{N+1})^\top)^\top$ belonging to a given set $\Theta \subset \mathbb{R}^M$ is made of the one-step-ahead mean ρ_{N+1} and the components of the one-step-ahead covariance matrix Q_{N+1}. The parameter dimension is thus $M = n + \frac{n(n+1)}{2}$. In

AMS 2000 subject classification: Primary: 62G05, 62M10; Secondary: 90C15
Key words and phrases: Adaptive estimation, weakly stationary process, stochastic optimization, Value-at-Risk, portfolio management

problem (1.2), the set X is bounded and closed and the objective function f_0 belongs to a class of functions introduced in the next section. The value of the parameter θ is not known and problem $\mathcal{P}(\theta)$ is the optimization problem providing the optimal decision x^*.

There are different approaches to deal with uncertainty in stochastic optimization problem (1.2). A possible approach is the worst case robust optimization methodology (see [BTN99] for instance). We propose instead to provide an estimate $\hat{\boldsymbol{\theta}}$ (from now on realizations of random vectors are in bold), for which the problem $\mathcal{P}(\hat{\boldsymbol{\theta}})$ approximates reasonably well problem $\mathcal{P}(\theta)$, with controlled accuracy. The estimation of the parameter θ is made on the basis of the N independent observations $\mathbf{r_t}$, $t = 1, \ldots, N$. We consider two special cases: the stationary case, where the functions ρ_t and Q_t are not time varying and the case in which these functions are slowly varying (in the sense of [MS04a]) on a given past interval. The problem of measuring the quality of approximate stochastic optimization problems appears for instance in [Sha89, Sha93, Sha94, Pfl03]. The originality of our approach is both the non-asymptotic study and the statistical framework (weakly stationary process).

This paper is organized as follows. In Section 2, we define the accuracy of the approximate problem and bound this accuracy from above using $\|\hat{\boldsymbol{\theta}} - \theta\|_\infty$. The estimation of θ and the quality of this estimation are discussed in details in Section 3. This work can be seen as a generalization of [MS04a], where an adaptive estimation of a slowly varying volatility is made within a one-dimensional and parametric framework. Here, we consider instead the multidimensional and distribution-free setting as well as the estimation of slowly varying means, variances and covariances.

In Section 4, we then give the accuracy of the approximate problem $\mathcal{P}(\hat{\boldsymbol{\theta}})$. In this section, we also specialize our results for a subclass of stochastic optimization problems of the form:

$$\tilde{\mathcal{P}}(\theta) \;=\; \begin{cases} \min\ \kappa\sqrt{x^\top Q_{N+1} x} - \rho_{N+1}^\top x, \\ x \in X \subset \mathbb{R}^p, \end{cases} \tag{1.3}$$

where κ is a fixed positive parameter. When the income linked with decision x is a linear random function of x given by $r_{N+1}^\top x$, the problem $\tilde{\mathcal{P}}(\theta)$ amounts to minimizing a trade-off between the mean cost $\mathbb{E}[-r_{N+1}^\top x] = -\rho_{N+1}^\top x$ and its standard deviation $\sigma(r_{N+1}^\top x) = \sqrt{x^\top Q_{N+1} x}$. The methodology introduced in this paper is assessed on a portfolio selection model (of the form $\tilde{\mathcal{P}}(\theta)$) using simulated and real data in Section 5. Proofs are given in the appendix.

2 How to control the accuracy of the approximate problem

2.1 Definition of the accuracy

Our objective is to construct for $\mathcal{P}(\theta)$ a data-driven approximate problem $\mathcal{P}(\hat{\boldsymbol{\theta}})$ that uses some specific estimation $\hat{\boldsymbol{\theta}}$ of θ, i.e. an estimation $\hat{\boldsymbol{\rho}}_{\boldsymbol{N+1}}$ of the mean and an estimation

$\hat{\boldsymbol{Q}}_{\boldsymbol{N+1}}$ of the covariance matrix, to solve $\mathcal{P}(\theta)$ with given accuracy. The corresponding approximate problem thus reads:

$$\mathcal{P}(\hat{\boldsymbol{\theta}}) = \begin{cases} \min\ f_0(x, \hat{\boldsymbol{\theta}}), \\ x \in X \subset \mathbb{R}^p. \end{cases}$$

We now define the notion of accuracy of the approximate problem. Let $\widehat{\boldsymbol{x}}$ be any optimal solution of the approximate problem $\mathcal{P}(\hat{\boldsymbol{\theta}})$ and let x^* be an optimal solution of $\mathcal{P}(\theta)$. The accuracy of the approximate problem $\mathcal{P}(\hat{\boldsymbol{\theta}})$ is given by:

$$\epsilon(\mathcal{P}(\hat{\boldsymbol{\theta}})) \equiv f_0(\widehat{\boldsymbol{x}}, \theta) - f_0(x^*, \theta).$$

2.2 Control of the accuracy

The control of the accuracy of $\mathcal{P}(\hat{\boldsymbol{\theta}})$ and $\tilde{\mathcal{P}}(\hat{\boldsymbol{\theta}})$ is based on Assumption 2.3 and Propositions 2.1 and 2.2 below.

Proposition 2.1 *The objective function f_0 of problem $\tilde{\mathcal{P}}(\theta)$ satisfies:*

$$|f_0(x, \theta) - f_0(x, \theta')| \leq \left(\|\theta - \theta'\|_\infty + \kappa \|\theta - \theta'\|_\infty^{1/2}\right) \|x\|_1. \tag{2.1}$$

Proposition 2.2 *The accuracy $\epsilon(\mathcal{P}(\hat{\boldsymbol{\theta}}))$ of $\mathcal{P}(\hat{\boldsymbol{\theta}})$ is bounded above as follows:*

$$\epsilon(\mathcal{P}(\hat{\boldsymbol{\theta}})) \leq 2 \sup_{x \in X} |f_0(x, \hat{\boldsymbol{\theta}}) - f_0(x, \theta)|. \tag{2.2}$$

We then make the following hypothesis on the objective function f_0 of problem $\mathcal{P}(\theta)$.

Assumption 2.3 *For every $x \in X$, and every $(\theta, \theta') \in \mathbb{R}^M \times \mathbb{R}^M$:*

$$|f_0(x, \theta) - f_0(x, \theta')| \leq C_0 \|\theta - \theta'\|_\infty^{\alpha_0} \|x\|_1^{p_0},$$

where $0 < \alpha_0 \leq 2$, $0 < p_0 \leq 2$ and $0 < C_0 < \infty$.

On the basis of Propositions 2.1 and 2.2 and under Assumption 2.3 for $\mathcal{P}(\theta)$, we thus see that to control the accuracy of $\mathcal{P}(\hat{\boldsymbol{\theta}})$ and $\tilde{\mathcal{P}}(\hat{\boldsymbol{\theta}})$, we need to define a statistical estimation $\hat{\boldsymbol{\theta}}$ of θ such that $\|\hat{\boldsymbol{\theta}} - \theta\|_\infty$ is controlled.

3 Adaptive estimation of the parameters

3.1 Parameter estimation problem

In this section, we address the problem of estimation of the parameter θ for $\mathcal{P}(\theta)$. We suppose the process r_t follows model (1.1) and satisfies the following assumption.

Assumption 3.1 *For some $\sigma > 0$, $\mathbb{E}\|r_t\|_\infty^4 \leq \sigma^4$, $t = 1, \ldots, N$.*

Our contribution is to determine estimations of the parameters ρ_{N+1} and Q_{N+1} which allow us to solve problem $\mathcal{P}(\theta)$ with a good accuracy. As mentioned in Section 2, we intend to define estimations $\hat{\boldsymbol{\rho}}_{\boldsymbol{N+1}}$ (associated to the estimator $\hat{\rho}_{N+1}$) and $\hat{\boldsymbol{Q}}_{\boldsymbol{N+1}}$ (associated to the estimator $\hat{Q}_{N+1}$), of respectively ρ_{N+1} and Q_{N+1} such that $\|\hat{\rho}_{N+1} - \rho_{N+1}\|_\infty$ and $\|\hat{Q}_{N+1} - Q_{N+1}\|_\infty$ are small with high probability. In the particular case where the mean $\rho_t = \rho$ and the covariance matrix $Q_t = Q$ are not time varying, the estimations $\hat{\boldsymbol{\rho}}_{\boldsymbol{N+1}}$ and $\hat{\boldsymbol{Q}}_{\boldsymbol{N+1}}$ use as many past available data as possible. Let us fix a positive parameter λ, and let $K_0(N)$ and $[\cdot]_K$, for $K > 0$, be the constant and the truncation operator defined by

$$K_0(N) = \sigma\left(\frac{N}{\ln n(n+1) + \lambda \ln N}\right)^{\frac{1}{4}};\ [x]_K = \begin{cases} K & \text{if } x > K, \\ -K & \text{if } x < -K, \\ x & \text{otherwise.} \end{cases} \tag{3.1}$$

Given the N observations $\mathbf{r_t}$, $t = 1, \ldots, N$, we choose for $\hat{\boldsymbol{\rho}}_{\boldsymbol{N+1}}$ and $\hat{\boldsymbol{Q}}_{\boldsymbol{N+1}}$ the empirical mean and covariance matrix of a process α_t derived from the process r_t:

$$\hat{\boldsymbol{\rho}}_{\boldsymbol{N+1}} = \frac{1}{N}\sum_{t=1}^{N} \boldsymbol{\alpha_t}, \quad \text{and} \quad \hat{\boldsymbol{Q}}_{\boldsymbol{N+1}} = \frac{1}{N}\sum_{t=1}^{N} (\boldsymbol{\alpha_t} - \hat{\boldsymbol{\rho}}_{\boldsymbol{N+1}})(\boldsymbol{\alpha_t} - \hat{\boldsymbol{\rho}}_{\boldsymbol{N+1}})^\top, \tag{3.2}$$

where for $i = 1, \ldots, n$, $t = 1, \ldots, N$, $\boldsymbol{\alpha_t}(i) = [\boldsymbol{r_t}(i)]_{K_0(N)}$.

Notice that for technical reasons, we do not use the empirical estimations directly (the accuracy of the approximate problem is more tightly controlled with our estimations and under less restrictive hypotheses). However, the results of this paper can be extended to the case where the empirical estimations of the mean and of the covariance matrix are used (see [Gui05]). Estimations (3.2) are all the closer to the empirical estimations as the number N of data used to compute them grows.

In the more general case where the parameters ρ_t and Q_t slowly vary on a given past interval (this notion of slowly varying functions is defined more precisely in the next subsection), there is a need for an adaptive procedure. In this case, using the terminology of [MS04a], we call *interval of local time homogeneity* (ILTH) an interval where ρ_t and Q_t slowly vary. The adaptive procedure determines an estimation $\hat{\boldsymbol{I}}$ of the best interval for parameter estimation i.e. of the largest ILTH. This question is addressed in Section 3.3. Once the interval $\hat{\boldsymbol{I}}$ is found, the estimations $\hat{\boldsymbol{\rho}}_{\boldsymbol{N+1}}$ and $\hat{\boldsymbol{Q}}_{\boldsymbol{N+1}}$ of ρ_{N+1} and Q_{N+1} are given by $\hat{\boldsymbol{\rho}}_{\hat{\boldsymbol{I}}}$ and $\hat{\boldsymbol{Q}}_{\hat{\boldsymbol{I}}}$ where for any nonempty interval I we define $\hat{\boldsymbol{\rho}}_{\boldsymbol{I}}$ and $\hat{\boldsymbol{Q}}_{\boldsymbol{I}}$ by:

$$\hat{\boldsymbol{\rho}}_{\boldsymbol{I}} = \frac{1}{|I|}\sum_{t\in I} \boldsymbol{\alpha}_{\boldsymbol{t}}^{\boldsymbol{I}}, \quad \text{and} \quad \hat{\boldsymbol{Q}}_{\boldsymbol{I}} = \frac{1}{|I|}\sum_{t\in I} (\boldsymbol{\alpha}_{\boldsymbol{t}}^{\boldsymbol{I}} - \hat{\boldsymbol{\rho}}_{\boldsymbol{I}})(\boldsymbol{\alpha}_{\boldsymbol{t}}^{\boldsymbol{I}} - \hat{\boldsymbol{\rho}}_{\boldsymbol{I}})^\top, \tag{3.3}$$

where for $i = 1, \ldots, n$, and $t \in I$, $\boldsymbol{\alpha}_{\boldsymbol{t}}^{\boldsymbol{I}}(i) = [\mathbf{r_t}(i)]_{K_0(|I|)}$ with $K_0(\cdot)$ defined in (3.1) and λ a positive parameter (of the adaptive algorithm).

3.2 Hypotheses on the means, variances and covariances

Under local time homogeneity, we suppose that there exists a past interval of right endpoint $N+1$ such that the means, variances and covariances slowly vary or are almost constant in this interval. The adaptive procedure we describe in the next subsection aims at finding

the largest interval satisfying this assumption. For a given past interval of right endpoint $N+1$, we should thus be able to decide, from a theoretical point of view, whether we can consider that the means, the variances and covariances slowly vary on this interval or not. If ρ_t and Q_t slowly vary on the interval $I = [N+1-m, N+1]$, where $m \in \mathbb{N}^*$, then the quantities:

$$\Delta_I^\rho = \sqrt{\frac{1}{|I|}\sum_{t\in I} \|\rho_t - \rho_{N+1}\|_\infty^2}, \quad \text{and} \quad \Delta_I^Q = \sqrt{\frac{1}{|I|}\sum_{t\in I} \|Q_t - Q_{N+1}\|_\infty^2},$$

should be small. Similarly, we expect Δ_J^ρ and Δ_J^Q to be small for all subintervals J of the interval I. In particular, if $\mathcal{I}(I)$ is a finite set of testing subintervals of the interval I (the choice of $\mathcal{I}(I)$ is discussed later) then Δ_J^ρ and Δ_J^Q should be small for every interval $J \in \mathcal{I}(I)$. To take into account the variance of the estimators, we then define for every interval I:

$$V_I^\rho = \mathbb{E}\|\hat{\rho}_I - \mathbb{E}\hat{\rho}_I\|_\infty, \quad V_I^Q = \mathbb{E}\|\hat{Q}_I - \mathbb{E}\hat{Q}_I\|_\infty.$$

Let us fix a small and non-negative constant D. If we set $\mathcal{I}_+(I) = \mathcal{I}(I) \cup I$, we say that ρ_t and Q_t are slowly varying on the interval I if:

$$\Delta_J^\rho \le DV_J^\rho, \quad \Delta_J^Q \le DV_J^Q, \text{ for } J \in \mathcal{I}_+(I). \tag{3.4}$$

Let now $\mathcal{I}$ be a family of candidate intervals. We suppose that (3.4) holds on the smallest candidate interval I. Notice that from a practical point of view, if no interval I satisfies the above relations (3.4), then the adaptive algorithm returns a minimal interval. The tests of homogeneity are thus in fact only made for intervals of length greater than the length of this minimal interval.

The ideal interval $\mathbb{I}$ of local time homogeneity that the oracle we build in Section 3.3 aims at approximating is then the largest interval (among the family of candidate intervals) such that (3.4) holds:

$$\mathbb{I} = \operatorname{argmax}\{|I| \mid I \in \mathcal{I}, \ \Delta_J^\rho \le DV_J^\rho, \ \Delta_J^Q \le DV_J^Q, \text{ for } J \in \mathcal{I}_+(I)\}, \tag{3.5}$$

where D is a fixed and non-negative constant. Our definition of ideal interval of local time homogeneity differs from that of [MS04b]. In [MS04b], condition (3.4) has to be satisfied only for $J = I$. However, suppose that a sufficiently large interval I is such that $Q_t = Q$ for all t in I and ρ_t is varying a lot only on the left of the interval. Using the definition of [MS04b] of an interval of local time homogeneity, we would probably accept I as an interval of local time homogeneity whereas our definition would probably conclude the contrary, which makes more sense in this case.

3.3 Adaptive method

We suppose that the mean ρ_t of the process r_t and the covariance matrix Q_t are slowly varying on an ILTH to determine. Under this hypothesis, there are intervals $\mathbb{I}_\rho = [N+1-m_\rho, N+1]$ and $\mathbb{I}_Q = [N+1-m_Q, N+1]$, such that the mean ρ_t does not vary much on $\mathbb{I}_\rho$ and Q_t does not vary much on $\mathbb{I}_Q$. From a theoretical point of view, there can be a change point in the mean; the variances and covariances being constant. However, in this case, we should take the change in the mean into account in the estimation of the

covariance matrix and the estimation of $\mathbb{I}_Q$ would be more difficult. We will thus look for the largest interval $\hat{\boldsymbol{I}}$ (an estimation of $\mathbb{I}$) such that the means, variances and covariances slowly vary on $\hat{\boldsymbol{I}}$ which amounts to finding an estimation of the intersection of $\mathbb{I}_\rho$ and $\mathbb{I}_Q$. An alternative method consists in first determining an estimation $\hat{\boldsymbol{I}}_{\boldsymbol{\rho}}$ of $\mathbb{I}_\rho$ and to further determine an estimation $\hat{\boldsymbol{I}}_{\boldsymbol{Q}}$ of the largest interval contained in $\mathbb{I}_\rho$ such that Q_t is slowly varying on this interval. We give the theoretical accuracy of the estimators obtained using one interval of homogeneity. The proofs can be directly adapted to show the accuracy of the resulting estimators $\hat{\rho}_{\hat{I}_\rho}$ and $\hat{Q}_{\hat{I}_Q}$ when two intervals of homogeneity $\hat{\boldsymbol{I}}_{\boldsymbol{\rho}}$ and $\hat{\boldsymbol{I}}_{\boldsymbol{Q}}$ are determined. To illustrate, we give the accuracy of the estimation of the mean by $\hat{\boldsymbol{\rho}}_{\hat{\boldsymbol{I}}_{\boldsymbol{\rho}}}$ when separate intervals of homogeneity are determined to estimate, on the one hand, the mean and on the other hand, the variances and covariances. We use an adaptive algorithm which is described in the next subsection to determine an estimation $\hat{\boldsymbol{I}}$ of I. The key question will be to decide, from a practical point of view, whether on a given interval, ρ_t and Q_t are slowly varying or not.

3.3.1 Algorithm description

The choice of the adaptive interval of time homogeneity is done as follows. Let $\mathcal{I}$ be a family of ordered candidate intervals and for every $I \in \mathcal{I}$, let $\mathcal{I}(I)$ be a family of testing subintervals. Notice that we suppose that $N+1$ belongs to the ILTH. This justifies the estimation of ρ_{N+1} and Q_{N+1} by empirical estimations of the mean and of the covariance matrix using the data of the ILTH. However, as we do not have observations for time step $N+1$, the candidate intervals are of the form $[N+1-m, N]$, where $m \in \mathbb{N}^*$ (the return $\boldsymbol{r_{N+1}}$ is in the ILTH but is not used for estimation as it is not available). We suppose we have a rule which allows us to know if we can consider that on a given candidate interval I, the means, variances and covariances slowly vary on I or not (this question is addressed next). The selected interval of time homogeneity $\hat{\boldsymbol{I}}$ is such that for all $I \in \mathcal{I}$ satisfying $I \subseteq \hat{\boldsymbol{I}}$, I is accepted as ILTH and the smallest interval $I \in \mathcal{I}$ such that $\hat{\boldsymbol{I}} \subsetneq I$ is rejected.

Now we should be able to decide, from a practical point of view, if on a given interval I, the mean ρ_t and the covariance matrix Q_t are slowly varying or not. If on $I = [N+1-m, N]$, ρ_t does not "vary much", then the mean value of ρ_t on I is close to the mean value of ρ_t on every subinterval $J \in \mathcal{I}(I)$. But the mean value of ρ_t on I is close to the mean value of $\boldsymbol{r_t}$ on I which is close to the mean value of $\boldsymbol{\alpha_t^I}$ on I for $|I|$ large enough. Thus, if ρ_t does not "vary much" on I, then for every subinterval $J \in \mathcal{I}(I)$, $\hat{\boldsymbol{\rho}}_{\boldsymbol{J}}$ is close to $\hat{\boldsymbol{\rho}}_{\boldsymbol{I}}$ where $\hat{\boldsymbol{\rho}}_{\boldsymbol{I}}$ is defined in (3.3). Similarly, if Q_t is nearly constant on I, then for every subinterval $J \in \mathcal{I}(I)$, $\hat{\boldsymbol{Q}}_{\boldsymbol{J}}$ is close to $\hat{\boldsymbol{Q}}_{\boldsymbol{I}}$. In fact, deciding whether a given interval I is of time homogeneity or not boils down to doing a test. For instance, in the particular case of piecewise constant functions ρ_t and Q_t, we have to do the test:

$$H_0^I : \forall (t,t') \in I^2,\ \rho_t = \rho_{t'}, Q_t = Q_{t'} \quad H_1^I :\ \exists (t,t') \in I^2 \ \mid\ \rho_t \neq \rho_{t'} \ \text{ or } \ Q_t \neq Q_{t'}.$$

We then need the following theorem which quantifies the proximity between two estimations $\hat{\boldsymbol{\rho}}_{\boldsymbol{I}}$ and $\hat{\boldsymbol{\rho}}_{\boldsymbol{J}}$ of the mean or $\hat{\boldsymbol{Q}}_{\boldsymbol{I}}$ and $\hat{\boldsymbol{Q}}_{\boldsymbol{J}}$ of the covariance matrix done on embedded intervals $I \in \mathcal{I}$ and $J \in \mathcal{I}(I)$.

Theorem 3.2 *Let r_t satisfy Assumptions 2.3 and 3.1, let I be a nonempty interval of local time homogeneity (which means that* (3.4) *holds) and let $J \in \mathcal{I}(I)$ with $|J| > 0$. Let $\hat{\rho}_I$ and $\hat{Q}_I$ be the estimators associated to the estimations in* (3.3)*. Then for every $\lambda > 0$ such that $\ln n(n+1) + \lambda \ln |I| \leq |I|$ and $\ln n(n+1) + \lambda \ln |J| \leq |J|$, we have:*

$$\mathbb{P}\left(\|\hat{\rho}_I - \hat{\rho}_J\|_\infty \geq \gamma_\rho(|I|, |J|, \lambda)\right) \leq \frac{1}{|I|^\lambda} + \frac{1}{|J|^\lambda}, \tag{3.6}$$

$$\mathbb{P}\left(\|\hat{Q}_I - \hat{Q}_J\|_\infty \geq \gamma_Q(|I|, |J|, \lambda)\right) \leq 2\left(\frac{1}{|I|^\lambda} + \frac{1}{|J|^\lambda}\right), \tag{3.7}$$

with

$$\gamma_\rho(|I|, |J|, \lambda) = 4\sqrt{\tfrac{2}{\ln 2}} D\sigma\left(\sqrt{\tfrac{\ln n}{|I|}} + \sqrt{\tfrac{\ln n}{|J|}}\right) + \left(\tfrac{7}{3} + \sqrt{2}\right)\sigma\left(f(|I|, \lambda) + f(|J|, \lambda)\right),$$

$$\gamma_Q(|I|, |J|, \lambda) = (k_Q D + k'_Q)\sigma^2(f(|I|, \lambda) + f(|J|, \lambda)),$$

where $f(|I|, \lambda) = \sqrt{\frac{\ln n(n+1) + \lambda \ln |I|}{|I|}}$, and k_Q and k'_Q are constants given in the appendix.

Remark 3.3 Conditions $\ln n(n+1) + \lambda \ln |I| \leq |I|$ and $\ln n(n+1) + \lambda \ln |J| \leq |J|$ in Theorem 3.2 above can be suppressed. In this case, we get more complicated expressions of γ_ρ and γ_Q. This more general case is studied in the appendix.

In virtue of Theorem 3.2, we will accept I as an interval of homogeneity if for every subinterval $J \in \mathcal{I}(I)$:

$$\|\hat{\boldsymbol{\rho}}_{\boldsymbol{I}} - \hat{\boldsymbol{\rho}}_{\boldsymbol{J}}\|_\infty \leq \gamma_\rho(|I|, |J|, \lambda) \text{ and } \|\hat{\boldsymbol{Q}}_{\boldsymbol{I}} - \hat{\boldsymbol{Q}}_{\boldsymbol{J}}\|_\infty \leq \gamma_Q(|I|, |J|, \lambda), \tag{3.8}$$

where λ is the positive parameter involved in the definition of K_0, and reject I otherwise. Thus, if I is indeed an interval of local time homogeneity then the probability that (3.8) holds (where we replaced in (3.8) the estimations by the estimators) can be controlled with an appropriate choice of λ. Notice that when the length of I and J grows, then the right hand sides of (3.6) and (3.7) rapidly go to 0, as expected. We now discuss the choice of the sets $\mathcal{I}$, $\mathcal{I}(I)$ and of the parameter λ.

3.3.2 Choice of the sets $\mathcal{I}$, $\mathcal{I}(I)$ and of the parameter λ

The simplest way to choose the sets $\mathcal{I}$ and $\mathcal{I}(I)$ is described in [MS04a] and [MS04b]. We briefly recall this choice. Let m_0 be the length of the smallest candidate interval (with the hypotheses of Subsection 3.2, the last m_0 time steps thus automatically belong to the ILTH). We choose a grid $\mathcal{G} = \{t_k = m_0 k, \quad k \in \mathbb{N}^*\}$ where $m_0 \in \mathbb{N}^*$ is the grid step. Let $N + 1 = k^* m_0 > 2m_0$, be a point where we want to carry out the adaptive algorithm to determine estimations of ρ_{N+1} and Q_{N+1}. The set $\mathcal{I}$ is defined as follows:

$$\mathcal{I} = \{[t_k, t_{k^*}[, \quad 1 \leq k < k^*\}.$$

In [MS04b], for every $I_k = [t_k, t_{k^*}[\in \mathcal{I}$, the set $\mathcal{I}(I_k)$ of testing subintervals of interval I_k is then the set of all smaller intervals whose endpoints belong to $\mathcal{G}$ and with either the same left endpoint or the same right endpoint as I_k:

$$\mathcal{I}(I_k) = \{[t_k, t_{k'}[\cup [t_{k'}, t_{k^*}[, \quad \text{with} \quad k < k' < k^*\}.$$

In [MS04a], only the subintervals of interval I_k with the same right endpoint are considered. When $N+1$ is not a point from the grid, we can use a dynamic grid adapted to the time $N+1$ of estimation and the intervals of the set $\mathcal{I}$ are of the kind $[N+1-m_0k,\, N+1[$ where $k \in \mathbb{N}^*$. If we decide to determine two intervals of homogeneity $\hat{\boldsymbol{I}}_{\boldsymbol{\rho}}$ and $\hat{\boldsymbol{I}}_{\boldsymbol{Q}}$, to estimate respectively the mean and the covariance matrix, the length $m_0(\rho)$ of the smallest testing subinterval chosen to find $\hat{\boldsymbol{I}}_{\boldsymbol{\rho}}$ and the length $m_0(Q)$ of the smallest testing subinterval chosen to find $\hat{\boldsymbol{I}}_{\boldsymbol{Q}}$, are not necessarily the same. These values depend on the variance of the components on $\mathbb{I}_\rho$ and $\mathbb{I}_Q$. We need less data to obtain a good estimation of the mean. However, to determine an estimation of $\mathbb{I}_Q$, we should take greater values for $m_0(Q)$, say at least n where n is the number of components (if the number of data is less than the number of components, the empirical covariance matrix is not invertible). This means that we will have a very rough estimate of $\mathbb{I}_Q$ if the step $m_0(Q)$ is large. The same will hold for $\mathbb{I}_\rho$ if the step $m_0(\rho)$ is large. We thus now intend to introduce a few modifications of the implementation choices proposed in [MS04b] (that we have just mentioned) to increase the algorithm performance or speed. In what follows, we drop the dependence of m_0 on ρ and Q (all further remarks on m_0 will be valid for $m_0(\rho)$ and $m_0(Q)$).

- First, we can introduce more flexibility in the choice of the set $\mathcal{I}$. Indeed, the length difference of two successive intervals in $\mathcal{I}$ is not necessarily m_0, where m_0 is the length of the smallest testing subinterval. Let the smallest interval from $\mathcal{I}$ be made of the last m_0 observations. Now, if $I = [t_\ell,\, N+1[$ is an interval from $\mathcal{I}$, the successor of I in $\mathcal{I}$ can be obtained adding k data instead of m_0 thus yielding the interval $[t_\ell - k,\, N+1[$ where $k \in \mathbb{N}^*$ and $k < m_0$. Even values of k as small as one (the successor of an interval from $\mathcal{I}$ has the length of its father plus one) can yield spectacular improvements (see [Gui05]). This simple modification of the set $\mathcal{I}$ allows us to find the optimal intervals of homogeneity with a higher probability.

- We can also modify the testing subintervals without changing much the performance of the algorithm while improving its rapidity. At a given iteration of the algorithm, if we have accepted an interval $I = [t_\ell,\, N+1[$ as an interval of homogeneity, the data from the next interval $I' = [t_\ell - k,\, N+1[$ the most prone to be before the break point (if there is one) in I' is the k left data. That's why a simple modification of the testing subintervals consists in only taking one subinterval $J = [t_\ell - k,\, t_\ell - k + m_0[$ of interval I' (in this case $\mathcal{I}(I')$ is reduced to J). If an interval is of homogeneity then this procedure will accept it with a higher probability than the adaptive algorithm as it is presented in [MS04b]. Indeed, the test done to know if this interval is accepted is one of the tests done by the adaptive algorithm used in [MS04b]. Now if there is a break point in I', as the previous interval I was accepted, the break point certainly lies on the left of the interval. Thus, the m_0 data lying on the left of interval I' is the m_0 data most prone to provide estimations far from those obtained using the data of the whole interval I' (this implementation choice is called choice (a)). Also, instead of checking the difference between the estimations on $I' = [t_\ell - k,\, N+1[$ and $J = [t_\ell - k,\, t_\ell - k + m_0[$, we can check the difference between the estimations on J and $J' = [N+1-m_0,\, N+1[$, which is thus a fixed interval (implementation choice (b)). In this manner, at least the mean and covariance matrix are slowly

varying on the whole of one of the two intervals (the interval J'). Having the same objective in mind, we can also check the difference between the estimators on J and $J' = [N+1-m_0, N+1[$ if the length of I' is less than or equal to $2m_0$ and on J and $J' = [t_\ell - k + m_0, N+1[$ otherwise (implementation choice (c)). For choices (b) and (c), $\mathcal{I}(I')$ is reduced to $J \cup J'$. From a practical point of view and using the above notation, I' is accepted as an ILTH for the mean using implementation choice (a) (resp. (b) or (c)) if $\|\hat{\rho}_{I'} - \hat{\rho}_J\|_\infty \leq \gamma_\rho(|I'|, |J|, \lambda)$ (resp. $\|\hat{\rho}_J - \hat{\rho}_{J'}\|_\infty \leq \gamma_\rho(|J|, |J'|, \lambda)$). Thus, at each iteration of the algorithm, we only have to perform one test, in which we are confident if m_0 is not too small.

We can show (see [Gui05]), both from a theoretical and practical point of view, the close behavior of these variants of the adaptive algorithm. Finally, we can calibrate λ using different techniques. We can for instance determine λ such that the type I error is controlled in a change point model. Parameter λ can also be chosen using Monte Carlo simulations.

3.3.3 Quality of the estimation

The key question we address now is to determine the quality of the approximation of ρ_{N+1} and Q_{N+1} by our adaptive estimators $\hat{\rho}_{\hat{I}}$ and $\hat{Q}_{\hat{I}}$. Recall that the ideal interval $\mathbb{I}$ is the interval which checks:

$$\mathbb{I} = \operatorname{argmax}\{|I| \ \mid \ I \in \mathcal{I}, \ \ \Delta_J^\rho \leq DV_J^\rho, \ \ \Delta_J^Q \leq DV_J^Q, \ \text{for } J \in \mathcal{I}_+(I)\}, \tag{3.9}$$

where D is a fixed and small constant.

We first give the quality of the estimation $\hat{\boldsymbol{\theta}}_{\mathbb{I}} = (\hat{\boldsymbol{\rho}}_{\mathbb{I}}{}^\top, \operatorname{svec}(\hat{\boldsymbol{Q}}_{\mathbb{I}})^\top)^\top$ that would be used if the ideal interval of local time homogeneity was known.

Theorem 3.4 *If $\lambda > 0$ is such that $\ln n(n+1) + \lambda \ln|\mathbb{I}| \leq |\mathbb{I}|$, then there is a constant $k(D)$ (given in the appendix) depending affinely on D, such that*

$$\mathbb{P}\left(\|\hat{\theta}_{\mathbb{I}} - \theta\|_\infty \geq k(D)\max(\sigma, \sigma^2)\sqrt{\frac{\ln n(n+1) + \lambda \ln|\mathbb{I}|}{|\mathbb{I}|}}\right) \leq \frac{3}{|\mathbb{I}|^\lambda}. \tag{3.10}$$

The following theorem gives the accuracy of the adaptive estimates.

Theorem 3.5 *Let $\hat{\boldsymbol{I}}$ be the interval selected by the adaptive algorithm and λ be the parameter involved in the definition of K_0. We suppose that $\ln n(n+1) + \lambda \ln m_0 \leq m_0$, where m_0 is the length of the smallest testing subinterval. Then there is a constant $k(D)$ (given in the appendix) depending affinely on D, such that if $\hat{\theta}_{\hat{I}} = (\hat{\rho}_{\hat{I}}^\top, \operatorname{svec}(\hat{Q}_{\hat{I}})^\top)^\top$, we get:*

$$\begin{aligned}\mathbb{P}\left(\|\hat{\theta}_{\hat{I}} - \theta\|_\infty \geq k(D)\max(\sigma, \sigma^2)\sqrt{\frac{\ln n(n+1) + \lambda \ln|\mathbb{I}|}{|\mathbb{I}|}}\right) \\ \leq \sum_{I \in \mathcal{I}} \sum_{|I \subseteq \mathbb{I}} \sum_{J \in \mathcal{I}_+(I)} \frac{3}{|J|^\lambda}.\end{aligned} \tag{3.11}$$

Notice that the estimation is all the closer to the estimated parameter as D and σ are small and as the length of $\mathbb{I}$ is large. Also, the number of terms on the right hand side of the inequality above in Theorem 3.5 depends on the choice of the candidate intervals and on the choice of the subintervals. This theorem shows in fact that the quality of the adaptive estimators is close to the quality of the estimators $\hat{\rho}_{\mathbb{I}}$ and $\hat{Q}_{\mathbb{I}}$ that would be used if the ideal interval $\mathbb{I}$ was known in advance. The adaptive algorithm can be viewed as an oracle which, receiving as input a collection of observations $\mathbf{r_t}$, $t = 1, \ldots, N$, of a process satisfying Assumption 3.1, gives estimations of ρ_{N+1} and Q_{N+1} that are close to the true (unknown) values. In order to determine two intervals of homogeneity $\hat{\boldsymbol{I}}_{\boldsymbol{\rho}}$ and $\hat{\boldsymbol{I}}_{\boldsymbol{Q}}$ for the mean and the covariance matrix respectively, we can adapt the definition of the optimal intervals of homogeneity. The optimal interval of time homogeneity $\mathbb{I}_\rho$ for ρ checks:

$$\mathbb{I}_\rho = \text{argmax}\,\{|I| \;\mid\; I \in \mathcal{I},\;\; \Delta_J^\rho \leq DV_J^\rho,\;\text{for}\;\; J \in \mathcal{I}_+(I)\}; \tag{3.12}$$

and the optimal interval of time homogeneity $\mathbb{I}_Q$ for Q:

$$\mathbb{I}_Q = \text{argmax}\,\{|I| \;\mid\; I \subseteq \mathbb{I}_\rho,\;\; I \in \mathcal{I},\;\; \Delta_J^Q \leq DV_J^Q,\;\text{for}\;\; J \in \mathcal{I}_+(I)\}. \tag{3.13}$$

An adaptive estimation $\hat{\boldsymbol{I}}_{\boldsymbol{\rho}}$ of $\mathbb{I}_\rho$ can then be determined using the same acceptance rule concerning the mean as for the determination of $\hat{\boldsymbol{I}}$. Following the proof of Theorem 3.5, we can then get the accuracy of $\hat{\rho}_{\hat{I}_\rho}$:

$$\mathbb{P}\left(\|\hat{\rho}_{\hat{I}_\rho} - \rho_{N+1}\|_\infty \geq k(D)\,\sigma\sqrt{\frac{\ln n(n+1) + \lambda \ln|\mathbb{I}_\rho|}{|\mathbb{I}_\rho|}}\right) \leq \sum_{I\in\mathcal{I}\;|I\subseteq\mathbb{I}_\rho}\;\sum_{J\in\mathcal{I}_+(I)}\frac{1}{|J|^\lambda};$$

where the constant $k(D) = 7 + 3\sqrt{2} + 12\sqrt{\frac{2}{\ln 2}}D$. A similar result can be obtained for the accuracy of the estimation of Q_{N+1} by $\hat{\boldsymbol{Q}}_{\hat{\boldsymbol{I}}_{\boldsymbol{Q}}}$ using the data of the interval $\hat{\boldsymbol{I}}_{\boldsymbol{Q}}$. Notice that the condition $\ln n(n+1) + \lambda \ln m_0 \leq m_0$ in the above Theorem 3.5 can be suppressed but this leads to more complicated left hand sides. However, this condition is not too restrictive. For instance for $n = 40$, if we take $m_0 = n$, then we can take values of λ as large as 8.84. If we choose $\lambda = 1$ and $m_0 = n$, then it suffices for r_t to have more than six components ($n \geq 6$) to get $\ln n(n+1) + \lambda \ln m_0 \leq m_0$. Notice that, following the proof of Theorem 3.5, we can give the accuracy of the adaptive estimators obtained using the different implementations of the adaptive algorithm described in Section 3.3.2.

Theorem 3.6 *Let $\mathbb{I}_\rho = [N + 1 - m_0 - k_{\mathbb{I}}, N + 1[$ ($k_{\mathbb{I}} \in \mathbb{N}$) be the optimal homogeneity interval for the mean, let $\lambda > 0$ satisfy $\ln n(n+1) + \lambda \ln m_0 \leq m_0$ and let*

$$\mathbb{P}_\rho = \mathbb{P}\left(\|\hat{\rho}_{\hat{I}_\rho} - \rho_{N+1}\|_\infty \geq k(D)\,\sigma\sqrt{\frac{\ln n(n+1) + \lambda \ln|\mathbb{I}_\rho|}{|\mathbb{I}_\rho|}}\right),$$

where $k(D) = 7 + 3\sqrt{2} + 12\sqrt{\frac{2}{\ln 2}}D$. Different choices of testing subintervals were envisaged in Section 3.3.2. If we want to test the homogeneity on $I = [N+1-k, N+1[$,

we can check the difference of the estimations on I and $J = [N+1-k,\ N-k+m_0]$ *(choice (a)), on J and* $J' = [N+1-m_0,\ N+1[$ *(choice (b)), or on J and (*$J' = [N+1-m_0,\ N+1[$ *if* $k \leq 2m_0$ *and* $J' = [N+1-k+m_0,\ N+1[$ *otherwise) (choice (c)). For these different implementation choices, we have:*

$$(a)\ \ \mathbb{P}_\rho \leq \frac{k_{\mathbb{I}}}{m_0^\lambda} + \sum_{k=1}^{k_{\mathbb{I}}} \frac{1}{(m_0+k)^\lambda}, \qquad (b)\ \ \mathbb{P}_\rho \leq \frac{k_{\mathbb{I}}+1}{m_0^\lambda} + \frac{1}{(m_0+k_{\mathbb{I}})^\lambda},$$

$$(c)\ \ \mathbb{P}_\rho \leq \frac{k_{\mathbb{I}}+1}{m_0^\lambda} + \frac{1}{(m_0+k_{\mathbb{I}})^\lambda} + \sum_{k=m_0+1}^{k_{\mathbb{I}}} \frac{1}{k^\lambda}.$$

A similar result can be given on the accuracy of $\hat{Q}_{\hat{I}_Q}$.

4 Accuracy of the approximate problem

4.1 The stationary case

We first consider the case of constant means, variances and covariances. We suppose that N returns $\mathbf{r_i}$, $i = 1, \ldots, N$, satisfying Assumption 3.1 are available to compute the empirical estimations $\hat{\boldsymbol{\rho}}_{\boldsymbol{N+1}}$ and $\hat{\boldsymbol{Q}}_{\boldsymbol{N+1}}$ of the mean and the covariance matrix of the process α_t defined in Subsection 3.1.

Definition 4.1 *For any* $n \times n$ *real symmetric matrix Q, let* $\beta(Q)$ *be such that the quadratic function* $x^\top Qx$ *is* $\beta(Q)$*-strongly convex w.r.t.* $\|\cdot\|_1$*, i.e.*

$$\beta(Q) = \inf_{x \neq 0} \frac{x^\top Qx}{\|x\|_1^2}.$$

Theorem 4.2 *Let the process* r_t *satisfy* (1.1) *with constant mean* $\rho_t = \rho$ *and covariance matrix* $Q_t = Q$ *and let Assumption 3.1 hold. Let* $\lambda > 0$ *and let* $(\hat{\boldsymbol{\rho}}_{\boldsymbol{N+1}}, \hat{\boldsymbol{Q}}_{\boldsymbol{N+1}})$ *be the estimations of* (ρ_{N+1}, Q_{N+1}) *given in* (3.2)*. If* $\ln n(n+1) + \lambda \ln N \leq N$*, then there is a constant k (given in the appendix) and a set* $S \subseteq \Omega$ *of probability at least* $1 - \frac{3}{N^\lambda}$ *such that for any* $\omega \in S$*, the accuracy of problem* $\mathcal{P}(\hat{\boldsymbol{\theta}})$ *is bounded as follows:*

$$\epsilon(\mathcal{P}(\hat{\boldsymbol{\theta}})) \leq k \max(\sigma^{\alpha_0}, \sigma^{2\alpha_0}) \left(\frac{\ln n(n+1) + \lambda \ln N}{N}\right)^{\frac{\alpha_0}{2}} \max_{x \in X} \|x\|_1^{p_0}. \tag{4.1}$$

Theorem 4.3 *Let the process* r_t *satisfy* (1.1) *with constant mean* $\rho_t = \rho$ *and covariance matrix* $Q_t = Q$ *and let Assumption 3.1 hold. Let* $\lambda > 0$ *and let* $(\hat{\boldsymbol{\rho}}_{\boldsymbol{N+1}}, \hat{\boldsymbol{Q}}_{\boldsymbol{N+1}})$ *be the estimations of* (ρ_{N+1}, Q_{N+1}) *given in* (3.2)*. If* $\ln n(n+1) + \lambda \ln N \leq N$*, then there are constants* k_1 *and* k_2 *(given in the appendix) and a set* $S \subseteq \Omega$ *of probability at least* $1 - \frac{3}{N^\lambda}$ *such that for any* $\omega \in S$ *the accuracy of problem* $\tilde{\mathcal{P}}(\hat{\boldsymbol{\theta}})$ *is bounded as follows:*

$$\epsilon(\tilde{\mathcal{P}}(\hat{\boldsymbol{\theta}})) \leq (k_1 + 2\kappa\sqrt{k_2})\sigma \left(\frac{\ln n(n+1) + \lambda \ln N}{N}\right)^{\frac{1}{4}} \max_{x \in X} \|x\|_1. \tag{4.2}$$

Further, if the matrix Q is non-degenerate, i.e. if $\beta(Q) > 0$, then

$$\epsilon(\tilde{\mathcal{P}}(\hat{\boldsymbol{\theta}})) \leq \left(k_1 + \frac{2\kappa k_2 \sigma}{\sqrt{\beta(Q)}}\right) \sigma \sqrt{\frac{\ln n(n+1) + \lambda \ln N}{N}} \max_{x \in X} \|x\|_1. \tag{4.3}$$

The accuracy of the approximate problem can also be bounded without the condition $\ln n(n+1) + \lambda \ln N \leq N$. In the general case, we get more complicated bounds (see the appendix).

4.2 Slowly varying parameters

In the case of slowly varying parameters, the accuracy of the approximate problem is roughly obtained by replacing, in the results given in the stationary case, the number of observations N with the length $|\mathbb{I}|$ of the ideal interval of local time homogeneity. Indeed, $\hat{\boldsymbol{\rho}}_{\hat{I}}$ and $\hat{\boldsymbol{Q}}_{\hat{I}}$ are close to $\hat{\boldsymbol{\rho}}_{\mathbb{I}}$ and $\hat{\boldsymbol{Q}}_{\mathbb{I}}$ (see Theorems 3.4 and 3.5).

Theorem 4.4 *Let $(\hat{\boldsymbol{\rho}}_{\boldsymbol{N+1}}, \hat{\boldsymbol{Q}}_{\boldsymbol{N+1}})$ be the estimations of (ρ_{N+1}, Q_{N+1}) given in Subsection 3.1 in the case of slowly varying parameters. Let $\lambda > 0$ be the parameter involved in the definition of K_0 and such that $\ln n(n+1) + \lambda \ln m_0 \leq m_0$, where m_0 is the length of the smallest testing subinterval. Then there is a constant $k(D)$ (given in the appendix) and a set $S \subseteq \Omega$ of probability at least $1 - \sum_{I \in \mathcal{I}} \sum_{|I \subseteq \mathbb{I}} \sum_{J \in \mathcal{I}_+(I)} \frac{3}{|J|^\lambda}$ such that for any $\omega \in S$*

$$\epsilon(\mathcal{P}(\hat{\boldsymbol{\theta}})) \leq k(D) \max\left(\sigma^{\alpha_0}, \sigma^{2\alpha_0}\right) \left(\frac{\ln n(n+1) + \lambda \ln |\mathbb{I}|}{|\mathbb{I}|}\right)^{\frac{\alpha_0}{2}} \max_{x \in X} \|x\|_1^{p_0}. \tag{4.4}$$

Theorem 4.5 *Let $(\hat{\boldsymbol{\rho}}_{\boldsymbol{N+1}}, \hat{\boldsymbol{Q}}_{\boldsymbol{N+1}})$ be the estimations of (ρ_{N+1}, Q_{N+1}) given in Subsection 3.1 in the case of slowly varying parameters. Let $\lambda > 0$ be the parameter involved in the definition of K_0 and such that $\ln n(n+1) + \lambda \ln m_0 \leq m_0$, where m_0 is the length of the smallest testing subinterval. Then there are constants $k_1(D)$ and $k_2(D)$ (given in the appendix) depending affinely on D and a set $S \subseteq \Omega$ of probability at least $1 - \sum_{I \in \mathcal{I}} \sum_{|I \subseteq \mathbb{I}} \sum_{J \in \mathcal{I}_+(I)} \frac{3}{|J|^\lambda}$ such that for any $\omega \in S$*

$$\epsilon(\tilde{\mathcal{P}}(\hat{\boldsymbol{\theta}})) \leq (k_1(D) + 2\kappa\sqrt{k_2(D)})\sigma \left(\frac{\ln n(n+1) + \lambda \ln |\mathbb{I}|}{|\mathbb{I}|}\right)^{\frac{1}{4}} \max_{x \in X} \|x\|_1. \tag{4.5}$$

Further, if the matrix Q_{N+1} is non-degenerate, i.e. if $\beta(Q_{N+1}) > 0$, then

$$\epsilon(\tilde{\mathcal{P}}(\hat{\boldsymbol{\theta}})) \leq \left(k_1(D) + \frac{2\kappa k_2(D)\sigma}{\sqrt{\beta(Q_{N+1})}}\right) \sigma \sqrt{\frac{\ln n(n+1) + \lambda \ln |\mathbb{I}|}{|\mathbb{I}|}} \max_{x \in X} \|x\|_1. \tag{4.6}$$

It is interesting to notice that the upper bound we obtain on the accuracy of the problem weakly increases with problem dimension. Thus, if $\mathbb{I}$ is sufficiently long, we can build an approximate problem of good quality even when the number n of components is

very large. For problem $(\mathcal{V})$ (see the application we consider in the next section), we can bound from above $\max_{x\in X} \|x\|_1$ (appearing in (4.2), (4.3), (4.5) and (4.6)) by $\|x^-\|_1$, where x^- is the portfolio before reallocation.

5 Numerical simulations: application in finance

5.1 Presentation of the application

We introduce in this subsection a portfolio selection model belonging to the class $\tilde{\mathcal{P}}(\theta)$ for which we build an approximate problem as explained before using both simulated and real data. Let H be the investment horizon, $N+1$ the date of the investment, and let $\mathbf{s_t} = (\mathbf{s_t}(1), \ldots, \mathbf{s_t}(n))$ be an observed asset process in discrete time, $t = 1, \ldots, N+1$, where n is the number of risky assets, whose prices at time t are collected in the vector $\mathbf{s_t} \in \mathbb{R}^n$. We define the corresponding H time steps return $\mathbf{r_t^H}$ at time t by $\mathbf{r_t^H} = \frac{\mathbf{s_{t+H}}}{\mathbf{s_t}}$, $t = 1, \ldots, N+1-H$, where the division should be understood componentwise: $\mathbf{r_t^H}(i) = \frac{\mathbf{s_{t+H}}(i)}{\mathbf{s_t}(i)}$, $i = 1, \ldots, n$. Notice that though we are at time $N+1$, the data $\mathbf{r_t^H}$, $t = N+2-H, \ldots, N+1$, is not available. However, the returns we are interested in are the returns $\mathbf{r_{N+1}^H}(i)$, $i = 1, \ldots, n$, over the investment period. We suppose that there is a past interval with right endpoint $N+1$ such that the parameters ρ_t and Q_t slowly vary on this interval. We also suppose that this interval is of a length greater than $H+k$ where k is the minimal number of data needed for estimation. Finally, to enter the framework specified in this article, we suppose the returns r_t are independent. We thus first simulate independent returns. However, in practice, the assumption of independence of the returns is not true (though it is a simplification commonly made) but we also test our procedure using real data to measure in practice the behavior of our methodology when applied to this portfolio selection model.

We fix an investment horizon $H = 1$ and denote from now on by r_t the returns at time t. In addition to the n risky assets, we have a risk-free asset and we take into account the transaction costs. We consider a single investment period and have to decide the amount of money to invest in the different assets over this investment period. The quantities referring to the risk-free asset are indexed by $n+1$. In this setting, a simplified portfolio selection problem ([DI93]) is as follows. Let x_i be the amount of asset i in the portfolio at the beginning of the period. The flow balance equations for the x_i are then given by:

$$\begin{cases} x_i &= x_i^- - y_i + z_i \text{ for the risky assets, } i = 1, \ldots, n, \\ x_{n+1} &= x_{n+1}^- + \sum_{i=1}^{n} (1-\mu_i)\, y_i - \sum_{i=1}^{n} (1+\nu_i)\, z_i \text{ for the risk-free asset,} \end{cases}$$

once we have defined

- x_i^- : the initial value of i-th asset before the rebalancing of the portfolio;
- y_i : the amount of risky asset i we sell at the beginning of the period, μ_i being the corresponding transaction cost (with $0 < \mu_i < 1$);
- z_i : the amount of risky asset i we buy at the beginning of the period, with the corresponding transaction cost ν_i (with $0 < \nu_i < 1$).

In what follows, a portfolio is thus a vector $x = (x_1, \ldots, x_n)^\top$ of n amounts invested in the different risky assets plus risk-free asset x_{n+1} and $\mathbf{r_{N+1}} = (\mathbf{r_{N+1}}(1), \ldots, \mathbf{r_{N+1}}(n))^\top$ is the vector of risky asset returns over the investment period. Note that the risk-free asset return $r(n+1)$ is known. The goal is then naturally to maximize the final total value of the portfolio given by $x^\top \mathbf{r_{N+1}} + r(n+1)x_{n+1}$.

Case of complete information. If we knew the returns $\mathbf{r_{N+1}}$, we could solve the following linear program:

$$\mathcal{ALLOC} \left\{ \begin{array}{lcl} \max & x^\top \mathbf{r_{N+1}} + r(n+1)x_{n+1}, & \\ x_i & = & x_i^- - y_i + z_i, \quad i = 1, \ldots, n, \\ x_{n+1} & = & x_{n+1}^- + \displaystyle\sum_{i=1}^{n} (1-\mu_i)\, y_i - \sum_{i=1}^{n} (1+\nu_i)\, z_i, \\ x \geq 0, & x_{n+1} \geq 0, & y \geq 0, \quad z \geq 0. \end{array} \right.$$

We denote by $X(\mu, \nu, x^-)$ the set of admissible portfolios defined by the above flow balance equations and the positivity constraints on x, x_{n+1}, y and z.

The problem of data uncertainty. In fact, the data which are known the moment we choose the portfolio are the transaction costs μ and ν and the return $r(n+1)$ of the risk-free asset. In order to solve the previous allocation problem, we could use a realization of the returns over the investment period and plug these values into $\mathcal{ALLOC}$. At first glance, this approach fails in ensuring a given target income with high probability. This means that we have to take into account the risk of our investment.

We use the Value-at-Risk technique, which is a modelling tool to make a decision in an uncertain environment. In the Value-at-Risk model, an investment is considered risky if its return has little chance of exceeding a certain reasonable value, fixed in advance. More precisely, given the distribution of the returns, if we fix a confidence level $0 < \varepsilon < \frac{1}{2}$, the maximal return that can be reached with probability greater than or equal to $1-\varepsilon$, for an admissible portfolio $(x_1, \ldots, x_{n+1})$, is the Value-at-Risk of level ε, $V_\varepsilon(x_1, \ldots, x_{n+1})$, of this portfolio:

$$V_\varepsilon(x, x_{n+1}) = \max\ \gamma \text{ subject to } \mathbb{P}\big(r_{N+1}^\top x + r(n+1)x_{n+1} \geq \gamma\big) \geq 1-\varepsilon.$$

A VaR asset allocation problem then amounts to solving:

$$\max\ \ V_\varepsilon(x, x_{n+1}) \text{ subject to } (x, x_{n+1}, y, z) \in X(\mu, \nu, x^-). \tag{5.1}$$

If the distribution of the risky asset returns r_{N+1} is Gaussian with given mean $\mathbb{E}[r_{N+1}] = \rho_{N+1}$ and covariance matrix $\mathbb{E}[(r_{N+1}-\rho_{N+1})(r_{N+1}-\rho_{N+1})^\top] = Q_{N+1}$, then we obtain $V_\varepsilon(x, x_{n+1}) = \rho_{N+1}^\top x + r(n+1)\, x_{n+1} - \Phi^{-1}(1-\varepsilon)\sqrt{x^\top Q_{N+1} x}$, where Φ is the CDF of the Gaussian density. Using the exact version of Chebyshev bound (see [BP05, Smi95]), we can show that if the returns are not Gaussian, an upper bound on the optimal value of (5.1) is given solving

$$(\mathcal{V}) \left\{ \begin{array}{l} \min\ \ \kappa(\varepsilon)\sqrt{x^\top Q_{N+1} x} - \rho_{N+1}^\top x - r(n+1)x_{n+1} \\ (x, x_{n+1}, y, z) \in X(\mu, \nu, x^-) = X, \end{array} \right. \tag{5.2}$$

where now $\kappa(\varepsilon) = \sqrt{\frac{1-\varepsilon}{\varepsilon}}$. Thus, in every case, the problem of maximizing the Value-at-Risk over all admissible portfolios reduces to (5.2), which is a problem of the form $\tilde{\mathcal{P}}(\theta)$, where $\kappa(\varepsilon)$ is a risk factor depending on the assumptions for the distribution of the returns:

- $\kappa(\varepsilon) = \Phi^{-1}(1-\varepsilon) > 0$ for Gaussian returns,
- $\kappa(\varepsilon) = \sqrt{\frac{1-\varepsilon}{\varepsilon}}$ for non-Gaussian random returns in $L_1(\mathbb{R}) \cap L_2(\mathbb{R})$.

We now intend to test the accuracy of the approximate problem $(\hat{\mathcal{V}})$ using the adaptive estimations of problem $(\mathcal{V})$ parameters. We use both simulated and real data for the returns $\mathbf{r_t}$, $t = 1, \ldots, N$, available the day $N+1$ of the investment. The efficiency of the adaptive algorithm itself (introduced in Section 3) is tested in [Gui05] (to detect break points in a change point model).

For more flexibility, we use the empirical adaptive estimators of the mean and of the covariance matrix. The empirical adaptive estimators are defined in [Gui05] and are obtained using the adaptive algorithm and replacing $\boldsymbol{\alpha}_t^{\boldsymbol{I}}$ by $\mathbf{r_t}$ in (3.3) (which leads to more standard definitions for $\hat{\boldsymbol{\rho}}_{\boldsymbol{I}}$ and $\hat{\boldsymbol{Q}}_{\boldsymbol{I}}$). In this manner, we can show (see [Gui05]) that the adaptive algorithm now depends on two positive parameters λ and μ. We then use the following acceptance rules for a given interval I. We accept the interval I as an interval of local time homogeneity if for all $J \in \mathcal{I}(I)$, $\|\hat{\rho}_I - \hat{\rho}_J\|_\infty \leq k_1\sigma\left(\sqrt{\frac{\ln n+\lambda\ln|I|}{|I|}} + \sqrt{\frac{\ln n+\lambda\ln|J|}{|J|}}\right)$ and $\|\hat{Q}_I - \hat{Q}_J\|_\infty \leq k_2\sigma^2\left(\sqrt{\frac{\ln n(n+1)+\mu\ln|I|}{|I|}} + \sqrt{\frac{\ln n(n+1)+\mu\ln|J|}{|J|}}\right)$ where (λ, μ, k_1, k_2) are positive constants. The parameter σ is estimated from the data. From a theoretical point of view, we can still bound from above the accuracy of the approximate problem which uses the empirical adaptive estimations (see [Gui05]). We did not choose to present this approach as it leads to more complicated and less precise upper bounds which are obtained under more restrictive conditions on the process r_t. However, the definition of the approximate problem as a function of the adaptive estimators, the dependence of the upper bounds on the quality of the approximate problem as a function of n and $|\mathbb{I}|$, and the tools used to show the results are the same (see [Gui05]).

To reduce the computational cost, we use the choice of testing subintervals denoted by choice (b) in Theorem 3.6. The intervals of the set $\mathcal{I}$ are of the form $I_k = [N+1-m_0-k, N+1[$, where $k \in \mathbb{N}$. Finally, the optimization problems are solved using Matlab and the Mosek optimization library.

5.2 Simulated data

In this section we are interested in the accuracy of the approximate VaR problem $(\hat{\mathcal{V}})$ in the particular case where the model for the returns is a change point model. The day of the investment, $N = T_1 + T_2$ independent observations of the returns are available where the first T_1 data is drawn from the Gaussian density $\mathcal{N}(\rho_1, Q_1)$ and the last T_2 observations from the Gaussian density $\mathcal{N}(\rho_2, Q_2)$. Assumption 3.1 is thus satisfied. We then assume that the mean return and the covariance matrix between the

returns over the investment period are respectively ρ_2 and Q_2. We thus know the optimal portfolio x^* that would be obtained solving the VaR problem $(\mathcal{V})$ with the values ρ_2 and Q_2 of the parameters. To use meaningful values of ρ_1 and Q_1 we choose $\rho_1 = \rho$ and $Q_1 = Q$ where ρ and Q are the empirical mean and covariance matrix obtained using the available 3 month returns of the assets of the Dow Jones we have (see the next subsection) on January 2, 1995. The matrix Q is thus an $n \times n$ matrix with $n = 30$. We consider a change point model with a change in the mean only: $\rho_2 = 1.25\rho_1$ and $Q_2 = Q_1$. We consider 200 realizations of such a change point time series and for each realization, we compute the portfolios obtained solving $(\hat{\mathcal{V}})$ using the empirical estimations of the parameters with different estimation horizons:

- a fixed estimation horizon using the last T_2 observations (method denoted by "Last"); a fixed estimation horizon using the first T_1 observations (method denoted by "First"); a fixed estimation horizon using all the data (method denoted by "Emp");
- the estimation horizon provided by the adaptive algorithm (method denoted by "Adap").

The risk-free rate chosen is the American federal bank loan rate the day of the investment (January 2, 1995). We choose $\kappa(\varepsilon) = 0.25$ and $T_1 = T_2$ with $T_1 = 50$, $T_1 = 100$ or $T_1 = 200$. We also test the procedure for different values of the number n of assets: $n = 30, 100, 500$ and $n = 1000$. When $n = 30$, we have just described how ρ and Q are computed. When $n = 100$, we choose $\rho = [\rho_{30}^\top, \rho_{30}^\top, \rho_{30}^\top, \rho_{30}^\top(1{:}10)]^\top$ and Q =blkdiag(Q_{30}, Q_{30}, Q_{30}, Q_{30}(1:10,1:10)) where ρ_{30} and Q_{30} are the mean ρ and covariance matrix Q computed when $n = 30$, and where blkdiag(Q_1,Q_2), for matrices Q_1 and Q_2, is the block diagonal matrix $\begin{pmatrix} Q_1 & 0 \\ 0 & Q_2 \end{pmatrix}$. When $n > 100$, we choose $\rho = 0.98 + 0.22\frac{i-1}{n-1}$, $i = 1, \ldots, n$, and Q =diag($0.25\rho^2$) (not to get an ill-conditioned matrix Q in high dimension). The parameters λ and μ of the adaptive algorithm are fixed: $\lambda = \mu = 1$. For fixed T_1 and T_2, the parameters k_1 and k_2 are those (among a grid of values for k_1 and k_2) providing the smallest type II error while providing a type I error of at most 5% for a stationary model with $\rho_2 = \rho_1 = 1.25\rho$ and $Q_2 = Q_1 = Q$. More precisely, we simulate 200 samples of size $T_1 + T_2$ drawn from the Gaussian density $\mathcal{N}(1.25\rho, Q)$. For each sample, a type I error is made when the whole interval of length $T_1 + T_2$ is not accepted as an interval of homogeneity. The same procedure is conducted with a change point model where $\rho_2 = 1.25\rho_1$ and $Q_2 = Q_1$. In this case, a type II error is made when the length of the adaptive interval is greater than T_2. The grids of values chosen for k_1 and k_2 are the same: [0.05;0.1;0.2;0.3;0.4;0.5;1]. We then compute the 90^{th} percentile of the accuracy. The results are given in Table 5.1.

We then conduct the same experiment with a stationary time series: $\rho_2 = \rho_1 = \rho$ and $Q_2 = Q_1 = Q$. The results are given in Table 5.2 which follows. The accuracy of the approximate problem is satisfying, close to the accuracy obtained using the true interval of homogeneity and much better than the accuracy obtained using all the data in the case of a change point time series.

n	N	Emp	Adap	First	Last	$\hat{\sigma}$	Mean($\|\hat{I}\|$)
30	100	29.12	25.82	26.90	25.82	2.03	49.40
30	200	23.04	19.57	21.03	18.92	2.077	98.8
30	400	20.48	15.89	17.83	14.09	2.104	195.85
100	100	19.00	11.11	12.20	11.35	2.084	49.8
100	200	16.05	11.25	11.48	11.39	2.14	100.05
100	400	9.01	8.46	6.48	7.51	2.22	199.25
500	100	10.22	6.47	6.32	6.47	2.22	50.00
500	200	8.50	4.64	5.23	4.64	2.26	100.1
500	400	6.06	2.84	2.96	2.84	2.26	199.3
1000	100	11.36	7.45	6.82	7.45	2.26	50.00
1000	200	8.04	4.28	4.40	4.28	2.22	100.00
1000	400	7.82	3.20	3.84	3.20	2.30	200.00

Table 5.1 Comparison of the 90^{th} percentile of the accuracy (defined in Section 2) of the approximate problem ($\hat{\mathcal{V}}$) using different estimation procedures for the parameters. Change point time series for the returns.

n	N	Emp	Adap	First	Last	$\hat{\sigma}$	Mean($\|\hat{I}\|$)
30	100	15.06	15.96	18.92	18.96	1.118	96.8
30	200	9.19	10.81	13.25	13.29	2.034	191.6
30	400	4.25	5.30	8.16	8.99	1.828	388.85
100	100	16.19	18.97	25.57	23.01	1.840	98.9
100	200	8.25	9.38	15.64	14.22	1.8703	196.05
100	400	3.46	3.46	7.74	7.92	2.047	383.9
500	100	14.07	14.07	29.59	22.46	1.923	98.8
500	200	2.54	2.54	5.63	4.53	2.029	195.45
500	400	1.45	1.45	2.44	2.47	2.047	388.45
1000	100	4.83	4.83	8.37	7.45	1.9376	100
1000	200	2.84	2.84	4.74	6.11	2.081	195.05
1000	400	1.70	1.70	3.50	2.95	2.048	391.1

Table 5.2 Comparison of the 90^{th} percentile of the accuracy (defined in Section 2) of the approximate problem ($\hat{\mathcal{V}}$) using different estimation procedures for the parameters. Stationary time series for the returns.

5.3 Simulations with real data

Our objective is now to compare the evolution of portfolios obtained with Value-at-Risk model $(\mathcal{V})$ rebalancing the portfolio for different dates and using different methods of calibration of the problem parameters. We are interested in the portfolio return and in the volatility of the portfolio return over the investment period. The different calibrations tested are the empirical estimations (using all the available data the day of the investment and computing the empirical estimations), the empirical estimations using a fixed given estimation horizon and the empirical adaptive estimations. The performances of the portfolios are also compared with a portfolio tracking the Dow Jones and with a risk-free investment (investing everything in the risk-free asset). We consider the 30 assets of the Dow Jones. We have the prices of these assets from January 2, 1992 to June 30, 2004. Notice that these prices are corrected and take into account the splits and capital growth. We invest \$1000 (this money is initially held in the risk-free asset) on January 2, 1995. We choose an investment horizon H. Different values of H are tested: $H = 90$ days of open stock markets (approximately 4 months and a half), $H = 60$ and $H = 20$. The portfolio is then regularly rebalanced every H days of open stock market, using model $(\mathcal{V})$. The risk-free rate used for an investment is the H-day American federal bank loan rate the day of the investment. The transaction costs amount to 0.5 %.

Different policies of choice of parameters of the adaptive algorithm will provide different estimation horizons. We now explain how the parameters of the adaptive algorithm are chosen.

5.3.1 A posteriori choice of parameters

The parameters of the adaptive algorithm are first determined a posteriori to show the influence of the estimation horizon on the performance of the portfolios. We determine one homogeneity interval $\hat{\boldsymbol{I}}$, the intersection of $\hat{\boldsymbol{I}}_{\boldsymbol{\rho}}$ and $\hat{\boldsymbol{I}}_{\boldsymbol{Q}}$. A grid of values is chosen for the parameters k_1, k_2 and m_0. The values of λ and μ are fixed to 0.5 to reduce the computational costs. We envisage all possible combinations of the values of the parameters k_1, k_2 and m_0 and choose the combination giving the maximal return over the investment period. Notice that the same parameters are used at each rebalancing. A dynamic choice of parameters could still improve the results. The results are summarized in Table 5.3 which follows. In this table (and in what follows) R_{Ad} is the return of the portfolio obtained with the adaptive method over the whole investment period and R_{Emp} is the return of the portfolio obtained with the empirical method over the same period. Knowing the evolution of the portfolio wealth over the investment period, we can compute the sample of H-day returns of the portfolio. The empirical mean of this sample (the H-day mean return) is R'_{Ad} (if the adaptive method is used) or R'_{Emp} (if the empirical method is used) and its empirical standard deviation (which measures the volatility of the portfolio return) is σ_{Ad} for the adaptive method and σ_{Emp} for the empirical method. We see that we can always find values of the parameters which provide a portfolio having a return larger than the "empirical portfolio" over the investment period. It seems that the shorter the investment horizon, the more interesting the adaptive algorithm. This would mean that the shorter the H, the more the means and variances of the returns vary (in particular for $H = 20$ where the use of

Horizon	k_1	k_2	m_0	R_{Ad}	R'_{Ad}	σ_{Ad}	R_{Emp}	R'_{Emp}	σ_{Emp}
$H = 20$	0.05	0.1	30	3.801	1.02	0.065	1.3425	1.0043	0.0003
$H = 60$	0.3	0.1	30	4.39	1.069	0.11	2.6217	1.0474	0.0765
$H = 90$	0.3	0.1	30	7.025	1.126	0.1806	4.1285	1.1019	0.1028

Table 5.3 Comparison between the adaptive and the empirical method. Static a posteriori choice of the parameters.

the empirical estimations gives a portfolio performing less well than a risk-free asset based investment). This method of choice of parameters is called Best Adap in what follows.

5.3.2 A priori choices of parameters

From a practical point of view, we have to determine a policy of choice of parameters using only the available information the day of the investment. We use the same grid of admissible values for the parameters of the adaptive algorithm as for the a posteriori determination of parameters. Parameters are chosen in a way to optimize a certain error criterion. We use three different dynamical techniques (the values of the parameters can change from one rebalancing to another) to solve this problem. We can first find the values of the parameters giving the smallest Mean Absolute Forecast Error (MAFE). If $\hat{\boldsymbol{\rho}}(\boldsymbol{t})$ is the forecasted adaptive mean for time step t, then using the observations $(\mathbf{r_1}, \ldots, \mathbf{r_{t-H}})$ available at time step t, the MAFE computed at time step t is:

$$\text{MAFE} = \frac{1}{t_0} \sum_{k=t-H-t_0+1}^{t-H} \|\mathbf{r_k} - \hat{\boldsymbol{\rho}}(k)\|_\infty,$$

where $t_0 > 0$ is a parameter. For our simulations, the dates chosen are the last $t_0 = 5$ dates for which the returns of the assets are known. Notice that if we decide to find two different intervals $\hat{\boldsymbol{I}}_{\boldsymbol{\rho}}$ and $\hat{\boldsymbol{I}}_{\boldsymbol{Q}}$, this method provides a calibration of k_1 and λ. This method will be tested if we only determine one interval of homogeneity. In this case, the values of k_2 and μ also influence the estimation of the interval of homogeneity and consequently influence the estimation of the mean. We call this method Adap 1. A second way of calibrating the parameters of the adaptive algorithm consists of simulating different investments in the past. We choose 5 different investment dates preceding the day of the rebalancing (for which we want to calibrate the parameters) and such that the prices of the assets are known for H days following these different dates. This allows us to compute the optimal portfolios for every value of parameters and to see the real evolution of the different portfolios. We then choose the values of parameters giving the maximal mean return over all testing dates (method Adap 2) or the maximal Sharpe ratio which is defined as the H-day mean return divided by the H-day standard deviation (method denoted by Adap 3). We call "Fix" the method using a fixed estimation horizon (different fixed estimation horizons are considered: 50, 100, 150, 200, 250 and 300 and the one providing

Model	R_{90}	R'_{90}	σ_{90}	Sharpe_{90}	R_{60}	R'_{60}	σ_{60}	Sharpe_{60}
Emp	4.13	1.100	0.103	10.72	2.62	1.047	0.077	13.69
Fix	5.88	1.131	0.148	7.64	4.56	1.070	0.128	8.37
Best Adap	7.02	1.126	0.181	6.23	4.39	1.069	0.110	9.74
Adap 1	2.64	1.069	0.116	9.25	3.63	1.061	0.097	15.23
Adap 2	2.85	1.075	0.155	6.95	1.94	1.034	0.131	7.89
Adap 3	4.45	1.107	0.157	7.04	2.04	1.034	0.120	10.34

Table 5.4 Comparison of different methods of estimation of the parameters of model ($\mathcal{V}$) for $H = 90$ and $H = 60$.

Model	R_{20}	R'_{20}	σ_{20}	Sharpe_{20}
Emp	1.34	1.004	0.0003	3348
Fix	5.37	1.026	0.065	15.76
Best Adap	3.80	1.020	0.065	15.62
Adap 1	1.08	1.002	0.032	31.49
Adap 2	2.66	1.018	0.048	21.21
Adap 3	2.92	1.019	0.050	20.38

Table 5.5 Comparison of different methods of estimation of the parameters of model ($\mathcal{V}$) for $H = 20$.

the best return over the investment period is chosen). Emp stands for the method using all the available data to compute the empirical estimations of the parameters. Two investment periods are chosen: the first one goes from January 2, 1995 to the beginning of May 2000; the second one goes from January 2, 1995 to June 30, 2004.

We first report in Tables 5.4 and 5.5 above, for the first investment period (from January 2, 1995 to the beginning of May 2000), the global return R_H for each method, the mean return R'_H, the standard deviation σ_H and the Sharpe ratio $\text{Sharpe}_H = \frac{R'_H}{\sigma_H}$ for the H-day return sample.

Not all the adaptive methods allow one to obtain either a better global return or a better Sharpe ratio over the investment period. Nevertheless, for every investment horizon H, there is always an adaptive method yielding a global return larger than the global return of method "Emp". In particular, method "Adap 3" beats method "Emp" for $H = 90$ and for $H = 20$ where the global return is more than doubled.

We now plot, in Figure 5.1 which follows, the evolution of the portfolios using the different investment methods and the two investment periods. Only the adaptive method (Adap 1, Adap 2 or Adap 3) providing the maximal return is shown in this figure.

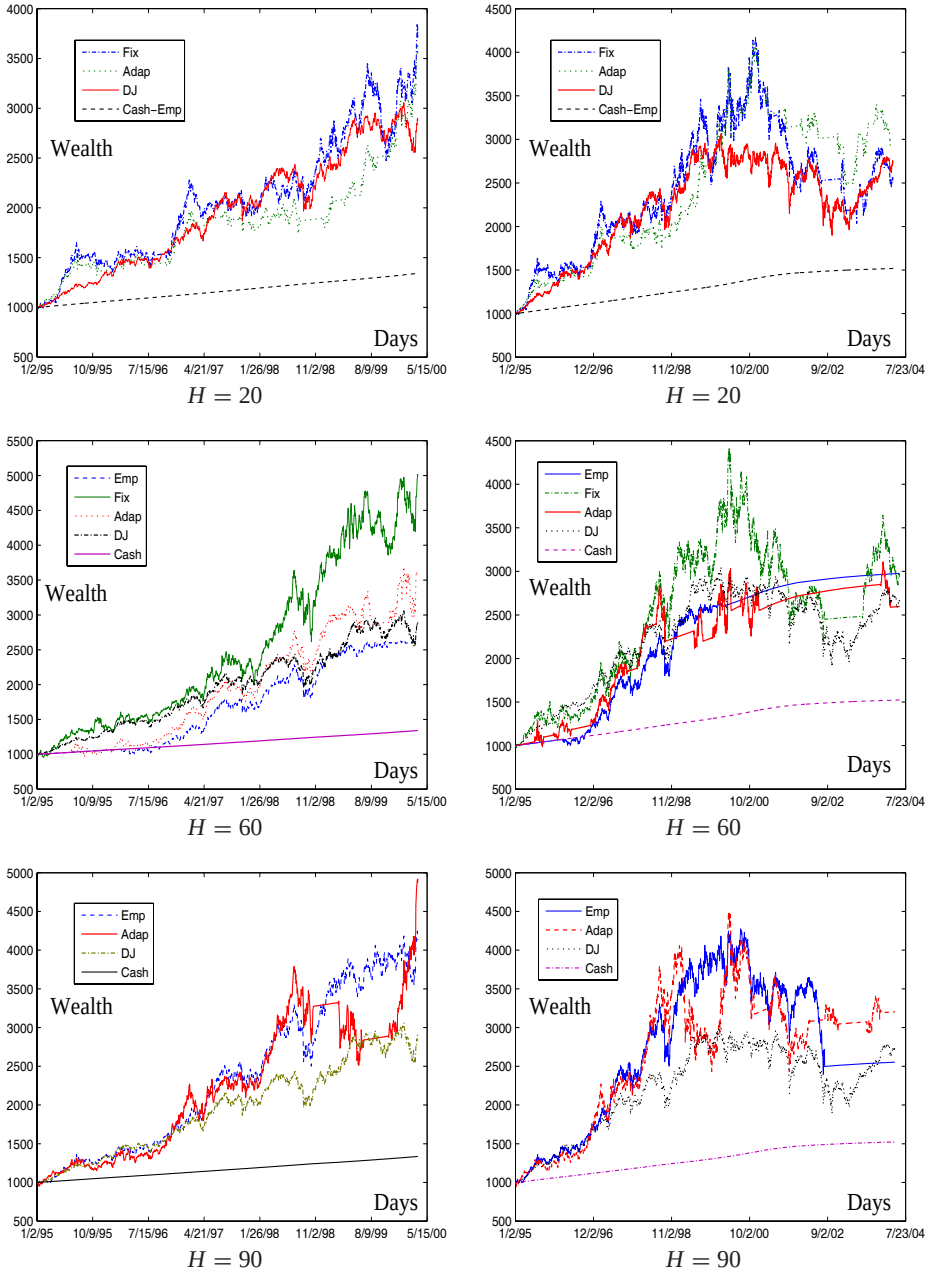

Figure 5.1 Comparison of the portfolios performances using asset allocation model ($\mathcal{V}$). On the left, the plots correspond to the first investment period going from January 2, 1995 to the beginning of May, 2000 and on the right, the plots correspond to the second investment period going from January 2, 1995 to June 30, 2004. "DJ" stands for a Dow Jones based portfolio and "Cash" for a risk-free asset based portfolio.

6 Concluding remarks

We generalized the work of [MS04a] to find the interval of local time homogeneity in a distribution-free and multidimensional context. We have then shown, using this procedure, how to treat the uncertainty in a class of stochastic optimization problems. The quality of the approximate problem we define to solve the stochastic optimization problem is theoretically controlled. The procedure has been tested on a portfolio selection model on both simulated and real data. In particular, on real data, one of the methods of calibration of the parameters of our adaptive algorithm has been shown to be competitive.

It remains to see how the adaptive estimations could be used to treat a broader class of stochastic optimization problems where the uncertainty is also in the constraints.

A Appendix

Proof of Proposition 2.1: For every $x \in X$, for every $\theta = (\rho_{N+1}^\top, \mathrm{svec}(Q_{N+1})^\top)^\top$ and every $\theta' = (\rho'^\top_{N+1}, \mathrm{svec}(Q'_{N+1})^\top)^\top$; the objective function f_0 of problem $\tilde{\mathcal{P}}(\theta)$ checks:

$$|f_0(x,\theta) - f_0(x,\theta')| \leq \kappa \left| \sqrt{x^\top Q_{N+1} x} - \sqrt{x^\top Q'_{N+1} x} \right| + |(\rho'_{N+1} - \rho_{N+1})^\top x|.$$

Notice that $|(\rho'_{N+1} - \rho_{N+1})^\top x| \leq \|\rho'_{N+1} - \rho_{N+1}\|_\infty \|x\|_1 \leq \|\theta - \theta'\|_\infty \|x\|_1$. If $Q'_{N+1} = Q_{N+1}$ we are done and else for any $\beta > 0$ and $x \neq 0$:

$$\begin{aligned}
\delta &= \left| \sqrt{x^\top Q_{N+1} x} - \sqrt{x^\top Q'_{N+1} x} \right| \\
&\leq \left| \frac{x^\top Q'_{N+1} x - x^\top Q_{N+1} x}{\beta} \right| k_\beta(x) + \beta \, 1\left(\max\left(\sqrt{x^\top Q'_{N+1} x},\ \sqrt{x^\top Q_{N+1} x} \right) < \beta \right),
\end{aligned}$$

where $k_\beta(x) = 1\left(\max(\sqrt{x^\top Q'_{N+1} x},\ \sqrt{x^\top Q_{N+1} x}\,) \geq \beta \right)$. We then choose $\beta = \sqrt{\|Q'_{N+1} - Q_{N+1}\|_\infty \|x\|_1}$ and obtain

$$\delta \leq \sqrt{\|Q'_{N+1} - Q_{N+1}\|_\infty \, \|x\|_1} \leq \sqrt{\|\theta' - \theta\|_\infty \, \|x\|_1}. \qquad \square$$

Proof of Proposition 2.2: Since $f_0(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\theta}}) \leq f_0(x^*, \hat{\boldsymbol{\theta}})$, we have:

$$\begin{aligned}
\epsilon(\mathcal{P}(\hat{\boldsymbol{\theta}})) &= |f_0(\hat{\boldsymbol{x}}, \theta) - f_0(x^*, \theta)| \\
&= f_0(\hat{\boldsymbol{x}}, \theta) - f_0(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\theta}}) + f_0(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\theta}}) - f_0(x^*, \hat{\boldsymbol{\theta}}) + f_0(x^*, \hat{\boldsymbol{\theta}}) - f_0(x^*, \theta) \\
&\leq 2 \sup_{x \in X} |f_0(x, \theta) - f_0(x, \hat{\boldsymbol{\theta}})|.
\end{aligned}$$

$\square$

Before showing the theorems of this paper, we need the three following Lemmas A.1, A.3 and A.4. The two last ones provide, for a process satisfying Assumption 3.1, new non-asymptotic bounds on the quality of the estimators of the mean and of the covariance

matrix given in Section 3.1 (Lemma A.4) and on the quality of two other close estimators (Lemma A.3). For short, we will sometimes write K_0 instead of $K_0(N)$.

Lemma A.1 *Let $X_t, t = 1, \ldots, N$, be N independent observations of a zero mean random vector in $\mathbb{R}^n$ with $n \geq 2$. If in addition, we have for every t : $\mathbb{E}\|X_t\|_\infty^2 \leq \sigma^2$, then:*

$$\mathbb{E}\Big\| \frac{1}{N} \sum_{t=1}^{N} X_t \Big\|_\infty \leq 2\sqrt{\frac{2}{\ln 2}}\sigma\sqrt{\frac{\ln n}{N}}. \tag{A.1}$$

Proof: Of the two integers $E\big[\frac{2\ln n}{\ln 2}\big]+1$ and $E\big[\frac{2\ln n}{\ln 2}\big]+2$, (where $E[x]$ is the integer part of x), let q be the one that is even and let $W(x) = \|x\|_q^2/2$. The proof is based on the following lemma.

Lemma A.2 *Let $n \geq 2$, and of the two integers $E\big[\frac{2\ln n}{\ln 2}\big]+1$ and $E\big[\frac{2\ln n}{\ln 2}\big]+2$, let q be the one that is even. Then the function*

$$W(x) = \frac{1}{2}\|x\|_q^2 : \mathbb{R}^n \to \mathbb{R}$$

satisfies for every $x, h \in \mathbb{R}^n$, the relation:

$$W(x+h) \leq W(x) + h^\top f(x) + c^*(n)\|h\|_\infty^2, \quad c^*(n) = \frac{4\ln n}{\ln 2}, \tag{A.2}$$

where $f : \mathbb{R}^n \to \mathbb{R}^n$ is defined by $f(x) = \nabla W(x)$ if $x \neq 0$ and $f(0) = 0$.

Proof of Lemma A.2: Let us fix $x, h \in \mathbb{R}^n$. We distinguish four cases for the pair $(x, x+h)$: first case: $(0, h)$; second case: $(x, 0)$; third case: 0 belongs to the open segment $]x, x+h[$; and fourth case: 0 does not belong to the segment $[x, x+h]$.

First notice that since q is even, $W(x) = \frac{1}{2}(\sum_{i=1}^n x_i^q)^{\frac{2}{q}}$ and for $x \neq 0$, $f_i(x) = \nabla_i W(x) = x_i^{q-1}\|x\|_q^{2-q}$.

In the first three cases, we can write x as $x = -kh$, with $k = 0$ in the first case, $k = 1$ in the second case, and $k < 1$ in the third case. In these three cases, we thus have

$$W(x+h) - W(x) - h^\top f(x) = \frac{1}{2}\|h\|_q^2. \tag{A.3}$$

Since $q \geq \frac{2\ln n}{\ln 2}$, we then obtain

$$\frac{1}{2}\|h\|_q^2 \leq \frac{1}{2}n^{\frac{2}{q}}\|h\|_\infty^2 \leq \|h\|_\infty^2 \leq \frac{4\ln n}{\ln 2}\|h\|_\infty^2. \tag{A.4}$$

Plugging (A.4) into (A.3) gives (A.2) in the first three cases.

In the fourth case, since W is continuously differentiable on $[x, x+h]$ and twice continuously differentiable on $]x, x+h[$, using Taylor formula, we have for some $z = x + \alpha h$, with $0 < \alpha < 1$,

$$W(x+h) = W(x) + h^\top \nabla W(x) + \frac{1}{2} h^\top \nabla^2 W(z) h. \tag{A.5}$$

Now for $x \neq 0$, we have $\nabla^2 W(x) = -(q-2)XX^\top + (q-1)\dfrac{\mathrm{diag}(x_i^{q-2})}{\|x\|_q^{q-2}}$, with $X = (X_1, \ldots, X_n)^\top$, where $X_i = \dfrac{x_i^{q-1}}{\|x\|_q^{q-1}}$. Observing that $q > 2$, and using (A.5) and Hölder's inequality, we then have

$$W(x+h) \;\leq\; W(x) + h^\top \nabla W(x) + \frac{q}{2}\,\|h\|_q^2 \leq W(x) + h^\top f(x) + q\|h\|_\infty^2.$$

We conclude bounding q from above by $\frac{4\ln n}{\ln 2}$, for $n \geq 2$. □

Let us now show (A.1). For $k = 0, \ldots, N-1$, we have, using Lemma A.2:

$$W\Big(\sum_{t=1}^{k+1} X_t\Big) \leq W\Big(\sum_{t=1}^{k} X_t\Big) + (X_{k+1})^\top f\Big(\sum_{t=1}^{k} X_t\Big) + c^*(n)\|X_{k+1}\|_\infty^2.$$

Hence, taking expectation, and since the random vectors X_t are independent and centered:

$$\mathbb{E}\Big[W\Big(\sum_{t=1}^{k+1} X_t\Big)\Big] \leq \mathbb{E}\Big[W\Big(\sum_{t=1}^{k} X_t\Big)\Big] + c^*(n)\,\sigma^2.$$

We then have: $\mathbb{E}\,[W(\sum_{t=1}^N X_t)\,] \leq N\,c^*(n)\,\sigma^2$, and since $W(z) \geq \frac{1}{2}\|z\|_\infty^2$,

$$\mathbb{E}\Big[\Big\|\sum_{t=1}^{N} X_t\Big\|_\infty^2\Big] \leq \frac{8}{\ln 2}\,\sigma^2\,N\,\ln n, \tag{A.6}$$

which achieves the proof of (A.1). □

Lemma A.3 *Let $\boldsymbol{r_t}$, $t = 1, \ldots, N$, be N independent observations of a random vector in $\mathbb{R}^n$ satisfying Assumption 3.1 and* (1.1) *where $\rho_t = \rho$ and $Q_t = Q$ are constant. Let us fix $\lambda > 0$ and let $K_0(N) = \sigma\Big(\frac{N}{\ln n(n+1) + \lambda \ln N}\Big)^{\frac{1}{4}}$. We suppose that N is sufficiently large to have $\ln(n(n+1)) + \lambda \ln N \leq N$. We define the following statistics: $\hat\rho_{N+1} = \frac{1}{N}\sum_{t=1}^N \alpha_t$,*

$$\hat Q_{N+1}^b = \frac{1}{N}\sum_{t=1}^{N} (\alpha_t - \rho)(\alpha_t - \rho)^\top, \text{ and } \hat Q_{N+1} = \frac{1}{N}\sum_{t=1}^{N} (\alpha_t - \hat\rho_{N+1})(\alpha_t - \hat\rho_{N+1})^\top,$$

with $\alpha_t = r_t\, 1(\|r_t\|_\infty \le K_0)$. *Then there exist constants* $m_1 = \frac{7}{3} + \sqrt{2}$, $m_2 = \frac{25}{3} + 4\sqrt{2}$, *and* $m_3 = m_1^2 + m_2$ *such that:*

$$\|\mathbb{E}\hat{\rho}_{N+1} - \rho\|_\infty \le \sigma\left(\frac{\ln n(n+1) + \lambda \ln N}{N}\right)^{\frac{3}{4}}, \tag{A.7}$$

$$\mathbb{P}\left(\|\hat{\rho}_{N+1} - \rho\|_\infty \ge m_1\sigma\sqrt{\frac{\ln n(n+1) + \lambda \ln N}{N}}\right) \le \frac{1}{N^\lambda}, \tag{A.8}$$

$$\|\mathbb{E}\hat{Q}^b_{N+1} - Q\|_\infty \le 3\sigma^2\sqrt{\frac{\ln n(n+1) + \lambda \ln N}{N}}, \tag{A.9}$$

$$\mathbb{P}\left(\|\hat{Q}^b_{N+1} - Q\|_\infty \ge m_2\sigma^2\sqrt{\frac{\ln n(n+1) + \lambda \ln N}{N}}\right) \le \frac{1}{N^\lambda}, \tag{A.10}$$

$$\mathbb{P}\left(\|\hat{Q}_{N+1} - Q\|_\infty \ge m_3\sigma^2\sqrt{\frac{\ln n(n+1) + \lambda \ln N}{N}}\right) \le \frac{2}{N^\lambda}. \tag{A.11}$$

Proof of Lemma A.3: Note first that for $k = 1, \ldots, n$ and $t = 1, \ldots, N$:

$$\begin{aligned} |\mathbb{E}\alpha_t(k) - \rho(k)| &= |\mathbb{E}\alpha_t(k) - \mathbb{E}r_t(k)| = |\mathbb{E}r_t(k)1(\|r_t\|_\infty > K_0)| \\ &\le \mathbb{E}\|r_t\|_\infty 1(\|r_t\|_\infty > K_0), \end{aligned}$$

and thus $|\mathbb{E}\hat{\rho}_{N+1}(k) - \rho(k)|$ is bounded above by $\frac{\sigma^4}{K_0^3}$. This shows (A.7). Now if $\delta_1 = \ln 2n + \lambda \ln N$, we show that:

$$\mathbb{P}\left(\|\hat{\rho}_{N+1} - \rho\|_\infty \ge \eta_1 = \frac{\sigma^4}{K_0^3} + \eta_1'\right) \le \frac{1}{N^\lambda},$$

with

$$\eta_1' = \sqrt{2}\sigma\sqrt{\frac{\delta_1}{N}} + \frac{4}{3}\frac{K_0\delta_1}{N},$$

and thus (A.8) will follow with $m_1 = \frac{7}{3} + \sqrt{2}$. We can bound from above $\mathbb{P}(\|\hat{\rho}_{N+1} - \rho\|_\infty \ge \eta_1)$ by

$$n \max_{k=1,\ldots,n} \mathbb{P}(|\hat{\rho}_{N+1}(k) - \mathbb{E}\hat{\rho}_{N+1}(k)| \ge \eta_1 - |\mathbb{E}\hat{\rho}_{N+1}(k) - \rho(k)|),$$

which itself is bounded from above by

$$n \max_{k=1,\ldots,n} \mathbb{P}\left(\Big|\sum_{i=1}^N \alpha_i(k) - \mathbb{E}\alpha_i(k)\Big| \ge N\eta_1'\right). \tag{A.12}$$

Further, notice that

$$\eta_1' \ge \eta_1'' = \frac{2}{3}\frac{K_0\delta_1}{N} + \sqrt{\Big(\frac{2}{3}\frac{K_0\delta_1}{N}\Big)^2 + 2\sigma^2\frac{\delta_1}{N}}.$$

Now, if $X_k^i = \alpha_i(k) - \mathbb{E}\alpha_i(k)$, for all fixed k, the random variables X_k^i, $i = 1, \ldots, N$, are independent with zero mean and such that

$$\mathrm{Var}(X_k^i) = \mathbb{E}(X_k^i)^2 \leq \mathbb{E}\alpha_i(k)^2 \leq \sigma^2 \text{ and } |X_k^i| \leq 2\,K_0.$$

Thus, when using the Bernstein inequality for (A.12), we get (A.8):

$$\mathbb{P}(\|\hat{\rho}_{N+1} - \rho\|_\infty \geq \eta_1) \leq 2n \exp\left(-\frac{1}{2}\frac{N\eta_1''^2}{\sigma^2 + \frac{2}{3}K_0\eta_1''}\right) = \frac{1}{N^\lambda}.$$

Now let $\tilde{\alpha}_t = \mathbb{E}(\alpha_t - \rho)(\alpha_t - \rho)^\top - Q$. We have, for all $1 \leq j, k \leq n$, and $t = 1, \ldots, N$:

$$\begin{aligned} |\tilde{\alpha}_t(j,k)| &\leq \mathbb{E}|r_t(j)r_t(k) - \alpha_t(j)\alpha_t(k)| + |\rho(k)||\mathbb{E}\alpha_t(j) - \rho(j)| \\ &\quad + |\rho(j)||\mathbb{E}\alpha_t(k) - \rho(k)| \qquad (A.13) \\ &\leq \mathbb{E}|r_t(j)r_t(k)1(\|r_t\|_\infty > K_0)| + |\rho(k)||\mathbb{E}r_t(j)1(\|r_t\|_\infty > K_0)| \\ &\quad + |\rho(j)|\,|\mathbb{E}r_t(k)1(\|r_t\|_\infty > K_0)| \leq \frac{\sigma^4}{K_0^2} + 2\frac{\sigma^5}{K_0^3}; \end{aligned}$$

and (A.9) follows. We now introduce $\delta_2 = \ln n(n+1) + \lambda \ln N$ and show that:

$$\mathbb{P}\left(\|\hat{Q}_{N+1}^b - Q\|_\infty \geq \eta_2' = \frac{\sigma^4}{K_0^2} + 2\frac{\sigma^5}{K_0^3} + \eta_2''\right) \leq \frac{1}{N^\lambda},$$

where

$$\eta_2''(N,\lambda) = \frac{4}{3}\frac{(K_0+\sigma)^2\delta_2}{N} + 4\sqrt{2}\sigma^2\sqrt{\frac{\delta_2}{N}}; \qquad (A.14)$$

which will prove (A.10). Reasoning as above and, for short, writing η_2'' for $\eta_2''(N,\lambda)$, we have

$$\begin{aligned} \mathbb{P}(\|\hat{Q}_{N+1}^b - Q\|_\infty \geq \eta_2') &\leq \frac{n(n+1)}{2}\max_{1\leq j\leq k\leq n}\mathbb{P}(|\hat{Q}_{N+1}^b(j,k) - \mathbb{E}\hat{Q}_{N+1}^b(j,k)| \geq \eta_2'') \\ &\leq \frac{n(n+1)}{2}\max_{1\leq j\leq k\leq n}\mathbb{P}\left(\left|\sum_{i=1}^N A_i^{j,k} - \mathbb{E}A_i^{j,k}\right| \geq N\tilde{\eta}\right) \end{aligned}$$

with

$$\tilde{\eta} = \frac{2}{3}\frac{(K_0+\sigma)^2\delta_2}{N} + \sqrt{\left(\frac{2}{3}\frac{(K_0+\sigma)^2\delta_2}{N}\right)^2 + 32\sigma^4\frac{\delta_2}{N}} \leq \eta_2'',$$

and $A_i^{j,k} = (\alpha_i(j) - \rho(j))(\alpha_i(k) - \rho(k))$. If we set $X_i^{j,k} = A_i^{j,k} - \mathbb{E}A_i^{j,k}$, for every (j,k), the random variables $(X_i^{j,k})_i$ are independent, with zero mean, and satisfy

$$|X_i^{j,k}| \leq 2(K_0+\sigma)^2, \text{ and } \mathrm{Var}(X_i^{j,k}) = \mathbb{E}(X_i^{j,k})^2 \leq \mathbb{E}(\|\alpha_i\|_\infty + \sigma)^4 \leq 16\sigma^4.$$

We then conclude the proof of (A.10) using Bernstein inequality. Further, since:

$$\|\hat{Q}^b_{N+1} - \hat{Q}_{N+1}\|_\infty = \|(\rho - \hat{\rho}_{N+1})(\rho - \hat{\rho}_{N+1})^\top\|_\infty = \|\rho - \hat{\rho}_{N+1}\|^2_\infty,$$

if $\eta_2 = m_3\sigma^2\sqrt{\frac{\delta_2}{N}}$, we have:

$$\begin{aligned}&\mathbb{P}(\|\hat{Q}_{N+1} - Q\|_\infty \geq \eta_2)\\ &\leq \mathbb{P}\left(\|\hat{\rho}_{N+1} - \rho\|^2_\infty \geq m_1^2\sigma^2\sqrt{\frac{\delta_2}{N}}\right) + \mathbb{P}\left(\|\hat{Q}^b_{N+1} - Q\|_\infty \geq m_2\sigma^2\sqrt{\frac{\delta_2}{N}}\right).\end{aligned}$$

Since $\sqrt{\frac{\delta_2}{N}} \geq \frac{\delta_2}{N}$, (A.8) allows us to bound from above the first term of the right hand side by $\frac{1}{N^\lambda}$ and from (A.10) the second term is also bounded above by $\frac{1}{N^\lambda}$. □

Lemma A.4 *Under the hypotheses of Lemma A.3, we define the following statistics:* $\hat{\rho}_{N+1} = \frac{1}{N}\sum_{t=1}^N \alpha_t$,

$$\hat{Q}^b_{N+1} = \frac{1}{N}\sum_{t=1}^N (\alpha_t - \rho)(\alpha_t - \rho)^\top, \text{ and } \hat{Q}_{N+1} = \frac{1}{N}\sum_{t=1}^N (\alpha_t - \hat{\rho}_{N+1})(\alpha_t - \hat{\rho}_{N+1})^\top,$$

where for $i = 1, \ldots, n$, $\alpha_t(i) = [r_t(i)]_{K_0(N)}$, *with* $[\cdot]_{K_0(N)}$ *a truncation operator (defined in* (3.1)*). The estimators* $\hat{\rho}_{N+1}$, $\hat{Q}^b_{N+1}$ *and* $\hat{Q}_{N+1}$ *satisfy* (A.7), (A.8), (A.10), (A.11) *(with* $m_1 = \frac{7}{3} + \sqrt{2}$, $m_2 = \frac{31}{3} + 4\sqrt{2}$, $m_3 = m_1^2 + m_2$*) and*

$$\|\mathbb{E}\hat{Q}^b_{N+1} - Q\|_\infty \leq 5\sigma^2\sqrt{\frac{\ln n(n+1) + \lambda \ln N}{N}}. \tag{A.15}$$

The condition $\ln(n(n+1)) + \lambda \ln N \leq N$ in the above Lemmas A.3 and A.4 can be suppressed but this leads to more complicated left hand sides.

Proof of Lemma A.4: If we notice that $|\alpha_t(i)| \leq |r_t(i)|$, we see, (following the proof of Lemma A.3), that (A.7), (A.8), (A.10) and (A.11) remain valid for Lemma A.4 provided that constants m_2 and m_3 are updated. For instance for $k = 1, \ldots, n$, and $t = 1, \ldots, N$,

$$\begin{aligned}|\mathbb{E}\alpha_t(k) - \rho(k)| &= |\mathbb{E}[K_0\frac{r_t(k)}{|r_t(k)|} - r_t(k)]1(|r_t(k)| > K_0)|\\ &\leq \mathbb{E}|r_t(k)|1(|r_t(k)| > K_0) \leq \frac{\sigma^4}{K_0^3},\end{aligned}$$

and (A.7) follows. Let us now bound $\|\mathbb{E}\hat{Q}^b_{N+1} - Q\|_\infty$ from above. Let us fix j, k in $1, \ldots, n$, and t in $1, \ldots, N$. First, note that (A.13) holds. Let us then denote by k^+ (resp. k^-) the quantity $1(|r_1(k)| > K_0)$ (resp. $1(|r_1(k)| \leq K_0)$) and by j^+ (resp. j^-) the quantity $1(|r_1(j)| > K_0)$ (resp. $1(|r_1(j)| \leq K_0)$). Further, we have

$$|\rho(k)||\mathbb{E}r_t(j) - \alpha_t(j)| \leq \sigma\mathbb{E}|r_t(j)|j^+ \leq \frac{\sigma^4}{K_0^2}.$$

We then bound from above $\mathbb{E}|\alpha_t(j)\alpha_t(k) - r_t(j)r_t(k)|$ by the sum of three terms $A_t(j,k)$, $B_t(j,k)$, and $C_t(j,k)$ defined by

$$\begin{aligned} A_t(j,k) &= \mathbb{E}|r_t(k)||r_t(j)|\Big|\frac{K_0}{|r_t(j)|} - 1\Big|k^- j^+, \\ B_t(j,k) &= \mathbb{E}|r_t(k)||r_t(j)|\Big|\frac{K_0}{|r_t(k)|} - 1\Big|k^+ j^-, \\ C_t(j,k) &= \mathbb{E}|r_t(j)r_t(k)|\Big|\frac{K_0^2}{|r_t(j)r_t(k)|} - 1\Big|k^+ j^+. \end{aligned}$$

Since each of these terms is bounded from above by $\frac{\sigma^4}{K_0^2}$, (A.15) follows. □

The following lemma will also be useful. Even if V_I^ρ and V_I^Q are unknown, this lemma allows us to have a (non-asymptotic) bound on these quantities: $V_I^\rho \leq 4\sqrt{\frac{2}{\ln 2}}\sigma\sqrt{\frac{\ln n}{|I|}}$ and $V_I^Q \leq k_Q\sigma^2\sqrt{\frac{\ln n(n+1)+\lambda\ln|I|}{|I|}}$, for some constant k_Q. If I is an interval of local time homogeneity, this lemma will also provide an upper bound for Δ_I^ρ and Δ_I^Q. In what follows, we denote $\frac{1}{|I|}\sum_{t\in I}\rho_t$ and $\frac{1}{|I|}\sum_{t\in I}Q_t$ by respectively $\bar{\rho}_I$ and $\bar{Q}_I$. For short, we will write α_t^I for α_t.

Lemma A.5 *Let $\lambda > 0$, let $n \geq 2$, let I be a nonempty interval such that $\ln n(n+1) + \lambda\ln|I| \leq |I|$ and let $\hat{\rho}_I$ and $\hat{Q}_I$ be the estimators defined in* (3.3)*. Then there is a constant $k_Q = 2 + \frac{64}{\ln 2} + \frac{16}{\sqrt{\ln 2}}(2+\sqrt{2})$ such that*

$$\mathbb{E}\|\hat{\rho}_I - \mathbb{E}\hat{\rho}_I\|_\infty^2 \leq \frac{32}{\ln 2}\sigma^2\left(\frac{\ln n}{|I|}\right), \tag{A.16}$$

$$\mathbb{E}\|\hat{Q}_I - \mathbb{E}\hat{Q}_I\|_\infty \leq k_Q\sigma^2\sqrt{\frac{\ln n(n+1)+\lambda\ln|I|}{|I|}}. \tag{A.17}$$

Proof: Notice that random vectors $X_k = \alpha_k - \mathbb{E}\alpha_k$, for $k \in I$, are independent with zero mean and check:

$$\mathbb{E}\|X_k\|_\infty^2 \leq \mathbb{E}(\|\alpha_k\|_\infty + \mathbb{E}\|\alpha_k\|_\infty)^2 \leq \mathbb{E}(\|r_k\|_\infty + \sigma)^2 \leq 4\sigma^2.$$

It then suffices to follow the proof of Lemma A.1 to show (A.16). Now, let $\tilde{Q}_I^b = \frac{1}{|I|}\sum_{t\in I}(\alpha_t - \bar{\rho}_I)(\alpha_t - \bar{\rho}_I)^\top$. We first prove that:

$$\mathbb{E}\|\tilde{Q}_I^b - \mathbb{E}\tilde{Q}_I^b\|_\infty \leq \frac{32}{\sqrt{\ln 2}}\sigma^2\sqrt{\frac{\ln n}{|I|}}. \tag{A.18}$$

Indeed, $\mathbb{E}\|\tilde{Q}_I^b - \mathbb{E}\tilde{Q}_I^b\|_\infty = \frac{1}{|I|}\,\mathbb{E}\|\sum_{t\in I}\tilde{\zeta}_t\|_\infty$, where $\tilde{\zeta}_t$ is the symmetric vectorisation of $(\alpha_t - \bar{\rho}_I)(\alpha_t - \bar{\rho}_I)^\top - \mathbb{E}(\alpha_t - \bar{\rho}_I)(\alpha_t - \bar{\rho}_I)^\top$. The random vectors $(\tilde{\zeta}_t)_{t\in I}$ in $\mathbb{R}^{\frac{n(n+1)}{2}}$ are independent with zero mean and $\mathbb{E}\|\tilde{\zeta}_t\|_\infty^2 \le \mathbb{E}((\|r_t\|_\infty + \sigma)^2 + 4\sigma^2)^2 \le 64\sigma^4$. Using Lemma A.1, we get (A.18). Now,

$$\begin{aligned}\mathbb{E}\|\hat{Q}_I - \mathbb{E}\hat{Q}_I\|_\infty &\le \mathbb{E}\|\hat{Q}_I - \tilde{Q}_I^b\|_\infty + \mathbb{E}\|\tilde{Q}_I^b - \mathbb{E}\tilde{Q}_I^b\|_\infty + \|\mathbb{E}\tilde{Q}_I^b - \mathbb{E}\hat{Q}_I\|_\infty \\ &\le 2\mathbb{E}\|\hat{Q}_I - \tilde{Q}_I^b\|_\infty + \frac{32}{\sqrt{\ln 2}}\sigma^2\sqrt{\frac{\ln n}{|I|}}. \qquad \text{(A.19)}\end{aligned}$$

Reasoning as in the proof of (A.7), we then have $\|\mathbb{E}\hat{\rho}_I - \bar{\rho}_I\|_\infty \le \sigma\sqrt{\frac{\ln n(n+1)+\lambda\ln|I|}{|I|}}$. Also, $\|\hat{Q}_I - \tilde{Q}_I^b\|_\infty = \|\hat{\rho}_I - \bar{\rho}_I\|_\infty^2$ and it follows that

$$\begin{aligned}\mathbb{E}\|\hat{\rho}_I - \bar{\rho}_I\|_\infty^2 &\le \mathbb{E}(\|\hat{\rho}_I - \mathbb{E}\hat{\rho}_I\|_\infty + \|\mathbb{E}\hat{\rho}_I - \bar{\rho}_I\|_\infty)^2 \\ &\le \mathbb{E}\left(\|\hat{\rho}_I - \mathbb{E}\hat{\rho}_I\|_\infty + \sigma\sqrt{\frac{\ln n(n+1) + \lambda\ln|I|}{|I|}}\right)^2. \qquad \text{(A.20)}\end{aligned}$$

We then use (A.16), (A.19) and (A.20) to prove (A.17). □

Proof of Theorem 3.2: Let r_t satisfy Assumption 3.1. We have:

$$\|\hat{\rho}_I - \hat{\rho}_J\|_\infty \le \|\hat{\rho}_I - \bar{\rho}_I\|_\infty + \|\bar{\rho}_I - \bar{\rho}_J\|_\infty + \|\bar{\rho}_J - \hat{\rho}_J\|_\infty. \qquad \text{(A.21)}$$

Now $\|\bar{\rho}_I - \bar{\rho}_J\|_\infty \le \|\bar{\rho}_I - \rho_{N+1}\|_\infty + \|\rho_{N+1} - \bar{\rho}_J\|_\infty$ and from the Cauchy–Schwartz inequality $\|\bar{\rho}_I - \rho_{N+1}\|_\infty \le \Delta_I^\rho$ and $\|\rho_{N+1} - \bar{\rho}_J\|_\infty \le \Delta_J^\rho$. Since I is an interval of local time homogeneity and $J \in \mathcal{I}(I)$, we have $\Delta_I^\rho \le DV_I^\rho$ and $\Delta_J^\rho \le DV_J^\rho$. Using Lemma A.5, we then get:

$$\|\bar{\rho}_I - \bar{\rho}_J\|_\infty \le 4\sqrt{\frac{2}{\ln 2}}D\sigma\left(\sqrt{\frac{\ln n}{|I|}} + \sqrt{\frac{\ln n}{|J|}}\right). \qquad \text{(A.22)}$$

We can then easily adapt the proof of (A.8) to show that for every nonempty interval I and $\lambda > 0$:

$$\mathbb{P}(\|\hat{\rho}_I - \bar{\rho}_I\|_\infty \ge \eta_1(|I|,\lambda)) \le \frac{1}{|I|^\lambda}, \qquad \text{(A.23)}$$

with

$$\eta_1(|I|,\lambda) = \frac{\sigma^4}{K_0^3(|I|)} + \sqrt{2}\sigma\sqrt{\frac{\ln 2n + \lambda\ln|I|}{|I|}} + \frac{4}{3}\frac{K_0(|I|)(\ln 2n + \lambda\ln|I|)}{|I|}, \qquad \text{(A.24)}$$

where $K_0(|I|) = \sigma\left(\frac{|I|}{\ln n(n+1)+\lambda\ln|I|}\right)^{\frac{1}{4}}$. Note that if $\ln n(n+1)+\lambda\ln|I| \leq |I|$, then (A.23) holds with $\eta_1(|I|,\lambda) = m_1\sigma\sqrt{\frac{\ln n(n+1)+\lambda\ln|I|}{|I|}}$, where $m_1 = \frac{7}{3}+\sqrt{2}$. Since $\ln n(n+1)+\lambda\ln|I| \leq |I|$ and $\ln n(n+1)+\lambda\ln|J| \leq |J|$, we then use (A.21), (A.22) and (A.23) with $\eta_1(|I|,\lambda) = (\frac{7}{3}+\sqrt{2})\sigma\sqrt{\frac{\ln n(n+1)+\lambda\ln|I|}{|I|}}$ to obtain (3.6). We now need the following lemma to prove (3.7).

Lemma A.6 *If I is a nonempty interval of local time homogeneity, then for all $\lambda > 0$,*

$$\mathbb{P}\Bigg(\|\hat{Q}_I - \bar{Q}_I\|_\infty \geq \eta(|I|,\lambda) \qquad (A.25)$$
$$= \eta_1^2(|I|,\lambda) + \eta_2''(|I|,\lambda) + k\sigma^2\sqrt{\frac{\ln n(n+1)+\lambda\ln|I|}{|I|}}\Bigg) \leq \frac{2}{|I|^\lambda},$$

where $k = 5 + 8\sqrt{\frac{2}{\ln 2}}D$ and the functions η_1 and η_2'' are defined in (A.24) *and* (A.14)*.*

Proof of Lemma A.6: Note first that:

$$\|\hat{Q}_I - \bar{Q}_I\|_\infty \leq \|\hat{\rho}_I - \bar{\rho}_I\|_\infty^2 + \frac{1}{|I|}\Big\|\sum_{t\in I}(\alpha_t - \bar{\rho}_I)(\alpha_t - \bar{\rho}_I)^\top - Q_t\Big\|_\infty. \qquad (A.26)$$

We now have to bound from above: $\|\Delta\|_\infty = \frac{1}{|I|}\Big\|\sum_{t\in I}\mathbb{E}(\alpha_t - \bar{\rho}_I)(\alpha_t - \bar{\rho}_I)^\top - Q_t\Big\|_\infty$. We have, for all $1 \leq j,k \leq n$:

$$\begin{aligned}
|\Delta(j,k)| \leq\ & \frac{1}{|I|}\sum_{t\in I}|\mathbb{E}[\alpha_t(j)\alpha_t(k) - r_t(j)r_t(k)]| + \frac{|\bar{\rho}_I(k)|}{|I|}\Big|\sum_{t\in I}\rho_t(j) - \mathbb{E}\alpha_t(j)\Big| \\
& + \frac{1}{|I|}\sum_{t\in I}|\rho_t(j)||\rho_t(k) - \mathbb{E}\alpha_t(k)| + \frac{1}{|I|}\sum_{t\in I}|\mathbb{E}\alpha_t(k)||\rho_t(j) - \bar{\rho}_I(j)| \\
\leq\ & \frac{5\sigma^4}{K_0^2(|I|)} + \frac{\sigma}{|I|}\left(\sum_{t\in I}|(\rho_t - \rho_{N+1})(j)| + |(\rho_{N+1} - \bar{\rho}_I)(j)|\right) \\
\leq\ & \frac{5\sigma^4}{K_0^2(|I|)} + 2\sigma\Delta_I^\rho \leq \left(5 + 8\sqrt{\frac{2}{\ln 2}}D\right)\sigma^2\sqrt{\frac{\ln n(n+1)+\lambda\ln|I|}{|I|}}.
\end{aligned}$$

We then use (A.26), (A.23) and follow the proof of (A.10) to conclude. □

Remark A.7 In the case when $\ln n(n+1)+\lambda\ln|I| \leq |I|$ and if $k'_Q = \frac{160}{9} + \frac{26}{3}\sqrt{2} + 8\sqrt{\frac{2}{\ln 2}}D$, then the above lemma holds with $\eta(|I|,\lambda) = k'_Q\,\sigma^2\sqrt{\frac{\ln n(n+1)+\lambda\ln|I|}{|I|}}$.

Similarly, we have:

$$\|\hat{Q}_I - \hat{Q}_J\|_\infty \leq \|\hat{Q}_I - \bar{Q}_I\|_\infty + \|\bar{Q}_I - \bar{Q}_J\|_\infty + \|\bar{Q}_J - \hat{Q}_J\|_\infty. \tag{A.27}$$

Using Cauchy–Schwartz inequality, Lemma A.5 and since I is of local time homogeneity we get:

$$\begin{aligned}\|\bar{Q}_I - \bar{Q}_J\|_\infty &\leq D(V_I^Q + V_J^Q) \qquad (A.28)\\ &\leq k_Q D\sigma^2\left(\sqrt{\frac{\ln n(n+1) + \lambda \ln|J|}{|J|}} + \sqrt{\frac{\ln n(n+1) + \lambda \ln|I|}{|I|}}\right).\end{aligned}$$

We conclude using (A.27), (A.29) and Remark A.7. □

Proof of Theorem 3.4: We show that (3.10) holds with $k(D) = k'_Q + k_Q D$, where k'_Q and k_Q are defined in Remark A.7 and Lemma A.5. If $\delta_2 = \ln n(n+1) + \lambda \ln|\mathbb{I}|$ and $k'(D) = \frac{7}{3} + \sqrt{2} + 4\sqrt{\frac{2}{\ln 2}}D$, since $k(D)\max(\sigma, \sigma^2) \geq k'(D)\sigma$, we can bound from above $\mathbb{P}\left(\|\hat{\theta}_\mathbb{I} - \theta\|_\infty \geq k(D)\max(\sigma, \sigma^2)\sqrt{\frac{\delta_2}{|\mathbb{I}|}}\right)$ by

$$\mathbb{P}\left(\|\hat{\rho}_\mathbb{I} - \rho_{N+1}\|_\infty \geq k'(D)\sigma\sqrt{\frac{\delta_2}{|\mathbb{I}|}}\right) + \mathbb{P}\left(\|\hat{Q}_\mathbb{I} - Q_{N+1}\|_\infty \geq k(D)\sigma^2\sqrt{\frac{\delta_2}{|\mathbb{I}|}}\right). \tag{A.29}$$

We then observe that

$$\|\hat{\rho}_\mathbb{I} - \rho_{N+1}\|_\infty \leq \frac{1}{|\mathbb{I}|}\sum_{t\in\mathbb{I}} \|\alpha_t - \rho_t\|_\infty + \frac{1}{|\mathbb{I}|}\sum_{t\in\mathbb{I}} \|\rho_t - \rho_{N+1}\|_\infty. \tag{A.30}$$

Then using Cauchy–Schwartz inequality, the definition of $\mathbb{I}$ and Lemma A.5:

$$\frac{1}{|\mathbb{I}|}\sum_{t\in\mathbb{I}} \|\rho_t - \rho_{N+1}\|_\infty \leq \Delta_\mathbb{I}^\rho \leq DV_\mathbb{I}^\rho \leq 4\sqrt{\frac{2}{\ln 2}}D\sigma\sqrt{\frac{\ln n}{|\mathbb{I}|}}. \tag{A.31}$$

Using (A.30), (A.31) and (A.23) with $\eta_1(|\mathbb{I}|, \lambda) = \left(\frac{7}{3} + \sqrt{2}\right)\sigma\sqrt{\frac{\delta_2}{|\mathbb{I}|}}$, we can bound from above the first term in (A.29) by $\frac{1}{|\mathbb{I}|^\lambda}$. Similarly for the covariance matrix:

$$\|\hat{Q}_\mathbb{I} - Q_{N+1}\|_\infty \leq \|\hat{Q}_\mathbb{I} - \bar{Q}_\mathbb{I}\|_\infty + \Delta_\mathbb{I}^Q, \tag{A.32}$$

with $\Delta_\mathbb{I}^Q \leq DV_\mathbb{I}^Q$. We then use (A.32) and Lemmas A.5 and A.6 to bound from above the second term in (A.29) by $\frac{2}{|\mathbb{I}|^\lambda}$. □

Proof of Theorem 3.5: We show that (3.11) is valid with $k(D) = 3(k_Q D + k'_Q)$ where k'_Q and k_Q are defined in Remark A.7 and Lemma A.5. To this end, we prove that the

union of the event:

$$\left\{ \|\hat{\rho}_{\hat{I}} - \rho_{N+1}\|_\infty > 3\eta_1(|\mathbb{I}|,\lambda) + 8\sqrt{\frac{2}{\ln 2}} D\sigma \sqrt{\frac{\ln n}{|\mathbb{I}|}} + \Delta_{\mathbb{I}}^{\rho} \right\} \tag{A.33}$$

and of the event

$$\left\{ \|\hat{Q}_{\hat{I}} - Q_{N+1}\|_\infty > 3\eta(|\mathbb{I}|,\lambda) + 2k_Q D\sigma^2 \sqrt{\frac{\ln n(n+1) + \lambda \ln |\mathbb{I}|}{|\mathbb{I}|}} + \Delta_{\mathbb{I}}^{Q} \right\} \tag{A.34}$$

implies the event

$$\bigcup_{I\in\mathcal{I} \mid I\subseteq\mathbb{I}} \bigcup_{J\in\mathcal{I}_+(I)} \left\{ \|\hat{\rho}_J - \bar{\rho}_J\|_\infty > \eta_1(|J|,\lambda) \ \cup \ \|\hat{Q}_J - \bar{Q}_J\|_\infty > \eta(|J|,\lambda) \right\}, \tag{A.35}$$

where $\eta(|I|,\lambda) = k'_Q\,\sigma^2 \sqrt{\frac{\ln n(n+1)+\lambda \ln |I|}{|I|}}$ and $\eta_1(|I|,\lambda) = m_1\sigma\sqrt{\frac{\ln n(n+1)+\lambda \ln |I|}{|I|}}$ with $m_1 = \frac{7}{3} + \sqrt{2}$. Since we easily check that the probability

$$\mathbb{P}\left(\|\hat{\theta}_{\hat{I}} - \theta\|_\infty \geq k(D) \max(\sigma,\sigma^2) \sqrt{\frac{\ln n(n+1) + \lambda \ln |\mathbb{I}|}{|\mathbb{I}|}} \right)$$

is bounded above by the probability of the union of the two events (A.33) and (A.34), and since every testing subinterval J satisfies $|J| \geq m_0$, this will prove the theorem.

Let us thus now prove that the union of the events (A.33) and (A.34) implies the event (A.35). Let us suppose that for all I in $\mathcal{I}$ such that $I \subseteq \mathbb{I}$ and $J \in \mathcal{I}_+(I)$, $\|\hat{\rho}_J - \bar{\rho}_J\|_\infty \leq \eta_1(|J|,\lambda)$ and $\|\hat{Q}_J - \bar{Q}_J\|_\infty \leq \eta(|J|,\lambda)$. We intend to prove that $\|\hat{\rho}_{\hat{I}} - \rho_{N+1}\|_\infty \leq 3\eta_1(|\mathbb{I}|,\lambda) + 8\sqrt{\frac{2}{\ln 2}} D\sigma \sqrt{\frac{\ln n}{|\mathbb{I}|}} + \Delta_{\mathbb{I}}^{\rho}$ and $\|\hat{Q}_{\hat{I}} - Q_{N+1}\|_\infty \leq 3\eta(|\mathbb{I}|,\lambda) + 2k_Q D\sigma^2 \sqrt{\frac{\ln n(n+1)+\lambda \ln |\mathbb{I}|}{|\mathbb{I}|}} + \Delta_{\mathbb{I}}^{Q}$. First, note that $\mathbb{I}$ is not rejected. Indeed, for all $I \in \mathcal{I}$ such that $I \subseteq \mathbb{I}$ and for all $J \in \mathcal{I}(I)$:

$$\begin{aligned} \|\hat{\rho}_I - \hat{\rho}_J\|_\infty &\leq \|\hat{\rho}_I - \bar{\rho}_I\|_\infty + \|\bar{\rho}_I - \rho_{N+1}\|_\infty + \|\rho_{N+1} - \bar{\rho}_J\|_\infty + \|\bar{\rho}_J - \hat{\rho}_J\|_\infty \\ &\leq \eta_1(|I|,\lambda) + \eta_1(|J|,\lambda) + \Delta_I^{\rho} + \Delta_J^{\rho}. \end{aligned}$$

Now due to the definition of $\mathbb{I}$, $\Delta_I^{\rho} \leq DV_I^{\rho}$, $\Delta_J^{\rho} \leq DV_J^{\rho}$ and using Lemma A.5 gives:

$$\|\hat{\rho}_I - \hat{\rho}_J\|_\infty \leq \eta_1(|I|,\lambda) + \eta_1(|J|,\lambda) + 4\sqrt{\frac{2}{\ln 2}} D\sigma \left(\sqrt{\frac{\ln n}{|I|}} + \sqrt{\frac{\ln n}{|J|}} \right).$$

Similarly we show that:

$$\begin{aligned} \|\hat{Q}_I - \hat{Q}_J\|_\infty \leq\ & \eta(|I|,\lambda) + \eta(|J|,\lambda) \\ & + k_Q D\sigma^2 \left(\sqrt{\frac{\ln n(n+1) + \lambda \ln |I|}{|I|}} + \sqrt{\frac{\ln n(n+1) + \lambda \ln |J|}{|J|}} \right). \end{aligned}$$

So for all $I \in \mathcal{I}$ such that $I \subseteq \mathbb{I}$, I is accepted and $\mathbb{I}$ is accepted so $\mathbb{I} \subseteq \hat{I}$. This implies

$$\begin{aligned}\|\hat{\rho}_{\mathbb{I}} - \hat{\rho}_{\hat{I}}\|_\infty &\leq \eta_1(|\mathbb{I}|, \lambda) + \eta_1(|\hat{I}|, \lambda) + 4\sqrt{\frac{2}{\ln 2}} D\sigma \left(\sqrt{\frac{\ln n}{|\mathbb{I}|}} + \sqrt{\frac{\ln n}{|\hat{I}|}}\right) \\ &\leq 2\eta_1(|\mathbb{I}|, \lambda) + 8\sqrt{\frac{2}{\ln 2}} D\sigma \sqrt{\frac{\ln n}{|\mathbb{I}|}},\end{aligned}$$

since $\eta_1(|I|, \lambda)$ is a decreasing function of $|I|$. Now

$$\begin{aligned}\|\hat{\rho}_{\hat{I}} - \rho_{N+1}\|_\infty &\leq \|\hat{\rho}_{\hat{I}} - \hat{\rho}_{\mathbb{I}}\|_\infty + \|\hat{\rho}_{\mathbb{I}} - \bar{\rho}_{\mathbb{I}}\|_\infty + \|\bar{\rho}_{\mathbb{I}} - \rho_{N+1}\|_\infty \\ &\leq 3\eta_1(|\mathbb{I}|, \lambda) + 8\sqrt{\frac{2}{\ln 2}} D\sigma \sqrt{\frac{\ln n}{|\mathbb{I}|}} + \Delta_{\mathbb{I}}^{\rho}.\end{aligned}$$

Since $\eta(|I|, \lambda)$ is also a decreasing function of $|I|$, we can show in the same fashion that

$$\|\hat{Q}_{\hat{I}} - Q_{N+1}\|_\infty \leq 3\eta(|\mathbb{I}|, \lambda) + 2k_Q D\sigma^2 \sqrt{\frac{\ln n(n+1) + \lambda \ln |\mathbb{I}|}{|\mathbb{I}|}} + \Delta_{\mathbb{I}}^{Q},$$

which achieves the proof. □

Proof of Theorems 4.2–4.5: In the stationary case, let $\lambda > 0$ be such that $\ln n(n+1) + \lambda \ln N \leq N$ and in the case of slowly varying parameters let λ be the parameter of the adaptive algorithm such that $\ln n(n+1) + \lambda \ln m_0 \leq m_0$. In the stationary case (resp. in the case of slowly varying parameters), on the basis of Lemma A.4 (resp. following the proof of Theorem 3.5) we can find a random set S of probability at least $1 - \frac{3}{N^\lambda}$ $\left(\text{resp. at least } 1 - \sum_{I \in \mathcal{I} \mid I \subseteq \mathbb{I}} \sum_{J \in \mathcal{I}_+(I)} \frac{3}{|J|^\lambda}\right)$ and functions η_ρ and η_Q depending on λ, σ, n and N (resp. λ, σ, n and $|\mathbb{I}|$) such that if $\omega \in S$: $\|\hat{\boldsymbol{\rho}}_{\boldsymbol{N+1}} - \rho_{N+1}\|_\infty \leq \eta_\rho$ and $\|\hat{\boldsymbol{Q}}_{\boldsymbol{N+1}} - Q_{N+1}\|_\infty \leq \eta_Q$. More precisely, in the stationary case we have $\eta_\rho = \left(\frac{7}{3} + \sqrt{2}\right)\sigma\sqrt{\frac{\ln n(n+1) + \lambda \ln N}{N}}$ and $\eta_Q = \left(\frac{160}{9} + \frac{26}{3}\sqrt{2}\right)\sigma^2\sqrt{\frac{\ln n(n+1) + \lambda \ln N}{N}}$, and in the case of slowly varying parameters, $\eta_\rho = \left(3\sqrt{2} + 7 + 12\sqrt{\frac{2}{\ln 2}} D\right)\sigma\sqrt{\frac{\ln n(n+1) + \lambda \ln |\mathbb{I}|}{|\mathbb{I}|}}$ and $\eta_Q = 3(k_Q D + k'_Q)\sigma^2\sqrt{\frac{\ln n(n+1) + \lambda \ln |\mathbb{I}|}{|\mathbb{I}|}}$. Also, if $\omega \in S$, then $\|\hat{\boldsymbol{\theta}} - \theta\|_\infty \leq \frac{\eta_Q}{\sigma^2} \max(\sigma, \sigma^2)$.

Using Proposition 2.2, we then see that (4.1) is valid with $k = 2C_0\left(\frac{160}{9} + \frac{26}{3}\sqrt{2}\right)^{\alpha_0}$ and that (4.4) holds with $k(D) = 2C_0\left(3(k_Q D + k'_Q)\right)^{\alpha_0}$.

Let us now study the accuracy of $\tilde{\mathcal{P}}(\hat{\boldsymbol{\theta}})$. Following the proof of Proposition 2.2, we have on S the bound:

$$\varepsilon(\tilde{\mathcal{P}}(\hat{\boldsymbol{\theta}})) \leq 2(\eta_\rho + \kappa\sqrt{\eta_Q}) \max_{x \in X} \|x\|_1;$$

which gives the bounds in (4.2) and (4.5). Note that if Q_{N+1} is definite positive, we can improve the bound for $\sup_{x\in X}\left|\sqrt{x^\top \hat{\boldsymbol{Q}}_{\boldsymbol{N+1}} x} - \sqrt{x^\top Q_{N+1} x}\right|$. Indeed, in this case, for any $x \in X$, $\sqrt{x^\top Q_{N+1} x} \geq \sqrt{\beta(Q_{N+1})}\,\|x\|_1$ and

$$\left|\frac{x^\top Q_{N+1} x - x^\top \hat{\boldsymbol{Q}}_{\boldsymbol{N+1}} x}{\sqrt{x^\top Q_{N+1} x} + \sqrt{x^\top \hat{\boldsymbol{Q}}_{\boldsymbol{N+1}} x}}\right| \leq \frac{\|\hat{\boldsymbol{Q}}_{\boldsymbol{N+1}} - Q_{N+1}\|_\infty \max_{x\in X}\|x\|_1}{\sqrt{\beta(Q_{N+1})}}.$$

This implies on S: $\sup_{x\in X}\left|\sqrt{x^\top \hat{\boldsymbol{Q}}_{\boldsymbol{N+1}} x} - \sqrt{x^\top Q_{N+1} x}\right| \leq \frac{\eta_Q \max_{x\in X}\|x\|_1}{\sqrt{\beta(Q_{N+1})}}$, what implies the estimation in (4.3) and (4.6). Notice that for (4.2) and (4.3), we have $k_1 = \frac{14}{3} + 2\sqrt{2}$ and $k_2 = \frac{160}{9} + \frac{26}{3}\sqrt{2}$, and for (4.5) and (4.6), we have $k_1(D) = 2\left(7 + 3\sqrt{2} + 12\sqrt{\frac{2}{\ln 2}D}\right)$ and $k_2(D) = 3(k_Q D + k'_Q)$. □

Acknowledgments. The author is grateful to Anatoli Juditski of the "Laboratoire de Modélisation et Calcul" of University Joseph Fourier for helpful advice and discussions.

References

[BP05] D. Bertsimas and I. Popescu. Optimal inequalities in probability theory: a convex optimization approach. *SIAM Journal on Optimization*, 15(3):780–804, 2005.

[BTN99] A. Ben-Tal and A. Nemirovski. Robust solutions of uncertain linear programs. *Operations Research Letters*, 25(1):1–13, 1999.

[DI93] G. B. Dantzig and G. Infanger. Multi-stage stochastic linear programs for portfolio optimization. *Annals of Operations Research*, 45(1):59–76, 1993.

[Gui05] V. Guigues. Inférence statistique pour l'optimisation stochastique. Applications en finance et en gestion de production. *PhD thesis, Université Joseph Fourier, http://tel.archives-ouvertes.fr/tel-00098287/en/*, 2005.

[MS04a] D. Mercurio and V. Spokoiny. Statistical inference for time-inhomogeneous volatility models. *Annals of Statistics*, 32(2):577–602, 2004.

[MS04b] D. Mercurio and V. Spokoiny. Statistical inference for time-inhomogeneous volatility models. *WIAS Report, Preprint 583*, 2004.

[Pfl03] G. Pflug. Stochastic optimization and statistical inference. *Chapter 7 of Stochastic Programming: Handbooks in Operations Research and Management Science, Vol. 10, Elsevier, 2003, ISBN 0-444-50854-6 (editors: A. Ruszczyński and A. Shapiro)*, 2003.

[Sha89] A. Shapiro. Asymptotic properties of statistical estimators in stochastic programming. *Annals of Statistics*, 17(2):841–858, 1989.

[Sha93] A. Shapiro. Asymptotic behavior of optimal solutions in stochastic programming. *Mathematics of Operations Research*, 18(4):829–845, 1993.

[Sha94] A. Shapiro. Quantitative stability in stochastic programming. *Mathematical Programming*, 67(1):99–108, 1994.

[Smi95] J. Smith. Generalized Chebyshev inequalities: theory and application in decision analysis. *Operations Research*, 43(5):807–825, 1995.

Vincent Guigues
Department of Statistics, LJK
BP 53
38 041 Grenoble Cedex 9
France
vincent.guigues@imag.fr