SFB 823

Discussion Paper

Some practical aspects of sequential change point detection

Sivanja Sivanesan, Holger Dette, Daniel Ziggel

Nr. 18/2021

Some practical aspects of sequential change point detection

Sivanja Sivanesan, Holger Dette
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum, Germany

Daniel Ziggel
Quasol GmbH
Hauptstraße 5b
46569 Hünxe, Germany

Abstract

In this report we investigate the finite sample properties of a new online monitoring scheme which was recently introduced by Gösmann et al. (2020) by means of a simulation study and a real data example. We also develop an algorithm which can be used in an active risk management.

We start with an introduction in the basic notation and an explanation of the monitoring procedure, continue with an extensive simulation study to provide recommendations for the choice of several tuning parameters. Finally we present some illustration analyzing the Standard & Poor's 500, MSCI World and MSCI Emerging Markets indices.

1 Testing problem and basic defintions

Let $(X_t)_{t\in\mathbb{Z}}$ be a d-dimensional time series with common mean $\mu := \mathbb{E}[X_t] \in \mathbb{R}^d$ and F_t the distribution function of the random variable X_t at time t. We want to detect changes in the variance $V_t := \mathrm{Var}(F_t) \in \mathbb{R}^{d\times d}$, where the functional is defined as

$$V(F) = \int_{\mathbb{R}^d} (x-\mu)(x-\mu)^\top dF(x)$$
$$= \int_{\mathbb{R}^d} xx^\top dF(x) - \int_{\mathbb{R}^d} x dF(x) \int_{\mathbb{R}^d} x^\top dF(x) \ .$$

We assume that there is a historical *"stable" data set* of length $m \in \mathbb{N}$, in the sense that

$$V_1 = V_2 = \ldots = V_m \ . \tag{1.1}$$

The sequence $X_1, \dots, X_m$ is usually called historical or initial training data set (see, for example, Chu et al., 1996; Horváth et al., 2004; Wied and Galeano, 2013; Kirch and Weber, 2018, among many others). One can also check with a retrospective test whether the historical data is stable. In order to develop a sequential procedure to detect changes in the variance in the future $m + k \geq m + 1, k \in \mathbb{N}$, we define the following null hypothesis for the monitoring data set $X_{m+1}, \dots, X_{(T+1)m}$

$$H_0 : V_1 = \dots = V_m = V_{m+1} = \dots = V_{(T+1)m} \ , \tag{1.2}$$

against the alternative that the variance changes (once) at the position $m + 1 \leq m + k^* \leq (T + 1)m$, that is

$$H_1 : \text{there exists } k^* \in \mathbb{N} \text{ such that } V_1 = \dots = V_{m+k^*-1} \neq V_{m+k^*} = \dots = V_{(T+1)m} \ . \tag{1.3}$$

To simplify the notation and to get some clarity we apply a linear matrix transformation $\text{vech}(\cdot)$ on the variance(matrix). The half-vectorization $\text{vech}(\cdot)$ transforms a symmetric $d \times d$-matrix in a $d^* := d(d + 1)/2$-dimensional column vector by vectorizing the upper triangle part of the matrix. That means we now test for a change in the vector $\text{vech}(V_t) := \text{vech}(V(F_t))$.

Note that the null hypothesis in (1.2) and the alternative hypothesis in (1.3) are equivalent to the corresponding hypotheses version for the vectorized variance $\text{vech}(V_t)$. Therefore, we always refer to (1.2) and (1.3) although we consider $\text{vech}(V_t)$ in the formulation of the statistics. Moreover, we also have to apply the operator $\text{vech}(\cdot)$ on the sample variance $\hat{V}_i^j$. In order to test these hypotheses we define the statistic

$$\hat{E}_m(k) = \frac{1}{\sqrt{m}} \max_{j=0}^{k-1} (k - j) \left\| \text{vech}\left(\hat{V}_1^{m+j}\right) - \text{vech}\left(\hat{V}_{m+j+1}^{m+k}\right) \right\|_{\hat{\Sigma}_m^{-1}} \ , \tag{1.4}$$

where the notation $\|v\|_A^2 = v^\top A v$ defines a *weighted norm* of the vector v induced by a positive-definite matrix A. Here we use canonical variance estimator from the observations $X_i, \dots, X_j$ defined by

$$\hat{V}_i^j := \text{Var}(\hat{F}_i^j) = \int_{\mathbb{R}^d} x x^\top d\hat{F}_i^j(x) - \int_{\mathbb{R}^d} x d\hat{F}_i^j(x) \int_{\mathbb{R}^d} x^\top d\hat{F}_i^j(x) \ ,$$

where

$$\hat{F}_i^j(z) = \frac{1}{j - i + 1} \sum_{t=i}^{j} I\{X_t \leq z\}$$

denotes the *empirical distribution function* of observations $X_i, \ldots, X_j$. The matrix $\hat{\Sigma}_m$ denotes a consistent estimator of the long-run variance matrix

$$\Sigma_F = \sum_{t \in \mathbb{Z}} \mathrm{Cov}\left(\mathcal{IF}(X_0, F, \mathrm{vech}(V)), \mathcal{IF}(X_t, F, \mathrm{vech}(V))\right) \in \mathbb{R}^{d^* \times d^*}$$

from the initial data set $X_1, \ldots, X_m$, where

$$\mathcal{IF}(x, F, V) = (x - \mathbb{E}_F[X])(x - \mathbb{E}_F[X])^\top - V(F) \ .$$

An explicit formula for an estimator in the case $d = d^* = 1$ is given in Section 2, see equation (2.1).

In our case of *sequential* change point analysis, the monitoring is performed, whenever a new observation, say X_{m+k}, arrives. We compute $\hat{E}_m(k)$ and multiply it with a weight function w, such that we obtain

$$w(k/m)\hat{E}_m(k) \ , \tag{1.5}$$

where the weight function is given by

$$w(t) = \frac{1}{1+t} \ .$$

The monitoring stops at the time $k \in \{1, \ldots, Tm\}$ if

$$w(k/m)\hat{E}_m(k) > c_{\alpha,1} \tag{1.6}$$

and in this case we reject the null hypothesis (1.2) in favor of the alternative hypothesis (1.3). The constant $c_{\alpha,1}$ has to be chosen such that the resulting test has *asymptotic level* α for $\alpha \in (0,1)$, that is

$$\limsup_{m \to \infty} \ \mathbb{P}_{H_0}\left(\sup_{k=1}^{Tm} w(k/m)\hat{E}_m(k) > c_{\alpha,1}\right) = \mathbb{P}_{H_0}\left(L_1(T) > c_{\alpha,1}\right) \leq \alpha, \tag{1.7}$$

where the statistic $L_1(T)$ is defined by

$$L_1(T) = \sup_{0 < t \leq T/(T+1)} \max_{0 \leq s \leq t} \left| W(t) - W(s) \right| \ ,$$

and W denotes a d^*-dimensional Brownian motion.

For dimension $d = d^* = 1$ we can determine $c_{\alpha,1}$ by an explicit formula for the distribution function of the statistic $L_1(T)$

$$F_{L_1(T)}(x) = 1 + 8\sum_{k=1}^{\infty} (-1)^k \cdot k \cdot \left(1 - \Phi(kx/\sqrt{q(T)})\right)$$

(see Borodin and Salminen, 2002, page 146), where Φ denotes the cumulative distribution function of a standard normal distribution and $q(T) = T/(T+1)$.
Two alternative statistics for sequential have been considered in the literature and are defined by

$$\hat{Q}_m(k) := \frac{k}{\sqrt{m}} \left\| \mathrm{vech}(\hat{V}_1^m) - \mathrm{vech}(\hat{V}_{m+1}^{m+k}) \right\|_{\hat{\Sigma}_m^{-1}} , \tag{1.8}$$

$$\hat{P}_m(k) := \max_{j=0}^{k-1} \frac{k-j}{\sqrt{m}} \left\| \mathrm{vech}(\hat{V}_1^m) - \mathrm{vech}(\hat{V}_{m+j+1}^{m+k}) \right\|_{\hat{\Sigma}_m^{-1}} . \tag{1.9}$$

The procedure based on $\hat{Q}_m$ was introduced by Horváth et al. (2004) and was then reconsidered for example by Aue et al. (2012), Wied and Galeano (2013) and Pape et al. (2016). The statistic $\hat{P}_m$ was investigated by Fremdt (2015) and Kirch and Weber (2018).
The monitoring scheme based on $\hat{Q}_m$ and $\hat{P}_m$ is similar to the one in (1.6) based on $\hat{E}_m$. After a new observation X_{t+k} arrives, one determines

$$w(k/m)\hat{Q}_m(k) , \tag{1.10}$$

$$w(k/m)\hat{P}_m(k) . \tag{1.11}$$

and monitoring stops at the time $k \in \{1, \dots, Tm\}$ if the $w(k/m)\hat{Q}_m(k)$ or $w(k/m)\hat{P}_m(k)$ exceed their corresponding (asymptotic) quantiles $c_{\alpha,2}$ and $c_{\alpha,3}$, respectively. More precisely, these quantiles are defined by

$$\limsup_{m\to\infty} \quad \mathbb{P}_{H_0}\left(\sup_{k=1}^{Tm} w(k/m)\hat{Q}_m(k) > c_{\alpha,2}\right) = \mathbb{P}_{H_0}\left(L_2(T) > c_{\alpha,2}\right) \le \alpha ,$$
$$\limsup_{m\to\infty} \quad \mathbb{P}_{H_0}\left(\sup_{k=1}^{Tm} w(k/m)\hat{P}_m(k) > c_{\alpha,3}\right) = \mathbb{P}_{H_0}\left(L_3(T) > c_{\alpha,3}\right) \le \alpha ,$$

where

$$L_2(T) = \sup_{0\le t< T/(T+1)} |W(t)| ,$$
$$L_3(T) = \sup_{0\le t< T/(T+1)} \max_{0\le s\le t} \left| W(t) - \frac{1-t}{1-s} W(s) \right| ,$$

and W denotes a d^*-dimensional Browninan motion.

2 Simulation study

In this section, we will investigate the finite sample properties of the developed methodologies by means of a detailed simulation study and apply the detection procedure to financial data like Standard & Poor's 500 (S&P 500), MSCI World and MSCI Emerging Markets (MSCI E.M.). For the sake of brevity we will restrict ourselves to the the case $d = 1$.
For a practical implementation of the detection procedure one has to determine the parameters m, T and $\alpha \in (0, 1)$ and an estimator $\hat{k}^*$ for the change point position. In order to archive the *"optimal"* setting for the monitoring and active risk management we will have a look at synthetic data first.

For the long-run variance estimation we use the Bartlett kernel estimator [see Andrews (1991)] which is implemented in the MATLAB-function *hac* and the bandwidth for the estimation is chosen as $b_m = \log_{10}(m)$. To be precise define

$$\bar{X_m^2} = \frac{1}{m} \sum_{t=1}^{m} X_t^2,$$

and let

$$\hat{\gamma}_i = \frac{1}{m-i} \sum_{t=i+1}^{m} (X_t^2 - \bar{X_m^2})(X_{t-i}^2 - \bar{X_m^2})$$

denote the estimator of the auto-covariances $\gamma_i = \mathrm{Cov}(X_0^2, X_{-i}^2)$ from the initial data set $X_1, \ldots, X_m$, then we use the estimator

$$\hat{\Sigma}_m = \hat{\gamma}_0 + 2 \sum_{i=1}^{m-1} k\left(\frac{i}{b_m}\right) \hat{\gamma}_i \ , \tag{2.1}$$

where $b_m = \log_{10}(m)$ denotes the bandwidth and

$$k(x) = \begin{cases} 1 - |x| & \text{for } |x| \leq 1 \ , \\ 0 & \text{otherwise} \ . \end{cases}$$

is the Bartlett kernel.

In the following section we will compare the different testing procedures based on the test statistics $\hat{E}_m$, $\hat{Q}_m$ and $\hat{P}_m$ defined in (1.4), (1.8) and (1.9), respectively. In the section thereafter we will have a closer look at the *"best"* test statistic among these three. All subsequent results presented in these sections are based on 1000 independent simulation runs.

2.1 Comparison of different monitoring schemes

For a comparison of the detection procedures based on the different test statistics we simulate the following models

(M1) $Y_t = \varepsilon_t$

(M2) $Y_t = \varepsilon_t + 0.5\varepsilon_{t-1}$

(M3) $Y_t = \varepsilon_t \sqrt{1 + 0.1 Y_{t-1}^2}$,

where $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a sequence of independent standard normal distributed random variables. The model (M2) is a MA(1) while (M3) is an ARCH(1) model. We generate a process $\{X_t\}$ with a sample of $Tm + m$ observations. With the intention to simulate the alternative hypothesis (1.3) we multiply the data points $Y_{m+k^*}, \dots, Y_{Tm+m}$ with the factor $\sqrt{V^*}$ and obtain the synthetic data set

$$X_t = \begin{cases} Y_t & \text{if } t < m + k^* \\ \sqrt{V^*} Y_t & \text{if } t \geq m + k^* \end{cases} ,$$

for each model such that we obtain

(M1) $\mathrm{Var}(X_{m+k^*}) = V^*$

(M2) $\mathrm{Var}(X_{m+k^*}) = 1.25V^*$

(M3) $\mathrm{Var}(X_{m+k^*}) = \frac{1}{1-0.1}V^*$

We start the monitoring procedure with the initial stable data set $X_1, \dots, X_m$ satisfying (1.1). Then for the data point X_{m+k} we compute for instance the statistic $\hat{Q}_m(k)$ starting with $k = 1$ and reject the null hypothesis (1.2) and set as *time of rejection* $\tilde{t} = m + k$, if

$$w(k/m)\hat{Q}_m(k) > c_{1-\alpha,2},$$

where $c_{1-\alpha,2}$ is the corresponding critical value from Table 1. If the null hypothesis (1.2) is not rejected we set $k \mapsto k + 1$ and keep monitoring until $k = Tm$. If the null hypothesis is not rejected we stop the monitoring procedure and assume that no change in the variance occurred within the monitoring period. We use the same method for the monitoring procedures based on the test statistics $\hat{E}_m$ and $\hat{P}_m$ with the quantiles $c_{1-\alpha,1}$ and $c_{1-\alpha,3}$, respectively. These quantiles are listed in Table 1 [see also Gösmann et al. (2020)].

α	$c_{1-\alpha,1}$	$c_{1-\alpha,2}$	$c_{1-\alpha,2}$
0.01	2.7043	2.5145	2.5572
0.05	2.2339	1.9826	2.0435
0.1	2.0046	1.7380	1.8019

Table 1: *$(1-\alpha)$-quantiles of the distributions $L_1(4), L_2(4)$ and $L_3(4)$.*

We apply the different testing procedures in (1.5), (1.10) and (1.11) with $V^* = 1$ for model (M1) in order to investigate the approximation of the nominal level of the tests under the null hypothesis (1.2). In Table 2 we show the type I error of the different methods, where data is simulated from model (M1) with $T = 4$ under the null hypothesis of no change. For small initial sample sizes the approximation of the nominal level is not accurate. For example, for $m = 50$, the type I error for the testing procedure with nominal level 0.05 based on $\hat{E}_m$ and $\hat{Q}_m$ is approximately 12% while it is 13.7% for the procedure based on $\hat{P}_m$. We also observe that the approximation of the nominal level improves with increasing sample size. For example, if $m = 350$ we observe for the test $\hat{E}_m$ an empirical type I error of 5.7%. In general the testing procedures based on the statistics $\hat{E}_m$ and $\hat{Q}_m$ yield a better approximation of the nominal level in comparison to the monitoring scheme based on the statistic $\hat{P}_m$.

	$\hat{E}_m(k)$	$\hat{Q}_m(k)$	$\hat{P}_m(k)$	$\hat{E}_m(k)$	$\hat{Q}_m(k)$	$\hat{P}_m(k)$
α	$m = 50$			$m = 100$		
0.01	6.6%	6%	6.6%	3.7%	3.6%	3.8%
0.05	12.3%	12%	13.7%	8.8%	9.1%	9.4%
0.1	17.9%	17.1%	19.7%	13.6%	13.6%	14.4%
	$m = 150$			$m = 350$		
0.01	2.6%	2.6%	2.6%	1.8%	1.6%	1.8%
0.05	7.1%	7.2%	7.3%	5.7%	6.2%	6.4%
0.1	12.3%	13%	13.6%	11.3%	12.1%	12.4%

Table 2: *Type I error of the tests in (1.5), (1.10) and (1.11) for simulated data from model (M1) with $T = 4$ and different sizes $m = 50, 100, 150$ and $m = 350$ for the initial data set.*

Corresponding results for models (M2) and (M3) can be found in Table 3 and 4, respectively. We observe here that the nominal level is not so well approximated, even for larger initial sample sizes. A potential explanation for this observation is a too small bandwidth in the long-run variance estimator.

	$\hat{E}_m(k)$	$\hat{Q}_m(k)$	$\hat{P}_m(k)$	$\hat{E}_m(k)$	$\hat{Q}_m(k)$	$\hat{P}_m(k)$
α	$m = 50$			$m = 100$		
0.01	9.8%	9.3%	10%	7%	6.2%	6.6%
0.05	15.9%	14.4%	16.1%	12.3%	12%	12.6%
0.1	23.6%	22.7%	23.6%	18.2%	17.1%	18.8%
	$m = 350$			$m = 450$		
0.01	2.8%	2.6%	2.6%	2.4%	2.1%	2.3%
0.05	9%	9.2%	9.6%	7.4%	8.2%	8.5%
0.1	14.4%	14.9%	15.8%	13.3%	14.9%	14.4%

Table 3: *Type I error of the tests in* (1.5),(1.10) *and* (1.11) *for simulated data from model (M2) with* $T = 4$ *and different sizes* $m = 50, 100, 150$ *and* $m = 350$ *for the initial data set.*

	$\hat{E}_m(k)$	$\hat{Q}_m(k)$	$\hat{P}_m(k)$	$\hat{E}_m(k)$	$\hat{Q}_m(k)$	$\hat{P}_m(k)$
α	$m = 50$			$m = 100$		
0.01	25.7%	23.9%	25.3%	19.8%	18.4%	19.4%
0.05	35%	32.7%	34.7%	27.8%	25.8%	27%
0.1	36.9%	36%	38.4%	32.3%	30.1%	32.8%
	$m = 350$			$m = 450$		
0.01	10.3%	9.5%	9.9%	9.9%	9.3%	10%
0.05	16.5%	16.4%	17%	16.8%	15.8%	16.6%
0.1	23.2%	22.4%	23.1%	21.3%	21.1%	21.8%

Table 4: *Type I error of the tests in* (1.5), (1.10) *and* (1.11) *for simulated data from model (M3) with* $T = 4$*and different sizes* $m = 50, 100, 150$ *and* $m = 350$ *for the initial data set.*

In Figure 1, 2, 3 we display the rejection probabilities of the different tests in models (M1), (M2), (M3), respectively, under the alternative, where we choose the level $\alpha = 0.05$ and different values of the factor $\sqrt{V^*}$ are investigated. For different values of m we choose the center of the monitoring period as the change point position which means $m + k^* = m + Tm/2$.

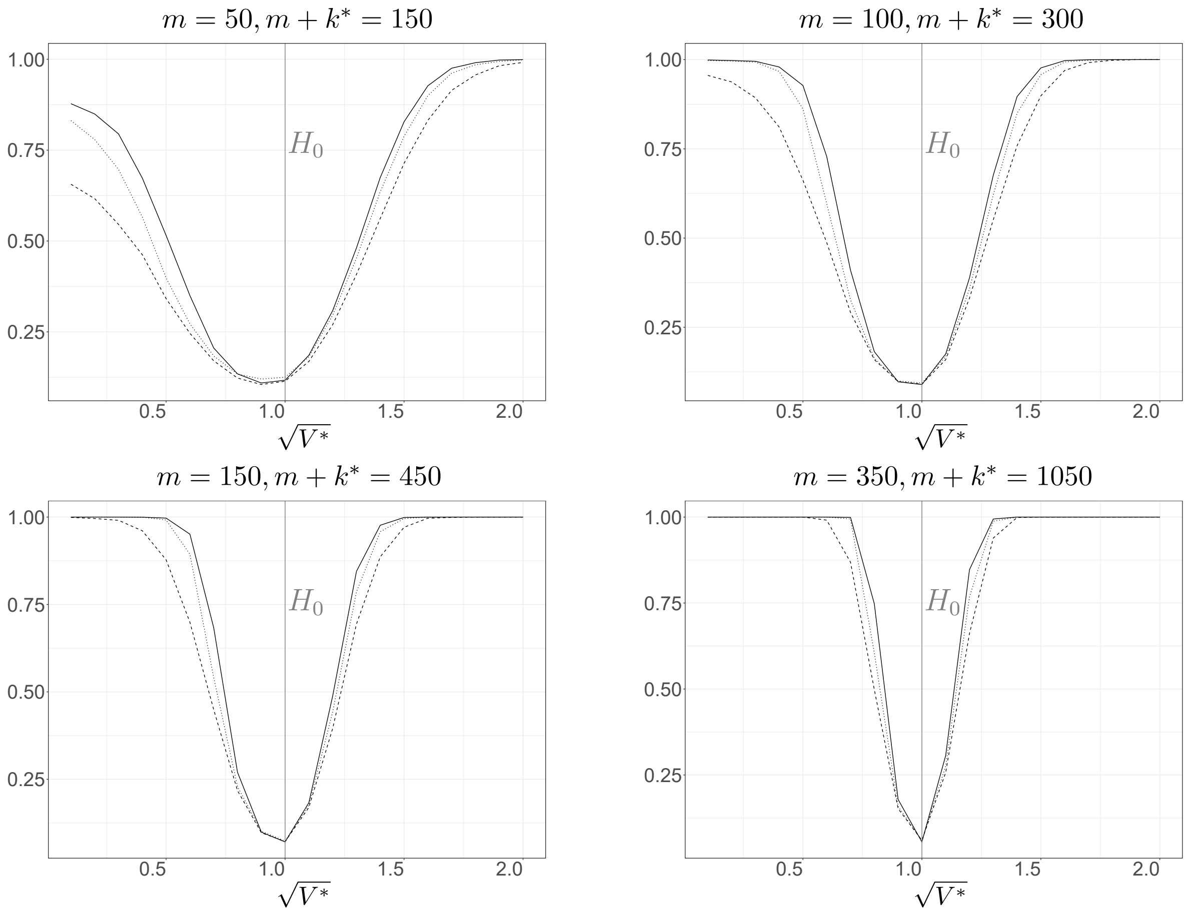


Figure 1: *Power of monitoring procedures for changes in the variance based on the statistics* $\hat{E}_m$ *(solid line),* $\hat{Q}_m$ *(dashed line) and* $\hat{P}_m$ *(dotted line) with different initial sample sizes (*$T = 4$*). Data is generated under model (M1) and* $\alpha = 0.05$.

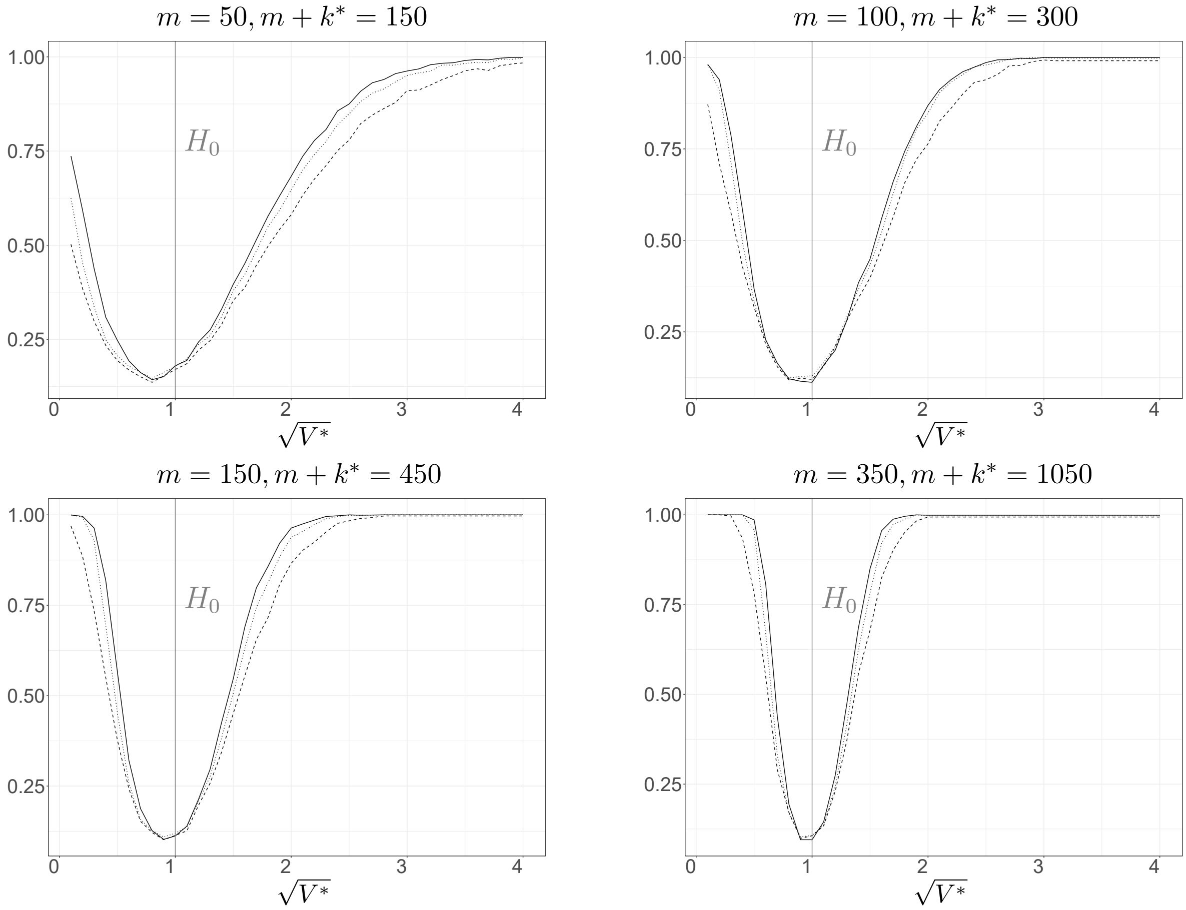


Figure 2: *Power of monitoring procedures for changes in the variance based on the statistics* $\hat{E}_m$ *(solid line),* $\hat{Q}_m$ *(dashed line) and* $\hat{P}_m$ *(dotted line) with different initial sample sizes (*$T = 4$*). Data is generated under model (M2) and* $\alpha = 0.05$.

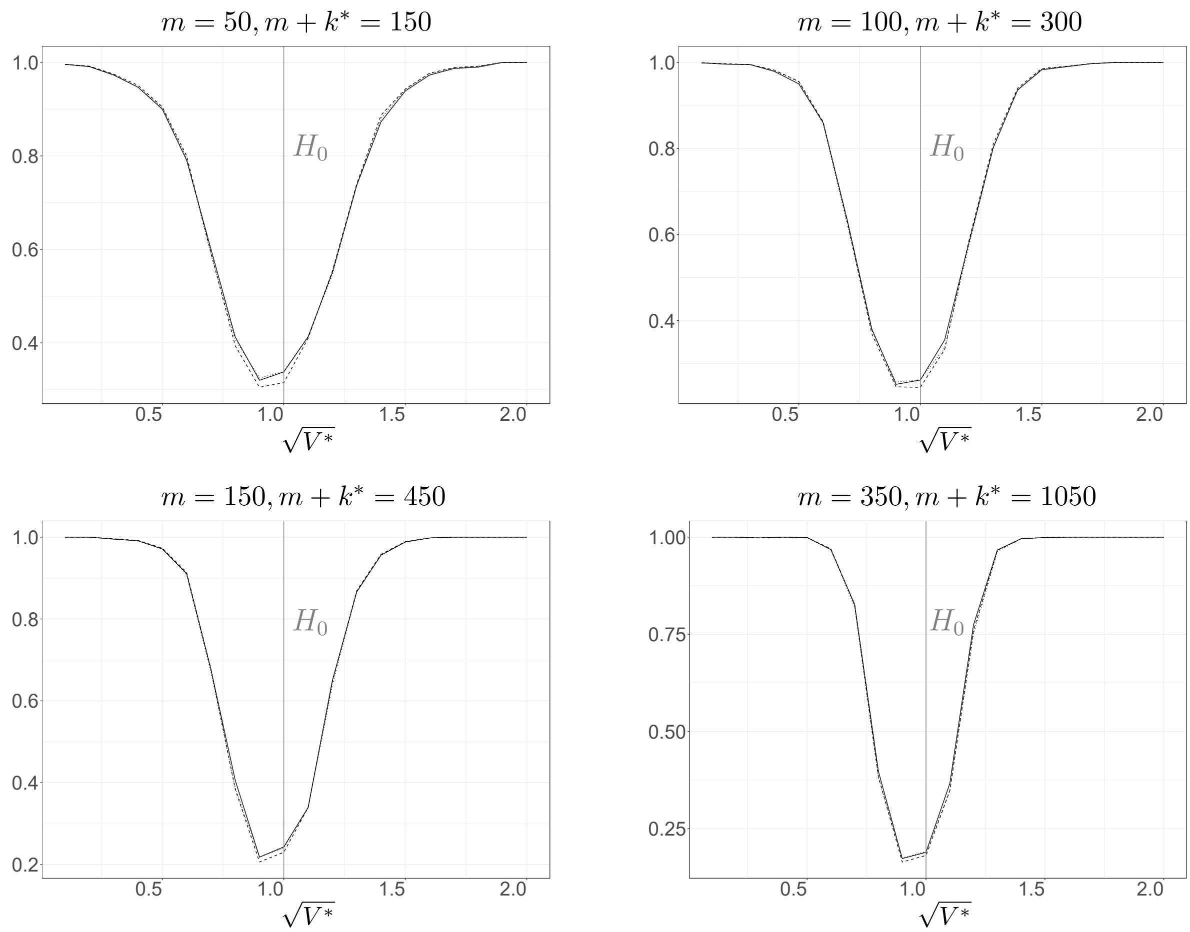


Figure 3: *Power of monitoring procedures for changes in the variance based on the statistics* $\hat{E}_m$ *(solid line),* $\hat{Q}_m$ *(dashed line) and* $\hat{P}_m$ *(dotted line) with different initial sample sizes (*$T = 4$*). Data is generated under model (M3) and* $\alpha = 0.05$.

From Figure 1 and 2 we see clearly that the monitoring scheme based on $\hat{E}_m$ outperforms the procedures based on $\hat{Q}_m$ and $\hat{P}_m$, while the test statistic $\hat{P}_m$ is almost as good as $\hat{E}_m$. For example in Figure 1 for $m = 150$ the procedure based on statistics $\hat{E}_m$ and $\hat{P}_m$ have for a variance change from 1 to $1.4^2 = 1.96$ already an empirical power close to 1, whereas the power of the test based on $\hat{Q}_m$ is approximately 0.88. Finally, for model (M3), we observe that the power for all three procedures is very similar, see Figure 3 .

In conclusion we see that the level and power of the detection procedure based on $\hat{E}_m$ proposed by Gösmann et al. (2020) is the best among the three competing monitoring schemes. Therefore, in the next section, we will have a closer look at the test statistic $\hat{E}_m$.

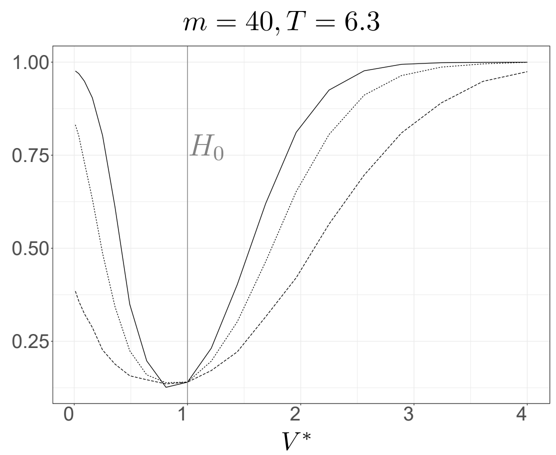
2.2 The monitoring scheme based on $\hat{E}_m$

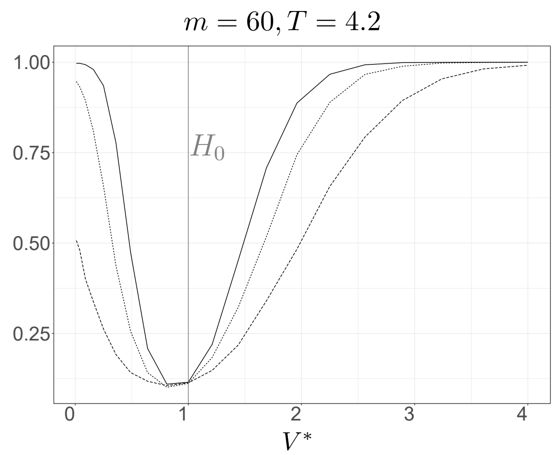
In order to analyze the procedure in (1.5) based on the statistic $\hat{E}_m$ in more detail we fix the test level to be $\alpha = 0.05$ and vary m and T such that the monitoring period mT is approximately 250, which nearly equals the amount of days in one trading year on the stock market. Table 5 shows that the approximation of the level by the test in (1.5) improves with an increasing initial sample size m and the best approximation of the nominal level is attained for $m = 200$ and $T = 1.2$.

$\boldsymbol{m}$	40	60	80	100	150	200
$\boldsymbol{T}$	6.3	4.2	3.1	2.5	1.7	1.2
	14%	11%	9.2%	8.3%	6.5%	5%

Table 5: *Empirical type I error of the test* (1.5) *(nominal level* $\alpha = 0.05$*) for different choices of* m *and* T*.*

Since we have so far only considered change point positions in the center of the monitoring procedure, we will now also have a look at two alternative positions of the change point. For this we simulate data from model (M1) with a change in the variance from 1 to $\mathrm{Var}(X_{m+k^*}) = V^*$ after one third and two thirds of the monitoring period. In other words we choose $k^* \approx 250/3$ and $k^* \approx 2 \cdot 250/3$ and present in Figure 4 the power curves of the procedure in (1.5) based on the statistic $\hat{E}_m$ for the different values of m, T and different change point positions.





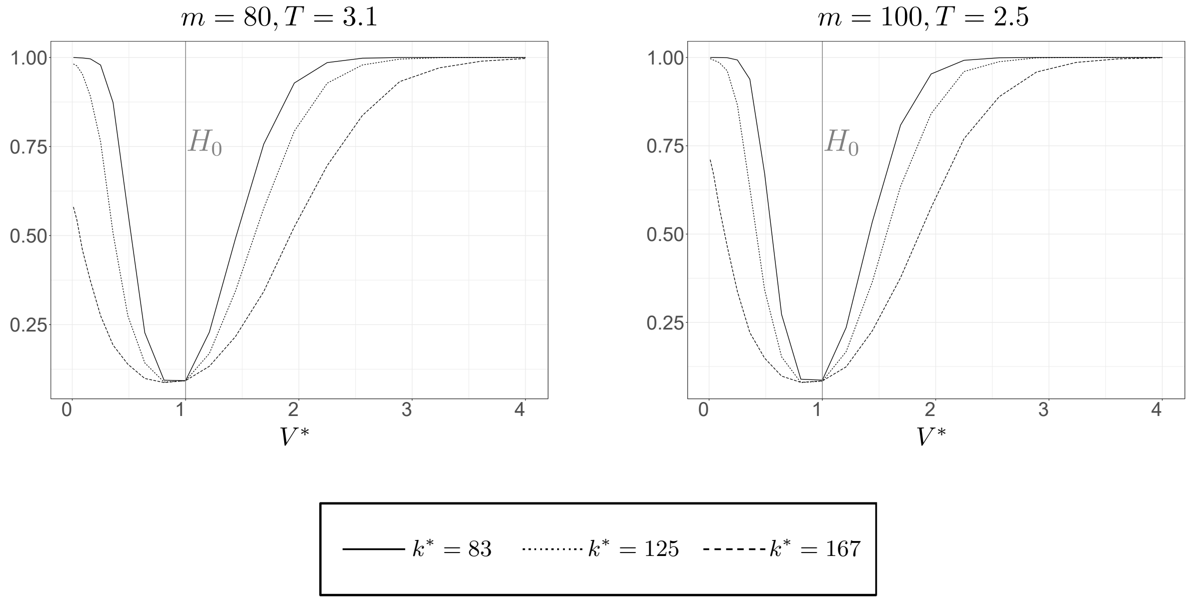


Figure 4: *Power of the monitoring procedure based on ther statistic $\hat{E}_m$ for different values of m and T with a break in the variance at the position $m + k^*$ for different k^*.*

We observe from Figure 4 that an earlier change point yields to larger power of the monitoring scheme. For example, in the case $m = 80$, $T = 3.1$ the power is approximately 1 when the variance changes after one third ($k^* = 83$) of the monitoring period from 1 to 2, while for $k^* = 125$ and $k^* = 167$ the power is 0.92 and 0.69, respectively.

In the following discussion we choose the change position to be in the center of the monitoring period with $k^* = 250/2 = 125$ which is also commonly used in the literature [see Gösmann et al. (2020)].

2.3 Comparison of different change point estimators

In practical applications besides the detection, the estimation of the change point is of particular importance and in this section we will investigate different estimators for the position of the change point which can be used after rejecting of the null hypothesis (1.2). For an empirical comparison of different change point estimators we simulate data $X_1, \ldots, X_{m+k^*}, \ldots, X_{m+mT}$ with the variance changing from 1 to V^* at time $m + k^*$ as described above for the following cases:

	m	T	$m+k^*$
case 1	40	6.3	165
case 2	60	4.2	185
case 3	80	3.1	205

If the monitoring procedure rejects the null hypothesis (1.2), we obtain a *time of rejection*, say $\tilde{t}$. Then we can estimate the change point with the help of an estimator, say $\hat{k}^*$, and stop the whole monitoring procedure. If there has been no rejection at time $m + mT$, we stop the procedure and conclude that no change in the variance has occurred within the monitoring period.
In order to compare the different parameters we repeat this monitoring procedure 1000 times and have a look at several key figures in Table 6-10. *No break* gives the proportion of procedures where the null hypothesis (1.2) is not rejected. *Early* stands for the percentage of procedures where the monitoring procedure stops before the actual change point position, i.e. if $\tilde{t} < m+k^*$. If the detection procedure stops after the actual change point one can determine the average and median time between $m+k^*$ and time of detection $\tilde{t}$ denoted by *average* and *median delay*, respectively. We observe that the results improve for increasing m. In particular, the percentage of too early stops becomes less for growing m. Note that the worst results are obtained for $V^* = 0.49$, which corresponds to the alternative, which is closest to the null hypothesis $V^* = 1$.

$\sqrt{V^*}$	**no break**	**early**	**average delay**	**median delay**
0.5	0.52	0.12	85.2	89
0.7	0.78	0.11	87.3	93.5
1.5	0.19	0.12	59.3	56
2	0	0.1	26.9	23

Table 6: *Basic properties of the monitoring scheme based on the statistic $\hat{E}_m$ for data simulated from model (M1), $m = 40$ and $T = 6.3$.*

$\sqrt{V^*}$	no break	early	average delay	median delay
0.5	0.34	0.07	86.5	89
0.7	0.75	0.09	88.5	93.5
1.5	0.11	0.07	52.7	48
2	0	0.08	22.9	20

Table 7: *Basic properties of the monitoring scheme based on the statistic* $\hat{E}_m$ *for data simulated from model (M1),* $m = 60$ *and* $T = 4.2$.

$\sqrt{V^*}$	no break	early	average delay	median delay
0.5	0.24	0.05	85.5	89
0.7	0.7	0.05	89.6	94
1.5	0.08	0.05	50.3	47
2	0	0.05	20.8	18

Table 8: *Basic properties of the monitoring scheme based on the statistic* $\hat{E}_m$ *for data simulated from model (M1),* $m = 80$ *and* $T = 3.1$

Corresponding results for models (M2) and (M3) are shown for the case $m = 40$ in Table 9 and 10, respectively. The results for model $(M2)$ are very similar to the results for $(M1)$ (see Table 6). However, for model (M3) more substantial differences can be observed. For example for $\sqrt{V^*} = 0.5$ the percentage of *no break* is quite low with 11% but from the 89% breaks approximately 30% are estimated *too early* (see Table 10).

$\sqrt{V^*}$	no break	early	average delay	median delay
0.5	0.48	0.16	84	90
0.7	0.74	0.16	78.6	83
1.5	0.2	0.16	54.9	51
2	0	0.14	27.9	22

Table 9: *Basic properties of the monitoring scheme based on the statistic* $\hat{E}_m$ *for data simulated from model (M2),* $m = 40$ *and* $T = 6.3$

$\sqrt{V^*}$	no break	early	average delay	median delay
0.5	0.11	0.3	16.8	11
0.7	0.39	0.33	25.8	19
1.5	0.05	0.33	17.4	10
2	0	0.32	4.1	3

Table 10: *Basic properties of the monitoring scheme based on the statistic $\hat{E}_m$ for data simulated from model (M3), $m = 40$ and $T = 6.3$*

In order to estimate the change point position we consider the following five estimators

$$
\begin{aligned}
\hat{k}_1^* &= m + \text{argmax}_{j=0}^{k-1}(m+j)(k-j)||\hat{V}_1^{m+j} - \hat{V}_{m+j+1}^{m+k}||_{\hat{\Sigma}_m^{-1}} \\
\hat{k}_2^* &= m + \text{argmax}_{j=0}^{k-1}\sqrt{(m+j)(k-j)}||\hat{V}_1^{m+j} - \hat{V}_{m+j+1}^{m+k}||_{\hat{\Sigma}_m^{-1}} \\
\hat{k}_3^* &= m + \text{argmax}_{j=0}^{k-1}\sqrt{(m+j)}(k-j)||\hat{V}_1^{m+j} - \hat{V}_{m+j+1}^{m+k}||_{\hat{\Sigma}_m^{-1}} \\
\hat{k}_4^* &= m + \text{argmax}_{j=0}^{k-1}(m+j)\sqrt{(k-j)}||\hat{V}_1^{m+j} - \hat{V}_{m+j+1}^{m+k}||_{\hat{\Sigma}_m^{-1}} \\
\hat{k}_5^* &= m + \text{argmax}_{j=0}^{k-1-b}||\hat{V}_1^{m+j} - \hat{V}_{m+j+1}^{m+k}||_{\hat{\Sigma}_m^{-1}} \quad \text{with } b = 10 \ ,
\end{aligned}
$$

where $k = \tilde{t} - m$ and we set $b = 0$ if $k - 1 - b - m \leq 0$.
A comparison of the different estimates is presented in Table 11, 12 and 13 for the model (M1) for different values of m. We consider the estimation of the change point as *correct* if the absolute difference of the actual and estimated change point is smaller than a given tolerance parameter W_{tol}. In our case we choose $W_{\text{tol}} = 50$ (days). If the estimated change point is not in the given tolerance range we classify the estimation as *wrong*. The mean and median of the absolute difference of the actual and estimated change point (in the case of no *early* break and no *wrong* estimation) is given through *average* and *median error*, respectively.

Note, that the estimators $\hat{k}_1^*$ and $\hat{k}_3^*$ do not perform well for estimating the change point. For example the percentage of *correct* estimation of $\hat{k}_1^*$ and $\hat{k}_3^*$ in the case $V^* = 4$ is 61% and 12%, respectively, while the other estimators produce the correct result in 85% of the cases (see Table 11). This trend can be also observed in Table 12 and Table 13 and accordingly the proportion of *wrong* estimations is higher for these estimators. Moreover, even if the estimators identify the change point *correctly*, the *error* they make is large, for example in Table 13 for $V^* = 2.25$ the *median error* for $\hat{k}_1^*$ and $\hat{k}_3^*$ are 41 and 86, respectively.
The estimator $\hat{k}_5^*$ is performs very good if the variance has a large variation, for instance in Table 13 we can see that for $V^* = 0.25$ and $V^* = 4$ the estimator yields the best results while for $V^* = 0.49$ and $V^* = 2.25$ (closer to H_0) it is not performing so well.

From these observations, we come to the conclusion that $\hat{k}_2^*$ and $\hat{k}_4^*$ are rather stable estimators for the various scenarios obtained by different choices of m, T and V^*. A comparison of $\hat{k}_2^*$ and $\hat{k}_4^*$ shows that in most cases the estimator $\hat{k}_4^*$ performs better. For example we observe from the results in Table 12 that $\hat{k}_4^*$ mostly estimates the change point *correctly* and also the estimated change points have the smallest distance to the actual change point.

	correct	**wrong**	**average error**	**median error**	**correct**	**wrong**	**average error**	**median error**
	0.5				**0.7**			
$\hat{k}_1^*$	0.31	0.06	25.4	18	0.08	0.02	33.4	25
$\hat{k}_2^*$	0.32	0.05	21	9.5	0.08	0.03	36.2	23
$\hat{k}_3^*$	0.11	0.26	68.1	69	0.01	0.1	91	96.5
$\hat{k}_4^*$	0.37	0	6.5	3	0.11	0	10.1	6
$\hat{k}_5^*$	0.32	0.05	19.1	4	0.05	0.06	61.8	64
	1.5				**2**			
$\hat{k}_1^*$	0.57	0.12	28.2	23	0.61	0.29	41.9	41
$\hat{k}_2^*$	0.65	0.04	13.1	6	0.86	0.04	9.1	3
$\hat{k}_3^*$	0.24	0.45	63.8	63	0.12	0.74	75.9	75
$\hat{k}_4^*$	0.66	0.03	13.6	7	0.89	0	5.9	3
$\hat{k}_5^*$	0.46	0.24	38.8	30	0.87	0.03	13	7

Table 11: *Performance of different estimators of the position of the change point for data generated from model (M1), where* $m = 40$ *and* $T = 6.3$.

	correct	**wrong**	**average error**	**median error**	**correct**	**wrong**	**average error**	**median error**
	0.5				**0.7**			
$\hat{k}_1^*$	0.48	0.11	28	21	0.10	0.06	42.8	40
$\hat{k}_2^*$	0.51	0.07	20.4	10	0.1	0.06	43.2	37.5
$\hat{k}_3^*$	0.19	0.4	70.3	74	0.02	0.15	88.8	90
$\hat{k}_4^*$	0.58	0	5.8	3	0.16	0	11.7	7
$\hat{k}_5^*$	0.56	0.02	8.9	3	0.08	0.09	56.1	56
	1.5				**2**			
$\hat{k}_1^*$	0.55	0.26	37.5	32	0.47	0.45	52	49
$\hat{k}_2^*$	0.74	0.08	16.1	7	0.87	0.05	11	3
$\hat{k}_3^*$	0.22	0.6	73	75	0.08	0.84	85	86
$\hat{k}_4^*$	0.79	0.03	11.9	6	0.92	0.01	6.1	3
$\hat{k}_5^*$	0.58	0.23	34.6	26.5	0.91	0.01	10.6	6

Table 12: *Performance of different estimators of the position of the change point for data generated from model (M1),* $m = 60$ *and* $T = 4.2$*.*

	correct	**wrong**	**average error**	**median error**	**correct**	**wrong**	**average error**	**median error**
	0.5				**0.7**			
$\hat{k}_1^*$	0.54	0.17	32.9	24	0.14	0.09	43.8	39
$\hat{k}_2^*$	0.63	0.08	20.3	10	0.15	0.08	41.4	32
$\hat{k}_3^*$	0.2	0.52	74.7	80	0.02	0.21	92.5	100
$\hat{k}_4^*$	0.71	0.01	6.5	3	0.22	0.01	13	8
$\hat{k}_5^*$	0.7	0.01	7.1	3	0.15	0.08	42.4	32
	1.5				**2**			
$\hat{k}_1^*$	0.52	0.35	44.5	41	0.36	0.59	60.2	58
$\hat{k}_2^*$	0.79	0.09	17	7	0.88	0.06	12.7	4
$\hat{k}_3^*$	0.15	0.72	81.3	86	0.04	0.90	91.3	94
$\hat{k}_4^*$	0.85	0.03	11.7	6	0.93	0.01	7.1	3
$\hat{k}_5^*$	0.66	0.22	32.4	26	0.94	0.01	9.3	6

Table 13: *Performance of different estimators of the position of the change point for data generated from model (M1), where* $m = 80$ *and* $T = 3.1$*.*

Corresponding results for model (M2) are displayed in Table 14 and we see that the estimator $\hat{k}_4^*$ provides the best results with respect to the estimation of the change position. For example for $\sqrt{V^*} = 2$ $\hat{k}_4^*$ estimates in 85% of the cases the change point correctly with 0% wrong estimations. On the other hand the alternative estimators produced some mistakes. Some results for model (M3) are depicted in Table 15, which shows that in this model the estimators $\hat{k}_4^*$ and $\hat{k}_5^*$ perform similar. For example in the case $\sqrt{V^*} = 0.7$. Both have almost the same number of correct and wrong results, while the average (median) error differ. Mostly the estimator $\hat{k}_5^*$ performs better than $\hat{k}_4^*$.

	correct	**wrong**	**average error**	**median error**	**correct**	**wrong**	**average error**	**median error**
			0.5				**0.7**	
$\hat{k}_1^*$	0.3	0.06	28.2	20	0.07	0.03	38.8	36
$\hat{k}_2^*$	0.3	0.05	28.2	20	0.07	0.04	43.3	33
$\hat{k}_3^*$	0.11	0.25	71.4	74	0.02	0.09	87.7	93
$\hat{k}_4^*$	0.35	0	6.8	3	0.1	0	14.6	8
$\hat{k}_5^*$	0.3	0.06	24.	6	0.04	0.06	64.4	65
			1.5				**2**	
$\hat{k}_1^*$	0.49	0.16	32.8	30	0.56	0.3	43.4	42
$\hat{k}_2^*$	0.58	0.07	18	9	0.8	0.05	11.8	4
$\hat{k}_3^*$	0.21	0.44	67	68	0.16	0.7	75.3	75
$\hat{k}_4^*$	0.61	0.04	15.8	10	0.85	0	7.14	3
$\hat{k}_5^*$	0.46	0.19	35.3	27	0.81	0.04	14.5	8

Table 14: *Performance of different estimators of the position of the change point for data generated from model (M2), where $m = 40$ and $T = 6.3$.*

	correct	wrong	average error	median error	correct	wrong	average error	median error
	0.5				**0.7**			
$\hat{k}_1^*$	0.21	0.38	60	56	0.09	0.2	62.3	60.5
$\hat{k}_2^*$	0.42	0.17	41.8	38.5	0.14	0.14	53.7	50
$\hat{k}_3^*$	0	0.6	85.3	84	0	0.27	88.1	87.5
$\hat{k}_4^*$	0.55	0.04	29.3	28	0.22	0.06	36.5	34
$\hat{k}_5^*$	0.53	0.06	23.1	16	0.2	0.08	37.5	29.5
	1.5				**2**			
$\hat{k}_1^*$	0.21	0.4	57.6	53	0.18	0.5	57.15	54
$\hat{k}_2^*$	0.5	0.13	39.5	37	0.65	0.02	23	21
$\hat{k}_3^*$	0	0.62	83.6	81	0	0.68	83.2	79
$\hat{k}_4^*$	0.59	0.03	26	24	0.68	0	16.8	15
$\hat{k}_5^*$	0.58	0.04	18.6	12	0.68	0	9.8	9

Table 15: *Performance of different estimators of the position of the change point for data generated from model (M3), where* $m = 40$ *and* $T = 6.3$.

After the empirical study we can sum up that in the following we will use the estimator $\hat{k}_4^*$ in order to estimate the change position. Thus, we will write

$$\hat{k}^* := \hat{k}_4^* = m + \text{argmax}_{j=0}^{k-1}(m+j)\sqrt{(k-j)}||\hat{V}_1^{m+j} - \hat{V}_{m+j+1}^{m+k}||_{\hat{\Sigma}_m^{-1}} \ . \tag{2.2}$$

3 Applications

Due to the current relatively low level of interest rates, traditionally safe securities such as government bonds (from countries with high credit ratings, e.g. Fitch AAA) have become much less attractive. In order to generate real returns, investors are forced to switch to riskier securities such as equities. However, these can lose value very quickly in crisis situations or more generally in phases of high volatility. For this reason, it is important to recognize phases of increased volatility as soon as possible to exit the market or to hedge with derivatives. Here, a risk overlay can be useful. In the optimal case, you then get a risk-return profile that safe government bonds used to have. For institutional investors, hedging is particularly interesting due to the regulatory framework. Institutional investors often shift their asset allocation to equities which are associated with high risk and over the last two decades several periods have been observed, where the market was not stable. Examples for such a situation are the Dot-Com-Bubble, the European debt crisis, the global financial crisis around and the recent COVID-19 pandemic. Because of such high-risk environments with extreme movements and high implied volatility most of the institutional investors consider an active risk overlay in their equity allocation (see, for example, Gösmann and Ziggel, 2018). With a strategy controlling the investors reduce the exposure to all the assets if the realized volatility exceeds a given level and increase it if the realized volatility is below the volatility level. Therefore it is of particular importance to locate phases when the markets enter a higher risk phase with the help of statistical testing procedures for detecting changes in the volatility and to react by an appropriate active risk management. Gösmann and Ziggel (2018) develop a methodology to detect changes in the volatility. They use a retrospective approach such that they perform their test every 10 trading days. In Section 3.3 we will have closer look on that methodology.

Before, we will apply the elaborated methodology to financial data to detect structural breaks such that we can divide our time series into different volatility phases. For this purpose we consider three frequently used financial time series: Standard & Poor's 500 (S&P 500), MSCI World and MSCI Emerging Markets (MSCI E.M.), which measure the market assessment for the most important companies from the United States, the whole world and emerging markets, respectively. We will then analyze these results in Section 3.2.

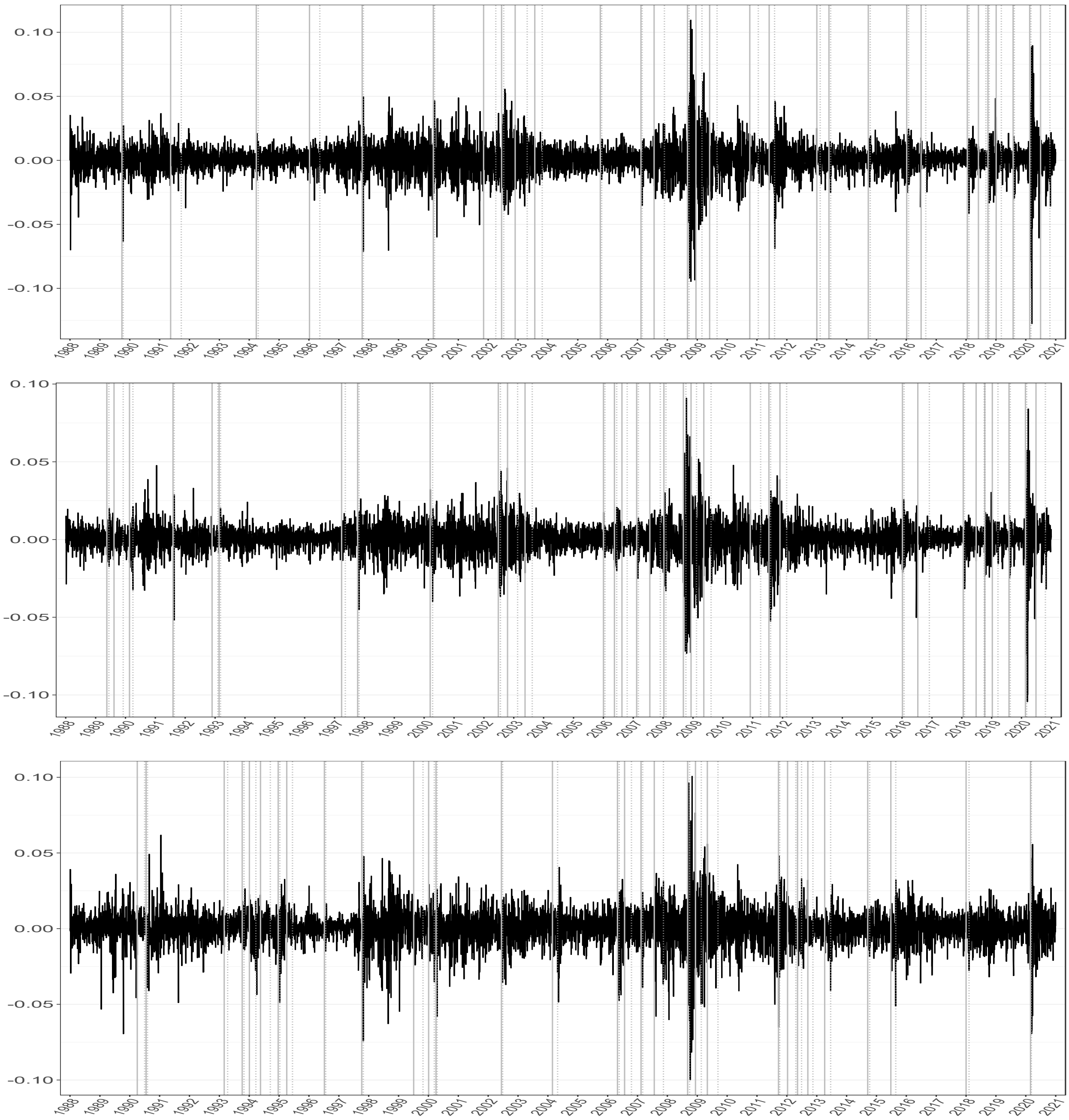


Figure 5: *Log-Return of the S&P 500 (upper panel), MSCI World (middle panel) and MSCI E.M. Index (lower panel)*

For each of the indices, we use the closing prices $\{P_t\}$ for $t \geq 0$ and determine the logarithmic returns $\{X_t\}$ with

$$X_t = \log\left(\frac{P_t}{P_{t-1}}\right) \quad \text{for} \quad t \geq 1 \ ,$$

which are displayed in Figure 5. In the application of the detection procedure we consider the logarithmic returns beginning at *day* $1 \mathrel{\widehat{=}}$ 1988-01-01 with X_1 and starting from day $m+1$ with the return X_{m+1} we apply the following algorithm:

Algorithm 1 Firstly choose T, the weight function w and the corresponding constant c_α such that (1.7) holds for an $\alpha \in (0, 1)$.

(*Initialization*) We take $X_1, \ldots, X_m$ as initial data set, which satisfies (or is assumed to satisfy) (1.1).

(*Monitoring*) If the data point X_{m+k} arrives for $k \in \{1, \ldots, Tm\}$, we compute the statistic $\hat{E}_m(k)$ and reject the null hypothesis in (1.2) if

$$w(k/m)\hat{E}_m(k) > c_{\alpha,1}$$

and set as time of rejection $\tilde{t} = m + k$. Then we estimate the time of change using the estimator $\hat{k}^*$ defined in (2.2). In the other case we continue *monitoring* the next observation which means we set $k \mapsto k+1$ until $k = Tm$.

(*Restart*) A restart of the algorithm is required if either a change point has been detected at $\hat{k}^*$ or the observation has been stopped at $k = Tm$ with no break. In both cases we take m new data points as initial training data set in order to continue the detecting procedure. Therefore we distinct two different cases:

(*Case of rejection*) The monitoring has been stopped at $\tilde{t}$ with an estimated change position $\hat{k}^*$ and we need two rules to determine the new stable data set:

i. If $\tilde{t} - \hat{k}^* \geq m$ we start the whole algorithm with the data starting at $X_{\tilde{t}-m+1}$ which means that we take m data points before the time of detection $\tilde{t}$ as initial training data set, such that $X_{\tilde{t}-m+1}, \ldots, X_{\tilde{t}}$ is the stable data. In this case $X_{\tilde{t}+1}$ is the first new monitoring observation.

ii. If $\tilde{t} - \hat{k}^* < m$ we start the whole algorithm with the data starting at $X_{\hat{k}^*+1}$. For this we have to wait until we have m observations after the estimated change point $\hat{k}^*$ and take these data points $X_{\hat{k}^*+1}, \ldots, X_{\hat{k}^*+m}$ as initial training data set. Then the first new monitoring observation is $X_{\hat{k}^*+m+1}$.

(*Case of no rejection*) If there has been no rejection of the null hypothesis such that the monitoring was stopped with the observation X_{Tm+m}, we take the last m data points of the monitoring period as new initial data set. That means $X_{Tm+1}, \dots, X_{Tm+m}$ is the stable data and we restart the whole algorithm with the first monitoring observation $X_{(T+1)m+1}$.

We continue applying this algorithm while new data arrives. However, as the data set ends with day 8610 $\widehat{=}$ 2020-12-31 the whole algorithm will be stopped completely with this last observation. During this whole algorithm the times of all rejections and all estimated change points will be saved.

In the following, we apply the algorithm with $m = 40$ and $T = 6.3$ to the data from the three indices. The detected change positions and time of detection are listed in Table 16 and in Figures 5. The algorithm detected for S&P 500 index 26 volatility breaks, while for MSCI World and MSCI E.M. 34 and 32 breaks, respectively.

A significant number of change points identified by our algorithm coincides with special phases on the financial markets. For example several breaks lay on the period of the gulf war (1990–1991), the rise and collapse of the Dot-com bubble (1995–2001), the global financial crisis (2007–2009), European debt crisis (2007–2009) and the COVID-19 pandemic (2019–today). The average time of delay to detect a break are 33 calendar days while the average median delay are 21 calendar days.

S&P 500		MSCI World		MSCI E.M.	
change position	detected at	change position	detected at	change position	detected at
1989-09-26	1989-10-12	1989-05-17	1989-06-14	1994-03-16	1994-04-01
1991-02-14	1991-03-28	1989-10-18	1989-12-19	1997-06-27	1997-08-14
1992-04-02	1992-04-06	1990-02-08	1990-02-23	1997-08-26	1997-09-01
1995-05-15	1995-05-30	1991-08-02	1991-08-19	1997-10-23	1997-10-28
1996-06-27	1996-07-12	1992-11-24	1993-02-05	1999-02-08	1999-07-06
1997-10-09	1997-10-24	1993-02-12	1993-03-05	1999-10-20	1999-12-24
2000-04-07	2000-04-24	1994-05-17	1994-08-30	1999-12-30	2000-01-06
2002-07-03	2002-07-23	1994-10-06	1994-11-21	2000-03-31	2000-04-17
2002-11-26	2003-02-26	1996-11-29	1997-01-30	2002-06-06	2002-06-25
2003-03-06	2003-03-21	1997-03-25	1997-05-09	2002-11-14	2003-02-07
2003-08-04	2003-12-16	1997-10-13	1997-10-28	2004-04-21	2004-05-14
2007-02-09	2007-02-26	2000-04-10	2000-05-31	2004-07-21	2004-09-28
2007-10-30	2008-03-17	2002-06-25	2002-07-19	2004-09-29	2004-10-11
2008-09-04	2008-10-06	2002-11-26	2003-04-15	2006-04-28	2006-06-07
2009-03-20	2009-05-28	2003-07-18	2003-10-09	2006-07-26	2006-09-27
2009-07-14	2009-09-29	2006-04-28	2006-06-07	2007-02-12	2007-03-05
2011-07-29	2011-08-26	2006-08-15	2006-10-10	2008-09-04	2008-10-03
2011-12-19	2012-03-20	2007-02-23	2007-03-12	2008-12-09	2009-01-29
2012-05-15	2012-06-05	2008-01-10	2008-03-10	2009-02-13	2009-03-27
2013-01-01	2013-03-29	2008-04-17	2008-09-01	2009-08-14	2009-12-10
2013-04-11	2013-04-15	2008-09-02	2008-09-05	2011-07-28	2011-08-22
2015-08-06	2015-08-21	2009-04-01	2009-06-09	2011-11-30	2012-02-29
2016-03-10	2016-09-05	2011-06-08	2011-07-20	2012-05-10	2012-07-26
2016-09-06	2016-09-08	2011-08-04	2011-08-08	2012-09-26	2012-12-25
2018-01-12	2018-01-29	2011-12-19	2012-02-20	2013-04-05	2013-06-13
2020-02-13	2020-02-26	2012-03-02	2012-03-07	2013-06-18	2013-06-25
		2012-11-16	2013-04-03	2014-09-01	2014-09-26
		2013-04-08	2013-04-12	2015-06-25	2015-08-26
		2013-06-03	2013-06-19	2017-11-13	2018-02-02
		2014-09-22	2014-10-08	2020-01-17	2020-02-03
		2018-01-17	2018-02-02	2020-05-01	2020-07-31
		2019-07-31	2019-08-15		
		2020-02-20	2020-03-06		
		2020-05-15	2020-07-29		

Table 16: *Estimated structural breaks: estimated change points and time of detection*

3.1 Active risk management

In this section we introduce our risk overlay which aims to avoid market phases of high volatility to achieve a more defensive risk profile than plain investments in the stock indices under consideration. We will compare our approach to the latter alternative by extensive backtesting.

Despite of testing for a change point, we compute on a daily basis the cumulative return P since the most recent change point. If the most recent change point is more than $W_{\max} = 250$ days ago, we consider the cumulative return of the last 250 days. Now the idea of the trading rule is to be completely invested or in cash. A naive approach is to invest if the cumulative return is positive ($P > 0$). However, to avoid unnecessary many transactions around the zero line and thus associated costs, we define an investment band, which is an interval $[-i, i]$ with $i \in (0, 1)$. In our particular case we take $i = 0.015$ and only generate trading signals if the cumulative return since the last change point of the last 250 days crosses the whole interval $[-0.015, 0.015]$. This means that we obtain the following cases:

- If we are completely invested and have

  $P > -i$ we stay invested.

  $P \leq -i$ we switch into cash.

- If we are completely in cash and have

  $P < i$ we stay in cash.

  $P \geq i$ we invest.

We apply this active risk management on the detected change points listed in Table 16.

3.2 Backtesting

In order to investigate the active risk management based on the new detection methodology we perform a backtest, comparing our strategy with the passive investment strategy. The idea of passive investment is the buy-and-hold portfolio strategy with the assumption that the market posts positive returns over a long time period. With this approach one can reduce costs which are incurred by transaction or risk management procedures (see, for example Hilliard and Hilliard, 2018).

As an alternative we consider a active risk management, where we switch from being invested to being in cash and vice versa. That means in this case we also have to imitate the regrouping for the backtest by taking trading costs for each transaction volume into account. For this purpose we take the average liquid stock market spread of 0.135% (see Wied et al., 2013) and declare the trading costs to be $0.135\%/2 = 0.0675\%$ (see Gösmann and Ziggel, 2018). Further, we assume the interest rate to be 0% which also means that cash assents do not gain any interest.We also assume that transactions are regulated in the local currency, i.e. fluctuations of exchange rates are not taken into account.

With this active risk management we can classify time regions in which we should whether invest or not. For the backtesting we still consider the three ETFs S&P 500, MSCI World and MSCI E.M. with the results from Table 16. We apply the above described active risk management and passive investment (no strategy) on the returns of the indices according to the detects change points. Now we can compute the new returns $\{X_t^*\}$ and the corresponding prices $\{P_t^*\}$ with the application of our strategy and compare these with the returns $\{X_t\}$ and prices $\{P_t\}$ with no strategy.

In order to compare these strategies we study some important financial key figures. Besides comparing the obvious key figures like *return* and *volatility*, we look at the *Sharpe ratio* (see Sharpe, 1994) and *maximum drawdown* (see Grossman and Zhou, 1993). The *Sharpe ratio* gives us the ratio of a (given) index's excess return to its corresponding volatility:

$$\frac{\textit{return} - \textit{riskfree return}}{\textit{volatility}} = \frac{\textit{excess return}}{\textit{volatility}} \; .$$

In our case we set *riskfree return* = 0%. *maximum drawdown* is the maximal possible loss of value of the index from its high point to its low point in the observing period and is defined with

$$\max_{t \in (1,8611)} \max_{s \in (0,t)} \frac{\textit{price}_s - \textit{price}_t}{\textit{price}_s} \; .$$

	S&P 500	MSCI World	MSCI E.M.
Return p.a.	6.52% (8.13%)	3.51% (5.49%)	4.18% (7.51%)
Volatility p.a.	11.26% (17.72%)	8.76% (14.76%)	10.12% (17.77%)
Sharpe ratio	0.58 (0.46)	0.4 (0.37)	0.41 (0.42)
Max. drawdown	21.36% (56.78%)	27.74% (59.07%)	50.13% (66.06%)
Regroupings	57	75	53

Table 17: *Backtest results for our active risk management and for the investment without any overlay (in the brackets).*

The corresponding results are displayed in Table 17 and show that the monitoring procedure with the active risk management reduces the risk of all three indices. The volatility decreased for example from 17.77% to 10.12% and the maximum drawdown from 66.06% to 50.13%. Since the great amount of regrouping of approximately 2 per year (57, 75 and 53) it is not surprising that the return decreased for S&P 500 1.6% p.a., MSCI World 2% p.a and MSCI E.M. 3.3% p.a.. Certainly the Sharpe ratio improved after applying the active risk management.

In the Figure 6 and 7 we display the effect of the active risk management for the three indices S&P 500, MSCI World and MSCI E.M., respectively. The red curves corresponds to the passive strategy and the blue curve to our strategy. Our active risk management is based on the new monitoring scheme, where $m = 40, T = 6.3$ and $\alpha = 0.05$ (the estimated change point are displayed in in Table 16). One can observe that the growth rate of the index without the application of the strategy is higher in comparison to the index with active risk management. This mostly leads to a lower return and volatility for the index with strategy than without the active risk management.

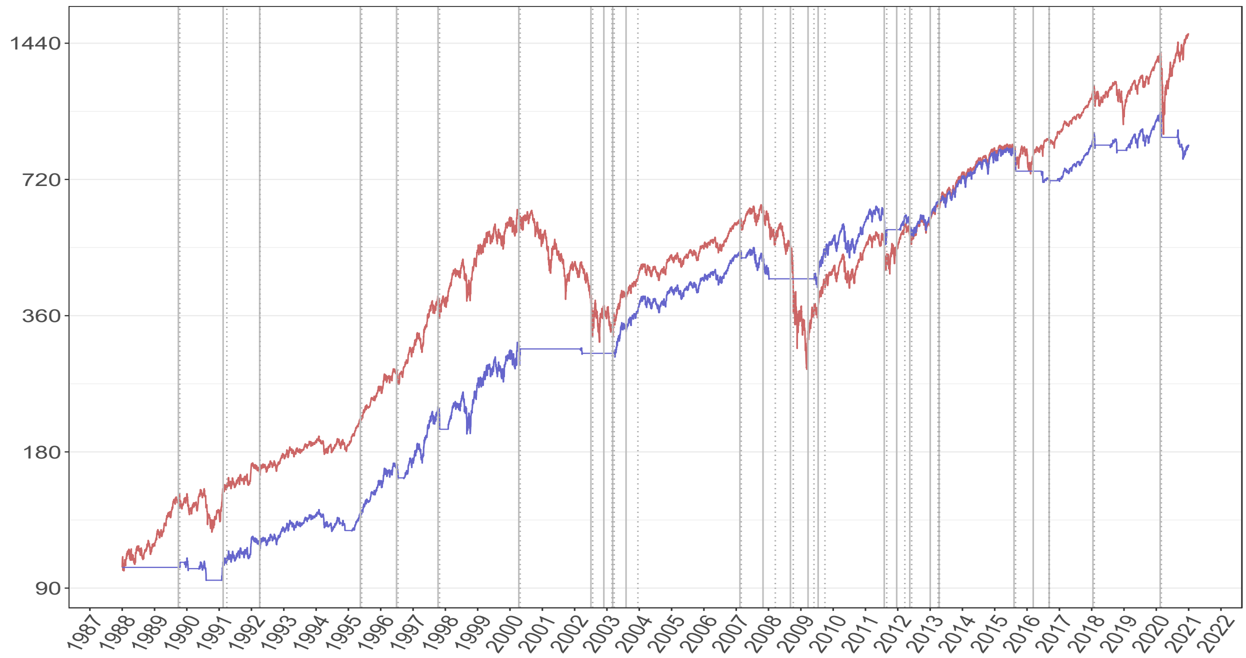


Figure 6: *Growth Rate of the S&P 500 Index (with basis 1988) without (red) and with active risk management (blue) while time of detection (dashed line) and estimated change point (solid line) are given.*

The change points of the index S&P 500 in Table 16 seem to be reasonable in many cases. The monitoring procedure detected the change point 1991-02-14 at 1991-03-28 which lays in the time period of the gulf war between the USA and Iraq (see Table 16). The active risk management and the change points 2000-04-07 and 2008-09-04 stopped us in investing around the high risk phases of the Dot-com bubble and the financial crisis, respectively. This leads to the over performance of the index with the active risk management during end of 2008 until end of 2012. Also in the period of the COVID-19 pandemic we did not invest.

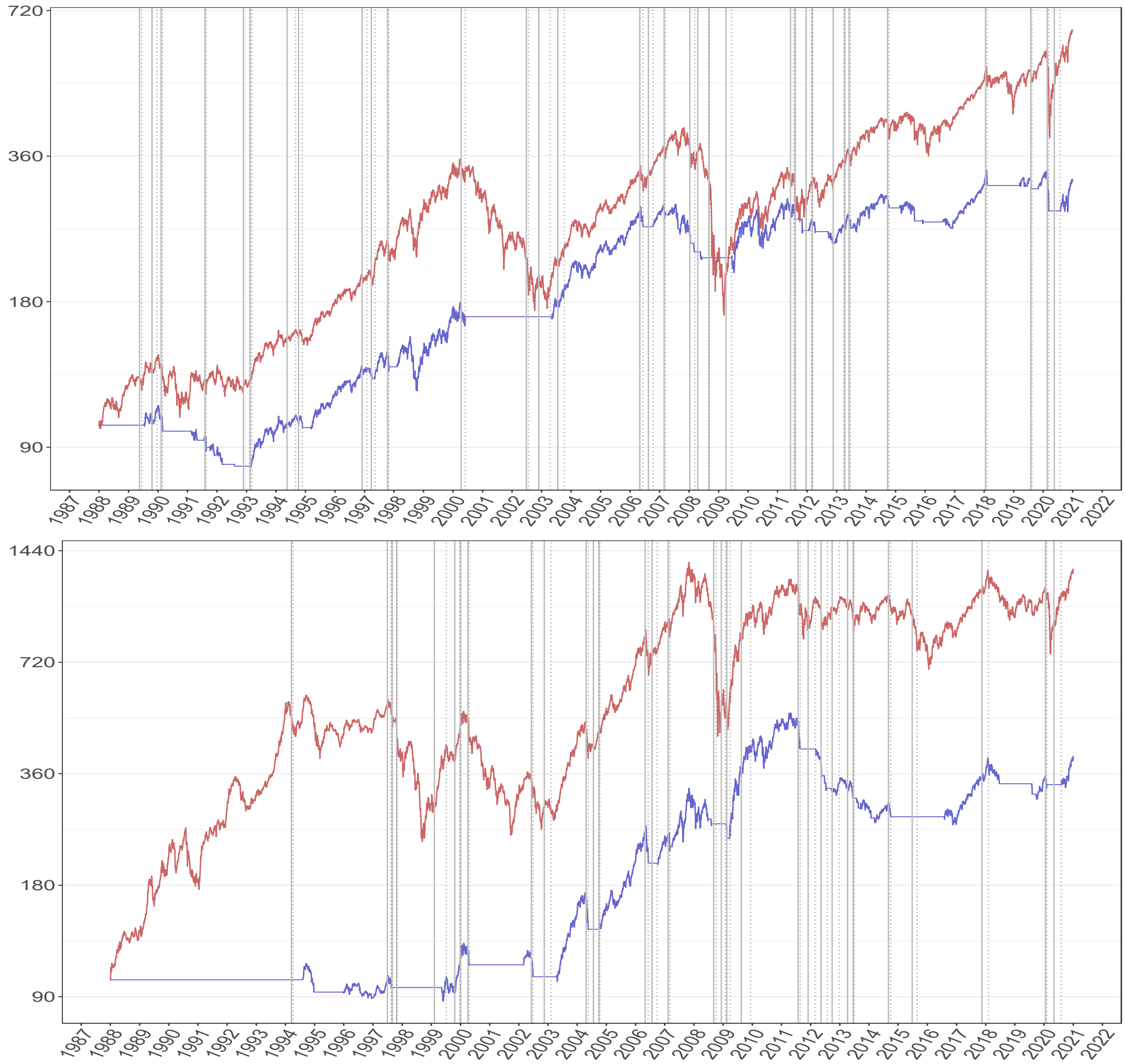


Figure 7: *Growth Rate of the MSCI World Index (upper panel) and MSCI E.M. Index (lower panel) with basis 1988. MSCI E.M. without (red) and with active risk management (blue) while time of detection (dashed line) and estimated change point (solid line) are given.*

For the MSCI World index 34 change points were detected by the monitoring procedure and we determined 75 regroupings. As one can see in Figure 7 most change points are localized at time periods corresponding to events with a strong impact on the market. For example the change point 2000-04-10 corresponds to the Dot-com bubble, 2008-09-02 to the global financial crisis, 2011-06-08 occurs after the Fukushima Daiichi nuclear disaster and 2020-02-20 during the COVID-19 pandemic. For the index MSCI E.M. we found 21 change points resulting in 53 regroupings. From Table 17 we observe that the MSCI E.M. is the index with the largest loss (of 44.3%) after the regrouping. Clearly the volatility decreased by 43% since we did not invest in high-risk periods. For example, we did not invest in the 1998 Russian financial crisis and because of the detected change points 2014-09-01 and 2015-06-25 we did not invest in the period of the Greek debt crisis.

3.3 Comparison with Methodology Gösmann and Ziggel (2018)

In this section, we will compare sequential change point detection methodology proposed in this paper (*methodology I*) with the sequential methodology introduced by Wied et al. (2012) and applied by Gösmann and Ziggel (2018) (*methodology II*). Gösmann and Ziggel (2018) proposed for log-returns $X_1, \ldots, X_K$ with $V_t = \mathrm{Var}(X_t)$ the hypotheses

$$
\begin{aligned}
&H_0 : V_1 = \ldots = V_K \\
&H_1 : \text{there exist } t \in \{1, \ldots, K-1\} \text{ with } V_t \neq V_{t+1} \ .
\end{aligned}
\tag{3.1}
$$

In order to test for a change in the volatility they employed the test statistic

$$
V_K = \max_{t=1}^{K} \left| \hat{D} \frac{t}{\sqrt{K}} (\hat{V}_1^t - \hat{V}_1^K) \right| ,
$$

while the factor $\hat{D}$ is important for the asymptotic distribution of V_K and is defined as

$$
\hat{D} = \left(\begin{pmatrix} 1 \\ -2\bar{X}_K \end{pmatrix}^{\top} \hat{D}_1 \begin{pmatrix} 1 \\ -2\bar{X}_K \end{pmatrix} \right)^{-1/2} .
$$

$\hat{D}_1$ is defined as

$$
\hat{D}_1 = \frac{1}{K} \sum_{t=1}^{K} \hat{U}_t \hat{U}_t^{\top} + 2 \sum_{j=1}^{K} k \left(\frac{j}{\sqrt{K}} \right) \frac{1}{K} \sum_{t=1}^{K-j} \hat{U}_t \hat{U}_{t+j}^{\top} ,
$$

while the vector $\hat{U}_t$ is defined as

$$
\hat{U}_t = \begin{pmatrix} X_t^2 - \frac{1}{K} \sum_{t=1}^{K} X_t^2 \\ X_t - \frac{1}{K} \sum_{t=1}^{K} X_t \end{pmatrix}
$$

and the function k is defined as

$$k(x) = \begin{cases} 1 - |x| & \text{for } |x| \leq 1 \ , \\ 0 & \text{otherwise} \ . \end{cases}$$

Under the null hypothesis (1.2) and some mild assumptions one can show (see Wied et al., 2012) that

$$V_K \xrightarrow{d} \sup_{t \in [0,1]} B(t) \ ,$$

where $B(t)$ is an one-dimensional Brownian Bridge such that with its quantiles we can provide an asymptotic test.
After observing 60 trading days such that we have log returns $X_1, \ldots, X_{60}$ we apply the test on day 61 on $X_1, \ldots, X_{61}$. We compute V_K and if $V_K \leq c_\alpha$ for an appropriate quantity c_α such that the test has significance level α, we do not reject the null hypothesis (3.1) and wait for 10 trading days to include these returns and apply the test once again. But if $V_K > c_\alpha$ we decide for the alternative hypothesis and estimate the change point with the following estimator

$$\hat{t} = \text{argmax}_{t=1}^{K} \left| \hat{D} \frac{t}{\sqrt{K}} (\hat{V}_1^t - \hat{V}_1^K) \right| \ .$$

Then the estimated change point is the new anchor point such that 10 trading days later we apply the test again. We continue the monitoring procedure in this way with the updated anchor point. But if the last anchor point is more that 250 trading days ago, we only consider the last 250 log-returns. We then stop the monitoring procedure with the last observation.

While detecting the change points we apply the active risk management defined in Section 3.1. To compare methodology I and II we apply the monitoring procedure with active risk management on the financial data and evaluate the results by means of backtesting. In the upper panel of Figure 8 we display the corresponding results for methodology I und II in the case of the S&P 500 index. The gray lines show the detected change points and time of detection of the methodology II with $\alpha = 0.05$. For methodology I we picked the parameter $m = 60, T = 4.2$ and $\alpha = 0.05$. With methodology I (methodology II) we detected 30 change points (34 change points) for S&P 500. Since methodology II is applied every 10 days the average time of delay to detect the change points are 88 trading days while for the other methodology the change point detection takes 42 trading days. Corresponding results for the two other indices are shown in the middle and lower panel of the figure. Further, in Table 18 we computed the key figures for backtesting.

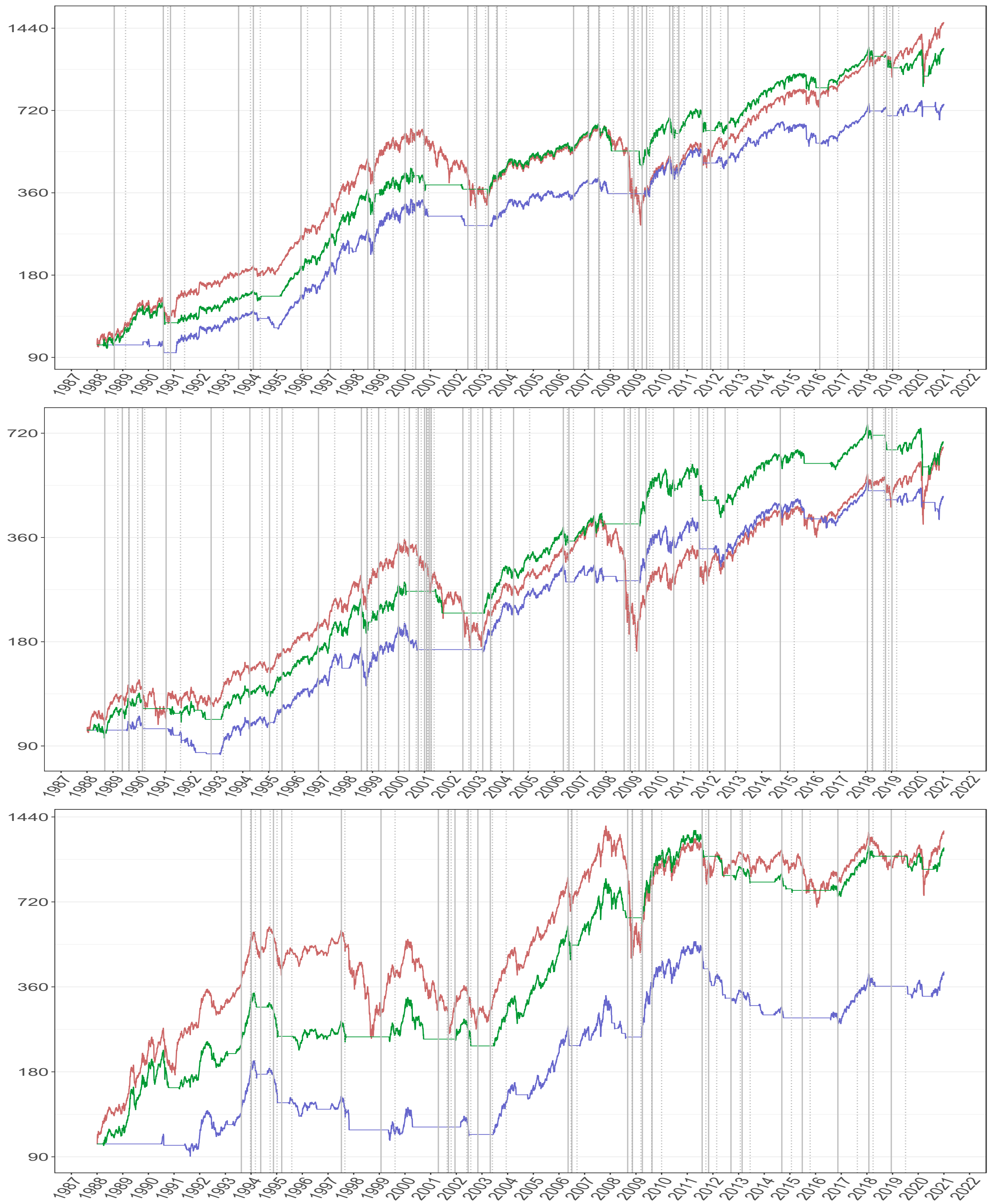


Figure 8: *Growth Rate of the S&P 500 Index (upper panel), MSCI World Index (middle panel) and MSCI E.M. (lower panel) without active risk management (red) and with the active risk management based on methodology I (blue) and the methodology II (green); time of detection (dashed gray line) and estimated change point (solid gray line) of methodology I*

	S&P 500			MSCI World		
	no strategy	Method. I	Method. II	no strategy	Method. I	Method. II
Return p.a.	8.15%	6.15%	7.58%	5.4%	4.95%	5.76%
Volatility p.a.	17.73%	11.5%	13.14%	14.78%	9.33%	10.29%
Sharpe ratio	0.46	0.53	0.58	0.37	0.53	0.56
Max. drawdown	56.78%	20.65%	32.07%	59.07%	27.82%	30%
Regroupings		59	35		65	33
	MSCI E.M.					
	no strategy	Method. I	Method. II			
Return p.a.	7.24%	6.14%	7.3%			
Volatility p.a.	17.78%	10.34%	12.19%			
Sharpe ratio	0.4	0.59	0.6			
Max. drawdown	66.06%	42.98%	42.4%			
Regroupings		63	43			

Table 18: *Backtest results for the investment without any overlay and our active risk management based on methodology I and II*

From Figure 8 we see that many change points are detected close to important financial events such as the Dot-com bubble, the global financial crisis and COVID-19 pandemic. The return of the backtest based on methodology II is better than for methodology I. But the volatility and maximum drawdown is better for methodology I compared to methodology II. We conclude that methodology I has advantages for a trading on a daily basis. Even though the return gets worse with methodology I it is a better to control the volatility. We also could identify high risk phases with methodology I. On the other hand methodology II is more useful in applications, where trading is performed every 10 days.

Acknowledgements This work has been supported in part by the Collaborative Research Center “Statistical modeling of nonlinear dynamic processes” (SFB 823, Teilprojekt A1) of the German Research Foundation (DFG).

References

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59(3):817–858.

Aue, A., Hörmann, S., Horváth, L., Hušková, M., and Steinebach, J. G. (2012). Sequential testing for the stability of high-frequency portfolio betas. *Econometric Theory*, 28(4):804–837.

Borodin, A. N. and Salminen, P. (2002). *Handbook of Brownian Motion - Facts and Formulae*. Probability and Its Applications. Birkhauser, Basel and Boston, second edition.

Chu, C.-S. J., Stinchcombe, M., and White, H. (1996). Monitoring structural change. *Econometrica*, 64(5):1045–1065.

Fremdt, S. (2015). Page's sequential procedure for change-point detection in time series regression. *Statistics*, 49(1):128–155.

Gösmann, J., Kley, T., and Dette, H. (2020). A new approach for open–end sequential change point monitoring. *Journal of Time Series Analysis*.

Gösmann, J. and Ziggel, D. (2018). An innovative risk management methodology for trading equity indices based on change points. *Journal of Asset Management*, 19(2):99–109.

Grossman, S. J. and Zhou, Z. (1993). Optimal investment strategies for controlling drawdowns. *Mathematical Finance*, 3(3):241–276.

Hilliard, J. E. and Hilliard, J. (2018). Rebalancing versus buy and hold: theory, simulation and empirical analysis. *Review of Quantitative Finance and Accounting*, 50(1):1–32.

Horváth, L., Hušková, M., Kokoszka, P., and Steinebach, J. (2004). Monitoring changes in linear models. *J. Stat. Plann. Inference*, 126(1):225–251.

Kirch, C. and Weber, S. (2018). Modified sequential change point procedures based on estimating functions. *Electronic Journal of Statistics*, 12(1):1579–1613.

Pape, K., Wied, D., and Galeano, P. (2016). Monitoring multivariate variance changes. *Journal of Empirical Finance*, 39:54–68.

Sharpe, W. F. (1994). The sharpe ratio. *The Journal of Portfolio Management*, 21(1):49–58.

Wied, D., Arnold, M., Bissantz, N., and Ziggel, D. (2012). A new fluctuation test for constant variances with applications to finance. *Metrika*, 75(8):1111–1127.

Wied, D. and Galeano, P. (2013). Monitoring correlation change in a sequence of random variables. *Journal of Statistical Planning and Inference*, 143(1):186–196.

Wied, D., Ziggel, D., and Berens, T. (2013). On the application of new tests for structural changes on global minimum-variance portfolios. *Statistical Papers*, 54(4):955–975.