SFB 823

Discussion Paper
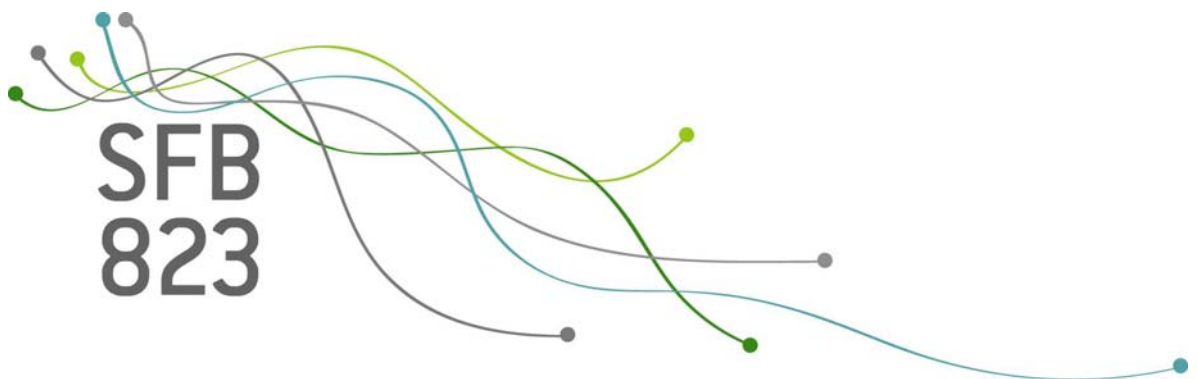
A new set of improved value-at-risk backtests

Daniel Ziggel, Tobias Berens, Gregor N.F. Weiß, Dominik Wied

Nr. 29/2013

SFB 823

A New Set of Improved Value-at-Risk Backtests[*]

Daniel Ziggel [†]

quasol GmbH

Tobias Berens [‡]

zeb/rolfes.schierenbeck.associates

Gregor N.F. Weiß[§]

Technische Universität Dortmund

Dominik Wied [¶]

Technische Universität Dortmund

August 26, 2013

ABSTRACT

We propose a new set of formal backtests for VaR-forecasts that significantly improve upon existing backtesting procedures. Our new test of unconditional coverage can be used for both directional and non-directional testing and is thus able to test separately whether a VaR-model is too conservative or underestimates the actual risk exposure. Second, we stress the importance of testing the property of independent and identically distributed (i.i.d.) VaR-exceedances and propose a simple approach that explicitly tests for the presence of clusters in VaR-violation processes. Results from a simulation study indicate that our tests significantly outperform competing backtests in several distinct settings. In addition, the empirical analysis of a unique data set consisting of asset returns of an asset manager's portfolios underline the usefulness of our new backtests especially in times of market turmoil.

Keywords: Value-at-Risk, backtesting, Monte Carlo simulation.

JEL Classification: C52, C53, C58.

[†]Marktallee 8, 48165 Münster, Germany, telephone: +49 2501 977 9662, e-mail: *daniel.ziggel@quasol.de*.

[‡]Hammer Str. 165, D-48153 Münster, Germany, telephone: +49 251 97128 355, e-mail: *tberens@zeb.de*.

[§]Otto-Hahn-Str. 6a, D-44227 Dortmund, Germany, telephone: +49 231 755 4608, e-mail: *gregor.weiss@tu-dortmund.de*.

[¶]CDI-Gebäude, D-44221 Dortmund, Germany, telephone: +49 231 755 5419, e-mail: *wied@statistik.tu-dortmund.de*.

A New Set of Improved Value-at-Risk Backtests

ABSTRACT

We propose a new set of formal backtests for VaR-forecasts that significantly improve upon existing backtesting procedures. Our new test of unconditional coverage can be used for both directional and non-directional testing and is thus able to test separately whether a VaR-model is too conservative or underestimates the actual risk exposure. Second, we stress the importance of testing the property of independent and identically distributed (i.i.d.) VaR-exceedances and propose a simple approach that explicitly tests for the presence of clusters in VaR-violation processes. Results from a simulation study indicate that our tests significantly outperform competing backtests in several distinct settings. In addition, the empirical analysis of a unique data set consisting of asset returns of an asset manager's portfolios underline the usefulness of our new backtests especially in times of market turmoil.

1 Introduction

Over the last two decades, Value-at-Risk (VaR) has become the de facto standard tool for measuring and managing risk in the financial services industry. Defined as the p-quantile of a relevant profit and loss (P/L) distribution where p is regularly set to 1% or 5%, it is now widely used by commercial banks and insurers as well as firms outside the financial industry to assess the risk exposure of single investments and portfolios.[1] A simple reason for this importance of VaR for the financial industry is given by the fact that under the 1996 Market Risk Amendment to the first Basel Accord, banks were allowed to employ internal VaR-models to calculate capital charges for their risky investments. Despite its popularity with practicioners, however, VaR has also received criticism from academia due to its lack of subadditivity (and thus coherence, see Artzner et al., 1999) in case of non-gaussian P/L distributions.[2] Even more importantly, commentators have blamed VaR in part for the severity of the recent financial crisis as the industry-wide use of VaR capital constraints enabled externalities to spread in financial markets through the pricing of risk (see Shin, 2010).[3] Consequently, both regulators and financial risk managers have recently taken an increased interest in model validation and backtests of VaR-forecasts.

Despite its importance for bank regulation, VaR-backtesting has received relatively little attention in the financial econometrics literature compared to the numerous studies on the estimation and forecasting of VaR. One of the first formal statistical backtests for VaR was proposed by Kupiec (1995) who tests the sequence of VaR-violations for the correct number of violations (i.e., unconditional coverage). Christoffersen (1998) and Christoffersen and Pelletier (2004) extend these first tests of unconditional coverage by additionally testing for the independence of the sequence of VaR-violations yielding a combined test of conditional coverage. Recently, an integrated framework for VaR-backtesting that includes the previously mentioned tests was proposed

---

[1] Extensive discussions of the properties of VaR and its use in practice are given, e.g., by Dowd (1998), Jorion (2006), and Alexander (2008).

[2] Note, however, that evidence by Daníelsson et al. (2005) points out the subadditivity of VaR for most practical applications.

[3] Similar arguments in favor of a destabilizing effect of bank regulation based on VaR on the economy are stated by Leippold et al. (2006) and Basak and Shapiro (2001).

by Berkowitz et al. (2011). Further examples of the few backtests for VaR that are available to regulators are due to Berkowitz (2001), Engle and Manganelli (2004), Haas (2005) and Candelon et al. (2011), although the test of unconditional coverage continues to be the industry standard mostly due to the fact that it is implicitly incorporated in the framework for backtesting internal models proposed by the Basel Committee on Banking Supervision (BCBS) (1996).[4]

In this paper, we propose a new set of backtests for VaR-forecasts that significantly improve upon existing formal VaR-backtests like, e.g., the benchmark models proposed by Christoffersen and Pelletier (2004). We first restate the definitions of the unconditional coverage property and propose a new test of the correct number of VaR-exceedances. Extending the current state-of-the-art, our new test can be used for both directional and non-directional testing and is thus able to test separately whether a VaR-model is too conservative or underestimates the actual risk exposure. Second, we stress the importance of testing both for the property of independent as well as the property of identically distributed VaR-exceedances and propose a simple approach for testing for both properties. While it has been noted in previous studies that VaR-violations should ideally be i.i.d., standard backtests focus solely on the independence of the violations.[5] In this paper, we argue that the property of identically distributed VaR-exceedances is of vital importance to regulators and risk managers. In particular, we show that traditional VaR-backtests that center around first-order autocorrelation in violation processes are often not able to detect misspecified VaR-models during calm boom and highly volatile bust cycles. The new test of the i.i.d. property of VaR-violations explicitly tests for the presence of clusters in VaR-violation processes. This new feature is highly economically relevant as our test for violation clusters can identify VaR-models that yield inaccurate risk forecasts when they are most undesirable: during economic busts and financial crises when extreme losses on investments cluster due to a persistent increase in the volatility level. Finally, we also propose a weighted backtest of conditional coverage that simultaneously tests for a correct number and the i.i.d. property of VaR-violations. Our proposed weighted back-

---

[4]   A review of backtesting procedures that have been proposed in the literature is given by Campbell (2007).

[5]   In fact, previous Markov- and duration-based tests of Christoffersen (1998), Christoffersen and Pelletier (2004) and Candelon et al. (2011) only consider autocorrelation in VaR-violations as one possible reason why VaR-violations could be clustered.

test is in the spirit of the original backtest of conditional coverage by Christoffersen and Pelletier (2004), but generalizes it by allowing the user to choose the weight with which the test of unconditional coverage enters the joint test of conditional coverage. Our newly proposed set of backtests are directly based on i.i.d. Bernoulli random variables making them very intuitive and easy to implement. By construction, these tests automatically keep their level, even for very small sample sizes as they are often found in VaR-backtesting.

We employ our proposed backtests in a simulation study using several sets of simulated data that mimic real-life settings in which the simulated data violate the unconditional coverage, i.i.d., and conditional coverage properties to different degrees. The results indicate that our tests significantly outperform competing backtests in several distinct settings. In addition, we present an empirical application of the new tests using a unique data set consisting of the asset returns of an asset manager's portfolios.

The paper is organized as follows. Section 2 introduces the notation, defines the properties of VaR-violations, and describes our new set of backtests. Section 3 evaluates the performance of the newly proposed backtests as well as several benchmark procedures for backtesting VaR-forecasts in a simulation study. Section 4 presents results from our empirical application study. Section 5 concludes the paper.

2   Methodology

In this section, we introduce the notation used throughout the paper, redefine the desirable properties of VaR-violations that are frequently discussed in the literature and present our new backtests.

2.1   Notation and VaR-Violation Properties

Let $\{y_t\}_{t=1}^n$ be a sample of a time series y_t corresponding to daily observations of the returns on an asset or a portfolio. We are interested in the accuracy of VaR-forecasts, i.e., an estimation of

confidence intervals. We denote the VaR-forecasts by $VaR_{t|t-1}(p)$, where p is the VaR coverage probability. In practice, the coverage probability p is typically chosen to be either 1% or 5% (see Christoffersen, 1998). This notation implies that information up to time $t - 1$ is used to obtain a forecast for time t. Moreover, we define the indicator variable $I_t(p)$ for a given VaR-forecast $VaR_{t|t-1}(p)$ as

$$I_t(p) = \begin{cases} 0, & \text{if } y_t \geq VaR_{t|t-1}(p); \\ 1, & \text{if } y_t < VaR_{t|t-1}(p). \end{cases} \tag{1}$$

If this indicator variable is equal to 1, we will call it a VaR-violation.

To backtest a given sequence of VaR-violations, Christoffersen (1998) state three desirable properties that the VaR-violation process should possess. First, the VaR-violations are said to have unconditional coverage (uc thereafter) if the probability of a VaR-violation is equal to p, i.e.,

$$\mathbb{P}[I_t(p) = 1] = \mathbb{E}[I_t(p)] = p. \tag{2}$$

Second, the independence (ind thereafter) property requires that the variable $I_t(p)$ has to be independent of $I_{t-k}(p), \forall k \neq 0$. Finally, the uc and ind properties are combined via $\mathbb{E}[I_t(p) - p|\Omega_{t-1}] = 0$ to the property of conditional coverage (cc thereafter). In detail, a sequence of VaR-forecasts is defined to have correct cc if

$$\{I_t(p)\} \stackrel{i.i.d.}{\sim} \text{Bern}(p), \forall t. \tag{3}$$

While we agree with the formulation of the cc property, we point out that the uc and the ind properties as defined above suffer from some serious restrictions. The uc property requires a test whether the expected coverage is p for each day t individually. To be more precise, the equation $\mathbb{P}[I_t(p) = 1] = \mathbb{E}[I_t(p)] = p$ holds only true if $\mathbb{P}[I_t(p) = 1] = p$ holds for all t. However, it is not feasible to verify if this assumption holds true for all t individually by means of a statistical test of uc. Moreover, it is quite likely that the sequence of VaR-violations is not stationary and that the actual p varies across different market phases even if $\frac{1}{n}\sum_{t=1}^{n} I_t$ equals p for the total sequence.

Evidence for this conjecture is found by Escanciano and Pei (2012). The practical relevance of this feature is demonstrated in our empirical study (see Section 4). Consequently, we redefine the uc property simply as

$$\mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n} I_t(p)\right] = p. \tag{4}$$

With respect to the ind property, it is interesting to note that the current state-of-the-art backtests in the financial econometrics literature completely neglect the necessity to test the property of VaR-violations being identically distributed. In fact, the sequence $\{I_t(p)\}$ could exhibit clusters of violations while still possessing the property of independence as defined above. In fact, unexpected temporal occurrences of clustered VaR-violations may have several potential reasons. On the one hand, $\{I_t(p)\}$ may be not identically distributed and p could vary over time. On the other hand, $I_t(p)$ may not be independent of $I_{t-k}(p), \forall k \neq 0$. We therefore reformulate the ind property as the i.i.d. property (i.i.d. thereafter). The hypothesis of i.i.d. VaR-violations holds true if

$$\{I_t(\tilde{p})\} \overset{i.i.d.}{\sim} \text{Bern}(\tilde{p}), \forall t, \tag{5}$$

where $\tilde{p}$ is an arbitrary probability.

In the following, we describe our new set of backtests that includes separate tests for all mentioned properties of VaR-violation processes. Pseudocode for all our new backtests is given in Appendix A.

2.2 A New Test of Unconditional Coverage

At this point, we are interested in testing the null hypothesis $\mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n} I_t(p)\right] = p$ against the alternative $\mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n} I_t(p)\right] \neq p$. In fact, as we will see later, our new test statistic also allows us to separately test against the alternatives $\mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n} I_t(p)\right] \geq p$ and $\mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n} I_t(p)\right] \leq p$. The most intuitive and commonly used test statistic for the test of uc is given by (see Christoffersen, 1998):

$$LR_{uc}^{kup} = -2\log[L(p; I_1, I_2, ..., I_n)/L(\tilde{p}; I_1, I_2, ..., I_n)] \overset{asy}{\sim} \chi^2(1), \tag{6}$$

5

where $\tilde{p} = \frac{n_1}{n_1+n_0}$, n_1 is the number of violations and $n_0 = n - n_1$. Moreover, we have

$$L(p; I_1, I_2, ..., I_n) = p^{n_1}(1 - p)^{n_0} \tag{7}$$

and

$$L(\tilde{p}; I_1, I_2, ..., I_n) = \tilde{p}^{n_1}(1 - \tilde{p})^{n_0}. \tag{8}$$

Candelon et al. (2011) recently introduced an alternative test for the uc hypothesis us-
ing orthonormal polynomials and the GMM test framework proposed by Bontemps (2006),
Bontemps and Meddahi (2005) and Bontemps and Meddahi (2012). Their test statistic is given
by

$$J_{uc} = J_{cc}(1) = \left(\frac{1}{\sqrt{m}} \sum_{i=1}^{m} M_1(d_i; p)\right)^2 \overset{asy}{\sim} \chi^2(1), \tag{9}$$

where M_1 is an orthonormal polynomial associated with a geometric distribution with a success
probability p and d_i denotes the duration between two consecutive violations (see Candelon et al.,
2011, for more details).

However, both tests suffer from significant drawbacks. First, without modifications, it is not
possible to construct one-sided confidence intervals and to test separately if a VaR-estimation is
too conservative or underestimates the actual risk exposure. Such an additional feature, on the
other hand, would be of particular interest to bank regulators and risk-averse investors who are
primarily interested in limiting downside risk. The second drawback is concerned with the be-
haviour of the tests in finite samples. As we deal with tail forecasts based on binary sequences,
the number of violations is comparatively small and discrete. Hence, ties between the sample
test value and those obtained from Monte Carlo simulation under the null hypothesis need to be
broken. Christoffersen and Pelletier (2004) propose to use the Dufour (2006) Monte Carlo test-
ing technique to break ties between test values. As their approach, however, is computationally
demanding and unnecessarily complex, we propose a different tie breaking procedure.

We address the latter problem by exploiting an idea used, among others, by

6

Podolskij and Ziggel (2009) and propose to use the test statistic

$$MCS_{uc} = \sum_{t=1}^{n} I_t(p) + \epsilon, \tag{10}$$

where ϵ is a small and continuously distributed random variable that serves to break ties between test values.[6] Critical values of the test statistic are computed via Monte Carlo simulations (MCS) as is done for all other backtests throughout this paper. For fixed n and p, the distribution of the test statistic is known. We then simulate a large number of realizations of the test statistic under the respective null hypothesis and use the resulting quantile for testing the uc hypothesis. Adding the random variable ϵ guarantees that the test exactly keeps its size if the number of Monte Carlo simulations for obtaining the critical value tends to infinity.[7] Note that without the addition of the random variable ϵ, the test statistic would have a discrete distribution and not all possible levels could be attained. Additionally, note that the choice of ϵ is not crucial for testing the uc hypothesis. Consequently, it is intuitive to use normally distributed random variables for ϵ. Nevertheless, one needs to assure that the test statistic for $v - 1$ violations is smaller then the test statistic for v violations. Followingly, we set $\epsilon \sim 0.001 \cdot N(0, 1)$ in our simulation study. Finally, it is instructive to see that our new approach allows for one-sided and two-sided testing for every desired test level.

Critical values for all our tests are then computed via MCS instead of, e.g., making use of explicit expressions of the exact or asymptotic distributions. Basically, all test statistics we consider are given as a sum of a discrete random variable (determined by Bernoulli distributed random variables) and a continuous random variable with known distribution that is independent from the discrete random variable. Thus, on the one hand, the distributions of the test statistics are uniquely determined for fixed n and p and additionally it is basically useful to consider MCS. On the other hand, due to the continuous part, the test statistics are also continuously distributed. This follows from the general fact that, for a discrete random variable X with support M_X and a continuous

---

[6]   Podolskij and Ziggel (2009) employ the idea of adding a small random variable to a test statistic to construct a
       new class of tests for jumps in semimartigale models.
[7]   The theoretical foundation of our approach is given by Dufour (2006) who considers a more general context and
       solves this problem by introducing randomized ranks according to a uniform distribution.

random variable Y such that X and Y are independent,

$$P(X + Y \leq a) = \sum_{x \in M_X} P(x + Y \leq a | X = x) P(X = x) = \sum_{x \in M_X} P(Y \leq a - x) P(X = x).$$

Thus, the cumulative distribution function of $X + Y$ can be written as a countable sum of continuous functions so that it is continuous as well. Using a result from Dufour (2006), the empirical critical values then yield a test that exactly keeps its size if the number of MCS tends to infinity.[8]

2.3 A New Test of I.I.D. VaR-Violations

As stated in Christoffersen (1998), testing solely for correct uc of a VaR-model neglects the possibility that violations might cluster over time. Consequently, Christoffersen (1998) propose a test of the violations being independent against an explicit first-order Markov alternative. The resulting test statisic is given by:

$$LR_{iid}^{mar} = -2 \log[L(\tilde{\Pi}_2; I_1, I_2, ..., I_n)/L(\tilde{\Pi}_1; I_1, I_2, ..., I_n)] \overset{asy}{\sim} \chi^2(1). \tag{11}$$

Here, the likelihood functions are given by:

$$L(\tilde{\Pi}_1; I_1, I_2, ..., I_n) = \left(1 - \frac{n_{01}}{n_{00} + n_{01}}\right)^{n_{00}} \left(\frac{n_{01}}{n_{00} + n_{01}}\right)^{n_{01}} \left(1 - \frac{n_{11}}{n_{10} + n_{11}}\right)^{n_{10}} \left(\frac{n_{11}}{n_{10} + n_{11}}\right)^{n_{11}} \tag{12}$$

and

$$L(\tilde{\Pi}_2; I_1, I_2, ..., I_n) = \left(1 - \frac{n_{01} + n_{11}}{n_{00} + n_{10} + n_{01} + n_{11}}\right)^{n_{00} + n_{10}} \left(\frac{n_{01} + n_{11}}{n_{00} + n_{10} + n_{01} + n_{11}}\right)^{n_{01} + n_{11}}, \tag{13}$$

---

[8] Instead of using MCS, one could basically also derive the exact distribution functions of the test statistics, although this would indubitably be a cumbersome task. It would also be possible to derive asymptotic results if the test statistics are appropriately standardized (for example by $1/n$) and if one imposes additional moment assumptions on the continuous random variable. However, this is not necessary in our setting as in practice, n and p are fixed. Since one typically deals with a low number of VaR violations, one could moreover expect the asymptotic approximation to be highly inaccurate, which is confirmed by several studies (see, e.g., Berkowitz et al., 2011).

where n_{ij} is the number of observations with value i followed by j. Note that this first-order Markov alternative has only limited power against general forms of clustering. Moreover, as shown in Christoffersen and Pelletier (2004), this test is not suited for several settings and has a poor behaviour in finite samples. The test can then be combined with the test of uc presented in the previous subsection to yield a full test of conditional coverage. Despite the aforementioned shortcomings, however, it is still one of the most frequently used backtests in practice (see Candelon et al., 2011).

In a subsequent work, Christoffersen and Pelletier (2004) introduce more flexible tests which are based on durations between the violations. The intuition behind these tests is that the clustering of violations will induce an excessive number of relatively short and long no-hit durations. Under the null hypothesis, the no-hit durations D should then be exponentially distributed with

$$f_{exp}(D; p) = pe^{-pD}, \tag{14}$$

where D is the no-hit duration. In their work, Christoffersen and Pelletier (2004) employ the Weibull and the gamma distribution to test for an exponential distribution of the no-hit durations. Nevertheless, we will only consider the Weibull test in our simulation study as it yields considerably better results than the gamma test (see Haas, 2005). In addition to the mentioned tests, the literature on VaR backtesting also includes the standard Ljung-Box test, the CAViaR test of Engle and Manganelli (2004) and spectral density tests. However, the level of most of these tests is poor for finite samples and therefore critical values need to be calculated based on the Dufour Monte Carlo testing technique (see Berkowitz et al., 2011).

Recently, Candelon et al. (2011) introduced a new test for the i.i.d. hypothesis. As described above, this test is based on orthonormal polynomials and the GMM test framework. The test statistic is given by

$$J_{iid}(q) = \left(\frac{1}{\sqrt{m}} \sum_{i=1}^{m} M(d_i; \tilde{p}) \right)^T \left(\frac{1}{\sqrt{m}} \sum_{i=1}^{m} M(d_i; \tilde{p}) \right) \overset{asy}{\sim} \chi^2(q), \tag{15}$$

where $M(d_i; \tilde{p})$ denotes a $(q, 1)$ vector whose components are the orthonormal polynomials $M_j(d_i; \tilde{p})$, for $j = 1, ..., q$, evaluated for the true violation rate $\tilde{p}$.

To introduce our new test statistic, we first define the set of points in time on which a VaR-violation occurs via

$$V = \{t | I_t = 1\} = (t_1, ..., t_m). \tag{16}$$

The test statistic for our new i.i.d. hypothesis is then given by

$$MCS_{iid,m} = t_1^2 + (n - t_m)^2 + \sum_{i=2}^{m}(t_i - t_{i-1})^2 + \epsilon. \tag{17}$$

Basically, the idea behind this test statistic follows the principle of the Run-Test proposed by Wald and Wolfowitz (1940). To be more precise, the sum of the squared durations between two violations is minimal if the violations are exactly equally spread across the whole sample period. If the violations are clustered and occur heaped, this sum increases. Additionally, $MCS_{iid,m}$ also allows to test for a too systematic occurrence of violations. For example, the process of VaR-violations could exhibit an undesirable cyclical or seasonal behaviour that is detected by our new test of the i.i.d. property.[9] However, for the purpose of this study we concentrate on testing for clustered VaR-violations.

As before, we waive a formal derivation of the distribution of our test statistic. Instead, we obtain the critical values of the test statistic by means of a Monte Carlo simulation (thus inspiring the abbreviation $MCS_{iid,m}$). The simulation is straightforward as only n and p have to be adapted to the specific situation. Note that the critical values need to be simulated separately for each value of m as we are solely interested in the durations between the violations and not in the absolute number of it. Again, we use a small continuously distributed random variable ϵ to break ties and for the MCS to yield a valid test. Moreover, the computational complexity of the test is negligible.

---

[9]   This feature is of particular interest, e.g., in commodity and weather risk management.

2.4 A New Test of Conditional Coverage

We now describe our new test of conditional coverage that combines the two new tests for the uc and the i.i.d. property. Starting point is again the standard test of conditional coverage as proposed by Christoffersen (1998) which utilizes the test statistic

$$LR_{cc}^{mar} = -2\log[L(p; I_1, I_2, ..., I_n)/L(\tilde{\Pi}_1; I_1, I_2, ..., I_n)] \stackrel{asy}{\sim} \chi^2(2),\tag{18}$$

and which is based on the first-order Markov alternative described above. In a related study, Berkowitz et al. (2011) extend their Weibull test for the i.i.d. property and derive an alternative test of conditional coverage. They postulate a Weibull distribution for the duration variable D with distribution

$$h(D; a, b) = a^b b D^{b-1} e^{-(aD)^b},\tag{19}$$

with $\mathbb{E}[D] = 1/p$. Then, the null hypothesis of their test of conditional coverage is given by

$$H_{0,cc} : b = 1, a = p.\tag{20}$$

Using orthonormal polynomials and the GMM test framework, Candelon et al. (2011) propose a competing test of the cc hypothesis. Their test statistic is given by

$$J_{cc}(q) = \left(\frac{1}{\sqrt{m}} \sum_{i=1}^{m} M(d_i; p)\right)^T \left(\frac{1}{\sqrt{m}} \sum_{i=1}^{m} M(d_i; p)\right) \stackrel{asy}{\sim} \chi^2(q).\tag{21}$$

Again, $M(d_i; p)$ denotes a $(q, 1)$ vector whose entries are the orthonormal polynomials $M_j(d_i; p)$, for $j = 1, ..., q$.

A drawback of both mentioned backtests is that neither allows for a weighting of the influence of the unconditional and i.i.d. tests in the combined test of conditional coverage. From the perspective of a risk manager, however, such a feature could be highly desirable as more weight could be assigned to one of the components of the test of conditional coverage. Hence, we are interested

in a test of the form

$$MCS_{cc,m} = a \cdot f(MCS_{uc}) + (1 - a) \cdot g(MCS_{iid,m}), 0 \le a \le 1, \tag{22}$$

where a is the weight of the test of uc in the combined cc test. The first component of our new cc test is then given by

$$f(MCS_{uc}) = \left| \frac{(\epsilon + \sum_{t=1}^{n} I_t)/n - p}{p} \right|. \tag{23}$$

This term measures (in percent) the deviation between the expected and observed proportion of violations. To allow for a one-sided testing within the uc component, the above term is multiplied by $1_{\{\sum_{t=1}^{n} I_t/n \ge p\}}$ or $1_{\{\sum_{t=1}^{n} I_t/n \le p\}}$, respectively.

The second component in the cc test in (22) is defined as

$$g(MCS_{iid,m}) = \frac{MCS_{iid,m} - \hat{r}}{\hat{r}} \cdot 1_{\{MCS_{iid,m} \ge \hat{r}\}}, \tag{24}$$

where $\hat{r}$ is an estimator of the expected value of the test statistic $MCS_{iid,m}$ under the null hypothesis, i.e., for $E(MCS_{iid,m}|H_0) =: r$ (see below and Appendix A for details). The second component measures the deviation (in percent) between the expected and observed sum of squared durations. Again, we use random variables ϵ to break ties. In line with the new uc and i.i.d. tests, we abstain from a formal derivation of the distribution of our test statistic and obtain the critical values by means of a Monte Carlo simulation for each combination of sample size n and weighting factor a.

Note that the estimator $\hat{r}$ is calculated in a prior step before calculating the actual test statistics and deriving critical values (cf. the pseudocode in Appendix A). Thus, for $MCS_{cc,m}$, the arguments regarding the correctness of the MCS from the end of Section 2.2 are also applicable.

As the weighting factor a can be chosen arbitrarily, a natural question to ask is how a should be chosen. On the one hand, small test samples (e.g., 250 days) and small values of p (e.g. $p = 1\%$) lead to a small expected number of VaR-violations. In these cases, a risk manager (or regulator) might be more interested in backtesting the VaR-violation frequency rather than the i.i.d. property

12

of, for instance, only two or three violations. On the other hand, large test samples (e.g., 1,000 days) may include calm bull and volatile bear markets. A VaR-model which is not flexible enough to adapt to these changes may lead to non-identically distributed VaR-violations while at the same time yielding a correct uc. Therefore, risk managers could be inclined to select a lower level of a to shift the sensitivity of the cc test to the test of the i.i.d. property. The selection of the optimal weighting facor a thus seems to be a practical rather than a mathematical task.

2.5 Multivariate Extension

In the last part of this section, we follow Christoffersen (1998) and Berkowitz et al. (2011) and shortly discuss the backtesting of multivariate VaR-forecasts. The extension of our newly proposed backtests to the multivariate case is straightforward and we again exploit the fact that our test statistics are considerably less computationally demanding than competing testing approaches (thus enabling the use of MCS). Here, we concentrate on exemplifying the extension of the test of uc to a multivariate setting (the extension of the other proposed tests works analogously). Assume that we are given a sample of an h-variate time series $\{Y_t\}_{t=1}^{n}$ as well as h sequences of VaR-forecasts, $VaR_{t,i|t-1}(p)$, and VaR-violations, $I_{t,i}(p)$, $i = 1, ..., h$, respectively. Then, the multivariate two-sided test for the uc hypothesis is defined by

$$MCS_{m-uc,h} = \epsilon + \sum_{i=1}^{h} \left| \frac{(\sum_{t=1}^{n} I_{t,i})/n - p}{p} \right|. \tag{25}$$

$MCS_{m-uc,h}$ is simply the sum of the single deviations (in percent) between the expected and observed proportion of violations. Multiplying each summand by $1_{\{\sum_{t=1}^{n} I_{t,i}/n \geq p\}}$ or $1_{\{\sum_{t=1}^{n} I_{t,i}/n \leq p\}}$, respectively, again yields a directional test. Moreover, simulation is straightforward as only n, h and p need to be specified in an empirical application.

13

3 Simulation Study

To examine the performance of our newly proposed backtests in finite samples, we perform a comprehensive simulation study in which we compare our new backtests to several different benchmark tests.

3.1 Tests of Unconditional Coverage

We analyze the performance of the different tests of uc by simulating $10,000$ samples and using different parameter combinations for p, γ, and n to analyze the size and power of the backtests in more detail. We directly simulate sequences of VaR-violations using the data generating process (DGP)

$$I_t \sim \text{Bern}(\gamma \cdot p), \ t = 1, ..., n. \tag{26}$$

To determine the size of the tests, we set the coverage parameter $\gamma = 1.0$. For the analysis of the tests' power, we increase the violation probability and set $\gamma = 1.1, \ 1.25, \ 1.50$ and 2.00.[10] Each sequence I_t of simulated VaR-violations is then backtested using the new upper-tail MCS_{uc}^{ut} and the two-tailed MCS_{uc}^{tt} backtest as described in Section 2.2. To evaluate each test's power, we compute the fraction of simulations in which the test is rejected (hereafter referred to as rejection rate). Critical values of the test statistics for different parameters p and n are computed using $10,000$ MC simulations. Complementing our new backtests, we also apply the LR_{uc}^{kup} test of Christoffersen (1998) and the GMM_{uc} test of Candelon et al. (2011) to the simulated violation sequences and compare the results of the tests. The results of the simulation study on the performance of the tests of uc are presented in Table I.

- Insert Table I about here -

Not surprisingly, due to the fact that the critical values for each of the tests are determined via simulation, the rejection frequencies for the setting $\gamma = 1.0$ are close to the nominal size of the

---

[10]   We calculate but do not report results for the setting $\gamma < 1$ and concentrate on the more practically relevant scenario of a VaR-model underestimating risk.

tests. With respect to the power of the uc tests, the results of the LR_{uc}^{kup} test, the GMM_{uc} test, and the two-tailed MCS_{uc}^{tt} test are very similar. Only in a few cases do the results of the GMM_{uc} test deviate from the rejection rates of the LR_{uc}^{kup} test and the two-tailed MCS_{uc}^{tt} test in a positive or negative direction. However, all of the three analyzed two-tailed tests are outperformed by the one-sided MCS_{uc}^{ut} test in the vast majority of settings. Consequently, in addition to being of high practical relevance to regulators, our new directional test of uc offers an increased test power compared to standard VaR-backtests from the literature.

3.2 Tests of the I.I.D. Property

As discussed in Section 2.1, a correctly specified VaR-model should yield i.i.d. violations. In this part of the simulation study, we analyze the power of the new backtests of i.i.d. VaR-violations using two data generating processes. First, we investigate the power of our new backtests and competing benchmark tests using dependent violations. Second, we repeat this analysis for non-identically distributed violation processes. In both settings, we perform the MCS_{iid} test and compare its finite sample behavior to that of the LR_{iid}^{mar} test of Christoffersen (1998), the LR_{iid}^{wei} test of Christoffersen and Pelletier (2004) and the GMM_{iid} test of Candelon et al. (2011).[11] Because clustering implies the occurance of at least two VaR-violations, the i.i.d. tests are not performed on samples where this minimum number is not achieved. To be more precise, $\sum_{t=1}^{n} \dot{I}_t \geq 2$ holds true for each of the samples simulated by the procedures below, where $\dot{I}_t$ denotes a simulated VaR-violation sequence. Basically, each of the utilized tests are feasible under this condition. Only the LR_{iid}^{wei} test statistic cannot be computed for some simulated samples containing two violations (for more details see Candelon et al., 2011). We classify these cases as *not rejected*.

---

[11]   As suggested in Candelon et al. (2011) we set $q = 3$ for $p = 5\%$ and $q = 5$ for $p = 1\%$ throughout the simulation study. Critical values for the MCS_{iid} test are obtained as outlined in Section 2.3 using $10,000$ MC simulations.

3.2.1 Independent VaR-Violations

In the first setting, we generate sequences of dependent VaR-violations with the degree of dependence inherent in the violation processes varying over time. For each λ and each n, we draw $10,000$ simulations from the distribution

$$\dot{y}_t = \begin{cases} \sim z_t, t = 1; \\ \sim \sigma_t z_t, t > 1, \end{cases} \tag{27}$$

where

$$\sigma_t^2 = \lambda \sigma_{t-1}^2 + (1 - \lambda) z_{t-1}^2, 0 \leq \lambda \leq 1, \tag{28}$$

and $z_t \sim N(0, 1), \forall t$. The distribution of $\dot{y}_t$ is based on the well-known exponentially weighted moving average (EWMA) type process. This approach allows for an easy regulation of the degree of dependence by determining λ as the single decay factor. To be more precise, λ controls the half-life interval of the observation weights (i.e., the interval in which the weight of an observed σ^2 decreases to half its original value) by $log(0.5)/log(\lambda)$. We apply the backtests to several different levels of λ representing half-life intervals of 5, 10, 20, 40, 60, and 80 days of data. This range of half-life intervals covers typical volatility persistence of asset return series.[12] Table II shows the half-life intervals and the corresponding λ level used to compute the power of the backtests.

- Insert Table II about here -

Dependent VaR-violations are ensured by setting a constant VaR for all $i = 1, \ldots, n$. For each decay factor λ the VaR is determined separately by the empirical p-quantile of $10,000$ random values simulated by Equation (27). The simulated VaR-violations $\dot{I}_t$ are computed as defined by Equation (1).

---

[12] The EWMA approach can be used for VaR-forecasting purposes (RiskMetrics) whereas λ is typically set to 0.94 for one-day and 0.97 for one-month forecasts (see Mina et al., 2001). This corresponds to half-life intervals of 11 and 23 days. Furthermore, Berkowitz et al. (2011) estimated variance persistences for actual desk-level daily P/Ls from several business lines from a large international bank. The determined values are 0.9140, 0.9230, 0.9882 and 0.9941 which correspond to half-life intervals of 8, 9, 58, and 117 days.

Table III shows the results of the power study concerning the independence property of VaR-violations. We apply each test to 60 different combinations of coverage probability p, decay factor λ and sample size n. Together with the three significance levels of 1%, 5%, and 10%, we thus obtain 180 different settings in our simulation study.

- Insert Table III about here -

In total, the MCS_{iid} test outperforms the remaining tests in 104 out of the 180 test settings. Compared to the other test methods, this test possesses a high statistical power in settings in which the half-life interval is relatively large. Furthermore, the superiority of the MCS_{iid} test increases with the significance level. The GMM test shows the best statistical power in almost one third of the considered settings. Compared to the remaining tests, the test performs well particularly for half-life intervals up to 20 days and for small significance levels. The LR_{iid}^{mar} test yields the best statistical power in 21 out of 150 settings. This result should be interpreted cautiously due to the fact that the vast majority of the top results are concentrated at the very short half-life interval of five days. It is to be expected that the LR_{iid}^{mar} test performs well in such circumstances, because short decay intervals lead to frequent occurrences of successive VaR-violations. Consequently, the power of this test deteriorates as the decay interval increases. For none of the 180 different settings does the LR_{iid}^{wei} test lead to the best statistical power of all analyzed test methods. Furthermore, for $p = 5\%$ and a half-life interval larger than 10 days, the test yields a statistical power below its nominal size and shows the undesired behavior of decreasing rejection rates as the sample size increases.

3.2.2 Identically Distributed VaR-Violations

The data generating process for the second part of the simulation study is given by:

$$\dot{I}_t = \begin{cases} \overset{i.i.d.}{\sim} \operatorname{Bern}(p - 2\delta), 1 \le t \le \frac{n}{4}; \\[2mm] \overset{i.i.d.}{\sim} \operatorname{Bern}(p + \delta), \frac{n}{4} < t \le \frac{n}{2}; \\[2mm] \overset{i.i.d.}{\sim} \operatorname{Bern}(p - \delta), \frac{n}{2} < t \le \frac{3n}{4}; \\[2mm] \overset{i.i.d.}{\sim} \operatorname{Bern}(p + 2\delta), \frac{3n}{4} < t \le n. \end{cases} \tag{29}$$

Here, we choose $\delta = 0p$ to analyze the size of a test and $\delta = 0.1p$, $0.2p$, $0.3p$, $0.4p$ and $0.5p$ for the power study. This setting leads to variations in the probability of obtaining a VaR-violation between the four equal-sized subsamples. Consequently, the violations will occur un-equally distributed. Note that the probability variations are determined in a way which ensures $\mathbb{E}\left(\sum_{t=1}^{n} \dot{I}_t\right) = n \cdot p$. The setup of this part of the simulation study covers a realistic scenario in which a VaR-model does not, or not fully, incorporate changes from calm market phases to highly volatile bear markets or financial crises and vice versa. This in turn leads to clustered VaR-violations re-gardless of the question whether the data might show signs of autocorrelation.

Alternatively, non-stationary VaR-violations could be identified by splitting a sample into sev-eral subsamples and applying the test for uc to each subsample. However, this approach suffers from two main drawbacks. First, for small subsamples the power of uc tests is relatively low (see Table I). Second, it remains unclear at which points real data samples have to be split into two or more subsamples.

Table IV shows the results of the power study concerning the property of identically distributed VaR-violations. We apply each test to 50 different combinations of coverage probability p, prob-ability variation factor δ, and sample size n. Furthermore, we compute rejection rates for signifi-cance levels of 1%, 5%, and 10% which leads to a total of 150 different test settings.

- Insert Table IV about here -

18

In total, the MCS_{iid} test possesses a high statistical power regarding non-identically distributed VaR-violations and its test results are comparable to or better than the performance of the remaining three approaches for 130 out of the 150 settings. Particularly for significance levels of 5% and 10%, it outperforms the competing tests in almost all cases, irrespective of the degree of probability variation or sample size. The GMM test yields rejection rates which are equal or better than the results of the competing models for 30 of the 150 simulation settings. The test particularly achieves its top results for a significance level of 1%. The LR_{iid}^{mar} test is able to match the results of the competing tests in only seven cases which are restricted to settings in which $p = 1\%$ and $\delta = 0.1p$. The results of the LR_{iid}^{wei} test falls short of the performance of the remaining tests in almost all settings. Finally, it is striking that the power of the LR_{iid}^{mar} test and the LR_{iid}^{wei} test significantly exceed the nominal size only for large shifts in the VaR-violation probability, i.e. $\delta \geq 0.4p$.

3.3 Conditional Coverage

Table VII illustrates the behavior of the MCS_{cc} test considering different levels of the weighting parameter a.

- Insert Table VII about here -

For reasons of space we present results only for a single parameter combination for each of the two settings. This includes $n = 1000$, a half-life interval of 20 days, and $\gamma = 1.25$ for setting 1 and $n = 1000$, $\delta = 0.3p$, $\gamma = 1.25$ for setting 2. Depending on the setting, the VaR probability p, and the significance level, the test yields the highest rejection rates for values of a between 0.5 and 0.8. This is consistent with our expectation that the maximum of the statistical power is achieved when $0 < a < 1$, i.e., when the cc test addresses both the unconditional coverage as well as the i.i.d. property of the violations.

We continue with a comparison of the size and the power of the conditional coverage test MCS_{cc} to the LR_{cc}^{mar} test of Christoffersen (1998), the LR_{cc}^{wei} test of Christoffersen and Pelletier (2004) and the GMM_{cc} test of Candelon et al. (2011). For this purpose, we combine each of the

two settings described in Section 3.2 with increased probabilities of a VaR-violation outlined in Section 3.1. Note that we use the two-tailed uc component and set the weighting factor to $a = 0.5$ in the entire study. For the determination of critical values we perform the procedure as explained in Section 2.4 using $10,000$ MC simulations. In line with the settings above, for each combination of γ, δ, volatility half-life, and n we repeat the simulation of VaR-violation sequences $10,000$ times. We present the results of the simulation study concerning an increased probability of a VaR-violation combined with non-independent occurrence of violations (setting 1) in Table V, and combined with non-identically distributed violations (setting 2) in Table VI. [13]

- Insert Tables V and VI about here -

Regarding both settings, the MCS_{cc} test yields the best rejection rates for the vast majority of test settings. To be precise, the MCS_{cc} test shows similar or better results compared to the competing tests in 157 out of 180 parameter combinations for setting 1 and 116 out of 150 parameter combinations for setting 2. With respect to setting 1, the LR_{cc}^{mar} test and the GMM_{cc} test achieve or exceed the rejection rates of the MCS_{cc} test in some cases in which the nominal VaR-level is set to 1%. The LR_{cc}^{wei} test does not achieve top rejection rates for any of the parameter combinations. Regarding setting 2, and parameter combinations for which the VaR-violation probability variation parameter is set to $\delta = 0.1p$, the LR_{uc}^{mar} test shows some superior results. In many cases, the rejection rates of the GMM_{cc} test show evidence of a good performance, but only in very few cases does it yield top results. For none of the reported parameter combinations does the LR_{cc}^{wei} test lead to results above the rejection rates of the remaining tests.

4   Empirical Application

To investigate the behavior of the new set of backtests and to illustrate their usefulness in a realistic risk management setting, we perform an empirical study using actual returns on a set of managed portfolios.

---

[13]   To save space, we do not present the rejection rates of all parameter combinations. The complete results are available from the authors upon request.

4.1 Data and Forecasting Scheme

We apply the new tests to a unique data set provided by a German asset manager.[14] The data set consists of 5,740 daily log-returns for each of four portfolios and covers a time period of 22 years (January 1, 1991 to December 31, 2012). While we exclude weekend days from our sample, it is not possible to easily eliminate holidays as well, because the portfolio assets are invested internationally and non-business days differ widely across the countries in our sample. To this end, we add the returns of these days (e.g., accrued interest) to the next trading day. Table VIII presents summary statistics for the portfolio log-returns we use in our empirical study.

- Insert Table VIII about here -

The summary statistics in Table VIII show evidence of the usual stylized facts of returns on financial assets. In addition to having negligible (daily) mean returns, the portfolio returns exhibit signs of typical properties like negative skewed and leptokurtic asset returns indicating fat tails particularly on the downside. Nevertheless, overall portfolio risk over the complete sample period appears to be only moderate as evidenced by the estimates of the (unconditional) return series volatility with all four portfolios having significant positive annualized returns.

We calculate the one-day VaRs for each portfolio by the use of two different VaR-models. First, we choose standard historical simulation as the most widely used model in practice (see Pérignon and Smith, 2010). This concept assumes no particular distribution of the returns. The VaR is rather estimated solely based on historical returns. For each VaR-estimation, we use the value of the 1% and 5% quantile of the last 250 data points as an estimate for the portfolio's VaR. Second, we employ a GARCH(1,1) process as a parametric model to forecast the VaR using the estimated conditional variance of the GARCH model. Compared to historical simulation, the GARCH model is more flexible because it accounts for autocorrelations in the return series' variances. We choose the simple GARCH(1,1) model rather than more sophisticated GARCH model specifications because Hansen and Lunde (2005) show that the GARCH(1,1) model is hard to beat

---

[14]   Due to confidentiality reasons, the asset manager wishes to remain anonymous.

in terms of volatility forecasting accuracy. For the sake of simplicity, we fit the GARCH param-eter for each portfolio separately to the total sample of 5,740 log-returns.[15] The next-day VaR is then calculated simply by the quantile of a normal distribution with a zero mean and the standard deviation forecasted by the GARCH model on the basis of the last 250 log-returns. Figure 1 plots the daily portfolio returns together with the corresponding VaR-forecasts of the historical VaR and the GARCH model.

- Insert Figure 1 about here -

In addition to the time-varying volatility of the returns, the charts illustrate the differences in the forecasts of the unconditional historical VaR approach and the conditional GARCH model. How-ever, it can be seen for both models that the VaR-violations cluster to some degree during certain subsamples.

After calculating the VaR-violation sequence $I_t(p)$, we validate the VaR-estimation by making use of the new set of MCS tests to compute p-values and check the uc, i.i.d. and cc hypotheses separately. With respect to the MCS_{cc} test, we use the two-tailed uc component and opt for a weighting factor of $a = 0.5$. For comparison purposes, we additionally present p-values of the uc, i.i.d., and cc version of the GMM test as the results of our simulation study indicate that the set of GMM tests is a suitable benchmark. Moreover, we repeat our analysis for four separate time periods. For the first time period, we include 5,740 log-returns of the whole available time span (January 1, 1991 - December 31, 2012). We then focus on a distinct market phase around the volatility shift from the highly volatile bear market at the later stage of the dotcom-bubble burst (250 log-returns from April 16, 2002 to March 31, 2003) to the early stage of the subsequent calm bull market (250 log-returns from April 1, 2003 to March 15, 2004). Additionally, we apply the tests to the 500 log-returns of the combination of the latter two periods from April 16, 2002 to March 15, 2004.

---

[15]   Of course, this procedure does not comply to the principle of out-of-sample forecasting. Nevertheless, as we focus on the performance of the backtests, the issue of optimally fitting the GARCH parameters to the data is not relevant for the purpose of this study.

4.2 Results

The results of applying the backtests to the total period data set are shown in Table IX.

- Insert Table IX about here -

First, we compute the VaR-violation ratios of each portfolio for each VaR-forecasting method and the nominal VaR levels of 5% and 1%. We define the VaR-violation ratio as the VaR-violation frequency divided by the number of VaR-forecasts. Both the historical VaR and the GARCH approach lead to VaR-violation ratios which deviate from the nominal VaR level of 5% and 1% to some degree. The p-values of the one-tailed MCS_{uc}^{lt} and MCS_{uc}^{ut} tests indicate that each of these deviations are statistically significant. However, some of the p-values yielded by the two-tailed MCS_{uc}^{tt} and the GMM_{uc} tests remain above the 10% significance level.

The MCS_{iid} test and the GMM_{iid} test reject the i.i.d. hypothesis for the violation sequences generated by the historical simulation VaR-model for the 5% and 1% VaR level. We expect a large sample like ours that consists of 22 years of data to suffer significantly from the stylized facts of financial returns (i.e., series of absolute or squared returns show profound serial correlation, volatility appears to vary over time, and extreme returns appear in clusters). Consequently, an inflexible and unconditional VaR-model like historical simulation should lead to non-i.i.d. VaR-violations. However, the p-values for the more flexible GARCH model suggest clustered VaR-violations only for the 5% VaR level. These findings are confirmed by significant p-values obtained for the MCS_{cc} and GMM_{cc} tests.

The test results for the bear and the bull market as well as for the combination of both market phases are reported in Table X. We restrict the presentation of the results to the VaR level of 5%, because it vividly illustrates the effects of a shift from a bear to a bull market.

- Insert Table X about here -

The differences in the VaR-violation ratios between the bear and the bull market are significant. On average, for the bear market the historical VaR approach yields VaR-violations in 8.45% of the

days whereas for the bull market the ratio amounts to 1.70%. Consequently, for both the bear and the bull market, the p-values of the relevant one-sided and the two-sided MCS_{uc} tests as well as the GMM_{uc} test are statistically significant in the vast majority of cases. With respect to the combined 500 trading days sample, the underestimated VaR of the bear market and the overestimated VaR of the bull market compensate each other and lead to an average VaR-violation ratio of 5.08%. Because this is very close to the nominal VaR level of 5%, all applied backtests imply a correct uc. This result underpins our redefinition of the uc property, because the backtests show no significant p-values although the probability for a VaR-violation is not equal to the nominal level p for all days t.

The i.i.d. tests show a remarkable behavior. Because the GARCH model accounts for auto-correlated volatility, it can be assumed that the VaR-violations are less dependent compared to the VaRs estimated with historical simulation. Consequently, the p-values regarding the GARCH model during the bear market and the bull market separately are statistically significant in only four out of twelve cases. These results are contrasted by the p-values for the sample where the bear and bull market are combined. Here, the i.i.d. tests attain p-values below the 1% level of significance in six out of eight cases. This result could be due to the large shift in the VaR-violation ratio. Only the p-values for Portfolio two reveal no significance which can be explained by a smaller drop of the violation ratio from the bear to the bull market compared to the remaining portfolios. This outcome demonstrates the necessity of testing the independence as well as the identical distribution hypothesis using a powerful test. Finally, the results of the cc tests reflect the implications of the corresponding uc and i.i.d. tests.

5 Conclusion

Comparatively little attention has been paid in the literature to the development of proper tools for backtesting VaR-forecasts. This paper provides three main contributions to the issue of backtesting the performance of VaR-models. First, we extend the discussion of the desirable properties

of violations originating from a correct VaR-model and restate the unconditional coverage property of a VaR-violation process. Furthermore, we stress the need to require the VaR-violations to be identically distributed to adequately backtest models across different market phases.

Second, we propose a new set of backtests that test VaR-violation processes for unconditional coverage, the i.i.d. property as well as conditional coverage. Compared to existing standard approaches, these backtests contain new desirable features like directional testing for unconditional coverage and a test of conditional coverage that allows for different weightings of the uc and i.i.d. parts. The new backtesting procedures are based on i.i.d. Bernoulli random variables obtained by Monte Carlo Simulation techniques and are very intuitive.

Third, we perform a simulation study using generated VaR-violation samples that specifically violate the uc, i.i.d., and cc property to different controllable degrees. Compared to existing classical and state-of-the-art backtests, the new backtests outperform these benchmarks in several distinct settings. In addition, we use the new backtests in an empirical application study using a unique dataset consisting of four portfolio return series provided by a German asset manager. We apply the backtests to samples of calm boom and highly volatile bust cycles. The obtained results demonstrate the need for a backtest that accounts for non-identically distributed VaR-violations and, moreover, support the reformulation of the uc hypotheses.

References

ALEXANDER, C. (2008): *Market Risk Analysis, Value at Risk Models*, vol. 4, Chichester: John Wiley & Sons Ltd.

ARTZNER, P., F. DELBAEN, J.-M. EBER, AND D. HEATH (1999): "Coherent Measures of Risk," *Mathematical Finance*, 9, 203–228.

BASAK, S. AND A. SHAPIRO (2001): "Value-at-Risk-Based Risk Management: Optimal Policies and Asset Prices," *Review of Financial Studies*, 14, 371–405.

BASEL COMMITTEE ON BANKING SUPERVISION (BCBS) (1996): "Supervisory framework for the use of 'backtesting' in conjunction with the internal models approach to market risk capital requirements," `http://www.bis.org/publ/bcbs22.pdf`.

BERKOWITZ, J. (2001): "Testing Density Forecasts, with Applications to Risk Management," *Journal of Business and Economic Statistics*, 19, 465–474.

BERKOWITZ, J., P. CHRISTOFFERSEN, AND D. PELLETIER (2011): "Evaluating Value-at-Risk Models with Desk-Level Data," *Management Science*, 57, 2213–2227.

BONTEMPS, C. (2006): "Moment-based tests for discrete distributions," *Working Paper*.

BONTEMPS, C. AND N. MEDDAHI (2005): "Testing normality: A GMM approach," *Journal of Econometrics*, 124, 149–186.

——— (2012): "Testing distributional assumptions: A GMM approach," *Journal of Applied Econometrics*, 27(6), 978–1012.

CAMPBELL, S. (2007): "A review of backtesting and backtesting procedures," *Journal of Risk*, 9(2), 1–18.

CANDELON, B., G. COLLETAZ, C. HURLIN, AND S. TOKPAVI (2011): "Backtesting Value-at-Risk: A GMM Duration-based Test," *Journal of Financial Econometrics*, 9(2), 314–343.

CHRISTOFFERSEN, P. (1998): "Evaluating Interval Forecasts," *International Economic Review*, 39, 841–862.

CHRISTOFFERSEN, P. AND D. PELLETIER (2004): "Backtesting Value-at-Risk: A Duration-Based Approach," *Journal of Financial Econometrics*, 2(1), 84–108.

DANÍELSSON, J., B. N. JORGENSEN, M. SARMA, G. SAMORODNITSKY, AND C. G. DE VRIES (2005): "Sub-additivity Re-Examined: the Case for Value-at-Risk," Discussion paper 549, LSE Financial Markets Group.

DOWD, K. (1998): *Beyond Value at Risk: The New Science of Risk Management*, Wiley Frontiers in Finance, Chichester: John Wiley & Sons Ltd.

DUFOUR, J.-M. (2006): "Monte Carlo Tests with Nuisance Parameters : A General Approach to Finite-Sample Inference and Nonstandard Asymptotics in Econometrics," *Journal of Econometrics*, 133, 443–477.

ENGLE, R. AND S. MANGANELLI (2004): "CAViaR: Conditional Autoregressive Value-at-Risk by Regression Quantiles," *Journal of Business and Economic Statistics*, 22, 367–381.

ESCANCIANO, J. C. AND P. PEI (2012): "Pitfalls in backtesting Historical Simulation VaR models," *Journal of Banking & Finance*, 36, 2233–2244.

HAAS, M. (2005): "Improved duration-based backtesting of Value-at-Risk," *Journal of Risk*, 8(2), 17–36.

HANSEN, P. R. AND A. LUNDE (2005): "A forecast comparison of volatility models: does anything beat a GARCH (1, 1)?" *Journal of applied econometrics*, 20, 873–889.

JORION, P. (2006): *Value at Risk: The New Benchmark for Managing Financial Risk*, Chicago: McGraw-Hill Professional, 3 ed.

KUPIEC, P. (1995): "Techniques for Verifying the Accuracy of Risk Measurement Models," *Journal of Derivatives*, 3, 73–84.

LEIPPOLD, M., F. TROJANI, AND P. VANINI (2006): "Equilibrium impact of value-at-risk regulation," *Journal of Economic Dynamics and Control*, 30, 1277–1313.

MINA, J., J. Y. XIAO, ET AL. (2001): "Return to RiskMetrics: the evolution of a standard," *RiskMetrics Group*.

PÉRIGNON, C. AND D. R. SMITH (2010): "The level and quality of Value-at-Risk disclosure by commercial banks," *Journal of Banking & Finance*, 34, 362–377.

PODOLSKIJ, M. AND D. ZIGGEL (2009): "New tests for jumps in semimartigale models," *Statistical Inference for Stochastic Processes*, 13(1), 15–41.

SHIN, H. S. (2010): *Risk and Liquidity*, Clarendon Lectures in Finance, New York: Oxford University Press Inc.

WALD, A. AND J. WOLFOWITZ (1940): "On a Test Whether Two Samples are from the Same Population," *The Annals of Mathematical Statistics*, 11(2), 147–162.

Figures and Tables

Table I: Unconditional Coverage - Size and Power of Tests

The table presents rejection rates obtained by applying unconditional coverage tests to 10,000 samples of Bernoulli simulated VaR-violation sequences. The VaR level p for panel A and B is set to 5% and 1%, respectively. Results are presented for various sets of sample sizes n and γ-factors which multiplies the probability of a VaR-violation by 1, 1.1, 1.25, 1.5, and 2. The results for $\gamma = 1p$ correspond to the evaluation of the size of the test. LR_{uc}^{kup} and GMM_{uc} refers to the unconditional coverage tests of Kupiec (1995) and Candelon et al. (2011). MCS_{uc}^{tt} and MCS_{uc}^{ut} refer to the new two-tailed and upper-tail Monte-Carlo-Simulation based tests. Top results are highlighted in bold type.

		Significance level: 1%				Significance level: 5%				Significance level: 10%			
$y \cdot p$	n	LR_{uc}^{kup}	GMM_{uc}	MCS_{uc}^{tt}	MCS_{uc}^{ut}	LR_{uc}^{kup}	GMM_{uc}	MCS_{uc}^{tt}	MCS_{uc}^{ut}	LR_{uc}^{kup}	GMM_{uc}	MCS_{uc}^{tt}	MCS_{uc}^{ut}
Panel A: 5% VaR													
	252	0.010	0.010	0.009	0.009	0.049	0.049	0.049	0.049	0.100	0.099	0.100	0.100
	500	0.011	0.011	0.010	0.010	0.049	0.049	0.050	0.047	0.099	0.103	0.099	0.097
5.00%	1,000	0.010	0.010	0.012	0.012	0.054	0.050	0.055	0.053	0.106	0.099	0.105	0.102
	1,500	0.010	0.012	0.009	0.009	0.047	0.052	0.049	0.048	0.098	0.099	0.097	0.101
	2,500	0.009	0.009	0.010	0.012	0.048	0.048	0.050	0.051	0.106	0.101	0.102	0.102
	252	0.015	0.005	0.015	**0.024**	0.062	0.059	0.064	**0.102**	0.111	0.128	0.124	**0.178**
	500	0.022	0.010	0.023	**0.036**	0.075	0.068	0.080	**0.128**	0.144	0.133	0.147	**0.223**
5.50%	1,000	0.033	0.020	0.034	**0.059**	0.105	0.099	0.118	**0.180**	0.195	0.190	0.191	**0.289**
	1,500	0.047	0.030	0.045	**0.076**	0.134	0.127	0.140	**0.215**	0.227	0.216	0.221	**0.345**
	2,500	0.083	0.055	0.082	**0.126**	0.201	0.186	0.204	**0.306**	0.336	0.296	0.310	**0.445**
	252	0.047	0.011	0.045	**0.072**	0.137	0.120	0.146	**0.223**	0.203	0.223	0.230	**0.338**
	500	0.089	0.048	0.095	**0.143**	0.211	0.215	0.240	**0.343**	0.331	0.331	0.346	**0.487**
6.25%	1,000	0.197	0.142	0.195	**0.281**	0.386	0.385	0.408	**0.530**	0.540	0.535	0.530	**0.667**
	1,500	0.342	0.268	0.328	**0.423**	0.549	0.542	0.560	**0.679**	0.672	0.666	0.679	**0.796**
	2,500	0.571	0.515	0.569	**0.661**	0.769	0.762	0.779	**0.859**	0.873	0.853	0.859	**0.922**
	252	0.196	0.061	0.192	**0.269**	0.377	0.349	0.396	**0.518**	0.481	0.510	0.519	**0.651**
	500	0.418	0.282	0.422	**0.516**	0.620	0.614	0.643	**0.754**	0.746	0.740	0.754	**0.852**
7.50%	1,000	0.761	0.700	0.769	**0.840**	0.894	0.898	0.907	**0.948**	0.951	0.950	0.948	**0.975**
	1,500	0.933	0.898	0.931	**0.958**	0.978	0.976	0.981	**0.992**	0.991	0.989	0.992	**0.997**
	2,500	0.996	0.993	0.996	**0.998**	0.999	**1.000**	0.999	0.999	**1.000**	**1.000**	0.999	**1.000**
	252	0.709	0.447	0.698	**0.777**	0.859	0.845	0.869	**0.922**	0.910	0.920	0.922	**0.960**
	500	0.961	0.924	0.961	**0.975**	0.988	0.988	0.988	**0.995**	0.996	0.996	0.995	**0.998**
10.00%	1,000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1,500	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
	2,500	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
Panel B: 1% VaR													
	252	0.010	0.012	0.009	0.010	0.051	0.050	0.049	0.050	0.101	0.103	0.100	0.104
	500	0.009	0.012	0.010	0.009	0.049	0.052	0.048	0.048	0.073	0.101	0.099	0.096
1.00%	1,000	0.014	0.009	0.012	0.011	0.048	0.050	0.053	0.051	0.105	0.102	0.103	0.107
	1,500	0.011	0.009	0.009	0.010	0.050	0.050	0.048	0.051	0.095	0.103	0.101	0.101
	2,500	0.010	0.008	0.010	0.011	0.054	0.047	0.052	0.051	0.106	0.099	0.100	0.100
	252	0.013	**0.017**	0.014	0.016	0.049	**0.074**	0.057	0.066	0.089	**0.138**	0.109	0.127
	500	0.010	0.014	0.015	**0.019**	0.046	0.054	0.062	**0.080**	0.082	0.135	0.115	**0.148**
1.10%	1,000	0.014	0.006	0.013	**0.023**	0.061	0.058	0.065	**0.089**	0.097	0.117	0.120	**0.166**
	1,500	0.015	0.010	0.017	**0.028**	0.069	0.058	0.070	**0.102**	0.136	0.132	0.127	**0.184**
	2,500	0.016	0.012	0.018	**0.036**	0.072	0.078	0.083	**0.130**	0.147	0.151	0.146	**0.221**
	252	0.026	**0.029**	0.020	**0.029**	0.058	0.108	0.076	**0.111**	0.095	0.187	0.134	**0.192**
	500	0.018	0.026	0.027	**0.039**	0.066	0.072	0.086	**0.136**	0.115	0.189	0.153	**0.234**
1.25%	1,000	0.032	0.003	0.039	**0.063**	0.112	0.119	0.131	**0.198**	0.164	0.207	0.207	**0.310**
	1,500	0.044	0.027	0.057	**0.091**	0.141	0.139	0.166	**0.253**	0.268	0.260	0.260	**0.371**
	2,500	0.082	0.050	0.087	**0.134**	0.220	0.219	0.232	**0.342**	0.334	0.335	0.344	**0.476**
	252	0.059	0.060	0.045	**0.069**	0.094	0.181	0.131	**0.192**	0.134	0.281	0.206	**0.305**
	500	0.054	0.081	0.072	**0.103**	0.137	0.160	0.186	**0.276**	0.220	0.339	0.282	**0.406**
1.50%	1,000	0.132	0.020	0.159	**0.220**	0.304	0.297	0.341	**0.447**	0.377	0.435	0.448	**0.580**
	1,500	0.194	0.140	0.227	**0.315**	0.401	0.401	0.439	**0.562**	0.573	0.569	0.563	**0.686**
	2,500	0.374	0.296	0.404	**0.506**	0.617	0.613	0.641	**0.747**	0.739	0.737	0.747	**0.848**
	252	0.182	**0.194**	0.143	**0.194**	0.238	**0.405**	0.291	0.401	0.281	0.518	0.405	**0.538**
	500	0.239	0.292	0.292	**0.358**	0.419	0.437	0.490	**0.605**	0.542	0.667	0.605	**0.721**
2.00%	1,000	0.533	0.213	0.583	**0.662**	0.747	0.749	0.778	**0.852**	0.810	0.845	0.852	**0.914**
	1,500	0.736	0.665	0.768	**0.831**	0.888	0.887	0.900	**0.941**	0.951	0.946	0.941	**0.969**
	2,500	0.944	0.911	0.947	**0.969**	0.988	0.984	0.987	**0.994**	0.992	0.993	0.994	**0.998**

Table II: Half-Life Interval and α-Level

The half-life interval is computed by $log(0.5)/log(\alpha)$ and refers to the time interval over which the weight of an observation decrease to one-half its original value. The corresponding λ refers to the decay factor of the EWMA type process of computing σ_t.

Half-Life Interval	5	10	20	40	60	80
λ	0.8706	0.9330	0.9659	0.9828	0.9885	0.9914

Table III: I.I.D. VaR-Violations - Setting 1: Independence - Power of Tests

The table presents rejection rates obtained by applying tests for i.i.d. VaR-violations to 10,000 samples of non-independent VaR-violation sequences simulated by Equation (27). The VaR level p for panel A and B is set to 5% and 1%, respectively. Results are presented for various sets of sample sizes n and half-life intervals which serve as a proxy for the degree of dependence. LR_{iid}^{mar}, LR_{iid}^{wei} and GMM_{iid} refers to the independence tests of Christoffersen (1998), Christoffersen and Pelletier (2004) and Candelon et al. (2011). MCS_{iid} refers to the new Monte-Carlo-Simulation based test. Top results are highlighted in bold type.

Half-Life Interval	n	Significance level: 1%				Significance level: 5%				Significance level: 10%			
		LR_{iid}^{mar}	LR_{iid}^{wei}	GMM_{iid}	MCS_{iid}	LR_{iid}^{mar}	LR_{iid}^{wei}	GMM_{iid}	MCS_{iid}	LR_{iid}^{mar}	LR_{iid}^{wei}	GMM_{iid}	MCS_{iid}
Panel A: 5% VaR													
	252	0.067	0.005	**0.108**	0.072	0.146	0.033	0.213	**0.220**	0.195	0.075	0.270	**0.339**
	500	0.093	0.016	**0.186**	0.142	0.170	0.081	**0.362**	0.354	0.224	0.153	0.451	**0.499**
5	1,000	0.126	0.047	**0.308**	0.264	0.217	0.160	**0.591**	0.552	0.308	0.260	0.689	**0.695**
	1,500	0.155	0.077	**0.423**	0.393	0.325	0.233	**0.741**	0.684	0.451	0.358	**0.823**	0.807
	2,500	0.308	0.170	**0.614**	0.611	0.515	0.396	**0.905**	0.858	0.631	0.535	**0.948**	0.933
	252	0.037	0.005	**0.086**	0.063	0.104	0.026	0.173	**0.188**	0.153	0.064	0.227	**0.296**
	500	0.047	0.006	**0.143**	0.120	0.098	0.038	0.281	**0.293**	0.145	0.080	0.357	**0.423**
10	1,000	0.049	0.014	**0.214**	0.211	0.104	0.065	0.454	**0.469**	0.168	0.122	0.556	**0.612**
	1,500	0.051	0.021	0.295	**0.315**	0.151	0.085	0.593	**0.600**	0.246	0.158	0.695	**0.732**
	2,500	0.096	0.033	0.425	**0.503**	0.234	0.134	**0.775**	0.774	0.338	0.223	0.860	**0.872**
	252	0.026	0.005	**0.061**	0.054	0.084	0.029	0.129	**0.149**	0.131	0.066	0.176	**0.236**
	500	0.029	0.005	**0.095**	0.092	0.073	0.029	0.195	**0.231**	0.112	0.062	0.262	**0.340**
20	1,000	0.025	0.004	0.135	**0.142**	0.067	0.027	0.300	**0.332**	0.119	0.058	0.392	**0.460**
	1,500	0.018	0.005	0.169	**0.202**	0.077	0.029	0.392	**0.438**	0.151	0.058	0.494	**0.578**
	2,500	0.034	0.005	0.228	**0.327**	0.107	0.027	0.536	**0.591**	0.181	0.055	0.645	**0.727**
	252	0.022	0.005	**0.052**	0.042	0.077	0.031	0.115	**0.128**	0.117	0.069	0.162	**0.210**
	500	0.022	0.008	**0.079**	0.077	0.064	0.030	0.163	**0.196**	0.103	0.068	0.226	**0.297**
40	1,000	0.018	0.003	0.095	**0.099**	0.052	0.024	0.219	**0.251**	0.103	0.051	0.293	**0.363**
	1,500	0.012	0.002	0.107	**0.129**	0.060	0.014	0.265	**0.307**	0.117	0.037	0.354	**0.430**
	2,500	0.017	0.002	0.128	**0.180**	0.073	0.010	0.324	**0.397**	0.132	0.025	0.424	**0.531**
	252	0.020	0.008	0.041	**0.042**	0.071	0.037	0.099	**0.130**	0.107	0.082	0.141	**0.211**
	500	0.023	0.005	**0.085**	0.080	0.059	0.032	0.164	**0.198**	0.095	0.070	0.224	**0.297**
60	1,000	0.016	0.005	0.093	**0.100**	0.049	0.024	0.204	**0.246**	0.098	0.049	0.275	**0.350**
	1,500	0.012	0.003	0.106	**0.119**	0.063	0.017	0.234	**0.280**	0.120	0.040	0.314	**0.396**
	2,500	0.016	0.001	0.110	**0.146**	0.065	0.009	0.269	**0.331**	0.122	0.026	0.363	**0.459**
	252	0.022	0009	0.032	**0.036**	0.072	0.041	0.089	**0.117**	0.107	0.086	0.130	**0.200**
	500	0.020	0.006	**0.085**	0.083	0.051	0.035	0.167	**0.206**	0.085	0.073	0.224	**0.305**
80	1,000	0.016	0.003	0.113	**0.119**	0.047	0.026	0.224	**0.263**	0.093	0.057	0.297	**0.371**
	1,500	0.014	0.002	0.113	**0.128**	0.065	0.021	0.250	**0.289**	0.122	0.045	0.323	**0.400**
	2,500	0.015	0.003	0.108	**0.150**	0.065	0.013	0.267	**0.323**	0.118	0.028	0.350	**0.436**
Panel B: 1% VaR													
	252	0.055	0.004	**0.068**	0.048	**0.181**	0.035	0.136	0.141	**0.237**	0.095	0.186	0.226
	500	0.065	0.010	**0.073**	0.047	**0.198**	0.065	0.152	0.148	**0.252**	0.132	0.212	0.241
5	1,000	**0.114**	0.038	0.099	0.055	**0.230**	0.137	0.211	0.182	**0.346**	0.224	0.285	0.296
	1,500	**0.141**	0.087	0.116	0.064	**0.283**	0.219	0.265	0.212	**0.388**	0.322	0.361	0.328
	2,500	**0.193**	0.179	0.149	0.083	**0.384**	0.362	0.363	0.255	**0.482**	0.475	0.470	0.393
	252	0.037	0.005	**0.076**	0.059	**0.156**	0.034	0.141	0.147	0.217	0.080	0.192	**0.227**
	500	0.039	0.009	**0.078**	0.051	0.151	0.051	**0.156**	0.150	0.225	0.104	0.211	**0.239**
10	1,000	0.064	0.026	**0.100**	0.058	0.152	0.100	0.205	0.187	0.265	0.173	0.281	**0.297**
	1,500	0.072	0.055	**0.111**	0.067	0.174	0.161	**0.250**	0.212	0.266	0.254	**0.343**	0.327
	2,500	0.094	0.117	**0.140**	0.098	0.236	0.275	**0.340**	0.273	0.324	0.384	**0.453**	0.404
	252	0.026	0.005	**0.084**	0.066	**0.158**	0.031	0.147	0.156	0.227	0.075	0.192	**0.237**
	500	0.028	0.008	**0.076**	0.052	0.114	0.049	0.144	**0.147**	0.198	0.099	0.194	**0.235**
20	1,000	0.040	0.020	**0.083**	0.067	0.103	0.078	0.173	**0.187**	0.209	0.137	0.244	**0.287**
	1,500	0.042	0.035	**0.098**	0.069	0.124	0.113	**0.216**	0.202	0.192	0.189	0.296	**0.320**
	2,500	0.048	0.071	**0.114**	0.084	0.149	0.181	**0.283**	0.258	0.225	0.271	0.380	**0.388**
	252	0.020	0.004	**0.079**	0.065	**0.199**	0.027	0.142	0.155	**0.266**	0.063	0.193	0.238
	500	0.023	0.010	**0.078**	0.070	0.107	0.048	0.135	**0.151**	0.204	0.093	0.187	**0.222**
40	1,000	0.031	0.026	**0.089**	0.068	0.083	0.077	0.154	**0.176**	0.181	0.136	0.216	**0.265**
	1,500	0.032	0.035	**0.087**	0.072	0.099	0.099	0.182	**0.195**	0.156	0.158	0.253	**0.295**
	2,500	0.031	0.050	**0.097**	0.088	0.119	0.126	0.223	**0.238**	0.180	0.195	0.308	**0.348**
	252	0.017	0.005	**0.077**	0.052	**0.257**	0.026	0.136	0.149	**0.330**	0.062	0.188	0.230
	500	0.024	0.010	**0.088**	0.074	0.116	0.045	0.142	**0.157**	0.212	0.095	0.189	**0.229**
60	1,000	0.031	0.030	**0.089**	0.073	0.081	0.084	0.155	**0.170**	0.174	0.135	0.213	**0.251**
	1,500	0.031	0.039	**0.092**	0.082	0.092	0.095	0.174	**0.189**	0.143	0.155	0.241	**0.280**
	2,500	0.029	0.052	**0.093**	0.091	0.109	0.118	0.199	**0.218**	0.162	0.183	0.277	**0.327**
	252	0.014	0.004	**0.064**	0.037	**0.302**	0.025	0.131	0.127	**0.374**	0.054	0.181	0.204
	500	0.023	0.006	**0.081**	0.071	0.112	0.039	0.135	**0.159**	0.211	0.084	0.182	**0.231**
80	1,000	0.030	0.031	**0.096**	0.083	0.083	0.085	0.157	**0.181**	0.171	0.135	0.211	**0.262**
	1,500	0.027	0.046	**0.090**	0.088	0.083	0.103	0.163	**0.193**	0.133	0.159	0.224	**0.279**
	2,500	0.033	0.054	0.097	**0.102**	0.116	0.118	0.194	**0.220**	0.175	0.177	0.265	**0.315**

Table IV: I.I.D. VaR-Violations - Setting 2: Identical Distribution - Size and Power of Tests

The table presents rejection rates obtained by applying tests for i.i.d. VaR-violations to 10,000 samples of non-identically distributed VaR-violation sequences simulated by Equation (29). The VaR level p for panel A and B is set to 5% and 1%, respectively. Results are presented for various sets of sample sizes n and probability variation factors δ. Results for $\delta = 0p$ correspond to the evaluation of the size of the test. LR_{iid}^{mar}, LR_{iid}^{wei} and GMM_{iid} refers to the independence tests of Christoffersen (1998), Christoffersen and Pelletier (2004) and Candelon et al. (2011). MCS_{iid} refers to the new simulation based i.i.d. test. Top results are highlighted in bold type.

δ	n	Significance level: 1%				Significance level: 5%				Significance level: 10%			
		LR_{iid}^{mar}	LR_{iid}^{wei}	GMM_{iid}	MCS_{iid}	LR_{iid}^{mar}	LR_{iid}^{wei}	GMM_{iid}	MCS_{iid}	LR_{iid}^{mar}	LR_{iid}^{wei}	GMM_{iid}	MCS_{iid}
Panel A: 5% VaR													
	252	0.010	0.010	0.011	0.010	0.048	0.053	0.049	0.053	0.095	0.104	0.101	0.101
	500	0.011	0.010	0.013	0.011	0.050	0.048	0.052	0.048	0.101	0.095	0.102	0.102
0p	1,000	0.009	0.010	0.010	0.008	0.046	0.046	0.046	0.050	0.097	0.096	0.097	0.097
	1,500	0.011	0.010	0.010	0.009	0.048	0.045	0.048	0.049	0.099	0.094	0.099	0.099
	2,500	0.010	0.009	0.009	0.010	0.051	0.049	0.049	0.051	0.101	0.102	0.101	0.101
	252	0.011	0.009	**0.014**	0.009	0.052	0.048	0.058	**0.060**	0.101	0.094	0.105	**0.111**
	500	0.011	0.009	**0.015**	0.015	0.050	0.044	0.054	**0.068**	0.100	0.087	0.102	**0.128**
0.1p	1,000	0.011	0.006	**0.019**	0.018	0.048	0.032	0.066	**0.074**	0.099	0.073	0.116	**0.136**
	1,500	0.009	0.007	**0.021**	0.020	0.047	0.036	0.071	**0.082**	0.094	0.076	0.124	**0.146**
	2,500	0.009	0.008	0.021	**0.023**	0.049	0.037	0.078	**0.093**	0.100	0.072	0.131	**0.170**
	252	0.015	0.008	**0.019**	0.015	0.060	0.037	0.068	**0.074**	0.111	0.075	0.117	**0.144**
	500	0.014	0.005	0.033	**0.035**	0.057	0.025	0.094	**0.124**	0.106	0.058	0.147	**0.208**
0.2p	1,000	0.011	0.002	0.055	**0.065**	0.049	0.020	0.140	**0.190**	0.094	0.044	0.204	**0.291**
	1,500	0.011	0.002	0.072	**0.090**	0.051	0.014	0.177	**0.238**	0.106	0.033	0.250	**0.344**
	2,500	0.012	0.001	0.096	**0.140**	0.057	0.008	0.243	**0.326**	0.111	0.019	0.329	**0.452**
	252	0.015	0.004	**0.037**	0.030	0.061	0.023	0.105	**0.130**	0.112	0.053	0.156	**0.227**
	500	0.020	0.003	0.094	**0.097**	0.061	0.018	0.202	**0.258**	0.106	0.050	0.275	**0.377**
0.3p	1,000	0.016	0.003	0.212	**0.241**	0.054	0.024	0.386	**0.456**	0.106	0.058	0.471	**0.579**
	1,500	0.015	0.005	0.297	**0.358**	0.063	0.028	0.504	**0.591**	0.130	0.068	0.593	**0.704**
	2,500	0.022	0.008	0.450	**0.549**	0.085	0.038	0.697	**0.771**	0.148	0.075	0.783	**0.856**
	252	0.027	0.001	**0.079**	0.053	0.080	0.017	0.181	**0.209**	0.131	0.043	0.240	**0.346**
	500	0.033	0.006	0.273	**0.283**	0.078	0.049	0.452	**0.540**	0.125	0.112	0.535	**0.664**
0.4p	1,000	0.032	0.043	0.613	**0.638**	0.079	0.164	0.783	**0.828**	0.140	0.275	0.838	**0.894**
	1,500	0.029	0.114	0.781	**0.838**	0.105	0.284	0.908	**0.943**	0.181	0.410	0.940	**0.971**
	2,500	0.053	0.248	0.942	**0.971**	0.158	0.482	0.987	**0.993**	0.250	0.616	0.993	**0.997**
	252	0.041	0.002	**0.158**	0.113	0.104	0.028	0.317	**0.378**	0.148	0.074	0.400	**0.552**
	500	0.053	0.040	0.688	**0.729**	0.109	0.213	0.863	**0.944**	0.157	0.376	0.915	**0.982**
0.5p	1,000	0.057	0.436	**1.000**	**1.000**	0.124	0.794	**1.000**	**1.000**	0.201	0.910	**1.000**	**1.000**
	1,500	0.061	0.892	**1.000**	**1.000**	0.186	0.986	**1.000**	**1.000**	0.299	0.998	**1.000**	**1.000**
	2,500	0.138	**1.000**	**1.000**	**1.000**	0.311	**1.000**	**1.000**	**1.000**	0.425	**1.000**	**1.000**	**1.000**
Panel B: 1% VaR													
	252	0.010	0.007	0.010	0.012	0.056	0.042	0.052	0.050	0.108	0.089	0.102	0.103
	500	0.009	0.009	0.009	0.011	0.050	0.050	0.048	0.053	0.101	0.097	0.099	0.101
0p	1,000	0.010	0.010	0.009	0.011	0.048	0.046	0.049	0.051	0.100	0.096	0.102	0.101
	1,500	0.010	0.010	0.010	0.010	0.047	0.047	0.050	0.052	0.098	0.095	0.100	0.096
	2,500	0.009	0.010	0.012	0.011	0.049	0.047	0.050	0.053	0.099	0.098	0.099	0.105
	252	**0.011**	0.008	0.009	0.009	**0.054**	0.042	0.049	0.050	**0.104**	0.087	0.099	0.098
	500	0.009	0.009	0.009	**0.012**	0.048	0.047	0.047	**0.054**	0.097	0.093	0.096	**0.103**
0.1p	1,000	0.011	0.011	0.012	0.012	0.053	0.049	0.054	**0.056**	0.102	0.099	0.107	**0.113**
	1,500	**0.013**	0.009	0.012	0.011	0.053	0.048	0.052	**0.059**	0.104	0.095	0.102	**0.113**
	2,500	**0.013**	0.008	0.012	**0.013**	0.055	0.042	0.056	**0.064**	0.104	0.088	0.111	**0.121**
	252	0.010	0.007	0.011	**0.014**	0.052	0.038	0.048	**0.056**	0.102	0.079	0.097	**0.106**
	500	0.010	0.009	0.011	**0.012**	**0.058**	0.042	0.052	0.053	**0.105**	0.086	0.098	0.105
0.2p	1,000	0.012	0.008	0.015	0.013	0.053	0.041	0.060	**0.065**	0.107	0.087	0.107	**0.126**
	1,500	0.012	0.007	**0.016**	0.016	0.056	0.039	0.064	**0.082**	0.114	0.085	0.122	**0.151**
	2,500	0.013	0.008	0.027	**0.031**	0.058	0.042	0.094	**0.120**	0.111	0.087	0.152	**0.204**
	252	0.013	0.006	0.014	**0.015**	0.057	0.033	0.054	**0.060**	0.105	0.073	0.102	**0.115**
	500	0.013	0.006	**0.017**	**0.017**	0.062	0.033	0.055	**0.067**	0.110	0.066	0.101	**0.124**
0.3p	1,000	0.015	0.005	0.022	0.020	0.064	0.034	0.076	**0.091**	0.123	0.078	0.132	**0.173**
	1,500	0.014	0.008	0.033	**0.034**	0.063	0.041	0.101	**0.143**	0.121	0.088	0.168	**0.250**
	2,500	0.017	0.011	0.077	**0.090**	0.070	0.058	0.193	**0.242**	0.125	0.119	0.278	**0.360**
	252	0.015	0.005	**0.017**	**0.017**	0.063	0.026	0.055	**0.070**	0.111	0.061	0.104	**0.121**
	500	0.016	0.003	**0.023**	0.022	0.069	0.023	0.066	**0.075**	0.114	0.057	0.111	**0.145**
0.4p	1,000	0.018	0.005	0.038	0.028	0.076	0.035	0.114	**0.138**	0.139	0.079	0.181	**0.253**
	1,500	0.020	0.014	**0.074**	0.073	0.074	0.065	0.191	**0.257**	0.139	0.129	0.280	**0.407**
	2,500	0.021	0.040	0.226	**0.259**	0.081	0.150	0.424	**0.522**	0.146	0.251	0.518	**0.645**
	252	0.018	0.001	0.022	**0.024**	0.069	0.011	0.066	**0.081**	0.114	0.039	0.108	**0.131**
	500	0.017	0.001	**0.031**	0.030	0.082	0.014	**0.091**	**0.091**	0.132	0.045	0.140	**0.164**
0.5p	1,000	0.025	0.007	0.079	0.053	0.087	0.051	0.197	**0.225**	0.157	0.113	0.277	**0.377**
	1,500	0.024	0.032	**0.174**	0.163	0.085	0.142	0.354	**0.487**	0.164	0.249	0.467	**0.670**
	2,500	0.027	0.167	0.597	**0.694**	0.099	0.437	0.822	**0.926**	0.172	0.602	0.893	**0.975**

Table V: Conditional Coverage - Setting 1: Independence - Power of Tests

The table presents rejection rates obtained by applying cc tests to 10,000 samples of non-independent VaR-violation sequences simulated by Equation (27) with an increased violation probability. The VaR level p for panel A and B is set to 5% and 1%, respectively. Results are presented for various sets of sample sizes n, γ-factors which increase the probability of a VaR-violation, and decay intervals which serve as a proxy for the degree of dependence. LR_{cc}^{mar}, LR_{cc}^{wei} and GMM_{cc} refers to the cc tests of Christoffersen (1998), Christoffersen and Pelletier (2004) and Candelon et al. (2011). MCS_{cc} refers to the new simulation based test. Top results are highlighted in bold type.

Decay Interval	p	n	Significance level: 1%				Significance level: 5%				Significance level: 10%			
			LR_{cc}^{mar}	LR_{cc}^{wei}	GMM_{cc}	MCS_{cc}	LR_{cc}^{mar}	LR_{cc}^{wei}	GMM_{cc}	MCS_{cc}	LR_{cc}^{mar}	LR_{cc}^{wei}	GMM_{cc}	MCS_{cc}
Panel A: 5% VaR														
10	5.50%	252	0.052	0.028	0.044	**0.093**	0.103	0.088	0.212	**0.237**	0.193	0.154	0.318	**0.344**
		500	0.059	0.033	0.063	**0.150**	0.128	0.108	0.287	**0.340**	0.208	0.177	0.415	**0.463**
		1,000	0.074	0.047	0.107	**0.251**	0.166	0.142	0.435	**0.493**	0.231	0.218	0.571	**0.613**
		1,500	0.104	0.061	0.183	**0.371**	0.199	0.168	0.558	**0.613**	0.280	0.256	0.686	**0.721**
		2,500	0.152	0.095	0.360	**0.565**	0.290	0.226	0.767	**0.783**	0.377	0.331	0.857	**0.860**
10	7.50%	252	0.204	0.109	0.060	**0.235**	0.302	0.222	0.364	**0.457**	0.433	0.307	0.488	**0.555**
		500	0.353	0.259	0.144	**0.417**	0.493	0.429	0.565	**0.661**	0.599	0.526	0.693	**0.762**
		1,000	0.591	0.524	0.387	**0.704**	0.747	0.693	0.825	**0.878**	0.804	0.770	0.893	**0.929**
		1,500	0.795	0.708	0.669	**0.886**	0.878	0.847	0.939	**0.967**	0.915	0.899	0.970	**0.984**
		2,500	0.946	0.909	0.932	**0.985**	0.979	0.961	0.994	**0.998**	0.988	0.980	0.997	**0.999**
20	6.25%	252	0.096	0.060	0.047	**0.128**	0.160	0.142	0.227	**0.285**	0.258	0.215	0.335	**0.382**
		500	0.127	0.083	0.052	**0.182**	0.218	0.178	0.285	**0.372**	0.306	0.253	0.418	**0.486**
		1,000	0.179	0.137	0.096	**0.299**	0.318	0.264	0.438	**0.539**	0.393	0.345	0.572	**0.651**
		1,500	0.272	0.191	0.175	**0.451**	0.403	0.343	0.575	**0.680**	0.486	0.437	0.699	**0.776**
		2,500	0.409	0.300	0.388	**0.678**	0.591	0.475	0.771	**0.853**	0.666	0.577	0.856	**0.907**
40	6.25%	252	0.142	0.119	0.094	**0.166**	0.201	0.212	0.280	**0.308**	0.289	0.290	0.385	**0.404**
		500	0.156	0.124	0.075	**0.189**	0.234	0.219	0.289	**0.366**	0.314	0.287	0.404	**0.471**
		1,000	0.200	0.174	0.098	**0.267**	0.329	0.292	0.399	**0.490**	0.399	0.367	0.525	**0.604**
		1,500	0.279	0.216	0.150	**0.372**	0.399	0.354	0.495	**0.597**	0.473	0.445	0.618	**0.700**
		2,500	0.397	0.301	0.289	**0.552**	0.571	0.460	0.669	**0.765**	0.643	0.552	0.775	**0.838**
80	5.50%	252	0.223	0.224	**0.256**	0.220	0.310	0.374	**0.458**	0.406	0.416	0.466	**0.546**	0.535
		500	0.173	0.175	0.193	**0.224**	0.252	0.288	0.391	**0.395**	0.335	0.369	0.486	**0.505**
		1,000	0.129	0.124	0.149	**0.217**	0.215	0.207	0.357	**0.394**	0.275	0.277	0.456	**0.502**
		1,500	0.122	0.104	0.139	**0.223**	0.194	0.183	0.343	**0.401**	0.253	0.248	0.446	**0.510**
		2,500	0.126	0.092	0.142	**0.250**	0.219	0.163	0.376	**0.454**	0.278	0.223	0.483	**0.557**
80	7.50%	252	0.278	0.249	0.218	**0.292**	0.336	0.348	0.423	**0.449**	0.413	0.417	0.513	**0.542**
		500	0.326	0.294	0.220	**0.362**	0.404	0.388	0.473	**0.540**	0.474	0.452	0.577	**0.633**
		1,000	0.491	0.477	0.313	**0.564**	0.625	0.614	0.676	**0.764**	0.685	0.681	0.770	**0.837**
		1,500	0.696	0.626	0.478	**0.713**	0.789	0.762	0.821	**0.874**	0.839	0.821	0.888	**0.919**
		2,500	0.908	0.874	0.807	**0.927**	0.957	0.937	0.966	**0.981**	0.970	0.960	0.982	**0.991**
Panel B: 1% VaR														
10	1.10%	252	0.038	0.017	0.093	**0.094**	0.140	0.066	**0.198**	0.191	**0.335**	0.128	0.273	0.266
		500	0.047	0.023	**0.092**	0.091	0.174	0.081	**0.201**	0.191	0.274	0.144	0.267	**0.274**
		1,000	0.044	0.037	0.023	**0.088**	0.158	0.129	0.194	**0.227**	0.242	0.210	0.303	**0.313**
		1,500	0.051	0.066	0.025	**0.094**	0.180	0.167	0.220	**0.253**	0.275	0.264	0.343	**0.359**
		2,500	0.057	0.120	0.042	**0.125**	0.194	0.271	**0.304**	**0.304**	0.326	0.383	**0.457**	0.426
10	1.50%	252	0.072	0.031	0.154	**0.162**	0.216	0.109	0.291	**0.297**	**0.455**	0.186	0.377	0.380
		500	0.127	0.059	0.177	**0.202**	0.341	0.147	0.327	**0.343**	**0.466**	0.220	0.402	0.436
		1,000	0.167	0.113	0.034	**0.229**	0.367	0.244	0.314	**0.426**	0.467	0.340	0.441	**0.528**
		1,500	0.225	0.210	0.048	**0.288**	0.439	0.366	0.413	**0.518**	0.568	0.476	0.553	**0.619**
		2,500	0.350	0.418	0.116	**0.424**	0.606	0.600	0.575	**0.672**	0.728	0.694	0.712	**0.771**
20	1.25%	252	0.069	0.040	**0.141**	0.135	0.182	0.104	**0.245**	0.238	**0.380**	0.168	0.317	0.312
		500	0.067	0.034	0.118	**0.130**	0.214	0.093	0.231	**0.233**	0.311	0.154	0.297	**0.316**
		1,000	0.074	0.051	0.023	**0.128**	0.207	0.133	0.219	**0.282**	0.289	0.207	0.329	**0.377**
		1,500	0.080	0.078	0.023	**0.150**	0.212	0.178	0.247	**0.321**	0.327	0.259	0.375	**0.423**
		2,500	0.107	0.141	0.038	**0.194**	0.277	0.284	0.324	**0.403**	0.409	0.389	0.466	**0.526**
40	1.25%	252	0.129	0.085	**0.183**	0.183	0.227	0.158	**0.273**	0.271	**0.387**	0.213	0.336	0.335
		500	0.099	0.064	0.135	**0.144**	0.230	0.124	0.228	**0.233**	**0.307**	0.183	0.294	0.306
		1,000	0.091	0.072	0.041	**0.146**	0.212	0.146	0.209	**0.271**	0.285	0.213	0.311	**0.356**
		1,500	0.095	0.089	0.035	**0.148**	0.206	0.172	0.221	**0.302**	0.312	0.247	0.334	**0.397**
		2,500	0.111	0.126	0.044	**0.190**	0.273	0.248	0.273	**0.377**	0.380	0.341	0.397	**0.491**
80	1.10%	252	0.243	0.192	**0.296**	0.296	0.342	0.273	0.373	**0.374**	**0.470**	0.329	0.424	0.427
		500	0.139	0.105	0.167	**0.174**	0.226	0.172	0.244	**0.244**	0.277	0.243	**0.307**	0.307
		1,000	0.109	0.103	0.085	**0.135**	0.198	0.190	**0.242**	0.236	0.278	0.263	**0.330**	0.321
		1,500	0.088	0.101	0.074	**0.128**	0.178	0.178	0.233	**0.248**	0.277	0.250	0.330	**0.339**
		2,500	0.077	0.098	0.068	**0.138**	0.182	0.183	0.222	**0.266**	0.260	0.257	0.320	**0.364**
80	1.50%	252	0.263	0.209	0.302	**0.316**	0.355	0.289	0.385	**0.388**	**0.480**	0.344	0.442	0.441
		500	0.198	0.141	0.217	**0.234**	**0.316**	0.209	0.308	0.313	**0.386**	0.267	0.367	0.378
		1,000	0.195	0.151	0.095	**0.222**	0.318	0.243	0.285	**0.352**	0.384	0.304	0.378	**0.440**
		1,500	0.213	0.172	0.078	**0.242**	0.351	0.272	0.296	**0.417**	0.464	0.351	0.409	**0.507**
		2,500	0.313	0.288	0.102	**0.331**	0.515	0.432	0.407	**0.560**	0.621	0.528	0.538	**0.658**

Table VI: Conditional Coverage - Setting 2: Identical Distribution - Size and Power of Tests

The table presents rejection rates obtained by applying cc tests to 10,000 samples of non-identically distributed VaR-violation sequences simulated by Equation (29) with an increased violation probability. The VaR level p for panel A and B is set to 5% and 1%, respectively. Results are presented for various sets of sample sizes n and γ-factors which increase the probability of a VaR-violation, and probability variation factors δ. The results for $\delta = 0$ correspond to the evaluation of the size of the test. LR_{cc}^{mar}, LR_{cc}^{wei} and GMM_{cc} refers to the cc tests of Christoffersen (1998), Christoffersen and Pelletier (2004) and Candelon et al. (2011). MCS_{iid} refers to the new simulation based test. Top results are highlighted in bold type.

δ	p	n	Significance level: 1%				Significance level: 5%				Significance level: 10%			
			LR_{cc}^{mar}	LR_{cc}^{wei}	GMM_{cc}	MCS_{cc}	LR_{cc}^{mar}	LR_{cc}^{wei}	GMM_{cc}	MCS_{cc}	LR_{cc}^{mar}	LR_{cc}^{wei}	GMM_{cc}	MCS_{cc}
Panel A: 5% VaR														
		252	0.010	0.010	0.009	0.011	0.049	0.049	0.051	0.051	0.093	0.099	0.103	0.100
		500	0.010	0.010	0.008	0.011	0.048	0.053	0.046	0.051	0.105	0.103	0.098	0.100
$0p$	5.00%	1,000	0.011	0.009	0.010	0.011	0.052	0.046	0.053	0.052	0.104	0.100	0.105	0.098
		1,500	0.011	0.009	0.009	0.011	0.049	0.047	0.052	0.053	0.098	0.101	0.102	0.102
		2,500	0.010	0.009	0.009	0.012	0.052	0.049	0.049	0.054	0.101	0.097	0.100	0.102
		252	**0.016**	0.008	0.004	0.011	0.061	0.044	0.046	**0.065**	0.115	0.086	0.096	**0.124**
		500	**0.019**	0.015	0.005	0.015	0.066	0.056	0.045	**0.078**	0.129	0.105	0.108	**0.148**
$0.1p$	5.50%	1,000	0.020	0.016	0.006	**0.021**	0.082	0.068	0.058	**0.103**	0.138	0.129	0.123	**0.186**
		1,500	0.030	0.021	0.008	**0.033**	0.092	0.084	0.071	**0.125**	0.154	0.148	0.151	**0.218**
		2,500	0.036	0.034	0.011	**0.048**	0.129	0.107	0.106	**0.174**	0.198	0.181	0.209	**0.281**
		252	**0.147**	0.073	0.008	0.103	0.280	0.193	0.220	**0.330**	0.431	0.296	0.372	**0.442**
		500	**0.309**	0.230	0.028	0.180	0.488	0.442	0.399	**0.501**	0.622	0.569	0.589	**0.659**
$0.1p$	7.50%	1,000	**0.609**	0.563	0.151	0.464	0.802	0.775	0.733	0.801	0.868	0.855	0.864	**0.902**
		1,500	**0.847**	0.808	0.407	0.733	0.932	0.928	0.896	**0.942**	0.962	0.961	0.961	**0.976**
		2,500	**0.979**	0.974	0.853	0.958	0.996	0.993	0.994	**0.997**	0.997	0.998	0.998	**0.999**
		252	0.038	0.010	0.003	**0.043**	0.100	0.048	0.095	**0.166**	0.193	0.096	0.187	**0.258**
		500	0.061	0.027	0.012	**0.101**	0.151	0.097	0.174	**0.291**	0.252	0.169	0.308	**0.419**
$0.3p$	6.25%	1,000	0.112	0.066	0.051	**0.237**	0.273	0.188	0.348	**0.492**	0.373	0.285	0.508	**0.634**
		1,500	0.199	0.113	0.114	**0.402**	0.367	0.281	0.515	**0.670**	0.477	0.402	0.670	**0.779**
		2,500	0.374	0.236	0.306	**0.667**	0.617	0.456	0.765	**0.873**	0.710	0.593	0.864	**0.929**
		252	0.017	0.002	0.014	**0.088**	0.045	0.023	0.177	**0.260**	0.105	0.060	0.298	**0.382**
		500	0.024	0.029	0.165	**0.477**	0.068	0.115	0.602	**0.733**	0.134	0.209	0.733	**0.824**
$0.5p$	5.50%	1,000	0.039	0.180	0.778	**0.892**	0.105	0.414	0.947	**0.963**	0.161	0.561	0.967	**0.981**
		1,500	0.063	0.429	0.951	**0.980**	0.148	0.682	0.992	**0.995**	0.230	0.791	0.997	**0.998**
		2,500	0.117	0.775	0.999	**1.000**	0.259	0.911	**1.000**	**1.000**	0.347	0.953	**1.000**	**1.000**
		252	0.137	0.044	0.022	**0.206**	0.256	0.148	0.320	**0.469**	0.418	0.240	0.478	**0.589**
		500	0.306	0.199	0.125	**0.491**	0.493	0.408	0.618	**0.759**	0.628	0.532	0.756	**0.853**
$0.5p$	7.50%	1,000	0.621	0.541	0.491	**0.849**	0.805	0.772	0.918	**0.961**	0.871	0.856	0.965	**0.983**
		1,500	0.864	0.794	0.820	**0.974**	0.939	0.924	0.987	**0.996**	0.965	0.959	0.996	**0.999**
		2,500	0.984	0.973	0.991	**1.000**	0.996	0.993	**1.000**	**1.000**	0.999	0.998	**1.000**	**1.000**
Panel B: 1% VaR														
		252	0.009	0.008	0.010	0.009	0.046	0.041	0.048	0.051	0.175	0.083	0.101	0.102
		500	0.011	0.012	0.009	0.010	0.061	0.052	0.050	0.053	0.115	0.098	0.094	0.105
$0p$	1.00%	1,000	0.012	0.011	0.009	0.010	0.048	0.053	0.048	0.050	0.091	0.102	0.098	0.100
		1,500	0.010	0.009	0.011	0.008	0.042	0.046	0.048	0.052	0.101	0.096	0.098	0.101
		2,500	0.010	0.009	0.011	0.010	0.047	0.046	0.051	0.050	0.093	0.097	0.103	0.104
		252	0.012	0.007	0.014	**0.016**	0.056	0.043	0.064	**0.066**	**0.211**	0.090	0.122	0.124
		500	0.012	0.007	0.013	**0.018**	**0.082**	0.043	0.058	0.065	**0.153**	0.092	0.106	0.125
$0.1p$	1.10%	1,000	**0.015**	0.008	0.006	0.012	0.062	0.041	0.038	**0.066**	0.110	0.086	0.090	**0.128**
		1,500	0.013	0.009	0.006	**0.015**	0.059	0.043	0.039	**0.071**	0.125	0.090	0.095	**0.134**
		2,500	0.013	0.011	0.005	**0.015**	0.069	0.050	0.041	**0.078**	0.142	0.102	0.099	**0.145**
		252	0.029	0.014	**0.053**	**0.053**	0.124	0.074	0.152	**0.158**	**0.385**	0.140	0.247	0.247
		500	0.055	0.015	0.050	**0.073**	**0.257**	0.071	0.167	0.203	**0.394**	0.128	0.233	0.309
$0.1p$	1.50%	1,000	**0.095**	0.037	0.002	0.084	**0.283**	0.129	0.120	0.259	0.387	0.221	0.241	**0.408**
		1,500	**0.148**	0.081	0.001	0.097	**0.355**	0.217	0.155	0.343	**0.528**	0.335	0.317	0.488
		2,500	**0.251**	0.222	0.006	0.170	**0.563**	0.445	0.269	0.506	**0.708**	0.576	0.457	0.646
		252	0.013	0.007	**0.027**	0.026	0.073	0.048	**0.098**	0.097	**0.269**	0.095	0.171	0.168
		500	0.021	0.007	0.023	**0.036**	**0.129**	0.040	0.097	0.116	**0.225**	0.082	0.152	0.198
$0.3p$	1.25%	1,000	0.028	0.008	0.002	**0.039**	0.115	0.047	0.062	**0.146**	0.185	0.096	0.141	**0.238**
		1,500	0.031	0.015	0.002	**0.050**	0.124	0.065	0.078	**0.188**	0.234	0.125	0.171	**0.293**
		2,500	0.046	0.033	0.006	**0.093**	0.188	0.122	0.129	**0.285**	0.312	0.203	0.253	**0.413**
		252	0.007	0.003	**0.022**	0.019	0.054	0.022	**0.082**	0.077	**0.212**	0.055	0.141	0.133
		500	0.011	0.004	**0.029**	0.029	0.087	0.025	0.102	**0.105**	0.167	0.059	0.161	**0.176**
$0.5p$	1.10%	1,000	0.010	0.006	0.008	**0.060**	0.062	0.037	0.119	**0.183**	0.117	0.088	0.219	**0.272**
		1,500	0.012	0.023	0.012	**0.127**	0.068	0.087	0.190	**0.321**	0.133	0.163	0.327	**0.450**
		2,500	0.010	0.082	0.109	**0.439**	0.077	0.238	0.551	**0.700**	0.162	0.365	0.715	**0.807**
		252	0.025	0.009	0.055	**0.059**	0.125	0.051	0.162	**0.170**	**0.380**	0.105	0.256	0.262
		500	0.058	0.011	0.075	**0.094**	**0.258**	0.060	0.206	0.224	**0.394**	0.111	0.279	0.332
$0.5p$	1.50%	1,000	0.091	0.033	0.006	**0.150**	0.293	0.118	0.199	**0.366**	0.395	0.201	0.339	**0.493**
		1,500	0.149	0.092	0.011	**0.250**	0.352	0.234	0.290	**0.518**	0.521	0.352	0.457	**0.641**
		2,500	0.263	0.258	0.046	**0.458**	0.569	0.474	0.500	**0.735**	0.719	0.605	0.674	**0.829**

Table VII: Conditional Coverage - Power of the MCS_{cc} Test under Different Level of a

The table presents rejection rates obtained by applying the MCS_{cc} test to 10,000 samples of non-i.i.d. distributed VaR-violation sequences. Panel A and B contain rejection rates for sequences simulated by Equation (27) and Equation (29) with an increased violation probability. The parameter combinations used for the simulations are described at the top of each panel. The top result for each combination of a, VaR level, and significance level is highlighted in bold type.

	5% VaR			1% VaR		
	Significance level:			Significance level:		
a	1%	5%	10%	1%	5%	10%
Panel A: $n = 1,000\,/\,\gamma = 1.25\,/$ half-life interval $= 20$ days						
0	0.107	0.294	0.440	0.056	0.171	0.283
0.1	0.123	0.329	0.482	0.053	0.184	0.295
0.2	0.149	0.376	0.535	0.068	0.219	0.337
0.3	0.169	0.449	0.607	0.082	0.232	0.356
0.4	0.231	0.511	0.649	0.106	0.265	0.378
0.5	0.310	**0.550**	**0.664**	0.128	0.277	0.372
0.6	0.350	0.545	0.641	**0.150**	**0.289**	**0.379**
0.7	**0.366**	0.539	0.621	0.144	0.254	0.340
0.8	0.343	0.511	0.604	0.140	0.256	0.330
0.9	0.318	0.468	0.553	0.149	0.264	0.342
1	0.306	0.455	0.536	0.125	0.224	0.300
Panel B: $n = 1,000\,/\,\gamma = 1.25\,/\,\delta = 0.3p$						
0	0.105	0.264	0.393	0.014	0.074	0.151
0.1	0.108	0.290	0.433	0.013	0.081	0.164
0.2	0.124	0.336	0.479	0.015	0.093	0.183
0.3	0.146	0.383	0.548	0.019	0.098	0.192
0.4	0.188	0.453	0.604	0.023	0.121	0.221
0.5	0.232	0.509	0.636	0.036	0.140	0.234
0.6	0.294	**0.542**	**0.657**	0.053	0.153	0.236
0.7	**0.299**	0.519	0.631	0.059	0.158	0.233
0.8	0.285	0.505	0.617	**0.067**	**0.163**	**0.238**
0.9	0.256	0.463	0.570	0.064	0.161	0.236
1	0.239	0.441	0.553	0.064	0.159	0.234

Table VIII: Summary Statistics

Summary statistics of the portfolio data set used for the empirical application of the *MCS* and *GMM* tests. The data set consists of 5,740 log-returns for each of the four portfolios covering a period from January 1, 1991 to December 31, 2012. Mean Return p.a. and Volatility p.a. are annualized with 250 days.

Portfolio	1	2	3	4
Minimum	-2.691%	-3.086%	-3.473%	-2.805%
5% quantile	-0.651%	-0.531%	-0.657%	-0.638%
Median Return	0.016%	0.011%	0.016%	0.016%
Mean Return	0.025%	0.020%	0.026%	0.027%
95% quantile	0.657%	0.564%	0.683%	0.648%
Maximum	3.705%	2.683%	3.621%	3.745%
Volatility	0.417%	0.369%	0.426%	0.425%
Skewness	-0.133	-0.467	-0.300	0.083
Kurtosis	6.67	8.94	6.85	7.80
Mean Return p.a.	6.24%	4.95%	6.43%	6.84%
Volatility p.a.	6.59%	5.84%	6.73%	6.71%
Maximum Drawdown	-23.46%	-24.51%	-23.80%	-24.62%

Table IX: Empirical Application - Total Period

The table contains Violation Ratios (i.e., VaR-violation frequency divided by the number of VaR-forecasts) of the total period consisting of 5,490 VaR-forecasts for each portfolio (17.12.1991 to 31.12.2012). In addition, the table contains p-values for the unconditional coverage tests MCS_{uc}^{lt} (lower tail), MCS_{uc}^{ut} (upper tail), MCS_{uc}^{tt} (two tailed), and GMM_{uc}, for the i.i.d. tests MCS_{iid} and GMM_{iid}, and for the conditional coverage tests MCS_{cc} and GMM_{cc}. The extensions *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level.

| VaR Model	Historical VaR				GARCH approach			
Portfolio	1	2	3	4	1	2	3	4
Panel A: 5% VaR								
Viol. Ratio	5.43%	5.37%	5.50%	5.66%	4.37%	4.12%	4.54%	4.12%
MCS_{uc}^{lt}	0.923	0.901	0.956	0.987	0.013**	0.001***	0.054*	0.001***
MCS_{uc}^{ut}	0.077*	0.099*	0.044**	0.013**	0.987	0.999	0.946	0.999
MCS_{uc}^{tt}	0.155	0.197	0.088*	0.025**	0.025**	0.002***	0.108	0.002***
GMM_{uc}	0.133	0.221	0.091*	0.035**	0.025**	0.002***	0.107	0.002***
MCS_{iid}	0.000***	0.000***	0.000***	0.000***	0.014**	0.000***	0.001***	0.001***
GMM_{iid}	0.000***	0.000***	0.000***	0.000***	0.005***	0.001***	0.005***	0.004***
MCS_{cc}	0.000***	0.000***	0.000***	0.000***	0.002***	0.000***	0.000***	0.000***
GMM_{cc}	0.001***	0.000***	0.001***	0.001***	0.002***	0.001***	0.004***	0.001***
Panel B: 1% VaR								
Viol. Ratio	1.20%	1.22%	1.35%	1.35%	1.53%	1.48%	1.46%	1.33%
MCS_{uc}^{lt}	0.924	0.949	0.994	0.992	1.000	1.000	0.999	0.994
MCS_{uc}^{ut}	0.076*	0.052*	0.006***	0.008***	0.000***	0.001***	0.001***	0.006***
MCS_{uc}^{tt}	0.151	0.103	0.012***	0.016***	0.000***	0.001***	0.001***	0.013**
GMM_{uc}	0.124	0.114	0.020**	0.020**	0.002***	0.003***	0.005***	0.026**
MCS_{iid}	0.022**	0.007***	0.003***	0.003***	0.130	0.204	0.578	0.057*
GMM_{iid}	0.022**	0.007***	0.004***	0.006***	0.439	0.012**	0.311	0.019**
MCS_{cc}	0.019**	0.006***	0.001***	0.001***	0.004***	0.009***	0.026**	0.011**
GMM_{cc}	0.034**	0.017**	0.008***	0.006***	0.019**	0.019**	0.051*	0.022**

Table X: Empirical Application - Bear, Bull, and Bear + Bull Market

For each portfolio, the table contains Violation Ratios (i.e., number of VaR-violations divided by VaR-forecasts) of the bear market period (250 VaR-forecasts from 16.04.2002 to 31.03.2003), the bull market period (250 VaR-forecasts from 01.04.2003 to 15.03.2004), and the combination of the bear and bull market period (500 VaR-forecasts from 16.04.2002 to 15.03.2004). The VaR level is set to 5%. In addition, the table contains the corresponding p-values for the unconditional coverage tests MCS_{uc}^{lt} (lower tail), MCS_{uc}^{ut} (upper tail), MCS_{uc}^{tt} (two tailed), and GMM_{uc}, for the i.i.d. tests MCS_{iid} and GMM_{iid}, and for the conditional coverage tests MCS_{cc} and GMM_{cc}. The extensions *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level.

VaR Model	Historical VaR				GARCH(1,1)			
Portfolio	1	2	3	4	1	2	3	4
Panel A: Bear Market / 5% VaR								
Viol. Ratio	7.60%	7.60%	8.40%	9.20%	8.80%	8.00%	9.20%	8.80%
MCS_{uc}^{lt}	0.967	0.964	0.988	0.998	0.995	0.982	0.998	0.995
MCS_{uc}^{ut}	0.033**	0.036**	0.012**	0.002***	0.005***	0.018**	0.002***	0.005***
MCS_{uc}^{tt}	0.065*	0.073*	0.023**	0.004***	0.011**	0.036**	0.003***	0.010**
GMM_{uc}	0.131	0.120	0.050*	0.019**	0.040**	0.064*	0.017**	0.045**
MCS_{iid}	0.000***	0.008***	0.025**	0.010**	0.033**	0.047**	0.207	0.051*
GMM_{iid}	0.000***	0.005***	0.042**	0.010**	0.078*	0.197	0.819	0.256
MCS_{cc}	0.001***	0.005***	0.007***	0.003***	0.006***	0.015**	0.015**	0.007***
GMM_{cc}	0.014**	0.037**	0.044**	0.020**	0.039**	0.090*	0.048**	0.048**
Panel B: Bull Market / 5% VaR								
Viol. Ratio	1.20%	2.00%	1.20%	1.60%	1.60%	2.80%	1.60%	1.60%
MCS_{uc}^{lt}	0.001***	0.007***	0.001***	0.004***	0.002***	0.046**	0.004***	0.004***
MCS_{uc}^{ut}	0.999	0.993	0.999	0.996	0.998	0.954	0.996	0.996
MCS_{uc}^{tt}	0.003***	0.013**	0.001***	0.007***	0.004***	0.093*	0.007***	0.008***
GMM_{uc}	0.001***	0.007***	0.001***	0.001***	0.004***	0.080*	0.003***	0.004***
MCS_{iid}	0.424	0.545	0.428	0.204	0.259	0.540	0.255	0.258
GMM_{iid}	0.657	0.634	0.659	0.787	0.770	0.643	0.757	0.761
MCS_{cc}	0.044**	0.095*	0.044**	0.025**	0.040**	0.237	0.036**	0.035**
GMM_{cc}	0.003***	0.013**	0.003***	0.003***	0.003***	0.193	0.005***	0.004***
Panel C: Bear + Bull Market / 5% VaR								
Viol. Ratio	4.40%	4.80%	4.80%	5.40%	5.20%	5.40%	5.40%	5.20%
MCS_{uc}^{lt}	0.269	0.404	0.457	0.687	0.580	0.666	0.684	0.591
MCS_{uc}^{ut}	0.731	0.596	0.543	0.313	0.420	0.334	0.316	0.409
MCS_{uc}^{tt}	0.538	0.807	0.914	0.627	0.841	0.668	0.633	0.818
GMM_{uc}	0.666	0.932	0.923	0.681	0.702	0.542	0.542	0.684
MCS_{iid}	0.003***	0.034**	0.004***	0.001***	0.003***	0.116	0.005***	0.003***
GMM_{iid}	0.001***	0.007***	0.003***	0.001***	0.003***	0.160	0.003***	0.005***
MCS_{cc}	0.003***	0.112	0.009***	0.004***	0.012**	0.306	0.011**	0.013**
GMM_{cc}	0.007***	0.030**	0.014**	0.010**	0.014**	0.374	0.024**	0.018**

Figure 1: Returns, VaR-Forecasts, and VaR-Violations

The figure presents returns, VaR-forecasts, and VaR-violations for the four portfolios considering a VaR-level of 1%. VaR-forecasts are plotted with lines whereas the dashes at the bottom of the charts mark the days on which a VaR-violation occurs.
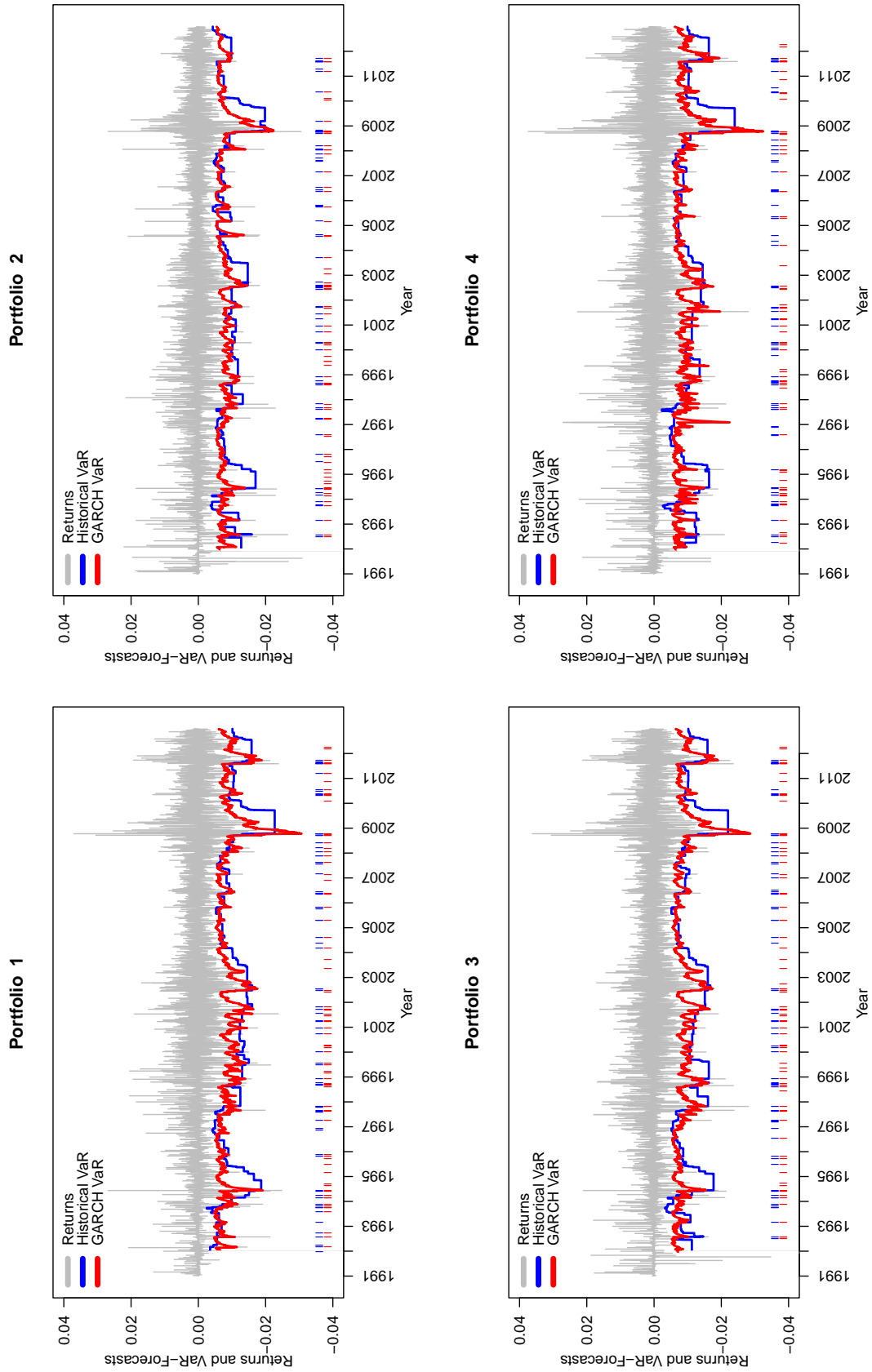
A   Appendix: Pseudocode

A.1   Test of Unconditional Coverage

(i) Generate the violation sequence resulting from the observed returns and the corresponding VaR forecasts by

$$I_i(p) = \begin{cases} 1, & \text{if } y_i < VaR_{i|i-1}(p); \\ 0, & \text{else.} \end{cases}$$

(ii) Draw $l + 1$ random variables by

$$\epsilon_j \sim N(0, 1) \cdot 0.001, \quad j = 1, ..., l + 1.$$

(iii) Calculate the test statistic for the observed violation sequence by

$$MCS_{uc} = \epsilon_{l+1} + \sum_{i=1}^{n} I_i.$$

(iv) Simulate violation sequences by drawing l-times n random variables with distribution

$$\hat{I}_{j,i}(p) \sim \text{Bern}(p), \quad i = 1, ..., n, \quad j = 1, ..., l.$$

(v) Calculate the test statistic for each simulated violation sequence by

$$\hat{MCS}_{uc,j} = \epsilon_j + \sum_{i=1}^{n} \hat{I}_{i,j}.$$

(vi) Sort the resulting values of the simulated statistic $\hat{MCS}_{uc,j}$ in descending order.

(vii) Compute the quantiles for the desired significance level and compare the test statistic for the observed violation sequence to the resulting critical values.

A.2 Test of the I.I.D. Property

(i) Generate the violation sequence resulting from the observed returns and the corresponding VaR forecasts by

$$I_i(p) = \begin{cases} 1, & \text{if } y_i < VaR_{i|i-1}(p); \\ 0, & \text{else.} \end{cases}$$

(ii) Calculate the sum of observed VaR violations by

$$m = \sum_{i=1}^{n} I_i.$$

(iii) Identify the time indexes where an observed VaR violation occurred by

$$V = \{i|I_i = 1\} = (t_1, ..., t_m).$$

(iv) Draw $l + 1$ random variables by

$$\epsilon_j \sim N(0, 1) \cdot 0.001, \quad j = 1, ..., l + 1.$$

(v) Calculate the test statistic for the observed violation sequence by

$$MCS_{iid,m} = t_1^2 + (n - t_m)^2 + \sum_{s=2}^{m}(t_s - t_{s-1})^2 + \epsilon_{l+1}.$$

(vi) Simulate violation sequences by drawing l-times n random variables with distribution

$$\hat{I}_{i,j}(p) \sim \text{Bern}(p), \quad i = 1, ..., n, \quad j = 1, ..., l,$$

under the condition that $\sum_{i=1}^{n} \hat{I}_{i,j} = m, \ \forall \ j$.

(vii) For each simulated violation sequence, identify the set of time indexes of the violations by

$$\hat{V}_j = \{t_j|\hat{I}_{i,j} = 1\} = (t_{j,1}, ..., t_{j,m}).$$

(viii) Calculate the test statistic for the simulated violation sequences by

$$\hat{MCS}_{iid,m,j} = t_{j,1}^2 + (n - t_{j,m})^2 + \sum_{s=2}^{m}(t_{j,s} - t_{j,s-1})^2 + \epsilon_j.$$

(ix) Sort the resulting values of the simulated statistic $\hat{MCS}_{iid,m,j}$ in descending order.

(x) Compute the quantile for the desired significance level and compare the test statistic for the observed violation sequence to the resulting critical value.

A.3 Test of Conditional Coverage

(i) Simulate violation sequences by drawing l-times n random variables with distribution

$$\hat{I}_{i,j}(p) \sim \text{Bern}(p), \ i = 1, ..., n, \ j = 1, ..., l,$$

under the condition that $\sum_{i=1}^{n} \hat{I}_{i,j} > 1, \ \forall \ j$.

(ii) For each simulated violation sequence, identify the set of time indexes of the violations by

$$\hat{V}_j = \{\hat{t}_j | \hat{I}_{j,i} = 1\} = (\hat{t}_{j,1}, ..., \hat{t}_{j,m}).$$

(iii) Draw $l + 1$ random variables by

$$\epsilon_j \sim N(0, 1) \cdot 0.001, \ j = 1, ..., l + 1.$$

(iv) Calculate the violation frequency of each of the simulated sequences

$$\hat{m}_j = \sum_{i=1}^{n} \hat{I}_{i,j}.$$

(v) Define $\hat{m} = (\hat{m}_1, ..., \hat{m}_l)$ and set $\hat{m}_{min} = \max(2, \min(\hat{m}))$ and $\hat{m}_{max} = \max(\hat{m})$ for the lower and upper bound of possible VaR violation frequencies.

(vi) For each $k = \hat{m}_{min}, \hat{m}_{min+1}, \ldots, \hat{m}_{max}$, simulate violation sequences by drawing l^*-times n random variables with distribution

$$\tilde{I}_{i,j}(k/n) \sim \text{Bern}(k/n), \ i = 1, ..., n, \ j = 1, ..., l^*,$$

under the condition that $\sum_{i=1}^{n} \tilde{I}_{i,j}(k/n) = k, \ \forall \ j$.

(vii) For k and each simulated violation sequence, identify the set of time indexes of the violations by

$$\tilde{V}_{j,k} = \{\tilde{t}_{j,k} | \tilde{I}_{i,j,k} = 1\} = (\tilde{t}_{j,1}, ..., \tilde{t}_{j,k}).$$

(viii) For each k, calculate r_k, an estimator for $E(MCS_{iid,k}|H_0)$, by

$$r_k = \frac{1}{l^*} \cdot \sum_{j=1}^{l^*} \left(\tilde{t}_{j,1}^2 + (n - \tilde{t}_{j,k})^2 + \sum_{s=2}^{k} (\tilde{t}_{j,s} - \tilde{t}_{j,s-1})^2 \right).$$

(ix) Calculate the test statistic for each violation sequence simulated in step (i) by

$$M\hat{C}S_{cc,k,j} = af(M\hat{C}S_{uc,j}) + (1 - a)g(M\hat{C}S_{iid,k,j}), 0 \le a \le 1,$$

where

$$f(M\hat{C}S_{uc,j}) = \left| \frac{\left(\epsilon_j + \sum_{i=1}^{n} \hat{I}_i \right)/n - p}{p} \right|,$$

and

$$g(\hat{MCS}_{iid,k,j}) = \frac{\hat{MCS}_{iid,k,j} - r_k}{r_k} \cdot 1_{\{\hat{MCS}_{iid,k,j} \geq r_k\}}, \quad k = \sum_{i=1}^{n} \hat{I}_{i,j}.$$

(x) Sort the resulting values of the simulated statistic $\hat{MCS}_{cc,k,j}$ in descending order.

(xi) Compute the quantile for the desired significance level.

(xii) Generate the violation sequence resulting from the observed returns and the corresponding VaR forecasts by

$$I_i(p) = \begin{cases} 1, & \text{if } y_i < VaR_{i|i-1}(p); \\ 0, & \text{else.} \end{cases}$$

(xiii) Calculate the sum of observed VaR violations by

$$m = \sum_{i=1}^{n} I_i.$$

(xiv) Identify the set of time indexes where an observed VaR violation occurred by

$$V = \{t | I_i = 1\} = (t_1, ..., t_m).$$

(xv) If $m \notin [\hat{m}_{min}, \hat{m}_{min+1}, \ldots, \hat{m}_{max}]$, determine r_m by repeating steps (vi) to (viii) where k is replaced by m.

(xvi) Calculate the test statistic for the observed violation sequence by

$$MCS_{cc,m} = af(MCS_{uc}) + (1 - a)g(MCS_{iid,m}), 0 \leq a \leq 1,$$

where

$$f(MCS_{uc}) = \left| \frac{(\epsilon_{l+1} + \sum_{i=1}^{n} I_i)/n - p}{p} \right|,$$

and

$$g(MCS_{iid,m}) = \frac{MCS_{iid,m} - r_m}{r_m} \cdot 1_{\{MCS_{iid,m} \geq r_m\}}.$$

(xvii) Compare the test statistic for the observed violation sequence to the critical value.