SFB 823

Evaluating value-at-risk forecasts: A new set of multivariate backtests

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

Nr. 9/2015

Discussion Paper

SFB 823

Evaluating Value-at-Risk Forecasts: A New Set of Multivariate Backtests*

Dominik Wied †

Technische Universität Dortmund

Gregor N.F. Weiß‡

Technische Universität Dortmund

Daniel Ziggel §

FOM Hochschule für Oekonomie & Management, Germany

April 10, 2015

ABSTRACT

We propose two new tests for detecting clustering in multivariate Value-at-Risk (VaR) forecasts. First, we consider CUSUM-tests to detect first-order instationarities in the matrix of VaR-violations. Second, we propose χ^2-tests for detecting cross-sectional and serial dependence in the VaR-forecasts. Moreover, we combine our new backtests with a test of unconditional coverage to yield two new backtests of multivariate conditional coverage. In all cases, a bootstrap approximation is possible, but not mandatory in terms of empirical size and power.

Keywords: Model Risk, Multivariate Backtesting, Value-at-Risk.

JEL Classification: C52, C53, C58.

---

*Financial support by the Collaborative Research Center "Statistical Modeling of Nonlinear Dynamic Processes" (SFB 823, projects A1 and A7) of the German Research Foundation (DFG) is gratefully acknowledged. We thank Robert Löser for calculating the general formulas of Σ_S.

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

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

§Feldstraße 88, 46535 D-Dinslaken, Germany, telephone: +49 2064 456 4394, e-mail: *daniel.ziggel@quasol.de*.

Evaluating Value-at-Risk Forecasts: A New Set of Multivariate Backtests

ABSTRACT

We propose two new tests for detecting clustering in multivariate Value-at-Risk (VaR) forecasts. First, we consider CUSUM-tests to detect first-order instationarities in the matrix of VaR-violations. Second, we propose χ^2-tests for detecting cross-sectional and serial dependence in the VaR-forecasts. Moreover, we combine our new backtests with a test of unconditional coverage to yield two new backtests of multivariate conditional coverage. In all cases, a bootstrap approximation is possible, but not mandatory in terms of empirical size and power.

1 Introduction

Over the past two decades, Value at Risk (VaR) has become the prevalent measure for assessing the risk of financial investments. Its widespread use in banking was recognized under the 1996 Market Risk Amendment to the first Basel Accord which allowed banks to employ internal forecasting models to calculate their required regulatory capital. Since then, VaR has become the industry standard for measuring and managing portfolio risk (not only for banks but also, e.g., for insurance companies due to Solvency II) even though it lacks the desirable property of a coherent risk measure (see Artzner et al., 1999) for non-Gaussian Profit & Loss (P/L) distributions. Consequently, not only regulators but also the firms that use VaR themselves have long been interested in assessing the forecasting accuracy of their VaR-models through formal backtesting. Nowadays, risk measures such as the Expected Shortfall, which explicitly take the amount of losses into account, are of increasing importance. Nevertheless, as these measures are still based on the VaR, appropriate backtesting has not lost its importance. In this paper, we address the highly important task of backtesting the VaR-forecasts of several business lines (or sub-portfolios) across several points in time and propose two new multivariate backtests. Such backtesting is the first step in risk analysis; if backtests indicate problems with a forecast model, appropriate actions by the investor are required. In this paper, we focus on this first step.

The backtesting of a VaR-model comprises a comparison of the model's out-of-sample VaR-forecasts and the investment's actual returns. If the investment is a single trading position or a portfolio it yields a univariate time series of VaR-forecasts and VaR-violations. In the last few years, several formal backtests have been proposed in the literature for the case of a univariate sequence of VaR-violations with tests concentrating on the correct number of violations (unconditional coverage, *uc* in short), the independence of the sample of violations, and both properties at the same time (tests of conditional coverage, *cc* in short) (see, e.g., Kupiec, 1995; Christoffersen, 1998; Berkowitz, 2001; Christoffersen and Pelletier, 2004; Engle and Manganelli, 2004; Haas, 2005; Candelon et al., 2011; Berkowitz et al., 2011). Most recently, Ziggel et al. (2014) proposed a set of tests that additionally test for identically distributed violations. None of these backtests, however, can be easily extended to the multivariate case in which VaR-violations might not only be correlated across time but also across business lines.

The key motivation for considering multivariate VaR backtesting is that financial institutions are usually interested in forecasts of their trading desk's aggregate P/L distribution in contrast to VaR-forecasts of isolated investments. However, aggregating individual VaR-forecasts often yields biased results as diversification effects between (sub-)portfolios are not adequately modeled. To tackle this problem, multivariate backtests need to account for cross-sectional dependence within the portfolio.[1] While it may also be possible to directly consider VaR for aggregate portfolios (i.e., for univariate additive combinations of different investments), the results of a (univariate) backtest for these always depends on the type of aggregation. Moreover, and more importantly, a multivariate backtest avoids the problem of multiple testing. Besides this relevance from a practical point of view, multivariate backtests process much more data at once thereby allowing further theoretical applications and improving power properties.

Despite the importance of multivariate backtesting, only few papers in the literature deal with this topic with most papers leaving the development of such tests for future research (see, e.g., Berkowitz et al., 2011; Ziggel et al., 2014). To the best of our knowledge, the only exception is Daniculescu (2010) who proposes a multivariate uc and independence test. The test is based on a multivariate Portmanteau statistic of Ljung-Box type that jointly tests for the absence of autocorrelations and cross-correlations in the vector of hits sequences for different business lines. However, the tests proposed by Daniculescu (2010) suffer from size distortions. Finally, to the best of the authors' knowledge, there currently exists no multivariate backtest that explicitly tests for the i.i.d. property (in contrast to the mere independence) of VaR-violations.[2]

In this paper, we suggest new multivariate backtests for clusters in VaR-violations which are easy to implement and have appealing properties under the null and the alternative. Moreover, these

[1] Acknowledging this need to backtest the VaR-forecasts of a bank holistically, the Basel guidelines explicitly demand that a bank should "[...] perform separate backtests on sub-portfolios using daily data on sub-portfolios subject to specific risk." (see Basel Committee on Banking Supervision (BCBS), 2009).

[2] Note that there exist some papers that deal with VaR backtests in miscellaneous multivariate settings. However, these backtests use multivariate approaches in order to investigate a univariate time series (see, e.g., Hurlin and Tokpavi, 2007).

tests can easily be extended to cc-versions. We essentially propose two different kinds of tests. First, we consider a CUSUM-test for detecting clusters that are caused by instationarities in the mean of the VaR-violations. To take the multidimensionality of the VaR-violations into account, we use the sums of the violations for different business lines and sub-portfolios for a single day. Second, we consider a χ^2-test for detecting clusters that are caused by cross-sectional and/or serial dependencies within the VaR-violations. Finally, we combine our new backtests with a test of unconditional coverage to yield two new backtests of multivariate conditional coverage. All tests are easy to implement and perform well in simulations. Additionally, all tests work without Monte Carlo simulations or bootstrap approximations. However, there are bootstrap approximations available: The one for the CUSUM-tests serves for making it more robust (which does not seem to be necessary, at least in our simulations), while the one for the χ^2-tests is potentially interesting with respect to the test's software implementation.

The rest of this paper is organized as follows. In Section 2, we introduce the notation and the new multivariate backtests. The performance of the new backtests in finite samples is analyzed in simulations in Section 3. Section 4 concludes.

2 Methodology

In this section, we introduce the notation used throughout the paper, define the desirable properties of VaR-violations, and present our new multivariate backtests.

2.1 Notation and VaR-violation Properties

First, we shortly discuss the univariate case in order to extend it in the following to a multivariate setting. Let $\{y_t\}_{t=1}^n$ be the observable part of a time series $\{y_t\}_{t\in\mathbb{Z}}$ corresponding to daily observations of the returns on an asset or a portfolio. We are interested in the accuracy of VaR-forecasts. Following Dumitrescu et al. (2012), the ex-ante VaR $VaR_{t|t-1}(p)$ (conditionally on an information set $\mathbb{F}_{t-1}$) is implicitly defined by $Pr(y_t < -VaR_{t|t-1}(p)) = p$, where p is the VaR coverage probability.

Note that we follow the actuarial convention of a positive sign for a loss. 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 ex-post 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, Ziggel et al. (2014) 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 on average, i.e.,

$$\mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n} I_t(p)\right] = p. \tag{2}$$

Second, VaR-violations should possess the i.i.d. property. Otherwise, the sequence $\{I_t(p)\}$ could exhibit clusters of violations. In fact, unexpected temporal occurrences of clustered VaR-violations may have several potential reasons. On the one hand, $\{I_t(p)\}$ may not be identically distributed and $\mathbb{E}(I_t(p))$ could vary over time. On the other hand, $I_t(p)$ may not be independent of $I_{t-k}(p), \forall k \neq 0$. The hypothesis of i.i.d. VaR-violations holds true if

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

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

Finally, the uc and i.i.d. 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

$$\{I_t(p)\} \overset{i.i.d.}{\sim} \text{Bern}(p), \forall t. \tag{4}$$

Note that in most related studies in the literature, the uc property is defined slightly differently than it is done in this paper, while the full i.i.d. hypothesis is not discussed at all with almost all papers concentrating on the independence property of VaR-violations (see, e.g., Christoffersen, 1998).[3]

At this point, we extend our analysis of VaR-violations to a multivariate setting. To this end, we assume that an m-dimensional time series $\{Y_{t,i}\}_{t=1,i=1}^{n,m}$ of returns exists as well as m sequences of VaR forecasts, $VaR_{t,i|t-1}(p_i)$. We then define the multivariate indicator variable $I_{t,i}(p_i)$ as

$$I_{t,i}(p_i) = \begin{cases} 0, \text{ if } Y_{t,i} \geq -VaR_{t,i|t-1}(p_i); \\ 1, \text{ if } Y_{t,i} < -VaR_{t,i|t-1}(p_i). \end{cases} \tag{5}$$

Here, p_i is the VaR coverage probability for sub-portfolio i. Note that p_i is explicitly allowed to vary among different sub-portfolios and we do not need to assume particular values of $p_i, i = 1, \ldots, m$. In each column, the resulting matrix contains information for a single business line or sub-portfolio (corresponding to the 1-dimensional case), while each row represents a single trading day. In Figure I, we illustrate a stylized matrix of VaR-violations across time and business lines.

[Place Figure I about here]

As can easily be seen from Figure I, clusters of VaR-violations can occur both across time and across sub-portfolios/business lines. Cluster across time indicate a misspecified VaR model, while cluster across sub-portfolios/business lines indicate low potential for diversification.

With this preliminary work, we start to define the desirable properties of VaR-violations in the multivariate case. For the uc hypothesis and most uc tests, an extension of the univariate to the multivariate case is straightforward. To this end, one simply needs to study the hit sequences of several business lines simultaneously and stack the series together. As doing so effectively increases the sample size, we expect the tests to have more power than in the univariate setting.

However, in this paper, we are interested in the multivariate distribution of VaR-violations and hence neglect this simple issue. In the present context, the VaR-violations should ideally exhibit

3 See Ziggel et al. (2014) for a critical discussion of previous treatments of the uc and the independence properties in the literature.

no clusters, i.e., neither in time (rows) nor across business lines (columns). For this, it is necessary that the matrix of VaR-violations fulfils the following multivariate independence hypothesis:

$$I_{t,i}(p_i) \text{ is independent of } I_{t-k,j}(p_j), \forall t, i, j \text{ and } \forall k > 0. \tag{6}$$

Property (6) implies that no information concerning VaR-violations available to the risk manager at the time the VaR is estimated is helpful in forecasting a VaR-violation. Thus, as stated in Berkowitz et al., 2011, past observations from the hit sequence of one business line do not help to predict violations of this or any other business line if the VaR model is correctly specified. In particular, property (6) postulates that lagged violations are not correlated. However, correlations within one row (trading day) are explicitly allowed. In this context, it is also natural to consider the more restrictive assumption

$$I_{t,i}(p_i) \text{ is independent of } I_{t-k,j}(p_j), \forall t, i, j \text{ and } \forall k \geq 0. \tag{7}$$

Although the VaR model is not necessarily wrongly calibrated if property 7 is not fulfilled, it may provide important information concerning diversification and aggregation of risks. For properties (6) and (7), we propose χ^2-tests in Section 2.3.

As stated in Ziggel et al. (2014), clusters of VaR-violations could also be caused by other reasons than simple correlation between violations. To be more precise, the probability of obtaining a VaR-violation may change over time. For example, the risk model could not be suited to incorporate changes from calm market phases to highly volatile bear markets or financial crises, and vice versa. This would in turn lead to clustered VaR-violations regardless of the question whether the data might show signs of dependence or autocorrelation. In Section (2.2), we consider CUSUM-tests for such instationarities. To be more precise, we consider the row sums

$$r_t := \sum_{i=1}^{m} I_{t,i}(p_i) \tag{8}$$

and test whether $\mathbb{E}(r_t)$ is constant over time (stationarity hypothesis). More precisely, we test for first-order instationarities caused by changes in $\mathbb{E}(I_{t,i}(p_i))$, resulting in the following assumption

$$\mathbb{E}\left[\sum_{i=1}^{m} I_{t,i}(p_i)\right] = c, \forall t, \tag{9}$$

where c is an arbitrary constant.

As in the univariate case, one can also define the cc-property in the multivariate setting. Here, properties (6) and (7) are modified to

$$\mathbb{E}(I_{t,i}(p_i)) = p_i \text{ and } I_{t,i}(p_i) \text{ is independent of } I_{t-k,j}(p_j), \forall t, i, j \text{ and } \forall k > 0. \tag{10}$$

and

$$\{I_{t,i}(p_i)\} \overset{i.i.d.}{\sim} \text{Bern}(p_i), \forall t, i. \tag{11}$$

Analogously, property (9) is modified to

$$\mathbb{E}\left[\sum_{i=1}^{m} I_{t,i}(p_i)\right] = \sum_{i=1}^{m} p_i, \forall t. \tag{12}$$

2.2 CUSUM-tests for the cc-property and first-order instationarities

In this subsection, we propose a backtest for first-order instationarities. The formal test problem which corresponds to property (9) is given by

$$H_0^s : \mathbb{E}(r_1) = \ldots = \mathbb{E}(r_n) \text{ vs. } H_1^s : \neg H_0^s.$$

with the row sums $r_1, \ldots, r_n$ being defined as in Equation (8). While the specific expectations are arbitrary in this test problem, this is different in the test problem which corresponds to property (12):

$$H_0^{s-cc} : \mathbb{E}(r_1) = \ldots = \mathbb{E}(r_n) = \sum_{i=1}^{m} p_i \text{ vs. } H_1^s : \neg H_0^s.$$

Before introducing the test statistics, we impose the following assumption:

Assumption 1 *Let r_i be defined as before. Then, we assume*

1. *$r_1, \dots, r_n$ are independent.*

2. *$Var(r_1) = \dots = Var(r_n)$.*

If the VaR model is correctly specified, Assumption 1.1 is a reasonable consequence. Assumption 1.2 may be violated if the cross-sectional dependence between $I_{t,1}, \dots, I_{t,m}$ is not constant over time. We will discuss this issue in detail below. Under H_0^s as well as under H_0^{s-cc} and Assumption 1, respectively, the row sums fulfill a functional central limit theorem, i.e., the process $(V_n, n \in \mathbb{N})$ with

$$V_n(s) = \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor sn \rfloor} (r_t - \mathbb{E}(r_t)), s \in [0, 1],$$

converges to a Brownian motion. Then, a suitable test statistic for H_0^{s-cc} is given by $RC_{cc,n} := D^{-1} C_{cc}$ (RC for "row CUSUM") with

$$C_{cc} := \max_{j=1,\dots,n} \frac{1}{\sqrt{n}} \left| \sum_{t=1}^{j} r_t - j \sum_{i=1}^{m} p_i \right| .$$

Here, D^2 is the usual variance estimator for independent observations, $D^2 = \frac{1}{n} \sum_{t=1}^{n} (r_t - \overline{r})^2$ with $\overline{r} = \frac{1}{n} \sum_{t=1}^{n} r_t$. Then, by means of the continuous mapping theorem we immediately obtain

Theorem 2 *Under H_0^{s-cc} and assumption 1, for $n \to \infty$, it holds that $RC_{cc,n} \to_d \sup_{s \in [0,1]} |W(s)|$, where W is a standard Brownian motion.*

With this preliminary work, we get the

Stat-m-cc-test. Reject H_0^{s-cc} whenever $RC_n > q_{1-\alpha,BM}$, where $q_{1-\alpha,BM}$ is the $1-\alpha$-quantile of the distribution of $\sup_{s\in[0,1]} |W(s)|$. The 0.95-quantile is given by 2.241.

For testing H_0^s, we do not consider any fixed values of p_i, but we use the test statistic $RC_{stat,n} := D^{-1} C_{stat}$ with

$$C_{stat} := \max_{j=1,\dots,n} \frac{1}{\sqrt{n}} \left| \sum_{t=1}^{j} r_t - \frac{j}{n} \sum_{t=1}^{n} r_t \right| .$$

Then, by means of the continuous mapping theorem we immediately obtain

Theorem 3 *Under H_0^s and assumption 1, for $n \to \infty$, it holds that $RC_{stat,n} \to_d \sup_{s\in[0,1]} |B(s)|$, where B is a standard Brownian bridge.*

With this preparatory work, we get the

Stat-m-test. Reject H_0^s whenever $RC_n > q_{1-\alpha,KS}$, where $q_{1-\alpha,KS}$ is the $1-\alpha$-quantile of the Kolmogorov-Smirnov-distribution. The 0.95-quantile is given by 1.358.

It can be shown that both tests are consistent, e.g., if, under the alternative,

$$\mathbb{E}(r_1) = \ldots = \mathbb{E}(r_{\lfloor kn \rfloor}) \neq \mathbb{E}(r_{\lfloor kn \rfloor + 1}) = \ldots = \mathbb{E}(r_n)$$

holds for a $k \in (0, 1)$. In this case, it is possible to estimate the location of a change point by the argmax estimator

$$C_{cc} := \text{argmax}_{j=1,\ldots,n} \frac{1}{\sqrt{n}} \left| \sum_{t=1}^{j} r_t - j \sum_{i=1}^{m} p_i \right|$$

or

$$C_{stat} := \text{argmax}_{j=1,\ldots,n} \frac{1}{\sqrt{n}} \left| \sum_{t=1}^{j} r_t - \frac{j}{n} \sum_{t=1}^{n} r_t \right|,$$

(see Aue and Horváth, 2013).

However, the empirical size is not close to the nominal size if there is either weak serial dependence within the $(r_t, t = 1, \ldots, n)$ (such as α-mixing under appropriate conditions as described in e.g. Billingsley (1968)) and/or if the $Var(r_t)$ are not constant over time. The test is not consistent in these cases. For the case of weak serial dependence, this is an immediate consequence of Slutzky's theorem. However, in this context, we will present a new χ^2-test for cross-sectional and serial dependence in Section 2.3. The issue of non-constant variances is discussed in detail in Zhou (2013). In particular, Zhou (2013) explicitly derives the limit distribution of a general CUSUM-statistic under the assumption of piecewise local stationarity. In our context, it is arguably most relevant in the case of a change in the cross-sectional dependence of the VaR-violations. However, there is a bootstrap approximation available which potentially makes the CUSUM-test more robust to changes in variances and weak serial dependence.

2.2.1 Bootstrap

In order to robustify the tests against changes in $Var(r_t)$ as well as against weak serial dependence, one can use a recently proposed approach by Zhou (2013). Here, we consider the quantity C, i.e., the test statistic without the variance estimator D^{-1}. Critical values are obtained by using a bootstrap approximation. This bootstrap is an extension of the wild bootstrap and relies on directly mimicking the behavior of the partial sum process V_n instead of mimicking the behavior of C. Despite the theoretical relevance, the bootstrap does not seem to be necessary in our specific situation as some robustness checks showed that also the common CUSUM-test is robust against changes in $Var(r_t)$ which are caused by changes in the cross-sectional dependence. Moreover, there is no power gain from the robust CUSUM-test.

2.3 χ^2-tests for the cc-property, cross-sectional and serial dependence

In this subsection, we propose a framework that can be used for testing independence as well as the corresponding cc-hypothesis taking into account arbitrary time lags and business lines. This test is somewhat similar to the test proposed by Daniculescu (2010) with the main difference that we explicitly allow for estimated violation probabilities in each business line and that we make use of explicit expressions for a certain covariance matrix. Note that due to the latter, our test has better size properties than the one proposed by Daniculescu (2010).

Denote with A the set of all triples $(i, j, l), i, j = 1, \ldots, m, l = 0, \ldots, K$, where (i, j) describes a pair of sub-portfolios and l the lag of interest. The convention is that we consider lags up to a fixed upper bound K, e.g. $K = 5$, corresponding to one week. We consider an arbitrary subset $A_s \subseteq A$ that has to be chosen by the analyst. For verifying Assumption (6), A_s could for example consist of all triples with $i \leq j$ and $l = 1$, while it could consist of all triples with $i < j$ and $l = 0$ for verifying (7). However, also other combinations are possible.

The test problem which corresponds to (6) and (7) is given by

$$\begin{aligned} H_0^{m-ind}: &\quad \mathbb{E}((I_{t,i}(p_i)-\tilde{p}_i)(I_{t+l,j}(p_j)-\tilde{p}_j)) = 0 \text{ for } (i,j,l)\in A_s \text{ and } t=1,\ldots,n-l \\ &\quad \text{and some } \tilde{p}_i := \mathbb{E}(I_{t,i}(p_i)), \tilde{p}_j := \mathbb{E}(I_{t,j}(p_j)) \text{ vs.} \\ H_1^{m-ind}: &\quad \neg H_0^{m-ind}. \end{aligned}$$

In the previous test problem, the expectations of $I_{t,i}(p_i)$ and $I_{t,j}(p_j)$ are arbitrary. If one is also interested in testing for them (i.e., for the correct number of VaR-violations), one can consider a modified test problem for the cc-hypothesis. With $f(i,j,l,t) := (I_{t,i}(p_i)-p_i)(I_{t+l,j}(p_j)-p_j)$ and the desired VaR coverage probabilities p_i and p_j,

$$H_0^{m-cc}: \mathbb{E}(f(i,j,l,t)) = 0 \text{ for } (i,j,l)\in A_s \text{ and } t=1,\ldots,n-l \text{ vs. } H_1^{m-cc}: \neg H_0^{m-cc}.$$

First, we consider the cc-test which is based on the vector

$$B_{s,n} := \left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n-l}(I_{t,i}(p_i)-p_i)(I_{t+l,j}(p_j)-p_j)\right)_{(i,j,l)\in A_s}.$$

We impose the following assumption:

Assumption 4 *Let the notation be as before. Then, we assume*

1. *The vectors* $f(i,j,l,1)_{(i,j,l)\in A_s},\ldots,f(i,j,l,n-l)_{(i,j,l)\in A_s}$ *are independent.*

2. $Cov(f(i,j,l,1)_{(i,j,l)\in A_s}) = \ldots = Cov(f(i,j,l,n-l)_{(i,j,l)\in A_s}) =: \Sigma_s$*, where* Σ_s *is a positive definite matrix.*

Under the assumption that the VaR model is correct, Assumption 4.1 is reasonable. Assumption 4.2 contains a higher-order stationarity assumption, as well as a regularity assumption on the matrix

$$\Sigma_s = (Cov((I_{1,i_1}(p_{i_1})-p_{i_1})(I_{1+l_1,j_1}(p_{j_1})-p_{j_1}),(I_{1,i_2}(p_{i_2})-p_{i_2})(I_{1+l_2,j_2}(p_{j_2})-p_{j_2})))_{(i_1,j_1,l_1),(i_2,j_2,l_2)\in A_s}.$$

This matrix can easily be calculated for each given set A_s. If, e.g., $A_s = \{(1,2,0)\}$, it holds that $\Sigma_s = p_1p_2 - p_1^2p_2 - p_1p_2^2 + p_1^2p_2^2$. If $A_s = \{(1,1,1),(2,2,1)\}$, it holds

$$\Sigma_s = \begin{pmatrix} p_1^2 - 2p_1^3 + p_1^4 & \rho_{12}^2 \\ \rho_{12}^2 & p_2^2 - 2p_2^3 + p_2^4 \end{pmatrix}$$

with $\rho_{12} = Cov(I_{t,1}(p_1), I_{t,2}(p_2))$. In these situations, Σ_s is for example positive definite for $0 < p_1 = p_2 < 1$ and $\rho_{12} = 0$.

In general, under Assumption (10) and in the situation in which it holds for all triples (i,j,l) that $i \leq j$ and $l \geq 1$,

$$\Sigma_s = (Cov(I_{1,i_1}(p_{i_1}), I_{1,i_2}(p_{i_2}))Cov(I_{1+l_1,j_1}(p_{j_1}), I_{1+l_2,j_2}(p_{j_2})))_{(i_1,j_1,l_1),(i_2,j_2,l_2)\in A_s}.$$

Under Assumption (11) and in the situation in which it holds for all triples (i,j,l) that $i < j$ and $l = 0$,

$$\Sigma_s = diag(p_ip_j - p_ip_j^2 - p_i^2p_j + p_i^2p_j^2)_{(i,j,0)\in A_s}.$$

Assumption 4.2 could be relaxed to the case in which the matrix $\Sigma_s^* := \lim_{n\to\infty} B_{s,n}$ exists, is positive definite and can be suitably estimated. An example for this would be the case in which $Cov(f(i,j,l,t)_{(i,j,l)\in A_s})$ is piecewise constant with a finite number of breaks and positive definite in all parts, which is a special case of the PLS setting discussed in Zhou (2013) and Section 2.2. In this case, the estimators described below are consistent for Σ_s^*.

Under H_0^{m-cc} and Assumption (4), it holds that $B_{s,n} \to_d N(0,\Sigma_s)$, while this quantity diverges if, e.g., under the alternative, $\mathbb{E}(f(i,j,l,t)) = c \neq 0$ for $(i,j,l) \in A_s$. Furthermore, with the continuous mapping theorem,

$$\Sigma_s^{-1/2} B_{s,n} \to_d N(0, I_{|A_s|})$$

and

$$B_{s,n}'\Sigma_s^{-1}B_{s,n} \to_d \chi^2_{|A_s|}.$$

Therefore, a suitable test statistic for the cc-test is given by $T_{s,n}^{m-cc} := B_{s,n}'(\hat{\Sigma}_s^{cc})^{-1}B_{s,n}$. Here, $\hat{\Sigma}_s^{cc}$ is an appropriate estimator of Σ_s. For $A_s = \{(1,1,1),(2,2,1)\}$, one would replace ρ_{12} with $\frac{1}{T}\sum_{t=1}^{T} I_{t,1}(p_1)I_{t,2}(p_2) - p_1p_2$ in Σ_s. Then, we get by the strong law of large numbers and Slutzky's theorem

Theorem 5 *Under H_0^{m-cc} and Assumption* (4)*, for $n \to \infty$, $T_{s,n}^{m-cc} \to_d \chi^2_{|A_s|}$.*

We obtain the

Ind-m-cc-test. Reject H_0^{m-cc} if $T_{s,n}^{m-cc} > q_{1-\alpha,\chi^2}$, where $q_{1-\alpha,\chi^2}$ is the $1-\alpha$-quantile of the χ^2-distribution with $|A_s|$ degress of freedom.

For testing the ind-property, we opt for a test statistic which is based on the quantity

$$C_{s,n} := \left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n-l} (I_{t,i}(p_i) - \hat{p}_i)(I_{t+l,j}(p_j) - \hat{p}_j) \right)_{(i,j,l)\in A_s}.$$

This is essentially the quantity $B_{s,n}$, whereas the summands $-p_i$ and $-p_j$ are replaced with $-\hat{p}_i$ and $-\hat{p}_j$, respectively. Here, $\hat{p}_i$ and $\hat{p}_j$ are the actual measured percentages of VaR-violations, $\hat{p}_i := \frac{1}{n}\sum_{t=1}^{n} I_{t,i}(p_i)$ and $\hat{p}_j := \frac{1}{n}\sum_{t=1}^{n} I_{t,j}(p_j)$. The test statistic is defined as $T_{s,n}^{m-ind} := C_{s,n}'(\hat{\Sigma}_s^{ind})^{-1}C_{s,n}$. Here, $\hat{\Sigma}_s^{ind}$ is an appropriate estimator of Σ_s. In contrast to $\hat{\Sigma}_s^{cc}$, within $\hat{\Sigma}_s^{ind}$, also the p_i would have to be estimated. Interestingly, the asymptotic behavior of $T_{s,n}^{m-cc}$ and $T_{s,n}^{m-ind}$ are the same, as the following theorem shows. Its proof is deferred to the Appendix.

Theorem 6 *Let H_0^{m-ind} and Assumption* (4) *be true. Then, as $n \to \infty$, $T_{s,n}^{m-ind} \to_d \chi^2_{|A_s|}$.*

We obtain the

Ind-m-Test. Reject H_0^{m-ind} if $T_{s,n}^{m-ind} > q_{1-\alpha,\chi^2}$, where $q_{1-\alpha,\chi^2}$ is the $1-\alpha$-quantile of the χ^2-distribution with $|A_s|$ degress of freedom.

Both the Ind-m-cc-Test and the Ind-m-Test are consistent if, e.g., under the alternative $\mathbb{E}(f(i,j,l,t)) = c \neq 0$ for $(i,j,l) \in A_s$ and if (for the Ind-m-Test) $\tilde{p}_i := \mathbb{E}(I_{t,i}(p_i))$, $\tilde{p}_j := \mathbb{E}(I_{t,j}(p_j))$ as well as Assumption 4 are fulfilled.

2.3.1 Bootstrap

To reduce the computational cost and to facilitate the tests' implementation in software, one can estimate the matrix Σ_s with a bootstrap approximation. The bootstrap is the same for testing H_0^{m-ind} and H_0^{m-cc}, respectively. We distinguish two cases, i.e., Assumptions (6)/(10) and (7)/(11). In the first one, cross-sectional dependence is allowed for, which is not true in the second one. Let B be a sufficiently large number.

Then, under Assumption (6) and given the observed matrix of VaR-violations, we generate, for $b = 1, \ldots, B$, a bootstrap sample $I_{t,i}^b, t = 1, \ldots, n, i = 1, \ldots, m$, by drawing n rows with replacement from the observed matrix. Thus, the generated bootstrap samples always fulfill Assumption (6).

When testing for cross-sectional dependence (that means, if Assumption (7) holds true under the null hypothesis), the bootstrap procedure from the previous paragraph has the drawback that there is no variation within each row in the bootstrap samples. Thus, in this case a bootstrap sample $I_{t,i}^b, t = 1, \ldots, n, i = 1, \ldots, m$, is obtained in a different way. In order to keep the information concerning $p_i, i = 1, \ldots, m$, for fixed i, $I_{t,i}^b, t = 1, \ldots, n$, is obtained by drawing n values with replacement of the respective business line from the observed matrix, whereas the draws are also independent with respect to i. Then, the generated bootstrap samples always fulfill Assumption (7).

Having obtained a bootstrap sample, we calculate the vector $C_{s,n}^b$ and consider the estimator

$$\Sigma_s^B := \frac{1}{B} \sum_{b=1}^{B} (C_{s,n}^b - \bar{C}_{s,n}^B)(C_{s,n}^b - \bar{C}_{s,n}^B)'$$

with $\bar{C}_{s,n}^B := \frac{1}{B} \sum_{b=1}^B C_{s,n}^b$. The test statistic for H_0^{m-cc} is then given by $T_{b,n}^{m-cc} := B_{s,n}'(\Sigma_s^B)^{-1} B_{s,n}$, the one for H_0^{m-iid} is given by $T_{b,n}^{m-iid} := C_{s,n}'(\Sigma_s^B)^{-1} C_{s,n}$. Both need to be compared with the $1-\alpha$-quantile of the $\chi^2_{|A_s|}$-distribution. The validity of this approach under the null hypothesis follows from standard bootstrap theory (bootstrap central limit theorem, uniform integrability), the validity under the alternative follows from the fact that the generated vectors $C_{s,n}^b$ remain stochastically bounded due to the arguments given in the previous paragraph.

Simulations show that the bootstrap tests for (6) and (7) have virtually the same size and power properties as the tests based on an explicit derivation of the matrix Σ_s. While in case of (10) and (11), the bootstrap does work in the sense of accuracy under the null hypothesis and consistency under the alternative, there is some power loss compared to the case in which the matrix Σ_s is calculated directly. Under Assumption (11), a better alternative is given by drawing the $I^b_{t,i}, t = 1, \ldots, n, i = 1, \ldots, m$, independently from Bernoulli distributions with the respective $p_i, i = 1, \ldots, m$.

3 Simulation study

To examine the performance of our newly proposed backtests in finite samples, we perform a simulation study. Within the study, we distinguish between different kinds of controllable violations concerning Assumptions (6), (7), (9), (10), (11), and (12). We compute all rejection rates for a significance level of 5%.

3.1 First-order instationarities

As a first step in our simulation study, we want to test if the new tests are able to detect first-order instationarities. We use the Stat-m-test as well as the Ind-m-test. With the latter, the subset A_s consists of the vectors $(1, 1, 1)$ and $(2, 2, 1)$ which corresponds to a time lag of 1. We expect the CUSUM-test to clearly outperform the χ^2-test in this setting. Basically, the data generating process used throughout the whole simulation study is given by:

$$I_{t,i} = \mathbb{I}(X_{t,i} \leq q_p), \forall i, t. \tag{13}$$

In the first case we consider, q_α is the α-quantile of a standard normal distribution. Moreover, $(X_{t,1}, \ldots, X_{t,m})$ follows a multivariate normal distribution with mean zero, marginal variances one and cross-correlation $\rho = 0$. Next, we set $m = 2$, $p = 0.05, 0.01$, $n = 250, 500, 1,000, 2,000$ and use $5,000$ Monte Carlo repetitions. For $p = 0.01$ and small n, $\hat{\Sigma}^{ind}_s$ is sometimes not invertible because there is not a single VaR-violation in at least one business line. In this and similar cases in the following, we repeat the respective simulation run. Finally, we modify equation (13) such that

$$I_{t,i} = \begin{cases} \mathbb{I}(X_{t,i} \leq q_{p-2\delta}), 1 \leq t \leq \frac{n}{4}, \forall i; \\ \mathbb{I}(X_{t,i} \leq q_{p+\delta}), \frac{n}{4} < t \leq \frac{n}{2}, \forall i; \\ \mathbb{I}(X_{t,i} \leq q_{p-\delta}), \frac{n}{2} < t \leq \frac{3n}{4}, \forall i; \\ \mathbb{I}(X_{t,i} \leq q_{p+2\delta}), \frac{3n}{4} < t \leq n, \forall i. \end{cases}$$

In this setting, VaR-violations are independent over time. Hence, clustering is solely based on changes of the probability of obtaining a VaR-violation. 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 unequally distributed. Note that the probability variations are determined in a way which ensures $\mathbb{E}\left(\sum_{t=1}^{n}\sum_{i=1}^{m} I_{t,i}\right) = n \cdot m \cdot p$. The setup of this part of the simulation study covers a realistic scenario in which VaR-models do 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 regardless of the question whether the data might show signs of dependence or autocorrelation. The results of the simulations are given in Table I.

[Place Table I about here]

The results show that the Stat-m-test clearly outperforms the Ind-m-test with rejection probabilities being regularly higher for the Stat-m-test than for the Ind-m-test. In fact, for the larger sample sizes of $n = 1,000$ and $n = 2,000$ and higher values of δ, the probability of the Stat-m-test to reject the matrix of VaR-violations with first-order instationarities is close to one. In contrast, the Ind-m-test rejects H_0 only in 4% to 46% of the simulations.

3.2 Cross-sectional dependence

In the second part of our simulation study, we want to investigate if the new tests are able to detect cross-sectional dependence within VaR-violations. Here, we expect the χ^2-test to clearly outperform the CUSUM-test. Again, we simulate random variables by

$$I_{t,i} = \mathbb{I}(X_{t,i} \leq q_p), \forall i, t,$$

where, q_α is the α-quantile of a standard normal distribution. Moreover, $(X_{t,1}, \ldots, X_{t,m})$ follows a multivariate normal distribution with mean zero, marginal variances one and cross-correlation ρ. In addition, we choose $m = 2$, $p = 0.05, 0.01$, $n = 250, 500, 1,000, 2,000$, and again use $5,000$ Monte Carlo repetitions. The cross-correlation ρ of the normally distributed random variables is set to be in $\{0, 0.2, 0.4, 0.6, 0.8\}$. Based on this setting, we analyze the Stat-m-test and the Ind-m-test with time lag 0, that means that we test for cross-sectional dependence. The subset A_s consists of the vector $(1, 2, 0)$.

The results are given in Table II.

[Place Table II about here]

As can be seen from the simulation results given in Table II, the probability to detect a matrix of VaR-violations suffering from cross-correlation is almost always lower for the Stat-m-test than for the Ind-m-test. While the Ind-m-test has appealing power properties in almost all settings, the Stat-m-test is not able to detect cross-sectional dependence. However, both tests have reasonable size properties.

Next, we also consider the situation with time lag 1 which implies that we test for serial dependence in the VaR-violations. Here, $(X_{t,1}, \ldots, X_{t,m})$ follows a MA(1)-process with autocorrelation parameter $\phi \in \{0, 0.25, 0.5, 0.75, 1\}$, i.e.,

$$(X_{t,1}, \ldots, X_{t,m}) = \epsilon_t + \phi\epsilon_{t-1}$$

for a sequence of i.i.d. bivariate normally distributed vectors $\epsilon_t, t = 1, \ldots, n$, with cross-correlation set to $\rho_t = 0.3$. The subset A_s consists of the vectors $(1, 1, 1)$ and $(2, 2, 1)$. The indicator variables are defined as $I_{t,i} = \mathbb{I}\left(X_{t,i} \leq q_p \sqrt{1 + \phi^2}\right)$. The results of the simulations in which both multivariate backtests are used on data with serially correlated VaR-violations are given in Table III.

[Place Table III about here]

The results given in Table III show that the Ind-m-test again performs significantly better than the Stat-m-test.

3.3 First-order instationarities and serial dependence

In the third part of our simulation study, we investigate the performance of our new multivariate backtests in a setting in which the data exhibit a combination of first-order instationarities and serial dependence. For this purpose, we define

$$I_{t,i} = \begin{cases} \mathbb{I}(X_{t,i} \leq q_{p-2\delta} \sqrt{1 + \phi^2}), 1 \leq t \leq \frac{n}{4}, \forall i; \\ \mathbb{I}(X_{t,i} \leq q_{p+\delta} \sqrt{1 + \phi^2}), \frac{n}{4} < t \leq \frac{n}{2}, \forall i; \\ \mathbb{I}(X_{t,i} \leq q_{p-\delta} \sqrt{1 + \phi^2}), \frac{n}{2} < t \leq \frac{3n}{4}, \forall i; \\ \mathbb{I}(X_{t,i} \leq q_{p+2\delta} \sqrt{1 + \phi^2}), \frac{3n}{4} < t \leq n, \forall i. \end{cases}$$

Here, $(X_{t,1}, \ldots, X_{t,m})$ follows the same MA(1) process as previously described above. Consequently, we use the same parametrization as before and investigate all parameter combinations of δ and ϕ. This setting ensures that we can draw correct conclusions concerning the characteristics of both tests in various situations. We consider the Stat-m-test and the Ind-m-test with $A_s = \{(1, 1, 1), (2, 2, 1)\}$.

The results are given in Tables IV and V.

[Place Tables IV and V about here]

Again, the results from the simulations show a clear picture. Except for $\phi = 0$, the Ind-m-test always performs significantly better than the Stat-m-test. In contrast, mean rejection probabilities are only higher for the Stat-m-test in the setting in which ϕ is set to zero.

3.4 Violations of the cc-property

Within the last setting of our simulation study, we concentrate on violations of the cc-property. To this end, we simulate data that exhibit serial dependence and also violations of the uc-property. We define

$$I_{t,i} = \mathbb{I}(X_{t,i} \leq q_{p+\delta} \sqrt{1+\phi^2}), \forall i, t.$$

To be more precise, $(X_{t,1}, \ldots, X_{t,m})$ follows the same MA(1) process as before and δ is set to $0.2p$, $0.4p$, $0.6p$, $0.8p$, and p, respectively. Apart from that, we use the same parametrization as before and investigate all parameter combinations of δ and ϕ. We consider the Stat-m-cc-test and the Ind-m-cc-test with $A_s = \{(1,1,1),(2,2,1)\}$. The results are given in Tables VI and VII.

[Place Tables VI and VII about here]

Just like in our simulations with data that exhibit cross-sectional dependence and first-order instationarities together with serial dependence in the matrix of VaR-violations, the Ind-m-test again performs significantly better than the Stat-m-test expect for the setting for $\phi = 0$.

In general, we observe that both the Ind-m-test and the Stat-m-test have sufficient power in almost all settings of our simulation study even for relatively small sample sizes of $n = 250$. Moreover, both tests also hold their nominal level in almost all simulation settings. The simulations thus underline the suitability of our newly proposed backtests for testing the adequacy of a multivariate VaR model.

4 Conclusion

In this paper, we have proposed two new multivariate backtests for clusters in VaR-violations. The first test is a CUSUM-test which is based on the sums of the violations for different business lines and sub-portfolios for a single day and which attempts to detect clusters in the matrix of VaR-violations that are caused by instationarities in the mean of the violations. Second, we consider a χ^2-test for detecting clusters that are caused by cross-sectional and/or serial dependencies within the VaR-violations. Both tests are easy to implement and work without Monte Carlo simulations or bootstrap approximations, although bootstrap approximations are readily available.

In simulations, we assess the performance of our new multivariate backtests in several distinct settings in which we consider simulated data that exhibit first-order instationarities, cross-sectional dependence, and serial dependence in the VaR-violations. Moreover, we also perform simulations in which the new backtests are used to test the simulated VaR-violations for the property of conditional coverage. With the exception of the setting in which the data only exhibit first-order instationarities, the χ^2-test performs better in our simulations than the CUSUM-test. Both tests hold their nominal level and, more importantly, have considerable power for testing the conditional coverage of the matrix of VaR-violations even for relatively small sample sizes.

While the multivariate backtests that we propose are intended for the use by risk managers in individual banks, one can easily think of further applications. For example, our multivariate tests could be used to backtest a whole banking sector with VaRs being estimated across time and individual banks (instead of business lines) with clusters in VaR-violations across banks indicating systemic risk in the sector. In this way, our backtests could be of significant help to regulators to forecast times of contagion in the financial system and thereby complement current endeavours to stress-test banking sectors (see, e.g., Acharya and Steffen, 2013).

Appendix

A.1 Proof of Theorem 6

First, we consider the process

$$\tilde{C}_{s,n} := \left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n-l} (I_{t,i}(p_i) - \tilde{p}_i)(I_{t+l,j}(p_j) - \tilde{p}_j) \right)_{(i,j,l)\in A_s},$$

and show that $\tilde{C}_{s,n} = C_{s,n} + o_p(1)$.

We define $\hat{p}_k := \frac{1}{n}\sum_{t=1}^{n} I_{t,k}(p_k)$. Then, it holds

$$\tilde{C}_{s,n} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n-l} I_{t,i}(p_i) I_{t+l,j}(p_j) - \frac{n-l}{\sqrt{n}} \hat{p}_i \tilde{p}_j - \frac{n-l}{\sqrt{n}} \hat{p}_j \tilde{p}_i + \frac{n-l}{\sqrt{n}} \tilde{p}_i \tilde{p}_j + o_p(1)$$

and

$$C_{s,n} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n-l} I_{t,i}(p_i) I_{t+l,j}(p_j) - \frac{n-l}{\sqrt{n}} \hat{p}_i \hat{p}_j - \frac{n-l}{\sqrt{n}} \hat{p}_j \hat{p}_i + \frac{n-l}{\sqrt{n}} \hat{p}_i \hat{p}_j + o_p(1)$$

such that

$$\begin{aligned}
\tilde{C}_{s,n} - C_{s,n} &= \frac{n-l}{\sqrt{n}}(-\hat{p}_i\tilde{p}_j + \hat{p}_i\hat{p}_j) + \frac{n-l}{\sqrt{n}}(-\hat{p}_j\tilde{p}_i + \hat{p}_j\hat{p}_i) + \frac{n-l}{\sqrt{n}}(\tilde{p}_i\tilde{p}_j - \hat{p}_i\hat{p}_j) \\
&= \frac{n-l}{\sqrt{n}}(\tilde{p}_i\tilde{p}_j - \hat{p}_i\tilde{p}_j + \hat{p}_j\hat{p}_i - \hat{p}_j\tilde{p}_i) = \frac{n-l}{\sqrt{n}}(\tilde{p}_j(\tilde{p}_i - \hat{p}_i) + \hat{p}_j(\hat{p}_i - \tilde{p}_i)) \\
&= \frac{n-l}{\sqrt{n}}(\hat{p}_i - \tilde{p}_i)(\hat{p}_j - \tilde{p}_j) \\
&= O_p(1)o_p(1) = o_p(1).
\end{aligned}$$

Then, the result from the theorem follows from the fact that, by uniform integrability, one directly obtains $\Sigma_s = \lim_{n\to\infty} Cov(C_{s,n})$ and $C_{s,n} \to_d N(0, \Sigma_s)$.

■

References

ACHARYA, V. V. AND S. STEFFEN (2013): "Falling Short of Expectations? Stress-Testing the European Banking System," *Working Paper*.

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

AUE, A. AND L. HORVÁTH (2013): "Structual Breaks in Time Series," *Journal of Time Series Analysis*, 34, 1–16.

BASEL COMMITTEE ON BANKING SUPERVISION (BCBS) (2009): "Revisions to the Basel II market risk framework," `http://www.bis.org/publ/bcbs158.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.

BILLINGSLEY, P. (1968): *Convergence of Probability Measures*.

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.

DANICULESCU, C. (2010): "Backtesting Value-at-Risk Models: A Multivariate Approach," *SSRN working paper (1591049)*.

DUMITRESCU, E., C. HURLIN, AND V. PHAM (2012): "Backtesting Value-at-Risk: from dynamic quantile to dynamic binary tests," *working paper*.

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

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

HURLIN, C. AND S. TOKPAVI (2007): "Backtesting value-at-risk accuracy: a simple new test," *Journal of Risk*, 9(2), 19–37.

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

Zhou, Z. (2013): “Heteroscedasticity and Autocorrelation Robust Structural Change Detection,” *Journal of the American Statistical Association*, 108, 726–740.

Ziggel, D., T. Berens, G. N. Weiss, and D. Wied (2014): “A New Set of Improved Value-at-Risk Backtests,” *Journal of Banking and Finance*, 48, 29–41.

Figures and Tables

Figure I: Multivariate Value-at-Risk hit matrix.

The Figure presents a stylized matrix of Value-at-Risk (VaR) violations for a given firm with m business lines (or sub-portfolios) and evaluations for n days. If the realized return in business line i on day j exceeds the corresponding VaR-forecast, the respective entry in the hit matrix is one, and zero otherwise. Stylized clusters of VaR-violations in time (third column) and across business lines (first row) are highlighted.

	Line 1	Line 2	Line 3	Line 4	Line 5	Line 6	···	Line m
Day 1	0	**1**	**1**	**1**	**1**	0	···	0
Day 2	0	0	**1**	0	1	0	···	0
Day 3	1	0	**1**	1	0	1	···	0
Day 4	0	0	**1**	0	1	0	···	1
Day 5	0	1	0	0	1	0	···	0
⋮	⋮							⋮
Day n	0	0	1	0	1	0	···	0

Table I: Simulated rejection probabilities for first-order instationarities.

The table presents the rejections probabilities of the Stat-m-test and the Ind-m-test based on simulated data with first-order instationarities with $p = 0.05$ (Panel A) and $p = 0.01$ (Panel B).

Panel A: $p = 0.05$

		Stat-m-test						Ind-m-test					
δ/p		0	0.1	0.2	0.3	0.4	0.5	0	0.1	0.2	0.3	0.4	0.5
n	250	0.04	0.06	0.13	0.26	0.49	0.76	0.05	0.05	0.06	0.08	0.11	0.15
	500	0.04	0.09	0.27	0.59	0.90	1.00	0.04	0.05	0.06	0.08	0.14	0.21
	1,000	0.05	0.14	0.54	0.93	1.00	1.00	0.04	0.05	0.07	0.11	0.18	0.29
	2,000	0.04	0.28	0.88	1.00	1.00	1.00	0.04	0.05	0.07	0.13	0.25	0.46

Panel B: $p = 0.01$

		Stat-m-Test						Ind-m-Test					
δ/p		0	0.1	0.2	0.3	0.4	0.5	0	0.1	0.2	0.3	0.4	0.5
n	250	0.03	0.03	0.04	0.06	0.08	0.11	0.05	0.04	0.05	0.05	0.05	0.07
	500	0.03	0.04	0.06	0.11	0.14	0.24	0.07	0.08	0.08	0.10	0.11	0.12
	1,000	0.04	0.05	0.10	0.20	0.35	0.56	0.09	0.09	0.11	0.12	0.15	0.16
	2,000	0.04	0.07	0.20	0.45	0.78	0.99	0.07	0.08	0.09	0.10	0.13	0.15

Table II: Simulated rejection probabilities for cross-sectional correlation.

The table presents the rejection probabilities of the Stat-m-test and the Ind-m-test based on simulated data with cross-sectional correlation ρ, $p = 0.05$ (Panel A) and $p = 0.01$ (Panel B).

Panel A: $p = 0.05$

		Stat-m-test					Ind-m-test				
	ρ	0	0.2	0.4	0.6	0.8	0	0.2	0.4	0.6	0.8
n	250	0.04	0.04	0.04	0.04	0.03	0.05	0.23	0.56	0.87	0.99
	500	0.04	0.04	0.04	0.04	0.04	0.04	0.29	0.77	0.98	1.00
	1000	0.05	0.05	0.04	0.04	0.05	0.04	0.43	0.94	1.00	1.00
	2000	0.04	0.05	0.04	0.04	0.05	0.05	0.68	1.00	1.00	1.00

Panel B: $p = 0.01$

		Stat-m-test					Ind-m-test				
	ρ	0	0.2	0.4	0.6	0.8	0	0.2	0.4	0.6	0.8
n	250	0.04	0.04	0.04	0.04	0.03	0.05	0.23	0.56	0.87	0.99
	500	0.04	0.04	0.04	0.04	0.04	0.04	0.29	0.77	0.98	1.00
	1000	0.05	0.05	0.04	0.04	0.05	0.04	0.43	0.94	1.00	1.00
	2000	0.04	0.05	0.04	0.04	0.05	0.05	0.68	1.00	1.00	1.00

Table III: Simulated rejection probabilities for serial dependence.

The table presents the rejections probabilities of the Stat-m-test and the Ind-m-test based on simulated data with autocorrelation ϕ, cross-correlation $\rho = 0.3$, $p = 0.05$ (Panel A) and $p = 0.01$ (Panel B).

Panel A: $p = 0.05$

		Stat-m-test					Ind-m-test				
	ϕ	0	0.25	0.5	0.75	1	0	0.25	0.5	0.75	1
n	250	0.04	0.06	0.09	0.10	0.11	0.05	0.33	0.73	0.86	0.90
	500	0.04	0.07	0.10	0.12	0.12	0.05	0.51	0.92	0.98	0.99
	1000	0.05	0.07	0.11	0.13	0.14	0.04	0.74	1.00	1.00	1.00
	2000	0.04	0.07	0.11	0.12	0.14	0.04	0.94	1.00	1.00	1.00

Panel B: $p = 0.01$

		Stat-m-Test					Ind-m-Test				
	ϕ	0	0.25	0.5	0.75	1	0	0.25	0.5	0.75	1
n	250	0.03	0.03	0.04	0.05	0.06	0.05	0.15	0.33	0.43	0.47
	500	0.03	0.04	0.05	0.06	0.07	0.08	0.27	0.52	0.65	0.69
	1000	0.04	0.05	0.07	0.07	0.08	0.09	0.39	0.71	0.84	0.87
	2000	0.04	0.05	0.07	0.08	0.08	0.08	0.49	0.87	0.96	0.98

Table IV: Simulated rejection probabilities for first-order instationarities and serial dependence with $p = 0.05$.

The table presents the rejections probabilities of the Stat-m-test and the Ind-m-test based on simulated data that exhibit a combination of first-order instationarities and serial dependence with autocorrelation ϕ and cross-correlation $\rho = 0.3$.

Panel A: $\phi = 0$

		Stat-m-test						Ind-m-test					
δ/p		0	0.1	0.2	0.3	0.4	0.5	0	0.1	0.2	0.3	0.4	0.5
n	250	0.04	0.06	0.13	0.23	0.42	0.66	0.06	0.05	0.07	0.08	0.10	0.15
	500	0.04	0.08	0.24	0.52	0.86	1.00	0.04	0.05	0.06	0.09	0.14	0.21
	1000	0.05	0.14	0.50	0.90	1.00	1.00	0.04	0.05	0.07	0.11	0.18	0.29
	2000	0.04	0.26	0.84	1.00	1.00	1.00	0.04	0.05	0.07	0.13	0.24	0.46

Panel B: $\phi = 0.25$

		Stat-m-test						Ind-m-test					
δ/p		0	0.1	0.2	0.3	0.4	0.5	0	0.1	0.2	0.3	0.4	0.5
n	250	0.06	0.09	0.16	0.28	0.45	0.68	0.33	0.35	0.38	0.42	0.48	0.56
	500	0.07	0.11	0.28	0.56	0.87	1.00	0.50	0.53	0.56	0.61	0.71	0.80
	1000	0.07	0.18	0.52	0.89	1.00	1.00	0.75	0.75	0.80	0.85	0.92	0.96
	2000	0.08	0.30	0.84	1.00	1.00	1.00	0.94	0.95	0.97	0.99	0.99	1.00

Panel C: $\phi = 0.5$

		Stat-m-test						Ind-m-test					
δ/p		0	0.1	0.2	0.3	0.4	0.5	0	0.1	0.2	0.3	0.4	0.5
n	250	0.09	0.11	0.19	0.32	0.49	0.69	0.71	0.71	0.73	0.77	0.81	0.86
	500	0.10	0.15	0.32	0.58	0.86	1.00	0.91	0.93	0.93	0.95	0.97	0.98
	1000	0.11	0.22	0.56	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	2000	0.11	0.35	0.85	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Panel D: $\phi = 0.75$

		Stat-m-test						Ind-m-test					
δ/p		0	0.1	0.2	0.3	0.4	0.5	0	0.1	0.2	0.3	0.4	0.5
n	250	0.11	0.12	0.20	0.32	0.50	0.71	0.86	0.86	0.87	0.90	0.92	0.94
	500	0.12	0.17	0.33	0.61	0.87	1.00	0.98	0.99	0.99	0.99	0.99	1.00
	1000	0.13	0.25	0.57	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	2000	0.14	0.37	0.86	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Panel E: $\phi = 1$

		Stat-m-test						Ind-m-test					
δ/p		0	0.1	0.2	0.3	0.4	0.5	0	0.1	0.2	0.3	0.4	0.5
n	250	0.11	0.14	0.21	0.34	0.50	0.70	0.89	0.91	0.90	0.92	0.94	0.96
	500	0.13	0.19	0.34	0.61	0.87	1.00	0.99	0.99	0.99	0.99	1.00	1.00
	1000	0.13	0.26	0.58	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	2000	0.14	0.39	0.85	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table V: Simulated rejection probabilities for first-order instationarities and serial dependence with $p = 0.01$.

The table presents the rejections probabilities of the Stat-m-test and the Ind-m-test based on simulated data that exhibit a combination of first-order instationarities and serial dependence with autocorrelation ϕ and cross-correlation $\rho = 0.3$.

Panel A: $\phi = 0$

		Stat-m-test						Ind-m-test					
δ/p		0	0.1	0.2	0.3	0.4	0.5	0	0.1	0.2	0.3	0.4	0.5
n	250	0.02	0.03	0.04	0.06	0.08	0.11	0.04	0.04	0.05	0.06	0.06	0.06
	500	0.03	0.04	0.06	0.10	0.15	0.22	0.08	0.07	0.08	0.09	0.12	0.11
	1000	0.04	0.06	0.10	0.19	0.33	0.52	0.10	0.09	0.11	0.11	0.15	0.15
	2000	0.04	0.08	0.21	0.42	0.74	0.99	0.07	0.08	0.09	0.10	0.12	0.15

Panel B: $\phi = 0.25$

		Stat-m-test						Ind-m-test					
δ/p		0	0.1	0.2	0.3	0.4	0.5	0	0.1	0.2	0.3	0.4	0.5
n	250	0.03	0.03	0.04	0.07	0.09	0.12	0.16	0.17	0.18	0.18	0.19	0.21
	500	0.04	0.05	0.07	0.11	0.18	0.23	0.28	0.28	0.29	0.31	0.34	0.35
	1000	0.05	0.07	0.12	0.21	0.35	0.54	0.40	0.40	0.42	0.43	0.46	0.50
	2000	0.05	0.09	0.22	0.45	0.76	0.99	0.47	0.49	0.51	0.54	0.58	0.64

Panel C: $\phi = 0.5$

		Stat-m-test						Ind-m-test					
δ/p		0	0.1	0.2	0.3	0.4	0.5	0	0.1	0.2	0.3	0.4	0.5
n	250	0.04	0.05	0.06	0.08	0.11	0.14	0.33	0.34	0.34	0.36	0.37	0.39
	500	0.05	0.07	0.09	0.12	0.18	0.26	0.54	0.52	0.53	0.55	0.56	0.60
	1000	0.07	0.07	0.13	0.24	0.37	0.55	0.71	0.71	0.71	0.75	0.76	0.80
	2000	0.06	0.11	0.23	0.46	0.76	0.99	0.86	0.86	0.88	0.89	0.91	0.93

Panel D: $\phi = 0.75$

		Stat-m-test						Ind-m-test					
δ/p		0	0.1	0.2	0.3	0.4	0.5	0	0.1	0.2	0.3	0.4	0.5
n	250	0.05	0.06	0.06	0.09	0.12	0.16	0.44	0.45	0.44	0.45	0.47	0.50
	500	0.06	0.07	0.09	0.14	0.20	0.28	0.66	0.66	0.66	0.67	0.69	0.72
	1000	0.07	0.09	0.16	0.25	0.39	0.56	0.86	0.85	0.86	0.87	0.87	0.90
	2000	0.09	0.13	0.26	0.49	0.76	0.99	0.96	0.96	0.96	0.96	0.97	0.98

Panel E: $\phi = 1$

		Stat-m-test						Ind-m-test					
δ/p		0	0.1	0.2	0.3	0.4	0.5	0	0.1	0.2	0.3	0.4	0.5
n	250	0.05	0.06	0.07	0.09	0.13	0.16	0.47	0.47	0.49	0.49	0.51	0.53
	500	0.07	0.08	0.11	0.14	0.20	0.29	0.69	0.67	0.69	0.70	0.73	0.75
	1000	0.09	0.10	0.16	0.25	0.39	0.57	0.87	0.87	0.88	0.89	0.91	0.91
	2000	0.08	0.14	0.27	0.48	0.76	0.98	0.97	0.97	0.97	0.98	0.98	0.99

Table VI: Simulated rejection probabilities for violation of the cc-property with $p = 0.05$.

The table presents the rejections probabilities of the Stat-m-test and the Ind-m-test based on simulated data that violate the cc-property with autocorrelation ϕ and cross-correlation $\rho = 0.3$.

Panel A: $\phi = 0$

		Stat-m-cc-test						Ind-m-cc-test					
δ/p		0	0.2	0.4	0.6	0.8	1.0	0	0.2	0.4	0.6	0.8	1.0
n	250	0.06	0.09	0.31	0.61	0.85	0.96	0.06	0.11	0.18	0.26	0.36	0.45
	500	0.05	0.19	0.63	0.91	0.99	1.00	0.05	0.11	0.19	0.29	0.40	0.53
	1000	0.06	0.38	0.92	1.00	1.00	1.00	0.04	0.12	0.22	0.33	0.47	0.61
	2000	0.05	0.71	1.00	1.00	1.00	1.00	0.05	0.13	0.23	0.39	0.58	0.77

Panel B: $\phi = 0.25$

		Stat-m-cc-test						Ind-m-cc-test					
δ/p		0	0.2	0.4	0.6	0.8	1.0	0	0.2	0.4	0.6	0.8	1.0
n	250	0.08	0.12	0.33	0.61	0.83	0.94	0.35	0.50	0.63	0.73	0.82	0.89
	500	0.08	0.22	0.62	0.90	0.99	1.00	0.52	0.69	0.80	0.90	0.96	0.98
	1000	0.07	0.41	0.91	1.00	1.00	1.00	0.74	0.88	0.96	0.99	1.00	1.00
	2000	0.07	0.69	1.00	1.00	1.00	1.00	0.94	0.98	1.00	1.00	1.00	1.00

Panel C: $\phi = 0.5$

		Stat-m-cc-test						Ind-m-cc-test					
δ/p		0	0.2	0.4	0.6	0.8	1.0	0	0.2	0.4	0.6	0.8	1.0
n	250	0.10	0.16	0.34	0.60	0.82	0.93	0.70	0.82	0.90	0.95	0.97	0.98
	500	0.10	0.23	0.62	0.89	0.98	1.00	0.90	0.96	0.99	1.00	1.00	1.00
	1000	0.10	0.41	0.89	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00
	2000	0.10	0.69	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Panel D: $\phi = 0.75$

		Stat-m-cc-test						Ind-m-cc-test					
δ/p		0	0.2	0.4	0.6	0.8	1.0	0	0.2	0.4	0.6	0.8	1.0
n	250	0.13	0.16	0.36	0.61	0.81	0.92	0.83	0.92	0.96	0.99	0.99	0.99
	500	0.11	0.26	0.61	0.88	0.98	1.00	0.98	0.99	1.00	1.00	1.00	1.00
	1000	0.12	0.43	0.89	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	2000	0.12	0.68	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Panel E: $\phi = 1$

		Stat-m-cc-test						Ind-m-cc-test					
δ/p		0	0.2	0.4	0.6	0.8	1.0	0	0.2	0.4	0.6	0.8	1.0
n	250	0.13	0.16	0.36	0.60	0.81	0.92	0.86	0.94	0.97	0.99	0.99	1.00
	500	0.12	0.24	0.61	0.89	0.98	1.00	0.98	1.00	1.00	1.00	1.00	1.00
	1000	0.12	0.43	0.88	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	2000	0.12	0.68	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table VII: Simulated rejection probabilities for violation of the cc-property with $p = 0.01$.

The table presents the rejections probabilities of the Stat-m-test and the Ind-m-test based on simulated data that violate the cc-property with autocorrelation ϕ and cross-correlation $\rho = 0.3$.

Panel A: $\phi = 0$

		Stat-m-cc-test						Ind-m-cc-test					
δ/p		0	0.2	0.4	0.6	0.8	1.0	0	0.2	0.4	0.6	0.8	1.0
n	250	0.03	0.02	0.03	0.06	0.12	0.20	0.05	0.07	0.10	0.12	0.14	0.18
	500	0.07	0.05	0.09	0.20	0.36	0.53	0.09	0.13	0.18	0.23	0.28	0.31
	1000	0.06	0.07	0.24	0.48	0.72	0.89	0.17	0.22	0.27	0.28	0.32	0.34
	2000	0.05	0.15	0.50	0.83	0.97	1.00	0.04	0.08	0.12	0.18	0.27	0.36

Panel B: $\phi = 0.25$

		Stat-m-cc-test						Ind-m-cc-test					
δ/p		0	0.2	0.4	0.6	0.8	1.0	0	0.2	0.4	0.6	0.8	1.0
n	250	0.03	0.03	0.04	0.08	0.13	0.21	0.19	0.26	0.30	0.35	0.41	0.46
	500	0.09	0.06	0.11	0.21	0.37	0.52	0.33	0.42	0.49	0.57	0.65	0.69
	1000	0.07	0.08	0.23	0.49	0.72	0.88	0.53	0.62	0.69	0.74	0.77	0.80
	2000	0.06	0.16	0.51	0.83	0.97	1.00	0.43	0.54	0.66	0.77	0.85	0.91

Panel C: $\phi = 0.5$

		Stat-m-cc-test						Ind-m-cc-test					
δ/p		0	0.2	0.4	0.6	0.8	1.0	0	0.2	0.4	0.6	0.8	1.0
n	250	0.03	0.03	0.06	0.10	0.15	0.23	0.37	0.44	0.51	0.56	0.62	0.68
	500	0.09	0.08	0.12	0.22	0.37	0.54	0.57	0.66	0.73	0.79	0.85	0.89
	1000	0.08	0.10	0.25	0.50	0.73	0.87	0.80	0.87	0.91	0.94	0.96	0.97
	2000	0.07	0.18	0.50	0.82	0.96	0.99	0.83	0.90	0.95	0.97	0.99	0.99

Panel D: $\phi = 0.75$

		Stat-m-cc-test						Ind-m-cc-test					
δ/p		0	0.2	0.4	0.6	0.8	1.0	0	0.2	0.4	0.6	0.8	1.0
n	250	0.03	0.04	0.06	0.10	0.16	0.24	0.48	0.56	0.60	0.68	0.71	0.75
	500	0.10	0.09	0.14	0.23	0.38	0.54	0.69	0.77	0.84	0.88	0.92	0.94
	1000	0.09	0.11	0.27	0.50	0.72	0.86	0.89	0.95	0.96	0.98	0.99	0.99
	2000	0.09	0.19	0.52	0.82	0.96	0.99	0.94	0.98	0.99	1.00	1.00	1.00

Panel E: $\phi = 1$

		Stat-m-cc-test						Ind-m-cc-test					
δ/p		0	0.2	0.4	0.6	0.8	1.0	0	0.2	0.4	0.6	0.8	1.0
n	250	0.04	0.04	0.06	0.11	0.17	0.24	0.51	0.57	0.64	0.70	0.74	0.78
	500	0.11	0.09	0.14	0.24	0.39	0.53	0.71	0.78	0.85	0.90	0.94	0.95
	1000	0.10	0.11	0.27	0.50	0.71	0.86	0.91	0.95	0.97	0.99	0.99	0.99
	2000	0.08	0.19	0.53	0.82	0.96	0.99	0.96	0.98	0.99	1.00	1.00	1.00