SFB 823

A simple and focused backtest of value at risk

Walter Krämer, Dominik Wied

Nr. 17/2015

Discussion Paper

A simple and focused backtest of value at risk[1]

by

Walter Krämer and Dominik Wied
Fakultät Statistik, Technische Universität Dortmund, Germany
Phone: 0231/755-3125 (W. Krämer), 0231/755-5419 (D. Wied), Fax: 0231/755-5284
e-mail: walterk@statistik.tu-dortmund.de, wied@statistik.tu-dortmund.de

Abstract

We suggest a simple improvement of recent VaR-backtesting procedures based on time intervals between VaR-exceedances and show via Monte Carlo that our test has more power than its competitors against empirically relevant clustering alternatives.

Keywords: backtesting, power, value at risk

JEL: C52, C53, C58

1 Introduction

Despite various well known shortcomings, value at risk (VaR) is still the most popular measure of portfolio risk in practice. Therefore, there is interest in the statistical properties of methods employed in its production.

Ideally,

$$P(y_t \leq VaR_t(p)) = p \ \forall t \tag{1}$$

[1]Research supported by DFG-Sonderforschungsbereich 823 (projects A1,Z)

where y_t is the sequence of returns of the security or the portfolio of securities in question, p is the probability (usually 1%) specified by the user of an extreme event, $VaR_t(p)$ is an estimator of the p-quantile of the return-distribution at t based on information available up to $t-1$, and

$$P(y_t \leq VaR_t(p), y_s \leq VaR_s(p)) = P(y_t \leq VaR_t(p)) \times P(y_s \leq VaR_s(p)). \tag{2}$$

Condition (1) (*unconditional coverage*) requires that VaR does what it is supposed to do, while condition (2) implies that information available up to $t-1$ is used efficiently. Both conditions combined can also be rephrased as

$$P_{t-1}(y_t \leq VaR_t(p)) = p \ \forall t, \tag{3}$$

where P_{t-1} is the probability conditional on information up to $t-1$ (*conditional coverage*).

While there is a large literature on how VaR_t is best produced (see Ardia et al. (2014)) or Herwartz et. al. (2015) for recent contributions), and various tests of (1) have also been around for quite a while, testing the condition (2) or (3) has received less methodological attention. Below we build upon Christoffersen (1998), Christoffersen and Pelletier (2004), Haas (2005), Candelon et al. (2011) and Ziggel et al. (2014) to construct a simple procedure to test this independence requirement which has high power to detect violations which occur in clusters. This appears important in practice, since a correct forecast of VaR is extremely important in periods of financial turmoil, where large losses often happen in succession. Then a cluster of exceedances implies that the particular VaR-measure employed has not been sufficiently adjusted downwards, risking losses that are even larger than expected. The practical relevance of weeding out VaR-procedures which are prone to this type of mistake is obvious.

2 A new test for independent VaR-violations

Let $y_1, ..., y_T$ be the returns under consideration, let $t_1, ..., t_n$ be the times where VaR-violations occur, and let $t_0 = 0$. Unconditional coverage requires that

$$E\left(\frac{n}{T}\right) = p. \tag{4}$$

Independence of VaR-violation requires, that, in addition, waiting times between violations (durations) follow a geometrical distribution, in particular, that

$$E(t_i - t_{i-1}) = \frac{1}{p}. \tag{5}$$

Christoffersen and Pelletier (2004) and Haas (2005) propose twin tests of (4) and (5) against parametric alternatives. Ziggel et al. (2014) improve upon these procedures by looking at squared durations, which are better able to detect various nonparametric deviation from the null. In this paper we propose another nonparametric improvement which is focused on condition (5). To that extent, let

$$d_i := t_i - t_{i-1}, i = 1, \ldots, n \tag{6}$$

be the n durations between successive VaR-violations. We include the waiting time up to t_1, but exclude the time elapsing from t_n to the end of the sample period, as this does not follow a geometric distribution. We suggest to look at the *inequality* of the d_i's (as measured by any inequality coefficient such as the Gini-index) as an indicator of possible violations of independence. As all conventional inequality measures g are both homogeneous of degree zero, i.e.

$$g(d_1, \ldots, d_n) = g(ad_1, \ldots, ad_n) \ \ (a > 0), \tag{7}$$

and population invariant, i.e.

$$g(d_1, \ldots, d_n) = g(d_1, \ldots, d_n, d_1, \ldots, d_n), \tag{8}$$

this test is asymptotically insensitive to violations of unconditional coverage, and focused on violations of independence. To the extent therefore that joint tests of unconditional coverage and independence involve some trade-off of power, our new test might be better able to detect violations of independence alone.

For concreteness, we argue in terms of the Gini-coefficient, which may be defined (among various equivalent definitions) as

$$g(d_1, ..., d_n) = \frac{\frac{1}{n^2}\sum_{i,j=1}^{n}(d_i - d_j)}{2\bar{d}}, \tag{9}$$

i.e. Gini's mean difference divided by twice the arithmetic mean. For geometrically distributed d's, the population Gini coefficient is

$$g(d) = \frac{1-p}{2-p} \tag{10}$$

(Dorfmann (1979)), which is between 0 ($p \to 1$) and 1/2 ($p \to 0$). Therefore we suggest a one-sided test of independence which rejects whenever $g(d_1, \ldots, d_n)$ is too large. While independence might also be violated by having the observed VaR-violation too equally spaced in the $[1, T]$-interval, this does not seem to induce a problem in practice, as illustrated by Figure 1.

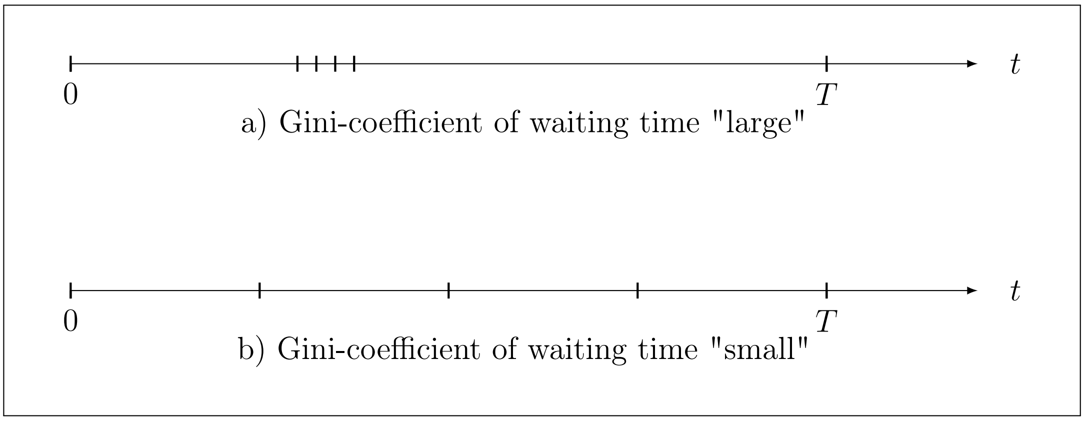

Figure 1: Two types of violation of the independence assumption

3 Finite sample power

It is certainly possible to base an asymptotically valid test on the limiting distribution of the difference between sample and population Gini (after plugging an estimator of p into formula (10)), i.e., to rely on the test statistic

$$\sqrt{T}\left(g(d_1,\ldots,d_n) - \frac{1-n/T}{2-n/T}\right).$$

Instead, we here use the simulated finite sample null distribution. This appears preferable as, even for large T, the number n of VaR-violations is rather small, and asymptotic arguments are hard to justify. Therefore, for fixed T and n, we obtain critical values by simulating the Gini-index $10,000$ times.

For ease of comparison with previous results, we repeat the simulation setup of Ziggel et al. (2014) who consider two types of alternatives; on the one hand dependent VaR-violations, on the other hand non-identical distributions. The first type of violation is the one our test is designed to detect. The second type of violation is added to check its performance also in situations where clusters are possibly induced by time-varying expectations of the durations.

For the dependence case, we have

$$y_t = \sigma_t z_t, \text{ with } \sigma_1 = 1 \text{ and} \tag{11}$$

$$\sigma_t^2 = \lambda\sigma_{t-1}^2 + (1-\lambda)z_{t-1}^2, 0 \leq \lambda \leq 1, t > 1, \tag{12}$$

where the z_t are i.i.d. standard normal and λ measures the degree of dependence: For $\lambda = 1$, we have serial independence (i.e., the null is true) and serial dependence increases as $\lambda \to 0$. We consider three different *half-life intervals* (the interval in which the weight of an observed value of σ^2 decreases to half its original value) $\log(0.5)/\log(\lambda)$. (Half-life intervals $5, 40, 80$ correspond to $\lambda = 0.8706, 0.9828, 0.9914$.) VaR-violations are the points where the y_t are smaller than their empirical p-quantile.

Table 1 gives the results, based on $10,000$ simulation runs. LR_{iid}^{mar} is the likelihood ratio test proposed by Christoffersen (1998) against a first-order Markov alternative, LR_{iid}^{wei} is the test in the framework of the Weibull and Gamma distribution proposed by Christoffersen and Pelletier (2004), GMM_{iid} is the test in the GMM framework proposed by Candelon et al. (2011) and MCS_{iid} is the test based on Monte Carlo simulations proposed by Ziggel et al. (2014). The empirical rejection probabilities of the latter tests are taken from Ziggel et al. (2014) (Table 4,5). It is seen that our test outperforms its competitors when dependence is large. As dependence decreases, this advantage drops and Ziggel et al. (2014) becomes the most powerful among the procedures considered here.

Table 1: Empirical rejection probabilities for dependence in VaR-violations, $\alpha = 0.05$,

p	Halflife	T	Our test	LR_{iid}^{mar}	LR_{iid}^{wei}	GMM_{iid}	MCS_{iid}
0.05	5	252	0.285	0.146	0.033	0.213	0.220
		1000	0.728	0.217	0.160	0.591	0.552
		2500	0.969	0.515	0.396	0.905	0.858
	40	252	0.077	0.077	0.031	0.115	0.128
		1000	0.180	0.052	0.024	0.219	0.251
		2500	0.330	0.073	0.010	0.324	0.397
	80	252	0.059	0.072	0.041	0.089	0.117
		1000	0.105	0.047	0.026	0.224	0.263
		2500	0.170	0.065	0.013	0.267	0.323
0.01	5	252	0.129	0.181	0.035	0.136	0.141
		1000	0.250	0.230	0.137	0.211	0.182
		2500	0.428	0.384	0.362	0.363	0.255
	40	252	0.061	0.199	0.027	0.142	0.155
		1000	0.135	0.083	0.077	0.154	0.176
		2500	0.246	0.119	0.126	0.223	0.238
	80	252	0.054	0.302	0.025	0.131	0.127
		1000	0.095	0.083	0.085	0.157	0.181
		2500	0.158	0.116	0.118	0.194	0.220

Next, we turn to non-identical distributions. This is not the alternative our test has been designed for, but it might still be interesting to know about its performance here as well. Following Ziggel et al. (2014), non-identical distributions are generated by

$$I_t = \begin{cases} \overset{i.i.d}{\sim} Bern(p - 2\delta), & 1 \leq t \leq \frac{T}{4} \\ \overset{i.i.d}{\sim} Bern(p + \delta), & \frac{T}{4} \leq t \leq \frac{T}{2} \\ \overset{i.i.d}{\sim} Bern(p - \delta), & \frac{T}{2} \leq t \leq \frac{3T}{4} \\ \overset{i.i.d}{\sim} Bern(p + 2\delta), & \frac{3T}{4} \leq t \leq T, \end{cases} \tag{13}$$

where I_t are indicator variables that are equal to 1 for a VaR-violation and where the degree of instationarity is measured by δ; $\delta = 0$ means identical distributions, i.e., the null hypothesis is true. As δ increases, the expected values of the I_t become more different over time.

Tables 2 gives the results. It shows that our test performs slightly worse than the Ziggel et al. (2014) test, but still outperforms all the others. The considerable power in these situations is reasonable because the changes in expectations, as simulated here, facilitate cluster effects.

4 Conclusion

We have shown that it is easy to improve upon existing tests for independent VaR violations when possible violations are highly clustered. As this is at the same time the situation most dangerous in applications, our procedure seems to have considerable practical appeal.

Table 2: Empirical rejection probabilities for instationarities in VaR-violations, $\alpha = 0.05$,

p	δ	T	Our test	LR_{iid}^{mar}	LR_{iid}^{wei}	GMM_{iid}	MCS_{iid}
0.05	$0 \cdot p$	252	0.049	0.048	0.053	0.049	0.053
		1000	0.049	0.046	0.046	0.046	0.050
		2500	0.051	0.051	0.049	0.049	0.051
	$0.1 \cdot p$	252	0.057	0.052	0.048	0.058	0.060
		1000	0.073	0.048	0.032	0.066	0.074
		2500	0.083	0.049	0.037	0.078	0.093
	$0.3 \cdot p$	252	0.127	0.061	0.023	0.105	0.130
		1000	0.397	0.054	0.024	0.386	0.456
		2500	0.704	0.085	0.038	0.697	0.771
	$0.5 \cdot p$	252	0.372	0.104	0.028	0.317	0.378
		1000	0.994	0.124	0.794	1.000	1.000
		2500	1.000	0.311	1.000	1.000	1.000
0.01	$0 \cdot p$	252	0.036	0.056	0.042	0.052	0.050
		1000	0.045	0.048	0.046	0.049	0.051
		2500	0.048	0.049	0.047	0.050	0.053
	$0.1 \cdot p$	252	0.036	0.054	0.042	0.049	0.050
		1000	0.049	0.053	0.049	0.054	0.056
		2500	0.061	0.055	0.042	0.056	0.064
	$0.3 \cdot p$	252	0.036	0.057	0.033	0.054	0.060
		1000	0.095	0.064	0.034	0.076	0.091
		2500	0.223	0.070	0.058	0.193	0.242
	$0.5 \cdot p$	252	0.046	0.069	0.011	0.066	0.081
		1000	0.233	0.087	0.051	0.197	0.225
		2500	0.786	0.099	0.437	0.822	0.926

References

Ardia, D. and Hoogerheide, L. F. (2014) GARCH Models for daily stock returns: Impact of estimation frequency on Value-at-Risk and Expected Shortfall forecasts. *Economic Letters* **123**, 187-190.

Candelon, B. and Colletaz, G. and Hurlin ,C. and Tokpavi, S. (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 PELLETIER, D. (2004). Backtesting Value-at-Risk: A duration-based approach. *Journal of Financial Econometrics* **2 (1)**, 84-108.

DORFMANN, R. (1979). A formula for the Gini-coefficient. *Review of Economics and Statistics* **61**, 146-149.

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

HERWARTZ, H. and RATERS, F. (2015), Copula-MGARCH with continuous covariance decomposition. *Economic Letters* **133**, 73-76.

ZIGGEL, D., BERENS, T.. WEISS, G.N. and WIED, D. (2014). A New Set of Improved Value-at-Risk Backtests. *Journal of Banking and Finance* **48**, 29-41.