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

by

Walter Krämer[2] **and Dominik Wied**[3]
TU Dortmund, University of Cologne

Abstract

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

Keywords: backtesting, power, value at risk

JEL: C52, C53, C58

[1]Research supported by DFG-Sonderforschungsbereich 823. We are grateful to Daniel Vogel and an anonymous referee for helpful discussions and comments.

[2]TU Dortmund, Faculty of Statistics, e-mail: walterk@statistik.tu-dortmund.de

[3]corresponding author, University of Cologne, Institute for Econometrics and Statistics, Albertus-Magnus-Platz, 50923 Cologne, Phone: 0221/470-4514, Fax: 0221/470-5084 / TU Dortmund, Faculty of Statistics, email: wied@statistik.tu-dortmund.de

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,

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

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 conditions (2) or (3) has received much 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 deviations from the null: Given the number of exceedances, the sum of squared durations increases as the inequality among the durations increases, which points to a violation of equation (5).

In this paper we propose another nonparametric improvement which is likewise 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 then suggest to directly 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 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.

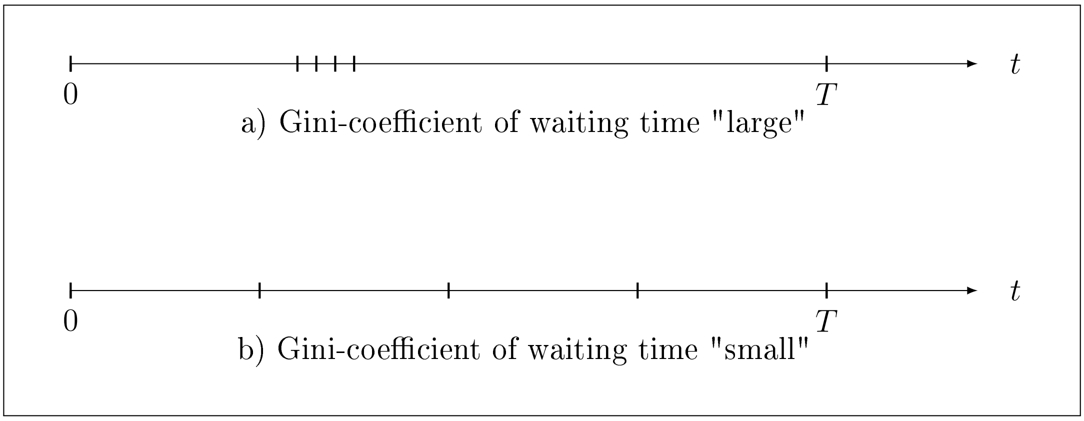

Figure 1: Two types of violation of the independence assumption

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 use a simulation setup similar to Ziggel et al. (2014) who consider two types of alternatives; on the one hand dependent VaR-violations, on the other hand non-identical distributions. Both alternatives tend to produce clusters in the t_i.

For the dependence case, we have

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

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

where the z_t are i.i.d. standard normal and λ measures the degree of dependence: For $\lambda = 0$, we have serial independence (i.e., the null is true) and serial dependence increases as $\lambda \to 1/2$. By construction, the unconditional variance remains equal to 1 and a VaR-violation occurs when y_t drops below the empirical p-quantile of all $y's$.

Table 1 gives the results, based on $10,000$ simulations. LR_{iid}^{mar} is the likelihood ratio test for independence proposed by Christoffersen (1998) against a first-order Markov alternative. (Here and elsewhere, we do not include results for additional tests for unconditional and conditional coverage, as we think it would be unfair to check these outside their home territory, so to speak.) GMM_{iid} is the χ^2-type test in the GMM framework proposed by Candelon et al. (2011) based on empirical moments of orthonormal polynomials, evaluated at the d_i. Under the null, these moments have expectation 0 and the test rejects if a suitable quadratic form in these empirical moments is too large. MCS_{iid} is the test based on squared durations proposed by Ziggel et al. (2014). For all the tests concerned, critical values are obtained by Monte Carlo to make their power more easily comparable (no size distortions). It is seen that, or this particular alternative, our test outperforms MCS_{iid}, but is dominated by LR_{iid}^{mar}.

Next, we consider non-identical distributions. Following Ziggel et al. (2014),

Table 1: Empirical rejection probabilities for dependent durations ($\alpha = 0.05$)

p	λ	T	Our test	LR_{iid}^{mar}	GMM_{iid}	MCS_{iid}
0.05	0	252	0.050	0.044	0.043	0.049
		1000	0.052	0.050	0.049	0.056
		2500	0.056	0.051	0.048	0.062
	0.1	252	0.078	0.111	0.056	0.060
		1000	0.108	0.086	0.067	0.084
		2500	0.158	0.253	0.079	0.105
	0.2	252	0.110	0.178	0.069	0.082
		1000	0.203	0.217	0.109	0.127
		2500	0.350	0.627	0.164	0.192
	0.3	252	0.156	0.243	0.091	0.102
		1000	0.339	0.394	0.190	0.189
		2500	0.611	0.858	0.336	0.327
	0.4	252	0.222	0.308	0.125	0.142
		1000	0.514	0.556	0.308	0.280
		2500	0.838	0.956	0.597	0.500
0.01	0	252	0.050	0.050	0.050	0.052
		1000	0.046	0.047	0.047	0.053
		2500	0.050	0.049	0.053	0.047
	0.1	252	0.063	0.088	0.056	0.060
		1000	0.067	0.152	0.053	0.060
		2500	0.081	0.228	0.065	0.061
	0.2	252	0.076	0.128	0.066	0.065
		1000	0.092	0.254	0.067	0.077
		2500	0.118	0.441	0.086	0.085
	0.3	252	0.090	0.165	0.068	0.079
		1000	0.113	0.313	0.082	0.095
		2500	0.167	0.562	0.126	0.123
	0.4	252	0.107	0.177	0.066	0.087
		1000	0.145	0.354	0.097	0.119
		2500	0.214	0.626	0.161	0.156

these 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, for this alternative, our test is slightly worse than the Ziggel et al. (2014) test, but outperforms all the others. The empirical rejection probabilities of our competitors are taken from Ziggel et al. (2014) (Table 4,5).

4 Conclusion

We have shown that it is easy to improve upon existing tests for independent VaR violations for certain alternatives 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. However, none of the procedures considered by us uniformly dominates the others.

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.

Table 2: Empirical rejection probabilities for non-identically distributed durations ($\alpha = 0.05$)

p	δ	T	Our test	LR_{iid}^{mar}	GMM_{iid}	MCS_{iid}
0.05	$0 \cdot p$	252	0.050	0.048	0.049	0.053
		1000	0.051	0.046	0.046	0.050
		2500	0.051	0.051	0.049	0.051
	$0.1 \cdot p$	252	0.058	0.052	0.058	0.060
		1000	0.076	0.048	0.066	0.074
		2500	0.083	0.049	0.078	0.093
	$0.3 \cdot p$	252	0.130	0.061	0.105	0.130
		1000	0.403	0.054	0.386	0.456
		2500	0.704	0.085	0.697	0.771
	$0.5 \cdot p$	252	0.378	0.104	0.317	0.378
		1000	0.995	0.124	1.000	1.000
		2500	1.000	0.311	1.000	1.000
0.01	$0 \cdot p$	252	0.038	0.056	0.052	0.050
		1000	0.045	0.048	0.049	0.051
		2500	0.049	0.049	0.050	0.053
	$0.1 \cdot p$	252	0.037	0.054	0.049	0.050
		1000	0.049	0.053	0.054	0.056
		2500	0.062	0.055	0.056	0.064
	$0.3 \cdot p$	252	0.037	0.057	0.054	0.060
		1000	0.095	0.064	0.076	0.091
		2500	0.224	0.070	0.193	0.242
	$0.5 \cdot p$	252	0.047	0.069	0.066	0.081
		1000	0.233	0.087	0.197	0.225
		2500	0.788	0.099	0.822	0.926

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.