Resampling vs. Parameter Correction: Ways of Dealing with Parameter Uncertainty in Testing VaR

Marta Małecka[1]

Abstract

Since 2010 there is a new line of research connected with testing VaR, which addresses the problem of the parameter uncertainty or, in other words, estimation risk. The presence of the estimation risk implies that the observed VaR violation process may not fulfill the standard postulates that underpin the testing framework. In consequence, there is a risk that statistical tests may reject correct VaR models due to the estimation errors committed when predicting VaR. In statistical terms, it means that the estimation risk distorts the test size. The previous studies suggested dealing with this issue by the resampling. We propose to solve the problem of parameter uncertainty in testing VaR without resorting to the time-consuming simulations. Our study focuses on the design of the forecasting scheme and its influence on the estimation risk. Specifically, we study the impact of the parameter correction frequency. We compare the effects of the parameter corrections with the effects of correcting for the estimation errors by the resampling methods. We show that our proposition is both less time-demanding and more efficient at ensuring the proper test size.

Keywords: VaR tests, parameter uncertainty, estimation risk, test size

JEL Classification: C12, G18, G32, D53

1. Introduction

Due to the current international standards of banking supervision, testing Value-at-Risk (VaR) continues to be a topical issue in financial literature. These standards, though reformed substantially in years 2012-2017 (Basel Committee on Banking Supervision, 2017), have left the VaR measure as a basis of the risk model evaluation. The important role of VaR in the banking supervision stimulates the inflow of new testing methods (eg. Berkowitz *et al.*, 2011; Candelon *et al.*, 2011; Ziggel *et al.*, 2014; Kramer and Wied, 2015; Pelletier and Wei, 2016; Pajhede, 2017). The new developments are mainly aimed at improving upon the statistical properties of the early VaR tests (Kupiec, 1996; Christoffersen, 1998). However, since 2010, there is a new line of research connected with testing VaR. This line treats the problem of the parameter uncertainty or, in other words, estimation risk. The studies dedicated to this problem

[1] Corresponding author: University of Łódź, Department of Statistical Methods, Rewolucji 1905 r. 41/43, 90-214 Łódź, Poland, marta.malecka@uni.lodz.pl.

focus on showing the influence of the parameter uncertainty on one of the two main properties of statistical tests – the test size.

The issue of the parameter uncertainty in the VaR testing context was defined by Escanciano and Olmo (2010). They studied the uncertainty inherent to the procedure of estimating a VaR model and noted that the estimation errors made in this procedure change the properties of the VaR violation process. This process indicates whether the forecasted VaR was exceeded or not and is a basis of the statistical inference about VaR. Assuming the absence of the estimation risk and the correct VaR model, the VaR violation process should satisfy two properties: the unconditional coverage and independence. The unconditional coverage property refers to the overall VaR violation probability and states that this probability should be equal to the chosen VaR level. The independence property requires the VaR violations to be independent of the past violations or, more generally, of all available market information. These two properties jointly mean that the VaR violation process should be iid Bernoulli. However, the presence of the estimation risk implies that the observed VaR violation process may not fulfill the iid Bernoulli postulate. This may happen not because of the VaR model incorrectness, but because of the estimation errors. In consequence, there is a risk that statistical tests may reject correct VaR models due to the estimation errors committed when predicting VaR. In statistical terms, it means that the estimation risk distorts the test size, which is the parameter that indicates the probability of rejecting correct models.

The previous studies on the parameter uncertainty in testing VaR suggested dealing with this issue by the resampling (Escanciano and Olmo, 2011; Candelon, 2011; Dumitrescu *et al.*, 2012). In particular, Candelon *et al.* (2011) advocated the subsampling method, which is a special case of the resampling. The idea behind this method is to replace the asymptotic distribution of the test statistic with the simulated distribution that takes account of the estimation risk. This simulated distribution is created by repeating many times the process of testing on smaller subsets of the initial data. At each repetition, the testing is preceded by the process of the VaR model parameter estimation. This mimics the real inference process and introduces the estimation risk. However, such simulations are very time-consuming. In practice, the implementation of the subsampling method would require conducting these simulations at each statistical test. The complexity and time demands of such an approach may impede its introduction to the business practice.

We propose to solve the issue of parameter uncertainty in testing VaR without resorting to the time-consuming simulations. Our study focuses on a different factor that potentially influences the estimation risk, which is the parameter correction frequency. We note that the previous studies are all based on the simplifying assumption that predicting VaR relies on the so-called fixed forecasting scheme. This scheme assumes that VaR forecasts are obtained from a model, whose parameters are estimated only once, on a fixed number of beginning observations. With this scheme, the parameter estimates are never corrected. Our intuition is that introducing the parameter correction could substantially reduce the negative effects of the estimation errors. Therefore, we study the impact of the parameter correction, done with a predefined frequency, on the estimation risk. We compare the effects of the parameter correction with the effects of correcting for the estimation errors by the subsampling method. We show that our proposition is both less time-demanding and more efficient at ensuring the proper test size.

Our study is organized as follows. In Section 2 we introduce the notion of the estimation risk in testing VaR and we present the ways of handling this issue. Section 3 compares the previously proposed subsampling technique with the newly proposed method, based on adjustments made to the forecasting scheme. The final section summarizes and concludes.

2. Parameter Uncertainty in Testing VaR

Influence of parameter uncertainty on statistical tests

The fundamental property of a statistical test, which is used to evaluate its quality, is the test size. It measures the probability of the type one error, which, in the VaR testing framework, corresponds to the probability of rejecting a correct VaR model. Ideally, this probability should be equal to the chosen significance level. Then such a test is called accurate.

The standard way of gauging the test size in the VaR evaluation framework does not take into account the estimation risk. It is based on simulations, whose starting point is the VaR violation process. This process compares the returns R_t to the *p*-level VaR forecast $VaR_{t|t-1,p}(\theta)$ and it is defined as $I_{t,p}(\theta)=1\{R_t<VaR_{t|t-1,p}(\theta)\}$. This definition is heavily dependent on the parameter vector θ. In reality, the "perfect" values of this parameter vector are unknown, which is the source of the estimation risk. What we really have is $VaR_{t|t-1,p}(\theta)$ and the observed violation process is $I_{t,p}(\theta)=1\{R_t<VaR_{t|t-1,p}(\theta)\}$.

In such a setting, even under the assumption of the correctness of the VaR model, the $I_{t,p}(\theta)$ process may not fulfill the postulates of the unconditional coverage and independence. Thus the tests that are built on these two postulates are likely to reject correct VaR models.

The negative impact of the estimation risk has motivated the development of methods that are aimed at dealing with this issue. The first such method was the derivation of the asymptotic distributions of the test statistics under the presence of the quantified estimation risk (Escanciano and Olmo, 2010). However, this method has been regarded as “difficult” by its authors, who pointed out technical problems. These problems are connected with the increasing complexity and the quick inflow of the new tests. As an alternative to deriving modified asymptotic distributions, it has been proposed to rely on the resampling methods (Escanciano and Olmo, 2010; Escanciano and Olmo, 2011; Candelon, 2011; Dumitrescu, 2012). Their idea is to replace the theoretical test statistic distributions with the simulated distributions that incorporate the estimation risk. On the contrary to this, our proposition is to come back to the asymptotic distributions, however, with the assumption of introducing frequent parameter corrections.

Dealing with parameter uncertainty through subsampling

One of the resampling methods advocated in the literature to ensure the correct VaR test size is the subsampling (Candelon, 2011). It consists in repeating the process of estimation and testing on smaller subsets of the initial data. Each of the subsets of the length b of the data $\{R_1, ..., R_T\}$ serves to mimic the real decision-making process.

The subsampling method requires the choice of the subsample length b, which then determines the number of possible repetitions N_b. Previous studies on this issue deliver some guidelines about the choice of b. Escanciano and Olmo (2011) suggest $b = KP^{2/5}$, $K \in \{65, 70, 75, 80\}$, $P = 1000$, where P is the size of the tesing sample. Using also the results of Candelon *et al.* (2011), we rely on $b = KP^{2/5}$, $K = 65$, $P = 1000$.

The subsampling method delivers the approximation of the test statistic distribution that incorporates the estimation risk. This method, however, has the drawback of being extremely time-consuming. More importantly, the previous studies have shown that the tests based on the subsampled distributions, though performing better than the asymptotic tests, are still not accurately sized. For this reason, we seek another alternative. We propose to deal with the issue of the parameter uncertainty by introducing their timely corrections. Intuitively, the parameter corrections have the potential to reduce estimation errors. We compare the influence of such corrections to the effects of using the subsampled distributions.

Dealing with parameter uncertainty through parameter corrections

The subsampling method described above assumes that the process of inference about a VaR model is based on the fixed forecasting scheme. It means that the model parameters θ are estimated only once from R beginning observations and then used to produce VaR forecasts throughout the out-of-sample period of the length P. Such a forecasting scheme obviously yields substantial estimation errors. These errors can be reduced by introducing the parameter corrections. We introduce them by replacing a fixed forecasting scheme with a rolling forecasting scheme.

In the rolling scheme, the estimation is based on a window consisting of a fixed number of observations. The estimates from such a window are used to produce the 1-step-ahead to f-step-ahead forecasts. At first, this window covers the R beginning observations. Then it is moved forward in the sample by adding more recent observations and dropping the oldest ones. This process of "rolling" is done across the whole sample. Such a rolling scheme is popular in practice as it enables institutions to continuously forecast VaR.

The rolling forecasting scheme requires the choice of the parameter correction frequency, measured by f. The more frequent the corrections (the lower f), the larger reduction of the estimation error is attainable. Under the sufficient frequency of the corrections, the parameter estimates should be close enough to their true values to allow for the use of the standard asymptotic distributions. Therefore, we suggest using the standard distributions instead of the subsampled ones. We study the impact of the parameter correction on the test size assuming a daily data with the correction frequency set subsequently to 10 days, 5 days and one day, i.e. $f = 1, 5, 10$.

3. Test Size under Parameter Uncertainty

We investigate the influence of the parameter uncertainty on the size of the VaR tests by means of the Monte Carlo (MC) study. The test size gives the probability of rejecting correct models. Ideally, it should be exactly equal to the chosen significance level. The influence of the parameter uncertainty on the test size results from the estimation process that precedes the testing procedure. As an outcome of the estimation process, we get the VaR forecasts and the sequence of violations. This sequence, $\{I_{t,p}(\theta)\}$, is not based on the "perfect" parameters θ_0, but on the ones subject to the estimation risk θ. Thus, even under the correct VaR model, it may not comply with the postulate of independence. As a consequence, the statistical tests based on the standard distributions are likely to reject correct models. This negative effect may be gauged by comparing the test size under the absence and presence of the parameter uncertainty.

We first study the compliance between the test size and the significance level under the absence of the parameter uncertainty. Such an exercise gives us the benchmark size to evaluate both its distortion caused by the estimation errors and its improvement gained from the resampling method or from introducing a rolling forecasting scheme. Since this initial study assumes the absence of the estimation risk, it is based on the abstract "perfect" VaR forecasts. Under the correct VaR model, such forecasts yield a violation process that fulfills the independence postulate. We draw this process from the Bernoulli trials with the probability of success set to the VaR level. We conduct our study for the commonly used 5% VaR level, three significance levels, 0.01, 0.05 and 0.1, and the sample size $T = 1000$. The test size is evaluated over 10 000 MC repetitions. We include two leading VaR independence tests – the LR_{ind}^{M} test based on the Markov-chain framework and the GMM_{VaR} test build within the duration-based approach.

Table 1. Corrected and uncorrected size of VaR independence tests

Significance level	LR_{ind}^{M}	GMM_{VaR}
Absence of parameter uncertainty		
0.01	0.021	0.022
0.05	0.108	0.053
0.1	0.204	0.088
Presence of parameter uncertainty, estimation risk uncorrected. fixed forecasting scheme		
0.01	0.118	0.141
0.05	0.223	0.291
0.1	0.287	0.372
Presence of parameter uncertainty, estimation risk corrected		
0.01	0.011	0.128
0.05	0.077	0.147
0.1	0.171	0.180
Presence of parameter uncertainty, estimation risk uncorrected. parameter correction every 10 days		
0.01	0.007	0.010
0.05	0.045	0.052
0.1	0.085	0.104
Presence of parameter uncertainty, estimation risk uncorrected. parameter correction every 5 days		
0.01	0.008	0.010
0.05	0.044	0.046
0.1	0.085	0.094
Presence of parameter uncertainty, estimation risk uncorrected. parameter correction every day		
0.01	0.009	0.009
0.05	0.043	0.043
0.1	0.086	0.090

The size results under the absence of the estimation risk (Table 1) show substantial differences between the tests with regard to their accuracy. The GMM_{VaR} test, built on the duration-based approach, largely improves upon the size properties of the standard Markov-chain LR_{ind}^{M} test. Its estimated size is much closer to the chosen significance levels.

The size results also show that the test accuracy is worst when operating on the lowest of the examined significance levels and it tends to improve for the higher levels of 0.05 and 0.1. For these significance levels, the differences between the estimated size and the nominal significance level of the GMM_{VaR} test may be regarded as negligible. Thus, under the absence of the estimation risk, the GMM_{VaR} test seems satisfactory in terms of its statistical accuracy.

Taking into account the parameter uncertainty requires substantial changes in our simulation study. The testing procedure needs to be preceded by the estimation process and the violation sequence needs to be modified accordingly. To this purpose, we need in-sample and out-of-sample data, which we assume to be of the lengths $R = 500$ and $P = 1000$, respectively. This accounts for the total sample length $T = R + P = 1500$. To mimic the estimation process, we employ a GARCH-class model[2] with the conditional mean of the form $R_t = \sqrt{h_t} Z_t,\ Z_t:\ N(0,1)$ and the conditional variance of the form $h_t = \omega + \alpha_1 R_{t-1}^2 + \beta_1 h_{t-1}$. Using VaR forecasts from this model, we obtain a sample of the violation process $\left\{ I_{t,p}(\theta) \right\}$. On this sample, we conduct the VaR tests. Evaluation of their size is done over 10 000 MC repetitions, where each repetition involves both the estimation and testing. These estimates are referred to as uncorrected sizes and they show the distortions to the initial sizes caused by the estimation errors.

The outcomes show that the test size under the presence of the estimation risk differs drastically from the initial size results. The only conclusions that remain unchained are those connected with relation between the tests and tendencies referring to the significance level. Still, the GMM_{VaR} test appears superior to the LR_{ind}^{M} procedure and the test properties improve for higher significance levels than 0.01. However, the estimated size substantially moves away from the nominal significance levels. For the 0.01 significance level, the observed size results are more than ten times higher. For the 0.1 level, when the tests are most accurate, the nominal levels are more than tripled. These results demonstrate the dramatic influence of the estimation

[2] We set the parameters of this model to the standard values $\omega_1 = 0.0001$, $\alpha_1 = 0.14$ and $\beta_1 = 0.85$, which mimic the behavior of the real financial returns, assuming the daily observation frequency.

risk on the test accuracy. They show, that omitting this issue in test evaluation, leads to the misleading conclusion that available procedures permit the control over the error of rejecting correct VaR models.

Correcting for the estimation errors by means of the subsampling, as proposed in previous studies, is based on a simulated distribution. This distribution replaces the standard asymptotic χ^2 distribution. In order to incorporate the estimation error correction, we conduct the simulations in a way that mimics all elements: estimation, finding the corrected distribution and testing. To this end, the subsets $\{R_k, ..., R_{k+b-1}\}$ are divided into the in-sample parts of the length R_b and the out-of-sample parts of the length P_b, where $\frac{R_b}{P_b} = \frac{R}{P}$. They are moved across the sample, giving $N_b = T - b + 1$ repetitions, which are used to approximate the estimation-error-corrected test statistic distribution. The assessment of the quality of such correction requires repeating the whole procedure a large number of times. In summary, we perform the simulations, where each involves the following steps:

(a) estimation of the VaR model parameters;
(b) producing a VaR violation process;
(c) generating the corrected distribution through the subsampling;
(d) testing;
(e) computing a decision from the simulated estimation-error-corrected distribution.

From a large number of repetitions, which involve steps (a)-(e), we can approximate the test size, which we call an estimation-error-corrected size. Since such nested simulations are very time-consuming, we set the number of MC repetitions to 1 000.

The size results obtained from the procedures based on the estimation-error-corrected distributions show the reduction of the negative effects of the estimation risk. All results are closer to the benchmark size estimates attainable under the absence of the estimation risk. This confirms the results from previous studies, which apply the subsampling method. However, the results still seem unsatisfactory. All size estimates, regardless of the test or significance level, more than double the nominal size.

As an alternative to the resampling methods, we consider dealing with the problem of size distortions by changing the forecasting scheme. We propose to shift from a fixed forecasting scheme to the rolling forecasting scheme. As explained in the previous sections, this scheme involves the parameter corrections, which are done with a predefined frequency. Since our study is based on the daily data, we set the frequencies f to 1, 5 and 10 days. This means that each set of the in-sample data of the length R is used to estimate the VaR model parameters and produce the 1- to f-day-ahead VaR forecasts. Then this in-sample data is moved by dropping the f last observations and adding f more recent ones. The resulting VaR forecasts give the violation process of the length P, which is then used for testing. The testing is done with the use of the standard asymptotic χ^2 distribution. We mimic such a procedure in simulations, whose number we set to 10 000. We conduct the simulations separately for each value of f. These simulations allow us to approximate the test size, which we call the estimation-error-uncorrected test size with the parameter correction.

The results from the final step of our simulation study, which involves the rolling forecasting scheme, show that adjustments made to the forecasting scheme have a larger impact on the size property than the subsampling method. The parameter corrections that we propose work better at reducing the effects of the estimation risk. Even for the lowest considered correction frequency, which is 10 days, the size results are closer to the benchmark ones than those obtained with subsampling. Increasing the correction frequency further improves the results. As expected, the everyday parameter correction yields the most accurate results. In particular, the results of the GMM_{VaR} for the significance levels 0.05 and 0.1 are in line with those obtained initially, under the absence of the estimation risk. This shows that the previous results about the necessity of replacing the asymptotic distribution with the simulated one are valid only for the case of the fixed forecasting scheme. This scheme, however, is rather rare in practice. For the rolling forecasting scheme, regular parameter corrections allow for accurate testing with the use of the standard asymptotic distributions.

4. Conclusions

The paper referred to the topic of the formal evaluation of VaR models. It focused on the problem of the parameter uncertainty in testing VaR. So far, this problem was dealt with by means of the resampling methods. We proposed to handle it by introducing the forecasting scheme that reduces the estimation errors. This method, contrary to the resampling, permits the use of the standard asymptotic distributions. In this way, is substantially reduces the time demands and complexity of the testing procedures. Our main result was that adjustments made to the forecasting scheme allow for better control over the test size than the resampling method. In particular, we showed that introducing a rolling forecasting scheme with frequent parameter corrections almost fully reduces the negative effects of the estimation risk. Thus the proposed approach outperforms the resampling method with regard to both practical aspects and statistical properties of the tests.

References

Basel Committee on Banking Supervision (2017). High-level summary of Basel III Reforms, online: www.bis.org/bcbs/publ/d424_hlsummary.pdf (accessed June 4, 2019).

Berkowitz, J., Christoffersen, P., Pelletier, D. (2011). Evaluating Value-at-Risk Models with Desk-Level Data. *Manage. Sci*., *12*(57), 2213-2227.

Candelon, B., Colletaz, G., Hurlin, C., Tokpavi, S. (2011). Backtesting Value-at-Risk: a GMM duration-based test. *J. Financ. Economet.*, *9*(2), 314-343.

Christoffersen, P. (1998). Evaluating Interval Forecasts. *Int. Econ. Rev.*, *39*(4), 841-862.

Dumitrescu, E.I., Hurlin, Ch., Pham, V. (2012). Backtesting Value-at-Risk: From Dynamic Quantile to Dynamic Binary Tests. *Finance*, *33*(1), 79-112.

Escanciano, J.C., Olmo, J. (2010). Backtesting Parametric Value-at-Risk With Estimation Risk. *J. Bus. Econ. Stat.*, *28*(1), 36-51.

Escanciano, J.C., Olmo, J. (2011). Robust Backtesting Tests for Value-at-risk Models, *J. Financ. Economet.*, *9*(1), 132-161.

Kramer, W., Wied, D. (2015). A simple and focused backtest of value at risk. *Econ. Lett.*, *137*, 29-31.

Kupiec, P. (1995), Techniques for Verifying the Accuracy of Risk Measurement Models, *J. Deriv.*, *3*, 73-84.

Pajhede, T. (2017). Backtesting Value-at-Risk: A Generalized Markov Test. *J. Forecast.*, *36*(5), 597-613.

Pelletier, D., Wei, W. (2016). The geometric-VaR backtesting method. *J. Financ. Economet.*, *14*(4), 725-745.

Ziggel, D., Berens, T., Weiß, G.N.F., Wied, D. (2014). A new set of improved value-at-risk backtests. *J. Bank. Financ.*, *48*, 29-41.