Journal of Banking and Finance 91 (2018) 133–145

Contents lists available at ScienceDirect

Journal of Banking and Finance

journal homepage: www.elsevier.com/locate/jbf

Unbiased estimation of risk

Marcin Pitera[a], Thorsten Schmidt[b,c,d,*]

[a] Institute of Mathematics, Jagiellonian University, Łojasiewicza 6, Cracow 30-348, Poland
[b] Freiburg University, Department of Mathematics, Ernst-Zermelo-Str. 1, Freiburg 79106, Germany
[c] Freiburg Research Institute of Advanced Studies (FRIAS), Germany
[d] University of Strasbourg Institute for Advanced Study (USIAS), France

ARTICLE INFO

Article history:
Received 20 December 2016
Accepted 23 April 2018
Available online 26 April 2018

Keywords:
Value-at-risk
Tail value-at-risk
Expected shortfall
Risk measure
Estimation of risk measures
Bias
Risk estimation
Elicitability
Backtest
Unbiased estimation of risk measures

ABSTRACT

The estimation of risk measures recently gained a lot of attention, partly because of the backtesting issues of expected shortfall related to elicitability. In this work we shed a new and fundamental light on optimal estimation procedures of risk measures in terms of bias. We show that once the parameters of a model need to be estimated, one has to take additional care when estimating risks. The typical plug-in approach, for example, introduces a bias which leads to a systematic *underestimation* of risk.

In this regard, we introduce a novel notion of *unbiasedness* to the estimation of risk which is motivated by economic principles. In general, the proposed concept does not coincide with the well-known statistical notion of unbiasedness. We show that an appropriate bias correction is available for many well-known estimators. In particular, we consider value-at-risk and expected shortfall (tail value-at-risk). In the special case of normal distributions, closed-formed solutions for unbiased estimators can be obtained.

We present a number of motivating examples which show the outperformance of unbiased estimators in many circumstances. The unbiasedness has a direct impact on backtesting and therefore adds a further viewpoint to established statistical properties.

1. Introduction

The estimation of risk measures is an area of highest importance in the financial industry as risk measures play a major role in risk management and in the computation of regulatory capital, see McNeil et al. (2010) for an in-depth treatment of the topic. Most notably, a major part of quantitative risk management is of statistical nature, as highlighted for example in Embrechts and Hofert (2014). This article takes this challenge seriously and does not target risk measures themselves, but *estimated* risk measures.

Statistical aspects in the estimation of risk measures recently raised a lot of attention: see the related articles Davis (2016) and Cont et al. (2010), Acerbi and Székely (2014), Ziegel (2016), Fissler et al. (2015) and Frank (2016). Surprisingly, it turns out that statistical properties of risk estimators - related to the presence of bias - have not yet been analysed thoroughly. Such properties are very important from the practical point of view, as the risk bias usually leads to a systematic underestimation of risk. It is our main goal to give a definition of *unbiasedness* that makes sense economically and statistically. The main motivation for this is the observation that the classical (statistical) definition of bias might be desirable from a theoretical point of view, while it might be not prioritised by financial institutions or regulators, for whom the backtests are currently the major source of evaluating the estimation.

There is an ongoing intensive debate in regulation and in science about the two most recognised risk measures: Expected Shortfall (ES) and Value-at-Risk (V@R). This debate is stimulated by Basel III project (BCBS - Basel Committee on Banking Supervision, 2009), which updates regulations responsible for capital requirements for initial market risk models (cf. BCBS - Basel Committee on Banking Supervision, 2006; BCBS - Basel Committee on Banking Supervision, 1996). In a nutshell, the old V@R at level 1% is replaced with ES at level 2.5%. In fact, such a correction may reduce the bias, however only in the right scenarios. The academic response to this fact is not unanimous: while ES is coherent and takes into consideration the whole tail distribution, it lacks some nice statistical properties characteristic to V@R. See e.g. Cont et al. (2010), Acerbi and Székely (2014), Ziegel (2016), Kellner and Rösch (2016), Yamai and Yoshiba (2005) and Emmer et al. (2015) for further details and interesting discussions. Also, the ES forecasts are believed to be much harder to backtest, a property essential from the regulator's point of view.

* Corresponding author.
E-mail addresses: marcin.pitera@im.uj.edu.pl (M. Pitera), thorsten.schmidt@stochastik.uni-freiburg.de (T. Schmidt).

https://doi.org/10.1016/j.jbankfin.2018.04.016
0378-4266/

A further argument in this debate emerges from the results in Gneiting (2011) (see also Weber, 2006), showing that ES is not *elicitable*. This interesting concept was originally developed in Osband and Reichelstein (1985) from an economic perspective; the main motivation is to ensure truthful reporting by penalizing false reports. In Section 7.1 we provide a detailed discussion of this topic.

The lack of elicitability led to the discussion whether or not (and how) it is possible to backtest ES and we refer to Carver (2014), Ziegel (2016), Acerbi and Székely (2014) and Fissler et al. (2015) for further details on this topic. Quite recently it was shown in Fissler et al. (2015) that ES is however jointly elicitable with V@R.[1] In particular, backtesting ES is possible; see Section 8.2 for a backtesting algorithm of ES in our setting; however, the results have to be interpreted with care.

Our article has two objectives. First, we introduce an economically motivated definition of unbiasedness: an estimation of risk capital is called *unbiased*, if adding the estimated amount of risk capital to the risky position makes the position acceptable; see Definition 4.1. It seems to be surprising that this is not the case for estimators considered so far. Second, we want to shed a new light on backtesting starting from the viewpoint of the standard Basel requirements. The starting point is the simple observation that a biased estimation naturally leads to a poor performance in backtests, such that the suggested bias correction should improve the results in backtesting.

In this regard, consider the standard (regulatory) backtest for V@R which is based on the rate of exception; see Giot and Laurent (2003). In Sections 7 and 8 we will show that a backtesting procedure based on the rate of exception will perform poorly if the estimation of the rate of exception is biased. Motivated by this, we systematically study bias of risk estimators and link our theoretical foundation to empirical evidence.

Let us start with an example: consider i.i.d. Gaussian data with unknown mean and variance and assume we are interested in estimating V@R at the level $\alpha \in (0, 1)$ ($V@R_\alpha$). Denote by $x = (x_1, \dots, x_n)$ the observed data. The *unbiased* estimator in this case is given by

$$\widehat{V@R}^{u}_{\alpha}(x_1, \dots, x_n) := -\left(\bar{x} + \bar{\sigma}(x)\sqrt{\frac{n+1}{n}}\,t^{-1}_{n-1}(\alpha)\right), \tag{1.1}$$

where t^{-1}_{n-1} is the inverse of the cumulative distribution function of the Student-*t*-distribution with $n-1$ degrees of freedom, $\bar{x}$ denotes the sample mean and $\bar{\sigma}(x)$ denotes the sample standard deviation. We call this estimator the *Gaussian unbiased estimator* and use this name throughout as reference to (1.1). Note that the *t*-distribution arises naturally by taking into account that variance has to be estimated and that the bias correction factor is $\sqrt{\frac{n+1}{n}}$. Comparing this estimator to standard estimators on NASDAQ data provides some motivating insights which we detail in the following paragraph.

Backtesting value-at-risk estimating procedures

To analyse the performance of various estimators of value-at-risk we performed a standard backtesting procedure. First, we estimated the risk measures using a learning period and then tested their adequacy in the backtesting period. The test was based on the standard *failure rate* (*exception rate*) procedure; see e.g. Giot and Laurent (2003) and BCBS - Basel Committee on Banking Supervision (1996). Given a data sample of size n, the first k observations were used for estimating the value-at-risk at level α. Afterwards it was counted how many times the actual loss in the following $n-k$ observations exceeded the estimate. For good estimators, we would expect that the number of exceptions divided by $(n-k)$ should be close to α.

More precisely, we considered returns based on (adjusted) closing prices of the NASDAQ100 index in the period from 1999-01-01 to 2014-11-25. The sample size is $n = 4000$, which corresponds to the number of trading days in this period. The sample was split into 80 separate subsets, each consisting of the consecutive 50 trading days. The backtesting procedure consisted in using the *i*th subset for estimating the value of $V@R_{0.05}$ and counting the number of exceptions in the $(i+1)$th subset. The total number of exceptions in the 79 periods was divided by $79 \cdot 50$. We compared the performance of the Gaussian unbiased estimator $\widehat{V@R}^{u}_{\alpha}$ to the three most common estimators of value-at-risk: the empirical sample quantile $\widehat{V@R}^{emp}_{\alpha}$ (sometimes called historical estimator[2]); the modified Cornish-Fisher estimator $\widehat{V@R}^{CF}_{\alpha}$; and the classical Gaussian estimator $\widehat{V@R}^{norm}_{\alpha}$, which is obtained by inserting mean and sample variance into the value-at-risk formula under normality:

$$\widehat{V@R}^{emp}_{\alpha}(x) := -\left(x_{(\lfloor h\rfloor)} + (h - \lfloor h\rfloor)(x_{(\lfloor h+1\rfloor)} - x_{(\lfloor h\rfloor)})\right), \tag{1.2}$$

$$\widehat{V@R}^{CF}_{\alpha}(x) := -\left(\bar{x} + \bar{\sigma}(x)\bar{Z}^{\alpha}_{CF}(x)\right), \tag{1.3}$$

$$\widehat{V@R}^{norm}_{\alpha}(x) := -\left(\bar{x} + \bar{\sigma}(x)\Phi^{-1}(\alpha)\right), \tag{1.4}$$

where $x_{(k)}$ is the kth order statistic of $x = (x_1, \dots, x_n)$, the value $\lfloor z\rfloor$ denotes the integer part of $z \in \mathbb{R}$, $h = \alpha(n-1) + 1$, Φ denotes the cumulative distribution function of the standard normal distribution and $\bar{Z}^{\alpha}_{CF}$ is a standard Cornish–Fisher α-quantile estimator (see e.g. (Alexander, 2009, Section IV.3.4.3) for details).

The results of the backtest are shown in the first part of Table 1. Surprisingly, the standard estimators show a rather poor performance. Indeed, one would expect a failure rate of 0.05 when using an estimator for the $V@R_{0.05}$ and the standard estimators show a clear *underestimation* of the risk, i.e. an exceedance rate higher than the expected rate. Only the Gaussian unbiased estimator is close to the expected rate, the empirical estimator having an exceedance rate which is 25% higher in comparison. One can also show that a Student-*t* (plug-in) estimator performs poorly, compared to Gaussian unbiased estimator.

To exclude possible disturbances of these findings by a bad fit of the Gaussian model to the data or possible dependences we additionally performed a simulation study: starting from an i.i.d. sample of normally distributed random variable with mean and variance fitted to the NASDAQ data we repeated the backtesting on this data; results are shown in the second column of Table 1. Let us first focus on the plug-in estimator $\widehat{V@R}^{norm}_{\alpha}$: expecting approximately 197 exceedances (5% out of 3.950) we experienced additional 36 exceedances on the NASDAQ data itself. On the simulated data, where we can exclude disturbances due to fat tails, correlation etc., still 19 unexpected exceedances were reported which is roughly 50% of the additional exceedances on the original data. These exceedances are due to the biasedness of the estimator and can be removed by considering the unbiased estimator $\widehat{V@R}^{u}_{\alpha}$ as may be seen from the last line of Table 1. The results on the other estimators confirm these findings,[3] the empirical estimator shows

[1] Another illustrative and self-explanatory example of this phenomena is variance. While not being elicitable, it is jointly elicitable with the mean; see Lambert et al. (2008).

[2] In fact there are numerous versions of the sample quantile estimator. We have decided to take the one used by default both in **R** and **S** statistical software for samples from continuous distribution.

[3] Further simulations show that these results are statistically significant: for example, repeating the simulation 10.000 times allows to compute the mean exception rates (with standard errors in parentheses) for estimators given in Table 1. They are equal to 0.055 (0.0027), 0.067 (0.0028), 0.056 (0.0026), and 0.050 (0.0027), respectively.