Unbiased estimation of risk

Marcin Pitera* Thorsten Schmidt†

First circulated: March 8, 2016
This version: March 8, 2016

Abstract

The estimation of risk measured in terms of a risk measure is typically done in two steps: in the first step, the distribution is estimated by statistical methods, either parametric or non-parametric. In the second step, the estimated distribution is considered as true distribution and the targeted risk-measure is computed. In the parametric case this is achieved by using the formula for the risk-measure in the model and inserting the estimated parameters. It is well-known that this procedure is not efficient because the highly nonlinear mapping from model parameters to the risk-measure introduces an additional biases. Statistical experiments show that this bias leads to a systematic *underestimation* of risk.

In this regard we introduce the concept of *unbiasedness* to the estimation of risk. We show that an appropriate bias correction is available for many well known estimators. In particular, we consider value-at-risk and tail value-at-risk (expected shortfall). In the special case of normal distributions, closed-formed solutions for unbiased estimators are given. For the general case we propose a bootstrapping algorithm and illustrate the outcomes by several data experiments.

Keywords: value-at-risk, tail value-at-risk, expected shortfall, risk measure, estimation of risk measures, risk estimation, back-test, bias.

1 Introduction

The estimation of measures of risk is an area of highest importance in the financial industry. Risk measures play a major role in the risk-management and the computation of regulatory capital. We refer to [23] for an in-depth treatment of the topic. In particular, in [13] the authors highlight that a major part of quantitative risk management is actually of statistical nature. 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 [12] and [29, 16, 2]. A careful analysis of the risk estimators shows that in general the estimators are biased, and systematically underestimate risk.

Surprisingly, it turns out that statistical properties of risk estimators have not yet been analysed thorougly. It is our main goal to give a definition of *unbiasedness* that has both economic and

*Institute of Mathematics, Jagiellonian University, Łojasiewicza 6, 30-348 Cracow, Poland
Email: marcin.pitera@im.uj.edu.pl
†Department of Mathematical Stochastics, University of Freiburg, Eckerstr.1, 79104 Freiburg, Germany
Email: thorsten.schmidt@stochastik.uni-freiburg.de

statistical sense for risk estimators. While the classical (statistical) definition of bias is always desirable from a theoretical point of view, it might be not prioritised by financial institutions or regulators, for whom the back-tests are the main source of estimation accuracy.

Not surprisingly, the occurrence of biasses in risk estimation plays an important role in practice: the Basel III project [7] suggests to change value-at-risk into expected shortfall and consider stressed scenarios where the risk level is replaced with a different level (97.5%). In fact, such a correction may reduce the bias, however only in the right scenarios. Our goal is to systematically study biases of risk estimators and provide a theoretical foundation together with empirical evidence propagating the use of unbiased estimators.

We begin by a motivation in the parametric case. Parametric estimation of risk measures in finance is typically done in two steps: starting from a parametric model for the financial market, the parameters of the model are estimated from historical data. Then, the estimator of the risk-measure is computed by using the formula for the risk-measure in the model and inserting the estimated parameters. In statistical theory it is well-known that this is not efficient because the highly nonlinear mapping from model parameters to the risk-measure introduces an additional biases. Statistical experiments show that this bias leads to a systematic underestimation of risk.

Consider i.i.d. Gaussian data with unknown mean and variance, and assume we were interested in estimating the value-at-risk (V@R) at the level $\alpha \in (0,1)$. Denote by $x = (x_1, \ldots, x_n)$ the observed data. The *unbiased* estimator under normality obtained by the methodology we propose is given by

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

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). 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 back-testing procedure. First, we estimated the risk measures using a learning period and then tested it's adequacy in the back-testing period. The test was based on the *failure rate* procedure as suggested in [20]. More precisely, given a data sample of size n, the first k observations were used for estimating the value-at-risk at level $\alpha = 95\%$, denoted by $\mathrm{V@R}_{0.95}$. 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 5%.

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 consecutive 50 trading days. The back-testing procedure consisted in using the i-th subset for estimating the value of $\mathrm{V@R}_{0.95}$ 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 to the three most common estimators of

Estimator		NASDAQ		NORMAL	
		exceeds	percentage	exceeds	percentage
Percentile	$\hat{\mathrm{V@R}}^{\mathrm{emp}}_\alpha(x)$	272	0.069	253	0.064
Modified C-F	$\hat{\mathrm{V@R}}^{\mathrm{mod}}_\alpha(x)$	249	0.063	230	0.058
Gaussian	$\hat{\mathrm{V@R}}^{\mathrm{norm}}_\alpha(x)$	241	0.061	221	0.056
Gaussian unbiased	$\hat{\mathrm{V@R}}^{\mathrm{u}}_\alpha(x)$	217	0.055	197	0.050

Table 1: *Estimates of* $\mathrm{V@R}_{0.95}$ *for NASDAQ 100 (first column) and for simulated i.i.d.normal distributed random variables with the mean and variance fitted to the NASDAQ data (second column), both for 4.000 data points.* Exceeds *reports the number of exceptions in the sample, where the actual loss exceeded the risk estimate. The expected rate of* 0.05 *is only reached by the Gaussian unbiased estimator.*

value-at-risk: the empirical 0.95%-percentile (sometimes called historical estimator); the modified Cornish-Fisher estimator (see [3]); and the classical Gaussian estimator, which is obtained by assuming that the log-returns are normally distributed and estimating the parameters with the maximum-likelihood estimators. These estimators are given by:

$$\hat{\mathrm{V@R}}^{\mathrm{emp}}_\alpha(x) := -x_{(\lfloor n\alpha\rfloor+1)}, \tag{1.2}$$

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

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

where $x_{(k)}$ is the k-th order statistic of $x = (x_1, \dots, x_n)$, the value $\lfloor z \rfloor$ denotes the integer part of $z \in \mathbb{R}$, and Φ denotes the cumulative distribution function of the standard normal distribution. For a deeper discussion of the empirical value-at-risk estimator see Section 5. Moreover,

$$\bar{Z}^{\alpha}_{CF}(x) = q_\alpha + \frac{1}{6}(q_\alpha^2 - 1)\bar{S}(x) + \frac{1}{24}(q_\alpha^3 - 3q_\alpha)\bar{K}(x) - \frac{1}{36}(2q_\alpha^3 - 5q_\alpha)\bar{S}^2(x),$$

for $\bar{S}(x)$ and $\bar{K}(x)$ being standard estimators of skewness and excess kurtosis and $q_\alpha = \Phi^{-1}(\alpha)$.

The results of the back-test are shown in 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 $\mathrm{V@R}_{0.95}$ 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 percentile estimator having an exceedance rate which is 25% higher in comparison. Also a Student-t-plug-in estimation performs poorly, compare Section 7.2.

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: in this regard, we simulated an i.i.d. sample of normally distributed random variable with mean and variance fitted to the NASDAQ data and repeated the back-testing on this data.

The results are shown in the second column of Table 1. The bias decreases in a moderate way, but is still present in all three alternative estimation procedures. The percentile estimator shows an exceedance rate being 28% percent higher compared to the Gaussian unbiased estimator, which in turn perfectly meets the level $\alpha = 0.05$. More empirical studies are provided in Section 7.

The structure of the paper is as follows: in Section 2, estimators of risk measures are formally introduced. Section 3 discusses the frequently used concept of plug-in estimators. Section 4 introduces the main concept of the paper, unbiasedness, and gives a number of first examples while Section 5 considers asymptotically unbiased estimators. When the estimator is not known, we proposed some bootstrapping procedures, which are given in Section 6. Section 7 gives a detailed empirical study of the proposed estimators and we conclude in Section 8.

2 Estimation of risk

In this section we study the estimation of risk, measured by a risk measure. Our focus lies on the most popular family of risk measures, so-called *law-invariant* risk measures. These measures solely depend on the distribution of the underlying losses, see [23] for an outline and practical applications of risk measurement. Law-invariant risk measures for example contain the special cases value-at-risk, tail value-at-risk or the spectral risk measure.

We consider the estimation problem in a parametric setup. If the parameter space is chosen infinite-dimensional, this also contains the non-parametric formulation of the problem. In this regard, let $(\Omega, \mathcal{A})$ be a measurable space and $(P_\theta : \theta \in \Theta)$ be a family of probability measures on this measurable space parametrized by θ, an element of the parameter space Θ. For simplicity, we assume that the measures P_θ are equivalent, such that their null-sets coincide. Otherwise it could be possible that some of the probability measures would be excluded almost surely by some observation which in turn would lead to unnecessarily complicated expressions. By $L^0 := L^0(\Omega, \mathcal{A})$ we denote the (equivalence classes of) real-valued and measurable functions. In our context, the space L^0 typically corresponds to discounted cash flows or financial positions return rates.

For the estimation, we assume that we have a sample $X_1, X_2, \ldots, X_n$ of observations at hand and we want to estimate the risk of a future position X. Sometimes, we consider $x = (x_1, \ldots, x_n)$ to distinguish specific realizations of $X_1, \ldots, X_n$ from the sample random variables. In particular, we know that $x_i = X_i(\omega)$ for some $\omega \in \Omega$.

Example 2.1 (The i.i.d.-case)**.** Assume that the future position X as well as the historical observations are independent and identically distributed (i.i.d.). This is the case, for example in the Black-Scholes model when one considers log-returns. More generally, this also holds in the case where the considered stock price S follows a geometric Lévy process (see [4, Section 5.6.2] and references therein). If t_i, $i = 0, \ldots, n+1$ denote equidistant times with $\Delta = t_i - t_{i-1}$, then the log-returns $X_i := \log(S_{t_i}) - \log(S_{t_{i-1}})$, $i = 1, \ldots, n$ are i.i.d. and the risk of the future position $S_{t_{n+1}}$ can be described in terms of $X := \log(S_{t_{n+1}}) - \log(S_{t_n})$.

Consider the problem of estimating the risk of a future position X. To quantify the risk associated with the position $X \in L^0$, we introduce a concept of *a risk measure*.

Definition 2.2. A risk measure ρ is a mapping from L^0 to $\mathbb{R} \cup \{+\infty\}$.

Typically, one assumes additional properties for ρ given in Definiton 2.2 such as *counter monotonicity*, *convexity* and *translation invariance*. For details, see [18] and references therein. For brevity, as we are interested in problems related to estimation of ρ, we have decided not to repeat detailed definitions here. Let us alone mention that from a financial point of view the value $\rho(X)$ is a quantification of risk for financial future position X and is often interpreted as the amount of money one has to add to the position X such that X becomes acceptable. Hence, positions X with $\rho(X) \le 0$ are considered acceptable (without adding additional money).

A priori, the definition of a risk measure is formulated without any relation to the underlying probability. However, in most practical applications this is not the case. One rather considers the following class of law-invariant risk-measures: roughly spoken, a risk-measure is called law-invariant with respect to a probability measure P, if $\rho(X) = \rho(\tilde{X})$ whenever the laws of X and $\tilde{X}$ coincide under P, see for example Section 5 in [17].

Hence, ρ typically depends on the underlying probability measure P_θ and consequently, we obtain a family of risk-measures $(\rho_\theta)_{\theta\in\Theta}$, which we again denote by ρ. Here, ρ_θ is the risk-measure obtained under P_θ. Being law-invariant, the risk-measure can be identified with a function of the cumulative distribution function of X. More precisely, we obtain the following definition. Denote by $\mathcal{D}$ the convex space of cumulative distribution functions of real-valued random variables.

Definition 2.3. The family of risk-measures $(\rho_\theta)_{\theta\in\Theta}$ is called *law-invariant*, if there exists a function $R : \mathcal{D} \to \mathbb{R} \cup \{+\infty\}$ such that for all $\theta \in \Theta$ and $X \in L^0$

$$\rho_\theta(X) = R(F_X(\theta)), \tag{2.1}$$

$F_X(\theta) = P_\theta(X \le \cdot)$ denoting the cumulative distribution function of X under the parameter θ.

We aim at estimating the risk of the future position where $\theta \in \Theta$ is unknown. If θ were known, we could directly compute the corresponding risk measure ρ_θ from P_θ and would not need to consider the family $(\rho_\theta)_{\theta\in\Theta}$. Various estimation methodologies are at hand, the most common one being the plug-in estimation (see Section 3 for details).

We recall that we are interested in estimating the risk-measure $\rho(X)$ of the future position X from a sample $x_1, \dots, x_n$. A typical setting is the i.i.d.-case from Example 2.1.

Definition 2.4. An *estimator* of a risk measure is a Borel function $\hat{\rho}_n : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$.

Sometimes we will call $\hat{\rho}_n$ also risk estimator. The value $\hat{\rho}_n(x_1, \dots, x_n)$ corresponds to the *estimated* amount of capital which should classify, after adding the capital to the position, the future position X acceptable.

Given random sample $X_1, X_2, \dots, X_n$, we denote as usual by $\hat{\rho}_n$ also the random function

$$\hat{\rho}_n(\omega) := \hat{\rho}_n(X_1(\omega), \dots, X_n(\omega)), \quad \omega \in \Omega,$$

corresponding to the estimator $\hat{\rho}$. By $\hat{\rho}$ we denote the sequence of risk estimators $\hat{\rho} = (\hat{\rho}_n)_{n\in\mathbb{N}}$. If there is no ambiguity, we will call $\hat{\rho}$ also risk estimator.

The concept of *estimator* given in Definition 2.4 is very general. One very common way in practical estimation of risk measures is to separate the estimation of the distribution of the underlying random variable from the estimation of the risk measure. This leads to the well established plug-in estimators, which we discuss in the following section.

3 Plug-in estimators

A common way to estimate risk are *plug-in estimators* (cf. [1, 11, 17] and references therein). The idea behind this approach is to use the highly developed tools for estimating the distribution function of X and plug in this estimate into the desired risk measure. For the plug-in procedure to work, one needs to consider law-invariant risk measures, see (2.1), which we assume for the following. Given a sample $x = (x_1, \dots, x_n)$, we denote the estimator of the unknown distribution by $\hat{F}_X$. Recall the function R from (2.1). Then the *plug-in estimator* $\hat{\rho}_{\text{plugin}}$ is given by

$$\hat{\rho}_{\text{plugin}}(x) := R(\hat{F}_X). \tag{3.1}$$

In particular, in the parametric case, first parameters are estimated, and then plugged into the formula for the risk-measure (3.1). In other words: first, given sample x, one estimates the parameter θ by a suitable estimator which we denote by $\hat{\theta}$. Second, one plugs the obtained estimate $\hat{\theta}$ into Equation (2.1) and obtains the *plug-in estimator*

$$\hat{\rho}_{\text{plugin}}(x) = R(F_X(\hat{\theta})) = \rho_{\hat{\theta}}(X).$$

Let us now present some specific examples, where we provide explicit formulas for the plug-in estimators of the considered risk both for non-parametric and parametric case.

Example 3.1 (Empirical distribution plug-in estimation)**.** As an example we could use the empirical distribution for the plug-in estimator. The assumption is that $X_1, \dots, X_n$ are independent, having the same distribution like X, and a sample $x = (x_1, \dots, x_n)$ is at hand. Then, the empirical distribution is given by

$$\hat{F}_X(t) := \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{x_i \le t\}}, \qquad t \in \mathbb{R},$$

where $\mathbb{1}_A$ is indicator of event A. It is a discrete distribution and hence $R(\hat{F}_X)$ is easy to compute. For example, for value-at-risk the resulting plug-in estimator using the empirical distribution is given in (1.2).

Example 3.2 (Kernel density estimation)**.** Assuming that X is (absolutely) continuous, one of the most popular non-parametric estimation techniques for the density of X is the so-called *kernel density estimation*, see for example [26, 25]. Instead of estimating the distribution itself, one focusses on estimating the probability density function, as in the continuous case we could recover one from another. Given the sample $x = (x_1, \dots, x_n)$, kernel function $K : \mathbb{R} \to \mathbb{R}$ and bandwidth parameter $h > 0$ (see [27] for details), the estimator $\hat{f}$ for the unknown density f is given by

$$\hat{f}(z) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{z - x_i}{h}\right), \quad z \in \mathbb{R}.$$

The most popular kernel functions are the Gaussian kernel K^1 and the Epanechnikov kernel K^2, given by

$$K^1(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} \quad \text{and} \quad K^2(u) = \frac{3}{4}(1 - u^2)\,\mathbf{1}_{\{|u| \le 1\}}.$$

The optimal value of the bandwidth parameter could also be estimated, but this depends on additional assumptions. For example, one could show that if the sample is Gaussian, then the optimal choice of bandwidth parameter is approximately $1.06\hat{\sigma}n^{-1/5}$, where $\hat{\sigma}$ is the standard deviation of the sample.

Example 3.3 (Plug-in estimators under normality)**.** Let us assume that X is normally distributed under P_θ, for any $\theta = (\theta_1, \theta_2) \in \Theta = \mathbb{R} \times \mathbb{R}_{>0}$, where θ_1 and θ_2 denote mean and variance, respectively. Given sample $x = (x_1, \ldots, x_n)$, let $\hat{\theta}_1$ and $\hat{\theta}_2$ denote the estimated parameters (obtained e.g. using MLE method). Then, assuming that ρ is *translation invariant* and *positively homogenous* (see [18] for details), the classical *plug-in estimator* $\hat{\rho}$ can be computed as follows

$$\hat{\rho}(x) = R(F_X(\hat{\theta})) = \rho_{\hat{\theta}}(X) = \rho_{\hat{\theta}}\left(\hat{\theta}_1 \frac{X - \hat{\theta}_2}{\hat{\theta}_1} + \hat{\theta}_2\right) = -\hat{\theta}_2 + \hat{\theta}_1 R(\Phi), \tag{3.2}$$

where Φ denotes the cumulative distribution function of the standard normal distribution. If we are interested in estimating the value-at-risk, then the estimator (3.2) coincides with the one defined in (1.4).

Example 3.4 (Plug-in estimator for the t-distribution)**.** Assume now that X has a generalized t-distribution under P_θ, for any $\theta = (\theta_1, \theta_2, \theta_3) \in \Theta = \mathbb{R} \times \mathbb{R}_{>0} \times \mathbb{N}_{>2}$, where θ_1, θ_2 and θ_3 denote mean, variance and degrees of freedom parameter, respectively. Given the sample $x = (x_1, \ldots, x_n)$, let $\hat{\theta}_1$, $\hat{\theta}_2$ and $\hat{\theta}_3$ denote the estimated parameters (obtained e.g. using Expectation-Maximization method; see [15] for details). Then, assuming that ρ is *translation invariant* and *positively homogenous*, the *plug-in estimator* can be expressed as

$$\hat{\rho}(x) = -\hat{\theta}_1 + \hat{\theta}_2 \sqrt{\frac{\hat{\theta}_3 - 2}{\hat{\theta}_3}} R(t_{\hat{\theta}_3}), \tag{3.3}$$

where t_v corresponds to the standard t-distribution with v degrees of freedom. In particular, for value-at-risk at level α we get $R(t_{\hat{\theta}_3}) = -t^{-1}_{\hat{\theta}_3}(\alpha)$.

Example 3.5 (Plug-in estimator using extreme-value theory)**.** Let us assume that X is absolutely continuous for any $\theta \in \Theta$. For any threshold level $u < 0$ we define the conditional excess loss distribution of X under $\theta \in \Theta$ as

$$[F_X]_u(\theta, t) = P_\theta(X \leq u + t | X < u), \quad \text{for } t \leq 0.$$

Roughly speaking, The Pickands–Balkema–de Haan theorem states that for any $\theta \in \Theta$, if $u \to -\infty$, then the conditional excess loss distribution should converge to some Generalized Pareto Distribution (GPD).[1] We refer to [24, 23] and references therein for more details. This result can be used to provide an approximative formula for the risk measure estimator, if the risk measure solely depends on the lower tail of X. This is the case e.g. for value-at-risk or expected shortfall, especially when small risk levels $\alpha \in (0, 1)$ are considered. Given threshold level $u < 0$, sample $x = (x_1, \ldots, x_n)$ and using so called *Historical-Simulation Method* (see e.g. [24] for details) we define $\hat{F}_X$ for any $t < u$ setting

$$\hat{F}_X(t) = \frac{k}{n}\left(1 + \hat{\xi}\frac{u - t}{\hat{\beta}}\right)^{-1/\hat{\xi}}, \tag{3.4}$$

[1] Under some mild condition imposed on distribution $F_X(\theta)$. This includes e.g. families of normal, lognormal, χ^2, t, F, gamma, exponential and uniform distributions.

where k is the number of observations that lie below threshold level u and $(\hat{\xi}, \hat{\beta})$ correspond to shape and scale estimators in the GPD family. The estimators $\hat{\xi}$ and $\hat{\beta}$ can be computed taking only (negative values of) observations that lie below threshold level u and using e.g. the *Probability Weighted Moments Method* (again, see [24] and references therein for details). Now, assuming that the function R given in (2.1) depends only on the tail of the distribution, i.e. for any $\theta \in \Theta$ we only need $F_X(\theta)|_{(-\infty,u)}$ to calculate $R(F_X(\theta))$, one could obtain the formula for the plug in estimator using (3.4). In particular, for value-at-risk at level $\alpha \in (0,1)$, if only $\alpha < \hat{F}_X(u)$, then we can set

$$\hat{\rho}(x) = \hat{F}_X^{-1}(\alpha) = -u + \frac{\hat{\beta}}{\hat{\xi}} \left(\left(\frac{\alpha n}{k} \right)^{-\hat{\xi}} - 1 \right). \tag{3.5}$$

Please note that this estimator might be in fact considered non-parametric, as it approximates the value of $\rho(X)$ for a large class of distribution including almost all ones used in practise.

4 Unbiased estimation of risk

As is well-known, under nonlinear mappings, the plug-in procedure introduces a bias. It is one of our goals to achieve a precise definition of unbiasedness in the context of risk-measures which reflects the economic meaning of a risk-measure. Let us now present the main concept of this paper, i.e. the definition of unbiasedness for an estimator of risk.

Definition 4.1. The estimator $\hat{\rho}_n$ will be called *unbiased* for $\rho(X)$, if for all $\theta \in \Theta$,

$$\rho_\theta(X + \hat{\rho}_n) = 0. \tag{4.1}$$

An unbiased estimator has the feature, that adding the estimated amount of risk capital $\hat{\rho}_n$ to the position X makes the position $X + \hat{\rho}_n$ acceptable under all possible scenarios $\theta \in \Theta$. Requiring equality in Equation (4.1) ensures that the estimated capital is not too high. It should be noted, that except in the i.i.d. case, the distribution of $X + \hat{\rho}_n$ does depend on the dependence of the estimator $\hat{\rho}_n$ and X. Hence the notion of unbiasedness does not only depend on the law of $X, X_1, \ldots, X_n$ but also on their dependence structure.

From the financial point of view, given a historical data set, or even a stress scenario $(x_1, \ldots, x_n)$, the number $\hat{\rho}_n(x_1, \ldots, x_n)$ is used to determine the capital reserve for position X, i.e. the minimal amount for which the risk of the secured position $\xi^n(x_1, \ldots, x_n) := X + \hat{\rho}_n(x_1, \ldots, x_n)$ is acceptable. As the parameter θ is unknown, it would be highly desirable to minimise the risk of the secured position ξ^n. If we do this for any $\theta \in \Theta$, then our estimated capital reserve would be close to the real (unknown) one. To do so, we want the (overall) risk of position ξ^n to be equal to 0, for any value of $\theta \in \Theta$. This is precisely the definition of unbiasedness presented in Definition 4.1.

Remark 4.2 (Relation to the statistical definition of unbiasedness)*.* Let us stress out that Definition 4.1 differs from the definition of unbiasedness in the classical sense, i.e. the condition

$$E_\theta[\hat{\rho}_n] = \rho_\theta(X), \qquad \text{for all } \theta \in \Theta, \tag{4.2}$$

where E_θ denotes the expectation operator under P_θ. An estimator satisfying (4.2) will be called *statistically unbiased*. While the condition (4.2) is always desirable from the theoretical point of view, it might not be prioritised by the institution interested in risk measurement, as the mean value of estimated capital reserve does not determine the risk of the secured position $X + \hat{\rho}_n$. Let us explain this in more detail: In practice, the main goal is to define an estimator in such a way, that it behaves well in various back-testing or stress-testing procedures. The types of conducted tests are usually given by the regulator (see for example Basel regulations [8]). In the case of value-at-risk the so-called *failure rate* procedure is often considered. As explained in Section 1, this procedure focus on the rate of exceptions, i.e. ratio of scenarios in which estimated capital reserve is insufficient. This number might not be directly connected to the average value of estimated capital reserve, creating the need for different definition of bias. See also Remark 4.3 for further explanation.

Remark 4.3. In [19], the authors introduced the concept of a *probability unbiased* estimation: denote by $F_X(\theta, t) = P_\theta(X \leq t)$, $t \in \mathbb{R}$ the cumulative distribution function of X under P_θ. Then the estimator $\hat{\rho}_n$ is called *probability unbiased*, if

$$E_\theta[F_X(\theta, -\hat{\rho}_n)] = F_X(\theta, -\rho_\theta(X)), \qquad \text{for all } \theta \in \Theta. \tag{4.3}$$

Intuitively, the left hand side corresponds to the average probability, that our estimated capital reserve would be insufficient, while the right hand side corresponds to the probability of insufficiency of the theoretical capital reserve. This approach is proper for value-at-risk in the strongly restricted setting of the i.i.d. Example 2.1. In fact, in that setting, it coincides with our definition of unbiasedness from Definition 4.1: indeed, assume that $F_X(\theta)$ is continuous and that $X_1, \ldots, X_n, X$ are i.i.d. Then $\hat{\rho}_n$ and X are independent and hence

$$E_\theta[F_X(\theta, -\hat{\rho}_n)] = P_\theta[X + \hat{\rho}_n < 0].$$

On the other hand we know that, for ρ_θ being value-at-risk at level α, we obtain $F_X(\theta, -\rho_\theta(X)) = \alpha$, so (4.3) is equivalent to

$$P_\theta[X + \hat{\rho}_n < 0] = \alpha.$$

Now it is easy to show that this is equivalent to

$$\rho_\theta(X + \hat{\rho}_n) = \inf\{x \in \mathbb{R} \colon P_\theta[X + \hat{\rho}_n + x < 0] \leq \alpha\} = 0.$$

In the general case we consider here, a more flexible concept is needed to define the risk estimator bias. In particular, the average probability of insufficiency does not contain information about the level of capital deficiency. This, however, is a key concept, e.g., when considering tail value-at-risk, compare Example 4.7.

Let us now present some specific examples, where we could explicitly find the unbiased estimators of the considered risk.

Example 4.4 (Unbiased estimation of the mean)**.** Assume that X is integrable for any $\theta \in \Theta$, and consider a position acceptable if it has non-negative mean. This corresponds to the family ρ of risk measures

$$\rho_\theta(X) = E_\theta[-X], \qquad \theta \in \Theta.$$

Clearly ρ is law-invariant. Corresponding to Equation (4.1), a risk estimator $\hat{\rho}$ is unbiased in this setting if

$$0 = \rho_\theta(X + \hat{\rho}) = E_\theta[-(X + \hat{\rho})] = \rho_\theta(X) - E_\theta[\hat{\rho}].$$

Therefore, the estimator $\hat{\rho}$ is unbiased if and only if it is statistically unbiased for any $\theta \in \Theta$. Hence, the negative of the sample mean, given by

$$\hat{\rho}_n(x_1, \ldots, x_n) = -\frac{\sum_{i=1}^n x_i}{n}, \qquad n \in \mathbb{N},$$

is an unbiased estimator of the risk measure of position X.

Example 4.5 (Unbiased estimation of value-at-risk under normality)**.** Let X be normally distributed with mean θ_1 and variance θ_2 under P_θ, for any $\theta = (\theta_1, \theta_1) \in \Theta = \mathbb{R} \times \mathbb{R}_{>0}$. For a fixed $\alpha \in (0, 1)$, let

$$\rho_\theta(X) = \inf\{x \in \mathbb{R}\colon P_\theta[X + x < 0] \leq \alpha\}, \qquad \theta \in \Theta, \tag{4.4}$$

denote value-at-risk at level α. As X is absolutely continuous, unbiasedness as defined in Equation (4.1) is equivalent to

$$P_\theta[X + \hat{\rho} < 0] = \alpha, \qquad \text{for all } \theta \in \Theta. \tag{4.5}$$

This concept coincides with the definition of a *probability unbiased* estimator of value-at-risk (see Remark 4.3 for details). We define estimator $\hat{\rho}$, as

$$\hat{\rho}(x_1, \ldots, x_n) = -\bar{x} - \bar{\sigma}(x)\sqrt{\frac{n+1}{n}}t_{n-1}^{-1}(\alpha), \tag{4.6}$$

where t_{n-1} stands for cumulative distribution function of the student-t distribution with $n-1$ degrees of freedom and

$$\bar{x} := \frac{1}{n}\sum_{i=1}^n x_i, \quad \bar{\sigma}(x) := \sqrt{\frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})^2},$$

denote the efficient estimators of mean and standard deviation, respectively. We show that the estimator $\hat{\rho}$ is an unbiased risk estimator: note that $X \sim \mathcal{N}(\theta_1, (\theta_2)^2)$ under P_θ. Using the fact that X, $\bar{X}$ and $\bar{\sigma}(X)$ are independent for any $\theta \in \Theta$ (see e.g. [9]), we obtain

$$T := \sqrt{\frac{n}{n+1}} \cdot \frac{X - \bar{X}}{\bar{\sigma}(X)} = \frac{X - \bar{X}}{\sqrt{\frac{n+1}{n}}\theta_2} \cdot \sqrt{\frac{n-1}{\sum_{i=1}^n (\frac{X_i - \bar{X}}{\theta_2})^2}} \sim t_{n-1}.$$

Thus, the random variable T is a pivotal quantity and

$$P_\theta[X + \hat{\rho} < 0] = P_\theta[T < q_{t_{n-1}}(\alpha)] = \alpha,$$

which concludes the proof.

Remark 4.6. It follows that the difference between Gaussian unbiased estimator defined in (4.6) and the classical plug-in Gaussian estimator given in (1.4) is equal to

$$\widehat{\mathrm{V@R}}{}^{\mathrm{u}}_\alpha(x) - \widehat{\mathrm{V@R}}{}^{\mathrm{norm}}_\alpha(x) = -\bar{\sigma}(x)\left(\sqrt{\frac{n+1}{n}}t_{n-1}^{-1}(\alpha) - \Phi^{-1}(\alpha)\right). \tag{4.7}$$

Consequently, as $\bar{\sigma}(x)$ is consistent, and

$$\sqrt{\frac{n+1}{n}}t_{n-1}^{-1}(\alpha) \xrightarrow{n\to\infty} \Phi^{-1}(\alpha),$$

we obtain that the bigger the sample, the closer the estimators are to each other – the bias of plug-in estimator decreases.

The procedure from the previous example can be applied to almost any (reasonable) coherent risk measure. We choose the famous coherent risk measure expected shortfall (or tail-value-at-risk) as an example to illustrate how this can be achieved.

Example 4.7 (Unbiased estimation of tail value-at-risk under normality)**.** As before, let X be normally distributed with mean θ_1 and variance θ_2 under P_θ, for any $\theta = (\theta_1, \theta_1) \in \Theta = \mathbb{R} \times \mathbb{R}_{>0}$. Let us fix $\alpha \in (0, 1)$. The tail value-at-risk at level α under a continuous distribution is given by

$$\rho_\theta(X) = E_\theta[-X|X \le q_X(\theta, \alpha)],$$ [2]

where $q_X(\theta, \alpha)$ is α-quantile of X under P_θ, that coincides with the negative of value-at-risk at level α from Equation (4.4). Due to translation invariance and positive homogeneity of ρ_θ, exploiting the fact that X, $\bar{X}$ and $\bar{\sigma}(X)$ are independent for normally distributed X, a good candidate for $\hat{\rho}$ is

$$\hat{\rho}(x_1, \dots, x_n) = -\bar{x} - \bar{\sigma}(x)a_n, \tag{4.8}$$

for some $(a_n)_{n\in\mathbb{N}}$, where $a_n \in \mathbb{R}$. There exists a sequence $(a_n)_{n\in\mathbb{N}}$ such that $\hat{\rho}$ is unbiased: As ρ_θ is positively homogeneous, we obtain for all $\theta \in \Theta$

$$\begin{aligned}
\rho_\theta(X + \hat{\rho}) &= \theta_2\sqrt{\frac{n+1}{n}}\,\rho_\theta\left(\frac{X - \bar{X} - a_n\bar{\sigma}(X)}{\theta_2\sqrt{\frac{n+1}{n}}}\right) \\
&= \theta_2\sqrt{\frac{n+1}{n}}\,\rho_\theta\left(\frac{X - \bar{X}}{\theta_2\sqrt{\frac{n+1}{n}}} - \frac{a_n\sqrt{n}}{\sqrt{(n-1)(n+1)}} \cdot \sqrt{n-1}\frac{\bar{\sigma}(X)}{\theta_2}\right) \\
&= \theta_2\sqrt{\frac{n+1}{n}}\,\rho_\theta\left(Z - \frac{a_n\sqrt{n}}{\sqrt{(n-1)(n+1)}}V_n\right),
\end{aligned} \tag{4.9}$$

where, $Z \sim \mathcal{N}(0, 1)$, $V_n \sim \chi_{n-1}$ and both being independent. Note that the distribution of (Z, V_n) does not depend on θ. Thus, it is enough to show that there exists $b_n \in \mathbb{R}$ such that

$$\rho_\theta\left(Z + b_nV_n\right) = 0. \tag{4.10}$$

As V_n is non-negative and the risk measure ρ_θ is counter-monotone, we obtain that (4.10) is decreasing with respect to b_n. Moreover, $0 < \rho_\theta(Z) = \rho_\theta(Z + 0V_n)$. For b_n large enough we get $\rho_\theta(Z + b_nV_n) < 0$, as $\rho_\theta(Z + b_nV_n) = b_n\rho_\theta(\frac{Z}{b_n} + V_n)$ and

$$\rho_\theta\left(\frac{Z}{b_n} + V_n\right) \xrightarrow{b_n\to\infty} \rho_\theta(V_n) < 0,$$

[2] See, for example, [23], Lemma 2.16.

due to the Lebesgue continuity property of tail value-at-risk on L^1 (see [22, Theorem 4.1]). Thus, again using continuity of ρ_θ, we conclude that there exists $b_n \in \mathbb{R}$ such that (4.10) holds. Moreover, the value of b_n is independent of θ, as the family $(\rho_\theta)_{\theta\in\Theta}$ is law-invariant (see Equation (2.1)) and Z, V_n are pivotal quantities. Note that we only needed positive homogeneity and monotonicity of ρ_θ as well as (4.10) to show the existence of an unbiased estimator. Moreover, the value of b_n in (4.10), and consequently a_n in (4.8), can be computed numerically without effort.

5 Asymptotically unbiased estimators

Even if the risk estimators from Examples 3.1, 3.2, 3.3, 3.4 and 3.5 are biased (cf. Table 1), one might still have nice properties in an asymptotic sense. This is what we study in the following.

Definition 5.1. A sequence of risk estimators $\hat{\rho} = (\hat{\rho}_n)_{n\in\mathbb{N}}$ will be called *unbiased* at $n \in \mathbb{N}$, if $\hat{\rho}_n$ is unbiased. If unbiasedness holds for all $n \in \mathbb{N}$, we call the sequence $\hat{\rho}$ unbiased. The sequence $\hat{\rho}$ is called *asymptotically unbiased*, if

$$\rho_\theta(X + \hat{\rho}_n) \xrightarrow{n\to\infty} 0, \qquad \text{for all } \theta \in \Theta.$$

In many cases the estimators of the distribution are consistent in the sense that $\hat{F}_X \to F_X(\theta)$. Indeed, the Glivenko-Cantelli theorem gives even uniform convergence over convex sets of the empirical distribution with probability one. Intuitively, if the underlying distribution-based risk measure admit some sort of continuity, then we could expect that the the plug-in estimator satisfies

$$\hat{\rho}_n \xrightarrow{n\to\infty} \rho_\theta(X)$$

almost surely for each $\theta \in \Theta$. Consequently, for any $\theta \in \Theta$ we also would get

$$\rho_\theta(X + \hat{\rho}_n) \xrightarrow{n\to\infty} \rho_\theta(X + \rho_\theta(X)) = 0,$$

which is exactly the definition of asymptotic unbiasedness. Let us now present two examples, which show asymptotic unbiasedness of the empirical value-at-risk estimator (1.2) and the plug-in Gaussian estimator for tail value-at-risk.

Remark 5.2. The proposed definition of asymptotical unbiasedness has similarities to the notion of consistency suggested in [12]. This notion of consistency requires that averages of the calibration errors converge suitable fast to 0 when the time period tends to infinity. Hence, asymptotically unbiased risk estimators will be consistent when the calibration error is measured with the risk measure itself. On the other side, it should be noted that our main goal is to obtain the optimal risk estimator without averaging out under- or overestimates as they have an asymmetric effect on the portfolio performance.

We obtain the following result. Recall that we study an i.i.d. sequence $X, X_1, X_2, \ldots$ Let $\alpha \in (0,1)$ and consider the negative of emprical α-quantile

$$\hat{\rho}_n(x_1, \ldots, x_n) = -x_{(\lfloor n\alpha \rfloor + 1)}, \quad n \in \mathbb{N}, \tag{5.1}$$

which we call empirical estimator of value-at-risk at level α (compare also (1.2)). By $\hat{\rho}_n$ we denote the random variable $\hat{\rho}_n(X_1, \ldots, X_n)$.

Proposition 5.3. *Assume that X is absolutely continuous under P_θ for any $\theta \in \Theta$. The sequence of empirical estimators of value-at-risk given in (5.1) is asymptotically unbiased.*

Proof. For any $\epsilon > 0$ and $\theta \in \Theta$, let $A_{n,\epsilon} := \big\{ |\hat{\rho}_n + F_X^{-1}(\theta, \alpha)| \geq \epsilon \big\}$, where $F_X^{-1}(\theta, \cdot)$ denotes the inverse of $F_X(\theta, \cdot)$. Then we have that

$$P_\theta[X + \hat{\rho}_n < 0] \geq P_\theta[A_{n,\epsilon}^c] F_X\big(\theta, F_X^{-1}(\theta, \alpha) - \epsilon\big) - P_\theta[A_{n,\epsilon}],$$
$$P_\theta[X + \hat{\rho}_n < 0] \leq P_\theta[A_{n,\epsilon}^c] F_X\big(\theta, F_X^{-1}(\theta, \alpha) + \epsilon\big) + P_\theta[A_{n,\epsilon}].$$

Using the fact that empirical value-at-risk estimator is consistent [11, Example 2.10], i.e. for any $\theta \in \Theta$, under P_θ, we get

$$\hat{\rho}_n \xrightarrow{n \to \infty} -F_X^{-1}(\theta, \alpha) \qquad a.s.,$$

we obtain that $P_\theta[A_{n,\epsilon}] \xrightarrow{n \to \infty} 0$, for any $\epsilon > 0$ and $\theta \in \Theta$. Consequently,

$$F_X\big(\theta, F_X^{-1}(\theta, \alpha) - \epsilon\big) \leq \lim_{n \to \infty} P_\theta[X + \hat{\rho}_n < 0] \leq F_X\big(\theta, F_X^{-1}(\theta, \alpha) + \epsilon\big),$$

for any $\epsilon > 0$ and $\theta \in \Theta$. Taking the limit, and noting that $F_X(\theta, \cdot)$ is continuous, we get

$$P_\theta[X + \hat{\rho}_n < 0] \xrightarrow{n \to \infty} F_X\big(\theta, F_X^{-1}(\theta, \alpha)\big) = \alpha,$$

for any $\theta \in \Theta$, which concludes the proof, due to (4.5). □

Slightly changing the proof of Proposition 5.3 one could show that under normality assumption the sequence of classical plug-in Gaussian estimators of value-at-risk given in (1.4) is asymptotically unbiased as well. See also Remark 4.6.

In a similar way we obtain asymptotic unbiasedness of the Gaussian plug-in tail value-at-risk estimator introduced in (3.2): In this regard let X be normally distributed with mean θ_1 and standard deviation θ_2 under P_θ, for any $(\theta_1, \theta_2) = \theta \in \Theta = \mathbb{R} \times \mathbb{R}_{>0}$. For a fixed $\alpha \in (0, 1)$, let

$$\rho_\theta(X) = E_\theta[-X | X \leq q_X(\theta, \alpha)],$$

denote the tail value-at-risk at level α. Following (3.2), set

$$\hat{\rho}_n(x_1, \ldots, x_n) = -\bar{x} + \bar{\sigma}(x) R(\Phi), \quad n \in \mathbb{N}, \tag{5.2}$$

where Φ is a Gaussian distribution and $R(\Phi)$ is the tail value-at-risk at level α under Φ. The estimator (5.2) corresponds to a standard MLE plug-in estimator, under the assumption that X is normally distributed

Proposition 5.4. *Assume that $X, X_1, X_2, \ldots$ are i.i.d. $\mathcal{N}(\theta_1, \theta_2^2)$ for any $\theta \in \Theta$. The sequence of estimators of tail value-at-risk given in (5.2) is asymptotically unbiased.*

Proof. First, Theorem 4.1 in [22] shows that the tail-value-at-risk is Lebesgue-continuous, which means that for a sequence Y_n converging to Y almost surely and such that all Y_n are dominated by a random variable being an element of L^p, $\lim_{n \to \infty} \rho(Y_n) = \rho(Y)$. Set $Y_n := -\bar{x} + \bar{\sigma}(x) R(\Phi)$, such that $Y_n \to \theta_1 + \theta_2 R(\Phi) =: Y$ almost surely as $n \to \infty$. But, it follows directly for the tail-value-at risk under a normal distribution, denoted by ρ_θ, that

$$\rho_\theta(-\theta_1 + \theta_2 R(\Phi)) = 0,$$

hence the claim.

□

6 Bootstrapping unbiased estimators

In this section we study the case where explicit unbiased estimators for risk measures are not available. We propose a bootstrap algorithm for this case. The examples provided in the empirical study in Section 7 underline the applicability of this approach.

We will utilize a Bayesian approach to effectively incorporate initial information for the minimiziation. This will allow as to effectively adjust the fitting quality in an area which is of high interest to us.

For any risk estimator $\hat{\rho}_n$ and any prior distribution π (on Θ), we define the *mean absolute deviation* by

$$\Psi_\pi(\hat{\rho}_n) := \int_\Theta |\rho_\theta(X + \hat{\rho}_n)| \, \pi(d\theta). \tag{6.1}$$

If $\hat{\rho}_n$ is unbiased, then certainly $\Psi_\pi(\hat{\rho}_n) = 0$. Unfortunately, finding the global minimizer of (6.1) is a very challenging task, often impossible to achieve. Nevertheless, the goal of this section is to find a local minimizer within a specific family of risk estimators and use this as a good approximation of the global minimizer.

The simplest choice of π is a Dirac-measure $\delta_{\hat{\theta}}$, where $\hat{\theta}$ is a point estimator, given the sample $(x_1, \ldots, x_n)$. In this case, to minimize (6.1), given a family of risk estimators, we can use a standard resampling bootstrap algorithm. Two exemplary algorithms, where the family of risk estimators is the family of plug-in estimators with possible shift in the parameters are presented below.

Algorithm 6.1 (Bootstrapping suggested by Franconi-Herzog for value-at-risk). In [19], the following algorithm was suggested. The main idea is to replace the level α by a suitable chosen level α' which minimizes the averaged distance of the bootstrapped estimators to α.

1. Input: sample $x = (x_1, \ldots, x_n)$ and number of boostrapping steps $B \in \mathbb{N}$.
2. Estimate $\hat{\theta}$ using the MLE approach.
3. For $i = 1, \ldots, B$, simulate a sample of size n from $\mathbb{P}_{\hat{\theta}}$ and calculate $\hat{\theta}_i$, using the MLE approach.
4. Set the bootstrapped risk measure estimator to

$$\hat{\rho}^{\text{boot1}} := F_{\hat{\theta}}^{-1} \left(\operatorname*{arg\,min}_{\alpha' \in (0,1)} \left| \frac{1}{B} \sum_{i=1}^{B} F_{\hat{\theta}}(F_{\hat{\theta}_i}^{-1}(\alpha')) - \alpha \right| \right). \tag{6.2}$$

For example, considering the unbiased estimator under normality from (1.1), the above routine approximates the level α^*, such that

$$\Phi^{-1}(\alpha^*) = \sqrt{\frac{n+1}{n}} t_{n-1}^{-1}(\alpha)$$

and achieves an approximation of the unbiased estimator through a modification of the confidence level.

From a theoretical viewpoint it is, however, much more appealing to distort the estimated parameters $\hat{\theta}$ instead of the confidence level α. Most suitable to the estimators considered here is a linear distortion and we propose a suitable bootstrapping algorithm in the following.

We provide an alternative bootstrap algorithm based on the ideas detailed above. Assume that $\Theta \subseteq \mathbb{R}^m$, for a fixed $m \in \mathbb{N}$, and denote by

$$a \circ \theta := \begin{pmatrix} a_1\theta_1 \\ \vdots \\ a_m\theta_m \end{pmatrix}$$

the Hadamard-product of the vectors a and θ. The goal of our algorithm is to find a such that the mean absolute deviation given by (6.1) is minimized. One step of complication is caused by the choice of the prior distribution π. We choose to set π equal to the Dirac measure at estimated value of θ. This procedure performs well and reasonably fast in the computations. Recall that we consider law-invariant risk measures, see (2.1), such that $\rho_\theta(X) = R(F_X(\theta))$ with a suitable function R.

Algorithm 6.2 (Local minimization bootstrap approach). The idea is very simple: We estimate parameter $\hat{\theta}$ and then try to define risk estimator which (locally) is as unbiased as possible. To reduce bias, we slightly move the estimated parameters from $\hat{\theta}$ to $a \circ \hat{\theta}$ and find a such that the bias is minimized.

1. Input: sample $x = (x_1, \ldots, x_n)$ and number of bootstrappings $B \in \mathbb{N}$.

2. Estimate $\hat{\theta}$ (e.g. using x and the MLE approach).

3. For $i = 1, \ldots, B$, simulate a sample of size n from $\mathbb{P}_{\hat{\theta}}$ and compute the estimator $\hat{\theta}_i$.

4. Simulate $\hat{x} = (\hat{x}_1, \ldots, \hat{x}_B)$ from $P_{\hat{\theta}}$.

5. For any $a \in \mathbb{R}^m$ let
$$y_a := (\hat{x}_1 + \rho_{a\circ\hat{\theta}^1}(X), \ldots, \hat{x}_B + \rho_{a\circ\hat{\theta}^B}(X)).$$

6. Calculate
$$a^* := \arg\min_{a\in\mathbb{R}^m} \left| R(\hat{F}_{y_a}) \right|,$$
where $\hat{F}_{y_a}(t) := \frac{1}{B}\sum_{i=1}^{B} \mathbf{1}_{\{\hat{x}_i + \rho_{a\circ\hat{\theta}^i}(X) \le t\}}$ is the empirical distribution function given observations y_a.

7. Set the bootstrapped risk estimator to
$$\hat{\rho}^{\text{boot2}} := \rho_{a^*\circ\hat{\theta}}(X). \tag{6.3}$$

Remark 6.3. If one wants to take into consideration also the variance of estimator, then other risk functions could be defined. For example, in Equation (6.1) we could considered a slightly modified version of mean square error given by

$$\widetilde{\Psi}_n(\hat{\rho}_n) := \int_\Theta \Big(D^2_\theta(\hat{\rho}_n) + |\rho_\theta(X + \hat{\rho}_n)| \Big) \pi(d\theta).$$

7 Empirical study

It is the aim of this section to analyse the performance of selected estimators on various sets of real market data. In the first section, to emphasise the practical aspect of our study we follow Basel II regulations [6] to design the appropriate back-test procedure (in fact the general outline of the back-testing did not change in Basel II.5 [8] and proposed Basel III [7] frameworks). In the second section, we show the performance of various bootstrap estimators for market data and for simulated data.

Remark 7.1. Recently, there is an intensive debate in regulation and in science about the statistical properties of risk measures, in particular expected shortfall and value-at-risk. The discussion started with the discovery that expected shortfall is not *elicitable* in [21], compare [29]. A statistic $\psi(Y)$ is called *elicitable*, if

$$\psi = \arg\min_x E[S(x, Y)],$$

i.e. it minimizes the expected value of a scoring function S, see the nice discussion in [2]. This work also points out, that elicitability rather refers to model selection that to model testing. Here we consider back-testing as a model testing tool and therefore move on safe grounds. As we stick mainly to the standard Basel back-testing, several alternatives discussed in [2] are at hand if specific characteristics of the risk measures should be tested.

All calculation were done using **R 3.2.2**. We have used libraries **fImport** and **zoo** for data handling, **PerformanceAnalytics** for standard value-at-risk estimators, **MASS** for fitting functions, **evir** for GPD, **fGarch** for t-student fit, **xtable** for tables and **plyr** for functions wrappers. The optimal values of parameters in the bootstrap algorithms were calculated using standard function **optim**. To speed up the calculations we have decided to estimate KDE density on 1000 element grid and do linear approximation for other points (see help of *density* function for details). Gaussian kernel was used. GPD fit was obtained using probability-weighted moments method (see function *gpd* from **evir** library) and the plug-in GPD estimator was obtained using historical-simulation method.

7.1 Backtesting according to Basel II-III

Consider the daily value-at-risk at level 99% ($\mathrm{V@R}_{0.99}$) applied to return rates. For a given portfolio, say X, and given date, say t, we compute the daily risk denoted by $[\mathrm{V@R}_{0.99}]_t(X)$, using past daily returns of the same portfolio. Note that in practice, to report the actual capital reserve, financial institutions need to use 10-day value-at-risk at level 99%, but typically the daily risks are computed and scaled using appropriate multiplication factors. Also, in practice, back-tests are performed using 1-day risks and the scaling is done afterwards (e.g. to avoid impact of the significant changes made to portfolio composition during 10 days, typically done by most major trading institutions).

Let us now describe the back-testing and its effect on the actual capital reserve (see [6, 5] for more details). According to Basel II regulations (for internal market risk models), we need to compute value-at-risk at level 99% on a daily basis. For a given day t and portfolio X (for simplicity we assume that X is fixed throughout the whole period), we need to report the number on exceedances (exceptions) in the last 250 trading days. In other words, assuming that $r_{t-k}(X)$ is a return rate of X at trading day $t-k$, we need to count how many times the event

$$[\mathrm{V@R}_{0.99}]_{t-k}(X) < -r_{t-k}(X),$$

has occurred, for $k = 1, 2, \ldots, 250$. Please note that in fact we need to compute $[\mathrm{V@R}_{0.99}]_{t-k}(X)$ for each trading day $t-k$ separately using only data available up to time $t-k-1$.

The number of exceptions in the last 250 days determine the accuracy of our risk model (at time t). The possible outcomes are classified by the regulator into three zones, distinguished by colours:

- **The green zone** corresponds to models which produce from 0 to 4 exceptions. Those models are believed to be accurate and no penalty (for the actual capital reserve) is imposed.

- **The yellow zone** corresponds to models which produce from 5 to 9 exceptions. The supervisor should encourage a financial institution to present additional information about its model before taking action. Usually, moderate penalty is imposed on the capital reserve. (In practice, so called scaling factor might be increased by 0.4 to 0.8, based on the actual number of exceptions).

- **The red zone** corresponds to models which produce 10 or more exceptions. Those models are believed to be not accurate and substantial penalty is enforced in almost all cases (scaling factor is increased by 1). Moreover, financial institution is encouraged to improve its model immediately.

In some cases the number of test trading days might be shortened and then the number of exceptions in each zone is adjusted appropriately.

Assuming that the model is correct, the mean number of exceptions in the 250 day test period should be equal to $0.01 \cdot 250 = 2.5$, so theoretically the correct model should be in the green zone. We want to stress out the fact, that the back-test procedure is just an intermediate step, in a process to determine the actual (10-days) capital reserve. Also, the practical implementations of Basel committee regulations might differ across different countries.

For example, following Basel II regulations [6], the actual (market risk) capital reserve for static portfolio X on day t, denoted by $R_t(X)$ might be given by

$$R_t(X) = \max\left\{ [\overline{\mathrm{V@R}}_{0.99}]_t(X), \frac{k}{60}\sum_{k=0}^{59}[\overline{\mathrm{V@R}}_{0.99}]_{t-k}(X) \right\},$$

where $[\overline{\mathrm{V@R}}_{0.99}]_t(X)$ is a 10-day value-at-risk at level 99% for portfolio X on day t (e.g. it might be 1-day $[\mathrm{V@R}_{0.99}]_t(X)$ multiplied by $\sqrt{10}$) and $k \in [3, 4]$ is a constant specified by the regulator and based on the back-testing.

In Basel II.5 [8], $R_t(X)$ in increased by the stressed equivalent of value-at-risk, i.e. $\mathrm{V@R}_{0.99}$ calculated under stressed scenarios (e.g. when the market was under significant stress, like in 2007-2008 crisis). The Basel III project [7] suggests to consider only stressed scenarios but the value-at-risk at level 99% is replaced by a different risk measure: tail value-at-risk at level 97.5%.

We are now ready to compute $[\mathrm{V@R}_{0.99}]_t(X)$ for various time periods and portfolios X. We check the performance of our estimators, counting the number of exceptions for the consecutive 250 trading days. Different data sets and lengths of learning period are considered. More precisely, we consider:

1. **NASDAQ100** in the period from 01.01.2005 to 01.01.2012. In this case, the situation is similar to the situation in Section 1. We present here a detailed study, for various learning periods and we consider all estimators given in (1.1)–(1.4), for comparison.

2. **WIG20**, **BUX**, **S&P500**, **DAX**, **FTSE250** and **CAC40** market stock indices in the period from 01.01.2005 to 01.01.2015. We present here the performance of the Gaussian unbiased estimator, defined in Equation (1.1), for each index, and for various learning periods.

3. **Fama & Fench 25 portfolios** formed on Book-to-Market and Operating Profitability in the period from 01.01.2005 to 01.01.2015. We present the performance of the Gaussian unbiased estimator, for each portfolio, and for various learning periods.

While, according to Basel II.5 regulations, the learning period (sample period) should not be shorter than one year (see [8, 718(Lxxvi) (d)]), our results show that using unbiased estimators already provide satisfactory results for much shorter periods.

We want to emphasize that in this small empirical study we only consider the *Gaussian unbiased estimator* defined in Equation (1.1). We have decided to do that for transparency and easier comparison. The results presented in this section could be improved using estimation combined with e.g. GARCH filtering, as is often done in practice [28, 10]. In this case, an increase in the length of learning period might increase the accuracy of the estimators.

Moreover, recall that the unbiased approach for estimating risk can also be adapted to the case of tail value-at-risk (as proposed in Example 4.7), such that a similar analysis within the Basel III framework could be done.

NASDAQ100

We consider simple returns based on (adjusted) closing prices of the NASDAQ 100 index in the period from 01.01.2005 to 01.01.2012.[3] For any trading day t in this period, we use the past 100 (or 4,5,6,10,20,50, respectively) days to estimate the value-at-risk at level 99%. While we consider these various lengths for learning period, the back-testing period is fixed and equal to 1. In other words, we compute value-at-risk for each day t using values from 4 to 100 previous trading days and check if the estimated value was sufficient on day t. Then, we split the sample data into (six) non-overlapping subsets of consequent 250 trading days and for each subset we sum the number of exceptions in it. This is done for all estimators given in (1.1)–(1.4). We also present the mean number of exceptions in the whole sample. The results are presented in Table 2.

The performance of the Gaussian unbiased estimator is surprisingly good when small learning periods are considered, i.e. for $n \in \{4, 5, 6\}$. In our opinion, the good performance of this estimator for small values of n is a result of the fact that the shorter the period, the more uniform our sample is, in the sense that the conditional volatility does not change much, making the sample almost i.i.d. Unfortunately, for longer periods, the clustering effects start to take place (note this obstacle could be overcome by applying e.g. GARCH filtering).

For example, we see that for $n = 4$ the mean number of exceptions is equal to 0.0105, which is very close to 0.01, and the number of exceptions does not exceed 4. This puts us in the green zone for all non-overlapping time periods. Please also note that the considered period include the 2007-2008 financial crisis, when many other models failed.

If the training period is large then the bias correction is very small, such that the Gaussian unbiased estimator almost equals the classical Gaussian estimators, such that we do not show results for this case.

[3] We have decided to consider this period, to show the performance of our estimator during the 2007-2008 financial crisis.

Number of exceptions in each non-overlapping subset of length 250

n	$\hat{\mathrm{V@R}}^{\mathrm{emp}}_{0.99}$						$\hat{\mathrm{V@R}}^{\mathrm{mod}}_{0.99}$						$\hat{\mathrm{V@R}}^{\mathrm{norm}}_{0.99}$						$\hat{\mathrm{V@R}}^{\mathrm{u}}_{0.99}$					
4	46	53	48	61	47	53	30	30	30	36	34	32	16	25	20	22	23	21	**3**	**2**	**3**	**2**	**3**	**3**
5	39	46	42	54	41	45	23	30	31	26	30	25	12	24	19	18	19	18	**2**	**2**	**4**	**2**	5	**4**
6	34	40	37	49	37	42	21	26	26	26	29	24	13	17	15	17	15	15	**3**	**3**	**3**	**2**	5	**2**
10	23	25	22	38	26	25	12	13	14	22	23	19	5	9	9	11	10	12	**4**	**4**	**3**	**2**	5	5
20	11	16	14	13	16	15	6	7	11	9	9	13	**4**	9	6	6	10	11	**3**	5	**4**	**4**	6	6
50	5	11	9	**3**	11	8	**4**	7	6	**1**	11	7	**4**	10	7	**4**	11	8	**4**	9	6	**4**	8	7
100	**3**	9	10	**1**	6	7	**2**	7	5	**1**	5	6	**4**	10	8	**2**	10	10	**3**	9	7	1	9	9

Total number of exceptions divided by total length

n	$\hat{\mathrm{V@R}}^{\mathrm{emp}}_{0.99}$	$\hat{\mathrm{V@R}}^{\mathrm{mod}}_{0.99}$	$\hat{\mathrm{V@R}}^{\mathrm{norm}}_{0.99}$	$\hat{\mathrm{V@R}}^{\mathrm{u}}_{0.99}$
4	0.2025	0.1257	0.0813	0.0102
5	0.1719	0.1081	0.0706	0.0120
6	0.1532	0.0979	0.0587	0.0108
10	0.1022	0.0656	0.0365	0.0137
20	0.0545	0.0367	0.0304	0.0166
50	0.0304	0.0234	0.0269	0.0228
100	0.0223	0.0162	0.0271	0.0235

Table 2: *We estimate the number and the total number of exceptions using NASDAQ100 data, from 2005-01-01 to 2012-01-01; n= number of days for learning period and* $\alpha = 0.99$. *The estimators are the empirical percentile, the classical Gaussian estimator, the modified Cornish-Fisher estimator and the proposed Gaussian unbiased estimator, see* (1.1)–(1.4). *The Basel rule will classify the outcomes in the green zone if the number of exceptions is between* 0 *and* 4. *We highlight those numbers in bold face. This is surprisingly often the case for the Gaussian unbiased estimator* $\hat{\mathrm{V@R}}^{u}_{\alpha}$ *with small learning periods.*

From a practical viewpoint, a natural question is to determine the size of the mean of the risk estimator, as this determines the (mean) actual size of our capital reserve. For $n = 4$, the mean value of the Gaussian risk unbiased estimator is equal to 6.4%, which is relatively high for equity portfolios. This mostly results from the two following facts. First, we have only used the basic estimator without additional improvements. Second, the considered period includes the 2007-2008 financial crisis. With some easy technical improvements, this number can be reduced substantially. For example, the initial mean value of the Gaussian risk unbiased estimator for $n = 6$ is equal to 4.7%. Putting additional upper constraint on risk estimator value equal to 8% increase the number of exceptions in the whole sample only by 1, but reduces the mean value of the Gaussian risk unbiased estimator to 4.2%. Moreover, the mean value for the pre-crisis period (first 500 trading days) is equal to 3.2%, which sounds like a reasonable starting point. Additional improvements could be made, for further reduction (e.g. applying GARCH filtering, using macroeconomic factors to forecast volatility, expert opinion, etc.).

The above findings suggest that the class of unbiased estimators are a valuable tool for estimation of risk in practice.

Index	n	Total number of exeptions divided by total length	Number of exeptions for non-overlapping subsets of length 250							
WIG20	4	0.0081	3	2	1	1	3	2	3	1
	6	0.0098	4	1	1	1	3	2	4	3
BUX	4	0.0098	1	2	4	2	3	**6**	2	0
	6	0.0119	0	4	2	2	3	**6**	4	1
S&P500	4	0.0132	1	**6**	2	4	4	**5**	2	2
	6	0.0149	4	**5**	2	2	7	2	4	5
DAX	4	0.0127	3	4	1	3	3	2	4	6
	6	0.0128	**5**	**5**	1	3	2	4	2	4
FTSE250	4	0.0098	3	2	2	**5**	2	1	2	2
	6	0.0094	**6**	2	0	2	1	2	3	4
CAC40	4	0.0093	1	3	1	3	2	1	4	3
	6	0.0106	3	2	1	2	2	2	4	4

Table 3: *We estimate the number and the total number of exceptions on various stock indices, ranging from 01.01.2005 to 01.01.2015, n=number of days for learning period; the computations are presented for the Gaussian unbiased estimator* $\hat{\mathrm{V@R}}^u_\alpha$ *defined in* (1.1)*. We concentrate on very small learning periods. As above, the Basel rule will put the outcomes in the green zone when the number of exceptions is between* 0 *and* 4*. The cases* **not** *in the green zone are marked in bold face.*

Selected major stock market indices

We take returns based on (adjusted) closing prices for six major (country) stock indices in the period from 01.01.2005 to 01.01.2015. Namely, we take **WIG20** (Poland), **BUX** (Hungary), **S&P500** (USA), **DAX** (Germany), **FTSE250** (England) and **CAC40** (France) stock indices. Then, we perform calculations similar to those presented for the previous dataset, but only for the Gaussian risk unbiased estimator given in (1.1) and learning period of 4 or 6 days. The results are presented in Table 3. The simulations show that only in rare cases the Gaussian unbiased estimator leaves the green zone.

At first sight it is quite surprising, that the estimator performs so well on this variety of indices and given the fact, that returns typically do not exhibit a normal distribution. However, the unbiased estimator corrects for large variances in small sample sizes via the t-distribution with small degree of freedom, which itself has fat tails. Furthermore, the short learning periods group days of similar character together which altogether explain the good performance.

Fama & French dataset

We take returns from Fama & French Data Library [14]. We take returns of 25 portfolios formed on book-to-market and operating profitability in the period from 01.01.2005 to 01.01.2015. Then, we calculate the mean number of exceptions in the whole set, for various learning periods (4 to 6 days). For transparency, for learning period of length 4, we also present the sum of exceptions in each of the 8 non-overlapping subsets, consisting of consequent 250 trading days. The results are presented in Table 4.

Again, the outcome is surprising: for such small learning periods and such a variety of portfolios, the unbiased estimator performs very well.

7.2 Analysis of bootstrapping algorithms

In this Section we do calculations similar to the ones done in the motivational example from Section 1. For various financial portfolios, we consider returns based on (adjusted) closing prices. For every portfolio, we take sample of size 1500 and split it into 30 separate subsets, each consisting of consecutive 50 trading days. Then, for $i = 1, 2, \ldots, 29$, we estimate the value of $\mathrm{V@R}_{0.95}$ using the i-th subset and test it's adequacy on $(i+1)$-th subset, counting the number of exceptions.

We take 25 Fama & French portfolios formed on book-to-market and operating profitability (see previous example for details) and consider the period from 16.01.2009 to 01.01.2015 to obtain exactly 1500 observations for each portfolio.

We present the results for value-at-risk estimators defined in (1.1)–(1.4), the student-t plug-in estimator (3.3) and GPD plug-in estimator (3.5). We have also decided to present the results for four different bootstrap estimators that base on algorithms introduced in Section 6:

1. The first estimator, $\hat{\rho}^{\text{boot1}}$, is a Franconi-Herzog value-at-risk bootstrap estimator defined in (6.2). It is computed under the assumption, that the data came from the normal distribution. We try to shift the risk level $\alpha \in (0, 1)$ to reduce the bias.

2. The second estimator, $\hat{\rho}^{\text{boot2}}$, is a local bias-minimizing bootstrap estimator given in (6.2). It is computed under the assumption, that the data came from normal distribution. We take normal plug-in estimator and perform a parameter shift procedure, to minimize the bias.

3. The third estimator, $\hat{\rho}^{\text{boot3}}$, is a non-parametric version of Franconi-Herzog value-at-risk bootstrap estimator defined in (6.2). Instead of estimating $\hat{\theta}$ and $\hat{\theta}_i$ in steps 2. and 3. (see Algorithm 6.1 for details), we estimate the whole distribution functions using kernel density estimation (see Example 3.2) and then plug the appropriate quantile value to (6.2).

4. The fourth estimator, $\hat{\rho}^{\text{boot4}}$, is a local bias-minimizing bootstrap estimator given in (6.2). It is computed under assumption that the data below a given threesshold u came from GPD family (see Example 3.5).[4]

For all four estimators, the bootstrap strong sample size is set to $B = 10.000$. Note that Estimators $\hat{\rho}^{\text{boot1}}$ and $\hat{\rho}^{\text{boot2}}$ are parametric estimators which base on the assumption that the data is from a normal model. Their performance should be similar to Gaussian unbiased estimator and we have decided to include them, to show how the bootstrap approach approximates Gaussian unbiased estimator (note that no analytical formula for bias correction is needed to define the bootstrap estimator). On the other hand, estimators $\hat{\rho}^{\text{boot3}}$ and $\hat{\rho}^{\text{boot4}}$ might be seen as a non-parametric estimators (see Example 3.2 and Example 3.5 for details) that could be applied to almost any (reasonable) data. The results are presented in Table 5.

[4] Given sample $x = (x_1, \ldots, x_n)$, the threeshold level is set to $u = x_{(\lfloor 0.3n \rfloor + 1)}$.

To exclude possible dependences or bad model fit we additionally performed a simulation study: in this regard, we simulated an i.i.d. sample of normally distributed random variables with mean and variance fitted to each of the 25 portfolios. The sample size was set to 1500 for each set of parameters. Then, we repeated the back-testing on this data. The results are presented in Table 6.

Similar study was conducted for t-distribution sample, i.e. we fitted parameters for each of 25 portfolios, performed simulation and repeated back-test procedure. See Table 7 for details.

To gain further insight about the performance of the Gaussian unbiased estimator, and to approximate the standard error for Gaussian data, we have replicated 10000 times the results from Tables 6 for Historical, plug-in Normal, Cornish-Fisher, GPD and Gaussian unbiased estimators. For brevity, we only show results for the first portfolio - the results for other cases were almost identical, at least when comparison between estimators was considered. For each estimator, say $\hat{\rho}$, we also include empirical mean and standard deviation of the 'higher in comparison' statistics given by

$$T_1(\hat{\rho})(x) := \frac{\hat{\rho}(x) - \hat{\mathrm{V@R}}^{\mathrm{u}}_\alpha(x)}{\hat{\mathrm{V@R}}^{\mathrm{u}}_\alpha(x)}.$$

We also calculate the probability, that the value of unbiased estimator is closer to 0.05 than the value of $\hat{\rho}$. In other words we present the mean value of statistic

$$T_2(\hat{\rho})(x) := \mathbf{1}_{\{|\hat{\rho}(x)-0.05|-|\hat{\mathrm{V@R}}^{\mathrm{u}}_\alpha(x)-0.05|>0\}}.$$

The outcome is listed in Table 8. It clearly shows that the competing estimators underestimate the risk systematically and exceed the targeted level on average up to 29% more times than the unbiased estimator. Moreover, as could be seen from values of T_2, the performance of the Gaussian unbiased estimator is better in almost all cases.

8 Conclusion

In this article, we proposed to consider *unbiased* risk estimators for the estimation of risk in practise. We could compute some estimators, for example for value-at-risk in the Gaussian case, in closed form. For more general cases we introduce appropriate bootstrapping algorithms. The performance of the Gaussian unbiased risk estimator on a variety of data sets is very surprising: in particular in small learning periods, he almost ever ends up in the green zone according to Basel II regulation (compare Tables 2 and 3, respectively).

References

[1] C. Acerbi, *Coherent measures of risk in everyday market practice*, Quantitative Finance **7** (2007), no. 4, 359–364.

[2] C. Acerbi and B. Székely, *Back-testing expected shortfall*, Risk magazine (2014).

[3] C. Alexander, *Market Risk Analysis, Value at Risk Models*, vol. 4, John Wiley & Sons, 2009.

[4] D. Applebaum, *Lévy processes and stochastic calculus*, Cambridge university press, 2009.

[5] Basel Committee on Banking Supervision, *Supervisory framework for the use of 'backtesting' in conjunction with the internal models approach to market risk capital requirements*, Tech. report, Bank for International Settlements, 1996.

[6] ______, *Basel II: International Convergence of Capital Measurement and Capital Standards: A Revised Framework - Comprehensive Version*, Tech. report, Bank for International Settlements, 2006.

[7] ______, *Fundamental review of the trading book: A revised market risk framework - Consultative document*, Tech. report, Bank for International Settlements, 2009.

[8] ______, *Revisions to the Basel II market risk framework - updated as of 31 December 2010*, Tech. report, Bank for International Settlements, 2011.

[9] D. Basu, *On statistics independent of a complete sufficient statistic*, Sankhyā: The Indian Journal of Statistics (1933-1960) **15** (1955), no. 4, pp. 377–380 (English).

[10] J. Berkowitz and J. O'Brien, *How accurate are Value-at-Risk models at commercial banks?*, Journal of Finance (2002), 1093–1111.

[11] R. Cont, R. Deguest, and G. Scandolo, *Robustness and sensitivity analysis of risk measurement procedures*, Quantitative Finance **10** (2010), no. 6, 593–606.

[12] M. Davis, *Consistency of risk measure estimates*, (2013).

[13] P. Embrechts and M. Hofert, *Statistics and quantitative risk management for banking and insurance*, Annual Review of Statistics and Its Application **1** (2014), 493–514.

[14] E. F. Fama and K. R. French, http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html, accessed 20.10.2015.

[15] C. Fernandez and M. Steel, *On bayesian modeling of fat tails and skewness*, Journal of the American Statistical Association **93** (1998), no. 441, 359–371.

[16] T. Fissler, J. F. Ziegel, and T. Gneiting, *Expected Shortfall is jointly elicitable with Value at Risk - Implications for backtesting*, Risk (2016).

[17] H. Föllmer and T. Knispel, *Convex risk measures: Basic facts, law-invariance and beyond, asymptotics for large portfolios*, Handbook of the Fundamentals of Financial Decision Making, Part II (2013), 507–554.

[18] H. Föllmer and A. Schied, *Stochastic finance: An introduction in discrete time*, de Gruyter Studies in Mathematics 27, 2002.

[19] I. Francioni and F. Herzog, *Probability-unbiased Value-at-Risk estimators*, Quantitative Finance **12** (2012), no. 5, 755–768.

[20] P. Giot and S. Laurent, *Value-at-risk for long and short trading positions*, Journal of Applied Econometrics **18** (2003), no. 6, 641–663.

[21] T. Gneiting, *Making and evaluating point forecasts*, Journal of the American Statistical Association **106** (2011), no. 494, 746–762.

[22] M. Kaina and L. Rüschendorf, *On convex risk measures on L^p-spaces*, Mathematical methods of operations research **69** (2009), no. 3, 475–495.

[23] A. McNeil, R. Frey, and P. Embrechts, *Quantitative risk management: Concepts, techniques and tools*, Princeton University Press, 2005.

[24] A. J. McNeil, *Extreme value theory for risk managers*, Internal Modelling and CAD II published by RISK Books (1999), 93–113.

[25] E. Parzen, *On estimation of a probability density function and mode*, The annals of mathematical statistics (1962), 1065–1076.

[26] M. Rosenblatt, *Remarks on some nonparametric estimates of a density function*, The Annals of Mathematical Statistics **27** (1956), no. 3, 832–837.

[27] B. W. Silverman, *Density estimation for statistics and data analysis*, vol. 26, CRC press, 1986.

[28] M.K.P. So and P.L.H. Yu, *Empirical analysis of GARCH models in value at risk estimation*, Journal of International Financial Markets, Institutions and Money **16** (2006), no. 2, 180–197.

[29] J. F. Ziegel, *Coherence and elicitability*, Forthcoming in Mathematical Finance (2014).

Portfolio	Total number of exceptions divided by total length			Number of exceptions for non-overlapping subsets of length 250, for $n = 4$								
	$n = 4$	$n = 5$	$n = 6$									
LoBM.LoOP	0.0127	0.0163	0.0139	3	4	4	2	1	3	4	2	**6**
BM1.OP2	0.0092	0.0123	0.0163	4	2	2	0	1	4	4	0	3
BM1.OP3	0.0119	0.0135	0.0127	1	**5**	3	3	**5**	**5**	1	1	3
BM1.OP4	0.0119	0.0155	0.0147	3	**5**	3	2	2	4	1	2	**6**
LoBM.HiOP	0.0107	0.0143	0.0151	3	4	1	2	4	**5**	2	1	3
BM2.OP1	0.0103	0.0107	0.0135	3	**6**	3	1	2	1	1	2	**6**
BM2.OP2	0.0099	0.0107	0.0115	4	**6**	2	1	1	**5**	0	3	3
BM2.OP3	0.0088	0.0119	0.0115	1	2	3	2	2	4	2	1	4
BM2.OP4	0.0092	0.0135	0.0147	2	3	1	1	3	2	4	4	1
BM2.OP5	0.0127	0.0163	0.0135	1	4	**5**	3	2	**5**	4	1	4
BM3.OP1	0.0092	0.0115	0.0119	**5**	3	1	2	1	4	1	1	3
BM3.OP2	0.0107	0.0104	0.0131	4	**7**	1	2	1	3	2	1	4
BM3.OP3	0.0127	0.0111	0.0115	3	**5**	3	3	**5**	3	3	2	3
BM3.OP4	0.0123	0.0131	0.0119	**5**	**5**	1	2	4	3	1	3	5
BM3.OP5	0.0103	0.0143	0.0115	3	2	**5**	0	3	**5**	3	2	2
BM4.OP1	0.0127	0.0115	0.0127	**7**	4	2	3	1	2	4	3	4
BM4.OP2	0.0135	0.0135	0.0135	3	**7**	3	**5**	1	3	**5**	1	3
BM4.OP3	0.0143	0.0139	0.0108	1	**5**	2	**5**	4	**5**	4	2	4
BM4.OP4	0.0103	0.0119	0.0135	2	**6**	1	3	1	2	**5**	2	4
BM4.OP5	0.0123	0.0127	0.0104	1	1	**6**	2	1	3	3	**5**	5
HiBM.LoOP	0.0092	0.0115	0.0092	2	3	1	3	3	4	0	2	3
BM5.OP2	0.0092	0.0119	0.0147	4	3	1	2	2	2	1	1	3
BM5.OP3	0.0092	0.0155	0.0143	1	**5**	2	2	2	2	1	**5**	2
BM5.OP4	0.0111	0.0100	0.0139	2	3	**5**	3	3	3	2	3	1
HiBM.HiOP	0.0088	0.0107	0.0115	3	**6**	0	0	1	4	1	1	3

Table 4: *We take returns of 25 portfolios formed on book-to-market and operating profitability in the period from 01.01.2005 to 01.01.2015 from the Fama & French dataset, see [14]. As in Section 7.1, we calculate the mean number of exceptions in the whole set, for various learning periods (4 to 6 days). The computations are presented for the Gaussian unbiased estimator* $\hat{\mathrm{V@R}}^u_\alpha$ *defined in* (1.1)*. The results show surpsingly often that the estimator is in the green zone (0-4 exceptions) while using only a small level of historical data. The cases* **not** *in the green zone are marked in bold face.*

Portfolio	Estimator type									
	Hist	Norm	Mod	GPD	Stud	Bia	Boot1	Boot2	Boot3	Boot4
LoBM.LoOP	0.066	0.066	0.061	0.061	0.068	0.063	0.062	0.061	0.052	0.054
BM1.OP2	0.066	0.063	0.060	0.061	0.066	0.059	0.058	0.058	0.052	0.054
BM1.OP3	0.069	0.061	0.060	0.063	0.065	0.059	0.057	0.057	0.052	0.053
BM1.OP4	0.064	0.068	0.065	0.063	0.070	0.066	0.063	0.063	0.051	0.056
LoBM.HiOP	0.063	0.065	0.063	0.063	0.072	0.061	0.061	0.061	0.051	0.061
BM2.OP1	0.067	0.070	0.063	0.060	0.073	0.065	0.063	0.063	0.056	0.058
BM2.OP2	0.068	0.066	0.063	0.059	0.074	0.063	0.061	0.061	0.050	0.053
BM2.OP3	0.068	0.068	0.066	0.059	0.069	0.063	0.061	0.061	0.056	0.055
BM2.OP4	0.067	0.069	0.063	0.063	0.071	0.066	0.066	0.063	0.054	0.056
BM2.OP5	0.061	0.063	0.053	0.054	0.067	0.056	0.054	0.054	0.049	0.050
BM3.OP1	0.060	0.054	0.054	0.057	0.059	0.054	0.054	0.054	0.048	0.050
BM3.OP2	0.061	0.063	0.061	0.061	0.068	0.058	0.057	0.057	0.047	0.050
BM3.OP3	0.070	0.067	0.063	0.063	0.068	0.061	0.060	0.060	0.055	0.057
BM3.OP4	0.061	0.063	0.061	0.059	0.068	0.060	0.059	0.058	0.049	0.052
BM3.OP5	0.060	0.058	0.055	0.054	0.063	0.055	0.055	0.054	0.042	0.047
BM4.OP1	0.068	0.066	0.061	0.060	0.070	0.061	0.061	0.062	0.053	0.054
BM4.OP2	0.063	0.066	0.062	0.057	0.072	0.059	0.057	0.059	0.050	0.053
BM4.OP3	0.068	0.063	0.059	0.058	0.068	0.059	0.058	0.058	0.054	0.052
BM4.OP4	0.058	0.057	0.054	0.058	0.062	0.054	0.052	0.052	0.048	0.050
BM4.OP5	0.072	0.066	0.059	0.056	0.068	0.061	0.059	0.058	0.055	0.052
HiBM.LoOP	0.067	0.063	0.061	0.060	0.068	0.059	0.058	0.057	0.055	0.052
BM5.OP2	0.064	0.064	0.057	0.054	0.067	0.059	0.057	0.057	0.046	0.051
BM5.OP3	0.066	0.063	0.059	0.061	0.068	0.059	0.058	0.058	0.052	0.054
BM5.OP4	0.064	0.057	0.059	0.061	0.066	0.054	0.054	0.052	0.048	0.053
HiBM.HiOP	0.072	0.066	0.061	0.063	0.068	0.062	0.059	0.059	0.054	0.055
MEAN	0.065	0.064	0.060	0.060	0.068	0.060	0.059	0.058	0.051	0.053
DIST	0.015	0.014	0.010	0.010	0.018	0.010	0.009	0.008	0.003	0.004

Table 5: *We take returns of 25 portfolios formed on book-to-market and operating profitability in the period from 16.01.2009 to 01.01.2015 from the Fama & French dataset, see [14]. We perform the standard back-test, splitting the sample into intervals of length 50. The table presents the average rate of exception for Value-at-Risk at level 5%. MEAN denote the mean value of all numbers in a given column, while DIST denotes average distance from* 0.05*. It can be seen that for the biased estimators, the average rate is significantly higher than the expected rate of* 0.05 *while the unbiased estimators perform very well.*

Portfolio	Estimator type									
	Hist	Norm	Mod	GPD	Stud	Bia	Boot1	Boot2	Boot3	Boot4
LoBM.LoOP	0.068	0.058	0.061	0.060	0.058	0.054	0.051	0.051	0.051	0.051
BM1.OP2	0.069	0.057	0.064	0.056	0.057	0.046	0.043	0.042	0.048	0.048
BM1.OP3	0.069	0.070	0.063	0.059	0.071	0.064	0.059	0.059	0.049	0.052
BM1.OP4	0.058	0.048	0.050	0.052	0.048	0.045	0.044	0.043	0.043	0.046
LoBM.HiOP	0.059	0.057	0.054	0.049	0.058	0.052	0.051	0.051	0.045	0.043
BM2.OP1	0.065	0.057	0.060	0.061	0.059	0.054	0.054	0.052	0.048	0.050
BM2.OP2	0.065	0.048	0.054	0.057	0.048	0.043	0.041	0.042	0.047	0.050
BM2.OP3	0.072	0.059	0.062	0.062	0.063	0.057	0.057	0.057	0.051	0.057
BM2.OP4	0.060	0.056	0.057	0.058	0.057	0.049	0.048	0.050	0.046	0.049
BM2.OP5	0.067	0.059	0.057	0.057	0.059	0.054	0.051	0.052	0.049	0.055
BM3.OP1	0.064	0.059	0.057	0.057	0.064	0.054	0.054	0.053	0.049	0.054
BM3.OP2	0.070	0.055	0.054	0.057	0.056	0.051	0.051	0.051	0.052	0.051
BM3.OP3	0.063	0.061	0.061	0.060	0.063	0.057	0.054	0.055	0.052	0.054
BM3.OP4	0.074	0.051	0.053	0.061	0.052	0.047	0.045	0.042	0.050	0.050
BM3.OP5	0.063	0.054	0.057	0.053	0.054	0.045	0.042	0.040	0.048	0.046
BM4.OP1	0.070	0.055	0.054	0.059	0.055	0.052	0.052	0.052	0.050	0.052
BM4.OP2	0.071	0.063	0.065	0.070	0.063	0.060	0.057	0.058	0.056	0.057
BM4.OP3	0.062	0.058	0.057	0.059	0.059	0.052	0.050	0.050	0.051	0.054
BM4.OP4	0.070	0.061	0.061	0.061	0.061	0.059	0.058	0.059	0.052	0.055
BM4.OP5	0.068	0.050	0.053	0.056	0.051	0.045	0.043	0.044	0.044	0.048
HiBM.LoOP	0.062	0.059	0.057	0.055	0.059	0.053	0.052	0.052	0.047	0.048
BM5.OP2	0.077	0.063	0.062	0.058	0.065	0.054	0.052	0.050	0.052	0.050
BM5.OP3	0.068	0.057	0.059	0.061	0.059	0.051	0.048	0.048	0.056	0.052
BM5.OP4	0.075	0.054	0.062	0.065	0.054	0.050	0.050	0.049	0.053	0.055
HiBM.HiOP	0.068	0.054	0.050	0.051	0.056	0.048	0.046	0.047	0.048	0.045
MEAN	0.067	0.057	0.058	0.058	0.058	0.052	0.050	0.050	0.050	0.051
DIST	0.017	0.007	0.008	0.008	0.008	0.004	0.004	0.004	0.003	0.003

Table 6: *We fit a normal distribution to each portfolio from the Fama & French dataset, see [14]. From this distributions we simulate samples of size n = 1500 and perform the standard back-test, splitting the sample into intervals of length 50. The table presents the average rate of exception for Value-at-Risk at level 5%. It can be seen that for the biased estimators, the average rate is significantly higher than the expected rate of* 0.05 *while the unbiased estimators perform very well. MEAN denote the mean value of all numbers in a given column, while DIST denotes average distance from* 0.05.

Portfolio	Estimator type									
	Hist	Norm	Mod	GPD	Stud	Bia	Boot1	Boot2	Boot3	Boot4
LoBM.LoOP	0.063	0.056	0.060	0.061	0.064	0.052	0.052	0.052	0.049	0.053
BM1.OP2	0.070	0.048	0.067	0.056	0.058	0.041	0.041	0.041	0.057	0.048
BM1.OP3	0.064	0.046	0.074	0.059	0.055	0.042	0.041	0.039	0.052	0.052
BM1.OP4	0.077	0.043	0.067	0.053	0.051	0.041	0.039	0.039	0.059	0.044
LoBM.HiOP	0.068	0.050	0.090	0.061	0.054	0.044	0.044	0.044	0.052	0.052
BM2.OP1	0.060	0.048	0.060	0.056	0.054	0.043	0.043	0.043	0.050	0.046
BM2.OP2	0.072	0.052	0.063	0.057	0.057	0.047	0.047	0.047	0.057	0.050
BM2.OP3	0.071	0.046	0.050	0.055	0.059	0.046	0.044	0.044	0.053	0.044
BM2.OP4	0.061	0.041	0.062	0.051	0.049	0.037	0.037	0.038	0.049	0.043
BM2.OP5	0.067	0.040	0.072	0.053	0.047	0.038	0.037	0.038	0.053	0.049
BM3.OP1	0.076	0.058	0.067	0.066	0.066	0.052	0.052	0.052	0.056	0.059
BM3.OP2	0.064	0.048	0.086	0.059	0.057	0.046	0.046	0.046	0.054	0.052
BM3.OP3	0.067	0.041	0.060	0.052	0.053	0.039	0.037	0.037	0.050	0.044
BM3.OP4	0.071	0.032	0.095	0.056	0.049	0.028	0.028	0.028	0.050	0.050
BM3.OP5	0.067	0.046	0.067	0.061	0.057	0.045	0.045	0.043	0.052	0.054
BM4.OP1	0.070	0.043	0.119	0.057	0.055	0.041	0.039	0.041	0.052	0.050
BM4.OP2	0.069	0.043	0.056	0.055	0.059	0.042	0.042	0.041	0.056	0.049
BM4.OP3	0.070	0.051	0.088	0.070	0.060	0.048	0.048	0.048	0.057	0.059
BM4.OP4	0.067	0.041	0.069	0.057	0.048	0.036	0.036	0.035	0.046	0.048
BM4.OP5	0.064	0.050	0.084	0.059	0.062	0.050	0.049	0.050	0.049	0.050
HiBM.LoOP	0.068	0.043	0.053	0.050	0.055	0.042	0.042	0.041	0.048	0.046
BM5.OP2	0.061	0.034	0.103	0.057	0.052	0.033	0.033	0.033	0.048	0.046
BM5.OP3	0.070	0.048	0.078	0.061	0.059	0.046	0.043	0.045	0.054	0.050
BM5.OP4	0.063	0.049	0.066	0.055	0.060	0.048	0.048	0.048	0.047	0.050
HiBM.HiOP	0.066	0.052	0.060	0.061	0.060	0.046	0.044	0.045	0.053	0.054
MEAN	0.067	0.046	0.073	0.058	0.056	0.043	0.042	0.042	0.052	0.050
DIST	0.017	0.006	0.023	0.008	0.006	0.007	0.008	0.008	0.003	0.003

Table 7: *In contrast to Table 6 where a normal distribution was used, here we fit a Student-t-distribution to the data. The table presents again the average rate of exception for Value-at-Risk at level 5%. In this case the unbiased estimators (which assume normality) overestimate the risk, i.e. are conservative. The proposed non-parametric bootstrap algorithms (Boot3 and Boot4) correct for the different distribution and perform well even in this case. MEAN denote the mean value of all numbers in a given column, while DIST denotes average distance from* 0.05*. It can be seen that for the biased estimators, the average rate is significantly higher than the expected rate of* 0.05 *while the unbiased estimators perform very well.*

Estimator		Exceeds		T_1		T_2
		mean	sd	mean	sd	mean
Percentile	$\hat{\text{V@R}}^{\text{emp}}_{\alpha}(x)$	0.067	0.0048	29.4%	11.7%	99%
Modified C-F	$\hat{\text{V@R}}^{\text{mod}}_{\alpha}(x)$	0.057	0.0046	9.7%	3.8%	82%
Gaussian	$\hat{\text{V@R}}^{\text{norm}}_{\alpha}(x)$	0.057	0.0044	11.1%	6.6%	82%
GPD	$\hat{\text{V@R}}^{\text{GPD}}_{\alpha}(x)$	0.058	0.0044	12.5%	8.3%	83%
Gaussian unbiased	$\hat{\text{V@R}}^{\text{u}}_{\alpha}(x)$	0.052	0.0045	-	-	-

Table 8: *We fit a normal distribution to the first portfolio from the Fama & French dataset, i.e. LoBM.LoOP portfolio. See [14] for details. From this distributions we simulate 10000 times strong sample of size* $n = 1500$ *and perform the standard back-test 10000 times. In other words, we replicate 10000 times the result from the first row of Table 7. The table presents the average rate of exception for Value-at-Risk at level 5%. It can be seen that for the biased estimators, the average mean exception rate is significantly higher than the expected rate of* 0.05 *while the Gaussian unbiased estimator perform very well. Statistic* T_1 *shows that the exceedance rate for Gaussian unbiased estimator is usually lower in comparison with other estimators, eliminating the effect of risk underestimation. Moreover, statistic* T_2 *shows that in almost all cases, the exception rate for Gaussian unbiased estimator is closer to 0.05, than the exception rate of any other of the considered estimators.*