Interval Estimation of Value-at-Risk Based on Nonparametric Models
Hussein Khraibani, Bilal Nehme, Olivier Strauss

To cite this version:

HAL Id: lirmm-01951982
https://hal-lirmm.ccsd.cnrs.fr/lirmm-01951982
Submitted on 11 Dec 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License
Interval Estimation of Value-at-Risk Based on Nonparametric Models

Hussein Khraibani 1,*, Bilal Nehme 2 and Olivier Strauss 3

1 Department of Applied Mathematics, Faculty of Sciences, Lebanese University, Beirut 2038 1003, Lebanon
2 Department of Economics, Faculty of Economic Sciences & Business Administration, Lebanese University, Beirut 2038 1003, Lebanon; Bilal.Nehme@ul.edu.lb
3 Department of Robotics, LIRMM University of Montpellier II, 61 rue Ada, 34392 Montpellier CEDEX 5, France; strauss@lirmm.fr
* Correspondence: hussein.khraibani@gmail.com; Tel.: +961-70453332

Received: 6 December 2018; Accepted: 6 December 2018; Published: 10 December 2018

Abstract: Value-at-Risk (VaR) has become the most important benchmark for measuring risk in portfolios of different types of financial instruments. However, as reported by many authors, estimating VaR is subject to a high level of uncertainty. One of the sources of uncertainty stems from the dependence of the VaR estimation on the choice of the computation method. As we show in our experiment, the lower the number of samples, the higher this dependence. In this paper, we propose a new nonparametric approach called maxitive kernel estimation of the VaR. This estimation is based on a coherent extension of the kernel-based estimation of the cumulative distribution function to convex sets of kernel. We thus obtain a convex set of VaR estimates gathering all the conventional estimates based on a kernel belonging to the above considered convex set. We illustrate this method in an empirical application to daily stock returns. We compare the approach we propose to other parametric and nonparametric approaches. In our experiment, we show that the interval-valued estimate of the VaR we obtain is likely to lead to more careful decision, i.e., decisions that cannot be biased by an arbitrary choice of the computation method. In fact, the imprecision of the obtained interval-valued estimate is likely to be representative of the uncertainty in VaR estimate.

Keywords: risk measures; quantile estimation; financial time series; Value-at-Risk; choquet integral; possibility theory; maxitive kernel; kernel estimation; parametric models

1. Introduction

Controlling financial risk is an important issue for financial institutions. For the necessity of risk management, the first task is to measure risk. Value-at-Risk (VaR) is probably the most widely used risk measurement in financial institutions. It has made its way into the Basel II Capital-Adequacy framework. VaR is an estimate of the largest loss over a specified time horizon at a particular probability level. For example, if the daily VaR of an investment portfolio is $10 million with a 95% confidence level, this means that we can be 95% confident that the portfolio will not lose more than $10 million over the next day. In a more formal way, the VaR of a portfolio at a probability level \( \alpha \) can be defined as McNeil et al. (2005, p. 38).

\[
\text{VaR}_\alpha = \inf \{ x, F_X(x) = P(X \leq x) \geq \alpha \} = F_X^{-1}(\alpha),
\]

where \( X \) is a random variable representing daily percent return for a total stock index with Cumulative Distribution Function (CDF) \( F_X \). As McNeil et al. (2005) point out, VaR is simply a quantile of the corresponding loss distribution, which makes its computation easy. Various methods have been proposed in the relevant literature to compute the VaR. Each method differs in the methodology used,
the hypotheses, and the way the models are implemented. A highly questionable fact is that each method leads to different results. Thus the choice of computational method is likely to have a large impact on the way a financial institution manages its credit portfolio.

The most commonly used approach is the nonparametric Historical Simulation (HS) method described in Linsmeier and Pearson (1997). It is based on the empirical CDF of the historically simulated returns by attributing an equal probability weight to each day’s return. The HS approach is easy to implement but suffers from two major drawbacks: First, the success of the approach depends on the ability to collect a large series of data; second, this is an unconditional model and thus we need a number of extreme scenarios in the historical record to provide more informative estimates of the tail of the loss distribution McNeil et al. (2005). Moreover, estimates of extreme quantiles are inefficient since extrapolation beyond past observations is impossible. To avoid these drawbacks, several parametric and nonparametric approaches have been developed. For example, Butler and Schachter (1998) propose the kernel estimators in conjunction with the historical simulation method. Also, Charpentier and Oulidi (2010) calculate the VaR by using several nonparametric estimators based on Beta kernel. This study shows that those estimators improve the efficiency of traditional ones, not only for light tailed, but also heavy tailed distributions.

Numerous recent VaR models are referred to as parametric. From the point of view of the risk manager, estimate VaR assuming normal distribution of asset returns is inappropriate and lead to underestimate the left tail at low probability level. For this reason, most recent research papers deal with going beyond the normal model and attempt to capture the related phenomena of heavy and long tails and asymmetric form of the returns series. EVT (Extreme Value Theory) and the so-called GHYP (Generalized HYPerbolic) distributions are among the most widely used. The main advantage of hypothesizing a GHYP distribution is its ability to account for the statistical properties of financial market data such as volatility clustering, asymmetry and heavy-tail phenomena (see McNeil et al. (2005) for an introduction and Paolella and Polak (2015) for a recent application). Kuester et al. (2006) use an EVT-based approach and focuses on the long tails of the return distribution. Braione and Scholtes (2016) study the performance of forecasting VaR under different parametric distributional assumptions and show the predominance and the predictive ability for the skewed and heavy-tailed distributions in the univariate case. The main drawback of using a parametric model is the high dependency of the obtained method to the hypothesized distribution model. To lower this dependency, some authors have proposed a group of semiparametric methods that are based on extreme value theory. These methods have been successfully used by financial analysts in estimating VaR Danielsson and De Vries (2000). However, all the above mentioned attempts have the common weakness of a low robustness to modelling leading to variations in the VaR estimation. To enable this method to be used in a prudent manner, it would be of instrumental interest to estimate the variation of the computed VaR w.r.t. the modelling. Some attempts have been proposed to achieve such a goal. For example, Butler and Schachter (1998) propose a method to measure the precision of a VaR estimate. Jorion (1996) suggests that VaR always be reported with confidence intervals and shows that it is possible to improve the efficiency of VaR estimates using their standard errors. Kupiec (1995) proposes a method for quantifying the uncertainty, in the estimated VaR, induced by the fact that the return distribution is unknown. Pritsker (1997) proposes to estimate a standard error by using a Monte Carlo VaR analysis.

In this paper, we propose a completely new approach to estimate the variations in the VaR induced by choosing a particular model in a kernel-based approach. The idea is that an estimate of the CDF of the daily percent return can be used to compute the VaR as suggested by Equation (1). Such an estimate can be obtained by using a kernel-based approach Silvermann (1986). In Kernel Density Estimation (KDE), the role of the kernel is to achieve an interpolation to lower the impact of sampling on the obtained estimation. Roughly speaking, it smoothes the empirical CDF. However, as in the above mentioned methods, there is a systematic bias induced by the choice of the kernel used to estimate the CDF. Our proposal is to focus on the new nonparametric approach developed by
Loquin and Strauss (2008a) to estimate the CDF. This approach makes use of the ability of a new kind of interpolating kernel, called maxitive kernel, to represent a convex set of conventional kernels—that we call summative kernels Loquin and Strauss (2008b). Within this approach, the estimate of the CDF is interval-valued. Such an interval-valued estimate of the CDF, also called a p-box Destercke et al. (2007), is the convex set of all the CDF that would have been computed by using the convex set of conventional summative kernel represented by the maxitive kernel. We show that this approach can advantageously be used to compute with accuracy the corresponding convex set of kernel-based VaR estimates. This approach has advantages over Monte Carlo based approaches since its complexity is comparable to that of classical kernel estimation. Moreover, within this approach the bounds are exact while Monte Carlo approaches provide an inner estimation of those bounds.

This paper is structured as follows. Section 2 reviews some nonparametric and parametric approaches and bootstrap methods used to compute the confidence interval of VaR. In Section 3, we introduce the empirical distribution function and the kernel cumulative estimator based on summative kernel. We define in Section 4 the maxitive kernel, which forms the basis of our approach, and it is shown how an interval-valued estimation of VaR, based on maxitive kernel, can be derived that has relevant properties in this context. Section 5 presents and discusses our empirical findings. Firstly, we show how the choice of kernel function affects the VaR estimates. Secondly, we investigate the performance of the interval-valued proposed in this paper by comparing it to three very competitive approaches: The simple Normal VaR, the HS VaR, and the GHYP VaR. Finally, Section 6 concludes this paper with some further remarks.

Throughout this paper, we consider that the observations belong to a convex and compact subset (universe) Ω of IR, called the reference set. P(Ω) is the collection of all measurable subsets of Ω. It naturally contains the empty set and is closed under complementation and countable unions. Then (Ω, P(Ω)) is a measurable space. Let L(Ω) be the set of functions defined on Ω with values in IR.


2.1. Historical Simulation

The HS method calculates the Value-at-Risk using real historical data of asset returns and captures the non-normal distribution of the returns. HS is a nonparametric method because it doesn’t make a specific assumption about distribution of returns. However HS method assumes that the distribution of past returns is a good and complete representation of expected future returns.

The VaR with α% of probability level is calculated as the α% percentile of the sorted data return values. For example, with a returns data with 1000 observations, the 1% VaR estimate is simply the negative of the 10th sample order statistic. This HS VaR can be defined as follows:

$$\text{VaR}_α = \text{percentile}\left(\{r_t\}_{t=1}^T, α\%ight),$$

where $r_t$ is the asset return at time $t$.

2.2. GHYP Parametric Distribution

The generalized hyperbolic distribution (GHYP) was introduced by Barndorff-Nielsen (1978) to fit financial returns. The GHYP is an asymmetric heavy-tailed distribution that can account for the extreme events and cater for skewness embedded in the data. It has since been applied in diverse disciplines such as physics, biology, financial mathematics (see Eberlein and Keller (1995); Sørensen and Bibby (1997); Paolella (2007, chp. 9)).

The probability density function (pdf) of univariate GHYP distribution with the parameterization of Eberlein et al. (1998) is given, for $x \in \mathbb{R}$, by:
\[ f_{\text{GHD}}(x; \lambda, \chi, \psi, \mu, \sigma^2, \gamma) = a K_{\lambda-1/2} \left( \sqrt{\frac{(x - \mu)^2}{\sigma^2}} \right) \exp \left( \frac{\gamma (x - \mu)}{\sigma^2} \left( \frac{(x - \mu)^2}{\sigma^2} \right) \right)^{\lambda/2-1/4}, \]

with the normalizing constant
\[ a = \left( \frac{\chi \psi}{2\pi} \right)^{1/2} \sigma K_{\lambda} \left( \sqrt{\chi \psi} \right), \]

where

(a) \( K_{\lambda} \) denotes a modified Bessel function of the third kind with index \( \lambda \).
(b) \( \lambda \) defines the subclasses of GHYP and is related to the fail flatness.
(c) \( \chi \) and \( \psi \) determine the distribution shape; in general, the larger those parameters are, the closer the distribution is to the normal distribution.
(d) \( \mu \) is the location parameter.
(e) \( \sigma \) is the dispersion parameter (standard deviation)
(f) \( \gamma \) is the skewness parameter (if \( \gamma = 0 \), the distribution reduces to the symmetric generalized hyperbolic distribution).

The GHYP family contains many special cases known under special names, listed as follows:

- **Hyperbolic Distribution (HYP)**
  If \( \lambda = 1 \), we get the hyperbolic distribution. However, HYP distribution is characterized by having a hyperbolic log-density function whereas the log-density for the normal distribution is a parabola. Thus, one may expect the HYP distribution to be a coherent alternative for heavy-tailed data.

- **Normal Inverse Gaussian (NIG) Distribution**
  If \( \lambda = -\frac{1}{2} \), then the distribution is known as normal inverse gaussian (NIG). NIG distribution is also widely used in financial modeling.

- **Variance Gamma (VG) Distribution**
  If \( \lambda > 0 \) and \( \chi = 0 \), then we obtain the limiting case which is known as variance gamma (VG) distribution.

- **The Skewed Student’s t-Distribution (St)**
  When \( \lambda < 0 \) and \( \chi = 0 \) we get another limiting case called the generalized hyperbolic skew student’s t distribution because it generalizes the usual Student t distribution, obtained from the skewed-t by setting the skewness parameter \( \gamma = 0 \).

In order to estimate the unknown parameters \((\lambda, \chi, \psi, \mu, \sigma^2, \gamma)\), one can use the maximum likelihood (ML) with numerical optimization method (EM-based algorithm).

2.3. Nonparametric Bootstrap Approach

The bootstrap method is used in an important number of statistics topics based on building a sampling distribution for a statistic by resampling from the data at hand. The term bootstrap was coined by Efron (1979), and is an allusion to the expression **pulling oneself up by one’s bootstraps**—in this case, using the sample data as a population from which repeated samples are drawn Fox (2002). The nonparametric bootstrap allows us to empirically estimate the sampling distribution of a statistic \( \theta \)—such as a mean, median, standard deviation, or quantile (VaR)—or an estimator \( \hat{\theta} \) without making assumptions about the form of the population, and without deriving the sampling distribution explicitly. The basic idea of the nonparametric bootstrap is as follows: Assuming a data set \( S = (x_1, \ldots, x_n) \) is available. We draw a sample of size \( n \) from among the elements of \( S \), sampling with
replacement. This result is called bootstrap sample \( S_1^* = (x_{11}^*, \ldots, x_{1n}^*) \). We repeat this procedure a large number of times, \( B \), selecting many bootstrap samples; the \( b \)th such bootstrap sample is denoted \( S_b^* = \{ x_{b1}^*, \ldots, x_{bn}^* \} \). Next, we compute the statistic \( \hat{\theta} \) for each of the bootstrap samples; that is \( \theta_b^* = f(x_b^*) \), with \( f \) denoting some function, that we can use to estimate from data. Then the distribution of \( \theta_b^* \) around the original estimate \( \hat{\theta} \) is analogous to the sampling distribution of the estimator \( \hat{\theta} \) around the population parameter. The variance is estimated by the sample variance (but for the bootstrap sample and the original \( \hat{\theta} \)). Then, the bootstrap estimates of bias and variance are given approximately by:

\[
\text{Bias}(\hat{\theta}^*) \approx \frac{1}{B} \sum_{b=1}^{B} \hat{\theta}_b^* - \hat{\theta} = \hat{\theta}^* - \hat{\theta} \quad \text{and} \quad \text{var}(\hat{\theta}^*) \approx \frac{1}{B-1} \sum_{b=1}^{B} (\hat{\theta}_b^* - \hat{\theta}^*)^2.
\]

There are various methods for constructing bootstrap confidence intervals. The normal-theory interval assumes that the distribution of \( \hat{\theta} \) is normally distributed (which is often approximately the case for statistics in sufficiently large samples), and uses the bootstrap estimate of sampling variance, and perhaps of bias, to construct a 100\((1 - \alpha)\)-percent confidence interval of the form:

\[
\left[ \hat{\theta} - z_{1-\alpha/2} \text{se}(\hat{\theta}), \hat{\theta} + z_{1-\alpha/2} \text{se}(\hat{\theta}) \right],
\]

where \( \text{se}(\hat{\theta}) = \sqrt{\text{var}(\hat{\theta})} \) is the bootstrap estimate of the standard error of \( \hat{\theta} \), and \( z_{1-\alpha/2} \) is the \( 1 - \alpha/2 \) quantile of the standard-normal distribution (e.g., \( z = 1.96 \) for a 95-percent confidence interval, where \( \alpha = 0.05 \)).

We can also estimate confidence intervals using percentiles of the sample distribution: The upper and lower bounds of the confidence interval would be given by the percentile points (or quantiles) of the sample distribution of parameter estimates. This alternative approach, called the bootstrap percentile interval, is to use the empirical quantiles of \( \hat{\theta}_b^* \) to form a confidence interval \( [\hat{\theta}_{(lo)}^*, \hat{\theta}_{(up)}^*] \), where \( \hat{\theta}_{(1)}, \ldots, \hat{\theta}_{(n)}^* \) are the ordered bootstrap replicates of the statistic; \( \text{lo} = [(r + 1)\alpha/2]; \text{up} = [(r + 1)(1 -\alpha/2)]; \) and \( \lfloor \cdot \rfloor \) is the nearest integer function. This basic percentile interval approach is limited itself, particularly if parameter estimators are biased (Dowd 2005). It is therefore often better to use the bias-corrected, accelerated (or BC\(_a\)) percentile intervals. In order to find the BC\(_a\), we first calculate:

\[
z = \Phi^{-1}(\alpha),
\]

where \( \Phi^{-1}(\cdot) \) is the standard-normal quantile function, and \( \rho = \frac{\sum_{b=1}^{B} (\hat{\theta}_b^* - \hat{\theta})}{B-1} \) is the adjusted proportion of bootstrap replicates at or below the original-sample estimate \( \hat{\theta} \). If the bootstrap sampling distribution is symmetric, and if \( \hat{\theta} \) is unbiased, then this proportion will be close to 0.5, and \( z \) will be close to 0. Now, let \( \hat{\theta}_{(-i)} \) be the value of \( \hat{\theta} \) produced when the \( i \)th observation is deleted from the sample there are \( n \) of these quantities. Let \( \bar{\theta} = \frac{1}{n-1} \sum_{i=1}^{n} \hat{\theta}_{(-i)} \) be the average of the \( \hat{\theta}_{(-i)} \). We calculate:

\[
\beta = \frac{n}{B} \sum_{i=1}^{n} (\hat{\theta}_{(-i)} - \bar{\theta})^3,
\]

\[
\beta_1 = \Phi \left[ z + \frac{z - z_{1-\alpha/2}}{1 - \beta(z - z_{1-\alpha/2})} \right] \quad \text{and} \quad \beta_2 = \Phi \left[ z + \frac{z + z_{1-\alpha/2}}{1 - \beta(z + z_{1-\alpha/2})} \right],
\]

where \( \Phi(\cdot) \) is the standard-normal cumulative distribution function. The values \( \beta_1 \) and \( \beta_2 \) are used to locate the endpoints of the corrected percentile confidence interval: \( [\hat{\theta}_{(lo)}^*, \hat{\theta}_{(up)}^*] \), where \( \text{lo}^* = [r\beta_1] \) and \( \text{up}^* = [r\beta_2] \). When \( \beta \) and \( z \) are both 0, \( \beta_1 = \Phi(-z_{1-\alpha/2}) = \Phi(z_{1-\alpha/2}) = \frac{1}{2} \) and \( \beta_2 = \Phi(z_{1-\alpha/2}) = 1 - \frac{1}{2} \), which corresponds to the (uncorrected) percentile interval.

For more background on the bootstrap approach and a broader array of applications see Efron and Tibshirani (1993); Davison and Hinkley (1997).
The bootstrap approach can be used to estimate the Value-at-Risk (VaR) and a confidence intervals for VaR. A resampling method based on the bootstrap and a bias-correction to improving the Value-at-Risk (VaR) forecasting ability of the normal-GARCH model has been developed in Hartz et al. (2006). As mentioned in Dowd (2005), if we have a data set of \(n\) observations, we create a new data set by taking \(n\) drawings, each taken from the whole of the original data set. Each new data set created in this way gives us a new VaR estimate. We then create a large number of such data sets and estimate the VaR of each. The resulting VaR distribution function enables us to obtain estimates of the confidence interval for our VaR. For estimate confidence intervals using a bootstrap approach, we produce a bootstrapped histogram of resample-based VaR estimates, and then read the confidence interval from the quantiles of this histogram. A very good discussion on this and other improvements can be seen in Dowd (2005, chp. 4). In Section 5.3, we use the bias corrected method to estimate the confidence interval of VaR obtained form HS, Normal and GHYP distribution. The purpose of this simulation exercise is to compare these methodologies with our maxitive kernel approach, in order to identify that our interval-valued estimation of VaR perform properly.

3. Summative Kernels

3.1. Summative Kernels and Probability Distributions

Kernels are used in signal processing and nonparametric statistics to define a weighted neighborhood around a location \(u \in \Omega\). In kernel regression, it is used to estimate the conditional expectation of a random variable (see e.g., Silvermann 1986; Wand and Jones 1995).

**Definition 1.** A summative kernel—or conventional kernel—is a positive function \(\kappa : \Omega \rightarrow \mathbb{R}^+\) that verifies the summativity property

\[
\int_{\Omega} \kappa(u) du = 1.
\]

From a basic summative kernel \(\kappa\), we can define a summative kernel \(\kappa_{\Delta}^x\) translated in \(x \in \Omega\) and dilated with a bandwidth \(\Delta > 0\) by

\[
\forall u \in \Omega, \kappa_{\Delta}^x(u) = \frac{1}{\Delta} \kappa\left(\frac{u - x}{\Delta}\right).
\]

By convention, \(\kappa(u) = \kappa_1^0(u)\).

The most used summative kernels in functional estimation are usually unimodal, symmetric, and centered (i.e., defining a weighted neighborhood around the origin). When \(\kappa(u)\) is symmetric and unimodal function, then the following conditions are fulfilled

\[
\int_{\Omega} u \kappa(u) du = 0 \text{ and } \int_{\Omega} u^2 \kappa(u) du = c > 0.
\]

For example, the Epanechnikov kernel proposed by Epanechnikov (1969) illustrated in Figure 1 is defined by

\[
\kappa(u) = \begin{cases} 
\frac{3}{4}(1 - u^2) & \text{if } |u| \leq 1, \\
0 & \text{otherwise.}
\end{cases}
\]

In the following, \(\mathcal{K}(\Omega)\) represents the set of unimodal, symmetric, and centered kernels. A summative kernel \(\kappa\) can be seen as a probability distribution inducing a probability measure \(P_\kappa : \mathcal{P}(\Omega) \rightarrow [0, 1]\) on the measurable space \((\Omega, \mathcal{P}(\Omega))\). \(P_\kappa\) is defined by

\[
\forall A \subseteq \Omega, P_\kappa(A) = \int_A \kappa(u) du.
\]
Considering the summative kernel \( \kappa \), we can define what we call the \textit{summative expectation} as follows:

\begin{definition}

Let \( s \) be a function of \( L(\Omega) \) and let \( \kappa \) be a summative kernel of \( K(\Omega) \). The \textit{summative expectation} of \( s \), in the neighborhood defined by \( \kappa \), is the classical expectation of \( s \) w.r.t. \( P_\kappa \):

\[
E_\kappa(s) = \int_\Omega s \, dP_\kappa = \int_\Omega s(u) \kappa(u) \, du,
\]

since \( dP_\kappa = \kappa(u) \, du \).

\end{definition}

### 3.2. Empirical Distribution Function

Let \( (x_1, \ldots, x_n) \) be a finite set of observations of \( n \) random variables \((X_1, \ldots, X_n)\) i.i.d with unknown pdf \( f \) and CDF \( F \). One can estimate \( F \) by the empirical CDF \( F_n \) defined by

\[
F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}(x_i \leq x),
\]

where \( \mathbb{I} \) is the indicator function, namely \( \mathbb{I}(x_i \leq x) = 1 \) if \( x_i \leq x \) and zero otherwise. Obviously \( \mathbb{I}(x_i \leq x) = 1 \) with probability \( P(x_i \leq x) = F(x) \), and \( \mathbb{I}(x_i > x) = 0 \) with probability \( P(x_i > x) = 1 - F(x) \). Then \( \mathbb{I}(x_i \leq x) \) is a Bernoulli random variable with success probability \( F(x) \). Since the \( x_i \) are independent, so are the \( \mathbb{I}(x_i \leq x) \). Thus \( Y_n = \sum_{i=1}^n \mathbb{I}(x_i \leq x) \) (a sum of independent \( \text{Ber}\{F(x)\} \) random variables) is a \( \text{Bin}\{n, F(x)\} \) random variable, and

\[
F_n(x) \sim \frac{1}{n} \text{Bin}\{n, F(x)\}.
\]

Then \( \forall x \in \Omega \),

\[
E(F_n(x)) = F(x) \text{ and } Var(F_n(x)) = \frac{F(x)(1 - F(x))}{n}.
\]

This implies that the empirical CDF is \textit{convergent in probability} to the true CDF:

\[
\forall x \in \Omega, \ P_\kappa(F_n(x) \rightarrow F(x)).
\]

Thus the empirical estimate of Value-at-Risk
\[
\text{VaR}_{x,n} = \inf \left\{ x, F_n(x) = P(X \leq x) \geq \alpha \right\},
\]

converges towards the true VaR \( \alpha \).

### 3.3. Summative Kernel Cumulative Estimator

The empirical distribution function (Equation (4)) is not smooth as it jumps by \( \frac{1}{n} \) at each point \( x_i \) \((i = 1 \ldots n)\). A kernel estimator of the CDF, introduced by authors such as Nadaraya (1964); and Watson and Leadbetter (1964), is a smoothed version of the empirical distribution estimator. Such an estimator arises as an integral of the Parzen-Rosenblatt kernel density estimator (see Rosenblatt 1956; Parzen 1962).

**Definition 3.** The summative-kernel cumulative estimator of the CDF (also called the Parzen-Rosenblatt kernel cumulative estimator), based on the summative kernel \( \kappa \), is given, in each point \( x \in \Omega \), by

\[
\hat{F}_n^\kappa(x) = \frac{1}{n} \sum_{i=1}^{n} \Gamma(x - x_i),
\]

where \( \Gamma(x) = \int_{-\infty}^{x} \kappa(u)du \).

Let \( \kappa \in K(\Omega) \) be a summative kernel function of the second order with support \([-1, 1]\), the function \( \Gamma(u) \) verifies the following properties

\[
\Gamma(u) = \begin{cases} 
0 & \text{if } u \in ]-\infty, -1], \\
1 & \text{if } u \in [1, \infty].
\end{cases}
\]

\[
\int_{-1}^{1} u^2 \kappa(u)du = \frac{1}{2}, \\
\int_{-1}^{1} u \kappa(u)du = \frac{1}{2}
\]

Note that if \( \kappa \) is the pdf of a probability measure \( P_\kappa \), then \( \Gamma \) is the cumulative distribution function of \( P_\kappa \). For example, \( \kappa \) being the Epanechnikov kernel defined by (2), the function \( \Gamma \), depicted in Figure 2, has the following expression:

\[
\Gamma(u) = \begin{cases} 
0 & \text{if } u \leq -1, \\
\frac{3}{4} u^3 + \frac{3}{4} u + \frac{1}{2} & \text{if } |u| \leq 1, \\
1 & \text{if } u \geq 1.
\end{cases}
\]

Let \( \kappa \in K(\Omega) \) be a summative kernel with support \([-1, 1]\), for a fixed \( x \), the bias and the variance of \( \hat{F}_n^\kappa(x) \) are given by

\[
\text{bias} \left( \hat{F}_n^\kappa(x) \right) = \frac{\Delta^2}{2} f''(x) \int_{-1}^{1} u^2 \kappa(u)du + o(\Delta^2),
\]

\[
\text{var} \left( \hat{F}_n^\kappa(x) \right) = \frac{1}{n} \left\{ F(x) \left( 1 - F(x) \right) + \Delta f(x) \left( \int_{-1}^{1} \kappa^2(u)du - 1 \right) + o(\Delta) \right\}.
\]

Some theoretical properties of the estimator \( \hat{F}_n^\kappa \) have been investigated, among others, by Winter (1973); Winter (1979); Sarda (1993); Yamato (1973); and Jones (1990). Some properties have long been known, e.g., the uniform convergence of \( \hat{F}_n^\kappa \) towards \( F \) when the pdf \( f = dF \) is
continuous Nadaraya (1964); and Yamato (1973) or without conditions on \( f \) Singh et al. (1983). The asymptotic expression of the Mean Integrated Squared Error (MISE) (i.e., \( \int_{\Omega} (\hat{F}_n(x) - F(x)) dx \)^2) has been studied in Swanepoel (1988). For a continuous pdf, it has been proved that the best kernel is the uniform kernel of bandwidth \( \Delta > 0 \) defined by \( \kappa(u) = \frac{1}{2\Delta} 1_{[-\Delta,\Delta]}(u) \) and for a discontinuous function \( f \), in a finite number of points, the best kernel is the exponential kernel of bandwidth \( \Delta > 0 \) defined by \( \kappa(u) = \frac{\Delta}{2} e^{-\Delta |u|}. \)

![Figure 2. Integrated Epanechnikov \( \kappa \) kernel with a bandwidth \( \Delta = 1 \).](image)

The method of choosing the optimal value of bandwidth in kernel estimation of the cumulative distribution function is of crucial interest. Many procedures—such as plug-in and cross validation—have been proposed in the relevant literature (see e.g., Sarda 1993; Polansky and Baker 2000; Altman and Leger 1995) to choose (estimate) the optimal bandwidth for estimating the CDF of the random process underlying a sample. As mentioned in Quintela Del Rio and Estevez-Perez (2013), the Polansky and Baker plug-in bandwidth is given by

\[
\Delta_{PB} = \left( \frac{\zeta(\kappa)}{-n\zeta_2(\kappa)\hat{V}(g_2)} \right)^{1/3},
\]

where

\[
\zeta_2(\kappa) = \int_{\Omega} u^2 \kappa(u) du ; \quad \zeta(\kappa) = 2 \int_{\Omega} u \kappa(u) \Gamma(u) du,
\]

\[
\hat{V}(g) = \frac{1}{n^2 g^{r+1}} \sum_{i=1}^{n} \sum_{j=1}^{n} H(r) \left( \frac{x_i - x_j}{g} \right)
\]

estimates

\[
V_r = \int_{\Omega} f^{(r)}(u) f(u) du,
\]

where \( r \geq 2 \) is an even integer and

\[
g_2 = \left( \frac{2H(0)}{-n\zeta_2(H)\hat{V}(4)} \right)^{1/5}.
\]

The kernel function \( H \) is not necessarily equal to \( \kappa \). An iterative method for calculating the plug-in bandwidth has been proposed by Polansky and Baker (2000). As mentioned in Quintela Del Rio and
Estevez-Perez (2013), the plug-in bandwidth is calculated as follows: let $b > 0$ be an integer, firstly, we calculate $\hat{V}_2 b + 2$ using
\[
\hat{V}_r = \frac{(-1)^{r/2} r!}{(2\hat{\sigma}(x_i))^{r+1}(r/2)! (3.14)^{1/2}},
\]
where $\sigma$ is the standard deviation of the data which can be estimated by $\hat{\sigma}(x_i) = \min\{\hat{s}, \frac{Q_3 - Q_1}{1.349}\}$, with $\hat{s}$ is the sample standard deviation, and $Q_1, Q_3$ denoting the first and third quartile, respectively.

Secondly, begin from $j = b$ to $j = 1$, we calculate $\hat{V}_2 j (\hat{g}_2 j)$, where
\[
\hat{g}_2 j = \left(\frac{2H^{(2)}(0)}{-n\xi^2(\kappa)\hat{V}_2 j (\hat{g}_2 j)}\right)^{1/3},
\]
with
\[
\hat{V}_2 j + 2 = \begin{cases} 
\hat{V}_2 b + 2, & \text{if } j = b, \\
\hat{V}_2 j + 2 (\hat{g}_2 j + 2), & \text{if } j < b.
\end{cases}
\]

Thirdly, the plug-in bandwidth is
\[
\hat{\Delta}_{PB} = \left(\frac{\zeta(\kappa)}{-n\xi^2(\kappa)\hat{V}_2 j (\hat{g}_2 j)}\right)^{1/3}.
\]

In practice, for most applications, we consider $b = 2$ Quintela Del Rio and Estevez-Perez (2013). Sarda (1993) introduced a cross-validation method to estimate the optimal bandwidth which minimize the MISE:
\[
CV(\Delta) = \sum_{i=1}^{n} \left(F_n(x_i) - \hat{F}^n_{\kappa_{i-1}}(x_i)\right)^2,
\]
where $F_n$ is the empirical distribution function (Equation (4)) and $\hat{F}^n_{\kappa_{i-1}}$ is the kernel cumulative distribution estimator computed by leaving out $x_i$:
\[
\hat{F}^n_{\kappa_{i-1}}(u) = \frac{1}{n-1} \sum_{j \neq i} \Gamma(u - x_j).
\]

Adrian et al. (1998) proposed a modified cross-validation method which minimizes the function
\[
CV(\Delta) = \frac{1}{n} \sum_{i=1}^{n} \int_{\Omega} \left(\mathbb{1}(u - x_i) - \hat{F}^n_{\kappa_{i-1}}(u)\right)^2 du.
\]

For more details about bandwidth selection of summative kernel CDF estimation, we refer our reader to Quintela Del Rio and Estevez-Perez (2013).

Now, if we suppose that all parameters have been chosen appropriately for $\hat{F}^n_{\kappa}$ to be a consistent estimate of the CDF, then a kernel estimator of VaR, denoted $\hat{\text{VaR}}_{\alpha, \kappa}$, can be easily obtained by inverting the equation $\hat{F}^n_{\kappa}(x) = \alpha$. In that case, $\hat{\text{VaR}}_{\alpha, \kappa}$ satisfies:
\[
\frac{1}{n} \sum_{i=1}^{n} \Gamma(\hat{\text{VaR}}_{\alpha, \kappa} - x_i) = \alpha.
\]

The kernel VaR estimator $\hat{\text{VaR}}_{\alpha, \kappa}$ can be seen as a smoothed version of the empirical distribution function of VaR ($\hat{\text{VaR}}_{\alpha, n}$) defined by (Equation (5)). For further details about the properties of the kernel VaR estimator (see e.g., Gourieroux et al. 2000; Chen and Tang 2005; Sheather and Marron 1990). However, the validity of such an estimate highly depends on the appropriateness of the chosen kernel and bandwidth. In the next Section, we propose another approach that allows to estimate the
dependance of the obtained estimate to the parameters of the method, and therefore to increase the robustness of the decision process based on the data.

4. Maxitive Kernels

4.1. Maxitive Kernels and Possibility Distributions

In many applications the summative kernels and their bandwidth are chosen in a very empirical way. As proposed in Loquin and Strauss (2008b), the empirical character of choosing a kernel could be taken in consideration by taking a family of kernels rather than one kernel. This family of kernels can be represented by using a maxitive kernel.

**Definition 4.** A maxitive kernel is a positive function \( \pi : \Omega \rightarrow [0, 1] \) that verifies the following maxitivity property

\[
\sup_{u \in \Omega} \pi(u) = 1.
\]

From a basic maxitive kernel \( \pi \), we can define a maxitive kernel \( \pi_{\Delta}^x \) translated in \( x \in \Omega \) and dilated with a bandwidth \( \Delta > 0 \) by

\[
\forall u \in \Omega, \quad \pi_{\Delta}^x(u) = \pi\left(\frac{u - x}{\Delta}\right).
\]

By convention, \( \pi(u) = \pi_{\Delta}^0(u) \).

For example, the triangular maxitive kernel proposed in Loquin and Strauss (2008b) illustrated in Figure 3 is defined by

\[
\pi(u) = \begin{cases} 
1 - |u| & \text{if } |u| \leq 1, \\
0 & \text{otherwise.}
\end{cases}
\]

\[ (9) \]

![Figure 3. Triangular maxitive kernel with a bandwidth \( \Delta = 1 \).](image)

A maxitive kernel can be seen as a possibility distribution Loquin and Strauss (2008b), inducing two dual non-additive confidence measures, a possibility measure \( \Pi_\pi \), and a necessity measure \( N_\pi \) Dubois and Prade (1988); and De Cooman (1997) defined, \( \forall A \subseteq \Omega \), by
\[ \Pi_\pi(A) = \sup_{x \in A} \pi(x) \quad \text{(possibility)}, \]
\[ N_\pi(A) = 1 - \Pi_\pi(A^c) \quad \text{(necessity)}, \]
with \(A^c\) being the complementary set of \(A\) in \(\Omega\).

A maxitive kernel \(\pi\) is said to dominate a summative kernel \(\kappa\) \cite{Loquin2008b}, if the possibility measure \(\Pi_\pi\) dominates the probability measure \(P_\kappa\), i.e.,
\[ \forall A \subseteq \Omega, P_\kappa(A) \leq \Pi_\pi(A). \]

In that sense, a maxitive kernel defines the convex set of summative kernels it dominates. This set is denoted \(\mathcal{M}(\pi)\) and defined by
\[ \mathcal{M}(\pi) = \left\{ \kappa \in \mathcal{K}(\Omega) / \forall A \subseteq \Omega, N_\pi(A) \leq P_\kappa(A) \leq \Pi_\pi(A) \right\}. \tag{10} \]

The specificity of a maxitive kernel \(\pi\), defined by its integral, i.e., \(\text{Sp}(\pi) = \int_{\Omega} \pi(u) du\), is a measure of the information contained by the possibility measure associated to \(\pi\) \cite{Loquin2008b}. A maxitive kernel \(\pi_1\) is at least as informative as another one \(\pi_2\) if \(\forall u \in \Omega, \pi_1(u) \leq \pi_2(u)\) \cite{Dubois2006}. In that case \(\pi_1\) is at least as specific as \(\pi_2\). It also characterizes the amount of summative kernels dominated by \(\pi\) in the sense that, if \(\forall u \in \Omega, \pi_1(u) \leq \pi_2(u)\), then \(\text{Sp}(\pi_1) \leq \text{Sp}(\pi_2)\) and \(\mathcal{M}(\pi_1) \subseteq \mathcal{M}(\pi_2)\). Moreover, if \(\exists u_0 \in \Omega\), such that \(\pi_1(u_0) < \pi_2(u_0)\), then \(\text{Sp}(\pi_1) < \text{Sp}(\pi_2)\) and \(\mathcal{M}(\pi_1) \subset \mathcal{M}(\pi_2)\).

In this context, Dubois et al. \cite{Dubois2004} proved that the triangular maxitive kernel, defined by (Equation (9)), with a support \([-\Delta, +\Delta]\) is the most specific maxitive kernel that dominates all summative symmetric and unimodal kernels whose support belongs to \([-\Delta, +\Delta]\).

### 4.2. Choquet Integrals and Maxitive Expectation

We start with the definition of a capacity or non-additive measure which generalize the notion of additive measure, i.e., probability. The notion of capacity was introduced by Gustave Choquet in 1953 and has played an important role in game theory, fuzzy set theory, Dempster-Shafer theory and many others (see Denneberg 1994; Shafer 1976; Choquet 1953).

**Definition 5.** Let \((\Omega, \mathcal{P}(\Omega))\) be the measurable space. A capacity \(v\) on \((\Omega, \mathcal{P}(\Omega))\) is a set function \(v : \mathcal{P}(\Omega) \to [0, 1]\) satisfying:

- \(v\) is normalized (i.e., \(v(\emptyset) = 0\) and \(v(\Omega) = 1\)).
- \(v\) is monotone (i.e., \(\forall A, B \in \mathcal{P}(\Omega), A \subseteq B \Rightarrow v(A) \leq v(B)\)).

One of the most important concepts closely related to additive measures is integration. It has a natural generalization to non-additive measure theory. Historically the first applied integral with respect to non-additive measures is the Choquet Integral formalized by Gustave Choquet.

**Definition 6.** Let \((\Omega, \mathcal{P}(\Omega))\) be the measurable space and \(v : \mathcal{P}(\Omega) \to [0, 1]\) a capacity. Let \(s\) be a bounded function of \(L(\Omega)\). The continuous Choquet integral of \(s\) w.r.t. \(v\) is defined by
\[ C_v(s) = \int_{\Omega} s \, dv = \int_{-\infty}^{0} \left( v(\{u \in \Omega : s(u) \geq z\}) - 1 \right) dz + \int_{0}^{\infty} v(\{u \in \Omega : s(u) \geq z\}) dz, \tag{11} \]
where the integral on the right hand side is a Riemann integral.
Definition 7. Let $\nu$ be a capacity on $(\Omega, \mathcal{P}(\Omega))$ and $x = (x_1, \ldots, x_n) \in \Omega^n$ be a discrete function $[1, \ldots, n] \to \Omega$ of $n$ samples. The discrete Choquet integral of $x$ w.r.t. $\nu$ is defined by:

$$
C_{\nu}(x) = \int_{\Omega} x d\nu = \sum_{i=1}^{n} (x_{\tau(i)} - x_{\tau(i-1)}) \nu(A_{\tau(i)}),
$$

where $\tau$ is the permutation on $[1, \ldots, n]$, such that $x_{\tau(1)} \leq \cdots \leq x_{\tau(n)}$, $A_{\tau(i)} := \{\tau(i), \ldots, \tau(n)\}$ and $x_{\tau(0)} = 0$ by convention.

Note that, a probability $P$ being a special case of (additive) capacity, the Choquet integral w.r.t. $P$ coincides with the classical expected value w.r.t. $P$.

Recall that, if $X$ is a random variable defined on a probability space $(\Omega, \mathcal{P}(\Omega), P)$, then the classical expected value of $X$ is given by

$$
\mathbb{E}_P(X) = \int_{\Omega} X dP = \int_{\Omega} X(u) dP(u),
$$

where $P: \mathcal{P}(\Omega) \to [0, 1]$ is a probability measure on the measurable space $(\Omega, \mathcal{P}(\Omega))$ and the integral is a Lebesgue-Stieltjes integral.

The notion of expectation w.r.t. a summative kernel can be extended to a maxitive kernel, as defined in Loquin and Strauss (2008b).

Definition 8. Let $\pi$ be a maxitive kernel and let $s$ be a bounded function of $L(\Omega)$. The expectation of $s$ w.r.t $\pi$ is defined by:

$$
\mathbb{E}_\pi(s) = \left[ \mathbb{E}_{\pi}(s), \mathbb{E}_{\pi}(s) \right] = \left[ C_{\Pi_\pi}(s), C_{\Pi_{\Pi_\pi}}(s) \right],
$$

where $\Pi_\pi$ (resp. $N_\pi$) is the possibility (resp. necessity) measure induced by the maxitive kernel $\pi$ and $C$ is the Choquet integral.

An important property, that will be used in the sequel, is that the maxitive interval-valued expectation of $s$ w.r.t. $\pi$ is the convex set of all the summative precise-valued expectations w.r.t. all the summative kernels dominated by $\pi$ Loquin and Strauss (2008b).

Property 1. Let $\pi$ be a maxitive kernel and let $\mathcal{M}(\pi)$ be the set of the summative kernels it dominates (see Equation (10)). Let $s$ be a bounded function of $L(\Omega)$. We have

$$
\forall g \in \mathbb{E}_{\pi}(s), \exists \kappa \in \mathcal{M}(\pi): \mathbb{E}_{\kappa}(s) = g,
$$

and

$$
\forall \kappa \in \mathcal{M}(\pi), \mathbb{E}_{\kappa}(s) \in \mathbb{E}_{\pi}(s),
$$

where $\mathbb{E}_{\kappa}(s)$ (resp. $\mathbb{E}_{\pi}(s)$) is the summative (resp. maxitive) expectation of $s$ w.r.t. $\kappa$ (resp. $\pi$).

4.3. Maxitive Kernel Cumulative and VaR Estimator

As mentioned in Loquin and Strauss (2008a) (Theorem 4), the summative kernel cumulative estimator $\hat{F}_n$, defined by (Equation (3)), in each point $x \in \Omega$, can be written as the summative expectation of the empirical CDF $F_n$, defined by (Equation (4)), in a probabilistic neighborhood defined by the summative kernel $\kappa$ translated in $x$

$$
\hat{F}_n^\kappa(x) = \mathbb{E}_{\kappa}(F_n).
$$

Based on this reformulation, Loquin and Strauss (2008a) propose an extension of the summative kernel estimator of the CDF. This extension called maxitive kernel estimator of the CDF or interval-valued estimation of the CDF.
Definition 9. Let $F_n$ be the empirical CDF based on the sample $(x_1, \ldots, x_n)$. Let $\pi$ be a maxitive kernel and let $\mathbb{E}_\pi(*)$ be the maxitive expectation w.r.t. $\pi$ (Expression (14)). The maxitive kernel cumulative estimator is defined, \(\forall x \in \Omega\), by

$$\mathbb{F}_\pi^n(x) = \mathbb{F}_\pi^n(F_n),$$

(18)

The computation of $\mathbb{F}_\pi^n$ involves two Choquet integrals. This computation Loquin and Strauss (2008a) is given, for all $x \in \Omega$, by

$$\mathbb{F}_\pi^n(x) = \mathbb{E}_\pi^n(F_n) = \mathbb{C}_{1\ldots p}(F_n) = \frac{1}{n} \sum_{i=1}^{n} \left( 1 - \pi^n(x_i) \right) \mathbb{I}(x_i \leq x),$$

(19)

$$\mathbb{F}_\pi^n(x) = \mathbb{E}_\pi^n(F_n) = \mathbb{C}_{1\ldots p}(F_n) = \frac{1}{n} \sum_{i=1}^{n} \left( \pi^n(x_i) \mathbb{I}(x \leq x_i) + \mathbb{I}(x_i \leq x) \right).$$

(20)

The estimation of the CDF is usually computed on $p$ regularly spaced points of $\Omega$. Let $\{y_j\}_{i=1,...,p}$ be those $p$ points. The algorithm of compute of $\{\mathbb{F}_\pi^n(y_j)\}_{j=1,...,p}$, in each point $\{y_j\}_{j=1,...,p}$ of $\Omega$, is given by Algorithm 1.

---

**Algorithm 1: Computation of $\{\mathbb{F}_\pi^n(y_j), \mathbb{F}_\pi^n(y_j)\}_{j=1,...,p}$**

**Data:** The considered location $\{y_j\}_{j=1,...,p}$, the observations $\{x_i\}_{i=1,...,n}$ and the maxitive kernel $\pi$.

**Result:** $\{\mathbb{F}_\pi^n(y_j), \mathbb{F}_\pi^n(y_j)\}_{j=1,...,p}$.

1. begin
2. for $j = 1$ to $p$ do
3.   $\mathbb{F}_\pi^n(y_j) \leftarrow 0$, $\mathbb{F}_\pi^n(y_j) \leftarrow 0$
4.   for $i = 1$ to $n$ do
5.     if $y_j <= x_i$ then
6.       $\mathbb{F}_\pi^n(y_j) \leftarrow \mathbb{F}_\pi^n(y_j) + 1$
7.     else
8.       $\pi^n_i = \pi^n(y_i)$
9.       $\mathbb{F}_\pi^n(y_j) \leftarrow \mathbb{F}_\pi^n(y_j) + \pi^n_i$
10. end
11. if $y_j >= x_i$ then
12.     $\mathbb{F}_\pi^n(y_j) \leftarrow \mathbb{F}_\pi^n(y_j) + 1$
13. else
14.     $\pi^n_i = \pi^n(y_i)$
15.     $\mathbb{F}_\pi^n(y_j) \leftarrow \mathbb{F}_\pi^n(y_j) + \pi^n_i$
16. end
17. end
18. $\mathbb{F}_\pi^n(y_j) = \frac{1}{n} \mathbb{F}_\pi^n(y_j)$
19. $\mathbb{F}_\pi^n(y_j) = 1 - \frac{1}{n} \mathbb{F}_\pi^n(y_j)$
20. end

The interval-values estimate $\mathbb{F}_\pi^n$ based on a maxitive kernel $\pi$ (Expression (9)) is the most specific interval containing all summative estimate $\hat{\mathbb{F}}_\kappa^n$ (Expression (6)) based on a summative kernel $\kappa$ dominated by $\pi$, i.e., such that $\kappa \in \mathcal{M}(\pi)$. 

Property 2. Let $\pi$ be a maxitive kernel on $\Omega$ and $\mathcal{M}(\pi)$ the set of all the summative kernels dominated by $\pi$. We have
\[ \forall x \in \Omega, \forall \kappa \in \mathcal{M}(\pi), \hat{F}_n^\kappa(x) \in \mathcal{F}_n^\pi(x), \]
and
\[ \forall g \in \mathcal{F}_n^\pi(x), \exists \kappa \in \mathcal{M}(\pi), \hat{F}_n^\kappa(x) = g. \]

Notice that, due to Property 2, both $\mathcal{F}_n^\pi$ and $\mathcal{F}_n^\kappa$ are also estimates of the sought after CDF. Clearly, $\mathcal{F}_n^\pi$ and $\mathcal{F}_n^\kappa$ are two strictly increasing continuous functions on $\Omega$, and so they both have an inverse. Then, we can infer immediately an maxitive interval-valued estimation of the VaR $\text{VaR}_{\alpha,\pi} = \left[ \text{VaR}_{\alpha,\pi}^{\leftarrow}, \text{VaR}_{\alpha,\pi}^{\rightarrow} \right]$: 

\[ \text{VaR}_{\alpha,\pi}^{\leftarrow} = \inf \left\{ x, F_n^\pi(x) = P(X \leq x) \geq \alpha \right\} = F_n^{-1}(\alpha), \]
\[ \text{VaR}_{\alpha,\pi}^{\rightarrow} = \inf \left\{ x, F_n^\pi(x) = P(X \leq x) \geq \alpha \right\} = F_n^{-1}(\alpha), \]

that inherits from the properties of the CDF: $\text{VaR}_{\alpha,\pi}^{\leftarrow} = \{ \text{VaR}_{\alpha,\pi}^{\leftarrow} | \kappa \in \mathcal{M}(\pi) \}$.

The MATLAB program computing the interval-valued estimation of VaR is outlined in Appendix A to the present paper.

5. Experiment: Empirical Results

5.1. Data and Experimental Process

The data used in the present study consisted of daily closing prices collected between January 2010 and December 2016 for four stock indexes from developed countries: S&P500, DJI, Nikkei225, and CAC40. The daily close prices were converted to daily log-returns. We took the first differences of the natural logarithm of the daily prices. For an observed price $P_t$, the corresponding one-day log-return on day $t$ was defined by: 

\[ r_t = \ln \left( \frac{P_t}{P_{t-1}} \right). \]

Table 1 contains descriptive statistics for the sample return of the four considered stock indexes. We observed that the returns were negatively skewed and characterized by heavy tails since the kurtoses were significantly greater than 3. The series had a distribution with tails that were significantly fatter than those of normal distribution. The Jarque-Bera test also indicated that hypothesizing the four data to be normally distributed can be rejected with a very low significance level.

Table 1. Descriptive statistics of daily returns on 4 stock indexes of a financial institution observed between 1 January 2010 and 1 September 2016.

<table>
<thead>
<tr>
<th>Index</th>
<th>S&amp;P500</th>
<th>Nikkei225</th>
<th>DJI</th>
<th>CAC40</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of observations</td>
<td>1702</td>
<td>1673</td>
<td>1702</td>
<td>1581</td>
</tr>
<tr>
<td>Min</td>
<td>-6.895</td>
<td>-11.15</td>
<td>-5.7</td>
<td>-5.634</td>
</tr>
<tr>
<td>Max</td>
<td>4.63</td>
<td>10.66</td>
<td>4.15</td>
<td>9.22</td>
</tr>
<tr>
<td>Mean</td>
<td>0.0379</td>
<td>0.027</td>
<td>0.032</td>
<td>0.0038</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>0.991</td>
<td>1.467</td>
<td>0.92</td>
<td>1.379</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.441</td>
<td>-0.468</td>
<td>-0.384</td>
<td>-0.015</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>7.08</td>
<td>9.77</td>
<td>6.416</td>
<td>5.85</td>
</tr>
<tr>
<td>Jarque-Bera</td>
<td>1239.5</td>
<td>3262.7</td>
<td>869.91</td>
<td>535.17</td>
</tr>
<tr>
<td>p-value</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Due to these properties, the normal distribution would be an inappropriate model for calculating the daily VaR—as mentioned in the introduction. We thus rather envisage a nonparametric asymmetric
approach based on kernel estimation of the CDF. However, as previously mentioned, estimating the VaR using a kernel based approach can be highly biased, since the choice of the kernel has an important impact on assessing the VaR estimate at a given probability level.

In this section, we propose to gauge this impact by estimating the VaR at several probability levels with different types of kernels, whose bandwidths have been adapted to the available datasets. We confirmed that no kernel can be considered as optimal in this context and show that the new maxitive interval-valued nonparametric approach we propose leads to a more cautious behavior when applied to real data. In fact, the bounds of the interval-valued estimate of the VaR, for a given probability level, can be instrumental in applications such as reserve requirements for banks. We then show that this approach allows discarding some parametric approaches since they are not adapted to heavy tailed data like those we consider.

5.2. Nonparametric Estimate of the VaR: Comparing Summative and Maxitive Approaches

In this experiment, we have considered estimating the VaR for low probability levels—namely at levels lower than 0.1—for each of the four datasets. In the first experiment, we have focused on DJI and Nikkei225 indexes. The CDF(s) have been estimated by considering four of the most used summative kernels whose bandwidth have been adapted using the rule of thumb method: Epanechnikov, normal, biweight, and triweight (linear)—see Silvermann (1986) for their analytical expressions. Figure 4 illustrates the impact of the kernel choice in the VaR estimation. This impact is more marked with high volatility stocks (Nikkei225). For example, referring to Figure 4a, it can be seen that the VaR estimate based on the Epanechnikov kernel is absolutely greater than the one based on the normal kernel at the probability level of 1% while the opposite situation occurs at the probability level of 2.5%. Referring to each plots of Figure 4, it is obvious that no kernel can be seen as always providing the upper or lower valued CDF. Since the CDF curves intersect each other for different probability levels then the risk of bias in the VaR estimate is relatively great, especially when the number of observations in the data set is low—compare Figure 4b–d. In this context, no kernel function can be considered as more or less conservative than the others. Also, other indexes would have produced similar results.

This first experiment shows that the high dependence of the CDF estimate to the chosen kernel shape can highly impact and bias the VaR estimate. The goal of the second experiment aims to show that none of the four considered kernel can be considered as optimal in this context. This experiment needs a ground truth which is not available. A resampling methodology was carried out to estimate such a ground truth. This methodology consists of five steps:

- **Step 1:** Fit a GARCH model to the four stock returns data. The fitted model considered for the returns $r_t$ is a AR(1)$^1$-GARCH(1,1) given by:

\[
\begin{align*}
    r_t &= \mu_t + \epsilon_t, \\
    \epsilon_t &= \sigma_t z_t, \quad \forall t = 1, \ldots, n, \\
    \sigma_t^2 &= \eta_0 + \eta_1 \epsilon_{t-1}^2 + \eta_2 \sigma_{t-1}^2,
\end{align*}
\]

where $r_t$ is the return value at time $t$, $z_t$ is a standard Normal random variable and $n$ is the sample size ($n = 200$ and 1500). The conditional mean $\mu_t$, is assumed to follow an AR(1) model given by: $\mu_t = \varphi_0 + \varphi_1 r_{t-1}$. By definition $\epsilon_t$ is serially uncorrelated with mean zero, but a time varying conditional variance equal to $\sigma_t^2$. The three positive parameters $\eta_0, \eta_1, \eta_2$ and the two parameters $\varphi_0, \varphi_1$ are, respectively, the parameters of the GARCH(1,1) and the AR(1) models. The ML method provides a systematic way to adjust the parameters of the model to give the best fit. Table 2 lists the fitted models for each daily stock returns data.

---

$^1$ Autoregressive of order 1.
Table 2. Fitted GARCH models.

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Index</th>
<th>$\mu_t \sim$ AR(1)</th>
<th>GARCH(1,1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>S&amp;P500</td>
<td>$r_1 = 0.032 - 0.132r_{t-1} + \epsilon_t$</td>
<td>$\sigma^2_t = 0.08 + 0.23c_{t-1} + 0.663\sigma^2_{t-1}$</td>
</tr>
<tr>
<td></td>
<td>Nikkei225</td>
<td>$r_1 = -0.05 - 0.085r_{t-1} + \epsilon_t$</td>
<td>$\sigma^2_t = 0.179 + 0.12c_{t-1} + 0.827\sigma^2_{t-1}$</td>
</tr>
<tr>
<td></td>
<td>DJI</td>
<td>$r_1 = 0.033 - 0.116r_{t-1} + \epsilon_t$</td>
<td>$\sigma^2_t = 0.08 + 0.25c_{t-1} + 0.63\sigma^2_{t-1}$</td>
</tr>
<tr>
<td></td>
<td>CAC40</td>
<td>$r_1 = -0.05 - 0.031r_{t-1} + \epsilon_t$</td>
<td>$\sigma^2_t = 0.032 + 0.044\sigma^2_{t-1} + 0.93\sigma^2_{t-1}$</td>
</tr>
<tr>
<td>1500</td>
<td>S&amp;P500</td>
<td>$r_1 = 0.046 - 0.063r_{t-1} + \epsilon_t$</td>
<td>$\sigma^2_t = 0.078 + 0.207c_{t-1} + 0.69\sigma^2_{t-1}$</td>
</tr>
<tr>
<td></td>
<td>Nikkei225</td>
<td>$r_1 = 0.068 - 0.0043r_{t-1} + \epsilon_t$</td>
<td>$\sigma^2_t = 0.088 + 0.17c_{t-1} + 0.814\sigma^2_{t-1}$</td>
</tr>
<tr>
<td></td>
<td>DJI</td>
<td>$r_1 = 0.039 - 0.069r_{t-1} + \epsilon_t$</td>
<td>$\sigma^2_t = 0.07 + 0.22c_{t-1} + 0.688\sigma^2_{t-1}$</td>
</tr>
<tr>
<td></td>
<td>CAC40</td>
<td>$r_1 = 0.025 - 0.027r_{t-1} + \epsilon_t$</td>
<td>$\sigma^2_t = 0.03 + 0.076\sigma^2_{t-1} + 0.906\sigma^2_{t-1}$</td>
</tr>
</tbody>
</table>

Figure 4. Cumulative distribution function (CDF) Kernel smoothed estimators for the DJI daily returns with sample size 200 (a) and 650 (b); and for the Nikkei225 daily returns with sample size 200 (c) and 650 (d). The black curve is the Epanechnikov distribution; red curve is the Triweight distribution. The blue curve corresponds to the Biweight distribution and the green curve is the Normal distribution.

- **Step 2:** Generate 1000 simulated samples from each returns data using the coefficients obtained from the above fitted models. The main reason behind proposing GARCH models to simulate
our real data is the dependence properties and the volatility phenomena of stock returns; see Engle (1982).

In a first step, we generate an i.i.d series, $z_t$, by the random generator in MATLAB (R2010a) and another series $\sigma_t$. The random numbers sampled were all assumed to be normally distributed with expectation zero and unit variance. Then, the innovations for the GARCH series have been obtained via the equation $\varepsilon_t = \sigma_t z_t$. Finally the generation of $r_t$ from the AR(1) process is straightforward. Therefore we run the series for 1500 and 200 times in each of the 1000 simulated samples.

- **Step 3:** Calculate the VaR estimates for each sample data generated in step two using the 4 kernel functions. These VaRs are calculated at five chosen levels of probability ($\alpha = 1\%, 2.5\%, 5\%, 7.5\%, 10\%$) with 2 times horizons ($n = 200$ and $n = 1500$).

- **Step 4:** Assess the performance of each kernel function by comparing their VaRs with the empirical VaRs. The performance criterion we examine is the mean square error (MSE), i.e.:

$$
MSE = \mathbb{E} \left( (\hat{Q}_n(p) - Q(p))^2 \right) \approx \frac{1}{m} \sum_{k=1}^{m} \left( \hat{Q}_n^{(k)}(p) - Q(p) \right)^2, \quad (21)
$$

where $\hat{Q}_n(p)$ is a vector of quantiles obtained from the simulated samples.

- **Step 5:** Tabulate the results that lead us to some important conclusions.

The results of the MSE of each $\alpha$ (%) VaR estimates are reported in Tables 3 and 4. From Tables 3 and 4, for S&P500, the best kernel function that estimate the VaR at level of 10%, 7.5%, 5% and 2.5% is the Epanechnikov kernel, while the Normal kernel is the best at 1% probabilities level. For Nikkei225, it seems that the Normal perform better than the other kernel functions. For DJI, the Epanechnikov kernel is more accurate to estimate the VaR, while for CAC40 the Normal and is matched more accurately. From Table 4, when $n = 200$, it appears that the Biwieight and Epanechnikov kernel are more likely to estimate VaR more accurately at several probability levels. However, the performance of kernel functions rapidly declines as $n$ and $\alpha$ get smaller.

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>Index</th>
<th>Epanechnikov</th>
<th>Normal</th>
<th>Biweight</th>
<th>Triweight</th>
<th>Best</th>
</tr>
</thead>
<tbody>
<tr>
<td>10%</td>
<td>S&amp;P500</td>
<td>0.01285</td>
<td>0.01287</td>
<td>0.01286</td>
<td>0.01286</td>
<td>Epanechnikov</td>
</tr>
<tr>
<td></td>
<td>Nikkei225</td>
<td>0.226</td>
<td>0.223</td>
<td>0.227</td>
<td>0.2265</td>
<td>Normal</td>
</tr>
<tr>
<td></td>
<td>DJI</td>
<td>0.00535</td>
<td>0.00533</td>
<td>0.00534</td>
<td>0.00535</td>
<td>Normal</td>
</tr>
<tr>
<td></td>
<td>CAC40</td>
<td>0.0256</td>
<td>0.0255</td>
<td>0.02551</td>
<td>0.02551</td>
<td>Normal</td>
</tr>
<tr>
<td>7.5%</td>
<td>S&amp;P500</td>
<td>0.03575</td>
<td>0.03580</td>
<td>0.035764</td>
<td>0.035763</td>
<td>Epanechnikov</td>
</tr>
<tr>
<td></td>
<td>Nikkei225</td>
<td>0.0316</td>
<td>0.0315</td>
<td>0.0318</td>
<td>0.0318</td>
<td>Normal</td>
</tr>
<tr>
<td></td>
<td>DJI</td>
<td>0.01075</td>
<td>0.01075</td>
<td>0.01074</td>
<td>0.01075</td>
<td>Biweight</td>
</tr>
<tr>
<td></td>
<td>CAC40</td>
<td>0.00339</td>
<td>0.034</td>
<td>0.0341</td>
<td>0.0342</td>
<td>Normal</td>
</tr>
<tr>
<td>5%</td>
<td>S&amp;P500</td>
<td>0.07288</td>
<td>0.07295</td>
<td>0.07289</td>
<td>0.07289</td>
<td>Epanechnikov</td>
</tr>
<tr>
<td></td>
<td>Nikkei225</td>
<td>0.0435</td>
<td>0.042</td>
<td>0.0435</td>
<td>0.0436</td>
<td>Normal</td>
</tr>
<tr>
<td></td>
<td>DJI</td>
<td>0.03877</td>
<td>0.03879</td>
<td>0.03878</td>
<td>0.03882</td>
<td>Epanechnikov</td>
</tr>
<tr>
<td></td>
<td>CAC40</td>
<td>0.078</td>
<td>0.0784</td>
<td>0.0782</td>
<td>0.0781</td>
<td>Normal</td>
</tr>
<tr>
<td>2.5%</td>
<td>S&amp;P500</td>
<td>0.01763</td>
<td>0.01764</td>
<td>0.017635</td>
<td>0.017637</td>
<td>Epanechnikov</td>
</tr>
<tr>
<td></td>
<td>Nikkei225</td>
<td>0.0858</td>
<td>0.0857</td>
<td>0.0859</td>
<td>0.08578</td>
<td>Normal</td>
</tr>
<tr>
<td></td>
<td>DJI</td>
<td>0.141</td>
<td>0.1409</td>
<td>0.1408</td>
<td>0.1409</td>
<td>Epanechnikov</td>
</tr>
<tr>
<td></td>
<td>CAC40</td>
<td>0.132</td>
<td>0.131</td>
<td>0.133</td>
<td>0.134</td>
<td>Normal</td>
</tr>
<tr>
<td>1%</td>
<td>S&amp;P500</td>
<td>0.05804</td>
<td>0.05801</td>
<td>0.05804</td>
<td>0.058017</td>
<td>Normal</td>
</tr>
<tr>
<td></td>
<td>Nikkei225</td>
<td>0.246</td>
<td>0.2467</td>
<td>0.2481</td>
<td>0.2482</td>
<td>Epanechnikov</td>
</tr>
<tr>
<td></td>
<td>DJI</td>
<td>0.2404</td>
<td>0.2405</td>
<td>0.2406</td>
<td>0.2407</td>
<td>Epanechnikov</td>
</tr>
<tr>
<td></td>
<td>CAC40</td>
<td>0.363</td>
<td>0.363</td>
<td>0.364</td>
<td>0.362</td>
<td>Triweight</td>
</tr>
</tbody>
</table>
Table 4. Mean square error results for several types of kernel functions and $n = 200$.  

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>Index</th>
<th>Epanechnikov</th>
<th>Normal</th>
<th>Biweight</th>
<th>Triweight</th>
<th>Best</th>
</tr>
</thead>
<tbody>
<tr>
<td>10%</td>
<td>S&amp;P500</td>
<td>0.0337</td>
<td>0.0335</td>
<td>0.0336</td>
<td>0.0338</td>
<td>Normal</td>
</tr>
<tr>
<td></td>
<td>Nikkei225</td>
<td>0.766</td>
<td>0.765</td>
<td>0.767</td>
<td>0.766</td>
<td>Normal</td>
</tr>
<tr>
<td></td>
<td>DJI</td>
<td>0.0441</td>
<td>0.0443</td>
<td>0.0442</td>
<td>0.0442</td>
<td>Epanechnikov</td>
</tr>
<tr>
<td></td>
<td>CAC40</td>
<td>0.0347</td>
<td>0.0348</td>
<td>0.0346</td>
<td>0.0348</td>
<td>Biweight</td>
</tr>
<tr>
<td>7.5%</td>
<td>S&amp;P500</td>
<td>0.0632</td>
<td>0.0634</td>
<td>0.0633</td>
<td>0.064</td>
<td>Epanechnikov</td>
</tr>
<tr>
<td></td>
<td>Nikkei225</td>
<td>0.938</td>
<td>0.937</td>
<td>0.939</td>
<td>0.936</td>
<td>Triweight</td>
</tr>
<tr>
<td></td>
<td>DJI</td>
<td>0.0614</td>
<td>0.0615</td>
<td>0.0612</td>
<td>0.0613</td>
<td>Biweight</td>
</tr>
<tr>
<td></td>
<td>CAC40</td>
<td>0.0588</td>
<td>0.0591</td>
<td>0.0592</td>
<td>0.0594</td>
<td>Epanechnikov</td>
</tr>
<tr>
<td>5%</td>
<td>S&amp;P500</td>
<td>0.1169</td>
<td>0.1173</td>
<td>0.1167</td>
<td>0.1183</td>
<td>Biweight</td>
</tr>
<tr>
<td></td>
<td>Nikkei225</td>
<td>1.131</td>
<td>1.129</td>
<td>1.131</td>
<td>1.127</td>
<td>Triweight</td>
</tr>
<tr>
<td></td>
<td>DJI</td>
<td>0.0975</td>
<td>0.0977</td>
<td>0.0974</td>
<td>0.0977</td>
<td>Biweight</td>
</tr>
<tr>
<td></td>
<td>CAC40</td>
<td>0.148</td>
<td>0.149</td>
<td>0.147</td>
<td>0.0148</td>
<td>Epanechnikov</td>
</tr>
<tr>
<td>2.5%</td>
<td>S&amp;P500</td>
<td>0.28</td>
<td>0.284</td>
<td>0.281</td>
<td>0.283</td>
<td>Biweight</td>
</tr>
<tr>
<td></td>
<td>Nikkei225</td>
<td>1.453</td>
<td>1.451</td>
<td>1.453</td>
<td>1.448</td>
<td>Triweight</td>
</tr>
<tr>
<td></td>
<td>DJI</td>
<td>0.252</td>
<td>0.255</td>
<td>0.253</td>
<td>0.254</td>
<td>Epanechnikov</td>
</tr>
<tr>
<td></td>
<td>CAC40</td>
<td>0.2905</td>
<td>0.2888</td>
<td>0.287</td>
<td>0.2889</td>
<td>Biweight</td>
</tr>
<tr>
<td>1%</td>
<td>S&amp;P500</td>
<td>0.869</td>
<td>0.821</td>
<td>0.87</td>
<td>0.881</td>
<td>Epanechnikov</td>
</tr>
<tr>
<td></td>
<td>Nikkei225</td>
<td>1.977</td>
<td>1.995</td>
<td>1.981</td>
<td>1.980</td>
<td>Epanechnikov</td>
</tr>
<tr>
<td></td>
<td>DJI</td>
<td>0.702</td>
<td>0.691</td>
<td>0.696</td>
<td>0.693</td>
<td>Normal</td>
</tr>
<tr>
<td></td>
<td>CAC40</td>
<td>0.794</td>
<td>0.791</td>
<td>0.788</td>
<td>0.793</td>
<td>Biweight</td>
</tr>
</tbody>
</table>

The simulations results showed that, for large sample sizes (here $n = 1500$), all kernel functions are consistent estimators, i.e., MSE values are close to 0. On the other hand, MSE values are larger for the smaller sample size (here $n = 200$). This demonstrates that choosing a particular kernel, when the sample size is low, is risky. Therefore a coherent interval-valued of VaR as we propose is likely to provide a more careful decision in this context.

5.3. Interval Estimation of Value-at-Risk and Some Numerical Comparisons

Here we apply the maxitive kernel estimator presented in Section 4.3 to obtain the lower and upper bound for VaR corresponding to each four stock indexes for three probability levels ($\alpha = 10\%, 5\%, 1\%$). To obtain the optimal bounds for the VaR, the bandwidth has been, once again, chosen using the most popular methods such as biased cross-validation method and plug-in method presented in Section 3.3. As shown in Table 5, the maximum available bandwidth is taken to insure that all kernel estimators are inside the interval given by the maxitive kernel estimator. In order to examine the performance of the maxitive kernel estimation method, we divide the data into three samples: the first sample corresponds to data with 6 year time horizons, the second sample is chosen with time horizons of 3 years and the third correspond to one year time horizon. In a first step, after choosing the best-fit distribution from the family of GHYP distribution by using the AIC² criteria, we compute the VaR across the candidate distribution. Next, we compute the VaR confidence interval based on these three distributions using the bias-corrected and accelerated (BCa) bootstrap method. Finally, we compare our proposed maxitive interval-valued with those based on the bootstrap technique under different samples sizes. So far, the bootstrap confidence interval of VaR for the three distributions: GHYP, Normal, and HS are presented together in Tables 6–8. With these punctual VaRs at hand for comparison purposes, we evaluate the performance of our approach. In this section, we have chosen to illustrate our results and to discuss the benefit of our approach in two ways. In Tables 6–8, we show the

---

The Akaike Information Criterial. $\text{AIC} = -2\ln \left( L(GHYP) \right) + 2\varsigma$, where $\varsigma$ is the number of estimated parameters and $L(GHYP)$ is the likelihood of the GHYP model.
explicit results for the minimum and maximum value, as well as the width of the interval estimation of VaR for several probability levels between 1% and 10%. Figure 5 shows the PP-plots, i.e., the \( \hat{F}(x_i) \) of all models against the \( F(x_i) \). Based on the numerical results we can formulate several conclusions: Evidently, the lower and upper bounds increase with the probability level \( \alpha \). However, it is important to note that the intervals estimation of VaRs for the two indexes Nikkei225 and CAC40 are larger than those of DJI and S&P500 indexes. Note also that the influence of the volatility stock market is much more important than the influence of the sample size. Also, the widths of maxitive VaR intervals are rather tight than those derived from the GHYP, HS, and normal distributions. For example, for the short time horizon 1 year with the high volatility stock (Nikkei225) and at the 1% of probability level the width of the interval-valued of VaR is 1.073 while the widths of the BCa (99%) confidence interval based on HS, normal, and GHYP distributions are respectively 4.236, 1.925, and 4.246 respectively (See Table 6). Furthermore, we can remark that the maxitive interval-valued estimation of the VaR is on the left side of the normal VaR and this shows the ability of our approach to model very dangerous financial risks while the normal distribution is not consistent with tail-thickness and right tail risk. This results indicate that our approach is more accurate and informative especially for the smaller sample size.

<table>
<thead>
<tr>
<th>Return Data</th>
<th>Time Horizon</th>
<th>Bandwidth Selection Method</th>
<th>Epanechnikov</th>
<th>Normal</th>
<th>Biweight</th>
<th>Triweight</th>
<th>Maxitive Kernel Bandwidth</th>
</tr>
</thead>
<tbody>
<tr>
<td>S&amp;P500</td>
<td>6 years</td>
<td>Plug-in, Cross validation</td>
<td>0.189</td>
<td>0.084</td>
<td>0.245</td>
<td>0.250</td>
<td>0.250</td>
</tr>
<tr>
<td></td>
<td>3 years</td>
<td>Plug-in, Cross validation</td>
<td>0.228</td>
<td>0.101</td>
<td>0.269</td>
<td>0.294</td>
<td>0.294</td>
</tr>
<tr>
<td></td>
<td>1 year</td>
<td>Plug-in, Cross validation</td>
<td>0.218</td>
<td>0.118</td>
<td>0.211</td>
<td>0.267</td>
<td>0.267</td>
</tr>
<tr>
<td>Nikkei225</td>
<td>6 years</td>
<td>Plug-in, Cross validation</td>
<td>0.332</td>
<td>0.147</td>
<td>0.405</td>
<td>0.440</td>
<td>0.440</td>
</tr>
<tr>
<td></td>
<td>3 years</td>
<td>Plug-in, Cross validation</td>
<td>0.330</td>
<td>0.110</td>
<td>0.331</td>
<td>0.412</td>
<td>0.412</td>
</tr>
<tr>
<td></td>
<td>1 year</td>
<td>Plug-in, Cross validation</td>
<td>0.406</td>
<td>0.180</td>
<td>0.480</td>
<td>0.540</td>
<td>0.540</td>
</tr>
<tr>
<td>DJI</td>
<td>6 years</td>
<td>Plug-in, Cross validation</td>
<td>0.173</td>
<td>0.077</td>
<td>0.205</td>
<td>0.224</td>
<td>0.224</td>
</tr>
<tr>
<td></td>
<td>3 years</td>
<td>Plug-in, Cross validation</td>
<td>0.228</td>
<td>0.101</td>
<td>0.270</td>
<td>0.294</td>
<td>0.294</td>
</tr>
<tr>
<td></td>
<td>1 year</td>
<td>Plug-in, Cross validation</td>
<td>0.360</td>
<td>0.160</td>
<td>0.425</td>
<td>0.468</td>
<td>0.468</td>
</tr>
<tr>
<td>CAC40</td>
<td>6 years</td>
<td>Plug-in, Cross validation</td>
<td>0.323</td>
<td>0.143</td>
<td>0.382</td>
<td>0.431</td>
<td>0.446</td>
</tr>
<tr>
<td></td>
<td>3 years</td>
<td>Plug-in, Cross validation</td>
<td>0.423</td>
<td>0.188</td>
<td>0.499</td>
<td>0.564</td>
<td>0.564</td>
</tr>
<tr>
<td></td>
<td>1 year</td>
<td>Plug-in, Cross validation</td>
<td>0.930</td>
<td>0.416</td>
<td>1.107</td>
<td>1.250</td>
<td>1.250</td>
</tr>
</tbody>
</table>
Table 6. A comparison between the bound estimates of the daily VaR based on the maxitive kernel estimation method and bootstrap confidence intervals of Value-at-Risk (VaRs) based on three different distributions: Historical simulation method (HS), Normal, and Generalized HYperbolic (GHYP) distributions. These VaR methods are applied to four daily stock returns at several probability levels α and time horizon of 1 year.

<table>
<thead>
<tr>
<th>Returns</th>
<th>Interval-Value of VaR</th>
<th>pdf</th>
<th>Estimated VaR</th>
<th>Bootstrap Confidence Intervals of VaRs</th>
</tr>
</thead>
<tbody>
<tr>
<td>10%</td>
<td>-1.092</td>
<td>-0.751</td>
<td>0.341</td>
<td></td>
</tr>
<tr>
<td>S&amp;P500</td>
<td>-1.605</td>
<td>-1.152</td>
<td>0.453</td>
<td>HS</td>
</tr>
<tr>
<td>5%</td>
<td>-3.240</td>
<td>-2.117</td>
<td>1.123</td>
<td>Normal</td>
</tr>
<tr>
<td>Nikkei225</td>
<td>5%</td>
<td>-6.760</td>
<td>-5.010</td>
<td>1.750</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>-1.099</td>
<td>-0.766</td>
<td>0.333</td>
</tr>
<tr>
<td>DJI</td>
<td>5%</td>
<td>-1.540</td>
<td>-1.094</td>
<td>0.446</td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>-2.949</td>
<td>-1.837</td>
<td>1.112</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>-3.555</td>
<td>-2.154</td>
<td>1.401</td>
</tr>
<tr>
<td>CAC40</td>
<td>5%</td>
<td>-4.202</td>
<td>-3.184</td>
<td>1.018</td>
</tr>
</tbody>
</table>

Table 7. A comparison between the bound estimates of the daily VaR based on the maxitive kernel estimation method and bootstrap confidence intervals of VaRs based on three different distributions: Historical simulation method (HS), Normal, and GHYP distributions. These VaR methods are applied to four daily stock returns at several probability levels α and time horizon of 3 years.

<table>
<thead>
<tr>
<th>Returns</th>
<th>α</th>
<th>Interval-Value of VaR</th>
<th>pdf</th>
<th>Estimated VaR</th>
<th>Bootstrap Confidence Intervals of VaRs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nikkei225</td>
<td>5%</td>
<td>-1.394</td>
<td>-1.130</td>
<td>0.264</td>
<td>HS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Normal</td>
<td></td>
<td></td>
<td>GHYP</td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>-2.230</td>
<td>-1.979</td>
<td>0.251</td>
<td>HS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Normal</td>
<td></td>
<td></td>
<td>GHYP</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>-3.024</td>
<td>-2.283</td>
<td>0.741</td>
<td>HS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Normal</td>
<td></td>
<td></td>
<td>GHYP</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>-4.324</td>
<td>-3.616</td>
<td>0.708</td>
<td>HS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Normal</td>
<td></td>
<td></td>
<td>GHYP</td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>-6.602</td>
<td>-6.052</td>
<td>0.550</td>
<td>HS</td>
</tr>
<tr>
<td>CAC40</td>
<td>5%</td>
<td>-3.095</td>
<td>-2.481</td>
<td>0.614</td>
<td>HS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Normal</td>
<td></td>
<td></td>
<td>GHYP</td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>-4.855</td>
<td>-4.137</td>
<td>0.718</td>
<td>HS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Normal</td>
<td></td>
<td></td>
<td>GHYP</td>
</tr>
<tr>
<td>Returns</td>
<td>α</td>
<td>Interval Value of VaR</td>
<td>pdf</td>
<td>Estimated VaR</td>
<td>Bootstrap Confidence Intervals of VaRs</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BCA (90%)</td>
<td>BCA (95%)</td>
</tr>
<tr>
<td>10%</td>
<td>-1.189</td>
<td>-0.941</td>
<td>0.248</td>
<td>HS</td>
<td>-1.070</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Normal</td>
<td>-1.233</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GHYP</td>
<td>-1.029</td>
</tr>
<tr>
<td>S&amp;P500</td>
<td>5%</td>
<td>-1.679</td>
<td>-1.443</td>
<td>0.236</td>
<td>HS</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Normal</td>
<td>-1.593</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GHYP</td>
<td>-1.548</td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>-2.939</td>
<td>-2.772</td>
<td>0.167</td>
<td>HS</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Normal</td>
<td>-2.269</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GHYP</td>
<td>-2.810</td>
</tr>
<tr>
<td>Nikkei225</td>
<td>5%</td>
<td>-3.698</td>
<td>-2.991</td>
<td>0.707</td>
<td>HS</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Normal</td>
<td>-3.286</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GHYP</td>
<td>-3.307</td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>-5.964</td>
<td>-5.284</td>
<td>0.680</td>
<td>HS</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Normal</td>
<td>-5.386</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GHYP</td>
<td>-5.959</td>
</tr>
<tr>
<td>DJI</td>
<td>5%</td>
<td>-1.459</td>
<td>-1.252</td>
<td>0.207</td>
<td>HS</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Normal</td>
<td>-1.482</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GHYP</td>
<td>-1.336</td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>-2.436</td>
<td>-2.192</td>
<td>0.244</td>
<td>HS</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Normal</td>
<td>-2.362</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>-2.501</td>
<td>-1.888</td>
<td>0.613</td>
<td>HS</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Normal</td>
<td>-1.764</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GHYP</td>
<td>-2.218</td>
</tr>
<tr>
<td>CAC40</td>
<td>5%</td>
<td>-3.577</td>
<td>-2.976</td>
<td>0.601</td>
<td>HS</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Normal</td>
<td>-2.265</td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>-5.477</td>
<td>-5.023</td>
<td>0.454</td>
<td>HS</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GHYP</td>
<td>-5.281</td>
</tr>
</tbody>
</table>
Next, in order to inspect the goodness-of-fit of the used models a graphical tool (PP-plot) is constructed (Figure 5) to compare the empirical cumulative function to the fitted cumulative functions. This plot confirms that the Epanechnikov kernel function and the GHYP distribution give a good global fit for the 4 returns data. The points of the PP-plot are close to the 45 degree line and they also lie within the interval estimation using maxitive kernel.

In contrast, from the same figure, the PP-plot of the normal cumulative function against the empirical cumulative function shows that the left end pattern is above the 45 degree line and the right end is below it. Thus, the normal distribution underestimates the VaR at low probability level. This is due to the fact that the normal distribution ignores the presence of fat tails in the actual distribution. Based on these results, we can conclude that the GHYP and the maxitive kernel method provide a better fit than a normal distribution to market return data. Thus, the maxitive kernel method seems to be a good choice to estimate the risk for VaR.

Figure 5. A graphical tool (PP-plot) of the theoretical CDF of Epanechnikov (in black), Generalized HYPerbolic (GHYP, in green), and normal (in violet) versus the empirical CDF for the four returns data sets: (a) Daily CAC40 (b) Daily DJI (c) Daily Nikkei225 and (d) Daily S&P500. In each of the four plots, the empirical CDF on the horizontal axis and the theoretical CDF on the vertical axis. The highest (in red) and lowest (in blue) dotted lines correspond respectively to the upper and lower bounds of the interval CDF using maxitive kernel.
6. Conclusions

Using an estimate of the Value-at-Risk (VaR) based on a small-sized sample may pose a risk to financial application due to the high dependence of this VaR estimate to the computation method. In fact, computing a VaR estimate can be performed in many ways including parametric and nonparametric approaches. We have shown, with experiments based on daily closing prices data of four stock indexes, that no method can be said to be optimal to achieve this estimate. Moreover, the bias induced by choosing a particular method is particularly sensitive for estimates based on small samples. In this paper, we have presented a new method for computing the VaR which highly differs from its competitors in the sense that it is interval-valued. This interval-valued VaR estimate is the set of all estimates that could have been obtained, with the same data sample, by using a set of kernel-based estimation methods. In our experiments, we noted that the output of parametric methods, like the GHYP VaR estimate, always belong to the interval-valued VaR we propose while others, like the normal VaR estimate, do not. It appears that VaR estimates belonging to the interval-valued VaR estimate we propose are likely to be less risky than those that do not. Moreover, a wide interval-valued VaR estimate is a marker of a high risk for a trader since it reflects thick tails, pronounced skewness, and excess kurtosis of financial asset price returns.

The interval-valued estimation, based on maxitive kernel, that we propose in this paper is a convex envelope of the kernel VaR estimation. Although the VaR is probably one of the most popular tool in risk management, an alternative measure for Value-at-Risk which satisfied the conditions for a coherent risk measure has been proposed (see e.g., Rockafellar and Uryasev 2000; Artzner et al. 1999). This risk measure is called Expected Shortfall (ES; also known as conditional VaR or average VaR). Indeed, the Basel Committee published, in January 2016 Basel Committee on Banking Supervision (2016), revised standards for minimum capital requirements for market risk which include a shift from Value-at-Risk to expected shortfall as the preferred risk measure. Then, the future research should be conducted into finding an \textit{interval-valued estimation of ES}, based on maxitive kernel, and to compare it with the \textit{interval-valued estimation of VaR} and the ES for many distributions. In this regard, Broda and Paolella (2011) presents easily-computed expressions for evaluating the expected shortfall for numerous distributions which are now commonly used for modelling asset returns. Another interesting avenue of research would be to construct a CoVaR \textit{interval} combining GARCH model with maxitive kernel.

Author Contributions: All authors contributed equally to the paper.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Matlab Code for Maxitive Interval-Valued Estimation of VaR

```matlab
% Data definition
Data = load(‘S&P500_daily.txt’);
NbData = length(Data);

% Definition of visualization samples
SampleSize = length(Data);
kernelBandwidth = 0.25;
MIN = min(Data);
MAX = max(Data);
delta = (MAX - MIN) / SampleSize;
MIN = MIN - 50*delta;
MAX = MAX + 50*delta;
delta = (MAX - MIN)/ SampleSize;
Absc = sort(Data);
SampleSize = length(Absc);
bandwidth = kernelBandwidth;

% Summative kernel estimation of CDF
```
% Epanechnikov kernel
for n=1:SampleSize
    FonctionEscaler1 = KernelIntegralEpanechnikov(Absc(n)–Data,0,0.305);
    CDFEpanechnikov(n) = sum(FonctionEscaler1) / NbData;
end

% Cosine kernel
for n=1:SampleSize
    FonctionEscaler6 = KernelIntegralCosine(Absc(n)–Data,0,0.7);
    CDFBCosine(n) = sum(FonctionEscaler6) / NbData;
end

% Biweight kernel
for n=1:SampleSize
    FonctionEscaler4 = KernelIntegralBiweight(Absc(n)–Data,0,0.05);
    CDFBiweight(n) = sum(FonctionEscaler4) / NbData;
end

% Triweight kernel
for n=1:SampleSize
    FonctionEscaler5 = KernelIntegralTriweight(Absc(n)–Data,0,0.06);
    CDFTriweight(n) = sum(FonctionEscaler5) / NbData;
end

% Interval–valued estimation of CDF

cumul_upp_plus = zeros(SampleSize,1);
cumul_low_plus = zeros(SampleSize,1);
cumul_upp_moins = zeros(SampleSize,1);
cumul_low_moins = zeros(SampleSize,1);

for n=1:SampleSize
    for k=1:NbData
        if Data(k) <= (Absc(n))
            cumul_upp_plus(n) = cumul_upp_plus(n) + 1;
        else
            Possibility = Max_kernel_Triangular(Data(k), Absc(n), bandwidth);
            cumul_upp_plus(n) = cumul_upp_plus(n) + Possibility;
        end
        if Data(k) >= (Absc(n))
            cumul_low_plus(n) = cumul_low_plus(n) + 1;
        else
            Possibility = Max_kernel_Triangular(Data(k), Absc(n), bandwidth);
            cumul_low_plus(n) = cumul_low_plus(n) + Possibility;
        end
    end
    cumul_low_plus = NbData–cumul_low_plus;
cumul_low_plus = cumul_low_plus/NbData;
cumul_upp_plus = cumul_upp_plus/NbData;
% Interval-valued estimation of VaR

\[
\text{w_low}=\text{cumul_low\_plus}; \\
\text{w_upp}=\text{cumul_upp\_plus}; \\
\text{dataCDF}=\left[\text{sort}\left(\text{Data}\right),\text{w_low},\text{w_upp}\right]; \\
\alpha=0.01; \\
\left[\text{NotNeeded1},\text{idx1}\right]=\min\left(\left|\text{dataCDF}\left(:,2\right)-\alpha\right|\right); \\
\text{dataCDF(idx1)}; \\
\left[\text{NotNeeded2},\text{idx2}\right]=\min\left(\left|\text{dataCDF}\left(:,3\right)-\alpha\right|\right); \\
\text{dataCDF(idx2)}; \\
\text{StandardMaxitiveVaR}=\left[\text{dataCDF(idx1)},\text{dataCDF(idx2)}\right]; \\
\text{VaRInterval}=\left[\text{mean}\left(\text{Absc}\right)+\text{std}\left(\text{Absc}\right)\times\text{dataCDF(idx1)},\text{mean}\left(\text{Absc}\right)+\text{std}\left(\text{Absc}\right)\times\text{dataCDF(idx2)}\right];
\]

Integrated Summative Kernels

% Integrated Epanechnikov kernel

\[
\text{function } [y] = \text{KernelIntegralEpanechnikov}(x, \text{centre}, \text{delta}) \\
x = (x-\text{centre})/\text{delta}; \\
y = \text{zeros(size}(x)); \\
xxx = x\times x\times x; \\
y = ( (x > -1) \times (x <= 0) \times (-0.25\times xxx + 0.75\times x + 0.5) ) ; \\
y = y + ( (x < 1) \times (x >= 0) \times (-0.25\times xxx + 0.75\times x + 0.5) ) + (x>=1); \\
\]

% Integrated Cosinus kernel

\[
\text{function } [y] = \text{KernelIntegralCosine}(x, \text{centre}, \text{delta}) \\
x = (x-\text{centre})/\text{delta}; \\
y = \text{zeros(size}(x)); \\
y = ( (x > -1) \times (x <= 0) \times (0.5\times \sin(0.5\times \pi\times x)+ 0.5) ) ; \\
y = y + ( (x < 1) \times (x >= 0) \times (0.5\times \sin(0.5\times \pi\times x)+ 0.5) ) + (x>=1)); \\
\]

% Integrated Triweight kernel

\[
\text{function } [y] = \text{KernelIntegralTriweight}(x, \text{centre}, \text{delta}) \\
x = (x-\text{centre})/\text{delta}; \\
y = \text{zeros(size}(x)); \\
x3 = x\times x\times x; \\
x5 = x\times x\times x\times x\times x; \\
x7 = x\times x\times x\times x\times x\times x\times x; \\
y = ((x > -1)\times (x <= 0) \times ((35/32)\times x -(35/32)\times x3 +(21/32)\times x5 -(5/32)\times x7 + 0.5) ) ; \\
y = y + ((x < 1)\times (x >= 0) \times ((35/32)\times x -(35/32)\times x3 +(21/32)\times x5 -(5/32)\times x7 + 0.5) ) + (x>=1)); \\
\]

% Integrated Biweight kernel

\[
\text{function } [y] = \text{KernelIntegralBiweight}(x, \text{centre}, \text{delta}) \\
x = (x-\text{centre})/\text{delta}; \\
\]
$y = \text{zeros}\left(\text{size}\left(x\right)\right);$

$xxx = x.*x.*x;$
$xxxxx = x.*x.*x.*x.*x;$

$y = \left( \left( x > -1 \right) .* \left( x <= 0 \right) .* \left( (15/16)*x - (5/8)*xxx + (3/16)*xxxxx + 0.5 \right) \right);$  
$y = y + \left( \left( x < 1 \right) .* \left( x >= 0 \right) .* \left( (15/16)*x - (5/8)*xxx + (3/16)*xxxxx + 0.5 \right) + \left(x>=1\right) \right);$  

Triangular Maxitive Kernel

% Triangulaire maxitive kernel

function [y] = Max_kernel_Triangular(x, centre, delta)
    x = (x − centre)/delta;
    y = zeros(size(x));
    x = abs(x);
    Mask = x <= 1;
    y = (1.0 − x) .* Mask;
    return;

References


Dubois, Didier. 2006. Possibility theory and statistical reasoning. *Computational Statistics and Data Analysis* 51: 47–69. [CrossRef]


© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).