

Efficient Density Estimation and Value at Risk Using Fejér-Type Kernel Functions

Olga Kosta¹, Natalia Stepanova²

¹Decision Economics Group, HDR, Inc., Ottawa, Canada

²School of Mathematics and Statistics, Carleton University, Ottawa, Canada

Email: Olga.Kosta@hdrinc.com, nstep@math.carleton.ca

Received 29 September 2015; accepted 27 November 2015; published 30 November 2015

Copyright © 2015 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

This paper presents a nonparametric method for computing the Value at Risk (VaR) based on efficient density estimators with Fejér-type kernel functions and empirical bandwidths obtained from Fourier analysis techniques. The kernel-type estimator with a Fejér-type kernel was recently found to dominate all other known density estimators under the \mathbb{L}_p -risk, $1 \leq p < \infty$. This theoretical finding is supported via simulations by comparing the quality of the density estimator in question with other fixed kernel estimators using the common \mathbb{L}_2 -risk. Two data-driven bandwidth selection methods, cross-validation and the one based on the Fourier analysis of a kernel density estimator, are used and compared to the theoretical bandwidth. The proposed nonparametric method for computing the VaR is applied to two fictitious portfolios. The performance of the new VaR computation method is compared to the commonly used Gaussian and historical simulation methods using a standard back-test procedure. The obtained results show that the proposed VaR model provides more reliable estimates than the standard VaR models.

Keywords

Value at Risk, Kernel-Type Density Estimator, Fejér-Type Kernel, Asymptotic Minimavity, Mean Integrated Squared Error, Fourier Analysis

1. Introduction

Financial institutions monitor their portfolios of assets using the Value at Risk (VaR) to mitigate their market risk exposure. The VaR was made popular in the early nineties by U.S. investment bank, J.P. Morgan, in response to the infamous financial disasters at the time and has since been implemented in the financial sector worldwide by the Basel Committee on Banking Supervision. By definition, the VaR is a risk measure of the

worst expected loss of a portfolio over a defined holding period at a given probability. The time horizon and the loss probability parameters are specified by the financial managers depending on the purpose at hand. Typically, the VaR is computed at short time horizons of one hour, two hours, one day, or a few days, while the loss probability can range from 0.001 to 0.1 depending on the risk averseness of the investors. Financial institutions then use the results of the VaR to determine the necessary capital and cash reserves to put aside for coverage against potential losses in the event of severe or prolonged adverse market movements.

Formally, the Value at Risk of a portfolio, $\text{VaR}_t(p)$, is the p -th quantile of the distribution of portfolio returns over a given time horizon h that satisfies the following expression:

$$F(\text{VaR}_t(p); h) = P(X_{t+h} \leq \text{VaR}_t(p)) = \int_{-\infty}^{\text{VaR}_t(p)} f(x; h) dx = p,$$

where X_{t+h} is the portfolio return between time t and $t+h$ and $f(x; h)$ is the probability density function (pdf) of returns. Equivalently,

$$\text{VaR}_t(p) = F^{-1}(p; h),$$

where $F^{-1}(\cdot; h)$ is the inverse of the distribution function $F(\cdot; h)$ that is continuous from the right. The time horizon h and loss probability $p \in (0, 1)$ are specified parameters. In our analysis, we use a time horizon of one day and probability levels ranging from 0.005 to 0.05. For a more in depth discussion on the origins of the VaR and its many uses see [1].

In practice, there exists a variety of computational methods for the VaR. The two most commonly used approaches are the parametric normal and the nonparametric historical simulation summarized below. The following models rely on the assumption of independent and identically distributed (iid) daily portfolio returns.

1. Normal method. For normally distributed returns, with μ as the expected return on a portfolio and σ^2 as the variance of portfolio returns, the VaR is the p -th quantile of the normal distribution function given by

$$\text{VaR}_t(p) = F^{-1}(p) = \sigma \Phi^{-1}(p) + \mu,$$

where $\Phi^{-1}(p)$ is the quantile function of the standard normal distribution, for any $0 < p < 1$. For estimators $\bar{X}_n = \sum_{j=1}^n X_j / n$ and $s_n = \sqrt{\sum_{j=1}^n (X_j - \bar{X}_n)^2 / (n-1)}$ of μ and σ based on $X_1, \dots, X_n \sim N(\mu, \sigma^2)$, a natural VaR estimator is

$$\widetilde{\text{VaR}}_{1,n}(p) = s_n \Phi^{-1}(p) + \bar{X}_n. \quad (1)$$

The normal method for estimating the VaR is widely used among financial institutions due to its familiar properties. It is not realistic, however, to assume that the portfolio returns are normally distributed since high frequency financial data have heavier tails than can be explained by the normal distribution. As a result, this method generally underestimates the true VaR.

2. Historical simulation. Let $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ denote the corresponding order statistics of the sample X_1, \dots, X_n of portfolio returns. For a given probability level $p \in (0, 1)$, the VaR estimator is the p -th sample quantile of portfolio returns:

$$\widetilde{\text{VaR}}_{2,n}(p) = X_{(\lfloor np \rfloor + 1)}, \quad (2)$$

where $\lfloor x \rfloor$ denotes the greatest integer strictly less than the real number x .

The main strengths of the historical simulation method are its simplicity and that it does not require any distributional assumptions on the portfolio returns as the VaR is determined by the actual price level movements. One has to be careful when selecting the data so as not to remove relevant or include irrelevant data. For instance, large samples of historical financial data can be disadvantageous. The portfolio composition is based on current circumstances; therefore, it may not be meaningful to evaluate the portfolio using data from the distant past since the distribution of past returns is not always a good approximation of expected future returns. Also, if new market risks are added, then there is not enough historical data to compute the VaR, which may underestimate it. Another drawback is that the discrete approximation of the true distribution at the extreme tails can cause biased results.

A more generalized and sophisticated nonparametric method for estimating the pdf of daily portfolio returns is kernel density estimation. Let X_1, X_2, \dots be a sequence of iid real-valued random variables from an absolutely continuous distribution with an unknown density f on \mathbb{R} , where f belongs to a suitable family \mathcal{F}

of densities. Density estimation then consists of constructing an estimator $f_n(x; X_1, \dots, X_n)$ of the true function $f(x)$ that would produce a good estimate $f_n(x; x_1, \dots, x_n)$, based on some performance criterion, of the underlying density $f(x)$ for the data x_1, \dots, x_n . The kernel density estimator contains a kernel function K and a smoothing parameter h . In most studies and applications, K is a fixed function and $h = h_n$ is a sample size dependent parameter. If K depends on n , then the corresponding estimator is called the *kernel-type density estimator*. In [2], a new kernel-type estimator of densities belonging to a class of infinitely smooth functions is shown to dominate in $\mathbb{L}_p(\mathbb{R})$, $1 \leq p < \infty$, all other estimators in the literature, in a strong locally asymptotically minimax sense. Moreover, it does the best under the \mathbb{L}_2 -risk. The estimator in [2] uses the Fejér-type kernel function and the common theoretical bandwidth, which is used by many authors in the case of estimating infinitely smooth density functions.

In this paper, we introduce a nonparametric approach for computing the VaR based on quantile estimation with the Fejér-type kernel and a nearly optimal bandwidth obtained from the Fourier analysis techniques. To do so, we first conduct a simulation study to support the theoretical finding that the kernel-type density estimator in hand has the best performance with respect to the \mathbb{L}_2 -risk. We then compare the new estimation technique for computing the VaR to the common Gaussian and historical simulation methods. Portfolio compositions can be rather complex therefore, for the purpose of empirically evaluating the VaR computation methods under consideration, we restrict ourselves to a portfolio consisting of only one stock. The VaR models are applied to two fictitious portfolios each consisting of a single stock represented by the stock market indices, the Dow Jones Industrial Average (DJIA) and the S&P/TSX Composite Index. The adequacy of each VaR model is then evaluated using a standard back-test procedure based on a likelihood ratio test. The kernel quantile estimation approach appears preferable to the two VaR computation methods mentioned above as no restrictive assumptions need to be made about the underlying distribution of returns, like in the case of the normal method. Also, smoothing the estimated quantile function using kernel density estimators can improve the precision of the VaR estimates.

The paper is organized as follows. Section 2 provides some background on assessing the goodness of a nonparametric estimator. Section 3 gives a brief overview of kernel density estimation and demonstrates how to obtain the empirically selected bandwidths. The density estimator with the Fejér-type kernel is presented in Section 4 along with its properties. Section 5 presents a simulation study comparing the kernel-type density estimator in question with other fixed kernel estimators in the literature. The proposed VaR computation method is introduced in Section 6. In Section 7, we use the new VaR model to estimate the VaR for two fictitious portfolios and compare the results to those of the commonly used VaR models by means of a back-test. Section 8 concludes the paper with a discussion and analysis of the results.

The following notation are used throughout the paper. We use the symbol $\mathbb{I}(A)$ for the indicator of a set A . The space of p -th power integrable functions on \mathbb{R} is denoted by $\mathbb{L}_p(\mathbb{R})$, for $1 \leq p < \infty$. Convergence almost surely is indicated by $\xrightarrow{a.s.}$. The expression $a_n \sim b_n$ means $\lim_{n \rightarrow \infty} a_n/b_n = 1$; whereas $a_n \asymp b_n$ means that there exist constants $0 < c < C < \infty$ and a number n^* such that $c < a_n/b_n < C$ for all $n \geq n^*$.

2. Common Approaches to Measuring the Quality of Density Estimators

Let X_1, X_2, \dots be a sequence of iid real-valued random variables with a common density f on \mathbb{R} that is unknown and is assumed to belong to a suitable family \mathcal{F} of smooth functions. For any function g that belongs to $\mathbb{L}_p(\mathbb{R})$, $1 \leq p < \infty$, we denote its \mathbb{L}_p -norm by $\|g\|_p = \left(\int_{\mathbb{R}} |g(x)|^p dx \right)^{1/p}$. Let $f_n(x) = f_n(x; X_1, \dots, X_n)$ be an arbitrary estimator of $f(x)$ at point $x \in \mathbb{R}$. One problem of interest is to construct an efficient estimator f_n of f in $\mathbb{L}_p(\mathbb{R})$, $1 \leq p < \infty$.

The performance of a density estimator can be evaluated through a risk function that measures the expected loss of choosing f_n as an estimator of f . For a given *loss function* $L(f, f_n)$, define the *risk* of f_n by

$$R(f, f_n) := \mathbb{E}L(f, f_n).$$

When $L(x) = l(\|x\|_p)$, for some function $l: [0, \infty) \rightarrow \mathbb{R}$ from a general class of loss functions \mathcal{L} , then we speak of the \mathbb{L}_p -risk given by

$$R_p(f, f_n) := \mathbb{E}l(\|f - f_n\|_p), \quad 1 \leq p < \infty. \tag{3}$$

The \mathbb{L}_p -risk with $p = 2$ is the measure used in this paper to judge the quality of a density estimator.

The quality of a density estimator is often measured by a minimax criterion. The idea is to protect statisticians from the worst that can happen. The *minimax risk* is given by

$$\mathcal{R}_n(\mathcal{F}) := \inf_{f_n} \sup_{f \in \mathcal{F}} R(f, f_n),$$

where the infimum is taken over all estimators f_n based on the random sample X_1, \dots, X_n and the supremum is over a given class \mathcal{F} of smooth density functions. In the nonparametric context, an asymptotic approach to minimax estimation is often used since exact minimaxity is rarely achievable. Asymptotic minimaxity, rate optimality, and local asymptotic minimaxity are three common criteria used in the statistical literature for measuring the asymptotic efficiency of a density estimator.

An estimator f_n^* is called *asymptotically minimax* if

$$\sup_{f \in \mathcal{F}} R(f, f_n^*) \sim \mathcal{R}_n(\mathcal{F}).$$

That is, for large sample sizes, the maximum risk of f_n^* over the class \mathcal{F} of estimated density functions is nearly equal to the minimax risk. Constructing asymptotically minimax estimators of f from some functional class \mathcal{F} is a difficult problem; instead, a large portion of the literature focuses on constructing rate optimal estimators. An estimator f_n^* is called *rate optimal* if as $n \rightarrow \infty$

$$\sup_{f \in \mathcal{F}} R(f, f_n^*) \asymp \mathcal{R}_n(\mathcal{F}).$$

In nonparametric regression analysis, work on asymptotically minimax estimators of smooth regression curves with respect to the \mathbb{L}_p -risk, $1 \leq p < \infty$, can be found in [3]-[5]. In connection with nonparametric density estimation, this is a more difficult problem and currently only solved for the \mathbb{L}_2 -risk (see Theorem 2 in [6]).

A more precise approach for finding efficient estimators is *local asymptotic minimaxity* (for a more detailed description on the origins of this method, see [7]). An estimator f_n^* of a density $f_0 \in \mathcal{F}$ is called *locally asymptotically minimax (LAM)* if as $n \rightarrow \infty$,

$$\sup_{f_0 \in \mathcal{V}} R(f_0, f_n^*) \sim \inf_{f_n} \sup_{f_0 \in \mathcal{V}} R(f_0, f_n),$$

where \mathcal{V} is a sufficiently small vicinity of f_0 with an appropriate distance defined on \mathcal{F} . Some examples of functions that admit LAM estimators can be found in [4] [8]. LAM estimators are preferred to asymptotically minimax ones since they are guaranteed to be globally efficient.

In kernel density estimation, the LAM ideology differs significantly from both the asymptotically minimax and rate optimality approaches. When constructing LAM estimators of f , one has to pay close attention to the choice of both the bandwidth h and the kernel K . Indeed, the usual *bias-variance tradeoff* approach, when the variance and the bias terms of an optimal estimator are to be balanced by a good choice of h , is no longer appropriate. In several papers, it is shown that with a careful choice of kernel the bias of f_n^* in the variance-bias decomposition becomes asymptotically negligible to its variance (see, for example, [2] [4]). Therefore, efficiency becomes achievable only with a careful choice of the kernel function.

In a recent paper of Stepanova [2], a kernel-type estimator for densities belonging to a class of infinitely smooth functions is shown to have the \mathbb{L}_p -risk coinciding with the minimax \mathbb{L}_p -risk as conjectured in Remark 5 of [5]. Moreover, following from Theorem 2 of [6], the estimator suggested in [2] cannot be improved with respect to the \mathbb{L}_2 -risk. The parameters used in the estimator in [2] are the Fejér-type kernel and the common theoretical bandwidth used for estimating infinitely smooth density functions. In this paper, we conduct a simulation study to show that the kernel-type density estimator in question cannot be improved with respect to the \mathbb{L}_2 -risk. We then show how to apply this efficient estimator to compute the VaR of portfolio returns.

3. Kernel Density Estimation

Let X_1, X_2, \dots be a sequence of iid real-valued random variables drawn from an absolutely continuous cumulative distribution function (cdf) $F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt$ in which the density function $f(x)$ is unknown. A kernel density estimator of $f(x)$ is defined by

$$f_n(x) = f_n(x; X_1, \dots, X_n) = \frac{1}{nh} \sum_{j=1}^n K\left(\frac{X_j - x}{h}\right), \quad x \in \mathbb{R},$$

where the parameter $h > 0$ is the *bandwidth*, the function K is the *kernel*, and $K_h(u) = h^{-1}K(h^{-1}u)$ is the *scaled kernel*. The bandwidth $h = h_n$, that typically depends on n , determines the smoothness of the estimator and satisfies

$$h_n \rightarrow 0 \quad \text{and} \quad nh_n \rightarrow \infty \quad \text{as} \quad n \rightarrow \infty.$$

Under certain nonrestrictive conditions on K , the above assumptions on h imply the consistency of $f_n(x)$ as an estimator of $f(x)$. The kernel is often a real-valued integrable function satisfying the following properties:

$$\int_{\mathbb{R}} K(u) du = 1, \quad K(u) = K(-u),$$

$$\max_{u \in \mathbb{R}} K(u) = K(0), \quad \text{and} \quad \int_{\mathbb{R}} K^2(u) du < \infty. \tag{4}$$

A more general class of density estimators includes the kernel-type estimators whose kernel functions, K_n , may depend on the sample size.

Some classical examples of kernels together with their Fourier transforms (see formula (6)) are listed in **Table 1** and presented graphically in **Figure 1**. These kernel functions are the most commonly applied in practice, most likely due to their additional nonnegative property as their corresponding estimators result in density functions. The group of kernels listed in **Table 2** and presented in **Figure 2**, along with their Fourier transforms, are well known in statistical theory and generally more asymptotically efficient than the standard kernels in **Table 1** since they were shown to achieve better rates of convergence in the works of [2] [9]. These kernel functions alternate between positive and negative values, except for the Fejér kernel. For these kernels, the *positive part estimator*

$$f_n^+(x) := \max\{0, f_n(x)\}$$

can be used to maintain the positivity of a density estimator. Throughout our analysis, we shall be using the positive part of all the kernel density estimators under study.

The most popular approach for judging the quality of an estimator in the literature and in practice is the Mean

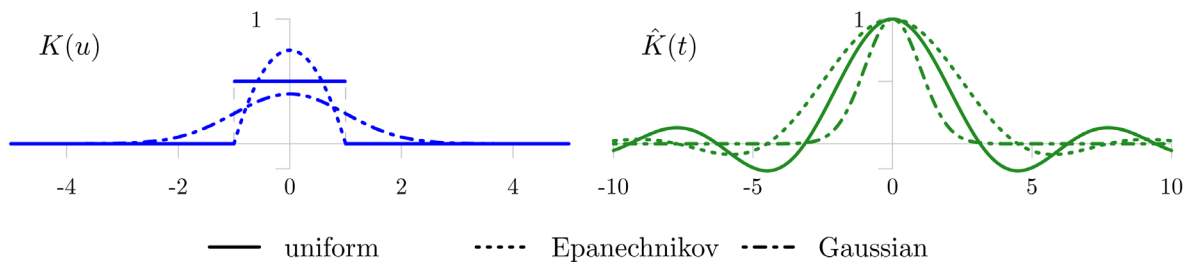


Figure 1. Some standard kernel functions and their Fourier transforms.

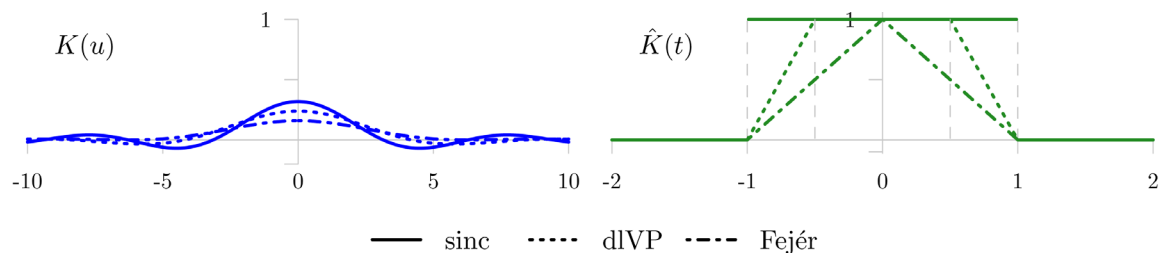


Figure 2. Some efficient kernel functions and their Fourier transforms.

Table 1. Some standard kernel functions and their Fourier transforms.

Kernel	$K(u)$	$\hat{K}(t) = \int_{\mathbb{R}} e^{itu} K(u) du$
uniform	$\frac{1}{2} \mathbb{I}(u \leq 1)$	$\begin{cases} \frac{\sin(t)}{t} & \text{if } t \neq 0, \\ 1 & \text{if } t = 0 \end{cases}$
Epanechnikov	$\frac{3}{4} (1-u^2) \mathbb{I}(u \leq 1)$	$\begin{cases} \frac{3}{t^3} (\sin(t) - t \cos(t)) & \text{if } t \neq 0, \\ 1 & \text{if } t = 0 \end{cases}$
Gaussian	$\frac{1}{\sqrt{2\pi}} e^{-u^2/2}$	$e^{-t^2/2}$

Table 2. Some efficient kernel functions and their Fourier transforms.

Kernel	$K(u)$	$\hat{K}(t) = \int_{\mathbb{R}} e^{itu} K(u) du$
sinc	$\begin{cases} \frac{\sin(u)}{\pi u} & \text{if } u \neq 0, \\ \frac{1}{\pi} & \text{if } u = 0 \end{cases}$	$\mathbb{I}(t \leq 1)$
de la Vallée Poussin	$\begin{cases} \frac{2(\cos(u/2) - \cos(u))}{\pi u^2} & \text{if } u \neq 0, \\ \frac{3}{4\pi} & \text{if } u = 0 \end{cases}$	$\mathbb{I}\left(t \leq \frac{1}{2}\right) + 2(1- t) \mathbb{I}\left(\frac{1}{2} < t < 1\right)$
Fejér	$\begin{cases} \frac{2\sin^2(u/2)}{\pi u^2} & \text{if } u \neq 0, \\ \frac{1}{2\pi} & \text{if } u = 0 \end{cases}$	$(1- t) \mathbb{I}(t \leq 1)$

Integrated Squared Error (MISE). Observe that the \mathbb{L}_2 -risk, as in (3), with $l(x) = x^2$ is simply the MISE defined as

$$\text{MISE}(f_n, f) := \mathbb{E}l(\|f_n - f\|_2^2) = \mathbb{E}\left[\int_{\mathbb{R}} (f_n(x) - f(x))^2 dx\right]. \tag{5}$$

By the Fubini theorem, the right-hand side (RHS) of (5) can be further expanded to represent the variance- bias decomposition of the density estimator:

$$\begin{aligned} \text{MISE}(f_n, f) &= \int_{\mathbb{R}} (\mathbb{E}[f_n^2(x)] - \mathbb{E}^2[f_n(x)]) dx + \int_{\mathbb{R}} (\mathbb{E}[f_n(x)] - f(x))^2 dx \\ &=: \int_{\mathbb{R}} \sigma_n^2(x) dx + \int_{\mathbb{R}} b_n^2(x) dx. \end{aligned}$$

Notice also that the MISE of the positive part estimator f_n^+ satisfies

$$\text{MISE}(f_n^+, f) \leq \text{MISE}(f_n, f).$$

3.1. Fourier Analysis of Kernel Density Estimators

In nonparametric estimation, the use of Fourier analysis makes it often easier to study statistical properties of estimators. It can be noted from **Figure 1** and **Figure 2** that the Fourier transforms of the efficient kernels have a

simpler form than those of the standard kernels. This simplifies the analysis of density estimators under certain settings when using efficient kernel functions. We begin by providing a few basic definitions and properties related to the Fourier transform (see, for example, Chapter 9 of [10]).

The *Fourier transform* \hat{g} of a function $g \in \mathbb{L}_1(\mathbb{R})$ is defined by

$$\hat{g}(t) := \int_{\mathbb{R}} e^{itx} g(x) dx, \quad t \in \mathbb{R}, \tag{6}$$

where $i = \sqrt{-1}$. The Plancherel theorem allows us to extend the definition of the Fourier transform to functions in $\mathbb{L}_2(\mathbb{R})$. Moreover, for any $g \in \mathbb{L}_2(\mathbb{R})$, the Parseval formula holds true:

$$\int_{\mathbb{R}} |g(x)|^2 dx = \frac{1}{2\pi} \int_{\mathbb{R}} |\hat{g}(t)|^2 dt. \tag{7}$$

Using also the notation $\widehat{g(\cdot)}(t)$ for the Fourier transform of $g \in \mathbb{L}_2(\mathbb{R})$ at t , for any $h > 0$ and $x \in \mathbb{R}$

$$\widehat{\frac{1}{h}g\left(\frac{\cdot}{h}\right)}(t) = \hat{g}(ht), \quad \widehat{g(x-\cdot)}(t) = e^{itx} \hat{g}(-t), \quad t \in \mathbb{R}. \tag{8}$$

The Fourier transform of a density is known to be the *characteristic function* defined by

$$\varphi(t) := \int_{\mathbb{R}} e^{itx} f(x) dx = \int_{\mathbb{R}} e^{itx} dF(x), \quad t \in \mathbb{R}.$$

The corresponding *empirical characteristic function* is

$$\varphi_n(t) := \varphi_n(t; X_1, \dots, X_n) = \int_{\mathbb{R}} e^{itx} dF_n(x) = \frac{1}{n} \sum_{j=1}^n e^{itX_j}, \quad t \in \mathbb{R},$$

where $F_n(x) = n^{-1} \sum_{j=1}^n \mathbb{I}(X_j \leq x)$, and has the following properties that follow one after the other:

$$\begin{aligned} \mathbb{E}[\varphi_n(t)] &= \varphi(t), \\ \mathbb{E}[|\varphi_n(t)|^2] &= \left(1 - \frac{1}{n}\right) |\varphi(t)|^2 + \frac{1}{n}, \\ \mathbb{E}[|\varphi_n(t) - \varphi(t)|^2] &= \frac{1}{n} (1 - |\varphi(t)|^2), \end{aligned} \tag{9}$$

for all $t \in \mathbb{R}$. Given the properties in (8) and the symmetry of K , the Fourier transform \hat{f}_n of the density estimator f_n can be expressed as follows:

$$\hat{f}_n(t) = \frac{1}{n} \sum_{j=1}^n \widehat{K_h(X_j - \cdot)}(t) = \frac{1}{n} \sum_{j=1}^n e^{itX_j} \hat{K}(ht) = \varphi_n(t) \hat{K}(ht). \tag{10}$$

In \mathbb{L}_2 -theory, the MISE can also be expressed using the Fourier analysis of kernel density estimators. Indeed according to (7), assuming that f and K are both in $\mathbb{L}_2(\mathbb{R})$ and that K is symmetric, the MISE satisfies

$$\text{MISE}(f_n, f) = \mathbb{E} \left[\int_{\mathbb{R}} (f_n(x) - f(x))^2 dx \right] = \frac{1}{2\pi} \mathbb{E} \left[\int_{\mathbb{R}} |\hat{f}_n(t) - \hat{f}(t)|^2 dt \right].$$

Continuing from (10) and relations (9), the MISE of the kernel estimator f_n of density f takes the form (see Theorem 1.4 of [11])

$$\begin{aligned} \text{MISE}(f_n, f) &= \frac{1}{2\pi} \left[\int_{\mathbb{R}} |1 - \hat{K}(ht)|^2 |\varphi(t)|^2 dt + \frac{1}{n} \int_{\mathbb{R}} |\hat{K}(ht)|^2 dt \right] \\ &\quad - \frac{1}{2\pi n} \int_{\mathbb{R}} |\varphi(t)|^2 |\hat{K}(ht)|^2 dt =: J_n(K, h, \varphi), \end{aligned} \tag{11}$$

where $f, K \in \mathbb{L}_2(\mathbb{R})$, $n \geq 1$, and $h > 0$.

Formula (11) provides a more suitable method for expressing the MISE than some classical approaches that derive upper bounds on the integrated squared risk (see [12], Section 2.1.1). Unlike the classical approaches, the assumptions required to obtain formula (11) are not very restrictive, which allows for the derivation of more

optimal kernels. Indeed, most of the general properties of a kernel in (4), such as K integrating to one or being an integrable function, do not need to be true. Also, the expression for $J_n(K, h, \varphi)$ in (11) makes it possible to easily determine inadmissible kernel functions in $\mathbb{L}_2(\mathbb{R})$ for any fixed n . Recall that a kernel is called *inadmissible* if there exist other kernels in $\mathbb{L}_2(\mathbb{R})$ that can improve $J_n(K, h, \varphi)$ for all characteristic functions in $\mathbb{L}_2(\mathbb{R})$. A simple method for detecting inadmissible kernels was presented by Cline [13]: if

$$\text{Leb}\left(t: \hat{K}(t) \notin [0,1]\right) > 0, \tag{12}$$

where $\text{Leb}(A)$ denotes the Lebesgue measure of a set A , then K is inadmissible. As seen from **Figure 1**, the Epanechnikov and uniform kernel functions are inadmissible since the set $\{t: \hat{K}(t) \notin [0,1]\}$ has a positive Lebesgue measure. This is another argument as to why the efficient kernels listed in **Table 2** as well as the family of Fejér-type kernels in (20) are preferred.

3.2. Bandwidth Selection Based on Unbiased Risk Estimation and Fourier Analysis Techniques

Selecting an appropriate bandwidth $h = h_n$, that is dependent on the sample size, is very important as it determines the smoothness of the kernel density estimator. A small bandwidth produces a peaky-like estimator indicative of high variability caused by under-smoothing. On the other hand, a large bandwidth increases the bias of the estimator and the important features of the distribution may be lost due to over-smoothing. The aim is to choose a bandwidth that minimizes the bias and the variance of an estimator to avoid over- or under-smoothing, a dilemma known as the bias-variance tradeoff.

In theory, an optimal bandwidth can be obtained by minimizing the MISE with respect to h :

$$h_{opt} = \arg \min_{h>0} \text{MISE}(h). \tag{13}$$

In practice, the RHS of (13) cannot be computed as the MISE depends on the unknown density f . Instead, an approximately unbiased estimator of the $\text{MISE}(h)$ computed from the random sample X_1, \dots, X_n is minimized. The idea is to consider the expansion of the MISE in (5) in the following way

$$\text{MISE}(h) = \mathbb{E}\left[\int_{\mathbb{R}} f_n^2(x) dx - 2\int_{\mathbb{R}} f_n(x) f(x) dx\right] + \int_{\mathbb{R}} f^2(x) dx =: \mathcal{J}_n(h) + \int_{\mathbb{R}} f^2(x) dx.$$

As we are only concerned with minimizing the MISE with respect to h , the term $\int_{\mathbb{R}} f^2(x) dx$ may be disregarded. The estimator

$$\text{CV}(h) := \int_{\mathbb{R}} f_n^2(x) dx - \frac{2}{n} \sum_{j=1}^n f_{n,-j}(X_j),$$

where

$$f_{n,-j}(x) := \frac{1}{(n-1)h} \sum_{k \neq j} K\left(\frac{X_k - x}{h}\right)$$

is the leave-one-out estimator of $f(x)$, is an unbiased estimator of $\mathcal{J}_n(h)$. The function $\text{CV}(\cdot)$ is called the unbiased *cross-validation* criterion, which can be further expanded to (see [14], p. 55)

$$\begin{aligned} \text{CV}(h) &= \frac{1}{n^2 h} \sum_{j=1}^n \sum_{k=1}^n (K * K) \left(\frac{X_j - X_k}{h}\right) - \frac{2}{n(n-1)h} \sum_{j=1}^n \sum_{k \neq j} K \left(\frac{X_j - X_k}{h}\right) \\ &= \frac{1}{nh} (K * K)(0) + \frac{2}{n^2 h} \sum_{j=1}^k \sum_{k=j+1}^n (K * K) \left(\frac{X_j - X_k}{h}\right) \\ &\quad - \frac{4}{nh(n-1)} \sum_{j=1}^k \sum_{k=j+1}^n K \left(\frac{X_j - X_k}{h}\right), \end{aligned} \tag{14}$$

where $*$ denotes the convolution. It follows that $\text{CV}(h) + \int_{\mathbb{R}} f^2(x) dx$ is an unbiased estimator of the MISE, where f^2 is independent of h , implying that $\mathbb{E}[\text{CV}(h)]$ and $\text{MISE}(h)$ would both obtain the same

minimums for the values of h_{opt} . In practice, it is expected that, for a random sample X_1, \dots, X_n , the minimizer of $CV(h)$ is close to the minimizer of $\mathbb{E}[CV(h)]$. The cross-validation bandwidth is therefore given by

$$h_{CV} := \arg \min_{h>0} CV(h; X_1, \dots, X_n), \tag{15}$$

yielding

$$f_{n,CV}(x) := \frac{1}{nh_{CV}} \sum_{j=1}^n K\left(\frac{X_j - x}{h_{CV}}\right)$$

as the kernel estimator of $f(x)$ specified by h_{CV} . Selecting bandwidths using the unbiased cross-validation criterion of the form above was first introduced by Rudemo in [15].

Cross-validation is perhaps the most common approach based on unbiased risk estimation for selecting h ; however, many authors have noted that it has a slow rate of convergence towards h_{opt} in (13) (see [16], Theorem 4.1). Another parallel method is to minimize an unbiased estimator of the MISE based on the Fourier analysis of kernel density estimators over h . The latter method for selecting a bandwidth is due to Golubev [17] and is shown to provide more reliable results in our simulation study in Section 5 than the cross-validation approach. For this reason, we shall use this method to select our bandwidth in the VaR computation model proposed in Section 6.

The MISE of interest is given in (11) and denoted by $J_n(K, h, \varphi)$. Golubev [17] found an approximately unbiased estimator of $J_n(K, h, \varphi)$ given by

$$\tilde{J}_n(h) := \int_{\mathbb{R}} \left(-2\hat{K}(ht) + \hat{K}^2(ht) \left(1 - \frac{1}{n}\right) \right) |\varphi_n(t)|^2 dt + \frac{2}{n} \int_{\mathbb{R}} \hat{K}(ht) dt, \tag{16}$$

where $h > 0$. Indeed, from relations (9) we get that, up to scaling and shifting, $\tilde{J}_n(h)$ is an unbiased estimator of $J_n(K, h, \varphi)$:

$$\mathbb{E}[\tilde{J}_n(h)] = 2\pi \left(1 - \frac{1}{n}\right) \left[J_n(K, h, \varphi) - \int_{\mathbb{R}} f^2(x) dx \right].$$

Hence, minimizing $\mathbb{E}[\tilde{J}_n(h)]$ is equivalent to minimizing $J_n(K, h, \varphi)$ over h . In practice, an *approximate minimizer* of $J_n(K, h, \varphi)$ is obtained by using the random sample X_1, \dots, X_n to compute $\tilde{J}_n(h)$ and then minimized with respect to h :

$$h_F := \arg \min_{h>0} \tilde{J}_n(h; X_1, \dots, X_n). \tag{17}$$

The corresponding kernel density estimator with the bandwidth h_F is then

$$f_{n,F}(x) := \frac{1}{nh_F} \sum_{j=1}^n K\left(\frac{X_j - x}{h_F}\right).$$

Under appropriate conditions, h_F and h_{CV} are asymptotically optimal as they are asymptotically equivalent to h_{opt} in (13). In other words, the MISE of the kernel density estimators $f_{n,F}$ and $f_{n,CV}$ is asymptotically equivalent to that of the estimator with the optimal bandwidth h_{opt} (see Section 1.4 of [11]).

4. Density Estimators with Fejér-Type Kernel Functions

Suppose that X_1, X_2, \dots is a sequence of iid random variables on \mathbb{R} with a common density function f from some functional class \mathcal{F} . Consider the functional class $\mathcal{F}_\gamma = \mathcal{F}(\gamma, L)$ of functions f in $\mathbb{L}_2(\mathbb{R})$ such that each $f \in \mathcal{F}_\gamma$ admits an analytic continuation to the strip $S_\gamma = \{x + iy : |y| \leq \gamma\}$ with $\gamma > 0$ such that $f(x + iy)$ is analytic on the interior of S_γ , bounded on S_γ , and for some $L > 0$

$$\|f(\cdot \pm iy)\|_2 \leq L.$$

We have for any $f \in \mathcal{F}_\gamma$

$$\frac{1}{2\pi} \int_{\mathbb{R}} \cosh^2(\gamma t) |\hat{f}(t)|^2 dt \leq L. \tag{18}$$

The functional class \mathcal{F}_γ is well known in approximation theory (see, for example, [18], Section 94) and widely used in nonparametric estimation (see [2]-[5] [19] [20]). For certain values of γ , the class \mathcal{F}_γ contains probability densities such as the normal, Student's t , and Cauchy as well as their analytic transformations and mixtures. The inequality in (18) is used in [12] to determine how large the values of γ can be chosen so that these probability densities belong to \mathcal{F}_γ . The normal, Student's t with odd degrees of freedom ν , and standard Cauchy density function are in the analytical class \mathcal{F}_γ with $\gamma > 0$, $0 < \gamma < \sqrt{\nu}$, and $0 < \gamma < 1$, respectively. For other examples of functions belonging to \mathcal{F}_γ we refer to Section 2.3 of [21].

The kernel-type estimator of $f \in \mathcal{F}_\gamma$ considered in this work has the form

$$f_n(x; \theta_n) := f_n(x; X_1, \dots, X_n, \theta_n) = \frac{1}{nh_n} \sum_{j=1}^n K_n\left(\frac{X_j - x}{h_n}; \theta_n\right), \quad x \in \mathbb{R}, \tag{19}$$

where K_n is the Fejér-type kernel given by

$$K_n(u; \theta_n) := \begin{cases} \frac{\cos(\theta_n u) - \cos(u)}{\pi(1 - \theta_n)u^2} & \text{if } u \neq 0, \\ \frac{1 + \theta_n}{2\pi} & \text{if } u = 0, \end{cases} \tag{20}$$

with

$$h_n = \frac{\theta_n}{M}, \quad \theta_n = 1 - \frac{1}{M}, \quad M = \frac{\log n}{2\gamma}. \tag{21}$$

It is easy to see that the Fejér-type kernel as in (20) satisfies the properties in (4). The parameters in (21) are chosen to have

$$h_n \rightarrow 0 \quad \text{and} \quad nh_n \rightarrow \infty \quad \text{as } n \rightarrow \infty,$$

ensuring the consistency of $f_n(x; \theta_n)$ in (19) as an estimator of $f(x)$. Moreover, the kernel-type density estimator as in (20) with the bandwidth h_n satisfying (21) is known to have very small \mathbb{L}_p -risk, $1 \leq p < \infty$ (see Theorem 1 of [2]). For some choices of $0 \leq \theta_n < 1$, **Table 3** shows how the Fejér-type kernel coincides with the well-known efficient kernels listed in **Table 2**. The sinc kernel is the limiting case of the Fejér-type kernel when $\theta_n \rightarrow 1$; in other words, when n approaches infinity. Additionally, choosing $\theta_n = 1/2$ and $\theta_n = 0$ leads to the de la Vallée Poussin and Fejér kernels, respectively.

The Fourier transform of the Fejér-type kernel is given by (see [18], p. 202)

$$\hat{K}_n(t; \theta_n) = \int_{\mathbb{R}} e^{itu} K_n(u; \theta_n) du = \mathbb{I}(|t| \leq \theta_n) + \left(\frac{1 - |t|}{1 - \theta_n}\right) \mathbb{I}(\theta_n \leq |t| \leq 1).$$

The Fejér-type kernel K_n and its Fourier transform \hat{K}_n are presented graphically in **Figure 3**. Observe the simple form of \hat{K}_n , which makes it very useful in studying analytically the properties of the estimator in (19). Also, \hat{K}_n is nonnegative and bounded by one making the Fejér-type kernel admissible according to the Cline criterion as in (12).

To apply the data-driven bandwidths given in (15) and (17), the unbiased estimators as in (14) and (16), denoted by $CV(h)$ and $\tilde{J}_n(h)$, need to be evaluated for the Fejér-type kernel. The unbiased cross-validation criterion $CV(h)$ includes the convolution of the kernel with itself. The self-convolution of the Fejér-type kernel is given by (see p. 44 of [12] for details)

Table 3. Cases of the Fejér-type kernel.

Kernel	θ_n
sinc	1
de la Vallée Poussin	1/2
Fejér	0

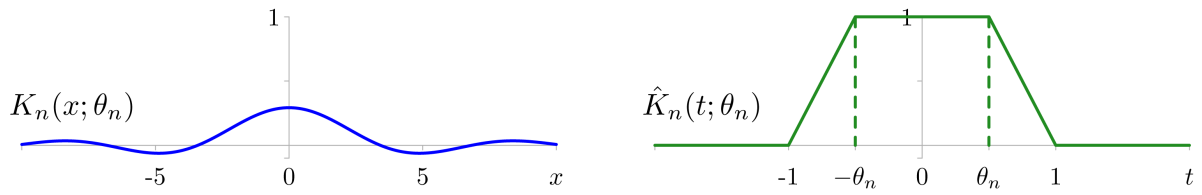


Figure 3. The Fejér-type kernel function and its Fourier transform.

$$(K_n * K_n)(u; \theta_n) = \begin{cases} \frac{2 \cos(\theta_n u)}{\pi(1-\theta_n)u^2} + \frac{2(\sin(\theta_n u) - \sin(u))}{\pi(1-\theta_n)^2 u^3} & \text{if } u \neq 0, \\ \frac{1+2\theta_n}{3\pi} & \text{if } u = 0, \end{cases}$$

where $0 \leq \theta_n < 1$ for $n > 1$. It can be shown (see [12], p. 48) that the unbiased risk estimator based on the Fourier analysis of a density estimator with the Fejér-type kernel is given by:

$$\begin{aligned} \tilde{J}_n(h; \theta_n) = & \frac{2}{n^2} \left[\frac{2h^2}{(1-\theta_n)^2} \left(\frac{n-1}{n} \right) \sum_{j \neq k} \frac{\sin(\theta_n(X_j - X_k)/h) - \sin((X_j - X_k)/h)}{(X_j - X_k)^3} \right. \\ & \left. + \frac{2h}{(1-\theta_n)} \sum_{j \neq k} \frac{\cos((X_j - X_k)/h) - \cos(\theta_n(X_j - X_k)/h)/n}{(X_j - X_k)^2} \right] \\ & - \frac{2(n+1)\theta_n}{n^2 h} + \frac{2(n-1)(1-\theta_n)}{3n^2 h} + \frac{4\theta_n}{nh}, \end{aligned}$$

for $n > 1$, $h > 0$, $0 \leq \theta_n < 1$, and $X_j \neq X_k$, $1 \leq j \neq k \leq n$. Note that in the case of sampling from a continuous distribution $P(X_j \neq X_k) = 1, 1 \leq j \neq k \leq n$.

In the following section, we demonstrate numerically that the positive part of the kernel-type estimator in (19) with the Fejér-type kernel in (20) works well with respect to the \mathbb{L}_2 -risk for both theoretical and empirical bandwidth selectors.

5. Simulation Study: Comparison of Kernel Density Estimators

A simulation study is carried out to assess the quality of the positive part of the density estimator in (19) with the Fejér-type kernel in (20) using the MISE criterion. The finite-sample performance of (19) is compared to other density estimators that use the sinc, de la Vallée Poussin, and Gaussian kernels. These kernel functions were chosen since the sinc and de la Vallée Poussin are efficient kernels that are specific cases of the Fejér-type, while the Gaussian kernel is the most commonly used in practice. Three bandwidth selectors are applied to the kernel density estimators in hand. The bandwidth selection methods include the empirical approaches from cross-validation and Fourier analysis and the theoretical smoothing parameter $h_n = (2\gamma\theta_n)/\log n$, which is used for density estimators with efficient kernels. From here on, we shall refer to the bandwidth h_F in (17) as the Fourier bandwidth.

We generated 200 random samples, of a wide range of sample sizes, from the following four distributions: standard normal $N(0,1)$, Student’s t_{15} with 15 degrees of freedom, chi-square χ_4^2 with 4 degrees of freedom, and normal mixture $0.4N(0,1) + 0.6N(1,0.4^2)$. These distributions were chosen as their density functions cover different shapes and characteristics such as: symmetry, skewness, unimodality, and bimodality. The chi-square is the only one out of these distributions whose density function f is not in the functional class \mathcal{F}_γ defined in Section 2; nonetheless, we are interested in observing the behaviour of $f_n(x; \theta_n)$ in (19) when estimating such densities. For each simulated dataset, density estimates are computed for every kernel function and bandwidth selection method under consideration. An appropriate smoothing parameter γ was manually selected for the Fejér-type kernel and the theoretical bandwidth. For further details on the methodology used to

conduct the experiments refer to Section 4.1 of [12].

Let us first assess the bandwidth selection methods under consideration. **Figure 4** and **Figure 5** capture the performance of each bandwidth selection method for each kernel density estimate under study by plotting the MISE estimates against samples of size 25 to 100 and 200 to 1000, respectively. The following can be observed from the figures. For a good choice of γ , the estimates of $f \in \mathcal{F}_\gamma$ with efficient kernel functions and theoretical bandwidths complement the results of Theorem 1 in [2] and Theorem 2 in [6] by outperforming the estimates with empirical bandwidths. The theoretical estimates do not perform as well, though, when estimating

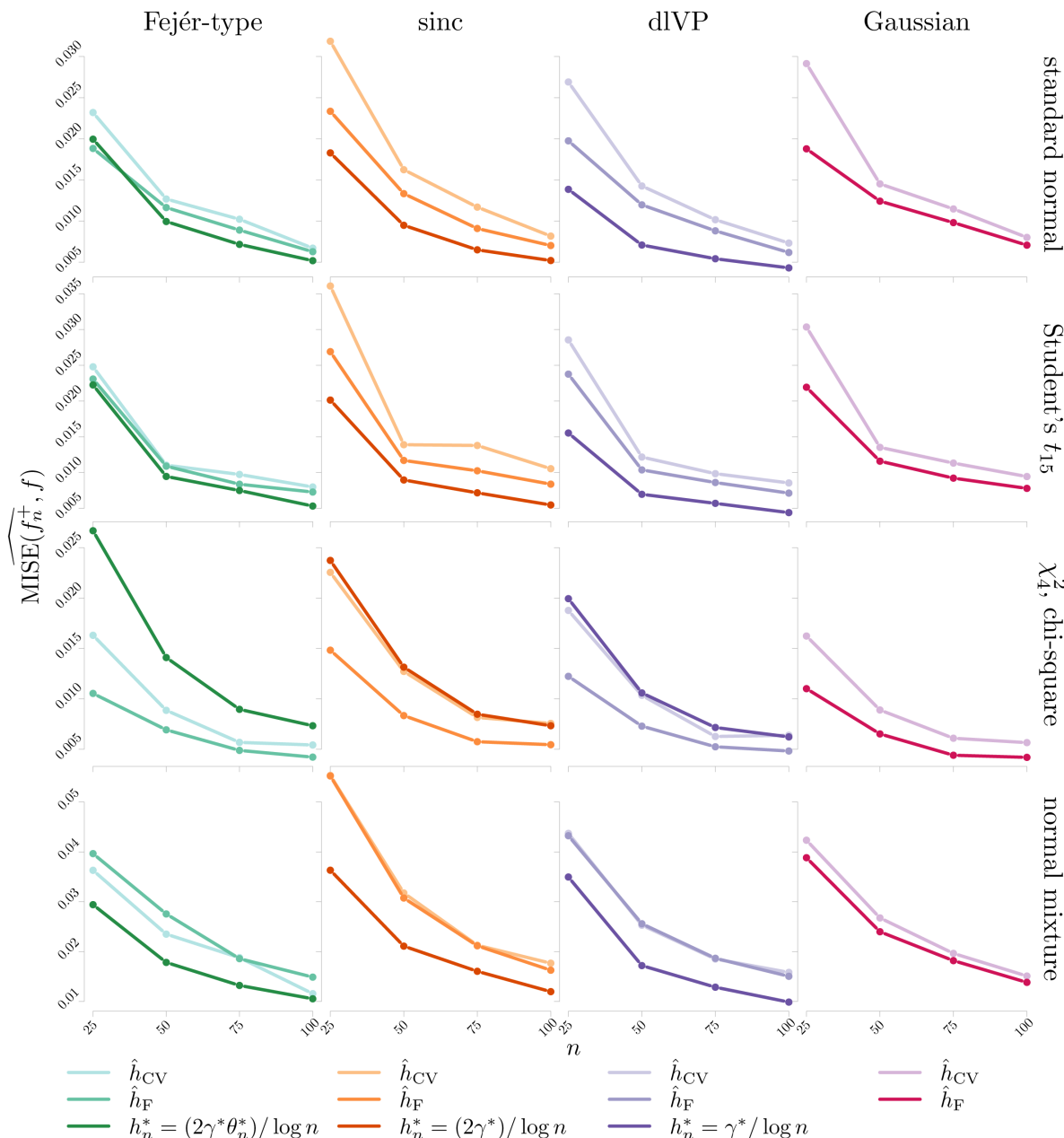


Figure 4. MISE estimates for the cross-validation, Fourier, and theoretical bandwidth selectors with Fejér-type, sinc, dIVP, and Gaussian kernels that estimate the standard normal, Student's t_{15} , chi-square χ^2 , and normal mixture, for small sample sizes. The MISE estimates are averaged over 200 replications. The symbol * denotes the manually-selected γ that provided good results.

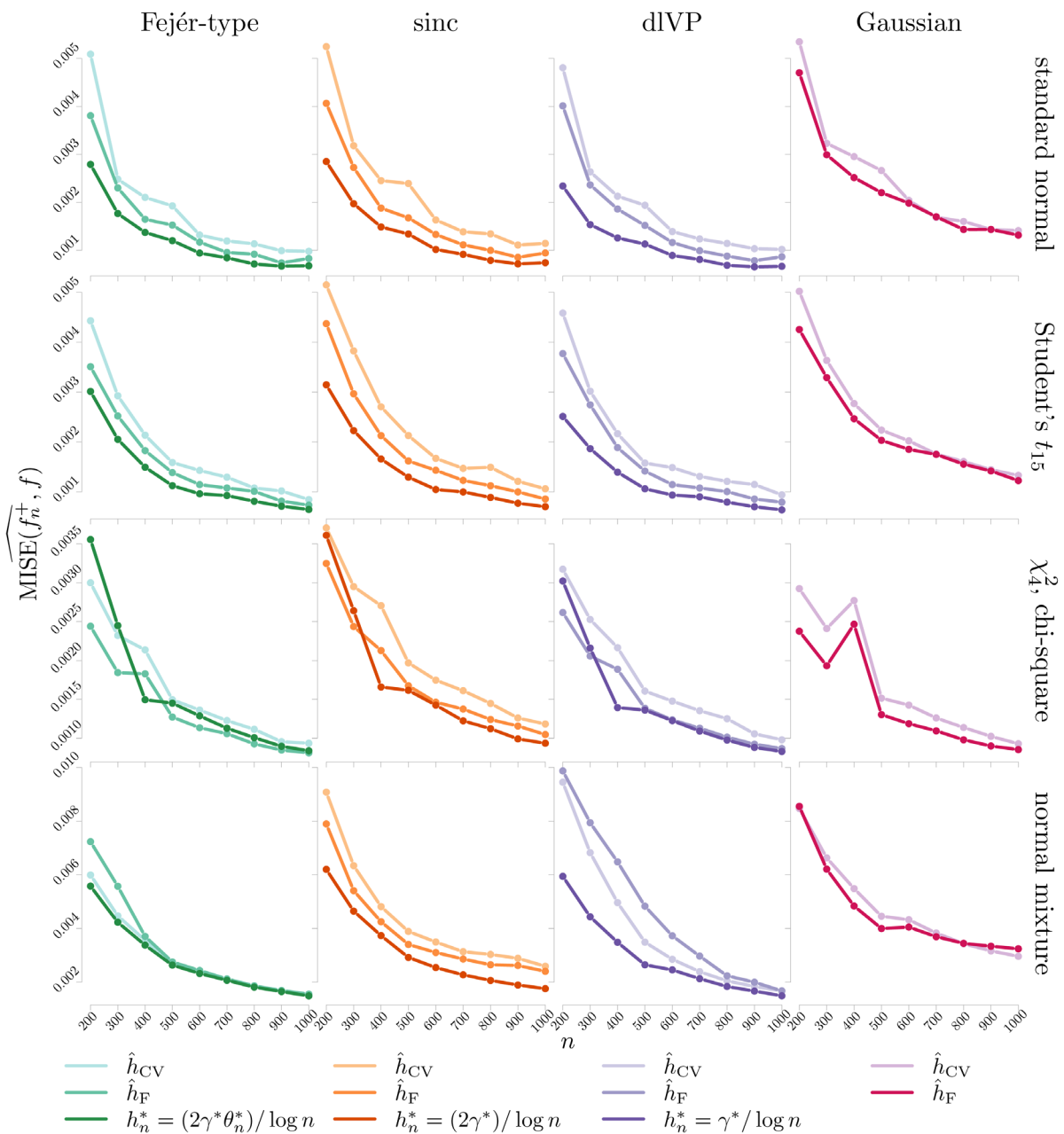


Figure 5. MISE estimates for the cross-validation, Fourier, and theoretical bandwidth selectors with Fejér-type, sinc, dlVP, and Gaussian kernels that estimate the standard normal, Student’s t_{15} , chi-square χ^2 , and normal mixture, for large sample sizes. The MISE estimates are averaged over 200 replications. The symbol $*$ denotes the manually-selected γ that provided good results.

the chi-square density, which is not in \mathcal{F}_γ , particularly for smaller sample sizes. Generally speaking, it can also be seen that, when estimating the unimodal densities, the bandwidth based on the Fourier analysis techniques is better than, or equal to, the cross-validation bandwidth. The difference in estimation error is especially noticeable for smaller sample sizes.

Now, we assess the quality of the Fejér-type kernel estimator when using empirical bandwidths. **Figure 6** and **Figure 7** capture the performance of the kernel functions for each density estimate under a specified empirical bandwidth by plotting the MISE estimates against samples of size 25 to 100 and 200 to 1000, respectively. For

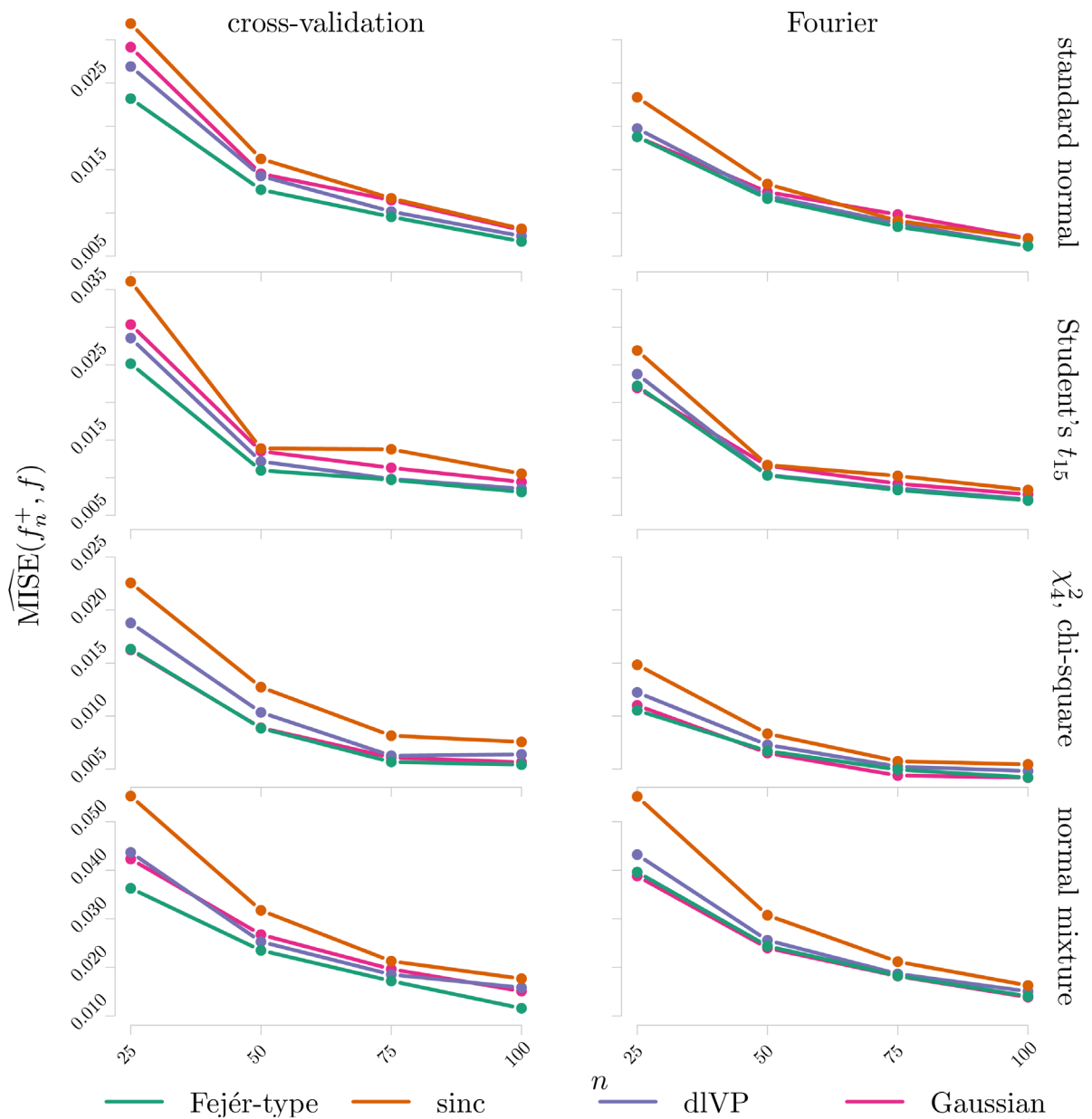


Figure 6. MISE estimates for the Fejér-type, sinc, dlVP, and Gaussian kernels with cross-validation and Fourier bandwidth selectors that estimate a standard normal, Student’s t_{15} , chi-square χ^2_4 , and normal mixture, for small sample sizes. The MISE estimates are averaged over 200 replications.

estimation of the unimodal densities with data-dependent bandwidth methods, the Fejér-type kernel slightly improves the other fixed efficient kernels and performs much better than the common Gaussian kernel for larger sample sizes. Also, we observe that, when estimating the bimodal density, the Fejér-type kernel performs much better than all of the competing kernels, especially for large sample sizes.

In summary, for an appropriate choice of γ , the estimates of $f \in \mathcal{F}_\gamma$ with efficient kernel functions and theoretical bandwidths outperform the estimates with empirical bandwidths. Between the data-dependent bandwidth selection methods, the method based on Fourier analysis techniques provided more accurate results than that of the cross-validation, regardless of the kernel function used. Moreover, the method based on Fourier analysis is easier to implement, more accurate for small sample sizes, and less time-consuming for large samples. The positive part of the kernel-type estimator of f as in (19) compares favourably, in terms of the estimated

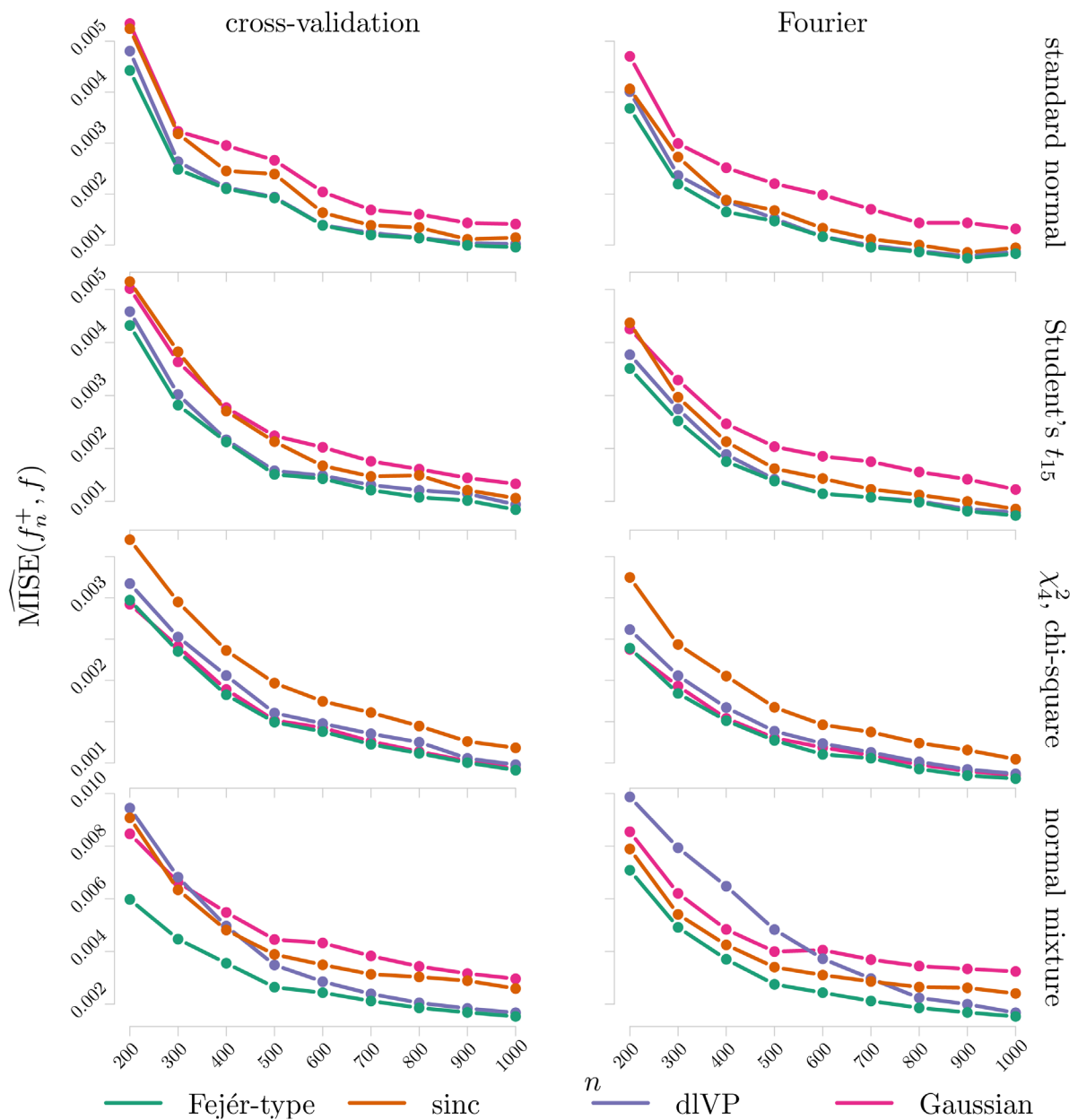


Figure 7. MISE estimates for the Fejér-type, sinc, dIVP, and Gaussian kernels with cross-validation and Fourier bandwidth selectors that estimate a standard normal, Student’s t_{15} , chi-square χ^2_4 , and normal mixture, for large sample sizes. The MISE estimates are averaged over 200 replications.

MISE, with competing kernel estimators, especially when estimating normal mixtures. The simulation results attest that, for a good choice of γ , the estimator with the Fejér-type kernel performs very well when using both empirical and theoretical bandwidths to estimate densities in \mathcal{F}_γ and therefore is reliable in application.

6. VaR Model with Fejér-Type Kernel Functions

Suppose that X_1, \dots, X_n is a random sample of iid portfolio returns with an absolutely continuous cdf F on \mathbb{R} , and let

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$$

denote the corresponding order statistics. Recall that VaR models are concerned with evaluating a quantile function $Q(p)$, $0 < p < 1$, defined as

$$Q(p) := F^{-1}(p) = \inf \{x : F(x) \geq p\}$$

for a general cdf $F(x)$ that is continuous from the right. We are interested in estimating a quantile function using the Fejér-type kernel function in (20).

Let $F_n(x)$ be the empirical distribution function given by

$$F_n(x) := \frac{1}{n} \sum_{j=1}^n \mathbb{I}(X_j \leq x), \quad x \in \mathbb{R}.$$

By Kolmogorov’s strong law of large numbers, the empirical distribution function is a strongly consistent estimator of the true distribution for any $x \in \mathbb{R}$, that is,

$$F_n(x) \xrightarrow[n \rightarrow \infty]{a.s.} F(x), \quad x \in \mathbb{R}.$$

Moreover, by the Glivenko-Cantelli theorem,

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \xrightarrow[n \rightarrow \infty]{a.s.} 0.$$

A common definition for the empirical quantile function is

$$Q_n(p) := F_n^{-1}(p) = X_{(j)}$$

for $(j-1)/n < p \leq j/n$, $j = 1, \dots, n$. In 1979, Parzen (see [22], p. 113) introduced the kernel quantile estimator

$$KQ_n(p) = \int_0^1 K_h(t-p) Q_n(t) dt = \int_0^1 K_h(t-p) F_n^{-1}(t) dt =: \sum_{j=1}^n \left[\int_{(j-1)/n}^{j/n} K_h(t-p) dt \right] X_{(j)}, \quad (22)$$

where $0 < p < 1$ and for a suitable kernel function K , $K_h(x) = h^{-1}K(h^{-1}x)$. Naturally, $KQ_n(p)$ puts most weight on the order statistic $X_{(j)}$, for which j/n is close to p . Sheather and Marron [23] showed that the following approximation to $KQ_n(p)$ as in (22) can be used in practice:

$$\widetilde{KQ}_n(p) = \frac{\sum_{j=1}^n K_h \left(\frac{j-1/2}{n} - p \right) X_{(j)}}{\sum_{j=1}^n K_h \left(\frac{j-1/2}{n} - p \right)}.$$

Therefore, for a probability level $p \in (0,1)$, we suggest that the VaR estimator can be computed as

$$\widehat{\text{VaR}}_{3,n}(p) = \frac{\sum_{j=1}^n K_{h_F} \left(\frac{j-1/2}{n} - p; \theta_n \right) X_{(j)}}{\sum_{j=1}^n K_{h_F} \left(\frac{j-1/2}{n} - p; \theta_n \right)}, \quad (23)$$

where K_{h_F} is the scaled Fejér-type kernel function based on (20) and h_F is the bandwidth in (17), referred to as the Fourier bandwidth. The bandwidth obtained from Fourier analysis methods was chosen as it provided good results in the simulation studies in Section 5.

7. Application to Value at Risk

We assess the proposed nonparametric VaR computation method given by formula (23) and compare it to the common normal and historical simulation approaches as in (1) and (2). Each VaR computation method is evaluated by means of a statistical back-test procedure based on a likelihood ratio test.

7.1. Evaluation of VaR Computation Methods

To evaluate the adequacy of each VaR computation method, we perform a statistical test that systematically compares the actual returns to the corresponding VaR estimates. The number of observations that exceed the

VaR of the portfolio should fall within a specified confidence level; otherwise, the model is rejected as it is not considered adequate for predicting the VaR of a portfolio. A back-test of this form was first used by Kupiec in 1995 (see [1], Chapter 6).

Let X_1, X_2, \dots be a sequence of iid random portfolio returns with a common density f on X_1, X_2, \dots . In our analysis, we consider two different samples; an estimation sample of size n for computing the VaR and an evaluation sample of size m for comparing the estimated VaR returns with the actual returns. Let Y_1, \dots, Y_m be iid random variables that indicate whether or not the realized return is worse than the predicted VaR measure, that is,

$$Y_{t+1} = \begin{cases} 1 & \text{if } X_{t+1} \leq \text{VaR}_t(p), \\ 0 & \text{if } X_{t+1} > \text{VaR}_t(p), \end{cases} \quad t = 0, \dots, m-1,$$

where $P(X_{t+1} \leq \text{VaR}_t(p)) = p$. Then, $N = \sum_{i=1}^m Y_i$ is the number of estimated VaR violations of a portfolio out of an evaluation sample of m observations and follows the binomial distribution $\text{Bin}(m, p)$ with probability mass function (pmf)

$$P(N = y; p) = \binom{m}{y} p^y (1-p)^{m-y},$$

for $y = 0, 1, \dots, m$. Suppose the probability level for the VaR is chosen to be p_0 . The ratio N/m represents the failure rate of the VaR model, which under the null hypothesis specified below converges to p_0 . The relevant null and alternative hypotheses for determining the fit of the VaR model are given by

$$H_0 : p = p_0 \quad \text{vs.} \quad H_1 : p \neq p_0.$$

A likelihood ratio test is carried out to determine whether or not to reject the null hypothesis that the model is adequate. The likelihood function for p given the observed values y_1, \dots, y_m of Y_1, \dots, Y_m is

$$L(p | y_1, \dots, y_m) = \prod_{i=1}^m P(Y_i = y_i; p) = \prod_{i=1}^m p^{y_i} (1-p)^{1-y_i} = p^y (1-p)^{m-y}, \quad (24)$$

where $y = \sum_{i=1}^m y_i$. Following from (24), the appropriate likelihood ratio test statistic is given by

$$\Lambda(p_0) = \Lambda(p_0 | y_1, \dots, y_m) = \frac{L(p_0 | y_1, \dots, y_m)}{L(y/m | y_1, \dots, y_m)} = \frac{p_0^y (1-p_0)^{m-y}}{(y/m)^y (1-y/m)^{m-y}},$$

where y_1, \dots, y_m are the observed values of Y_1, \dots, Y_m , and $y = \sum_{i=1}^m y_i$ is the outcome of $N = \sum_{i=1}^m Y_i$. Under mild regularity conditions, the asymptotic distribution of the log-likelihood ratio statistic $-2 \log(\Lambda(p_0)) = -2 \log(\Lambda(p_0 | Y_1, \dots, Y_m))$ under H_0 is $\chi_1^2(1-\alpha)$ as m approaches infinity (see [24], Chapter 13, Theorem 6). Thus, if

$$-2 \log(\Lambda(p_0)) > \chi_1^2(1-\alpha),$$

we would reject H_0 that the failure rate of the model is reasonable at level α . Typically, the α is set at 0.05. Therefore, we reject the null hypothesis if

$$-2 \log(\Lambda(p_0)) > \chi_1^2(0.95) \approx 3.841459.$$

In this study, we evaluate the test statistic $-2 \log(\Lambda(p_0))$ for the 0.05, 0.025, 0.01, and 0.005 probability levels and evaluation samples of size 250, 500, 750, and 1000. The acceptable number of failures in a VaR model are displayed in **Table 4**. The VaR model can be rejected when the number of failures is both high and low. If there are too many exceptions, the model underestimates the VaR. On the other hand, if there are too few exceptions, then the model is too conservative and can harm profit opportunities.

7.2. Comparative Study of VaR Computation Methods

We apply the normal, historical simulation, and newly proposed VaR computation methods defined in (1), (2), and (23), respectively, to estimate 1000 daily VaR forecasts from two portfolios. Probability levels of 0.05,

0.025, 0.01, and 0.005 are considered. Each VaR model is estimated using samples of 252, 504, and 1000 trading days. A back-test is then performed to evaluate the adequacy of each VaR model under consideration over an evaluation sample of 1000 trading days.

We have two imaginary investment portfolios each consisting of a single well-known stock index, the Dow Jones Industrial Average (DJIA) and the S&P/TSX Composite Index. These indices were chosen to be in our fictitious portfolios as they have abundant publicly available historical data. Here, they are used as representative stocks since, in reality, an index cannot be invested directly being that it is a mathematical construct. The raw values of the daily DJIA and S&P/TSX Composite indices are displayed in **Figure 8** from June 28, 2007 to March 11, 2015. The effect of the 2008 financial crisis is indicated by both indices, where the DJIA can be seen to have a large decrease in points with a low level of approximately 6500 in the early months of 2009. This is followed by an increase in the level of both indices in the recent years, particularly for the DJIA.

The index values are used to evaluate the daily logarithmic returns as follows. If P_{t+1} and P_t are the index values at time $t + 1$ and t , respectively, then the return X_{t+1} at time $t + 1$ is given by

Table 4. 95% nonrejection confidence regions for the likelihood ratio test under different VaR confidence levels and evaluation sample sizes.

Probability Level p_0	VaR Confidence Level	Nonrejection Region for Number of Failures N			
		$m = 250$ days	$m = 500$ days	$m = 750$ days	$m = 1000$ days
0.05	95.0%	$7 \leq N \leq 19$	$17 \leq N \leq 35$	$27 \leq N \leq 49$	$38 \leq N \leq 64$
0.025	97.5%	$3 \leq N \leq 11$	$7 \leq N \leq 19$	$12 \leq N \leq 17$	$16 \leq N \leq 35$
0.01	99.0%	$1 \leq N \leq 6$	$2 \leq N \leq 9$	$3 \leq N \leq 13$	$5 \leq N \leq 16$
0.005	99.5%	$0 \leq N \leq 5$	$1 \leq N \leq 6$	$1 \leq N \leq 8$	$2 \leq N \leq 9$

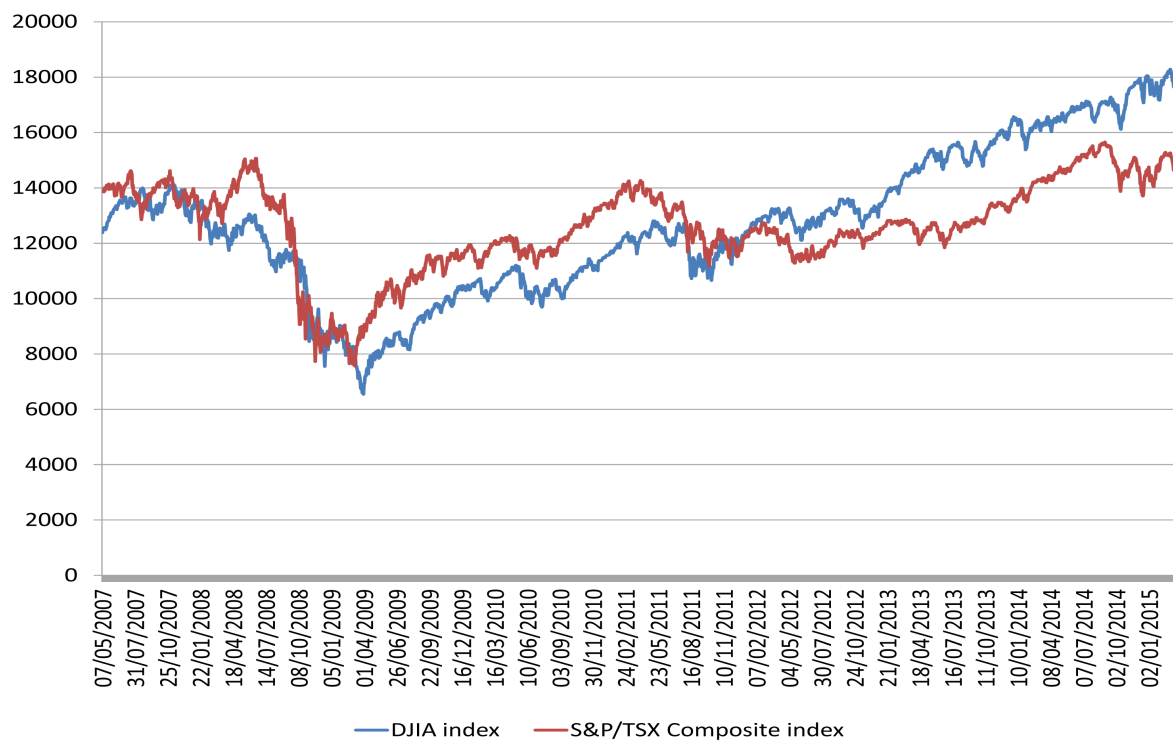


Figure 8. Daily DJIA and S&P/TSX Composite indices from June 28, 2007 through March 11, 2015.

$$X_{t+1} = \log(P_{t+1}/P_t).$$

The autocorrelation of the daily log returns are plotted in **Figure 9** for each index. We can observe that there are no significant autocorrelations as almost all of them fall within the 95% confidence limits. A few lags slightly outside of the limits do not necessarily indicate non-randomness as this can be expected due to random fluctuations. In addition, there is absence of a pattern. Therefore, both portfolios may be considered random, and thus all the VaR computation methods in hand may be applied.

The daily log returns and VaR estimates for every model under consideration are displayed in **Figure 10** and **Figure 11** for each stock index over a time period of one thousand trading days. Each row of plots corresponds to the VaR confidence level, while each column provides the results of the estimation sample used. **Table 5** displays the back-test results of all the VaR models in question for each stock market index. The outcome of each test, that is whether or not to reject the model given the observed number of VaR violations, is reported for every VaR model. These outcomes are determined by the 95% nonrejection regions indicated in **Table 4** when the evaluation sample size is 1000 days.

Table 5. Back-test results of all the VaR models under consideration applied to each stock index over an evaluation sample of 1000 days and 95% confidence regions.

Probability Level	No. Exceptions/No. Obs.			Test Outcome		
	normal	historical	Fejér-type	normal	historical	Fejér-type
DJIA Portfolio						
0.005	20/252	8/252	5/252	Reject	-	-
0.01	26/252	10/252	10/252	Reject	-	-
0.025	40/252	30/252	29/252	Reject	-	-
0.05	57/252	57/252	57/252	-	-	-
0.005	14/504	3/504	3/504	Reject	-	-
0.01	23/504	11/504	10/504	Reject	-	-
0.025	35/504	24/504	25/504	-	-	-
0.05	47/504	49/504	47/504	-	-	-
0.005	3/1000	0/1000	0/1000	-	Reject	Reject
0.01	4/1000	1/1000	1/1000	Reject	Reject	Reject
0.025	14/1000	6/1000	6/1000	Reject	Reject	Reject
0.05	24/1000	23/1000	23/1000	Reject	Reject	Reject
S&P/TSX Composite Index Portfolio						
0.005	22/252	11/252	9/252	Reject	Reject	-
0.01	29/252	15/252	15/252	Reject	-	-
0.025	51/252	38/252	34/252	Reject	Reject	-
0.05	63/252	59/252	59/252	-	-	-
0.005	16/504	7/504	7/504	Reject	-	-
0.01	22/504	12/504	12/504	Reject	-	-
0.025	40/504	30/504	28/504	Reject	-	-
0.05	59/504	51/504	51/504	-	-	-
0.005	5/1000	0/1000	0/1000	-	Reject	Reject
0.01	6/1000	3/1000	3/1000	-	Reject	Reject
0.025	12/1000	6/1000	6/1000	Reject	Reject	Reject
0.05	23/1000	21/1000	20/1000	Reject	Reject	Reject

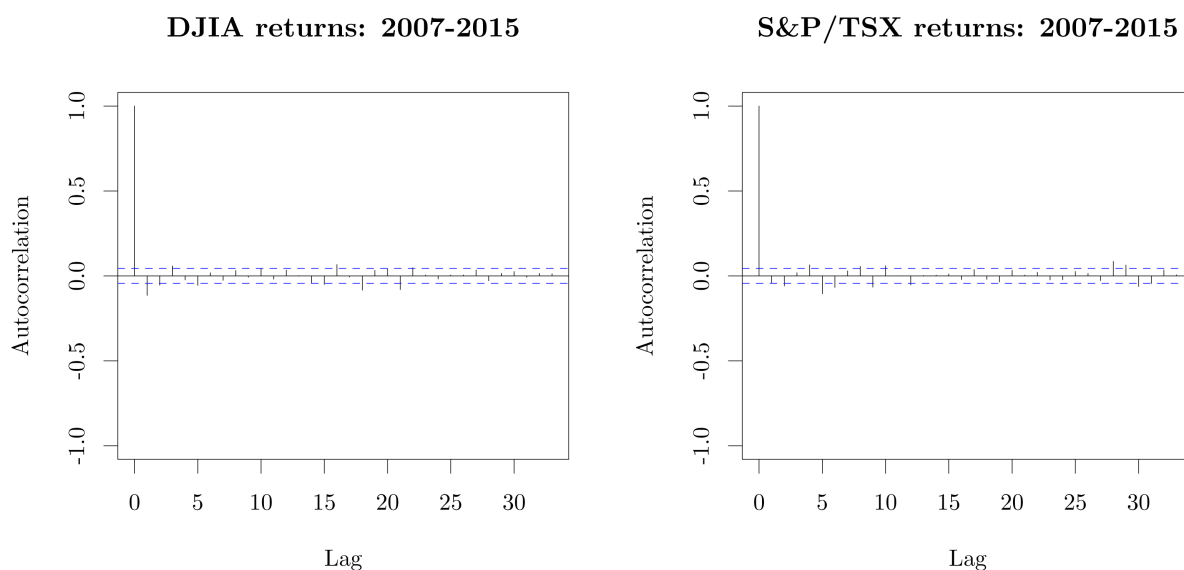


Figure 9. Autocorrelation plots of the daily logarithmic returns for each stock market index.

The following can be observed from the aforementioned figures and tables. Overall, the empirical results of both stock indices are fairly similar. The back-test results in [Table 5](#) show that the normal model has the poorest performance as it is not considered adequate in most cases. The observed number of VaR violations is quite high for smaller probability levels, meaning that the mass in the tails of the distribution is underestimated. The only case when the normal model is not consistently rejected is when the probability level is 0.05 for estimation samples of 252 and 504 observations. The historical simulation method generally performs well and shares similar results with the newly proposed VaR estimation method. It is, however, rejected for probability levels 0.005 and 0.025 when the number of observations is 252 in the S&P/TSX Composite Index portfolio. Finally, the VaR model of interest based on the Fejér-type kernel quantile estimation is the most reliable as it has the least number of rejections for all the tests considered.

Overall, it can be seen that none of the models perform well when the estimation sample is large, except for sometimes the normal method when probability levels are small. This is expected as financial data from four years ago may no longer be relevant to the current market situation. Moreover, the performance of all the VaR computation methods is similar at the 95% confidence level.

For an illustration of the density of portfolio returns on a specific day see [Figure 12](#) and [Figure 13](#). The Fejér-type kernel density estimates with Fourier bandwidths are represented by the green curves while the normal densities have the red curves. The images are consistent with the assertion that the stock returns are heavy tailed. It can be clearly seen that the density estimates with Fejér-type kernels can account for heavy tails of the return distributions better than the normal densities.

In summary, the proposed method for computing the VaR based on density estimation with Fejér-type kernel functions and Fourier analysis bandwidth selectors provides more reliable results than the commonly used VaR computation methods. Density estimates with Fejér-type kernel functions can account for the heavy tails of the return distributions, unlike the normal density. The normal method for computing the VaR tends to underestimate the risk, especially for higher confidence levels. For the nonparametric models, one has to be careful in choosing a relevant estimation period; otherwise, they tend to overestimate the risk for large estimation samples.

8. Conclusion

The paper introduces a nonparametric method of VaR computation on portfolio returns. The approach relies on the kernel quantile estimator introduced by Parzen [22]. The kernel functions employed are Fejér-type kernel functions. We use these functions because they are known to produce asymptotically efficient kernel density estimators with respect to the \mathbb{L}_2 -risk. A simulation study in support of this theoretical result is first conducted, and a new VaR estimator is then introduced. In the simulation study, several bandwidths are used, including the

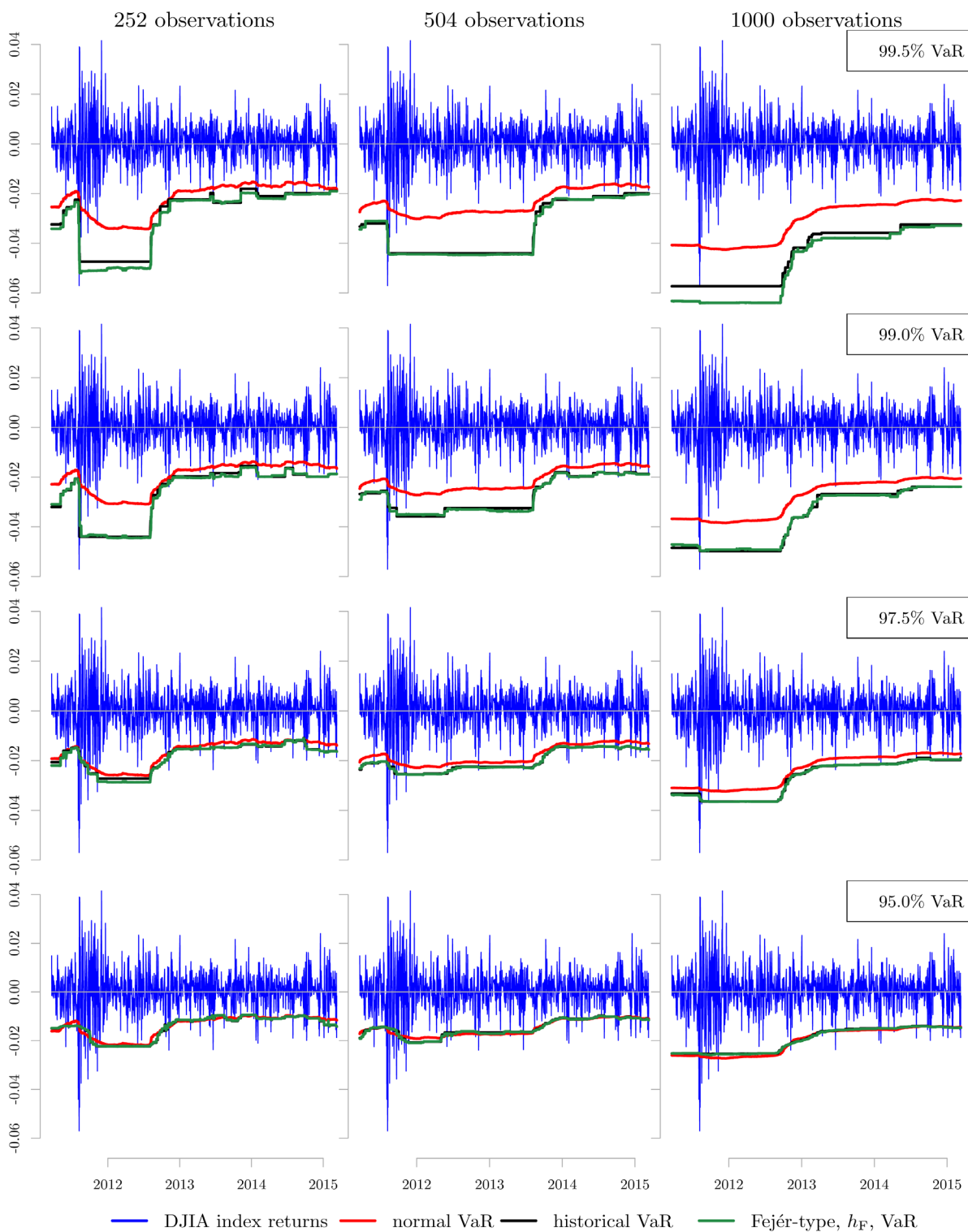


Figure 10. Daily log returns and VaR estimates of the DJIA at 95%, 97.5%, 99%, and 99.5% confidence levels under 252, 504, and 1000 observations over 1000 trading days for the VaR computation methods in consideration.

data-driven bandwidth obtained from the Fourier analysis of a kernel density estimator. The latter bandwidth is chosen for constructing the new VaR estimator, $\widehat{\text{VaR}}_{3,n}(p)$, based on the analytical arguments and obtained

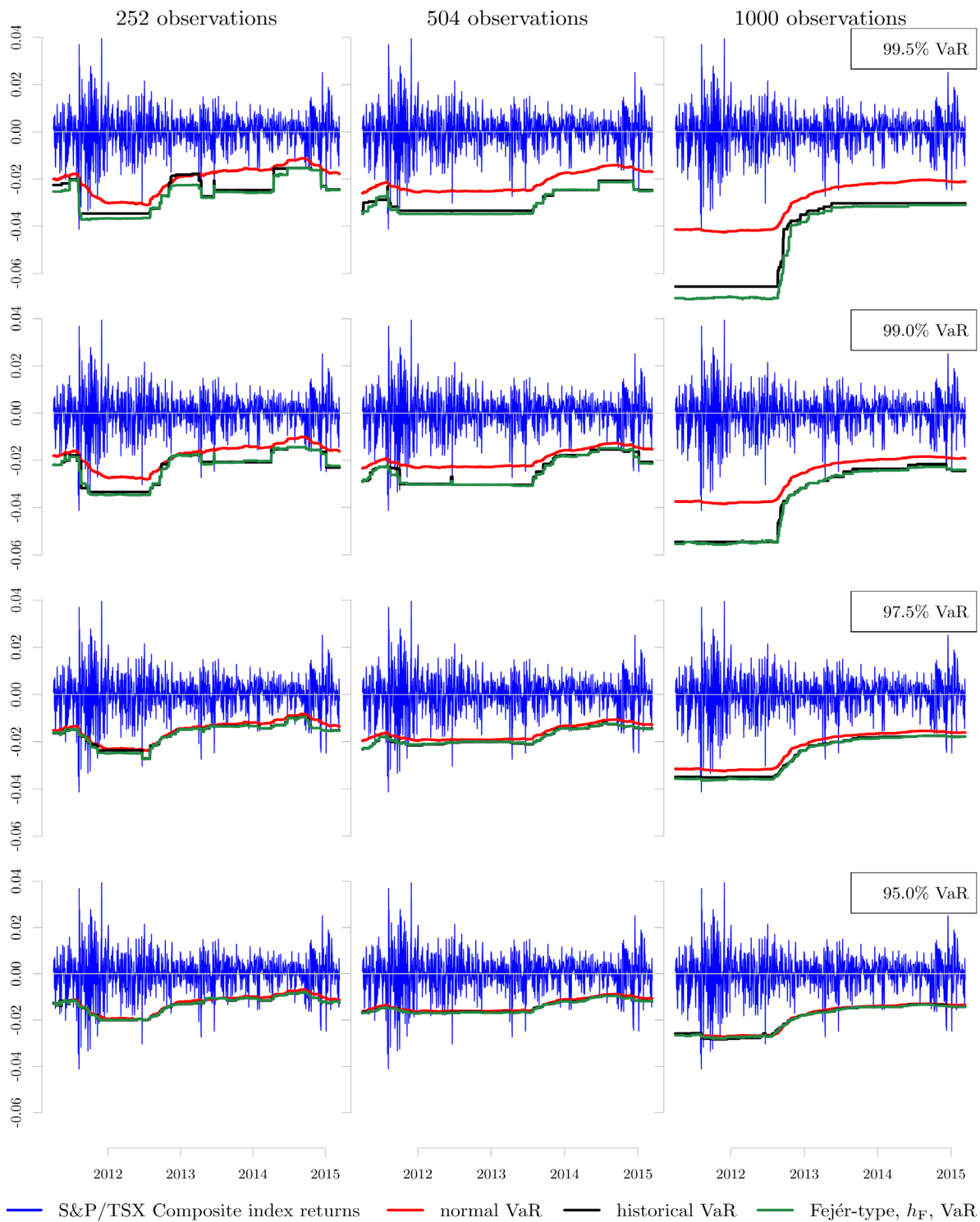


Figure 11. Daily log returns and VaR estimates of the S&P/TSX Composite Index at 95%, 97.5%, 99%, and 99.5% confidence levels under 252, 504, and 1000 observations over 1000 trading days for the VaR computation methods in consideration.

numerical results. The resulting estimator is compared numerically with the two standard VaR estimators, $\widetilde{\text{VaR}}_{1,n}(p)$ and $\widetilde{\text{VaR}}_{2,n}(p)$, and is found to be more reliable. The proposed method of VaR computation is

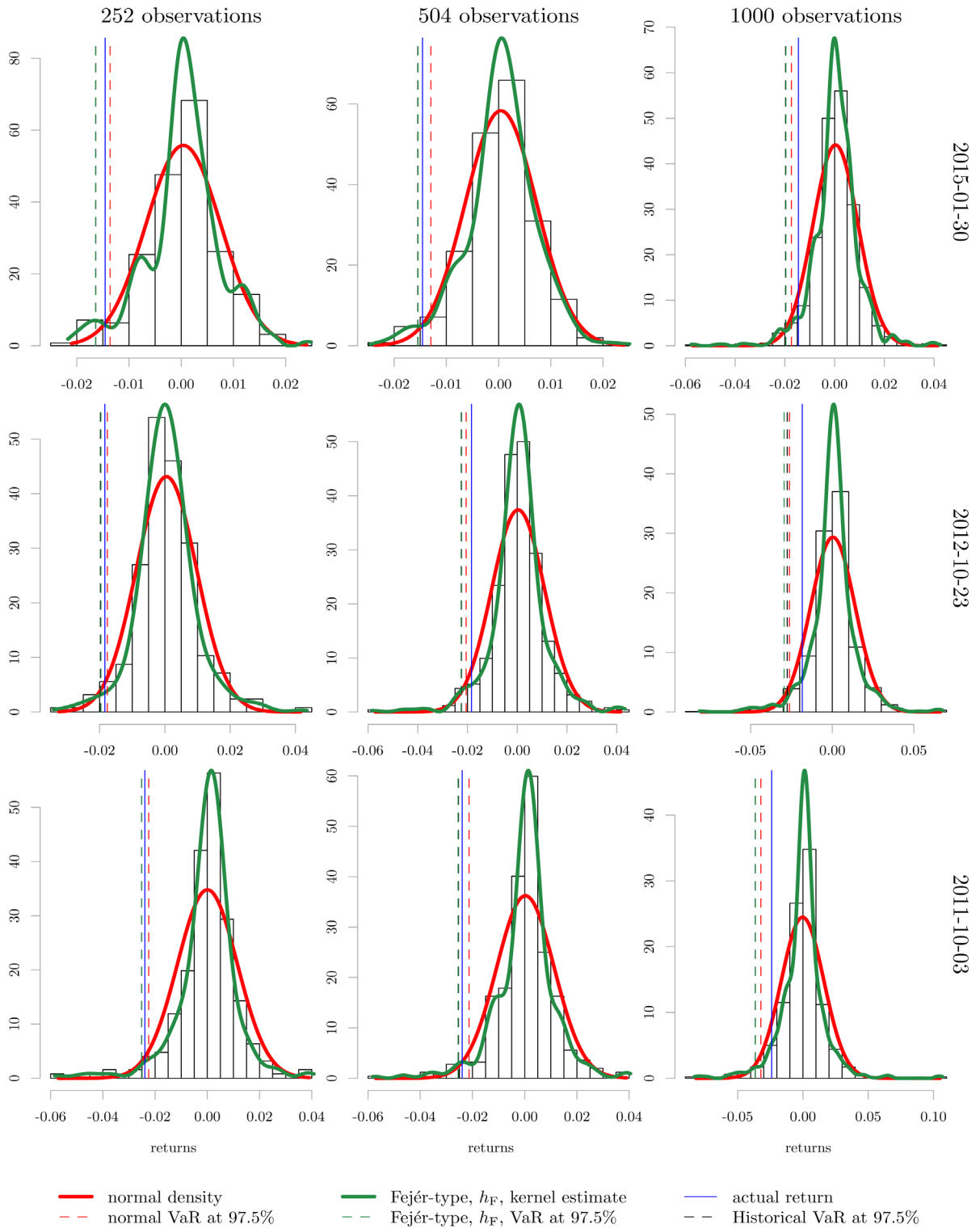


Figure 12. Normal, empirical, and positive part Fejér-type kernel densities of the DJIA daily returns based on 252, 504, and 1000 observations for the days 29/06/2011, 25/06/2013, and 30/09/2014. The 97.5% VaR of each model is illustrated along with the actual daily return.

convenient for practitioners because it does not require restrictive assumptions on the underlying distribution, as

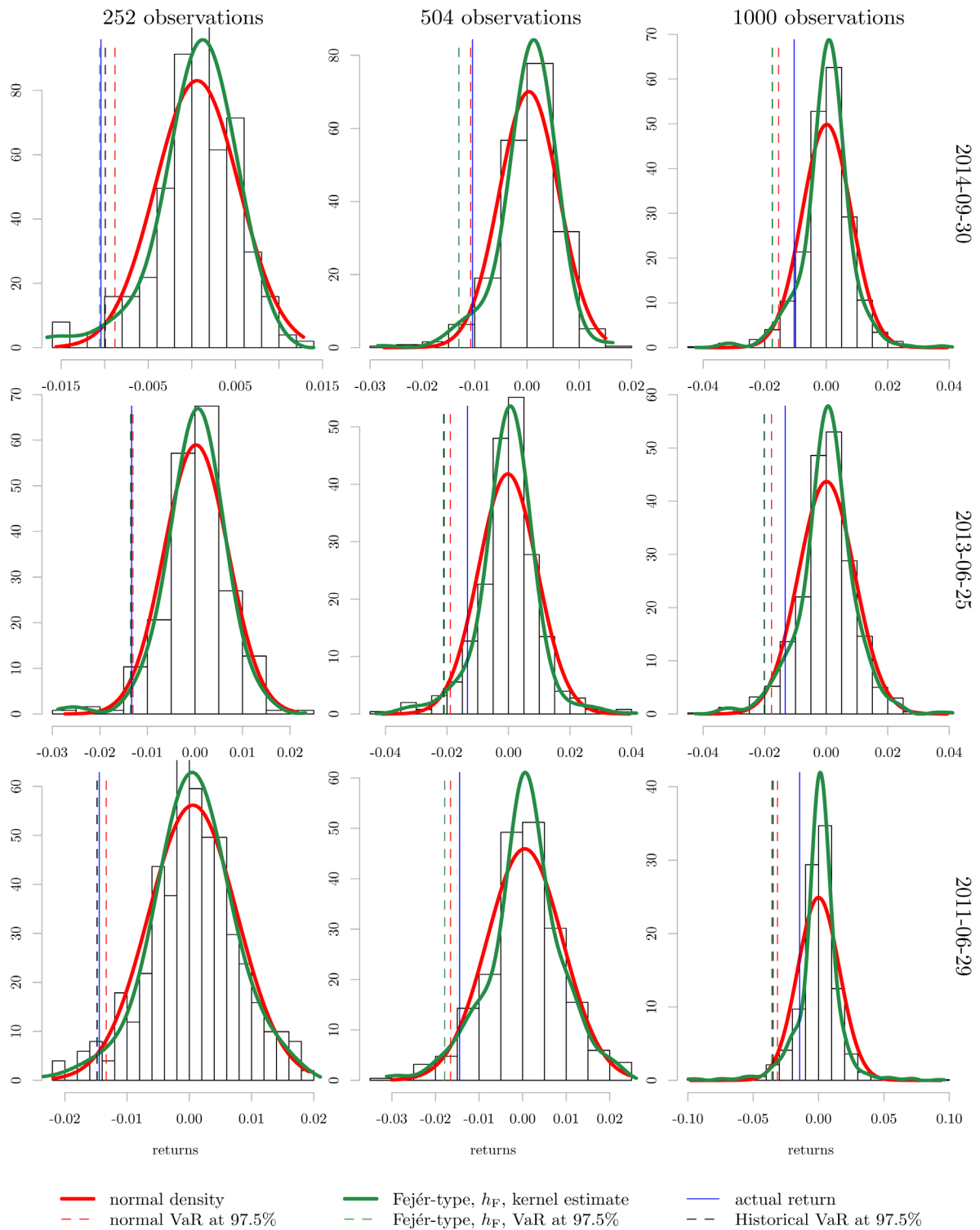


Figure 13. Normal, empirical, and positive part Fejér-type kernel densities of the S&P/TSX Composite daily returns based on 252, 504, and 1000 observations for the days 29/06/2011, 25/06/2013, and 30/09/2014. The 97.5% VaR of each model is illustrated along with the actual daily return.

the normal method does. Our method also provides more accurate VaR estimates than the historical simulation method due to its smooth structure.

Acknowledgements

This research was supported by an NSERC grant held by Natalia Stepanova at Carleton University.

References

- [1] Jorion, P. (2001) Value at Risk: The New Benchmark for Managing Financial Risk. 2nd Edition, McGraw-Hill, United States of America
- [2] Stepanova, N. (2013) On Estimation of Analytic Density Functions in L_p . *Mathematical Methods of Statistics*, **22**, 114-136. <http://dx.doi.org/10.3103/S1066530713020038>
- [3] Levit, B. and Stepanova, N. (2004) Efficient Estimation of Multivariate Analytic Functions in Cube-Like Domains. *Mathematical Methods of Statistics*, **13**, 253-281.
- [4] Golubev, G.K., Levit, B.Y. and Tsybakov, A.B. (1996) Asymptotically Efficient Estimation of Analytic Functions in Gaussian Noise. *Bernoulli*, **2**, 167-181. <http://dx.doi.org/10.2307/3318549>
- [5] Guerre, E. and Tsybakov, A.B. (1998) Exact Asymptotic Minimax Constants for the Estimation of Analytic Functions in L_p . *Probability Theory and Related Fields*, **112**, 33-51. <http://dx.doi.org/10.1007/s004400050182>
- [6] Schipper, M. (1996) Optimal Rates and Constants in L_2 -Minimax Estimation. *Mathematical Methods of Statistics*, **5**, 253-274.
- [7] Hájek, J. (1972) Local Asymptotic Minimax and Admissibility in Estimation. *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, 175-194.
- [8] Belitser, E. (1998) Efficient Estimation of Analytic Density under Random Censorship. *Bernoulli*, **4**, 519-543. <http://dx.doi.org/10.2307/3318664>
- [9] Ibragimov, I.A. and Hasminskii, R.Z. (1983) On Estimation of the Density Function. *Journal of Soviet Mathematics*, **25**, 40-57. <http://dx.doi.org/10.1007/BF01091455>
- [10] Rudin, W. (1987) Real and Complex Analysis. 3rd Edition, McGraw-Hill, Singapore.
- [11] Tsybakov, A.B. (2009) Introduction to Nonparametric Estimation. Springer Science, United States of America. <http://dx.doi.org/10.1007/b13794>
- [12] Kosta, O. (2015) Efficient Density Estimation Using Fejér-Type Kernel Functions. M.Sc. Thesis, Carleton University, Ottawa.
- [13] Cline, D.B.H. (1988) Admissible Kernel Estimators of a Multivariate Density. *Annals of Statistics*, **16**, 1421-1427. <http://dx.doi.org/10.1214/aos/1176351046>
- [14] Härdle, W., Müller, M., Sperlich, S. and Weratz, A. (2004) Nonparametric and Semiparametric Models. Springer, Heidelberg. <http://dx.doi.org/10.1007/978-3-642-17146-8>
- [15] Rudemo, M. (1982) Empirical Choice of Histograms and Kernel Density Estimators. *Scandinavian Journal of Statistics*, **9**, 65-78.
- [16] Scott, D.W. and Terrell, G.R. (1987) Biased and Unbiased Cross-Validation in Density Estimation. *Journal of the American Statistical Association*, **82**, 1131-1146. <http://dx.doi.org/10.1080/01621459.1987.10478550>
- [17] Golubev, G.K. (1992) Nonparametric Estimation of Smooth Densities of a Distribution in L_2 . *Problems of Information Transmission*, **23**, 57-67.
- [18] Achieser, N.I. (1956) Theory of Approximation. Frederick Ungar Publishing, New York.
- [19] Golubev, G.K. and Levit, B.Y. (1996) Asymptotically Efficient Estimation for Analytic Distributions. *Mathematical Methods of Statistics*, **5**, 357-368.
- [20] Mason, D.M. (2010) Risk Bounds for Kernel Density Estimators. *Journal of Mathematical Sciences*, **163**, 238-261. <http://dx.doi.org/10.1007/s10958-009-9671-0>
- [21] Lepski, O.V. and Levit, B.Y. (1998) Adaptive Minimax Estimation of Infinitely Differentiable Functions. *Mathematical Methods of Statistics*, **7**, 123-156.
- [22] Parzen, E. (1979) Nonparametric Statistical Data Modeling. *Journal of the American Statistical Association*, **74**, 105-121. <http://dx.doi.org/10.1080/01621459.1979.10481621>
- [23] Marron, J.S. and Sheather, S.J. (1990) Kernel Quantile Estimators. *Journal of the American Statistical Association*, **85**, 410-416. <http://dx.doi.org/10.1080/01621459.1990.10476214>
- [24] Roussas, G. (1997) A Course in Mathematical Statistics. 2nd Edition, Academic Press, United States of America.