

Efficient Density Estimation using Fejér-Type Kernel Functions

by

Olga Kosta

A Thesis submitted to
the Faculty of Graduate Studies and Postdoctoral Affairs
in partial fulfilment of
the requirements for the degree of
Master of Science

in

Mathematics

Ottawa-Carleton Institute for
Mathematics and Statistics

Carleton University
Ottawa, Ontario, Canada

© 2015, Olga Kosta

Abstract

The problem of estimating an unknown probability density function (pdf) is of fundamental importance in statistics and required for many statistical applications. In recent years, efficient nonparametric estimation has had greater focus on the problem of nonparametric regression, while the more challenging problem of density estimation has been given much less attention. In this thesis, we consider a class of kernel-type density estimators with Fejér-type kernels and theoretical smoothing parameters $h_n = (2\gamma\theta_n)/\log n$, where the parameter $\gamma > 0$ describes the class of underlying pdfs and $0 \leq \theta_n < 1$. In theory, the estimator under consideration dominates in \mathbb{L}_p , $1 \leq p < \infty$, all other known estimators from the literature in the locally asymptotic minimax (LAM) sense. We demonstrate via simulations that the estimator in question is good by comparing its performance to other fixed kernel estimators. The kernel-type estimator is also studied under empirical bandwidth selection methods such as the common cross-validation and the less-known method based on the Fourier analysis of kernel density estimators. The common \mathbb{L}_2 -risk is used to assess the quality of estimation. The estimator of interest is then tried to real financial data for a risk measure that is widely used in many applications. The simulation results testify that, for a good choice of γ , the theoretical estimator under study provides very good finite sample performance compared to the other kernel estimators. The study also suggests that the bandwidth obtained by using the Fourier analysis techniques performs better than the one from cross-validation in most settings.

Acknowledgements

I would like to first thank my supervisor, Dr. Natalia Stepanova, for all of her help during my research and for introducing me to the interesting world of nonparametric statistics. Her knowledge, guidance, and support has been very valuable to me and has assisted me throughout my studies in the Master of Science program at Carleton University.

I would also like to thank the employees at the Bank of Canada for providing me with a practical example in which to apply my research.

I am especially grateful to my family and my boyfriend Eric for all of their love and encouragement throughout my life.

Dedication

I would like to dedicate this thesis to the loving memory of my father whose passion for Mathematics has been the inspiration that has led me to study one of its many realms.

Contents

List of Tables	vii
List of Figures	ix
Nomenclature	xi
Introduction	1
1 Problem Statement and Motivation	4
2 Kernel Density Estimators	9
2.1 Performance Criterion	14
2.1.1 MISE Evaluation under the Classical Approach	15
2.1.2 MISE Evaluation using Fourier Analysis Techniques	17
2.2 Kernel Functions	22
2.3 Bandwidth Selection	28
2.3.1 Unbiased Cross-Validation Method	30
2.3.2 Unbiased Risk Estimation based on the Fourier Analysis of Kernel Density Estimators	32

3	Density Estimators with Fejér-Type Kernels	35
3.1	Choice of Smoothing Parameter γ	39
3.2	Evaluation of the Unbiased Risk Estimators	41
3.2.1	Unbiased Cross-Validation Criterion	42
3.2.2	Unbiased Risk Estimator based on the Fourier Analysis of Kernel Density Estimators	46
4	Simulation Studies	50
4.1	Methodology	51
4.2	Numerical Results	55
5	Value at Risk using Density Estimators with Fejér-type Kernels	66
5.1	Evaluation of VaR for One Stock	67
5.1.1	Models for Estimating VaR	68
5.1.2	Model Evaluation	73
5.2	Results	76
	Conclusions	86
	APPENDICES	89
	A Definitions and Theorems	90
	B R code	95
	C Simulation Results	102
	References	111

List of Tables

2.1	<i>Some common kernel functions and their Fourier transforms</i>	24
3.1	<i>Cases of the Fejér-type kernel</i>	37
5.1	<i>95% nonrejection confidence regions for the likelihood ratio test under different VaR confidence levels and evaluation sample sizes</i>	76
5.2	<i>Backtest results for all the VaR models under consideration</i>	79
C.1	<i>Estimated values for the MISE of density estimates with bandwidths \hat{h}_{CV} and \hat{h}_F that use the Fejér-type, sinc, dlVP, and Gaussian kernels based on the realizations generated from $N(0, 1)$ for a variety of sample sizes</i>	103
C.2	<i>Estimated values for the MISE of density estimates with the theoretical bandwidth h_n that use the Fejér-type, sinc, and dlVP kernels based on the realizations generated from $N(0, 1)$ for a variety of sample sizes</i>	104
C.3	<i>Estimated values for the MISE of density estimates with bandwidths \hat{h}_{CV} and \hat{h}_F that use the Fejér-type, sinc, dlVP, and Gaussian kernels based on the realizations generated from t_{15} for a variety of sample sizes</i>	105
C.4	<i>Estimated values for the MISE of density estimates with the theoretical bandwidth h_n that use the Fejér-type, sinc, and dlVP kernels based on the realizations generated from t_{15} for a variety of sample sizes</i>	106

C.5	<i>Estimated values for the MISE of density estimates with bandwidths \hat{h}_{CV} and \hat{h}_F that use the Fejér-type, sinc, dlVP, and Gaussian kernels based on the realizations generated from χ_4^2 for a variety of sample sizes</i>	107
C.6	<i>Estimated values for the MISE of density estimates with the theoretical bandwidth h_n that use the Fejér-type, sinc, and dlVP kernels based on the realizations generated from χ_4^2 for a variety of sample sizes</i>	108
C.7	<i>Estimated values for the MISE of density estimates with bandwidths \hat{h}_{CV} and \hat{h}_F that use the Fejér-type, sinc, dlVP, and Gaussian kernels based on the realizations generated from $0.4N(0, 1) + 0.6N(1, 0.4^2)$ for a variety of sample sizes</i>	109
C.8	<i>Estimated values for the MISE of density estimates with the theoretical bandwidth h_n that use the Fejér-type, sinc, and dlVP kernels based on the realizations generated from $0.4N(0, 1) + 0.6N(1, 0.4^2)$ for a variety of sample sizes</i>	110

List of Figures

2.1	<i>Comparison of three kernel density estimates with different smoothing bandwidths</i>	13
2.2	<i>Some common kernel functions and their Fourier transforms</i>	25
3.1	<i>Fejér-type kernel and its Fourier transform</i>	39
3.2	<i>Fourier transform of the self-convolution of the Fejér-type kernel</i>	45
4.1	<i>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_4^2, and normal mixture, for small sample sizes.</i>	60
4.2	<i>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_4^2, and normal mixture, for large sample sizes.</i>	61
4.3	<i>MISE estimates for the Fejér-type, sinc, dlVP, and Gaussian kernels with cross-validation, Fourier, and theoretical bandwidth selectors that estimate a standard normal, Student's t_{15}, chi-square χ_4^2, and normal mixture, for small sample sizes.</i>	62

4.4	<i>MISE estimates for the Fejér-type, sinc, dIVP, and Gaussian kernels with cross-validation, Fourier, and theoretical bandwidth selectors that estimate a standard normal, Student's t_{15}, chi-square χ_4^2, and normal mixture, for large sample sizes.</i>	63
4.5	<i>Optimal values of θ_n for the Fejér-type kernel with the cross-validation, Fourier, and theoretical bandwidths when estimating the standard normal, Student's t_{15}, chi-square χ_4^2, and normal mixture for different sample sizes</i>	64
4.6	<i>Kernel density estimation of a standard normal, Student's t_{15}, chi-square χ_4^2, and normal mixture for different sample sizes</i>	65
5.1	<i>DJIA and S&P/TSX Composite indices from June 28, 2007 through March 11, 2015</i>	77
5.2	<i>Daily log returns and VaR estimates of the DJIA</i>	82
5.3	<i>Daily log returns and VaR estimates of the S&P/TSX Composite Index</i>	83
5.4	<i>Normal, empirical, and Fejér-type kernel densities of the DJIA daily returns</i>	84
5.5	<i>Normal, empirical, and Fejér-type kernel densities of the S&P/TSX Composite daily returns</i>	85

Nomenclature

Abbreviations

a.s.	almost surely
cdf	cumulative distribution function
CMC	crude Monte Carlo
CV	cross-validation
DJIA	Dow Jones Industrial Average
dLVP	de la Vallée Poussin
i.i.d.	independent and identically distributed
ISE	integrated squared error
LAM	locally asymptotically minimax
MISE	mean integrated squared error
pdf	probability density function
pmf	probability mass function
VaR	value at risk

Notation

\mathbb{C}	set of the complex numbers
$\overline{f(\cdot)}$	the complex conjugate of function $f(\cdot)$
$\xrightarrow{a.s.}$	convergence almost surely
\xrightarrow{P}	convergence in probability
$\mathbb{E}_f(\cdot)$	expectation with respect to the distribution with pdf $f(\cdot)$
$\Gamma(\cdot)$	the Gamma function
$\hat{g}(\cdot)$	Fourier transform of a function $g(\cdot)$
$\mathbb{I}(A)$	indicator function of the set A
i	imaginary unit, $i = \sqrt{-1}$
$\text{Leb}(A)$	Lebesgue measure of the set A
$\lim_{n \rightarrow \infty} a_n = a$	limit of the sequence a_n is a
$a_n \sim b_n$	means $\lim_{n \rightarrow \infty} a_n/b_n = 1$
$\liminf_{n \rightarrow \infty} a_n$	limit inferior of the sequence a_n , $\sup_{n \geq 1} \inf_{k \geq n} a_k$
$\limsup_{n \rightarrow \infty} a_n$	limit superior of the sequence a_n , $\inf_{n \geq 1} \sup_{k \geq n} a_k$
$\log a$	the natural logarithm with base e of the number a
$\mathbb{L}_p(\mathbb{R})$	the space of the p th power integrable functions on \mathbb{R} for $1 \leq p < \infty$
$\ \cdot\ _p$	the $\mathbb{L}_p(\mathbb{R})$ -norm for $1 \leq p < \infty$
$N(\mu, \sigma^2)$	normal distribution with mean μ and variance σ^2
$f = O(g)$	“big O ” means: $f(x)/g(x) \rightarrow C$, constant $C \neq 0$, as $x \rightarrow a$

$f = o(g)$	“small o ” means: $f(x)/g(x) \rightarrow 0$ as $x \rightarrow a$
$\Phi(\cdot)$	cdf of the standard normal distribution
$\phi(\cdot)$	pdf of the standard normal distribution
\mathbb{R}	set of the real numbers
$\mathcal{S}(\beta, B)$	Sobolev ball of radius B of β -smooth functions on \mathbb{R}
t_ν	Student’s t distribution with ν degrees of freedom
$\varphi(\cdot)$	characteristic function
χ_ν^2	chi-square distribution with ν degrees of freedom
$\lfloor x \rfloor$	the largest integer not greater than x
$X_{(k)}$	the k^{th} order statistic
\mathbb{Z}_+	set of the positive integers

Introduction

The problem of estimating an unknown probability density function (pdf) is of fundamental importance in statistics. Let X_1, X_2, \dots be a sequence of independent and identically distributed (i.i.d.) real-valued random variables from an absolutely continuous distribution with an unknown pdf f on \mathbb{R} . To understand the behaviour of the distribution of these random variables, f needs to be specified. As f cannot be precisely reconstructed from the data, one can try to estimate f using the observed values x_1, \dots, x_n of a random sample X_1, \dots, X_n of size $n \geq 1$. The idea behind density estimation is to construct 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 . Unlike a similar problem of estimating an unknown function in a regression model with normal errors, the problem of density estimation is much more challenging in view of the lack of the normality assumption on the observations. The techniques used for density estimation can be classified as parametric and nonparametric.

Traditional parametric methods for density estimation, dating back to Fisher, assume the random sample belongs to a *known* parametric family of distributions. Density estimation is thus equivalent to that of the finite-dimensional parameter. Parametric methods are generally more efficient and powerful than their nonparametric counterparts if the assumption of the underlying distribution is valid. The parametric form is often presupposed by experience and may not always be available in practice. If there is insufficient prior knowledge about the random sample, then parametric estimation is not always appropriate

as misspecified forms of f may induce invalid inference rendering the results meaningless.

Nonparametric estimation makes much less rigid assumptions about the distribution of the random sample. It simply assumes that the distribution of the random variables has a density function belonging to some sufficiently large class \mathcal{F} of functions. Nonparametric methods then consist of estimating, from the observations, the unknown function. The fact that the estimate of f is not constrained under an assumed parametric form makes the resulting inference more universal and robust. It is the data, not the experimenter as in the parametric case, that provides the information about the estimate of f . Therefore, the validity of the statistical inference will not be challenged by questioning whether the data actually follows the assumed distribution.

Nonparametric density estimation typically involves histogram-based or data-smoothing methods. The histogram is the oldest and most widely used nonparametric density estimator. It is a step-like discontinuous function of the number of observations that fall into different bins. The rough discontinuous form of the histogram makes it difficult to properly capture all of the characteristics of the underlying density and work with when derivatives are required. A generalized and more sophisticated version of the histogram is the kernel density estimator. It uses a kernel function K and a smoothing parameter h to construct a smooth density function that more accurately reflects the underlying distribution. 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 we call the corresponding estimator the *kernel-type estimator*. This thesis examines a specific class of kernel-type estimators, with the Fejér-type kernel and the common theoretical smoothing parameter, that are theoretically known to dominate in \mathbb{L}_p , $1 \leq p < \infty$, all other estimators in the literature in the locally asymptotic minimax sense.

The thesis is organised as follows. Chapter 1 provides a brief overview of previous research on kernel density estimation followed by the motivation behind studying Fejér-type kernel estimators. Chapter 2 comprises a detailed description of the theory surrounding a

kernel density estimator, how it is measured, its parameters, and its properties. The kernel-type density estimator in question is defined in Chapter 3 along with its basic properties. The simulation study supporting the theoretical finding that the estimator under study is good is found in Chapter 4. In Chapter 5, we use real financial data to illustrate the advantages the kernel-type estimator may serve in practice to a common risk measure. The results of the study are summarised in the Conclusions.

Chapter 1

Problem Statement and Motivation

Let X_1, X_2, \dots be a sequence of i.i.d. real-valued random variables with a common density f , where f belongs to a suitable family \mathcal{F} of densities, and let $f_n(x) = f_n(x; X_1, \dots, X_n)$ be an estimator of f at point $x \in \mathbb{R}$. Let f belong to the function space $\mathbb{L}_p = \mathbb{L}_p(\mathbb{R})$ with the norm $\|f\|_p = \left(\int_{\mathbb{R}} |f(x)|^p dx\right)^{1/p} < \infty$, $1 \leq p < \infty$. One problem of interest is to construct an asymptotically efficient estimator f_n of f in \mathbb{L}_p , $1 \leq p < \infty$, by selecting the appropriate kernel K and bandwidth h . In doing so, one often assumes that both K and h depend on the sample size n .

The notion of the kernel density estimator, initiated by Rosenblatt [25] and Parzen [22] in the late 1950s, has stimulated substantial interest in the area of nonparametric estimation. Statisticians have, since then, continuously strived to find the best K and h for constructing a good f_n of the true f in terms of some predetermined criterion.

There exist common misconceptions surrounding the optimal choice of K . For starters, nonnegative kernels are often used due to the argument that a density estimator inherits the properties of its corresponding kernel. Nonnegativity of density estimators, however, is not necessarily achieved through nonnegative kernels. Golubev and Levit [8] and Hasminskii and Ibragimov [14] have provided examples of optimal density estimators with kernels that do not have the nonnegative property. Furthermore, many believe the Epanechnikov kernel

to be an optimal choice of K due to an optimization argument derived by Bartlett [2] and Epanechnikov [6] in the early sixties. It has more recently been brought to light, as expressed in Section 1.2.4 of Tsybakov [34], that this is not a consistent concept of optimality since it is based on the minimisation of the asymptotic risk, with respect to $K \geq 0$, for a fixed f . In our study, we compare the behaviour of the estimators over a class of densities instead of for a fixed density. Finally, it is generally viewed that the choice of K is considered less crucial for the quality of the estimator than the choice of h ; yet, we will soon see that this is not the case when we attempt to construct asymptotically efficient estimators.

In statistical literature, the efficiency of a density estimator is often measured by a *minimax criterion*, which assesses the quality of estimation so that statisticians may protect themselves from the worst that can happen. For a given *loss function* $L(f, f_n)$, define the *risk* of f_n by

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

The *minimax risk* is then given by

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

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

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

and we speak of the \mathbb{L}_p -*risk*. Note that the \mathbb{L}_2 -risk as above with $l(x) = x^2$ is simply the

mean integrated squared error (MISE) defined as

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

In the nonparametric context, exact minimaxity is rarely achievable. Instead, an asymptotic approach, as the sample size grows indefinitely, to minimax estimation is widely used.

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

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

where the notation $a_n \sim b_n$ means $\lim_{n \rightarrow \infty} a_n/b_n = 1$. That is, for large sample sizes, the maximum risk of f_n^* over the class \mathcal{F} of estimated density functions is the smallest possible among the risks of all estimators of f based on X_1, \dots, X_n . In general, the problem of constructing asymptotically minimax estimators of f from some functional class \mathcal{F} is a difficult one; therefore, it remains somewhat unexplored. Instead, a large portion of the literature focuses on a less precise problem known as *rate optimality*. Here, one claims the existence of positive constants $0 < \alpha < \beta < \infty$ such that for some sequence $a_n \rightarrow 0$ as $n \rightarrow \infty$

$$\alpha \leq \liminf_{n \rightarrow \infty} a_n^{-1} \mathcal{R}_n(\mathcal{F}) \leq \limsup_{n \rightarrow \infty} a_n^{-1} \mathcal{R}_n(\mathcal{F}) \leq \beta. \quad (1.1)$$

The sequence a_n in relation (1.1) describes the *optimal rate of convergence* of estimators on \mathcal{F} . An estimator f_n^* satisfying

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

where the notation $a_n \asymp b_n$ means that there exist constants $0 < c < C < \infty$ and a number n^* such that $c < a_n/b_n < C$ for all $n \geq n^*$, is called a *rate optimal estimator* on \mathcal{F} .

There has been considerable work done for asymptotic minimaxity and rate optimality in nonparametric regression analysis. Refer to [9], [11], and [18] for instances of asymptotically minimax estimators of smooth regression curves with respect to the \mathbb{L}_p -risk, $1 \leq p < \infty$. In connection with nonparametric density estimation, this problem is much more difficult and currently only solved for the \mathbb{L}_2 -risk (see [29], Theorem 2). Mason [20] took the initial step towards establishing the problem of asymptotically minimax estimation of $f \in \mathcal{F}$ with general \mathbb{L}_p -norms by providing a partial solution to a conjecture in Remark 5 of [11]; however, the result of Mason is not uniform over the class of density functions, and therefore is not minimax.

Another, more precise, approach for finding efficient estimators is *local asymptotic minimaxity* (see [12] for a more detailed description on the origins of this method). An estimator f_n^* of density $f_0 \in \mathcal{F}$ is called *locally asymptotically minimax* (LAM) if

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

as $n \rightarrow \infty$, where \mathcal{V} is a sufficiently small vicinity of f_0 with an appropriate distance defined on \mathcal{F} . Examples of functions that admit LAM estimators can be found in [3] and [9]. LAM estimators are preferred to asymptotically minimax ones since the former class of estimators cannot be guaranteed to be globally efficient. In connection with estimating density functions, the LAM ideology differs significantly from both asymptotically minimax and rate optimality approaches. The key difference is that, when constructing an LAM estimator of f one has to pay attention to the choice of both h and K . Also, the usual *bias-variance tradeoff* approach, when the variance and the bias terms of an optimal estimator are to be balanced, is no longer appropriate. In several papers concerned with the construction of LAM estimators, efficiency becomes achievable only with a careful choice of the kernel function; and with such a choice of the kernel, the bias of f_n^* in the variance-bias decomposition is shown to be asymptotically negligible to its variance

(see [9], for example).

The problem of LAM density estimation with general \mathbb{L}_p -risks, $1 \leq p < \infty$, has remained largely unsolved. In a recent paper of Stepanova [32], a new kernel-type estimator for densities belonging to a class of infinitely smooth functions is proposed and it is shown that, locally asymptotically, its maximum \mathbb{L}_p -risk coincides with the minimax \mathbb{L}_p -risk as conjectured in Remark 5 of [11]. Moreover, as follows from [29], the estimator suggested in [32] cannot be improved with respect to the \mathbb{L}_2 -risk. The estimator in [32] uses the Fejér-type kernel and the common theoretical bandwidth $h_n = (2\gamma\theta_n)/\log n$, $0 \leq \theta_n < 1$, used by many authors in the case of estimating infinitely smooth density functions. Here, $\gamma > 0$ is the smoothness parameter that describes the class \mathcal{F} of infinitely differentiable functions of interest and $n > 1$ is the sample size.

In this thesis, our goal is three-fold. First, we support via simulations the theoretical finding that the kernel-type estimator in [32] is the best one with respect to the \mathbb{L}_2 -risk. Second, we compare kernel estimators in hand by using the common theoretical bandwidth h_n and two empirical bandwidth selection methods; the well-known cross-validation and a less known method based on the Fourier analysis of kernel density estimators. Third, we apply the kernel-type estimator with the Fejér-type kernel to a practical situation with real financial data.

Chapter 2

Kernel Density Estimators

Let X_1, X_2, \dots be a sequence of i.i.d. 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. The problem is to estimate f from the random sample X_1, \dots, X_n by means of an estimator f_n under the assumption that f belongs to some class \mathcal{F} of smooth functions. One approach to density estimation is to use estimators with kernel functions.

Rosenblatt [25] and Parzen [22] pioneered the formulation of the kernel density estimator in the late fifties. Their approach of finding an appropriate kernel estimator was based on the relationship between the density and its distribution. Given that f is the derivative of F , then for some arbitrarily small $h > 0$ we have an approximate equality

$$f(x) = \frac{dF(x)}{dx} \approx \frac{F(x+h) - F(x-h)}{2h}.$$

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

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

Due to Kolmogorov's strong law of large numbers (see Theorem 5 in Appendix A),

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

that is, the empirical distribution function is a strongly consistent estimator of the true distribution for any $x \in \mathbb{R}$. Moreover, by the Glivenko-Cantelli theorem (see Theorem 4 in Appendix A),

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

Thus, for a large sample size n , a natural estimator of $f(x)$ is

$$f_n(x) = \frac{F_n(x+h) - F_n(x-h)}{2h}, \quad (2.1)$$

where $h = h_n$ is a function of n that tends to zero as n approaches infinity. Substituting the expression for $F_n(x)$ in (2.1) we get the *Rosenblatt estimator*

$$\begin{aligned} f_n^R(x) &:= \frac{1}{2nh} \sum_{j=1}^n \mathbb{I}(x-h \leq X_j \leq x+h) \\ &= \frac{1}{2nh} \sum_{j=1}^n \mathbb{I}\left(\left|\frac{X_j - x}{h}\right| \leq 1\right) \\ &= \frac{1}{nh} \sum_{j=1}^n K_0\left(\frac{X_j - x}{h}\right), \end{aligned}$$

where K_0 is the *uniform kernel*, that is, $K_0(u) = \mathbb{I}(|u| \leq 1)/2$. Later on, Parzen naturally extended the Rosenblatt idea by allowing K_0 to be any appropriate kernel function.

Formally, the *kernel density estimator* of $f(x)$ is defined by

$$f_n(x) := f_n(x; X_1, \dots, X_n) = \frac{1}{nh} \sum_{j=1}^n K\left(\frac{X_j - x}{h}\right), \quad x \in \mathbb{R}, \quad (2.2)$$

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

$$h_n \rightarrow 0 \quad \text{and} \quad nh_n \rightarrow \infty \quad \text{as} \quad n \rightarrow \infty. \quad (2.3)$$

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

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

$$\max_{u \in \mathbb{R}} K(u) = K(0), \quad \text{and} \quad \int_{\mathbb{R}} K^2(u) du < \infty. \quad (2.4)$$

If $K \geq 0$, then f_n is a probability density; however, we do not put any restrictions on the non-negativity of a kernel function since it allows us to use functions that lead to asymptotically efficient density estimators as shown, for example, in [8], [18], and [14]. A more general class of density estimators includes the so-called kernel-type estimators whose kernel functions, K_n , may depend on the sample size.

A kernel estimator distributes the mass that is allocated to X_j smoothly around X_j , not homogeneously, according to the kernel and the bandwidth. For a more intuitive description, $f_n(x)$ is the average of n small “mountains” centred around each observation X_j , $j = 1, \dots, n$. These mountains are given by the scaled kernel functions $K_h(X_j - x)$, $j = 1, \dots, n$. The overall efficiency and smoothness of a kernel density estimator depends on these individual mountains. The shape of the mountains is determined by K while their width is determined by h . The smaller the value of h , the more concentrated the mountains, resulting in a peaky-like density estimator. Contrarily, the larger the value of h , the flatter the mountains, leading to a lower and smoother density estimator. Figure 2.1

displays the construction of three kernel density estimates with different bandwidths from a random sample of size 10 drawn from a standard normal distribution. We choose $n = 10$ to clearly show how the mountains add up to a single estimate. Typically, such a small sample size would not be appropriate to construct a kernel density estimate as larger samples are required to boost the precision of the estimation. A *Gaussian kernel*, $K(u) = (2\pi)^{-1/2}e^{-u^2/2}$, is placed on each observation resulting in 10 mountains that are indicated by the blue dashed lines. The average of these mountains produces the kernel estimates, which are represented by the solid red curves. The true density, which is the standard normal, is depicted by the solid black curves. The different bandwidths in each plot illustrate the smoothing effect for the choice of h on a kernel density estimate. The smaller bandwidth in plot (a) causes the kernel mountains to spike at the observations thus increasing the variability of the density estimate; the larger bandwidth in plot (c) obscures the nature of the distribution. Plots (a) and (c) are examples of *under-* and *over-smoothing*.

The remainder of this chapter analyses the manner in which a kernel density estimator is assessed in $\mathbb{L}_2(\mathbb{R})$ and examines the choices of various kernels and bandwidths.

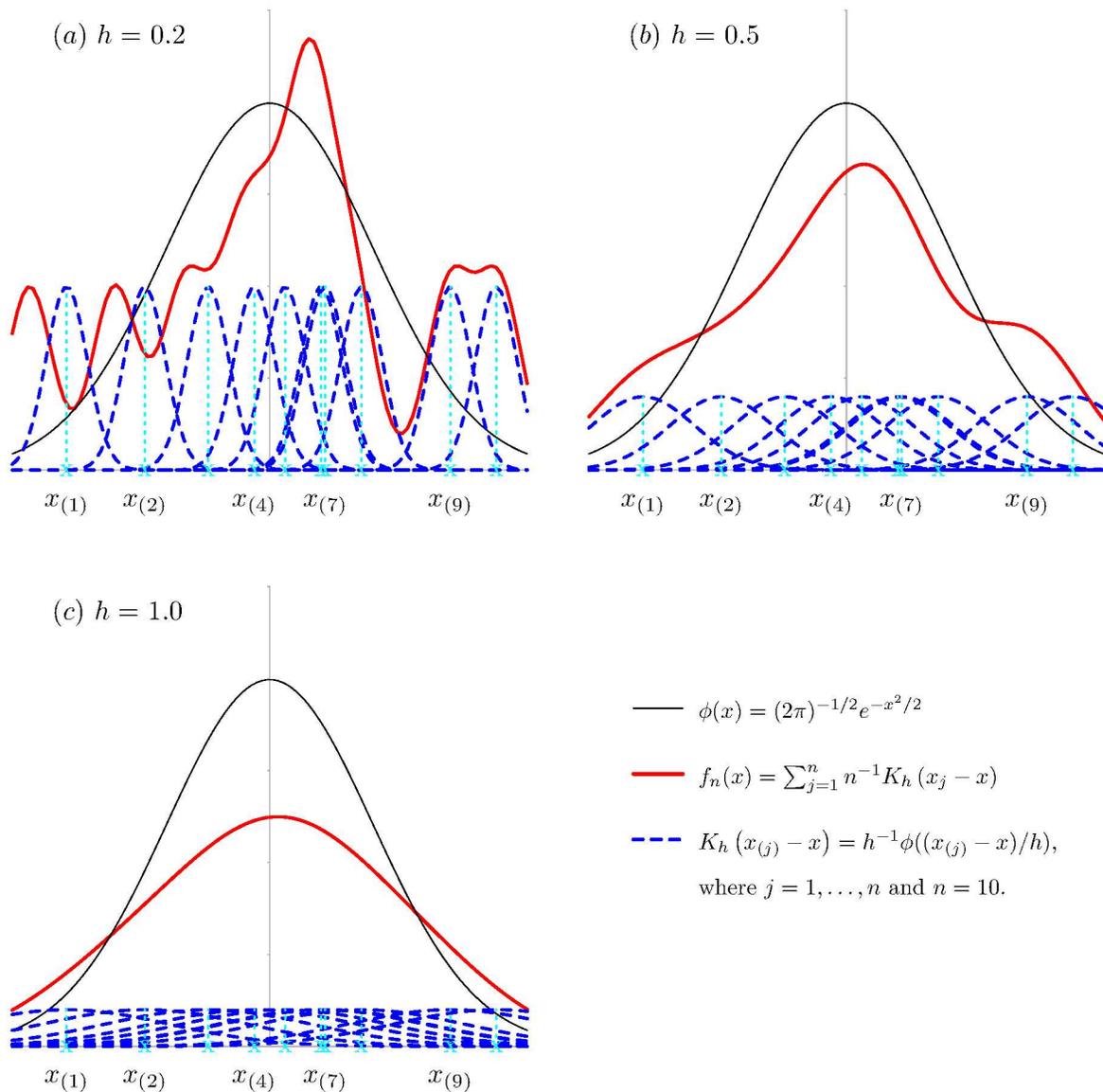


Figure 2.1: Comparison of three kernel density estimates with different smoothing bandwidths and a Gaussian kernel constructed from a standard normally distributed random sample of size $n = 10$. The individual “mountains” centred around each data point are the blue dashed curves, the kernel density estimates are the red curves, and the true densities are the black curves. Plots (a) and (c) are examples of under- and over-smoothing, respectively.

2.1 Performance Criterion

The problem of density estimation consists of selecting the best kernel estimator, with “good” properties, from a wide class of density estimators based on some performance criterion. A performance criterion is a measure of the distance between the true density function and the estimator. Ideally, this would be as small as possible. There are several ways to assess the proximity of these two functions so that the estimator is “good”, that is, nearly unbiased, consistent, and efficient. One common approach considers judging the quality of an estimator by using the $\mathbb{L}_p(\mathbb{R})$ -risk, $1 \leq p < \infty$.

Let X_1, X_2, \dots be a sequence of i.i.d. real-valued random variables with an unknown density f on \mathbb{R} that is assumed to belong to a class \mathcal{F} of smooth functions. For any function g in $\mathbb{L}_p = \mathbb{L}_p(\mathbb{R})$, $1 \leq p < \infty$, we denote its \mathbb{L}_p -norm by $\|g\|_p = (\int_{\mathbb{R}} |g(x)|^p dx)^{1/p}$. Let $f_n(\cdot) = f_n(\cdot; X_1, \dots, X_n)$ be an arbitrary estimator of $f(\cdot)$ based on the observations X_1, \dots, X_n . The *expected loss* of choosing f_n as an estimator of f is represented by the \mathbb{L}_p -risk given by

$$R_p(f_n, f) := \mathbb{E}_f l(\|f_n - f\|_p). \quad (2.5)$$

In application, it is difficult to measure the risk with $p \neq 2$, so for the purpose of our simulations we will focus our analysis on the \mathbb{L}_2 -risk, a global criterion for measuring the overall quality of f_n in \mathbb{L}_2 . We stress here again that when $l(x) = x^2$, the \mathbb{L}_2 -risk, $R_2(f_n, f)$ as defined in (2.5), coincides with the *mean integrated squared error* (MISE) which is given by

$$\text{MISE}(f_n, f) := \mathbb{E}_f l(\|f_n - f\|_2^2) = \mathbb{E}_f \left[\int_{\mathbb{R}} (f_n(x) - f(x))^2 dx \right]. \quad (2.6)$$

The MISE is, perhaps, the most popular criterion for measuring the quality of estimation in the literature and in practice. Note that, by the Fubini theorem (see Theorem 3 in

Appendix A),

$$\begin{aligned}
\text{MISE}(f_n, f) &= \int_{\mathbb{R}} \mathbb{E}_f [(f_n(x) - f(x))^2] dx \\
&= \int_{\mathbb{R}} (\mathbb{E}_f[f_n^2(x)] - 2f(x)\mathbb{E}_f[f_n(x)] + f^2(x) + \mathbb{E}_f^2[f_n(x)] - \mathbb{E}_f^2[f_n(x)]) dx \\
&= \int_{\mathbb{R}} (\mathbb{E}_f[f_n^2(x)] - \mathbb{E}_f^2[f_n(x)]) dx + \int_{\mathbb{R}} (\mathbb{E}_f[f_n(x)] - f(x))^2 dx \\
&=: \int_{\mathbb{R}} \sigma_n^2(x) dx + \int_{\mathbb{R}} b_n^2(x) dx, \tag{2.7}
\end{aligned}$$

where the terms $\int_{\mathbb{R}} \sigma_n^2(x) dx$ and $\int_{\mathbb{R}} b_n^2(x) dx$ represent the integrated variance and the integrated squared bias of $f_n(x)$. The above representation of the MISE is called the *variance-bias decomposition*. The aim is to choose a nearly unbiased $f_n(x)$ with small variance so that the estimation error of $f(x)$ is as small as possible.

We proceed to discuss the derivation of the MISE using two different approaches. The first approach, well-established in the literature, focuses on deriving an upper bound for the MISE by obtaining upper bounds on the variance and the bias terms. The second, less restrictive, approach utilises Fourier analysis of kernel density estimators. The latter is a more suitable approach, in spite of it being limited to \mathbb{L}_2 -theory, as it admits more general and elegant results for the MISE. The results for both methods shall be used and compared throughout this study.

2.1.1 MISE Evaluation under the Classical Approach

We are interested in evaluating the MISE of a kernel estimator f_n of an unknown density f to help us determine the accuracy of f_n . Recall from (2.7) that the structure of the MISE is composed of the variance and the bias terms of the estimator. To evaluate the MISE, an upper bound is obtained for each of these terms.

To begin, choose a kernel function $K : \mathbb{R} \rightarrow \mathbb{R}$ with properties (2.4). Then, for any f ,

the following upper bound on the variance term holds true (see [34], Proposition 1.4):

$$\int_{\mathbb{R}} \sigma_n^2(x) dx \leq \frac{1}{nh} \int_{\mathbb{R}} K^2(u) du. \quad (2.8)$$

Although the upper bound for the variance term in (2.8) does not require any conditions on the density, this is not the case when evaluating the upper bound for the bias term. Some additional assumptions on the kernel and the density functions need to be made to control for the bias term. Let us first define the order of a kernel function and the criterion for smoothness of a functional density class.

A kernel K is said to be of *order* l , where $l \geq 1$ is an integer, if the functions $u \mapsto u^j K(u)$, $j = 0, \dots, l$, are integrable and satisfy

$$\int_{\mathbb{R}} u^j K(u) du = 0, \quad j = 1, \dots, l. \quad (2.9)$$

For a more restrictive definition often used in the literature, K is of order $l + 1$ if the conditions of (2.9) hold and

$$0 \neq \int_{\mathbb{R}} u^{l+1} K(u) du < \infty, \quad (2.10)$$

is satisfied.

Typically, f is assumed to belong to a smooth class \mathcal{F} of probability density functions; however, there are many criteria for smoothness. A class consisting of a stronger notion of smoothness is considered. As the risk in hand corresponds to the \mathbb{L}_2 -norm, the density f is naturally assumed to be smooth with respect to the \mathbb{L}_2 -norm. A common way to describe smoothness in \mathbb{L}_2 is by using Sobolev classes.

Given an integer $\beta \geq 1$ and a number $B > 0$, consider the class $\mathcal{S}(\beta, B)$ of functions $f : \mathbb{R} \rightarrow \mathbb{R}$ such that any $f \in \mathcal{S}(\beta, B)$ is $\beta - 1$ times differentiable, has absolutely continuous

derivative $f^{(\beta-1)}$ on \mathbb{R} , and satisfies

$$\int_{\mathbb{R}} (f^{(\beta)}(x))^2 dx \leq B^2.$$

The function class $\mathcal{S}(\beta, B)$ is usually referred to as the (non-periodic) *Sobolev class* or *Sobolev ball* (of radius B). Then, the Sobolev class of pdfs is defined by

$$\mathcal{F}_S(\beta, B) = \{f \in \mathcal{S}(\beta, B) : f \text{ is a density on } \mathbb{R}\}.$$

Now, let K be of order $\beta - 1$ and f belong to $\mathcal{F}_S(\beta, B)$. Then, it follows from (2.10) that $\int_{\mathbb{R}} |u|^\beta |K(u)| du < \infty$. Using the Taylor expansion of the density estimator, one can estimate the bias term from above as

$$\int_{\mathbb{R}} b_n^2(x) dx \leq \frac{h^{2\beta}}{((\beta - 1)!)^2} \left(\int_{\mathbb{R}} |u|^\beta K(u) du \right)^2 \int_{\mathbb{R}} (f^{(\beta)}(x))^2 dx < \infty. \quad (2.11)$$

For the derivation of (2.11), see Section 1.2.3 of [34].

Finally, under the above assumptions on the kernel and the density, combining (2.8) and (2.11) yields the following upper bound on the MISE:

$$\text{MISE}(f_n, f) \leq \frac{1}{nh} \int_{\mathbb{R}} K^2(u) du + \frac{h^{2\beta}}{((\beta - 1)!)^2} \left(\int_{\mathbb{R}} |u|^\beta |K(u)| du \right)^2 \int_{\mathbb{R}} (f^{(\beta)}(x))^2 dx. \quad (2.12)$$

2.1.2 MISE Evaluation using Fourier Analysis Techniques

In this section, the MISE is evaluated using the results of Fourier analysis, where the kernel density estimators are viewed as inverse Fourier transforms of the sample characteristic function. Unlike the classical approach outlined in the previous section, the assumptions here are much less restrictive. This method is not as prominent as the classical one, perhaps due to its limitation to \mathbb{L}_2 -theory, but it is more suitable for studying the \mathbb{L}_2 -risk as it allows

much easier derivation of the optimal kernels. In fact, the results of Fourier analysis, in closely related settings, have been actively employed since the times of Parzen [21], where it was used to estimate unknown spectral densities of a stationary time series. Watson and Leadbetter [36] later adopted this approach to the nonparametric density framework, which has then been used by others, such as Cline [5], to help determine optimal kernels. We will now show how the exact MISE is evaluated using Fourier analysis of kernel density estimators.

We begin by recalling a few basic definitions and properties that can be found in any standard real and complex analysis textbook (see [28], for example). The *Fourier transform* \hat{g} of a function $g \in \mathbb{L}_1$ is defined by

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

where $i = \sqrt{-1}$ is the *imaginary unit*. Sometimes we will also write $\widehat{g(\cdot)}(t)$ for the Fourier transform of g at t . By the Fourier inversion theorem (see Theorem 2 in Appendix A),

$$g(x) := \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \hat{g}(t) dt, \quad x \in \mathbb{R},$$

that is, g is the *inverse Fourier transform* of \hat{g} . The Plancherel theorem (see Theorem 6 in Appendix A) allows us to extend the definition of the Fourier transform to functions in \mathbb{L}_2 . Moreover, for any $g \in \mathbb{L}_2$, the Parseval formula (see Theorem 7 in Appendix A) holds true:

$$\int_{\mathbb{R}} |g(x)|^2 dx = \frac{1}{2\pi} \int_{\mathbb{R}} |\hat{g}(t)|^2 dt. \quad (2.13)$$

Consequently, the Plancherel theorem enables us to evaluate the MISE of density estimators

that use kernels that are not in \mathbb{L}_1 , for instance, the *sinc kernel*

$$K(u) = \begin{cases} \sin u/(\pi u) & \text{if } u \neq 0, \\ 1/\pi & \text{if } u = 0. \end{cases} \quad (2.14)$$

For a function $g \in \mathbb{L}_2$ and for any $h > 0$ and $x \in \mathbb{R}$, the following scaling and translation properties of the Fourier transform are well known:

$$\frac{1}{h} \widehat{g\left(\frac{\cdot}{h}\right)}(t) = \widehat{g}(ht), \quad \widehat{g(x - \cdot)}(t) = e^{itx} \widehat{g}(-t), \quad t \in \mathbb{R}. \quad (2.15)$$

Note that the Fourier transform of a density f is the *characteristic function*, which is 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

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

Using the notation $K_h(x) = h^{-1}K(h^{-1}x)$, formulas (2.15), and symmetry of K , we can express the Fourier transform \widehat{f}_n of the density estimator f_n as follows:

$$\widehat{f}_n(t) = \frac{1}{n} \sum_{j=1}^n K_h(\widehat{X_j - \cdot})(t) = \frac{1}{n} \sum_{j=1}^n e^{itX_j} \widehat{K}(ht) = \varphi_n(t) \widehat{K}(ht). \quad (2.16)$$

Now we give some properties of $\varphi_n(t)$. First, it is easy to see that the empirical characteristic function is an unbiased estimator of the true characteristic function, that is, for all $t \in \mathbb{R}$,

$$\mathbb{E}[\varphi_n(t)] = \varphi(t). \quad (2.17)$$

The notation \mathbb{E} without index f will be used for brevity for the remainder of this section.

Next, using (2.17) we arrive at the relation

$$\begin{aligned}
\mathbb{E} [|\varphi_n(t)|^2] &= \mathbb{E} [\varphi_n(t)\varphi_n(-t)] \\
&= \frac{1}{n^2} \sum_{j,k:j \neq k} \mathbb{E} [e^{iX_j t}] \mathbb{E} [e^{-iX_k t}] + \frac{1}{n} \\
&= \frac{n-1}{n^2} |\varphi(t)|^2 + \frac{1}{n},
\end{aligned} \tag{2.18}$$

followed by

$$\mathbb{E} [|\varphi_n(t) - \varphi(t)|^2] = \mathbb{E} [|\varphi_n(t)|^2] - |\varphi(t)|^2 = \frac{1}{n} (1 - |\varphi(t)|^2). \tag{2.19}$$

Throughout this section, we assume that f and K are both in \mathbb{L}_2 . Recall that K is symmetric, hence its Fourier transform, \hat{K} , is real-valued. According to (2.13), the MISE can be expressed as

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

Applying (2.16), we get

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

where the last two equalities are due to relations (2.17) and (2.19). Finally, the form of the exact MISE of the kernel estimator f_n of density f based on the methods of Fourier

analysis is (see [34], Theorem 1.4),

$$\begin{aligned} \text{MISE}(f_n, f) = & \frac{1}{2\pi} \left[\int_{\mathbb{R}} |1 - \hat{K}(ht)|^2 |\varphi(t)|^2 dt + \frac{1}{n} \int_{\mathbb{R}} |\hat{K}(ht)|^2 dt \right] \\ & - \frac{1}{2\pi n} \int_{\mathbb{R}} |\varphi(t)|^2 |\hat{K}(ht)|^2 dt =: J_n(K, h, \varphi), \quad (2.20) \end{aligned}$$

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

We can see that the conditions required for (2.20) to be true are much milder than those for the upper bound on MISE in (2.12). Recall that under classical assumptions, K is assumed to be of order $\beta - 1$, $\beta \in \mathbb{Z}_+$, satisfying the following conditions:

$$\int_{\mathbb{R}} K^2(u) du < \infty \quad \text{and} \quad \int_{\mathbb{R}} |u|^\beta |K(u)| du < \infty.$$

Here, the only condition required for K is that it be symmetric and in \mathbb{L}_2 . Indeed, $\int_{\mathbb{R}} K(u) du = 1$ does not need to be true nor does K need to be integrable, despite it not making sense for a kernel function to not integrate to one. Also, (2.20) can be used to construct estimators that use kernels that are not in \mathbb{L}_1 , such as the well-known efficient sinc kernel in (2.14). These mild restrictions allow for the analysis of a larger pool of kernel candidates thus increasing the chances of discovering more improved density estimators. Also, the exact MISE given by (2.20) has less restrictions on f , while the classical MISE given by (2.12) is only valid for densities whose derivatives of given order satisfy certain conditions. Consequently, formula (2.20) allows for the analysis of more density estimators than (2.12) due to its much milder constraints on K and f .

The two MISE results, (2.12) and (2.20), are similar; however (2.20) can provide slightly more accurate results than (2.12); the latter holds true only under the classical assumptions mentioned above. The right-hand side (RHS) of (2.20) consists of a main term, the expression in square brackets, and a negative term. The main term is similar to the RHS of (2.12) as it is a combination of the variance and the bias terms. Recall that the upper

bound on the variance term in (2.12) did not have any conditions on f , so by the Plancherel theorem and (2.15), the bound on the variance under classical assumptions is the same as

$$\frac{1}{2\pi n} \int_{\mathbb{R}} |\hat{K}(ht)|^2 dt = \frac{1}{nh} \int_{\mathbb{R}} K^2(u) du.$$

On the other hand, the bias term

$$\frac{1}{2\pi} \int_{\mathbb{R}} |1 - \hat{K}(ht)|^2 |\varphi(t)|^2 dt,$$

differs from the upper bound of the bias term in (2.12) as it does not require that K be of some specified order nor does it have any conditions on f other than belonging to \mathbb{L}_2 . The additional negative term

$$-\frac{1}{2\pi n} \int_{\mathbb{R}} |\varphi(t)|^2 |\hat{K}(ht)|^2 dt,$$

included in the exact MISE in (2.20), serves to correct the variance term above rendering it somewhat more accurate than the upper bound in (2.12). In most cases, however, the correction term is of smaller order than the variance term, particularly for small h , which results in the variance term asymptotically coinciding with that of (2.12). Therefore, the MISE obtained by the Fourier analysis approach provides greater flexibility in the choice of the kernel and the density functions as well as the possibility of more accurate results.

2.2 Kernel Functions

A kernel function K is one of two necessary parameters of a density estimator f_n . Kernels are defined in the beginning of this chapter as any real-valued integrable functions that have properties (2.4). We are interested in choosing a kernel whose corresponding density estimator has the smallest global \mathbb{L}_p -risk. Here, we discuss the efficiency of various kernel

functions that are well known in theory and in application.

Table 2.1 lists some well-known kernel functions in the literature on nonparametric estimation and their associated Fourier transforms. A graphical representation of these functions is provided in Figure 2.2. The first group of kernels listed in Table 2.1a are the most commonly applied in practice and have the additional property $K \geq 0$. We have already seen the uniform kernel used in the original Rosenblatt estimator at the start of this chapter. The Epanechnikov and the Gaussian kernels are popular choices in practice since the former is considered optimal under certain settings while the latter has familiar properties from the normal density. The second group of kernels presented in Table 2.1b alternate between positive and negative values, except for the Fejér kernel. Although less common, they are generally more asymptotically efficient than the standard kernels in Table 2.1a since they were shown to achieve better rates of convergence and smaller exact asymptotic constants in the works of [8], [11], and [14]. The sinc kernel has been used by many authors since the early seventies, but it is not absolutely integrable, that is, not in \mathbb{L}_1 ; the latter fact complicates the analytical studies of the corresponding estimator. The de la Vallée Poussin kernel, which is well known in approximation theory, has been successfully used in the field of nonparametric estimation starting the work of Ibragimov and Hasminskii [14]. Therefore, the kernel functions in Table 2.1b are preferred to those in Table 2.1a as their corresponding density estimators result in smaller MISE.

The nonnegative kernels in Table 2.1a are usually chosen for constructing estimators since they maintain the nonnegative property of the density functions. In fact, nonnegativity of density estimators is not necessarily achieved from using nonnegative kernels as we can always use the *positive part estimator*

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

whose risk is smaller than or equal to the risk of f_n :

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

Therefore, for the efficient kernels in Table 2.1b we can use the estimator f_n^+ , which is nonnegative and attains fast convergence rates.

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	$\frac{1}{\sqrt{2\pi}}e^{-u^2/2}$	$e^{-t^2/2}$

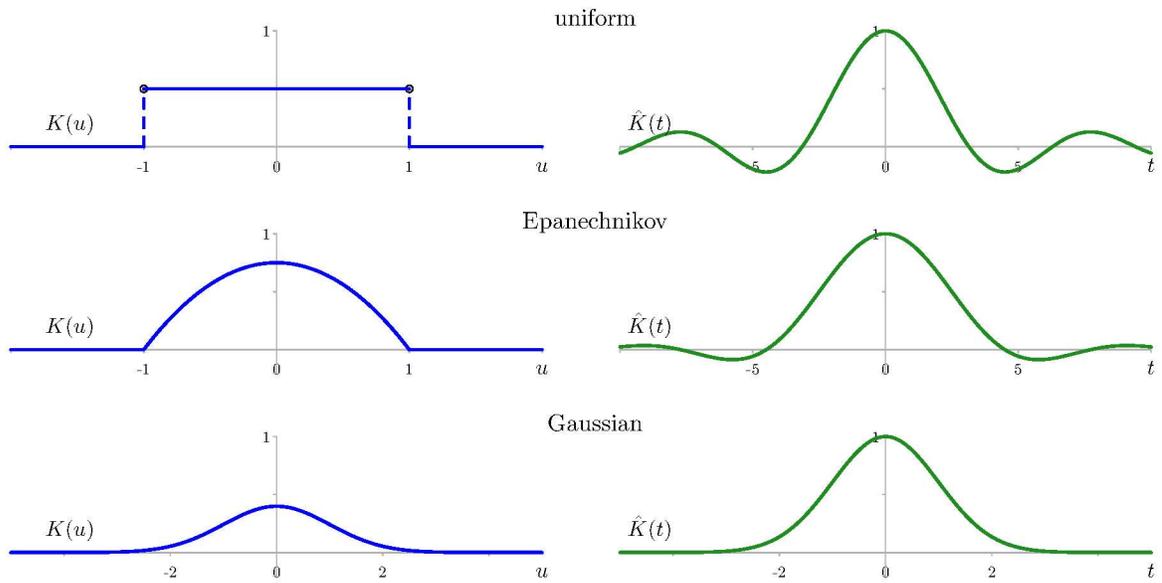
(a) Some standard kernels

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	$\begin{cases} \frac{2 \sin^2(u/2)}{\pi u^2} & \text{if } u \neq 0, \\ \frac{1}{2\pi} & \text{if } u = 0 \end{cases}$	$(1 - t)\mathbb{I}(t \leq 1)$

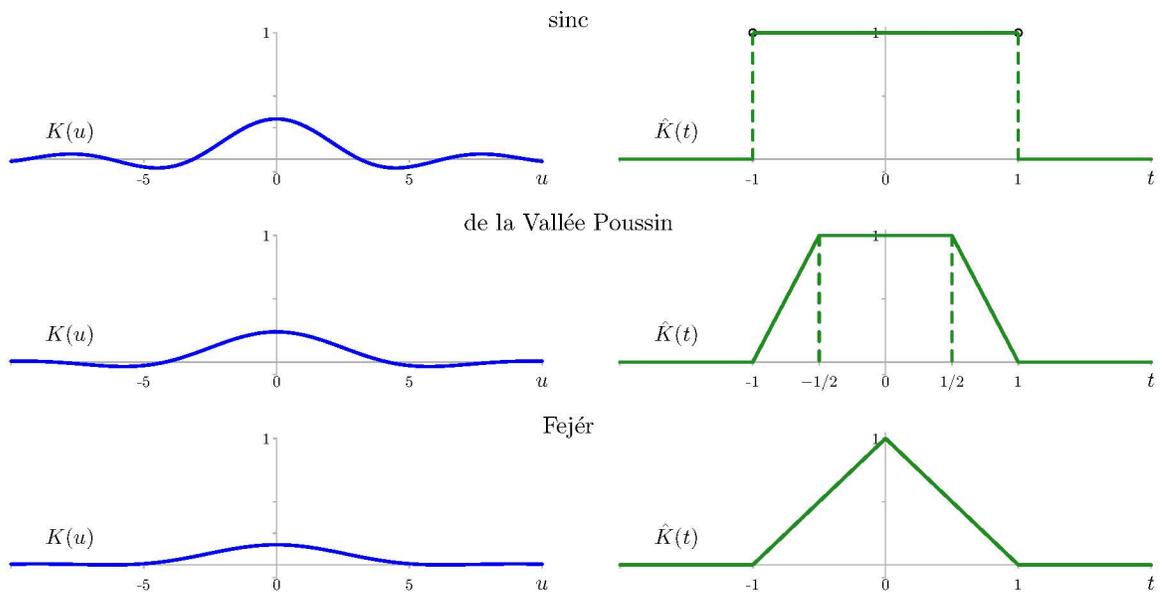
(b) Some efficient kernels

Table 2.1: Some common kernel functions and their Fourier transforms

There is a common misconception that the Epanechnikov kernel is optimal due to an old, yet popular, approach for deriving optimal density estimators. In the early sixties, Bartlett [2] derived an approach for optimally choosing $K \geq 0$ and $h > 0$ based on



(a) *Some standard kernels*



(b) *Some efficient kernels*

Figure 2.2: *Some common kernel functions and their Fourier transforms*

asymptotic optimality for a fixed density. Soon after, Epanechnikov [6] applied this method to determine the optimal kernel now recognised under his name. It was not until the late 1990s that the inconsistency of asymptotic risk for a fixed density was brought to light by Brown et al. [4] in their asymptotic analysis of superefficiency in nonparametric

estimation. We briefly explain how such asymptotic statements leading to the optimality of the Epanechnikov kernel are misleading according to the propositions supplied in Section 1.2.4 of [34].

Consider a version of this result where $K \geq 0$ is assumed to be of order 1 and the density f belongs to the Sobolev class $\mathcal{S}(\beta, B)$ with $\beta = 2$. Proposition 1.6 in [34] shows that the upper bound in (2.12) is asymptotically sharp in the sense that

$$\text{MISE}(f_n, f) = \left[\frac{1}{nh} \int_{\mathbb{R}} K^2(u) du + \frac{h^4}{4} \left(\int_{\mathbb{R}} |u|^2 |K(u)| du \right)^2 \int_{\mathbb{R}} (f''(x))^2 dx \right] (1 + o(1)), \quad (2.21)$$

where the density dependent term $o(1)$ is independent of n and tends to zero as h approaches zero. Minimising in (2.21) the main term in square brackets in $K \geq 0$ and in h results in the “optimal” *Epanechnikov kernel* K^* , displayed in Table 2.1a, and the “optimal” bandwidth h^* for given K^* ,

$$h^*(K^*) = \left(\frac{\int_{\mathbb{R}} (K^*(u))^2 du}{n S_{K^*}^2 \int_{\mathbb{R}} (f''(x))^2 dx} \right)^{1/5} = \left(\frac{15}{n \int_{\mathbb{R}} (f''(x))^2 dx} \right)^{1/5}, \quad (2.22)$$

where $S_K^2 = \int_{\mathbb{R}} u^2 K(u) du \neq 0$. These “optimal” parameters, however, cannot define a valid kernel estimator since the choices of h^* in (2.22) are dependent on the unknown f . As a result, the optimal density estimator, which we denote $f_n^E(x)$, is a random variable consisting of K^* and h^* that satisfies the best achievable MISE given by (see (1.27) of [34])

$$\lim_{n \rightarrow \infty} n^{4/5} \mathbb{E}_f \int_{\mathbb{R}} (f_n^E(x) - f(x))^2 dx = \frac{3^{4/5}}{5^{1/54}} \left(\int_{\mathbb{R}} (f''(x))^2 dx \right)^{1/5}. \quad (2.23)$$

A counterexample in Section 1.2.4 of [34] shows that, under the same assumptions on fixed f but different conditions on K , there exist a multitude of estimators that can attain smaller asymptotic MISE than $f_n^E(x)$. Now, consider a case where the conditions on the density remain the same but the kernel is no longer nonnegative and has order 2. According

to (2.10), this implies that $S_K^2 = \int_{\mathbb{R}} u^2 K(u) du = 0$ and that the density estimator f_n can take on both positive and negative values. Then, for any small $\epsilon > 0$, the kernel estimator f_n with bandwidth

$$h = n^{-1/5} \epsilon^{-1} \int_{\mathbb{R}} K^2(u) du$$

satisfies

$$\limsup_{n \rightarrow \infty} n^{4/5} \mathbb{E}_f \int_{\mathbb{R}} (f_n(x) - f(x))^2 dx \leq \epsilon \quad (2.24)$$

and

$$\limsup_{n \rightarrow \infty} n^{4/5} \mathbb{E}_f \int_{\mathbb{R}} (f_n^+(x) - f(x))^2 dx \leq \epsilon, \quad (2.25)$$

where $f_n^+ = \max(0, f_n)$. A proof of this result can be found in Appendix A.2 of [34]. Hence, for all $\epsilon > 0$ that are smaller than the RHS of (2.23), the estimators f_n and f_n^+ in (2.24) and (2.25) outperform the “optimal” estimator f_n^E that utilizes the Epanechnikov kernel. This does not mean, however, that the estimators in (2.24) and (2.25) are themselves good. In fact, they could be quite counterintuitive if the performance criterion is not based on a fixed density. This is because, for fixed f , the condition $S_K^2 = 0$ eliminates the bias term in (2.21) for large n , while the arbitrary factor ϵ^{-1} in their bandwidth h serves to reduce the variance term in (2.21). However, the bias cannot be eliminated uniformly over f in $\mathcal{S}(\beta, B)$ with $\beta = 2$. Therefore, arguments (2.24) and (2.25) capture the inconsistency of the asymptotic optimality approach for a fixed density.

Fourier analysis of kernel functions is another way of showing that the Epanechnikov kernel is not good. Formula (2.20) makes it possible to compare the MISE, which we have denoted by $J_n(K, h, \varphi)$, of different kernel estimators, nonasymptotically, for any fixed n . In particular, one can eliminate “bad” kernels by using the following notion of admissibility.

A kernel $K \in \mathbb{L}_2(\mathbb{R})$ is called *inadmissible* if there exists another kernel $K_0 \in \mathbb{L}_2(\mathbb{R})$ such that for all characteristic functions $\varphi \in \mathbb{L}_2(\mathbb{R})$,

$$J_n(K_0, h, \varphi) \leq J_n(K, h, \varphi), \quad h > 0, n \geq 1,$$

and there exists a characteristic function $\varphi_0 \in \mathbb{L}_2(\mathbb{R})$ such that

$$J_n(K_0, h, \varphi_0) < J_n(K, h, \varphi_0), \quad h > 0, n \geq 1;$$

otherwise, the kernel K is called *admissible*.

In 1988, Cline [5] presented a way to generally detect inadmissible kernels. A simple method is as follows: if

$$\text{Leb}(t : \hat{K}(t) \notin [0, 1]) > 0, \tag{2.26}$$

where $\text{Leb}(A)$ denotes the Lebesgue measure of a set A , then K is inadmissible (for details, see Theorem 1 in [5] or Proposition 1.8 in [34]). As seen from Figure 2.2a and Table 2.1a, the Epanechnikov kernel is inadmissible since the set $\{t : \hat{K}(t) \notin [0, 1]\}$ has a positive Lebesgue measure. Indeed, this is also true for the uniform kernel in Table 2.1a, rendering it inadmissible. The efficient kernels in Table 2.1b are characterized as admissible since they have nonnegative Fourier transforms bounded by 1. This is another argument as to why the efficient kernels as well as the family of Fejér-type kernels studied in this thesis are preferred.

2.3 Bandwidth Selection

The bandwidth $h = h_n$ is a smoothing parameter of a kernel density estimator f_n that depends on the sample size n with properties (2.3). We want to choose an appropriate h

such that its corresponding f_n is as close to the true density f as possible. Bandwidth selection is very important in kernel density estimation as it affects the quality of the estimator.

A good estimator must have both the variance and the bias terms small so that the overall error in density estimation remains negligible. The choice of h determines the balance of the variance and the bias. From the upper bounds on the integrated risk presented in (2.12), it can be seen that the choice of small h corresponds to a larger variance leading to under-smoothing. On the other hand, for large values of h , over-smoothing occurs due to a higher bias. The aim is to minimize the bias and the variance of f_n to avoid under- or over-smoothing, a dilemma known as the *bias-variance tradeoff*. Being that the MISE is a sum of these two error terms, minimising the MISE with respect to h yields the optimal value of the smoothing parameter:

$$h_{opt} = \arg \min_{h>0} \text{MISE}(h). \quad (2.27)$$

As the MISE depends on the unknown f , the result (2.27) remains purely theoretical. In practice, an appropriate unbiased estimator of the $\text{MISE}(h)$, computed from the random sample X_1, \dots, X_n , is minimised instead with respect to h . Cross-validation, developed by Rudemo [27] and Stone [33] in the early 1980s, is 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 the h_{opt} (see, for example, Theorem 4.1 in [30]). There exists another method proposed in 1992 by Golubev [7] where the optimal bandwidth minimises the exact MISE in (2.20), obtained by using Fourier analysis of kernel density estimators, as a function of h . Although this latter method is not well known to the general public, the simulation studies conducted in Chapter 4 show that it works very well. These two data-dependent methods based on unbiased risk estimation for the choice of h will be discussed in detail in the successive subsections.

2.3.1 Unbiased Cross-Validation Method

Cross-validation is a popular data-driven approach that attempts to choose h by minimising an approximately unbiased estimator of the MISE. We begin by describing how the unbiased estimator of the \mathbb{L}_2 -risk is characterized. Observe that the MISE defined in (2.6) can be expanded as

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

When minimising $\text{MISE}(h)$ with respect to h we may disregard the term $\int_{\mathbb{R}} f^2(x) dx$ since it does not depend on h . Minimising $\mathcal{J}_n(h)$ with respect to h produces equivalent results as (2.27); hence it suffices to find an unbiased estimator of the function $\mathcal{J}_n(h)$.

Note that an unbiased estimator for the quantity $\mathbb{E}_f \left[\int_{\mathbb{R}} f_n^2(x) dx \right]$ is $\int_{\mathbb{R}} f_n^2(x) dx$. Now, consider the *leave-one-out estimator* of $f(x)$ given by

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

then $n^{-1} \sum_{j=1}^n f_{n,-j}(X_j)$ is an unbiased estimator of $\mathbb{E}_f \left[\int_{\mathbb{R}} f_n(x)f(x) dx \right]$. Indeed, since X_1, \dots, X_n are i.i.d., it follows that

$$\begin{aligned} \mathbb{E}_f \left[\frac{1}{n} \sum_{j=1}^n f_{n,-j}(X_j) \right] &= \mathbb{E}_f \left[\frac{1}{n} \sum_{j=1}^n \frac{1}{(n-1)h} \sum_{k \neq j} K \left(\frac{X_k - X_j}{h} \right) \right] \\ &= \mathbb{E}_f \left[\frac{1}{(n-1)h} \sum_{k \neq 1} \int_{\mathbb{R}} K \left(\frac{X_k - u}{h} \right) f(u) du \right] \\ &= \frac{1}{h} \int_{\mathbb{R}} \int_{\mathbb{R}} K \left(\frac{x - u}{h} \right) f(u)f(x) du dx \\ &= \int_{\mathbb{R}} f(x) \mathbb{E}_f [f_n(x)] dx \\ &= \mathbb{E}_f \left[\int_{\mathbb{R}} f_n(x)f(x) dx \right]. \end{aligned}$$

Finally, when combining the two unbiased estimators, we arrive at the following unbiased estimator for $\mathcal{J}_n(h)$:

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

where CV stands for *cross-validation*. We call the function $\text{CV}(\cdot)$ the *unbiased cross-validation criterion*. It can be shown (see [13], p. 55) that the cross-validation criterion can further be expanded to

$$\begin{aligned} \text{CV}(h) &= \frac{1}{n^2 h} \sum_{j=1}^n \sum_{k=1}^n (K * K) \left(\frac{X_j - X_k}{h} \right) - \frac{2}{n(n-1)h} \sum_{j=1}^n \sum_{k \neq j}^n K \left(\frac{X_j - X_k}{h} \right) \\ &= \frac{1}{nh} (K * K)(0) + \frac{2}{n^2 h} \sum_{j=1}^k \sum_{k=j+1}^n (K * K) \left(\frac{X_j - X_k}{h} \right) \\ &\quad - \frac{4}{n^2 h(n-1)} \sum_{j=1}^k \sum_{k=j+1}^n K \left(\frac{X_j - X_k}{h} \right), \end{aligned} \tag{2.28}$$

where $*$ denotes convolution (see Definition 1 in Appendix A).

Now, since $\text{CV}(h)$ is an unbiased estimator of $\mathcal{J}_n(h)$, then it is also an unbiased estimator of $\text{MISE}(h)$ shifted by $\int_{\mathbb{R}} f^2(x) dx$:

$$\mathbb{E}_f[\text{CV}(h)] = \mathcal{J}_n(h) = \text{MISE}(h) - \int_{\mathbb{R}} f^2(x) dx,$$

where f^2 is independent of h . This implies that $\mathbb{E}[\text{CV}(h)]$ and $\text{MISE}(h)$ would both obtain the same minimums for the values of h_{opt} . Furthermore, the function $\text{CV}(h)$ would obtain approximate results to those of $\mathbb{E}[\text{CV}(h)]$ when minimised with respect to h . In practice, the random sample X_1, \dots, X_n is used to compute the function $\text{CV}(h)$ so that the cross-validation bandwidth, h_{CV} , may be characterized as

$$h_{CV} := \arg \min_{h>0} \text{CV}(h; X_1, \dots, X_n). \tag{2.29}$$

The *cross-validation kernel estimator* of f specified by h_{CV} is then given by

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

Unbiased cross-validation, in the form described in this section, was introduced by Rudemo [27] in 1982. Stone [33] then showed that, under appropriate conditions, h_{CV} is asymptotically equivalent to h_{opt} and therefore asymptotically optimal. In turn, the *integrated squared error* (ISE) of the estimator $f_{n,\text{CV}}(x)$, defined as $\int_{\mathbb{R}} (f_{n,\text{CV}}(x) - f(x))^2 dx$, is asymptotically close to that of the estimator specified by h_{opt} . Scott and Terrell (see [30], Theorem 4.1) later found that, for nonnegative kernels and sufficiently large n , the convergence of h_{CV} to h_{opt} is of order $O(n^{-1/10})$, namely, as $n \rightarrow \infty$

$$n^{1/10} \frac{(h_{\text{CV}} - h_{\text{opt}})}{h_{\text{opt}}} \xrightarrow{P} T,$$

where T is some random variable independent of n . This is a slow convergence rate compared to the best achievable \sqrt{n} -convergence.

The method of cross-validation has good theoretical properties, but in application it can be unreliable for very small sample sizes leading to higher variability of the density estimator. In addition, this method is computationally intensive and, as a result, its implementation can be time-consuming.

2.3.2 Unbiased Risk Estimation based on the Fourier Analysis of Kernel Density Estimators

The data-dependent bandwidth selection method studied in this section is parallel to that of cross-validation except that the MISE is expressed using the Fourier analysis of a kernel density estimator. The exact MISE of interest is given in (2.20) and denoted by $J_n(K, h, \varphi)$.

We are interested in minimising the unbiased estimator of $J_n(K, h, \varphi)$ so that we may obtain a value for h that would be close to h_{opt} .

In 1992, Golubev [7] introduced the following function of $h > 0$,

$$\tilde{J}_n(h) := \int_{\mathbb{R}} \left(-2\hat{K}(ht) + \hat{K}^2(ht) \left(1 - \frac{1}{n} \right) \right) |\varphi_n(t)|^2 dt + \frac{2}{n} \int_{\mathbb{R}} \hat{K}(ht) dt, \quad (2.30)$$

which is, up to scaling and shifting, an unbiased estimator of $J_n(K, h, \varphi)$. Indeed, recalling relations (2.17) and (2.18), we get that the expected value of $\tilde{J}_n(h)$ as in (2.30) will be

$$\begin{aligned} \mathbb{E}_f[\tilde{J}_n(h)] &= \int_{\mathbb{R}} \left(-2\hat{K}(ht) + \hat{K}^2(ht) \left(1 - \frac{1}{n} \right) \right) \mathbb{E}_f[|\varphi_n(t)|^2] dt + \frac{2}{n} \int_{\mathbb{R}} \hat{K}(ht) dt, \\ &= \int_{\mathbb{R}} \left(-2\hat{K}(ht) + \hat{K}^2(ht) \left(1 - \frac{1}{n} \right) \right) \left(\left(1 - \frac{1}{n} \right) |\varphi(t)|^2 + \frac{1}{n} \right) dt + \frac{2}{n} \int_{\mathbb{R}} \hat{K}(ht) dt \\ &= \int_{\mathbb{R}} \left(-2\hat{K}(ht) + \hat{K}^2(ht) \left(1 - \frac{1}{n} \right) \right) \left(1 - \frac{1}{n} \right) |\varphi(t)|^2 dt + \frac{1}{n} \left(1 - \frac{1}{n} \right) \int_{\mathbb{R}} \hat{K}^2(ht) dt \\ &= \left(1 - \frac{1}{n} \right) \left[\int_{\mathbb{R}} \left(-2\hat{K}(ht) + \hat{K}^2(ht) - \frac{1}{n} \hat{K}^2(ht) \right) |\varphi(t)|^2 dt + \frac{1}{n} \int_{\mathbb{R}} \hat{K}^2(ht) dt \right] \\ &= \left(1 - \frac{1}{n} \right) \left[\int_{\mathbb{R}} \left(-2\hat{K}(ht) + \hat{K}^2(ht) \right) |\varphi(t)|^2 dt + \frac{1}{n} \int_{\mathbb{R}} (1 - |\varphi(t)|^2) \hat{K}^2(ht) dt \right] \\ &= \left(1 - \frac{1}{n} \right) \left[\int_{\mathbb{R}} (1 - \hat{K}(ht))^2 |\varphi(t)|^2 dt + \frac{1}{n} \int_{\mathbb{R}} (1 - |\varphi(t)|^2) \hat{K}^2(ht) dt - \int_{\mathbb{R}} |\varphi(t)|^2 dt \right] \\ &= 2\pi \left(1 - \frac{1}{n} \right) \left[J_n(K, h, \varphi) - \int_{\mathbb{R}} f^2(x) dx \right]. \end{aligned}$$

Hence, $\tilde{J}_n(h)$ yields an unbiased estimator of $J_n(K, h, \varphi)$ shifted by the quantity $\int_{\mathbb{R}} f^2(x) dx$, where f^2 is independent of h , and amplified with constant $2\pi(1 - 1/n)$. For application purposes, the function $\tilde{J}_n(h)$ as in (2.30) can be simplified even further by applying the inverse Fourier transform property $\int_{\mathbb{R}} \hat{K}(t) dt = 2\pi K(0)$, which results in

$$\tilde{J}_n(h) = \int_{\mathbb{R}} \left(-2\hat{K}(ht) + \hat{K}^2(ht) \left(1 - \frac{1}{n} \right) \right) |\varphi_n(t)|^2 dt + \frac{4\pi K(0)}{nh}. \quad (2.31)$$

Minimising $\mathbb{E}_f[\tilde{J}_n(h)]$ with respect to h is equivalent to minimising $J_n(K, h, \varphi)$ as in (2.20) in h . In order to get an *approximate minimiser* of $J_n(K, h, \varphi)$, the function $\tilde{J}_n(h)$

is first computed using the observations X_1, \dots, X_n and then minimised with respect to h :

$$h_{\text{F}} := \arg \min_{h>0} \tilde{J}_n(h; X_1, \dots, X_n). \quad (2.32)$$

The corresponding kernel density estimator with the data-dependent bandwidth h_{F} is then defined as

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

Under appropriate conditions, h_{F} given by (2.32) behaves like h_{CV} in (2.29), in that, it is asymptotically optimal as it is asymptotically equivalent to h_{opt} in (2.27). In other words, the MISE of the estimator $f_{n,\text{F}}$ is asymptotically equivalent to that of the estimator corresponding to the optimal bandwidth h_{opt} (see without proof [34], Section 1.4).

Chapter 3

Density Estimators with Fejér-Type Kernels

The following functional class is well known in approximation theory (see, for example, Section 94 of [1]) and is widely used in nonparametric estimation (see, for example, [8], [9], [11], [18], [20], [32]). Consider the functional class $\mathcal{F} = \mathcal{F}(\gamma, L)$ of functions f in $\mathbb{L}_2(\mathbb{R})$ such that each $f \in \mathcal{F}$ 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 $L > 0$

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

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

$$\frac{1}{2\pi} \int_{\mathbb{R}} \cosh^2(\gamma t) |\hat{f}(t)|^2 dt \leq L. \tag{3.1}$$

For this and other facts about the class $\mathcal{F}(\gamma, L)$ consult [1] and [8].

Let X_1, X_2, \dots be a sequence of i.i.d. random variables in \mathbb{R} with common density

function f in the functional class \mathcal{F} . The *kernel-type estimator* of $f \in \mathcal{F}$ that is studied in this thesis has the form

$$f_n(x; \theta_n) := f_n(x; X_1, \dots, X_n, \theta_n) = \frac{1}{nh_n} \sum_{j=1}^n K_n \left(\frac{X_j - x}{h_n}; \theta_n \right), \quad x \in \mathbb{R}, \quad (3.2)$$

where K_n is the *Fejér-type kernel* given by

$$K_n(u; \theta_n) := \begin{cases} \frac{\cos(\theta_n u) - \cos(u)}{\pi(1 - \theta_n)u^2} & \text{if } u \neq 0, \\ \frac{1 + \theta_n}{2\pi} & \text{if } u = 0, \end{cases} \quad (3.3)$$

with

$$h_n = \frac{\theta_n}{N}, \quad \theta_n = 1 - \frac{1}{N}, \quad N = \frac{\log n}{2\gamma}. \quad (3.4)$$

The parameters h_n , θ_n , and N are chosen to satisfy

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

implying the consistency of $f_n(x; \theta_n)$ as an estimator of $f(x)$. The Fejér-type kernel, as it is referred to in [1], is a classical family of functions well known in approximation theory. The kernel-type estimator given in (3.2) was introduced by Stepanova [32] in connection with the problem of estimating f in a more general class of analytic functions using the \mathbb{L}_p -risk, $1 \leq p < \infty$. In theory, the kernel-type estimator based on the Fejér-type kernel outperforms, with respect to the \mathbb{L}_p -risk, all the other estimators existing in the literature. Moreover, it is known to be locally asymptotically minimax with respect to the \mathbb{L}_2 -risk (refer to Theorem 2 of [29] and Theorem 1 of [32]).

For some choices of the parameter θ_n , the Fejér-type kernel coincides with the well-known efficient kernels listed in Table 2.1b. 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.

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

Table 3.1: Cases of the Fejér-type kernel

The Fejér-type kernel in (3.3), like other kernels, satisfies properties (2.4). Let us, for example, show that K_n integrates to 1. For this, we will use the symmetry of K_n about zero, the trigonometric identity (see [10], formula 1.314.5)

$$\cos(x) - \cos(y) = 2 \sin\left(\frac{x+y}{2}\right) \sin\left(\frac{y-x}{2}\right), \quad (3.5)$$

and the identities (see [10], formulas 2.641.1 and 2.641.4)

$$\begin{aligned} \int \frac{\sin(kx)}{a+bx} dx &= -\frac{1}{b} \left[\cos\left(\frac{ka}{b}\right) si\left(\frac{k}{b}(a+bx)\right) - \sin\left(\frac{ka}{b}\right) ci\left(\frac{k}{b}(a+bx)\right) \right], \\ \int \frac{\cos(kx)}{(a+bx)^2} dx &= -\frac{1}{b} \frac{\cos(kx)}{a+bx} - \frac{k}{b} \int \frac{\sin(kx)}{a+bx} dx. \end{aligned} \quad (3.6)$$

We shall also use a *sine integral* (see [10], formula 8.232.1)

$$si(x) = - \int_x^\infty \frac{\sin(t)}{t} dt,$$

with the property

$$si(x) = \begin{cases} -\pi/2 & \text{if } x = 0, \\ 0 & \text{if } x = \infty. \end{cases} \quad (3.7)$$

Then, by applying (3.5), (3.6), and (3.7), we obtain

$$\begin{aligned}
\int_{\mathbb{R}} K_n(u; \theta_n) du &= \int_{\mathbb{R}} \frac{\cos(\theta_n u) - \cos(u)}{\pi(1 - \theta_n)u^2} du \\
&= \frac{2}{\pi(1 - \theta_n)} \left[\int_0^\infty \frac{\cos(\theta_n u)}{u^2} du - \int_0^\infty \frac{\cos(u)}{u^2} du \right] \\
&= \frac{2}{\pi(1 - \theta_n)} \left[-\frac{\cos(\theta_n u)}{u} - \theta_n \text{si}(\theta_n u) + \frac{\cos(u)}{u} + \text{si}(u) \right]_0^\infty \\
&= \frac{2}{\pi(1 - \theta_n)} \left[\frac{\pi}{2} (1 - \theta_n) \right] = 1.
\end{aligned}$$

It can be shown that the Fourier transform of the Fejér-type kernel is given by (see [1], p. 202)

$$\hat{K}_n(t; \theta_n) = \int_{\mathbb{R}} e^{itu} K_n(u; \theta_n) du = \mathbb{I}(|t| \leq \theta_n) + \left(\frac{1 - |t|}{1 - \theta_n} \right) \mathbb{I}(\theta_n \leq |t| \leq 1). \quad (3.8)$$

Figure 3.1 provides a graphical representation of the Fejér-type kernel and its Fourier transform. The simple form of \hat{K}_n makes it very useful in studying analytically the properties of the kernel-type estimator in (3.2). According to the simple criterion for inadmissibility defined in (2.26), the Fejér-type kernel is admissible since its Fourier transform is nonnegative and bounded by 1.

We would like to demonstrate numerically that the kernel-type estimator in (3.2) with the Fejér-type kernel in (3.3) and the theoretical bandwidth $h_n = (2\gamma\theta_n)/\log n$ cannot be improved with respect to the \mathbb{L}_2 -risk. To do so, we need to select an appropriate value for γ and then compare the finite-sample performance of the estimator in question with other estimators. First, we compare (3.2) against Fejér-type kernel estimators that use empirically selected bandwidths. Two data-driven bandwidth selection methods based on the unbiased estimation of the \mathbb{L}_2 -risk are used; namely the cross-validation and a method based on the Fourier analysis techniques. This requires the minimisation of the unbiased estimators (2.28) and (2.31), denoted by $\text{CV}(h)$ and $\tilde{J}_n(h)$, with respect to h . Next, we

compare the performance of these kernel-type estimators against that of other density estimators that use some of the most well-known (fixed) kernel functions in the literature; namely, the sinc, de la Vallée Poussin, and Gaussian kernels. The reason being that, as indicated in Table 3.1, the sinc and de la Vallée Poussin kernels are specific cases of the Fejér-type kernel, while the Gaussian kernel is the most commonly applied in practice. The remainder of this chapter considers appropriate choices of the smoothing parameter $\gamma > 0$ in $\mathcal{F}(\gamma, L)$ and the evaluation of the $\text{CV}(h)$ and $\tilde{J}_n(h)$ estimators for the Fejér-type kernel and its competing kernels.

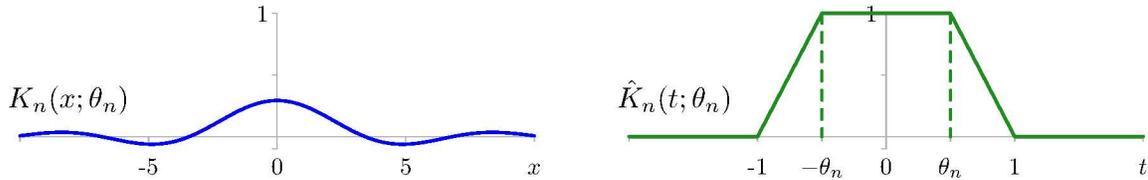


Figure 3.1: Fejér-type kernel and its Fourier transform with $\theta_n = 1 - (2\gamma)/\log n$ and $\gamma = 1/2$.

3.1 Choice of Smoothing Parameter γ

The class $\mathcal{F} = \mathcal{F}(\gamma, L)$ of the infinitely differentiable functions of interest is described by the smoothing parameter $\gamma > 0$. For appropriate values of γ , the class \mathcal{F} contains probability densities such as the normal, Student's t , and Cauchy as well as their analytic transformations and mixtures. We shall use (3.1) to see how large the values of γ can be chosen so that the normal, Student's t , and Cauchy densities belong to \mathcal{F} .

The left-hand side of inequality (3.1) can be expanded using the standard algebraic expression of $\cosh(\gamma t) = (e^{\gamma t} + e^{-\gamma t})/2$ as follows:

$$\frac{1}{2\pi} \int_{\mathbb{R}} \cosh^2(\gamma t) |\hat{f}(t)|^2 dt = \frac{1}{8\pi} \left[\int_{\mathbb{R}} (e^{2\gamma t} + e^{-2\gamma t}) |\varphi(t)|^2 dt + 2 \int_{\mathbb{R}} |\varphi(t)|^2 dt \right]. \quad (3.9)$$

The characteristic function for a normal distribution with mean μ and variance σ^2 is known to be $e^{it\mu - \frac{1}{2}\sigma^2 t^2}$. The absolute square of the characteristic function is then $e^{-\sigma^2 t^2}$. Plugging in $e^{-\sigma^2 t^2}$ into (3.9) gives

$$\begin{aligned} \frac{1}{2\pi} \int_{\mathbb{R}} \cosh^2(\gamma t) |\hat{f}(t)|^2 dt &= \frac{1}{8\pi} \left[\int_{\mathbb{R}} (e^{2\gamma t} + e^{-2\gamma t}) e^{-\sigma^2 t^2} dt + 2 \int_{\mathbb{R}} e^{-\sigma^2 t^2} dt \right] \\ &= \frac{1}{8\pi} \left[\frac{2\sqrt{\pi}}{\sigma} e^{\gamma^2/\sigma^2} + \frac{2\sqrt{\pi}}{\sigma} \right]. \end{aligned}$$

Hence, in the normal case, the integral

$$\frac{1}{2\pi} \int_{\mathbb{R}} \cosh^2(\gamma t) |\hat{f}(t)|^2 dt = \frac{1}{4\sigma\sqrt{\pi}} (e^{\gamma^2/\sigma^2} + 1)$$

is finite for all $\gamma > 0$. Therefore, the normal density is in the analytical class $\mathcal{F}(\gamma, L)$ with $\gamma > 0$ and $L = (4\sigma\sqrt{\pi})^{-1}(e^{\gamma^2/\sigma^2} + 1)$.

The characteristic function $\varphi_\nu(t)$ of the Student's t distribution with ν degrees of freedom is of a more complex form. Hurst [15] provides the following general result:

$$\varphi_\nu(t) = \frac{B_{\nu/2}(\sqrt{\nu}|t|)(\sqrt{\nu}|t|)^{\nu/2}}{\Gamma(\nu/2)2^{\nu/2-1}}, \quad t \in \mathbb{R},$$

where $\Gamma(\cdot)$ is the *Gamma function* given by $\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dt$, $x > 0$, and $B_{\nu/2}(\cdot)$ is the modified Bessel function of the third kind with the property (see, for example, (4.7) of [15])

$$\lim_{t \rightarrow \pm\infty} B_{\nu/2}(\sqrt{\nu}|t|) = \sqrt{\frac{\pi}{2\sqrt{\nu}|t|}} e^{-\sqrt{\nu}|t|}.$$

When the degrees of freedom ν are odd, the characteristic function is given by (see [15], p. 5)

$$\varphi_\nu(t) = \frac{\sqrt{\pi}\Gamma((\nu+1)/2)}{\Gamma(\nu/2)} \frac{e^{-\sqrt{\nu}|t|}}{2^{2(m-1)}(m-1)!} \sum_{j=0}^{m-1} (2j)! \binom{m-1+j}{2j} (2\sqrt{\nu}|t|)^{m-1-j}$$

$$= C_\nu e^{-\sqrt{\nu}|t|} \sum_{j=0}^{m-1} a_{\nu,j} |t|^{m-1-j}, \quad (3.10)$$

where $m = (\nu + 1)/2$ is a positive integer. We have by (3.10) that

$$|\varphi_\nu(t)|^2 = C_\nu^2 e^{-2\sqrt{\nu}|t|} \left(\sum_{j=0}^{m-1} a_{\nu,j}^2 |t|^{2(m-1-j)} + 2 \sum_{j < k} a_{\nu,j} a_{\nu,k} |t|^{2(m-1)-j-k} \right).$$

To avoid technicalities, we restrict our attention to the case of odd ν . From (3.9), we note that for $p = 0, \dots, 2m - 2$, the integral

$$\int_{\mathbb{R}} e^{2\gamma t} e^{\pm 2\sqrt{\nu}|t|} |t|^p dt$$

is finite when $2\sqrt{\nu} - 2\gamma > 0$, implying that $0 < \gamma < \sqrt{\nu}$. Therefore, the Student's t density with odd degrees of freedom ν is in $\mathcal{F}(\gamma, L)$ when $0 < \gamma < \sqrt{\nu}$.

The standard Cauchy distribution with density function

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad x \in \mathbb{R},$$

is in $\mathcal{F}(\gamma, L)$ when $0 < \gamma < 1$ (see [17], Section 2.3). This is consistent with the fact that the standard Cauchy distribution coincides with the Student's t -distribution with one degree of freedom. For other examples of functions belonging to $\mathcal{F}(\gamma, L)$ we refer to Section 2.3 of [17].

3.2 Evaluation of the Unbiased Risk Estimators

The smoothing parameters h_{CV} and h_F , obtained by the methods of cross-validation and Fourier analysis, respectively, are derived by minimising the estimators $CV(h)$ and $\tilde{J}_n(h)$ (see (2.28) and (2.31)) with respect to h . Here, we evaluate the estimators $CV(h)$ and

$\tilde{J}_n(h)$ for the Fejér-type kernel as well as the sinc, de la Vallée Poussin, and Gaussian kernels. The Fourier transform of each kernel, formula 2.633.2 in [10]:

$$\int x^n \cos(ax) dx = \sum_{k=0}^n k! \binom{n}{k} \frac{x^{n-k}}{a^{k+1}} \sin\left(ax + \frac{1}{2}k\pi\right), \quad a > 0, \quad (3.11)$$

and formula 3.323.2 in [10]:

$$\int_{\mathbb{R}} e^{-p^2 x^2 \pm qx} dx = \frac{\sqrt{\pi}}{p} e^{q^2/(4p^2)}, \quad p^2 > 0, \quad q \in \mathbb{R}, \quad (3.12)$$

are used to calculate $\text{CV}(h)$ and $\tilde{J}_n(h)$.

3.2.1 Unbiased Cross-Validation Criterion

The unbiased cross-validation criterion $\text{CV}(h)$ in (2.28) is made up of the convolution of the kernel with itself. The *convolution theorem* (see Theorem 1 in Appendix A) allows us to use Fourier transforms to calculate the convolution of two functions. Since all kernel functions K are symmetric in \mathbb{R} , then so are their Fourier transforms \hat{K} . Hence, by the convolution and Fourier inversion theorems, the convolution of a kernel function with itself can be expressed as

$$\begin{aligned} (K * K)(u) &= \frac{1}{2\pi} \int_{\mathbb{R}} (\widehat{K * K})(t) e^{-itu} dt \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} (\widehat{K * K})(t) (\cos(ut) - i \sin(ut)) dt \\ &= \frac{1}{\pi} \int_0^{\infty} (\widehat{K * K})(t) \cos(ut) dt \\ &= \frac{1}{\pi} \int_0^{\infty} \hat{K}(t) \cdot \hat{K}(t) \cos(ut) dt. \end{aligned} \quad (3.13)$$

We proceed to evaluate the self-convolution of each kernel in the study starting with the Fejér-type kernel.

Following from (3.13), we use the Fourier transform in (3.8) of the Fejér-type kernel to obtain the Fourier transform of its self-convolution:

$$(\widehat{K_n * K_n})(t; \theta_n) = \hat{K}_n(t; \theta_n) \cdot \hat{K}_n(t; \theta_n) = \mathbb{I}(|t| \leq \theta_n) + \frac{(1 - |t|)^2}{(1 - \theta_n)^2} \mathbb{I}(\theta_n \leq |t| \leq 1), \quad (3.14)$$

where $0 \leq \theta_n < 1$. The graphical representation of (3.14) is given in Figure 3.2. Using (3.14), we continue from (3.13):

$$\begin{aligned} (K_n * K_n)(u; \theta_n) &= \frac{1}{\pi} \int_0^\infty (\widehat{K_n * K_n})(t; \theta_n) \cos(ut) dt \\ &= \frac{1}{\pi} \left[\int_0^{\theta_n} \cos(ut) dt + \frac{1}{(1 - \theta_n)^2} \int_{\theta_n}^1 (1 - t)^2 \cos(ut) dt \right]. \end{aligned}$$

Applying equality (3.11) we obtain

$$\begin{aligned} (K_n * K_n)(u; \theta_n) &= \\ &= \frac{\sin(\theta_n u)}{\pi u} + \left(\frac{\sin(ut)}{\pi(1 - \theta_n)^2 u} - 2 \frac{ut \sin(ut) + \cos(ut)}{\pi(1 - \theta_n)^2 u^2} + \frac{(u^2 t^2 - 2) \sin(ut) + 2ut \cos(ut)}{\pi(1 - \theta_n)^2 u^3} \right) \Big|_{\theta_n}^1. \end{aligned}$$

Simple algebraic manipulation allows us to simplify the above relation to the following

$$(K_n * K_n)(u; \theta_n) = \frac{2 \cos(\theta_n u)}{\pi(1 - \theta_n)u^2} + \frac{2(\sin(\theta_n u) - \sin(u))}{\pi(1 - \theta_n)^2 u^3}, \quad (3.15)$$

for $u \neq 0$. We now use the Taylor series expansion to evaluate the limit of (3.15) as u approaches 0. Expanding the cosine and sine functions as $u \rightarrow 0$ gives

$$\begin{aligned} \cos(\theta_n u) &= 1 - \frac{(\theta_n u)^2}{2} + o((\theta_n u)^3), \\ \sin(\theta_n u) &= \theta_n u - \frac{(\theta_n u)^3}{6} + o((\theta_n u)^4), \\ \sin(u) &= u - \frac{u^3}{6} + o(u^4). \end{aligned} \quad (3.16)$$

Substituting (3.16) into (3.15) we obtain by continuity that

$$(K_n * K_n)(0; \theta_n) = \frac{1 + 2\theta_n}{3\pi}.$$

Finally, the self-convolution of the Fejér-type kernel is given by

$$(K_n * K_n)(u; \theta_n) = \begin{cases} \frac{2 \cos(\theta_n u)}{\pi(1 - \theta_n)u^2} + \frac{2(\sin(\theta_n u) - \sin(u))}{\pi(1 - \theta_n)^2 u^3} & \text{if } u \neq 0, \\ \frac{1 + 2\theta_n}{3\pi} & \text{if } u = 0, \end{cases} \quad (3.17)$$

where $0 \leq \theta_n < 1$ for $n > 1$.

Now, we can easily obtain the self-convolutions of the de la Vallée Poussin and sinc kernels as they are special cases of the Fejér-type kernel. Recall that for $\theta_n = 1/2$ the Fejér-type kernel coincides with the de la Vallée Poussin kernel. Therefore, the convolution of the de la Vallée Poussin kernel is given by

$$(K * K)(u) = \begin{cases} \frac{4 \cos(u/2)}{\pi u^2} + \frac{8(\sin(u/2) - \sin(u))}{\pi u^3} & \text{if } u \neq 0, \\ \frac{2}{3\pi} & \text{if } u = 0. \end{cases} \quad (3.18)$$

The sinc kernel is the limiting case of the Fejér-type kernel when θ_n approaches 1. So, the limiting case of (3.17) when θ_n approaches 1 gives the self-convolution of the sinc kernel, which is, in fact, the sinc kernel

$$(K * K)(u) = \begin{cases} \frac{\sin u}{\pi u} & \text{if } u \neq 0, \\ \frac{1}{\pi} & \text{if } u = 0. \end{cases} \quad (3.19)$$

It remains to find the convolution of the Gaussian kernel $K(u) = (1/\sqrt{2\pi})e^{-u^2/2}$.

From (3.13) with $\hat{K}(t) = e^{-t^2/2}$ and using the formula

$$\int_0^\infty e^{-a^2 x^2} \cos(bx) dx = \frac{\sqrt{\pi}}{2a} e^{-b^2/4a^2}, \quad a > 0,$$

the self-convolution of the Gaussian kernel is given by

$$(K * K)(u) = \frac{1}{\sqrt{4\pi}} e^{-u^2/4}, \quad u \in \mathbb{R}. \quad (3.20)$$

Therefore, the cross-validated bandwidth h_{CV} can be determined for the Fejér-type, sinc, de la Vallée Poussin, and Gaussian kernels by using their self-convolutions (3.17) – (3.20) to estimate $\widehat{CV}(h; x_1, \dots, x_n)$ as in (2.28) from the observed values x_1, \dots, x_n of the random sample X_1, \dots, X_n . The estimate \hat{h}_{CV} is the value of h that minimises $\widehat{CV}(h; x_1, \dots, x_n)$ and is selected as the smoothing parameter of the cross-validation kernel estimate $f_{n,CV}(x; x_1, \dots, x_n)$.

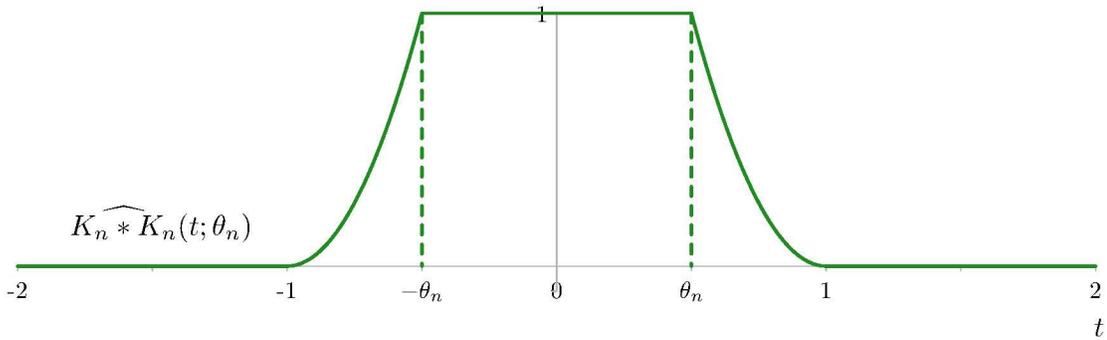


Figure 3.2: *Fourier transform of the self-convolution of the Fejér-type kernel, where $\theta_n = 1 - (2\gamma)/\log n$ and $\gamma = 1/2$*

3.2.2 Unbiased Risk Estimator based on the Fourier Analysis of Kernel Density Estimators

The approximately unbiased estimator $\tilde{J}_n(h)$ in (2.31) consists of the Fourier transform of the kernel and its square, as well as the absolute square of the empirical characteristic function given by

$$\begin{aligned} |\varphi_n(t)|^2 &= \varphi_n(t)\varphi_n(-t) = \frac{1}{n^2} \sum_{j \neq k} e^{i(X_j - X_k)t} + \frac{1}{n} \\ &= \frac{1}{n^2} \sum_{j \neq k} [\cos((X_j - X_k)t) + i \sin((X_j - X_k)t)] + \frac{1}{n}. \end{aligned} \quad (3.21)$$

We evaluate $\tilde{J}_n(h)$ for each kernel in the study starting with the Fejér-type kernel. Note that $P_f(X_j = X_k) = 0$ for $1 \leq j \neq k \leq n$.

Plugging the Fourier transform as in (3.8) into (2.31) gives

$$\begin{aligned} \tilde{J}_n(h; \theta_n) &= \int_{\mathbb{R}} \left[-2 \left(\mathbb{I}(|ht| \leq \theta_n) + \frac{1 - |ht|}{1 - \theta_n} \mathbb{I}(\theta_n \leq |ht| \leq 1) \right) \right. \\ &\quad \left. + \left(\mathbb{I}(|ht| \leq \theta_n) + \left(\frac{1 - |ht|}{1 - \theta_n} \right)^2 \mathbb{I}(\theta_n \leq |ht| \leq 1) \right) \left(1 - \frac{1}{n} \right) \right] |\varphi_n(t)|^2 dt + \frac{4\pi(1 + \theta_n)}{nh \cdot 2\pi}, \end{aligned}$$

which can be further simplified to

$$\begin{aligned} \tilde{J}_n(h; \theta_n) &= - \left(1 + \frac{1}{n} \right) \int_{|t| \leq \theta_n/h} |\varphi_n(t)|^2 dt - \frac{2}{1 - \theta_n} \int_{\theta_n/h \leq |t| \leq 1/h} (1 - |ht|) |\varphi_n(t)|^2 dt \\ &\quad + \frac{n-1}{n(1 - \theta_n)^2} \int_{\theta_n/h \leq |t| \leq 1/h} (1 - |ht|)^2 |\varphi_n(t)|^2 dt + \frac{2(1 + \theta_n)}{nh}. \end{aligned}$$

Using (3.21) and the fact that the cosine function is symmetric about 0, we may continue

$$\begin{aligned}
\tilde{J}_n(h; \theta_n) = & - \left(1 + \frac{1}{n}\right) \left[\int_{|t| \leq \theta_n/h} \left(\frac{1}{n^2} \sum_{j \neq k} \cos((X_j - X_k)t) + \frac{1}{n} \right) dt \right] \\
& - \frac{2}{1 - \theta_n} \left[\int_{\theta_n/h \leq |t| \leq 1/h} (1 - |ht|) \left(\frac{1}{n^2} \sum_{j \neq k} \cos((X_j - X_k)t) + \frac{1}{n} \right) dt \right] \\
& + \frac{n-1}{n(1 - \theta_n)^2} \left[\int_{\theta_n/h \leq |t| \leq 1/h} (1 - |ht|)^2 \left(\frac{1}{n^2} \sum_{j \neq k} \cos((X_j - X_k)t) + \frac{1}{n} \right) dt \right] + \frac{2(1 + \theta_n)}{nh}.
\end{aligned}$$

Further expansion and formula (3.11) give

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

Plugging in the endpoints results in

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

$$+ \frac{(n-1)}{n(1-\theta_n)^2} \left\{ \frac{2}{n^2} \sum_{j \neq k} \left(\frac{2h^2 (\sin(\theta_n(X_j - X_k)/h) - \sin((X_j - X_k)/h))}{(X_j - X_k)^3} \right. \right. \\ \left. \left. + \frac{2h(1-\theta_n) \cos(\theta_n(X_j - X_k)/h)}{(X_j - X_k)^2} - \frac{(1-\theta_n)^2 \sin(\theta_n(X_j - X_k)/h)}{X_j - X_k} \right) + \frac{2(1-\theta_n)^3}{3nh} \right\}.$$

Finally, after simple algebra, the unbiased risk estimator based on the Fourier analysis of a density estimator with the Fejér-type kernel is given by:

$$\begin{aligned} \tilde{J}_n(h; \theta_n) &= \frac{2}{n^2} \left[\frac{2h^2}{(1-\theta_n)^2} \left(\frac{n-1}{n} \right) \sum_{j \neq k} \frac{\sin(\theta_n(X_j - X_k)/h) - \sin((X_j - X_k)/h)}{(X_j - X_k)^3} \right. \\ &\quad \left. + \frac{2h}{(1-\theta_n)} \sum_{j \neq k} \frac{\cos((X_j - X_k)/h) - \cos(\theta_n(X_j - X_k)/h)/n}{(X_j - X_k)^2} \right] \\ &\quad - \frac{2(n+1)\theta_n}{n^2h} + \frac{2(n-1)(1-\theta_n)}{3n^2h} + \frac{4\theta_n}{nh}, \end{aligned} \quad (3.22)$$

for $n > 1$, $h > 0$, $0 \leq \theta_n < 1$, and $X_j \neq X_k$, $1 \leq j < k \leq n$.

We may now use (3.22) to obtain $\tilde{J}_n(h; \theta_n)$ for the de la Vallée Poussin and sinc kernels. Substituting $\theta_n = 1/2$ into (3.22) gives the following unbiased risk estimator for the de la Vallée Poussin kernel:

$$\begin{aligned} \tilde{J}_n(h; 1/2) &= \frac{2}{n^2} \left[8h^2 \left(\frac{n-1}{n} \right) \sum_{j \neq k} \frac{\sin((X_j - X_k)/2h) - \sin((X_j - X_k)/h)}{(X_j - X_k)^3} \right. \\ &\quad \left. + 4h \sum_{j \neq k} \frac{\cos((X_j - X_k)/h) - \cos((X_j - X_k)/2h)/n}{(X_j - X_k)^2} \right] - \frac{2(n+2)}{3n^2h} + \frac{2}{nh}, \end{aligned} \quad (3.23)$$

for all $n > 1$, $h > 0$, and $X_j \neq X_k$, $1 \leq j < k \leq n$. In the case of the sinc kernel we have by (2.31) the following unbiased risk estimator:

$$\tilde{J}_n(h; 1) = - \left(1 + \frac{1}{n} \right) \frac{2}{n^2} \sum_{j \neq k} \frac{\sin((X_j - X_k)/h)}{X_j - X_k} - \frac{2}{n^2h} + \frac{2}{nh}, \quad (3.24)$$

for all $n > 1$, $h > 0$, and $X_j \neq X_k$, $1 \leq j < k \leq n$.

We now evaluate $\tilde{J}_n(h)$ for the Gaussian kernel. Plugging the Fourier transform of the

kernel into (2.31) gives

$$\begin{aligned}\tilde{J}_n(h) &= \int_{\mathbb{R}} \left(-2e^{-(ht)^2/2} + e^{-(ht)^2} \left(1 - \frac{1}{n} \right) \right) |\varphi_n(t)|^2 dt + \frac{4\pi}{\sqrt{2\pi}nh} \\ &= -2 \int_{\mathbb{R}} e^{(ht)^2/2} |\varphi_n(t)|^2 dt + \left(1 - \frac{1}{n} \right) \int_{\mathbb{R}} e^{-(ht)^2} |\varphi_n(t)|^2 dt + \frac{2\sqrt{2\pi}}{nh}.\end{aligned}$$

From this, applying (3.21), we get

$$\begin{aligned}\tilde{J}_n(h) &= -\frac{2}{n^2} \sum_{j \neq k} \int_{\mathbb{R}} e^{-(ht)^2/2 + i(X_j - X_k)t} dt - \frac{2}{n} \int_{\mathbb{R}} e^{-(ht)^2/2} dt \\ &+ \frac{n-1}{n^3} \sum_{j \neq k} \int_{\mathbb{R}} e^{-(ht)^2 + i(X_j - X_k)t} dt + \frac{n-1}{n^2} \int_{\mathbb{R}} e^{-(ht)^2} dt + \frac{2\sqrt{2\pi}}{nh},\end{aligned}$$

which, in view of formula (3.12), becomes

$$\begin{aligned}\tilde{J}_n(h) &= -\frac{2\sqrt{2\pi}}{n^2h} \sum_{j \neq k} e^{-(X_j - X_k)^2/2h^2} - \frac{2\sqrt{2\pi}}{nh} \\ &+ \frac{(n-1)\sqrt{\pi}}{n^3h} \sum_{j \neq k} e^{-(X_j - X_k)^2/4h^2} + \frac{(n-1)\sqrt{\pi}}{n^2h} + \frac{2\sqrt{2\pi}}{nh}.\end{aligned}$$

Finally, the unbiased risk estimator for the Gaussian kernel is equal to

$$\tilde{J}_n(h) = \frac{\sqrt{\pi}}{n^2h} \left[(n-1) \left(1 + \frac{1}{n} \sum_{j \neq k} e^{-(X_j - X_k)^2/4h^2} \right) - 2\sqrt{2} \sum_{j \neq k} e^{-(X_j - X_k)^2/2h^2} \right], \quad (3.25)$$

for all $n > 1$ and $h > 0$.

Therefore, the bandwidth h_F can be determined for the Fejér-type, sinc, de la Vallée Poussin, and Gaussian kernels by estimating $\tilde{J}_n(h)$ in (3.22) – (3.25) from the observed values x_1, \dots, x_n of the random sample X_1, \dots, X_n . The estimate \hat{h}_F is the value of h that minimises the estimated $\hat{J}_n(h; x_1, \dots, x_n)$ and is selected as the smoothing parameter of the kernel estimate $f_{n,F}(x; x_1, \dots, x_n)$.

Chapter 4

Simulation Studies

We carried out simulation experiments to assess the performance of the proposed kernel-type estimator given by (3.2) with the Fejér-type kernel and the theoretical bandwidth, as in (3.3) and (3.4). The MISE criterion is used to compare the efficiency of the kernel-type estimator with that of other estimators that use some of the most well-known (fixed) kernel functions in the literature; namely the sinc, de la Vallée Poussin, and Gaussian. 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 is the most commonly applied kernel in practice. We use three different smoothing parameters for the kernel estimators in hand: the theoretical bandwidth and two empirical bandwidths. The two data-dependent bandwidths are selected using the cross-validation criterion and a method obtained from Fourier analysis techniques. Various combinations of the parameter $\gamma > 0$ and the sample size $n > 1$ are also explored. The simulation studies are run in the R statistical programming environment. The R code of all the functions is provided in Appendix B. The methodology used to conduct the simulations is discussed followed by a summary of the numerical results.

4.1 Methodology

We are interested in comparing the performance of the kernel-type estimator against that of the other fixed-kernel density estimators using empirical and theoretical bandwidths. In order to construct the estimators and compute their risks, we use the following algorithm.

1. Choose a large positive number M , such as 100, 200, \dots , and generate M random samples $x_1^{(i)}, \dots, x_n^{(i)}$, $i = 1, \dots, M$, of sample size n from some probability density $f \in \mathcal{F}(\gamma, L)$.
2. Choose an appropriate positive value for γ that satisfies condition (3.1) so that the probability density f selected in step 1 belongs to the functional class $\mathcal{F}(\gamma, L)$.
3. Create a uniform grid for the values of h by specifying a minimum value $h_{(1)}$ and a maximum value $h_{(J)}$ over the interval $[h_{(1)}, h_{(J)}]$ with the distance between the grid points $h_{(j)}$ and $h_{(j+1)}$ equal to some small specified number such as 0.01. For each M generated sample, find the optimal bandwidth by the cross-validation criterion using the data obtained in step 1, the value of γ from step 2, and the grid of bandwidths $[h_{(1)}, \dots, h_{(J)}]$, for each kernel under consideration:

$$\hat{h}_{\text{CV}}^{(i)} = \arg \min_{h \in [h_{(1)}, \dots, h_{(J)}]} \widehat{\text{CV}}(h; x_1^{(i)}, \dots, x_n^{(i)}), \quad i = 1, \dots, M.$$

Recall that the $\text{CV}(\cdot)$ for each kernel is evaluated using the convolution formulas (3.17) – (3.20). Do the same for the optimal bandwidth obtained from the Fourier analysis techniques:

$$\hat{h}_{\text{F}}^{(i)} = \arg \min_{h \in [h_{(1)}, \dots, h_{(J)}]} \hat{J}_n(h; x_1^{(i)}, \dots, x_n^{(i)}), \quad i = 1, \dots, M,$$

where $\hat{J}_n(\cdot)$ is evaluated using (3.22) – (3.25) for each kernel of interest.

4. Use the sample size n and the value of γ from step 2 to evaluate the theoretical smoothing parameter $h_n = (2\gamma\theta_n)/\log n$, $\theta_n = 1 - (2\gamma)/\log n$, for the Fejér-type estimator and the other estimators in hand. Note that the theoretical bandwidths for the sinc and de la Vallée Poussin estimators are $h_n = (2\gamma)/\log n$ and $h_n = \gamma/\log n$, respectively.
5. Use the data from step 1, the γ value from step 2, and the results of the bandwidths from steps 3 and 4 to compute, for each of the M generated samples, the density estimates

$$f_{n,\text{CV}}^{(i)}(x; x_1^{(i)}, \dots, x_n^{(i)}) = \frac{1}{n\hat{h}_{\text{CV}}^{(i)}} \sum_{j=1}^n K\left(\frac{x_j^{(i)} - x}{\hat{h}_{\text{CV}}^{(i)}}\right),$$

$$f_{n,\text{F}}^{(i)}(x; x_1^{(i)}, \dots, x_n^{(i)}) = \frac{1}{n\hat{h}_{\text{F}}^{(i)}} \sum_{j=1}^n K\left(\frac{x_j^{(i)} - x}{\hat{h}_{\text{F}}^{(i)}}\right),$$

$$f_n^{(i)}(x; x_1^{(i)}, \dots, x_n^{(i)}) = \frac{1}{nh_n} \sum_{j=1}^n K\left(\frac{x_j^{(i)} - x}{h_n}\right),$$

$i = 1, \dots, M$, for each kernel under consideration.

6. Get the positive part for each kernel density estimate obtained in step 5:

$$f_{n,\text{CV}}^{+(i)}(x) = \max(0, f_{n,\text{CV}}^{(i)}(x; x_1^{(i)}, \dots, x_n^{(i)})),$$

$$f_{n,\text{F}}^{+(i)}(x) = \max(0, f_{n,\text{F}}^{(i)}(x; x_1^{(i)}, \dots, x_n^{(i)})),$$

$$f_n^{+(i)}(x) = \max(0, f_n^{(i)}(x; x_1^{(i)}, \dots, x_n^{(i)})),$$

for $i = 1, \dots, M$.

7. To compare the performances of each kernel density estimate for each bandwidth method, compute the MISE estimate for each quantity obtained in step 6, which is

defined by

$$\begin{aligned} \text{MISE}(f_n^+, f) &= \mathbb{E}_f \left[\int_{\mathbb{R}} (f_n^+(x) - f(x))^2 dx \right] \\ &\cong \frac{1}{M} \sum_{i=1}^M \int_{\mathbb{R}} (f_n^{+(i)}(x) - f(x))^2 dx =: \widehat{\text{MISE}}(f_n^+, f). \end{aligned} \quad (4.1)$$

Crude Monte Carlo (CMC) integration methods (see [24], Section 3.2) are used to evaluate the MISE estimate in (4.1). The basic idea behind CMC integration is to compute complicated definite integrals using random numbers. This is accomplished by evaluating the expected value of a function g of random variable X given that X follows a standard uniform distribution $U(0, 1)$, that is,

$$\mathbb{E}_f[g(X)] = \int_{\mathbb{R}} g(x)f(x) dx = \int_0^1 g(x) dx. \quad (4.2)$$

Let X_1, \dots, X_W, \dots be a sequence of random variabls from $U(0, 1)$. Then, by the strong law of large numbers (see Theorem 5 in Appendix A), the average \bar{g}_W converges almost surely to $\mathbb{E}_f[g(X)]$:

$$\bar{g}_W = \sum_{w=1}^W \frac{g(X_w)}{W} \xrightarrow[W \rightarrow \infty]{a.s.} \mathbb{E}_f[g(X)],$$

provided $\text{Var}_f(g(X_1)) < \infty$. For the functions employed in our simulation study, the latter condition is fulfilled.

The integrated squared error (ISE) of the estimator f_n^+ can be expressed in the form of (4.2). First, we separate the ISE into the following two integrals

$$\begin{aligned} \text{ISE}(f_n^+, f) &= \int_{\mathbb{R}} (f_n^+(x) - f(x))^2 dx \\ &= \int_{-\infty}^0 (f_n^+(x) - f(x))^2 dx + \int_0^{\infty} (f_n^+(s) - f(s))^2 ds. \end{aligned}$$

We then obtain (0,1) bounds by defining the variables $y = 1/(1-x)$ and $z = 1/(1+s)$. Substituting $x = 1 - 1/y$ and $dx/dy = 1/y^2$ into the first integral and $s = 1/z - 1$ and $ds/dz = -1/z^2$ into the second integral gives

$$\text{ISE}(f_n^+, f) = \int_0^1 g(y) dy + \int_0^1 g(z) dz,$$

where

$$g(y) = \left(f_n^+ \left(1 - \frac{1}{y} \right) - f \left(1 - \frac{1}{y} \right) \right)^2 \frac{1}{y^2} \quad (4.3)$$

and

$$g(z) = \left(f_n^+ \left(\frac{1}{z} - 1 \right) - f \left(\frac{1}{z} - 1 \right) \right)^2 \frac{1}{z^2}. \quad (4.4)$$

Now we may apply the CMC method to evaluate the MISE estimate in (4.1) using the following algorithm:

- I. Choose a very large positive number W (i.e. 5000, 10000, ...) and generate two random samples, y_1, \dots, y_W and z_1, \dots, z_W , of size W from a standard uniform distribution $U(0, 1)$.
- II. Evaluate the functions $g(y_w)$ and $g(z_w)$, $w = 1, \dots, W$, defined in (4.3) and (4.4), respectively, for each of the W realized values.
- III. Find the average of $g(y_w)$ and $g(z_w)$ over W replications,

$$\text{ISE}_W(\widehat{f}_n^+, f) = \frac{1}{W} \sum_{w=1}^W [g(y_w) + g(z_w)].$$

IV. Evaluate the CMC MISE estimate:

$$\begin{aligned} \text{MISE}_W(\widehat{f}_n^+, f) &:= \frac{1}{M} \sum_{i=1}^M \left\{ \frac{1}{W} \sum_{w=1}^W \left[\left(f_n^{+(i)} \left(1 - \frac{1}{y_w} \right) - f \left(1 - \frac{1}{y_w} \right) \right)^2 \frac{1}{y_w^2} \right. \right. \\ &\quad \left. \left. + \left(f_n^{+(i)} \left(\frac{1}{z_w} - 1 \right) - f \left(\frac{1}{z_w} - 1 \right) \right)^2 \frac{1}{z_w^2} \right] \right\} \xrightarrow[W \rightarrow \infty]{a.s.} \text{MISE}(\widehat{f}_n^+, f). \end{aligned} \quad (4.5)$$

The numerical results obtained from the above methodology are summarised in the next section.

4.2 Numerical Results

To demonstrate the good performance of the kernel-type estimator given by (3.2) with the Fejér-type kernel (3.3) and the theoretical bandwidth (3.4), we apply the methodology of Section 4.1. 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 $0.4N(0, 1) + 0.6N(1, 0.4^2)$. We chose these distributions 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 $\mathcal{F}(\gamma, L)$ defined in Chapter 3; nonetheless, we are interested in observing the behaviour of the proposed kernel-type estimator when estimating such densities.

For each simulated dataset, density estimates are computed for every kernel function and bandwidth selection method under consideration. The kernels of interest include the efficient Fejér-type, sinc, and de la Vallée Poussin (dVLP) kernels, as well as the classical Gaussian kernel. The bandwidth selection methods include the empirical ap-

proaches from cross-validation and Fourier analysis and the theoretical smoothing parameter $h_n = (2\gamma\theta_n)/\log n$, which we use for the efficient estimators. For concision, in this section we shall refer to the bandwidth h_F based on the Fourier analysis of kernel density estimators as the Fourier bandwidth. The parameter γ that appears in the Fejér-type kernel and the theoretical bandwidth was manually selected. Several choices of γ were explored. The values of γ that provided the smallest MISE estimates were chosen. The density estimates are then compared in terms of their MISE estimates, calculated from (4.5). All the numerical results obtained in the study are given in Appendix C.

The MISE estimates of each kernel estimate of each density under study are plotted in Figures 4.1 and 4.3, against sample sizes 25 to 100, and Figures 4.2 and 4.4, against sample sizes 200 to 1000. Figures 4.1 and 4.2 capture the performance of each bandwidth selection method for each kernel density estimate, while Figures 4.3 and 4.4 compare the efficiency of the kernel functions for each density estimate under a specified bandwidth selector. The figures are presented as an array of plots. Each row of plots corresponds to the density function being estimated. The first, second, third, and fourth rows estimate, respectively, the standard normal, Student's t_{15} , chi-square χ_4^2 , and normal mixture. On the other hand, each column of plots corresponds to the kernel density estimate being examined. For Figures 4.1 and 4.2, density estimates that use the Fejér-type, sinc, dVP, and Gaussian kernels are indicated, respectively, by the the first, second, third, and fourth columns; whereas, for Figures 4.3 and 4.4, kernel estimates that use the cross-validation, Fourier, and theoretical bandwidths are indicated, respectively, by the first, second, and third columns. For easier detection, each density estimate pertains to a hue: green, orange, purple, and pink for estimates with Fejér-type, sinc, dVP, and Gaussian kernels, respectively. The values of γ that provide good results for the MISE are denoted by γ^* in the figures to follow. From studying Figures 4.1 – 4.4, the estimates of the MISE decrease as the sample size increases as expected.

Let us first assess the bandwidth selection methods under consideration. From Fig-

ures 4.1 and 4.2 we observe the following. For a good choice of γ , the estimates of $f \in \mathcal{F}(\gamma, L)$ with efficient kernel functions and theoretical bandwidths complement the results of Theorem 1 in [32] and the theorems in [9] and [29] by outperforming the estimates with empirical bandwidths. The theoretical estimates do not perform as well, though, when estimating the chi-square density, which is not in $\mathcal{F}(\gamma, L)$, particularly for smaller sample sizes. It can also be seen that, when estimating the unimodal densities, the method based on Fourier analysis for estimating h is better than, or equal to, the cross-validation method. The difference in estimation error is especially noticeable for smaller sample sizes. When estimating the bimodal density, the bandwidth from Fourier analysis is not as favourable for estimates with Fejér-type and dIVP kernels; nonetheless, we see that all three bandwidth methods eventually converge as the sample size increases. For estimates with Gaussian kernels, the cross-validation method outperforms the Fourier method at $n > 800$.

We next examine the observations in Figures 4.3 and 4.4 to compare the efficiency of the kernel functions. Let us first compare the results of the fixed kernel functions. Between the sinc and dIVP kernels, which are asymptotically efficient and special cases of the Fejér-type kernel, the performance of the dIVP is superior regardless of the bandwidth method, except when estimating the bimodal density for certain n . The sinc kernel does not perform very well for small sample sizes and is notably poorer when estimating the chi-square density. From the theory, we know that the sinc kernel behaves very well for large sample sizes when estimating $f \in \mathcal{F}(\gamma, L)$, but, according to our results, the dIVP is generally more favourable for $n \leq 1000$. The Gaussian kernels have similar results as the other kernels for $n \leq 100$, or sometimes better like in the case of the chi-square distribution, but, as can be seen from the plots, it is surpassed by the asymptotically efficient kernels as the sample size increases. Indeed, the slow rate of convergence in the case of the Gaussian kernel can clearly be seen when compared to the more efficient kernel functions. Generally speaking, for sample sizes $n < 1000$, the dIVP kernel performs best, followed by the sinc kernel, and

then the Gaussian kernel.

Now, we assess the Fejér-type kernel, which is dependent on the sample size n . For estimation of the unimodal densities with data-dependent bandwidth methods, the Fejér-type kernel is slightly more improved than the other fixed efficient kernels and much better than the common Gaussian kernel for larger sample sizes. Observe from Figure 4.5 that the parameter $0 \leq \theta_n < 1$, which is dependent on n and the choice of γ , has a somewhat positive relationship with the sample size. In other words, the value of θ_n increases with n , save for very small samples ($n \leq 100$) where there is a bit of variability. It can also be seen that, for sample sizes $n \leq 1000$, the values of θ_n are generally less than 0.5 except when estimating the normal mixture with the theoretical bandwidth. This explains why the results of the Fejér-type kernel are similar to those of the dlVP and why the dlVP kernel performs better than the sinc. For estimation of the unimodal densities with theoretical bandwidths and a good choice of γ , it is the dlVP kernel that performs best followed closely by the Fejér-type and then the sinc. This is perhaps due to the sample sizes $n < 1000$ being considered too small for the kernel-type estimator to be asymptotically more efficient than the others. Finally, we observe that, when estimating the bimodal density, the Fejér-type kernel performs much better than all of its competing kernels with both the empirical and theoretical bandwidths, especially for large sample sizes.

Figure 4.6 displays the estimations of the standard normal, Student's t_{15} , chi-square χ_4^2 , and normal mixture $0.4N(0, 1) + 0.6N(1, 0.4^2)$ with density estimates that use the Fejér-type, sinc, dlVP, and Gaussian kernels, for sample sizes of 50, 250, 500, and 750. The estimates with asymptotically efficient kernels use the theoretical bandwidth h_n , while the estimates with Gaussian kernels use the cross-validation bandwidth \hat{h}_{CV} . All the kernel estimates are obtained from averaging over 200 simulated estimates,

$$\bar{f}_n^+(x) = \frac{1}{200} \sum_{i=1}^{200} f_n^{+(i)}(x; x_1^{(i)}, \dots, x_n^{(i)}).$$

The density functions are illustrated by the solid black curves while the kernel estimates \bar{f}_n^+ are represented by the coloured dashed lines. Note that the kernel-type estimator of interest captures the symmetric, skewed, and bimodal characteristics of the densities in quite a satisfactory manner.

In summary, as expected from the theory, for a good choice of γ , the estimates of $f \in \mathcal{F}(\gamma, L)$ with efficient kernel functions and theoretical bandwidths outperform the estimates with empirical bandwidths. If one were to use a data-driven bandwidth technique, then we recommend the method based on the Fourier analysis of a kernel density estimator as it was shown to provide more accurate results than the cross-validation approach regardless of the kernel function. Also, the method from Fourier analysis is easier to implement, more accurate for small sample sizes, and less time-consuming when computing large samples than the cross-validation method. Moreover, the kernel-type estimator as in (3.2) compares favourably with competing kernel density estimators in terms of the estimated MISE, especially when estimating the normal mixture. Estimates that use dIVP kernels and theoretical h_n perform the best when estimating unimodal densities for $25 \leq n \leq 1000$ followed very closely by the kernel-type estimates with Fejér-type kernels. With a good choice of γ , the latter also performs very well when using the empirical bandwidths and improves the results of the fixed kernel functions, particularly those of the sinc kernel for smaller sample sizes. The value of γ should be chosen accordingly so that θ_n is in conjunction with the sample size; in other words, if n is small, then θ_n should be small and vice versa. The simulation results therefore indicate that, for a good choice of γ , the kernel-type estimator with the Fejér-type kernel performs very well when using both empirical and theoretical bandwidths to estimate densities in $\mathcal{F}(\gamma, L)$ and is reliable in application.

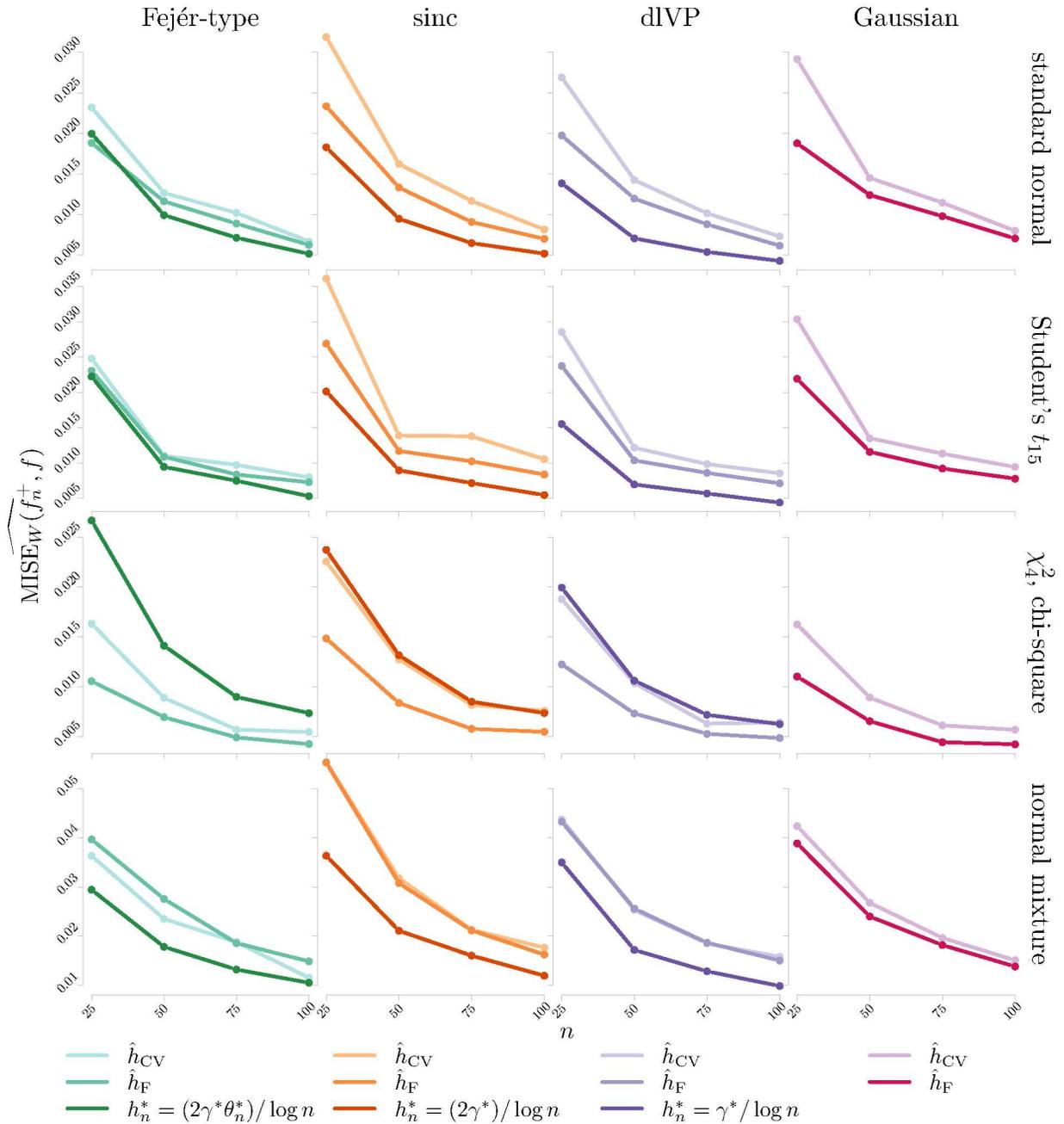


Figure 4.1: MISE estimates for the cross-validation, Fourier, and theoretical bandwidth selectors with Fejér-type, sinc, dLVP, and Gaussian kernels that estimate the standard normal, Student's t_{15} , chi-square χ_4^2 , and normal mixture, for small sample sizes. The MISE estimates are averaged over 200 replications.

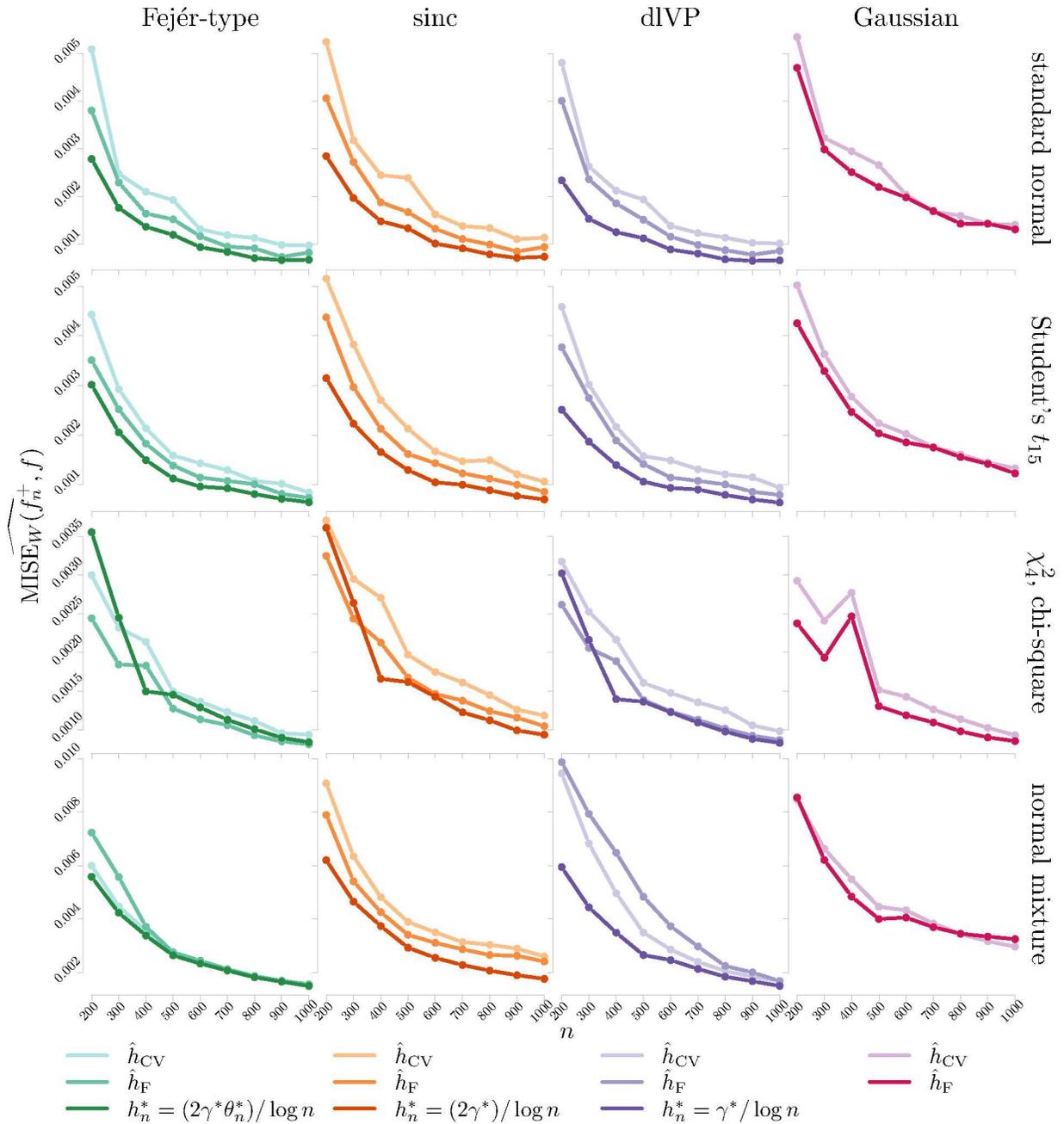


Figure 4.2: MISE estimates for the cross-validation, Fourier, and theoretical bandwidth selectors with Fejér-type, sinc, dLVP, and Gaussian kernels that estimate the standard normal, Student's t_{15} , chi-square χ^2_4 , and normal mixture, for large sample sizes. The MISE estimates are averaged over 200 replications

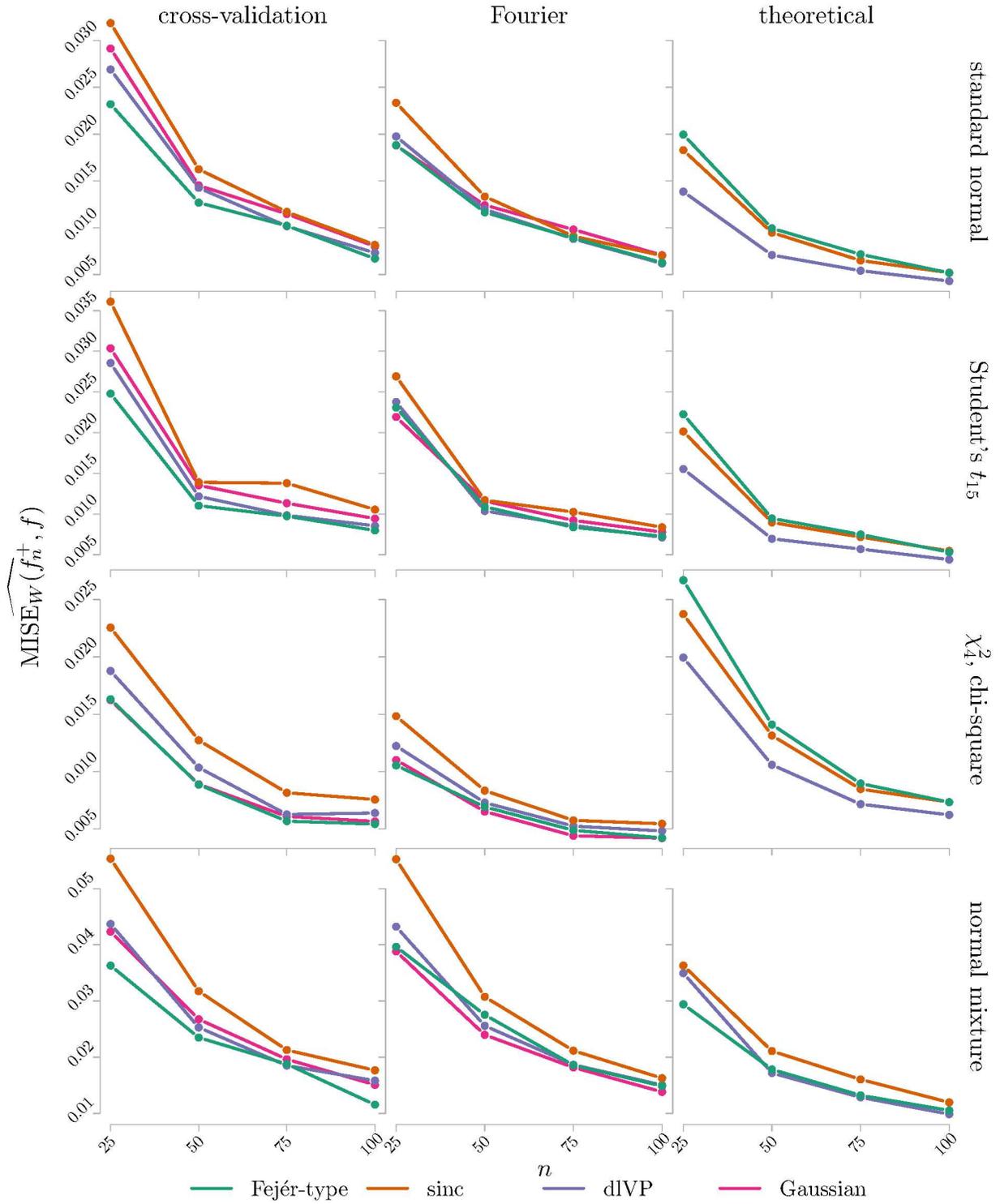


Figure 4.3: MISE estimates for the Fejér-type, sinc, dLVP, and Gaussian kernels with cross-validation, Fourier, and theoretical bandwidth selectors that estimate a standard normal, Student's t_{15} , chi-square χ^2_4 , and normal mixture, for small sample sizes. The MISE estimates are averaged over 200 replications.

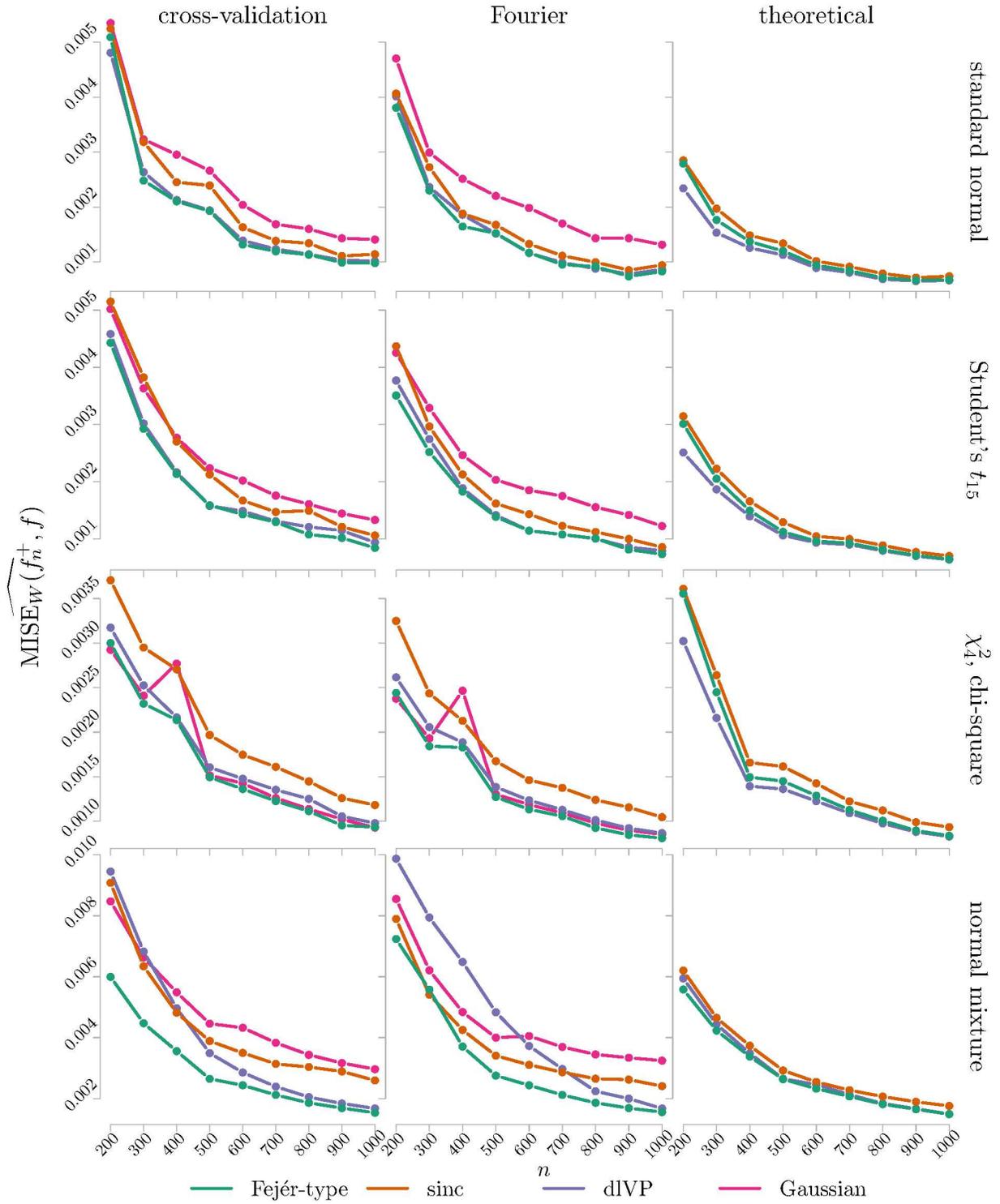


Figure 4.4: MISE estimates for the Fejér-type, sinc, dLVP, and Gaussian kernels with cross-validation, Fourier, and theoretical bandwidth selectors that estimate a standard normal, Student's t_{15} , chi-square χ^2_4 , and normal mixture, for large sample sizes. The MISE estimates are averaged over 200 replications.



Figure 4.5: Optimal values of θ_n for the Fejér-type kernel with the cross-validation, Fourier, and theoretical bandwidths when estimating the standard normal, Student's t_{15} , chi-square χ_4^2 , and normal mixture for different sample sizes. For comparison, the fixed $\theta_n = 1$ and $\theta_n = 0.5$ for the sinc and dIVP kernels are indicated by the orange and purple lines, respectively.

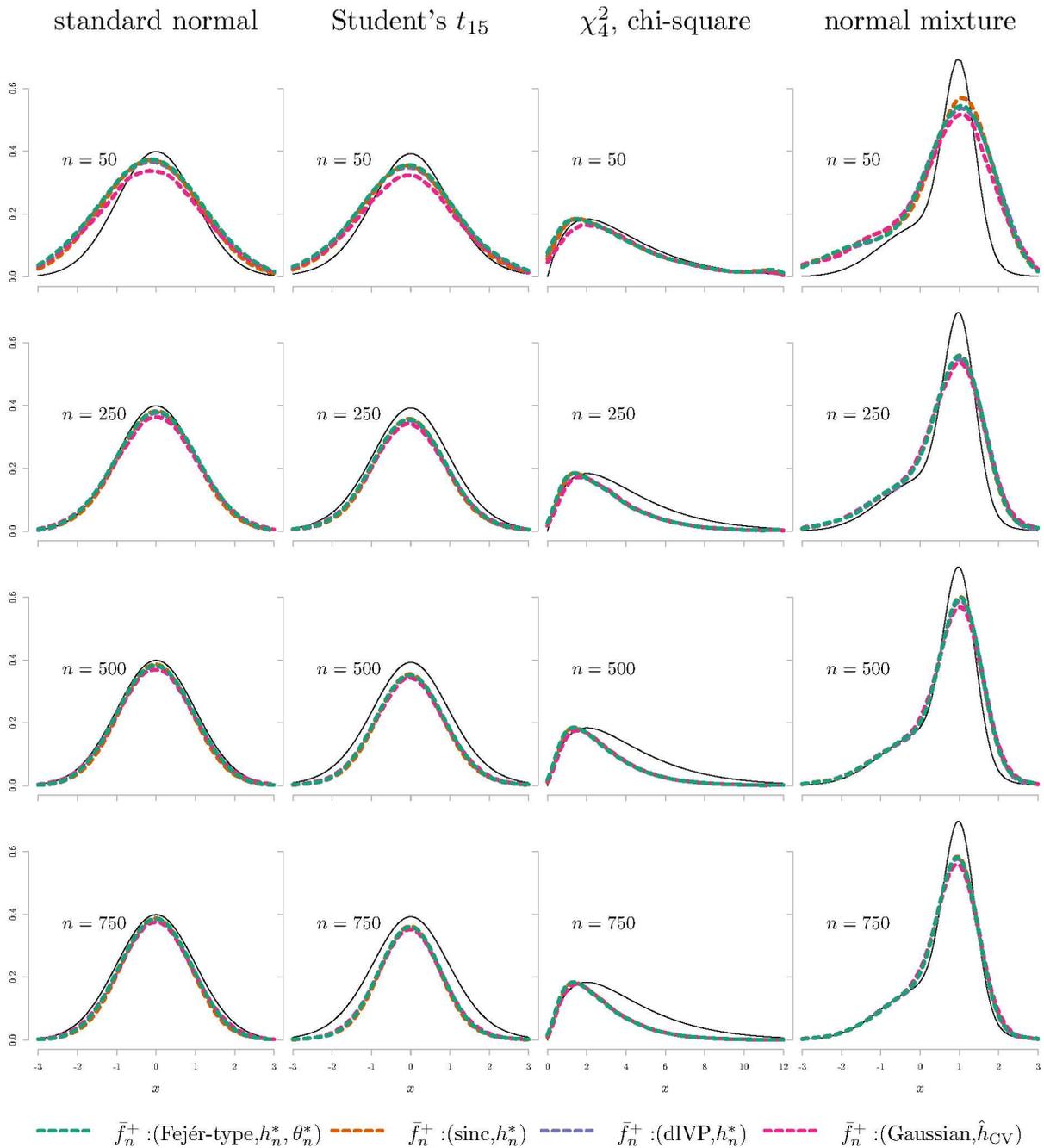


Figure 4.6: Kernel density estimation of a standard normal, Student's t_{15} , chi-square χ_4^2 , and normal mixture for sample sizes of 50, 250, 500, and 750. Density estimates with Fejér-type kernels and theoretical bandwidths, sinc kernels and theoretical bandwidths, dLVP kernels and theoretical bandwidths, and Gaussian kernels and cross-validation bandwidths, are indicated, respectively, by the green, orange, purple, and pink dashed lines. The number of replications is 200.

Chapter 5

Value at Risk using Density

Estimators with Fejér-type Kernels

Financial institutions have complex portfolios containing stocks, bonds, and other instruments that are susceptible to market risk. Market risk is a category of financial risk caused by changes in the prices of financial variables. Fluctuations in price may positively or negatively affect the investment portfolio of an institution bringing about huge gains or losses. It is important for institutions to monitor their portfolio risks so that they can mitigate any damages that may be incurred from adverse financial market movements. A standard way to assess risk is to use *Value at Risk* (VaR), which is a method that measures the worst possible losses on a portfolio that occur within a given probability.

In this chapter, we use the kernel-type estimator under study to estimate the probability densities and predict the daily VaR estimates of the daily portfolio returns. We compare our nonparametric method for estimating the VaR to other commonly used VaR techniques. All the VaR models are applied to two simple portfolios each consisting of a well-known stock market index, the Dow Jones Industrial Average (DJIA) and the S&P/TSX Composite Index. The adequacy of each VaR model is evaluated using a backtest procedure. We conclude this chapter with a discussion and analysis of the empirical results.

5.1 Evaluation of VaR for One Stock

The *Value at Risk* (VaR) is a commonly used method for risk assessment. It is a risk measure of the worst expected loss over a defined holding period at a given confidence level. It 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 remained one of the most common risk measurement techniques in finance. Financial institutions use VaR to measure their exposure to market risk by capturing the potential maximum loss in the value of a portfolio over a defined holding period. Shareholders and managers then compare the results of the VaR to their available capital and cash reserves to ensure that the losses can be covered without putting the firms at risk. If they are not comfortable with the level of risk, then they must determine ways to mitigate it. An in depth discussion on the origins of VaR and its many uses are provided in Jorion [16].

Suppose the portfolio consists of a single stock and we wish to compute the VaR over the time horizon h . If P_{t+h} and P_t are the portfolio values at time $t+h$ and t , respectively, then the theoretical return over the horizon h is given by

$$X_h = \log \left(\frac{P_{t+h}}{P_t} \right). \quad (5.1)$$

Let $f(x; h)$ be the density of returns. We denote the return $x_h(p)$ as the VaR of the portfolio, which is defined by

$$F(x_h(p); h) = P(X_h \leq x_h(p)) = \int_{-\infty}^{x_h(p)} f(x; h) dx = p,$$

where $p \in (0, 1)$ is chosen by the financial managers or investors. Equivalently,

$$x_h(p) := F^{-1}(p; h),$$

where $F^{-1}(\cdot; h)$ is the inverse of $F(\cdot; h)$. Hence, the VaR is the p -th quantile of the return distribution over the horizon h .

The time horizon h and the probability level p need to be specified by the financial managers. Both parameters depend on the purpose at hand. Typically, the time horizon is short, one day to a couple of weeks, and the VaR confidence level $100(1-p)\%$ is within the range of 95 to 99.5 percent. Investors or financial managers that are more risk averse tend to choose higher confidence levels. In our analysis, we estimate the VaR using 95, 97.5, 99, and 99.5 percent confidence levels and a time horizon of one day. Hence, we denote the VaR as $x(p) = x_1(p)$ for the remainder of this chapter.

Two sample sizes are considered in our study: a sample of size n , used for estimating the VaR measure, and a sample of size m , used for evaluating the adequacy of the VaR model. We shall refer to these samples as the *estimation sample* and the *evaluation sample*, respectively.

In the following sections, we discuss two commonly used VaR models and our non-parametric model with Fejér-type kernels. We show how to estimate the VaR using these models and how to evaluate the adequacy of the VaR models.

5.1.1 Models for Estimating VaR

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 distribution function $F(x)$ that is continuous from the right. Two basic approaches for estimating the VaR, with numerous variations within each approach, include parametric variance methods and the nonparametric historical simulation. The former make assumptions concerning the distribution of the portfolio returns and relate the VaR

to the variances of these returns. The latter methods are nonparametric and do not require any distributional assumptions on the portfolio returns. The vast majority of variance methods assume the returns are normally distributed while historical simulation is usually based on the empirical distribution of portfolio returns. The VaR can also be computed using kernel density estimation by evaluating the kernel quantile estimators. Below, we discuss how the VaR is estimated using the normal variance, simple historical simulation, and kernel quantile estimation methods.

For the following VaR models, it is assumed that the stock returns are i.i.d., which is implicit in the current risk measures used by practitioners. This assumption is consistent with the random walk version of the *efficient-market hypothesis* (EMH), which asserts that financial markets are informationally efficient. That is, current market prices include all relevant information about a particular asset. Prices, therefore, move only when new, unanticipated, information is received. In addition, the EMH states that investors are rational, knowing what information is important and what is not. In turn, the collective consciousness of the market finds an equilibrium price. Yesterday's news is no longer important and today's return is unrelated to yesterday's return, thus implying that the returns are independent. If the daily stock returns are independent, then they are random variables and follow a random walk. Furthermore, it can be reasonably assumed that the returns are identically distributed over time.

Normal Method

The normal VaR model is widely used among financial institutions since it involves some very convenient properties and a straightforward expression for VaR. Let X_1, \dots, X_n be i.i.d. normally distributed random variables, that is,

$$X_1, \dots, X_n \stackrel{iid}{\sim} N(\mu, \sigma^2),$$

where the mean μ and the variance σ^2 are unknown. To compute the VaR, one uses the quantile function of the normal distribution, in other words, the inverse of its cdf. The quantile function can be derived from a simple transformation of the probit function $\Phi^{-1}(p)$, which is the quantile function of the standard normal distribution. Thus, for a normally distributed return with mean μ and variance σ^2 , the VaR is given by

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

for any $0 < p < 1$. The parameters μ and σ are estimated using the sample mean and the unbiased sample standard deviation, that is,

$$\tilde{\mu}_n = \bar{X}_n = \frac{1}{n} \sum_{j=1}^n X_j \quad \text{and} \quad \tilde{\sigma}_n = s_n = \sqrt{\frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X}_n)^2}.$$

Hence, for some specified probability level $p_0 \in (0, 1)$, a natural VaR estimator is

$$\tilde{x}_n(p_0) = \tilde{\sigma}_n\Phi^{-1}(p_0) + \tilde{\mu}_n. \tag{5.2}$$

In reality, normally distributed returns are not a sound assumption as it is believed that high frequency financial return data have heavier tails than can be explained by the normal distribution. As a result, if the returns are not normal, then the normal VaR model will underestimate the true VaR. The current financial crisis demonstrated that the overly optimistic estimates of the normal VaR do not properly account for extreme events.

Historical Simulation

Consider a nonparametric approach that does not assume normality. Let X_1, \dots, X_n be i.i.d. random returns with a continuous distribution function F on \mathbb{R} , and let

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

denote the corresponding order statistics. For a given probability level $p_0 \in (0, 1)$ the VaR estimator is the p_0 -th sample quantile of the random sample of portfolio returns:

$$\tilde{x}_n(p_0) = X_{(\lfloor np_0 \rfloor + 1)}. \quad (5.3)$$

The main strengths of the historical simulation method are its simplicity and the absence of distributional assumptions as the VaR is determined by the actual price level movements. The method simply uses real historical data to build an empirical density by means of a histogram for the portfolio of returns. Yet, this method is not without its drawbacks. Typically, larger sample sizes are required to improve the accuracy of nonparametric estimation; however, 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 representation 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. One has to be careful on how the data is selected so as not to remove relevant or include irrelevant data. Another drawback is that the VaR is sensitive to the historical period chosen. The empirical distribution is a step function so its discrete approximation of the true distribution at the extreme tails can cause biased results. Smoothing the estimated quantile function using kernel density estimators can improve the precision of the VaR estimates.

Kernel Quantile Estimators

As before, let X_1, \dots, X_n be i.i.d. random portfolio returns with a continuous distribution function F on \mathbb{R} , and let

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

denote the corresponding order statistics. Having the empirical distribution function

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

a common definition for the *empirical quantile function* is

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

for $(j-1)/n < p \leq j/n$, $j = 1, \dots, n$. Parzen introduced the *kernel quantile estimator* (see [23], p. 113)

$$\begin{aligned} KQ_n(p) &= \int_0^1 K_h(t-p)Q_n(t) dt = \int_0^1 K_h(t-p)F_n^{-1}(t) dt \\ &=: \sum_{j=1}^n \left[\int_{(j-1)/n}^{j/n} K_h(t-p) dt \right] X_{(j)}, \end{aligned} \tag{5.4}$$

where, for a suitable kernel function K , $K_h(x) = h^{-1}K(h^{-1}x)$ and $0 < p < 1$. Naturally, $KQ_n(p)$ puts most weight on the order statistic $X_{(j)}$, for which j/n is close to p . Sheather and Marron [19] showed that the following approximation to (5.4) can be used in practice:

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

We choose the bandwidth h_F obtained by the methods of Fourier analysis as it was shown to provide good results from the simulation studies in Chapter 4. Therefore, for a specified probability level $p_0 \in (0, 1)$, the VaR estimator using a kernel density estimator with bandwidth h_F is

$$\tilde{x}_n(p_0) = \frac{\sum_{j=1}^n K_{h_F} \left(\frac{j-1/2}{n} - p_0 \right) X_{(j)}}{\sum_{j=1}^n K_{h_F} \left(\frac{j-1/2}{n} - p_0 \right)}. \quad (5.5)$$

The main purpose here is to observe how well the estimator in (5.5) with Fejér-type kernels estimates the VaR of the portfolio returns. To compare, we also apply the sinc, de la Vallée Poussin, and Gaussian kernels to (5.5).

5.1.2 Model Evaluation

We are interested in verifying the adequacy of a VaR model by means of a *backtest*. Backtesting is a statistical procedure that consists of systematically comparing 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 a good fit for measuring the VaR of a portfolio. Therefore, we perform a statistical test to determine whether the number of exceptions for each model is reasonable or not.

Let Y_1, \dots, Y_m be i.i.d. random variables indicating whether or not the realized return is worse than the forecasted VaR measure, that is,

$$Y_i = \begin{cases} 1 & \text{if } X \leq x(p), \\ 0 & \text{if } X > x(p), \end{cases}$$

where $P(X \leq x(p)) = p$. Then, $N = \sum_{i=1}^m Y_i$ is the number of VaR violations of a portfolio in an evaluation sample of size m and follows a binomial distribution $Bin(m, p)$ with probability mass function (pmf)

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

for $y = 0, 1, \dots, m$. We want to know what is the likelihood that the model is adequate given the observed number of VaR violations of a portfolio in an evaluation sample. The likelihood function for p given the observed values y_1, \dots, y_m of Y_1, \dots, Y_m is

$$L(p|y_1, \dots, y_m) = \prod_{i=1}^m P(Y_i = y_i; p) = \prod_{i=1}^m p^{y_i} (1 - p)^{1-y_i} = p^y (1 - p)^{m-y}, \quad (5.6)$$

where $y = \sum_{i=1}^m y_i$. Suppose the probability level for the VaR is chosen to be p_0 . The ratio N/m represents the model's *failure rate* of measuring VaR, which ideally should converge to p_0 . Thus, the relevant null and alternative hypotheses for determining the fit of the VaR model are:

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

We use (5.6) to carry out a likelihood ratio test to determine whether or not to reject the null hypothesis that the model is adequate. The appropriate likelihood ratio test statistic is given by

$$\begin{aligned} \Lambda(p_0) = \Lambda(p_0|y_1, \dots, y_m) &= \frac{L(p_0|y_1, \dots, y_m)}{\sup_{0 < p < 1} L(p|y_1, \dots, y_m)} \\ &= \frac{L(p_0|y_1, \dots, y_m)}{L(y/m|y_1, \dots, y_m)} =: \frac{p_0^y (1 - p_0)^{m-y}}{(y/m)^y (1 - y/m)^{m-y}}, \end{aligned} \quad (5.7)$$

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 $N = \sum_{i=1}^m Y_i$. We use the log-likelihood ratio statistic $-2 \log(\Lambda(p_0))$, where $\Lambda(p_0)$

is given in (5.7), as, under certain regularity conditions, the asymptotic distribution of $-2\log(\Lambda(p_0))$ is $\chi_1^2(1 - \alpha)$ as m approaches infinity (see [26], 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 . In accordance with convention, the α is set at 0.05. Therefore, we reject the null hypothesis if

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

A backtest of the form discussed in this section was first used by Kupiec in 1995 to evaluate the adequacy of the VaR Models (see [16], Chapter 6).

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 since the confidence level $100(1 - p)\%$ of the VaR is typically in the range of 95 to 99.5 percent. The $100(1 - \alpha)\%$ nonrejection regions for the number of VaR failures N based on the likelihood ratio test with $\alpha = 0.05$ are displayed in Table 5.1 over evaluation samples of 250, 500, 750, and 1000 trading days. Notice that the VaR model can be rejected when the number of failures is both high and low. This is because 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, which leads to inefficient allocation of capital across units, potentially harming profit opportunities.

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%	$3 \leq N \leq 11$	$7 \leq N \leq 19$	$12 \leq N \leq 27$	$16 \leq N \leq 35$
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%	$0 \leq N \leq 5$	$1 \leq N \leq 6$	$1 \leq N \leq 8$	$2 \leq N \leq 9$

Table 5.1: 95% nonrejection confidence regions for the likelihood ratio test under different VaR confidence levels and evaluation sample sizes

5.2 Results

We use publicly available historical stock market data to estimate the daily VaR forecasts from the distribution of daily portfolio returns. The objective is to assess the performance of the density estimator with Fejér-type kernels and bandwidth h_F when estimating the portfolio return densities and their VaR. This involves comparing the kernel-type estimator to other commonly used density and VaR estimation methods. Throughout the analysis, a one day time horizon is used and probability levels 0.05, 0.025, 0.01, and 0.005 are considered. The normal, historical simulation, and kernel quantile VaR models are employed using formulas (5.2), (5.3) and (5.5), respectively, where the latter model is evaluated for the Fejér-type, sinc, de la Vallée Poussin (dlVP), and Gaussian kernels. All the VaR models are applied to 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. For each VaR model, estimation samples of 252, 504, and 1000 trading days are used and a backtest is performed using an evaluation sample of 1000 trading days.

A stock market index is a measurement of the value of a section of the stock market and is used by financial managers and investors as a tool to describe a particular market or sector. The DJIA is a stock index that represents the standard trading sessions of 30 large publicly owned companies in the stock market that are based in the United States. It is price-weighted, meaning that only the price of each component stock is considered when determining the value of the index. Hence, any price shift from even the smallest security

heavily influences the value of the index. The S&P/TSX Composite Index is an index of the stock prices of the largest companies on the Toronto Stock Exchange (TSX). Unlike the DJIA, the S&P/TSX is market capitalization-weighted, that is, the size of the company is factored in. Hence, a price shift of a large company heavily influences the value of the index, even if the shift is relatively small. The DJIA and S&P/TSX Composite indices are compiled to track the performance of certain sections within the American and Canadian economy, respectively.

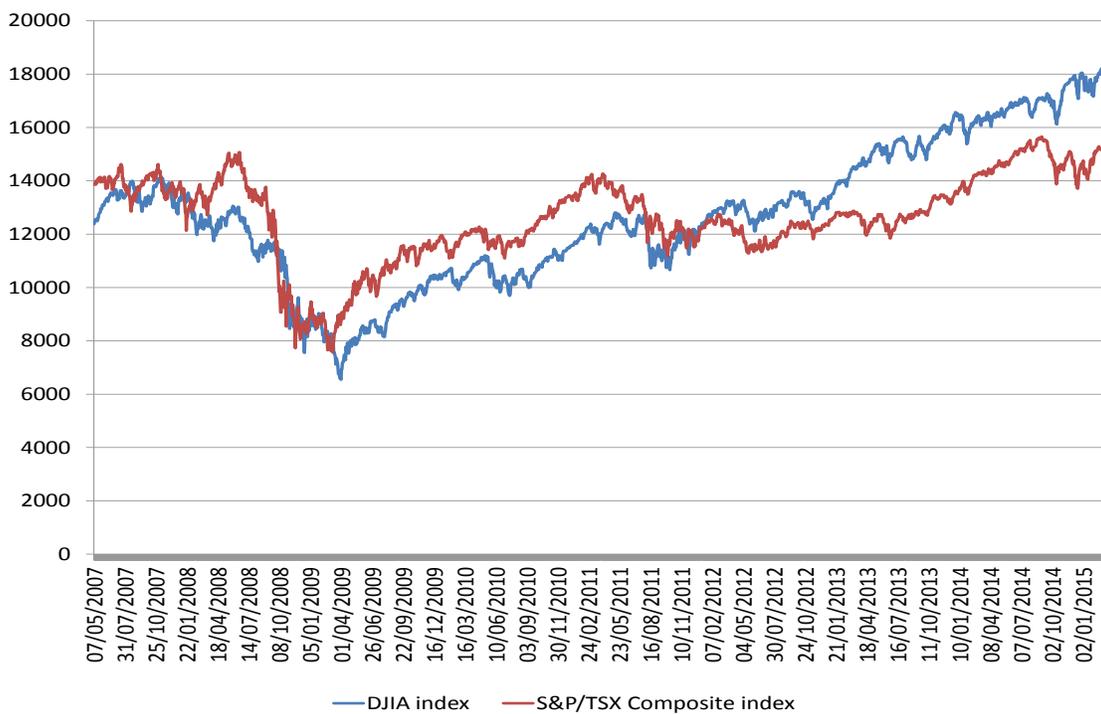


Figure 5.1: DJIA and S&P/TSX Composite indices from June 28, 2007 through March 11, 2015

In reality, an index cannot be invested directly as it is a mathematical construct, so the DJIA and S&P/TSX Composite indices are each used as a representative stock. We chose these indices to be in our fictitious investments due to the abundance of publicly available historical data. The values of the DJIA and S&P/TSX Composite Index used in this study are displayed in Figure 5.1 from June 28, 2007 to March 11, 2015. The effect of the recent financial crisis in 2008 is indicated by both indices. A large decrease in points can be seen

with the DJIA having a low of an approximately 6500 level in the early months of 2009. The level of both indices has increased in the recent years, particularly for the DJIA.

The daily logarithmic returns for each stock index are evaluated as in (5.1) and displayed in the graphs of Figure 5.2 for the DJIA and Figure 5.3 for the S&P/TSX Composite index, over a time period of one thousand trading days. One thousand daily VaR estimates from the normal, historical simulation, and Fejér-type kernel quantile models are also plotted in the figures. The figures are presented as an array of plots. Each row of plots corresponds to the VaR confidence level, while each column provides the estimation sample used. The plots in the first, second, third, and fourth rows display, respectively, the daily VaR figures at 99.5%, 99%, 97.5%, and 95%. On the other hand, the plots in the first, second, and third columns display, respectively, the VaR estimates under an estimation sample of 252, 504, and 1000 observations. In the business calendar, one year is equivalent to 252 trading days, which means that we predicted VaR estimates based on daily data from the past year, two years, and approximately four years.

To assess the adequacy of each VaR model, a backtest is performed over an evaluation period of one thousand trading days. Table 5.2 displays the backtest results of all the VaR models under consideration for the DJIA and S&P/TSX Composite Index portfolios. The table reports the observed number of exceptions of each VaR model based on estimation samples of 252, 504, and 1000 observations for a range of probability levels. The table also reports the outcome of the backtest, that is, whether or not to reject the model given the observed number of VaR exceptions. These outcomes are determined by the 95% nonrejection regions indicated in Table 5.1 when the evaluation sample size is 1000 days.

Figures 5.4 and 5.5 display the 97.5% VaR estimates of portfolio return probability densities under estimation samples of 252, 504, and 1000 observations, for three specified days, as well as the actual returns of these days. The figures are arranged in a matrix of plots. Each row of plots corresponds to a specific day from the evaluation period, while each column provides the estimation sample used. In Figure 5.4, the first, second, and

Probability Level	Observed Number of Exceptions/Number of Observations						Test Outcome					
	normal	historical	Fejér-type	sinc	dIVP	Gaussian	normal	historical	Fejér-type	sinc	dIVP	Gaussian
DJIA Portfolio												
0.005	20/252	8/252	5/252	3/252	5/252	6/252	Reject	–	–	–	–	–
0.01	26/252	10/252	10/252	10/252	10/252	10/252	Reject	–	–	–	–	–
0.025	40/252	30/252	29/252	29/252	30/252	30/252	Reject	–	–	–	–	–
0.05	57/252	57/252	57/252	58/252	57/252	56/252	–	–	–	–	–	–
0.005	14/504	3/504	3/504	3/504	3/504	3/504	Reject	–	–	–	–	–
0.01	23/504	11/504	10/504	11/504	10/504	10/504	Reject	–	–	–	–	–
0.025	35/504	24/504	25/504	24/504	25/504	25/504	–	–	–	–	–	–
0.05	47/504	49/504	47/504	47/504	47/504	47/504	–	–	–	–	–	–
0.005	3/1000	0/1000	0/1000	0/1000	0/1000	0/1000	–	Reject	Reject	Reject	Reject	Reject
0.01	4/1000	1/1000	1/1000	1/1000	1/1000	1/1000	Reject	Reject	Reject	Reject	Reject	Reject
0.025	14/1000	6/1000	6/1000	6/1000	6/1000	6/1000	Reject	Reject	Reject	Reject	Reject	Reject
0.05	24/1000	23/1000	23/1000	25/1000	23/1000	23/1000	Reject	Reject	Reject	Reject	Reject	Reject
S&P/TSX Composite Index Portfolio												
0.005	22/252	11/252	9/252	9/252	9/252	9/252	Reject	Reject	–	–	–	–
0.01	29/252	15/252	15/252	15/252	15/252	15/252	Reject	–	–	–	–	–
0.025	51/252	38/252	34/252	33/252	33/252	34/252	Reject	Reject	–	–	–	–
0.05	63/252	59/252	59/252	58/252	59/252	59/252	–	–	–	–	–	–
0.005	16/504	7/504	7/504	6/504	7/504	7/504	Reject	–	–	–	–	–
0.01	22/504	12/504	12/504	12/504	12/504	11/504	Reject	–	–	–	–	–
0.025	40/504	30/504	28/504	29/504	28/504	29/504	Reject	–	–	–	–	–
0.05	59/504	51/504	51/504	51/504	51/504	51/504	–	–	–	–	–	–
0.005	5/1000	0/1000	0/1000	0/1000	0/1000	0/1000	–	Reject	Reject	Reject	Reject	Reject
0.01	6/1000	3/1000	3/1000	3/1000	3/1000	3/1000	–	Reject	Reject	Reject	Reject	Reject
0.025	12/1000	6/1000	6/1000	6/1000	6/1000	6/1000	Reject	Reject	Reject	Reject	Reject	Reject
0.05	23/1000	21/1000	20/1000	20/1000	20/1000	20/1000	Reject	Reject	Reject	Reject	Reject	Reject

Table 5.2: *Backtest results of all the VaR models under consideration applied to each stock index over an evaluation sample of 1000 days and 95% confidence regions*

third rows display, respectively, the results of the DJIA index returns on December 3, 2011, December 23, 2012, and January 30, 2015. In Figure 5.4, the first, second, and third rows display, respectively, the results of the S&P/TSX Composite index returns on June 29, 2011, June 25, 2013, and September 30, 2015. These days were chosen as they include a few examples of some VaR violations that occurred in the evaluation period. On the other hand, the plots in the first, second, and third columns display, respectively, the results of the stock index returns under an estimation sample of 252, 504, and 1000 observations, for both figures. The normal density is represented by the red curve and the kernel-type estimator of interest is represented by the green curve. The 97.5% predicted VaR estimates of the normal, historical simulation, and Fejér-type kernel quantile models are indicated, respectively, by the vertical red, black, and green dashed lines, while the value of the actual daily return is the solid blue line. The purpose of the figures is to compare how well the normal density, histogram, and density estimator with the Fejér-type kernel and the

bandwidth h_F estimate the density of index returns.

From the figures and table presented in this study we observe the following. Overall, the empirical results applied to both stock indices are fairly similar. The backtest results in Table 5.2 show that the normal model has the poorest performance as it is not considered an adequate model in most cases, while the models based on kernel quantile estimation are the most reliable. Historical simulation is also quite reliable, but when applied to the S&P/TSX Composite Index we see that it is not adequate for probability levels 0.005 and 0.025 when the number of observations is 252. For the normal model, 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 cases in which the normal model is not consistently rejected is at a probability level of 0.05 for estimation samples of 252 and 504 observations. Furthermore, none of the models perform well when the estimation sample is large, except the normal for very small probability levels. This is expected as financial data from four years ago might no longer be relevant to the current market situation.

The backtest results are consistent with what we see in Figures 5.2 and 5.3. The normal model tends to underestimate the risk, specifically for higher confidence levels, while the nonparametric models overestimate the risk when their estimation sample is too large. The normal VaR estimates do not differ much across the probability levels, another indication that the normal underestimates the tails of the distribution. The two nonparametric models, historical simulation and Fejér-type kernel quantile estimator, generally have similar outcomes. The performance of all the VaR models is similar at the 95% confidence level.

The images in Figures 5.4 and 5.5 are also consistent with the assertion that stock returns are heavy tailed. The density estimates with Fejér-type kernels and bandwidth h_F can account for heavy tails of the return distributions better than the normal. The VaR violations of the normal model, that is when the actual daily return exceeds the VaR estimate, can be seen for estimation samples of 252 and 504 observations. One VaR exception can also be seen from the historical simulation under 252 observations,

besides that, the VaR figures of the historical simulation and the Fejér-type kernel quantile estimates are similar.

Therefore, VaR models that use density estimators with Fejér-type kernel functions are more reliable than the commonly used normal and historical simulation methods as they can account for heavy tails of the return distributions. One just has to be careful in choosing a relevant estimation period.

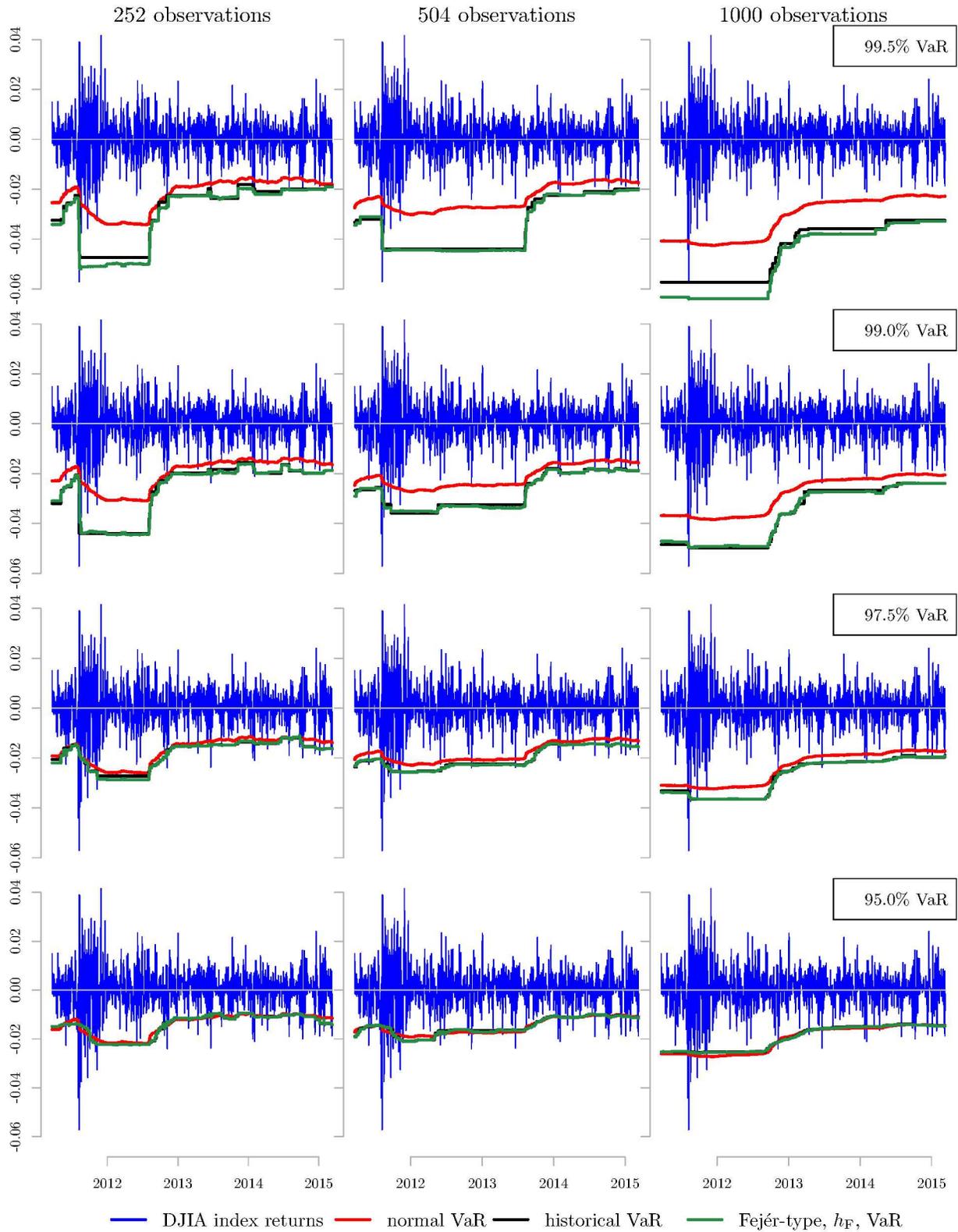


Figure 5.2: Daily log returns and VaR estimates of the DJIA at 95%, 97.5%, 99%, and 99.5% confidence levels under 252, 504, and 1000 observations over an evaluation sample of 1000 days for various VaR techniques.

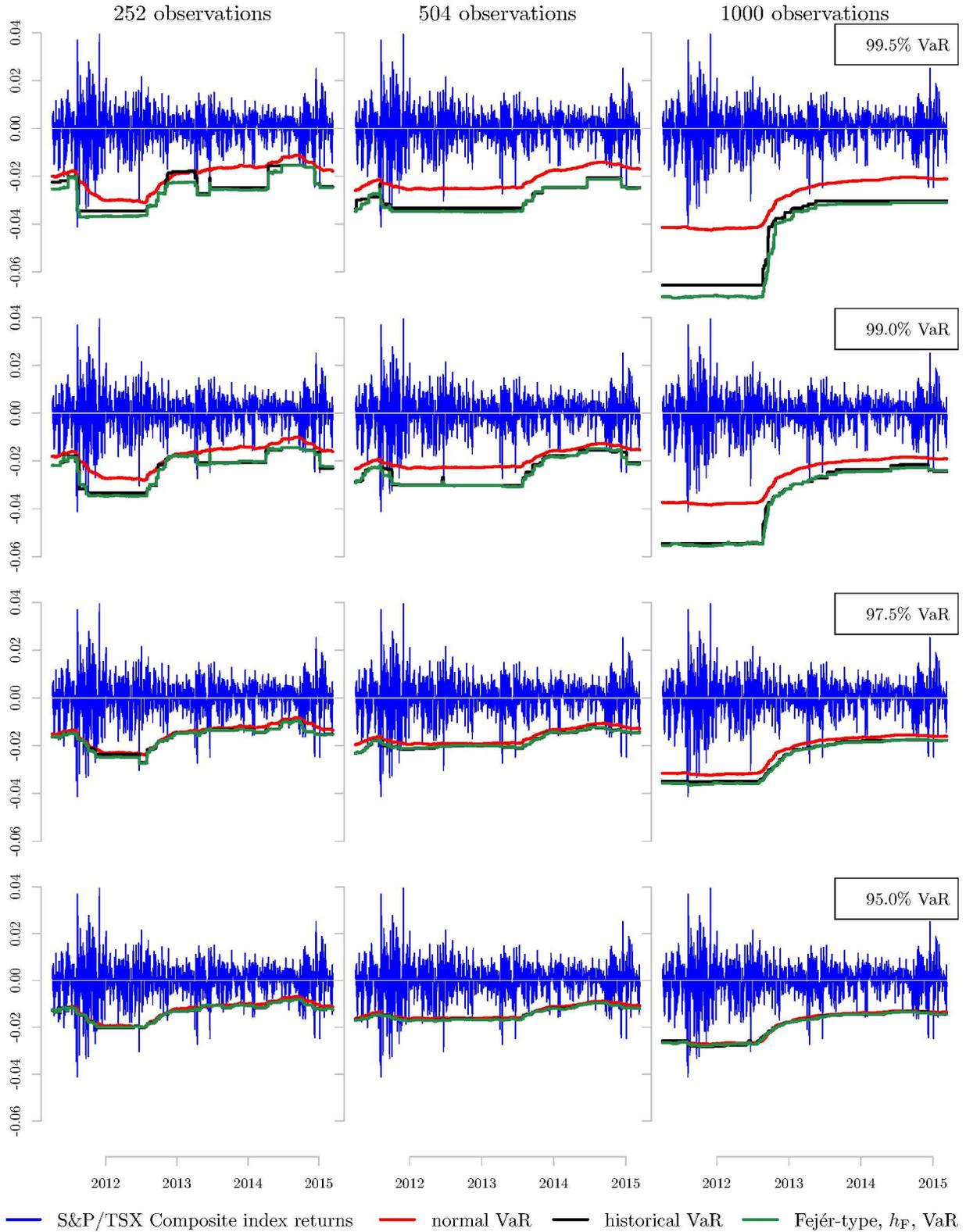


Figure 5.3: Daily log returns and VaR estimates of the S&P/TSX Composite Index at 95%, 97.5%, 99%, and 99.5% confidence levels under 252, 504, and 1000 observations over an evaluation sample of 1000 days for various VaR techniques.

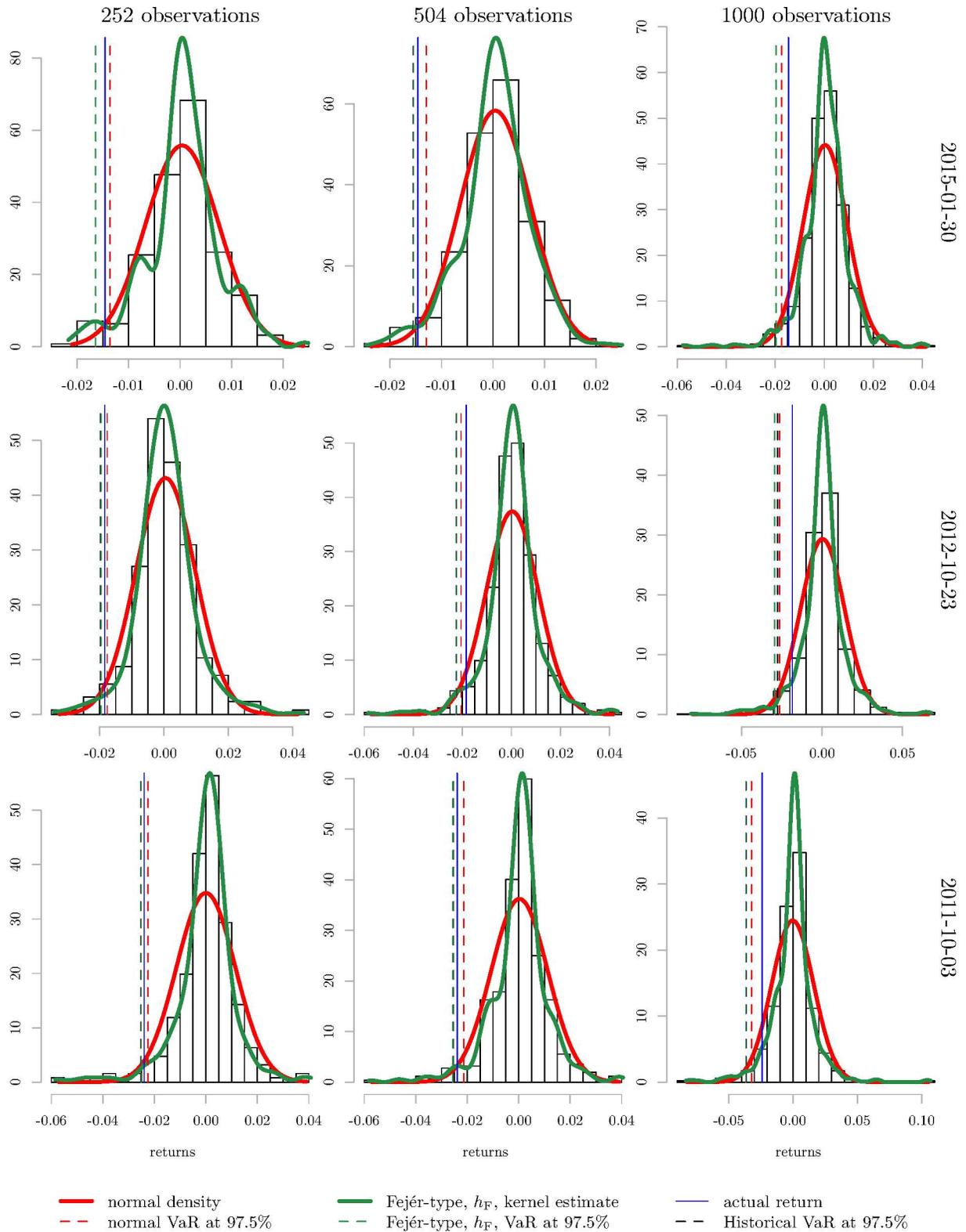


Figure 5.4: Normal, empirical, and Fejér-type kernel densities of the DJIA daily returns based on 252, 504, and 1000 observations for the days 29/06/2011, 25/06/2013, and 30/09/2014. The 97.5% VaR of each model is illustrated along with the actual daily return.

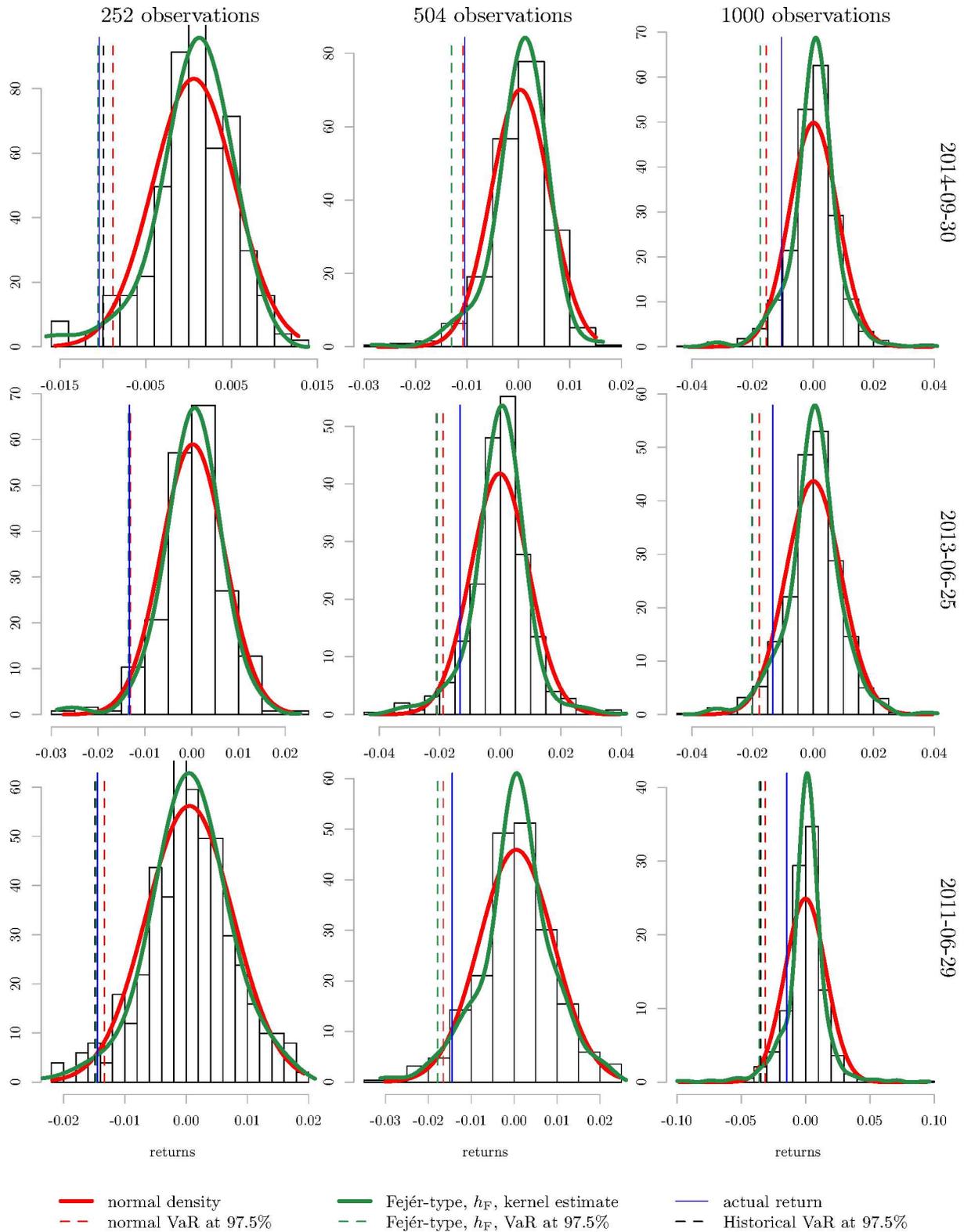


Figure 5.5: Normal, empirical, and Fejér-type kernel densities of the S&P/TSX Composite daily returns based on 252, 504, and 1000 observations for the days 29/06/2011, 25/06/2013, and 30/09/2014. The 97.5% VaR of each model is illustrated along with the actual daily return.

Conclusions

In this thesis, we conducted a simulation study to support the theoretical finding in [32] that the class of kernel-type estimators with the Fejér-type kernel and the theoretical bandwidth h_n is optimal for estimating smooth densities in \mathbb{L}_2 . We also studied the performance of the density estimator with Fejér-type kernels using data-dependent bandwidth selection methods; namely, the cross-validation and the approach based on the Fourier analysis of kernel density estimators. The kernel-type estimator was then applied to a practical problem using real financial data.

In Chapter 1, a brief overview of the literature on density estimation is provided outlining the few misconceptions behind commonly used kernel functions and the criteria in which to measure the asymptotic efficiency of an estimator. It is believed that the choice of kernel K is not considered as crucial as the choice of the bandwidth h for the quality of a density estimator, yet, this is not the case for the so-called LAM estimators. When constructing an LAM estimator of a density f one has to carefully choose both h and K . Also, many kernel functions recommended in the literature may not always be the best choice. One example is the Epanechnikov kernel. Although it is considered optimal in some settings, we show in Chapter 2 that it is, in fact, inadmissible. We stress that the estimation of infinitely smooth densities can be further improved by using the Fejér-type kernels.

The theory behind a kernel density estimator and its properties are given in Chapter 2, along with an analysis of its \mathbb{L}_2 -risk. To assess the behaviour of the Fejér-type kernels with

empirically-selected bandwidths, the minimisation of an unbiased estimator of the MISE in h is required. We analysed the MISE in both the classical sense and using the Fourier analysis techniques; the latter approach proved to provide nicer results in theory and in practice as we later saw in our simulations. Unbiased risk estimators for both MISE approaches were derived. In Chapter 3, we evaluated the asymptotically unbiased risk estimators for cross-validation and Fourier analysis for all of the kernels under consideration.

In Chapter 4, we presented the methodology and results of our simulation studies, which ascertain the good theoretical properties of the kernel-type estimator in hand. We found numerically that, for a good choice of constant γ , the kernel-type estimator of densities $f \in \mathcal{F}$ as in (3.2) performed favourably as compared to most of the estimators under consideration, particularly when estimating the normal mixture. For unimodal densities, the efficient estimator with de la Vallée Poussin kernels and theoretical bandwidths performed slightly better for $n < 1000$ than the kernel-type estimator of interest. The Fejér-type kernel estimators also performed very well with the empirical bandwidth selectors compared to other density estimators with such bandwidths. We found that the bandwidths based on the Fourier analysis of kernel density estimators provided more favourable results than the ones from cross-validation, especially for small sample sizes and unimodal densities. Therefore, in practice we recommend the use of the bandwidth h_F based on the Fourier analysis techniques.

In Chapter 5, we applied the density estimator with the Fejér-type kernel and the bandwidth h_F to the common risk management concept of value at risk using historical stock market data. The probability densities of the daily portfolio returns and their quantiles, i.e., VaR, were estimated using the kernel-type estimator and the commonly applied normal density and histogram approaches. As a result, our kernel-type estimator proved to be more reliable as it can better account for the heavy tails of the return distributions.

For future work, one can try to compare the kernel-type estimator to other kernel density estimators using different \mathbb{L}_p -risks, $1 \leq p < \infty$. Another possible direction of

future studies would be to produce confidence bands for f where we compute the mean and the variance of the kernel density estimators. It would also be interesting to explore the application of the kernel-type estimator of interest to multivariate densities.

APPENDICES

Appendix A

Definitions and Theorems

Definition 1 (convolution). For any two functions $f, g \in \mathbb{L}_1(\mathbb{R})$ and any $x \in \mathbb{R}$, the *convolution* of f and g is given by

$$(f * g)(x) := \int_{\mathbb{R}} f(t)g(x - t) dt.$$

It is well known that $f * g \in \mathbb{L}_1(\mathbb{R})$.

Definition 2 (Cauchy sequence). Let μ be an arbitrary positive measure. If $\{f_n\}$ is a sequence in $\mathbb{L}_p(\mu)$, if $f \in \mathbb{L}_p(\mu)$, and if $\lim_{n \rightarrow \infty} \|f_n - f\|_p = 0$, we say that $\{f_n\}$ converges to f in $\mathbb{L}_p(\mu)$. If for every $\epsilon > 0$ there exists an integer N such that $\|f_n - f_m\|_p < \epsilon$ as soon as $n > N$ and $m > N$, we call $\{f_n\}$ a *Cauchy sequence* in $\mathbb{L}_p(\mu)$.

Definition 3 (metric space). A *metric space* is a non-empty set X in which a *distance function* (or *metric*) ρ is defined, with the following properties:

- (a) $0 \leq \rho(x, y) < \infty$ for all x and $y \in X$.
- (b) $\rho(x, y) = 0$ if and only if $x = y$.
- (c) $\rho(x, y) = \rho(y, x)$ for all x and $y \in X$.

(d) $\rho(x, y) \leq \rho(x, z) + \rho(z, y)$ for all x, y , and $z \in X$.

Property (d) is called the *triangle inequality*.

Definition 4 (inner product space). A complex vector space H is called an *inner product space* if to each ordered pair of vectors x and $y \in H$ there is a complex number $\langle x, y \rangle$, the so-called “inner product” (or “scalar product”) of x and y , such that the following rules hold:

(a) $\langle y, x \rangle = \overline{\langle x, y \rangle}$.

(b) $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$ if x, y , and $z \in H$.

(c) $\langle \alpha x, y \rangle = \alpha \langle x, y \rangle$ if x and $y \in H$ and α is a scalar.

(d) $\langle x, x \rangle \geq 0$ for all $x \in H$.

(e) $\langle x, x \rangle = 0$ only if $x = 0$.

By (d), we may define $\|x\|$, the norm of the vector $x \in H$, to be the nonnegative square root of $\langle x, x \rangle$. Thus

$$\|x\|^2 := x^T x = \langle x, x \rangle.$$

Definition 5 (Hilbert space). Let H be an inner product space. Then, by the triangle inequality

$$\|x - z\| \leq \|x - y\| + \|y - z\|, \quad x, y, z \in H.$$

If we define the distance between x and y to be $\|x - y\|$, all the axioms for a metric space are satisfied. Thus, H is now a metric space. If this metric space is *complete*, i.e., if every Cauchy sequence converges in H , then H is called a *Hilbert space*.

Definition 6 (Hilbert space isomorphism). Two Hilbert spaces H_1 and H_2 are isomorphic if there is a one-to-one linear mapping Δ of H_1 onto H_2 which also preserves inner products: $\langle \Delta x, \Delta y \rangle$ for all x and $y \in H_1$. Such a Δ is a *Hilbert space isomorphism* of H_1 onto H_2 .

Theorem 1 (convolution theorem (see [28], Theorem 9.2)). *For any two functions $f, g \in \mathbb{L}_1(\mathbb{R})$ with respective Fourier transforms \hat{f} and \hat{g} and any $x \in \mathbb{R}$*

$$\widehat{(f * g)}(x) = \hat{f}(x) \cdot \hat{g}(x).$$

Theorem 2 (Fourier inversion theorem (see [28], p. 182)). *If a function $f : \mathbb{R} \rightarrow \mathbb{C}$ and its Fourier transform \hat{f} given by*

$$\hat{f}(t) = \int_{\mathbb{R}} f(x) e^{itx} dx,$$

are both in $\mathbb{L}_1(\mathbb{R})$, then for almost all $x \in \mathbb{R}$

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \hat{f}(t) e^{-itx} dt.$$

Theorem 3 (Fubini theorem (see [28], Theorem 7.8)). *Finite double integrals can be computed using iterated integrals:*

$$\int_X \left(\int_Y f(x, y) dy \right) dx = \int_Y \left(\int_X f(x, y) dx \right) dy = \int_{X \times Y} f(x, y) d(x, y).$$

Theorem 4 (Glivenko-Cantelli theorem (see [35], Theorem 19.1)). *Suppose that X_1, X_2, \dots is a sequence of independent and identically distributed random variables with a common distribution function F on \mathbb{R} . Put $F_n(x) = n^{-1} \sum_{i=1}^n \mathbb{I}(X_i \leq x)$. Then as $n \rightarrow \infty$*

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

Theorem 5 (Kolmogorov's strong law of large numbers (see [31], Theorem 2.3.10)). *Let*

X_1, X_2, \dots be a sequence of independent random variables such that $\mathbb{E}[X_i] = \mu_i$ and $\text{Var}(X_i) = \sigma_i^2$ exist for every $i \geq 1$. Also, let $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$ and $\mathbb{E}[\bar{X}_n] = n^{-1} \sum_{i=1}^n \mu_i$. If $\sum_{i \geq 1} i^{-2} \sigma_i^2 < \infty$, then

$$\bar{X}_n - \mathbb{E}[\bar{X}_n] \xrightarrow[n \rightarrow \infty]{a.s.} 0.$$

Theorem 6 (Plancherel theorem (see [28], p. 186)). *To each $f \in \mathbb{L}_2(\mathbb{R})$, one can associate a function $\hat{f} \in \mathbb{L}_2(\mathbb{R})$ so that the following properties hold:*

- I. *If $f \in \mathbb{L}_1(\mathbb{R}) \cap \mathbb{L}_2(\mathbb{R})$, then \hat{f} is its Fourier transform, as above,*
- II. $\|\hat{f}\|_2 \leq \|f\|_2$,
- III. *The mapping $f \mapsto \hat{f}$ is a Hilbert space isomorphism of $\mathbb{L}_2(\mathbb{R})$ onto $\mathbb{L}_2(\mathbb{R})$.*
- IV. *The following symmetric relation exists between f and \hat{f} : if for some $A > 0$ we define*

$$\varphi_A(t) = \int_{-A}^A f(x) e^{ixt} dx, \quad t \in \mathbb{R},$$

and

$$\psi_A(x) = \frac{1}{2\pi} \int_{-A}^A \hat{f}(t) e^{-ixt} dt, \quad x \in \mathbb{R},$$

then $\|\varphi_A - \hat{f}\|_2 \rightarrow 0$ and $\|\psi_A - f\|_2 \rightarrow 0$ as $A \rightarrow \infty$,

- V. *If $\hat{f} \in \mathbb{L}_1(\mathbb{R}) \cap \mathbb{L}_2(\mathbb{R})$, and we define*

$$g(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \hat{f}(t) e^{-ixt} dt, \quad x \in \mathbb{R},$$

then $g = f$ almost everywhere.

Theorem 7 (Parseval's theorem (see [28], p. 187)). *For all functions f and g in $\mathbb{L}_2(\mathbb{R})$,*

$$\int_{\mathbb{R}} f(x)\overline{g(x)} dx = \frac{1}{2\pi} \int_{\mathbb{R}} \hat{f}(t)\overline{\hat{g}(t)} dt.$$

In particular,

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

Appendix B

R code

```
1 #####
2 #Olga Kosta
3 #2015, Carleton University
4 #
5 #   Script contains functions used for constructing and measuring the
6 #   efficiency of kernel density estimators for univariate pdfs.
7 #####
8 #-----
9 # Random samples:
10 #-----
11 #
12 # Generate M random samples of a density f with size n
13 M.rs<-function(M, f, n) {
14   rs.mtx<-switch(f, replicate(M, rnorm(n)), replicate(M, rt(n, 15)), replicate(M,
15     rchisq(n, 4)), replicate(M, c(rnorm(0.4*n), rnorm(0.6*n, 1, 0.4))))
16   rs.mtx
17 }
18 #-----
19 # Kernels:
20 #-----
21 #
22 # Fejer-type
23 ker.Fejer.type<-function(n, x, gamma) {
24   # compute estimate and convolution
25   N=log(n) / (2*gamma)
26   theta.n=1-1/N
27   x.1=x[which(x!=0)]
28   x.2=x[which(x==0)]
29   est.1<-(cos(theta.n*x.1)-cos(x.1)) / (pi*(1-theta.n)*x.1^2)
30   est.2<-rep((1+theta.n) / (2*pi), length(x.2))
31   conv.1<-2*cos(theta.n*x.1) / (pi*(1-theta.n)*x.1^2)+2*(sin(theta.n*x.1)-sin(
    x.1)) / (pi*(1-theta.n)^2*x.1^3)
```

```

32 conv.2<-rep((1+2*theta.n)/(3*pi),length(x.2))
33 estimate<-c(est.1,est.2)
34 convolution<-c(conv.1,conv.2)
35
36 # output
37 result<-cbind(estimate,convolution)
38 result
39 }
40 # Sinc
41 ker.sinc<-function(n,x,g){
42 # compute estimate and convolution
43 x.1=x[which(x!=0)]
44 x.2=x[which(x==0)]
45 est.1<-(1/(pi*x.1))*sin(x.1)
46 est.2<-rep(1/pi,length(x.2))
47 estimate<-c(est.1,est.2)
48 convolution<-estimate
49
50 # output
51 result<-cbind(estimate,convolution)
52 result
53 }
54 # De La Vallee Poussin
55 ker.dlVP<-function(n,x,g){
56 # compute estimate and convolution
57 x.1=x[which(x!=0)]
58 x.2=x[which(x==0)]
59 est.1<-2*(cos(x.1/2)-cos(x.1))/(pi*(x.1^2))
60 est.2<-rep((3)/(4*pi),length(x.2))
61 conv.1<-(4*cos(x.1/2))/(pi*x.1^2)+8*(sin(x.1/2)-sin(x.1))/(pi*x.1^3)
62 conv.2<-rep(2/(3*pi),length(x.2))
63 estimate<-c(est.1,est.2)
64 convolution<-c(conv.1,conv.2)
65
66 # output
67 result<-cbind(estimate,convolution)
68 result
69 }
70 # Gaussian
71 ker.Gauss<-function(n,x,g){
72 # compute estimate, convolution, and Fourier transform
73 estimate<-(1/sqrt(2*pi))*exp((-1/2)*x^2)
74 convolution<-(1/sqrt(4*pi))*exp((-1/4)*x^2)
75 Fourier.transform<-exp((-1/2)*x^2)
76
77 # output
78 result<-cbind(estimate,convolution,Fourier.transform)
79 result
80 }
81 #-----
82 # Bandwidths:
83 #-----
84 #
85 # cross-validation

```

```

86 h.cv<-function(data, kernel, gamma, hmin, hmax, incr) {
87   n<-length(data)
88   k<-switch(kernel, fun<-ker.Fejer.type, fun<-ker.sinc, fun<-ker.dlVP, fun<-ker.
      Gauss)
89   h<-seq(hmin, hmax, by=incr)
90   m<-length(h)
91
92   f<-function(i) {data[(i+1):n]-data[i]}
93   list.d.ij<-sapply(1:(n-1), f)
94   d.ij=unlist(list.d.ij)
95   hf<-function(l) {d.ij/h[l]}
96   list.d.ijl<-sapply(1:m, hf)
97   d.ijl=unlist(list.d.ijl)
98
99   cf<-function(l) {
100     s=sum(k(n, d.ijl[, l], gamma)[, 2]-k(n, d.ijl[, l], gamma)[, 1]*2*n/(n-1))
101     a=k(n, 0, gamma)[2]*n/2+s
102     return=2/((n^2)*h[l])*a
103     return
104   }
105   cv=sapply(1:m, cf)
106   index.min<-which(cv==min(cv))
107
108   # output
109   h.return<-h[index.min]
110   h.return
111 }
112 # M sized vector of h.cv, i=1, ..., M
113 h.cv.i<-function(data, kernel, gamma, hmin, hmax, incr) {
114   M=dim(data)[2]
115   h.cv.i=sapply(1:M, function(i) {h.cv(data[, i], kernel, gamma, hmin, hmax, incr)})
116   h.cv.i
117 }
118 #
119 # Fourier-based
120 h.Fourier<-function(data, kernel, gamma, hmin, hmax, incr) {
121
122   # remove duplicates in data
123   data.unique=unique(data)
124   n<-length(data.unique)
125   k<-switch(kernel, fun<-ker.Fejer.type, fun<-ker.sinc, fun<-ker.dlVP, fun<-ker.
      Gauss)
126   h<-seq(hmin, hmax, by=incr)
127   m<-length(h)
128
129   f<-function(j) {data.unique[j]-data.unique[(j+1):n]}
130   list.d.jk<-sapply(1:(n-1), f)
131   d.jk=unlist(list.d.jk)
132
133   # Unbiased estimators for the different kernels
134   Jh.ft<-function(i) {
135     N=log(n)/(2*gamma)
136     theta.n=1-1/N

```

```

137     s1=2*(h[i]/(1-theta.n))^2*((n-1)/n)*sum((sin((theta.n*d.jk)/h[i])-sin(d.
      jk/h[i]))/d.jk^3)
138     s2=((2*h[i]/(1-theta.n))*sum((cos(d.jk/h[i])-cos((theta.n*d.jk)/h[i])/n
      )/d.jk^2)
139     Jh=(2/n^2)*(s1+s2)-(2*(n+1)*theta.n)/(n^2*h[i])+(2*(n-1)*(1-theta.n))/(3
      *n^2*h[i])+(4*theta.n)/(n*h[i])
140     Jh
141   }
142   Jh.sinc<-function(i){
143     s=sum(sin(d.jk/h[i])/d.jk*(1+1/n))
144     Jh=2/(n*h[i])-2/(n^2*h[i])-(2/n^2)*s
145     Jh
146   }
147   Jh.dlVP<-function(i){
148     theta.n=1/2
149     s1=2*(h[i]/(1-theta.n))^2*((n-1)/n)*sum((sin((theta.n*d.jk)/h[i])-sin(d.
      jk/h[i]))/d.jk^3)
150     s2=((2*h[i]/(1-theta.n))*sum((cos(d.jk/h[i])-cos((theta.n*d.jk)/h[i])/n
      )/d.jk^2)
151     Jh=(2/n^2)*(s1+s2)-(2*(n+1)*theta.n)/(n^2*h[i])+(2*(n-1)*(1-theta.n))/(3
      *n^2*h[i])+(4*theta.n)/(n*h[i])
152     Jh
153   }
154   Jh.Gauss<-function(i){
155     s1=sum(exp(-(d.jk^2/(4*h[i]^2))))
156     s2=sum(exp(-(d.jk^2/(2*h[i]^2))))
157     Jh=sqrt(pi)/(n^2*h[i])*((n-1)*(1+(1/n)*s1)-2^(3/2)*s2)
158     Jh
159   }
160   Jh<-switch(kernel, fun<-Jh.ft, fun<-Jh.sinc, fun<-Jh.dlVP, fun<-Jh.Gauss)
161   ft=sapply(1:m, Jh)
162   index.min<-which(ft==min(ft))
163
164   # output
165   h.return<-h[index.min]
166   h.return
167 }
168 # M sized vector of h.Fourier, i=1,...,M
169 h.Fourier.i<-function(data, kernel, gamma, hmin, hmax, incr) {
170   M=dim(data)[2]
171   h.Fourier.i=sapply(1:M, function(i){h.Fourier(data[,i], kernel, gamma, hmin,
      hmax, incr)})
172   h.Fourier.i
173 }
174 #
175 # theoretical
176 h.theory<-function(data, kernel, gamma) {
177   g=gamma
178   n=length(data)
179   N=log(n)/(2*g)
180   theta.n=switch(kernel, 1-1/N, 1, 1/2)
181
182   # output
183   h.n=theta.n/N

```

```

184   h.n
185 }
186 # M sized vector of h.theory, i=1,...,M
187 h.theory.i<-function(data, kernel, gamma, NA.1, NA.2, NA.3) {
188   M=length(data[,1])
189   h.n=h.theory(data[,1], kernel, gamma)
190   h.n.i=rep(h.n, M)
191   h.n.i
192 }
193 #
194 #-----
195 # Kernel density estimator:
196 #-----
197 #
198 # Univariate kernel density estimator
199 kde.n<-function(data, kernel, h, gamma, points) {
200   n<-length(data)
201
202   # calculate equally spaced grid points on the interval (gridmin, gridmax)
203   range<-range(data)
204   gridmin<-range[1]-h
205   step=(2.0*h+abs(range[2]-range[1]))/(points-1)
206   grid=sapply(1:points, function(j){gridmin+step*j })
207
208   # calculate kernel density estimates at each grid point
209   k<-switch(kernel, fun<-ker.Fejer.type, fun<-ker.sinc, fun<-ker.dlVP, fun<-ker.
      Gauss)
210   nh<-h*n
211   fh=sapply(1:points, function(j){(1/nh)*sum(k(n, (data-grid[j])/h, gamma)[,1])
      })
212   fh=pmax(0, fh)
213
214   # output
215   return.mtx<-cbind(grid, fh)
216   return.mtx
217 }
218 # n x M matrix of Kernel density estimator, i=1,...,M
219 kde.n.i<-function(data, kernel, h, gamma, points) {
220   M=dim(data)[2]
221   f.i.n=sapply(1:M, function(i){kde.n(data[,i], kernel, h[i], gamma, points)
      [,2]})
222   f.i.n
223 }
224 #
225 # Average of all M kernel density estimators
226 kde.bar<-function(kde.n.i) {
227   n=dim(kde.n.i)[1]
228   M=dim(kde.n.i)[2]
229   kde.bar=sapply(1:n, function(j){(1/M)*sum(kde.n.i[j,])})
230   kde.bar
231 }
232 #
233 #-----
234 # MISE

```

```

235 #-----
236 y<-runif(10000)
237 z<-runif(10000)
238 #
239 # MISE estimate evaluated using CMC
240 MISE.CMC<-function(density,data, kernel,h, gam,y,z) {
241   n=length(data[,1])
242   M=length(data[1,])
243   k<-switch(kernel, fun<-fejer.type, fun<-sinc, fun<-dlvp, fun<-gauss)
244   x=1-1/y
245   s=1/z-1
246
247   int<-function(i) {
248     x.vec<-sapply(1:length(y), function(k) {matrix(rep(x[k],n), ncol=n, byrow=T)
249       })
250     d.ijk=(x.vec-data[,i])/h[i]
251     K=sapply(1:length(y), function(k) {1/(n*h[i]) *sum(k(n,d.ijk[,k], gam)[,1])
252       })
253     f<-switch(density, (1/sqrt(2*pi)) *exp(-(x^2)/2), (gamma(8)/(sqrt(15*pi) *
254       gamma(15/2))) * (1+x^2/15)^-(8), (1/4) *x*exp(-x/2), 0.4 * ((1/sqrt(2*pi)) *
255       exp(-x^2/2)) + 0.6 * ((1/sqrt(2*pi*0.16)) *exp(-(x-1)^2/(2*0.16))))
256     g.y=(K-f)^2 * (1/y^2)
257     if(is.finite(mean(g.y))) {int1=mean(g.y)} else {int1=0}
258
259     s.vec<-sapply(1:length(z), function(k) {matrix(rep(s[k],n), ncol=n, byrow=T)
260       })
261     d.ijk=(s.vec-data[,i])/h[i]
262     K=sapply(1:length(z), function(k) {1/(n*h[i]) *sum(k(n,d.ijk[,k], gam)[,1])
263       })
264     f<-switch(density, (1/sqrt(2*pi)) *exp(-(s^2)/2), (gamma(8)/(sqrt(15*pi) *
265       gamma(15/2))) * (1+s^2/15)^-(8), (1/4) *s*exp(-s/2), 0.4 * ((1/sqrt(2*pi)) *
266       exp(-s^2/2)) + 0.6 * ((1/sqrt(2*pi*0.16)) *exp(-(s-1)^2/(2*0.16))))
267     g.z=(K-f)^2 * (1/z^2)
268     int2=mean(g.z)
269
270     int=int1+int2
271     int
272   }
273   M.ints=sapply(1:M, int)
274
275   # output
276   MISE=(1/M) *sum(M.ints)
277   MISE
278 }

```

scripts/fun-kdes.R

```

1 #####
2 #Olga Kosta
3 #2015, Carleton University
4 #
5 # Script contains the VaR functions for the following models:
6 # (1) Normal VaR
7 # (2) historical simulation

```

```

8 # (3) kernel quantile estimators (KQE)
9 #####
10 #-----
11 # (1) Normal VaR
12 #-----
13 VaR.normal<-function(data,p.0){
14   sigma=sd(data)
15   mu=mean(data)
16   VaR=sigma*qnorm(p.0)+mu
17   VaR
18 }
19 #-----
20 # (2) historical simulation
21 #-----
22 VaR.hist<-function(data,p.0){
23   n=length(data)
24   X=data[order(data)]
25   np=n*p.0
26   VaR=X[floor(np)+1]
27   VaR
28 }
29 #-----
30 # (3) kernel quantile estimators (KQE)
31 #-----
32 VaR.KQE<-function(data,p.0,kernel,h,gamma){
33   k<-switch(kernel,fun<-ker.Fejer.type,fun<-ker.sinc,fun<-ker.dlVP,fun<-ker.
      Gauss)
34   n=length(data)
35   X=data[order(data)]
36   num=sum(sapply(1:n,function(i){k(n,((i-1/2)/n-p.0)/h,gamma)[,1]*X[i]}))
37   den=sum(sapply(1:n,function(j){k(n,((j-1/2)/n-p.0)/h,gamma)[,1]}))
38   VaR=num/den
39   VaR
40 }

```

scripts/fun-VaR.R

Appendix C

Simulation Results

The following tables give the results from the simulation studies based on the realizations generated from the following probability 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 $0.4N(0, 1) + 0.6N(1, 0.4^2)$. Each table displays the estimated values for the MISE, as in (4.5), of density estimates with a specified bandwidth selector that use the Fejér-type, sinc, de la Vallée Poussin (dVVP), and Gaussian kernels, for a variety of sample sizes n . The MISE estimates are averaged over 200 simulations. The bandwidth selection methods employed include the cross-validation, the approach obtained from the Fourier analysis techniques, and the theoretical bandwidth. The tables also include the optimal choices of the smoothing parameter γ , denoted by γ^* , and the parameters $\theta_n^* = 1 - (2\gamma^*)/\log n$ and $h_n^* = (2\gamma^*\theta_n^*)/\log n$ corresponding to the optimal γ^* for each of the asymptotically efficient kernels.

n	γ^*	θ_n^*	$\widehat{\text{MISE}}(f_n^+, f), f = \phi$			
			Fejér-type	sinc	dLVP	Gaussian
			$f_n^+ = f_{n,\text{CV}}^+, h = \hat{h}_{\text{CV}}$			
25	1.50	0.068	0.023191	0.031854	0.026902	0.029137
50	1.60	0.182	0.012677	0.016241	0.014262	0.014520
75	1.60	0.259	0.009548	0.011693	0.010161	0.011471
100	1.60	0.305	0.006701	0.008170	0.007322	0.008009
200	1.60	0.396	0.004425	0.005245	0.004806	0.005346
250	1.60	0.420	0.003056	0.003671	0.003396	0.004079
300	1.60	0.439	0.002483	0.003182	0.002635	0.003231
400	1.60	0.466	0.002102	0.002453	0.002128	0.002952
500	1.60	0.485	0.001921	0.002392	0.001939	0.002661
600	1.62	0.494	0.001384	0.001632	0.001390	0.002040
700	1.50	0.542	0.001193	0.001385	0.001236	0.001687
750	1.50	0.547	0.001072	0.001228	0.001111	0.001559
800	1.50	0.551	0.001133	0.001341	0.001141	0.001601
900	1.50	0.559	0.000990	0.001109	0.001033	0.001432
1000	1.50	0.566	0.000955	0.001142	0.001018	0.001409
			$f_n^+ = f_{n,\text{F}}^+, h = \hat{h}_{\text{F}}$			
25	1.20	0.254	0.018773	0.023342	0.019749	0.018778
50	1.40	0.284	0.011641	0.013329	0.011970	0.012417
75	1.30	0.398	0.008403	0.009095	0.008810	0.009808
100	1.30	0.435	0.006175	0.007022	0.006180	0.007068
200	1.70	0.358	0.003681	0.004062	0.004010	0.004702
250	1.70	0.384	0.002641	0.003132	0.002994	0.003586
300	1.70	0.404	0.002193	0.002726	0.002364	0.002991
400	1.70	0.433	0.001646	0.001881	0.001860	0.002514
500	1.60	0.485	0.001468	0.001676	0.001517	0.002202
600	1.59	0.503	0.001159	0.001328	0.001162	0.001983
700	1.60	0.512	0.000953	0.001114	0.000990	0.001700
750	1.60	0.517	0.000889	0.001068	0.000915	0.001552
800	1.60	0.521	0.000863	0.000997	0.000879	0.001433
900	1.60	0.530	0.000738	0.000852	0.000783	0.001433
1000	1.60	0.537	0.000829	0.000943	0.000864	0.001314

Table C.1: Estimated values for the MISE of density estimates with bandwidths \hat{h}_{CV} and \hat{h}_{F} that use the Fejér-type, sinc, dLVP, and Gaussian kernels based on the realizations generated from $N(0, 1)$ for a variety of sample sizes n . The MISE estimates are averaged over 200 replications.

standard normal, $N(0, 1)$										
n	Fejér-type				sinc			dlVP		
	γ^*	θ_n^*	h_n^*	$\widehat{\text{MISE}}$	γ^*	h_n^*	$\widehat{\text{MISE}}$	γ^*	h_n^*	$\widehat{\text{MISE}}$
25	1.00	0.379	0.235	0.019949	0.70	0.435	0.018285	1.10	0.342	0.013853
50	1.10	0.438	0.246	0.009940	0.80	0.409	0.009483	1.30	0.332	0.007084
75	1.20	0.444	0.247	0.007163	0.90	0.417	0.006500	1.40	0.324	0.005416
100	1.30	0.435	0.246	0.005182	0.90	0.391	0.005200	1.40	0.304	0.004307
200	1.40	0.472	0.249	0.002790	1.00	0.377	0.002849	1.60	0.302	0.002339
250	1.50	0.457	0.248	0.002073	1.00	0.362	0.002281	1.65	0.299	0.001821
300	1.60	0.439	0.246	0.001765	1.05	0.368	0.001971	1.70	0.298	0.001534
400	1.70	0.433	0.245	0.001371	1.10	0.367	0.001486	1.75	0.292	0.001257
500	1.80	0.421	0.244	0.001200	1.10	0.354	0.001338	1.80	0.290	0.001129
600	1.80	0.437	0.246	0.000941	1.15	0.360	0.001017	1.80	0.281	0.000892
700	1.80	0.450	0.248	0.000842	1.20	0.366	0.000915	1.80	0.275	0.000807
750	1.80	0.456	0.248	0.000776	1.20	0.363	0.000847	1.80	0.272	0.000750
800	1.80	0.461	0.249	0.000714	1.20	0.359	0.000790	1.80	0.269	0.000688
900	1.80	0.471	0.249	0.000669	1.20	0.353	0.000716	1.80	0.265	0.000654
1000	1.80	0.479	0.250	0.000677	1.20	0.347	0.000741	1.80	0.261	0.000665

Table C.2: *Estimated values for the MISE of density estimates with the theoretical bandwidth h_n that use the Fejér-type, sinc, and dlVP kernels based on the realizations generated from $N(0, 1)$ for a variety of sample sizes n . The MISE estimates are averaged over 200 replications.*

n	γ^*	θ_n^*	$\widehat{\text{MISE}}(f_n^+, f), f : t_{15}$			
			Fejér-type	sinc	dLVP	Gaussian
$f_n^+ = f_{n,\text{CV}}^+, h = \hat{h}_{\text{CV}}$						
25	1.50	0.068	0.025156	0.036092	0.028555	0.030356
50	1.40	0.284	0.010987	0.013895	0.012167	0.013525
75	1.60	0.259	0.009733	0.013788	0.009843	0.011324
100	2.00	0.131	0.008108	0.010539	0.008545	0.009439
200	2.20	0.170	0.004322	0.005149	0.004584	0.005018
250	2.25	0.185	0.003059	0.003784	0.003148	0.003868
300	2.30	0.194	0.002819	0.003825	0.003018	0.003634
400	2.30	0.232	0.002124	0.002704	0.002165	0.002770
500	2.35	0.244	0.001510	0.002128	0.001577	0.002238
600	2.40	0.250	0.001429	0.001672	0.001487	0.002022
700	2.40	0.267	0.001214	0.001471	0.001310	0.001758
750	2.40	0.275	0.001134	0.001492	0.001291	0.001754
800	2.40	0.282	0.001077	0.001493	0.001210	0.001608
900	2.40	0.294	0.001018	0.001211	0.001147	0.001444
1000	2.40	0.305	0.000844	0.001061	0.000941	0.001332
$f_n^+ = f_{n,\text{F}}^+, h = \hat{h}_{\text{F}}$						
25	1.35	0.161	0.022199	0.026909	0.023763	0.021931
50	1.10	0.438	0.010315	0.011697	0.010372	0.011590
75	1.40	0.351	0.008367	0.010248	0.008604	0.009220
100	1.40	0.392	0.006985	0.008374	0.007122	0.007787
200	1.70	0.358	0.003509	0.004370	0.003770	0.004255
250	1.70	0.384	0.002711	0.003024	0.002867	0.003498
300	1.70	0.404	0.002519	0.002966	0.002746	0.003290
400	1.70	0.433	0.001752	0.002128	0.001886	0.002465
500	1.70	0.453	0.001386	0.001619	0.001416	0.002033
600	1.60	0.500	0.001144	0.001433	0.001144	0.001852
700	1.64	0.500	0.001077	0.001229	0.001077	0.001750
750	1.60	0.517	0.000958	0.001141	0.000974	0.001622
800	1.60	0.521	0.000986	0.001123	0.001004	0.001556
900	1.60	0.530	0.000815	0.000999	0.000857	0.001419
1000	1.50	0.566	0.000734	0.000857	0.000793	0.001225

Table C.3: Estimated values for the MISE of density estimates with bandwidths \hat{h}_{CV} and \hat{h}_{F} that use the Fejér-type, sinc, dLVP, and Gaussian kernels based on the realizations generated from t_{15} for a variety of sample sizes n . The MISE estimates are averaged over 200 replications.

Student's t_{15}										
n	Fejér-type				sinc			dlVP		
	γ^*	θ_n^*	h_n^*	$\widehat{\text{MISE}}$	γ^*	h_n^*	$\widehat{\text{MISE}}$	γ^*	h_n^*	$\widehat{\text{MISE}}$
25	1.00	0.379	0.235	0.022253	0.70	0.435	0.020137	1.10	0.342	0.015529
50	1.10	0.438	0.246	0.009457	0.80	0.409	0.008962	1.30	0.332	0.006956
75	1.20	0.444	0.247	0.007487	0.90	0.417	0.007166	1.40	0.324	0.005692
100	1.30	0.435	0.246	0.005313	0.90	0.391	0.005464	1.40	0.304	0.004396
200	1.40	0.472	0.249	0.003013	1.00	0.377	0.003146	1.60	0.302	0.002511
250	1.50	0.457	0.248	0.002084	1.00	0.362	0.002314	1.65	0.299	0.001861
300	1.60	0.439	0.246	0.002052	1.05	0.368	0.002228	1.70	0.298	0.001867
400	1.70	0.433	0.245	0.001495	1.10	0.367	0.001659	1.75	0.292	0.001394
500	1.80	0.421	0.244	0.001124	1.10	0.354	0.001294	1.80	0.290	0.001061
600	1.80	0.437	0.246	0.000961	1.15	0.360	0.001047	1.80	0.281	0.000935
700	1.80	0.450	0.248	0.000926	1.20	0.366	0.000998	1.80	0.275	0.000901
750	1.80	0.456	0.248	0.000785	1.20	0.363	0.000865	1.80	0.272	0.000763
800	1.80	0.461	0.249	0.000811	1.20	0.359	0.000888	1.80	0.269	0.000794
900	1.80	0.471	0.249	0.000713	1.20	0.353	0.000772	1.80	0.265	0.000702
1000	1.80	0.479	0.250	0.000646	1.20	0.347	0.000703	1.80	0.261	0.000639

Table C.4: *Estimated values for the MISE of density estimates with the theoretical bandwidth h_n that use the Fejér-type, sinc, and dlVP kernels based on the realizations generated from t_{15} for a variety of sample sizes n . The MISE estimates are averaged over 200 replications.*

n	γ^*	θ_n^*	$\widehat{\text{MISE}}(f_n^+, f), f : \chi_4^2$			
			Fejér-type	sinc	dLVP	Gaussian
			$f_n^+ = f_{n,\text{CV}}^+, h = \hat{h}_{\text{CV}}$			
25	1.50	0.068	0.016304	0.022553	0.018774	0.016232
50	1.70	0.131	0.008857	0.012722	0.010346	0.008890
75	1.80	0.166	0.005675	0.008150	0.006261	0.006087
100	1.80	0.218	0.005418	0.007560	0.006385	0.005652
200	2.10	0.207	0.002972	0.003705	0.003173	0.002924
250	2.10	0.239	0.002535	0.003349	0.002975	0.002590
300	2.10	0.264	0.002352	0.002950	0.002526	0.002410
400	2.20	0.266	0.001826	0.002363	0.002059	0.001890
500	2.20	0.292	0.001494	0.001968	0.001606	0.001515
600	2.20	0.312	0.001380	0.001747	0.001477	0.001426
700	2.20	0.328	0.001226	0.001610	0.001354	0.001261
750	2.20	0.335	0.001009	0.001340	0.001079	0.001094
800	2.40	0.282	0.001117	0.001447	0.001252	0.001137
900	2.40	0.294	0.001004	0.001260	0.001056	0.001023
1000	2.40	0.305	0.000912	0.001183	0.000979	0.000928
			$f_n^+ = f_{n,\text{F}}^+, h = \hat{h}_{\text{F}}$			
25	1.30	0.192	0.010536	0.014825	0.012229	0.011008
50	1.50	0.233	0.006699	0.008335	0.007286	0.006515
75	1.40	0.351	0.004953	0.005743	0.005239	0.004399
100	1.90	0.175	0.004211	0.005444	0.004820	0.004189
200	2.25	0.151	0.002390	0.003248	0.002616	0.002375
250	2.30	0.167	0.002012	0.002806	0.002330	0.002088
300	2.30	0.194	0.001843	0.002435	0.002057	0.001932
400	2.42	0.194	0.001513	0.002052	0.001671	0.001540
500	2.50	0.195	0.001273	0.001674	0.001385	0.001304
600	2.50	0.218	0.001104	0.001462	0.001236	0.001187
700	2.60	0.206	0.001058	0.001375	0.001129	0.001094
750	2.50	0.245	0.000928	0.001212	0.001019	0.000979
800	2.50	0.252	0.000927	0.001241	0.001013	0.000979
900	2.40	0.294	0.000847	0.001156	0.000923	0.000899
1000	2.20	0.363	0.000811	0.001047	0.000867	0.000853

Table C.5: Estimated values for the MISE of density estimates with bandwidths \hat{h}_{CV} and \hat{h}_{F} that use the Fejér-type, sinc, dLVP, and Gaussian kernels based on the realizations generated from χ_4^2 for a variety of sample sizes n . The MISE estimates are averaged over 200 replications.

chi-square, χ_4^2										
n	Fejér-type				sinc			dlVP		
	γ^*	θ_n^*	h_n^*	$\widehat{\text{MISE}}$	γ^*	h_n^*	$\widehat{\text{MISE}}$	γ^*	h_n^*	$\widehat{\text{MISE}}$
25	1.00	0.379	0.235	0.026684	0.70	0.435	0.023735	1.10	0.342	0.019941
50	1.10	0.438	0.246	0.014097	0.80	0.409	0.013141	1.30	0.332	0.010585
75	1.30	0.398	0.240	0.008955	0.90	0.417	0.008472	1.40	0.324	0.007155
100	1.30	0.435	0.246	0.007329	0.90	0.391	0.007330	1.40	0.304	0.006218
200	1.40	0.472	0.249	0.003554	1.00	0.377	0.003610	1.60	0.302	0.003022
250	1.50	0.457	0.248	0.002840	1.00	0.362	0.003039	1.65	0.299	0.002425
300	1.60	0.439	0.246	0.002448	1.05	0.368	0.002639	1.70	0.298	0.002160
400	1.70	0.433	0.245	0.001843	1.10	0.367	0.002014	1.75	0.292	0.001689
500	1.80	0.421	0.244	0.001451	1.10	0.354	0.001615	1.80	0.290	0.001362
600	1.80	0.437	0.246	0.001287	1.15	0.360	0.001426	1.80	0.281	0.001226
700	1.80	0.450	0.248	0.001128	1.20	0.366	0.001224	1.80	0.275	0.001091
750	1.80	0.456	0.248	0.001024	1.20	0.363	0.001120	1.80	0.272	0.000995
800	1.80	0.461	0.249	0.001007	1.20	0.359	0.001122	1.80	0.269	0.000976
900	1.80	0.471	0.249	0.000896	1.20	0.353	0.000990	1.80	0.265	0.000881
1000	1.80	0.479	0.250	0.000839	1.20	0.347	0.000935	1.80	0.261	0.000829

Table C.6: *Estimated values for the MISE of density estimates with the theoretical bandwidth h_n that use the Fejér-type, sinc, and dlVP kernels based on the realizations generated from χ_4^2 for a variety of sample sizes n . The MISE estimates are averaged over 200 replications.*

n	γ^*	θ_n^*	$\widehat{\text{MISE}}(f_n^+, f), f = 0.4\phi(x) + 1.5\phi((x-1)/0.4)$			
			Fejér-type	sinc	dlVP	Gaussian
$f_n^+ = f_{n,\text{CV}}^+, h = \hat{h}_{\text{CV}}$						
25	1.50	0.068	0.036297	0.055314	0.043700	0.042332
50	1.80	0.080	0.023501	0.031746	0.025296	0.026751
75	1.90	0.120	0.017189	0.021277	0.018516	0.019636
100	2.00	0.131	0.011563	0.017667	0.015827	0.015093
200	2.10	0.207	0.005978	0.009081	0.009451	0.008472
250	2.10	0.239	0.005229	0.007314	0.007983	0.007515
300	2.10	0.264	0.004473	0.006343	0.006826	0.006631
400	2.10	0.299	0.003559	0.004819	0.004967	0.005489
500	2.10	0.324	0.002647	0.003891	0.003491	0.004459
600	2.10	0.343	0.002438	0.003498	0.002856	0.004327
700	2.10	0.359	0.002121	0.003138	0.002393	0.003832
750	2.10	0.366	0.001893	0.003221	0.002122	0.003629
800	2.10	0.372	0.001862	0.003036	0.002051	0.003436
900	2.10	0.383	0.001689	0.002892	0.001841	0.003167
1000	2.10	0.392	0.001536	0.002597	0.001677	0.002965
$f_n^+ = f_{n,\text{F}}^+, h = \hat{h}_{\text{F}}$						
25	1.40	0.130	0.039646	0.055221	0.043233	0.038826
50	1.40	0.284	0.024390	0.030745	0.025586	0.023980
75	1.20	0.444	0.018260	0.021165	0.018660	0.018184
100	2.10	0.088	0.014047	0.016267	0.015046	0.013830
200	2.10	0.207	0.007090	0.007897	0.009873	0.008551
250	2.10	0.239	0.005843	0.006444	0.008957	0.007433
300	2.10	0.264	0.004919	0.005412	0.007942	0.006210
400	2.20	0.266	0.003707	0.004253	0.006483	0.004840
500	2.10	0.324	0.002755	0.003408	0.004834	0.004000
600	2.10	0.343	0.002438	0.003107	0.003728	0.004054
700	2.10	0.359	0.002121	0.002865	0.002972	0.003695
750	2.10	0.366	0.001893	0.002870	0.002455	0.003602
800	2.10	0.372	0.001862	0.002651	0.002241	0.003449
900	2.10	0.383	0.001689	0.002623	0.001999	0.003341
1000	2.10	0.392	0.001535	0.002408	0.001677	0.003246

Table C.7: Estimated values for the MISE of density estimates with bandwidths \hat{h}_{CV} and \hat{h}_{F} that use the Fejér-type, sinc, dlVP, and Gaussian kernels based on the realizations generated from $0.4N(0, 1) + 0.6N(1, 0.4^2)$ for a variety of sample sizes n . The MISE estimates are averaged over 200 replications.

normal mixture, $0.4N(0, 1) + 0.6N(1, 0.4^2)$										
n	Fejér-type				sinc			dlVP		
	γ^*	θ_n^*	h_n^*	$\widehat{\text{MISE}}$	γ^*	h_n^*	$\widehat{\text{MISE}}$	γ^*	h_n^*	$\widehat{\text{MISE}}$
25	0.70	0.565	0.246	0.029405	0.50	0.311	0.036312	0.90	0.280	0.034952
50	0.65	0.668	0.222	0.017827	0.40	0.204	0.021094	0.80	0.204	0.017194
75	0.58	0.731	0.196	0.013214	0.40	0.185	0.016047	0.80	0.185	0.012860
100	0.60	0.739	0.193	0.010536	0.50	0.217	0.011966	0.80	0.174	0.009872
200	0.60	0.774	0.175	0.005581	0.50	0.189	0.006203	0.90	0.170	0.005943
250	0.60	0.783	0.170	0.005008	0.50	0.181	0.005476	0.90	0.163	0.005234
300	0.60	0.790	0.166	0.004233	0.50	0.175	0.004650	0.90	0.158	0.004437
400	0.60	0.800	0.160	0.003381	0.50	0.167	0.003736	0.90	0.150	0.003487
500	0.60	0.807	0.156	0.002638	0.50	0.161	0.002926	0.90	0.145	0.002650
600	0.60	0.812	0.152	0.002327	0.50	0.156	0.002547	0.90	0.141	0.002461
700	0.60	0.817	0.150	0.002071	0.50	0.153	0.002277	0.90	0.137	0.002131
750	0.65	0.804	0.158	0.001851	0.50	0.151	0.002142	0.90	0.136	0.001869
800	0.65	0.806	0.157	0.001816	0.50	0.150	0.002066	0.90	0.135	0.001839
900	0.65	0.809	0.155	0.001651	0.50	0.147	0.001895	0.90	0.132	0.001669
1000	0.65	0.812	0.153	0.001490	0.50	0.145	0.001760	0.90	0.130	0.001493

Table C.8: *Estimated values for the MISE of density estimates with the theoretical bandwidth h_n that use the Fejér-type, sinc, and dlVP kernels based on the realizations generated from $0.4N(0, 1) + 0.6N(1, 0.4^2)$ for a variety of sample sizes n . The MISE estimates are averaged over 200 replications.*

References

- [1] Achieser, N. (1956). *Theory of Approximation*. Frederick Ungar Publishing, New York.
- [2] Bartlett, M. (1963). Statistical estimation of density functions. *Sankhyā: The Indian Journal of Statistics, Series A*, **25**(3), 245–254.
- [3] Belitser, E. (1998). Efficient estimation of analytic density under random censorship. *Bernoulli*, **4**, 519–543.
- [4] Brown, L., Low, M., and Zhao, L. (1997). Superefficiency in nonparametric function estimation. *Annals of Statistics*, **25**, 898–924.
- [5] Cline, D. (1988). Admissible kernel estimators of a multivariate density. *Annals of Statistics*, **16**, 1421–1427.
- [6] Epanechnikov, V. (1969). Non-parametric estimation of a multivariate probability density. *Theory of Probability and Its Applications*, **14**, 153–158.
- [7] Golubev, G. (1992). Nonparametric estimation of smooth densities of a distribution in L_2 . *Problems of Information Transmission*, **23**, 57–67.
- [8] Golubev, G. and Levit, B. (1996). Asymptotically efficient estimation for analytic distributions. *Mathematical Methods of Statistics*, **5**, 357–368.
- [9] Golubev, G., Levit, B., and Tsybakov, A. (1996). Asymptotically efficient estimation of analytic functions in gaussian noise. *Bernoulli*, **2**, 167–181.
- [10] Gradshteyn, I. and Ryzhik, I. (2007). *Tables of Integrals, Series, and Products*. Elsevier Inc., United States of America, 7th edition.
- [11] Guerre, E. and Tsybakov, A. (1998). Exact asymptotic minimax constants for the estimation of analytic functions in L_p . *Probability Theory and Related Fields*, **112**, 33–51.
- [12] Hájek, J. (1972). Local asymptotic minimax and admissibility in estimation. *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, **1**, 175–194.

- [13] Härdle, W., Müller, M., Sperlich, S., and Weratz, A. (2004). *Nonparametric and Semiparametric Models*. Springer, Heidelberg.
- [14] Hasminskii, R. and Ibragimov, I. (1984). On estimation of the value of a linear function in gaussian white noise. *Theory of Probability and Its Applications*, **29**, 18–32.
- [15] Hurst, S. (1995). The characteristic function of the student t distribution. Financial Mathematics Research Report No. FMRR 006-95, Statistics Research Report No. SRR044-95.
- [16] Jorion, P. (2001). *Value at Risk: The New Benchmark for Managing Financial Risk*. McGraw-Hill, United States of America, 2nd edition.
- [17] Lepski, O. and Levit, B. (1998). Adaptive minimax estimation of infinitely differentiable functions. *Mathematical Methods of Statistics*, **7**(2), 123–156.
- [18] Levit, B. and Stepanova, N. (2004). Efficient estimation of multivariate analytic functions in cube-like domains. *Mathematical Methods of Statistics*, **13**, 253–281.
- [19] Marron, J. and Sheather, S. (1990). Kernel quantile estimators. *Journal of the American Statistical Association*, **85**, 410–416.
- [20] Mason, D. (2010). Risk bounds for kernel density estimators. *Journal of Mathematical Sciences*, **163**, 238–261.
- [21] Parzen, E. (1958). On asymptotically efficient consistent estimates of the spectral density function of a stationary time series. *Journal of the Royal Statistical Society*, **20**, 303–322.
- [22] Parzen, E. (1962). On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, **33**, 1065–1076.
- [23] Parzen, E. (1979). Nonparametric statistical data modeling. *Journal of the American Statistical Association*, **74**, 105–121.
- [24] Robert, C. and Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer, United States of America, 2nd edition.
- [25] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, **27**, 832–837.
- [26] Roussas, G. (1997). *A Course in Mathematical Statistics*. Academic Press, United States of America, 2nd edition.
- [27] Rudemo, M. (1982). Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics*, **9**, 65–78.
- [28] Rudin, W. (1987). *Real and Complex Analysis*. McGraw-Hill, Singapore, 3rd edition.

- [29] Schipper, M. (1996). Optimal rates and constants in L_2 -minimax estimation. *Mathematical Methods of Statistics*, **5**, 253–274.
- [30] Scott, D. and Terrell, G. (1987). Biased and unbiased cross-validation in density estimation. *Journal of the American Statistical Association*, **82**(400), 1131–1146.
- [31] Sen, P. and Singer, J. (1993). *Large Sample Methods in Statistics*. Chapman & Hall, Inc., United States of America.
- [32] Stepanova, N. (2013). On estimation of analytic density functions in L_p . *Mathematical Methods of Statistics*, **22**(2), 114–136.
- [33] Stone, C. (1980). Optimal rates of convergence for nonparametric estimators. *Annals of Statistics*, **8**, 1348–1360.
- [34] Tsybakov, A. (2009). *Introduction to Nonparametric Estimation*. Springer Science, United States of America.
- [35] van der Vaart, A. (1998). *Asymptotic Statistics*. Cambridge University Press, New York.
- [36] Watson, G. and Leadbetter, M. (1963). On the estimation of the probability density. *Annals of Mathematical Statistics*, **34**, 480–491.