METHODS FOR DISTRIBUTIONAL FORM AND SHAPE

VALIDATION AND INFERENTIAL METHODS FOR DISTRIBUTIONAL FORM AND SHAPE

By KIRILL MAYOROV,
B.Sc., M.Sc., M.Sc., M.Sc.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Doctor of Philosophy

McMaster University

DOCTOR OF PHILOSOPHY (2017)	McMaster University
(Mathematics)	Hamilton, Ontario

TITLE:	Validation and Inferential Methods for Distributional Form and Shape
AUTHOR:	Kirill Mayorov B.Sc. in Mathematics (Lobachevsky State University of Nizhni Novgorod, Russia) M.Sc. in Mathematics (Lobachevsky State University of Nizhni Novgorod, Russia) M.Sc. in Mathematics (McMaster University) M.Sc. in Statistics (McMaster University)
SUPERVISOR:	Professor N. Balakrishnan
NUMBER OF PAGES:	xvii, 147

Abstract

This thesis investigates some problems related to the form and shape of statistical distributions with the main focus on goodness of fit and bump hunting.

A bump is a distinctive characteristic of distributional shape. A search for bumps, or bump hunting, in a probability density function (PDF) has long been an important topic in statistical research. We introduce a new definition of a bump which relies on the notion of the curvature of a planar curve. We then propose a new method for bump hunting which is based on a kernel density estimator of the unknown PDF. The method gives not only the number of bumps but also the location of their centers and base points.

In quantitative risk applications, the selection of distributions that properly capture upper tail behavior is essential for accurate modeling. We study tests of distributional form, or goodness-of-fit (GoF) tests, that assess simple hypotheses, i.e., when the parameters of the hypothesized distribution are completely specified. From theoretical and practical perspectives, we analyze the limiting properties of a family of weighted Cramér-von Mises GoF statistics $W_n^{2,\beta}$ with weight function $\psi(t) = 1/(1-t)^{\beta}$ (for $\beta \leqslant 2$) which focus on the upper tail. We demonstrate that $W_n^{2,2}$ has no limiting distribution. For this reason, we provide a normalization of $W_n^{2,2}$ that leads to a non-degenerate limiting distribution.

Further, we study $W_n^{2,\beta}$ for composite hypotheses, i.e., when distributional parameters must be estimated from a sample at hand. When the hypothesized distribution is heavy-tailed, we examine the finite sample properties of $W_n^{2,\beta}$ under the Chen-Balakrishnan transformation that reduces the original GoF test (the direct test) to a test for normality (the indirect test). In particular, we compare the statistical level and power of the pairs of direct and indirect tests. We observe that decisions made by the direct and indirect tests agree well, and in many cases they become independent as sample size grows.

Acknowledgements

This work could never have been completed without ample guidance, assistance and encouragement. In this regard, I would like to thank the following persons:

My supervisor, Professor N. Balakrishnan, who knows virtually everything and finds joy in sharing. Thank you for sharing some of your knowledge and wisdom with me. I am indeed privileged to learn from a scholar of your calibre.

Professor Aaron Childs, Professor Shui Feng, and Professor Angelo Canty for valuable advice and comments which led to important improvements of this thesis.[1]

James Hristoskov, a colleague from Royal Bank of Canada: It is your optimistic turn of mind and constant encouragement to overcome difficulties that helped me move smoothly toward the completion of this endeavor.

Andrei Biryuk, a bright mathematician and a long-standing friend, for shedding light on some important mathematical paradigms.

A special word of thanks to my parents, Irina and Vyacheslav, for all the effort you put in to my education, both as a person and a student. I greatly appreciate all the time you invested and the example you set.

My wife, Anna, for bearing with me through late nights and the many hours of being a computer widow. Thank you for your love, support and companionship.

Finally, I wish to thank my wonderful children, Dmitry and Maria. Thank you for bearing with my mood swings and for your patience during daddy's studies.

[1] All remaining errors are mine.

To my family

Contents

List of Figures

List of Tables

List of Acronyms and Abbreviations

AD	Anderson-Darling
ALD	Aggregate Loss Distribution
AMA	Advanced Measurement Approach
AMISE	Asymptotic Mean Integrated Squared Error
BCV	Biased Cross-Validation
BIA	Basic Indicator Approach
CBT	Chen-Balakrishnan Transformation
CDF	Cumulative Distribution Function
CvM	Cramér-von Mises
DPI	Direct Plug-In
EDF	Empirical Distribution Function
GCM	Greatest Convex Minorant
GoF	Goodness-of-Fit
IID	Independent and Identically Distributed
KDE	Kernel Density Estimation (or Estimator)
LDA	Loss Distribution Approach
LSCV	Least Squares Cross-Validation
MC	Monte Carlo
MISE	Mean Integrated Squared Error
MLE	Maximum Likelihood Estimation (or Estimator)
NCM	Nonparametric Clustering Method
OpRisk	Operational Risk
OU	Ornstein-Uhlenbeck
PDF	Probability Density Function
RAM	Random-Access Memory
RoT	Rule-of-Thumb
SJ	Sheather-Jones
STE	Solve-the-Equation
TBH	Targeted Bump Hunting
TSA	The Standardized Approach
UEP	Uniform Empirical Process
WCvM	Weighted Cramér-von Mises
wp1	With Probability One

List of Symbols and Notations

AD^2_{up}	The same as $W_n^{2,2}$
AD^2_{down}	WCvM with weight function $1/t^{\beta}$ for a simple null hypothesis
$B(t)$	Standard Brownian bridge at t
$\widehat{B}(t)$	Estimated Brownian bridge at t
$C(x)$	Curvature of a planar curve at x
$\widehat{C}_h(x)$	KDE of the curvature of planar curve with bandwidth h
$C^{2,\beta}_{n,\alpha}$	Critical value of $W_n^{2,\beta}$ at significance level α
$Cov(X,Y)$	Covariance of random variables X and Y
$\stackrel{d}{=}$	Has the same distribution as
$\stackrel{d}{\to}$	Converges in distribution
$\mathbb{E}[X]$	Mathematical expectation of a random variable X
$E_n(x)$	$F_n(x)$ for a sample of size n at x from the Uniform distribution on $[0,1]$
$e_n(t)$	UEP for a sample of size n at t
$\hat{e}_n(t)$	Estimated UEP for a sample of size n at t
$F(x)$	CDF at x
$F_n(x)$	Empirical CDF for a sample of size n at x
$f(x)$	PDF at x
$\widehat{f}_h$	KDE of f with bandwidth h
$\widehat{F}_n(x)$	Estimated empirical CDF for a sample of size n at x
$\mathbb{I}_A(x)$	Indicator function on a set A at x
$Logn(x;\mu,\sigma)$	Random variable with PDF $N(\ln x;\mu,\sigma)/x$
$N(x;\mu,\sigma)$	$(1/\sigma)\varphi((x-\mu)/\sigma)$
$N(\mu,\sigma)$	Random variable with PDF $N(x;\mu,\sigma)$
$\mathbb{N}_0$	$\{0,1,2,\dots\}$
$\stackrel{P}{\to}$	Converges in probability
$X_n = O_P(Y_n)$	For any $\varepsilon>0$, there exist $M_\varepsilon\in(0,\infty)$ and $N_\varepsilon\in\mathbb{N}$, such that $P(\lvert X_n/Y_n\rvert>M_\varepsilon)<\varepsilon$ for all $n\geqslant N_\varepsilon$
$X_n = o_P(Y_n)$	$X_n/Y_n \stackrel{P}{\to} 0$ as $n\to\infty$
$\mathscr{R}_\beta$	Class of regularly varying functions with parameter β
$\mathbb{R}_+$	$[0,+\infty)$
$\mathscr{S}$	Class of subexponential functions
$VaR_\delta(L)$	Value-at-risk at confidence level δ for a random variable L
$\mathbb{V}[X]$	Variance of a random variable X
$W(t)$	Standard Brownian motion at t
$W_n^{2,\beta}$	WCvM with weight function $1/(1-t)^\beta$ for a simple null hypothesis
$\widehat{W}_n^{2,\beta}$	WCvM with weight function $1/(1-t)^\beta$ for a composite null hypothesis
α	Significance level
$\varphi(x)$	$(1/\sqrt{2\pi})\exp(-x^2/2)$
$\Phi(x)$	$\int_{-\infty}^{x}\varphi(x)dx$
∇_θ	Gradient with respect to θ
$\overline{i,j}$	$i,i+1,i+2,\dots,j$
$\langle\theta_1^T,\dots,\theta_m^T\rangle$	A row vector with components $\theta_1\in\mathbb{R}^{k_1},\dots,\theta_m\in\mathbb{R}^{k_m}$
$X\sim F(\cdot;\boldsymbol{\theta})$	Random variable X has CDF $F(\cdot;\boldsymbol{\theta})$
$\sqcup$	Disjoint union
$\blacksquare$	The end of a proof

Declaration of Academic Achievement

I hereby declare that I am responsible for the content of the research in this thesis, for writing this thesis, for the formulation and proof of theoretical results, for the design of numerical experiments and for the programming.

I recognize the contributions of Dr. Narayanaswamy Balakrishnan, James Hristoskov and Dr. Andrei Biryuk in the research process.

Dr. Narayanaswamy Balakrishnan, my supervisor and Professor of Statistics in the Department of Mathematics & Statistics at McMaster University, provided comprehensive guidance and insightful advice on statistical methodology in the course of the research for each topic in the thesis.

James Hristoskov, Senior Director of Risk Methodologies at Royal Bank of Canada, introduced me to the topic of multimodality and bump hunting in probability density functions. He assisted me with the development of an early prototype version of the bump-hunting method (Chapter 2) and provided ample guidance to the context of applications of a class of weighted Cramér-von Mises test statistics in the modeing of operational risk (Chapter 3).

Dr. Andrei Biryuk, Associate Professor of Mathematics in the Department of Mathematics & Computer Sciences at Kuban State University in the Russian Federation, provided in-depth advice on the theory associated with several methods used in Chapter 2 and Chapter 3.

Chapter 1

Introduction

1.1 Bump Hunting in Probability Density Functions

A search for bumps, or bump-hunting, in a probability density function (PDF) has long been an important topic in statistical research. In a PDF without flat regions,[1] a bump is defined as the part lying between two inflection points and that is concave if viewed from below. A bump may not necessarily correspond to a local maximum. As Good and Gaskins (1980) and Dazard and Rao (2010) rightly point out, even if a bump is not a local maximum, it still indicates some features of a random variable requiring an explanation, and these features often reveal some useful characteristics of the underlying phenomena leading to scientific discoveries.

If a bump does correspond to a local maximum, it is referred to as a modal bump. Bump hunting is then closely related to tests for multimodality in PDFs. The latter has developed into a vast subject (e.g., Silverman, 1980; Hartigan and Hartigan, 1985; Müller and Sawitzki, 1991; Minnotte, 1997; Cheng and Hall, 1999; Andryushkiw et al., 2008). Mode analysis is also paramount in some clustering procedures – the ones which view modes as centers of clusters in a data set (e.g., Wishart, 1969; Azzalini and Torelli, 2007; Menardi and Azzalini, 2014).

The modal bump estimation procedures are roughly of the following two types (Dharmadhikari and Joag-Dev, 1988): direct and indirect. The direct methods are based on the clustering of observations and go back to Chernoff (1964). The indirect methods obtain estimators of the unknown PDF from the data at hand and

[1] Flat parts play a special role; see Müller and Sawitzki (1991, p. 740) and Appendix A.1.3.

derive estimators of a bump as a by-product. This group of methods is termed nonparametric and was originated by Parzen (1962b). A principal representative of nonparametric methods, kernel density estimation (KDE), has given rise to a number of methods of bump-hunting (Silverman, 1980, 1986; Azzalini and Torelli, 2007; Menardi and Azzalini, 2014). A detailed overview of nonparametric various bump hunting methods can be found in Hall et al. (2004).

Bump-hunting tasks arise in a variety of areas ranging from physics and astronomy (Good and Gaskins, 1980), epidemiology (Jaffe et al., 2012) to financial risk modeling. In operational risk modeling, for example, a prior knowledge of modal structure of loss severities allows the modeler to select a suitable set of distributions that accommodate shapes with multiple bumps.

In this thesis, we propose a new method for bump-hunting which is based on a kernel density estimator[2] of the unknown PDF. Motivated by geometrical considerations, we introduce a new definition of a bump. The proposed method borrows from the notion of curvature of a planar curve and utilizes it to make "surgeries" to the KDE. A surgery is defined as an operation of cutting the "cap" of a bump (the part under the bump confined between the KDE and the chord connecting the bump's two base points). The cap is then evaluated as follows: If its area exceeds a set threshold, the bump is counted as a prominent one, otherwise, a number of steps are performed to determine if the bump is part of an adjacent bump(s) whose cap's area is greater than the threshold or not. As the proposed method aims to identify bumps with the cap's area greater than a certain threshold, it will be referred to as Targeted Bump Hunting (TBH).

1.2 Goodness-of-Fit Tests in Operational Risk Modeling

This section uses material from the following article:

- K. Mayorov, J. Hristoskov and N. Balakrishnan. On a family of weighted Cramér-von Mises goodness-of-fit tests in operational risk modeling. *Journal of Operational Risk*, forthcoming 2017.

[2]A kernel density estimator will be abbreviated as KDE throughout this thesis when there will be no confusion with kernel density estimation.

Financial institutions bear a number of operational risks while conducting business activities. These risks can often result in potentially significant loss events that can degrade the firm's assets to levels that are lower than liabilities. Under such circumstances financial institutions will not have adequate backing to compensate losses of depositors and lenders, and will usually be faced with imminent default. In order to mitigate the potential for default from adverse operational risk events, firms are required to keep sufficient equity capital that can foreseeably absorb all significant losses without jeopardizing liabilities and causing bankruptcy.

Definition. *Operational risk* (OpRisk) is defined as the risk of loss resulting from inadequate or failed internal processes, people and systems or from external events. This definition includes legal risk, but excludes strategic and reputational risk.

Some examples of OpRisk events (McNeil et al., 2015) are:

- Fraud (internal and external), losses due to IT failure, errors in settlements of transactions, losses due to external events like flooding, fire, earthquake or terrorism;
- System errors, such as the estimated $440 million loss from a computer-trading glitch at Knight Capital Group in 2012;
- Legal losses and penalties such as Bank of America's August 2014 record fine of $16.65 billion for the misselling of financial products on the basis of inaccurate or misleading information about their risks.

Regulators typically require financial institutions to hold equity capital for OpRisk at a level that can ensure that the total yearly loss from extreme or on aggregate large operational events not exceed capital with a chance of 0.1 percent over one year time horizon. As such, to ensure capital adequacy, financial institutions attempt to determine the 1-year, 99.9 percentile aggregate loss that is possible to occur and ensure that equity capital is at a level that is sufficient to cover such loss.

The 99.9 percentile aggregate loss is specified by the Basel Committee on Banking Supervision (Basel Committee on Banking Supervision, 2004) as representing the target for minimum capitalization under Pillar 1 of its three pillar system. At present, Pillar 1 minimum capitalization for OpRisk can be determined via three different approaches, a basic indicator approach (BIA), standardized approach (TSA) and the advanced measurement approach (AMA). The AMA is driven by regulatory principles prescribing a general framework which requires the application of frequency and severity distributions to complete an aggregate loss distribution under the loss distribution approach (LDA).

Under Basel II, Pillar 2 directives, financial institutions can apply internal models to complement and test the minimum capital requirements set under Pillar 1 (either through Economic Capital or Stress Testing frameworks). Being typically more advanced, Pillar 2 models for OpRisk generally follow LDA prescriptions and require well calibrated severity models as the loss targets for capitalization can exceed the 99.95 percentile.

Given the high percentile loss targets for capitalization, under both pillars, a major challenge for LDA modeling has been to accurately model loss severity with limited publicly available or internal operational loss data for more extreme events. Often the underlying problem is that less than a minimum set of historic losses have been collected for adequate statistical modeling. In addition, wrong candidate severity distributions may be improperly selected. The combination of these problems can lead to significant error in modeling for capital adequacy.

Motivated by these practical considerations, in Chapter 3, we address the problem of more accurate severity distribution selection under an LDA approach in the case of simple hypotheses of goodness of fit (GoF). A simple GoF hypothesis is that a random variable X is distributed according to a CDF with completely specified parameters, i.e., H_0: $X \sim F(\cdot;\boldsymbol{\theta}_0)$ with both the distributional form of F and its parameters $\boldsymbol{\theta}_0$ being known. We discuss distributional properties of a family of weighted Cramér-von Mises (WCvM) GoF test statistics, $W_n^{2,\beta}$, with weight function $\psi(t) = 1/(1-t)^{\beta}$, that are suitable for more accurate selection of best-fit severity distributions. We demonstrate that for weights where $\beta \geqslant 2$ the WCvM test statistic does not have limiting distributions, which may limit its practical utility. In order to rectify the problem we provide a normalization for $\beta = 2$, which leads to a non-degenerate asymptotic distribution. Notwithstanding the normalization, our analysis will show that the tests provide greater utility when $\beta < 1.5$ and that for $\beta \geqslant 1.5$ utility is questionable as only Monte Carlo schemes, which are shown to be very slow in approaching the asymptotic regime, are practical even for very large samples.

1.3 The Chen-Balakrishnan Transformation to Normality

GoF tests play an important role in statistics and its applications. The theory of GoF tests is well established (D'Agostino and Stephens, 1986; Thas, 2010; Bagdonavičius et al., 2011). The range of applications of GoF tests spans from reliability theory (Chen and Balakrishnan, 1995; Barros et al., 2014) to medicine (Kuss, 2002) and risk modeling (see Chapter 3).

In applications, testing whether the distribution of a random variable belongs to a given family of distributions with unspecified parameters, i.e., testing of a composite hypothesis H_0: $X \sim F(\cdot;\boldsymbol{\theta})$, is not uncommon. In this case, the distributional form of F is known, but the parameter vector $\boldsymbol{\theta} \in \Omega \subseteq \mathbb{R}^m$ (for some $m \in \mathbb{N}$ and open set Ω) of the CDF $F(\cdot;\boldsymbol{\theta})$ must be estimated from a sample at hand.

If a statistic is chosen to assess this type of hypotheses, one often has to resort to a Monte Carlo (MC) method, such as a parametric bootstrap, to calculate a p-value or critical value of a given statistic. Although the processing power and available amounts of random-access memory (RAM) of modern computers have significantly increased computational tractability of Monte Carlo methods, evaluation of critical values may still face problems of long execution time as each MC trial includes distributional parameter estimation. This is the case in OpRisk modeling wherein dozens of truncated, multi-parameter severity distributions are tested via MC simulations.

In Chen (1991) and Chen and Balakrishnan (1995), a transformation was proposed that serves to avoid MC simulations. We refer to this transformation as the Chen-Balakrishnan transformation (CBT) - the term suggested in Goldmann et al. (2015, p. 60). CBT reduces the initial hypothesis H_0 (the direct test) to an auxiliary composite hypothesis for normality $H_0^\star$: $Y \sim N(\mu,\sigma)$ (the indirect test), with $N(\mu,\sigma)$ denoting a normally distributed random variable with mean μ and variance σ^2 whose CDF is $\Phi((\cdot-\mu)/\sigma)$, where $\Phi(x) = (1/\sqrt{2\pi})\int_{-\infty}^{x} \exp(-t^2/2)dt$. CBT is presented in Algorithm 1.1.

CBT requires efficient estimation of the unknown parameter. It is known that

Algorithm 1.1 The Chen-Balakrishnan Transformation

Input: Random sample $X_1, \dots, X_n$, CDF $F(\cdot; \boldsymbol{\theta})$ with unknown $\boldsymbol{\theta} \in \Omega \subseteq \mathbb{R}^m$ for $m \in \mathbb{N}$ in $H_0 : X \sim F(\cdot; \boldsymbol{\theta})$.

Output: Transformed null $H_0^\star$: $Y \sim N(\mu, \sigma)$ for unknown μ and σ.

1: Let $X_{(1)} \leqslant \cdots \leqslant X_{(n)}$ be order statistics of the sample;

2: Efficiently estimate $\boldsymbol{\theta}$ by $\widehat{\boldsymbol{\theta}}_n$ based on $\{X_{(j)}\}_{j=1}^n$;

3: Calculate $Y_{(j)} = \Phi^{-1}(F(X_{(j)}; \widehat{\boldsymbol{\theta}}_n))$;

4: Calculate $Z_j = (Y_{(j)} - \bar{Y})/s_Y$, where $\bar{Y} = (\sum_{j=1}^n Y_{(j)})/n$ and $s_Y^2 = (\sum_{j=1}^n (Y_{(j)} - \bar{Y})^2)/(n-1)$;

5: **return** $H_0^\star : Z \sim N(\mu, \sigma)$, where μ and σ have to be estimated from $\{Z_j\}_{j=1}^n$.

under certain regularity conditions, maximum likelihood (ML) estimators are asymptotically efficient (see Casella and Berger, 2002, p. 472 and p. 516).

The transformation to and assessment of $H_0^\star$ are performed in a fraction of a second (see Section 4.6). The speed and applicability to arbitrary families $F(\cdot; \boldsymbol{\theta})$ of null distributions makes this transformation approach attractive in practice.

In the literature, the auxiliary normality hypothesis has mainly been studied on a standalone basis (Chen and Balakrishnan, 1995; Meintanis, 2009) with rare attempts to examine the quality of the approximation (Goldmann et al., 2015). However, due to the fact the mechanics of CBT relies on some assumptions which may only be approximately valid in practice, it is natural to ask to what degree decisions made by assessing $H_0^\star$ are reflective of those made by assessing H_0 directly. Secondly, since CBT is a nonlinear transformation, it is not clear a priori whether switching from H_0 to $H_0^\star$ leads to a gain in the level and power.

Motivated by these considerations, for a selected family of WCvM GoF test statistics, in this thesis, we examine agreement between the decisions made based on H_0 and $H_0^\star$ on a given data set for distributional families with varying degree of tail heaviness. We compare the level and power of the selected WCvM GoF statistics for H_0 and $H_0^\star$. In many situations, decisions of direct and indirect tests turn out to become independent as sample size grows.

1.4 Scope of This Thesis

The rest of this thesis is organized as follows.

In Chapter 2, we introduce and study the targeted bump-hunting method. Therein, Section 2.3 describes TBH. To substantiate the theoretical foundation of TBH, in Section 2.3.2, Section 2.3.3, and Appendix A.1, we establish a number of mathematical results which are also of independent interest. The method is tested on selected distributions under various settings and its performance is compared to the performance of some existing methods in Section 2.4. We discuss the performance of the method and make recommendations for the test's use in practice. We then apply TBH to a real dataset from Good and Gaskins (1980) in Section 2.5.

In Chapter 3, we study a family of weighted Cramér-von Mises test statistics. In Section 3.3, we collect results from the literature related to the supremum and quadratic classes of upper-tail test statistics. This section shows that some previously defined defined test statistics do not have well-defined limiting distributions. The notion of the spectrum of an integral operator and some of its properties associated with weighted Cramér-von Mises test statistics are presented in Section 3.4. In Section 3.6, we provide a positive affine normalization under which the statistic $W_n^{2,2}$ tends in distribution to a standard Gaussian random variable. We also obtain some results on the spectrum of the integral operator associated with $W_n^{2,2}$. In Section 3.8, we provide evidence to demonstrate that for $W_n^{2,\beta}$, $1.5 \leqslant \beta \leqslant 2$, the tests' practical utility may be limited due to a very slow rate of convergence of the finite sample distribution to the asymptotic regime.

In Chapter 4, we study the Chen-Balakrishnan transformation to normality in selected goodness-of-fit tests. In Section 4.2, we present several WCvM statistics for the purposes of the subsequent study. In Section 4.3, we discuss the convergence of $W_n^{2,\beta}$, $\beta \leqslant 2$, in distribution. In Section 4.4, on the selected WCvM statistics, we perform a simulation study on several distributions some of which are flexible to allow for a varying degree of tail heaviness. In Section 4.5, we discuss the findings of the study. An application to a set of real data of severities of operational risk losses is presented in Section 4.6.

Finally, Chapter 5 provides concluding remarks and a discussion of possible directions for future research.

Auxiliary results are relegated to appendices for conciseness in the presentation of the main text of this thesis.

Chapter 2

A Targeted Bump-Hunting Method for PDFs in a Nonparametric Setting

2.1 Introduction

In this chapter, we propose a new method for bump hunting, Targeted Bump Hunting, or TBH, which is based on KDE of the unknown PDF. The proposed method utilizes a new definition of a bump. The method performs sequentially "surgeries" to the KDE. A surgery is a series of operations of cutting the "cap" of a bump (the part between the bump and the chord connecting the bump's two base points).

The cap is then evaluated as follows: If its area exceeds a set threshold, the potential bump is counted as a true bump and not an artifact of the particular sample drawn; otherwise, certain actions are performed to determine whether the bump is part of an adjacent bump(s) whose cap's area is greater than the threshold or not.

An application to a real data set is presented to illustrate the proposed method.

2.2 Facts from Theory of Kernel Density Estimation

In this section, we give a brief overview of some key results on univariate KDE which are pertinent to the developments in subsequent sections. A comprehensive account of KDE methods can be found in Silverman (1986), Wand and Jones (1995) and Henderson and Parmeter (2015), and the results presented below are from these sources, unless stated otherwise.

Let $\{X_i : i = 1, \ldots, n\}$ be a collection of independent and identically distributed (IID) random variables with common PDF $f : \mathbb{R} \to \mathbb{R}_+$. In particular, this means $\kappa_0(f) = 1$ and $f(x) \geqslant 0$ for all $x \in \mathbb{R}$. Here, we denote $\kappa_j(g) = \int_{-\infty}^{\infty} x^j g(x)dx$, $j \in \mathbb{N}_0$, for a function $g : \mathbb{R} \to \mathbb{R}$.

Consider also a function $K : \mathbb{R} \to \mathbb{R}_+$ such that it is an even PDF and $\kappa_2(K) < \infty$. It follows that $\kappa_0(K)$ and $\kappa_1(K)$ equal 1 and 0, respectively. The function $K(\cdot)$ is referred to as the kernel function and the estimator

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{j=1}^{n} K\left(\frac{x - X_j}{h}\right) \tag{2.1}$$

is referred to as the kernel density estimator of the unknown PDF f. Here, the positive parameter $h = h(n)$, called a smoothing window, or bandwidth, is such that $h(n) \to 0$ as $n \to \infty$. In what follows, dependence of h on n will be suppressed for notational simplicity.

If $f(\cdot)$ is twice continuously differentiable, $h \to 0$ and $nh \to \infty$ as $n \to \infty$, then (2.1) is a consistent and asymptotically unbiased estimator of $f = f(x)$. More generally, if $f(\cdot)$ is $r + 2$-times continuously differentiable, where $r \in \mathbb{N}_0$, and if the derivative $K^{(r)}(\cdot)$ exists and is square integrable,[1] $h \to 0$ and $nh^{1+2r} \to \infty$ as $n \to \infty$, then, taking $\widehat{f_h^{(r)}}(\cdot) = \hat{f}_h^{(r)}(\cdot)$,

$$\hat{f}_h^{(r)}(x) = \frac{1}{nh^{r+1}} \sum_{j=1}^{n} K^{(r)}\left(\frac{x - X_j}{h}\right) \tag{2.2}$$

is a consistent and asymptotically unbiased estimator of $f = f^{(r)}(x)$.

Although there are other plausible KDEs of the unknown PDF and its derivatives (Singh, 1977), it is (2.1)-(2.2) that will be used throughout our work here.

It has been noted in the literature that although the choice of the kernel is important, the overall effect on the results obtained from the statistical analysis is limited when compared to the effect of the bandwidth. Indeed, (2.1) based on too large a bandwidth will likely erase important features of the unknown PDF; on the other hand, for too small h, $\hat{f}_h(\cdot)$ will result in a wiggly estimator thereby introducing artificial features.

The literature dedicated to bandwidth selection is abundant. Thorough discus-

[1] In Section 2.3.2 and Appendix A, some of these assumptions are relaxed.

sions are given by Henderson and Parmeter (2015, pp. 30–50) and Jones et al. (1996). There are roughly two large classes of bandwidth selection methods: optimal and data-driven. Although the classes are different, they are all based to some degree on estimation or approximation of Asymptotic Mean Integrated Squared Error (AMISE) which arises as an approximation of MISE for large sample size: $MISE(h) \approx AMISE(h)$, where $MISE(\cdot)$ is

$$MISE(h) = \mathbb{E}\left[\int_{-\infty}^{+\infty} \left(\hat{f}_h^{(r)}(x) - f^{(r)}(x)\right)^2 dx\right].$$

Then, $AMISE(\cdot)$ can be shown to equal

$$AMISE(h) = \frac{h^4}{4} R\left(f^{(r+2)}\right) \kappa_2^2(K) + \frac{R\left(K^{(r)}\right)}{nh^{1+2r}} \text{ as } n \to \infty,$$

where $R(g) = \int_{-\infty}^{\infty} g^2(x)dx$ for a function $g : \mathbb{R} \to \mathbb{R}$.

By routine calculations, the bandwidth that minimizes $AMISE(\cdot)$, known as the $AMISE$-optimal bandwidth, is found to be

$$h_{AMISE} = \left(\frac{(1+2r)R\left(K^{(r)}\right)}{R\left(f^{(r+2)}\right)\kappa_2^2(K)}\right)^{\frac{1}{5+2r}} n^{-\frac{1}{5+2r}}.$$

Let $N(\cdot;\mu,\sigma)$ denote a Gaussian PDF with mean μ and standard deviation σ, i.e., $N(x;\mu,\sigma) = (1/\sigma)\varphi((x-\mu)/\sigma)$, where $\varphi(x) = (1/\sqrt{2\pi})\exp(-x^2/2)$. Then, for the PDF[2] being $N(\mu,\sigma)$, the so-called Silverman rule-of-thumb bandwidth equals

$$h_{RoT} = c_r(K)\widehat{\sigma} n^{-1/(5+2r)},$$

where the constant $c = c_r(K)$ depends on both r and $K(\cdot)$ and $\widehat{\sigma}$ is a sample estimate of the standard deviation. If the kernel $K(x) = \varphi(x)$ is standard Gaussian, then $c_0(K) = 4/3$. The rule-of-thumb bandwidth tends to oversmooth data, especially, for data from multimodal, skewed or heavy-tailed distributions. However, the rule-of-thumb bandwidth is usually sufficient for a preliminary analysis for gaining insights from the data at hand.

The data-driven selection methods are mainly represented by plug-in (PI), least-

[2] $N(\cdot;\mu,\sigma)$ will be abbreviated to just $N(\mu,\sigma)$ when no confusion would arise.

squares cross-validation (LSCV) and biased cross-validation (BCV). There is no single data-driven estimator which can be preferred over the others in all circumstances. As discussed in Jones et al. (1996), h_{LSCV} is often too variable (especially in the direction of undersmoothing) and h_{BCV} tends to oversmooth (and has some instability problems as well). A useful compromise between h_{LSCV} and h_{BCV} is given by the PI bandwidth from Sheather and Jones (1991), h_{SJPI}, which is found to be a stable performer in estimating $f(\cdot)$. A similar conclusion has been reached in Chacón and Duong (2013) in the context of bandwidth selection for estimation of $f^{(r)}(\cdot)$: They generalize h_{SJPI} to the derivative case and, based on their study, recommend the resulting bandwidth for practical use.

2.3 The Method

In this section, we utilize one of the fundamental concepts from differential geometry, namely, the curvature of a planar curve to introduce a new definition of a bump. We then present some results on KDE of curvature. We introduce the notion of a λ-surgery and then describe the algorithm of the proposed method for bump-hunting.

2.3.1 Curvature and its Extrema, and New Definition of a Bump

We first borrow some known facts about curvature. The reader is referred to Gibson (2001) for a detailed introduction into the subject.

Let $y = g(x)$ be the graph of a twice differentiable function $g : \mathbb{R} \to \mathbb{R}$. It can be viewed as the planar curve $z(t) = (x(t), y(t))$, where $x(t) = t$ and $y(t) = g(t)$ for $t \in \mathbb{R}$. In this case, the curvature of $z = z(t)$, or, equivalently, of $y = g(x)$ is given by

$$C(x) = \frac{g''(x)}{\left(1 + (g'(x))^2\right)^{\frac{3}{2}}}.$$

Curvature is invariant in the sense that speed (i.e., $\sqrt{1 + (g'(x))^2}$) and curvature determine the curve uniquely up to the relation of congruence (Gibson, 2001, p. 87). In particular, it implies that curves of constant curvature C are either line segments ($C = 0$) or arcs of a circle ($C \neq 0$).

The points where $C(\cdot)$ changes sign are known inflection points. In particular, inflection points must satisfy[3] $C(x) = 0$, or $g''(x) = 0$.

Remark. It is worthwhile mentioning that it has been common in the KDE literature to view the second-order derivative as a measure of curvature or even refer to it as curvature. While curvature and the second-order derivative are indeed very similar when the curve does not have sharp features, the two may be vastly different otherwise. See Figure 2.2 below.

In a PDF without flat regions, Good and Gaskins (1980, p. 42) and Silverman (1986, p. 137) define a bump as the part lying between two inflection points and that is concave if viewed from below. That is, the base of the bump is formed by two inflection points. We argue that, in general, this definition may not reflect the common perception of the base of a hill which one encounters in nature, i.e., perception of the foot of a hill. Indeed, let us consider the PDF of a symmetric "claw" $y = 0.5\varphi(t) + \sum_{j=0}^{4} \varphi(10(t-(j/2-1)))$ taken from Marron and Wand (1992). It has five distinct bumps which are also local maxima. Figure 2.1 shows that inflection points (marked as squares) appear to be too high. Viable candidates for the base of the bumps are the points marked by circles and diamonds (located much lower down the bumps than inflection points), out of which the points marked by diamonds seem to be the most natural candidates to serve as base points.

These special points are the points at which the second order derivative and curvature attain positive maxima. The first and second order derivatives as well as the curvature of the "claw" PDF are depicted in Figure 2.2. The importance of curvature extrema, known as vertices, has been appreciated in image recognition and shape analysis (Eberly, 1996; Hahmann et al., 2008). Vertices characterize sharp turning points of the curve. A vertex is called a course if $C'(x) = 0, C''(x) < 0$ and $C(x) > 0$ and is called a ridge if $C'(x) = 0$, $C''(x) > 0$ and $C(x) < 0$. In this terminology, the diamonds in Figure 2.2 correspond to courses.

Remark. It may be tempting to guess that stationary points (and, in particular, maxima and minima) of the graph of a function $g(\cdot)$ will coincide with vertices. However, this is generally not the case. It is known (Gibson, 2001, p. 128) that a stationary point $x = b$ is a vertex if and only if $g'''(b) = 0$. This condition is satisfied if, for example, $y = g(x)$ is locally symmetric around $x = b$.

[3]This is a necessary but not sufficient condition for inflection. The so-called undulation points x satisfy $C(x) = 0$, but curvature is sign-constant in a vicinity of x.

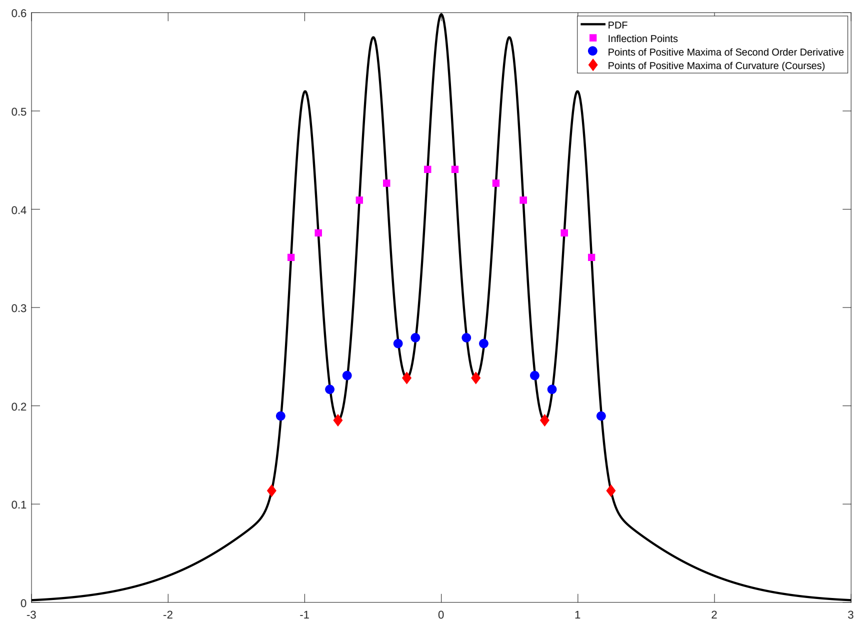

Figure 2.1: The PDF of a symmetric "claw" defined by the equation $y = 0.5\varphi(t) + \sum_{j=0}^{4} \varphi(10(t-(j/2-1)))$ and special points.

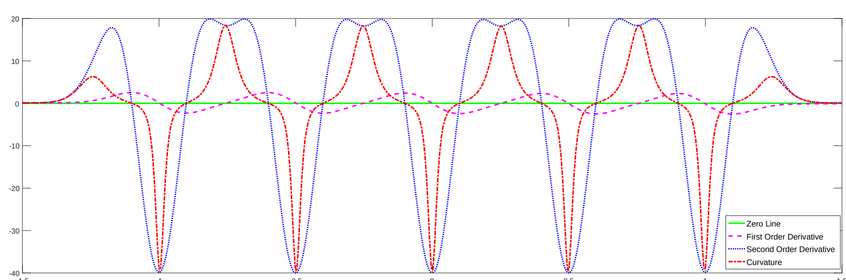

Figure 2.2: The first and second order derivatives, and the curvature of the "claw" PDF.

Motivated by these geometric considerations, we propose the following definition of a bump.

Definition 2.1 (Definition of a Bump)**.** Let $g : \mathbb{R} \to \mathbb{R}$ be a twice differentiable function and let $x = c_1$ and $x = c_2$ be two consecutive courses, such that the interval (c_1, c_2) contains a ridge $x = r_{1,2}$. Then, a *bump* is defined as the part of the graph of $g(\cdot)$ for $x \in (c_1, c_2)$.

Remark. A bump from Definition 2.1 is well defined as in Appendix A.1.3, for a continuous function $g : \mathbb{R} \to \mathbb{R}$ without flat parts, we establish that between its any two consecutive minima (maxima), there exists only one maximum (minimum).

The bump defined above will always be "wider" than the bump between two inflection points. Indeed, the Intermediate Value Theorem guarantees the existence of an inflection point $i_1 \in (c_1, r_{1,2})$ and, similarly, there exists another inflection point $i_2 \in (r_{1,2}, c_2)$. Thus, $(i_1, i_2) \subset (c_1, c_2)$.

Remark. In the rest of this thesis, a bump will be understood in the sense of Definition 2.1, unless stated otherwise.

Remark. In the spirit of Definition 2.1, a dip can be defined as the part of the graph of $g(\cdot)$ between two consecutive ridges that contains a course.

It seems convenient to classify bumps of a function $g : \mathbb{R} \to \mathbb{R}$ containing no flat pieces into the following three types:

Type 1. Function $g(\cdot)$ has a local maximum, or mode, $m \in (c_1, c_2)$, i.e., $g'(m) = 0$, $g''(m) < 0$;

Type 2. Function $g(\cdot)$ has a "shoulder" $s \in (c_1, c_2)$, i.e., $g'(s) = 0$, $g''(s) = 0$, but $g'''(s) \neq 0$ (Cheng and Hall, 1999);

Type 3. Function $g(\cdot)$ is strictly increasing or strictly decreasing for all $x \in (c_1, c_2)$, i.e., $g'(x) > 0$ or $g'(x) < 0$ throughout the interval.

Type 1 is a modal bump, Type 3 is a purely non-modal bump, and Type 2 is a non-modal bump, transitionary between modal and purely non-modal bumps. These three types are illustrated in Figure 2.3. The left panel exemplifies a PDF with two bumps, the left one of which is of Type 1, and the other bump is of Type 3. The right panel shows a PDF (borrowed from Cheng and Hall, 1999) with two bumps, too, but the bump on the left is now of Type 2 while the other one is of Type 1.

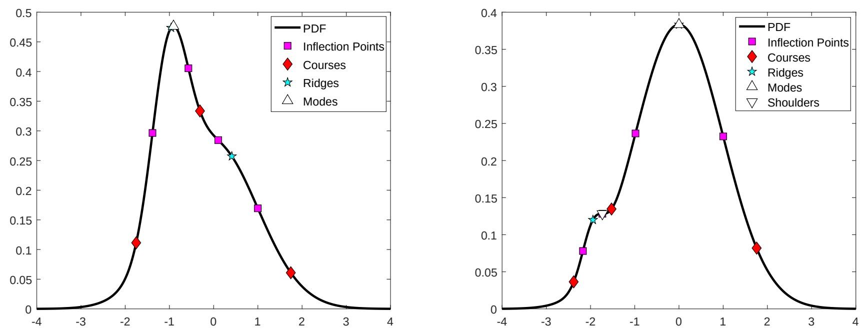

Figure 2.3: The left panel shows the plot of the PDF $y = 0.7N(0,1)+0.3N(-1,0.4)$, and the right panel shows $y = pN(0,1)+(1-p)N(-9\sqrt{3}/8, 0.25)$, where $p = 8\exp(9/8)/(1+8\exp(9/8))$. Inflection points, courses, ridges, modes, and shoulders are all shown where applicable.

2.3.2 Kernel Density Estimation of Curvature

Let us consider a collection of IID random variables $\{X_i : i = 1,\dots,n\}$ with a common unknown twice differentiable PDF $f : \mathbb{R} \to \mathbb{R}$. We wish to estimate $f(\cdot)$ and its curvature $C(\cdot)$ by KDE methods. To this end, for a fixed kernel function $K : \mathbb{R} \to \mathbb{R}$, the KDE of $f(\cdot)$ is obtained as in (2.1) while $C(\cdot)$ can be estimated as

$$\widehat{C}_{h_1,h_2}(x) = \frac{\hat{f}''_{h_1}(x)}{\left(1+\left(\hat{f}'_{h_2}(x)\right)^2\right)^{\frac{3}{2}}}, \tag{2.3}$$

where $h_1 = h_1(n)$ and $h_2 = h_2(n)$ are bandwidths corresponding to the estimation of $f''(\cdot)$ and $f'(\cdot)$. When h_1 and h_2 are equal to some common value h_3, $\widehat{C}_{h_1,h_2}(\cdot)$ will be denoted by $\widehat{C}_{h_3}(\cdot)$.

Below we present theorems for the expressions of the MISE, AMISE and AMISE-optimal bandwidth h_{AMISE} for $\widehat{C}_{h_3}(\cdot)$. The theorems are valid under a wide range of assumptions on $f(\cdot)$ and $K(\cdot)$. Under these conditions, $\widehat{C}_{h_3}(\cdot)$ is shown to be a consistent (see Corollary 2.6) and asymptotically unbiased (see Corollary 2.3) estimator of the unknown curvature function $C(\cdot)$. The proofs rely on some auxiliary lemmas which are relegated to Appendix A.

We begin by introducing the following two assumptions.

Assumption 2.1.

1. *Derivatives: Let* $g : \mathbb{R} :\to \mathbb{R}$ *be a function which possesses the second order derivative at a point* $x_0 \in \mathbb{R}$*;*

2. *Boundedness:*

   (a) $g(x) \leqslant b_1 \exp(b_2 x^2)$ *for some constants* $b_1, b_2 > 0$ *and all* $x \in \mathbb{R}$*; or*

   (b) $g(x) \leqslant c_1 + c_2 x^2$ *for some constants* $c_1, c_2 > 0$ *and all* $x \in \mathbb{R}$*.*

Consider also a function $V : \mathbb{R} :\to \mathbb{R}$ such that it satisfies the following assumption.

Assumption 2.2.

1. $\int_{-\infty}^{\infty} |V(x)|dx < \infty$*;*

2. $\int_{-\infty}^{\infty} x^2 |V(x)|dx < \infty$*;*

3. *Support and Growth:*

   (a) $\operatorname{supp}|V| \Subset \mathbb{R}$*; or*

   (b) $|V(x)| \leqslant d_1 \exp(-d_2 x^2)$ *for some constants* $d_1, d_2 > 0$ *and all* $x \in \mathbb{R}$*.*

Remark. Conditions 1 and 2 of Assumption 2.2, together with an application of the Cauchy-Schwarz inequality, imply that $\int_{-\infty}^{\infty} x|V(x)|dx < \infty$. Therefore, each $\kappa_j(V) < \infty$ for $j \in \{0, 1, 2\}$, where $\kappa_j(V) = \int_{-\infty}^{\infty} u^j V(u)du$.

Theorem 2.2. The asymptotic bias of the KDE estimator of curvature of a PDF f from (2.3) equals

$$Bias\left(\widehat{C}_h(x)\right) = \frac{h^2}{2}\left(\frac{f^{(4)}(x)}{\left(1+\left(f'(x)\right)^2\right)^{\frac{3}{2}}} - \frac{3f'(x)f''(x)f'''(x)}{\left(1+\left(f'(x)\right)^2\right)^{\frac{5}{2}}}\right)\kappa_2(K) + o(h^2) + O\left(\frac{1}{nh^4}\right)$$

as $n \to \infty$.

Proof. Indeed, from Lemma A.7, Lemma A.10 and (A.9), we get:

$$\begin{aligned}
\mathbb{E}\left[\widehat{C}_h(x)\right] &= \frac{f''(x)+\frac{h^2}{2}f^{(4)}(x)\kappa_2(K)+o(h^2)}{\left(1+\left(f'(x)+\frac{h^2}{2}f'''(x)\kappa_2(K)+o(h^2)\right)^2\right)^{\frac{3}{2}}}+O\left(\frac{1}{nh^4}\right)\\
&= C(x)\left(1+\frac{h^2}{2}\frac{f^{(4)}(x)\kappa_2(K)}{f''(x)}+o(h^2)\right)\left(1-\frac{h^2}{2}\frac{3f'(x)f'''(x)\kappa_2(K)}{\left(1+\left(f'(x)\right)^2\right)}+o(h^2)\right)\\
&+O\left(\frac{1}{nh^4}\right)
\end{aligned}$$

Then,

$$\mathbb{E}\left[\widehat{C}_h(x)\right]=C(x)\left(1+\frac{h^2}{2}\frac{f^{(4)}(x)\kappa_2(K)}{f''(x)}-\frac{h^2}{2}\frac{3f'(x)f'''(x)\kappa_2(K)}{\left(1+\left(f'(x)\right)^2\right)}\right)+o(h^2)+O\left(\frac{1}{nh^4}\right).$$

∎

Corollary 2.3. The curvature estimator is asymptotically unbiased.

Theorem 2.4. The variance of the KDE estimator of curvature of a PDF f from (2.3) equals

$$\mathbb{V}\left[\widehat{C}_h(x)\right]=\frac{1}{nh^5}\frac{f(x)R\left(K''\right)}{\left(1+\left(f'(x)\right)^2\right)^3}+O\left(\frac{1}{nh^5}\right)+O\left(\frac{1}{n^{3/2}h^7}\right)$$

as $n\to\infty$.

Proof. Using (A.9), (A.11), Lemmas A.7, A.12 and A.13, we get

$$\begin{aligned}
\mathbb{V}\left[\widehat{C}_h(x)\right] &= \frac{\mathbb{V}[X]}{\left(1+(\mathbb{E}[Y])^2\right)^3}+O(1)\,O\left(\frac{1}{nh^3}\right)-O(1)\,O\left(\frac{1}{nh^3}\right)+O\left(\frac{1}{n^{3/2}h^7}\right)\\
&= \left(\frac{f(x)R\left(K''\right)}{nh^5}+O\left(\frac{1}{nh^5}\right)\right)\frac{1}{\left(1+\left(f'(x)\right)^2\right)^3}(1+O(1))+O\left(\frac{1}{nh^3}\right)+O\left(\frac{1}{n^{3/2}h^7}\right)\\
&= \frac{1}{nh^5}\frac{f(x)R\left(K''\right)}{\left(1+\left(f'(x)\right)^2\right)^3}+O\left(\frac{1}{nh^5}\right)+O\left(\frac{1}{n^{3/2}h^7}\right).
\end{aligned}$$

∎

Thus, we obtain the following theorem regarding the MSE and MISE of the curvature estimator.

Theorem 2.5. (1) The mean squared error of the curvature estimator (A.6) when $h_1 = h_2 = h$ equals

$$MSE\left(\widehat{C}_h\right) = \frac{1}{4}h^4\alpha^2(x) + \frac{1}{nh^5}\beta(x) + o(h^4) + O\left(\frac{1}{nh^5}\right) + O\left(\frac{1}{n^{3/2}h^7}\right);$$

(2) The mean squared integrated error of the curvature estimator (A.6) when $h_1 = h_2 = h$ equals

$$MISE\left(\widehat{C}_h\right) = \frac{1}{4}h^4 R(\alpha) + \frac{1}{nh^5}\int_{-\infty}^{\infty}\beta(x)\,dx + o(h^4) + O\left(\frac{1}{nh^5}\right) + O\left(\frac{1}{n^{3/2}h^7}\right),$$

where

$$\alpha(x) = \frac{1}{\left(1+\left(f'(x)\right)^2\right)^{\frac{3}{2}}}\left(f^{(4)}(x) - \frac{3f'(x)f''(x)f'''(x)}{1+\left(f'(x)\right)^2}\right)\kappa_2(K)$$

and

$$\beta(x) = \frac{f(x)R\left(K''\right)}{\left(1+\left(f'(x)\right)^2\right)^3}.$$

Consistency of $\widehat{C}_h(x)$ follows immediately from Theorem 2.5 and Chebyshev's inequality.

Corollary 2.6. Suppose $h \to 0$ and $nh^5 \to \infty$ as $n \to \infty$. Then, under Assumptions 2.1 and 2.2, $\widehat{C}_h(x) \xrightarrow{P} C(x)$ as $n \to \infty$.

Proof. By Chebyshev's inequality,

$$\begin{aligned} P\left(\left|\widehat{C}_h(x) - C(x)\right| > \varepsilon\right) &\leqslant \frac{\mathbb{E}\left[\left(\widehat{C}_h(x)-C(x)\right)^2\right]}{\varepsilon^2} \\ &= \frac{1}{\varepsilon^2}\left(\frac{1}{4}h^2\alpha^2(x) + \frac{1}{nh^5}\beta(x)\right) + o(h^4) + O\left(\frac{1}{nh^5}\right) + O\left(\frac{1}{n^{3/2}h^7}\right) \\ &\to 0 \end{aligned}$$

as $n \to \infty$. ∎

As yet another corollary, we obtain the expression of the AMISE of the curvature estimator and the corresponding optimal bandwidth.

Corollary 2.7. (1) The asymptotic mean squared integrated error of the curvature estimator (A.6) when $h_1 = h_2 = h$ equals

$$AMISE\left(\widehat{C}_h\right) = \frac{1}{4}h^4 R(\alpha) + \frac{1}{nh^5}\int_{-\infty}^{\infty}\beta(x)\,dx;$$

(2) The asymptotic optimal bandwidth equals

$$h_{\text{AMISE}} = \left(5\frac{\int_{-\infty}^{\infty}\beta(x)\,dx}{R(\alpha)}\right)^{\frac{1}{9}} n^{-\frac{1}{9}}; \tag{2.4}$$

(3) AMISE at h_{opt} is

$$AMISE\left(\widehat{C}_{h_{\text{opt}}}\right) = \frac{9}{4}\left(\left(\frac{R(\alpha)}{5}\right)^5\left(\int_{-\infty}^{\infty}\beta(x)\,dx\right)^4\right)^{\frac{1}{9}} n^{-\frac{4}{9}}.$$

Proof. While (1) is trivial, (2) and (3) are proved by routine calculations. ■

One can observe that if $f(\cdot)$ does not have sharp features, then h_{AMISE} is close to the AMISE-optimal bandwidth for the KDE $\hat{f}''_{h_1}(\cdot)$ for the second order derivative $f''(\cdot)$ (see Henderson and Parmeter, 2015, p. 48) given by

$$h_{\text{AMISE}} = \left(5\frac{R(K'')}{R(f^{(4)})\kappa_2^2(K)}\right)^{\frac{1}{9}} n^{-\frac{1}{9}}.$$

Remark. It is worth noting that if $\hat{f}'_{h_1}(\cdot)$ and $\hat{f}''_{h_1}(\cdot)$ are consistent estimators of $f'(\cdot)$ and $f''(\cdot)$, respectively, then $\hat{C}_{h_1,h_2}(\cdot)$, as a continuous function of $\hat{f}'_{h_1}(\cdot)$ and $\hat{f}''_{h_1}(\cdot)$, will be a consistent estimator of $C(\cdot)$.

2.3.3 The Number of Modal and Purely Non-Modal Bumps for KDE with a Gaussian Kernel

In Silverman (1980), it has been shown that the number of maxima as x varies in $\hat{f}_h^{(r)}(x)$ in (2.2) for a Gaussian $K(\cdot)$ is a right continuous decreasing function of h for each $r \in \mathbb{N}_0$. The results imply that the number of maxima for $\hat{f}_h(\cdot)$ and $\hat{f}''_h(\cdot)$ is at most n and $2n$, respectively.

Below we show that similar results can be obtained for the number of minima. The proofs rely on some auxiliary lemmas which are relegated to Section A.1.3 in Appendix A.

Instrumental to our discussion is the following definition.

Definition 2.8. Let $g : \mathbb{R} \to \mathbb{R}$ be a continuous function such that $\lim_{x\to\pm\infty} g(x) = 0$ and is monotonic in a neighbourhood of both $+\infty$ and $-\infty$. We say that g is a function of type $(\nearrow \cdots \searrow)$ if it increases in a neighbourhood of $-\infty$ and decreases in a neighbourhood of $+\infty$. Functions of types $(\searrow \cdots \searrow)$, $(\nearrow \cdots \nearrow)$, and $(\searrow \cdots \nearrow)$ are defined analogously.

From Lemmas A.16 and A.17 with the help of Lemma A.18, we conclude that for functions of type $(\nearrow \cdots \searrow)$ the number of local minima Min equals the number of maxima Max minus one. Similarly, for functions of type $(\searrow \cdots \nearrow)$, $Min = Max + 1$. But, for functions of types $(\searrow \cdots \searrow)$ and $(\nearrow \cdots \nearrow)$, $Min = Max$.

Theorem 2.9. For $\hat{f}_h^{(r)}(\cdot)$ from (2.2) with a Gaussian $K(\cdot)$, the number of minima, Min, possesses the following properties:

1. It is a right continuous decreasing function of h;
2. It equals the number of maxima, Max, of $\hat{f}_h^{(r)}(\cdot)$ or equals $Max - 1$ according to whether r is odd or even.

Proof. First, let us show that $\hat{f}_h(x) = (\sum_{i=1}^n K((x - X_i)/h))/(nh)$ in (2.1) and each derivative $\hat{f}_h^{(r)}(x)$ in (2.2) for a Gaussian kernel function $K(\cdot)$ do not have flat parts. Let us assume that, on the contrary, there exists a function $\ell(x) = ax + b$ and interval (c,d) such that $\ell(x) \equiv \hat{f}_h^{(r)}(x)$ for all $x \in (c,d)$. Then there exists a polynomial $P(x)$ such that $P(x) \equiv \hat{f}_h(x)$ for all $x \in (c,d)$. Since both functions $\hat{f}_h$ and P are real analytic, then $P \equiv \hat{f}_h$ for all x (see Krantz and Parks, 2002, p. 14).

Next, as $\lim_{x\to\pm\infty} \hat{f}_h(x) = 0$, then $P(x) \equiv 0$. However, $\hat{f}_h \not\equiv 0$ (e.g., due to $\int_{-\infty}^{+\infty} \hat{f}_h(x)dx = 1$). This contradiction establishes the absence of flat parts in (2.1) and (2.2).

From the fact that (2.1) is a finite sum it follows that if K is any of the types of Definition 2.8, then the function $\hat{f}$ has the same type. But for the Gaussian kernel function, $K^{(r)}$ has type $(\nearrow \cdots \searrow)$ if r is even and type $(\nearrow \cdots \nearrow)$ if r is odd. Thus, for odd r, the number of minima, Min, of $\hat{f}_h^{(r)}$ is the same as Max and equals

$Max - 1$ for even r. The fact that the number of minima of $\hat{f}_h^{(r)}(x)$, as x varies, is a right continuous decreasing function of h follows from the main theorem in Silverman (1980, p. 97). ∎

Corollary 2.10. The number of minima for $\hat{f}_h''(\cdot)$ is at most $2n-1$.

Corollary 2.11. For $\hat{f}_h(\cdot)$ with a Gaussian kernel, the total number of modal bumps and purely non-modal (i.e., excluding shoulders) bumps implied by $\hat{f}_h''(\cdot)$ (i.e., if $f''(\cdot)$ is taken as a measure of curvature) is at most $3n-1$.

We conclude this section by proposing the following conjecture which we leave for future work.

Conjecture 2.12. The number of minima of curvature estimator (2.3) is at most $2n-1$.

Corollary 2.13. The total number of modal and purely non-modal (i.e., excluding shoulders) bumps for $\hat{f}_h(\cdot)$ with a Gaussian kernel is at most $3n-1$.

2.3.4 Definition of a Surgery

For subsequent developments, we need to introduce some definitions.

Definition 2.14. If a function $g : \mathbb{R} \to \mathbb{R}$ has a bump in (c_1, c_2), such that the graph of $g(\cdot)$ lies entirely above the chord connecting $(c_1, g(c_1))$ and $(c_2, g(c_2))$, i.e., $g(x) > ((g(c_2) - g(c_1))/(c_2 - c_1))(x - c_1) + g(c_1)$ for all $x \in (c_1, c_2)$, then c_1 and c_2 will be said to be *aligned*.

Definition 2.15. If a function $g : \mathbb{R} \to \mathbb{R}$ has a bump in (c_1, c_2), where c_1 and c_2 are aligned, then the part confined between the chord and the graph of $f(\cdot)$ will be referred to as the *cap* of the bump.

Definition 2.16 (Definition of a Surgery)**.** The process of determination and removal of caps whose area exceeds a certain fixed threshold $\lambda \in [0, 1)$ will be referred to as a λ*-surgery*.

Definition 2.16 implies that a λ-surgery is also a λ_1-surgery for any $\lambda_1 \geqslant \lambda$. A λ-surgery produces a partition $\sqcup_{j=1}^{m}(a_j, b_j) \subseteq \operatorname{supp} g$ into m (for $m \in \mathbb{N}$) disjoint intervals (a_j, b_j) such that each pair $\langle a_j, b_j \rangle$ is aligned. The support line $S = S(x)$

comprises of arcs of the graph of $g(\cdot)$ and chords connecting $\langle a_j, b_j \rangle$ lying no higher than the graph of $g(\cdot)$. Since the partition is disjoint, if for a given $x \in \mathbb{R}$, there exists an index $j_0 \in \overline{1,m}$, such that $x \in \left(a_{j_0}, b_{j_0}\right)$, it is unique and $S(x) = ((g(b_{j_0}) - g(a_{j_0}))/(b_{j_0} - a_{j_0}))(x - a_{j_0}) + g(a_{j_0})$. However, if no such j_0 exists, then $S(x) = g(x)$. Definition 2.16 implies that the area between the axis of the abscissas and the support line is at most $1 - m\lambda$.

If each $\langle a_j, b_j \rangle$ is a pair of consecutive inflection points, then this pair satisfies Definition 2.14 while the partition $\sqcup_{j=1}^{m}(a_j, b_j)$ trivially satisfies Definition 2.16 for $\lambda = 0$. However, Definition 2.14 may be satisfied by other points too. See Figure 2.4 where the result of a 0.01-surgery is illustrated for $g(x) = \hat{f}_h(x)$ from (2.1) with the standard Gaussian kernel function $K(\cdot)$ for a random sample of size $n = 1{,}000$ drawn from the mixture PDF $f(x) = 0.1N(0, 0.4) + 0.9N(1.3, 1.8)$.

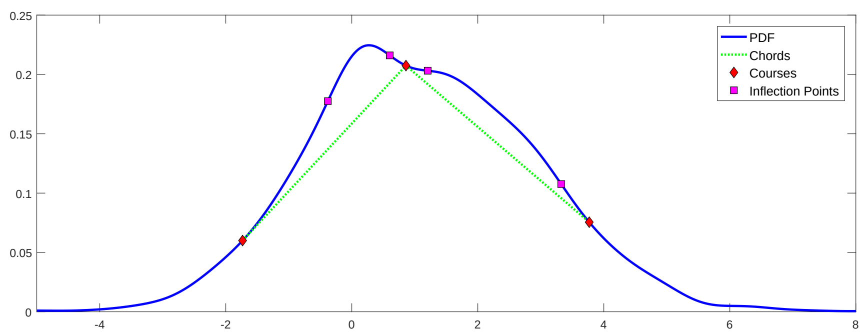

Figure 2.4: A 0.01-surgery for $g(x) = \hat{f}_h(x)$ from (2.1) with the standard Gaussian kernel function $K(\cdot)$ for a random sample of size $1{,}000$ drawn from the mixture PDF $f(x) = 0.1N(0, 0.4) + 0.9N(1.3, 1.8)$. The areas of the left and right caps equal 0.07 and 0.05, respectively. Chords, inflection points and courses are all shown.

2.3.5 Algorithm for Bump Hunting

We are now in a position to present the method for bump-hunting. The main steps of TBH are presented in Algorithm 2.1, where Step 26 performs a λ-surgery.

Algorithm 2.1 Main Steps of Bump Hunting

Input: Data set $\{X_i : i = 1, \ldots, n\}$ of size n, threshold λ

Output: Number m, locations of bumps $b_1, \ldots, b_m$ and values of caps' areas $s_1, \ldots, s_m$ no smaller than λ

1: Fix a kernel function $K(\cdot)$
2: Compute bandwidth h for KDE (2.1) of $f(\cdot)$ with kernel $K(\cdot)$
3: Compute points of maxima and minima of $\hat{f}_h(\cdot)$
4: **if** $UseCurvature = 1$ **and** $UseSecondDerivative = 0$ **then** ▷ Use curvature
5: Compute bandwidths h_1 and h_2 for KDE (2.3) of $C(\cdot)$ with kernel $K(\cdot)$
6: Compute points of maxima and minima $\hat{C}_{h_1,h_2}(\cdot)$
7: Remove points in Step 6 resulting in negative maxima and positive minima of $\hat{C}_{h_1,h_2}(\cdot)$
8: Merge points of maxima of $\hat{f}_h(\cdot)$ with points of minima of $\hat{C}_{h_1,h_2}(\cdot)$
9: Merge points of minima of $\hat{f}_h(\cdot)$ with points of maxima of $\hat{C}_{h_1,h_2}(\cdot)$
10: Retain points from Step 9 where each pair consecutive values contains a point from Step 8 between them. Refer to them collectively as "troughs"
11: Retain troughs with lowest values of $\hat{f}_h(\cdot)$
12: Retain points from Step 8 which fall between a pair of troughs with highest values of $\hat{f}_h(\cdot)$. Refer to them collectively as "peaks"
13: **else if** $UseCurvature = 0$ **and** $UseSecondDerivative = 1$ **then** ▷ Use second derivative
14: Compute bandwidth h_2 for KDE (2.2) of $f''(\cdot)$ with kernel $K(\cdot)$
15: Compute points of maxima and minima $\hat{f}''_{h_2}(\cdot)$
16: Remove points in Step 15 resulting in negative maxima and positive minima of $\hat{f}''_{h_2}(\cdot)$
17: Merge points of maxima of $\hat{f}_h(\cdot)$ with points of minima of $\hat{f}''_{h_2}(\cdot)$
18: Retain points from Step 21 where each pair consecutive values contains a point from Step 17 between them. Refer to them collectively as "troughs"
19: Retain troughs with lowest values of $\hat{f}_h(\cdot)$
20: Retain points from Step 17 which fall between a pair of troughs with highest values of $\hat{f}_h(\cdot)$. Refer to them collectively as "peaks"
21: Merge points of minima of $\hat{f}_h(\cdot)$ with points of maxima of $\hat{f}''_{h_2}(\cdot)$
22: **end if**

Algorithm 2.1 Main Steps of Bump Hunting (continued)

23: Ensure that a trough-peak-trough pattern is obeyed
24: Ensure that each pair of consecutive troughs is aligned
25: Compute areas of the caps of bumps between consecutive troughs from Step 24
26: Run λ-surgery 2.2 in attempt to merge bumps whose cap's area is smaller than λ with adjacent bumps
27: Return the number m and locations of bumps $b_1,\dots,b_m$ with cap's areas $s_1,\dots,s_m$ no smaller than λ

This Algorithm relies on an auxiliary procedure given by Algorithm 2.2.

Algorithm 2.2 λ-Surgery

Input: M *peaks*, $2M$ *troughs*, $(X,Y)=\{\langle x_k,\hat{f}_h(x_k)\rangle\}_{k=1,\dots,N}$, area threshold λ
Output: *peaks*, *troughs*, *areas*, *supportLine*

$supportLine \leftarrow [\quad]$, $troughsTemp \leftarrow troughs$, $areas \leftarrow [\quad]$
$i \leftarrow 0$, $C \leftarrow 1$, $Flag \leftarrow 0$
while Flag=0 **do**
  if $C > 2M$ **or** $C+i > 2M$ **then** $Flag \leftarrow 1$
  else
    $supportLineTemp \leftarrow [\quad]$, $L \leftarrow troughsTemp(C)$, $R \leftarrow troughsTemp(C+i)$
    Connect L and R by a chord
    if The graph of (X,Y) between L and R is no lower than the chord **then**
      Measure the area S of this bump's cap
      Append *supportLineTemp* by L and R
      if $S \geqslant \lambda$ **then**
        Append *supportLine* by *supportLineTemp*, append *areas* by S
        Append *peaks* by the highest *peak* between L and R
        $troughsTemp(C) \leftarrow L$, $troughsTemp(C+i) \leftarrow R$, $C \leftarrow C+i+1$, $i \leftarrow 0$
      else
        $i \leftarrow i+1$, $troughsTemp(C) \leftarrow L$ ▷ Keep the left trough
      end if
    else
      Check if there is a screening trough between L and R
      for $j \leftarrow i$ **by** -1 **to** 1 **do**
        if troughs(C+j) lies lower than the chord between L and R **then**
          $L \leftarrow troughs(C+j)$, $C \leftarrow C+j$ ▷ Switch to the rightmost screening bump
          $i \leftarrow 0$

Algorithm 2.2 λ-Surgery (continued)

```
                break
            end if
        end for
        Connect L and R by a chord
        if The graph of (X,Y) between L and R is no lower than the chord then
            Measure the area S of this bump's cap
            Append supportLineTemp by L and R
        else
            Compute the GCM between L and R
            Append supportLineTemp by the values of the GCM
            Measure the area S confined between the graph (X,Y) and the GCM
        end if
        if S ⩾ λ then
            Append supportLine by supportLineTemp, append areas by S
            Append peaks by the highest peak between L and R
            troughsTemp(C) ← L, troughsTemp(C+i) ← R
            C ← C+i+1, i ← 0
        else
            i ← i+1, supportLineTemp ← [ ], troughsTemp(C) ← L    ▷ Keep the left trough
        end if
    end if
  end if
end while
```

To complete Step 24 in Algorithm 2.1, the alignment of any two base points $\langle c_1, c_2\rangle$ of a bump is accomplished by fitting the isotonic regression[4] (Robertson et al., 1988, pp. 4-11) computed by the pool-adjacent-violators algorithm on raw slopes of those KDE points $(x_j, \hat{f}_h(x_j))$, $x_1 \leqslant \cdots \leqslant x_M$, $j \in \overline{1,M}$, for which $c_1 \leqslant x_j \leqslant c_2$. Figure 2.5 shows a PDF where a bump is selected with its two courses being not aligned. Figure 2.6 shows the result of alignment. It is known (Robertson et al., 1988, p. 7) that graphically the (unique) solution to the isotonic regression will coincide with the greatest convex minorant (GCM). If the original function is piecewise linear, the GCM will be such as well.

[4] We have also developed an alternative method of alignment by iterative rotations of the arc around its left end. This led to results identical to those from the isotonic regression.

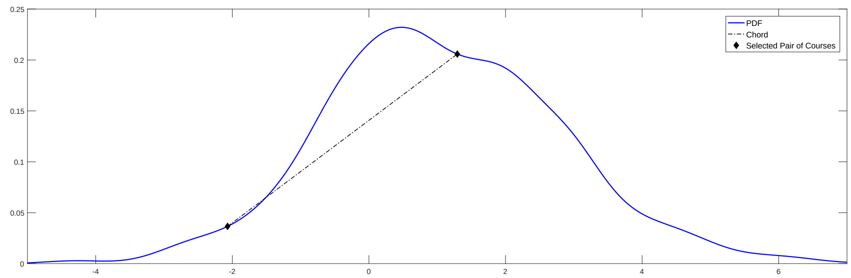

Figure 2.5: A bump for $g(x) = \hat{f}_h(x)$ from (2.1) with the standard Gaussian kernel function $K(\cdot)$ for a random sample of size $1{,}000$ drawn from the mixture PDF $f(x) = 0.1N(0, 0.4) + 0.9N(1.3, 1.8)$. The courses corresponding to this bump are not aligned.

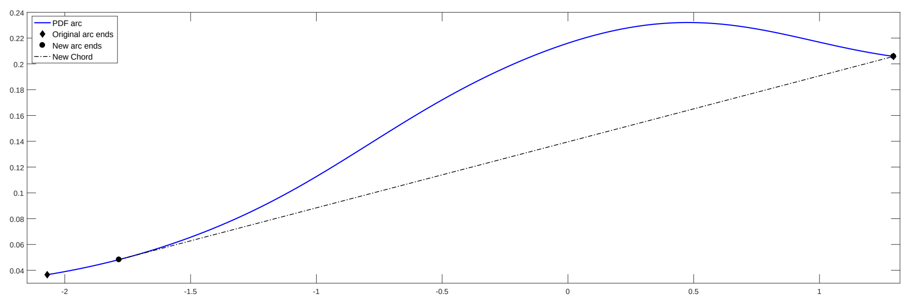

Figure 2.6: The demonstration of alignment of the two base points of the bump from Figure 2.5. The figure displays the arc together with the original and aligned base points.

Figure 2.7 demonstrates the output of the bump-hunting procedure.

2.4 Simulation Study

2.4.1 Preliminary Setup

We have assessed the performance of TBH by conducting extensive Monte Carlo simulations on a variety of distributions under various settings. The method has been implemented in MATLAB®[5] software (MATLAB, 2016a). TBH has been compared with the Hartigans' DIP method (Hartigan and Hartigan, 1985) and the nonparametric clustering method ("NCM") from Azzalini and Torelli (2007) and

[5]MATLAB is a registered trademark of The MathWorks, Inc. For more information, see http://www.mathworks.com.

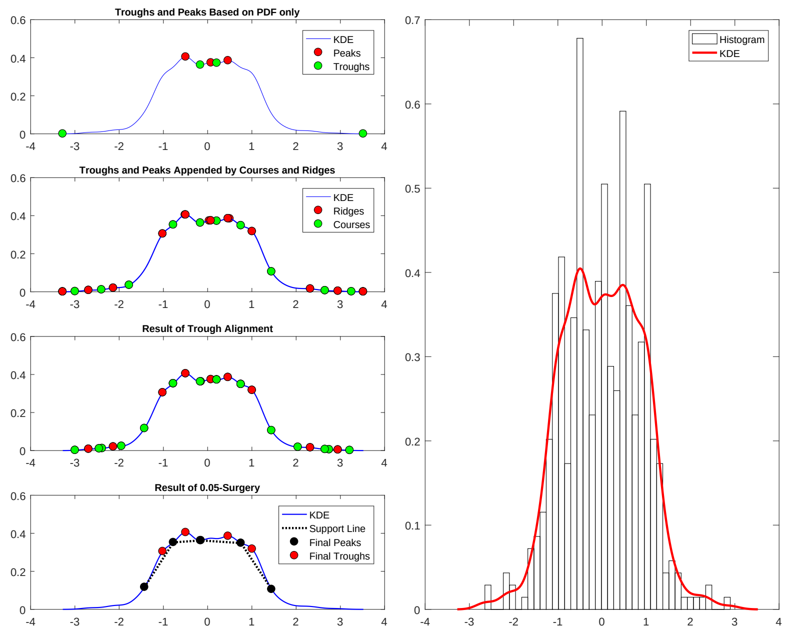

Figure 2.7: A 0.05-surgery for a KDE with the standard Gaussian kernel function $K(\cdot)$ on a random sample of size 500 drawn from distribution number 10 of Marron and Wand (1992) (the plot of its PDF is presented in Figure 2.1).

Menardi and Azzalini (2014) whose implementations are available as R (R Core Team, 2016) (version 3.3.0) packages in Maechler (2015) and Azzalini and Menardi (2014), respectively. To link MATLAB and R, an R package by Bengtsson (2016) was employed. Computations were facilitated by the the use of parallel processing (Calaway et al., 2015) in R and (MATLAB, 2016b) in MATLAB. All computations were performed on a 12-core HP Z800 workstation with the Xeon 2.93 GHz processor and 48 GB DDR3 of random-access memory (RAM).

We have selected 22 distributions for testing purposes. Among those, 15 are taken from Marron and Wand (1992, p. 720); see Table 2.1 where the corresponding PDF expressions are listed. The PDFs for distributions 16-22, are given in Table 2.2.

The plots of all the 22 PDFs are given in Figures 2.8-2.9. The distributions are Gaussian or LogNormal mixtures with and without sharp features. The distributions also aim to imitate regular and asymmetric, kurtotic and long-tailed distributional shapes.

We have selected several combinations of bandwidths which are all listed in

Table 2.1: Parameter values for the PDFs of distributions 1-15 taken from Marron and Wand (1992). Adapted with permission from the Institute of Mathematical Statistics.

No	$\sum_{i=1}^{m} N(\mu_i, \sigma_i)$
1	$N(0,1)$
2	$\frac{1}{5}N(0,1)+\frac{1}{5}N\left(\frac{1}{2},\frac{2}{3}\right)+\frac{3}{5}N\left(\frac{13}{12},\frac{5}{9}\right)$
3	$\sum_{i=0}^{7}\frac{1}{8}N\left(3\left(\left(\frac{2}{3}\right)^i-1\right),\left(\frac{2}{3}\right)^i\right)$
4	$\frac{2}{3}N(0,1)+\frac{1}{3}N\left(0,\frac{1}{10}\right)$
5	$\frac{1}{10}N(0,1)+\frac{9}{10}N\left(0,\frac{1}{10}\right)$
6	$\frac{1}{2}N\left(-1,\frac{2}{3}\right)+\frac{1}{2}N\left(1,\frac{2}{3}\right)$
7	$\frac{1}{2}N\left(-\frac{3}{2},\frac{1}{2}\right)+\frac{1}{2}N\left(\frac{3}{2},\frac{1}{2}\right)$
8	$\frac{3}{4}N(0,1)+\frac{1}{4}N\left(\frac{3}{2},\frac{1}{3}\right)$
9	$\frac{9}{20}N\left(-\frac{6}{5},\frac{3}{5}\right)+\frac{9}{20}N\left(\frac{6}{5},\frac{3}{5}\right)+\frac{1}{10}N\left(0,\frac{1}{4}\right)$
10	$\frac{1}{2}N(0,1)+\sum_{i=0}^{4}\frac{1}{10}N\left(\frac{i}{2}-1,\frac{1}{10}\right)$
11	$\frac{49}{100}N\left(-1,\frac{2}{3}\right)+\frac{49}{100}N\left(1,\frac{2}{3}\right)+\sum_{i=0}^{6}\frac{1}{350}N\left(\frac{i-3}{2},\frac{1}{100}\right)$
12	$\frac{1}{2}N(0,1)+\sum_{i=-2}^{2}\frac{2^{1-i}}{31}N\left(i+\frac{1}{2},\frac{2^{-i}}{10}\right)$
13	$\sum_{i=0}^{1}\frac{46}{100}N\left(2i-1,\frac{2}{3}\right)+\sum_{i=1}^{3}\frac{1}{300}N\left(-\frac{i}{2},\frac{1}{100}\right)+\sum_{i=1}^{3}\frac{7}{300}N\left(\frac{i}{2},\frac{7}{100}\right)$
14	$\sum_{i=0}^{5}\frac{2^{5-i}}{63}N\left(\frac{1}{21}\left(65-96\left(\frac{1}{2}\right)^i\right),\frac{32}{63}\left(\frac{1}{2}\right)^i\right)$
15	$\sum_{i=0}^{2}\frac{2}{7}N\left(\frac{1}{7}(12i-15),\frac{2}{7}\right)+\sum_{i=8}^{10}\frac{1}{21}N\left(\frac{2i}{7},\frac{1}{21}\right)$

Table 2.2: Parameter values for the PDFs of distributions 16-22. For No 16-20 the PDF is $f(x) = pN(x;\mu_1,\sigma_1) + (1-p)N(x;\mu_2,\sigma_2)$, and for No 21-22 $f(x) = pLogn(x;\mu_1,\sigma_1) + (1-p)Logn(x;\mu_2,\sigma_2)$, where $Logn(x;\mu_1,\sigma_1) = N(\ln x;\mu_1,\sigma_1)/x$. Mixtures 16 and 17 are from Cheng and Hall (1999): Used with permission from the Institute of Mathematical Statistics.

No	Parameters			Description
	p	$\mu = \langle\mu_1,\mu_2\rangle$	$\sigma = \langle\sigma_1,\sigma_2\rangle$	
16	$8e^{9/8}/(1+8e^{9/8})$	$\langle 0.0,-9\sqrt{3}/8\rangle$	$\langle 1,0.25\rangle$	Mixture with a mode and a shoulder
17	100/109	$\langle 0.0,1.3\rangle$	$\langle 1.0,0.3\rangle$	Mixture with a mode and a shoulder
18	0.7	$\langle 0.0,-1.0\rangle$	$\langle 1.0,0.4\rangle$	Mixture with two non-modal bumps
19	0.1	$\langle 0.0,1.3\rangle$	$\langle 0.4,1.8\rangle$	Mixture with two non-modal bumps
20	0.9	$\langle 3.0,20.0\rangle$	$\langle 1.0,3.0\rangle$	Mixture with two widely separated modal bumps
21	1.0	$\langle 0.0,0.0\rangle$	$\langle 1.0,1.0\rangle$	Logn(0,1) distribution
22	0.7	$\langle 0.0,3.5\rangle$	$\langle 1.0,0.2\rangle$	Mixture with two widely separated modal bumps

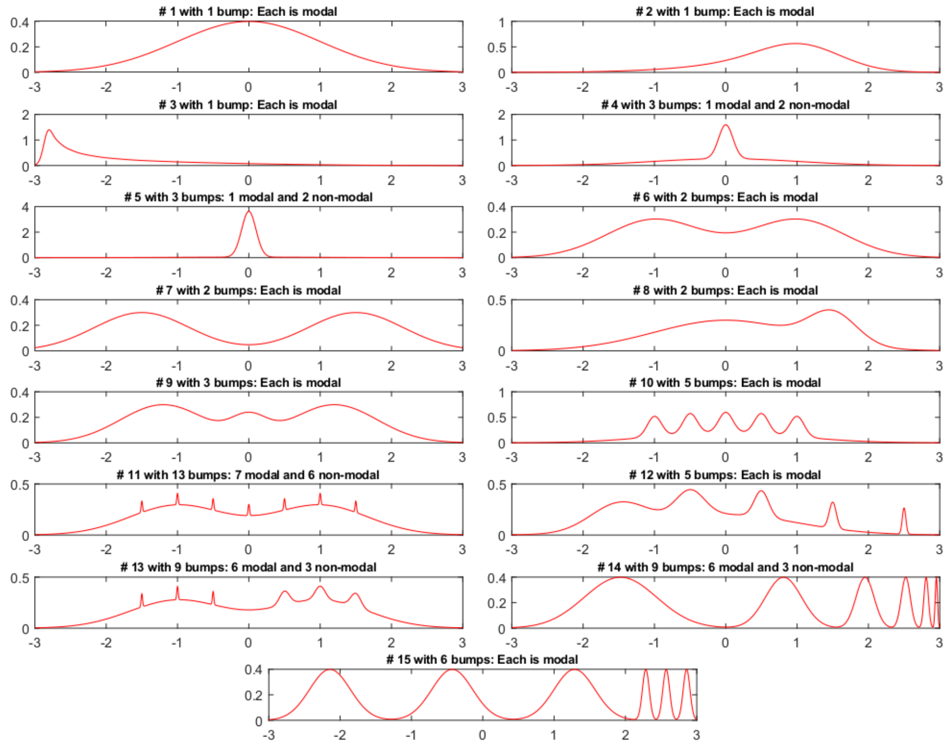

Figure 2.8: Mixtures of Gaussian distributions 1-15 from Marron and Wand (1992). Adapted with permission from the Institute of Mathematical Statistics.

Table 2.3. Among them are the Sheather-Jones direct-plugin bandwidths for the KDEs of f, f' and f'' implemented in Duong (2016), the Sheather-Jones solve-the-equation bandwidth implemented in Raykar et al. (2010), the 3/4 of Silverman's rule of thumb suggested in Menardi and Azzalini (2014), and the rule-of-thumb bandwidth obtained for curvature by formula (2.4). Because no expression is available in a closed analytical form for the Gaussian distribution based on (2.4), we have numerically computed the latter bandwidth for $N(0,\sigma_i)$ in the Maple™ software[6] (MAPLE, 2016) on a fine grid of σ_i between 0.01 and 100 and applied linear spline interpolation and extrapolation to determine the bandwidth for standard deviations outside this range.

For bandwidth combinations 1 and 7 in Table 2.3, curvature information is not used and the search is for modal bumps only. On the other hand, for combinations

[6] Maple is a trademark of Waterloo Maple Inc. For more information, see http://www.maplesoft.com.

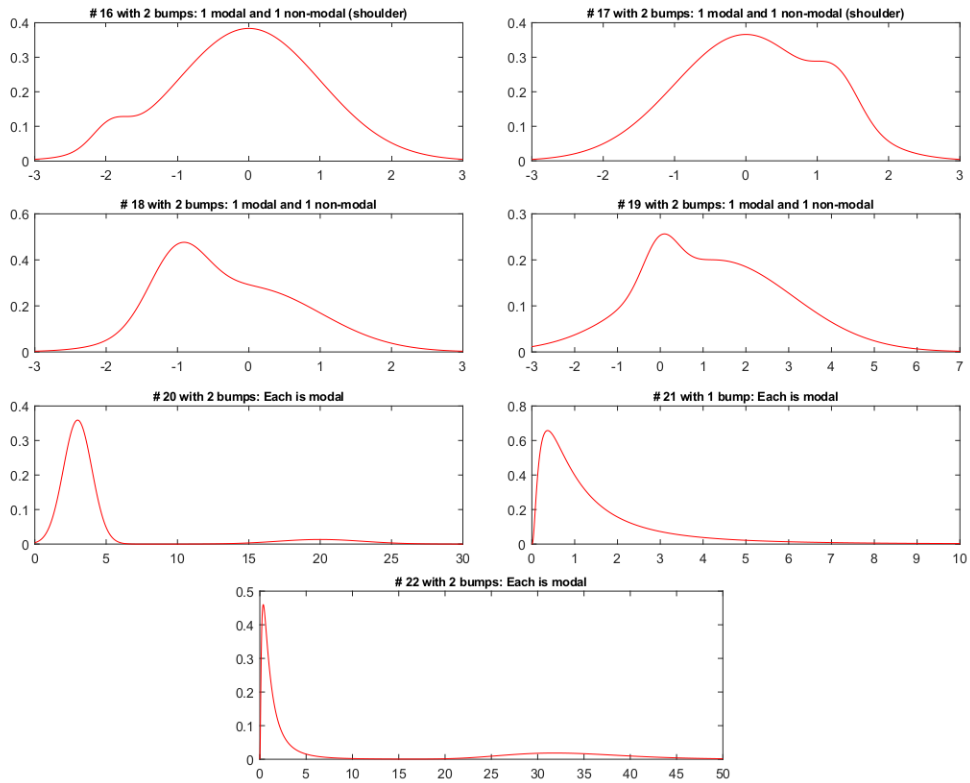

Figure 2.9: Mixture of Gaussian distributions. Distributions 16-17 are from Cheng and Hall (1999). Adapted with permission from the Institute of Mathematical Statistics.

3 and 5, $\hat{f}_h''(\cdot)$ plays the role of curvature. Where both $h_{\hat{f}'}$ and $h_{\hat{f}''}$ are present, they are used as, respectively, h_2 and h_1 in (2.3).

Remark. It is important to keep in mind that, as stressed in Jones et al. (1996, p. 403), "when using the optimal bandwidth for estimation of f, the estimate $\hat{f}''$ is asymptotically inconsistent for f'' and $R(\hat{f}'')$ is only barely consistent for $R(f'')$, with much better performance available from better bandwidths". The same caveat applies to the estimate $\hat{f}'$. However, we have observed that the performance of TBH is satisfactory when h, an optimal bandwidth of $\hat{f}_h$, is fed into $\hat{C}_h$ which is not a standalone object, but is a function of $\hat{f}_h'$ and $\hat{f}_h''$.

As kurtotic, asymmetric and long-tailed distributions have been known to pose significant problems to kernel density estimation (Fisher et al., 1994, p. 510) due to the inducement of outliers, a number of techniques have been suggested that

Table 2.3: Combinations of bandwidths for $\hat{f}_h(\cdot)$, $\hat{f}'_h(\cdot)$, $\hat{f}''_h(\cdot)$ used for testing purposes. Here, SJSTE0 is the Sheather-Jones solve-the-equation bandwidth for the KDE of f. SJDPI0, SJDPI1 and SJDPI2 stand for the Sheather-Jones direct-plugin bandwidths for the KDEs of f, f' and f'', respectively. A0 denotes the 3/4 of Silverman's rule of thumb for the KDE of f. Finally, ACURV stands for the rule-of-thumb bandwidth for curvature.

Bandwidth	Combination								
	1	2	3	4	5	6	7	8	9
$h_{\hat{f}}$	SJSTE0	SJSTE0	SJSTE0	SJDPI0	SJDPI0	SJSTE0	A0	SJSTE0	A0
$h_{\hat{f}'}$		SJSTE0		SJDPI1		SJDPI1		ACURV	ACURV
$h_{\hat{f}''}$		SJSTE0	SJDPI2	SJDPI2	SJDPI2	SJDPI2		ACURV	ACURV

attempt to deal with this issue. For example, one may apply an outlier removal method. Alternatively, a suitable transformation may be performed on the data, such as ln (on positively supported data), arctan (Markovitch and Krieger, 2000), Champernowne (Buch-Larsen et al., 2005), or a transform from Johnson's family (Yang and Marron, 1999). A systematic treatise on transformations in the context of nonparametric modeling can be found in Tarter and Lock (1993).

We have chosen to apply an efficient outlier rejection rule called X84 (Hampel et al., 1986, p. 68) which, given a data set $X = \{x_1, \ldots, x_n\}$, removes the observations with the property $|x_i - med| > 5.2MAD$, where med is the median of X and $MAD = median\{|x_i - med|\}$ is the median absolute deviation. However, in many applications (especially, risk modeling), it is not uncommon that observations beyond this cut-off $5.2MAD$ from the median do describe a true phenomenon (see Pachamanova and Fabozzi, 2010, p. 288). Distributions 20 and 22, with two widely separated modal bumps, fall into this class as the rightmost bump is located in the tail region affected by X84. Where the relative frequency of observations located beyond $5.2MAD$ from the median exceeds a chosen area threshold λ, automated application of X84 is not recommended without a prior inspection of the data set in question. Out of the 22 distributions, 8 distributions may be affected by X84: 2-5, 21 (due to asymmetry and/or peakedness) and 20 and 22 (due to the presence of remote modal bumps). Therefore, we have run the simulations for these 8 distributions without invoking the X84 outlier removal rule.

Two competing approaches to TBH are evaluated: DIP and NCM. As the DIP test is a statistical test of the null hypothesis $H_0 : m = 1$ vs $H_1 : m > 1$, where m is the number of modes of the underlying distribution, to assess H_0, we have chosen a value of 0.05 as the significance level. NCM is a clustering procedure based on a Voronoi diagram, Delaunay triangulation, modal trees and KDE of the unknown PDF for underlying data. We preserved all default parameter values for NCM, in particular, the bandwidth equal to 3/4 of the Silverman rule of thumb. For all the distributions, in TBH, we have fixed a common value of area threshold $\lambda = 0.05$. We have tested other values for λ, but $\lambda = 0.05$ proved to be the most stable performer across all of the 22 distributions considered. Table 2.4 shows theoretical values of bumps' cap areas together with the number of them exceeding $\lambda = 0.05$.

Table 2.4: Theoretical values of the areas of caps corresponding to bumps for the selected 22 distributions.

No	# Mixture Components	# Modes	Total # Bumps	Values of Areas of Bump Caps													# Areas Exceeding 0.05
1	1	1	1	0.6188													1
2	3	1	1	0.4767													1
3	8	1	1	0.9121													1
4	2	1	3	0.0034	0.3249	0.0034											1
5	2	1	3	0.0003	0.8899	0.0003											1
6	2	2	2	0.1994	0.2007												2
7	2	2	2	0.3317	0.3317												2
8	2	2	2	0.1621	0.1675												2
9	3	3	3	0.1771	0.0276	0.1771											2
10	6	5	5	0.0891	0.0838	0.0845	0.0838	0.0891									5
11	9	7	13	0.0000	0.0028	0.0031	0.0028	0.0025	0.0028	0.0028	0.0028	0.0025	0.0028	0.0031	0.0028	0.0000	0
12	6	5	5	0.1070	0.0980	0.0713	0.0316	0.0157									3
13	8	6	9	0.0000	0.0033	0.0029	0.0032	0.0023	0.0033	0.0223	0.0318	0.0233					0
14	6	6	6	0.4925	0.2415	0.1207	0.0604	0.0302	0.0154								4
15	6	6	6	0.2325	0.2705	0.2722	0.0457	0.0451	0.0449								3
16	2	1	2	0.0150	0.5082												1
17	2	1	2	0.2177	0.0401												1
18	2	1	2	0.1942	0.0217												1
19	2	1	2	0.0669	0.0615												2
20	2	2	2	0.6688	0.0736												2
21	1	1	1	0.3220													1
22	2	2	2	0.6892	0.1962												2

We have selected a grid of 13 sample sizes ranging from 30 to 2,000. For each sample size n from the grid, 1,000 samples $\{X_i\}_{i=1,\dots,n}$ were generated from each of the 22 distributions and the three methods were then applied to them. To compare the performance of TBH to those of NCM and DIP, we calculated the following metrics on the number of detected bumps/modes: mean, median, standard deviation, median absolute deviation and mean absolute deviation from the theoretical number of modes. The percentage of pairwise-identical decisions made by TBH and NCM, TBH and DIP, and NCM and DIP on each generated data set were tabulated. The DIP test was compared to TBH and NCM as follows: If the number of bumps (clusters) identified by TBH (NCM) was equal to 1, we regarded this as a manifestation of unimodality, and multimodality otherwise. Finally, we also recorded the number of errors occurred during simulations for each test.

2.4.2 Discussion of Results

For the sake of brevity, we only present partial results[7] for the subset of 10 distributions: number 1, 3, 5, 7, 9, 10, 11, 17, 20 and 21.

Inspection of Tables 2.5 and 2.6, which provide the average number of bumps discovered by TBH with the use of X84, suggests that in majority of cases TBH is capable of identifying the correct number of bumps whose cap's areas exceed 0.05 (see Table 2.4) as sample size grows. Notable exceptions are distributions 11 and 20. For the latter one, the rightmost modal bump is removed by X84 and it therefore becomes unidentifiable by TBH.

As Table 2.7 shows, removal of X84 rectifies the situation: there are 2 bandwidth combinations which capture the 2 bumps. The two combinations correspond to the use of $A0$ for $\hat{f}_h$ with and without $ACURV$ for $\hat{C}_h$. For distribution 11, the method reports two bumps, on average, as the 7 tiny spikes out of the 13 bumps altogether have areas about 3.1% which is smaller than the set value of λ. Should the λ be decreased, TBH still would have difficulty in capturing the spikes as the size of the samples necessary to realize the spikes is so large that it falls beyond the sample sizes commonly used in practice (cf. Marron and Wand, 1992). TBH aligns with NCM for distribution 11 which also reports 2 bumps.

In terms of speed of convergence, as measured by the standard deviation (see Tables 2.8, 2.9 and 2.10), TBH is at least as good as NCM and in some cases outperforming it. For example, it performs much better than NCM for distributions 1 and 20 (when X84 is not employed). Also, TBH, when used with X84, has an apparent advantage over NCM for skewed and/or kurtotic distributions 3, 5 and 21. The number of bumps identified by NCM seems to diverge to infinity for distributions 5 and 21.

For these two distributions as well as for the "claw" distribution 10, NCM results in execution failures whose number rises with sample size. See Table 2.11. As documentation (Azzalini and Menardi, 2014) suggests, in such circumstances, default parameters in NCM need to be adjusted.

We have also compared the ability of TBH to discover the unimodal nature of a distribution to that of DIP and NCM. These results are reported in Tables 2.12, 2.13 and 2.14.

[7] The complete set of simulation results is available on author's website `http://www.kmayorov.ca/supp_materials/TBH/` as spreadsheets in the Microsoft® Excel format.

Table 2.5: The average number of identified bumps by TBH and NCM for distributions 1, 3, 5, 7, and 9. TBH was used with $\lambda = 0.05$ and the X84 rule.

ID	True Bumps	True Modes	Sample Size	TBH								NCM
				Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8	
1	1	1	30	1.19	1.22	1.26	1.17	1.18	1.25	1.25	1.18	1.29
1	1	1	50	1.13	1.18	1.17	1.12	1.13	1.14	1.18	1.14	1.30
1	1	1	70	1.11	1.15	1.13	1.11	1.10	1.11	1.18	1.09	1.30
1	1	1	100	1.08	1.12	1.13	1.08	1.07	1.08	1.17	1.06	1.30
1	1	1	150	1.07	1.11	1.09	1.07	1.06	1.07	1.16	1.06	1.29
1	1	1	200	1.04	1.09	1.08	1.04	1.05	1.05	1.11	1.05	1.27
1	1	1	250	1.04	1.06	1.08	1.04	1.03	1.04	1.11	1.06	1.27
1	1	1	300	1.04	1.06	1.08	1.04	1.02	1.03	1.08	1.03	1.27
1	1	1	400	1.03	1.07	1.05	1.03	1.01	1.03	1.10	1.04	1.25
1	1	1	500	1.03	1.05	1.06	1.03	1.02	1.02	1.08	1.02	1.24
1	1	1	750	1.02	1.04	1.03	1.02	1.01	1.01	1.06	1.02	1.25
1	1	1	1000	1.01	1.04	1.03	1.01	1.00	1.00	1.07	1.02	1.24
1	1	1	2000	1.01	1.01	1.01	1.01	1.00	1.00	1.04	1.01	1.22
3	1	1	30	1.59	1.50	1.55	1.36	1.36	1.57	1.17	1.56	1.37
3	1	1	50	1.57	1.54	1.54	1.36	1.34	1.62	1.09	1.65	1.55
3	1	1	70	1.61	1.49	1.53	1.34	1.33	1.59	1.06	1.59	1.62
3	1	1	100	1.57	1.49	1.47	1.31	1.34	1.58	1.06	1.54	1.67
3	1	1	150	1.53	1.41	1.41	1.26	1.27	1.52	1.02	1.51	1.74
3	1	1	200	1.40	1.33	1.34	1.19	1.22	1.48	1.01	1.42	1.77
3	1	1	250	1.34	1.29	1.27	1.16	1.18	1.42	1.01	1.36	1.80
3	1	1	300	1.31	1.23	1.22	1.12	1.14	1.35	1.01	1.29	1.76
3	1	1	400	1.23	1.14	1.17	1.08	1.11	1.32	1.00	1.22	1.78
3	1	1	500	1.16	1.12	1.10	1.05	1.06	1.24	1.00	1.19	1.77
3	1	1	750	1.09	1.04	1.04	1.02	1.04	1.16	1.00	1.08	1.76
3	1	1	1000	1.04	1.02	1.02	1.01	1.03	1.09	1.00	1.05	1.72
3	1	1	2000	1.00	1.00	1.00	1.00	1.00	1.02	1.00	1.01	1.58
5	3	1	30	1.01	1.01	1.02	1.07	1.02	1.01	1.01	1.01	1.03
5	3	1	50	1.01	1.01	1.02	1.03	1.01	1.00	1.00	1.00	1.12
5	3	1	70	1.00	1.01	1.01	1.02	1.00	1.00	1.00	1.01	1.31
5	3	1	100	1.00	1.01	1.01	1.03	1.00	1.00	1.00	1.00	1.57
5	3	1	150	1.00	1.01	1.02	1.02	1.00	1.00	1.00	1.00	2.08
5	3	1	200	1.00	1.01	1.02	1.01	1.00	1.00	1.00	1.00	2.43
5	3	1	250	1.00	1.01	1.01	1.01	1.00	1.00	1.00	1.01	2.78
5	3	1	300	1.00	1.01	1.01	1.02	1.00	1.00	1.00	1.00	3.10
5	3	1	400	1.00	1.01	1.01	1.01	1.00	1.00	1.00	1.00	3.62
5	3	1	500	1.00	1.01	1.01	1.01	1.00	1.00	1.00	1.00	3.98
5	3	1	750	1.00	1.00	1.01	1.00	1.00	1.00	1.00	1.00	4.94
5	3	1	1000	1.00	1.01	1.01	1.01	1.00	1.00	1.00	1.01	5.62
5	3	1	2000	1.00	1.00	1.01	1.02	1.00	1.00	1.00	1.00	7.34
7	2	2	30	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.01	2.00
7	2	2	50	2.01	2.01	2.01	2.00	2.00	2.00	2.00	2.01	2.00
7	2	2	70	2.01	2.01	2.00	2.00	2.00	2.00	2.00	2.01	2.00
7	2	2	100	2.01	2.01	2.01	2.00	2.00	2.00	2.00	2.00	2.00
7	2	2	150	2.00	2.01	2.01	2.00	2.00	2.00	2.00	2.00	2.00
7	2	2	200	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00
7	2	2	250	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00
7	2	2	300	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00
7	2	2	400	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00
7	2	2	500	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00
7	2	2	750	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00
7	2	2	1000	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00
7	2	2	2000	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00
9	3	3	30	1.85	1.92	1.92	1.89	1.89	1.90	1.96	1.87	1.95
9	3	3	50	1.92	1.97	1.97	1.95	1.93	1.94	1.98	1.93	2.02
9	3	3	70	1.94	1.99	1.98	1.96	1.96	1.94	1.98	1.95	2.06
9	3	3	100	1.97	2.01	1.99	1.96	1.97	1.98	1.98	1.97	2.15
9	3	3	150	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.18
9	3	3	200	2.00	2.01	2.01	2.00	2.00	2.00	2.00	2.00	2.29
9	3	3	250	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.31
9	3	3	300	2.01	2.00	2.01	2.00	2.00	2.00	2.00	2.01	2.40
9	3	3	400	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.49
9	3	3	500	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.57
9	3	3	750	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.70
9	3	3	1000	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.83
9	3	3	2000	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.97

Table 2.6: The average number of identified bumps by TBH and NCM for distributions 10, 11, 17, 20, and 21. TBH was used with $\lambda = 0.05$ and the X84 rule.

ID	True Bumps	True Modes	Sample Size	TBH								NCM
				Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8	
10	5	5	30	1.19	1.28	1.25	1.23	1.22	1.25	1.38	1.19	1.37
10	5	5	50	1.21	1.32	1.31	1.24	1.24	1.26	1.42	1.23	1.44
10	5	5	70	1.22	1.41	1.36	1.30	1.26	1.28	1.41	1.22	1.51
10	5	5	100	1.28	1.42	1.43	1.32	1.27	1.32	1.47	1.28	1.59
10	5	5	150	1.31	1.55	1.51	1.37	1.30	1.30	1.51	1.30	1.71
10	5	5	200	1.35	1.61	1.59	1.40	1.36	1.36	1.55	1.36	1.83
10	5	5	250	1.41	1.69	1.71	1.44	1.36	1.41	1.59	1.45	1.98
10	5	5	300	1.50	1.77	1.79	1.48	1.39	1.43	1.63	1.48	2.14
10	5	5	400	1.73	1.92	1.95	1.56	1.42	1.47	1.81	1.69	2.57
10	5	5	500	2.06	2.25	2.23	1.68	1.47	1.60	1.86	2.06	3.03
10	5	5	750	3.42	3.37	3.36	1.91	1.60	1.87	1.98	3.36	4.16
10	5	5	1000	4.40	4.29	4.31	2.00	1.75	2.16	2.00	4.35	4.84
10	5	5	2000	4.98	4.91	4.90	2.40	2.36	4.61	2.00	4.98	5.21
11	13	7	30	1.67	1.76	1.73	1.70	1.72	1.74	1.83	1.70	1.84
11	13	7	50	1.76	1.84	1.85	1.81	1.80	1.78	1.89	1.79	1.93
11	13	7	70	1.84	1.93	1.95	1.92	1.88	1.89	1.95	1.86	2.01
11	13	7	100	1.89	1.96	1.95	1.92	1.90	1.92	1.96	1.92	2.04
11	13	7	150	1.94	1.98	1.99	1.96	1.95	1.97	1.99	1.96	2.07
11	13	7	200	1.95	1.99	1.99	1.98	1.96	1.96	2.00	1.96	2.08
11	13	7	250	1.98	2.00	2.00	1.99	1.98	1.98	2.00	1.99	2.08
11	13	7	300	1.98	2.00	2.00	2.00	1.99	1.99	2.00	1.99	2.08
11	13	7	400	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.07
11	13	7	500	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.07
11	13	7	750	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.06
11	13	7	1000	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.05
11	13	7	2000	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.04
17	2	1	30	1.21	1.27	1.27	1.24	1.23	1.26	1.30	1.23	1.36
17	2	1	50	1.21	1.29	1.28	1.21	1.23	1.23	1.33	1.20	1.44
17	2	1	70	1.19	1.27	1.28	1.23	1.22	1.22	1.37	1.21	1.47
17	2	1	100	1.17	1.28	1.25	1.20	1.17	1.22	1.35	1.18	1.47
17	2	1	150	1.15	1.26	1.28	1.19	1.15	1.16	1.32	1.18	1.49
17	2	1	200	1.17	1.29	1.28	1.20	1.17	1.18	1.31	1.18	1.53
17	2	1	250	1.17	1.28	1.28	1.19	1.15	1.17	1.33	1.17	1.54
17	2	1	300	1.16	1.29	1.31	1.21	1.20	1.17	1.31	1.18	1.57
17	2	1	400	1.13	1.26	1.27	1.19	1.16	1.19	1.33	1.16	1.56
17	2	1	500	1.14	1.28	1.29	1.19	1.14	1.17	1.25	1.14	1.56
17	2	1	750	1.13	1.31	1.30	1.17	1.15	1.16	1.29	1.14	1.60
17	2	1	1000	1.14	1.30	1.34	1.18	1.14	1.17	1.26	1.15	1.60
17	2	1	2000	1.09	1.37	1.35	1.17	1.14	1.17	1.22	1.09	1.61
20	2	2	30	1.00	1.00	1.00	1.04	1.06	1.00	1.00	1.00	1.21
20	2	2	50	1.00	1.00	1.00	1.04	1.04	1.00	1.00	1.00	1.96
20	2	2	70	1.00	1.00	1.00	1.03	1.04	1.00	1.00	1.00	2.00
20	2	2	100	1.00	1.00	1.01	1.02	1.02	1.00	1.00	1.00	2.07
20	2	2	150	1.00	1.00	1.00	1.02	1.02	1.00	1.00	1.00	2.12
20	2	2	200	1.00	1.01	1.01	1.02	1.02	1.00	1.00	1.00	2.13
20	2	2	250	1.00	1.01	1.01	1.02	1.01	1.00	1.00	1.00	2.15
20	2	2	300	1.00	1.01	1.01	1.02	1.01	1.00	1.00	1.00	2.16
20	2	2	400	1.00	1.01	1.01	1.02	1.01	1.00	1.00	1.00	2.14
20	2	2	500	1.00	1.01	1.01	1.01	1.01	1.00	1.00	1.00	2.15
20	2	2	750	1.01	1.01	1.00	1.01	1.00	1.00	1.00	1.00	2.15
20	2	2	1000	1.00	1.00	1.01	1.01	1.01	1.00	1.00	1.00	2.14
20	2	2	2000	1.00	1.00	1.01	1.00	1.01	1.00	1.00	1.00	2.13
21	1	1	30	1.22	1.24	1.21	1.26	1.27	1.22	1.05	1.23	1.20
21	1	1	50	1.18	1.19	1.16	1.21	1.20	1.22	1.02	1.21	1.26
21	1	1	70	1.19	1.14	1.15	1.17	1.17	1.16	1.01	1.17	1.36
21	1	1	100	1.13	1.11	1.12	1.13	1.11	1.12	1.00	1.15	1.46
21	1	1	150	1.07	1.04	1.06	1.07	1.08	1.08	1.00	1.08	1.58
21	1	1	200	1.04	1.04	1.03	1.04	1.05	1.05	1.00	1.05	1.70
21	1	1	250	1.03	1.03	1.02	1.02	1.03	1.04	1.00	1.04	1.76
21	1	1	300	1.03	1.01	1.01	1.02	1.01	1.02	1.00	1.02	1.87
21	1	1	400	1.01	1.01	1.01	1.01	1.01	1.01	1.00	1.00	1.99
21	1	1	500	1.01	1.00	1.00	1.00	1.00	1.00	1.00	1.01	2.08
21	1	1	750	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	2.31
21	1	1	1000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	2.48
21	1	1	2000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	2.85

Table 2.7: The average number of identified bumps by TBH for distributions 20 and 22. TBH was used with $\lambda = 0.05$ without the X84 rule.

ID	True Bumps	True Modes	Sample Size	Bandwidth Combinations								
				Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8	Case 9
20	2	2	30	1.96	1.85	1.84	2.00	2.00	1.97	1.99	1.97	1.99
20	2	2	50	1.92	1.75	1.72	1.87	1.87	1.92	2.00	1.91	2.00
20	2	2	70	1.89	1.67	1.66	1.82	1.82	1.87	2.00	1.88	2.00
20	2	2	100	1.84	1.60	1.61	1.73	1.75	1.83	2.00	1.85	2.00
20	2	2	150	1.80	1.52	1.51	1.71	1.70	1.76	2.00	1.80	2.00
20	2	2	200	1.76	1.49	1.44	1.62	1.62	1.74	2.00	1.74	2.00
20	2	2	250	1.72	1.44	1.43	1.59	1.58	1.71	2.00	1.70	2.00
20	2	2	300	1.67	1.38	1.42	1.55	1.58	1.67	2.00	1.68	2.00
20	2	2	400	1.61	1.33	1.35	1.51	1.51	1.65	2.00	1.62	2.00
20	2	2	500	1.57	1.29	1.27	1.46	1.47	1.58	2.00	1.60	2.00
20	2	2	750	1.52	1.26	1.22	1.40	1.41	1.52	1.99	1.51	2.00
20	2	2	1000	1.48	1.21	1.21	1.37	1.38	1.48	2.00	1.49	2.00
20	2	2	2000	1.42	1.12	1.12	1.31	1.27	1.39	1.99	1.38	1.99
22	2	2	30	2.51	2.38	2.37	2.19	2.21	2.48	2.00	2.45	2.00
22	2	2	50	2.48	2.39	2.40	2.13	2.13	2.51	2.00	2.48	2.00
22	2	2	70	2.49	2.38	2.37	2.12	2.10	2.47	2.00	2.46	2.00
22	2	2	100	2.45	2.31	2.31	2.06	2.07	2.46	2.00	2.46	2.00
22	2	2	150	2.39	2.16	2.16	2.04	2.04	2.35	2.00	2.38	2.00
22	2	2	200	2.31	2.05	2.04	2.04	2.03	2.28	2.00	2.30	2.00
22	2	2	250	2.23	1.93	1.95	2.02	2.03	2.20	2.00	2.24	2.00
22	2	2	300	2.12	1.80	1.81	2.02	2.02	2.14	2.00	2.16	2.00
22	2	2	400	2.00	1.63	1.64	2.01	2.01	2.03	2.00	2.02	2.00
22	2	2	500	1.88	1.47	1.48	2.01	2.01	1.87	2.00	1.86	2.00
22	2	2	750	1.65	1.30	1.28	2.00	2.00	1.64	2.00	1.65	2.00
22	2	2	1000	1.50	1.15	1.16	1.99	1.99	1.47	2.00	1.49	2.00
22	2	2	2000	1.18	1.04	1.04	1.96	1.96	1.18	2.00	1.22	2.00

Table 2.8: The standard deviation of the number of identified bumps by TBH and NCM for distributions 1, 3, 5, 7, and 9. TBH was used with $\lambda = 0.05$ and the X84 rule.

ID	True Bumps	True Modes	Sample Size	TBH								NCM
				Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8	
1	1	1	30	0.41	0.43	0.46	0.38	0.39	0.45	0.44	0.41	0.46
1	1	1	50	0.35	0.39	0.39	0.33	0.34	0.35	0.39	0.36	0.47
1	1	1	70	0.32	0.36	0.35	0.31	0.31	0.32	0.38	0.29	0.48
1	1	1	100	0.27	0.32	0.33	0.27	0.26	0.27	0.38	0.23	0.48
1	1	1	150	0.25	0.31	0.29	0.25	0.23	0.26	0.36	0.24	0.48
1	1	1	200	0.19	0.29	0.27	0.20	0.22	0.22	0.31	0.22	0.47
1	1	1	250	0.21	0.24	0.26	0.21	0.18	0.20	0.32	0.23	0.47
1	1	1	300	0.18	0.24	0.26	0.20	0.15	0.16	0.28	0.18	0.47
1	1	1	400	0.16	0.25	0.22	0.18	0.10	0.17	0.30	0.19	0.46
1	1	1	500	0.16	0.22	0.23	0.16	0.14	0.13	0.28	0.13	0.45
1	1	1	750	0.14	0.20	0.16	0.14	0.08	0.09	0.24	0.14	0.45
1	1	1	1000	0.12	0.19	0.17	0.11	0.04	0.06	0.25	0.15	0.44
1	1	1	2000	0.08	0.09	0.11	0.08	0.00	0.03	0.20	0.11	0.43
3	1	1	30	0.58	0.59	0.57	0.50	0.49	0.56	0.38	0.59	0.49
3	1	1	50	0.60	0.59	0.59	0.49	0.48	0.56	0.28	0.61	0.52
3	1	1	70	0.60	0.56	0.59	0.49	0.48	0.59	0.24	0.58	0.55
3	1	1	100	0.60	0.56	0.57	0.48	0.50	0.59	0.23	0.58	0.57
3	1	1	150	0.58	0.55	0.53	0.46	0.46	0.59	0.13	0.57	0.60
3	1	1	200	0.54	0.51	0.51	0.40	0.42	0.57	0.11	0.55	0.63
3	1	1	250	0.52	0.48	0.47	0.37	0.39	0.54	0.07	0.53	0.64
3	1	1	300	0.49	0.43	0.42	0.33	0.34	0.52	0.09	0.48	0.66
3	1	1	400	0.43	0.36	0.37	0.27	0.32	0.50	0.05	0.42	0.67
3	1	1	500	0.38	0.33	0.30	0.22	0.25	0.44	0.00	0.40	0.67
3	1	1	750	0.28	0.19	0.19	0.13	0.18	0.38	0.00	0.27	0.69
3	1	1	1000	0.20	0.13	0.14	0.11	0.16	0.29	0.00	0.21	0.70
3	1	1	2000	0.06	0.04	0.05	0.00	0.03	0.13	0.00	0.07	0.66
5	3	1	30	0.07	0.12	0.13	0.25	0.15	0.07	0.08	0.10	0.16
5	3	1	50	0.09	0.09	0.12	0.18	0.09	0.04	0.04	0.06	0.33
5	3	1	70	0.04	0.12	0.11	0.14	0.00	0.00	0.00	0.07	0.48
5	3	1	100	0.05	0.10	0.10	0.17	0.00	0.00	0.00	0.00	0.59
5	3	1	150	0.04	0.11	0.14	0.14	0.00	0.00	0.00	0.06	0.69
5	3	1	200	0.06	0.08	0.12	0.12	0.00	0.00	0.00	0.06	0.72
5	3	1	250	0.06	0.11	0.11	0.11	0.00	0.00	0.00	0.07	0.78
5	3	1	300	0.05	0.10	0.12	0.15	0.00	0.00	0.00	0.05	0.83
5	3	1	400	0.05	0.11	0.11	0.08	0.00	0.00	0.00	0.07	0.91
5	3	1	500	0.07	0.12	0.09	0.11	0.00	0.00	0.00	0.07	0.96
5	3	1	750	0.07	0.07	0.08	0.07	0.00	0.00	0.00	0.05	1.10
5	3	1	1000	0.00	0.08	0.12	0.10	0.00	0.00	0.00	0.08	1.19
5	3	1	2000	0.00	0.00	0.09	0.15	0.00	0.00	0.00	0.00	1.30
7	2	2	30	0.18	0.12	0.08	0.07	0.06	0.09	0.03	0.15	0.06
7	2	2	50	0.09	0.08	0.10	0.00	0.00	0.04	0.00	0.11	0.02
7	2	2	70	0.09	0.08	0.04	0.00	0.03	0.00	0.00	0.08	0.02
7	2	2	100	0.08	0.08	0.07	0.03	0.00	0.00	0.00	0.06	0.00
7	2	2	150	0.04	0.08	0.07	0.00	0.00	0.00	0.00	0.06	0.00
7	2	2	200	0.04	0.04	0.03	0.03	0.00	0.03	0.00	0.03	0.00
7	2	2	250	0.04	0.04	0.03	0.00	0.00	0.00	0.00	0.03	0.00
7	2	2	300	0.00	0.03	0.04	0.00	0.00	0.00	0.00	0.00	0.00
7	2	2	400	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.04	0.00
7	2	2	500	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00
7	2	2	750	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00
7	2	2	1000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
7	2	2	2000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
9	3	3	30	0.39	0.30	0.31	0.31	0.31	0.32	0.20	0.39	0.26
9	3	3	50	0.31	0.21	0.21	0.22	0.25	0.24	0.15	0.30	0.25
9	3	3	70	0.28	0.19	0.19	0.20	0.19	0.24	0.14	0.23	0.30
9	3	3	100	0.21	0.15	0.17	0.19	0.17	0.18	0.13	0.25	0.38
9	3	3	150	0.11	0.10	0.07	0.06	0.08	0.10	0.04	0.13	0.39
9	3	3	200	0.08	0.09	0.09	0.06	0.05	0.11	0.03	0.10	0.46
9	3	3	250	0.05	0.03	0.04	0.03	0.04	0.05	0.04	0.04	0.46
9	3	3	300	0.09	0.04	0.08	0.00	0.00	0.05	0.03	0.07	0.49
9	3	3	400	0.06	0.03	0.05	0.00	0.00	0.00	0.00	0.04	0.50
9	3	3	500	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.03	0.50
9	3	3	750	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.46
9	3	3	1000	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.38
9	3	3	2000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.16

Table 2.9: The standard deviation of the number of identified bumps by TBH and NCM for distributions 10, 11, 17, 20, and 21. TBH was used with $\lambda = 0.05$ and the X84 rule.

ID	True Bumps	True Modes	Sample Size	TBH								NCM
				Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8	
10	5	5	30	0.40	0.45	0.44	0.42	0.42	0.44	0.48	0.40	0.49
10	5	5	50	0.41	0.47	0.46	0.43	0.43	0.44	0.49	0.42	0.52
10	5	5	70	0.41	0.50	0.48	0.46	0.44	0.45	0.49	0.41	0.54
10	5	5	100	0.45	0.49	0.50	0.47	0.45	0.47	0.50	0.45	0.57
10	5	5	150	0.46	0.50	0.51	0.48	0.46	0.46	0.50	0.46	0.62
10	5	5	200	0.48	0.49	0.49	0.49	0.48	0.49	0.50	0.49	0.66
10	5	5	250	0.50	0.47	0.46	0.50	0.48	0.49	0.49	0.50	0.70
10	5	5	300	0.52	0.46	0.42	0.50	0.49	0.50	0.48	0.53	0.74
10	5	5	400	0.61	0.45	0.46	0.50	0.49	0.52	0.39	0.62	0.81
10	5	5	500	0.85	0.69	0.67	0.47	0.50	0.58	0.35	0.79	0.88
10	5	5	750	1.13	1.12	1.11	0.29	0.49	0.69	0.13	1.12	0.86
10	5	5	1000	0.82	0.88	0.87	0.08	0.43	0.78	0.04	0.86	0.70
10	5	5	2000	0.14	0.29	0.31	0.49	0.48	0.60	0.05	0.16	0.45
11	13	7	30	0.49	0.45	0.46	0.46	0.45	0.44	0.38	0.48	0.39
11	13	7	50	0.45	0.39	0.38	0.39	0.40	0.41	0.31	0.43	0.34
11	13	7	70	0.40	0.29	0.25	0.28	0.33	0.32	0.22	0.36	0.30
11	13	7	100	0.32	0.21	0.24	0.28	0.30	0.28	0.19	0.29	0.30
11	13	7	150	0.24	0.17	0.11	0.19	0.21	0.19	0.12	0.20	0.28
11	13	7	200	0.22	0.13	0.13	0.13	0.20	0.19	0.07	0.20	0.29
11	13	7	250	0.14	0.05	0.09	0.12	0.13	0.13	0.06	0.10	0.29
11	13	7	300	0.14	0.07	0.05	0.05	0.12	0.11	0.03	0.09	0.27
11	13	7	400	0.03	0.03	0.03	0.03	0.00	0.05	0.00	0.00	0.27
11	13	7	500	0.00	0.03	0.00	0.00	0.03	0.04	0.00	0.00	0.26
11	13	7	750	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.23
11	13	7	1000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.22
11	13	7	2000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.19
17	2	1	30	0.43	0.47	0.48	0.44	0.44	0.46	0.46	0.46	0.49
17	2	1	50	0.43	0.46	0.46	0.41	0.43	0.45	0.47	0.42	0.52
17	2	1	70	0.41	0.45	0.46	0.42	0.41	0.42	0.48	0.42	0.53
17	2	1	100	0.38	0.45	0.44	0.40	0.37	0.42	0.48	0.39	0.54
17	2	1	150	0.36	0.44	0.45	0.39	0.36	0.37	0.47	0.39	0.56
17	2	1	200	0.38	0.46	0.45	0.40	0.38	0.39	0.47	0.39	0.57
17	2	1	250	0.37	0.45	0.45	0.40	0.36	0.38	0.47	0.38	0.56
17	2	1	300	0.37	0.45	0.46	0.40	0.40	0.38	0.46	0.38	0.58
17	2	1	400	0.34	0.44	0.44	0.39	0.37	0.39	0.47	0.37	0.58
17	2	1	500	0.34	0.45	0.46	0.39	0.34	0.37	0.43	0.35	0.59
17	2	1	750	0.34	0.46	0.46	0.37	0.36	0.36	0.45	0.34	0.60
17	2	1	1000	0.34	0.46	0.47	0.39	0.34	0.38	0.44	0.36	0.60
17	2	1	2000	0.28	0.48	0.48	0.37	0.35	0.37	0.42	0.29	0.62
20	2	2	30	0.00	0.00	0.00	0.20	0.25	0.00	0.00	0.00	0.41
20	2	2	50	0.00	0.03	0.03	0.19	0.20	0.00	0.00	0.00	0.19
20	2	2	70	0.00	0.05	0.04	0.16	0.20	0.03	0.00	0.00	0.20
20	2	2	100	0.00	0.00	0.07	0.13	0.15	0.00	0.00	0.03	0.25
20	2	2	150	0.00	0.05	0.04	0.14	0.14	0.03	0.00	0.00	0.32
20	2	2	200	0.00	0.08	0.08	0.13	0.12	0.00	0.00	0.03	0.34
20	2	2	250	0.04	0.11	0.09	0.14	0.11	0.00	0.00	0.00	0.36
20	2	2	300	0.04	0.09	0.09	0.12	0.09	0.03	0.00	0.04	0.37
20	2	2	400	0.03	0.09	0.10	0.13	0.08	0.04	0.00	0.04	0.35
20	2	2	500	0.00	0.08	0.08	0.07	0.08	0.03	0.00	0.05	0.36
20	2	2	750	0.07	0.07	0.04	0.10	0.06	0.03	0.00	0.03	0.36
20	2	2	1000	0.04	0.05	0.08	0.07	0.07	0.04	0.00	0.05	0.35
20	2	2	2000	0.04	0.03	0.08	0.03	0.08	0.03	0.00	0.04	0.34
21	1	1	30	0.45	0.45	0.42	0.46	0.46	0.44	0.23	0.46	0.40
21	1	1	50	0.41	0.41	0.38	0.42	0.41	0.44	0.13	0.45	0.45
21	1	1	70	0.41	0.35	0.37	0.38	0.40	0.38	0.11	0.39	0.50
21	1	1	100	0.35	0.32	0.33	0.35	0.32	0.34	0.04	0.36	0.55
21	1	1	150	0.26	0.20	0.25	0.25	0.28	0.28	0.00	0.27	0.60
21	1	1	200	0.20	0.21	0.18	0.21	0.21	0.22	0.00	0.21	0.63
21	1	1	250	0.16	0.17	0.15	0.14	0.17	0.20	0.00	0.20	0.64
21	1	1	300	0.16	0.10	0.11	0.13	0.12	0.14	0.00	0.15	0.67
21	1	1	400	0.10	0.10	0.07	0.10	0.11	0.11	0.00	0.06	0.69
21	1	1	500	0.07	0.05	0.04	0.06	0.00	0.06	0.00	0.10	0.71
21	1	1	750	0.05	0.05	0.05	0.03	0.00	0.03	0.00	0.00	0.75
21	1	1	1000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.79
21	1	1	2000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.87

Table 2.10: The standard deviation of the number of identified bumps by TBH for distributions 20 and 22. TBH was used with $\lambda = 0.05$ without the X84 rule.

ID	True Bumps	True Modes	Sample Size	Bandwidth Combinations								
				Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8	Case 9
20	2	2	30	0.19	0.36	0.37	0.31	0.29	0.17	0.11	0.18	0.10
20	2	2	50	0.27	0.44	0.45	0.40	0.40	0.27	0.05	0.28	0.04
20	2	2	70	0.31	0.47	0.48	0.43	0.43	0.34	0.00	0.33	0.03
20	2	2	100	0.37	0.50	0.49	0.48	0.46	0.38	0.03	0.36	0.05
20	2	2	150	0.40	0.51	0.51	0.47	0.49	0.43	0.03	0.40	0.03
20	2	2	200	0.43	0.51	0.50	0.50	0.51	0.44	0.03	0.44	0.05
20	2	2	250	0.45	0.50	0.50	0.51	0.51	0.45	0.05	0.46	0.03
20	2	2	300	0.47	0.49	0.52	0.52	0.50	0.47	0.06	0.47	0.06
20	2	2	400	0.49	0.47	0.48	0.52	0.50	0.48	0.04	0.48	0.05
20	2	2	500	0.49	0.46	0.45	0.50	0.52	0.50	0.06	0.49	0.00
20	2	2	750	0.50	0.44	0.42	0.49	0.50	0.50	0.08	0.50	0.06
20	2	2	1000	0.50	0.41	0.41	0.49	0.49	0.50	0.06	0.50	0.07
20	2	2	2000	0.50	0.32	0.32	0.46	0.45	0.49	0.11	0.49	0.09
22	2	2	30	0.56	0.52	0.52	0.43	0.45	0.55	0.06	0.54	0.04
22	2	2	50	0.55	0.55	0.53	0.39	0.37	0.56	0.00	0.55	0.00
22	2	2	70	0.55	0.54	0.53	0.35	0.33	0.55	0.00	0.54	0.00
22	2	2	100	0.55	0.56	0.54	0.28	0.28	0.54	0.00	0.54	0.00
22	2	2	150	0.54	0.59	0.56	0.20	0.22	0.55	0.00	0.55	0.00
22	2	2	200	0.55	0.60	0.58	0.21	0.20	0.55	0.00	0.53	0.00
22	2	2	250	0.55	0.61	0.60	0.17	0.20	0.53	0.00	0.52	0.00
22	2	2	300	0.53	0.62	0.60	0.16	0.15	0.56	0.00	0.56	0.00
22	2	2	400	0.57	0.58	0.59	0.13	0.14	0.55	0.00	0.56	0.00
22	2	2	500	0.59	0.56	0.55	0.13	0.11	0.57	0.00	0.58	0.00
22	2	2	750	0.57	0.48	0.47	0.11	0.11	0.57	0.00	0.56	0.00
22	2	2	1000	0.55	0.36	0.38	0.12	0.13	0.53	0.00	0.54	0.00
22	2	2	2000	0.38	0.19	0.19	0.21	0.20	0.39	0.00	0.42	0.00

Table 2.11: Number of execution failures for NCM out of 1,000 simulations.

ID	Sample Size												
	30	50	70	100	150	200	250	300	400	500	750	1000	2000
5	0	0	0	0	1	18	45	91	203	336	566	705	884
10	0	0	0	0	0	0	0	0	0	1	2	10	56
21	0	0	0	0	0	0	1	4	11	14	31	51	68

Table 2.12: The percentage of the number of decisions toward unimodality made by TBH, NCM, and DIP for distributions 1, 3, 5, 7, and 9. TBH was used with $\lambda = 0.05$ and the X84 rule.

ID	True Bumps	True Modes	Sample Size	TBH								NCM	DIP
				Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8		
1	1	1	30	82%	78%	75%	84%	83%	76%	75%	83%	71%	99%
1	1	1	50	88%	82%	83%	88%	87%	86%	82%	87%	71%	100%
1	1	1	70	90%	86%	87%	89%	90%	89%	82%	91%	71%	100%
1	1	1	100	92%	89%	88%	92%	93%	92%	83%	94%	71%	100%
1	1	1	150	93%	89%	91%	94%	94%	93%	84%	94%	72%	100%
1	1	1	200	96%	91%	92%	96%	95%	95%	89%	95%	74%	100%
1	1	1	250	96%	94%	93%	96%	97%	96%	89%	95%	75%	100%
1	1	1	300	97%	94%	93%	96%	98%	97%	92%	97%	74%	100%
1	1	1	400	97%	94%	95%	97%	99%	97%	90%	96%	76%	100%
1	1	1	500	98%	95%	95%	97%	98%	98%	92%	98%	76%	100%
1	1	1	750	98%	96%	98%	98%	99%	99%	94%	98%	76%	100%
1	1	1	1000	99%	96%	97%	99%	100%	100%	93%	98%	77%	100%
1	1	1	2000	99%	99%	99%	99%	100%	100%	96%	99%	79%	100%
3	1	1	30	46%	55%	49%	65%	65%	47%	83%	50%	63%	100%
3	1	1	50	49%	51%	51%	65%	67%	42%	91%	42%	47%	100%
3	1	1	70	45%	54%	52%	66%	68%	46%	94%	46%	42%	100%
3	1	1	100	49%	55%	56%	70%	67%	47%	94%	50%	38%	100%
3	1	1	150	51%	62%	61%	75%	74%	53%	98%	53%	34%	100%
3	1	1	200	63%	69%	68%	81%	79%	56%	99%	61%	34%	100%
3	1	1	250	68%	73%	75%	84%	83%	61%	100%	67%	33%	100%
3	1	1	300	70%	78%	79%	88%	86%	68%	99%	72%	36%	100%
3	1	1	400	78%	86%	84%	92%	89%	69%	100%	79%	35%	100%
3	1	1	500	85%	89%	91%	95%	94%	77%	100%	82%	36%	100%
3	1	1	750	91%	97%	96%	98%	97%	84%	100%	92%	38%	100%
3	1	1	1000	96%	98%	98%	99%	97%	91%	100%	95%	41%	100%
3	1	1	2000	100%	100%	100%	100%	100%	98%	100%	100%	51%	100%
5	3	1	30	100%	99%	99%	94%	98%	100%	99%	99%	97%	100%
5	3	1	50	99%	99%	99%	97%	99%	100%	100%	100%	88%	100%
5	3	1	70	100%	99%	99%	98%	100%	100%	100%	100%	69%	100%
5	3	1	100	100%	99%	99%	97%	100%	100%	100%	100%	48%	100%
5	3	1	150	100%	99%	98%	98%	100%	100%	100%	100%	19%	100%
5	3	1	200	100%	99%	98%	99%	100%	100%	100%	100%	8%	100%
5	3	1	250	100%	99%	99%	99%	100%	100%	100%	99%	3%	100%
5	3	1	300	100%	99%	99%	98%	100%	100%	100%	100%	1%	100%
5	3	1	400	100%	99%	99%	99%	100%	100%	100%	100%	0%	100%
5	3	1	500	100%	99%	99%	99%	100%	100%	100%	100%	0%	100%
5	3	1	750	100%	100%	99%	100%	100%	100%	100%	100%	0%	100%
5	3	1	1000	100%	99%	99%	99%	100%	100%	100%	99%	0%	100%
5	3	1	2000	100%	100%	99%	98%	100%	100%	100%	100%	0%	100%
7	2	2	30	2%	1%	0%	1%	0%	0%	0%	1%	0%	40%
7	2	2	50	0%	0%	0%	0%	0%	0%	0%	0%	0%	18%
7	2	2	70	0%	0%	0%	0%	0%	0%	0%	0%	0%	7%
7	2	2	100	0%	0%	0%	0%	0%	0%	0%	0%	0%	2%
7	2	2	150	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
7	2	2	200	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
7	2	2	250	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
7	2	2	300	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
7	2	2	400	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
7	2	2	500	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
7	2	2	750	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
7	2	2	1000	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
7	2	2	2000	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
9	3	3	30	16%	9%	10%	11%	11%	11%	4%	15%	6%	85%
9	3	3	50	10%	4%	4%	5%	7%	6%	2%	9%	2%	85%
9	3	3	70	7%	3%	3%	4%	4%	6%	2%	5%	1%	83%
9	3	3	100	4%	1%	2%	4%	3%	3%	2%	4%	1%	84%
9	3	3	150	1%	0%	0%	0%	1%	1%	0%	1%	0%	71%
9	3	3	200	0%	0%	0%	0%	0%	1%	0%	1%	0%	68%
9	3	3	250	0%	0%	0%	0%	0%	0%	0%	0%	0%	56%
9	3	3	300	0%	0%	0%	0%	0%	0%	0%	0%	0%	52%
9	3	3	400	0%	0%	0%	0%	0%	0%	0%	0%	0%	39%
9	3	3	500	0%	0%	0%	0%	0%	0%	0%	0%	0%	27%
9	3	3	750	0%	0%	0%	0%	0%	0%	0%	0%	0%	9%
9	3	3	1000	0%	0%	0%	0%	0%	0%	0%	0%	0%	3%
9	3	3	2000	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%

Table 2.13: The percentage of the number of decisions toward unimodality made by TBH, NCM, and DIP for distributions 10, 11, 17, 20, and 21. TBH was used with $\lambda = 0.05$ and the X84 rule.

ID	True Bumps	True Modes	Sample Size	TBH								NCM	DIP
				Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8		
10	5	5	30	81%	72%	76%	77%	78%	75%	63%	81%	63%	100%
10	5	5	50	79%	68%	69%	76%	76%	74%	59%	78%	57%	100%
10	5	5	70	79%	59%	64%	70%	74%	73%	59%	78%	51%	100%
10	5	5	100	72%	58%	57%	68%	73%	68%	53%	72%	45%	99%
10	5	5	150	69%	45%	50%	63%	70%	70%	49%	70%	38%	99%
10	5	5	200	66%	40%	41%	60%	64%	64%	45%	64%	31%	98%
10	5	5	250	59%	32%	30%	56%	64%	59%	41%	56%	24%	96%
10	5	5	300	51%	25%	22%	52%	62%	58%	37%	54%	18%	94%
10	5	5	400	33%	13%	11%	44%	58%	54%	19%	38%	7%	87%
10	5	5	500	21%	4%	4%	32%	53%	45%	14%	20%	3%	75%
10	5	5	750	1%	0%	0%	9%	40%	31%	2%	1%	0%	38%
10	5	5	1000	0%	0%	0%	1%	25%	20%	0%	1%	0%	10%
10	5	5	2000	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
11	13	7	30	34%	25%	28%	30%	28%	26%	18%	31%	17%	91%
11	13	7	50	25%	17%	16%	19%	20%	22%	11%	22%	9%	90%
11	13	7	70	17%	8%	6%	8%	12%	11%	5%	14%	4%	86%
11	13	7	100	11%	4%	5%	9%	10%	8%	4%	8%	3%	84%
11	13	7	150	6%	3%	1%	4%	5%	4%	1%	4%	1%	79%
11	13	7	200	5%	1%	2%	2%	4%	4%	1%	4%	1%	79%
11	13	7	250	2%	0%	1%	1%	2%	2%	0%	1%	0%	75%
11	13	7	300	2%	0%	0%	0%	1%	1%	0%	1%	0%	67%
11	13	7	400	0%	0%	0%	0%	0%	0%	0%	0%	0%	56%
11	13	7	500	0%	0%	0%	0%	0%	0%	0%	0%	0%	44%
11	13	7	750	0%	0%	0%	0%	0%	0%	0%	0%	0%	26%
11	13	7	1000	0%	0%	0%	0%	0%	0%	0%	0%	0%	14%
11	13	7	2000	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%
17	2	1	30	80%	74%	75%	76%	77%	75%	70%	79%	64%	99%
17	2	1	50	79%	72%	73%	79%	78%	78%	67%	80%	57%	99%
17	2	1	70	81%	73%	72%	77%	78%	78%	64%	79%	55%	100%
17	2	1	100	83%	73%	75%	80%	83%	79%	65%	83%	55%	100%
17	2	1	150	85%	74%	73%	81%	85%	84%	68%	82%	53%	100%
17	2	1	200	83%	71%	72%	80%	83%	82%	69%	82%	51%	100%
17	2	1	250	84%	72%	73%	81%	85%	83%	67%	83%	49%	100%
17	2	1	300	84%	71%	70%	80%	81%	83%	69%	83%	48%	100%
17	2	1	400	87%	74%	73%	81%	84%	81%	67%	84%	49%	100%
17	2	1	500	86%	72%	71%	82%	86%	83%	75%	86%	49%	100%
17	2	1	750	87%	69%	70%	83%	85%	84%	71%	86%	46%	100%
17	2	1	1000	87%	70%	66%	82%	86%	83%	74%	85%	46%	100%
17	2	1	2000	91%	63%	65%	83%	86%	83%	78%	91%	47%	100%
20	2	2	30	100%	100%	100%	96%	95%	100%	100%	100%	79%	100%
20	2	2	50	100%	100%	100%	96%	96%	100%	100%	100%	4%	100%
20	2	2	70	100%	100%	100%	97%	96%	100%	100%	100%	2%	100%
20	2	2	100	100%	100%	100%	98%	98%	100%	100%	100%	0%	100%
20	2	2	150	100%	100%	100%	98%	98%	100%	100%	100%	0%	100%
20	2	2	200	100%	99%	99%	98%	99%	100%	100%	100%	0%	100%
20	2	2	250	100%	99%	99%	98%	99%	100%	100%	100%	0%	100%
20	2	2	300	100%	99%	99%	99%	99%	100%	100%	100%	0%	99%
20	2	2	400	100%	99%	99%	98%	99%	100%	100%	100%	0%	90%
20	2	2	500	100%	99%	99%	100%	99%	100%	100%	100%	0%	60%
20	2	2	750	100%	100%	100%	99%	100%	100%	100%	100%	0%	3%
20	2	2	1000	100%	100%	99%	100%	100%	100%	100%	100%	0%	0%
20	2	2	2000	100%	100%	99%	100%	99%	100%	100%	100%	0%	0%
21	1	1	30	79%	77%	79%	75%	74%	79%	95%	79%	80%	100%
21	1	1	50	83%	82%	84%	80%	80%	79%	98%	80%	74%	100%
21	1	1	70	82%	86%	86%	84%	84%	85%	99%	84%	64%	100%
21	1	1	100	88%	89%	89%	87%	89%	89%	100%	86%	56%	100%
21	1	1	150	93%	96%	94%	93%	92%	92%	100%	92%	47%	100%
21	1	1	200	96%	96%	97%	96%	95%	95%	100%	95%	39%	100%
21	1	1	250	97%	97%	98%	98%	97%	96%	100%	96%	35%	100%
21	1	1	300	97%	99%	99%	98%	99%	98%	100%	98%	29%	100%
21	1	1	400	99%	99%	99%	99%	99%	99%	100%	100%	23%	100%
21	1	1	500	99%	100%	100%	100%	100%	100%	100%	99%	19%	100%
21	1	1	750	100%	100%	100%	100%	100%	100%	100%	100%	12%	100%
21	1	1	1000	100%	100%	100%	100%	100%	100%	100%	100%	8%	100%
21	1	1	2000	100%	100%	100%	100%	100%	100%	100%	100%	2%	100%

DIP is the best performer for unimodal distributions 1, 3, 5, 17 and 21, but

Table 2.14: The percentage of the number of decisions toward unimodality made by TBH, NCM, and DIP for distributions 20 and 22. TBH was used with $\lambda = 0.05$ without the X84 rule.

ID	True Bumps	True Modes	Sample Size	TBH									NCM	DIP
				Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8	Case 9		
20	2	2	30	4%	15%	16%	5%	4%	3%	1%	4%	1%	80%	100%
20	2	2	50	8%	25%	28%	15%	15%	8%	0%	9%	0%	4%	100%
20	2	2	70	11%	33%	35%	20%	20%	13%	0%	12%	0%	2%	100%
20	2	2	100	17%	40%	39%	29%	27%	17%	0%	15%	0%	0%	100%
20	2	2	150	20%	48%	49%	30%	32%	24%	0%	20%	0%	0%	100%
20	2	2	200	24%	52%	56%	39%	39%	26%	0%	26%	0%	0%	100%
20	2	2	250	28%	57%	57%	42%	43%	29%	0%	30%	0%	0%	100%
20	2	2	300	33%	62%	59%	46%	42%	33%	0%	33%	0%	0%	100%
20	2	2	400	39%	68%	65%	50%	50%	36%	0%	38%	0%	0%	90%
20	2	2	500	43%	71%	73%	54%	54%	42%	0%	40%	0%	0%	62%
20	2	2	750	48%	75%	78%	60%	59%	48%	1%	49%	0%	0%	2%
20	2	2	1000	53%	79%	80%	63%	62%	52%	0%	52%	1%	0%	0%
20	2	2	2000	58%	88%	88%	70%	73%	61%	1%	62%	1%	0%	0%
22	2	2	30	0%	1%	0%	0%	0%	0%	0%	0%	0%	1%	64%
22	2	2	50	0%	2%	1%	1%	0%	0%	0%	0%	0%	0%	45%
22	2	2	70	0%	2%	2%	1%	1%	0%	0%	0%	0%	0%	29%
22	2	2	100	1%	4%	3%	1%	1%	0%	0%	1%	0%	0%	13%
22	2	2	150	1%	10%	9%	0%	0%	2%	0%	2%	0%	0%	2%
22	2	2	200	4%	15%	15%	0%	1%	4%	0%	3%	0%	0%	0%
22	2	2	250	6%	22%	21%	1%	1%	6%	0%	4%	0%	0%	0%
22	2	2	300	9%	32%	29%	1%	0%	9%	0%	8%	0%	0%	0%
22	2	2	400	16%	42%	42%	0%	1%	14%	0%	15%	0%	0%	0%
22	2	2	500	24%	56%	55%	1%	0%	24%	0%	25%	0%	0%	0%
22	2	2	750	40%	71%	73%	1%	0%	41%	0%	39%	0%	0%	0%
22	2	2	1000	53%	85%	84%	1%	1%	55%	0%	53%	0%	0%	0%
22	2	2	2000	83%	96%	96%	4%	4%	82%	0%	78%	0%	0%	0%

it performs poorly on multimodal distributions 7, 9, 10, 11 and 20. On the other hand, TBH is the second best for distributions 1, 3, 5, 17, 20 (without X84) and 21. However, TBH is somewhat inferior to NCM in recognizing the multimodal nature of distributions 9, 10 and 11 for smaller samples with bandwidth combination 2 from Table 2.3 being close to NCM.

In terms of execution time, as demonstrated in Table 2.15, TBH is an intermediate performer between DIP (the fastest) and NCM (slowest).

Table 2.15: Execution time, in seconds.

Procedure	Sample Size						
	50	100	250	500	1000	2000	5000
TBH	0.18	0.21	0.21	0.39	0.57	0.95	2.33
NCM	0.07	0.10	0.25	0.77	2.27	7.24	64.09
DIP	0.07	0.09	0.11	0.15	0.21	0.23	0.24

While NCM is fast for small samples, the execution time grows rapidly with sample size. Taking into consideration the fact that DIP (Maechler, 2015) aims to merely decide if the data at hand come from a unimodal or multimodal distribution without providing the specifics on the number of bumps, TBH is definitely a fairly fast method for doing the bump identification job although further code optimization is possible.

We conclude this section by suggesting the following iterative strategy for using the TBH in practice:

Step 1 Fix a value of λ that best reflects one's belief in the degree of bump prominence.

Step 2 Decide which bump types are of interest: modal only or non-modal as well. Curvature or second derivative information will be used accordingly.

Step 3 Visually examine the data at hand for indications of long tails, peakedness, sharp features and the presence of dense regions in the tails.

Step 4 Fit a KDE $\hat{f}_h(\cdot)$ (and $\hat{C}_{h_1,h2}(\cdot)$ or $\hat{f}_h''(\cdot)$ according to Step 2) with a few choices of bandwidths.

- For data without long tails, which may be either regularly shaped or with apparent sharp features, Combinations 1, 2, 7, 8, or 9 for bandwidths may be attempted.
- For long tailed or/and kurtotic distributions, Combinations with cruder bandwidths are recommended such as Combinations 7 or 9. Also, the outlier removal rule X84 may be invoked for better KDE performance if there is no evidence of the dense region in the tails. If, however, the relative frequency of observations affected by X84 is comparable to or exceeds λ, one may consider applying TBH without X84.

Step 5 Assess the overall quality of the KDE $\hat{f}_h(\cdot)$ fits and select the one which fits the data reasonably well. This can be achieved by examining the Cox-Snell residuals as exemplified in Section 2.5.

Step 6 Run TBH.

Step 7 Examine the resultant number and location of bumps. Compare these with prior beliefs, if any. If not satisfied by the results, then return to Step 1 and adjust relevant parameters.

2.5 Application to Real Data

In this section, we apply the proposed method to the scaled Chondrite data from Good and Gaskins (1980, p. 44). The data originate from the distribution of silica in 22 chondrite meteors, and are presented in Table 2.16. The scatterplot and histogram are displayed in Figure 2.10.

Table 2.16: Scaled Percentages of Silica in 22 Chondrites (Good and Gaskins, 1980, p. 44). Used with Permission from Taylor & Francis Group.

Scaled Percentages of Silica in 22 Chondrites										
0.04	0.15	0.16	0.18	0.4	0.44	0.45	0.46	0.47	0.49	0.54
0.59	0.64	0.75	0.81	0.83	0.84	0.84	0.85	0.87	0.87	0.92

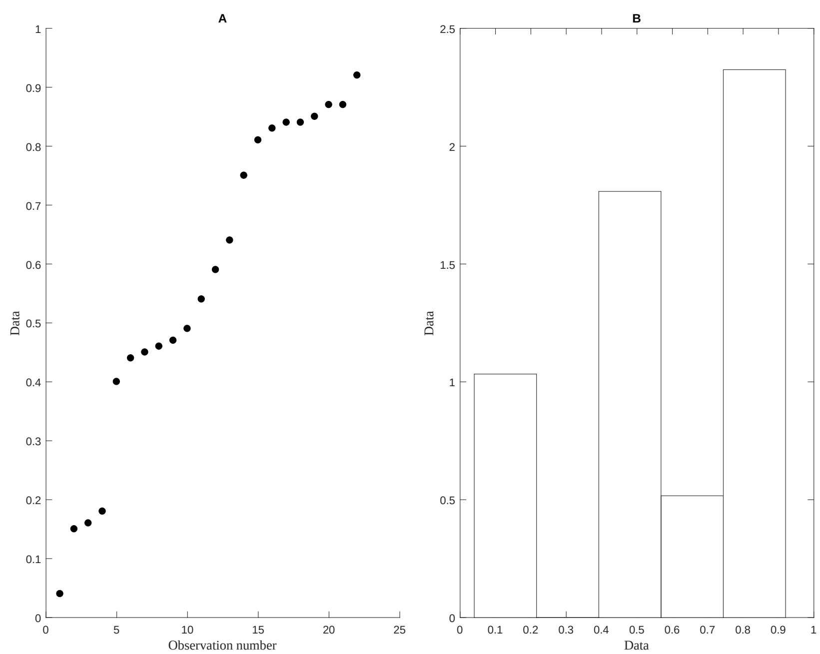

Figure 2.10: The scatterplot (Panel A) and histogram (Panel B) of the chondrite data.

We considered four KDE (2.1) fits, namely, with bandwidth equal to the Sheather-Jones solve-the-equation, Sheather-Jones direct plugin, Silverman's rule-of-thumb and the 3/4-th of the Silverman's rule of thumb. The overall quality of the fits were assessed by using the Cox-Snell residuals (Collett, 1994). Recall that the Cox-Snell residual for a data set $X = \{X_i\}_{i=1,\dots,n}$ is defined as $r(x) = -\ln(1-\hat{F}(x))$, where the $\hat{F}(\cdot)$ is the CDF of X estimated based on the model. In the case of KDE (2.1), with a standard Gaussian kernel, the CDF is given by

$$\hat{F}_h(x) = \frac{1}{n}\sum_{j=1}^{n}\Phi\left(\frac{x-X_j}{h}\right).$$

If the model is exactly right, the Cox-Snell residuals should follow a unit exponential distribution, and so the quantiles of $r(X_i)$ should be close to the line passing through the origin with unit slope. Therefore, in Figure 2.11, we plot the estimated residuals against the quantiles of the unit-exponential distribution having unit mean and variance. Figure 2.11 suggests that none of the candidate bandwidths performs particularly well. This may be attributed to the asymptotic nature of the considered bandwidths while the size of the data set under examination may be too small.

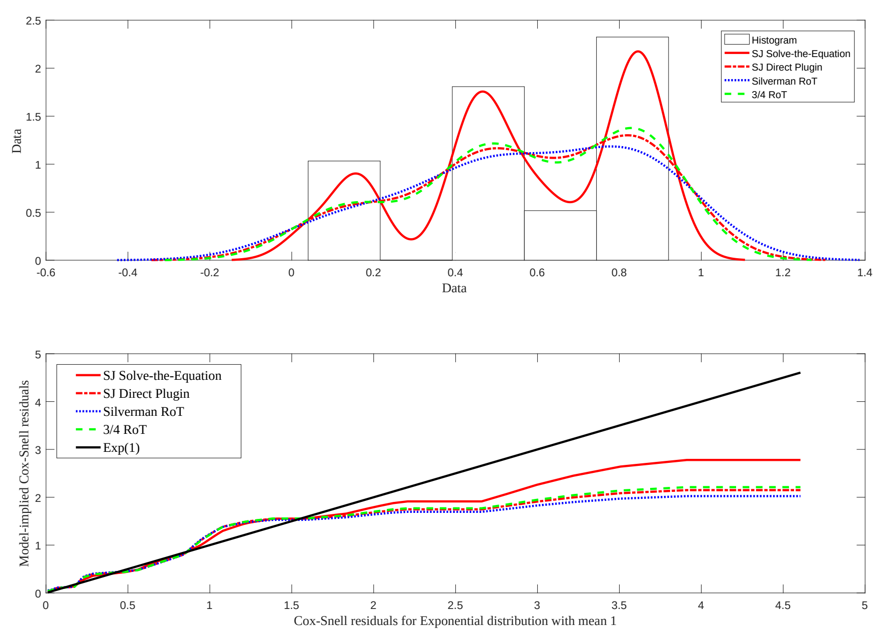

Figure 2.11: The KDE fits (Panel A) and Cox-Snell residuals (Panel B) of the chondrite data.

Having said this, we note that based on the descriptive statistics from Table 2.17, the Sheather-Jones solve-the-equation bandwidth provides the best fit. For this reason, we have chosen this bandwidth for the subsequent analysis.

Table 2.17: Bandwidths and descriptive statistics for Cox-Snell residuals for the KDE fits on the Chondrite data. Deviations are calculated with respect to the quantiles of the unit-exponential distribution.

Bandwidth Type	Bandwidth Value	Mean	Variance	Maximum Absolute Deviation	Mean Absolute Deviation
SJ Solve-the-Equation	0.06	0.91	0.52	1.83	0.15
SJ Direct Plugin	0.13	0.87	0.41	2.46	0.21
Silverman Rule-of-Thumb	0.16	0.86	0.38	2.58	0.23
3/4 of Silverman Rule-of-Thumb	0.12	0.88	0.42	2.39	0.21

We have applied TBH to the chondrite data with the choice of $\lambda = 0.05$. The data exhibit neither long tails nor peakedness. In fact, no observation lies more than $5.2MAD$ from the median. Hence, X84 was not considered. To get an estimate for curvature, we used the rule-of-thumb based on (2.4) for the bandwidth, namely, $h_C = 0.18$. An application of TBH to the chondrite data then resulted in identification of three bumps each of which is modal (see Table 2.18).[8] We note that the negative value in Table 2.18 is a consequence of using the Gaussian kernel.

Table 2.18: Results of an application of TBH to the chondrite data: the values of the areas of the caps of the identified bumps are provided. Also, the locations of the corresponding peaks and troughs are shown. Each peak $p_i = \langle x_{p_i}, y_{p_i} \rangle$, $i \in \{1,2,3\}$, has a pair of troughs $t_i^{left} = \langle x_{t_i^l}, y_{t_i^l} \rangle$ and $t_i^{right} = \langle x_{t_i^r}, y_{t_i^r} \rangle$ containing it.

No	Areas	Peaks		Troughs			
		x_p	y_p	x_{t^l}	y_{t^l}	x_{t^r}	y_{t^r}
1	0.13	0.16	0.90	-0.12	0.01	0.29	0.22
2	0.24	0.47	1.76	0.30	0.22	0.68	0.60
3	0.30	0.85	2.17	0.68	0.60	1.06	0.04

Both Good and Gaskins (1980) and Silverman (1980) also arrived at the same conclusion of the trimodal nature of the chondrite data. On the other hand, the DIP test suggests unimodality (the p-value equals 0.18) while the NCM procedure discovers just 2 clusters. This behavior of the latter two methods may be attributed to the small number of observations in the data set.

[8]The same conclusion is reached if h_C is taken to be the SJ solve-the-equation bandwidth or when ridges and courses are not used in TBH.

Chapter 3

On a Family of Weighted Cramér-von Mises Goodness-of-Fit Tests in Operational Risk Modeling

3.1 Introduction

Measurement of operational risk, through a loss distribution approach (LDA), for bank capitalization purposes offers significant modeling challenges. Under LDA, the severity of losses characterizing the monetary impact of potential operational risk events is modelled via a severity distribution. The selection of best-fit severity distributions that properly capture tail behavior is essential for accurate modeling.

In this chapter, we analyze limiting properties of a family of weighted Cramér-von Mises (WCvM) goodness-of-fit test statistics, with weight function $\psi(t) = 1/(1-t)^{\beta}$, that are suitable for more accurate selection of severity distributions. Specifically, we apply classical theory to determine if limiting distributions exist for these WCvM test statistics under a simple null hypothesis. We show that limiting distributions do not exist for $\beta \geqslant 2$. For $\beta = 2$, we provide a normalization that leads to a non-degenerate limiting distribution. Where limiting distributions originally exist, for $\beta < 2$, or were obtained through the normalization, we show that for $1.5 \leqslant \beta \leqslant 2$, the tests' practical utility may be limited due to a very slow convergence of the finite sample distribution to the asymptotic regime.

Our results suggest that the tests provide greater utility when $\beta < 1.5$ and that for $\beta \geqslant 1.5$ utility is questionable as only Monte Carlo schemes are practical even for very large samples.

This chapter uses material from the following article:

- K. Mayorov, J. Hristoskov and N. Balakrishnan. On a family of weighted Cramér-von Mises goodness-of-fit tests in operational risk modeling. *Journal of Operational Risk*, forthcoming 2017.

The material is reproduced with permission from Incisive Media.

3.2 Suitable GoF Tests for OpRisk Modeling

Under LDA, financial institutions model frequencies and magnitudes of losses that can result from potential future OpRisk events. The various magnitudes of potential losses and their relative probabilities of occurrence are assumed to be best captured and modeled through parametric severity distributions. In general, the profile of historic losses is highly leptokurtic which makes weighted GoF test statistics that are more sensitive to tails, such as the WCvM test statistic, popular for selection of candidate severity distributions. In order to convey the importance of severity distribution selection on capital modeling, below we show how under a generalized capital approximation regime, the upper quantiles of the severity distribution directly affect capital requirements.

Definition 3.1. The *Aggregate Loss Distribution (ALD)* is the distribution of the aggregate (compound) loss: $L = X_1 + ... + X_N$, where the frequency N is a discrete random variable and $X_1, ..., X_N$ are positive continuous random severities.

OpRisk capital is then a quantile of the ALD, given by

$$VaR_{\delta}(L) = F_L^{-1}(\delta),$$

where $F_L^{-1}(t) = \inf\{x \in \mathbb{R} \mid P(L > x) \leqslant 1 - t\}$ and confidence level $\delta \in (0,1)$. Typically $\delta \in \{0.999, 0.9995, 0.9997\}$.

Frequencies of losses are usually assumed to follow the Poisson distribution and generally are well-behaved and do not exhibit leptokurtic features. Therefore, with the exception of cases where frequencies of losses are extremely low and jumps may occur (e.g., where the Poisson lambda is less than 0.25 per annum), new loss event occurrences usually do not change the profile of frequency distributions and the impact of new occurrences on capital requirements is typically modest.

This is due to the so-called Single Loss Approximation (Böcker and Klüppelberg, 2005) capital $VaR_{\delta}(L) \approx F^{-1}\left(1-\frac{1-\delta}{\lambda}\right)$, where λ is the mean of the Poisson-distributed frequency of loss event occurrence, and $F(\cdot)$ is the severity cumulative distribution function (CDF).

For regulatory capital, $\delta = 0.999$. Even a λ of moderate magnitude, say, 25, 50 and 100 leads to calculation of 99.996, 99.998 and 99.999%-quantiles of the severity distribution.

Consequently, the final choice of a severity distribution is crucial and capital requirements are highly sensitive to upper tail behaviour of the selected severity distribution. This supports the notion that use of appropriately tuned GoF tests statistics for severity distribution selection is of significant importance. Our analysis, therefore, will focus on presenting findings which allow for the development of practical utility properties for WCvM test statistics. Based on the developed properties we are able to indicate weight boundaries for the test statistics that ensure that practical utility exists.

To this end, below we provide a brief description of some concepts and techniques that are key to the developments presented in subsequent sections.

Definition. Given a random sample $x_1, \ldots, x_n$ of independent observations with CDF F. The *Empirical (Sample) Distribution Function* (EDF) is defined as $F_n(t) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{(-\infty,t]}(x_i)$, where $\mathbb{I}_A(\cdot)$ is the indicator function on event A.

Some basic properties of the EDF are as follows:

- $\mathbb{E}(F_n(t)) = F(t)$ and $Cov(F_n(t_1), F_n(t_2)) = \frac{1}{n}F(t_1)(1-F(t_2))$ for $0 \leqslant t_1 \leqslant t_2 \leqslant 1$.
- $\sqrt{n}(F_n(t) - F(t)) \xrightarrow{d} N(0, \sqrt{F(t)(1-F(t))})$ as $n \to \infty$.

We wish to test the null hypothesis

$$H_0 : F(x) = F_0(x) \text{ vs } H_1 : F(x) \neq F_0(x) \tag{3.1}$$

for a continuous CDF F_0.

The parameters of the CDF F_0 may be completely known or have to be estimated. The two cases are known as a simple and composite hypothesis, respectively. In this chapter, we will consider the case of simple hypotheses, and composite hypotheses are discussed in Chapter 4.

Definition. *EDF-based GoF tests* are tests for assessing H_0 that are based on a comparison of $F_0(x)$ with $F_n(x)$.

The Kolmogorov-Smirnov and Cramér-von Mises (CvM) test statistics proposed in this regard are

$$D_n = \sqrt{n} \sup_{x \in \mathbb{R}} |F_n(x) - F_0(x)| \text{ and } \omega_n^2 = n \int_{-\infty}^{\infty} (F_n(x) - F_0(x))^2 \, dF_0(x).$$

These statistics are classical representatives of two classes of EDF-based GoF tests: the supremum and integral-based tests.

If a test statistic is adopted, denoted by K_n, the hypothesis is rejected for the samples for which K_n is greater than some $C_{n,\alpha}$.

The value $C_{n,\alpha}$ is to be chosen so that when the hypothesis H_0 is true the probability of rejection is some specified number, referred to as a significance level and denoted by α. Usually α takes values from $\{0.01, 0.05, 0.10\}$. Then $C_{n,\alpha}$ is called the critical value of the test statistic at significance level α. The critical value $C_{n,\alpha}$ is usually determined by Monte Carlo simulations.

We now define the limiting distribution of a test statistic.

Definition 3.2. Let $K_n \xrightarrow{d} K_\infty$ as $n \to \infty$, i.e., $F_{K_n}(t) \to F_{K_\infty}(t)$ at continuity points of $F_{K_\infty}(\cdot)$. If K_∞ is non-degenerate (i.e., with probability one (wp1) not a constant or $\pm\infty$), then $F_{K_\infty}(\cdot)$ is referred to as the (non-degenerate) *limiting distribution*.

Definition 3.3. Under the notations of Definition 3.2,

- Where K_∞ does not exist or exists but is degenerate, K_n will be said *not to possess a limiting distribution*.
- If there is no limiting distribution, but there exist deterministic sequences $a_n \in \mathbb{R}$ and $b_n > 0$, such that $(K_n - a_n)/b_n \xrightarrow{d} K_\infty^*$ as $n \to \infty$, and K_∞^* is non-degenerate, then K_n will be said to possess an *asymptotic distribution under positive affine normalization*.

In sufficiently large samples, it is common practice to approximate the critical values of K_n by their analogues from limiting (or asymptotic) distributions:[1] $C_{n,\alpha} \approx C_{\infty,\alpha}$ (respectively, $C_{n,\alpha} \approx b_n C^*_{\infty,\alpha} + a_n$), where $C_{n,\alpha} = F^{-1}_{K_n}(1-\alpha)$ and $C_{\infty,\alpha} = F^{-1}_{K_\infty}(1-\alpha)$ (respectively, $C^*_{\infty,\alpha} = F^{-1}_{K^*_\infty}(1-\alpha)$).

Therefore, limiting properties of distributions of GoF test statistics are of importance as they are instrumental in computing critical values.

The classic Cramér-von Mises test places equal weight on the main part and the tails of the distribution under H_0. To weight the deviations between the EDF and F_0 according to the importance attached to various portions of the CDF, let us consider weighted Cramér-von Mises test (WCvM) statistics:

$$W_n^2(\psi) = n\int_{-\infty}^{\infty} \psi(F_0(x))(F_n(x) - F_0(x))^2\, dF_0(x), \tag{3.2}$$

where weight $\psi : [0,1] \to \mathbb{R}_+$.

Notable examples include the following choices of the weighting function: $\psi(t) \equiv 1$ (Cramér, 1928; von Mises, 1928), $\psi(t) = 1/(t(1-t))$ (Anderson and Darling, 1952), $\psi(t) = 1/t$, $\psi(t) = 1/(1-t)$ (Sinclair et al., 1990; Scott, 1999), $\psi(t) = 1/t^\beta$ for $\beta < 2$ (Deheuvels and Martynov, 2003), and $\psi(t) = 1/(1-t)^\beta$ for $\beta < 2$ (Feuerverger, 2015).

In a series of works (see Chernobai et al., 2005; Chernobai, 2006; Luceño, 2006; Chernobai et al., 2015), the following test statistics have been proposed:

$$AD^* = \sqrt{n}\sup_{x\in\mathbb{R}} \frac{|F_n(x) - F_0(x)|}{\sqrt{F_0(x)(1-F_0(x))}}, \tag{3.3}$$

$$AD_{up} = \sqrt{n}\sup_{x\in\mathbb{R}} \frac{|F_n(x) - F_0(x)|}{(1-F_0(x))}, \tag{3.4}$$

and

$$AD_{up}^2 = n\int_{-\infty}^{\infty} \frac{(F_n(x) - F_0(x))^2}{(1-F_0(x))^2}\, dF_0(x). \tag{3.5}$$

A related statistic that is useful will be denoted by AD^2_{down}:

$$AD_{down}^2 = n\int_{-\infty}^{\infty} \frac{(F_n(x) - F_0(x))^2}{(F_0(x))^2}\, dF_0(x). \tag{3.6}$$

[1] As $n \to \infty$, $F_{K_n}(t) \to F_{K_\infty}(t)$ at all continuity points of $F_{K_\infty}(\cdot)$ is equivalent to $F^{-1}_{K_n}(\alpha) \to F^{-1}_{K_\infty}(\alpha)$ at all continuity points of $F^{-1}_{K_\infty}(\cdot)$ (van der Vaart, 2007, p. 305).

Statistic AD^2_{up} is $W^2_n(\psi)$ with $\psi(t) = 1/(1-t)^2$. The AD^2_{up} test statistic is a popular (see Mignola and Ugoccioni, 2006; Turk, 2009; Wahlström, 2013; Lavaud and Lehérissé, 2014; Feuerverger, 2015; Chernobai et al., 2015) test statistic in OpRisk modeling for severity distribution selection. It is also known in engineering and meteorological applications under the name of the right-tail Anderson-Darling test of second degree (Zhou, 2013; Drobinski et al., 2015).

Until recently, to the best of our knowledge, the OpRisk literature has not questioned limiting properties of any of (3.3)-(3.6). In Footnote 2, Chernobai et al. (2015, p. 585), when briefly speaking about the mean of the asymptotic distribution of AD^2_{up}, takes for granted the very existence of such an asymptotic distribution. In fact, in the terminology of the above Definition 3.2, Chernobai et al. (2015) means the limiting distribution. In this thesis, however, we will show that no limiting distribution exists for the AD^2_{up} statistic.

Moreover, we discuss limiting properties for (3.3) and (3.4) and contribute to the analysis of the limiting properties of a subfamily of WCvM GoF test statistics (3.2), with weight function $\psi(t) = 1/(1-t)^\beta$, which will be denoted throughout by

$$W_n^{2,\beta} = n\int_{-\infty}^{\infty} \frac{(F_n(x)-F_0(x))^2}{(1-F_0(x))^\beta}\, dF_0(x),$$

for $\beta \in \mathbb{R}$. Statistics $W_n^{2,\beta}$ are suitable for more accurate selection of severity distributions due to the emphasis they put on the right tail of the distributions.

3.3 Facts from Theory of Empirical Processes

3.3.1 Preliminaries

Let $U_1,\ldots,U_n$ be independent $U(0,1)$ random variables with corresponding order statistics $U_{(1)} \leqslant \ldots \leqslant U_{(n)}$.

Definition. For $E_n(s) = \begin{cases} 0, & 0 \leqslant s < U_{(1)}, \\ k/n, & U_{(k)} \leqslant s < U_{(k+1)}, k = 1,..,n-1, \\ 1, & U_{(n)} \leqslant s \leqslant 1. \end{cases}$ the *uniform empirical process* is defined $e_n(s) = \sqrt{n}\,(E_n(s)-s)$, $0 \leqslant s \leqslant 1$.

As shown in Shorack and Wellner (2009, pp. 3-5) and Csörgő and Horváth (1993, pp. 365-367), for a continuous $F_0(\cdot)$, the probabilistic study of

$$\sqrt{n}\left(F_n(x) - F_0(x)\right), x \in \mathbb{R},$$

is equivalent to that of $e_n(F_0(x))$, $x \in \mathbb{R}$ under the null hypothesis (3.1) as in this case $U = F_0(X)$ is $U(0,1)$-distributed.

It is known (see Csörgő and Horváth, 1993, formula (5.1.65)) that

$$\{e_n(t)\psi(1-t), t \in (0,1)\} \stackrel{d}{=} \{e_n(1-t)\psi(1-t), t \in (0,1)\}$$

for each $n \geqslant 1$, where $\stackrel{d}{=}$ denotes equality in distribution.

Hence, in particular,

$$\int_0^1 \left(e_n^2(t)/t^2\right) dt \stackrel{d}{=} \int_0^1 \left(e_n^2(t)/(1-t)^2\right) dt$$

and

$$\int_0^1 \left(e_n^2(t)/t\right) dt \stackrel{d}{=} \int_0^1 \left(e_n^2(t)/(1-t)\right) dt.$$

Thus, $AD^2_{down} \stackrel{d}{=} AD^2_{up}$.

A standard Brownian motion will be denoted by $W = \{W(t) \mid W(0) = 0, t \geqslant 0\}$ with state space $\mathbb{R}$ and standard Brownian Bridge $B \stackrel{d}{=} W(t) - tW(1)$ for $t \in [0,1]$.

Of key interest is an adaptation of Theorem 1.1 from Csörgő et al. (1993):

Proposition 3.4. $W_n^2(\psi) \xrightarrow{d} \int_0^1 \psi(t) B^2(t)\, dt$ as $n \to \infty$ if and only if

$$\int_0^1 \psi(t)\, t\,(1-t)\, dt < \infty.$$

It follows that not all weights are suitable to produce a well-defined limiting distribution for the corresponding test statistic.

3.3.2 Classical Results

Throughout, the limits are taken as $n \to \infty$ unless otherwise stated. We refer to known and proven results as "Propositions" and to important new results emerging from this work as "Theorems".

Proposition 3.5. (i) $\sup\limits_{0\leqslant t\leqslant 1}\frac{|e_n(t)|}{(1-t)^\beta}\xrightarrow{d}\infty$ for any $\beta\geqslant 1/2$;

(ii) $\sup\limits_{0\leqslant t\leqslant 1}\frac{|e_n(t)|}{\sqrt{t(1-t)}}\xrightarrow{d}\infty$.

Proof. Part (i) follows from Mason (1985) and Part (ii) from Chibisov (1964, Theorem 2). ■

Corollary 3.6. Test statistics AD^* (3.3) and AD_{up} (3.4) do not have well-defined limiting distributions.

Proposition 3.7. (i) $\int_0^1 \frac{e_n^2(t)}{(1-t)^\beta}\,dt \xrightarrow{d} \int_0^1 \frac{B^2(t)}{(1-t)^\beta}\,dt$ if and only if $\beta<2$;

(ii) $\int_0^1 \frac{e_n^2(t)}{(1-t)^\beta}\,dt \xrightarrow{d}\infty$ for any $\beta\geqslant 2$.

Proof. Part (i) follows from Proposition 3.4, while Part (ii) is a consequence of Csörgő et al. (1993, Theorem 1.3) or, for $\beta=2$, from the results from Shepp (1966, p. 353). ■

Corollary 3.8. Test statistics AD^2_{up} (3.5) and AD^2_{down} (3.6) do not have well-defined limiting distributions.

As discussed in Csörgő and Horváth (1988) and Csörgő and Horváth (1993), $\psi(t)=1/\sqrt{t(1-t)}$ and $\psi(t)=1/(t(1-t))^2$ are special weight functions in the sense that they separate light weights from heavy ones in the supremum and integral class, respectively. In Chapter 5 of Csörgő and Horváth (1993) it is emphasized that asymptotic behaviour of a weighted EDF-based test statistic will be determined by that of the uniform empirical process $e_n(t)$ on a subinterval $I_n\subseteq[0,1]$ for t: in the tails, if I_n is either $(0,a_n]$ or $[1-a_n,1)$, or in the middle, i.e., if I_n is $[a_n,1-a_n]$, where $a_n\downarrow 0$.

Propositions 3.9 and 3.10 make this specific for weight function $\psi(t)=1/(1-t)^\beta$.

Proposition 3.9. For any sequence of real numbers $\{a_n\}_{n=1}^\infty$ such that $a_n\to 0$ and $na_n\to\infty$, we have:

(i) If $-\infty<\beta<1/2$, then

$$a_n^{\beta-\frac{1}{2}}\sup_{1-a_n\leqslant t<1}\frac{|e_n(t)|}{(1-t)^\beta}\xrightarrow{d}\sup_{t\in[0,1]}\frac{|W(t)|}{t^\beta};$$

(ii) If $\beta = 1/2$, then

$$A\left(\frac{1}{2}\ln(na_n)\right)\sup_{1-a_n\leqslant t<1}\frac{|e_n(t)|}{\sqrt{t(1-t)}} - D\left(\frac{1}{2}\ln(na_n)\right) \xrightarrow{d} Y,$$

where $A(x) = \sqrt{2\ln x}$, $D(x) = 2\ln x + (\ln\ln x)/2 - (\ln\pi)/2$ and Y is a Gumbel random variable, whose PDF is $\exp(-2\exp(-x))$, $x \in \mathbb{R}$;

(iii) If $\beta > 1/2$, then

$$n^{\frac{1}{2}-\beta}\sup_{1-a_n\leqslant t<1}\frac{|e_n(t)|}{(1-t)^{\beta}} \xrightarrow{d} \sup_{t\in(0,\infty)}\frac{|N(t)-t|}{t^{\beta}},$$

where $\{N(t), t \geqslant 0\}$ is a Poisson process with $\mathbb{E}[N(t)] = t$;

(iv) If $\beta > 1/2$, then

$$n^{\frac{1}{2}-\beta}\sup_{1-a_n\leqslant t<U_{(n)}}\frac{|e_n(t)|}{(1-t)^{\beta}} \xrightarrow{d} \sup_{t\in[S(1),\infty)}\frac{|N(t)-t|}{t^{\beta}},$$

where $S(1)$ is the time of the first jump of $N(t)$.

Proof. See in Csörgő and Horváth (1993, p. 265, Theorem 1.2). ∎

Remark. It is known (Mason, 1985) that

$$\sup_{t\in(0,\infty)}\frac{N(t)-t}{t} \stackrel{d}{=} X,$$

where X has CDF $P(X \leqslant x) = x/(1+x)$ for $x \geqslant 0$ and 0 otherwise. This is a Lomax distribution (Johnson et al., 1994).

Proposition 3.10. For any sequence of real numbers $\{a_n\}_{n=1}^{\infty}$ such that $a_n \to 0$ and $na_n \to \infty$, we have

(i) If $-\infty < \beta < 2$, then

$$a_n^{\beta-2}\int_{1-a_n}^{1}\frac{e_n^2(t)}{(1-t)^{\beta}}\,dt \xrightarrow{d} \int_0^1 \frac{W^2(t)}{t^{\beta}}\,dt;$$

(ii) If $\beta = 2$, then

$$\frac{1}{2\sqrt{\ln(na_n)}}\left(\int_{1-a_n}^{1}\frac{e_n^2(t)}{t^2(1-t)^2}\,dt - \ln(na_n)\right) \xrightarrow{d} N(0,1)$$

and, as $n \to \infty$,

$$\frac{1}{2\sqrt{2\ln n}}\left(\int_{0}^{1}\frac{e_n^2(t)}{t^2(1-t)^2}\,dt - 2\ln n\right) \xrightarrow{d} N(0,1);$$

(iii) If $\beta > 2$, then

$$n^{2-\beta}\int_{1-a_n}^{1}\frac{e_n^2(t)}{(1-t)^{\beta}}\,dt \xrightarrow{d} \int_{0}^{\infty}\frac{(N(t)-t)^2}{t^{\beta}}\,dt.$$

Proof. The proof follows from Csörgő and Horváth (1993, p. 325, Theorem 3.2) and Csörgő and Horváth (1993, p. 335, Equation (5.3.113)). ■

3.4 Facts from Spectral Theory

In this section, we define the spectrum of an integral operator and present some known results related to the integral operators corresponding to the $W_n^{2,\beta}$ test statistics

Definition. (Reed and Simon, 1980, p. 188)

1. Let T be a bounded linear operator from a Banach space X to itself. A complex number λ is said to be in the *resolvent set* $\rho(T)$ of T if $T - \lambda I$ is a bijection with a bounded inverse. $R_\lambda(T) = (T - \lambda I)^{-1}$ is called the resolvent of T at λ. If $\lambda \notin \rho(T)$, then λ is said to be in the spectrum $\sigma(T)$ of T.

2. An $x \neq 0$ which satisfies $Tx = \lambda x$ for some $\lambda \in \mathbb{C}$ is called an *eigenvector* of T; λ is called the corresponding *eigenvalue*.

Definition. (Pugachev and Sinitsyn, 1999, p. 419) If $R_T(\lambda)$ is the range of $T - \lambda I$,

1. A set of eigenvalues of the operator T is called a *point spectrum* of T and is denoted by $\sigma_p(T)$. The resolvent does not exist for $\lambda \in \sigma_p(T)$.

2. A subset $\sigma_c(T)$ of $\sigma(T)$ for which the resolvent exists and its domain is dense in X, $R_T(\lambda) \neq X$, $\overline{R_T(\lambda)} = X$, is called a *continuous spectrum* of T.

3. A subset $\sigma_r(T)$ of $\sigma(T)$ for which the resolvent exists and its domain is not dense in X, $\overline{R_T(\lambda)} \neq X$, is called the *residual spectrum* of T.

Thus, $\sigma(T) = \sigma_p(T) \sqcup \sigma_c(T) \sqcup \sigma_r(T)$, where $\sqcup$ denotes a disjoint union of sets. The spectrum of a bounded linear operator is a bounded closed set (Pugachev and Sinitsyn, 1999, p. 423). If X, being a Banach space, is also Hilbert space (as in the case of $X = L^2(S)$, $S \subseteq \mathbb{R}$, of square integrable functions) and if T is self-adjoint, then $\sigma(T) \subset \mathbb{R}$ and $\sigma_r(T) = \varnothing$ (Pugachev and Sinitsyn, 1999, pp. 451-452).

If $\int_0^1 \psi(t)\, t\,(1-t)\, dt < \infty$, an application of Mercer's theorem and Karhunen-Loève theorem leads to the classic representation (Shorack and Wellner, 2009, pp. 201-213): $\int_0^1 \psi(t)\, B^2(t)\, dt \overset{d}{=} \sum\limits_{k=1}^{\infty} \lambda_k X_k$, where X_k are independent random variables distributed as χ_1^2, i.e., chi-squared distribution with one degree of freedom.

Here, λ_k are eigenvalues of the following integral operator acting in $L^2((0,1))$:

$$(Kg)(x) = \int_0^1 k(x,y) g(y) dy,\, x \in (0,1),\, g \in L^2((0,1)), \tag{3.7}$$

with kernel $k(x,y) = (\min(x,y) - xy)\sqrt{\psi(x)\psi(y)}$. That is, λ_k solve $(Kg)(x) = \lambda g(x)$.

For $\psi(t) = 1/(1-t)^{\beta}$, $\beta < 2$, Deheuvels and Martynov (2003) demonstrates that $\lambda_k = \left(2\nu/z_{\nu,k}\right)^2$, where $z_{\nu,k}$ is the k-th positive zero of the Bessel function $J_\nu(\cdot)$ of the first kind and $\nu = 1/(2-\beta)$.

In other words, the spectrum of integral operator (3.7) for $\psi(t) = 1/(1-t)^{\beta}$ with $\beta < 2$ contains eigenvalues. In contrast, in Section 3.6, we show that for $\beta = 2$ the spectrum is purely continuous.

3.5 Ornstein-Uhlenbeck Stochastic Process

In this section, we define the Ornstein-Uhlenbeck stochastic process. Some properties of this process will be used in the sequel.

We borrow from Section 1.1 of Fatalov (2014) the definition of the process.

Definition. The *Ornstein-Uhlenbeck ("OU") process* $\{\varsigma_\theta(t), t \geqslant 0\}$ of order $\theta > 0$ is a Gaussian homogeneous diffusion process satisfying the stochastic differential equation

$$d\varsigma_\theta(t) = -\theta \varsigma_\theta(t)\, dt + dW(t), \tag{3.8}$$

where $W(t)$, $W(0) = 0$, is the standard Brownian motion.

The state space for $\{\varsigma_\theta(t), t \geqslant 0\}$ is the whole real axis, and the rate measure is the one-dimensional Gaussian probability measure

$$m_\theta(dx) = \sqrt{\theta/\pi} \exp\left(-\theta x^2\right) dx, x \in \mathbb{R}.$$

There are two types of OU processes: stationary and nonstationary. Both of them satisfy (3.8) but with different initial distributions.

Definition. The *stationary OU process* is the process

$$Y_\theta(t) = \left\{\varsigma_\theta(t) \mid \varsigma_\theta(t) \stackrel{d}{=} m_\theta\right\}$$

for $\theta > 0$ and $t \geqslant 0$.

Definition. The *nonstationary OU process* is the process

$$X_{\theta,x}(t) = \{\varsigma_\theta(t) \mid \varsigma_\theta(0) = x\}$$

for $\theta > 0$ and $t \geqslant 0$ and some $x \in \mathbb{R}$.

Both $\{Y_\theta(t), t \geqslant 0\}$ and $\{X_{\theta,x}(t), t \geqslant 0\}$ are Markov processes while $\{Y_\theta(t), t \geqslant 0\}$ is also ergodic (Pavliotis, 2014; Iakus, 2008) with invariant measure $m_\theta(dx)$. The stationary OU process has mean zero and covariance function

$$\mathbb{E}\left[Y_\theta(t)Y_\theta(s)\right] = \frac{1}{2\theta} \exp\left(-\theta\,|t-s|\right)$$

for $t, s \geqslant 0$. On the other hand, the mean and covariance function of the nonstationary OU process are $\mathbb{E}\left[X_{\theta,x}(t)\right] = x\exp(-\theta t)$ and

$$cov\left(X_{\theta,x}(t), X_{\theta,x}(s)\right) = \frac{1}{2\theta}\left(\exp\left(-\theta\,|t-s|\right) - \exp\left(-\theta\,(t+s)\right)\right)$$

for $t, s \geqslant 0$. The stationary and nonstationary processes admit the following representations (Iakus, 2008, p. 45):

$$Y_\theta(t) = \frac{1}{\sqrt{2\theta}} \exp(-\theta t) W(\exp(2\theta t))$$

and

$$X_{\theta,x}(t) = x \exp(-\theta t) + \frac{1}{\sqrt{2\theta}} \exp(-\theta t) W(\exp(2\theta t) - 1) \tag{3.9}$$

for $t \geqslant 0$.

Central to our discussion will be the functional $Z_{\theta,x}(T) = \int_0^T X^2_{\theta,x}(t)\,dt$ for $0 \leqslant T < \infty$. For convenience, the zero-start $\{X_{\theta,0}(t), t \geqslant 0\}$ and corresponding $Z_{\theta,0}(T)$ will be denoted by $\{X_\theta(t), t \geqslant 0\}$ and $Z_\theta(T)$, respectively.

3.6 Main Results

In this section, we establish the following main theorems.

Theorem 3.11. We have

(i) Statistics AD^2_{up} and AD^2_{down} possess asymptotic distributions under positive affine normalization:

$$\frac{AD^2_{up} - \ln n}{2\sqrt{\ln n}} \xrightarrow{d} N(0,1) \text{ and } \frac{AD^2_{down} - \ln n}{2\sqrt{\ln n}} \xrightarrow{d} N(0,1);$$

(ii) For all real x and y

$$P\left(\frac{AD^2_{up} - \ln n}{2\sqrt{\ln n}} \leqslant x, \frac{AD^2_{down} - \ln n}{2\sqrt{\ln n}} \leqslant y\right) \to \Phi(x)\Phi(y).$$

That is, the random variables

$$\frac{AD^2_{up} - \ln n}{2\sqrt{\ln n}} \text{ and } \frac{AD^2_{down} - \ln n}{2\sqrt{\ln n}}$$

are asymptotically independent.

Corollary 3.12. We have

$$\frac{\sqrt{\ln n}}{2}\ln\left(\frac{AD_{up}^2}{\ln n}\right)\xrightarrow{d} N(0,1) \text{ and } \frac{\sqrt{\ln n}}{2}\ln\left(\frac{AD_{down}^2}{\ln n}\right)\xrightarrow{d} N(0,1).$$

Corollary 3.13. We have

$$\frac{AD_{up}^2-\ln n}{2\sqrt{\ln n}}+\frac{AD_{down}^2-\ln n}{2\sqrt{\ln n}}\xrightarrow{d}\sqrt{2}N(0,1).$$

Theorem 3.14. Statistic $W_n^{2,2}$ admits the decomposition $W_n^{2,2}=A_n+B_n$, such that

$$\frac{A_n+2n}{2\sqrt{n}}\xrightarrow{d} -X \text{ and } \frac{B_n-2n}{2\sqrt{n}}\xrightarrow{d} X,$$

with X being distributed as the standard Gaussian distribution.

Theorem 3.15. Let $k(x,y)=(\min(x,y)-xy)/((1-x)(1-y))$ for x and $y\in[0,1)$. Define the integral operator K with kernel $k(\cdot,\cdot)$ as $(Kg)(x)=\int_0^1 k(x,y)g(y)dy$ for $g\in L^2([0,1))$ and $x\in[0,1)$. Then, the spectrum of this operator is purely continuous filling in the interval $[0,4]$.

3.7 Proofs of Main Results

The proofs presented below rely on some auxiliary lemmas. In order not to interrupt the flow of the exposition, these lemmas are relegated to Appendix B.

Proof of Theorem 3.11. Part (i). First, let us observe that

$$\int_0^1\frac{e_n^2(t)}{(1-t)^2}dt=\int_0^{U_{(n)}}\frac{e_n^2(t)-B_n^2(t)}{(1-t)^2}dt+\int_0^{U_{(n)}}\frac{B_n^2(t)}{(1-t)^2}dt+\int_{U_{(n)}}^1\frac{e_n^2(t)}{(1-t)^2}dt,$$

where $B_n=B_n(s)$, $s\in[0,1]$, is the sequence of Brownian bridges from Csörgő and Horváth (1988, Theorem 3.1).

In Csörgő and Horváth (1988), it is shown that $\int_{U_{(1)}}^{U_{(n)}}\frac{\left|e_n^2(t)-B_n^2(t)\right|}{t^2(1-t)^2}dt=O_P(1)$. Since, for $t\in(0,1)$, $1/(1-s)^2<1/\left(s^2(1-s)^2\right)$, then $\int_{U_{(1)}}^{U_{(n)}}\frac{\left|e_n^2(t)-B_n^2(t)\right|}{(1-t)^2}dt=O_P(1)$. It follows then Lemma B.3 that $\int_{U_{(n)}}^1\frac{e_n^2(t)}{(1-t)^2}dt=O_P(1)$.

As in Lemma B.4, on account that $U_{(1)} \xrightarrow{P} 0$, we conclude that $\int_0^{U_{(1)}} \frac{e_n^2(t)}{(1-t)^2}\,dt = o_P(1)$. It is well known (Shepp, 1966; Csörgő et al., 1993) that $\int_0^1 B_n^2(t)\,dt < \infty$ wp1. Consequently, as in Lemma B.4, $\int_0^{U_{(1)}} \frac{B_n^2(t)}{(1-t)^2}\,dt = o_P(1)$.

Then, $\int_0^{U_{(n)}} \frac{e_n^2(t)-B_n^2(t)}{(1-t)^2}\,dt \leqslant \int_{U_{(1)}}^{U_{(n)}} \frac{\left|e_n^2(t)-B_n^2(t)\right|}{(1-t)^2}\,dt + \int_0^{U_{(1)}} \frac{e_n^2(t)}{(1-t)^2}\,dt + \int_0^{U_{(1)}} \frac{B_n^2(t)}{(1-t)^2}\,dt = O_P(1)$, so that $\int_0^{U_{(n)}} \frac{e_n^2(t)-B_n^2(t)}{(1-t)^2}\,dt = O_P(1)$.

Observe that

$$B(t) \overset{d}{=} (1-t)W\left(\frac{t}{1-t}\right).$$

Then for all $0 < b < 1$,

$$\int_0^b \frac{B^2(t)}{(1-t)^2}\,dt \overset{d}{=} \int_0^b W^2\left(\frac{t}{1-t}\right)dt.$$

The change of variable $t/(1-t) = \exp(s) - 1$ leads to

$$\int_0^b \frac{B^2(t)}{(1-t)^2}\,dt \overset{d}{=} \int_0^{-\ln(1-b)} \left(\exp\left(-\frac{s}{2}\right) W\left(\exp(s)-1\right)\right)^2 ds.$$

From representation (3.9), we recognize that

$$\int_0^b \frac{B^2(t)}{(1-t)^2}\,dt \overset{d}{=} \int_0^{-\ln(1-b)} X_{\frac{1}{2}}^2(t)\,dt \tag{3.10}$$

Now the proof proceeds analogous to the proof of Corollary 2.2 of Csörgő and Horváth (1988), but with the use of Lemma B.2 to conclude that as $n \to \infty$

$$\frac{1}{2\sqrt{\ln(n)}}\left(\int_0^1 \frac{e_n^2(t)}{(1-t)^2}\,dt - \ln n\right) \xrightarrow{d} N(0,1).$$

Since $AD_{up}^2 \overset{d}{=} AD_{down}^2$ (see Section 3.3), the proof of Part (i) is complete.

Part (ii). Let $k(n) = n/\ln^2 n$. It is known (Csörgő and Horváth, 1993, p. 335) that

$$\int_{U_{(k(n))}}^{U_{(n-k(n))}} \frac{e_n^2(t)}{t^2(1-t)^2}\,dt = O_P(\ln\ln n)$$

This implies that both

$$\int_{U_{(k(n))}}^{U_{(n-k(n))}} \frac{e_n^2(t)}{t^2}\,dt = O_P(\ln\ln n) \text{ and } \int_{U_{(k(n))}}^{U_{(n-k(n))}} \frac{e_n^2(t)}{(1-t)^2}\,dt = O_P(\ln\ln n). \quad (3.11)$$

From (3.11), Lemma B.4, and the results obtained above in proving Part (i) we conclude that the following representations are valid as $n \to \infty$:

$$AD_{down}^2 = \int_{U_{(1)}}^{U_{(k(n))}} \frac{e_n^2(t)}{t^2}\,dt + O_P(\ln\ln n)$$

and

$$AD_{up}^2 = \int_{U_{(n-k(n))}}^{U_{(n)}} \frac{e_n^2(t)}{(1-t)^2}\,dt + O_P(\ln\ln n).$$

By Csörgő and Horváth (1993, p. 272, Lemma 1.4), the random variables

$$\frac{\left(\int_{U_{(1)}}^{U_{(k(n))}} \frac{e_n^2(t)}{t^2}\,dt - \ln n\right)}{2\sqrt{\ln n}} \text{ and } \frac{\left(\int_{U_{(n-k(n))}}^{U_{(n)}} \frac{e_n^2(t)}{(1-t)^2}\,dt - \ln n\right)}{2\sqrt{\ln n}}$$

are asymptotically independent. This completes the proof of Theorem 3.11.

■

From (B.1), it follows[2] that $\lim_{T\to+\infty} \mathbb{E}\left[\exp\left(-Z_\theta\left(T\right)\right)\right] = 0$ for all $\theta > 0$. Consequently, $P\left(\int_0^\infty X_\theta^2(t)\,dt = +\infty\right) = 1$ (see Liptser and Shiryaev, 2001, p. 234, Equation (17.57)). This, together with (3.10), gives another proof of Part (ii) in Proposition 3.7 for $\beta = 2$.

Although the above proof is similar to that of Corollary 2.2 in Csörgő and Horváth (1988), there is a substantial difference. In Csörgő and Horváth (1988), the OU $\{\xi(t), t \geqslant 0\}$ process (B.5) is stationary and hence

$$\int_a^b \xi^2(t)\,dt \stackrel{d}{=} \int_0^{b-a} \xi^2(t)\,dt$$

for all $b \geqslant a \geqslant 0$. However, the zero-start OU $\{X_{1/2}(t), t \geqslant 0\}$ is nonstationary, and

[2]In the proof of Lemma B.1, we show that the moment generating function, Laplace transform, and characteristic function of $Z_\theta(T)$ can be obtained from each other by a corresponding change in the independent variable.

it is readily checked that

$$\int_a^b X_{\frac{1}{2}}^2(t)\,dt \overset{d}{\neq} \int_0^{b-a} X_{\frac{1}{2}}^2(t)\,dt$$

in general (clearly, the equality trivially holds when either $a = 0$ or $a = b$). It is sufficient to show that $\mathbb{E}\left[\int_a^b X_{1/2}^2(t)\,dt\right] \neq \mathbb{E}\left[\int_0^{b-a} X_{1/2}^2(t)\,dt\right]$. We have

$$\mathbb{E}\left[\int_a^b X_{\frac{1}{2}}^2(t)\,dt\right] = b - a + \exp(-b) - \exp(-a)$$

and

$$\mathbb{E}\left[\int_0^{b-a} X_{\frac{1}{2}}^2(t)\,dt\right] = b - a + \exp(a-b) - 1.$$

So

$$\mathbb{E}\left[\int_a^b X_{\frac{1}{2}}^2(t)\,dt\right] - \mathbb{E}\left[\int_0^{b-a} X_{\frac{1}{2}}^2(t)\,dt\right] = (\exp(a)-1)(\exp(-a)-\exp(-b)),$$

which equals 0 if and only if $a = 0$ or $a = b$.

Proof of Corollary 3.12. As $AD_{up}^2 \overset{d}{=} AD_{down}^2$, we shall only prove the statement for AD_{up}^2. Let us observe that $\left(AD_{up}^2 - \ln n\right) / \left(2\sqrt{\ln n}\right) \equiv \left(\sqrt{\ln n}/2\right)\left(AD_{up}^2/\ln n - 1\right)$. Since $y = \ln x$ is differentiable at $x = 1$ and $(\ln x)'|_{x=1} = 1$ is not zero, the Delta Method (van der Vaart, 2007, p. 26, Theorem 3.1) yields the result. ■

Remark. 1. This corollary can be proved directly by resorting to the fact that, as $T \to \infty$, the CDF and PDF of $\int_0^T X_{1/2}^2(t)\,dt$ are equivalent to those of an inverse Gaussian random variable (Dankel, 1991) whose PDF is

$$f(x;\mu,\lambda) = \sqrt{\lambda/(2\pi x^3)}\exp\left(-\lambda(x-\mu)^2/(2\mu^2 x)\right)$$

with parameters $\mu = T$ and $\lambda = T^2/4$ for positive T and x. Then, the statement of the corollary follows from the following property of an inverse Gaussian distribution given in Jørgenssen (1982, p. 23): $\sqrt{\varphi}\ln(X/\mu) \xrightarrow{d} N(0,1)$ as $\varphi \to \infty$ (for $\varphi = \lambda/\mu$ and any $\mu > 0$) where $\varphi = T/4$. As $T = \ln n$ in the case under consideration, the result follows.

2. The main result of Whitmore and Yalovsky (1978) $1/(2\sqrt{\varphi})+\sqrt{\varphi}\ln(X/\mu) \xrightarrow{d} N(0,1)$, as $\varphi \to \infty$, leads to a slightly different normalization:

$$\frac{\sqrt{\ln n}}{2}\ln\left(\frac{AD_{up}^2}{\ln n}\right)+\frac{1}{\sqrt{\ln n}} \xrightarrow{d} N(0,1). \tag{3.12}$$

   This normalization immediately follows from Corollary 3.12 and Slutsky's Theorem. The reader is referred to Jørgenssen (1982, p. 24) for a discussion of the nature of the additive term $1/(2\sqrt{\varphi}) = 1/\sqrt{\ln n}$.

3. Whitmore and Yalovsky (1978) suggests that for the normalizations from Theorem 3.11 and normalization (3.12) to adequately approximate $N(0,1)$, φ must be much greater than 1000 and 10, respectively. This is equivalent to requiring that sample size should exceed $e^{4000} \approx 1.5\times 10^{1737}$ and $e^{40} \approx 2.4\times 10^{17}$.

Proof of Corollary 3.13. The statement follows from the Continuous Mapping Theorem (van der Vaart, 2007, p. 7, Theorem 2.3). ■

Luceño (2006) introduced the statistic $a_n^2 = l_n^2 + r_n^2$, where, in the notation of this paper, $l_n^2 = AD_{down}^2$ and $r_n^2 = AD_{up}^2$. The above results yield that a_n^2 has no well-defined limiting distribution but (cf. (ii) in Proposition 3.10 and Corollary 3.13)

$$\frac{1}{2\sqrt{2\ln(n)}}\left(a_n^2 - 2\ln n\right) \xrightarrow{d} N(0,1).$$

Proof of Theorem 3.14. From the proof of Theorem 3.11, we deduce that

$$\frac{W_n^{2,2}}{2\sqrt{n}} = o_p(1) + \left(\frac{\ln n}{2\sqrt{n}}\right)\left(\frac{1}{\ln n}\int_0^{\ln n} X_{\frac{1}{2}}^2(t)\,dt\right).$$

Therefore, from Lemma B.5 and Slutsky's Theorem, it follows that

$$\frac{W_n^{2,2}}{2\sqrt{n}} \xrightarrow{P} 0. \tag{3.13}$$

In Chernobai et al. (2005), the following computational formula is obtained:

$$W_n^{2,2} = 2\sum_{j=1}^{n} \ln\left(1-U_{(j)}\right) + \frac{1}{n}\sum_{j=1}^{n} \frac{1+2(n-j)}{1-U_{(j)}}.$$

Since $\sum_{j=1}^{n} \ln\left(1-U_{(j)}\right) = \sum_{j=1}^{n} \ln\left(1-U_j\right)$, the latter decomposition can be rewritten as

$$W_n^{2,2} = 2\sum_{j=1}^{n} \ln\left(1-U_j\right) + \frac{1}{n}\sum_{j=1}^{n} \frac{1+2(n-j)}{1-U_{(j)}}.$$

Let

$$A_n = 2\sum_{j=1}^{n} \ln\left(1-U_j\right) \text{ and } B_n = \frac{1}{n}\sum_{j=1}^{n} \frac{1+2(n-j)}{1-U_{(j)}}.$$

It is a straightforward exercise to see that $-A_n$ is distributed as a random variable with χ^2_{2n} distribution, i.e., the chi-squared distribution with $2n$ degrees of freedom. It is well-known that $(\chi^2_m - m)/\sqrt{2m} \xrightarrow{d} N(0,1)$ as $m \to \infty$. Hence, as $n \to \infty$, we have

$$\frac{A_n + 2n}{2\sqrt{n}} \to -X,$$

where X is distributed as the standard Gaussian distribution. Hence, from (3.13) and Slutsky's Theorem, we conclude that as $n \to \infty$,

$$\frac{B_n - 2n}{2\sqrt{n}} \to X.$$

This completes the proof of Theorem 3.14. ∎

The fact that $(B_n - 2n)/(2\sqrt{n})$ converges in distribution to a standard Gaussian random variable can be obtained directly by noticing that

$$\frac{1}{n}\sum_{j=1}^{n} \frac{1+2(n-j)}{1-U_{(j)}} = O_P(1) + \frac{1}{n}\sum_{j=2}^{n-1} \frac{1+2(n-j)}{1-U_{(j)}}. \tag{3.14}$$

In (3.14), the term

$$T_n = \frac{1}{n}\sum_{j=2}^{n-1} \frac{1+2(n-j)}{1-U_{(j)}}$$

is an example of a trimmed L-statistic. Then, Mason and Shorack (1992, Theorem

1.1) implies that, as $n \to \infty$,

$$\frac{T_n - 2n}{2\sqrt{n}} \xrightarrow{d} N(0,1).$$

Before proceeding to the proof of Theorem 3.15, we make a few observations. In Deheuvels and Martynov (2003), it is demonstrated that eigenvalues of

$$\int_0^1 \frac{\min(x,y) - xy}{(1-x)^{\beta/2}(1-y)^{\beta/2}} \varphi(y)dy = \lambda \varphi(x), \, x \in [0,1), \, \varphi \in L^2([0,1)),$$

for $\beta < 2$ are $\lambda_k = \left(\frac{2\nu}{z_{\nu,k}}\right)^2$, $k \in \mathbb{N}$, where $z_{\nu,k}$ is the k-th positive zero of the Bessel function $J_\nu(\cdot)$ of the first kind and $\nu = 1/(2-\beta)$. In Elbert and Laforgia (1984), the following result of Tricomi is cited: $z_{\nu,k} = \nu + a_k \nu^{1/3} + O\left(\nu^{1/3}\right)$ as $\nu \to \infty$, where a_k is independent of ν. Hence $\lambda_k \to 4$ as $\beta \uparrow 2$. From Abramovitz and Stegun (1972), the following interlacing property of $z_{\nu,k}$ is known: $0 < z_{\nu,1} < z_{\nu+1,1} < z_{\nu,2} < z_{\nu+1,2} < z_{\nu,3} < \ldots$. Hence, this provides a heuristic evidence that λ_k gradually fill in $[0,4]$ as $\beta \uparrow 2$ as stated in Theorem 3.15.

Proof of Theorem 3.15. Let $k(x,y) = (\min(x,y) - xy)/((1-x)(1-y))$ for $x,y \in [0,1)$. Let us define the integral operator K with kernel $k(\cdot,\cdot)$ as

$$(Kg)(x) = \int_0^1 k(x,y)g(y)dy$$

for any $g \in L^2([0,1))$ and $x \in [0,1)$. In $(Kg)(x) = \lambda g(x)$, $x \in [0,1)$, we can make the following change of variables:

$$x = 1 - \exp(-t), \, y = 1 - \exp(-s) \text{ and } g^*(s) = \exp(-s/2)g(1 - \exp(-s)),$$

where $t,s \in [0,\infty)$. With the help of two identities,

$$\min(x,y) = \frac{1}{2}(x + y - |x-y|), \, x,y \in \mathbb{R},$$

and

$$\min(\exp(x), \exp(y)) = \exp(\min(x,y)), \, x,y \in \mathbb{R},$$

after some algebra, we arrive at the equation

$$\int_0^\infty \left(\exp\left(-\frac{|t-s|}{2} \right) - \exp\left(-\frac{t+s}{2} \right) \right) g^*(s)\, ds = \lambda g^*(t)$$

for $t \in [0,\infty)$. An application of Lemma B.6 for $\theta = 1/2$ completes the proof. ■

3.8 Practical Utility of $W_n^{2,\beta}$ Test Statistics

Commonly used GoF test statistics have been observed to possess the following properties, or stylized facts:

Property 1 The statistic has a finite mean and variance in finite samples;

Property 2 It possesses a non-degenerate limiting distribution;

Property 3 The limiting distribution has a finite mean and variance;

Property 4 The finite-sample distributions rapidly converge to the limiting distribution.

For example, the Kolmogorov-Smirnov, Cramér-von Mises $W_n^{2,0}$, Anderson-Darling and modified Anderson-Darling $W_n^{2,1}$ test statistics satisfy the properties (see Marsaglia et al., 2003; Csörgő and Faraway, 1996; Stephens, 1974; Sinclair et al., 1990) for practically important upper-tail probabilities (0.8 and higher).

One may argue that the existence of the first two moments may be of little to no importance in practical applications, where one is primarily concerned with computing a critical value being a quantile. However, for a nonnegative random variable the finiteness of its mean ensures that the random variable is well defined, i.e., finite wp1 (Gut, 2013, Theorem 4.4, p. 52).

A conventional way to determine a finite sample critical value is via Monte Carlo experiments. Although the processing power and available amounts of RAM of modern computers have significantly increased computational tractability of Monte Carlo methods, evaluation of critical values for sufficiently large samples still faces problems of long execution time. In these circumstances, it is common to approximate the critical value by an asymptotic one.

In this section, we investigate if the class of $W_n^{2,\beta}$ test statistics for $\beta > 1$ possess Properties 1-4.

Consider testing $H_0 : F(x) = F_0(x)$ vs. $H_1 : F(x) \neq F_0(x)$ for a continuous CDF $F_0(\cdot)$ with specified parameters. Given a sample of independent random variables $X_1, \dots, X_n$ distributed according to $F_0(\cdot)$. If $X_{(1)} \leqslant \dots \leqslant X_{(n)}$ are corresponding order statistics, let $z_j = F_0(X_{(j)})$.

In Feuerverger (2015), the following computational formula for $W_n^{2,\beta}$ is established:[3]

$$W_n^{2,\beta} = Const + \frac{2}{2-\beta}\sum_{j=1}^{n}(1-z_j)^{2-\beta} + \frac{1}{(\beta-1)n}\sum_{j=1}^{n}\frac{1+2(n-j)}{(1-z_j)^{\beta-1}}, \tag{3.15}$$

where $Const = n\left(\frac{1}{3-\beta} - \frac{2}{2-\beta} + \frac{1}{1-\beta}\right)$.

In Chernobai et al. (2005), the computational formula is given for $W_n^{2,2}$:

$$W_n^{2,2} = 2\sum_{j=1}^{n}\ln(1-z_j) + \frac{1}{n}\sum_{j=1}^{n}\frac{1+2(n-j)}{(1-z_j)}. \tag{3.16}$$

Observe that $W_n^{2,2}$ can be viewed as a limit of $W_n^{2,\beta}$ as $\beta \uparrow 2$ (cf. Feuerverger, 2015).

Under the assumption that $F_0(\cdot)$ is fully specified, z_j, $j = 1 \dots n$, will follow a uniform distribution in $[0,1]$. As such, $W_n^{2,\beta}$ is independent of the null distribution. After some straightforward algebra, it is not hard to determine the first two moments of $W_n^{2,\beta}$. Indeed, $\mathbb{E}\left[W_n^{2,\beta}\right] = 1/\left((2-\beta)(3-\beta)\right)$ for $\beta < 2$ and $+\infty$ otherwise. Similarly, $\mathbb{V}\left[W_n^{2,\beta}\right] = 2/\left((2-\beta)(5-2\beta)(3-\beta)^2\right)$ for $\beta < 2$ and $+\infty$ otherwise. In Deheuvels and Martynov (2003), it shown that the limiting distributions of $W_n^{2,\beta}$ ($\beta < 2$) have the same moments as their finite sample analogues. As mentioned above, this means that $W_n^{2,\beta}$ (for $\beta < 2$) and their limiting random variables are well defined.

Although $W_n^{2,2}$ has no moments, it is still well defined as demonstrated in Lemma 3.16 below.

Lemma 3.16. Under the null hypothesis, $P(W_n^{2,2} < \infty) = 1$.

[3]The expression in Feuerverger (2015) must be multiplied by n to yield the correct result (3.15).

Proof. Observe that
$$\bigcup_{m=0}^{\infty}\{W_n^{2,2}\leqslant m\}=\{W_n^{2,2}<\infty\}$$
for all $n\in\mathbb{N}$. Then (see Gut, 2013, Theorem 3.1, p. 11) $P(W_n^{2,2}\leqslant m)\to P(W_n^{2,2}<\infty)$ as $m\to\infty$. Clearly $P(W_n^{2,2}\leqslant 0)=0$. From (3.16) it is readily seen that $W_n^{2,2}\leqslant n/(1-z_n)$ for all $n\in\mathbb{N}$. Hence $P(W_n^{2,2}\leqslant m)\geqslant P(n/(1-z_n)\leqslant m)$. But $P(n/(1-z_n)\leqslant m)=(1-n/m)^n$ implies $P(n/(1-z_n)\leqslant m)\to 1$ as $m\to\infty$. This completes the proof. ■

The results of Section 3.6 imply that since no limiting distribution exists for $W_n^{2,2}$, one may speak of no moments of the limiting distribution whatsoever. Therefore, while $W_n^{2,\beta}$ ($\beta<2$) enjoys the first three GoF properties identified above, $W_n^{2,2}$ albeit well defined fails each of them.

To study convergence of the finite sample distributions of $W_n^{2,\beta}$ ($\beta<2$) and $(W_n^{2,2}-\ln n)/(2\sqrt{\ln n})$ to their limiting distributions, we considered testing $H_0: F(x)=\Phi(x)$ vs. $H_1: F(x)\neq\Phi(x)$ for the standard Gaussian CDF $\Phi(\cdot)$ at significance level $\alpha=0.05$.

For a range of sample size n from 10 to $50,000$ and various β from 1 to 2 with an increment of 0.05, we have tabulated[4] $W_n^{2,\beta}$ using (3.15)-(3.16) by extensive Monte Carlo simulations (10^6 trials for each sample size). We then computed critical values $C_{n,\alpha}^{2,\beta}$ for $W_n^{2,\beta}$. See Figure 3.1.

The asymptotic critical values[5] $C_{\infty,\alpha}^{2,\beta}$, as a function of β, are shown in Figure 3.2. It is readily seen that $C_{\infty,\alpha}^{2,\beta}$ explode as they approach $\beta=2$.

It is also informative to plot the finite-sample and asymptotic critical values together for the normalized version of the $W_n^{2,2}$ test statistic. It is seen that the finite-sample critical values $(C_{n,\alpha}^{2,2}-\ln n)/(2\sqrt{\ln n})$ converge extremely slowly to the asymptotic critical values $\Phi^{-1}(1-\alpha)$. See Figure 3.3 where the case of $\alpha=0.05$ is depicted.

We have considered a partition of the range of sample sizes n from 10 to $50,000$ into disjoint intervals $[n_1,n_2)$. We fitted a simple linear regression to the finite sample critical values $C_{n,\alpha}^{2,\beta}$: $C_{n,\alpha}^{2,\beta}=an+b+\varepsilon$ in each interval. Here ε is the error

[4]Note that (3.15) is not applicable for $\beta=1$. For this case, the computational formula is given in Sinclair et al. (1990).

[5]The computation is based on the limiting distribution established in Deheuvels and Martynov (2003, Formula (1.4), p. 67). Also, see Table 1.1 therein.

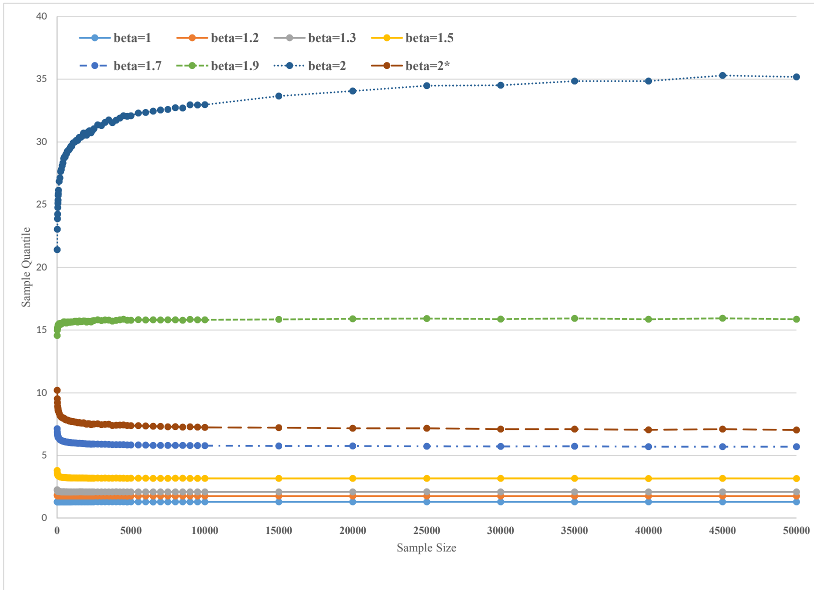

Figure 3.1: Critical values $C_{n,\alpha}^{2,\beta}$ for $W_n^{2,\beta}$ by Monte Carlo simulations for selected values of $\beta \in [1,2)$ and for selected sample sizes (10^6 trials for each sample size). Here $\alpha = 0.05$ and the 2.0* exponent refers to the normalized version of $W_n^{2,\beta}$.

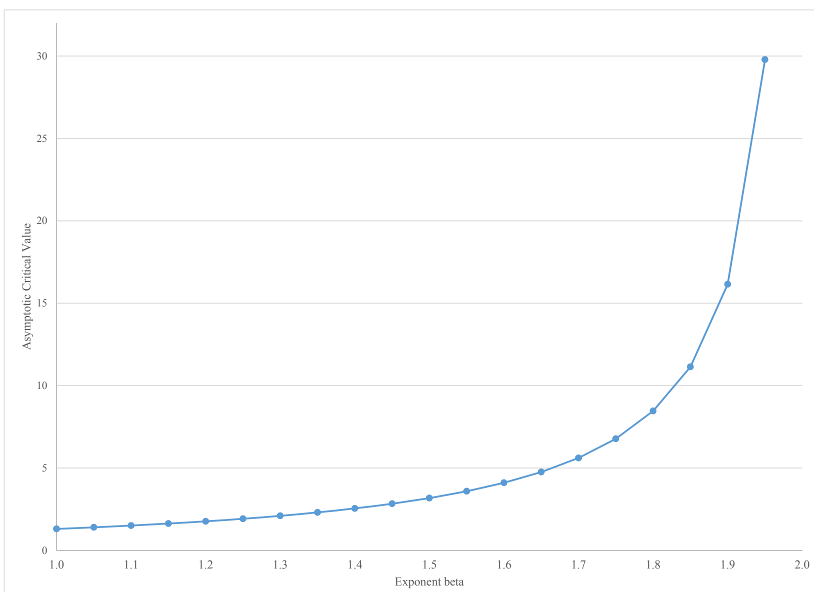

Figure 3.2: Asymptotic critical values $C_{\infty,\alpha}^{2,\beta}$ as a function of β. Here $\beta \in [1,2)$ and $\alpha = 0.05$.

term.

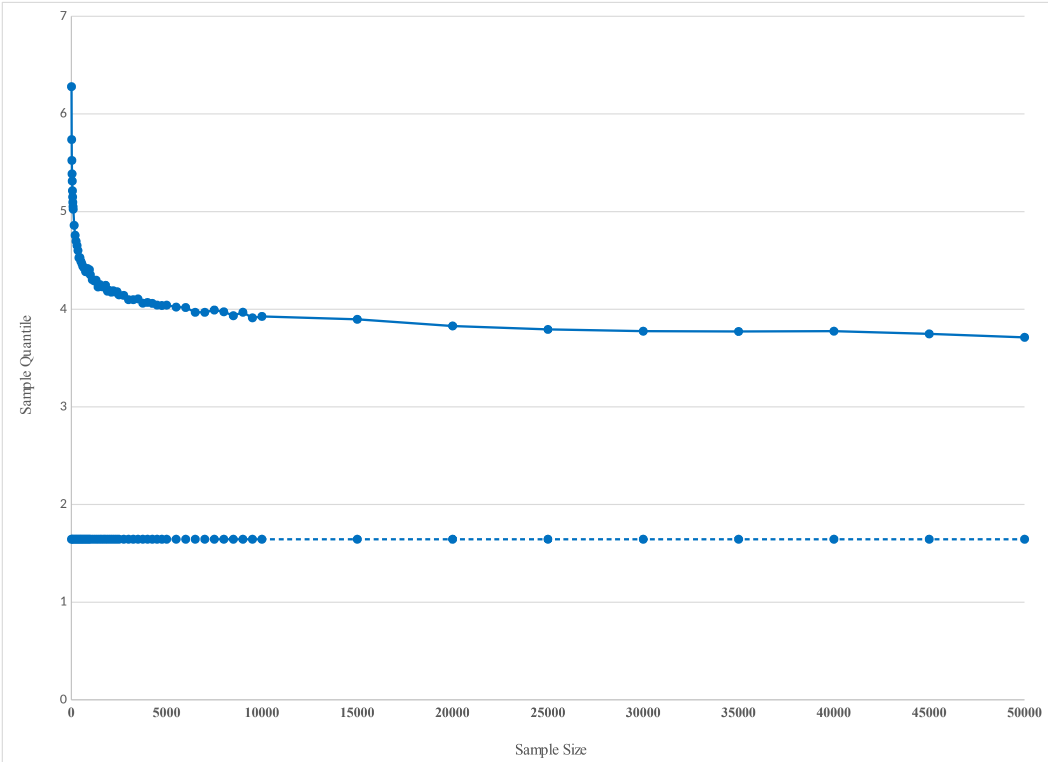

Figure 3.3: Monte Carlo finite-sample critical values and asymptotic critical value $\Phi^{-1}(1-\alpha)$ (straight line) for the normalized version of $W_n^{2,2}$ for $\alpha = 0.05$.

Then $H_0 : a = 0$ vs $H_1 : a \neq 0$ was assessed. The p-values are presented in Table 3.1 for $\alpha = 0.05$ and selected β values.

Table 3.1: p-values for testing $H_0 : a = 0$ vs $H_1 : a \neq 0$ at $\alpha = 0.05$ for $C_{n,\alpha}^{2,\beta} = an + b + \varepsilon$. Samples whose size n falls in $[n_1, n_2)$ are considered. The 2.0* exponent refers to the normalized version of $W_n^{2,\beta}$.

		β											
n_1	n_2	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0	2.0*
10	1200	0.39	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1200	8000	0.19	0.20	0.12	0.09	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00
8000	50000	0.66	0.56	0.31	0.21	0.15	0.01	0.00	0.00	0.00	0.00	0.00	0.00

It is noteworthy that in all cases but one the p-values behave as expected: increase rowwise and decrease columnwise. In practice, for a given event type, the size of available internal loss data of financial institutions typically does not exceed $1,000$ events while for external data the size does not exceed $50,000$ points. Thus,

Table 3.1 suggests that there is insufficient evidence that $C_{n,\alpha}^{2,\beta}$, $1.5 \leqslant \beta \leqslant 2$ and the normalized version $(C_{n,\alpha}^{2,2} - \ln n)/(2\sqrt{\ln n})$ flatten over the entire range of sample sizes important in OpRisk applications.

The overall quality of the approximation to normality provided by a normalization to $W_n^{2,2}$ can be conveniently assessed by the overlapping coefficient (OVL) introduced in Weitzman (1970). Given two PDFs $f_1(\cdot)$ and $f_2(\cdot)$, OVL is equal to the area under $f_1(\cdot)$ and $f_2(\cdot)$ simultaneously:

$$OVL = \int_{-\infty}^{\infty} \min\{f_1(x), f_2(x)\}\, dx, \text{ or } OVL = 1 - \frac{1}{2}\int_{-\infty}^{\infty} |f_1(x) - f_2(x)|\, dx.$$

OVL is a measure of agreement or similarity of the two PDFs. It takes on values in $[0,1]$, where 0 and 1 correspond to no and full overlap, respectively.

Figure 3.4 shows OVLs between the PDFs of three normalizations

$$(W_n^{2,2} - \ln n)/(2\sqrt{\ln n}),\ (\sqrt{\ln n}/2)\ln(W_n^{2,2}/\ln n),$$

$$(\sqrt{\ln n}/2)\ln(W_n^{2,2}/\ln n) + 1/\sqrt{\ln n} \text{ (see (3.12))}$$

and the standard Gaussian PDF $\varphi(x) = (1/\sqrt{2\pi})\exp(-x^2/2)$. The three PDFs converge to $\varphi(\cdot)$ (Whitmore and Yalovsky, 1978; Dankel, 1991). The OVLs were calculated nonparametrically as in Schmid and Schmidt (2006), where kernel density estimators of the PDFs were based on the Gaussian kernel and the Silverman rule-of-thumb bandwidth (see p. 10). This convergence appears to be rather slow for all the three normalizations. The second normalization seems to perform slightly better than the third one and much better than the first one. From Whitmore and Yalovsky (1978), we can infer that for the first and third normalizations to provide an adequate approximation to normality, it would require sample size of order $n \gg e^{4000} \approx 1.5 \times 10^{1737}$ and $n \gg e^{40} \approx 2.4 \times 10^{17}$, respectively; see p. 64.

As such, the knowledge of the theoretical limiting distribution is of little utility. Instead, a Monte Carlo method should be applied to determine critical values even for large sample sizes.

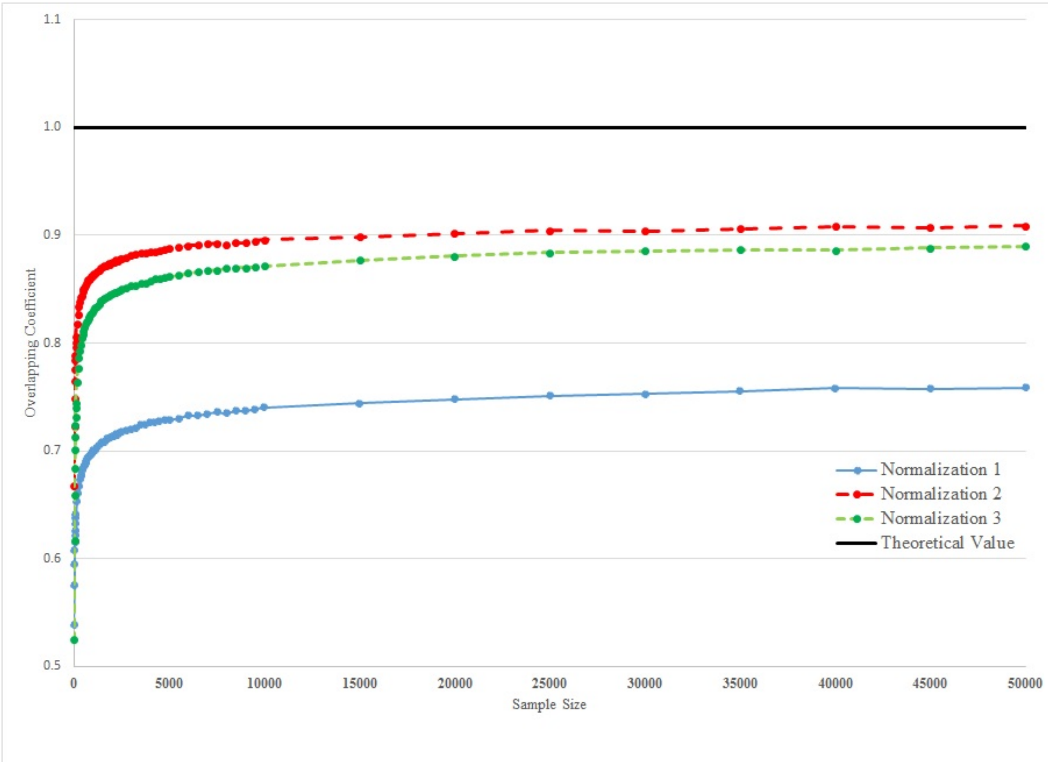

Figure 3.4: OVLs between the PDF of $N(0,1)$ and the normalizations $(W_n^{2,2} - \ln n)/(2\sqrt{\ln n})$, $(\sqrt{\ln n}/2)\ln(W_n^{2,2}/\ln n)$, $(\sqrt{\ln n}/2)\ln(W_n^{2,2}/\ln n) + 1/\sqrt{\ln n}$.

Chapter 4

On the Chen-Balakrishnan Transformation to Normality in Weighted Cramér-von Mises Goodness-of-Fit Tests

4.1 Introduction

Let $X = \{X_1, \ldots, X_n\}$ be a sample of IID random variables with common CDF $G(\cdot)$. In applications, tests of composite hypotheses

$$H_0 : G \in \mathscr{F} = \{F(\cdot;\boldsymbol{\theta}), \boldsymbol{\theta} \in \Omega \subseteq \mathbb{R}^m\} \tag{4.1}$$

for goodness of fit (GoF) are not uncommon. Here the distributional class $\mathscr{F}$ is given, but the parameter vector $\boldsymbol{\theta} \in \Omega \subseteq \mathbb{R}^m$ (with $m \in \mathbb{N}$ and Ω being an open set) is unknown and has to be estimated from a sample at hand.

In this case, one often has to resort to Monte Carlo (MC) simulations to calculate a p-value or critical value. In general, this may require significant and time consuming computational effort as each MC trial includes parameter estimation.

One possibility to overcome this is to transform H_0 to a hypothesis easier to deal with. The transformation suggested by Chen and Balakrishnan reduces the initial hypothesis H_0 to an auxiliary composite hypothesis for normality $H_0^\star$: $Y \sim N(\mu, \sigma)$. The transformation to and assessment of $H_0^\star$ are performed quite quickly.

The speed and applicability to arbitrary families $F(\cdot;\boldsymbol{\theta})$ of null distributions make the transformation attractive in practice.

In previous works, $H_0^\star$ was mainly studied on a standalone basis. In this chapter, for a selected family of weighted Cramér-von Mises (WCvM) GoF test statistics, we examine agreement between the decisions made based on H_0 (direct tests) and $H_0^\star$ (indirect tests) on the same data set for distributional families with a varying degree of tail heaviness. We compare the level and power of the selected WCvM GoF statistics for the direct and indirect tests. In many situations, decisions of direct and indirect tests become independent as sample size grows. We present some results on convergence in distribution. The work here is motivated by potential applications in risk modeling. In particular, we apply the tests on a real data set of operational risk losses.

4.2 Selected Weighted Cramér-von Mises GoF Tests

For the sake of convenience, in this section we will reproduce some key definitions from Chapter 3.

We consider GoF test statistics which belong to the quadratic class of statistics which are based on the empirical distribution function (EDF). EDF is defined as $F_n(t) = (1/n)\sum_{i=1}^{n} \mathbb{I}_{(-\infty,t]}(X_i)$, where $\{X_i\}_{i=1}^{n}$ is a random sample with common CDF $F(x;\boldsymbol{\theta})$ and $\mathbb{I}_A(\cdot)$ is the indicator function on event A. The prominent members of the quadratic class are the Cramér-von Mises (CvM) and Anderson-Darling (AD) test statistics which are defined in (4.2) and (4.3) below:

$$W_n^2 = n\int_{-\infty}^{\infty} \left(F_n(x) - F(x;\boldsymbol{\theta})\right)^2 dF(x;\boldsymbol{\theta}), \tag{4.2}$$

$$A_n^2 = n\int_{-\infty}^{\infty} \frac{\left(F_n(x) - F(x;\boldsymbol{\theta})\right)^2}{F(x;\boldsymbol{\theta})\left(1 - F(x;\boldsymbol{\theta})\right)} dF(x;\boldsymbol{\theta}). \tag{4.3}$$

The CvM test places equal weight on the main part and the tails of the distribution under H_0. On the other hand, AD, while emphasizing on the left and right tails, treats them equally. To weigh the deviations between the EDF and F_0 according to the importance attached to various portions of the CDF, let us consider weighted

Cramér-von Mises test statistics of the form

$$W_n^2(\psi) = n\int_{-\infty}^{\infty} \psi\left(F\left(x;\boldsymbol{\theta}\right)\right)\left(F_n\left(x\right) - F\left(x;\boldsymbol{\theta}\right)\right)^2 dF\left(x;\boldsymbol{\theta}\right), \tag{4.4}$$

where weight $\psi : [0,1] \to \mathbb{R}_+$.

For the case of simple null hypotheses, limiting properties of the WCvM statistics (4.4) with $\psi(t) = 1/(1-t)^\beta$ (which will be denoted by $W_n^{2,\beta}$, where n is the sample size) have been treated in Chapter 3 for $\beta \leqslant 2$. In particular, it is shown there that for $\beta < 2$, $W_n^{2,\beta}$ has a well-defined limiting distribution while $W_n^{2,2} \xrightarrow{d} +\infty$ as $n \to +\infty$. Unfortunately, convergence to the limiting distribution of $W_n^{2,\beta}$ with $1.5 \leqslant \beta < 2$ appears to be rather slow to be useful in practice as one has to resort to MC simulations even for very large samples. This fact makes $W_n^{2,\beta}$ with $1.5 \leqslant \beta < 2$ natural candidates for the study in this chapter. Therefore, we have selected both standard and specialized test statistics: CvM $\widehat{W}_n^2$, AD $\widehat{A}_n^2$ and $\widehat{W}_n^{2,\beta}$ for $\beta \in \{1.5, 1.8, 1.95\}$. Here, the hat indicates the fact that an estimator $\hat{\boldsymbol{\theta}}_n$ of the unknown distributional parameter $\boldsymbol{\theta}$ is used to evaluate the statistic.

4.3 Theoretical Results on Asymptotic Behavior of $\widehat{W}_n^{2,\beta}$

In this section, we discuss convergence in distribution of $\widehat{W}_n^{2,\beta}$ for $\beta \leqslant 2$ and $(\widehat{W}_n^{2,2} - \ln n)/(2\sqrt{\ln n})$. For $\beta < 2$, we present conditions under which $\widehat{W}_n^{2,\beta}$ converges in distribution and verify them by an example of the Gaussian distribution. We also provide conditions for $\widehat{W}_n^{2,2} \xrightarrow{d} +\infty$ as $n \to +\infty$, where $\xrightarrow{d}$ denotes convergence in distribution. For example, the conditions are satisfied by the Gaussian distribution. Finally, we sketch a proof that the normalization $(\widehat{W}_n^{2,2} - \ln n)/(2\sqrt{\ln n})$ leads $\widehat{W}_n^{2,2}$ to a non-degenerate limiting distribution.

The results presented here rely on some auxiliary lemmas which are relegated to Appendix C.

4.3.1 On Convergence in Distribution of $\widehat{W}_n^{2,\beta}$

Let $X = \{X_1, \ldots, X_n\}$ be a sample of IID random variables with common CDF $G(\cdot)$. We wish to test the following null hypothesis:

$$H_0 : G \in \mathscr{F} = \{F(\cdot;\boldsymbol{\theta}), \boldsymbol{\theta} \in \Omega \subseteq \mathbb{R}^m\} \text{ versus } H_1 : G \notin \mathscr{F}, \tag{4.5}$$

where the distributional class $\mathscr{F}$ is given, but the parameter vector[1] $\boldsymbol{\theta} \in \Omega \subseteq \mathbb{R}^m$ (with $m \in \mathbb{N}$ and Ω being an open set) is unknown and has to be estimated.

To assess the null hypothesis (4.5), let us consider weighted Cramér-von Mises test statistics of the form

$$\widehat{W}_n^2(\psi) = n \int_{-\infty}^{\infty} \psi\left(F(x;\hat{\boldsymbol{\theta}}_n)\right)\left(\hat{F}_n(x) - F(x;\hat{\boldsymbol{\theta}}_n)\right)^2 dF(x;\hat{\boldsymbol{\theta}}_n), \tag{4.6}$$

where weight $\psi : [0,1] \to \mathbb{R}_+$, $\hat{\boldsymbol{\theta}}_n$ is an estimator of $\boldsymbol{\theta}$, and $\hat{F}_n(\cdot)$ is the empirical CDF of $X_1, \ldots, X_n$.

Throughout this section, we shall assume that $\hat{\boldsymbol{\theta}}_n$ is the maximum likelihood (ML) estimator. The true value of $\boldsymbol{\theta}$ will be denoted by $\boldsymbol{\theta}_0$. Further, let us consider the Fisher Information $\mathscr{I}(\boldsymbol{\theta}_0) = \mathbb{E}\left[\nabla_{\boldsymbol{\theta}} \ln f(x;\boldsymbol{\theta}) \nabla_{\boldsymbol{\theta}}^T \ln f(x;\boldsymbol{\theta}) \mid \boldsymbol{\theta}_0\right]$ of $\hat{\boldsymbol{\theta}}_n$, where f is the PDF and $\nabla_{\boldsymbol{\theta}}(\cdot)$ is the operator of differentiation with respect to $\boldsymbol{\theta}$. Recall that the standard Brownian bridge $\{B(t), t \in [0,1]\}$ is a Gaussian process with mean 0 and covariance function $Cov(B(t), B(s)) = K_0(t,s)$, where $K_0(t,s) = \min(t,s) - ts$ for all $t, s \in [0,1]$. We shall define the estimated Brownian bridge as follows.

Definition. The *estimated Brownian bridge* $\{\hat{B}(t), t \in [0,1]\}$ is a Gaussian process with mean 0 and covariance $Cov(\hat{B}(t), \hat{B}(s)) = K(t,s;F,\boldsymbol{\theta}_0)$, and

$$K(t,s;F,\boldsymbol{\theta}_0) = K_0(t,s) - r(t,s;F,\boldsymbol{\theta}_0),$$

where

$$r(t,s;F,\boldsymbol{\theta}_0) = \nabla_{\boldsymbol{\theta}}^T F(x;\boldsymbol{\theta})|_{t=F(x;\boldsymbol{\theta}_0),\boldsymbol{\theta}=\boldsymbol{\theta}_0} \mathscr{I}^{-1}(\boldsymbol{\theta}_0) \nabla_{\boldsymbol{\theta}} F(y;\boldsymbol{\theta})|_{s=F(y;\boldsymbol{\theta}_0),\boldsymbol{\theta}=\boldsymbol{\theta}_0}$$

for all $t, s \in [0,1]$.

Without proof, Neuhaus (1979, p. 491), Shorack and Wellner (2009, p. 233),

[1] All vectors in Section 4.3 are column ones. We use the notation $\langle \cdot, \ldots, \cdot \rangle$ for row vectors.

Thas (2010, p. 140) and Martynov (2011, p. 3572) mention that, under suitable regularity conditions on F, $\nabla_{\boldsymbol{\theta}} F$, $\hat{\boldsymbol{\theta}}_n$ and ψ, the following result on convergence of (4.6) in distribution holds.

Proposition 4.1. Under the null hypothesis, $\widehat{W}_n^2(\psi) \xrightarrow{d} \int_0^1 \hat{B}^2(t)\psi(t)dt$ as $n \to \infty$.

Proving convergence in distribution of (4.6) for $\psi(t) \not\equiv 1$ is, in general, a delicate process. For continuous weight functions $\psi(\cdot)$, Durbin (1973b, p. 56) established Proposition 4.1. Viollaz (1995) developed a method for proving Proposition 4.1 for a class of weight functions $\psi = \psi(t)$ which includes certain non-integrable on $[0,1]$ functions, such as $\psi(t) = 1/(t(1-t))$, $\psi(t) = 1/(t(2-t))$ and $\psi(t) = 1/(1-t^2)$. However, as we shall see below, the case of $\psi(t) = 1/(1-t)^\beta$ for $3/2 \leqslant \beta < 2$ is not covered. The remark of Viollaz (1995, p. 2829) that, in the literature, no rigourous proof is available for the general case, is still valid today, to the best of our knowledge.

In Chapter 3, for simple hypotheses, we have considered $\psi(t) = 1/(1-t)^\beta$ with $\beta \in [1,2]$ and the corresponding statistic $W_n^{2,\beta}$, i.e., $W_n^2(1/(1-t)^\beta)$. Below, we specialize some known results on convergence in distribution to $\widehat{W}_n^{2,\beta}$.

For Proposition 4.2, we shall need some regularity conditions. Specifically, we shall assume that the family of distributions $\mathscr{F}$ and the estimator $\hat{\boldsymbol{\theta}}_n$ of $\boldsymbol{\theta}$ satisfy conditions 4.1-4.2.

Regularity Condition 4.1. In a neighbourhood of $\boldsymbol{\theta}$, for $\beta < 3/2$, we have

$$\begin{aligned} F(x;\boldsymbol{\theta}^*) - F(x;\boldsymbol{\theta}) - (\boldsymbol{\theta}^* - \boldsymbol{\theta})^T \nabla_{\boldsymbol{\theta}} F(x;\boldsymbol{\theta}) &= O\Big((F(x;\boldsymbol{\theta})(1-F(x;\boldsymbol{\theta})))^{\beta/3} \\ &\quad \times (\boldsymbol{\theta}^* - \boldsymbol{\theta})^T (\boldsymbol{\theta}^* - \boldsymbol{\theta}) \Big), \end{aligned} \tag{4.7}$$

$$\frac{\partial F(x;\boldsymbol{\theta})}{\partial \theta_{nj}} = O\Big((F(x;\boldsymbol{\theta})(1-F(x;\boldsymbol{\theta})))^{\beta/3} \Big), \tag{4.8}$$

where $j = \overline{1,m}$, $\boldsymbol{\theta} = \langle \theta_{n1}, \ldots, \theta_{nm} \rangle^T$ and the O terms hold uniformly in a neighbourhood of $\boldsymbol{\theta}$ for both $x \to +\infty$ and $x \to -\infty$.

Regularity Condition 4.2. The estimator $\hat{\boldsymbol{\theta}}_n = \langle \hat{\theta}_{n1}, \ldots, \hat{\theta}_{nm} \rangle^T$ of $\boldsymbol{\theta} = \langle \theta_{n1}, \ldots, \theta_{nm} \rangle^T$ is asymptotically linear in the sense that, for each $j = \overline{1,m}$,

$$\sqrt{n}(\hat{\theta}_{nj} - \theta_j) = \frac{1}{\sqrt{n}} \sum_{i=1}^n h_j(\xi_{ni}) + o_P(1),$$

as $n \to \infty$, where $h_j(\xi_{ni})$ for each j are IID with mean 0 and variance σ_j^2.

Remark. We note that

1. ML estimators usually satisfy the Regularity Condition 4.2 (see van der Vaart, 2007, p. 65, Theorem 5.39);

2. As in Viollaz (1995, p. 2837), it can be verified that distributions satisfying the Regularity Conditions 4.1-4.2 exist. For example, the conditions are satisfied by the Exponential and Gumbel families of distributions. Indeed, for these distributions, Regularity Condition 4.2 is known to hold from Cramér (1999, pp. 505) and Smith (1985, p. 88). Part (4.7) of Regularity Condition 4.1 is checked by an application of Taylor's formula with the remainder in Lagrange's form, while (4.8) is verified directly.

We are now in a position to establish the following theorem.

Theorem 4.2. Under the null hypothesis and the regularity conditions 4.1-4.2, $\widehat{W}_n^{2,\beta} \xrightarrow{d} \int_0^1 \hat{B}^2(t)/(1-t)^\beta dt$ for $\beta < 3/2$, as $n \to \infty$.

Proof. For $\beta \leqslant 0$, the proposition follows from Durbin (1973b, p. 56) as $\psi(t) = 1/(1-t)^\beta$ is continuous on $[0,1]$. For $\beta \in (0,3/2)$, it can be established by adapting the approach of Viollaz (1995) to $\psi(t) = 1/(1-t)^\beta$. Indeed, the proof of their main theorem (i.e., Theorem 3.1 on p. 2833) relies on

(i) The regularity conditions 4.1-4.2 (with $\beta/3$ replaced by $1/4$);

(ii) Three lemmas, which require that $\psi(\cdot)$ should be differentiable, $\psi^{3/4}(\cdot)$ and $(q(\cdot))^{-2}$ be integrable, where $q(t) = \psi^{-1/4}(t)$ is strictly increasing on $(0,c)$ and strictly decreasing on $(c,1)$ for some $c \in [0,1]$;

(iii) $q(t) = O(\sqrt[4]{t})$ as $t \downarrow 0$ and $q(t) = O(\sqrt[4]{1-t})$ as $t \uparrow 1$.

On a close examination of the proof of Lemma 3.2 on p. 2835, we conclude that instead of (iii) a different condition was used:

$$1/q(t) = O(1/\sqrt[4]{t}) \text{ and } 1/q(t) = O(1/\sqrt[4]{1-t}), \tag{4.9}$$

as $t \downarrow 0$ and $t \uparrow 1$, respectively. Therefore, (iii) should be replaced by (4.9).

Under the null hypothesis, regularity conditions (i)-(ii) and (4.9), for $\psi(t) = (1-t)^{-\beta}$, Viollaz (1995, Theorem 3.1) only implies Proposition 4.2 for $0 < \beta \leqslant 1$. This result can be improved to $0 < \beta < 3/2$ by observing that the assumptions of integrability of $\psi^{3/4}(\cdot)$ and $q^{-2}(\cdot)$ may be modified as follows. It is sufficient to require that there should exist an $r \in (0,1)$, such that ψ^r and $q^{-2}(\cdot)$ are integrable on $[0,1]$, where $q(t) = \psi^{r-1}(t)$. In this case,

$$\begin{aligned} \beta &< \sup_{r\in(0,1)} \min\left\{\tfrac{1}{r}, \tfrac{1}{2(1-r)}\right\} \\ &= 2\left(\inf_{r\in(0,1)}(2-r+|3r-2|)\right)^{-1} \\ &= \tfrac{3}{2} \end{aligned}$$

for the optimal $r_{opt} = 2/3$.

If $\beta \in (0,3/2)$, then $\psi^{r_{opt}}(t) = (1-t)^{2\beta/3}$ and $q(t) = (1-t)^{\beta/3}$ satisfy the integrability assumptions, $1/q(t) = O\left((1-t)^{-\beta/3}\right)$ as $t \uparrow 1$, $1/q(t) = O\left(t^{-\beta/3}\right)$ as $t \downarrow 0$ and $q(\cdot)$ is strictly decreasing on $[0,1]$, i.e., $c = 0$. Thus, Theorem 3.1 and Lemmas 4.1-4.2 of Viollaz (1995) are applicable mutatis mutandis to conclude the proof of this proposition. ■

Unfortunately, as it stands, the method of Viollaz (1995) is not suitable for proving the convergence of $\widehat{W}_n^{2,\beta}$ in distribution for $\beta \in [3/2,2)$.

However, a general regularity condition on the weight function in Proposition 4.2 is given in Neuhaus (1979) and Martynov (2011) without proof. The condition requires

$$\int_0^1 K(t,t;F,\boldsymbol{\theta}_0)\psi(t)dt < \infty.$$

As follows from Theorem C.1, under this condition, $\int_0^1 \hat{B}^2(t)\psi(t)dt$ is finite wp1.

The next proposition establishes the condition for the almost sure finiteness of $\int_0^1 \hat{B}^2(t)\psi(t)dt$.

Proposition 4.3. $P(\int_0^1 \hat{B}^2(t)\psi(t)dt < \infty) = 1$ if and only if $\int_0^1 K(t,t;F,\boldsymbol{\theta}_0)\psi(t)dt < \infty$.

Proof. The proof follows from an application of Corollary C.2 with $p = 2$. ■

The following corollary provides a sufficient condition for $\int_0^1 K(t,t;F,\boldsymbol{\theta}_0)\psi(t)dt < \infty$.

Corollary 4.4. If

$$\int_0^1 K_0(t,t)\psi(t)dt < \infty \text{ and } \int_0^1 r(t,t;F,\boldsymbol{\theta}_0)\psi(t)dt < \infty,$$

then $\int_0^1 K(t,t;F,\boldsymbol{\theta}_0)\psi(t)dt < \infty$.

Similarly, we we can get a simple condition for $P(\int_0^1 \hat{B}^2(t)\psi(t)dt = \infty) = 1$.

Corollary 4.5. If

$$\int_0^1 K_0(t,t)\psi(t)dt = \infty \text{ and } \int_0^1 r(t,t;F,\boldsymbol{\theta}_0)\psi(t)dt < \infty,$$

then $P(\int_0^1 \hat{B}^2(t)\psi(t)dt = \infty) = 1$.

We have emphasized that $r(\cdot,\cdot;F,\boldsymbol{\theta}_0)$ is, in general, a function of both the CDF and the true value of the unknown distributional parameter. Therefore, a separate assessment has to be done for each distribution in the null hypothesis (4.5).

It is, however, known (Bagdonavičius et al., 2011, pp. 92-96) that for location-scale $\mathscr{G} = \{G((\cdot-\mu)/\sigma) \mid \mu \in \mathbb{R}, \sigma > 0)\}$ and the particular type of scale-shape $\mathscr{W} = \{W((\cdot/\beta)^\alpha) \mid \alpha,\beta > 0\}$ families of distributions, r does not depend on the distributional parameter. In particular, for the Gaussian distribution, we have (Sukhatme, 1972; del Barrio et al., 2007)

$$r^\star(t,s) = \varphi(\Phi^{-1}(t))\varphi(\Phi^{-1}(s)) + 0.5\Phi^{-1}(t)\Phi^{-1}(s)\varphi(\Phi^{-1}(t))\varphi(\Phi^{-1}(s)),$$

where $r^\star(t,s) = r(t,s;\Phi,\boldsymbol{\theta}_0)$, $\varphi(x) = (1/\sqrt{2\pi})\exp(-x^2/2)$ and $\Phi(x) = \int_{-\infty}^x \varphi(t)dt$.

For the weight function $\psi(t) = 1/(1-t)^\beta$ with $\beta \in [1,2]$, $\int_0^1 K_0(t,t)\psi(t)dt = 1/((\beta-2)(\beta-3))$ (cf. p. 68) if $\beta < 2$, and ∞ if $\beta = 2$. A similar fact holds for $\widehat{W}_n^{2,\beta}$ for the case of the Gaussian distribution.

Theorem 4.6. For the Gaussian distribution as null in (4.5), the limiting distribution of $\widehat{W}_n^{2,\beta}$ admits the following dichotomy: it is finite wp1 when $\beta < 2$, and is infinite wp1 when $\beta = 2$.

Proof. An application of Corollaries 4.4, 4.5 and Lemma C.4 yield the result. ■

4.3.2 On a Normalization of $\widehat{W}_n^{2,2}$

We now proceed to constructing a normalization that would lead $\widehat{W}_n^{2,2}$ to a non-degenerate random variable as $n \to \infty$. Let us recall that Theorem 3.11 on p. 59 for simple hypotheses, established $(W_n^{2,2} - \ln n)/(2\sqrt{\ln n}) \xrightarrow{d} N(0,1)$ as $n \to \infty$. In what follows, by an example of the Gaussian distribution as null in (4.5), we provide evidence suggesting that the same normalization appears to work for $\widehat{W}_n^{2,2}$.

Next, we recall the notion of the uniform empirical process (see Section 3.3.1) and, following Durbin (1973a) and Durbin (1973b), introduce the notion of the estimated uniform empirical process.

Definition. Let $\boldsymbol{\theta}_0$ be the true value of the unknown parameter $\boldsymbol{\theta}$ in (4.5). Let further $t_j = F(x_j; \boldsymbol{\theta}_0)$ and $F_n(t)$ be the sample CDF, that is, the proportion of t_j, $j = \overline{1,n}$, which are smaller than or equal to t. Then, $e_n(t) = \sqrt{n}(F_n(t) - t)$ is referred to as the *uniform empirical process* (UEP).

Definition. Let $\hat{t}_j = F(x_j; \hat{\boldsymbol{\theta}}_n)$ and let $\hat{F}_n(t)$ be the estimated sample CDF, that is, the proportion of $\hat{t}_j$, $j = 1, \ldots, n$ which are smaller than or equal to t. Then $\hat{e}_n(t) = \sqrt{n}(\hat{F}_n(t) - t)$ is referred to as the *estimated uniform empirical process*.

Let

$$g(t) = \nabla_{\boldsymbol{\theta}} F(x; \boldsymbol{\theta})|_{t=F(x;\boldsymbol{\theta}_0)} \text{ and } w_n = (1/\sqrt{n})\mathscr{I}^{-1}(\boldsymbol{\theta}_0) \sum_{j=1}^{n} \nabla_{\boldsymbol{\theta}} \ln f(x_j; \boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}.$$

Then, Lemma 2 of Durbin (1973a) (for the MLE $\hat{\boldsymbol{\theta}}_n$ and under the null hypothesis (4.5)) establishes that the estimated UEP $\{\hat{e}_n(t), t \in [0,1]\}$ admits the following decomposition as $n \to \infty$: $\hat{e}_n(t) = e_n(t) - w_n^T g(t) + \varepsilon_n(t)$, where $\varepsilon_n \xrightarrow{P} 0$ as $n \to \infty$ and $\xrightarrow{P}$ stands for convergence in probability.

Hence,

$$\frac{(\widehat{W}_n^{2,2} - \ln n)}{2\sqrt{\ln n}} = \frac{(W_n^{2,2} - \ln n)}{2\sqrt{\ln n}} + \frac{2\int_0^1 \frac{e_n(t)(-w_n^T g(t) + \varepsilon_n(t))}{(1-t)^2} dt}{2\sqrt{\ln n}} + \frac{\int_0^1 \frac{(-w_n^T g(t) + \varepsilon_n(t))^2}{(1-t)^2} dt}{2\sqrt{\ln n}}. \tag{4.10}$$

By Slutsky's Theorem, $-w_n^T g(t) + \varepsilon_n(t)$ and $-w_n^T g(t)$ will have the same limiting distribution.

From Shorack and Wellner (2009, p. 229), we get $e_n \xrightarrow{d} B$ and $w_n^T g \xrightarrow{d} \sum_{j=1}^m Z_j F_j$ as $n \to \infty$, where $F_j(t) = \frac{\partial F(x;\boldsymbol{\theta})}{\partial \theta_j}|_{t=F(x;\boldsymbol{\theta}_0), \boldsymbol{\theta}=\boldsymbol{\theta}_0}$. The vector $Z^T = \langle Z_1, \ldots, Z_m \rangle$ and the Brownian bridge process $\{B(t), t \in [0,1]\}$ are jointly Gaussian (see Shorack and Wellner, 2009, pp. 229-230) with zero means[2] and covariances $Cov(Z) = \mathscr{I}^{-1}(\boldsymbol{\theta}_0)$ and $Cov(Z, B(t)) = \int_0^t h(s)ds$, where the vector function

$$h(s) = \mathscr{I}^{-1}(\boldsymbol{\theta}_0)\nabla_{\boldsymbol{\theta}} \ln f(x;\boldsymbol{\theta})|_{x=F(s;\boldsymbol{\theta}_0),\boldsymbol{\theta}=\boldsymbol{\theta}_0}.$$

This follows (e.g., refer to Sukhatme (1972, p. 1921) and Durbin (1973a, p. 285)) from an application of the Multivariate Central Limit Theorem to $< e_n, w_n^T >^T$. This also implies $\langle e_n, w_n^T g\rangle^T \xrightarrow{d} Y$ as $n \to \infty$, where $Y^T = \langle B, \sum_{j=1}^m Z_j F_j\rangle$ is a Gaussian process with mean zero for all $t \in [0,1]$.

By the Continuous Mapping Theorem (van der Vaart, 2007, p. 7, Theorem 2.3), $-e_n w_n^T g \xrightarrow{d} -B\sum_{j=1}^m Z_j F_j$ and $(w_n^T g)^2 \xrightarrow{d} \left(\sum_{j=1}^m Z_j F_j\right)^2$ as $n \to \infty$.

Thus, under suitable regularity conditions, the last two integrals in (4.10) will converge in distribution as follows:

$$2\int_0^1 \frac{e_n(t)(-w_n^T g(t) + \varepsilon_n(t))}{(1-t)^2} dt \xrightarrow{d} -\int_0^1 \frac{B(t)\sum_{j=1}^m Z_j F_j(t)}{(1-t)^2} dt \tag{4.11}$$

and

$$\int_0^1 \frac{(-w_n^T g(t) + \varepsilon_n(t))^2}{(1-t)^2} dt \xrightarrow{d} \int_0^1 \frac{\left(\sum_{j=1}^m Z_j F_j(t)\right)^2}{(1-t)^2} dt. \tag{4.12}$$

For right-hand sides in (4.11) and (4.12) to be nondegenerate, the two integrals

$$-\int_0^1 \frac{B(t)\sum_{j=1}^m Z_j F_j(t)}{(1-t)^2} dt \text{ and } \int_0^1 \frac{\left(\sum_{j=1}^m Z_j F_j(t)\right)^2}{(1-t)^2} dt \tag{4.13}$$

should be finite wp1.

In this case, as $n \to \infty$,

$$\frac{\widehat{W}_n^{2,2} - \ln n}{2\sqrt{\ln n}} = \frac{W_n^{2,2} - \ln n}{2\sqrt{\ln n}} + o_P(1)$$

[2]Throughout this section, by mean zero, we understand the zero vector $\langle 0,\dots,0\rangle^T$ of a corresponding dimension.

and by Slutsky's Theorem

$$\frac{\widehat{W}_n^{2,2} - \ln n}{2\sqrt{\ln n}} \xrightarrow{d} N(0,1).$$

For the Gaussian distribution with CDF $\Phi((\cdot - \mu)/\sigma)$, we now verify that the integrals (4.13) are indeed finite wp1. In this case, $m = 2$, $Z^T = \langle Z_1, Z_2 \rangle$, $F_1(t) = -(1/\sigma)\varphi(\Phi^{-1}(t))$, $F_2(t) = -(1/\sigma)\Phi^{-1}(t)\varphi(\Phi^{-1}(t))$ and (see del Barrio et al., 2007, p. 39)

$$\mathscr{I}^{-1}(\mu, \sigma) = \sigma^2 \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{pmatrix}. \tag{4.14}$$

By the Cauchy-Schwarz inequality, $\mathbb{E}[B(t)Z_j] \leqslant \sqrt{t(1-t)}\sqrt{\mathbb{V}[Z_j]}$ for $j \in \{1,2\}$. From (4.14), $\mathbb{V}[Z_1] = \sigma^2$ and $\mathbb{V}[Z_2] = 0.5\sigma^2$.

Upon using Lemma C.6, Theorem C.1, Corollary C.2, Lemma C.4 (for $\beta = 2$) and Lemma C.5, we now conclude that the two integrals in (4.13) are finite wp1. These considerations suggest the following proposition.

Proposition 4.7. Under suitable regularity conditions,

$$\frac{\widehat{W}_n^{2,2} - \ln n}{2\sqrt{\ln n}} \xrightarrow{d} N(0,1)$$

as $n \to \infty$.

Remark. Clearly, a formal proof of Proposition 4.7 will require the following:

1. A set of regularity conditions under which the statements (4.11) and (4.12) are valid;

2. A rigorous proof of (4.11) and (4.12) under these regularity conditions.

4.4 Simulation Study

In this section, we explore agreement between H_0 and $H_0^\star$ for selected distributions. All computations were performed on a 12-core HP Z800 workstation with the Xeon 2.93 GHz processor and 48 GB DDR3 of random-access memory (RAM) with the help of MATLAB®[3] software (MATLAB, 2016a) and the Parallel Computing Toolbox (MATLAB, 2016b).

4.4.1 Tabulation of Critical Values for Testing for Normality

CBT relies on the fact that for a given GoF statistic, the mechanism of the assessment of a hypothesis for normality is readily available. In Chen and Balakrishnan (1995, p. 155), critical values for the CvM and AD statistics are presented. To tabulate critical values for $\widehat{W}_n^{2,\beta}$ for $\beta \in \{1.5, 1.8, 1.95\}$, we have to use MC simulations.

For location-scale distributional families to which the normal distribution belongs, it is known (Bagdonavičius et al., 2011, pp. 92-96) that both the finite-sample distribution and the limiting distribution (if it exists) of $W_n^2(\psi)$ are independent of the unknown parameter $\boldsymbol{\theta}$. Therefore, for the purposes of the tabulation, it is sufficient to consider the standard Gaussian distribution $N(0,1)$.

We considered a grid of sample sizes from 10 to 5000 and ran ten million trials for each sample size. In each simulation trial for a given sample size n, we generated a random sample $\{x_i\}_{i=1}^n$ from $N(0,1)$, estimated the mean μ and standard deviation σ using the parameter estimates $\hat{\mu} = \left(\sum_{i=1}^n x_i\right)/n$ and $\hat{\sigma} = \sqrt{\left(\sum_{i=1}^n (x_i - \hat{\mu})^2\right)/(n-1)}$.

Remark 4.8. $\hat{\sigma}$ is not an MLE estimate of σ; while $\hat{\sigma}^2$ is an unbiased estimator of σ^2, the bias of $\hat{\sigma}$ tends to 0 as $n \to \infty$ (Cramér, 1999, p. 484). Also, see Chen (1991, p. 129) for relevance of this estimate to CBT.

Each statistic was evaluated at $\{x_i\}_{i=1}^n$ as follows.

Let $z_{(j)} = F(x_{(i)}; \hat{\boldsymbol{\theta}}_n)$, where $x_{(1)} \leqslant \cdots \leqslant x_{(n)}$, F is the null CDF of interest (e.g., $\Phi((\cdot - \mu)/\sigma)$ for tabulation purposes) and $\hat{\boldsymbol{\theta}}_n$ is an (efficient) estimate of the unknown parameter (a maximum likelihood (ML) estimate throughout this thesis). Let us recall the following computational formula for $W_n^{2,\beta}$ which is valid for $\beta < 2$ (see p. 68):

$$\widehat{W}_n^{2,\beta} = Const + \frac{2}{2-\beta} \sum_{j=1}^n (1 - z_{(j)})^{2-\beta} + \frac{1}{(\beta-1)n} \sum_{j=1}^n \frac{1 + 2(n-j)}{(1 - z_{(j)})^{\beta-1}}, \qquad (4.15)$$

where $Const = n\left(\frac{1}{3-\beta} - \frac{2}{2-\beta} + \frac{1}{1-\beta}\right)$.

[3] MATLAB is a registered trademark of The MathWorks, Inc. For more information, see http://www.mathworks.com.

The critical values $C_{n,\alpha}^{2,\beta}$ were calculated as $100(1-\alpha)\%$-sample quantiles of the simulated distributions of the statistics for significance levels $\alpha \in \{0.01, 0.05, 0.10\}$. In Table 4.1, the critical values are presented for sample size $n \in \{50, 250, 1000\}$.

Table 4.1: Monte Carlo based critical values $C_{n,\alpha}^{2,\beta}$ for $H_0 : X \sim N(\mu, \sigma)$ with unknown μ and σ for $W_n^{2,\beta}$ for $\beta \in \{1.5, 1.8, 1.95\}$, sample size $n \in \{50, 250, 1000\}$, at significance level $\alpha \in \{0.01, 0.05, 0.10\}$. Ten million simulations were used for each sample size.

Sample Size	$\beta = 1.5$			$\beta = 1.8$			$\beta = 1.95$		
	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
50	2.13	1.23	0.97	11.52	3.99	2.60	30.55	8.34	4.71
250	2.05	1.31	1.06	15.02	5.25	3.49	52.01	13.22	7.46
1000	1.89	1.30	1.07	14.33	5.60	3.95	59.47	15.87	9.42

To be able to compute the critical values for arbitrary sample sizes in $[10, 5000]$, a nonlinear function can be fitted (e.g., D'Agostino and Stephens (1986, p. 105) and MacKinnon (1991)) to the tabulated critical values $C_{n,\alpha}^{2,\beta}$. For example, for $\alpha = 0.10$, we have devised the following functional form that worked very well: $1/(a_0 + a_1 n^{-1/3} + a_2 n^{-1/2} + a_3 n^{-1})$; see Tables 4.2 and 4.3. The coefficients $\{a_k\}_{k=0}^{3}$ were estimated by linear least-squares method. In Table 4.3, the following error metrics are presented that compare $C_{\cdot,\alpha}^{2,\beta}$ with the fitted values $\widehat{C_{\cdot,\alpha}^{2,\beta}}$: Mean Squared Error (MSE), $A = \max \left| 1 - \widehat{C_{\cdot,\alpha}^{2,\beta}} / C_{\cdot,\alpha}^{2,\beta} \right|$ and $B = \text{mean} \left| 1 - \widehat{C_{\cdot,\alpha}^{2,\beta}} / C_{\cdot,\alpha}^{2,\beta} \right|$ with mean $\left\{ r_{i=1}^{N} \right\}$ defined as $\left(\sum_{i=1}^{N} r_i \right) / N$.

Table 4.2: Fitted coefficients $\{\hat{a}_i\}_{i=1}^{3}$ in the functional form $1/(a_0 + a_1 n^{-1/3} + a_2 n^{-1/2} + a_3 n^{-1})$ for the case of $\widehat{C_{\cdot,0.10}^{2,\beta}}$ with $\beta \in \{1.5, 1.8, 1.95\}$.

β	1.5	1.8	1.95
$\hat{a}_0$	0.973	0.221	0.070
$\hat{a}_1$	-1.157	0.042	0.300
$\hat{a}_2$	2.345	0.802	0.127
$\hat{a}_3$	1.914	1.918	2.165

Table 4.3: Descriptive statistics for errors implied by the fits from Table 4.2.

β	1.5	1.8	1.95
$A \times 100\%$	0.05%	0.09%	0.48%
$B \times 100\%$	0.03%	0.04%	0.19%
$\mathrm{MSE} \times 100\%$	0.00%	0.00%	0.04%

4.4.2 Selected Distributions

As explained in Section 3.2, in financial applications, statistical distributions which accurately describe upper tail behavior of a data set at hand are important. Such distributions are usually heavy-tailed. In particular, heavy-tailed distributions have been known to play a prominent role in operational risk (OpRisk) modeling (Peters and Shevchenko, 2015). In this thesis, we will call a distribution heavy-tailed when its CDF belongs to the class of subexponential distributions. Subexponentiality and other notions which allow us to further classify heavy-tailed distributions are defined below.

We begin by recalling the definition of the tail index of a distribution. It is well defined for regularly varying distributions as follows (Degen, 2010, p. 9).

Definition 4.9. A positive measurable function f is *regularly varying* with parameter $\beta \in \mathbb{R}$ if f satisfies $\lim_{t\to\infty} f(tx)/f(t) = x^{\beta}$ for all $x > 0$.[4] This is denoted by $f \in \mathscr{R}_{\beta}$. In the case of a probability density function (PDF) f, one in particular has that if $f \in \mathscr{R}_{-1/\xi-1}$, then $\overline{F} \equiv 1 - F \in \mathscr{R}_{-1/\xi}$ and ξ is referred to as the tail index.

The larger ξ is, the heavier the upper tail is. A distribution with $\xi > 0$ is considered to be heavy-tailed, while it belongs to the so-called class of rapidly varying functions if $\xi = 0$, or, equivalently, $\overline{F} \in \mathscr{R}_{-\infty}$. The precise definition of a rapidly varying function is given below (Embrechts et al., 1997, p. 570).

Definition 4.10. A positive measurable function h is *rapidly varying* with index $-\infty$ ($h \in \mathscr{R}_{-\infty}$) if $\lim_{t\to\infty} h(xt)/h(t)$ equals 0 if $x > 1$ and equals ∞ if $0 < x < 1$.

Distributions with $0 \leqslant \xi < 1/2$, $1/2 \leqslant \xi < 1$ and $\xi \geqslant 1$ will have finite mean and variance, finite mean but infinite variance, and infinite mean and variance, respectively.

[4] Definitions 4.9-4.11 are generalizable to functions defined on $\mathbb{R}$ (see Embrechts et al., 1997, p. 571).

Another class of distributions extensively used in financial applications is the class of subexponential distributions which are defined as follows (see Embrechts et al., 1997, p. 13 and p. 39).

Definition 4.11. A CDF F with support $(0,\infty)$ is *subexponential* if, for all $n \geqslant 2$, $\lim_{n\to\infty} \overline{F^{n\star}}(x)/\overline{F}(x) = n$, where $F^{n\star}$ is the n-fold convolution of itself.

The class of subexponential distributions is denoted by $\mathscr{S}$. The upper tail of distributions from $\mathscr{S}$ decays slower than any exponential tail (Böcker and Klüppelberg, 2005, p. 91). It is known that $\mathscr{R}_\beta \subset \mathscr{S}$. Distributions which are subexponential and belong to the maximum domain of attraction of the Gumbel distribution (which is denoted by $MDA(\Lambda)$; for pertinent details, refer to Embrechts et al. (1997).) are referred to as semi-heavy-tailed. We will call distributions light-tailed if they are neither heavy-tailed nor semi-heavy-tailed.

The subexponential distributions are realistic models for heavy-tailed random variables. In risk modeling, they are the candidates for loss distributions in the heavy-tailed case. If $X_1,\ldots,X_n$ are independent random variables distributed according to $F \in \mathscr{S}$, then $\lim_{n\to\infty} P\left(\sum_{i=1}^{n} X_i > x\right)/P\left(\max_{i=1,\ldots,n} X_i > x\right) = 1$. In the modeling of OpRisk, subexponential distributions find their applications as part of describing the aggregate loss distribution which was defined on p. 48 as follows.

Definition. The *Aggregate Loss Distribution* (ALD) is the distribution of the aggregate (compound) loss $L = \sum_{i=1}^{N} X_i$, where the frequency N is a discrete random variable and $X_1,...,X_N$ are positive continuous random severities with a common CDF F_X.

As we have discussed in Section 1.2, regulators typically require financial institutions to hold equity capital for OpRisk. The capital is a quantile of the ALD, given by

$$VaR_\delta(L) = \inf\{x \in \mathbb{R} \mid P(L > x) \leqslant 1-\delta\},$$

where confidence level $\delta \in (0,1)$. Normally, δ is very close to one. In particular, value-at-risk for $\delta = 0.999$ is known as regulatory capital.

If $F_X \in \mathscr{S}$ and if $\lambda = \mathbb{E}[N]$ with N independent of $\{X_i\}_{i=1}^{n}$, then $VaR_\delta(L) \approx F_X^{-1}\left(1-(1-\delta)/\lambda\right)$ is the single loss approximation (SLA) to the capital (Böcker and Klüppelberg, 2005).

Motivated by the above considerations, we have selected seven distributions for testing purposes which are commonly used in applications: Gaussian, Gamma, Weibull, LogNormal, Generalized Pareto (GP), LogGamma and Fréchet. These distributions range from light-tailed (e.g., Gaussian) to heavy-tailed (e.g., GP). Also, four of them have rapidly varying CDFs while the remaining ones are regularly varying. At last, subexponentiality is a property of all but Gaussian and Gamma distributions. The PDFs of these distributions, domains of definition and corresponding tail indices are given in Table 4.4. A detailed discussion of tail behavior of Gamma, Weibull and LogNormal distributions can be found in Papalexiou et al. (2013), while the remaining distributions are part of numerous examples throughout Embrechts et al. (1997).

Table 4.4: PDFs, domains of definition and tail indices for selected distributions

No	Distribution	PDF $f = f(x)$	Domain of Definition	Tail Index	Tail Heaviness	Subexponential
1	Gaussian	$\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	$x \in \mathbb{R}$, $\mu \in \mathbb{R}$, $\sigma > 0$	0	Light	No
2	LogNormal	$\frac{1}{x\sqrt{2\pi}\sigma}\exp\left(-\frac{(\ln x-\mu)^2}{2\sigma^2}\right)$	$x > 0$, $\mu \in \mathbb{R}$, $\sigma > 0$	0	Semi-Heavy	Yes
3	Gamma	$\frac{1}{\beta^\alpha \Gamma(\alpha)}x^{\alpha-1}\exp\left(-\frac{x}{\beta}\right)$	$x > 0$, $\alpha, \beta > 0$	0	Light	No
4	Weibull	$\frac{b}{a}\left(\frac{x}{a}\right)^{b-1}\exp\left(-\left(\frac{x}{a}\right)^b\right)$	$x \geqslant 0$, $a, b > 0$	0	Semi-Heavy ($b \in (0,1)$), Light ($b \geqslant 1$)	Yes ($b \in (0,1)$), No ($b \geqslant 1$)
5	Generalized Pareto	$\frac{1}{\theta}\left(1+\xi\frac{x}{\theta}\right)^{-\frac{1}{\xi}-1}$	$x \geqslant 0$, $\xi > 0$, $\theta > 0$	ξ	Heavy	Yes
6	LogGamma	$\frac{1}{x^{\beta+1}\Gamma(\alpha)}(\ln x)^{\alpha-1}\beta^\alpha$	$x \geqslant 1$, $\alpha, \beta > 0$	$\frac{1}{\beta}$	Heavy	Yes
7	Fréchet	$\frac{a}{b}\left(\frac{b}{x}\right)^{a+1}\exp\left(-\left(\frac{b}{x}\right)^a\right)$	$x > 0$, $a, b > 0$	$\frac{1}{a}$	Heavy	Yes

Remark. For the purposes of this thesis, we classify the Gamma distribution as light-tailed for all $\alpha > 0$. The Gamma distribution, albeit belonging to $MDA(\Lambda)$, is not subexponential for all $\alpha > 0$. In Papalexiou et al. (2013, p. 856), it is mistakenly asserted that the Gamma CDF belongs to the class $\mathscr{S}(\alpha)$ where $\mathscr{S}(0) = \mathscr{S}$. In fact, the CDF belongs to the different class $\mathscr{L}(1/\beta)$ (for the definitions of the two classes, see Embrechts and Goldie, 1982, pp. 263-265).

Remark. We have restricted the range of the parameter ξ of the GP distribution from $\mathbb{R}$ to $(0, +\infty)$ as the latter is the most typical range used in OpRisk modeling (Peters and Shevchenko, 2015, p. 56). The case of $\xi = 0$ corresponds to the exponential distribution which is already covered by the Gamma and Weibull distributions when $\alpha = 1$ and $b = 1$, respectively.

For each of the selected distributions, we have considered three sets of parameters;[5] see Table 4.5. These sets were constructed in such a way that they lead to

[5]The parameters of the last three distributions in the first set in Table 4.5 were suggested to the author by J.D. Opdyke as part of personal communication.

approximately the same $VaR_{0.999}(L)$ with mean frequency $\lambda = 25$. To compute the value of $VaR_{0.999}(L)$, we used interpolated single loss approximation (ISLA) approximation described in Opdyke (2014). The degree of tail heaviness increases from the first to the third set.

Table 4.5: Three sets of parameters for each distribution. VaR is calculated for $\delta = 0.999$ and mean frequency $\lambda = 25$.

Distribution/Parameters	First Set				Second Set				Third Set			
	Parameter 1	Parameter 2	Tail Index	VaR	Parameter 1	Parameter 2	Tail Index	VaR	Parameter 1	Parameter 2	Tail Index	VaR
Gaussian	3.694	1.000	0.00	100	4384605.980	62.000	0.000	114,000,000	Not Used			
LogNormal	0.070	1.000	0.00	100	10.000	2.155	0.000	113,889,353				
Gamma	2.000	1.590	0.00	100	39480000.000	0.111	0.000	113,942,032				
Weibull	4.000	2.203	0.00	100	45000.000	0.299	0.000	114,321,912				
GP	0.250	1.260	0.25	100	0.990	4954.250	0.990	114,500,114	1.50	8716.46	1.50	22,939,448,519.56
LogGamma	2.960	4.000	0.25	100	4.890	1.010	0.990	114,161,591	3.45	0.67	1.50	23,319,433,810.44
Fréchet	4.000	2.310	0.25	100	1.010	5000.000	0.990	114,583,740	0.67	5811.16	1.50	22,939,503,003.08

Table 4.6 summarizes the combinations of null and alternative distributions considered in this study.

Table 4.6: Combinations of distributions for the power study. Samples are simulated from distribution F_1 with a corresponding set of parameters from Table 4.5. The parameters are estimated according to distribution F_0.

First and Second Sets				Third Set			
F_0	F_1			F_0	F_1		
Gaussian	LogNormal	Gamma	Weibull	Not Used			
LogNormal	Gaussian	GP	LogGamma				
Gamma	Weibull	LogNormal	LogGamma				
Weibull	Gamma	LogNormal	LogGamma				
GP	LogNormal	LogGamma	Fréchet	GP	LogNormal	LogGamma	Fréchet
LogGamma	LogNormal	GP	Fréchet	LogGamma	LogNormal	GP	Fréchet
Fréchet	LogNormal	GP	LogGamma	Fréchet	LogNormal	GP	LogGamma

4.4.3 Simulations

For a given sample $x_1, \ldots, x_n$ and its order statistics $x_{(1)} \leqslant \cdots \leqslant x_{(n)}$, let the null CDF be $F(\cdot;\boldsymbol{\theta})$, an ML estimate $\widehat{\boldsymbol{\theta}}_n$ of the unknown parameter $\boldsymbol{\theta}$, $\hat{z}_{(j)} = F(x_{(j)};\widehat{\boldsymbol{\theta}}_n)$, the Cramér-von Mises $\widehat{W}_n^2$ and Anderson-Darling AD $\widehat{A}_n^2$ statistics were calculated (see Chen and Balakrishnan, 1995, p. 155) as $\widehat{W}_n^2 = \sum_{i=1}^n \left(\hat{z}_{(j)} - (2i-1)/(2n)\right)^2 + 1/(12n)$ and $\widehat{A}_n^2 = -n - n^{-1}\sum_{i=1}^n \{(2i-1)\ln \hat{z}_{(i)} + (2n+1-2i)\ln(1-\hat{z}_{(i)})\}$, respectively. For $\widehat{W}_n^{2,\beta}$, we used (4.15).

For each selected sample size $n \in \{50, 250, 1000\}$, we ran MC experiments as detailed in Algorithm 4.1, where we chose $nSim = 10^3$, $nSim_1 = 10^4$ and significance level $\alpha \in \{0.01, 0.05, 0.10\}$, distributional parameters from Table 4.5, and distribution names F_0, F_1 from Table 4.6.

The method of maximum likelihood is used to estimate the unknown parameters for all distribution types from Table 4.4 except for the Gaussian distribution for which parameter estimates $\hat{\mu} = \left(\sum_{i=1}^{n} x_i\right)/n$ and $\hat{\sigma} = \sqrt{\left(\sum_{i=1}^{n}(x_i - \hat{\mu})^2\right)/(n-1)}$ were used; see Remark 4.8.

Algorithm 4.1 Steps for the MC Experiments

Input: CDF $F_1(\cdot;\boldsymbol{\theta})$ with $\boldsymbol{\theta} \in \mathbb{R}^d$ specified in Table 4.5, CDF $F_0(\cdot;\boldsymbol{\eta})$ with unknown $\boldsymbol{\eta}$), GoF statistic T, sample size n, number *nSim* of simulations for checking normality by CBH, number $nSim_1$ of simulations for assessing the null by MC simulations and significance level α.

Output: Rejection rates (level if $F_0 \equiv F_1$ or power if $F_0 \not\equiv F_1$) for the five statistics on H_0 and $H_0^\star$, metrics of agreement between decisions made by T for H_0 and $H_0^\star$.

1: **for** $K \leftarrow 1$ **to** *nSim*; **do**
2: Generate a pseudo-random sample $\{x_j\}_{j=1}^n$ from $F_1(\cdot;\boldsymbol{\theta})$. ▷ Base sample for H_0 and $H_0^\star$;
3: Fit $F_0(\cdot;\boldsymbol{\eta})$ to $\{x_j\}_{j=1}^n$ by determining an ML estimate $\widehat{\boldsymbol{\eta}}_0$;
4: Calculate values T_0 of T for $\{x_j\}_{j=1}^n$ and $z_{(j)} = F_0(x_{(j)};\widehat{\boldsymbol{\eta}}_0)$, $j = 1,\ldots,n$;
5: **for** $i \leftarrow 1$ **to** $nSim_1$ **do** ▷ Parametric bootstrap to determine p-values for H_0;
6: Generate a pseudo-random sample $\{y_j\}_{j=1}^n$ from $F_0(\cdot;\widehat{\boldsymbol{\eta}}_0)$;
7: Determine an ML estimate $\widehat{\boldsymbol{\eta}}_i$;
8: Calculate values T_i of T for $\{y_j\}_{j=1}^n$ and $z_{(j)} = F_0(y_{(j)};\widehat{\boldsymbol{\eta}}_i)$, $j = 1,\ldots,n$;
9: Based on T_0 from Step 4, record whether T rejects H_0;
10: **end for**;
11: Based on the result of Steps 5-10, calculate the p-value p_K as the proportion of rejections in the $nSim_1$ trials;
12: Run Algorithm 1.1 to determine value T_A for testing for normality;
13: With a critical value from Table 4.1 corresponding to significance level α and T_A from Step 12, record $R_K = 1$ if T rejects $H_0^\star$ and $R_K = 0$, otherwise;

Algorithm 4.1 Steps for the MC Experiments (continued)

14: **end for**
15: **return** Rate P_{rr} of simultaneous rejection of H_0 and $H_0^\star$: $(1/nSim)\sum_{K=1}^{nSim}\mathbb{I}_{[0,\alpha]\times\{1\}}(p_K,R_K)$;
16: **return** Rate P_{aa} of simultaneous failure to reject H_0 and $H_0^\star$: $(1/nSim)\sum_{K=1}^{nSim}\mathbb{I}_{(\alpha,1]\times\{0\}}(p_K,R_K)$;
17: **return** Rate P_{ra} of simultaneous rejection of H_0 and failure to reject $H_0^\star$: $(1/nSim)\sum_{K=1}^{nSim}\mathbb{I}_{[0,\alpha]\times\{0\}}(p_K,R_K)$;
18: **return** Rate P_{ar} of simultaneous failure to reject H_0 and rejection $H_0^\star$ is $1-P_{rr}-P_{aa}-P_{ar}$;
19: **return** Rate of rejection of H_0 is $P_r = P_{rr}+P_{ra}$. ▷ Level if $F_0 \equiv F_1$ or power if $F_0 \not\equiv F_1$;
20: **return** Rate of rejection of $H_0^\star$ is $P_r^\star = P_{rr}+P_{ar}$. ▷ Level if $F_0 \equiv F_1$ or power if $F_0 \not\equiv F_1$.

4.5 Discussion of Results

Since the results are fairly similar across the three parameter sets considered in Table 4.5, we only present a discussion of results for the second set.[6]

In this section, we assess the performance of both direct and indirect tests according to a number of metrics.

4.5.1 Level and Power

Tables 4.7 and 4.8 summarize the rejection rates of the direct and indirect tests based on the CvM, AD and $\widehat{W}_n^{2,\beta}$ ($\beta \in \{1.5, 1.8, 1.95\}$) test statistics under the null hypotheses H_0 and $H_0^\star$. By using the Wald 95%-confidence interval

$$\alpha \pm 1.96\sqrt{\alpha(1-\alpha)/1000}$$

(Newcombe, 2013, pp. 55-58), it can be readily seen that the nominal level of α is largely maintained by both direct and indirect tests. While indirect CvM, AD and $\widehat{W}_n^{2,1.5}$ are somewhat conservative under the GP null distribution, both $\widehat{W}_n^{2,1.8}$ and $\widehat{W}_n^{2,1.95}$ maintain the nominal level in this case.

[6]The complete set of simulation results is available on the author's website `http://www.kmayorov.ca/supp_materials/CBT/` as spreadsheets in the Microsoft® Excel format.

Table 4.7: Observed empirical Type 1 error rates for a given significance level of the direct (indirect) tests based on CvM and AD under the null hypothesis H_0 ($H_0^\star$). The parameters $\boldsymbol{\theta} = \langle \theta_1, \theta_2 \rangle$ are from the second set of Table 4.5.

Distribution	θ_1	θ_2	Tail index	Sample size	Direct CvM			Indirect CvM			Direct AD			Indirect AD		
					Significance Level			Significance Level			Significance Level			Significance Level		
					0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
Gaussian	4384605.98	62.00	0	50	0.010	0.047	0.104	0.011	0.049	0.104	0.010	0.046	0.099	0.011	0.045	0.101
Gaussian	4384605.98	62.00	0	250	0.002	0.043	0.089	0.002	0.042	0.087	0.002	0.039	0.095	0.003	0.039	0.096
Gaussian	4384605.98	62.00	0	1000	0.011	0.045	0.090	0.010	0.046	0.088	0.012	0.045	0.094	0.012	0.045	0.092
LogNormal	10.00	2.16	0	50	0.011	0.053	0.101	0.011	0.052	0.097	0.011	0.048	0.097	0.011	0.049	0.094
LogNormal	10.00	2.16	0	250	0.017	0.058	0.103	0.015	0.059	0.105	0.018	0.059	0.108	0.016	0.059	0.110
LogNormal	10.00	2.16	0	1000	0.013	0.057	0.098	0.012	0.057	0.098	0.015	0.060	0.097	0.014	0.060	0.096
Gamma	0.27	8000000.00	0	50	0.011	0.047	0.098	0.009	0.044	0.089	0.011	0.042	0.103	0.012	0.045	0.091
Gamma	0.27	8000000.00	0	250	0.016	0.045	0.083	0.013	0.038	0.080	0.014	0.049	0.084	0.009	0.041	0.084
Gamma	0.27	8000000.00	0	1000	0.007	0.044	0.096	0.006	0.042	0.086	0.009	0.044	0.096	0.008	0.037	0.089
Weibull	45000.00	0.30	0	50	0.010	0.059	0.105	0.011	0.056	0.099	0.012	0.051	0.115	0.012	0.053	0.108
Weibull	45000.00	0.30	0	250	0.013	0.060	0.106	0.011	0.054	0.102	0.014	0.061	0.110	0.012	0.061	0.105
Weibull	45000.00	0.30	0	1000	0.009	0.052	0.101	0.006	0.051	0.088	0.008	0.046	0.104	0.005	0.045	0.087
GP	0.99	4954.25	0.99	50	0.011	0.045	0.097	0.006	0.026	0.053	0.009	0.047	0.095	0.006	0.028	0.057
GP	0.99	4954.25	0.99	250	0.006	0.042	0.082	0.004	0.026	0.060	0.008	0.038	0.080	0.004	0.027	0.055
GP	0.99	4954.25	0.99	1000	0.010	0.050	0.094	0.003	0.034	0.072	0.006	0.053	0.101	0.002	0.026	0.073
LogGamma	4.89	1.01	0.99	50	0.010	0.056	0.103	0.012	0.054	0.100	0.009	0.052	0.097	0.009	0.051	0.099
LogGamma	4.89	1.01	0.99	250	0.010	0.056	0.115	0.009	0.058	0.116	0.013	0.059	0.114	0.012	0.059	0.115
LogGamma	4.89	1.01	0.99	1000	0.010	0.045	0.098	0.010	0.046	0.097	0.011	0.050	0.100	0.011	0.048	0.102
Fréchet	1.01	5000.00	0.99	50	0.011	0.045	0.104	0.010	0.045	0.083	0.015	0.051	0.091	0.012	0.046	0.084
Fréchet	1.01	5000.00	0.99	250	0.016	0.057	0.102	0.013	0.053	0.096	0.016	0.052	0.104	0.013	0.049	0.093
Fréchet	1.01	5000.00	0.99	1000	0.014	0.067	0.120	0.009	0.060	0.112	0.013	0.065	0.119	0.012	0.058	0.117

Table 4.8: Observed empirical Type 1 error rates for a given significance level of the direct (indirect) tests based on $\widehat{W}_n^{2,\beta}$ for $\beta \in \{1.5, 1.8, 1.95\}$ under the null hypothesis H_0 ($H_0^\star$). The parameters $\boldsymbol{\theta} = \langle \theta_1, \theta_2 \rangle$ are from the second set of Table 4.5.

Distribution	θ_1	θ_2	Tail index	Sample size	Direct $\widehat{W}_n^{2,1.5}$			Indirect $\widehat{W}_n^{2,1.5}$			Direct $\widehat{W}_n^{2,1.8}$			Indirect $\widehat{W}_n^{2,1.8}$			Direct $\widehat{W}_n^{2,1.95}$			Indirect $\widehat{W}_n^{2,1.95}$		
					Significance Level			Significance Level			Significance Level			Significance Level			Significance Level			Significance Level		
					0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
Gaussian	4384605.98	62.00	0	50	0.010	0.064	0.115	0.011	0.058	0.112	0.012	0.059	0.106	0.012	0.059	0.107	0.013	0.058	0.105	0.013	0.058	0.106
Gaussian	4384605.98	62.00	0	250	0.008	0.045	0.091	0.008	0.044	0.091	0.009	0.045	0.098	0.011	0.045	0.097	0.010	0.049	0.102	0.010	0.053	0.102
Gaussian	4384605.98	62.00	0	1000	0.008	0.041	0.091	0.006	0.040	0.090	0.011	0.049	0.093	0.011	0.048	0.091	0.009	0.045	0.097	0.009	0.046	0.101
LogNormal	10.00	2.16	0	50	0.011	0.054	0.106	0.010	0.052	0.103	0.013	0.051	0.100	0.011	0.051	0.101	0.012	0.051	0.101	0.012	0.052	0.100
LogNormal	10.00	2.16	0	250	0.020	0.069	0.116	0.021	0.070	0.118	0.019	0.069	0.121	0.018	0.069	0.122	0.019	0.069	0.112	0.017	0.070	0.113
LogNormal	10.00	2.16	0	1000	0.009	0.045	0.099	0.009	0.047	0.096	0.009	0.040	0.094	0.010	0.041	0.093	0.009	0.046	0.093	0.008	0.045	0.092
Gamma	0.27	8000000.00	0	50	0.012	0.059	0.103	0.016	0.054	0.101	0.019	0.057	0.112	0.019	0.059	0.106	0.019	0.061	0.117	0.019	0.056	0.116
Gamma	0.27	8000000.00	0	250	0.009	0.053	0.093	0.007	0.054	0.093	0.010	0.051	0.100	0.008	0.050	0.100	0.011	0.054	0.107	0.009	0.053	0.104
Gamma	0.27	8000000.00	0	1000	0.009	0.043	0.098	0.008	0.041	0.092	0.014	0.050	0.091	0.016	0.053	0.091	0.013	0.058	0.095	0.016	0.054	0.095
Weibull	45000.00	0.30	0	50	0.010	0.044	0.119	0.009	0.049	0.114	0.009	0.056	0.101	0.009	0.054	0.101	0.009	0.059	0.101	0.009	0.052	0.102
Weibull	45000.00	0.30	0	250	0.008	0.051	0.094	0.006	0.051	0.092	0.004	0.038	0.097	0.004	0.038	0.099	0.004	0.036	0.094	0.003	0.040	0.096
Weibull	45000.00	0.30	0	1000	0.012	0.050	0.095	0.011	0.048	0.087	0.010	0.047	0.108	0.010	0.050	0.104	0.011	0.060	0.106	0.010	0.057	0.103
GP	0.99	4954.25	0.99	50	0.012	0.048	0.092	0.009	0.041	0.076	0.009	0.039	0.099	0.010	0.044	0.088	0.009	0.043	0.093	0.007	0.049	0.090
GP	0.99	4954.25	0.99	250	0.013	0.042	0.102	0.007	0.033	0.073	0.010	0.045	0.102	0.010	0.038	0.087	0.011	0.046	0.107	0.009	0.041	0.090
GP	0.99	4954.25	0.99	1000	0.007	0.053	0.114	0.003	0.033	0.078	0.007	0.044	0.104	0.007	0.043	0.085	0.009	0.041	0.105	0.008	0.040	0.098
LogGamma	4.89	1.01	0.99	50	0.009	0.062	0.113	0.008	0.061	0.112	0.011	0.061	0.115	0.010	0.061	0.116	0.012	0.059	0.110	0.011	0.059	0.112
LogGamma	4.89	1.01	0.99	250	0.013	0.056	0.109	0.013	0.058	0.108	0.013	0.059	0.127	0.013	0.059	0.124	0.014	0.058	0.126	0.013	0.058	0.127
LogGamma	4.89	1.01	0.99	1000	0.015	0.053	0.106	0.016	0.054	0.102	0.009	0.054	0.108	0.009	0.053	0.109	0.010	0.051	0.109	0.010	0.050	0.109
Fréchet	1.01	5000.00	0.99	50	0.008	0.061	0.107	0.006	0.052	0.090	0.009	0.053	0.116	0.008	0.045	0.095	0.010	0.051	0.109	0.008	0.045	0.099
Fréchet	1.01	5000.00	0.99	250	0.007	0.049	0.106	0.008	0.042	0.092	0.010	0.059	0.103	0.008	0.049	0.103	0.009	0.055	0.110	0.007	0.059	0.106
Fréchet	1.01	5000.00	0.99	1000	0.016	0.062	0.108	0.015	0.055	0.109	0.013	0.050	0.106	0.011	0.047	0.101	0.011	0.047	0.104	0.011	0.051	0.098

In Tables 4.9 and 4.10, we present the powers of the direct and indirect tests based on the CvM, AD and $\widehat{W}_n^{2,\beta}$ ($\beta \in \{1.5, 1.8, 1.95\}$) test statistics.

Table 4.9: Powers of the direct and indirect tests based on CvM and AD under the alternative hypothesis. The parameters $\boldsymbol{\theta} = \langle \theta_1, \theta_2 \rangle$ are from the Second Set of Table 4.5.

Simulated as	θ_1	θ_2	Tail index	Estimated as	Sample size	Direct CvM			Indirect CvM			Direct AD			Indirect AD		
						Significance level			Significance level			Significance level			Significance level		
						0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
					50	0.49	0.76	0.86	0.43	0.72	0.83	0.58	0.86	0.94	0.51	0.80	0.90
				GP	250	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
					1000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
					50	0.15	0.32	0.44	0.16	0.33	0.44	0.18	0.37	0.48	0.19	0.37	0.48
Fréchet	1.01	5000.00	0.99	LogGamma	250	0.85	0.95	0.98	0.84	0.95	0.98	0.90	0.97	0.99	0.90	0.97	0.99
					1000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
					50	0.34	0.54	0.66	0.34	0.54	0.66	0.40	0.59	0.71	0.40	0.60	0.72
				LogNormal	250	0.99	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
					1000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
					50	50	0.98	0.99	0.94	0.98	0.99	0.98	1.00	1.00	0.97	1.00	1.00
				LogGamma	250	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
					1000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
					50	0.82	0.93	0.96	0.81	0.93	0.96	0.87	0.95	0.98	0.87	0.95	0.98
Gamma	0.27	8000000.00	0	LogNormal	250	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
					1000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
					50	0.13	0.27	0.38	0.11	0.24	0.35	0.16	0.32	0.44	0.13	0.26	0.37
				Weibull	250	0.78	0.91	0.95	0.72	0.88	0.94	0.86	0.96	0.98	0.81	0.93	0.97
					1000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
					50	0.46	0.64	0.71	0.40	0.59	0.68	0.52	0.68	0.74	0.46	0.65	0.72
				Fréchet	250	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
					1000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
					50	0.16	0.29	0.39	0.17	0.30	0.40	0.18	0.32	0.42	0.19	0.33	0.43
GP	0.99	4954.25	0.99	LogGamma	250	0.71	0.83	0.88	0.71	0.83	0.88	0.76	0.86	0.91	0.77	0.87	0.91
					1000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
					50	0.04	0.12	0.20	0.04	0.13	0.22	0.05	0.14	0.22	0.05	0.15	0.24
				LogNormal	250	0.20	0.39	0.51	0.21	0.40	0.52	0.24	0.45	0.56	0.25	0.46	0.57
					1000	0.83	0.95	0.98	0.84	0.95	0.98	0.90	0.97	0.99	0.90	0.97	0.99
					50	0.02	0.06	0.11	0.01	0.06	0.11	0.01	0.06	0.11	0.01	0.06	0.11
				Fréchet	250	0.02	0.07	0.13	0.02	0.07	0.13	0.02	0.07	0.15	0.02	0.08	0.14
					1000	0.07	0.24	0.35	0.08	0.25	0.35	0.08	0.27	0.40	0.11	0.29	0.41
					50	0.02	0.08	0.16	0.03	0.12	0.20	0.02	0.09	0.19	0.03	0.12	0.22
LogGamma	4.89	1.01	0.99	GP	250	0.10	0.32	0.49	0.36	0.67	0.81	0.32	0.68	0.82	0.51	0.81	0.91
					1000	0.86	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
					50	0.23	0.42	0.55	0.23	0.43	0.55	0.27	0.48	0.61	0.28	0.49	0.61
				LogNormal	250	0.97	0.99	1.00	0.96	0.99	1.00	0.99	1.00	1.00	0.99	1.00	1.00
					1000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
					50	0.05	0.16	0.25	0.01	0.07	0.16	0.06	0.17	0.28	0.02	0.08	0.15
				GP	250	0.43	0.69	0.79	0.25	0.49	0.64	0.50	0.74	0.84	0.31	0.57	0.71
					1000	1.00	1.00	1.00	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
					50	0.06	0.17	0.25	0.06	0.17	0.25	0.08	0.19	0.28	0.08	0.19	0.28
LogNormal	10.00	2.16	0	LogGamma	250	0.34	0.55	0.66	0.34	0.55	0.65	0.40	0.61	0.72	0.40	0.61	0.72
					1000	0.97	0.99	1.00	0.97	0.99	1.00	0.99	1.00	1.00	0.99	1.00	1.00
					50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
				Gaussian	250	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
					1000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
					50	0.01	0.05	0.11	0.01	0.05	0.11	0.01	0.06	0.11	0.01	0.06	0.12
				Gamma	250	0.01	0.06	0.10	0.01	0.06	0.10	0.01	0.06	0.11	0.01	0.06	0.11
					1000	0.01	0.04	0.10	0.01	0.04	0.10	0.01	0.05	0.10	0.01	0.05	0.10
					50	0.00	0.04	0.10	0.01	0.04	0.10	0.01	0.05	0.10	0.01	0.05	0.10
Gaussian	4384605.98	62.00	0	LogNormal	250	0.01	0.06	0.12	0.01	0.06	0.12	0.01	0.06	0.11	0.01	0.06	0.11
					1000	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.06	0.10
					50	0.28	0.50	0.63	0.24	0.48	0.61	0.34	0.55	0.67	0.30	0.55	0.66
				Weibull	250	0.99	1.00	1.00	0.98	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00
					1000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
					50	0.40	0.62	0.72	0.21	0.43	0.56	0.41	0.62	0.72	0.26	0.48	0.62
				Gamma	250	1.00	1.00	1.00	0.97	0.99	1.00	1.00	1.00	1.00	0.98	1.00	1.00
					1000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
					50	0.60	0.80	0.88	0.59	0.80	0.88	0.80	0.91	0.94	0.80	0.91	0.94
Weibull	45000.00	0.30	0	LogGamma	250	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
					1000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
					50	0.31	0.54	0.65	0.32	0.54	0.66	0.37	0.61	0.71	0.37	0.61	0.71
				LogNormal	250	0.99	1.00	1.00	0.99	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00
					1000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 4.10: Powers of the direct and indirect tests based on on $\widehat{W}_n^{2,\beta}$ for $\beta \in \{1.5, 1.8, 1.95\}$ under the alternative hypothesis. The parameters $\boldsymbol{\theta} = \langle\theta_1, \theta_2\rangle$ are from the Second Set of Table 4.5.

Simulated as	θ_1	θ_2	Tail index	Estimated as	Sample size	Direct $\widehat{W}_n^{2,1.5}$			Indirect $\widehat{W}_n^{2,1.5}$			Direct $\widehat{W}_n^{2,1.8}$			Indirect $\widehat{W}_n^{2,1.8}$			Direct $\widehat{W}_n^{2,1.95}$			Indirect $\widehat{W}_n^{2,1.95}$		
						Significance level			Significance level			Significance level			Significance level			Significance level			Significance level		
						0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
					50	0.20	0.43	0.59	0.39	0.78	0.90	0.11	0.34	0.50	0.27	0.70	0.88	0.10	0.30	0.47	0.25	0.64	0.86
				GP	250	0.93	1.00	1.00	1.00	1.00	1.00	0.26	0.84	0.97	0.93	1.00	1.00	0.17	0.61	0.89	0.73	1.00	1.00
					1000	1.00	1.00	1.00	1.00	1.00	1.00	0.96	1.00	1.00	1.00	1.00	1.00	0.50	1.00	1.00	1.00	1.00	1.00
					50	0.25	0.47	0.59	0.24	0.46	0.57	0.19	0.45	0.60	0.20	0.45	0.59	0.18	0.44	0.58	0.18	0.44	0.57
Fréchet	1.01	5000.00	0.99	LogGamma	250	0.89	0.98	0.99	0.88	0.97	0.99	0.67	0.94	0.98	0.67	0.93	0.98	0.57	0.91	0.97	0.57	0.90	0.97
					1000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.98	1.00	1.00	0.98	1.00	1.00
					50	0.44	0.65	0.75	0.43	0.65	0.74	0.36	0.63	0.75	0.35	0.63	0.74	0.33	0.61	0.74	0.33	0.61	0.73
				LogNormal	250	0.99	1.00	1.00	0.99	1.00	1.00	0.90	0.99	1.00	0.90	0.99	1.00	0.82	0.99	1.00	0.82	0.99	1.00
					1000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
					50	0.83	0.99	1.00	0.88	0.99	1.00	0.00	0.85	0.99	0.00	0.91	1.00	0.00	0.39	0.96	0.00	0.58	0.98
				LogGamma	250	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.10	1.00	1.00	0.11	1.00	1.00
					1000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
					50	0.55	0.92	0.97	0.61	0.94	0.98	0.00	0.55	0.91	0.00	0.67	0.95	0.00	0.12	0.79	0.00	0.22	0.87
Gamma	0.27	8000000.00	0	LogNormal	250	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	0.00	1.00	1.00	0.00	1.00	1.00
					1000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
					50	0.03	0.20	0.32	0.02	0.19	0.34	0.00	0.04	0.19	0.00	0.02	0.19	0.00	0.00	0.09	0.00	0.00	0.09
				Weibull	250	0.72	0.93	0.98	0.74	0.95	0.99	0.00	0.67	0.91	0.00	0.72	0.95	0.00	0.16	0.79	0.00	0.18	0.85
					1000	1.00	1.00	1.00	1.00	1.00	1.00	0.96	1.00	1.00	0.99	1.00	1.00	0.00	1.00	1.00	0.00	1.00	1.00
					50	0.11	0.45	0.62	0.11	0.40	0.55	0.00	0.10	0.39	0.00	0.12	0.37	0.00	0.01	0.21	0.00	0.02	0.24
				Fréchet	250	1.00	1.00	1.00	0.97	1.00	1.00	0.37	0.99	1.00	0.07	0.95	0.99	0.00	0.90	0.99	0.00	0.61	0.96
					1000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	0.49	1.00	1.00
					50	0.04	0.16	0.26	0.05	0.19	0.28	0.02	0.07	0.17	0.02	0.08	0.19	0.02	0.05	0.13	0.02	0.06	0.16
GP	0.99	4954.25	0.99	LogGamma	250	0.37	0.66	0.78	0.39	0.67	0.79	0.02	0.28	0.54	0.03	0.30	0.57	0.02	0.10	0.35	0.02	0.11	0.37
					1000	0.99	1.00	1.00	0.99	1.00	1.00	0.32	0.96	0.99	0.34	0.96	0.99	0.03	0.64	0.93	0.03	0.67	0.94
					50	0.08	0.17	0.24	0.08	0.17	0.24	0.08	0.16	0.25	0.08	0.16	0.25	0.08	0.16	0.25	0.08	0.17	0.25
				LogNormal	250	0.28	0.45	0.57	0.28	0.46	0.57	0.22	0.43	0.55	0.22	0.43	0.56	0.20	0.41	0.54	0.20	0.41	0.54
					1000	0.83	0.93	0.97	0.83	0.93	0.97	0.57	0.85	0.93	0.57	0.85	0.93	0.48	0.79	0.88	0.48	0.79	0.88
					50	0.00	0.03	0.07	0.00	0.03	0.06	0.00	0.02	0.06	0.00	0.01	0.05	0.00	0.02	0.05	0.00	0.01	0.04
				Fréchet	250	0.01	0.04	0.10	0.01	0.05	0.11	0.00	0.01	0.04	0.00	0.01	0.06	0.00	0.01	0.03	0.00	0.01	0.03
					1000	0.05	0.20	0.32	0.07	0.23	0.34	0.00	0.05	0.18	0.00	0.05	0.20	0.00	0.01	0.09	0.00	0.01	0.09
					50	0.00	0.03	0.09	0.01	0.06	0.14	0.00	0.03	0.07	0.01	0.05	0.12	0.00	0.03	0.06	0.01	0.05	0.12
LogGamma	4.89	1.01	0.99	GP	250	0.00	0.06	0.13	0.03	0.20	0.40	0.00	0.01	0.05	0.01	0.04	0.16	0.00	0.01	0.03	0.01	0.03	0.12
					1000	0.07	0.31	0.51	0.81	0.99	1.00	0.00	0.04	0.15	0.01	0.35	0.77	0.00	0.01	0.06	0.00	0.04	0.32
					50	0.31	0.55	0.65	0.31	0.54	0.63	0.23	0.52	0.65	0.23	0.52	0.64	0.22	0.49	0.65	0.22	0.49	0.64
				LogNormal	250	0.96	0.99	1.00	0.96	0.99	1.00	0.76	0.97	0.99	0.77	0.97	0.99	0.65	0.95	0.99	0.65	0.95	0.99
					1000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
					50	0.00	0.08	0.17	0.00	0.04	0.12	0.00	0.01	0.07	0.00	0.00	0.05	0.00	0.00	0.03	0.00	0.00	0.01
				GP	250	0.24	0.64	0.78	0.20	0.58	0.74	0.00	0.17	0.56	0.00	0.17	0.56	0.00	0.00	0.31	0.00	0.00	0.34
					1000	1.00	1.00	1.00	1.00	1.00	1.00	0.21	1.00	1.00	0.30	1.00	1.00	0.00	0.83	1.00	0.00	0.91	1.00
					50	0.00	0.07	0.16	0.01	0.09	0.19	0.00	0.01	0.06	0.00	0.01	0.10	0.00	0.00	0.03	0.00	0.01	0.05
LogNormal	10.00	2.16	0	LogGamma	250	0.17	0.48	0.63	0.18	0.50	0.64	0.00	0.12	0.41	0.00	0.14	0.44	0.00	0.01	0.22	0.00	0.01	0.26
					1000	0.97	1.00	1.00	0.97	1.00	1.00	0.14	0.95	0.99	0.14	0.96	0.99	0.00	0.65	0.97	0.00	0.68	0.97
					50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
				Gaussian	250	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
					1000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
					50	0.01	0.05	0.10	0.01	0.06	0.10	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.04	0.10	0.01	0.04	0.10
				Gamma	250	0.01	0.05	0.10	0.01	0.06	0.10	0.01	0.04	0.09	0.01	0.04	0.09	0.01	0.04	0.09	0.01	0.05	0.09
					1000	0.01	0.06	0.10	0.01	0.06	0.10	0.01	0.04	0.10	0.01	0.04	0.11	0.01	0.04	0.10	0.01	0.05	0.10
					50	0.01	0.06	0.10	0.01	0.05	0.11	0.01	0.06	0.11	0.01	0.06	0.11	0.01	0.06	0.11	0.01	0.06	0.11
Gaussian	4384605.98	62.00	0	LogNormal	250	0.02	0.06	0.10	0.02	0.06	0.10	0.01	0.06	0.10	0.01	0.06	0.10	0.01	0.06	0.10	0.01	0.07	0.10
					1000	0.00	0.05	0.09	0.01	0.05	0.09	0.01	0.04	0.10	0.01	0.04	0.10	0.01	0.04	0.10	0.01	0.04	0.10
					50	0.35	0.54	0.66	0.40	0.62	0.73	0.28	0.53	0.66	0.32	0.60	0.73	0.26	0.51	0.66	0.30	0.57	0.72
				Weibull	250	0.97	0.99	1.00	0.99	1.00	1.00	0.81	0.99	0.99	0.91	0.99	1.00	0.70	0.96	0.99	0.84	0.99	1.00
					1000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00
					50	0.44	0.64	0.73	0.47	0.70	0.77	0.35	0.60	0.72	0.43	0.69	0.80	0.32	0.57	0.71	0.41	0.67	0.79
				Gamma	250	1.00	1.00	1.00	1.00	1.00	1.00	0.93	0.99	1.00	0.96	1.00	1.00	0.84	0.98	1.00	0.94	0.99	1.00
					1000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
					50	0.34	0.79	0.90	0.40	0.83	0.92	0.00	0.35	0.79	0.00	0.49	0.85	0.00	0.04	0.62	0.00	0.11	0.74
Weibull	45000.00	0.30	0	LogGamma	250	1.00	1.00	1.00	1.00	1.00	1.00	0.84	1.00	1.00	0.89	1.00	1.00	0.00	1.00	1.00	0.00	1.00	1.00
					1000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
					50	0.06	0.39	0.59	0.08	0.45	0.64	0.00	0.07	0.36	0.00	0.11	0.46	0.00	0.00	0.19	0.00	0.01	0.28
				LogNormal	250	0.99	1.00	1.00	0.99	1.00	1.00	0.08	0.99	1.00	0.09	0.99	1.00	0.00	0.79	0.99	0.00	0.84	1.00
					1000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.76	1.00	1.00	0.82	1.00	1.00

The AD test generally has greater power than CvM. Among the three $\widehat{W}_n^{2,\beta}$ statistics, $\widehat{W}_n^{2,1.5}$ provides the greatest power in most of the cases. Because CBT is a nonlinear transformation, it is not obvious a priori whether its application will result in a loss or gain in power. Tables 4.9 and 4.10 suggest that in 31%, 42% and 28% of the considered cases, indirect tests resulted in the same, greater, or smaller power, respectively, than that for the corresponding direct test. The average gain and loss in power amount to 7.5% and -4.5%, respectively. A general observation is that any advantage or disadvantage of an indirect test tend to disappear as sample size increases. There are three notable exceptions to this observation. Firstly, for $\widehat{W}_n^{2,\beta}$ ($\beta \in \{1.5, 1.8, 1.95\}$), when the null distribution is GP and alternative distribution is LogGamma, then the amount of gain in power increases (up to 74%) with sample size. For the first and third sets of distributional parameters from Table 4.5, when the null distribution is Fréchet and alternative distribution is LogGamma, then the indirect tests progressively lose power (up to -33%). For the Third Set, when the null distribution is GP and alternative distribution is Fréchet, indirect tests based on $\widehat{W}_n^{2,\beta}$ ($\beta \in \{1.8, 1.95\}$) tend to gain a considerable amount of power (up to 77%) with sample size growing.

We note each of the considered tests has weak power to discern the Gamma and LogNormal null distributions from the Gaussian distribution in the alternative. It is known (Johnson et al., 1994, p. 340 and p. 215) that the Gamma distribution and LogNormal distributions may very well mimic the Gaussian distribution. Indeed, under proper normalizations, they tend to the standard Gaussian distribution as the shape parameter $\alpha \to \infty$ for Gamma and as $\sigma \downarrow 0$ for LogNormal distribution.

4.5.2 Agreement of Decisions

For pairs of direct and indirect tests, we recorded the proportions of simultaneous rejection (of H_0 by the direct and of $H_0^\star$ by the indirect tests), failure to reject and rejection by direct combined with a failure to reject by indirect denoted by P_{rr}, P_{aa} and P_{ra}, respectively. Clearly, $P_{ar} = 1 - P_{rr} - P_{aa} - P_{ra}$ and the rejection rates of the direct and indirect tests equal, respectively, $P_r = P_{rr} + P_{ra}$ and $P_r^\star = P_{rr} + P_{ar}$.

We observe that the level of agreement is very good. P_{ra} and P_{ar}, measures of disagreement, are generally negligible, or close to 0. They are observed to tend to 0 with increasing sample size. Under the null, P_{rr} is small while P_{aa} is high. Under an alternative, $P_{rr} \to 1$ while $P_{aa} \to 0$ as sample size increases; see Tables 4.11-4.12.

Table 4.11: Rates of agreement of decisions made by the direct and indirect tests based on CvM and AD. The parameters $\boldsymbol{\theta} = \langle \theta_1, \theta_2 \rangle$ are from the Second Set of Table 4.5.

Simulated as	θ_1	θ_2	Tail index	Estimated as	Sample size	CvM												AD											
						T_{aa}			T_{rr}			T_{ra}			T_{ar}			T_{aa}			T_{rr}			T_{ra}			T_{ar}		
						Significance level			Significance level			Significance level			Significance level			Significance level			Significance level			Significance level			Significance level		
						0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
Fréchet	1.01	5000.00	0.99	Fréchet	50	0.99	0.95	0.89	0.01	0.04	0.08	0.00	0.01	0.03	0.00	0.01	0.00	0.98	0.95	0.90	0.01	0.04	0.08	0.00	0.01	0.01	0.00	0.00	0.01
					250	0.98	0.94	0.89	0.01	0.05	0.09	0.00	0.01	0.01	0.00	0.01	0.01	0.98	0.95	0.89	0.01	0.05	0.08	0.01	0.01	0.02	0.00	0.00	0.01
					1000	0.99	0.93	0.88	0.01	0.05	0.11	0.01	0.02	0.01	0.00	0.01	0.01	0.99	0.93	0.88	0.01	0.05	0.11	0.00	0.01	0.01	0.00	0.00	0.01
				GP	50	0.45	0.19	0.10	0.36	0.67	0.80	0.12	0.09	0.07	0.06	0.05	0.04	0.37	0.11	0.05	0.46	0.78	0.89	0.12	0.09	0.05	0.05	0.02	0.01
					250	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
					1000	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
				LogGamma	50	0.84	0.67	0.56	0.15	0.32	0.43	0.00	0.00	0.00	0.01	0.01	0.01	0.81	0.62	0.52	0.18	0.37	0.47	0.00	0.00	0.00	0.00	0.00	0.01
					250	0.15	0.05	0.02	0.84	0.95	0.98	0.01	0.00	0.00	0.00	0.00	0.00	0.10	0.03	0.01	0.90	0.97	0.99	0.00	0.00	0.00	0.00	0.00	0.00
					1000	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
				LogNormal	50	0.65	0.45	0.34	0.33	0.54	0.66	0.01	0.00	0.00	0.01	0.01	0.01	0.59	0.40	0.28	0.39	0.59	0.71	0.01	0.01	0.00	0.01	0.01	0.01
					250	0.01	0.00	0.00	0.99	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
					1000	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
Gamma	0.27	8000000.00	0	Gamma	50	0.99	0.95	0.89	0.01	0.04	0.08	0.01	0.01	0.02	0.00	0.01	0.01	0.98	0.95	0.88	0.01	0.03	0.08	0.00	0.01	0.03	0.01	0.01	0.01
					250	0.98	0.95	0.90	0.01	0.03	0.06	0.01	0.02	0.02	0.00	0.01	0.02	0.98	0.95	0.90	0.01	0.04	0.07	0.01	0.01	0.02	0.00	0.01	0.02
					1000	0.99	0.94	0.89	0.01	0.03	0.07	0.00	0.02	0.03	0.00	0.01	0.02	0.99	0.95	0.89	0.01	0.03	0.08	0.00	0.02	0.02	0.00	0.01	0.01
				LogGamma	50	0.06	0.02	0.01	0.93	0.98	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.01	0.00	0.97	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
					250	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
					1000	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
				LogNormal	50	0.18	0.07	0.04	0.81	0.93	0.96	0.01	0.00	0.00	0.00	0.00	0.00	0.13	0.05	0.02	0.87	0.95	0.98	0.00	0.00	0.00	0.01	0.00	0.00
					250	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
					1000	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
				Weibull	50	0.87	0.72	0.61	0.11	0.23	0.34	0.02	0.04	0.04	0.01	0.01	0.01	0.84	0.67	0.56	0.13	0.26	0.37	0.04	0.06	0.07	0.00	0.00	0.01
					250	0.22	0.08	0.05	0.72	0.88	0.94	0.06	0.04	0.02	0.00	0.00	0.00	0.14	0.04	0.02	0.81	0.93	0.97	0.05	0.03	0.01	0.00	0.00	0.00
					1000	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
GP	0.99	4954.25	0.99	Fréchet	50	0.54	0.36	0.28	0.40	0.59	0.68	0.06	0.05	0.03	0.00	0.01	0.01	0.48	0.31	0.25	0.46	0.64	0.71	0.06	0.04	0.03	0.00	0.01	0.01
					250	0.00	0.00	0.00	0.99	1.00	1.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
					1000	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
				GP	50	0.99	0.95	0.89	0.01	0.02	0.04	0.01	0.03	0.05	0.00	0.01	0.01	0.99	0.95	0.90	0.01	0.02	0.05	0.00	0.03	0.05	0.00	0.01	0.01
					250	0.99	0.96	0.90	0.00	0.02	0.05	0.00	0.02	0.04	0.00	0.00	0.02	0.99	0.96	0.91	0.00	0.02	0.05	0.01	0.02	0.04	0.00	0.01	0.01
					1000	0.99	0.95	0.89	0.00	0.03	0.06	0.01	0.02	0.04	0.00	0.00	0.01	0.99	0.95	0.89	0.00	0.03	0.06	0.00	0.03	0.04	0.00	0.00	0.01
				LogGamma	50	0.83	0.70	0.60	0.16	0.29	0.39	0.00	0.00	0.00	0.00	0.01	0.01	0.81	0.67	0.57	0.18	0.32	0.41	0.00	0.00	0.00	0.01	0.01	0.01
					250	0.29	0.17	0.12	0.71	0.83	0.88	0.00	0.00	0.00	0.00	0.00	0.00	0.23	0.13	0.09	0.76	0.86	0.91	0.00	0.00	0.00	0.01	0.00	0.00
					1000	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
				LogNormal	50	0.96	0.87	0.78	0.04	0.12	0.20	0.00	0.00	0.00	0.00	0.01	0.02	0.95	0.85	0.76	0.05	0.14	0.22	0.00	0.00	0.00	0.01	0.01	0.02
					250	0.79	0.60	0.48	0.20	0.39	0.51	0.00	0.00	0.00	0.01	0.01	0.01	0.75	0.54	0.43	0.24	0.45	0.56	0.00	0.00	0.00	0.01	0.01	0.01
					1000	0.16	0.05	0.02	0.83	0.95	0.98	0.00	0.00	0.00	0.01	0.00	0.00	0.10	0.03	0.01	0.90	0.97	0.99	0.00	0.00	0.00	0.00	0.00	0.00
LogGamma	4.89	1.01	0.99	Fréchet	50	0.99	0.94	0.88	0.01	0.05	0.10	0.00	0.01	0.01	0.00	0.00	0.01	0.99	0.93	0.89	0.01	0.05	0.10	0.00	0.01	0.01	0.00	0.01	0.01
					250	0.97	0.92	0.85	0.02	0.07	0.12	0.00	0.01	0.01	0.00	0.01	0.01	0.98	0.91	0.84	0.02	0.07	0.13	0.00	0.01	0.01	0.00	0.01	0.01
					1000	0.91	0.74	0.62	0.07	0.22	0.33	0.00	0.02	0.03	0.02	0.03	0.03	0.89	0.70	0.58	0.08	0.25	0.38	0.00	0.01	0.02	0.03	0.03	0.03
				GP	50	0.97	0.87	0.79	0.01	0.07	0.15	0.00	0.01	0.02	0.01	0.05	0.05	0.97	0.87	0.77	0.02	0.08	0.18	0.00	0.01	0.02	0.02	0.04	0.04
					250	0.64	0.33	0.19	0.10	0.32	0.49	0.00	0.00	0.00	0.26	0.35	0.32	0.49	0.18	0.08	0.32	0.67	0.82	0.00	0.00	0.00	0.19	0.14	0.09
					1000	0.00	0.00	0.00	0.86	0.99	1.00	0.00	0.00	0.00	0.14	0.01	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
				LogGamma	50	0.99	0.94	0.89	0.01	0.05	0.10	0.00	0.00	0.01	0.00	0.00	0.00	0.99	0.95	0.90	0.01	0.05	0.10	0.00	0.00	0.00	0.00	0.00	0.00
					250	0.99	0.94	0.88	0.01	0.06	0.11	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.94	0.88	0.01	0.06	0.11	0.00	0.00	0.00	0.00	0.00	0.00
					1000	0.99	0.95	0.90	0.01	0.05	0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.95	0.90	0.01	0.05	0.10	0.00	0.00	0.00	0.00	0.00	0.00
				LogNormal	50	0.77	0.57	0.45	0.23	0.41	0.55	0.00	0.01	0.00	0.00	0.01	0.01	0.72	0.51	0.39	0.27	0.48	0.60	0.00	0.00	0.01	0.01	0.01	0.01
					250	0.03	0.01	0.00	0.96	0.99	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.99	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
					1000	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
LogNormal	10.00	2.16	0	GP	50	0.95	0.84	0.74	0.01	0.07	0.16	0.03	0.09	0.10	0.00	0.00	0.00	0.94	0.83	0.72	0.02	0.08	0.15	0.04	0.09	0.12	0.00	0.00	0.00
					250	0.57	0.31	0.21	0.25	0.49	0.64	0.18	0.20	0.15	0.00	0.00	0.00	0.49	0.26	0.16	0.31	0.57	0.71	0.20	0.17	0.13	0.01	0.00	0.00
					1000	0.00	0.00	0.00	0.98	1.00	1.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
				LogGamma	50	0.94	0.83	0.74	0.06	0.16	0.25	0.00	0.01	0.01	0.00	0.01	0.00	0.92	0.81	0.71	0.08	0.18	0.28	0.00	0.01	0.01	0.00	0.01	0.01
					250	0.66	0.45	0.34	0.34	0.55	0.65	0.01	0.00	0.01	0.00	0.00	0.00	0.60	0.39	0.28	0.40	0.61	0.71	0.00	0.00	0.00	0.00	0.00	0.00
					1000	0.03	0.01	0.00	0.97	0.99	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.99	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
				LogNormal	50	0.99	0.95	0.90	0.01	0.05	0.10	0.00	0.00	0.01	0.00	0.00	0.00	0.99	0.95	0.90	0.01	0.05	0.09	0.00	0.00	0.00	0.00	0.00	0.00
					250	0.98	0.94	0.89	0.02	0.06	0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.98	0.94	0.89	0.02	0.06	0.11	0.00	0.00	0.00	0.00	0.00	0.00
					1000	0.99	0.94	0.90	0.01	0.06	0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.94	0.90	0.01	0.06	0.10	0.00	0.00	0.00	0.00	0.00	0.00
				Gaussian	50	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
					250	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
					1000	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
Gaussian	4384605.98	62.00	0	Gamma	50	0.99	0.95	0.89	0.01	0.05	0.10	0.00	0.00	0.01	0.00	0.00	0.01	0.99	0.94	0.88	0.01	0.05	0.11	0.00	0.00	0.00	0.00	0.00	0.01
					250	0.99	0.94	0.90	0.01	0.06	0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.94	0.89	0.01	0.06	0.11	0.00	0.00	0.00	0.00	0.00	0.00
					1000	0.99	0.96	0.90	0.01	0.04	0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.95	0.90	0.01	0.05	0.10	0.00	0.00	0.00	0.00	0.00	0.00
				LogNormal	50	1.00	0.96	0.90	0.00	0.04	0.09	0.00	0.00	0.00	0.00	0.01	0.00	1.00	0.95	0.90	0.01	0.05	0.10	0.00	0.00	0.00	0.00	0.00	0.01
					250	0.99	0.94	0.88	0.01	0.06	0.12	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.94	0.89	0.01	0.05	0.11	0.00	0.00	0.00	0.00	0.00	0.00
					1000	0.99	0.95	0.90	0.01	0.05	0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.95	0.90	0.01	0.05	0.09	0.00	0.00	0.00	0.00	0.00	0.00
				Gaussian	50	0.99	0.95	0.89	0.01	0.05	0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.95	0.90	0.01	0.04	0.10	0.00	0.00	0.00	0.00	0.00	0.00
					250	1.00	0.96	0.91	0.00	0.04	0.09	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.96	0.90	0.00	0.04	0.09	0.00	0.00	0.00	0.00	0.00	0.00
					1000	0.99	0.95	0.91	0.01	0.05	0.09	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.96	0.91	0.01	0.05	0.09	0.00	0.00	0.00	0.00	0.00	0.00
				Weibull	50	0.71	0.49	0.36	0.23	0.46	0.60	0.05	0.04	0.03	0.01	0.02	0.01	0.65	0.43	0.32	0.29	0.53	0.65	0.05	0.02	0.02	0.01	0.02	0.01
					250	0.01	0.00	0.00	0.98	1.00	1.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
					1000	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
Weibull	45000.00	0.30	0	Gamma	50	0.59	0.38	0.28	0.21	0.43	0.55	0.20	0.19	0.17	0.00	0.00	0.00	0.59	0.38	0.27	0.26	0.48	0.61	0.15	0.15	0.11	0.00	0.00	0.00
					250	0.00	0.00	0.00	0.97	0.99	1.00	0.03	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.98	1.00	1.00	0.02	0.00	0.00	0.00	0.00	0.00
					1000	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
				LogGamma	50	0.40	0.20	0.12	0.59	0.80	0.88	0.01	0.01	0.00	0.00	0.00	0.00	0.20	0.09	0.06	0.80	0.91	0.94	0.00	0.00	0.00	0.00	0.00	0.00
					250	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
					1000	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
				LogNormal	50	0.68	0.45	0.34	0.31	0.53	0.65	0.00	0.01	0.00	0.01	0.01	0.01	0.62	0.39	0.29	0.37	0.60	0.70	0.00	0.01	0.00	0.01	0.01	0.01
					250	0.01	0.00	0.00	0.99	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.99	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
					1000	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
				Weibull	50	0.99	0.93	0.89	0.01	0.05	0.09	0.00	0.01	0.02	0.00	0.01	0.01	0.99	0.94	0.88	0.01	0.05	0.10	0.00	0.01	0.02	0.00	0.01	0.01
					250	0.99	0.94	0.89	0.01	0.05	0.10	0.00	0.01	0.01	0.00	0.00	0.01	0.98	0.93	0.89	0.01	0.05	0.10	0.00	0.01	0.01	0.00	0.01	0.01
					1000	0.99	0.94	0.90	0.01	0.04	0.09	0.00	0.01	0.01	0.00	0.01	0.00	0.99	0.95	0.89	0.01	0.04	0.08	0.00	0.00	0.02	0.00	0.00	0.00

Table 4.12: Rates of agreement of decisions made by the direct and indirect tests based on $\widehat{W}_n^{2,\beta}$ for $\beta \in \{1.5, 1.8, 1.95\}$ at significance level $\alpha = 0.1$. The parameters $\boldsymbol{\theta} = \langle \theta_1, \theta_2 \rangle$ are from the Second Set of Table 4.5.

Simulated as	θ_1	θ_2	Tail index	Estimated as	Sample size	$\widehat{W}_n^{2,1.5}$				$\widehat{W}_n^{2,1.8}$				$\widehat{W}_n^{2,1.95}$			
						Taa	Trr	Tra	Tar	Taa	Trr	Tra	Tar	Taa	Trr	Tra	Tar
Fréchet	1.01	5000.00	0.99	Fréchet	50	0.89	0.08	0.02	0.01	0.87	0.09	0.03	0.01	0.88	0.09	0.02	0.01
					250	0.88	0.07	0.03	0.02	0.89	0.09	0.01	0.01	0.88	0.10	0.01	0.01
					1000	0.88	0.09	0.01	0.02	0.88	0.09	0.02	0.01	0.89	0.10	0.01	0.00
				GP	50	0.10	0.59	0.00	0.31	0.12	0.50	0.00	0.38	0.14	0.47	0.00	0.39
					250	0.00	1.00	0.00	0.00	0.00	0.97	0.00	0.03	0.00	0.89	0.00	0.11
					1000	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00
				LogGamma	50	0.42	0.57	0.02	0.00	0.41	0.59	0.01	0.00	0.42	0.57	0.00	0.00
					250	0.01	0.99	0.00	0.00	0.02	0.98	0.00	0.00	0.03	0.97	0.00	0.00
					1000	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00
				LogNormal	50	0.25	0.74	0.01	0.00	0.25	0.74	0.01	0.00	0.26	0.73	0.01	0.00
					250	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00
					1000	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00
Gamma	0.27	8000000.00	0	Gamma	50	0.89	0.09	0.01	0.01	0.88	0.10	0.02	0.01	0.87	0.10	0.01	0.01
					250	0.90	0.09	0.01	0.01	0.89	0.09	0.01	0.01	0.88	0.09	0.01	0.01
					1000	0.90	0.09	0.01	0.00	0.90	0.09	0.01	0.01	0.90	0.09	0.01	0.01
				LogGamma	50	0.00	1.00	0.00	0.00	0.00	0.99	0.00	0.01	0.02	0.96	0.00	0.03
					250	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00
					1000	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00
				LogNormal	50	0.02	0.97	0.00	0.01	0.05	0.91	0.00	0.04	0.13	0.79	0.00	0.08
					250	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00
					1000	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00
				Weibull	50	0.65	0.30	0.02	0.04	0.78	0.17	0.02	0.03	0.89	0.07	0.02	0.01
					250	0.02	0.98	0.00	0.01	0.05	0.91	0.00	0.04	0.14	0.78	0.00	0.07
					1000	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00
GP	0.99	4954.25	0.99	Fréchet	50	0.37	0.54	0.08	0.01	0.58	0.34	0.04	0.03	0.74	0.19	0.02	0.05
					250	0.00	1.00	0.00	0.00	0.00	0.99	0.01	0.00	0.01	0.96	0.03	0.00
					1000	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00
				GP	50	0.89	0.06	0.03	0.02	0.88	0.07	0.03	0.02	0.89	0.07	0.02	0.02
					250	0.89	0.07	0.04	0.01	0.89	0.08	0.03	0.01	0.89	0.08	0.02	0.01
					1000	0.88	0.07	0.05	0.01	0.89	0.08	0.02	0.00	0.89	0.09	0.01	0.01
				LogGamma	50	0.72	0.26	0.00	0.03	0.81	0.17	0.00	0.02	0.84	0.13	0.00	0.03
					250	0.21	0.78	0.00	0.02	0.43	0.54	0.00	0.03	0.63	0.35	0.00	0.02
					1000	0.00	1.00	0.00	0.00	0.01	0.99	0.00	0.00	0.06	0.93	0.00	0.01
				LogNormal	50	0.76	0.23	0.00	0.01	0.74	0.25	0.01	0.00	0.74	0.25	0.00	0.01
					250	0.43	0.57	0.00	0.01	0.45	0.55	0.00	0.00	0.46	0.53	0.01	0.00
					1000	0.03	0.97	0.00	0.00	0.07	0.93	0.00	0.00	0.12	0.88	0.00	0.01
LogGamma	4.89	1.01	0.99	Fréchet	50	0.92	0.05	0.02	0.01	0.93	0.04	0.01	0.01	0.95	0.04	0.02	0.00
					250	0.89	0.09	0.01	0.02	0.94	0.04	0.00	0.02	0.97	0.03	0.00	0.01
					1000	0.65	0.31	0.01	0.03	0.80	0.17	0.01	0.03	0.90	0.08	0.01	0.01
				GP	50	0.85	0.08	0.01	0.06	0.88	0.06	0.01	0.06	0.88	0.06	0.00	0.06
					250	0.59	0.12	0.01	0.28	0.83	0.05	0.01	0.12	0.88	0.03	0.01	0.09
					1000	0.00	0.51	0.00	0.49	0.23	0.15	0.00	0.62	0.68	0.05	0.01	0.26
				LogGamma	50	0.88	0.11	0.00	0.00	0.88	0.11	0.00	0.00	0.89	0.11	0.00	0.00
					250	0.89	0.10	0.01	0.00	0.87	0.12	0.00	0.00	0.87	0.13	0.00	0.00
					1000	0.89	0.10	0.00	0.00	0.89	0.11	0.00	0.00	0.89	0.11	0.00	0.00
				LogNormal	50	0.35	0.63	0.01	0.00	0.35	0.64	0.01	0.00	0.36	0.64	0.01	0.00
					250	0.00	1.00	0.00	0.00	0.01	0.99	0.00	0.00	0.01	0.99	0.00	0.00
					1000	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00
LogNormal	10.00	2.16	0	GP	50	0.83	0.11	0.06	0.01	0.92	0.04	0.03	0.00	0.97	0.01	0.02	0.00
					250	0.21	0.73	0.05	0.01	0.41	0.52	0.03	0.03	0.63	0.28	0.03	0.05
					1000	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00
				LogGamma	50	0.81	0.16	0.00	0.02	0.90	0.06	0.00	0.04	0.95	0.03	0.00	0.01
					250	0.36	0.63	0.00	0.02	0.56	0.41	0.00	0.03	0.74	0.22	0.00	0.04
					1000	0.00	1.00	0.00	0.00	0.01	0.99	0.00	0.00	0.03	0.97	0.00	0.00
				LogNormal	50	0.89	0.10	0.01	0.00	0.90	0.10	0.00	0.00	0.90	0.10	0.00	0.00
					250	0.88	0.12	0.00	0.00	0.88	0.12	0.00	0.00	0.89	0.11	0.00	0.00
					1000	0.90	0.10	0.00	0.00	0.91	0.09	0.00	0.00	0.91	0.09	0.00	0.00
				Gaussian	50	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00
					250	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00
					1000	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00
Gaussian	4384605.98	62.00	0	Gamma	50	0.90	0.10	0.00	0.00	0.89	0.10	0.00	0.00	0.90	0.10	0.00	0.00
					250	0.90	0.10	0.00	0.00	0.91	0.09	0.00	0.00	0.91	0.09	0.00	0.00
					1000	0.90	0.10	0.00	0.00	0.89	0.10	0.00	0.00	0.90	0.09	0.00	0.00
				LogNormal	50	0.89	0.10	0.00	0.01	0.88	0.11	0.01	0.00	0.89	0.11	0.00	0.00
					250	0.90	0.10	0.00	0.00	0.90	0.10	0.00	0.00	0.90	0.10	0.00	0.00
					1000	0.90	0.09	0.00	0.00	0.90	0.10	0.00	0.00	0.90	0.10	0.00	0.00
				Gaussian	50	0.88	0.11	0.01	0.00	0.89	0.10	0.00	0.00	0.89	0.10	0.00	0.00
					250	0.91	0.09	0.00	0.00	0.90	0.10	0.00	0.00	0.90	0.10	0.00	0.00
					1000	0.91	0.09	0.00	0.00	0.91	0.09	0.00	0.00	0.90	0.10	0.00	0.00
				Weibull	50	0.27	0.66	0.00	0.08	0.27	0.66	0.00	0.07	0.28	0.66	0.00	0.07
					250	0.00	1.00	0.00	0.00	0.00	0.99	0.00	0.01	0.00	0.99	0.00	0.01
					1000	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00
Weibull	45000.00	0.30	0	Gamma	50	0.23	0.73	0.00	0.03	0.20	0.72	0.00	0.08	0.21	0.71	0.00	0.09
					250	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.99	0.00	0.01
					1000	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00
				LogGamma	50	0.08	0.90	0.00	0.02	0.16	0.79	0.00	0.06	0.26	0.62	0.00	0.12
					250	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00
					1000	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00
				LogNormal	50	0.36	0.59	0.00	0.05	0.54	0.36	0.00	0.10	0.72	0.19	0.00	0.09
					250	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.01	0.99	0.00	0.00
					1000	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00
				Weibull	50	0.87	0.10	0.02	0.01	0.89	0.09	0.01	0.01	0.89	0.09	0.01	0.01
					250	0.90	0.09	0.01	0.01	0.89	0.09	0.01	0.01	0.90	0.09	0.01	0.01
					1000	0.90	0.08	0.01	0.00	0.89	0.10	0.01	0.01	0.89	0.10	0.01	0.00

4.5.3 Independence of Decisions

Let us consider $P_{rr} - P_r P_r^\star$. It is readily seen to be a measure of independence of the decisions made by a pair of the direct and indirect tests. This quantity has been tabulated in Table 4.13.

We observe that generally, when the null and alternative distributions are different, $P_{rr} - P_r P_r^\star \to 0$ as n grows. That is, the decisions made by the direct and indirect tests appear to be asymptotically independent.

4.6 Application to Real Data

We apply the direct and indirect goodness-of-fit tests to a set of real data of losses caused by operational risk events. The source of the data is Operational Risk Exchange (ORX), the world's leading OpRisk loss data consortium for financial institutions. ORX has over 80 members and a database of more than 500,000 worldwide loss events.[7] The data set was accessed through Royal Bank of Canada with permission from ORX.[8]

This data set contains 3,034 severities of losses incurred by financial institutions, members of ORX in North America and Western Europe, from January 01, 2002 to June 30, 2016. This, on average, corresponds to approximately $\lambda = 204.227$ events per year. The business line and event type which we chose to consider are Capital Markets and "Clients, Products & Business Practices", respectively. The latter includes, for example, fiduciary breaches, aggressive sales, account churning, market manipulation, insider trading (on firm's account) and unlicensed activity (Basel Committee on Banking Supervision, 2004). The heavytail nature of the data set is apparent from Table 4.14, which contains some descriptive characteristics of the data.

Due to this fact, from Table 4.4, we consider all but the Gaussian and Gamma distributions. As ORX only records losses above €$20,000$, we also include the left-truncated versions of the distributions. One could also subtract the threshold from data and fit regular distributions. This results in using the so-called "shifted distributions" which, according to the regulatory note Board of Governors of the

[7]For more information on ORX, see `https://managingrisktogether.orx.org/`.

[8]Disclaimer: The views expressed in this thesis are the author's own and do not represent the views of ORX, Royal Bank of Canada or any other institution.

Table 4.13: Measure $P_{rr} - P_r P_r^{\star}$ of the independence of decisions made by the direct and indirect tests based on $\widehat{W}_n^{2,\beta}$ for $\beta \in \{1.5, 1.8, 1.95\}$ at significance level $\alpha = 0.1$. The parameters $\boldsymbol{\theta} = \langle \theta_1, \theta_2 \rangle$ are from the Second Set of Table 4.5.

Simulate as	Parm1	Parm2	Tail index	Estimate as	Sample size	CvM	AD	$\widehat{W}_n^{2,1.5}$	$\widehat{W}_n^{2,1.8}$	$\widehat{W}_n^{2,1.95}$
Fréchet	1.01	5000.00	0.99	Fréchet	50	0.070	0.070	0.073	0.074	0.075
					250	0.081	0.074	0.064	0.082	0.084
					1000	0.094	0.097	0.082	0.080	0.085
				GP	50	0.076	0.043	0.062	0.062	0.066
					250	0.000	0.000	0.000	0.000	0.000
					1000	0.000	0.000	0.000	0.000	0.000
				LogGamma	50	0.240	0.245	0.236	0.237	0.243
					250	0.023	0.012	0.008	0.020	0.032
					1000	0.000	0.000	0.000	0.000	0.000
				LogNormal	50	0.219	0.199	0.183	0.184	0.191
					250	0.000	0.000	0.000	0.000	0.002
					1000	0.000	0.000	0.000	0.000	0.000
Gamma	0.27	8000000.00	0	Gamma	50	0.066	0.069	0.083	0.083	0.090
					250	0.055	0.058	0.077	0.083	0.082
					1000	0.063	0.067	0.081	0.078	0.081
				LogGamma	50	0.006	0.000	0.000	0.003	0.017
					250	0.000	0.000	0.000	0.000	0.000
					1000	0.000	0.000	0.000	0.000	0.000
				LogNormal	50	0.035	0.020	0.021	0.046	0.102
					250	0.000	0.000	0.000	0.000	0.000
					1000	0.000	0.000	0.000	0.000	0.000
				Weibull	50	0.205	0.203	0.194	0.132	0.065
					250	0.043	0.017	0.015	0.047	0.113
					1000	0.000	0.000	0.000	0.000	0.000
Gaussian	4384605.98	62.00	0	Gamma	50	0.089	0.097	0.085	0.090	0.089
					250	0.087	0.094	0.089	0.084	0.083
					1000	0.090	0.085	0.088	0.092	0.084
				Gaussian	50	0.090	0.088	0.096	0.092	0.092
					250	0.077	0.083	0.079	0.086	0.089
					1000	0.078	0.082	0.082	0.083	0.087
				LogNormal	50	0.082	0.085	0.088	0.095	0.096
					250	0.102	0.098	0.086	0.091	0.091
					1000	0.090	0.084	0.082	0.086	0.087
				Weibull	50	0.215	0.208	0.174	0.178	0.183
					250	0.001	0.000	0.000	0.001	0.002
					1000	0.000	0.000	0.000	0.000	0.000
GP	0.99	4954.25	0.99	Fréchet	50	0.191	0.178	0.196	0.198	0.135
					250	0.001	0.001	0.001	0.001	0.010
					1000	0.000	0.000	0.000	0.000	0.000
				GP	50	0.038	0.042	0.054	0.057	0.061
					250	0.040	0.041	0.059	0.068	0.073
					1000	0.051	0.054	0.059	0.072	0.083
				LogGamma	50	0.233	0.236	0.185	0.139	0.111
					250	0.101	0.080	0.160	0.234	0.219
					1000	0.000	0.000	0.000	0.010	0.054
				LogNormal	50	0.156	0.167	0.176	0.183	0.184
					250	0.243	0.239	0.242	0.246	0.244
					1000	0.021	0.014	0.028	0.065	0.101
LogGamma	4.89	1.01	0.99	Fréchet	50	0.086	0.086	0.050	0.038	0.034
					250	0.104	0.111	0.079	0.035	0.024
					1000	0.202	0.219	0.200	0.134	0.075
				GP	50	0.113	0.135	0.066	0.052	0.049
					250	0.093	0.068	0.071	0.038	0.022
					1000	0.000	0.000	0.002	0.035	0.034
				LogGamma	50	0.087	0.085	0.096	0.101	0.097
					250	0.099	0.099	0.092	0.107	0.110
					1000	0.086	0.088	0.091	0.095	0.095
				LogNormal	50	0.243	0.232	0.224	0.224	0.227
					250	0.002	0.001	0.004	0.008	0.013
					1000	0.000	0.000	0.000	0.000	0.000
LogNormal	10.00	2.16	0	Gaussian	50	0.000	0.000	0.000	0.000	0.000
					250	0.000	0.000	0.000	0.000	0.000
					1000	0.000	0.000	0.000	0.000	0.000
				GP	50	0.118	0.110	0.092	0.040	0.012
					250	0.134	0.116	0.153	0.214	0.178
					1000	0.000	0.000	0.000	0.000	0.000
				LogGamma	50	0.182	0.195	0.133	0.056	0.031
					250	0.221	0.201	0.224	0.230	0.163
					1000	0.003	0.002	0.003	0.006	0.030
				LogNormal	50	0.086	0.085	0.089	0.088	0.089
					250	0.090	0.095	0.102	0.105	0.097
					1000	0.087	0.087	0.086	0.083	0.083
Weibull	45000.00	0.30	0	Gamma	50	0.153	0.167	0.171	0.145	0.145
					250	0.000	0.000	0.000	0.000	0.000
					1000	0.000	0.000	0.000	0.000	0.000
				LogGamma	50	0.105	0.057	0.074	0.122	0.160
					250	0.000	0.000	0.000	0.000	0.000
					1000	0.000	0.000	0.000	0.000	0.000
				LogNormal	50	0.219	0.203	0.211	0.195	0.137
					250	0.002	0.001	0.001	0.001	0.005
					1000	0.000	0.000	0.000	0.000	0.000
				Weibull	50	0.080	0.088	0.089	0.080	0.084
					250	0.084	0.088	0.077	0.079	0.079
					1000	0.078	0.075	0.075	0.088	0.088

Table 4.14: Descriptive statistics of the OpRisk loss data set

Min	Max	Mean	StDev	Q25	Q50	Q75	Skewness	Kurtosis
20,000	3,834,093,911	11,952,929	124,517,475	45,300	134,495	698,088	19.98	482.98

Federal Reserve System (2014, p. 13), "may introduce bias into capital estimates by not capturing the probability distribution for loss severities below the modeling threshold." For this reason, we chose not to include shifted distributions in the pool of candidates.

The significance level α was chosen to be 0.10. The p-values for the direct tests were calculated by means of $10,000$ trials of a parametric bootstrap. The critical values of the indirect tests based on $\widehat{W}_n^{2,\beta}$ ($\beta \in \{1.5, 1.8, 1.95\}$) were determined by substituting coefficients $\{\hat{a}_i\}_{j=0}^{3}$ taken from Table 4.2 and sample size $n = 3,034$ into $1/(\hat{a}_0 + \hat{a}_1 n^{-1/3} + \hat{a}_2 n^{-1/2} + \hat{a}_3 n^{-1})$. In Table 4.15, the results of the assessment is presented.

Table 4.15: Decisions made by the direct (indirect) tests on assessing the null hypothesis H_0 ($H_0^\star$) at significance level $\alpha = 0.10$. Here "A" and "R" stand for "fails to reject" and "rejects", respectively. The last column contains the ratio of execution time between that for the direct and indirect tests.

Distribution	Type	$\hat{\theta}_1$	$\hat{\theta}_2$	NLL		Decision										Execution Time, sec		Time Ratio
						CvM		AD		$\widehat{W}_n^{2,1.5}$		$\widehat{W}_n^{2,1.8}$		$\widehat{W}_n^{2,1.95}$				
				Direct	Indirect	Direct	Indirect	Direct	Indirect	Direct	Indirect	Direct	Indirect	Direct	Indirect	Direct	Indirect	
Weibull	Truncated	0.002	0.084	43267.161	4291.519	A	A	A	A	A	A	A	A	A	A	607.207	0.004	172458.24
LogNormal	Truncated	4.605	4.861	43268.354	4298.583	A	A	A	A	A	A	A	A	A	A	105.220	0.003	32779.81
LogGamma	Truncated	9.320	1.009	43271.372	4308.848	R	A	A	A	R	R	A	A	A	A	159.054	0.003	45954.20
Fréchet	Truncated	0.497	15828.319	43280.904	4296.954	R	R	R	R	R	R	R	R	R	R	213.579	0.002	86648.04
GP	Truncated	2.158	23641.743	43288.314	4298.993	R	R	R	R	R	R	R	R	R	R	144.017	0.003	45210.95
Fréchet	Regular	0.680	93889.730	43641.949	4437.091	R	R	R	R	R	R	R	R	R	R	6.615	0.013	506.78
GP	Regular	1.676	132454.272	43902.535	3971.351	R	R	R	R	R	R	R	R	R	R	22.357	0.011	1981.30
LogGamma	Regular	37.995	3.071	43927.598	4304.934	R	R	R	R	R	R	R	R	R	R	8.038	0.004	1937.86
LogNormal	Regular	12.372	2.113	44110.295	4305.060	R	R	R	R	R	R	R	R	R	R	4.239	0.027	157.78
Weibull	Regular	754480.374	0.373	44876.111	3853.222	R	R	R	R	R	R	R	R	R	R	6.924	0.015	455.10

The order of the distributions corresponds to the ascending order of the values of the negative log-likelihood function (NLL) for the direct tests. As each distribution has two parameters, this order corresponds to the ascending order implied by the Akaike Information Criterion as well. Only the left-truncated Weibull, LogNormal and LogGamma distributions appear to be suitable models for the given data set. Table 4.15 suggests that according to both direct and indirect tests, the left-truncated Weibull distribution is the best at describing the data in the body and tails. The decisions made by the direct and indirect tests agree very well. The only exception is CvM for LogGamma where the direct test rejects the null while the indirect test fails to do so.

The last three columns in Table 4.15 provide evidence on clear advantage of the indirect tests over their direct counterparts in terms of execution time. The indirect tests for the Weibull distribution are $172,458$ times faster than the direct tests. Without using parallel computing, the advantage will be even more drastic. This may be a decisive factor in applications.

We conclude this section by demonstrating the fact that the selection of best-fit severity distributions properly capturing the upper tail behavior is essential for accurate modeling. In Table 4.16, *VaR* of the aggregate loss distribution is determined for $\delta = 0.999$ and $\lambda = 204.227$ by two different methods.

Table 4.16: Value-at-Risk (in €) of the aggregate loss distribution for $\delta = 0.999$ and $\lambda = 204.227$ computed for the severity distributions of Table 4.12 by the methods of Opdyke (2014) ($VaR^{\star}$) and Hernández et al. (2014) ($VaR^{\star\star}$). The relative discrepancy is $100(VaR^{\star\star}/VaR^{\star} - 1)$ (in %).

Distribution	Type	$\hat{\theta}_1$	$\hat{\theta}_2$	$VaR^{\star}$	$VaR^{\star\star}$	Relative Difference
Weibull	Truncated	0.002	0.084	369,057,122,035.81	367,897,668,700.84	-0.314
LogNormal	Truncated	4.605	4.861	1,612,936,041,395.56	1,601,895,740,617.92	-0.684
LogGamma	Truncated	9.320	1.009	16,946,010,489,032.10	16,644,246,530,583.70	-1.781
Fréchet	Truncated	0.497	15828.319	2,172,587,392,340,770.00	2,172,516,886,433,050.00	-0.003
GP	Truncated	2.158	23641.743	8,934,481,137,897,670.00	8,930,541,905,501,710.00	-0.044
Fréchet	Regular	0.680	93889.730	6,090,425,793,574.64	6,107,774,521,521.65	0.285
GP	Regular	1.676	132454.272	62,835,740,830,873.70	62,918,519,959,436.90	0.132
LogGamma	Regular	37.995	3.071	14,190,399,916.16	14,150,245,634.38	-0.283
LogNormal	Regular	12.372	2.113	3,143,387,786.26	3,138,064,038.88	-0.169
Weibull	Regular	754480.374	0.373	1,246,514,105.86	1,245,671,937.88	-0.068

The first method is ISLA of Opdyke (2014) and the second one is a first-order perturbative approximation of Hernández et al. (2014). As Table 4.16 suggests, the two methods agree very well.

The range of the *VaR* numbers is extremely wide, and even for the top three distributions the variability is significant: the largest *VaR* of the three is 46 times larger than the smallest one. Clearly, an erroneous final choice of the best-fit distribution may cause significant under- or over-estimation of the capital.

These *VaR* numbers reflect broad capitalization required for all the financial institutions in the ORX pool for the particular risk event type. An average capital requirement per institution is about €8 billion (i.e., €369 billion divided by 46, where 46 is the number of financial institutions which experienced OpRisk losses in North America and Western Europe during the considered period). External data, however, are usually not used on their own, but rather taken as an auxiliary source of information by a financial institution in the course of modeling its own, internal, losses. There are approaches to combine the internal and external experience that aim to produce a reasonable final severity model for the internal data. One is referred to Bolancé et al. (2012, Chapter 4) for a detailed exposition of this topic.

Chapter 5

Conclusions

5.1 Summary

In this thesis, we have investigated some important aspects of distributional form and shape. To this end, we have used inferential methods, both parametric and nonparametric. In particular, we used linear and isotonic regression, hypothesis testing, point estimation and kernel density estimation. We also extensively relied on methods of asymptotic statistics and empirical processes.

In Chapter 2, we have presented a method for bump hunting for probability density functions, TBH, using kernel density estimators. TBH is featured by a new definition of a bump together with a classification of bumps into types. The development is based on the notion of curvature of a planar curve. Under different settings, we have applied the method to a wide range of distributions distinguished by their varying bump patterns, and compared it to two procedures known in the literature. Extensive Monte Carlo simulations have demonstrated the viability of the proposed methodology and its satisfactory performance. We have also considered a real data example where TBH identified three bumps, a conclusion supported by previous works.

In Chapter 3, motivated by the problem of selecting parametric distributions for modeling OpRisk loss severities more adequately in the upper tails, we have examined a class of weighted EDF-based GoF test statistics $W_n^{2,\beta}$ when $\beta \leqslant 2$. As β gets closer to 2, $W_n^{2,\beta}$ puts more weight on the upper tail of the distribution under the null hypothesis. The statistic $W_n^{2,2}$ is used for severity distribution selection by some OpRisk modelers in the financial industry. Given the importance of this class of test statistics in the severity selection process, we have more carefully studied the behavior of $W_n^{2,\beta}$ for $1.5 \leqslant \beta \leqslant 2$ and discovered that the practical utility of the statistics is likely limited. In fact, it has been established that $W_n^{2,\beta}$ for $\beta \geqslant 2$ has no well-defined limiting distribution. Therefore, OpRisk modelers applying LDA principles should be highly cautious when using $W_n^{2,\beta}$ for $1.5 \leqslant \beta < 2$ and exercise extreme caution when attempting to use the statistic for $\beta \geqslant 2$.

In Chapter 4, we have studied a transformation that reduces a general composite hypothesis to a test for normality. CBT is fast and easy to implement. In the literature, the indirect tests based on CBT have been studied on a standalone basis. We have examined the performance of the direct and indirect tests based on selected WCvM test statistics in parallel. The study confirmed a good degree of agreement between the two tests. WCvM tests with a focus on the upper tail, $\widehat{W}_n^{2,\beta}$ ($\beta \in [1,2)$), seem to benefit best from the reduction to normality while indirect tests based on the classical CvM and AD tests may suffer from a loss in power. The study also suggested that, in many situations, decisions made by the direct and indirect tests become independent. An application of the tests to an OpRisk loss data set has demonstrated excellent agreement between direct and indirect tests and has also shown a considerable advantage of the indirect tests in terms of execution time. Overall, we feel that the transformation deserves practitioners' attention and warrants further study.

5.2 Future Work

As part of future work related to Chapter 2, we plan to enhance the method in a number of ways. First, we can make the area threshold, λ, data driven. This may be attempted, for example, along the same lines as the critical bandwidth paradigm of Silverman (1980) which also may help us to (re)define TBH as a statistical test. Indeed, the theoretical results of Section 2.3.3 regarding minima of derivatives of the KDE with a Gaussian kernel indicate potential for making progress in this direction. In order to deal with outliers, we proposed the use of X84 rejection rule. However, this rule may be avoided or used in conjunction with data transformation methods mentioned in Section 2.4.1. Alternatively, we can rely on an adaptive bandwidth approach such as the one described by Botev et al. (2010) has shown promise when coupled with TBH in our Monte Carlo experiments. Finally, a generalization of the method to the multidimensional case may be considered which would use the concepts of multidimensional KDE (Silverman, 1986, pp. 75-94), results on the number of its maxima from Carreira-Perpiñán and Williams (2003), multidimensional ridges (Eberly, 1996; Belyaev et al., 1998) and multivariate isotonic regression (Sasabuchi et al., 1983).

Work on the material of Chapter 3 may be continued to establish analogous theoretical statements about test statistics in the composite hypothesis case by expanding results of Section 4.3.

As part of future work for Chapter 4, we plan to enhance CBT. For example, one possibility is to leverage the fact that for location-scale $\mathscr{G} = \{G((\cdot - \mu)/\sigma) \mid \mu \in \mathbb{R}, \sigma > 0)\}$ and a subclass of scale-shape $\mathscr{W} = \{W((\cdot/\beta)^{\alpha}) \mid \alpha, \beta > 0\}$ families of distributions, the limiting distribution of the WCvM test statistics (when it exists) is independent of the unknown distributional parameters (Bagdonavičius et al., 2011, pp. 92-96). Therefore, the base distribution, instead of the Gaussian distribution, can be taken to be the Gumbel, Weibull, or Pareto distributions. This may provide greater power against heavy-tailed alternative distributions. Considerations of Section 4.3 evidenced the validity of $(\widehat{W}_n^{2,2} - \ln n)/(2\sqrt{\ln n}) \xrightarrow{d} N(0,1)$ as $n \to \infty$. We aim to work on developing a rigourous proof of this statement.

Bibliography

M. Abramovitz and I. Stegun, editors. *Handbook of Mathematical Functions, With Formulas, Graphs, and Mathematical Tables*. Dover Publications, London, 1972.

T. W. Anderson and D. A. Darling. Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *The Annals of Mathematical Statistics*, 23 (2):193–212, June 1952. doi:10.1214/aoms/1177729437.

R. I. Andryushkiw, D. D. Klyushin, and Y. I. Petunin. A new test for unimodality. *Theory of Stochastic Processes*, 14(30):1–6, 2008.

A. Azzalini and G. Menardi. Clustering via nonparametric density estimation: The R package pdfCluster. *Journal of Statistical Software*, 57(11):1–26, 2014. doi:10.18637/jss.v057.i11. URL `http://www.jstatsoft.org/v57/i11/`.

A. Azzalini and N. Torelli. Clustering via nonparametric density estimation. *Statistics and Computing*, 17:71–80, 2007. doi:10.1007/s11222-006-9010-y.

V. Bagdonavičius, J. Kruopis, and M. Nikulin. *Non-parametric Tests for Complete Data*. John Wiley & Sons, Hoboken, NJ, 2011. doi:10.1002/9781118557716.

M. Barros, V. Leiva, R. Ospina, and A. Tsuyuguchi. Goodness-of-fit tests based on the Birnbaum-Saunders distribution for censored reliability data analysis. *IEEE Transactions on Reliability*, 63(2):543–554, June 2014. doi:10.1109/TR.2014.2313707.

Basel Committee on Banking Supervision. International convergence of capital measurement and capital standards: A revised framework. Report, Bank for International Settlements, 2004. URL `https://www.bis.org/publ/bcbs107.htm`.

A. G. Belyaev, A. A. Pasko, and T. L. Kunii. Ridges and ravines on implicit surfaces. In F.-E. Wolter and N. M. Patrikalakis, editors, *Proceedings of the 1998 Computer Graphics International (CGI'98), June 22-24, 1998 - Hannover, Germany*, pages 530–535, Los Alamitos, CA, 1998. IEEE Computer Society.

H. Bengtsson. *R.matlab: Read and Write MAT Files and Call MATLAB from Within R*, 2016. URL `https://CRAN.R-project.org/package=R.matlab`. R package version 3.6.0.

R. N. Bhattacharya. On the functional central limit theorem and the law of the iterated logarithm for Markov processes. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 60:185–201, 1982. doi:10.1007/BF00531822.

R. N. Bhattacharya and E. C. Waymire. *Stochastic Processes with Applications*. Society for Industrial and Applied Mathematics, Philadalphia, 2009.

Board of Governors of the Federal Reserve System. Supervisory guidance for data, modeling, and model risk management under the operational risk advanced measurement approaches. Basel Coordination Committee (BCC) Bulletin, Board of Governors of the Federal Reserve System, June 2014. URL `https://www.federalreserve.gov/bankinforeg/basel/basel-coordination-committee-bulletins.htm`.

K. Böcker and C. Klüppelberg. Operational VAR: a closed form approximation. *Risk*, pages 90–93, December 2005.

C. Bolancé, M. Guillén, J. Gustafsson, and J. P. Nielsen. *Quantitative Operational Risk Models*. CRC Press, Boca Raton, FL, 2012.

Z. I. Botev, J. F. Grotowski, and D. P. Kroese. Kernel density estimation via diffusion. *The Annals of Statistics*, 38(5):2916–2957, 2010.

T. Buch-Larsen, J. P. Nielsen, M. Guillén, and C. Bolanceé. Kernel density estimation for heavy-tailed distributions using the Champernowne transformation. *Statistics. A Journal of Theoretical and Applied Statistics*, 39(6):503–518, 2005. doi:10.1080/02331880500439782.

R. Calaway, Revolution Analytics, S. Weston, and D. Tenenbaum. *doParallel: Foreach Parallel Adaptor for the 'parallel' Package*, 2015. URL https://CRAN.R-project.org/package=doParallel. R package version 1.0.10.

M. Á. Carreira-Perpiñán and C. K. I. Williams. *On the Number of Modes of a Gaussian Mixture*, pages 625–640. Springer Berlin Heidelberg, Berlin, Heidelberg, 2003. doi:10.1007/3-540-44935-3_44.

G. Casella and R. L. Berger. *Statistical Inference*. Duxbury Press, Pacific Grove, CA, second edition, 2002.

J. E. Chacón and T. Duong. Data-driven density derivative estimation, with applications to nonparametric clustering and bump hunting. *Electronic Journal of Statistics*, 7:499–532, 2013. doi:10.1214/13-EJS781.

G. Chen. *Empirical processes based on regression residuals: Theory and Applications*. PhD thesis, Simon Fraser University, Canada, August 1991. URL http://summit.sfu.ca/item/4526.

G. Chen and N. Balakrishnan. A general purpose approximate goodness-of-fit test. *Journal of Quality Technology*, 27(2):154–151, 1995.

M.-Y. Cheng and P. Hall. Mode testing in difficult cases. *The Annals of Statistics*, 27(4):1294–1315, 1999. doi:10.1214/aos/1017938927.

A. Chernobai. *Contributions to Modeling of Operational Risk in Banks*. PhD thesis, University of California, Santa Barbara, June 2006. URL http://zanran_storage.s3.amazonaws.com/whitman.syr.edu/ContentPages/314994662.pdf.

A. Chernobai, S. Rachev, and F. Fabozzi. Composite goodness-of-fit tests for left truncated loss samples. Technical report, University of California, Santa Barbara, June 2005. URL http://www.pstat.ucsb.edu/research/papers/KSmissing20050604-JFE.pdf.

A. Chernobai, S. Rachev, and F. Fabozzi. Composite goodness-of-fit tests for left truncated loss samples. In C.-F. Lee and J.C. Lee, editors, *Handbook of Financial Econometrics and Statistics*, pages 575–596. Springer, New York, NY, 2015. doi:10.1007/978-1-4614-7750-1.

H. Chernoff. Estimation of the mode. *Annals of the Institute of Statistical Mathematics*, 16:31–41, 1964. doi:10.1007/BF02868560.

D. M. Chibisov. Some theorems on the limiting behaviour of an empirical distribution function. *Trudy Mat. Inst. Steklov*, 71:104–112, 1964.

D. Collett. *Modelling Survival Data in Medical Research*. Chapman & Hall, London, 1994.

S. Corlay and G. Pagès. Functional quantization-based stratified sampling methods. *Monte Carlo Methods and Applications*, 21(1):1–32, 2015.

H. Cramér. On the composition of elementary errors. *Scandinavian Actuarial Journal*, 1928(1):13–74, 1928.

H. Cramér. *Mathematical methods of statistics*. Princeton University Press, Princeton, NJ, 1999. Nineteenth Printing.

M. Csörgő and L. Horváth. On the distribution of L_p norms of weighted uniform empirical and quantile processes. *The Annals of Probability*, 16(1):142–161, 1988. doi:10.1214/aop/1176991890.

M. Csörgő and L. Horváth. *Weighted Approximations in Probability and Statistics*. John Wiley & Sons, New York, NY, 1993.

M. Csörgő, L. Horváth, and Q.-M. Shao. Convergence of integrals of uniform empirical and quantile processes. *Stochastic Processes and their Applications*, 45:283–294, 1993. doi:10.1016/0304-4149(93)90075-F.

S. Csörgő and J. J. Faraway. The exact and asymptotic distributions of the Cramér-von Mises statistics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 58(1):221–234, 1996.

R. D'Agostino and M. Stephens, editors. *Goodness-of-fit techniques*. Marcel Dekker Inc, New York, NY, 1986.

T. Dankel, Jr. On the distribution of the integrated square of the Ornstein-Uhlenbeck process. *SIAM Journal on Applied Mathematics*, 51(2):568–574, April 1991. doi:10.1137/0151029.

J. E. Dazard and J. S. Rao. Local sparse bump hunting. *Journal of Computational and Graphical Statistics*, 19(4):900–929, 2010. doi:10.1198/jcgs.2010.09029.

M. Degen. The calculation of minimum regulatory capital using single-loss approximations. *The Journal of Operational Risk*, 5(4):3–17, 2010. doi:10.21314/JOP.2010.084.

P. Deheuvels and G. V. Martynov. Karhunen-Loéve expansions for weighted Wiener processes and Brownian bridges via Bessel functions. *Progress in Probability*, 5:57–93, 2003. doi:10.1007/978-3-0348-8059-6.

E. del Barrio, P. Deheuvels, and S. van de Geer. *Lectures on Empirical Processes: Theory and Statistical Applications*. European Mathematical Society, Zürich, Switzerland, 2007. doi:10.4171/027.

S. W. Dharmadhikari and K. Joag-Dev. *Unimodality, Convexity, and Applications*. Academic Press, Inc., San Diego, CA, 1988.

P. Drobinski, C. Coulais, and B. Jourdier. Surface wind-speed statistics modelling: Alternatives to the Weibull distribution and performance evaluation. *Boundary-Layer Meteorology*, 157(1):97–123, October 2015. doi:10.1007/s10546-015-0035-7.

T. Duong. *ks: Kernel Smoothing*, 2016. URL `https://CRAN.R-project.org/package=ks`. R package version 1.10.4.

J. Durbin. Weak convergence of the sample distribution function when parameters are estimated. *The Annals of Statistics*, 1(2):279–290, March 1973a.

J. Durbin. *Distribution theory for tests based on the sample distribution function*. SIAM, Philadelphia, PA, 1973b.

D. Eberly. *Ridges in Image and Data Analysis*. SAS Institute, Inc., Cary, NC, 1996. doi:10.1007/978-94-015-8765-5.

A. Elbert and A. Laforgia. An asymptotic relation for the zeros of Bessel functions. *Journal of Mathematical Analysis and Applications*, 98:502–511, 1984. doi:10.1016/0022-247X(84)90265-8.

P. Embrechts and C. M. Goldie. On convolution tails. *Stochastic Processes and their Applications*, 13:263–278, 1982. doi:10.1016/0304-4149(82)90013-8.

P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag, Berlin, 1997. doi:10.1007/978-3-642-33483-2.

V. R. Fatalov. Gaussian Ornstein-Ulhenbeck and Bogoliubov processes: Asymptotics of small deviations for L^p-functionals, $0 < p < \infty$. *Problems of Information Transmission*, 50(4):371–389, 2014. doi:10.1134/S0032946014040073.

A. Feuerverger. On goodness of fit for operational risk. *International Statistical Review*, pages 1–22, July 2015. doi:10.1111/insr.12112.

N. I. Fisher, E. Mammen, and J. S. Marron. Testing for multimodality. *Computational Statistics & Data Analysis*, 18:499–512, 1994. doi:10.1016/0167-9473(94)90080-9.

F. Gao, J. Hannig, T.-Y. Lee, and F. Torcaso. Laplace transforms via Hadamard factorization. *Electronical Journal of Probability*, 8:1–20, August 2003. doi:10.1214/EJP.v8-151.

C. G. Gibson. *Elementary Geometry of Differentiable Curves: An Undergraduate Introduction*. Cambridge University Press, Cambridge, 2001. doi:10.1017/CBO9781139173377.

G. Goldmann, B. Klar, and S. G. Meintanis. Data transformations and goodness-of-fit tests for type-II right censored samples. *Metrika*, 78:59–83, 2015. doi:10.1007/s00184-014-0490-z.

I. J. Good and R. A. Gaskins. Density estimation and bump-hunting by the penalized likelihood method exemplified by scattering and meteorite data. *Journal of the American Statistical Association*, 75(369):42–56, 1980. doi:10.2307/2287385.

G. R. Grimmett and D. R. Stirzaker. *Probability and Random Processes*. Oxford University Press, New York, NY, 3rd edition, 2001.

A. Gut. *Probability: A Graduate Course*. Springer, New York, NY, 2nd edition, 2013. doi:10.1007/978-1-4614-4708-5.

S. Hahmann, A. Belyaev, L. Busé, G. Elber, B. Mourrain, and C. Rössl. Shape interrogation. In L. De Floriani and M. Spagnuola, editors, *Shape Analysis and Structuring*, pages 1–53. Springer-Verlag, Berlin, 2008. doi:10.1007/978-3-540-33265-7_1.

P. Hall, M. C. Minnotte, and C. Zhang. Bump hunting with non-Gaussian kernels. *The Annals of Statistics*, 32(5):2124–2141, 2004. doi:10.1214/009053604000000715.

F. Hampel, P. Rousseeuw, E. Ronchetti, and W. Stahel. *Robust Statistics: The Approach Based on Influence Functions*. Wiley, New York, NY, 1986.

J. A. Hartigan and P. M. Hartigan. The dip test of unimodality. *The Annals of Statistics*, 13(1):70–84, 1985. doi:10.1214/aos/1176346577.

D. J. Henderson and C. F. Parmeter. *Applied Nonparametric Econometrics*. Cambridge University Press, New York, NY, 2015. doi:10.1017/cbo9780511845765.

L. Hernández, J. Tejero, A. Suárez, and S. Carrillo-Menéndez. Percentiles of sums of heavy-tailed random variables: beyond the single-loss approximation. *Statistics and Computing*, 24(3):377–397, May 2014. doi:10.1007/s11222-013-9376-6.

S. M. Iakus. *Simulation and Inference for Stochastic Differential Equations*. Springer, New York, NY, 2008. doi:10.1007/978-0-387-75839-8.

A. E. Jaffe, Murakami, P. H. Lee, J. T. Leek, M. D. Fallin, A. P. Feinberg, and R. A. Irizarry. Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. *International Journal of Epidemiology*, 41(1):200–209, 2012. doi:10.1093/ije/dyr238.

N. L. Johnson, S. Kotz, and N. Balakrishnan. *Continuous univariate distributions - Vol. 1*. John Wiley & Sons, New York, NY, 2nd edition, 1994.

M. C. Jones, J. S. Marron, and S. J. Sheather. A brief survey of bandwidth selection for density estimation. *Journal of the American Statistical Association*, 91(433): 401–407, 1996. doi:10.1080/01621459.1996.10476701.

B. Jørgenssen. *Statistical Properties of the Generalized Inverse Gaussian Distribution*, volume 9 of *Lecture Notes in Statistics*. Springer-Verlag, New York, NY, 1982. doi:10.1007/978-1-4612-5698-4.

A. N. Kolmogorov and S. V. Fomin. *Introductory Real Analysis*. Prentice-Hall, Inc., Englewood Cliffs, NJ, 1970.

V. Kostrikin and K. A. Makarov. On Krein's example. *Proceedings of the American Mathematical Society*, 136(6):2067–2071, June 2008. doi:10.1090/S0002-9939-08-09141-7.

P. A. Kozhukhar. On the discrete spectrum of the perturbed Wiener-Hopf operator. *Matematicheskie Zametki*, 51(1):66–73, 1992. English translation.

S. G. Krantz and H. R. Parks. *A Primer of Real Analytic Functions*. Birkhäuser Boston, Boston, MA, 2nd edition, 2002. doi:10.1007/978-0-8176-8134-0.

M. G. Krein. On a trace formula in perturbation theory. *Matem. Sbornik*, 33:597–626, 1953. (In Russian).

O. Kuss. Global goodness-of-fit tests in logistic regression with sparse data. *Statistics in Medicine*, 21(24):3789–3801, December 2002. doi:10.1002/sim.1421.

S. Lavaud and V. Lehérissé. Goodness-of-fit tests and selection methods for operational risk. *Journal of Operational Risk*, 9(3):21–50, 2014. doi:10.21314/JOP.2014.139.

R. S. Liptser and A. N. Shiryaev. *Statistics of Random Processes II. Applications*. Springer, New York, NY, 2nd, revised and expanded edition, 2001. doi:10.1007/978-3-662-10028-8.

A. Luceño. Fitting the generalized Pareto distribution to data using maximum goodness-of-fit estimators. *Computational Statistics and Data Analysis*, 51(2): 904–917, November 2006. doi:1016/j.csda.2005.09.011.

J. G. MacKinnon. Critical values for cointegration tests. In R. Engle and C. Granger, editors, *Long-Run Economic Relationships: Readings in Cointegration*, chapter 13, pages 267–276. Oxford University Press, Oxford, 1991. URL `http://EconPapers.repec.org/RePEc:oxp:obooks:9780198283393`.

M. Maechler. *diptest: Hartigan's Dip Test Statistic for Unimodality - Corrected*, 2015. URL `https://CRAN.R-project.org/package=diptest`. R package version 0.75-7.

P. Mandl. *Analytical Treatment of One-Dimensional Markov Processes*. Springer, New York, NY, 1968.

MAPLE. *MAPLE version 2016.0, build ID 1113130*. Maplesoft, a division of Waterloo Maple Inc, Waterloo, ON, 2016.

N. M. Markovitch and U. R. Krieger. Nonparametric estimation of long-tailed density functions and its application to the analysis of world wide web traffic. *Performance Evaluation*, 42:205–222, 2000. doi:10.1016/S0166-5316(00)00031-6.

J. S. Marron and M. P. Wand. Exact mean integrated squared error. *The Annals of Statistics*, 20(2):712–736, 1992. doi:10.1214/aos/1176348653.

G. Marsaglia, W. W. Tsang, and J. Wang. Evaluating Kolmogorov's distribution. *Journal of Statistical Software*, 8(18):1–4, 2003. doi:10.18637/jss.v008.i18.

G. Martynov. Weighted Cramér-von Mises test with estimated parameters. *Communications in Statistics - Theory and Methods*, 40:3569–3586, 2011. doi:10.1080/03610926.2011.581179.

D. M. Mason. The asymptotic distribution of generalized Rényi statistics. *Acta Mathematica Scientia*, 48:315–323, 1985.

D. M. Mason and G. R. Shorack. Necessary and sufficient conditions for asymptotic normality of *L*-statistics. *The Annals of Probability*, 20(4):1779–1804, 1992. doi:10.1214/aop/1176989529.

MATLAB. *MATLAB version 9.1.0.441655 (R2016b)*. The MathWorks Inc, Natick, MA, 2016a.

MATLAB. *Parallel Computing Toolbox Version 6.9 (R2016b)*. The MathWorks Inc, Natick, MA, 2016b.

A. J. McNeil, R. Frey, and P. Embrechts. *Quantitative Risk Management*. Princeton University Press, Princeton, New Jersey, Revised edition, 2015.

S. G. Meintanis. Goodness-of-fit testing by transforming to normality: comparison between classical and characteristic function-based methods. *Journal of Statistical Computation and Simulation*, 79(2):205–212, February 2009. doi:10.1080/00949650701730547.

G. Menardi and A. Azzalini. An advancement in clustering via nonparametric density estimation. *Statistics and Computing*, 24:753–767, 2014. doi:10.1007/s11222-013-9400-x.

G. Mignola and R. Ugoccioni. Sources of uncertainty in modeling operational risk losses. *Journal of Operational Risk*, 1(2):33–50, 2006. doi:10.21314/JOP.2006.008.

M. C. Minnotte. Nonparametric testing of the existence of modes. *The Annals of Statistics*, 25(4):1646–1660, 1997. doi:10.1214/aos/1031594735.

D. W. Müller and G. Sawitzki. Excess mass estimates and tests for multimodality. *Journal of the American Statistical Association*, 86(415):738–746, 1991. doi:10.1080/01621459.1991.10475103.

S. Nabeya. Absolute moments in 2-dimensional normal distributions. *Annals of the Institute of Statistical Mathematics*, 3:2–6, 1951. doi:10.1007/BF02949771.

G. Neuhaus. Asymptotic theory of goodness of fit tests when parameters are present: a survey. *Mathematische Operationsforschung und Statistik. Series Statistics*, 10(3):479–494, 1979.

R. G. Newcombe. *Confidence Intervals for Proportions and Related Measures of Effect Size*. CRC Press, Boca Raton, FL, 2013.

J. D. Opdyke. Estimating operational risk capital with greater accuracy, precision and robustness. *Journal of Operational Risk*, 9:3–79, 2014. doi:10.21314/JOP.2014.137.

D. A. Pachamanova and F. J. Fabozzi. *Simulation and Optimization in Finance: Modeling with MATLAB, @RISK, or VBA*. John Wiley & Sons, Inc, Hoboken, NJ, 2010. doi:10.1002/9781118267752.

S. M. Papalexiou, D. Koutsoyiannis, and C. Makropoulos. How extreme is extreme? An assessment of daily rainfall distribution tails. *Hydrology and Earth System Sciences*, 17(2):851–862, 2013. doi:10.5194/hess-17-851-2013.

E. Parzen. Spectral analysis of asymptotically stationary time series. *Bulletin of the International Statistical Institute*, 39(2):87–103, 1962a.

E. Parzen. On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33:1065–1076, 1962b. doi:10.1214/aoms/1177704472.

G. A. Pavliotis. *Stochastic Processes and Applications*. Springer, New York, NY, 2014. doi:10.1007/978-1-4939-1323-7.

G. W. Peters and P. V. Shevchenko. *Advances in heavy tailed risk modeling: a handbook of operational risk*. John Wiley & Sons, Inc., Hoboken, NJ, 2015. doi:10.1002/9781118909560.

V. S. Pugachev and I. N. Sinitsyn. *Lectures on Functional Analysis and Applications*. World Scientific, Singapore, 1999.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016. URL `https://www.R-project.org/`. R version 3.3.0.

V. C. Raykar, R. Duraiswami, and L. H. Zhao. Fast computation of kernel estimators. *Journal of Computational and Graphical Statistics*, 19(1):205–220, 2010. doi:10.1198/jcgs.2010.09046.

M. Reed and B. Simon. *Methods of Modern Mathematical Physics, vol. I: Functional Analysis*. Academic Press, Inc., San Diego, CA, Revised and enlarged edition, 1980.

T. Robertson, F. T. Wright, and R. L. Dykstra. *Order Restricted Statistical Inference*. John Wiley & Sons, Chichester, 1988.

S. Sasabuchi, M. Inutsuka, and D. D. S. Kulatunga. A multivariate version of isotonic regression. *Biometrika*, 70(2):465–472, August 1983. doi:10.1093/biomet/70.2.465.

F. Schmid and A. Schmidt. Nonparametric estimation of the coefficient of overlapping - theory and empirical application. *Computational Statistics & Data Analysis*, 50:1583–1596, 2006. doi:10.1016/j.csda.2005.01.014.

W. F. Scott. A weighted Cramér-von Mises statistic, with some applications to clinical trials. *Communications in Statistics - Theory and Methods*, 28:3001–3008, 1999. doi:10.1080/03610929908832461.

S. J. Sheather and M. C. Jones. A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, 53:683–690, 1991.

L. A. Shepp. Radon-Nykodim derivatives of Gaussian measures. *The Annals of Mathematical Statistics*, 37:321–354, 1966. doi:10.1214/aoms/1177699516.

G. R. Shorack and J. A. Wellner. *Empirical processes with applications to statistics*. The Society for Industrial and Applied Mathematics, Philadelphia, 2009. doi:10.1137/1.9780898719017.

B. W. Silverman. Using kernel density estimates to investigate multimodality. *Journal of the Royal Statistical Society. Series B (Methodological)*, 43(1):97–99, 1980.

B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, New York, NY, 1986. doi:10.1007/978-1-4899-3324-9.

C. D. Sinclair, B. D. Spurr, and M. I. Ahmad. Modified Anderson-Darling test. *Communications in Statistics - Theory and Methods*, 19(10):3677–3686, 1990. doi:10.1080/03610929008830405.

R. S. Singh. Improvement on some known nonparametric uniformly consistent estimators of derivatives of a density. *The Annals of Statistics*, 5:394–359, 1977. doi:10.1214/aos/1176343805.

C. G. Small. *Expansions and Asymptotics for Statistics*. Chapman & Hall/CRC, Boca Raton, FL, 2010. doi:10.1201/9781420011029.

R. L. Smith. Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72(1):67–90, 1985. doi:10.1093/biomet/72.1.67.

M. A. Stephens. EDF statisics for goodness of fit and some comparisons. *Journal of the American Statistical Association*, 69(347):730–737, 1974. doi:10.1080/01621459.1974.10480196.

S. Sukhatme. Fredholm determinant of a positive definite kernel of a special type and its application. *The Annals of Mathematical Statistics*, 43(6):1914–1926, 1972. doi:10.1214/aoms/1177690862.

M. E. Tarter and M. D. Lock. *Model-Free Curve Estimation*. Chapman & Hall, New York, NY, 1993.

O. Thas. *Comparing distributions*. Springer Science + Business Media LLC, New York, NY, 2010. doi:10.1007/978-0-387-92710-7.

A. B. Turk. A quantitative operational risk management model. *WSEAS Transactions on Business and Economics*, 6(5):241–253, 2009.

A. W. van der Vaart. *Asymptotic statistics*. Cambridge University Press, Cambridge, England, 2007.

A. J. Viollaz. Convergence in weighted supremum norms of the Skorokhod representation of the estimated empirical process. *Communications in Statistics - Theory and Methods*, 24(11):2829–2839, 1995. doi:10.1080/03610929508831653.

R. E. von Mises. *Wahrscheinlichkeit, Statistik und Wahrheit*. Julius Springer, 1928. doi:10.1007/978-3-662-36230-3.

J. Wahlström. Operational risk modeling: Theory and practice. Master's thesis, Royal Institute of Technology, Stockholm, Sweden, 2013. URL `https://www.math.kth.se/matstat/seminarier/reports/M-exjobb13/131218b.pdf`.

M. P. Wand and M. C. Jones. *Kernel Smoothing*. Chapman and Hall, New York, NY, 1995. doi:10.1007/978-1-4899-4493-1.

M. S. Weitzman. *Measures of overlap of income distributions of white and Negro families in the United States*. U.S. Bureau of the Census; [for sale by the Supt. of Docs., U.S. Govt. Print. Off.], Washington D.C., 1970.

G. A. Whitmore and M. Yalovsky. A normalizing logarithmic transformation for inverse Gaussian random variables. *Technometrics*, 20(2):207–208, May 1978. doi:10.1080/00401706.1978.10489648.

D. Wishart. Mode analysis: a generalization of nearest neighbor which reduces chaining effects. In A. J. Cole, editor, *Numerical Taxonomy*, pages 282–308, New York, NY, 1969. Academic Press.

L. Yang and J. S. Marron. Iterated transformation-kernel density estimation. *Journal of the American Statistical Association*, 94(446):580–589, 1999.

X. Y. Zhou. Statistical analysis of traffic loads and traffic load effects on bridges. mechanics of the structures. Université Paris - EST, 2013. URL `https://tel.archives-ouvertes.fr/tel-00949929/`. (In English).

Appendix A

For Chapter 2

A.1 Auxiliary Results

A.1.1 Facts from Real Analysis

Let us now consider the following integral:

$$I_x(h) = \int_{-\infty}^{\infty} V(u)g(x-hu)du \tag{A.1}$$

for $h \to 0$.

Condition 1 of Assumption 2.1, in particular, implies that the function g has the first order derivative in a neighbourhood of point x_0 and the Taylor's expansion

$$g(x_0+\Delta x) = g(x_0) + g'(x_0)\Delta x + \frac{1}{2}g''(x_0)(\Delta x)^2 + o\left((\Delta x)^2\right) \tag{A.2}$$

holds as $\Delta x \to 0$. The $o\left((\Delta x)^2\right)$ term can be represented as $(\Delta x)^2 r^*_{2,x_0}(\Delta x)$ with function $r^*_{2,x_0} : \mathbb{R} \to \mathbb{R}$ being such that $r^*_{2,x_0}(\Delta x) \to 0$ as $\Delta x \to 0$.

By plugging (A.2) into (A.1) with $\Delta x = -hu$, we obtain the following formal expansion:

$$I_{x_0}(h) = g(x_0)\kappa_0(V) - g'(x_0)\kappa_1(V)h + \frac{1}{2}g''(x_0)\kappa_2(V)h^2 + h^2R_2(h), \tag{A.3}$$

where $R_2(h) = \int_{-\infty}^{\infty} u^2V(u)r_{2,x_0}(hu)du$ and $r_{2,x_0}(hu) = r^*_{2,x_0}(-hu)$.

Remark. Function $R_2(\cdot)$, of course, depends on x_0. However, for the sake of notational simplicity, we will suppress this dependance and assume that the number $x_0 \in \mathbb{R}$ is fixed. Moreover, in what follows, we will denote the function $r_{2,x_0}(\cdot)$ by just $r(\cdot)$.

We are interested in establishing conditions under which, in (A.3), the remainder term $R_2(h) \to 0$ as $h \to 0$, or, equivalently, $h^2 R_2(h) = o\left(h^2\right)$. The following three theorems provide sufficient conditions for this to hold true.

Theorem A.1. Suppose a function $g : \mathbb{R} \to \mathbb{R}$ satisfies condition 1 of Assumption 2.1. Also, let a function $V : \mathbb{R} \to \mathbb{R}$ satisfy conditions 1, 2 and 3a of Assumption 2. Then, $h^2 R_2(h) = o\left(h^2\right)$ as $h \to 0$.

Theorem A.2. Suppose a function $g : \mathbb{R} \to \mathbb{R}$ satisfies conditions 1 and 2b of Assumption 2.1. Also, let a function $V : \mathbb{R} \to \mathbb{R}$ satisfy conditions 1 and 2 of Assumption 2. Then, $h^2 R_2(h) = o\left(h^2\right)$ as $h \to 0$.

Theorem A.3. Suppose a function $g : \mathbb{R} \to \mathbb{R}$ satisfies conditions 1 and 2a of Assumption 2.1. Also, let a function $V : \mathbb{R} \to \mathbb{R}$ satisfy conditions 1, 2 and 3b of Assumption 2. Then, $h^2 R_2(h) = o\left(h^2\right)$ as $h \to 0$.

To prove the theorems, we will need several auxiliary results. We will summarize them as lemmas.

Lemma A.4. If $\operatorname{supp}|V| \Subset \mathbb{R}$, then $R_2(h) \to 0$ as $h \to 0$.

Proof. There exists a number $0 < A < \infty$, such that $V(x) \equiv 0$ when $|x| > A$. Then,

$$\begin{aligned}
|R_2(h)| &\leqslant \int_{-A}^{A} u^2 |V(u)||r(hu)|du \\
&\leqslant \sup_{u\in[-A,A]} |r(hu)| \int_{-A}^{A} u^2 |V(u)|du \\
&= \sup_{x\in[-hA,hA]} |r(x)| \int_{-A}^{A} u^2 |V(u)|du.
\end{aligned}$$

As $h \to 0$, the right-hand side of the last row goes to 0 because $r(x) \to 0$ and $\int_{-A}^{A} u^2 |V(u)|du$ is bounded. ■

Lemma A.5. Suppose there exists a constant $0 < C < \infty$ such that for any $x \in \mathbb{R}$ $r(x) \leqslant C$. Assume further that $r(x) \to 0$ as $x \to 0$. Then, $R_2(h) \to 0$ as $h \to 0$.

Proof. The claim of the lemma follows from Lebesgue's Dominated Convergence Theorem (Kolmogorov and Fomin, 1970, p. 303). Indeed,

$$|R_2(h)| \leqslant \int_{-\infty}^{\infty} u^2|V(u)||r(hu)|du,$$

where

- $u^2|V(u)||r(hu)| \leqslant Cu^2|V(u)|$ and $\int_{-\infty}^{\infty} u^2|V(u)|du < \infty$;
- For any fixed u, we have $u^2|V(u)||r(hu)| \to 0$ as $h \to 0$.

∎

Lemma A.6. Suppose there exist constants $0 < d_1, d_2 < \infty$ such that for any $x \in \mathbb{R}$ $|V(x)| \leqslant d_1 \exp(-d_2x^2)$. Further, let us assume that

1. For any $h \in \mathbb{R}$, the integral $\int_{-\infty}^{\infty} u^2|V(u)||r(hu)|du$ exists and is finite;
2. There exist constants $0 < a_1, a_2 < \infty$ such that $r(x) \leqslant a_1 \exp(a_2x^2)$;
3. $r(x) \to 0$ as $x \to 0$.

Then, $R_2(h) \to 0$ as $h \to 0$.

Proof. First, observe that condition 1 follows from 2 when $|h| < \sqrt{d_2/a_2}$. Let us now fix a number $h_1 \in (0, \sqrt{d_2/a_2})$ and only consider those $h \in [-h_1, h_1]$. As before, note that $|R_2(h)| \leqslant \int_{-\infty}^{\infty} u^2|V(u)||r(hu)|du$ and apply Lebesgue's Dominated Convergence Theorem to obtain

- $u^2|V(u)||r(hu)| \leqslant d_1a_1u^2 \exp(-(d_2 - a_2h^2)u^2) \leqslant d_1a_1u^2 \exp(-(d_2 - a_2h_1^2)u^2)$. Since $d_2 - a_2h_1^2 > 0$, then $\int_{-\infty}^{\infty} u^2 \exp(-(d_2 - a_2h_1^2)u^2)du < \infty$;
- For any fixed u, we have $u^2|V(u)||r(hu)| \to 0$ as $h \to 0$.

∎

Let us now proceed to proving Theorems A.1, A.2 and A.3.

Proof of Theorem A.1. Directly follows from Lemma A.4. ∎

Proof of Theorem A.2. Since the function $g : \mathbb{R} \to \mathbb{R}$ has the second order derivative at some point x_0, we have seen above that $r_{2,x_0}(x) \to 0$ as $x \to 0$. Consequently, there exists a $\delta > 0$ such that $\sup_{x\in[-\delta,\delta]} |r_{2,x_0}(x)| \leqslant 1$. The statement of the theorem will follow from Lemma A.5 if we show that there exists a constant $0 < A_3 < \infty$ such that for any $x \in \mathbb{R}$ $|r(x)| \leqslant A_3$.

To this end, consider Δx outside of $[-\delta, \delta]$. Let $g(x) \leqslant c_1 + c_2 x^2$ for some positive constants c_1 and c_2. Then,

$$\begin{aligned} |r(\Delta x)| &\leqslant \tfrac{|g(x_0+\Delta x)|+|g(x_0)|+|\Delta x||g'(x_0)|+(\Delta x)^2|g''(x_0)|/2}{(\Delta x)^2} \\ &\leqslant \tfrac{A_2(\Delta x)^2+A_1|\Delta x|+A_0}{(\Delta x)^2} \\ &\leqslant A_2 + A_1/\delta + A_0/\delta^2 \end{aligned}$$

for $|\Delta x| \geqslant \delta$ and some positive constants A_0, A_1 and A_2. Then, for any $x \in \mathbb{R}$, $|r(x)| \leqslant A_3$, where $A_3 = \max\{1, A_2 + A_1/\delta + A_0/\delta^2\}$; that is, $|r(\cdot)|$ is bounded, as desired. ■

Proof of Theorem A.3. $g(x) \leqslant b_1 \exp(b_2 x^2)$ for some positive constants b_1 and b_2. Analogous to the proof of Theorem A.2, it can be readily shown that there exist constants $0 < C_1, C_2 < \infty$ such that, for any $x \in \mathbb{R}$, $|r(x)| \leqslant C_1 \exp(C_2 x^2)$. Then, the statement of the theorem follows from Lemma A.6. ■

We conclude this section with the following lemma.

Lemma A.7. If the assumptions of either Theorem A.1, A.2 or A.3 hold, and if, additionally, $V(-x) = V(x)$ for all $x \in \mathbb{R}$, then the expansion

$$\begin{aligned} \left(1 + I_x^2(h)\right)^{-\beta} &= \left(1 + (g(x)\kappa_0(V))^2\right)^{-\beta} \\ &\quad -2\beta g(x)\kappa_0(V)(1 + (g(x)\kappa_0(V))^2)^{-(\beta+1)} \\ &\quad \times \left(\tfrac{1}{2} g''(x)\kappa_2(V)h^2\right) + o\left(h^2\right) \end{aligned}$$

is valid as $h \to 0$ and $\beta \in \mathbb{R}$.

Proof. The fact that $\left(\left(1+x^2\right)^{-\beta}\right)' = -2\beta x \left(1+x^2\right)^{-(\beta+1)}$ means, in particular, that

$$\left(1 + (x_0 + \Delta x)^2\right)^{-\beta} = \left(1 + x_0^2\right)^{-\beta} - 2\beta x_0 \left(1 + x_0^2\right)^{-(\beta+1)} \Delta x + o(\Delta x) \qquad \text{(A.4)}$$

as $\Delta x \to 0$. Under the assumptions of either Theorem A.1, A.2 or A.3,

$$I_x(h) = g(x)\kappa_0(V) + \frac{1}{2}g''(x)\kappa_2(V)h^2 + o\left(h^2\right)$$

as $h \to 0$. Here we have made use of the fact that $\kappa_1(V) = 0$ due to the symmetry of V. Having made the substitutions $x_0 = g(x)\kappa_0(V)$ and $\Delta x = g''(x)\kappa_2(V)h^2/2 + o\left(h^2\right)$ in (A.4), we arrive at the conclusion of the lemma. ■

A.1.2 Lemmas for Section 2.3.2

In this section, present several lemmas which are used in the proofs of the theorems in Section 2.3.2.

We assume that $\{X_i : i = 1, \ldots, n\}$ is a collection of IID random variables with common PDF $f : \mathbb{R} \to \mathbb{R}_+$. In particular, this means $\kappa_0(f) = 1$.

Consider also a function $K : \mathbb{R} \to \mathbb{R}_+$ such that it is an even PDF and $\kappa_2(K) < \infty$. It follows that $\kappa_0(K)$ and $\kappa_1(K)$ equal 1 and 0, respectively. The function $K(\cdot)$ is referred to as the kernel function and the estimator

$$\hat{f}_h(x) = \frac{1}{nh}\sum_{j=1}^{n} K\left(\frac{x - X_j}{h}\right)$$

is referred to as the kernel density estimator of the unknown PDF f. Here, the bandwidth $h = h(n)$ is such that $h(n) \to 0$ as $n \to \infty$.

If $f(\cdot)$ is twice differentiable, we let

$$C(x) = \frac{f''(x)}{\left(1 + \left(f'(x)\right)^2\right)^{\frac{3}{2}}}$$

be the curvature of the PDF and

$$\widehat{C}_{h_1,h_2}(x) = \frac{\widehat{f''_{h_1}}(x)}{\left(1 + \left(\widehat{f'_{h_2}}(x)\right)^2\right)^{\frac{3}{2}}}$$

be its kernel density estimator. In what follows, we take $\widehat{f_h^{(r)}}(x) = \hat{f}_h^{(r)}(x)$, and

$$\hat{f}_h^{(r)}(x) = \frac{1}{nh^{r+1}} \sum_{j=1}^{n} K^{(r)} \left(\frac{x - X_j}{h} \right). \tag{A.5}$$

When $h_1 = h_2 = h$, $\widehat{C}_{h_1,h_2}(\cdot)$ will be denoted by $\widehat{C}_h(\cdot)$.

We wish to derive $MISE\left(\widehat{C}_{h_1,h_2}\right)$. Recall that

$$MISE\left(\widehat{C}_{h_1,h_2}\right) = \mathbb{E}\left[\int_{-\infty}^{+\infty} \left(\widehat{C}_{h_1,h_2}(x) - C(x)\right)^2 dx\right],$$

or, since the integrand is nonnegative,

$$MISE\left(\widehat{C}_{h_1,h_2}\right) = \int_{-\infty}^{+\infty} MSE\left(\widehat{C}_{h_1,h_2}\right) dx,$$

where

$$MSE\left(\widehat{C}_{h_1,h_2}\right) = \mathbb{E}\left[\left(\widehat{C}_{h_1,h_2}(x) - C(x)\right)^2\right].$$

By standard properties of expectation and variance, the latter can be rewritten as

$$MSE\left(\widehat{C}_{h_1,h_2}(x)\right) = \left(Bias\left(\widehat{C}_{h_1,h_2}(x)\right)\right)^2 + \mathbb{V}\left[\widehat{C}_{h_1,h_2}(x)\right]$$

and so

$$MISE\left(\widehat{C}_{h_1,h_2}\right) = \int_{-\infty}^{+\infty} \left(Bias\left(\widehat{C}_{h_1,h_2}(x)\right)\right)^2 dx + \int_{-\infty}^{+\infty} \mathbb{V}\left[\widehat{C}_{h_1,h_2}(x)\right] dx, \tag{A.6}$$

where $Bias\left(\widehat{C}_{h_1,h_2}(x)\right) = \mathbb{E}\left[\widehat{C}_{h_1,h_2}(x)\right] - C(x)$.

We will need the following lemmas.

Lemma A.8. For any $s \in \mathbb{N}$ and IID random variables $Z_1, \ldots, Z_n$ with mean μ there exists a positive constant α_s depending only on s such that the following inequality holds:

$$\mathbb{E}\left[\left(\frac{1}{n}\sum_{j=1}^{n}\left(Z_j - \mu\right)\right)^{2s}\right] \leqslant \frac{\alpha_s}{n^s}\mathbb{E}\left[Z_1^{2s}\right].$$

Proof. First, let us note that

$$\mathbb{E}\left[\left(\frac{1}{n}\sum_{j=1}^{n}\left(Z_j-\mu\right)\right)^{2s}\right]=\left(\frac{1}{n}\right)^{2s}\mathbb{E}\left[\left(\sum_{j=1}^{n}\left(Z_j-\mu\right)\right)^{2s}\right].$$

We will now need the following corollary to the Marcinkiewicz-Zygmund Inequality (see Small, 2010, Inequality (4.3) on p. 103) valid for independent random variables $Y_1,\ldots,Y_n$ with zero mean which establishes the existence of some constant $\beta_m>0$ depending on m only:

$$\mathbb{E}\left[\left|\sum_{j=1}^{n}Y_j\right|^m\right]\leqslant\beta_m n^{(m-2)/2}E\left[\sum_{j=1}^{n}\left|Y_j\right|^m\right],$$

This inequality, when applied to $Y_j=Z_j-\mu_n$, implies that

$$\mathbb{E}\left[\left(\frac{1}{n}\sum_{j=1}^{n}\left(Z_j-\mu\right)\right)^{2s}\right]\leqslant\frac{1}{n^{s+1}}\beta_{2s}\mathbb{E}\left[\sum_{j=1}^{n}\left(Z_j-\mu\right)^{2s}\right]=\frac{1}{n^s}\beta_{2s}\mathbb{E}\left[\left(Z_1-\mu\right)^{2s}\right].$$

From the binomial theorem, we get

$$\mathbb{E}\left[\left(Z_1-\mu\right)^{2s}\right]\leqslant\sum_{k=0}^{2s}\binom{2s}{k}\mathbb{E}\left[\left|Z_1\right|^{2s-k}\right]\left|\mu\right|^k.$$

By the Lyapounov inequality (Gut, 2013, p. 129), we have

$$\mathbb{E}\left[\left|Z_1\right|^{2s-k}\right]\left|\mu\right|^k\leqslant\mathbb{E}\left[Z_1^{2s}\right]^{(2s-k)/(2s)}\mathbb{E}\left[Z_1^{2s}\right]^{k/(2s)}=\mathbb{E}\left[Z_1^{2s}\right]$$

for each $k=\overline{0,2s}$. As the sum of binomial coefficients equals 2^{2s}, defining $\alpha_s=2^{2s}\beta_{2s}$ completes the proof of the lemma. ■

With the help of Lemma A.8, the following bound on the moments of the KDE (A.5) is valid.

Lemma A.9. Let $X_1,\ldots,X_n$ be IID random variables with PDF $f=f(x)$ possessing finite second order derivative at some $x\in\mathbb{R}$. Further, suppose $K^{(r)}(u)$ exists for all $u\in\mathbb{R}$. If $f(\cdot)$ and $K^{(r)}(\cdot)$ satisfy assumptions of either Theorem A.1, A.2 or A.3, and if $\int_{-\infty}^{\infty}\left(K^{(r)}(z)\right)^{2s}dz<\infty$ for some $s\in\mathbb{N}$, then

$$\mathbb{E}\left[\left(\hat{f}_h^{(r)}(x)-\mathbb{E}\left[\hat{f}_h^{(r)}(x)\right]\right)^{2s}\right]=O\left(\frac{1}{n^s h^{2s(r+1)-1}}\right)$$

as $n\to\infty$, where $\hat{f}_h^{(r)}(\cdot)$ is the estimator (A.5).

Proof. Indeed, Lemma A.8 implies that

$$\mathbb{E}\left[\left(\hat{f}_h^{(r)}(x)-\mathbb{E}\left[\hat{f}_h^{(r)}(x)\right]\right)^{2s}\right]\leqslant\frac{\alpha_s}{n^s}\mathbb{E}\left[\left(\frac{1}{h^{r+1}}K^{(r)}\left(\frac{x-X_1}{h}\right)\right)^{2s}\right].$$

The expectation on the right-hand side can be written as

$$\begin{aligned}\mathbb{E}\left[\left(\tfrac{1}{h^{r+1}}K^{(r)}\left(\tfrac{x-X_1}{h}\right)\right)^{2s}\right]&=\tfrac{1}{h^{2s(r+1)-1}}\int_{-\infty}^{\infty}\left(K^{(r)}(z)\right)^{2s}f(x-zh)\,dz\\&=\tfrac{1}{h^{2s(r+1)-1}}\left(f(x)\int_{-\infty}^{\infty}\left(K^{(r)}(z)\right)^{2s}dz+o(1)\right),\end{aligned}$$

where the $o(1)$ step is justified by Theorem A.1, A.2 or A.3. This, when combined with the previous inequality, proves the lemma. ■

Lemma A.9 allows us to demonstrate the following result.

Lemma A.10. Let $X_1,\dots,X_n$ be IID random variables with PDF $f=f(x)$ possessing finite fourth order derivative at some $x\in\mathbb{R}$. Suppose also that $K^{(r)}(u)\to 0$ as $|u|\to\infty$ for $r\in\{0,1,2\}$ and that $f^{(4)}(\cdot)$ and $K(\cdot)$ satisfy the assumptions of either Theorem A.1, A.2 or A.3. If $nh^4\to\infty$ as $n\to\infty$, then

$$\mathbb{E}\left[\widehat{C}_h(x)\right]=\frac{\mathbb{E}\left[\hat{f}_h''(x)\right]}{\left(1+\left(\mathbb{E}\left[\hat{f}_h'(x)\right]\right)^2\right)^{\frac{3}{2}}}+O\left(\frac{1}{nh^4}\right).$$

Proof. Taylor's formula applied to function $g(X,Y)=X/\left(1+Y^2\right)^{3/2}$ in a neighbourhood of a point $(X_0,Y_0)\in\mathbb{R}^2$ yields

$$\frac{X}{(1+Y^2)^{\frac{3}{2}}}=\frac{X_0}{\left(1+Y_0^2\right)^{\frac{3}{2}}}+\frac{X-X_0}{\left(1+Y_0^2\right)^{\frac{3}{2}}}-\frac{3X_0Y_0(Y-Y_0)}{\left(1+Y_0^2\right)^{\frac{5}{2}}}+R_2, \tag{A.7}$$

where the remainder term R_2, if written in the Lagrange form, equals

$$R_2 = -\frac{3\eta}{(1+\eta^2)^{\frac{5}{2}}}(X-X_0)(Y-Y_0) + \frac{3\left(4\eta^2-1\right)\xi}{2(1+\eta^2)^{\frac{7}{2}}}(Y-Y_0)^2 \tag{A.8}$$

with $\xi \in (X_0, X)$ and $\eta \in (Y_0, Y)$.

Therefore,

$$|R_2| \leqslant |X-X_0|\,|Y-Y_0| + (|X_0|+|X-X_0|)(Y-Y_0)^2$$

as $\max\left(3\eta/\left(1+\eta^2\right)^{5/2}\right) = 48\sqrt{5}/125 < 1$, $\max\left(3\left(4\eta^2-1\right)/\left(2\left(1+\eta^2\right)^{\frac{7}{2}}\right)\right) = 384\sqrt{7}/2401 < 1$ and $|\xi| \leqslant \max(|X_0|, |X|) \leqslant |X_0| + |X-X_0|$. Now let $(X,Y) = \left(\hat{f}_h{}''(x), \hat{f}_h'(x)\right)$ and $(X_0, Y_0) = \left(\mathbb{E}\left[\hat{f}_h{}''(x)\right], \mathbb{E}\left[\hat{f}_h'(x)\right]\right)$. Then,

$$|\mathbb{E}[R_2]| \leqslant \sqrt{\mathbb{V}\left[\hat{f}_h{}''(x)\right]}\left(\sqrt{\mathbb{V}\left[\hat{f}_h'(x)\right]} + \sqrt{\mathbb{E}\left[\hat{f}_h'(x) - \mathbb{E}\left[\hat{f}_h'(x)\right]\right]^4}\right) + O(1)\mathbb{V}\left[\hat{f}_h'(x)\right],$$

where we have used the Cauchy-Schwarz inequality. The O-term above can be justified as follows. Observe first that

$$\mathbb{E}\left[\hat{f}_h^{(r)}(x)\right] = \frac{1}{h^r}\int_{-\infty}^{\infty} K^{(r)}(u) f(x-uh)\,du.$$

By repeated integration by parts of the integral, one can observe that if $K^{(r)}(u) \to 0$ as $|u| \to \infty$, then the latter expression simplifies to

$$\mathbb{E}\left[\hat{f}_h^{(r)}(x)\right] = \int_{-\infty}^{\infty} K(u) f^{(r)}(x-uh)\,du.$$

Under the conditions of Theorem A.1, A.2 or A.3 on $f^{(r)}(\cdot)$ and $K(\cdot)$ (for $r=2$),

$$\mathbb{E}\left[\hat{f}_h^{(r)}(x)\right] = f^{(r)}(x) + \frac{h^2}{2} f^{(r+2)}(x)\kappa_2(K) + o(h^2) \tag{A.9}$$

as $h \to 0$ (cf. Henderson and Parmeter, 2015, Section 2.6). Hence, $\mathbb{E}\left[\hat{f}_h^{(r)}(x)\right] =$

$O(1)$. By Lemma A.9, we obtain

$$\begin{aligned}|\mathbb{E}[R_2]| &\leqslant O\left(\tfrac{1}{nh^4}\right)+O\left(\tfrac{1}{n^{3/2}h^6}\right)+O\left(\tfrac{1}{nh^3}\right),\\ &= O\left(\max\left\{\tfrac{1}{nh^4},\tfrac{1}{n^{3/2}h^6},\tfrac{1}{nh^3}\right\}\right)\\ &= O\left(\tfrac{1}{nh^4}\right),\end{aligned}$$

which concludes the proof of the lemma. ■

Let us now turn to deriving an expression for the variance $\mathbb{V}\left[\widehat{C}_h(x)\right]$. We will need a few lemmas again.

Lemma A.11. If $nh^4 \to \infty$ as $n \to \infty$, $\mathbb{V}\left[\widehat{C}_h(x)\right]$ admits the following decomposition:

$$\mathbb{V}\left[\widehat{C}_h(x)\right]=\mathbb{E}\left[\left(\widehat{C}_h(x)-\frac{\mathbb{E}\left[\hat{f}_h{}''(x)\right]}{\left(1+\left(\mathbb{E}\left[\hat{f}_h{}'(x)\right]\right)^2\right)^{\frac{3}{2}}}\right)^2\right]+O\left(\frac{1}{n^2h^8}\right).$$

Proof. The proof is straightforward based on linear properties of expectations and Lemma A.10. ■

Lemma A.12. Let PDF $f=f(x)$ possesses finite fourth order derivative at some $x\in\mathbb{R}$. Suppose also that $K^{(r)}(u)\to 0$ as $|u|\to\infty$ for $r\in\{0,1,2\}$ and that $f^{(4)}(\cdot)$ and $K(\cdot)$ satisfy the assumptions of either Theorem A.1, A.2 or A.3. If $nh^5\to\infty$ as $n\to\infty$, then $\mathbb{V}\left[\widehat{C}_h(x)\right]$ can also be decomposed as follows:

$$\begin{aligned}\mathbb{V}\left[\widehat{C}_h(x)\right] &= \tfrac{\mathbb{V}[X]}{\left(1+(\mathbb{E}[Y])^2\right)^3}+\tfrac{9(\mathbb{E}[X])^2(\mathbb{E}[Y])^2}{\left(1+(\mathbb{E}[Y])^2\right)^5}\mathbb{V}[Y]-\tfrac{6\mathbb{E}[X]\mathbb{E}[Y]}{\left(1+(\mathbb{E}[Y])^2\right)^4}Cov[X,Y]\\ &\quad+O\left(\tfrac{1}{n^{3/2}h^7}\right),\end{aligned}$$

where $X=\hat{f}_h{}''(x)$ and $Y=\hat{f}_h{}'(x)$.

Proof. With the help of (A.7), we can write

$$\mathbb{E}\left[\left(\widehat{C}_h(x)-\frac{X_0}{\left(1+Y_0^2\right)^{\frac{3}{2}}}\right)^2\right]=\mathbb{E}\left[\left(\frac{X-X_0}{\left(1+Y_0^2\right)^{\frac{3}{2}}}-\frac{3X_0Y_0\,(Y-Y_0)}{\left(1+Y_0^2\right)^{\frac{5}{2}}}+R_2\right)^2\right],$$

where $(X_0,Y_0)=\left(\mathbb{E}\left[\hat{f}''_h(x)\right],\mathbb{E}\left[\hat{f}'_h(x)\right]\right)$. Expanding the brackets, the arrive at

$$\begin{aligned}\mathbb{E}\left[\left(\widehat{C}_h(x)-\frac{X_0}{\left(1+Y_0^2\right)^{\frac{3}{2}}}\right)^2\right] &= \mathbb{E}\left[\left(\frac{X-X_0}{\left(1+Y_0^2\right)^{\frac{3}{2}}}-\frac{3X_0Y_0(Y-Y_0)}{\left(1+Y_0^2\right)^{\frac{5}{2}}}\right)^2\right]\\ &+2\mathbb{E}\left[R_2\left(\frac{X-X_0}{\left(1+Y_0^2\right)^{\frac{3}{2}}}-\frac{3X_0Y_0(Y-Y_0)}{\left(1+Y_0^2\right)^{\frac{5}{2}}}\right)\right]\\ &+\mathbb{E}\left[R_2^2\right].\end{aligned} \tag{A.10}$$

From the proof of Lemma A.10, we infer that

$$\mathbb{E}\left[R_2^2\right]\leqslant\mathbb{E}\left[\left(|X-X_0|\,|Y-Y_0|+(|X_0|+|X-X_0|)\,(Y-Y_0)^2\right)^2\right].$$

Recall that $\mathbb{E}\left[\hat{f}_h^{(r)}(x)\right]=O(1)$ and observe that $\mathbb{E}\left[\left(\hat{f}_h^{(r)}(x)\right)^2\right]=O(1)$. The latter can be established as follows. By routine calculations, upon using (A.9), we arrive at

$$\mathbb{V}\left[\hat{f}_h^{(r)}(x)\right]=\frac{1}{nh^{1+2r}}\int_{-\infty}^{\infty}\left(K^{(r)}(u)\right)^2 f(x-uh)du+o(n^{-1}).$$

If $f(\cdot)$ and $(K^{(r)}(\cdot))^2$ satisfy the assumptions of either Theorem A.1, A.2 or A.3, then the latter integral can be extended to result in Henderson and Parmeter (2015, pp. 45-50):

$$\mathbb{V}\left[\hat{f}_h^{(r)}(x)\right]=\frac{f(x)R(K^{(r)})}{nh^{1+2r}}+O\left(\frac{1}{nh^{1+2r}}\right), \tag{A.11}$$

where $R(W)=\int_{-\infty}^{\infty}W^2(u)du$.

As $nh^5\to 0$ by assumption, then it follows that $\mathbb{E}\left[\left(\hat{f}_h^{(r)}(x)\right)^2\right]=O(1)$ as desired. Then, by applying the Cauchy-Schwarz inequality, we get

$$\begin{aligned}\mathbb{E}\left[R_2^2\right] &\leqslant \sqrt{\mathbb{E}\left[|X-X_0|^4\right]}\sqrt{\mathbb{E}\left[|Y-Y_0|^4\right]}+\sqrt{\mathbb{E}\left[|X-X_0|^4\right]}\sqrt{\mathbb{E}\left[|Y-Y_0|^8\right]}\\ &+2\,|X_0|\,\sqrt{\mathbb{V}[X]}\sqrt{\mathbb{E}\left[|Y-Y_0|^8\right]}+X_0^2\mathbb{E}\left[|Y-Y_0|^4\right]\\ &+2\sqrt{\mathbb{E}\left[|X-X_0|^4\right]}\sqrt{\mathbb{E}\left[|Y-Y_0|^6\right]}+2X_0\sqrt{\mathbb{V}[X]}\sqrt{\mathbb{E}\left[|Y-Y_0|^6\right]}.\end{aligned}$$

The moments can be estimated as in Lemma A.9:

$$\mathbb{E}\left[R_2^2\right] \leqslant O\left(\max\left\{\frac{1}{n^2h^9},\frac{1}{n^3h^{13}},\frac{1}{n^{5/2}h^{10}},\frac{1}{n^2h^7},\frac{1}{n^{5/2}h^{11}},\frac{1}{n^2h^8}\right\}\right) = O\left(\frac{1}{n^2h^9}\right). \tag{A.12}$$

Similarly, the upper bound is obtained for the middle term in (A.10):

$$\begin{aligned}\mathbb{E}\left[R_2\left(\tfrac{X-X_0}{\left(1+Y_0^2\right)^{\frac{3}{2}}}-\tfrac{3X_0Y_0(Y-Y_0)}{\left(1+Y_0^2\right)^{\frac{5}{2}}}\right)\right] &\leqslant \sqrt{\mathbb{E}\left[R_2^2\right]}\sqrt{\mathbb{E}\left[\left(|X-X_0|+|X_0|\,|Y-Y_0|\right)^2\right]}\\ &= O\left(\tfrac{1}{nh^{9/2}}\right)\left(\sqrt{\mathbb{V}(X)}+\sqrt{\mathbb{V}(Y)}\right)\\ &= O\left(\tfrac{1}{nh^{9/2}}\right)O\left(\max\left\{\tfrac{1}{n^{1/2}h^{5/2}},\tfrac{1}{n^{1/2}h^{3/2}}\right\}\right)\\ &= O\left(\tfrac{1}{n^{3/2}h^7}\right).\end{aligned} \tag{A.13}$$

On account of Lemma A.11, (A.12) and (A.13), we get

$$\begin{aligned}\mathbb{V}\left[\widehat{C}_h(x)\right] &= \tfrac{\mathbb{V}[X]}{\left(1+(\mathbb{E}[Y])^2\right)^3}+\tfrac{9(\mathbb{E}[X])^2(\mathbb{E}[Y])^2}{\left(1+(\mathbb{E}[Y])^2\right)^5}\mathbb{V}[Y]-\tfrac{6\mathbb{E}[X]\mathbb{E}[Y]}{\left(1+(\mathbb{E}[Y])^2\right)^4}Cov[X,Y]\\ &+O\left(\tfrac{1}{n^2h^8}\right)+O\left(\tfrac{1}{n^{3/2}h^7}\right)+O\left(\tfrac{1}{n^2h^9}\right).\end{aligned}$$

Noticing that the sum of the three O-terms equals $O\left(1/\left(n^{3/2}h^7\right)\right)$ completes the proof of the lemma. ■

Lemma A.13. Let the kernel $K(\cdot)$ be twice differentiable. Suppose further that the PDF $f(\cdot)$ and function $V(\cdot) = K'(\cdot)K''(\cdot)$ satisfy the assumptions of either Theorem A.1, A.2 or A.3. Then, the covariance of the KDE estimators $\hat{f}_h^{'}(x)$ and $\hat{f}_h^{''}(x)$ from (A.5) of $f^{'}(x)$ and $f''(x)$, respectively, equals

$$Cov\left[\hat{f}_h^{'}(x),\hat{f}_h^{''}(x)\right] = O\left(\frac{1}{nh^3}\right).$$

Proof. Observe that $Cov\left[\hat{f}_h^{'}(x),\hat{f}_h^{''}(x)\right] = \mathbb{E}\left[\hat{f}_h^{'}(x)\hat{f}_h^{''}(x)\right]-\mathbb{E}\left[\hat{f}_h^{'}(x)\right]\mathbb{E}\left[\hat{f}_h^{''}(x)\right]$ simplifies to

$$Cov\left[\hat{f}_h^{'}(x),\hat{f}_h^{''}(x)\right] = \frac{1}{nh^5}\int_{-\infty}^{\infty}K^{'}\left(\frac{x-y}{n}\right)K^{''}\left(\frac{x-y}{n}\right)f(y)\,dy-\frac{1}{n}\mathbb{E}\left[\hat{f}_h^{'}(x)\right]\mathbb{E}\left[\hat{f}_h^{''}(x)\right].$$

The integral can be written as

$$\begin{aligned}\frac{1}{nh^5}\int_{-\infty}^{\infty}K^{'}\left(\frac{x-y}{n}\right)K^{''}\left(\frac{x-y}{n}\right)f(y)\,dy &= \frac{1}{nh^4}\int_{-\infty}^{\infty}K^{'}(w)\,K^{''}(w)\,f(x-hw)\,dw\\ &= \frac{1}{nh^4}\int_{-\infty}^{\infty}K^{'}(w)\,K^{''}(w)\Big(f(x)-hwf^{'}(x)\\ &\quad+\tfrac{1}{2}h^2w^2f^{''}(x)+o\left(h^2\right)\Big)\,dw\\ &= -\frac{1}{nh^3}f^{'}(x)\left(\int_{-\infty}^{\infty}wK^{'}(w)K^{''}(w)dw\right)+o\left(\frac{1}{nh^2}\right).\end{aligned}$$

Here the o-term is justified by Theorems A.1, A.2 or A.3. We have also used the fact that $K(\cdot)$ is symmetric around zero and hence its derivatives $K^{'}(\cdot)$ and $K^{''}(\cdot)$ are odd and even functions, respectively. Hence, both $\int_{-\infty}^{\infty}K^{'}(w)K^{''}(w)dw = 0$ and $\int_{-\infty}^{\infty}w^2K^{'}(w)K^{''}(w)dw = 0$. As $o\left(\frac{1}{nh^2}\right) = O\left(\frac{1}{nh^3}\right)$ and $\frac{1}{n}\mathbb{E}\left[\hat{f}_h^{'}(x)\right]\mathbb{E}\left[\hat{f}_h^{''}(x)\right] = O\left(\frac{1}{n}\right)$, the conclusion of the lemma follows. ∎

A.1.3 Lemmas for Section 2.3.3

In this section, we assume that function $g:\mathbb{R}\to\mathbb{R}$ does not have flat parts. Importance of the role played by flat parts has already been recognized in the literature (Müller and Sawitzki, 1991, p. 740).

Lemma A.14. Let $g:\mathbb{R}\to\mathbb{R}$ be a continuous function. Suppose $a_1 < a_2$ are points of local maxima of g. Then, between them there exists at least one point b of local minimum.

Proof. By the definition of a local maximum, there exist points $x_1, x_2 \in (a_1, a_2)$ such that $g(x_1) < g(a_1)$ and $g(x_2) < g(a_2)$. Here, x_i is located in a neighbourhood of a_i for $i \in \{1,2\}$. Since g is a continuous function, then in $[a_1, a_2]$ it attains its minimum at some point b. Moreover, $g(b) \leqslant g(x_i) < g(a_i)$, and consequently, $b \neq a_i$, $i \in \{1,2\}$. Thus, point b is an interior point of (a_1, a_2). ∎

Lemma A.15. Let $g:\mathbb{R}\to\mathbb{R}$ be a continuous function. Suppose $a_1 < a_2$ are points of local minima of g. Then, between them there exists at least one point b of local maximum.

Proof. Proof is analogous to that of Lemma A.14. ∎

Lemma A.16. Let $g:\mathbb{R}\to\mathbb{R}$ be a continuous function. Suppose $a_1 < a_2$ are two consecutive points of local maxima of g. Then, between them there exists only one point b of local minimum.

Proof. By Lemma A.14, there exists at least one such minimum. If there were more than one minimum, then between them by Lemma A.15 there would be yet another local maximum. In this case, a_1 and a_2 will not be consecutive local maxima. ■

Lemma A.17. Let $g : \mathbb{R} \to \mathbb{R}$ be a continuous function. Suppose $a_1 < a_2$ are two consecutive points of local minima of g. Then, between them there exists only one point b of local maximum.

Proof. Proof is analogous to that of Lemma A.16. ■

Lemma A.18. Let $g : \mathbb{R} \to \mathbb{R}$ be such that $\lim_{x\to+\infty} g(x) = 0$. Let g be monotonically decreasing in a neighbourhood of $+\infty$, i.e., there exists an $M \in \mathbb{R}$ such that g decreases in $(M, +\infty)$. Then, the last (the rightmost) extremum is a local maximum.

Proof. Let us define $\xi = \inf\{m \in \mathbb{R} : \text{function } g \text{ decreases in } (m, +\infty)\}$. Then, ξ is a point of maximum followed by no extrema. ■

Appendix B

For Chapter 3

This appendix uses material from the following article:

- K. Mayorov, J. Hristoskov and N. Balakrishnan. On a family of weighted Cramér-von Mises goodness-of-fit tests in operational risk modeling. *Journal of Operational Risk*, forthcoming 2017.

The material is reproduced with permission from Incisive Media.

B.1 Auxiliary Results

Lemma B.1. Let $\Lambda_\theta(a) = \sqrt{\theta^2 - 2ai}$,[1] $a \in \mathbb{R}$, $\theta \in (0,\infty)$ and $T \in [0,\infty)$. We have

(i) The characteristic function of $\int_0^T Y_\theta^2(t)\,dt$ is equal to

$$\mathbb{E}\left[\exp\left(ia\int_0^T Y_\theta^2(t)\,dt\right)\right] = \sqrt{\frac{\Lambda_\theta(a)\exp(T\theta)}{\left(\theta - \frac{ai}{\theta}\right)\sinh(T\Lambda_\theta(a)) + \Lambda_\theta(a)\cosh(T\Lambda_\theta(a))}};$$

(ii) The characteristic functions of $Z_{\theta,x}(t)$ and $Z_\theta(t)$ are equal to

$$\begin{aligned}\mathbb{E}\left[\exp\left(iaZ_{\theta,x}(T)\right)\right] &= \mathbb{E}\left[\exp\left(iaZ_\theta(T)\right)\right] \\ &\quad\times\exp\left(\frac{iax^2\sinh(T\Lambda_\theta(a))}{\theta\sinh(T\Lambda_\theta(a))+\Lambda_\theta(a)\cosh(T\Lambda_\theta(a))}\right)\end{aligned}$$

[1]The branch of the square root is chosen consistent with the branch chosen for $a = 0$.

and

$$\mathbb{E}\left[\exp\left(iaZ_\theta(T)\right)\right] = \sqrt{\frac{\Lambda_\theta(a)\exp\left(T\theta\right)}{\theta\sinh\left(T\Lambda_\theta(a)\right) + \Lambda_\theta(a)\cosh\left(T\Lambda_\theta(a)\right)}}, \qquad \text{(B.1)}$$

respectively.

Proof. Let us establish Part (ii). The proof of Part (i) is analogous. Let

$$\varphi_S(a) = \mathbb{E}\left[\exp\left(iaS\right)\right] \text{ and } M_S(a) = \mathbb{E}\left[\exp\left(aS\right)\right],$$

$a \in \mathbb{R}$, be the characteristic function and moment-generating function, respectively, of a random variable S.

Let us consider an OU process $\{X^*_{1,x^*}(t), t \geqslant 0\}$ with parameter $\theta = 1$ and a starting value $x^* \in \mathbb{R}$ and the corresponding $Z^*_{1,x^*}(T) = \int_0^T (X^*_{1,x^*}(t))^2\,dt$ for $0 \leqslant T < \infty$.

Dankel (1991, pp. 569-571) derived expressions of the Laplace transforms

$$M_{\int_0^T Y_1^2(t)\,dt}(-a) \text{ and } M_{Z^*_{1,x^*}(T)}(-a)$$

for $a \geqslant 0$. The expressions remain well-defined for $a > -1/2$ and, in particular, for any $a \in (-\delta, \delta)$ with $\delta \in (0, 1/2)$. In this case, it is known (Grimmett and Stirzaker, 2001, p. 184) that

$$M_{Z^*_{1,x^*}(T)}(ia) = \varphi_{Z^*_{1,x^*}(T)}(a)$$

for all $a \in \mathbb{R}$. Equation (3.9) implies $(1/\sqrt{\theta})X^*_{1,x^*}(\theta t) \overset{d}{=} X_{\theta,x}(t)$ for all $t \geqslant 0$, where $x = x^*/\sqrt{\theta}$. Then, the claim of Part (ii) follows from the results of Dankel (1991) after replacing a by $-ia/\theta^2$, T by θT, and x^* by $\sqrt{\theta}x$. ∎

Remark. The Laplace transform of $Z_{\theta,0}(T)$ was also given in Liptser and Shiryaev (2001, p. 232). For $T = 1$, the Laplace transforms were obtained in Gao et al. (2003, pp. 7-8) using the Hadamard Factorization Theorem.

The next lemma establishes a useful central limit theorem.

Lemma B.2. We have, as $T \to \infty$:

(i)

$$\frac{\int_0^T Y_\theta^2(t)\,dt - T/(2\theta)}{\sqrt{T/(2\theta^3)}} \xrightarrow{d} N(0,1);$$

(ii)
$$\frac{\int_0^T X_\theta^2(t)\,dt - T/(2\theta)}{\sqrt{T/(2\theta^3)}} \xrightarrow{d} N(0,1).$$

Proof. Let us prove Part (ii). Let $R_\theta(T) = \frac{Z_\theta(T)-T/(2\theta)}{\sqrt{T/(2\theta^3)}}$. The result will follow by Lévy's continuity theorem (Gut, 2013, pp. 237-239) if we show

$$\mathbb{E}[\exp(iaR_\theta(T))] \to \exp(-a^2/2) \tag{B.2}$$

as $T \to \infty$, for all $a \in \mathbb{R}$.

With the help of the previous lemma, we compute

$$\begin{aligned}\mathbb{E}[\exp(iaR_\theta(T))] &= \exp\left(-\tfrac{1}{2}ia\sqrt{2T\theta}+\tfrac{1}{2}T\theta\right)\\ &\times\sqrt{\frac{2w_{a,\theta}(T)\exp\left((T\theta)^{3/4}w_{a,\theta}(T)\right)}{\left((T\theta)^{1/4}+w_{a,\theta}(T)\right)\exp\left(2(T\theta)^{3/4}w_{a,\theta}(T)\right)-(T\theta)^{1/4}+w_{a,\theta}(T)}},\end{aligned}$$

where $w_{a,\theta}(T) = \sqrt{\sqrt{T\theta}-2\sqrt{2}ia}$.[2]

Firstly, let us demonstrate that

$$\lim_{T\to\infty}\left(-\frac{1}{2}ia\sqrt{2T\theta}+\frac{1}{2}T\theta-\frac{1}{2}(T\theta)^{3/4}w_{a,\theta}(T)\right) = -\frac{a^2}{2}. \tag{B.3}$$

For brevity, let L denote the left-hand side of (B.3). Then, upon rewriting the expression under the limit sign as

$$\begin{aligned}-\tfrac{1}{2}ia\sqrt{2T\theta}+\tfrac{1}{2}T\theta-\tfrac{1}{2}(T\theta)^{3/4}w_{a,\theta}(T) &= \tfrac{1}{2}\Big(\left(T\theta-ia\sqrt{2T\theta}\right)\\ &\quad-\sqrt{(T\theta)^2-2\sqrt{2}iaT\theta\sqrt{T\theta}}\Big)\end{aligned}$$

and multiplying and dividing the right-hand side of the last equality by $\left(T\theta-ia\sqrt{2T\theta}\right)+\sqrt{(T\theta)^2-2\sqrt{2}iaT\theta\sqrt{T\theta}}$, we arrive at

[2] The branch of the square root is chosen consistent with the branch chosen for $a=0$.

$$
\begin{aligned}
L &= \frac{1}{2}\lim_{T\to\infty}\frac{-2a^2T\theta}{T\theta-ia\sqrt{2T\theta}+\sqrt{(T\theta)^2-2\sqrt{2}iaT\theta\sqrt{T\theta}}} \\
&= -\frac{a^2}{2}\lim_{T\to\infty}\frac{1}{\left(\frac{1}{2}-ia\frac{1}{\sqrt{2T\theta}}\right)+\frac{1}{2}\sqrt{1-\frac{2\sqrt{2}ia}{\sqrt{T\theta}}}} \\
&= -\frac{a^2}{2},
\end{aligned}
$$

as desired.

Secondly, let us demonstrate that

$$
\lim_{T\to\infty}\sqrt{\frac{2w_{a,\theta}(T)}{\left(1-w^*_{a,\theta}(T)\right)(T\theta)^{1/4}+\left(1+w^*_{a,\theta}(T)\right)w_{a,\theta}(T)}}=1, \qquad \text{(B.4)}
$$

where $w^*_{a,\theta}(T)=\exp\left(-2(T\theta)^{3/4}w_{a,\theta}(T)\right)$.

Indeed, the fact that $w^*_{a,\theta}(T)\to 0$ as $T\to\infty$ and

$$
\lim_{T\to\infty}\frac{(T\theta)^{1/4}}{\sqrt{\sqrt{T\theta}-2\sqrt{2}ia}}=1,
$$

yield (B.4).

On combining (B.3) and (B.4), we readily obtain (B.2). This concludes the proof of Part (ii). The proof of Part (i) is analogous. ■

Remark. We observe that Part (i) is a particular case of a result of Bhattacharya (refer to Bhattacharya (1982, p. 188, Theorem 2.1) or Bhattacharya and Waymire (2009, p. 440, Proposition 13.2)). Indeed, the result states that

$$
\frac{\int_0^T\left(Y^2_\theta(t)-1/(2\theta)\right)dt}{\sqrt{T}}\xrightarrow{d}N(0,\sigma),
$$

where $\sigma^2=2\int_{-\infty}^{\infty}(x^2-1/(2\theta))\mathscr{L}_\theta^{-1}(1/(2\theta)-x^2/(2\theta))m_\theta(x)dx$, where $\mathscr{L}_\theta=\frac{1}{2}\frac{d^2}{dx^2}-\theta x\frac{d}{dx}$ is the infinitesimal generator for (3.8). But, $\mathscr{L}_\theta(x^2/(2\theta))=1/(2\theta)-x^2$. By simple calculations, then $\sigma^2=1/\left(2\theta^3\right)$.

Remark. As part of Csörgő and Horváth (1988, Theorem 3.4), Part (i) is established for the stationary OU process parameterized differently from (3.8) (the proof is based on Mandl (1968, p. 94, Theorem 9)). In fact, in Csörgő and Horváth (1988) the OU process actually solves

$$d\xi(t) = -\xi(t)\,dt + \sqrt{2}dW(t). \tag{B.5}$$

This is a zero-mean process with covariance function $\mathbb{E}[\xi(t), \xi(s)] = \exp(-|t-s|)$. It is not hard to see that $\xi(t) \stackrel{d}{=} \sqrt{2}Y_1(t)$, where $\{Y_1(t), t \geqslant 0\}$ is the OU process solving (3.8) for $\theta = 1$.

We will need the following two lemmas.

Lemma B.3. We have

$$\int_{U_{(n)}}^{1} \frac{e_n^2(t)}{(1-t)^2}\,dt \xrightarrow{d} X,$$

where X is an exponentially distributed random variable with mean 1.

Proof. Since $\int_{U_{(n)}}^1 \frac{e_n^2(t)}{(1-t)^2}\,dt = n\left(1-U_{(n)}\right)$, the result follows. ■

Lemma B.4. For $k(n) = n/\ln^2 n$, as $n \to \infty$, we have

(i) $\int_0^{U_{(k(n))}} \frac{e_n^2(t)}{(1-t)^2}\,dt = o_P(1)$.

(ii) $\int_{U_{(n-k(n))}}^1 \frac{e_n^2(t)}{t^2}\,dt = o_P(1)$.

Proof. Observe that

$$\int_0^{U_{(k(n))}} \frac{e_n^2(t)}{(1-t)^2}\,dt \leqslant \frac{1}{(1-U_{(k(n))})^2} \int_0^{U_{(k(n))}} e_n^2(t)\,dt$$

and

$$\int_{U_{(n-k(n))}}^1 \frac{e_n^2(t)}{t^2}\,dt \leqslant \frac{1}{U_{(n-k(n))}^2} \int_{U_{(n-k(n))}}^1 e_n^2(t)\,dt.$$

Since $nU_{(k(n))}/k(n) \xrightarrow{P} 1$ and $nU_{(n-k(n))}/(n-k(n)) \xrightarrow{P} 1$ (Csörgő and Horváth, 1993, p. 335, Equation (5.3.111)), then $U_{(k(n))} \xrightarrow{P} 0$, $U_{(n-k(n))} \xrightarrow{P} 1$. With this and Proposition 3.4 we have

$$\int_0^{U_{(k(n))}} \frac{e_n^2(t)}{(1-t)^2}\,dt \leqslant O_P(1)o_P(1) \text{ and } \int_{U_{(n-k(n))}}^1 \frac{e_n^2(t)}{t^2}\,dt \leqslant O_P(1)o_P(1),$$

which establish Part (i) and Part (ii), respectively. ■

Lemma B.5. We have, as $T \to \infty$, wp1:

(i)
$$\frac{\int_0^T Y_\theta^2(t)\,dt}{T} \to \frac{1}{2\theta};$$

(ii)
$$\frac{\int_0^T X_\theta^2(t)\,dt}{T} \to \frac{1}{2\theta}.$$

Proof. Part (i). This follows from the fact that $\{Y_\theta(t), t \geqslant 0\}$ is ergodic and the expectation $\mathbb{E}_{m_\theta}\left[Y_\theta^2(t)\right] = 1/(2\theta)$.

Part (ii). Every ergodic process is stationary. Hence, $\{X_\theta(t), t \geqslant 0\}$ is not ergodic. However, the process is asymptotically stationary in the sense of Parzen Parzen (1962a) as

$$\lim_{t\to\infty} \mathbb{E}\left[X_\theta(t) X_\theta(t+v)\right] = \frac{1}{2\theta}\exp(-\theta|v|)$$

for any $v \geqslant 0$ and $\exp(-\theta|v|)/(2\theta) = \mathbb{E}\left[Y_\theta(t)Y_\theta(t+v)\right]$. The validity of the statement of Part (ii) will follow (Parzen, 1962a) if we prove that for all $v \geqslant 0$ there exists a positive number q such that

$$\lim_{t\to\infty} t^q C_t(v) = 0, \tag{B.6}$$

where
$$C_t(v) = \frac{1}{t}\int_0^t cov\left[X_\theta(s)X_\theta(s+v), X_\theta(t)X_\theta(t+v)\right] ds.$$

But,

$$\begin{aligned} cov[X_\theta(s)X_\theta(s+v), X_\theta(t)X_\theta(t+v)] = \; & cov[X_\theta(s)X_\theta(t), X_\theta(s+v)X_\theta(t+v)] \\ & + cov[X_\theta(s)X_\theta(t+v), X_\theta(t)X_\theta(s+v)], \text{ or} \end{aligned}$$

$$\begin{aligned} cov[X_\theta(s)X_\theta(s+v), X_\theta(t)X_\theta(t+v)] = \; & \tfrac{1}{4\theta^2}(\exp(-2\theta|t-s|) \\ & -\exp(-\theta(|t-s|+t+s))) \\ & -\exp(-\theta(|t-s|+t+s+2v)) \\ & +2\exp(-2\theta(t+s+v)) \\ & -\exp(-\theta(|t-s+v|+t+s+v)) \\ & +\exp(-\theta(|t-s+v|+|t-s-v|)) \\ & -\exp(-\theta(|t-s-v|+t+s+v))). \end{aligned}$$

Assuming $v \leqslant t$, we readily obtain that for any $q \in [0,1)$ (B.6) holds true. For $t \to \infty$, the case $v > t$ is only possible when $v = \infty$. In this case, for any $q \in [0,1)$,

$$\lim_{t,v \to \infty} t^q C_t(v) = 0.$$

Thus, (B.6) is true for all $v \geqslant 0$ and any $q \in [0,1)$. This completes the proof of Part (ii) of Lemma B.5. ∎

Lemma B.6. Define the integral operator K with kernel $k(\cdot,\cdot)$ as

$$(Kg)(t) = \int_0^T k(t,s)g(s)\,ds$$

with kernel

$$k(t,s) = \frac{1}{2\theta}\left(\exp\left(-\theta\,|t-s|\right) - \exp\left(-\theta\,(t+s)\right)\right),$$

where $\theta > 0$, $g \in L^2([0,T])$ and $t \in [0,T]$. The spectrum $\sigma(K)$ of K satisfies

(i) $\sigma(K) \setminus \{0\} = \{\lambda_k \mid \lambda_k > 0 \text{ is an eigenvalue and } k \in \mathbb{N}\}$ if $T > 0$ is finite;

(ii) $\sigma(K) = [0, 1/\theta^2]$ is purely continuous if $T = +\infty$.

Proof. First note that K, a linear operator acting in the separable Hilbert space $L^2([0,T])$, is self-adjoint as the kernel $k(t,s) = k(s,t)$ is symmetric (Pugachev and Sinitsyn, 1999, p. 521).

Part (i). It is known (Reed and Simon, 1980, p. 210) that K is Hilbert-Schmidt (and, in particular, bounded) if and only if

$$\int_0^T \int_0^T (k(t,s))^2\,dt\,ds < \infty.$$

The evaluation of the left-hand side yields

$$\int_0^T \int_0^T (k(t,s))^2\,dt\,ds = \frac{4T\theta + 4\exp(-2T\theta) + \exp(-4T\theta) + 8T\theta\exp(-2T\theta) - 5}{16\theta^4}, \tag{B.7}$$

which is finite for $0 \leqslant T < \infty$. It is readily seen that the operator K corresponds to $Z_{\theta,x}(T)$.

As the $k(t,s) = k(s,t)$ for all $0 \leqslant t,s \leqslant T$, K is a self-adjoint Hilbert-Schmidt operator, its spectrum $\sigma(K)$ consists of no more than a countable set of eigenvalues and the point 0 (Pugachev and Sinitsyn, 1999, p. 507).

In Corlay and Pagès (2015), the expressions of the eigenvalues of K are determined, so

$$\sigma(K)\setminus\{0\} = \begin{aligned}&\Big\{\lambda_n > 0,\, n\in\mathbb{N} \mid \lambda_n = \tfrac{1}{\omega_n^2+\theta^2}, \text{where}\\ &\theta\sin(T\omega_n)+\omega_n\cos(T\omega_n)=0,\\ &0<\omega_1<\dots<\omega_n<\dots\Big\}.\end{aligned}$$

This proves the first part of the Lemma.

Part (ii). It follows from (B.7) that $\int_0^\infty\int_0^\infty (k(t,s))^2\,dt\,ds$ is infinite. Hence, the operator K is not Hilbert-Schmidt on $L^2([0,\infty))$. However, it is still bounded. This can be demonstrated as follows.

Let us observe that the integral equation $(Kg)(t)=\lambda g(t)$ can be rewritten as $\int_0^\infty k^*(t,s)g^*(s)\,ds = \lambda^* g^*(t)$, where $k^*(t,s)=\frac{1}{2}\left(\exp(-|t-s|)-\exp(-(t+s))\right)$, $g^*(t)=g(t/\theta)$, and $\lambda^*=\theta^2\lambda$. From Krein (1953, pp. 622-623) and Kostrikin and Makarov (2008, p. 2068), we know that to $k^*(t,s)$ there corresponds a bounded in $L^2([0,\infty))$ operator $K^*g^*(t)=\int_0^\infty k^*(t,s)g^*(s)\,ds$.

The operator K^* is the resolvent of the Dirichlet one-dimensional Laplacian $-d^2/dt^2$ at the point -1 which has purely continuous spectrum[3] filling in $[0,1]$. Then, by the Spectral Mapping Theorem (Reed and Simon, 1980, p. 222) we conclude that the spectrum of K is

$$\sigma(K) = \left[0, \frac{1}{\theta^2}\right]. \tag{B.8}$$

This completes the proof of Lemma B.6. ■

Another result from Corlay and Pagès (2015) is that

$$\lim_{n\to\infty}\left(\omega_n - \frac{\pi}{T}\left(n-\frac{1}{2}\right)\right) = 0.$$

For large n, $\omega_n \approx \frac{\pi}{T}\left(n-\frac{1}{2}\right)$ which implies that as $T\to\infty$ and $n\to\infty$, $\lambda_n \to 1/\theta^2$.

Since $\{X_{\theta,x}(t), t\in[0,\infty)\}$ is a Gaussian process with continuous covariance $k(\cdot,\cdot)$, it is known (see Shorack and Wellner (2009, pp. 201–203) and Pavliotis

[3] In fact, it is purely absolutely continuous. For the definition of the purely absolutely continuous spectrum, see Reed and Simon (1980, p. 231).

(2014, pp. 17–23)) that its Karhunen-Loéve expansion is

$$X_{\theta,x}(t) = \sum_{n=1}^{\infty} \sqrt{\lambda_n}\xi_n b_n(t), \tag{B.9}$$

where $t \in [0,T]$ and $\{\xi_n\}_{n=1}^{\infty}$ are IID distributed standard Gaussian random variables and $b_n(t)$ is the normalized eigenfunction corresponding to eigenvalue λ_n.

Due to the orthonormality of $\{b_n(t)\}_{n=1}^{\infty}$, from (B.9), we readily get:

$$Z_{\theta,x}(T) = \sum_{n=1}^{\infty} \lambda_n \xi_n^2. \tag{B.10}$$

As byproducts, for an important case of the zero-start $Z_\theta(T)$ from (B.10) and (B.1), we can determine the sum of the eigenvalues and the sum of their squares. Indeed, as in Anderson and Darling (1952, p. 200) and Shorack and Wellner (2009, p. 213), we get

$$\sum_{n=1}^{\infty} \lambda_n = \int_0^T k(t,t)dt$$

and

$$\sum_{n=1}^{\infty} \lambda_n^2 = \int_0^T \int_0^T (k(t,s))^2 \, dt \, ds.$$

Using the fact that $\int_0^T k(t,t)dt = (2\theta T + \exp(-2\theta T) - 1)/(4\theta^2)$ and (B.7), we arrive at

$$\sum_{n=1}^{\infty} \lambda_n = \frac{2\theta T + \exp(-2\theta T) - 1}{4\theta^2},$$

$$\sum_{n=1}^{\infty} \lambda_n^2 = \frac{4T\theta + 4\exp(-2T\theta) + \exp(-4T\theta) + 8T\theta\exp(-2T\theta) - 5}{16\theta^4}.$$

Remark. We also note that conclusion (B.8) is readily reached by applying results of Kozhukhar (1992, Example 1 and Example 3) on the spectrum of compact perturbations of Wiener-Hopf operators.

Appendix C

For Chapter 4

C.1 Auxiliary Results

Theorem C.1 (Theorem 2.1 in Csörgő et al. (1993))**.** Let $-\infty \leqslant a < b \leqslant \infty$, μ be a measure on (a,b) and $\{\xi(t), a < t < b\}$ be a real-valued μ-measurable stochastic process with finite mean. We assume that there are constants $r > 1$ and $C > 0$ such that

$$\mathbb{E}[|\xi(t)|^r] \leqslant C(\mathbb{E}[|\xi(t)|])^r \text{ for all } a < t < b.$$

Then,

$$\int_a^b |\xi(t)|\mu(dt) < \infty \text{ wp1}$$

if and only if

$$\int_a^b \mathbb{E}[|\xi(t)|]\,\mu(dt) < \infty.$$

Corollary C.2 (Corollary 2.2 in Csörgő et al. (1993))**.** Let $0 < p < \infty$, $-\infty \leqslant a < b \leqslant \infty$, and $\{\eta(t), a < t < b\}$ be a Gaussian process with $E[\eta(t)] = 0$ and $\sigma^2(t) = \mathbb{V}[\eta(t)]$. Then,

$$P\left(\int_a^b |\eta(t)|^p dt < \infty\right) = \begin{cases} 1, & \text{if } \int_a^b \sigma^p(t)dt < \infty, \\ 0, & \text{if } \int_a^b \sigma^p(t)dt = \infty. \end{cases}$$

Lemma C.3. Let $h = h(x)$ be the hazard rate of the standard Gaussian distribution $N(0,1)$, i.e., $h(x) = \varphi(x)/(1-\Phi(x))$. Then,

1. $h(x)/x \to 1$ as $x \to \infty$;

2. $h = h(x)$ is a strictly increasing function for all $x \in \mathbb{R}$.

Proof. The first part follows by an application of L'Hôpital's rule.
Let us now establish the second part. Since $(x+t)^2/2 \geqslant x^2/2 + xt$, we have

$$e^{-\frac{(x+t)^2}{2}} \leqslant e^{-\frac{x^2}{2}} e^{-xt}$$

with equality holding only if $t = 0$. Therefore, for $x > 0$,

$$\int_x^\infty e^{-\frac{\tau^2}{2}} d\tau = \int_0^\infty e^{-\frac{(x+t)^2}{2}} dt < e^{-\frac{x^2}{2}} \int_0^\infty e^{-xt} dt = \frac{1}{x} e^{-\frac{x^2}{2}}.$$

Thus, for $x > 0$, we have $1 - \Phi(x) < \varphi(x)/x$, or equivalently,

$$\varphi(x) > x(1 - \Phi(x)). \tag{C.1}$$

Inequality (C.1) can also be written as $h(x) > x$. Obviously, inequality (C.1) holds for $x \leqslant 0$ as well. Hence, $w(x) = \varphi(x) - x(1 - \Phi(x))$ is strictly positive for all $x \in \mathbb{R}$. Since

$$h'(x) = \frac{h(x)w(x)}{1 - \Phi(x)},$$

we conclude that the map $x \mapsto h(x)$ is a strictly increasing function for all $x \in \mathbb{R}$. ∎

Remark. 1. The function $w(x)$ introduced in the proof is strictly decreasing for $x \in \mathbb{R}$ as $w'(x) = \Phi(x) - 1 < 0$ for all $x \in \mathbb{R}$;

2. We also observe that $\lim_{x \to +\infty} w(x) = 0$. Indeed, for $x > 0$, the inequality $0 < w(x) < \varphi(x)$ and the Squeeze Theorem imply the result.

Lemma C.4. For all $\beta \leqslant 2$,

1. $I_1(\beta) = \int_0^1 \varphi^2(\Phi^{-1}(t))/(1-t)^\beta dt < \infty$;

2. $I_2(\beta) = \int_0^1 (\Phi^{-1}(t))^2 \varphi^2(\Phi^{-1}(t))/(1-t)^\beta dt < \infty$.

Proof. To establish convergence of both integrals, due to nonnegativity of the integrands and the fact that $1/(1-t)^\beta \leqslant 1/(1-t)^2$ for all $t \in [0, 1)$, it is sufficient to consider $\beta = 2$. Then, the claim of the Lemma will follow from the Comparison Test of convergence of improper integrals.

Part 1. Making the change of variable $x = \Phi^{-1}(t)$, the integral, for $\beta = 2$, becomes $I_1(2) = \int_{-\infty}^{\infty} h^2(x)\varphi(x)dx$. Due to Part 2 of Lemma C.3 and the fact that $h(0) = \sqrt{2/\pi}$, the integral $\int_{-\infty}^{0} h^2(x)\varphi(x)dx < 1/\pi < \infty$.

As $\lim_{x\to\infty}(h^2(x)\varphi(x))/(x^2\varphi(x)) = 1$, where we have used Part 1 of Lemma C.3, by the Limit Comparison Test, $\int_0^{\infty} h^2(x)\varphi(x)dx$ converges or diverges depending on whether $\int_0^{\infty} x^2\varphi(x)dx$ converges or diverges. But, the latter integral is simply the variance of the standard Gaussian distribution, and is equal to 1. This completes the proof of Part 1 of the Lemma.

Part 2. The proof is analogous to that of Part 1, except $\int_0^{\infty} x^4\varphi(x)dx$ should be considered in the last step. This integral converges and is equal to $\kappa/2 = 3/2$, where $\kappa = 3$ is the kurtosis of the standard Gaussian distribution. ■

Lemma C.5. The following three integrals are finite:

1. $\int_0^1 \frac{\sqrt{t}\varphi(\Phi^{-1}(t))}{(1-t)^{3/2}}dt$;
2. $\int_0^1 \frac{\sqrt{t}\Phi^{-1}(t)\varphi(\Phi^{-1}(t))}{(1-t)^{3/2}}dt$;
3. $\int_0^1 \frac{\Phi^{-1}(t)\varphi^2(\Phi^{-1}(t))}{(1-t)^2}dt$;

Proof. Part 1. By the Comparison Test, it is sufficient to show that $\int_0^1 \varphi(\Phi^{-1}(t))/(1-t)^{3/2}dt$ converges. The latter integral equals $\int_{-\infty}^{\infty}\sqrt{h^3(x)\varphi(x)}dx$. As in the proof of Lemma C.4, it can be readily shown that $\int_{-\infty}^{0}\sqrt{h^3(x)\varphi(x)}dx < \infty$. Since

$$\lim_{x\to\infty}\sqrt{h^3(x)\varphi(x)}/\sqrt{x^3\varphi(x)} = 1,$$

by the Limit Comparison Test, $\int_0^{\infty} h^{3/2}(x)\sqrt{\varphi(x)}dx$ will converge or diverge depending on whether $\int_0^{\infty}\sqrt{x^3\varphi(x)}dx$ converges or diverges. We can partition

$$\int_0^{\infty}\sqrt{x^3\varphi(x)}dx = \int_0^{1}\sqrt{x^3\varphi(x)}dx + \int_1^{\infty}\sqrt{x^3\varphi(x)}dx.$$

The first integral in the sum is finite as the integrand is a bounded continuous function on $[0,1]$. Also, $\int_1^{\infty}\sqrt{x^3\varphi(x)}dx \leqslant \int_{-\infty}^{\infty} x^2\sqrt{\varphi(x)}dx$, and the integral on the right-hand side equals $2^{1/2}(2\pi)^{1/4}\mathbb{V}[X] = \sqrt[4]{128\pi}$, where X is an $N(0,\sqrt{2})$ distributed random variable. Thus, $\int_1^{\infty}\sqrt{x^3\varphi(x)}dx$ converges by the Comparison Test. This completes the proof of Part 1.

Parts 2 and 3. The proof is analogous to that of Part 1. ■

Lemma C.6. Let $Z = (X,Y)$ be a bivariate Gaussian random variable with $\mathbb{E}[X] = 0$, $\mathbb{E}[Y] = 0$. Then, there exists a constant $C > 0$ such that $\mathbb{E}\left[|XY|^2\right] \leqslant C\left(\mathbb{E}\left[|XY|\right]\right)^2$.

Proof. In Nabeya (1951), it is shown that if $\mathbb{V}[X] = \sigma^2$, $\mathbb{V}[Y] = \gamma^2$ and correlation ρ, then $\mathbb{E}\left[|XY|\right] = 2\sigma\gamma\Omega(\rho)/\pi$, where $\Omega(\rho) = \rho \arcsin\rho + \sqrt{1-\rho^2}$. It is known that $\mathbb{E}\left[|XY|^2\right] = \sigma^2\gamma^2(1+2\rho^2)$. Since $\max_{\rho\in[0,1]}\{1/\Omega^2(\rho)\} = 1$ and $\max_{\rho\in[0,1]}\{1+2\rho^2\} = 3$, it is readily seen that if $C = 3\pi^2/4$, then $\mathbb{E}\left[|XY|^2\right] \leqslant C\left(\mathbb{E}\left[|XY|\right]\right)^2$. ■