André Filipe Correia Gomes

Estimação não paramétrica da Perda Esperada

Dissertação de Mestrado em Métodos Quantitativos em Finanças, orientada pelo Professor Doutor Paulo Eduardo Oliveira e apresentada ao Departamento de Matemática da Faculdade de Ciências e Tecnologia e à Faculdade de Economia da Universidade de Coimbra.

Abril de 2017

· U C ·

UNIVERSIDADE DE COIMBRA

Nonparametric estimation of Expected Shortfall

André Filipe Correia Gomes

· U C ·

UNIVERSIDADE DE COIMBRA

Master in Quantitative Methods in Finance

Mestrado em Métodos Quantitativos em Finanças

MSc Dissertation | Dissertação de Mestrado

April 2017

Acknowledgements

It was good as a student to be working on a thesis under Kac - you knew with Kac, you could come in and talk with him if you wanted to, and you would get good advice.

Murray Rosenblatt, about his thesis advisor, Marc Kac

I am most thankful to my thesis advisor, Professor Paulo Eduardo Oliveira, for his guidance, availability, patience and immense knowledge. Much like Kac, there was not a single time in which I left his office without a new idea. I am also thankful for his support. At the seemingly tougher times, when I was not able to make progress in my dissertation, his belief in me and his kind attitude gave me motivation to keep moving forward.

I am thankful to all the professors at DMUC for their teachings. Without a doubt, mathematics has made me a happier person, and I owe it to them.

I am thankful to all my friends for their suggestions and continuous support.

Abstract

The Expected Shortfall is an increasingly popular risk measure in financial risk management. This work seeks to study the asymptotic statistical properties of two nonparametric estimators of Expected Shortfall, under the assumption of dependence in the time series of study. The first estimator can be seen as an average of values that satisfy a certain property, whereas the second estimator is a kernel smoothed version of the first. The assumption of dependence is considered one of weakest (α-mixing), for which reason the control of the presented random variables (namely their variances and covariances) has a big emphasis on this work. Due to this control we are able to present a Central Limit Theorem for each estimator, from which we are able to draw relevant conclusions about the efficiency of both estimators.

Table of contents

Chapter 1

Introduction

Value-at-Risk (VaR) and Expected Shortfall (ES) are popular measures of financial risk associated with an asset or a portfolio of assets. In Artzner et al. (1999), four desirable properties for measures of risk are introduced, and if a measure of risk fulfills those properties it is considered to be coherent. One of those properties is subadditivity, and in Artzner et al. (1999) it is proven that VaR is not subadditive, and consequently, it is not considered a coherent risk measure. On the other hand, ES fulfills all four properties, making it a coherent risk measure. To introduce both these risk measures, let $(X_t, t = 0, ..., n)$ be the market values of an asset or a portfolio of assets over $n+1$ periods of a time unit, and let $Y_t = -\log(X_t/X_{t-1})$, $t = 1, ..., n$, be the negative log return on the t-th time period. Suppose $(Y_t, t = 1, ..., n)$ is a dependent and (strictly) stationary family of random variables, with marginal distribution function F. Given a positive value p close to zero, the $1-p$ level VaR, denoted as ν_p, is defined as

$$\nu_p = \inf\{u : F(u) \geq 1-p\}.$$

Simply put, VaR is the $(1-p)$ quantile of the loss distribution F. This value specifies the smallest amount of loss such that the probability of a loss being larger than ν_p, is at most p. Besides the incoherence as a risk measure that we have pointed before, there is another obvious shortcoming of VaR - it provides no information about the loss when it surpasses ν_p. That is not the case with ES. The ES associated with the $1-p$ level, denoted as μ_p, is defined as

$$\mu_p = \mathrm{E}(Y_t | Y_t > \nu_p).$$

ES is the conditional expectation of a loss, given that the loss is larger than ν_p. It is a value of great importance in the financial and actuarial contexts, and its estimation is commonly done by assuming a parametric loss distribution. A popular parametric method to estimate the ES, is the extreme value theory approach (Embrechts et al., 1997), which uses the asymptotic distribution of exceedance over a high threshold to model the excessive losses and carries out parametric inference within the framework of the Generalized Pareto distributions.

Besides the well known disadvantage of parametric estimation - the underlying assumption being too rigid - data is generally sparse in the tail part of the loss distribution, which is the part that is more relevant in this framework, and that makes the choosing of a loss model for parametric estimation not trivial. Moreover, the extreme value theory approach presented in Embrechts et al. (1997)

considers conditions under which the high exceedances are asymptotically independent and identically distributed, and a more recent empirical study by Bellini and Figá-Talamanca (2004) has shown that financial returns can exhibit strong tail dependence even for large threshold levels. This is a strong reason to consider the dependence in financial returns. The nonparametric approach allows for a wide range of data dependence, and has the advantage of being free of distributional assumptions on $(Y_t, t = 1, ..., n)$ while being able to capture fat-tailed and asymmetric distribution of returns.

Before introducing the ES estimators that will be studied in this work, it is relevant to introduce some definitions and theorems to give more motivation for this subject.

This work is heavily based on Chen (2008) and is structured as follows. In Chapter 1 we present definitions and theorems that will be needed in this work and we introduce the nonparametric ES estimators we will study and the assumptions of this work. In Chapter 2 we derive the results that lead to two Central Limit Theorems, which are the main focus of this work, and in Chapter 3 we analyze the obtained results and suggest future work.

1.1 Notation

$(\Omega, \mathscr{A}, \mathrm{P})$ Probability space: Ω non-empty set, $\mathscr{A}$ σ-Algebra of subsets of Ω, P Probability measure on $\mathscr{A}$,

$\int f(x)dx$ integral of f over $\mathbb{R}$,

$L^q(P)$ space of real measurable functions f such that

$||f||_q = \left(\int |f|^q dP\right)^{1/q} < \infty \; (1 \leq q < \infty)$,

$||f||_\infty = \inf\{a : P(f > a) = 0\} < \infty \; (q = \infty)$,

$\mathrm{E}(X), \mathrm{Var}(X)$ Expectation and Variance of X,

$\mathrm{Cov}(X,Y), \mathrm{Corr}(X,Y)$ Covariance and Correlation coefficient of X and Y,

$P_X = P_Y$ distribution of X and Y is the same,

$[x]$ integer part of x,

$I(x \in A)$ indicator function of A: $I(x \in A) = 1, x \in A; = 0, x \notin A$,

$\mathscr{N}(\mu, \sigma^2)$ a normally distributed random variable with mean μ and variance σ^2,

$\sigma(X_i, i \in T)$ sigma-algebra generated by the random variables $X_i, i \in T$,

:= definition.

1.2 Definitions and Motivation

1.2.1 Kernel Density Estimator

Suppose that $Y_1,...,Y_n$ is a set of absolutely continuous random variables with common density f, and let Y be a random variable also having density f. Perhaps the most widely known nonparametric estimator for the density function f is the histogram. The histogram is created by dividing the real line into equally sized intervals, called bins. If b denotes the width of the bins, then the histogram estimate at a point x is given by

$$\hat{f}_H(x;b) = \frac{\text{number of observations in bin containing } x}{nb}.$$

The value of the binwidth b affects the shape of the histogram. A lower value leads to a more 'spiked' histogram, while a higher value leads to a smoother looking histogram. This is why the binwidth b is called a 'smoothing parameter', as its selection controls the amount of smoothing in the obtained histogram. There are several well known problems with using the histogram as a density estimator, such as the placement of the bin edges, and the fact that most density functions are not step functions. Moreover, it can be shown that in a certain sense, the histogram does not use the available data efficiently. We will discuss about this matter later.

We now introduce the kernel density estimator (see Rosenblatt (1956) and Parzen (1962) to get insight on how it came about). It is defined as

$$\hat{f}(x;h) = \frac{1}{nh}\sum_{t=1}^{n} K\left(\frac{x-Y_t}{h}\right), \tag{1.1}$$

where K is a function satisfying $\int K(u)\,du = 1$, which we call the kernel, and h is a positive number, usually called the bandwidth. It serves as the smoothing parameter. Setting $K_h(u) = h^{-1}K(u/h)$ allows us to write

$$\hat{f}(x;h) = \frac{1}{n}\sum_{t=1}^{n} K_h(x-Y_t). \tag{1.2}$$

Usually K is chosen to be a unimodal probability density function that is symmetric around zero. This ensures that $\hat{f}(x;h)$ is a density itself. Moreover, it is clear that if K is continuous and differenciable, then the kernel density estimator inherits those important properties, which makes it a more interesting density estimator than the histogram. There are several interesting topics about the kernel density estimator, such as how the selection of K and its shape can influence our estimate, or how one should choose the value h. We refer to Wand and Jones (1995) for the interested reader.

1.2.2 The O, o notation.

To study some basic properties of the kernel density estimator, it is convenient to introduce order and asymptotic notation. We will follow the approach in Bishop (1975) and Wand and Jones (1995).

Let $(a_n, n \in \mathbb{N})$ and $(b_n, n \in \mathbb{N})$ be two sequences of real numbers.

Definition 1. *We say that a_n is of order b_n (or a_n is 'big oh' b_n), and write $a_n = O(b_n)$ as $n \to \infty$, if the ratio $|a_n/b_n|$ remains bounded for large n. That is, if there exists a number C and an integer $n(C)$ such that if n is larger than $n(C)$ then $|a_n| \leq C|b_n|$, or equivalently, if $\lim_{n\to\infty} |a_n/b_n| < \infty$.*

Definition 2. *We say that a_n is of small order b_n (or a_n is 'small oh' b_n), and write $a_n = o(b_n)$ as $n \to \infty$, if the ratio $|a_n/b_n|$ converges to zero. That is, if for any $\varepsilon > 0$, there exists an integer $n(\varepsilon)$ such that if n is larger than $n(\varepsilon)$ then $|a_n| \leq \varepsilon|b_n|$, or equivalently, if $\lim_{n\to\infty} |a_n/b_n| = 0$.*

In the context of our work, we will not write 'as $n \to \infty$' when referring to the order of a sequence, as that will always be the case of interest. There are several trivial properties that are consequence of this notation, such as $O(a_n)o(b_n) = o(a_n b_n)$ that will not be proven, but will be used in this work. The final definition concerning asymptotic notation is presented now. It will be useful to compare rates of convergence.

Definition 3. *We say that a_n is asymptotically equivalent to b_n, and write $a_n \sim b_n$ if $\lim_{n\to\infty} a_n/b_n = 1$.*

From these definitions, we may state the following version of Taylor's theorem, which we will not be proving.

Theorem 1 (Taylor's Theorem)**.** *Suppose that g is a real-valued function defined on $\mathbb{R}$ and let $x \in \mathbb{R}$. Assume that g has p continuous derivatives in an interval $(x-\delta, x+\delta)$ for some $\delta > 0$. Then for any sequence $(\alpha_n, n \in \mathbb{N})$ converging to zero,*

$$g(x+\alpha_n) = \sum_{j=0}^{p} \frac{\alpha_n^j}{j!} g^{(j)}(x) + o(\alpha_n^p).$$

With these definitions and this version of Taylor's theorem we are ready to study a few interesting statistical properties of the kernel density estimator. Let us consider the estimation of f at $x \in \mathbb{R}$. From (1.2) it is clear that

$$\mathrm{E}\left(\hat{f}(x;h)\right) = \int K_h(x-y) f(y)\, dy. \tag{1.3}$$

Now let us assume the following conditions. Most of them are common in kernel smoothing theory, and will also be hypothesis in this work.

(i) The density f is such that its second derivative f'' is continuous and square integrable.

(ii) The bandwidth $h = h_n$ is a non-random sequence of positive numbers that satisfies

$$\lim_{n\to\infty} h = 0 \quad \text{and} \quad \lim_{n\to\infty} nh = \infty,$$

which is equivalent to saying that h approaches zero, but at a rate slower than n^{-1}.

(iii) The kernel K is a bounded probability density function having finite fourth moment and symmetry about the origin.

Under these conditions, setting $S(K) = \int u^2 K(u) du$ and doing a change of variables in (1.3) and a Taylor expansion, it is easily shown that

$$\text{Bias}(\hat{f}(x;h)) = \text{E}(\hat{f}(x;h)) - f(x) = \frac{1}{2} h^2 S(K) f''(x) + o(h^2).$$

This means that the bias of $\hat{f}(x;h)$ is of order h^2, which implies that $\hat{f}(x;h)$ is asymptotically unbiased, making the kernel density estimator a decent alternative to the histogram, as a density function estimator. Another good reason to consider the kernel density estimator is the following. Recall the definition of Mean Integrated Square Error (MISE), which is the expected value of the Integrated Square Error (ISE), that serves as an error criterion that globally measures the distance between the functions f and its estimator. It is defined as

$$\text{MISE}(\hat{f}(\cdot;h)) = \text{E}[\text{ISE}(\hat{f}(\cdot;h))] = \text{E} \int [\hat{f}(x;h) - f(x)]^2 dx.$$

Let $R(g) = \int g^2(x) dx$. It can be shown that (Wand and Jones, 1995)

$$\inf_{h>0} \text{MISE}(\hat{f}(\cdot;h)) \sim \frac{5}{4} [S(K)^2 R(K)^4 R(f'')]^{1/5} n^{-4/5}.$$

If b satisfies assumption (ii), it can also be shown that (Scott, 1979)

$$\inf_{b>0} \text{MISE}(\hat{f}_H(\cdot;h)) \sim \frac{1}{4} [36 R(f')]^{1/3} n^{-2/3}.$$

What this means is that the MISE of the histogram is asymptotically inferior to the kernel density estimator's, in the sense that its convergence rate is $O(n^{-2/3})$ compared to the kernel estimator's $O(n^{-4/5})$ rate. This is a quantification of the inefficiency of the histogram that was mentioned before, and it is a convincing argument to consider estimation of density functions using the kernel density estimator. Later in this work, we will compare two estimators for ν_p, and we shall see how the kernel estimator helps in variance reduction.

1.2.3 The O_p, o_p notation. Convergence in distribution and in probability.

To generalize the O, o notation for non-random sequences, we now introduce the O_p, o_p notation. The idea is to maintain the concept behind each definition, and try to adapt it to stochastic processes.

Let $(A_n, n \in \mathbb{N})$ and $(B_n, n \in \mathbb{N})$ be two stochastic processes.

Definition 4. *We write $A_n = o_p(B_n)$ if for every $\varepsilon > 0$,*

$$\lim_{n\to\infty} P\left(\left| \frac{A_n}{B_n} \right| < \varepsilon \right) = 1,$$

or equivalently, if for every $\varepsilon > 0$ and every $\eta > 0$ there exists an integer $n(\varepsilon, \eta)$ such that if $n \geq n(\varepsilon, \eta)$, then

$$P\left(\left| \frac{A_n}{B_n} \right| < \varepsilon \right) \geq 1 - \eta.$$

Informally, this definition means that with arbitrarily high probability, $|A_n/B_n| = o(1)$. To preserve the idea behind this description, we wish to define $A_n = O_p(B_n)$ to mean that with high probability, $|A_n/B_n| = O(1)$. We accomplish this using the following definition.

Definition 5. *We write $A_n = O_p(B_n)$ if for every $\eta > 0$ there exist a constant $C(\eta)$ and an integer $n(\eta)$ such that if $n \geq n(\eta)$, then*

$$P\left(\left|\frac{A_n}{B_n}\right| < C(\eta)\right) \geq 1 - \eta.$$

Let us note the case when $B_n = 1$ a.s., for all $n \in \mathbb{N}$. There is not any generality loss because $A_n = O_p(B_n)$ if and only if $A_n/B_n = O_p(1)$, and the same argument is valid for o_p. In both $O_p(1)$ and $o_p(1)$, events are required to hold with a probability arbitrarily close to 1, replacing the certainty of the O, o definitions. In both $o(1)$ and $o_p(1)$, the sequences are required to be less than any arbitrarily small ε, for n suficiently large, whereas in both $O(1)$ and $O_p(1)$, the sequences are required to be bounded by some constant C, for n suficiently large. This is the reason why we refer to $A_n = O_p(1)$ by saying that A_n is bounded in probability, and to $A_n = o_p(1)$ by saying that A_n converges to zero in probability.

The obvious question now is how to identify the stochastic order of a process. Chebyshev's inequality provides a good starting point.

Theorem 2. *Let $(A_n, n \in \mathbb{N})$ be a stochastic process with $\mu_n = E(A_n)$ and $\sigma_n^2 = Var(A_n) < \infty$, for all $n \in \mathbb{N}$. Then*

$$A_n - \mu_n = O_p(\sigma_n).$$

Proof. From Chebyshev's inequality, for all $n \in \mathbb{N}$ and $h > 0$,

$$\mathrm{P}\left(\frac{|A_n - \mu_n|}{\sigma_n} < h\right) \geq 1 - h^{-2}.$$

Setting $h = \eta^{-1/2}$ for any $0 < \eta < 1$, and $C(\eta) = \eta^{-1/2}$, Definition 5 is fulfilled.

From this, it is also convenient to define two kinds of convergence of sequences of random variables that will be relevant in this work, and make the parallelism with the notation we have introduced.

Definition 6. *Let $(X_n, n \in \mathbb{N})$ be a stochastic process such that X_n has distribution function F_n and let X be a random variable with distribution function F. We say $(X_n, n \in \mathbb{N})$ converges in distribution to X, and write $X_n \xrightarrow{d} X$, if for all x which are continuity points of F,*

$$\lim_{n\to\infty} F_n(x) = F(x).$$

Convergence in distribution is perhaps the most frequently used in practice. It often arises from the application of a Central Limit Theorem, and it is provides a tool for the approximation of probabilities when F_n is not known, and the construction of confidence intervals.

Definition 7. *Let $(X_n, n \in \mathbb{N})$ be a stochastic process. We say that $(X_n, n \in \mathbb{N})$ converges in probability to a random variable X, and write $X_n \xrightarrow{P} X$, if for every $\varepsilon > 0$,*

$$\lim_{n\to\infty} P(|X_n - X| \leq \varepsilon) = 1.$$

If $X = c$ with probability 1 (i.e, X is a degenerate random variable), we simply say that X_n converges to c in probability.

Remark 1. *Saying that $(X_n, n \in \mathbb{N})$ converges in probability to 0 is the same as saying $X_n = o_p(1)$.*

To end this section, we present three famous theorems that will be useful in this work. The proofs of the last two will be ommited, as they are far more technical than the other. We start by presenting a sufficient condition for convergence in probability.

Theorem 3. *Let $(X_n, n \in \mathbb{N})$ be a stochastic process such that $E(X_n)$ converges to $c \in \mathbb{R}$, and $Var(X_n)$ converges to 0. Then $X_n \xrightarrow{P} c$.*

Proof. Fix $\varepsilon > 0$. For n suficiently large, $|\mathrm{E}(X_n) - c| \leq \varepsilon/2$, and consequently, by Chebyshev's inequality,

$$\mathrm{P}(|X_n - c| \geq \varepsilon) \leq \mathrm{P}\left(|X_n - \mathrm{E}(X_n)| \geq \frac{\varepsilon}{2}\right) \leq \frac{4\mathrm{Var}(X_n)}{\varepsilon^2} \xrightarrow{n\to\infty} 0,$$

which is a condition equivalent to the one in Definition 7. ■

The following two theorems will be of great importance when we need to prove the two Central Limit Theorems of this work.

Theorem 4 (Slutsky's Theorem)**.** *Let $(X_n, n \in \mathbb{N})$ and $(Y_n, n \in \mathbb{N})$ be two stochastic process, and X be a random variable. If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{P} c$, where c is some constant, then $X_n + Y_n \xrightarrow{d} X + c$.*

The original version of Slutsky's theorem is far more general than the one we presented. However, for the purpose of this work, this statement of Slutsky's theorem will suffice.

Theorem 5 (Lyapounov's Theorem)**.** *Let $(\xi_n, n \in \mathbb{N})$ be a stochastic process, and let k_n be a sequence of positive integers. For each n, let $\xi_1, ..., \xi_{k_n}$ be independent random variables with mean 0 and finite variance σ_i^2, for $i = 1, ..., k_n$. Let $S_n = \xi_1 + ... + \xi_{k_n}$ and suppose its variance $s_n^2 = \sigma_1^2 + ... + \sigma_{k_n}^2$ is positive. If, for some positive δ,*

$$\frac{1}{s_n^{2+\delta}} \sum_{k=1}^{k_n} E(|\xi_k|^{2+\delta}) \xrightarrow{n\to\infty} 0,$$

then

$$\frac{S_n}{s_n} \xrightarrow{d} \mathcal{N}(0,1).$$

1.2.4 Mixing

Let us recall the definitions of independence between two σ-algebras and between two random variables.

Definition 8. *Consider the probability space* $(\Omega, \mathscr{A}, P)$ *and let* $\mathscr{U}$ *and* $\mathscr{V}$ *be two* σ*-algebras of* $\mathscr{A}$*. We say that* $\mathscr{U}$ *and* $\mathscr{V}$ *are independent if*

$$\forall U \in \mathscr{U}, \forall V \in \mathscr{V} \qquad P(U \cap V) = P(U)P(V).$$

Definition 9. *We say that the random variables X and Y of the same probability space* $(\Omega, \mathscr{A}, P)$ *are independent if the* σ*-algebras* $\sigma(X)$ *and* $\sigma(Y)$ *are independent.*

To move away from the concept of independence, Rosenblatt (1956) introduced a fairly intuitive idea, which is now known as the α-mixing coefficient. It measures the dependence between two σ-algebras, and is defined as follows.

Definition 10. *Consider the probability space* $(\Omega, \mathscr{A}, P)$*. Given two* σ*-algebras* $\mathscr{U}$ *and* $\mathscr{V}$ *of* $\mathscr{A}$*, their* α*-mixing coefficient is defined by*

$$\alpha(\mathscr{U}, \mathscr{V}) = \sup_{\substack{U \in \mathscr{U} \\ V \in \mathscr{V}}} |P(U \cap V) - P(U)P(V)|.$$

The extension of this concept to random variables is quite straightforward.

Definition 11. *Let Y and Z be two random variables defined on* $(\Omega, \mathscr{A}, P)$*. The* α*-mixing coefficient of Y and Z, which we denote as* $\alpha(Y, Z)$ *is defined as*

$$\alpha(Y, Z) = \alpha(\sigma(Y), \sigma(Z)).$$

That is,

$$\alpha(Y, Z) = \sup_{\substack{A \in \sigma(Y) \\ B \in \sigma(Z)}} |P(A \cap B) - P(A)P(B)|.$$

Furthermore, if $X = (X_t, t \in \mathbb{Z})$ *is a stochastic process, the* α*-mixing coefficient of order* $k \geq 1$ *is defined as*

$$\alpha(k) = \sup_{t \in \mathbb{N}} \alpha(\sigma(X_s : s \leq t), \sigma(X_s : s \geq t + k)).$$

That is,

$$\alpha(k) = \sup_{t \in \mathbb{N}} \sup_{\substack{A \in \sigma(X_s : s \leq t) \\ B \in \sigma(X_s : s \geq t+k)}} |P(A \cap B) - P(A)P(B)|.$$

The same definition applies if the index set of the process is $\mathbb{N}$. It can be shown that if X is a stationary process, then $\alpha(k)$ does not depend on t. In this case, $\alpha(k)$ can simply be written as

$$\alpha(k) = \sup_{\substack{A \in \sigma(X_s : s \leq t) \\ B \in \sigma(X_s : s \geq t+k)}} |\mathrm{P}(A \cap B) - \mathrm{P}(A)\mathrm{P}(B)|.$$

We say that the process X is strong mixing (or α-mixing) if $\lim_{k\to\infty} \alpha(k) = 0$. This condition specifies a form of asymptotic independence of the past and future of X, that is, the statistical dependence between X_{t_1} and X_{t_2} is arbitrarily close to zero when $|t_1 - t_2|$ is suficiently large.

Since the introduction of the α-mixing, several other coefficients have been introduced. We refer to Doukhan (1994) for the interested reader. Here, we would only like to point that α-mixing is the weakest type of mixing, in the sense that it is implied by other types of mixing. Just as a curiosity, another measure that quantifies the extent to which given random variables are dependent was introduced by Kolmogorov and Rozanov in 1960. It is now known as the ρ-mixing coefficient, and its definition of order $k \geq 1$ for a stationary stochastic process $X = (X_t, t \in \mathbb{Z})$ is

$$\rho(k) = \sup_{\substack{Y \in L^2(\sigma(X_s : s \leq t)) \\ Z \in L^2(\sigma(X_s : s \geq t+k))}} |\text{Corr}(Y, Z)|.$$

In the same sense, we say that X is ρ-mixing if $\lim_{k\to\infty} \rho(k) = 0$, and it can be shown that if X is ρ-mixing then it is also α-mixing. This is one of the implications we mentioned in the above paragraph.

Remark 2. *If $(X_t, t \in \mathbb{Z})$ is a stationary α-mixing process, since $\sigma(X_s : s \geq t+k+1) \subseteq \sigma(X_s : s \geq t+k)$ it is clear from the construction of the α-mixing coeficient that $\alpha(k)$ is a non-increasing sequence.*

Remark 3. *If $X = (X_t, t \in \mathbb{Z})$ is a stationary α-mixing process and $Z = (Z_t, t \in \mathbb{Z})$ is defined as $Z_t = g(X_t), \forall t \in \mathbb{Z}$, with g a measurable transformation, then Z is also a stationary process and each α-mixing coefficient of Z is smaller or equal than the corresponding α-mixing coefficient of X.*

In this work, we will be assuming a condition stronger than α-mixing, to guarantee the summability of the α-mixing coefficients.

Definition 12. *A process is said to be geometric α-mixing if there exist constants $C > 0$ and $\rho \in (0,1)$ such that $\alpha(k) \leq C\rho^k$ for $k \geq 1$.*

The obvious question now, is if the condition of a stochastic process being geometric α-mixing is too restrictive to be applied in the financial fields. Fortunately, it is not. Many commonly used financial time series models have been proven to be geometric α-mixing, with some restrictions to their parameters. These include ARMA (Pham and Tran, 1985), ARCH (Masry and Tjøstheim, 1995), GARCH (Carrasco and Chen, 2002) and Diffusion and Stochastic Volatility models (Genon-Catalot, Jeantheau and Laredo, 2000).

On the subject of mixing, it is worth noting that under certain mixing conditions, one may still derive important Central Limit Theorems. A common technique used in these types of proofs is to consider different 'blocks' of random variables, which are no more than sums of those random variables, that satisfy certain imposed conditions. We will see more about this blocking technique later.

It is now convenient to introduce the theorems concerning α-mixing that we will be using. We start by presenting an inequality that is often used to derive limit theorems for strong mixing processes. The following can be found in Bosq (1998) and Yokoyama (1980).

Theorem 6 (Bosq's Theorem)**.** *Let $X = (X_t, t \in \mathbb{Z})$ be a zero-mean process such that $\sup_{1\le t\le n} ||X_t||_\infty \le b$, and $S_n = \sum_{t=1}^{n} X_t$. Then, for each integer $q \in \left[1, \frac{n}{2}\right]$ and each $\varepsilon > 0$,*

$$P\left(|S_n| > n\varepsilon\right) \le 4\exp\left(-\frac{\varepsilon^2}{8\tau^2(q)}q\right) + 22\left(1+\frac{4b}{\varepsilon}\right)^{1/2} q\alpha\left(\left[\frac{n}{2q}\right]\right),$$

where

$$\tau^2(q) = 2m^{-2}\sigma^2(q) + \frac{b\varepsilon}{2},$$

with $m = \frac{n}{2q}$, $\sigma^2(q) = \max_{0\le j\le 2q-1} E\Big(([jm]+1-jm)X_{[jm]+1} + X_{[jm]+2} + \ldots + X_{[(j+1)m]} + ((j+1)m - [(j+1)m])X_{[(j+1)m+1]}\Big)^2$, and $\alpha\left(\left[\frac{n}{2q}\right]\right)$ is the α-mixing coefficient of order $\left[\frac{n}{2q}\right]$ of X.

Remark 4. *In the context of this work, we will be applying Bosq's theorem with two small differences. The first difference concerns the upper bound given in Bosq's theorem for $P(|S_n| > n\varepsilon)$. Using the same arguments that are present in this theorem's proof, we can deduce that this bound is also valid for $P(|S_n| \ge n\varepsilon)$. The second difference concerns an argument in the proof itself. Bosq considers an auxiliary continuous time process to complete his proof. As a consequence, he has to define $\sigma^2(q)$ in the form we see above. A simpler quantity can be deduced if we don't consider such an auxiliary process, but the original discrete time process itself. In this case, since n is a multiple of m (which is the size of each block used in the proof), the coefficient of $X_{[jm]+1}$ is 1, and the coefficient of $X_{[(j+1)m+1]}$ is 0, which allows us to get a simpler definition of $\sigma^2(q)$, that equates to*

$$\sigma^2(q) = \max_{0\le j\le 2q-1} E\left(\sum_{l=[jm]+1}^{[(j+1)m]} X_l\right)^2.$$

The next theorem is perhaps the most interesting one. It provides an upper bound for the covariance between two random variables of an α-mixing process as a function of how far apart their indexes are.

Theorem 7 (Davydov's Inequality)**.** *Let X and Y be two random variables such that $X \in L^q(P)$, $Y \in L^r(P)$ where $q > 1$, $r > 1$ and $q^{-1} + r^{-1} + s^{-1} = 1$. Then*

$$|Cov(X,Y)| \le 12||X||_q||Y||_r[\alpha(\sigma(X),\sigma(Y))]^{1/s}.$$

For instance, if $(X_t, t \in \mathbb{N})$ is a stationary α-mixing process with $X_1 \in L^q(\mathrm{P})$ for some $q > 1$, the following inequality holds for appropriate choice of $q = r$:

$$|\mathrm{Cov}(X_1, X_{t+1})| \le 12||X_1||_q^2\alpha^{1/s}(t).$$

Clearly, as $t \to \infty$, $\mathrm{Cov}(X_1, X_{t+1})$ approaches zero. This is not a surprise, given the formulation of the α-mixing coefficient.

Remark 5. *Let $\gamma(k) = Cov\left[(Y_1 - \nu_p)I(Y_1 \geq \nu_p), (Y_{k+1} - \nu_p)I(Y_{k+1} \geq \nu_p)\right]$, for positive integers k. A relevant quantity in this work that will show up several times is defined as*

$$\sigma_0^2(p;n) := Var[(Y_1 - \nu_p)I(Y_1 \geq \nu_p)] + 2\sum_{k=1}^{n-1}\left(1 - \frac{k}{n}\right)\gamma(k).$$

We can prove that $\lim_{n\to\infty}\sigma_0^2(p;n)$ is finite using Davydov's inequality. For that purpose, assume that $Y = (Y_t, t \in \mathbb{N})$ is a stationary geometric α-mixing process with $Y_1 \in L^{4+\delta}(P)$ for some $\delta > 0$. As we will see, these will be underlying assumptions in our work. Taking into consideration Remark 3, and letting $q = 2+\delta'$, where $\delta' \in (0,\delta)$. Davydov's inequality gives us that

$$\lim_{n\to\infty}\sum_{k=1}^{n-1}\left|\left(1 - \frac{k}{n}\right)\gamma(k)\right| \leq \lim_{n\to\infty}\sum_{k=1}^{n-1}|\gamma(k)| \leq 12||(Y_1 - \nu_p)I(Y_1 \geq \nu_p)||_q^2 \lim_{n\to\infty}\sum_{k=1}^{n-1}\alpha^{\frac{\delta'}{2+\delta'}}(k) < \infty.$$

As of right now it may not be clear why $||(Y_1 - \nu_p)I(Y_1 \geq \nu_p)||_q$ is finite, but soon it will.

We now present Bradley's Lemma, a result that will be crucial to prove both Central Limit Theorems that we will study.

Theorem 8 (Bradley's Lemma)**.** *Let (X,Y) be a $\mathbb{R}^d \times \mathbb{R}$-valued random vector such that $Y \in L^p(P)$ for some $p \in [1,+\infty]$. Let c be a real number such that $||Y+c||_p > 0$, and $\zeta \in (0, ||Y+c||_p]$. Then, there exists a random variable Y^* such that*

(i) $P_{Y^*} = P_Y$ and Y^* is independent of X,

(ii) $P(|Y^* - Y| > \zeta) \leq 18\left(\zeta^{-1}||Y+c||_p\right)^{p/(2p+1)}[\alpha(\sigma(X),\sigma(Y))]^{2p/(2p+1)}$.

In a certain sense, this lemma tells us about the price to pay if we wish to replace dependent random variables by independent ones, with the same distribution. To end the note on theorems concerning α-mixing, we present a powerful inequality, due to Yokoyama. The following is achieved in a broader context than just the hypothesis of a stochastic process being α-mixing, but since this is the case that concerns us, it is the one we present.

Theorem 9 (Yokoyama's Inequality)**.** *Let $(X_t, t \in \mathbb{N})$ be a stationary strong mixing process, with $E(X_1) = 0$, $E\left(|X_1|^{r+\delta}\right) < \infty$ for some $r > 2$ and $\delta > 0$, and $S_n = \sum_{t=a+1}^{a+n} X_t$, for $a \geq 0$. If*

$$\sum_{i=1}^{\infty}(i+1)^{r/2-1}[\alpha(i)]^{\delta/(r+\delta)} < \infty,$$

then there exists a constant C such that

$$E\left(|S_n|^r\right) \leq Cn^{r/2}, \quad n \geq 1, \quad a \geq 0. \tag{1.4}$$

Also, under the condition that $(X_t, t \in \mathbb{N})$ is a stationary strong mixing process, with $E(X_1) = 0$ and $|X_1| \leq C < \infty$ a.s., and letting $r > 0$, if

$$\sum_{i=1}^{\infty} (i+1)^{r/2-1} \alpha(i) < \infty,$$

then (1.4) still holds.

Yokoyama's inequality provides us a way to bound expected values of powers of sums. As we will see, that will save us some work when we need to verify the hypothesis of Lyapounov's theorem.

1.3 Some well known auxiliary results

In this brief chapter, we present some results that will be useful in our work. They will mostly be relevant to provide upper bounds to quantities that will come up. We start with Hoeffding's covariance identity.

Theorem 10 (Hoeffding's Identity). *Let X and Y be two random variables with finite second moments, $F_X(x)$ and $F_Y(y)$ be the marginal distribution function of X and Y respectively, and $F_{X,Y}(x,y)$ be the joint distribution function of the pair (X,Y). In this case,*

$$Cov(X,Y) = \int \int F_{X,Y}(x,y) - F_X(x) F_Y(y) dx dy. \tag{1.5}$$

Remark 6. *Note that $F_{X,Y}(x,y) - F_X(x)F_Y(y) = P(X > x, Y > y) - P(X > x)P(Y > y)$, for all $(x,y) \in \mathbb{R}^2$. This means that (1.5) can be rewritten as*

$$Cov(X,Y) = \int \int P(X > x, Y > y) - P(X > x)P(Y > y) dx dy.$$

Next we present a special case of Hölder's inequality. The more general inequality will not be presented as it is not needed in this work.

Theorem 11 (Hölder's Inequality). *Let X and Y be two random variables with finite p-th and q-th moments respectively, with $p, q \in [1, +\infty]$ such that $p^{-1} + q^{-1} = 1$. In this case, the following inequality holds*

$$E\big(|XY|\big) \leq E^{\frac{1}{p}}\big(|X|^p\big) E^{\frac{1}{q}}\big(|Y|^q\big).$$

The case $p = q = 2$ is the famous Cauchy-Schwarz inequality.

Remark 7. *In particular, if X has finite p-th moment, for $p \geq 1$, then $E\big(|X|\big) \leq E^{\frac{1}{p}}\big(|X|^p\big)$.*

Finally, we present the c_r inequality.

Theorem 12 (c_r inequality). *Let X and Y be two real random variables and let r be a positive number such that X and Y have finite r-th moments. In this case, the following inequality holds*

$$E\big(|X+Y|^r\big) \leq c_r \Big[E\big(|X|^r\big) + E\big(|Y|^r\big)\Big],$$

where

$$c_r = \begin{cases} 1, & \text{if } 0 < r \leq 1 \\ 2^{r-1}, & \text{if } r > 1 \end{cases}.$$

1.4 Nonparametric estimators and assumptions

In order to present the two nonparametric estimators of the ES we will be studying in this work, we must first introduce two VaR estimators. The first one, which we denote as $\hat{v}_p$, is simply an order statistics. It is defined as

$$\hat{v}_p = Y_{([n(1-p)]+1)},$$

where $Y_{(r)}$ is the r-th order statistic of $(Y_t, t = 1, ..., n)$.

To introduce the second estimator, we must first extend the concept of kernel density estimator. For instance, in the context of Section 1.2.1, if we are interested in estimating the distribution function of Y, which we denote by F, a natural estimator involving the kernel density estimator would be obtained by integrating both sides of (1.1). That would lead us to

$$\hat{F}(x;h) = \frac{1}{n}\sum_{t=1}^{n}\int_{-\infty}^{\frac{x-Y_t}{h}} K(u)du.$$

In a more general context, we say that this is a kernel estimator because it involves the usage of a kernel. Another way to get to this estimator would be to consider the empirical distribution function, which is defined as

$$\hat{F}_n(x) = \frac{1}{n}\sum_{t=1}^{n} I(Y_t \leq x),$$

and replace the indicator function I by the smoother H, which we define as $H(x) = \int_{-\infty}^{x} K(u)du$, where K is a kernel. That is,

$$\hat{F}(x;h) = \frac{1}{n}\sum_{t=1}^{n} H\left(\frac{x-Y_t}{h}\right),$$

where h is a smoothing parameter that controls the amount of smoothing in the estimation of F.

Let $G(x) = \int_{x}^{+\infty} K(u)du$ and $G_h(x) = G(x/h)$. In this work we will be interested in estimating the survival function $S(x) := 1 - F(x)$. For that purpose we consider the following estimator for $S(x)$, that can be derived in the same manner that $\hat{F}(x;h)$ was, and is defined as

$$\hat{S}(x;h) = \frac{1}{n}\sum_{t=1}^{n} G_h(x - Y_t).$$

The second estimator of ν_p, which we denote as $\hat{\nu}_{p,h}$, is a kernel estimator, introduced by Gourieroux, Laurent and Scaillet (2000), and can be viewed as a smoothed version of $\hat{\nu}_p$. It is the solution of $\hat{S}(z;h) = p$, that is,

$$\frac{1}{n}\sum_{t=1}^{n} G_h\left(\hat{\nu}_{p,h} - Y_t\right) = p.$$

We are now ready to present the two nonparametric estimators of the ES. The first one, which we denote as $\hat{\mu}_p$, is an average of the losses larger than $\hat{\nu}_p$. It is defined as

$$\hat{\mu}_p = \frac{\sum_{t=1}^{n} Y_t I(Y_t \geq \hat{\nu}_p)}{\sum_{t=1}^{n} I(Y_t \geq \hat{\nu}_p)}. \tag{1.6}$$

The second estimator, which we denote as $\hat{\mu}_{p,h}$, is a kernel estimator proposed by Scaillet (2004). It is obtained by replacing the indicator function I and $\hat{\nu}_p$ by the smoother G and $\hat{\nu}_{p,h}$ respectively in (1.6). It is defined as

$$\hat{\mu}_{p,h} = \frac{1}{np}\sum_{t=1}^{n} Y_t G_h\left(\hat{\nu}_{p,h} - Y_t\right).$$

Now that the ES estimators have been introduced, we will present the assumptions for this work. They are:

(i) $Y = (Y_t, t \in \mathbb{N})$ is a stationary α-mixing process, and there exist $C > 0$ and $\rho \in (0,1)$ such that $\alpha(k) \leq C\rho^k$ for all $k \geq 1$; Y_t is absolutely continuous, with f and F as its density and distribution functions, respectively.

(ii) f has continuous second derivatives in $\mathscr{B}_{\nu_p}$, a neighbourhood of ν_p, and satisfies $f(\nu_p) > 0$; The joint distribution function of (Y_1, Y_{k+1}), $k \geq 1$, which we denote by F_k, has all its second partial derivatives bounded in $\mathscr{B}_{\nu_p}$ by a constant B, say; There exist $C, \delta > 0$ such that $\mathrm{E}(|Y_t|^{4+\delta}) \leq C$.

(iii) K is an univariate symmetric and bounded probability density function satisfying the moment conditions $\int u^2 K(u)du = \sigma_K^2$ and $\int u^4 K(u)du = \sigma_K^4$, and K has bounded and Lipschitz continuous derivative. K also satisfies $\int K^{2+\delta'}(u)du < C$ and $\int uK^{2+\delta'}(u)du < C$, for some $C > 0$, $\delta' \in (0, \delta)$.

(iv) The smoothing bandwidth h satisfies $h \to 0$, $nh^{3.5-\beta} \to \infty$ for any $\beta \geq 0$ and $nh^4 \log^2(n) \to 0$ as $n \to \infty$.

Condition (i) means that the time series of study, $(Y_t, t \in \mathbb{N})$, is geometric α-mixing, and as we have seen in Section 1.2.4, that is not too restrictive of an assumption. Condition (ii) contains standard assumptions for quantile estimation, which require underlying smoothness for f and F_k, as well as finite moments for the absolute returns. Conditions (iii) and (iv) are commonly imposed conditions in kernel smoothing, with some particular conditions for this work. For instance, condition (iv) specifies a range for the bandwidth h.

Under slightly different assumptions, Chen and Tang (2005) showed that

$$\mathrm{Var}\left(\hat{\nu}_{p,h}\right) = n^{-1} f^{-2}(\nu_p)\Delta^2(p;n) - 2n^{-1} h f^{-1}(\nu_p) c_K + o(n^{-1}h), \tag{1.7}$$

with $\Delta^2(p;n) = \left[p(1-p) + 2\sum_{k=1}^{n-1}\left(1-\frac{k}{n}\right)\mathrm{Cov}\left[I(Y_1 \geq \nu_p), I(Y_{k+1} \geq \nu_p)\right] \right]$ and $c_K = \int uK(u)G(u)du$. Using a similar argument to the one in Remark 5, it is easy to prove that $\lim_{n\to\infty} \Delta^2(p;n)$ is finite.

Yoshihara (1995) had already established that, under the α-mixing assumption,

$$\mathrm{Var}(\hat{\nu}_p) = n^{-1} f^{-2}(\nu_p)\Delta^2(p;n)\left[1+o(1)\right]. \tag{1.8}$$

Equations (1.7) and (1.8) indicate that both VaR estimators have the same leading asymptotic variance term. However, the VaR kernel estimator $\hat{\nu}_{p,h}$ reduces the variance in the second order of $n^{-1}h$ as $c_K > 0$. This second-order reduction can be of significance, not only because the available data in the tail is usually low, but also because even a low reduction in variance can translate to a large reduction in absolute financial losses.

Based on this particular improvement that was achieved with kernel smoothing, it is reasonable to expect that the kernel ES estimator, $\hat{\mu}_{p,h}$, will also bring some sort of improvement over the unsmoothed ES estimator, $\hat{\mu}_p$. Confirming this or otherwise is one of the objectives of this work, and to do so, we must first prove some results, which we state in form of four lemmas, that will facilitate the proof of the two Central Limit Theorems.

Chapter 2

Main Results

In this chapter we present the results we will derive and their proofs. For that purpose, we will let C denote generic positive constants. We start with Lemma 1, where we will prove that a certain probability converges to zero faster than any polynomial. By this we mean that said probability is $O\left(n^{-C}\right)$, for any real number C. This result will be helpful to prove Lemma 2.

2.1 Auxiliary Lemmas

Lemma 1 (Chen, 2008). *Let $\varepsilon_n = n^{-1/2}\log n$. Under condition* (i), *$P(|\hat{\nu}_p - \nu_p| \geq \varepsilon_n)$ converges to zero faster than any polynomial.*

Proof. Fix $n \in \mathbb{N}$. Taking into consideration that $\hat{\nu}_p$ is an absolutely continuous random variable, it is clear that $\mathrm{P}(|\hat{\nu}_p - \nu_p| \geq \varepsilon_n) = \mathrm{P}(\hat{\nu}_p > \nu_p + \varepsilon_n) + \mathrm{P}(\hat{\nu}_p < \nu_p - \varepsilon_n)$. Define $C_1(n) = \inf_{x\in[\nu_p-\varepsilon_n,\nu_p+\varepsilon_n]} f(x)$. We will now find upper bounds for the two probabilities.

Assume that $\hat{\nu}_p > \nu_p + \varepsilon_n$. Given the definitions of the empirical distribution function and $\hat{\nu}_p$, it is clear that for any value smaller than $\hat{\nu}_p$, such as $\nu_p + \varepsilon_n$, $\hat{F}_n$ satisfies $\hat{F}_n(\hat{\nu}_p) > \hat{F}_n(\nu_p + \varepsilon_n)$.

Note that

$$\hat{F}_n(\nu_p + \varepsilon_n) \leq \frac{1}{n}\sum_{t=1}^{n} I(Y_t \leq \nu_p + \varepsilon_n) \leq \frac{[n(1-p)]}{n} \leq 1-p, \tag{2.1}$$

and

$$F(\nu_p + \varepsilon_n) = \int_{-\infty}^{\nu_p+\varepsilon_n} f(x)dx = 1-p+\int_{\nu_p}^{\nu_p+\varepsilon_n} f(x)dx \geq 1-p+C_1(n)\varepsilon_n. \tag{2.2}$$

(2.1) and (2.2) allow us to conclude that $|\hat{F}_n(\nu_p+\varepsilon_n) - F(\nu_p+\varepsilon_n)| \geq C_1(n)\varepsilon_n$.

Now consider the case $\hat{\nu}_p < \nu_p - \varepsilon_n$. As $\hat{F}_n$ is a nonincreasing function, $\hat{F}_n(\hat{\nu}_p) \leq \hat{F}_n(\nu_p - \varepsilon_n)$, and since

$$\hat{F}_n(\hat{\nu}_p) = \frac{[n(1-p)]+1}{n} \geq 1-p,$$

we get that

$$\hat{F}_n(\nu_p - \varepsilon_n) \geq 1-p \tag{2.3}$$

Also note that

$$F(\nu_p-\varepsilon_n)=\int_{-\infty}^{\nu_p-\varepsilon_n} f(x)dx = 1-p-\int_{\nu_p-\varepsilon_n}^{\nu_p} f(x)dx \leq 1-p-C_1(n)\varepsilon_n. \tag{2.4}$$

Combine (2.3) and (2.4) to conclude that $|\hat{F}_n(\nu_p-\varepsilon_n)-F(\nu_p-\varepsilon_n)|\geq C_1(n)\varepsilon_n$.

Hence

$$\begin{aligned} &\mathrm{P}(\hat{\nu}_p>\nu_p+\varepsilon_n)+\mathrm{P}(\hat{\nu}_p<\nu_p-\varepsilon_n)\\ &\leq \mathrm{P}(|\hat{F}_n(\nu_p+\varepsilon_n)-F(\nu_p+\varepsilon_n)|\geq C_1(n)\varepsilon_n)+\mathrm{P}(|\hat{F}_n(\nu_p-\varepsilon_n)-F(\nu_p-\varepsilon_n)|\geq C_1(n)\varepsilon_n). \end{aligned} \tag{2.5}$$

In order to determine an upper bound for the two probabilities in (2.5), consider the auxiliary stochastic process $Z=(Z_t, t\in\mathbb{N})$ defined as $Z_t=I(Y_t\leq\nu_p+\varepsilon_n)-F(\nu_p+\varepsilon_n)$, for $t\in\mathbb{N}$. Clearly, for any t, $\mathrm{E}(Z_t)=0$ and $|Z_t|\leq 2$ a.s.. Also note that, from Remark 3, Z is a stationary process with smaller or equal α-mixing coefficients than those of Y. Note that

$$\begin{aligned} \mathrm{P}(|\hat{F}_n(\nu_p+\varepsilon_n)-F(\nu_p+\varepsilon_n)|\geq C_1(n)\varepsilon_n) &= \mathrm{P}\left(\left|n^{-1}\sum_{t=1}^{n}\left[I\left(Y_t\leq\nu_p+\varepsilon_n\right)-F\left(\nu_p+\varepsilon_n\right)\right]\right|\geq C_1(n)\varepsilon_n\right)\\ &=\mathrm{P}(|S_n|\geq nC_1(n)\varepsilon_n), \end{aligned}$$

where $S_n=\sum_{t=1}^{n} Z_t$. Set $q=b_0 n\varepsilon_n$, where b_0 is such that q is a positive integer smaller or equal to $n/2$ for each n, and $m=n/2q$. We seek to apply Bosq's Theorem with the alteration we mentioned in Remark 4. Remember that, in this case,

$$\sigma^2(q)=\max_{0\leq j\leq 2q-1}\mathrm{E}\left(\sum_{l=[jm]+1}^{[(j+1)m]} Z_l\right)^2.$$

Since $E(Z_t)=0$ and $|Z_t|\leq 2$ a.s., for each $j\in\{0,...,2q-1\}$, Yokoyama's Inequality gives us the following upper bound

$$\mathrm{E}\left|\sum_{l=[jm]+1}^{[(j+1)m]} Z_l\right|^r \leq C([(j+1)m]-[jm])^{r/2}\leq C(m+1)^{r/2}\leq Cm^{r/2},$$

provided that $\sum_{k=1}^{\infty}(k+1)^{r/2-1}\alpha(k)$ is finite, which under assumption (i) is trivially true for $r=2$, the case we are interested in. Hence,

$$\sigma^2(q)=\max_{0\leq j\leq 2q-1}\mathrm{E}\left(\sum_{l=[jm]+1}^{[(j+1)m]} Z_l\right)^2=\max_{0\leq j\leq 2q-1}\mathrm{E}\left|\sum_{l=[jm]+1}^{[(j+1)m]} Z_l\right|^2\leq Cm.$$

We now apply the modified version of Bosq's theorem, to conclude that

$$\mathrm{P}(|S_n|\geq nC_1(n)\varepsilon_n)\leq 4\exp\left(-\frac{(C_1(n))^2\,\varepsilon_n^2}{8\tau^2(q)}q\right)+22\left(1+\frac{8}{C_1(n)\varepsilon_n}\right)^{1/2} q\alpha\left(\left[\frac{n}{2q}\right]\right),$$

where $\alpha\left(\left[\frac{n}{2q}\right]\right)$ is the α-mixing coefficient of order $\left[\frac{n}{2q}\right]$ of the process Z and

$$\tau^2(q) = 2m^{-2}\sigma^2(q) + C_1(n)\varepsilon_n.$$

Since $\sigma^2(q) \leq Cm$, we get that $\tau^2(q) \leq 4b_0C\varepsilon_n + C_1(n)\varepsilon_n = C_2(n)\varepsilon_n$, which allows us to conclude that

$$4\exp\left(-\frac{(C_1(n))^2\varepsilon_n^2}{8\tau^2(q)}q\right) \leq 4\exp\left(-C_3(n)\varepsilon_n q\right),$$

which converges to zero faster than any polynomial, since $C_3(n)\varepsilon_n q$ tends to infinity.

On the other hand, for n suficiently large,

$$22\left(1+\frac{8}{C_1(n)\varepsilon_n}\right)^{1/2} q\alpha\left(\left[\frac{n}{2q}\right]\right) \leq 22\left(\frac{16}{C_1(n)\varepsilon_n}\right)^{1/2} q\alpha\left(\left[\frac{n}{2q}\right]\right) = C_4(n)n^{\frac{3}{4}}(\log n)^{\frac{1}{2}}\rho^{\left[\frac{Cn^{1/2}}{\log n}\right]},$$

which converges to zero faster than any polynomial too. Hence $\mathrm{P}(|\hat{F}_n(\nu_p+\varepsilon_n) - F(\nu_p+\varepsilon_n)| \geq C_1(n)\varepsilon_n)$ converges to zero faster than any polynomial. Analogous arguments can be used to prove that $\mathrm{P}(|\hat{F}_n(\nu_p-\varepsilon_n) - F(\nu_p-\varepsilon_n)| \geq C_1(n)\varepsilon_n)$ also converges to zero faster than any polynomial, which concludes the proof. ■

In Lemma 2 we will prove that a certain sum of random variables is $o_p\left(n^{-3/4+\kappa}\right)$, for $\kappa > 0$ arbitrarily small. This result will help us get a simple, yet practical expansion of $\hat{\mu}_p$ as a function of μ_p.

Lemma 2 (Chen, 2008)**.** *Under the conditions* (i) – (ii) *and for* $\kappa > 0$ *arbitrarily small,*

$$\frac{1}{n}\sum_{t=1}^{n}(Y_t-\nu_p)\left[I(Y_t \geq \hat{\nu}_p) - I(Y_t \geq \nu_p)\right] = o_p\left(n^{-3/4+\kappa}\right).$$

Proof. We seek to apply theorem 2. Let $W_t = (Y_t-\nu_p)[I(Y_t \geq \hat{\nu}_p) - I(Y_t \geq \nu_p)]$ for $t = 1,...,n$. Note that, since Y is stationary, $\mathrm{E}(W_t)$ and $\mathrm{Var}(W_t)$ do not depend on t. We start by computing $\mathrm{E}(W_1)$ by noting that

$$W_1 = \begin{cases} Y_1-\nu_p, & \text{if } \hat{\nu}_p \leq Y_1 < \nu_p \\ \nu_p - Y_1, & \text{if } \nu_p \leq Y_1 < \hat{\nu}_p\ , \\ 0, & \text{otherwise} \end{cases} \qquad \text{a.s..}$$

Defining

$$I_1 = (Y_1-\nu_p)I(\nu_p \leq Y_1 < \hat{\nu}_p)I(\hat{\nu}_p > \nu_p),$$
$$I_2 = (Y_1-\nu_p)I(\hat{\nu}_p \leq Y_1 < \nu_p)I(\hat{\nu}_p < \nu_p),$$

it is clear that $W_1 = -I_1 + I_2$ a.s., hence $\mathrm{E}(W_1) = -\mathrm{E}(I_1) + \mathrm{E}(I_2)$. Moreover, for $a \in (0, 1/2)$ and $\eta > 0$, if we define

$$\begin{aligned} I_{11} &= (Y_1 - \nu_p) I(\nu_p \le Y_1 < \hat{\nu}_p) I(\hat{\nu}_p \ge \nu_p + n^{-a}\eta), \\ I_{12} &= (Y_1 - \nu_p) I(\nu_p \le Y_1 < \hat{\nu}_p) I(\nu_p < \hat{\nu}_p < \nu_p + n^{-a}\eta), \\ I_{21} &= (Y_1 - \nu_p) I(\nu_p > Y_1 \ge \hat{\nu}_p) I(\hat{\nu}_p \le \nu_p - n^{-a}\eta), \\ I_{22} &= (Y_1 - \nu_p) I(\nu_p > Y_1 \ge \hat{\nu}_p) I(\nu_p > \hat{\nu}_p > \nu_p - n^{-a}\eta), \end{aligned}$$

it is also clear that $\mathrm{E}(I_1) = \mathrm{E}(I_{11}) + \mathrm{E}(I_{12})$ and $\mathrm{E}(I_2) = \mathrm{E}(I_{21}) + \mathrm{E}(I_{22})$. Applying the Cauchy-Schwarz inequality, we get

$$|\mathrm{E}(I_{11})|^2 \le \mathrm{E}\left[(Y_1 - \nu_p)^2 I(\nu_p \le Y_1 < \hat{\nu}_p)\right] \mathrm{E}\left[I(\hat{\nu}_p \ge \nu_p + n^{-a}\eta)\right].$$

Since $(Y_1 - \nu_p)^2 I(\nu_p \le Y_1 < \hat{\nu}_p) \le (\hat{\nu}_p - \nu_p)^2$ a.s.,

$$|\mathrm{E}(I_{11})|^2 \le \mathrm{E}(\hat{\nu}_p - \nu_p)^2 \mathrm{P}(\hat{\nu}_p \ge \nu_p + n^{-a}\eta) \le \mathrm{E}(\hat{\nu}_p - \nu_p)^2 \mathrm{P}(|\hat{\nu}_p - \nu_p| \ge n^{-a}\eta),$$

which implies that

$$|\mathrm{E}(I_{11})| \le [\mathrm{E}(\hat{\nu}_p - \nu_p)^2 \mathrm{P}(|\hat{\nu}_p - \nu_p| \ge n^{-a}\eta)]^{1/2}. \tag{2.6}$$

Since $n^{-1/2}\log n = o(n^{-a}\eta)$, Lemma 1 guarantees that

$$\mathrm{P}(|\hat{\nu}_p - \nu_p| \ge n^{-a}\eta) \to 0 \qquad \text{faster than any polynomial.} \tag{2.7}$$

Moreover, Yoshihara (1995) showed that

$$\mathrm{Var}(\hat{\nu}_p) = O\left(n^{-1}\right) \qquad \text{and} \qquad \mathrm{E}(\hat{\nu}_p) - \nu_p = O\left(n^{-3/4}\right),$$

which means that

$$\mathrm{MSE}(\hat{\nu}_p) = \mathrm{E}(\hat{\nu}_p - \nu_p)^2 = \mathrm{Var}(\hat{\nu}_p) + (\mathrm{E}(\hat{\nu}_p) - \nu_p)^2 = O\left(n^{-1}\right). \tag{2.8}$$

From (2.6), (2.7) and (2.8), it is clear that $|\mathrm{E}(I_{11})|$ converges to zero faster than any polynomial. Using the same arguments and the inequality $(Y_1 - \nu_p)^2 I(\nu_p > Y_1 \ge \hat{\nu}_p) \le (\hat{\nu}_p - \nu_p)^2$ a.s., it can be shown that $|\mathrm{E}(I_{21})|$ converges to zero faster than any polynomial.

In order to evaluate $\mathrm{E}(I_{12})$, we note that $0 \le I_{12} \le (Y_1 - \nu_p) I(\nu_p \le Y_1 < \nu_p + n^{-a}\eta)$, a.s.. Therefore

$$\begin{aligned} \mathrm{E}(I_{12}) &\le \int_{\nu_p}^{\nu_p + n^{-a}\eta} (z - \nu_p) f(z) dz \le \int_{\nu_p}^{\nu_p + n^{-a}\eta} (z - \nu_p) \max_{x \in [\nu_p, \nu_p + \eta]} f(x) dz \\ &= C \int_{\nu_p}^{\nu_p + n^{-a}\eta} (z - \nu_p) dz = \frac{C\eta^2}{2} n^{-2a} = O\left(n^{-2a}\right). \end{aligned}$$

Finally, to evaluate $\mathrm{E}(I_{22})$, we note that $0 \geq I_{22} \geq (Y_1 - \nu_p) I(\nu_p > Y_1 > \nu_p - n^{-a}\eta)$ a.s.. Hence

$$\begin{aligned}
|\mathrm{E}(I_{22})| &\leq \left| \int_{\nu_p - n^{-a}\eta}^{\nu_p} (z - \nu_p) f(z) dz \right| \leq \left| \int_{\nu_p - n^{-a}\eta}^{\nu_p} (z - \nu_p) \max_{x \in [\nu_p - \eta, \nu_p]} f(x) dz \right| \\
&= C \left| \int_{\nu_p - n^{-a}\eta}^{\nu_p} (z - \nu_p) dz \right| = \frac{C\eta^2}{2} n^{-2a} = O\left(n^{-2a}\right).
\end{aligned}$$

From all this, we conclude that $\mathrm{E}(W_1) = O\left(n^{-2a}\right)$. Choosing $a = 1/2 - \gamma$, where $\gamma > 0$ is arbitrarily small, we conclude that $\mathrm{E}(W_1) = O\left(n^{-1+\kappa'}\right)$, for an arbitrarily small $\kappa' > 0$. This and the fact that $n^{-1+\kappa'} \to 0$ imply that $\mathrm{E}(W_1) = o\left(n^{-1+\kappa}\right)$, for an arbitrarily small $\kappa > 0$, which in turn implies that

$$\mathrm{E}\left[\frac{1}{n}\sum_{t=1}^{n} (Y_t - \nu_p)\left[I(Y_t \geq \hat{\nu}_p) - I(Y_t \geq \nu_p)\right]\right] = o\left(n^{-1+\kappa}\right).$$

Let us now evaluate $\mathrm{Var}(W_1)$. For $a \in (0, 1/2)$ and $\eta > 0$,

$$\begin{aligned}
\mathrm{E}(W_1^2) &= \mathrm{E}\left[(Y_1 - \nu_p)^2 \{I(Y_1 \geq \hat{\nu}_p) - 2I(Y_1 \geq \hat{\nu}_p) I(Y_1 \geq \nu_p) + I(Y_1 \geq \nu_p)\}\right] \\
&= \mathrm{E}\left[(Y_1 - \nu_p)^2 \{I(\nu_p > Y_1 \geq \hat{\nu}_p) + I(\hat{\nu}_p > Y_1 \geq \nu_p)\}\right] \\
&= \mathrm{E}\left[(Y_1 - \nu_p)^2 I(\hat{\nu}_p \leq Y_1 < \nu_p) I(\hat{\nu}_p \geq \nu_p - n^{-a}\eta)\right] \\
&\quad + \mathrm{E}\left[(Y_1 - \nu_p)^2 I(\hat{\nu}_p \leq Y_1 < \nu_p) I(\hat{\nu}_p < \nu_p - n^{-a}\eta)\right] \\
&\quad + \mathrm{E}\left[(Y_1 - \nu_p)^2 I(\hat{\nu}_p > Y_1 \geq \nu_p) I(\hat{\nu}_p \geq \nu_p + n^{-a}\eta)\right] \\
&\quad + \mathrm{E}\left[(Y_1 - \nu_p)^2 I(\hat{\nu}_p > Y_1 \geq \nu_p) I(\hat{\nu}_p < \nu_p + n^{-a}\eta)\right] \\
&:= \mathrm{E}(A) + \mathrm{E}(B) + \mathrm{E}(C) + \mathrm{E}(D).
\end{aligned}$$

Hölder's inequality allows us to bound $\mathrm{E}(B)$. Setting $p = 1 + \frac{\delta}{2}$, $q = \frac{p}{p-1}$, we get that

$$\begin{aligned}
|\mathrm{E}(B)| &\leq \mathrm{E}\left[\left|(Y_1 - \nu_p)^2 I(\hat{\nu}_p \leq Y_1 < \nu_p) I(\hat{\nu}_p < \nu_p - n^{-a}\eta)\right|\right] \\
&\leq \mathrm{E}^{\frac{1}{p}}\left[|(Y_1 - \nu_p)^2|^p\right] \mathrm{E}^{\frac{1}{q}}\left[|I(\hat{\nu}_p \leq Y_t < \nu_p) I(\hat{\nu}_p < \nu_p - n^{-a}\eta)|^q\right] \\
&\leq \mathrm{E}^{\frac{1}{p}}(|(Y_1 - \nu_p)|^{2+\delta}) \{\mathrm{P}(|\hat{\nu}_p - \nu_p| \geq n^{-a}\eta)\}^{\frac{1}{q}} \\
&\leq C^{\frac{2}{2+\delta}} \left[\mathrm{P}(|\hat{\nu}_p - \nu_p| \geq n^{-a}\eta)\right]^{\frac{\delta}{2+\delta}},
\end{aligned}$$

which converges to zero faster than any polynomial, by Lemma 1. The same argument can be used to prove that $|\mathrm{E}(C)|$ also converges to zero faster than any polynomial.

To compute $\mathrm{E}(A)$, we note that $0 \leq (Y_1 - \nu_p)^2 I(\hat{\nu}_p \leq Y_1 < \nu_p) I(\hat{\nu}_p \geq \nu_p - n^{-a}\eta) \leq (Y_1 - \nu_p)^2 I(\nu_p - n^{-a}\eta < Y_1 < \nu_p)$ a.s., so

$$\mathrm{E}(A) \leq \int_{\nu_p - n^{-a}\eta}^{\nu_p} (z - \nu_p)^2 f(z) dz \leq \max_{x \in [\nu_p - \eta, \nu_p]} f(x) \int_{\nu_p - n^{-a}\eta}^{\nu_p} (z - \nu_p)^2 dz = O\left(n^{-3a}\right).$$

Using the same arguments, and noting that $0 \leq (Y_1 - \nu_p)^2 I(\hat{\nu}_p > Y_1 \geq \nu_p) I(\hat{\nu}_p < \nu_p - n^{-a}\eta) \leq (Y_1 - \nu_p)^2 I(\nu_p < Y_1 < \nu_p + n^{-a}\eta)$, it can be shown that $\mathrm{E}(D) = O(n^{-3a})$, which finally implies that $\mathrm{E}\left(W_t^2\right) = O\left(n^{-3a}\right)$.

Similarly to what was done before, by choosing $a = \frac{1}{2} - \gamma$, for $\gamma > 0$ arbitrarily small, we conclude that $\mathrm{E}\left(W_1^2\right) = o\left(n^{-\frac{3}{2}+\kappa}\right)$, for an arbitrarily small $\kappa > 0$, which implies that $\mathrm{Var}(W_1) = o(n^{-\frac{3}{2}+\kappa})$.

Finally, we will find a bound for $\mathrm{Cov}(W_i, W_j)$, with $i, j = 1, ..., n, i \neq j$. Cauchy-Schwarz inequality gives us that

$$\left|\mathrm{Cov}(W_i, W_j)\right| \leq \mathrm{Var}^{\frac{1}{2}}(W_i)\mathrm{Var}^{\frac{1}{2}}(W_j) = o\left(n^{-\frac{3}{2}+\kappa}\right),$$

for an arbitrarily small $\kappa > 0$. Hence,

$$\begin{aligned}
\left|\mathrm{Var}\left(n^{-1}\sum_{t=1}^{n} W_t\right)\right| &= n^{-2}\left|\left[\sum_{t=1}^{n}\mathrm{Var}(W_t) + 2\sum_{i,j=1,i>j}^{n}\mathrm{Cov}\left(W_i, W_j\right)\right]\right| \\
&\leq n^{-2}\left[\sum_{t=1}^{n}\left|\mathrm{Var}(W_t)\right| + 2\sum_{i,j=1,i>j}^{n}\left|\mathrm{Cov}\left(W_i, W_j\right)\right|\right] \\
&= n^{-2}\left[o\left(n^{-\frac{3}{2}+\kappa}\right) + o\left(n^{-\frac{3}{2}+\kappa}\right)\right] \\
&= o\left(n^{-\frac{3}{2}+\kappa}\right).
\end{aligned}$$

Thus far, we have shown that $\mathrm{E}\left(n^{-1}\sum_{t=1}^{n} W_t\right) = o\left(n^{-1+k}\right)$ and $\mathrm{Var}\left(n^{-1}\sum_{t=1}^{n} W_t\right) = o\left(n^{-\frac{3}{2}+k}\right)$, for an arbitrarily small $\kappa > 0$. Theorem 2 allows us to conlude that

$$\frac{1}{n}\sum_{t=1}^{n} W_t - o\left(n^{-1+k}\right) = O_p\left(o\left(n^{-\frac{3}{4}+k}\right)\right), \text{for } \kappa > 0 \text{ arbitrarily small.}$$

We finish the proof if we can show that

1. $O_p\left(o\left(n^{-\frac{3}{4}+k}\right)\right) = o_p\left(n^{-\frac{3}{4}+k}\right)$,
2. $o\left(n^{-1+k}\right) = o_p\left(n^{-1+k}\right)$,
3. $o_p\left(n^{-1+k}\right) + o_p\left(n^{-\frac{3}{4}+k}\right) = o_p\left(n^{-\frac{3}{4}+k}\right)$.

The proofs of 2. and 3. are rather trivial: to prove 2. simply consider $o\left(n^{-1+k}\right)$ as a sequence of random variables that take their value with probability 1, and to prove 3. use the fact that if a stochastic process is $o_p\left(n^{-1+k}\right)$, then it also is $o_p\left(n^{-\frac{3}{4}+k}\right)$. The proof of 1. is as follows:

Name the $O_p\left(o\left(n^{-\frac{3}{4}+k}\right)\right)$ as T_n. From the assumption, for all $\eta > 0$, there exist $C(\eta)$ and $n(\eta)$ such that for all $n \geq n(\eta)$,

$$\mathrm{P}\left(\frac{|T_n|}{a_n} < C(\eta)\right) \geq 1 - \eta,$$

where $a_n = o\left(n^{-\frac{3}{4}+k}\right)$. Let $\varepsilon^* > 0, \eta^* > 0$ be arbitrarily fixed. Again, from our assumption, there exist $C(\eta^*)$ and $n(\eta^*)$ such that for all $n \geq n(\eta^*)$,

$$\mathrm{P}\left(\frac{|T_n|}{a_n} < C(\eta^*)\right) \geq 1 - \eta^* \Leftrightarrow \mathrm{P}\left(\frac{|T_n|}{n^{-\frac{3}{4}+k}} < \frac{a_n}{n^{-\frac{3}{4}+k}}C(\eta^*)\right) \geq 1 - \eta^*.$$

When $n \to \infty$, $\frac{a_n}{n^{-\frac{3}{4}+k}} K(\eta^*) \to 0 < \varepsilon^*$, which implies that, for n large enough,

$$\mathrm{P}\left(\frac{|T_n|}{n^{-\frac{3}{4}+k}} < \varepsilon^*\right) \geq 1 - \eta^*,$$

that is, $T_n = o_p\left(n^{-\frac{3}{4}+k}\right)$. This allows us to write

$$\frac{1}{n}\sum_{t=1}^{n} W_t = o_p\left(n^{-\frac{3}{4}+k}\right),$$

which finishes the Lemma's proof. ■

Out of all the four Lemmas, it is only Lemma 2 that will help us prove the asymptotic normality of $\hat{\mu}_p$. Lemmas 3 and 4, which we present next, will aid us with the proof of asymptotic normality of $\hat{\mu}_{p,h}$.

In Lemma 3 we will show that some covariances that will appear when we deduce $\mathrm{Var}(\hat{\mu}_{p,h})$ are simply $o\left(n^{-1}h\right)$, which will facilitate our computations when we prove the convergence of the smoothed estimator.

Lemma 3 (Chen, 2008). *Let* $\hat{\beta} = (np)^{-1}\sum_{t=1}^{n} Y_t G_h(\nu_p - Y_t)$, $\hat{\eta} = (np)^{-1}\sum_{t=1}^{n} Y_t K_h(\nu_p - Y_t)$, $\beta = E\left(\hat{\beta}\right)$ *and* $\eta = E(\hat{\eta})$. *Under conditions* (i) – (iv),

$$\begin{aligned}
&(a) \quad Cov\left[\hat{\beta}, \left(p - \hat{S}(\nu_p;h)\right)\left(\hat{f}(\nu_p;h) - f(\nu_p)\right)\right] = o\left(n^{-1}h\right), \\
&(b) \quad Cov\left[\hat{\beta}, \left(\hat{\eta} - \eta\right)\left(p - \hat{S}(\nu_p;h)\right)\right] = o\left(n^{-1}h\right), \\
&(c) \quad Cov\left[\left(p - \hat{S}(\nu_p;h)\right), \left(\hat{\eta} - \eta\right)\left(p - \hat{S}(\nu_p;h)\right)\right] = o\left(n^{-1}h\right).
\end{aligned}$$

Proof. From (1.2), it is clear that

$$\mathrm{E}\left(\hat{f}(\nu_p;h)\right) = \mathrm{E}\left(K_h(\nu_p - Y_t)\right) = f(\nu_p) + O\left(h^2\right).$$

Moreover, by means of Fubini's and Taylor's theorems,

$$\begin{aligned}
\mathrm{E}(\hat{S}(\nu_p;h)) &= \int \int_{\frac{\nu_p - z}{h}}^{\infty} K(u)\,du f(z)\,dz = \int K(u)\left(\int_{\nu_p}^{\infty} f(z)\,dz + \int_{\nu_p - hu}^{\nu_p} f(z)\,dz\right)du \\
&= p + \int K(u)\left[F(\nu_p) - F(\nu_p - hu)\right]du = p - \frac{1}{2}f'(\nu_p)\sigma_K^2 h^2 + o\left(h^2\right) = p + O\left(h^2\right).
\end{aligned}$$

Using these definitions and results, let

$$\hat{\beta}-\beta=n^{-1}\sum_{t=1}^{n}\left[p^{-1}\left(Y_tG_h\left(\nu_p-Y_t\right)-\mathrm{E}\left(Y_tG_h\left(\nu_p-Y_t\right)\right)\right)\right]=:n^{-1}\sum_{t=1}^{n}\psi_1\left(Y_t\right),$$

$$\hat{f}\left(\nu_p;h\right)-f\left(\nu_p\right)=n^{-1}\sum_{t=1}^{n}\left[K_h\left(\nu_p-Y_t\right)-\mathrm{E}\left(K_h\left(\nu_p-Y_t\right)\right)\right]+O\left(h^2\right)=:n^{-1}\sum_{t=1}^{n}\psi_2\left(Y_t\right)+O\left(h^2\right),$$

$$p-\hat{S}\left(\nu_p;h\right)=n^{-1}\sum_{t=1}^{n}\left[\mathrm{E}\left(G_h\left(\nu_p-Y_t\right)\right)-G_h\left(\nu_p-Y_t\right)\right]+O\left(h^2\right)=:n^{-1}\sum_{t=1}^{n}\psi_3\left(Y_t\right)+O\left(h^2\right),$$

$$\hat{\eta}-\eta=n^{-1}\sum_{t=1}^{n}\left[p^{-1}\left(Y_tK_h\left(\nu_p-Y_t\right)-\mathrm{E}\left(Y_tK_h\left(\nu_p-Y_t\right)\right)\right)\right]=:n^{-1}\sum_{t=1}^{n}\psi_4\left(Y_t\right).$$

Note that for $j=1,2,3,4$ and for $t=1,\ldots,n$, $\mathrm{E}\left(\psi_j\left(Y_t\right)\right)=0$.

We start with the proof of (a). Using the definitions above, we may write

$$\begin{aligned}
&\left|\mathrm{Cov}\left[\hat{\beta},\left(p-\hat{S}(\nu_p;h)\right)\left(\hat{f}(\nu_p;h)-f\left(\nu_p\right)\right)\right]\right|=\left|\mathrm{E}\left[\left(\hat{\beta}-\beta\right)\left(p-\hat{S}(\nu_p;h)\right)\left(\hat{f}(\nu_p;h)-f\left(\nu_p\right)\right)\right]\right|\\
=&\left|\mathrm{E}\left[\left(n^{-1}\sum_{t=1}^{n}\psi_1\left(Y_t\right)\right)\left(n^{-1}\sum_{t=1}^{n}\psi_2\left(Y_t\right)+O\left(h^2\right)\right)\left(n^{-1}\sum_{t=1}^{n}\psi_3\left(Y_t\right)+O\left(h^2\right)\right)\right]\right|\\
\leq&\left|\mathrm{E}\left[n^{-3}\sum_{t=1}^{n}\psi_1\left(Y_t\right)\sum_{t=1}^{n}\psi_2\left(Y_t\right)\sum_{t=1}^{n}\psi_3\left(Y_t\right)\right]\right|+\left|\mathrm{E}\left[\sum_{t=1}^{n}\psi_1\left(Y_t\right)\sum_{t=1}^{n}\psi_2\left(Y_t\right)\right]\right|O\left(n^{-2}h^2\right)\\
&+\left|\mathrm{E}\left[\sum_{t=1}^{n}\psi_1\left(Y_t\right)\sum_{t=1}^{n}\psi_3\left(Y_t\right)\right]\right|O\left(n^{-2}h^2\right).
\end{aligned}\tag{2.9}$$

Let us start by studying the first term in (2.9). As we have pointed out, $\psi_1(Y_t),\psi_2(Y_t)$ and $\psi_3(Y_t)$ for any $t=1,\ldots,n$ are centered random variables. This together with the stationary of $Y=(Y_t,t\in\mathbb{N})$ allows us to write, in an analogous way as done in Billingsley (1968, p 173),

$$\left|\mathrm{E}\left[n^{-3}\sum_{t=1}^{n}\psi_1\left(Y_t\right)\sum_{t=1}^{n}\psi_2\left(Y_t\right)\sum_{t=1}^{n}\psi_3\left(Y_t\right)\right]\right|\leq 3!n^{-2}\sum_{i\geq1,j\geq1,i+j\leq n}\left|\mathrm{E}\left(\psi_1\left(Y_1\right)\psi_2\left(Y_i\right)\psi_3\left(Y_{i+j}\right)\right)\right|\tag{2.10}$$

Setting $p=2+\delta'$, $q=2+\delta'$ for some $\delta'\in(0,\delta)$ and $s^{-1}=1-p^{-1}-q^{-1}$, Davydov's inequality gives us that

$$\begin{aligned}
|\mathrm{E}(\psi_1(Y_1)\psi_2(Y_i)\psi_3(Y_{i+j}))|&=\mathrm{Cov}(\psi_1(Y_1),\psi_2(Y_i)\psi_3(Y_{i+j}))\\
&\leq 12||\psi_1(Y_1)||_p||\psi_2(Y_i)\psi_3(Y_{i+j})||_q\alpha^{1/s}(i).
\end{aligned}\tag{2.11}$$

We start by bounding $||\psi_1(Y_1)||_p=\mathrm{E}^{1/p}\left(|\psi_1(Y_1)|^p\right)$. Write

$$\mathrm{E}\left(|\psi_1(Y_1)|^p\right)=p^{-2-\delta'}\mathrm{E}\left(\left|(Y_1G_h(\nu_p-Y_1)-\mathrm{E}(Y_1G_h(\nu_p-Y_1))\right|^{2+\delta'}\right).\tag{2.12}$$

Applying the c_r-inequality in (2.12), it is clear that

$$\mathrm{E}(|\psi_1(Y_1)|^p)\leq p^{-2-\delta'}2^{1+\delta'}\left[\mathrm{E}\left(|Y_1G_h(\nu_p-Y_1)|^{2+\delta'}\right)+|\mathrm{E}(Y_1G_h(\nu_p-Y_1))|^{2+\delta'}\right].\tag{2.13}$$

On the other hand, Hölder's inequality gives us that

$$|\mathrm{E}\left(Y_1 G_h(\nu_p - Y_1)\right)| \leq \mathrm{E}^{\frac{1}{2+\delta'}}\left(|Y_1 G_h(\nu_p - Y_1)|^{2+\delta'}\right). \tag{2.14}$$

Combine (2.12), (2.13) and (2.14) to obtain

$$\mathrm{E}(|\psi_1(Y_1)|^p) \leq p^{-2-\delta'} 2^{2+\delta'} \mathrm{E}\left(|Y_1 G_h(\nu_p - Y_1)|^{2+\delta'}\right). \tag{2.15}$$

Finally, set $\xi_1 = \frac{2+\delta}{2+\delta'} > 1$ and $\xi_2 = (1 - \xi_1^{-1})^{-1}$. Applying Hölder's inequality once more in (2.15), we conclude that

$$\mathrm{E}\left(|Y_1|^{2+\delta'} G_h^{2+\delta'}(\nu_p - Y_1)\right) \leq \mathrm{E}^{\frac{1}{\xi_1}}\left(|Y_1|^{2+\delta}\right) \mathrm{E}^{\frac{1}{\xi_2}}\left(|G_h^{2+\delta'}(\nu_p - Y_1)|^{\xi_2}\right) \leq C. \tag{2.16}$$

Hence

$$||\psi_1(Y_1)||_p \leq C. \tag{2.17}$$

In order to bound $||\psi_2(Y_i)\psi_3(Y_{i+j})||_q$, we start by noting that, since $|\psi_3(Y_{i+j})| \leq 2$, we get that $||\psi_2(Y_i)\psi_3(Y_{i+j})||_q \leq 2||\psi_2(Y_i)||_q$. Applying the same arguments present in (2.13) and (2.14) as above and remembering that K is a non-negative function, we get that

$$\begin{aligned}
\mathrm{E}(|\psi_2(Y_i)|^q) &= \mathrm{E}\left(|K_h(\nu_p - Y_i) - \mathrm{E}(K_h(\nu_p - Y_i))|^{2+\delta'}\right) \\
&\leq 2^{1+\delta'}\left[\mathrm{E}\left(K_h^{2+\delta'}(\nu_p - Y_i)\right) + \mathrm{E}^{2+\delta'}\left(K_h(\nu_p - Y_i)\right)\right] \\
&\leq 2^{2+\delta'} \mathrm{E}\left(K_h^{2+\delta'}(\nu_p - Y_i)\right) \\
&= 2^{2+\delta'} \int h^{-2-\delta'} K^{2+\delta'}\left(\frac{\nu_p - z}{h}\right) f(z) dz \\
&= 2^{2+\delta'} h^{-1-\delta'} \int K^{2+\delta'}(u) f(\nu_p - uh) du
\end{aligned} \tag{2.18}$$

Assumption (iii) guarantees that the integral in (2.18) is finite. Therefore

$$||\psi_2(Y_i)\psi_3(Y_{i+j})||_q \leq 2\mathrm{E}^{\frac{1}{2+\delta'}}\left(|\psi_2(Y_i)|^{2+\delta'}\right) \leq C h^{-\frac{1+\delta'}{2+\delta'}}. \tag{2.19}$$

Consequently, combining (2.11), (2.17) and (2.19), we obtain

$$|\mathrm{E}(\psi_1(Y_1)\psi_2(Y_i)\psi_3(Y_{i+j}))| \leq 12 C h^{-\frac{1+\delta'}{2+\delta'}} \alpha^{\frac{\delta'}{2+\delta'}}(i). \tag{2.20}$$

Applying Davydov's inequality in a different manner, we can also deduce that

$$\begin{aligned}
|\mathrm{E}(\psi_1(Y_1)\psi_2(Y_i)\psi_3(Y_{i+j}))| &= \mathrm{Cov}(\psi_1(Y_1)\psi_2(Y_i), \psi_3(Y_{i+j})) \\
&\leq 12||\psi_1(Y_1)\psi_2(Y_i)||_p ||\psi_3(Y_{i+j})||_q \alpha^{1/s}(j).
\end{aligned} \tag{2.21}$$

Employing Hölder's inequality and the same arguments present in (2.15), (2.16) and (2.18), it can be shown that $||\psi_1(Y_1)\psi_2(Y_i)||_p \leq Ch^{-\frac{1+\delta'}{2+\delta'}}$, and since $||\psi_3(Y_{i+j})||_q \leq 2$, we conclude that

$$|\mathrm{E}(\psi_1(Y_1)\psi_2(Y_i)\psi_3(Y_{i+j}))| \leq 12Ch^{-\frac{1+\delta'}{2+\delta'}}\alpha^{\frac{\delta'}{2+\delta'}}(j), \tag{2.22}$$

and from (2.20) and (2.22), we derive

$$|\mathrm{E}(\psi_1(Y_1)\psi_2(Y_i)\psi_3(Y_{i+j}))| \leq 12Ch^{-\frac{1+\delta'}{2+\delta'}}\min\{\alpha^{\frac{\delta'}{2+\delta'}}(i), \alpha^{\frac{\delta'}{2+\delta'}}(j)\}. \tag{2.23}$$

To control the second term in (2.9) we note that, by stationarity,

$$\begin{aligned}\left|\mathrm{E}\left[\sum_{t=1}^{n}\psi_1(Y_t)\sum_{t=1}^{n}\psi_2(Y_t)\right]\right| &= \left|\mathrm{Cov}\left(\sum_{t=1}^{n}Y_tG_h(\nu_p-Y_t), \sum_{t=1}^{n}K_h(\nu_p-Y_t)\right)\right| \\ &= \left|\sum_{t=0}^{n-1}(n-t)\mathrm{Cov}(Y_1G_h(\nu_p-Y_1), K_h(\nu_p-Y_{t+1}))\right| \leq n\sum_{t=0}^{n-1}|\mathrm{Cov}(Y_1G_h(\nu_p-Y_1), K_h(\nu_p-Y_{t+1}))|.\end{aligned} \tag{2.24}$$

The bounding of this quantity becomes much simpler if we define $\alpha(0)$ in the same manner as we have defined the other α coefficients. This is not an uncommon practice, as can be seen in Yokoyama (1980), and for the purpose of this proof, we only need to realize that $\alpha(0)$ is obviously a finite value. We can then use Davydov's inequality in a much more direct way. For that purpose, setting $p = q = 2+\delta'$ for some $\delta' \in (0,\delta)$ and $s^{-1} = 1-p^{-1}-q^{-1}$, we can conclude that

$$|\mathrm{Cov}(Y_1G_h(\nu_p-Y_1), K_h(\nu_p-Y_{t+1}))| \leq 12||Y_1G_h(\nu_p-Y_1)||_p||K_h(\nu_p-Y_1)||_q\alpha^{\frac{\delta'}{2+\delta'}}(t). \tag{2.25}$$

Using analogous arguments to the ones we have used to prove (2.17) and (2.19), it can easily be shown that $||Y_1G_h(\nu_p-Y_1)||_p \leq C$ and $||K_h(\nu_p-Y_1)||_q \leq Ch^{-\frac{1+\delta'}{2+\delta'}}$. This fact, the summability of the α-mixing coefficients, (2.24) and (2.25) imply that

$$\left|\mathrm{E}\left[\sum_{t=1}^{n}\psi_1(Y_t)\sum_{t=1}^{n}\psi_2(Y_t)\right]\right| \leq Cnh^{-\frac{1+\delta'}{2+\delta'}}\sum_{t=0}^{n-1}\alpha^{\frac{\delta'}{2+\delta'}}(t) = O\left(nh^{-\frac{1+\delta'}{2+\delta'}}\right). \tag{2.26}$$

Analogous arguments to (2.24) and (2.25) can be used to show that

$$\left|\mathrm{E}\left[\sum_{t=1}^{n}\psi_1(Y_t)\sum_{t=1}^{n}\psi_3(Y_t)\right]\right| \leq Cn = O(n). \tag{2.27}$$

Taking into account that $\alpha(k)$ is a non-increasing sequence (Remark 2), and noticing that $O\left(n^{-2}h^2\right)O\left(nh^{-\frac{1+\delta'}{2+\delta'}}\right) = o\left(n^{-1}h\right)$, combine (2.9), (2.10), (2.23), (2.26) and (2.27) to conclude that

$$\left|\mathrm{Cov}\left[\hat{\beta}, \left(p-\hat{S}(\nu_p;h)\right)\left(\hat{f}(\nu_p;h)-f(\nu_p)\right)\right]\right| \leq Cn^{-2}h^{-\frac{1+\delta'}{2+\delta'}}\sum_{j=1}^{n-1}(2j-1)\alpha^{\frac{\delta'}{2+\delta'}}(j) + o\left(n^{-1}h\right). \tag{2.28}$$

Since $\sum_{j=1}^{\infty} j\alpha^{\frac{\delta}{2+\delta}}(j) < \infty$ is implied by Condition (i), the first term of the right hand side of (2.28) is simply $O\left(n^{-2}h^{-\frac{1+\delta'}{2+\delta'}}\right) = o\left(n^{-1}h\right)$, which concludes the proof of (a).

Let us proceed with the proof of (b). Using analogous arguments as the ones presented in (2.9) and (2.10),

$$\begin{aligned}
&\left|\mathrm{Cov}\left[\hat{\beta},(\hat{\eta}-\eta)\left(p-\hat{S}(\nu_p;h)\right)\right]\right| = \left|\mathrm{E}\left[\left(\hat{\beta}-\beta\right)(\hat{\eta}-\eta)\left(p-\hat{S}(\nu_p;h)\right)\right]\right| \\
&= \left|\mathrm{E}\left[n^{-3}\sum_{t=1}^{n}\psi_1(Y_t)\sum_{t=1}^{n}\psi_3(Y_t)\sum_{t=1}^{n}\psi_4(Y_t)\right] + \mathrm{E}\left[\sum_{t=1}^{n}\psi_1(Y_t)\sum_{t=1}^{n}\psi_4(Y_t)\right]\right| O\left(n^{-2}h^2\right) \\
&\le 3!n^{-2}\sum_{i\ge1,j\ge1,i+j\le n}\left|\mathrm{E}\left(\psi_1(Y_1)\psi_3(Y_i)\psi_4(Y_{i+j})\right)\right| + \left|\mathrm{E}\left[\sum_{t=1}^{n}\psi_1(Y_t)\sum_{t=1}^{n}\psi_4(Y_t)\right]\right| O\left(n^{-2}h^2\right). \quad (2.29)
\end{aligned}$$

Applying Davydov's inequality with the same choices for p,q and s as before, we get that

$$|\mathrm{E}(\psi_1(Y_1)\psi_3(Y_i)\psi_4(Y_{i+j}))| \le 12||\psi_1(Y_1)||_p||\psi_3(Y_i)\psi_4(Y_{i+j})||_q\alpha^{1/s}(i). \quad (2.30)$$

Applying both Hölder's and C_r-inequality,

$$\mathrm{E}(|\psi_4(Y_{i+j})|^q) \le p^{-2-\delta'}2^{2+\delta'}\mathrm{E}\left(|Y_{i+j}K_h(\nu_p - Y_{i+j})|^{2+\delta'}\right) \quad (2.31)$$

Applying Hölder's inequality once again with ξ_1 and ξ_2 as above, we conclude that

$$\begin{aligned}
\mathrm{E}\left(|Y_{i+j}|^{2+\delta'}K_h^{2+\delta'}(\nu_p - Y_{i+j})\right) &\le \mathrm{E}^{\frac{1}{\xi_1}}\left(|Y_{i+j}|^{2+\delta}\right)\mathrm{E}^{\frac{1}{\xi_2}}\left(K_h^{(2+\delta')\xi_2}(\nu_p - Y_{i+j})\right) \\
&\le C^{\frac{1}{\xi_1}}\left[\int h^{-(2+\delta')\xi_2}K^{(2+\delta')\xi_2}\left(\frac{\nu_p - z}{h}\right)f(z)dz\right]^{\frac{1}{\xi_2}} \\
&= C\left[h^{-(2+\delta')\xi_2+1}\int K^{(2+\delta')\xi_2}(u)f(\nu_p - hu)du\right]^{\frac{1}{\xi_2}} \\
&\le Ch^{-(2+\delta')+\frac{1}{\xi_2}}. \quad (2.32)
\end{aligned}$$

Hence

$$||\psi_4(Y_{i+j})||_q \le Ch^{-1+\frac{\xi_2^{-1}}{2+\delta'}}. \quad (2.33)$$

Consequently, using the bound we derived in (2.17) for $||\psi_1(Y_1)||_p$ and the fact that $|\psi_3(Y_i)| \le 2$, we get that

$$|\mathrm{E}(\psi_1(Y_1)\psi_3(Y_i)\psi_4(Y_{i+j}))| \le 12Ch^{-1+\frac{\xi_2^{-1}}{2+\delta'}}\alpha^{\frac{\delta'}{2+\delta'}}(i). \quad (2.34)$$

Applying Davydov's inequality in a different manner, we also conclude that

$$|\mathrm{E}(\psi_1(Y_1)\psi_3(Y_i)\psi_4(Y_{i+j}))| \le 12||\psi_1(Y_1)\psi_3(Y_i)||_p||\psi_4(Y_{i+j})||_q\alpha^{1/s}(j), \quad (2.35)$$

where p and q are the same constants as before. Applying the same arguments and bounds presented in (2.17) and (2.33) to inequality (2.35), we conclude that

$$|\mathrm{E}(\psi_1(Y_1)\psi_3(Y_i)\psi_4(Y_{i+j}))| \leq 12Ch^{-1+\frac{\xi_2^{-1}}{2+\delta'}}\alpha^{\frac{\delta'}{2+\delta'}}(j), \tag{2.36}$$

which, coupled with (2.34) finally implies that

$$|\mathrm{E}(\psi_1(Y_1)\psi_2(Y_i)\psi_3(Y_{i+j}))| \leq 12Ch^{-1+\frac{\xi_2^{-1}}{2+\delta'}}\min\{\alpha^{\frac{\delta'}{2+\delta'}}(i), \alpha^{\frac{\delta'}{2+\delta'}}(j)\}. \tag{2.37}$$

As was done for the proof of (a), and using analogous arguments as the ones presented in (2.24), (2.25), (2.17) and (2.33) it can easily be shown that

$$\left|\mathrm{E}\left[\sum_{t=1}^{n}\psi_1(Y_t)\sum_{t=1}^{n}\psi_4(Y_t)\right]\right| \leq Cnh^{-1+\frac{\xi_2^{-1}}{2+\delta'}},$$

which implies that

$$\left|\mathrm{E}\left[\sum_{t=1}^{n}\psi_1(Y_t)\sum_{t=1}^{n}\psi_4(Y_t)\right]O\left(n^{-2}h^2\right)\right| = o\left(n^{-1}h\right). \tag{2.38}$$

Combining (2.29), (2.37) and (2.38), we conclude that

$$\left|\mathrm{Cov}\left[\hat{\beta}, (\hat{\eta}-\eta)\left(p-\hat{S}(\nu_p;h)\right)\right]\right| \leq Cn^{-2}h^{-1+\frac{\xi_2^{-1}}{2+\delta'}}\sum_{j=1}^{n-1}(2j-1)\alpha^{\frac{\delta'}{2+\delta'}}(j) + o\left(n^{-1}h\right) = o\left(n^{-1}h\right).$$

Lastly, to prove (c), we note that, using the same argument that was presented in (2.10),

$$\begin{aligned}
&\left|\mathrm{Cov}\left[\left(p-\hat{S}(\nu_p;h)\right), (\hat{\eta}-\eta)\left(p-\hat{S}(\nu_p;h)\right)\right]\right| \\
&= \left|\mathrm{E}\left[\left(p-\hat{S}(\nu_p;h)\right)^2(\hat{\eta}-\eta)\right] - \mathrm{E}\left(p-\hat{S}(\nu_p;h)\right)\mathrm{E}\left((\hat{\eta}-\eta)\left(p-\hat{S}(\nu_p;h)\right)\right)\right| \\
&\leq \left|\mathrm{E}\left[\left(\left(n^{-1}\sum_{t=1}^{n}\psi_3(Y_t)\right)^2 + O\left(h^4\right) + n^{-1}\sum_{t=1}^{n}\psi_3(Y_t)O\left(h^2\right)\right)\left(n^{-1}\sum_{t=1}^{n}\psi_4(Y_t)\right)\right]\right| \\
&+ \left|O\left(h^2\right)\mathrm{E}\left[n^{-2}\sum_{t=1}^{n}\psi_3(Y_t)\sum_{t=1}^{n}\psi_4(Y_t) + n^{-1}\sum_{t=1}^{n}\psi_4(Y_t)O\left(h^2\right)\right]\right| \\
&\leq \left|\mathrm{E}\left[n^{-3}\sum_{t=1}^{n}\psi_3(Y_t)\sum_{t=1}^{n}\psi_3(Y_t)\sum_{t=1}^{n}\psi_4(Y_t)\right]\right| + \left|\mathrm{E}\left[\sum_{t=1}^{n}\psi_3(Y_t)\sum_{t=1}^{n}\psi_4(Y_t)\right]\right|O\left(n^{-2}h^2\right) \\
&\leq 3!n^{-2}\sum_{i\geq 1, j\geq 1, i+j\leq n}\left|\mathrm{E}\left(\psi_3(Y_1)\psi_3(Y_i)\psi_4(Y_{i+j})\right)\right| + \left|\mathrm{E}\left[\sum_{t=1}^{n}\psi_3(Y_t)\sum_{t=1}^{n}\psi_4(Y_t)\right]\right|O\left(n^{-2}h^2\right).
\end{aligned} \tag{2.39}$$

Once more, we apply Davydov's inequality with the same constants, and recall (2.33) to conclude that

$$|\mathrm{E}(\psi_3(Y_1)\psi_3(Y_i)\psi_4(Y_{i+j}))| \leq 12||\psi_3(Y_1)||_p||\psi_3(Y_i)\psi_4(Y_{i+j})||_q\alpha^{1/s}(i)$$
$$\leq C||\psi_4(Y_{i+j})||_q\alpha^{1/s}(i) \leq Ch^{-1+\frac{\xi_2^{-1}}{2+\delta'}}\alpha^{1/s}(i), \tag{2.40}$$

and that

$$|\mathrm{E}(\psi_3(Y_1)\psi_3(Y_i)\psi_4(Y_{i+j}))| \leq 12||\psi_3(Y_1)\psi_3(Y_i)||_p||\psi_4(Y_{i+j})||_q\alpha^{1/s}(j)$$
$$\leq C||\psi_4(Y_{i+j})||_q\alpha^{1/s}(j) \leq Ch^{-1+\frac{\xi_2^{-1}}{2+\delta'}}\alpha^{1/s}(j). \tag{2.41}$$

Combining these last two inequalities leads to

$$|\mathrm{E}(\psi_3(Y_1)\psi_3(Y_i)\psi_4(Y_{i+j}))| \leq Ch^{-1+\frac{\xi_2^{-1}}{2+\delta'}}\min\{\alpha^{\frac{\delta'}{2+\delta'}}(i),\alpha^{\frac{\delta'}{2+\delta'}}(j)\}. \tag{2.42}$$

As was done before, using the arguments presented in (2.24), (2.25) and (2.33), it is easily shown that

$$\left|\mathrm{E}\left[\sum_{t=1}^{n}\psi_3\left(Y_t\right)\sum_{t=1}^{n}\psi_4\left(Y_t\right)\right]\right| \leq Cnh^{-1+\frac{\xi_2^{-1}}{2+\delta'}},$$

which implies that

$$\left|\mathrm{E}\left[\sum_{t=1}^{n}\psi_3\left(Y_t\right)\sum_{t=1}^{n}\psi_4\left(Y_t\right)\right]O\left(n^{-2}h^2\right)\right| = o\left(n^{-1}h\right). \tag{2.43}$$

Combine (2.39), (2.42) and (2.43) to conclude that

$$\left|\mathrm{Cov}\left[\left(p-\hat{S}(\nu_p;h)\right),(\hat{\eta}-\eta)\left(p-\hat{S}(\nu_p;h)\right)\right]\right| \leq Cn^{-2}h^{-1+\frac{\xi_2^{-1}}{2+\delta'}}\sum_{j=1}^{n-1}(2j-1)\,\alpha^{\frac{\delta'}{2+\delta'}}(j)+o\left(n^{-1}h\right)$$
$$= o\left(n^{-1}h\right).$$

This completes the proof of Lemma 3. ■

Lemma 4 will be useful for two purposes: It will serve to identify the small order of some quantities, but most importantly, it will provide a mean of comparison of the asymptotic variance of both estimators.

Lemma 4 (Chen, 2008). *Under conditions* (i) − (iv), *for* $i = 0,1$ *and* $j = 0,1$,

$$\sum_{k=1}^{n-1}\left(1-\frac{k}{n}\right)\left[Cov\left(Y_1^iG_h(\nu_p-Y_1),Y_{k+1}^jG_h(\nu_p-Y_{k+1})\right)\right.$$
$$\left.-\,Cov\left(Y_1^iI\left(Y_1>\nu_p\right),Y_{k+1}^jI\left(Y_{k+1}>\nu_p\right)\right)\right]=o\left(h\right).$$

Proof. We start by proving the case $i = j = 0$ much like is done in Cai and Roussas (1998). As we shall see later, the proof of the other cases could be extended to the case $i = j = 0$. Nevertheless, we opt to present a different proof for this first case as it brings some interesting insight. Set $\gamma_h^{i,j}(k) = \text{Cov}\left(Y_1^iG_h\left(\nu_p-Y_1\right),Y_{k+1}^jG_h\left(\nu_p-Y_{k+1}\right)\right)$ and $\gamma^{i,j}(k) = \text{Cov}\left(Y_1^iI\left(Y_1>\nu_p\right),Y_{k+1}^jI\left(Y_{k+1}>\nu_p\right)\right)$. We shall first prove that

$$\left|\sum_{k=1}^{n-1}\left(1-\frac{k}{n}\right)\left[\gamma_h^{0,0}(k)-\gamma^{0,0}(k)\right]\right|=o\left(h\right).$$

Fix $k \in \mathbb{N}$. It is clear that

$$\gamma^{0,0}(k)=\text{Cov}(I(Y_1\leq\nu_p),I(Y_{k+1}\leq\nu_p))=F_k(\nu_p,\nu_p)-F^2(\nu_p). \tag{2.44}$$

On the other hand, Hoeffding's Covariance Identity gives us that $\gamma_h^{0,0}(k)$ is equal to

$$\int_{\mathbb{R}^2}\text{P}\left(G\left(\frac{\nu_p-Y_1}{h}\right)>u,G\left(\frac{\nu_p-Y_{k+1}}{h}\right)>v\right)$$
$$-\,\text{P}\left(G\left(\frac{\nu_p-Y_1}{h}\right)>u\right)\text{P}\left(G\left(\frac{\nu_p-Y_{k+1}}{h}\right)>v\right)dudv.$$

Let $r = G^{-1}(u), s = G^{-1}(v)$, where G^{-1} denotes the generalized inverse of G. Then $du = -K(r)dr$, $dv = -K(s)ds$. These two changes of variables allow us to write the last integral as

$$\int_{\mathbb{R}^2}\left[F_k\left(\nu_p-hr,\nu_p-hs\right)-F\left(\nu_p-hr\right)F\left(\nu_p-hs\right)\right]K\left(r\right)K\left(s\right)drds. \tag{2.45}$$

Finally, note that, from a change of variables,

$$\text{E}(\hat{F}(\nu_p;h))=\int F(\nu_p-hu)K(u)du,$$

and from Cai and Roussas (1998),

$$\left|\text{E}^2\left(\hat{F}\left(\nu_p;h\right)\right)-F^2\left(\nu_p\right)\right|\leq Ch^2.$$

We will now provide an upper bound for $\left|\gamma_h^{0,0}(k)-\gamma^{0,0}(k)\right|$. By means of a Taylor expansion to $F_k(\nu_p-hr,\nu_p-hs)$ around (ν_p,ν_p) we obtain, for some $\theta_1\in(\nu_p-hr,\nu_p)$, $\theta_2\in(\nu_p-hs,\nu_p)$,

$$\begin{aligned}
&\left|\gamma_h^{0,0}(k)-\gamma^{0,0}(k)\right|\\
&=\left|\int_{\mathbb{R}^2}\left[F_k(\nu_p-hr,\nu_p-hs)-F(\nu_p-hr)F(\nu_p-hs)\right]K(r)K(s)\,drds-F_k(\nu_p,\nu_p)-F^2(\nu_p)\right|\\
&=\left|\int_{\mathbb{R}^2}\left[(F_k(\nu_p-hr,\nu_p-hs)-F_k(\nu_p,\nu_p))-\left(F(\nu_p-hr)F(\nu_p-hs)-F^2(\nu_p)\right)\right]K(r)K(s)\,drds\right|\\
&\le\left|\int_{\mathbb{R}^2}\left[F_k(\nu_p-hr,\nu_p-hs)-F_k(\nu_p,\nu_p)\right]K(r)K(s)\,drds\right|+\left|\mathrm{E}^2\left(\hat F(\nu_p;h)\right)-F^2(\nu_p)\right|\\
&\le\left|\int_{\mathbb{R}^2}\left[\frac{\partial F_k}{\partial x}(\nu_p,\nu_p)(-hr)+\frac{\partial F_k}{\partial y}(\nu_p,\nu_p)(-hs)+\frac{1}{2}\left(\frac{\partial^2F_k}{\partial x^2}(\theta_1,\theta_2)h^2r^2+2\frac{\partial^2F_k}{\partial x\partial y}(\theta_1,\theta_2)h^2rs\right.\right.\right.\\
&\quad\left.\left.\left.+\frac{\partial^2F_k}{\partial y^2}(\theta_1,\theta_2)h^2s^2\right)\right]K(r)K(s)\,drds\right|+Ch^2\le|Bh^2\sigma_K^2|+Ch^2=Ch^2.
\end{aligned}\tag{2.46}$$

Now, from the definition of $\alpha(k)$, it is clear that $\sup_{(x,y)\in\mathbb{R}^2}|F_k(x,y)-F(x)F(y)|\le\alpha(k)$. From this, (2.44) and (2.45) allow us to conclude that $\left|\gamma_h^{0,0}(k)-\gamma^{0,0}(k)\right|\le 2\alpha(k)$, which coupled with (2.46) implies that

$$|\gamma_h^{0,0}(k)-\gamma^{0,0}(k)|=|\gamma_h^{0,0}(k)-\gamma^{0,0}(k)|^{2/3}|\gamma_h^{0,0}(k)-\gamma^{0,0}(k)|^{1/3}\le Ch^{4/3}\alpha^{1/3}(k),$$

and consequently,

$$\left|\sum_{k=1}^{n-1}\left(1-\frac{k}{n}\right)\left[\gamma_h^{0,0}(k)-\gamma^{0,0}(k)\right]\right|\le\sum_{k=1}^{n-1}\left|\gamma_h^{0,0}(k)-\gamma^{0,0}(k)\right|\le Ch^{4/3}\sum_{k=1}^{\infty}\alpha^{1/3}(k)=o(h),$$

which concludes the proof for the case $i=j=0$.

The proof for the other choices of i and j is quite similar, so we only present the proof for the case $i=j=1$. In this case, as we shall see later,

$$\mathrm{E}(Y_1I(Y_1>\nu_p))=p\mu_p\qquad\text{and}\qquad\mathrm{E}(Y_1G_h(\nu_p-Y_1))=p\mu_p+O(h^2).$$

From this, it is clear that

$$\begin{aligned}
&\left|\gamma_h^{1,1}(k)-\gamma^{1,1}(k)\right|\\
&\le\left|\mathrm{E}\left[Y_1Y_{k+1}G_h(\nu_p-Y_1)G_h(\nu_p-Y_{k+1})\right]-\mathrm{E}\left[Y_1Y_{k+1}I(Y_1>\nu_p)I(Y_{k+1}>\nu_p)\right]\right|+O\left(h^2\right).
\end{aligned}\tag{2.47}$$

Let us compute $\mathrm{E}\left[Y_1Y_{k+1}G_h(\nu_p-Y_1)G_h(\nu_p-Y_{k+1})\right]$. Write

$$\mathrm{E}\left[Y_1Y_{k+1}G_h(\nu_p-Y_1)G_h(\nu_p-Y_{k+1})\right]=\int\int\int_{\frac{\nu_p-x}{h}}^{\infty}K(u)\,du\int_{\frac{\nu_p-y}{h}}^{\infty}K(v)\,dv\;xyf_k(x,y)\,dxdy,\tag{2.48}$$

where $f_k(x,y)$ is the density function of the pair (Y_1, Y_{k+1}). To compute this integral, we start by changing the order of integration. We want that the last integral to be computed to be the one with respect to u. For that purpose, realize that for a given u, when v ranges from $-\infty$ to u, then both x and y range from $\nu_p - hv$ to $+\infty$, and when v ranges from u to $+\infty$, then both x and y range from $\nu_p - hu$ to $+\infty$. This allows us to write (2.48) as

$$\int K(u)\left[\int_{-\infty}^{u} K(v)\left\{\int_{\nu_p}^{\infty}\int_{\nu_p}^{\infty} xyf_k(x,y)dxdy + \int_{\nu_p-hv}^{\nu_p}\int_{\nu_p-hv}^{\nu_p} xyf_k(x,y)dxdy\right\}dv\right.$$
$$\left. + \int_{u}^{\infty} K(v)\left\{\int_{\nu_p}^{\infty}\int_{\nu_p}^{\infty} xyf_k(x,y)dxdy + \int_{\nu_p-hu}^{\nu_p}\int_{\nu_p-hu}^{\nu_p} xyf_k(x,y)dxdy\right\}dv\right]du. \qquad (2.49)$$

It is clear that $\mathrm{E}[Y_1Y_{k+1}I(Y_1 > \nu_p)I(Y_{k+1} > \nu_p)] = \int_{\nu_p}^{\infty}\int_{\nu_p}^{\infty} xyf_k(x,y)dxdy$. Putting together the appropriate integrals so that we may use the previous representation, (2.49) can be rewritten as

$$\int K(u)\left[\int_{-\infty}^{u} K(v)\int_{\nu_p-hv}^{\nu_p}\int_{\nu_p-hv}^{\nu_p} xyf_k(x,y)dxdydv + \int_{u}^{\infty} K(v)\left\{\int_{\nu_p-hu}^{\nu_p}\int_{\nu_p-hu}^{\nu_p} xyf_k(x,y)dxdy\right\}dv\right]du$$
$$+\mathrm{E}[Y_1Y_{k+1}I(Y_1 > \nu_p)I(Y_{k+1} > \nu_p)]. \qquad (2.50)$$

Set $d_K = \int u^2K(u)\int_{-\infty}^{u} K(v)dvdu$. Apply Taylor's theorem to rewrite $xyf_k(x,y)$ as $\nu_p^2 f_k(\nu_p,\nu_p) + [\theta_2 f_k(\theta_1,\theta_2) + \theta_1\theta_2\frac{\partial f_k}{\partial x}(\theta_1,\theta_2)](x-\nu_p) + [\theta_1 f_k(\theta_1,\theta_2) + \theta_1\theta_2\frac{\partial f_k}{\partial y}(\theta_1,\theta_2)](y-\nu_p)$, for some $\theta_1 \in (x,\nu_p)$, $\theta_2 \in (y,\nu_p)$. This allows us to write the integral present in (2.50) as

$$h^2\nu_p^2 f_k(\nu_p,\nu_p)\left(\int K(u)\int_{-\infty}^{u} v^2K(v)dvdu + \int u^2K(u)\int_{u}^{\infty} K(v)dvdu\right) + O(h^3). \qquad (2.51)$$

One more application of Fubini's theorem allows us to realize that what is inside the parenthesis in (2.51) is simply $2(\sigma_K^2 - d_K)$. Consequently, from (2.48), (2.49), (2.50) and (2.51),

$$\mathrm{E}[Y_1Y_{k+1}G_h(\nu_p - Y_1)G_h(\nu_p - Y_{k+1})] = \mathrm{E}[Y_1Y_{k+1}I(Y_1 > \nu_p)I(Y_{k+1} > \nu_p)] + O(h^2). \qquad (2.52)$$

Combine (2.47) and (2.52) to conclude that $\left|\gamma_h^{1,1}(k) - \gamma^{1,1}(k)\right| \le Ch^2$. Now from Davydov's inequality, and as was done in the proof of Lemma 3, setting $p = q = 2 + \delta'$, we obtain

$$\begin{aligned}\left|\gamma_h^{1,1}(k) - \gamma^{1,1}(k)\right| &\le \left|\gamma_h^{1,1}(k)\right| + \left|\gamma^{1,1}(k)\right| \\ &\le 12||Y_1G_h(\nu_p - Y_1)||_p^2\alpha^{\frac{\delta'}{2+\delta'}}(k) + 12||Y_1I(Y_1 > \nu_p)||_p^2\alpha^{\frac{\delta'}{2+\delta'}}(k) \\ &\le C\alpha^{\frac{\delta'}{2+\delta'}}(k),\end{aligned}$$

and consequently

$$\left|\sum_{k=1}^{n-1}\left(1-\frac{k}{n}\right)\left[\gamma_h^{1,1}(k) - \gamma^{1,1}(k)\right]\right| \le Ch^{4/3}\sum_{k=1}^{\infty}\alpha^{\frac{\delta'}{3(2+\delta')}}(k) = o(h).$$

This concludes the proof of Lemma 4. ■

We are now ready to present the two main theorems of this work. We start by proving the asymptotic normality of the unsmoothed estimator, $\hat{\mu}_p$.

2.2 Convergence of the empirical estimator

Theorem 13 (Chen, 2008, Central Limit Theorem for $\hat{\mu}_p$)**.** *Under conditions* (i) − (ii),

$$\sqrt{n}p\sigma_0^{-1}(p;n)(\hat{\mu}_p - \mu_p) \xrightarrow{d} N(0,1).$$

Proof. Let $\phi_1(z) = n^{-1}\sum_{t=1}^{n} Y_t I(Y_t \geq z)$ and $\phi_2(z) = n^{-1}\sum_{t=1}^{n} I(Y_t \geq z)$. Clearly $\hat{\mu}_p = \phi_1(\hat{\nu}_p)/\phi_2(\hat{\nu}_p)$. Let us start by noting that

$$\mathrm{E}(\phi_1(\nu_p)) = p\mu_p \qquad \text{and} \qquad \mathrm{E}(\phi_2(\nu_p)) = p.$$

Also note that

$$\phi_2(\hat{\nu}_p) = \begin{cases} \frac{[np]+1}{n} & \text{, if } [np] \neq np \\ p & \text{, if } [np] = np \end{cases},$$

which means that

$$\frac{[np]+1}{n} \geq \phi_2(\hat{\nu}_p) \geq p.$$

Consequently,

$$|\phi_2(\hat{\nu}_p) - p| = \phi_2(\hat{\nu}_p) - p \leq \frac{[np]+1}{n} - p \leq \frac{1}{n}.$$

By virtue of this fact, and using the same arguments that we presented in the end of Lemma's 2 proof,

$$\phi_2(\hat{\nu}_p) - E(\phi_2(\nu_p)) = o\left(n^{-1}\right) = o_p\left(n^{-3/4}\right). \tag{2.53}$$

From Lemma 2, for an arbitrarily small positive κ,

$$\begin{aligned} &\frac{1}{n}\sum_{t=1}^{n}(Y_t - \nu_p)\left[I(Y_t \geq \hat{\nu}_p) - I(Y_t \geq \nu_p)\right] = o_p\left(n^{-3/4+\kappa}\right) \\ &\Leftrightarrow \phi_1(\hat{\nu}_p) = \phi_1(\nu_p) + \nu_p(\phi_2(\hat{\nu}_p) - \phi_2(\nu_p)) + o_p\left(n^{-3/4+\kappa}\right). \end{aligned} \tag{2.54}$$

Let $h(x,y) = \frac{x}{y}$. Expand $h(\phi_1(\hat{\nu}_p), \phi_2(\hat{\nu}_p))$ around $(\mathrm{E}(\phi_1(\nu_p)), \mathrm{E}(\phi_1(\nu_p)))$ and consider (2.53) to obtain

$$\hat{\mu}_p = \mu_p + p^{-1}(\phi_1(\hat{\nu}_p) - p\mu_p) + o_p\left(n^{-3/4+k}\right). \tag{2.55}$$

Taking into consideration (2.53), plug (2.54) into (2.55) to conclude

$$\hat{\mu}_p = \mu_p + p^{-1}(\phi_1(\nu_p) - p\mu_p) + p^{-1}\nu_p(p - \phi_2(\nu_p)) + o_p\left(n^{-3/4+\kappa}\right) \tag{2.56}$$

$$= \mu_p + p^{-1}\left[n^{-1}\sum_{t=1}^{n}(Y_t - \nu_p)I(Y_t \geq \nu_p) - p(\mu_p - \nu_p)\right] + o_p\left(n^{-3/4+\kappa}\right). \tag{2.57}$$

Now set $T_{i,n} = \sigma_0^{-1}(p;n)\, p^{-1}\left[(Y_i - \nu_p) I(Y_i \geq \nu_p) - p(\mu_p - \nu_p)\right]$, for $i = 1,...,n$. From (2.57) we obtain the following equality

$$\sigma_0^{-1}(p;n)(\hat{\mu}_p - \mu_p) = n^{-1}\sum_{i=1}^{n} T_{i,n} + o_p\left(n^{-3/4+\kappa}\right). \tag{2.58}$$

We will prove the asymptotic normality of this random variable by employing the blocking technique.

Let k and k' be two sequences of positive integers such that $k' \to \infty$, $k'/k \to 0$ and $k/n \to 0$ as $n \to \infty$. Let r be a sequence of positive integers such that $r(k+k') \leq n < r(k+k'+1)$. Define the large blocks

$$V_{j,n} = T_{(j-1)(k+k')+1,n} + ... + T_{(j-1)(k+k')+k,n} \quad \text{for} \quad j = 1,2,\cdots,r,$$

the small blocks

$$V'_{j,n} = T_{(j-1)(k+k')+k+1,n} + ... + T_{(j-1)(k+k')+k+k',n} \quad \text{for} \quad j = 1,2,\cdots,r,$$

and the residual block

$$\delta_n = T_{r(k+k')+1,n} + ... + T_{n,n}.$$

Set

$$S_n := n^{-1/2}\sum_{i=1}^{n} T_{i,n} = n^{-1/2}\sum_{j=1}^{r} V_{j,n} + n^{-1/2}\sum_{j=1}^{r} V'_{j,n} + n^{-1/2}\delta_n =: S_{n,1} + S_{n,2} + S_{n,3}. \tag{2.59}$$

We will start by proving that $S_{n,2}$ and $S_{n,3}$ converge to 0 in probability. Start by noting that, using the definition of μ_p, for $i = 1,...,n$,

$$\begin{aligned}
\mathrm{E}(T_{i,n}) &= \mathrm{E}\left[\sigma_0^{-1}(p;n)\, p^{-1}\left((Y_i - \nu_p) I(Y_i \geq \nu_p) - p(\mu_p - \nu_p)\right)\right] \\
&= \sigma_0^{-1}(p;n)\, p^{-1}\left[\mathrm{E}\left((Y_i - \nu_p) I(Y_i \geq \nu_p)\right) - p(\mu_p - \nu_p)\right] \\
&= \sigma_0^{-1}(p;n)\, p^{-1}\left[\mathrm{E}(Y_i - \nu_p | Y_i \geq \nu_p)\, p - p(\mu_p - \nu_p)\right] \\
&= 0,
\end{aligned}$$

which means that $\mathrm{E}(S_{n,2}) = \mathrm{E}(S_{n,3}) = 0$. We proceed by computing $\mathrm{Var}(S_{n,3})$, which will give us insight on how to compute $\mathrm{Var}(S_{n,2})$. Set $\gamma(k) = \mathrm{Cov}\left[(Y_1 - \nu_p) I(Y_1 \geq \nu_p), (Y_{k+1} - \nu_p) I(Y_{k+1} \geq \nu_p)\right]$, with k a non-negative integer. Trivially,

$$\begin{aligned}
\mathrm{Var}(S_{n,3}) &= \mathrm{Var}\left(n^{-1/2}\delta_n\right) = \mathrm{Var}\left(n^{-1/2}\left(T_{r(k+k')+1,n} + \cdots + T_{n,n}\right)\right) \\
&= n^{-1}\left[\left(n - r\left(k+k'\right)\right)\mathrm{Var}(T_{i,n}) + 2\sum_{i,j=r(k+k')+1,j>i}^{n} \mathrm{Cov}(T_{i,n}, T_{j,n})\right].
\end{aligned} \tag{2.60}$$

Now, for $i = 1, ..., n$,

$$\begin{aligned}\mathrm{Var}(T_{i,n}) &= \mathrm{Var}\left[\sigma_0^{-1}(p;n)\,p^{-1}\left((Y_i-\nu_p)I(Y_i\geq\nu_p)-p(\mu_p-\nu_p)\right)\right]\\ &= \sigma_0^{-2}(p;n)\,p^{-2}\mathrm{Var}\left[\left((Y_i-\nu_p)I(Y_i\geq\nu_p)\right)-p(\mu_p-\nu_p)\right]\\ &= \sigma_0^{-2}(p;n)\,p^{-2}\mathrm{Var}\left[(Y_i-\nu_p)I(Y_i\geq\nu_p)\right]\\ &= \sigma_0^{-2}(p;n)p^{-2}\gamma(0). \end{aligned} \tag{2.61}$$

And for $i, j = 1, ..., n$, with $j > i$, taking into account that the $T_{i,n}$ are centered, it is clear that

$$\begin{aligned}&\mathrm{Cov}(T_{i,n},T_{j,n}) = \mathrm{E}(T_{i,n}T_{j,n})\\ &= \mathrm{E}\left[\sigma_0^{-2}(p;n)\,p^{-2}\left[(Y_i-\nu_p)I(Y_i\geq\nu_p)-p(\mu_p-\nu_p)\right]\left[(Y_j-\nu_p)I(Y_j\geq\nu_p)-p(\mu_p-\nu_p)\right]\right]\\ &= \sigma_0^{-2}(p;n)\,p^{-2}\left[\mathrm{E}\left[(Y_i-\nu_p)I(Y_i\geq\nu_p)(Y_j-\nu_p)I(Y_j\geq\nu_p)\right]-p^2(\mu_p-\nu_p)^2\right]\\ &= \sigma_0^{-2}(p;n)\,p^{-2}\mathrm{Cov}\left[(Y_i-\nu_p)I(Y_i\geq\nu_p),(Y_j-\nu_p)I(Y_j\geq\nu_p)\right] = \sigma_0^{-2}(p;n)\,p^{-2}\gamma(j-i).\end{aligned} \tag{2.62}$$

Hence, plugging (2.61) and (2.62) into (2.60), and due to stationarity, we obtain

$$\begin{aligned}\mathrm{Var}(S_{n,3}) &= n^{-1}p^{-2}\sigma_0^{-2}(p;n)\left[\left(n-r\left(k+k'\right)\right)\gamma(0)+2\sum_{i,j=r(k+k')+1,j>i}^{n}\gamma(j-i)\right]\\ &= n^{-1}p^{-2}\sigma_0^{-2}(p;n)\left[\left(n-r\left(k+k'\right)\right)\gamma(0)+2\sum_{i,j=1,j>i}^{n-r(k+k')}\gamma(j-i)\right]\\ &= n^{-1}p^{-2}\sigma_0^{-2}(p;n)\left[\left(n-r\left(k+k'\right)\right)\gamma(0)+2\sum_{l=1}^{n-r(k+k')-1}\left(n-r(k+k')-l\right)\gamma(l)\right]\\ &= n^{-1}p^{-2}\sigma_0^{-2}(p;n)\left(n-r\left(k+k'\right)\right)\left[\gamma(0)+2\sum_{l=1}^{n-r(k+k')-1}\left(1-\frac{l}{n-r(k+k')}\right)\gamma(l)\right]\\ &= \frac{(n-r(k+k'))\,\sigma_0^2(p;n-r(k+k'))}{np^2\sigma_0^2(p;n)}.\end{aligned} \tag{2.63}$$

The exact computation of $\mathrm{Var}(S_{n,2})$ is more complex than that of $\mathrm{Var}(S_{n,3})$, as $S_{n,2}$ is composed by more than one block of random variables. We will not be exhibiting it, as we only need to prove that it converges to zero. Let us assume for a moment that any two variables $T_{i,n}$, taken from two different small blocks, are independent. If that was the case, taking into consideration that $S_{n,2}$ is composed of r blocks, each with k' variables, and using the same arguments present in the computation of $\mathrm{Var}(S_{n,3})$, we would conclude that

$$\mathrm{Var}(S_{n,3}) = r\frac{k'\sigma_0^2(p;k')}{np^2\sigma_0^2(p;n)}.$$

In this situation, we are not taking into consideration the covariances between any of the variables between two different blocks. However, since any of those two random variables are apart by at least k other random variables, and k tends to infinity, Davydov's inequality guarantees that their covariance goes to zero, due to the fact that our original process is α-mixing. Consequently we conclude that

$$\mathrm{Var}(S_{n,2}) = \frac{rk'\left[\sigma_0^2(p;k') + o(1)\right]}{np^2\sigma_0^2(p;n)} \to 0,$$
$$\mathrm{Var}(S_{n,3}) = \frac{(n - r(k+k'))\sigma_0^2(p;n-r(k+k'))}{np^2\sigma_0^2(p;n)} \to 0.$$

Therefore, by theorem 3,

$$S_{n,2} \xrightarrow{\mathrm{p}} 0 \qquad \text{and} \qquad S_{n,3} \xrightarrow{\mathrm{p}} 0. \tag{2.64}$$

We now prove the asymptotic normality of $S_{n,1}$. By choosing $r \sim n^{2/3}$ (for instance, $r = [n^{2/3}]$), since $\sqrt{n}/r \to 0$, recursive applications of Bradley's Lemma guarantee that there exist independent and identically distributed random variables $W_{j,n}$, $j = 1, ..., r$, such that each $W_{j,n}$ has the same distribution as $V_{j,n}$ and

$$\mathrm{P}(|V_{j,n} - W_{j,n}| > \varepsilon\sqrt{n}/r) \leq 18\varepsilon^{-2/5}r^{2/5}n^{-1/5}(\mathrm{E}(V_{j,n}^2))^{1/5}\alpha^{4/5}(k'),$$

where $\varepsilon > 0$. Using Yokoyama's inequality and noticing that $rk \sim n$, we get that

$$\mathrm{P}(|V_{j,n} - W_{j,n}| > \varepsilon\sqrt{n}/r) \leq C_1\varepsilon^{-2/5}n^{-1/5}r^{2/5}k^{1/5}\alpha^{4/5}(k') \leq C_2\varepsilon^{-2/5}r^{1/5}\alpha^{4/5}(k').$$

Let $\triangle_n = S_{n,1} - n^{-1/2}\sum_{j=1}^{r} W_{j,n}$. Clearly,

$$\mathrm{P}\left(|\triangle_n| > \varepsilon\right) \leq \sum_{j=1}^{r}\mathrm{P}\left(\left|V_{j,n} - W_{j,n}\right| > \varepsilon\sqrt{n}/r\right) \leq C_3\varepsilon^{-2/5}r^{6/5}\rho^{4k'/5}. \tag{2.65}$$

By choosing $k' \sim n^c$ such that $c \in (0, 1/3)$, it is clear that the right hand side of (2.65) converges to 0 as $n \to \infty$. Hence,

$$\triangle_n \xrightarrow{\mathrm{p}} 0, \tag{2.66}$$

which means that

$$S_{n,1} = n^{-1/2}\sum_{j=1}^{r} W_{j,n} + o_p(1). \tag{2.67}$$

By virtue of this fact and Slutsky's Theorem, we only need to derive the asymptotic normality of $n^{-1/2}\sum_{j=1}^{r} W_{j,n}$, which will be a simpler task, as the $W_{j,n}$ are mutually independent.

Taking into account the construction of the $W_{j,n}$, $j = 1, ..., r$, applying Yokoyama's inequality, it follows that $\mathrm{E}(|W_{j,n}|^4) = \mathrm{E}(|V_{j,n}|^4) \leq C_1k^2$ and $\mathrm{Var}(W_{j,n}) = \mathrm{E}(V_{j,n}^2) = O(k)$. Thus, as $n \to \infty$,

$$\frac{\sum_{j=1}^{r}\mathrm{E}\left(\left|W_{j,n}\right|^4\right)}{(r\mathrm{Var}(W_{1,n}))^2} \leq \frac{C_3rk^2}{r^2k^2} \to 0.$$

This is Lyapounov's condition for the Central Limit Theorem. It allows us to conclude that

$$\frac{W_{1,n}+\ldots+W_{r,n}}{\sqrt{r\mathrm{Var}(W_{1,n})}} \xrightarrow{\mathrm{d}} N(0,1).$$

Using anlogous arguments to the ones presented in (2.63), it is easy to show that $\mathrm{Var}(W_{1,n}) = \mathrm{Var}(V_{1,n}) = k\sigma_0^2(p;k)/p^2\sigma_0^2(p;n)$. Consequently, by choosing $k \sim n^{1/3}$ (which goes in acordance with $rk \sim n$), we get that

$$n^{-1/2}p(W_{1,n}+\ldots+W_{r,n}) = \sqrt{\frac{rk}{n}\frac{\sigma_0^2(p;k)}{\sigma_0^2(p;n)}}\frac{W_{1,n}+\ldots+W_{r,n}}{\sqrt{\frac{rk}{p^2}\frac{\sigma_0^2(p;k)}{\sigma_0^2(p;n)}}} \xrightarrow{\mathrm{d}} N(0,1),$$

That is,

$$n^{-1/2}p\sum_{j=1}^{r} W_{j,n} \xrightarrow{\mathrm{d}} N(0,1). \tag{2.68}$$

Consequently, from (2.59), (2.64), (2.66), (2.67), (2.68) and Slutsky's Theorem,

$$n^{-1/2}p\sum_{i=1}^{n} T_{i,n} \xrightarrow{\mathrm{d}} N(0,1), \tag{2.69}$$

From (2.58) and (2.69) and one more application of Slutsky's Theorem we conclude that

$$\sqrt{n}p\sigma_0^{-1}(p;n)(\hat{\mu}_p-\mu_p) \xrightarrow{\mathrm{d}} N(0,1).$$

■

2.3 Convergence of the smoothed estimator

Theorem 14 (Chen, 2008, Central Limit Theorem for $\hat{\mu}_{p,h}$)**.** *Under conditions* (i) − (iv)*,*

$$\sqrt{n}p\sigma_0^{-1}(p;n)(\hat{\mu}_{p,h}-\mu_p) \xrightarrow{d} N(0,1).$$

Furthermore,

$$Bias(\hat{\mu}_{p,h}) = -\frac{1}{2}p^{-1}\sigma_K^2 h^2 f(\nu_p)+o(h^2), \tag{2.70}$$

$$Var(\hat{\mu}_{p,h}) = p^{-2}n^{-1}\sigma_0^2(p;n)+o(n^{-1}h). \tag{2.71}$$

Proof. First we will characterize the behaviour of $\mathrm{Bias}\left(\hat{\mu}_{p,h}\right)$ and $\mathrm{Var}\left(\hat{\mu}_{p,h}\right)$. From Chen and Tang (2005), $\hat{\nu}_{p,h}$ admits the expansion

$$\hat{\nu}_{p,h}-\nu_p = \frac{\hat{S}(\nu_p;h)-p}{f(\nu_p)}+o_p\left(n^{-1/2}\right), \tag{2.72}$$

and its bias satisfies

$$\text{Bias}\left(\hat{\nu}_{p,h}\right) = \text{E}\left(\hat{\nu}_{p,h}\right) - \nu_p = -\frac{1}{2}\sigma_K^2 f'(\nu_p) f^{-1}(\nu_p) h^2 + o\left(h^2\right). \tag{2.73}$$

Also from Chen and Tang (2005),

$$\hat{S}(\nu_p; h) - p = o_p(n^{-1/2}\log n). \tag{2.74}$$

Applying a Taylor expansion around ν_p

$$\begin{aligned}\hat{\mu}_{p,h} = {} & (np)^{-1}\sum_{t=1}^{n}\left[Y_t G_h(\nu_p - Y_t) - Y_t K_h(\nu_p - Y_t)\left(\hat{\nu}_{p,h} - \nu_p\right)\right] \\ & + \frac{1}{2}(np)^{-1} h^{-2}\sum_{t=1}^{n} Y_t K'\left(\frac{\nu_p + \theta\left(\hat{\nu}_{p,h} - \nu_p\right) - Y_t}{h}\right)\left(\hat{\nu}_{p,h} - \nu_p\right)^2,\end{aligned} \tag{2.75}$$

for some $\theta \in (0,1)$. It is not relevant in the context of this work, but it can be shown that $(2nph)^{-1}\sum_{t=1}^{n} Y_t K'\left[h^{-1}(\nu_p + \theta\left(\hat{\nu}_{p,h} - \nu_p\right) - Y_t)\right]$ converges to a constant ϕ, say. Taking into consideration (2.72), (2.74) and assumption (iv), it can be shown that (2.75) can be written as

$$\hat{\mu}_{p,h} = (np)^{-1}\sum_{t=1}^{n}\left[Y_t G_h(\nu_p - Y_t) - Y_t K_h(\nu_p - Y_t)\left(\hat{\nu}_{p,h} - \nu_p\right)\right] + o_p\left(h^2\right). \tag{2.76}$$

Before proceeding with the proof, we would like to make a note. For sake of simplicity, we will not exhibit the calculations of expectations, variances and covariances of random variables that involve O_p or o_p. For instance, it can be shown that under reasonable conditions, $E(o_p(n^{-1})) = o(n^{-1})$. Nonetheless, the computations we make already take these considerations into account.

Let us continue by computing $\text{E}\left(\hat{\mu}_{p,h}\right)$. Note that

$$\text{E}\left[(np)^{-1}\sum_{t=1}^{n} Y_t G_h(\nu_p - Y_t)\right] = p^{-1}\int z G_h(\nu_p - z) f(z)\,dz = p^{-1}\int\int_{\frac{\nu_p - z}{h}}^{\infty} K(u)\,du\,z f(z)\,dz. \tag{2.77}$$

Applying Fubini's theorem, (2.77) can be written as

$$p^{-1}\int K(u)\left(\int_{\nu_p}^{\infty} z f(z)\,dz + \int_{\nu_p - hu}^{\nu_p} z f(z)\,dz\right)du = \mu_p + p^{-1}\int K(u)\int_{\nu_p - hu}^{\nu_p} z f(z)\,dz\,du. \tag{2.78}$$

Applying Taylor's theorem once more, we get that $f(z) = f(\nu_p) + f'(\nu_p)(z - \nu_p) + \frac{f''(\theta)}{2}(z - \nu_p)^2$, for some $\theta \in (z, \nu_p)$. We use this expansion to compute $\int_{\nu_p - hu}^{\nu_p} z f(z)\,dz$. Taking into consideration that $\int uK(u)\,du = 0$ and $\int u^3 K(u)\,du = 0$, trivial calculations lead to

$$p^{-1}\int K(u)\int_{\nu_p - hu}^{\nu_p} z f(z)\,dz\,du = -\frac{1}{2}p^{-1}h^2\sigma_K^2\left(\nu_p f'(\nu_p) + f(\nu_p)\right) + o(h^3). \tag{2.79}$$

Combine (2.73), (2.78) and (2.79) to obtain

$$\mathrm{E}\left[(np)^{-1}\sum_{t=1}^{n}Y_tG_h(\nu_p-Y_t)\right]=\mu_p-\frac{1}{2}p^{-1}h^2\sigma_K^2\left(\nu_pf'(\nu_p)+f(\nu_p)\right)+o(h^3). \tag{2.80}$$

Using the same notation as in Lemma 3, let us now compute $\eta:=\mathrm{E}\left[(np)^{-1}\sum_{t=1}^{n}Y_tK_h\left(\nu_p-Y_t\right)\right]$. Clearly,

$$\begin{aligned}\eta&=p^{-1}\int h^{-1}K\left(\frac{\nu_p-z}{h}\right)zf(z)dz=p^{-1}\int(\nu_p-hu)K(u)f(\nu_p-hu)du\\&=p^{-1}\nu_p\int K(u)f(\nu_p-hu)du-p^{-1}h\int uK(u)f(\nu_p-hu)du\end{aligned} \tag{2.81}$$

Expand $f(\nu_p-hu)$ as $f(\nu_p)-huf'(\nu_p)+2^{-1}(hu)^2f''(\nu_p)+O(h^2)$ in the first integral of (2.81), and as $f(\nu_p)-huf'(\nu_p)+o(h)$ in the second to conclude that

$$\eta=p^{-1}\nu_pf(\nu_p)+O(h^2). \tag{2.82}$$

Recalling (2.72) and applying Davydov's inequality in an analogous way as we did in (2.25), it can be shown that

$$\mathrm{Cov}\left[(np)^{-1}\sum_{t=1}^{n}Y_tK_h\left(\nu_p-Y_t\right),\hat{\nu}_{p,h}-\nu_p\right]=o\left(h^2\right). \tag{2.83}$$

Hence, from (2.76), (2.82) and (2.83), it follows that

$$\begin{aligned}\mathrm{E}\left((np)^{-1}\sum_{t=1}^{n}Y_tK_h\left(\nu_p-Y_t\right)\left(\hat{\nu}_{p,h}-\nu_p\right)\right)&=\mathrm{E}\left((np)^{-1}\sum_{t=1}^{n}Y_tK_h\left(\nu_p-Y_t\right)\right)\mathrm{E}\left(\hat{\nu}_{p,h}-\nu_p\right)+o\left(h^2\right)\\&=-\frac{1}{2}p^{-1}\nu_pf'\left(\nu_p\right)h^2\sigma_K^2+o\left(h^2\right).\end{aligned} \tag{2.84}$$

Combining (2.76), (2.80) and (2.84), and noticing that $O\left(n^{-1}\right)=o\left(h^2\right)$ we conclude that

$$\mathrm{E}\left(\hat{\mu}_{p,h}\right)=\mu_p-\frac{1}{2}p^{-1}\sigma_K^2h^2f\left(\nu_p\right)+o\left(h^2\right),$$

which establishes the bias given in (2.70).

Let us now describe $\mathrm{Var}\left(\hat{\mu}_{p,h}\right)$. Let $A:=(np)^{-1}\sum_{t=1}^{n}\left[Y_tG_h\left(\nu_p-Y_t\right)-Y_tK_h\left(\nu_p-Y_t\right)\left(\hat{\nu}_{p,h}-\nu_p\right)\right]$ be the leading order term of the expansion (2.76). Trivially,

$$\begin{aligned}\mathrm{Var}\left(A\right)=\mathrm{Var}&\left[(np)^{-1}\sum_{t=1}^{n}Y_tG_h\left(\nu_p-Y_t\right)\right]+\mathrm{Var}\left[\hat{\eta}\left(\hat{\nu}_{p,h}-\nu_p\right)\right]\\&-2\mathrm{Cov}\left[(np)^{-1}\sum_{t=1}^{n}Y_tG_h\left(\nu_p-Y_t\right),\hat{\eta}\left(\hat{\nu}_{p,h}-\nu_p\right)\right].\end{aligned} \tag{2.85}$$

We shall first compute $\text{Var}\left[(np)^{-1}\sum_{t=1}^{n} Y_t G_h(\nu_p - Y_t)\right]$. Using the same arguments we presented in (2.60) and (2.63), it is clear that

$$\begin{aligned}
&\text{Var}\left[(np)^{-1}\sum_{t=1}^{n} Y_t G_h(\nu_p - Y_t)\right] \\
&= (np)^{-2}\left[n\text{Var}(Y_t G_h(\nu_p - Y_t)) + 2\sum_{1\le i<j\le n}\text{Cov}\left[Y_i G_h(\nu_p - Y_i), Y_j G_h(\nu_p - Y_j)\right]\right] \\
&= n^{-1}p^{-2}\left[\text{Var}(Y_t G_h(\nu_p - Y_t)) + 2\sum_{k=1}^{n-1}\left(1-\frac{k}{n}\right)\text{Cov}\left[Y_1 G_h(\nu_p - Y_1), Y_{k+1} G_h(\nu_p - Y_{k+1})\right]\right].
\end{aligned} \tag{2.86}$$

Let us consider $\text{Var}(Y_t G_h(\nu_p - Y_t))$. Applying Fubini's Theorem,

$$\begin{aligned}
\text{Var}(Y_t G_h(\nu_p - Y_t)) &= \text{E}(Y_t G_h(\nu_p - Y_t))^2 - \text{E}^2(Y_t G_h(\nu_p - Y_t)) \\
&= \int z^2 G_h^2(\nu_p - z) f(z)dz - p^2\mu_p^2 + O(h^2) \\
&= \int z^2 \int_{\frac{\nu_p - z}{h}}^{\infty} K(u)du \int_{\frac{\nu_p - z}{h}}^{\infty} K(v)dv f(z)dz - p^2\mu_p^2 + O(h^2).
\end{aligned} \tag{2.87}$$

To compute this integral, we use analogous arguments to the ones we presented in (2.49) to conclude that (2.87) can be written as

$$\begin{aligned}
&\int K(u)\left[\int_{-\infty}^{u} K(v)\left\{\int_{\nu_p}^{\infty} z^2 f(z)dz + \int_{\nu_p - hv}^{\nu_p} z^2 f(z)dz\right\}dv\right. \\
&\qquad \left. + \int_{u}^{\infty} K(v)\left\{\int_{\nu_p}^{\infty} z^2 f(z)dz + \int_{\nu_p - hu}^{\nu_p} z^2 f(z)dz\right\}dv\right]du - p^2\mu_p^2 + O(h^2).
\end{aligned} \tag{2.88}$$

It is clear that $\text{Var}(Y_t I(Y_t \ge \nu_p)) = \int_{\nu_p}^{\infty} z^2 f(z)dz + p^2\mu_p^2$. Putting together the appropriate integrals so that we may use the previous representation, (2.88) may be rewritten as

$$\begin{aligned}
&\int K(u)\left[\int_{-\infty}^{u} K(v)\int_{\nu_p - hv}^{\nu_p} z^2 f(z)dzdv + \int_{u}^{\infty} K(v)\left\{\int_{\nu_p - hu}^{\nu_p} z^2 f(z)dz\right\}dv\right]du \\
&\qquad + \text{Var}\{Y_t I(Y_t \ge \nu_p)\} + O(h^2).
\end{aligned} \tag{2.89}$$

Apply Taylor's theorem to rewrite $z^2 f(z)$ as $\nu_p^2 f(\nu_p) + \left(2\theta f(\theta) + \theta^2 f'(\theta)\right)(z - \nu_p)$, for some $\theta \in (z, \nu_p)$. This allows us to write (2.89) as

$$h\nu_p^2 f(\nu_p)\left(\int K(u)\int_{-\infty}^{u} vK(v)dvdu + \int uK(u)\int_{u}^{\infty} K(v)dvdu\right) + \text{Var}\{Y_t I(Y_t \ge \nu_p)\} + O\left(h^2\right). \tag{2.90}$$

Now set $c_K = \int uK(u)\int_{-\infty}^{u} K(v)dvdu$. One more application of Fubini's theorem allows us to realize that what is inside the parenthesis of (2.90) is simply $-2c_K$. Consequently,

$$\mathrm{Var}\left(Y_t G_h\left(\nu_p - Y_t\right)\right) = \mathrm{Var}\left(Y_t I\left(Y_t \geq \nu_p\right)\right) - 2h\nu_p^2 f\left(\nu_p\right)c_K + O\left(h^2\right). \tag{2.91}$$

To conclude the computation of $\mathrm{Var}\left[(np)^{-1}\sum_{t=1}^{n} Y_t G_h\left(\nu_p - Y_t\right)\right]$, note that, using the notation of $\phi_1(t)$ and $\phi_2(t)$,

$$\begin{aligned}\mathrm{Var}\left(\phi_1\left(\nu_p\right)\right) &= \mathrm{Var}\left(n^{-1}\sum_{t=1}^{n} Y_t I\left(Y_t \geq \nu_p\right)\right)\\ &= n^{-1}\left[\mathrm{Var}\left(I\left(Y_t \geq \nu_p\right)\right) + 2\sum_{k=1}^{n-1}\left(1-\frac{k}{n}\right)\mathrm{Cov}\left[Y_1 I\left(Y_1 \geq \nu_p\right), Y_{k+1} I\left(Y_{k+1} \geq \nu_p\right)\right]\right].\end{aligned} \tag{2.92}$$

That is,

$$\mathrm{Var}\left(I\left(Y_t \geq \nu_p\right)\right) = n\mathrm{Var}\left(\phi_1\left(\nu_p\right)\right) - 2\sum_{k=1}^{n-1}\left(1-\frac{k}{n}\right)\mathrm{Cov}\left[Y_1 I\left(Y_1 \geq \nu_p\right), Y_{k+1} I\left(Y_{k+1} \geq \nu_p\right)\right]. \tag{2.93}$$

Combining (2.86), (2.91) and (2.93) and applying Lemma 4, we deduce that

$$\mathrm{Var}\left[(np)^{-1}\sum_{t=1}^{n} Y_t G_h\left(\nu_p - Y_t\right)\right] = p^{-2}\mathrm{Var}\left(\phi_1\left(\nu_p\right)\right) - 2n^{-1}p^{-2}h\nu_p^2 f\left(\nu_p\right)c_K + o\left(n^{-1}h\right). \tag{2.94}$$

Now, the second term of (2.85) is

$$\begin{aligned}\mathrm{Var}\left[\hat{\eta}\left(\hat{\nu}_{p,h} - \nu_p\right)\right] &= \mathrm{Var}\left[\eta\left(\hat{\nu}_{p,h} - \nu_p\right) + \left(\hat{\eta} - \eta\right)\left(\hat{\nu}_{p,h} - \nu_p\right)\right]\\ &= \eta^2\mathrm{Var}\left(\hat{\nu}_{p,h}\right) + 2\eta\mathrm{Cov}\left[\hat{\nu}_{p,h} - \nu_p, \left(\hat{\eta} - \eta\right)\left(\hat{\nu}_{p,h} - \nu_p\right)\right]\\ &\quad + \mathrm{Var}\left[\left(\hat{\eta} - \eta\right)\left(\hat{\nu}_{p,h} - \nu_p\right)\right].\end{aligned} \tag{2.95}$$

From Chen and Tang (2005),

$$\mathrm{Var}\left(\hat{\nu}_{p,h}\right) = n^{-1}f^{-2}\left(\nu_p\right)\Delta^2\left(p;n\right) - 2n^{-1}hf^{-1}\left(\nu_p\right)c_K + o\left(n^{-1}h\right), \tag{2.96}$$

where

$$\Delta^2\left(p;n\right) = \left(p\left(1-p\right) + 2\sum_{k=1}^{n-1}\left(1-\frac{k}{n}\right)\mathrm{Cov}\left[I\left(Y_1 \geq \nu_p\right), I\left(Y_{k+1} \geq \nu_p\right)\right]\right).$$

Hence, from (2.82) and (2.96),

$$\begin{aligned}\eta^2\mathrm{Var}\left(\hat{\nu}_{p,h}\right) &= \left[p^{-1}\nu_p f\left(\nu_p\right) + O\left(h^2\right)\right]^2\left[n^{-1}f^{-2}\left(\nu_p\right)\Delta^2\left(p;n\right) - 2n^{-1}hf^{-1}\left(\nu_p\right)c_K + o\left(n^{-1}h\right)\right]\\ &= p^{-2}\nu_p^2 n^{-1}\Delta^2\left(p;n\right) - 2p^{-2}n^{-1}h\nu_p^2 f\left(\nu_p\right)c_K + o\left(n^{-1}h\right).\end{aligned} \tag{2.97}$$

Finally, taking into notice that

$$\begin{aligned}\mathrm{Var}(\phi_2(\nu_p)) &= \mathrm{Var}\left(n^{-1}\sum_{t=1}^{n} I(Y_t \geq \nu_p)\right)\\ &= n^{-2}\left[n\mathrm{Var}(I(Y_t \geq \nu_p)) + 2\sum_{k=1}^{n-1}(n-k)\,\mathrm{Cov}\left[I(Y_1 \geq \nu_p), I(Y_{k+1} \geq \nu_p)\right]\right]\\ &= n^{-1}\left[p(1-p) + 2\sum_{k=1}^{n-1}\left(1-\frac{k}{n}\right)\mathrm{Cov}\left[I(Y_1 \geq \nu_p), I(Y_{k+1} \geq \nu_p)\right]\right]\\ &= n^{-1}\Delta^2(p;n), \end{aligned} \tag{2.98}$$

allows us to write

$$\eta^2 \mathrm{Var}\left(\hat{\nu}_{p,h}\right) = p^{-2}\nu_p^2 \mathrm{Var}(\phi_2(\nu_p)) - 2p^{-2}n^{-1}h\nu_p^2 f(\nu_p)\, c_K + o\left(n^{-1}h\right). \tag{2.99}$$

On the other hand, taking into consideration (2.72), the second term of (2.95) is equal to

$$\begin{aligned}&2\eta \mathrm{Cov}\left[\hat{\nu}_{p,h} - \nu_p, (\hat{\eta} - \eta)\left(\hat{\nu}_{p,h} - \nu_p\right)\right]\\ &= 2\eta \mathrm{Cov}\left[\left(\hat{S}(\nu_p;h) - p\right) f^{-1}(\nu_p) + o_p\left(n^{-1/2}\right), (\hat{\eta} - \eta)\left(\left(\hat{S}(\nu_p;h) - p\right) f^{-1}(\nu_p) + o_p\left(n^{-1/2}\right)\right)\right]\\ &= 2\eta f^{-2}(\nu_p)\,\mathrm{Cov}\left[\hat{S}(\nu_p;h) - p, (\hat{\eta} - \eta)\left(\hat{S}(\nu_p;h) - p\right)\right] + o\left(n^{-1}h\right) \end{aligned} \tag{2.100}$$

which, by Lemma 3, is just $o\left(n^{-1}h\right)$.

To study the final term of (2.95), we apply Yokoyama's inequality. It allows us to write

$$\mathrm{E}\left(\hat{\nu}_{p,h} - \nu_p\right)^4 = O\left(n^{-2}\right) \quad \text{and} \quad \mathrm{E}(\hat{\eta} - \eta)^4 = O\left(n^{-2}h^{-4}\right).$$

Applying the Cauchy-Schwarz inequality three times,

$$\begin{aligned}\left|\mathrm{Var}\left((\hat{\eta} - \eta)\left(\hat{\nu}_{p,h} - \nu_p\right)\right)\right| &= \left|\mathrm{E}\left[(\hat{\eta} - \eta)^2\left(\hat{\nu}_{p,h} - \nu_p\right)^2\right] - \mathrm{E}^2\left[(\hat{\eta} - \eta)\left(\hat{\nu}_{p,h} - \nu_p\right)\right]\right|\\ &\leq \left|\mathrm{E}\left[(\hat{\eta} - \eta)^2\left(\hat{\nu}_{p,h} - \nu_p\right)^2\right]\right| + \left|\mathrm{E}^2\left[(\hat{\eta} - \eta)\left(\hat{\nu}_{p,h} - \nu_p\right)\right]\right|\\ &\leq \sqrt{\mathrm{E}(\hat{\eta} - \eta)^4\,\mathrm{E}\left(\hat{\nu}_{p,h} - \nu_p\right)^4} + \mathrm{E}(\hat{\eta} - \eta)^2\,\mathrm{E}\left(\hat{\nu}_{p,h} - \nu_p\right)^2\\ &\leq 2\sqrt{\mathrm{E}(\hat{\eta} - \eta)^4\,\mathrm{E}\left(\hat{\nu}_{p,h} - \nu_p\right)^4} = 2\sqrt{O(n^{-2}h^{-4})\,O(n^{-2})}\\ &= O\left(n^{-2}h^{-2}\right) = o\left(n^{-1}h\right). \end{aligned} \tag{2.101}$$

Combine (2.99), (2.100) and (2.101) to obtain

$$\mathrm{Var}\left[\hat{\eta}\left(\hat{\nu}_{p,h} - \nu_p\right)\right] = p^{-2}\nu_p^2 \mathrm{Var}(\phi_2(\nu_p)) - 2p^{-2}n^{-1}h\nu_p^2 f(\nu_p)\, c_K + o\left(n^{-1}h\right). \tag{2.102}$$

Now consider the final term of (2.85). Recall the definition of $\hat{\beta} = (np)^{-1}\sum_{t=1}^{n} Y_t G_h(\nu_p - Y_t)$. From (2.72) and an application of Lemma 3, it is clear that

$$\begin{aligned}\operatorname{Cov}\left[\hat{\beta}, \hat{\eta}(\hat{\nu}_{p,h}-\nu_p)\right] &= \operatorname{Cov}\left[\hat{\beta}, (\hat{\eta}-\eta)\left(\hat{\nu}_{p,h}-\nu_p\right)+\eta\left(\hat{\nu}_{p,h}-\nu_p\right)\right] \\ &= \operatorname{Cov}\left[\hat{\beta}, (\hat{\eta}-\eta)\left(\left(\hat{S}(\nu_p;h)-p\right)f^{-1}(\nu_p)+o_p\left(n^{-1/2}\right)\right)\right] \qquad (2.103)\\ &\quad+\operatorname{Cov}\left[\hat{\beta}, \eta\left(\left(\hat{S}(\nu_p;h)-p\right)f^{-1}(\nu_p)+o_p\left(n^{-1/2}\right)\right)\right] \\ &= f^{-1}(\nu_p)\operatorname{Cov}\left[\hat{\beta}, (\hat{\eta}-\eta)\left(\hat{S}(\nu_p;h)-p\right)\right]+\operatorname{Cov}\left[\hat{\beta}, (\hat{\eta}-\eta)\,o_p\left(n^{-1/2}\right)\right] \\ &\quad+\operatorname{Cov}\left[\hat{\beta}, \eta\left(\hat{S}(\nu_p;h)-p\right)f^{-1}(\nu_p)\right]+\operatorname{Cov}\left[\hat{\beta}, \eta\, o_p\left(n^{-1/2}\right)\right] \\ &= \operatorname{Cov}\left[\hat{\beta}, \eta f^{-1}(\nu_p)\hat{S}(\nu_p;h)\right]+o\left(n^{-1}h\right). \qquad (2.104)\end{aligned}$$

From (2.82), it is clear that

$$\begin{aligned}&\operatorname{Cov}\left[\hat{\beta}, \eta f^{-1}(\nu_p)\hat{S}(\nu_p;h)\right] \\ &= \operatorname{Cov}\left[(np)^{-1}\sum_{t=1}^{n} Y_t G_h(\nu_p - Y_t), \left(p^{-1}\nu_p f(\nu_p)+O\left(h^2\right)\right)f^{-1}(\nu_p)n^{-1}\sum_{t=1}^{n} G_h(\nu_p - Y_t)\right] \\ &= \left[(np)^{-2}\nu_p + n^{-2}O\left(h^2\right)\right]\operatorname{Cov}\left[\sum_{t=1}^{n} Y_t G_h(\nu_p - Y_t), \sum_{t=1}^{n} G_h(\nu_p - Y_t)\right] \qquad (2.105)\end{aligned}$$

Similarly to what was done in (2.26) or in (2.38), it may be shown that

$$n^{-2}O\left(h^2\right)\operatorname{Cov}\left[\sum_{t=1}^{n} Y_t G_h(\nu_p - Y_t), \sum_{t=1}^{n} G_h(\nu_p - Y_t)\right] = o\left(n^{-1}h\right).$$

This means that (2.105) can be written as

$$\operatorname{Cov}\left[\hat{\beta}, \eta f^{-1}(\nu_p)\hat{S}(\nu_p;h)\right] = (np)^{-2}\nu_p \operatorname{Cov}\left[\sum_{t=1}^{n} Y_t G_h(\nu_p - Y_t), \sum_{t=1}^{n} G_h(\nu_p - Y_t)\right] + o\left(n^{-1}h\right). \qquad (2.106)$$

Now note that

$$\begin{aligned}&\operatorname{Cov}\left[\sum_{t=1}^{n} Y_t G_h(\nu_p - Y_t), \sum_{t=1}^{n} G_h(\nu_p - Y_t)\right] \\ &= n\left[\operatorname{Cov}\left[Y_t G_h(\nu_p - Y_t), G_h(\nu_p - Y_t)\right] + 2\sum_{t=1}^{n-1}\left(1-\frac{k}{n}\right)\operatorname{Cov}\left[Y_1 G_h(\nu_p - Y_1), G_h(\nu_p - Y_{k+1})\right]\right]. \qquad (2.107)\end{aligned}$$

Let us first compute $\operatorname{Cov}\left[Y_t G_h(\nu_p - Y_t), G_h(\nu_p - Y_t)\right]$. Clearly,

$$\mathrm{E}(Y_t G_h(\nu_p - Y_t)) = p\mu_p + O(h^2) \qquad \text{and} \qquad \mathrm{E}(G_h(\nu_p - Y_t)) = p + O(h^2).$$

Taking this fact into consideration, and using the same arguments that were presented in the computation of $\mathrm{Var}(Y_t G_h(\nu_p - Y_t))$, it is clear that

$$\begin{aligned}
\mathrm{Cov}\left[Y_t G_h(\nu_p - Y_t), G_h(\nu_p - Y_t)\right] &= \mathrm{E}\left(Y_t G_h^2(\nu_p - Y_t)\right) - p^2\mu_p + O\left(h^2\right) \\
&= \int z^2 \int_{\frac{\nu_p - z}{h}}^{\infty} K(u)\,du \int_{\frac{\nu_p - z}{h}}^{\infty} K(v)\,dv f(z)\,dz - p^2\mu_p + O\left(h^2\right) \\
&= \int K(u)\left[\int_{-\infty}^{u} K(v)\left\{\int_{\nu_p}^{\infty} zf(z)\,dz + \int_{\nu_p - hv}^{\nu_p} zf(z)\,dz\right\}dv\right. \\
&\quad \left. + \int_{u}^{\infty} K(v)\left\{\int_{\nu_p}^{\infty} zf(z)\,dz + \int_{\nu_p - hu}^{\nu_p} zf(z)\,dz\right\}dv\right]du \\
&\quad - p^2\mu_p + O\left(h^2\right) \qquad (2.108) \\
&= p(1-p)\mu_p - 2\nu_p f(\nu_p) h c_K + o(h). \qquad (2.109)
\end{aligned}$$

Now, notice that

$$\begin{aligned}
\mathrm{Cov}\left[\phi_1(\nu_p), \phi_2(\nu_p)\right] &= n^{-2}\mathrm{Cov}\left[\sum_{t=1}^{n} Y_t I(Y_t \geq \nu_p), \sum_{t=1}^{n} I(Y_t \geq \nu_p)\right] \\
&= n^{-1}\left[\mathrm{Cov}\left[Y_t I(Y_t \geq \nu_p), I(Y_t \geq \nu_p)\right] + 2\sum_{k=1}^{n-1}\left(1 - \frac{k}{n}\right)\mathrm{Cov}\left[Y_1 I(Y_1 \geq \nu_p), I(Y_{k+1} \geq \nu_p)\right]\right] \\
&= n^{-1}\left[p(1-p)\mu_p + 2\sum_{k=1}^{n-1}\left(1 - \frac{k}{n}\right)\mathrm{Cov}\left[Y_1 I(Y_1 \geq \nu_p), I(Y_{k+1} \geq \nu_p)\right]\right]. \qquad (2.110)
\end{aligned}$$

Plug (2.109) into (2.107), and take into consideration (2.110) and Lemma 4 to conlude that

$$\begin{aligned}
&\mathrm{Cov}\left[\hat{\beta}, \eta f^{-1}(\nu_p)\hat{S}(\nu_p; h)\right] = \\
&= \left(np^2\right)^{-1}\nu_p\left[p(1-p)\mu_p - 2\nu_p f(\nu_p) h c_K + o(h)\right. \\
&\quad + 2\sum_{k=1}^{n-1}\left(1 - \frac{k}{n}\right)\mathrm{Cov}\left[Y_1 G_h(\nu_p - Y_1), G_h(\nu_p - Y_{k+1})\right] \\
&\quad - 2\sum_{k=1}^{n-1}\left(1 - \frac{k}{n}\right)\mathrm{Cov}\left[Y_1 I(Y_1 \geq \nu_p), I(Y_{k+1} \geq \nu_p)\right] \\
&\quad \left. + 2\sum_{k=1}^{n-1}\left(1 - \frac{k}{n}\right)\mathrm{Cov}\left[Y_1 I(Y_1 \geq \nu_p), I(Y_{k+1} \geq \nu_p)\right]\right] \\
&\quad + o\left(n^{-1}h\right) \\
&= p^{-2}\nu_p \mathrm{Cov}\left[\phi_1(\nu_p), \phi_2(\nu_p)\right] - 2n^{-1}p^{-2}\nu_p^2 f(\nu_p) h c_K + o\left(n^{-1}h\right). \qquad (2.111)
\end{aligned}$$

Combine (2.104) and (2.111) to conclude that

$$\operatorname{Cov}\left[(np)^{-1}\sum_{t=1}^{n} Y_t G_h(\nu_p - Y_t), \hat{\eta}\left(\hat{\nu}_{p,h} - \nu_p\right)\right]$$
$$= p^{-2}\nu_p \operatorname{Cov}\left[\phi_1(\nu_p), \phi_2(\nu_p)\right] - 2n^{-1}p^{-2}\nu_p^2 f(\nu_p) h c_K + o\left(n^{-1}h\right). \qquad (2.112)$$

Finally, substituting (2.94), (2.102) and (2.112) in (2.85), we note that the terms that are $O(n^{-1}h)$ cancel and lead us to

$$\operatorname{Var}(A) = p^{-2}\operatorname{Var}(\phi_1(\nu_p)) + p^{-2}\nu_p^2 \operatorname{Var}(\phi_2(\nu_p)) - 2p^{-2}\nu_p \operatorname{Cov}(\phi_1(\nu_p), \phi_2(\nu_p)) + o(n^{-1}h). \quad (2.113)$$

Recalling (2.92), (2.98) and (2.110) into (2.113), simple covariance computations lead to

$$\operatorname{Var}(A) = n^{-1}p^{-2}\Bigg[\operatorname{Var}(Y_t I(Y_t \geq \nu_p)) + \nu_p^2 p(1-p) - 2\nu_p p(1-p)\mu_p$$
$$+ 2\sum_{t=1}^{n-1}\left(1 - \frac{k}{n}\right)\operatorname{Cov}((Y_1 - \nu_p)I(Y_1 \geq \nu_p), (Y_{k+1} - \nu_p)I(Y_{k+1} \geq \nu_p))\Bigg] + o(n^{-1}h).$$
$$(2.114)$$

By adequately adding and subtracting ν_p, it can be easily shown that $\operatorname{Var}[Y_t I(Y_t \geq \nu_p)] = \operatorname{Var}((Y_t - \nu_p)I(Y_t \geq \nu_p)) - \nu_p^2 p(1-p) + 2\nu_p p(1-p)\mu_p$. This fact in combination with (2.114) lets us conclude that

$$\operatorname{Var}(\hat{\mu}_{p,h}) = n^{-1}p^{-2}\sigma_0^2(p;n) + o(n^{-1}h),$$

which establishes (2.71).

We shall now prove the asymptotic normality of $\hat{\mu}_{p,h}$, much like was done in the previous theorem. Let $A_i := Y_i G_h(\nu_p - Y_i) - Y_i K_h(\nu_p - Y_i)(\hat{\nu}_{p,h} - \nu_p)$. From (2.76) and (2.70) we may write

$$\sigma_0^{-1}(p;n)(\hat{\mu}_{p,h} - \mu_p) = n^{-1}\sum_{i=1}^{n} T_{i,n} + o_P(h), \qquad (2.115)$$

where $T_{i,n} = \sigma_0^{-1}(p;n)p^{-1}\left[A_i - \operatorname{E}(A_i)\right]$. We let k, k' and r be three sequences of positive integers satisfying the same assumptions as those presented in the previous theorem, and we define the same large, small and residual blocks as before. Setting

$$S_n := n^{-1/2}\sum_{i=1}^{n} T_{i,n} = n^{-1/2}\sum_{j=1}^{r} V_{j,n} + n^{-1/2}\sum_{j=1}^{r} V'_{j,n} + n^{-1/2}\delta_n =: S_{n,1} + S_{n,2} + S_{n,3}, \qquad (2.116)$$

it is clear that $\operatorname{E}(S_{n,2}) = \operatorname{E}(S_{n,3}) = 0$. Moreover,

$$\operatorname{Var}(S_{n,3}) = n^{-1}p^{-2}\sigma_0^{-2}(p;n)\operatorname{Var}\left[\sum_{i=r(k+k')+1}^{n}\Big(Y_i G_h(\nu_p - Y_i) - Y_i K_h(\nu_p - Y_i)\left(\hat{\nu}_{p,h} - \nu_p\right)\Big)\right],$$

and using the same arguments that led us to (2.71), we conclude that

$$\mathrm{Var}(S_{n,3}) = \frac{(n-r(k+k')\left[\sigma_0^2(p;n-r(k+k'))+o(n^{-1}h)\right]}{np^2\sigma_0^2(p;n)} \to 0. \tag{2.117}$$

Using the above arguments and the argument presented in previous theorem to justify that the covariances of variables of different blocks tend to zero, it is clear that

$$\mathrm{Var}(S_{n,2}) = \frac{rk'\left[\sigma_0^2(p;k')+o(n^{-1}h)+o(1)\right]}{np^2\sigma_0^2(p;n)} \to 0. \tag{2.118}$$

Theorem 2 lets us conclude that

$$S_{n,2} \xrightarrow{\mathrm{p}} 0 \qquad \text{and} \qquad S_{n,3} \xrightarrow{\mathrm{p}} 0. \tag{2.119}$$

The proof of the asymptotic normality of $S_{n,1}$ is similar. By choosing $r \sim n^{2/3}$ (for instance, $r = [n^{2/3}]$), $k' \sim n^c$ such that $c \in (0, 2/3)$, $k \sim n^{1/3}$, and setting $\triangle_n = S_{n,1} - n^{-1/2}\sum_{j=1}^{r} W_{j,n}$, recursive applications of Bradley's lemma allow us to conclude that

$$S_{n,1} = n^{-1/2}\sum_{j=1}^{r} W_{j,n} + o_p(1). \tag{2.120}$$

Yokoyama's inequality leads to

$$\frac{\sum_{j=1}^{r} \mathrm{E}\left(\left|W_{j,n}\right|^4\right)}{(r\mathrm{Var}(W_{1,n}))^2} \le \frac{C_3 rk^2}{r^2k^2} \to 0,$$

which is Lyapounov's condition for the Central Limit Theorem. Using the arguments that lead us to (2.118), it is easy to show that $\mathrm{Var}(W_{1,n}) = \mathrm{Var}(V_{1,n}) = \frac{k\left[\sigma_0^2(p;k)+o(n^{-1}h)\right]}{p^2\sigma_0^2(p;n)}$, which implies that

$$n^{-1/2}p(W_{1,n}+\ldots+W_{r,n}) = \sqrt{\frac{rk}{n}\frac{\left[\sigma_0^2(p;k)+o(n^{-1}h)\right]}{\sigma_0^2(p;n)}}\,\frac{W_{1,n}+\ldots+W_{r,n}}{\sqrt{\frac{rk}{p^2}\frac{\left[\sigma_0^2(p;k)+o(n^{-1}h)\right]}{\sigma_0^2(p;n)}}} \xrightarrow{\mathrm{d}} N(0,1),$$

That is,

$$n^{-1/2}p\sum_{j=1}^{r} W_{j,n} \xrightarrow{\mathrm{d}} N(0,1). \tag{2.121}$$

Combine (2.115), (2.116), (2.119), (2.120), (2.121) and apply Slutsky's Theorem to conclude that

$$\sqrt{n}p\sigma_0^{-1}(p;n)(\hat{\mu}_{p,h} - \mu_p) \xrightarrow{\mathrm{d}} N(0,1).$$

■

Chapter 3

Conclusion

3.1 Results Analysis

In this section we revise and comment the main results of this work. Let us start with the unsmoothed estimator, $\hat{\mu}_p$. We showed that

Theorem 1 (Chen, 2008) *Under conditions* (i) – (ii),

$$\sqrt{n}p\sigma_0^{-1}(p;n)(\hat{\mu}_p - \mu_p) \xrightarrow{\text{d}} N(0,1).$$

This theorem indicates that the asymptotic variance of $\hat{\mu}_p$ is $(np^2)^{-1}\sigma_0^2(p;n)$. As we noted in Remark 5, $\lim_{n\to\infty} \sigma_0^2(p;n)$ is finite. We note that the dependence in the original time series is reflected in the asymptotic variance through the covariance terms in $\sigma_0^2(p;n)$. This knowledge is of crucial importance if one wishes to do further statistical inference (such as confidence intervals estimation and hypothesis testing) concerning expected shortfall estimation, as it would be an error (financially speaking) to ignore this term. We also note that the effective sample size for the ES estimation is only np^2. As we mentioned before, financial risk management is generally concerned with values of p that range from 0.01 to 0.05, which is a reason why ES estimation is subject to high volatility. This is a general challenge for statistic inference of risk measures.

For the kernel estimator, $\hat{\mu}_{p,h}$, we proved the following theorem.

Theorem 2 (Chen, 2008) *Under conditions* (i) – (iv),

$$\sqrt{n}p\sigma_0^{-1}(p;n)(\hat{\mu}_{p,h} - \mu_p) \xrightarrow{\text{d}} N(0,1).$$

Furthermore,

$$Bias(\hat{\mu}_{p,h}) = -\frac{1}{2}p^{-1}\sigma_K^2 h^2 f(\nu_p) + o(h^2),$$
$$Var(\hat{\mu}_{p,h}) = p^{-2}n^{-1}\sigma_0^2(p;n) + o(n^{-1}h).$$

Comparing both theorems we find that both ES estimators have the same asymptotic distribution. We also note that, unlike the VaR estimation, the kernel estimator does not grant a variance reduction, at least up to the term of order $n^{-1}h$, as the term of this order vanishes, instead of taking a negative value. At the same time, the kernel estimator brings in a bias, which leads to an overall increase in the mean square error. Therefore, for the purpose of ES estimation, the kernel smoothing is counter-productive. One of the possible reasons why kernel smoothing is not efficient in this context is the fact that the ES is effectively a mean parameter, which can be estimated accurately by simple averaging.

It should also be noted that the above statement is only applicable for point estimation of ES. For constructing confidence intervals and hypothesis testing on μ_p in the presence of data dependence, the kernel smoothing can play a significant role in estimating $\sigma_0^2(p;n)$ via the spectral density estimation approach. This approach is presented in Chen and Tang (2005), but it serves the purpose of estimating another important quantity that shows up in VaR estimation, but plays the same roll as $\sigma_0^2(p;n)$.

3.2 Future Work

Our work was based on the assumption that the underlying stochastic process is α-mixing. There are several other interesting ways to model dependence, such as association (see Esary, Proschan and Walkup, (1967)) or a stronger type of mixing, that can be used as a premise to study the asymptotic statistical properties of both presented estimators.

References

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3):203–228.

Bellini, F. and Figà-Talamanca, G. (2004). Detecting and modeling tail dependence. *International Journal of Theoretical and Applied Finance*, 07(03):269–287.

Billingsley, P. (1968). *Convergence of probability measures*. Wiley, New York.

Bishop, Y., Fienberg, S., and Holland, P. (1975). *Discrete multivariate analysis*. MIT Press, Cambridge, Mass. [u.a.].

Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes : Estimation and Prediction*. Springer New York, New York, NY.

Cai, Z. and Roussas, G. G. (1998). Efficient estimation of a distribution function under quadrant dependence. *Scandinavian Journal of Statistics*, 25(1):211–224.

Carrasco, M. and Chen, X. (2002). Mixing and moment properties of various garch and stochastic volatility models. *Econometric Theory*, 18(1):17–39.

Chen, S. X. (2008). Nonparametric estimation of expected shortfall. *Journal of Financial Econometrics*, 6(1):87.

Chen, S. X. and Tang, C. Y. (2005). Nonparametric inference of value-at-risk for dependent financial returns. *Journal of Financial Econometrics*, 3(2):227.

Dedecker, J., Doukhan, P., and Lang, G. (2007). *Weak dependence : with examples and applications*. Lecture notes in statistics. Springer, New York.

Doukhan, P. (1994). *Mixing : Properties and Examples*. Springer New York, New York, NY.

Embrechts, P. (1997). *Modelling extremal events for insurance and finance*. Springer-Verlag, Berlin.

Esary, J. D., Proschan, F., and Walkup, D. W. (1967). Association random variables, with applications. *The Annals of Mathematical Statistics*, 38:1466–1474.

Genon-Catalot, V., Jeantheau, T., and Larédo, C. (2000). Stochastic volatility models as hidden markov models and statistical applications. *Bernoulli*, 6(6):1051–1079.

Gourieroux, C., Laurent, J., and Scaillet, O. (2000). Sensitivity analysis of values at risk. *Journal of Empirical Finance*, 7(3–4):225 – 245. Special issue on Risk Management.

Masry, E. and Tjøstheim, D. (1995). Nonparametric estimation and identification of nonlinear arch time series: Strong convergence and asymptotic normality. *Econometric Theory*, 11(2):258–289.

Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.*, 33(3):1065–1076.

Pham, T. D. and Tran, L. T. (1985). Some mixing properties of time series models. *Stochastic Processes and their Applications*, 19(2):297 – 303.

Rosenblatt, M. (1956a). A central limit theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences of the United States of America*, 42(1):43–47.

Rosenblatt, M. (1956b). Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, 27(3):832–837.

Scaillet, O. (2004). Nonparametric estimation and sensitivity analysis of expected shortfall. *Mathematical Finance*, 14(1):115–129.

Scott, D. W. (1979). On optimal and data-based histograms. *Biometrika*, 66(3):605–610.

Wand, M. (1995). *Kernel smoothing*. Chapman and Hall, London.

Yokoyama, R. (1980). Moment bounds for stationary mixing sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 52(1):45–57.

Yoshihara, K. (1995). The bahadur representation of sample quantiles for sequences of strongly mixing random variables. *Statistics & Probability Letters*, 24(4):299 – 304.