Kink estimation in stochastic regression with dependent errors and predictors[*]

Justin Wishart

School of Mathematics and Statistics, F07
University of Sydney NSW 2006
Australia
e-mail: justin.wishart@sydney.edu.au

Rafał Kulik[†]

University of Ottawa
e-mail: rkulik@uottawa.ca

Abstract: In this article we study the estimation of the location of jump points in the first derivative (referred to as kinks) of a regression function μ in two random design models with different long-range dependent (LRD) structures. The method is based on the zero-crossing technique and makes use of high-order kernels. The rate of convergence of the estimator is contingent on the level of dependence and the smoothness of the regression function μ. In one of the models, the convergence rate is the same as the minimax rate for kink estimation in the fixed design scenario with i.i.d. errors which suggests that the method is optimal in the minimax sense.

AMS 2000 subject classifications: Primary 62G08; secondary 62G05, 62G20.
Keywords and phrases: Change point, Kink, High-order kernel, Zero-crossing technique, Long-range dependence, Random design, Separation rate lemma.

Contents

1. Introduction

Assume that we observe a bivariate dataset $\{X_i, Y_i\}_{i=1}^n$ that follows the regression model,

$$Y_i = \mu(X_i) + \sigma(X_i)\varepsilon_i, \tag{1}$$

where μ is the regression function and σ is a deterministic scale function. Also, ε_i and X_i are the error and random design variables respectively (both being possibly long-range dependent) and X_i has cumulative distribution function $F = F_X : \mathbb{R} \longrightarrow [0,1]$ that is strictly increasing.

We are interested in testing the presence of a change point in the slope of a regression function μ and if one exists, estimating its location. We describe this jump in the first derivative of μ as a kink and denote the change point by θ. Knowledge of this change point will allow us to identify change in trends in the underlying regression function of a non-parametric model. This could explain the change in qualitative or quantitative behaviour of an underlying process.

1.1. Existing results

Before examining the kink estimation under the random design regression model (1), we first look at other non-parametric and parametric models and their link to the existing theory for kink point estimation. A change point estimation technique was pioneered by Goldenshluger, Tsybakov and Zeevi (2006) for estimating change points in the regression function itself, not the kink scenario. The underlying model assumed for their framework was the indirect model with fixed design. The indirect model assumes that the regression function is not observed in practice but a so called 'blurred' version of the regression function is observed whereby the regression function has been transformed by a convolution operator. More specifically, the indirect model assumes that observations are realisations of the asymptotic model,

$$dY(x) = K\,\mu(x)\,dx + \epsilon dB(x). \tag{2}$$

In the above model the function $K\,\mu(x) = \int_{\mathbb{R}} K(t-x)\mu(x)\,dx$ represents the convolution of μ and K and the noise is driven by a regular Brownian motion, $B(x)$ and controlled by $\epsilon \asymp n^{-\frac{1}{2}}$ where the statement $a_n \asymp b_n$ means that the ratio a_n/b_n is bounded above and below by positive constants. The fixed design implies that the design variables $x_i = \frac{i}{n}$ are equally spaced points on the unit

interval. The asymptotic model (2) is considered due to a result by Brown and Low (1996) that shows (2) is asymptotically equivalent to the model,

$$Y_i = K \mu(x_i) + z_i, \tag{3}$$

where z_i is an i.i.d. sequence of error variables.

The specific estimation technique that Goldenshluger, Tsybakov and Zeevi (2006) formulated was the zero-crossing technique and it used a particular class of kernel functions to identify the change point. Their technique will be adapted for use in this article and is pursued in further detail in Section 4.1. At this stage it will suffice to say that the main result of their paper established that the zero-crossing technique is optimal in the minimax sense under the framework given in (2).

The zero-crossing technique has been applied by Cheng and Raimondo (2008) to estimate a kink instead of a jump point and was done in the direct model in the fixed design setting. In this framework the observations are assumed to follow a fixed design and realisations derived from the following asymptotic model,

$$dY(x) = \mu(x)dx + \epsilon dB(x). \tag{4}$$

Model (3) and its asymptotic equivalent is usually appropriate in practice when a variable is observed at regular intervals indexed by time and the errors are i.i.d. homoscedastic random variables.

More recently, Wishart (2009) extended the technique further to include long-range dependent (LRD) noise observations instead of independent noise. The kink estimation technique was extended to include the model,

$$dY(x) = \mu(x)dx + \epsilon^{\alpha} dB_H(x), \tag{5}$$

where $B_H(x)$ is a fractional Brownian motion with self-similarity index $H \in [\frac{1}{2}, 1)$. The noise process was normalised by ϵ^{α} where $\alpha = 2 - 2H$. Wang (1996) has shown that Model (5) is the asymptotic equivalent to the discrete model,

$$y_i = \mu(x_i) + e_i, \tag{6}$$

where e_i is a LRD sequence of random variables.

In this paper we are interested in model (1), which extends the fixed design cases given in models (3) and (6) above. They are extended in the sense that the design points are no longer restricted to a uniform grid of points and the scale function $\sigma(\cdot)$ allows heteroscedasticity for the error terms in the regression model. The analysis of this random design model needs to be considered quite carefully, since the asymptotic behaviour of the estimators will depend on the behaviour of the scale function and on the level of dependence present in the design variables and errors themselves. It has been shown by Reiß (2008) that there exists an asymptotic equivalence between model (1) and (4) when $\sigma(\cdot) \equiv$ constant, and the design variables are independent uniform random variables. However, this is not the case in general. As noted in Kulik and Raimondo (2009a), with LRD design variables, model (1) cannot be equivalent to any asymptotic model,

which is in contrast to model (5) being the asymptotic equivalent to model (6) in the fixed design case.

There is an extensive treatment in the literature on both parametric and non-parametric methods for regression models with a random design framework that assume i.i.d. design and error variables. The methodologies used include, but are not limited to, kernel smoothing, wavelet decompositions and orthogonal series. The methods of change point estimation for the random design case have been considered in Gijbels, Hall and Kneip (1999); Huh and Park (2004); Korostelëv and Tsybakov (1993)

There is also literature on the fixed design scenario in the presence of long-range dependent errors and the introduction of dependence in the errors always has a detrimental effect on estimation in this scenario. In the context of function estimation some recent treatments of this topic include Cavalier (2004); Csörgő and Mielniczuk (1995); Johnstone (1999); Johnstone and Silverman (1997); Kulik and Raimondo (2009a); Wang (1996). For change point estimation work has been done by Wang (1999); Wishart (2009).

Then there is a new emerging literature that attempts to combine the two scenarios with random design regression models where the design variables and/or the error variables are LRD. When the framework includes a random design and possibly LRD variables there is a more subtle asymptotic theory that is based on a delicate balance between the behaviour of the σ function and the level of dependence present. This is evident in a number of papers in the area and will be the case here as well. The interested reader is referred to work by Guo and Koul (2008); Robinson and Hidalgo (1997) for a parametric linear model approach in this context and to Csörgő and Mielniczuk (1999); Kulik and Raimondo (2009b); Mielniczuk and Wu (2004); Yang (2001) for regression estimation in a non-parametric framework. Finally some studies to estimate change points in the non-parametric context include Lin, Li and Chen (2008); Wang (2008).

1.2. Article outline

Some preliminary framework is outlined in Section 2, setting up the class of functions that are considered and specific dependence assumptions made in the random design model. The main result of the paper is described in Section 3, along with a brief discussion. The estimation method is explained in detail in Section 4, with a brief outline of the zero-crossing technique in the fixed design and its extension to the random design case. All the necessary proofs of the results are given in Section 5.

2. Preliminaries

2.1. Smoothness assumptions and kernels

First we look at the smoothness of the regression function μ and the properties of the kernel function that was constructed to use the zero-crossing technique by

Cheng and Raimondo (2008). We define a class of functions that have domain $\mathcal{X} \subseteq \mathbb{R}$, a kink at $\theta \in \mathcal{X}$ and $s \geq 3$ derivatives that exist in the neighbourhood of θ.

Definition 1. *We say that* $\mu \in \mathscr{F}_s(\mathcal{X}, \theta)$ *if,*

1. $\mu : \mathcal{X} \longrightarrow \mathbb{R}$
2. μ *has a kink, that is, there exists a* $\theta \in \mathcal{X}$ *and* $a_\mu \in \mathbb{R}$ *with* $a_\mu \neq 0$ *such that,*

$$[\mu^{(1)}](\theta) = \mu^{(1)}(\theta_+) - \mu^{(1)}(\theta_-) = a_\mu,$$

where $\mu^{(1)}(\theta_+)$ *and* $\mu^{(1)}(\theta_-)$ *are the right and left first derivatives of* μ *respectively.*
3. *The higher order derivatives* $\mu^{(i)}$ *exist and are finite everywhere and satisfy,*

$$\mu^{(i)}(\theta_+) = \mu^{(i)}(\theta_-) \qquad for \ i = 2,3,\ldots,s\text{-}1. \tag{7}$$

4. *For all* $x_+ \in (0, \sup \mathcal{X} - \theta)$ *and* $x_- \in (\inf \mathcal{X} - \theta, 0)$,

$$\mu^{(1)}(\theta_\pm + x_\pm) - \mu^{(1)}(\theta_\pm) = \sum_{i=1}^{s-2} \frac{x_\pm^i \mu^{(i+1)}(\theta_\pm)}{i!} + \mathcal{O}(x_\pm^{s-1}). \tag{8}$$

Condition *4.* should be interpreted in the sense that $\mu^{(1)}$ has a separate Taylor expansion for points to the left and right of θ respectively. Condition *3.* of Definition 1 might seem overly restrictive but is required to exploit the class of kernel functions that are introduced later in this Section. We will also denote $\mathscr{F}_s(\theta) = \mathscr{F}_s(\mathbb{R}, \theta)$. For completeness and comparison purposes we will also introduce another smoothness class $\mathscr{G}_s$ to denote the class of functions that do not have a kink. This class is identical to $\mathscr{F}_s(\theta)$ except conditions *2.* and *3.* are relaxed in Definition 1 in the sense that there does not exist a $\theta \in \mathcal{X}$ such that, $[\mu^{(1)}](\theta) \neq 0$.

In the fixed design setting, we can assume that the domain of the regression function is $[0, 1]$ since any finite interval, $[a, b]$ can be mapped to the $[0, 1]$ interval by an affine transformation. However this assumption is not always valid in the general random design case. In particular, if the design variables are LRD then it is required that they have a domain across the whole real line.

To use the zero-crossing technique for this class of regression functions Cheng and Raimondo (2008) constructed a class of kernel functions via Legendre polynomials and we will denote this class of functions by $\mathscr{K}_s$. The full description of the zero-crossing technique and the consequent technical details required of the kernel functions are not covered here and the reader is referred to Goldenshluger, Tsybakov and Zeevi (2006) and Cheng and Raimondo (2008) respectively for full treatment. However, some key aspects will be given and for our case we will say $K \in \mathscr{K}_s$, where $s = 2k + 1$ and $k \in \mathbb{Z}^+$ if,

$$K(x) = K(k, x) = a_k \sum_{j=k-1}^{2k+2} b_{j,k} x^{2j-2k+2} \mathbb{1}_{[-1,1]}(x) \,,$$

where the polynomial coefficients are defined by

$$a_k := \frac{(4k+5)!}{2^{4k+5}(2k)!(2k+2)!} \,, \quad b_{j,k} := \frac{(-1)^{k+j+1}(2j)!}{j!(2k-j+2)!(2j-2k+2)!}.$$

This class of kernel functions is indexed by the level of smoothness s and is constructed to exploit the extra smoothness of the class $\mathscr{F}_s(\theta)$. To save on notation we denote $K_i = K^{(i)}$, to represent the ith order derivative of K. The kernels have the following properties:

$$K_i(-1) = K_i(1) = 0 \quad \text{for } i = 1, 2, 3. \quad \text{and} \quad K_1(0) = 0. \tag{9}$$

$$\int_{-1}^{1} u^j K_3(u) \, du = 0 \,, \quad j = 0, 1, \ldots, 2k. \tag{10}$$

Property (10) of $\mathscr{K}_s$ ensures that the smoothness of $\mathscr{F}_s(\theta)$ can be exploited to obtain faster rates of convergence of the estimator $\widehat{\theta}$ in estimating θ. For our purposes of estimation assume that $\mu \in \mathscr{F}_s(\theta)$ and $\sigma \in \mathscr{G}_r$ where $s \wedge r \geq 3$.

2.2. Dependence assumptions

Throughout the paper there will be a dependence assumption either among the design random variables or in the error random variables. In particular, the assumed dependence structure is a causal LRD linear process that is defined below.

Definition 2. *Let c_i be a set of square summable constant coefficients that are defined,*

$$c_i := \begin{cases} 1, & \text{if } i = 0, \\ i^{-(1+\alpha)/2}L(i), & \text{if } i \geq 1, \end{cases}$$

where $L : \mathbb{R}^+ \longrightarrow \mathbb{R}^+$ is a slowly varying function and $0 < \alpha \leq 1$. Then, a random variable ξ_i, is said to be a causal LRD linear process if,

$$\xi_i = \mu_\xi + \sum_{j=0}^{\infty} c_j \eta_{i-j}$$

where $|\mu_\xi| < \infty$ and η_i are i.i.d. random variables with density f_η and moments $\mathbb{E}\eta_t = 0$ and $\mathbb{E}\eta_t^2 = \left(\sum_{i=0}^{\infty} c_j^2\right)^{-1} =: \sigma_\eta^2$.

Furthermore, a random variable ξ_i is said to be a causal LRD Gaussian linear process if ξ_i satisfies Definition 2 and $\{\ldots, \eta_{i-1}, \eta_i\}$ are i.i.d $\mathcal{N}\left(0, \sigma_\eta^2\right)$. The case of $\alpha = 1$ is to be interpreted as a short range dependent case and by the construction the random variable has $\mathbb{E}\xi_i = \mu_\xi$ and $\mathbb{V}\text{ar}\xi_i = 1$. Moreover, it can be shown that ξ_i is a second-order stationary process and has asymptotic covariance structure $\mathbb{C}\text{ov}\left(\xi_0, \xi_k\right) \sim C_0^2 k^{-\alpha} L^2(k)$ where $C_0^2 = \sigma_\eta^2 \int_0^\infty (x^2 + x)^{-(1+\alpha)/2} \, dx$.

Therefore the process exhibits long-range dependence and a consequence of this asymptotic covariance structure is that,

$$\mathbb{Var}\left(\sum_{i=1}^{n}\xi_i\right) \sim C_1^2 n^{2-\alpha}L^2(n),$$

$$\mathbb{Var}\left(\sum_{i=1}^{n}\xi_i^2\right) \sim \begin{cases} C_2^2 n^{2-2\alpha}L^4(n), & \text{if } 0 < \alpha < \frac{1}{2}, \\ \\ C_3^2 n, & \text{if } \frac{1}{2} < \alpha < 1, \end{cases} \tag{11}$$

where $C_1^2 := 2C_0^2/((1-\alpha)(2-\alpha))$, $C_2^2 := 4C_0^4/((1-2\alpha)(2-2\alpha))$ and when $1/2 < \alpha < 1$, the sequence $\mathbb{Cov}\left(\xi_0^2, \xi_i^2\right)$ is summable and $C_3^2 = 1 + 2\sum_{i=0}^{\infty}\mathbb{Cov}\left(\xi_0^2, \xi_i^2\right)$. Also, when $\alpha = 1/2$, $\mathbb{Var}\left(\sum_{i=1}^{n}\xi_i^2\right)$ is asymptotically proportional to a term of order n times another term involving slowly-varying functions. Now throughout the paper, the design variables and error variables are assumed to follow one of the following dependence conditions:

(A) The design variables, $\{X_i\}_{i=1}^{n}$ are i.i.d. random variables with domain $\mathcal{X} \subseteq \mathbb{R}$ and common density f such that $f(x) > 0$ for all $x \in \mathcal{X}$ and $\sup_{x \in \mathcal{X}}|f^{(s \wedge r)}(x)| < \infty$. The error variables $\{\varepsilon_i\}_{i=1}^{n}$ are a causal LRD process with parameter α_ε. Furthermore, the random variables $\{\varepsilon_i\}_{i=1}^{n}$ are assumed to be independent of $\{X_i\}_{i=1}^{n}$. Under (A), define the associated set of σ-fields,

$$\mathcal{G}_i := \sigma(\ldots, \eta_{i-1}, \eta_i; X_1, X_2, \ldots, X_i).$$

(B) The design variables, $\{X_i\}$ are a causal LRD linear process with parameter α_x where $f_\eta^{(j)}$ is a Lipschitz continuous function for $j = 0, 1, \ldots, s+2$ with $f_X(x) > 0$ for all $x \in \mathbb{R}$. The error variables $\{\varepsilon_i\}_{i=1}^{n}$, are centred and i.i.d., with a finite variance, independent of $\{X_i\}_{i=1}^{n}$. Similarly, define the associated set of σ-fields,

$$\mathcal{F}_i = \sigma(\ldots, \eta_{i-1}, \eta_i; \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_i).$$

In both cases, the support of the design variables will be denoted $\mathcal{X}$. Let $F = F_X$ be the cumulative distribution function of X which is strictly increasing and denote by $F_n(x) = n^{-1}\sum_{i=1}^{n}\mathbb{1}_{\{X_i \leq x\}}$ the empirical distribution function of X. Also let $Q = F^{-1}$ and $Q_n = F_n^{-1}$ be the quantile and empirical quantile functions respectively. We require that Q is Lipschitz, that is, there exists an $L_Q > 0$ such that

$$|Q(x) - Q(y)| \leq L_Q|x - y|.$$

Finally, we need to impose some mild restrictions on σ. We assume σ is bounded away from 0 and ∞ in the sense that,

$$0 < \inf_{t \in \mathcal{X}} \sigma(t) < \sup_{t \in \mathcal{X}} \sigma(t) < \infty$$

and that $\sigma \in \mathscr{G}_r$ where $r \geq 3$. Throughout the article we denote by C a general constant that is assumed to be positive and finite but which possibly changes from line to line.

3. Main result

The main result of the paper is concerned with the construction and analysis of an estimator, $\widehat{\theta}$, of the kink location θ. The analysis of the estimator is given in Theorem 1 and concerns the rate of convergence of $\widehat{\theta}$ to the true the kink location θ. The estimator, $\widehat{\theta}$, will be constructed in Section 4 along with the motivations and analysis.

Theorem 1. *Suppose a bivariate sequence of observations $\{X_i, Y_i\}$ that follow model (1) are observed such that $\mu \in \mathscr{F}_s(\theta)$ with $s \geq 3$. Then an estimator, $\widehat{\theta}$ of the change point, θ, can be constructed such that,*

$$
|\widehat{\theta} - \theta| = \begin{cases} \mathcal{O}_p(n^{-\frac{s}{2s+1}}) & \text{if Assumption (A) holds and } \sigma \equiv C, \\ \\ \mathcal{O}_p\left(n^{-\frac{s}{2s+1}} \vee \left(n^{-\frac{\alpha_x}{2}} L(n)\right)\right) & \text{if Assumption (B) holds and } \sigma \in \mathscr{G}_r. \end{cases}
$$

where C is an arbitrary positive constant and $s \wedge r \geq 3$.

It is worth noting at this stage that the further restriction that $\sigma \equiv C$ under Assumption (A) is unnecessary for the specific estimation technique and is only required in the maximal deviation result to ensure that a kink can be detected in practice. Further detailed discussion of this matter along with the proof of Theorem 1 is given at the end of Section 4.

The minimax optimality of this result is not pursued in this paper since the lower bounds on the convergence rate of $\widehat{\theta}$ for the functional class $\mathscr{F}_s(\theta)$ are not determined in the framework of random design. However, it is worth making the specific point that the obtained rate of convergence under Assumption (A) is the same as the minimax rates for the fixed design case with i.i.d. errors (see Cheng and Raimondo (2008)). Consequently, it seems reasonable to conjecture that the rates of our estimator are optimal in the minimax sense.

4. Kink estimation method

In this section, the basis of the zero-crossing technique is studied and a brief overview given. Firstly, the zero-crossing technique pioneered by Goldenshluger, Tsybakov and Zeevi (2006) and applied by Cheng and Raimondo (2008); Wishart (2009) will be described briefly in Section 4.1 and then an adaptation for the random design case constructed in Sections 4.2 - 4.7.

4.1. Approximation of the third derivative for the fixed design model

In the fixed design setting (cf. model (6)) it can be assumed without loss of generality that the regression function μ has domain $[0, 1]$. More specifically, assume that $\mu \in \mathscr{F}_s([0, 1], \lambda)$ and estimate $\mu^{(3)}(t)$ by,

$$
\kappa_h(t) = \kappa_h(t, \mu) := h^{-4} \int_0^1 K_3\left(\frac{x - t}{h}\right) \mu(x)\,dx.
$$

where $h = h(n)$ is the bandwidth that depends on n. Throughout the article it will be assumed that the bandwidth satsifies, at the very least, $h + \frac{1}{nh} \to 0$, as $n \to \infty$. This is a standard regularity condition for kernel smoothing techniques and additional conditions on the bandwidth will be stated as needed. Using the functional class $\mathscr{F}_s(\theta)$ and the properties of the kernel function it can be shown that for $t \in (h, 1 - h)$,

$$\kappa_h(t) = h^{-2} K_1 \left(\frac{\lambda - t}{h} \right) [\mu^{(1)}](\lambda) + \mathcal{O}(h^{s-3}) =: L_h(t) + \mathcal{O}(h^{s-3}), \qquad (12)$$

where $L_h(t)$ is the localisation term. Indeed, by exploiting the conditions of K_3 we can express $\kappa_h(t)$ as follows. Change variable of integration to obtain,

$$\kappa_h(t) = h^{-4} \int_0^1 K_3 \left(\frac{x - t}{h} \right) \mu(x)\, dx$$

$$= h^{-3} \int_{-1}^1 K_3(x)\, \mu(t + hx)\, dx.$$

The last equality follows because the domain of K is $[-1, 1]$ and the values of t are restricted to $t \in (h, 1 - h)$. This restriction is used to avoid possible edge bias effects from the two sided kernel function. Using integration by parts and exploiting the boundary condition (9),

$$\kappa_h(t) = -h^{-2} \int_{-1}^1 K_2(x)\, \mu^{(1)}(t + hx)\, dx. \qquad (13)$$

Let $D = \{t : |\lambda - t| < h\}$ and $\tau = (\lambda - t)/h$. Then $|\tau| < 1$ for all $t \in D$. We now split (13) into two integrals,

$$\kappa_h(t) = -h^{-2} \int_{-1}^\tau K_2(x)\, \mu^{(1)}(t + hx)\, dx - h^{-2} \int_\tau^1 K_2(x)\, \mu^{(1)}(t + hx)\, dx.$$

To exploit $\mathscr{F}_s([0, 1], \lambda)$ define,

$$J_h(t) := -h^{-2} \left(\int_{-1}^\tau K_2(x) \left(\mu^{(1)}(t + hx) - \mu^{(1)}(\lambda^-) \right) dx \right.$$

$$\left. + \int_\tau^1 K_2(x) \left(\mu^{(1)}(t + hx) - \mu^{(1)}(\lambda^+) \right) dx \right) = \mathcal{O}(h^{s-3}).$$

The order bound follows by using (7) and (8) in combination with (10). Therefore, this allows us to express $\kappa_h(t)$ in the following way,

$$\kappa_h(t) = -h^{-2} \int_{-1}^\tau K_2(x)\, \mu^{(1)}(\lambda^-)\, dx - h^{-2} \int_\tau^1 K_2(x)\, \mu^{(1)}(\lambda^+)\, dx + J_h(t)$$

$$= h^{-2} K_1(\tau)[\mu^{(1)}](\lambda) + J_h(t) = L_h(t) + J_h(t).$$

Since the method is based on estimating a smoothed third derivative of μ, it is assumed that $s \geq 3$. This will guarantee that $\mu^{(3)}$ exists and is finite and the

method makes sense. Then the expansion in the above equation ensures that $\kappa_h(\cdot) = \mathcal{O}(h^{-2})$. More specifically we have the following,

$$\kappa_h(t) = \begin{cases} \mathcal{O}(h^{-2}), & \text{if } \mu \in \mathscr{F}_s(\theta) \text{ and } t \in D \\ \mathcal{O}(h^{s-3}), & \text{if } \mu \in \mathscr{G}_s \text{ with } t \in [0,1] \text{ or } \mu \in \mathscr{F}_s(\theta) \text{ with } t \notin D. \end{cases} \tag{14}$$

As seen in all three of the aforementioned papers that use the zero-crossing technique, the $\delta-$separation rate Lemma given below is the technical result that explains why the above representation is effective.

Lemma 1 (δ-separation rate)**.** *Let $K \in \mathscr{K}_s$ and $\mu \in \mathscr{F}_s([0,1],\theta)$. In what follows the constant $0 < q < 1$ is the abscissa of the global minimum of K_1. Let $h > 0$, $\delta > 0$ be such that $\delta < qh$. Let $A_{\delta,h} = \{t : \delta < |t - \theta| < qh\}$. Then for $\kappa_h(t) = \kappa_h(t,\mu)$:*

(a) $|\kappa_h(\theta)| \leq Ch^{s-3}$,
(b) for all $t \in A_{\delta,h}$ and $\delta \geq Ch^s$, $|\kappa_h(t)| \geq C\delta h^{-3}$,
(c) for all $t \in (0,1)$ such that $|\theta - t| > h$, $|\kappa_h(t)| \leq Ch^{s-3}$.

The proof of this Lemma is given in Cheng and Raimondo (2008). Their proof requires a minor correction as the extra regularity condition 3. is needed in the smoothness class $\mathscr{F}_s(\theta)$.

The main idea of Lemma 1 allows us to exploit the expansion given in (12) and focus in on the location of the kink. The kernel function has specific properties to guarantee that a unique global maximum and minimum occurs within order h of the kink point. Furthermore, the estimator was constructed so that the rate of convergence of kink location estimation is minimax for model (4). We will seek to adapt these results to the random design setting.

4.2. Adapted random design Estimator of the third derivative

Now consider $\mu \in \mathscr{F}_s(\mathcal{X}, \theta)$ in model (1). An estimator is constructed to exploit the smoothed third derivative of μ and the argument built around Lemma 1 discussed in Section 4.1. The most natural extension would be to use the estimator,

$$\widetilde{\kappa}_h(t) = \frac{1}{nh^4 \widehat{f}_X(t)} \sum_{i=1}^{n} K_3\left(\frac{X_i - t}{h}\right) Y_i, \tag{15}$$

where $\widehat{f}_X(t)$ is the estimate for the density of X_i at the point t given by,

$$\widehat{f}_X(t) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{X_i - t}{h}\right).$$

Unfortunately, from a brief computational investigation, the estimator given in (15) appears to suffer from poor numerical performance. Instead of using (15), another estimator with better numerical performance is constructed by rescaling the design variables by the distribution function F and defining a

rescaled regression function $\mu_F(\cdot) = \mu(Q(\cdot))$. This new estimator of $\kappa_h(t)$ in the random design setting is given by,

$$\widehat{\kappa}_h(t) = \frac{1}{nh^4} \sum_{i=1}^n Y_i K_3 \left(\frac{F(X_i) - t}{h} \right). \tag{16}$$

Apart from the gain in numerical computation benefits, the estimator $\widehat{\kappa}_h(t)$ has some properties that can be exploited. It reduces the general random design problem into a somewhat simpler framework. To see this, consider $F(X_i)$ to be the new random design variables of the regression problem. These new design variables follow the theoretically easier uniform distribution on $[0, 1]$. The price paid for this simplification is that the regression function that corresponds to the new design variables is now μ_F. This simpler framework is useful for a couple of reasons. Firstly, (16) is an unbiased estimate of the smoothed third derivative of the rescaled regression function μ_F,

$$\begin{aligned}
\mathbb{E}\widehat{\kappa}_h(t) &= h^{-4}\mathbb{E}\mu(X_1)K_3 \left(\frac{F(X_1) - t}{h} \right) \\
&= h^{-4} \int_{\mathbb{R}} \mu(u)K_3 \left(\frac{F(u) - t}{h} \right) dF(u) \\
&= h^{-4} \int_0^1 \mu_F(x)K_3 \left(\frac{x - t}{h} \right) dx = \kappa_h(t, \mu_F),
\end{aligned} \tag{17}$$

Thus, if $\mu \in \mathscr{F}_s(\theta)$, then $\mu_F \in \mathscr{F}_s([0, 1], \lambda)$ where $\theta = Q(\lambda)$. So, the smoothed third derivative of μ_F given in (17) can be exploited by Lemma 1 and the argument shown in Section 4.1. Then the problem is equivalent to estimating a kink location λ for the function μ_F in the fixed design setting.

With the previous argument in mind, an estimator $\widehat{\theta}$ of a kink location of the regression function μ in the random design setting is constructed that is approximately the same as the estimator for kink location λ of μ_F in the fixed design setting. This is done by estimating the value of λ by $\widehat{\lambda}$ using the established zero-crossing technique in the fixed design setting and then rescaling λ back by the quantile function to obtain an estimate of θ. Thus to assess the performance of our estimator we need to check that the convergence of $\widehat{\kappa}_h(t)$ to $\kappa_h(t)$ is sufficiently fast. To do this consider the two following processes,

$$\begin{aligned}
\gamma_i(t) &= \mu(X_i)K_3 \left(\frac{F(X_i) - t}{h} \right) \\
\zeta_i(t) &= \sigma(X_i)K_3 \left(\frac{F(X_i) - t}{h} \right).
\end{aligned} \tag{18}$$

With these definitions, the overall accuracy of the estimator can be decomposed into,

$$\widehat{\kappa}_h(t) = \kappa_h(t) + b_h(t) + Z_h(t), \tag{19}$$

where $b_h(t)$ and $Z_h(t)$ represent respectively the stochastic bias and stochastic error contributions to the estimator and are given by,

$$b_h(t) = n^{-1}h^{-4}\sum_{i=1}^{n}\left(\gamma_i(t) - \mathbb{E}\gamma_1(t)\right), \qquad Z_h(t) = n^{-1}h^{-4}\sum_{i=1}^{n}\zeta_i(t)\varepsilon_i.$$

The analysis of the above terms are given in the next subsection.

4.3. Probabilistic behaviour for the adapted estimator

In this section the analysis of the stochastic bias and stochastic error terms are considered before proceeding to the next stage of the zero-crossing technique to ensure that the stochastic contributions do not overwhelm the signal generated by the $\kappa_h(t)$ term. The proofs of the claims in this section will be deferred to Section 5.

The first term to be considered is the stochastic bias term which did not appear in previous kink analyses pursued by Cheng and Raimondo (2008); Wishart (2009) since there is some stochastic contribution by adapting the fixed design estimator to the the random design framework. Therefore, this term needs to be appropriately dealt with and the next Lemma is a useful tool that considers this term.

Lemma 2. *Consider a function* $\mu : \mathcal{X} \longrightarrow \mathbb{R}$ *such that* μ' *exists and is bounded. Then define the function*

$$\gamma_i^*(t) = \left(\mu(X_i) - \mu_F(t)\right)K_3\left(\frac{F(X_i) - t}{h}\right).$$

If the design variables follow Assumption (A) then,

$$\sup_{t\in(0,1)}\left|\sum_{i=1}^{n}\left(\gamma_i^*(t) - \mathbb{E}\gamma_i^*(t)\right)\right| = o_{\mathrm{a.s.}}\left(\sqrt{nh^3\left|\log h\right|}\right).$$

If the design variables follow Assumption (B) then,

$$\sup_{t\in(0,1)}\left|\sum_{i=1}^{n}\left(\gamma_i^*(t) - \mathbb{E}\left[\gamma_i^*(t)\middle|\mathcal{F}_{i-1}\right]\right)\right| = \mathcal{O}_p\left(\sqrt{nh^3\left|\log h\right|}\right).$$

Note that the two claims in given in Lemma 2 are respectively a uniform law of iterated logarithms for independent variables and a similar type of iterated logarithm result for martingale difference sequences.

We now state some central and non-central limit theorems for the estimator, $\widehat{\kappa}_h(t)$. The convergence of the estimator $\widehat{\kappa}_h(t)$ under both Assumption (A) and (B) is contingent on the size of the bandwidth relative to the level of dependence α. The specific details of this relationship between h and n^α will be shown in detail inside the Theorems. Roughly speaking, if the bandwidth is too

'large' compared to α then the dependence of the random variables dominate and the estimator converges to a process that needs to be normed by a sequence that relies on α. Conversely, if the bandwidth is 'small' compared to α then the dependence of the random variables is negligible and a regular central limit theorem holds with a norming sequence that is not reliant on α. In the forthcoming Theorems the extra smoothness of the regression and variance functions are exploited to be able to obtain an estimator that is not as sensitive to the level of dependence. In practice, this extra level of smoothness will most likely be unknown. Due to its common occurrence in the subsequent Theorems, define the asymptotic variance term, $\upsilon^2(t) := \left(\sigma_F^2(t) + \mu_F^2(t)\right) \int_{-1}^1 K_3^2(x)\,dx$. The following Theorem deals with the case of Assumption (A).

Theorem 2. *Let $K \in \mathscr{K}_{s\wedge r}$, $\mu \in \mathscr{F}_s$, $\sigma \in \mathscr{G}_r$ with $s \wedge r \geq 3$ and $t \in (h, 1-h)$. Also if the design variables and error random variables follow Assumption (A) and the bandwidth $h = h(n)$ also satisfies,*

$$h^{2(s\wedge r)+1} n^{1-\alpha_\varepsilon} L^2(n) \to 0 \quad as\ n \to \infty, \tag{A1}$$

then the following convergence result holds,

$$\sqrt{nh^7}\left(\widehat{\kappa}_h(t) - \kappa_h(t)\right) \xrightarrow{\mathscr{D}} \mathcal{N}\left(0, \upsilon^2(t)\right). \tag{20}$$

Conversely, if the bandwidth $h = h(n)$ satisfies,

$$h^{2(s\wedge r)+1} n^{1-\alpha_\varepsilon} L^2(n) \to \infty \quad as\ n \to \infty, \tag{A2}$$

then,

$$\frac{n^{\frac{\alpha_\varepsilon}{2}} h^{3-(s\wedge r)}}{L(n)}\left(\widehat{\kappa}_h(t) - \kappa_h(t)\right) \xrightarrow{\mathscr{D}} \mathcal{N}\left(0, C_1^2 \upsilon_*^2(t)\right).$$

where

$$\upsilon_*(t) = \frac{\sigma_F^{(s\wedge r)}(t)}{(s\wedge r)!} \int_{-1}^1 x^{s\wedge r} K_3(x)\,dx.$$

Theorem 3 and Theorem 4 deal with case under Assumption (B) and give the central limit theorems when there is a 'small' or 'large' bandwidth respectively. In the 'large' bandwidth scenario a stronger assumption is used whereby the design variables are a causal LRD Gaussian linear process.

Theorem 3. *Let $K \in \mathscr{K}_{s\wedge r}$, $\mu \in \mathscr{F}_s$, $\sigma \in \mathscr{G}_r$ with $s \wedge r \geq 3$ and $t \in (h, 1-h)$. If the design variables and error random variables follow Assumption (B) and the bandwidth $h = h(n)$ satisfies,*

$$h^7 n^{1-\alpha_x} L^2(n) \to 0 \quad as\ n \to \infty, \tag{B1}$$

then the estimator obeys the following law,

$$\sqrt{nh^7}\left(\widehat{\kappa}_h(t) - \kappa_h(t)\right) \xrightarrow{\mathscr{D}} \mathcal{N}\left(0, \upsilon^2(t)\right). \tag{21}$$

Theorem 4. *Let $K \in \mathscr{K}_{s \wedge r}$, $\mu \in \mathscr{F}_s$, $\sigma \in \mathscr{G}_r$ with $s \wedge r \geq 3$ and $t \in (h, 1-h)$. Assume the design variables and error random variables follow Assumption (B) and that the design variables are a causal LRD Gaussion linear process. If the bandwidth $h = h(n)$ satisfies,*

$$h^7 n^{1-\alpha_x} L^2(n) \to \infty \quad as\ n \to \infty, \tag{B2}$$

and the estimator $\widehat{\kappa}_h(t)$ has a Hermite rank of 1 then the the estimator obeys the following law,

$$\frac{n^{\frac{\alpha_x}{2}}}{L(n)} \left(\widehat{\kappa}_h(t) - \kappa_h(t) \right) \xrightarrow{\mathscr{D}} \mathcal{N}(0, C_1^2 \mathcal{H}_1(t))$$

where

$$\mathcal{H}_1(t) = \frac{\kappa_h(t)}{s_X^3 \sigma_\eta \phi \left(\Phi^{-1}(t) \right)} \int_{\mathbb{R}} \phi \left(\frac{\Phi^{-1}(t) - u}{s_X} \right) \left(\Phi^{-1}(t) - u \right) \phi \left(\frac{u}{\sigma_\eta} \right)\, du,$$

$s_X^2 = 1 - \sigma_\eta^2$ and ϕ and Φ are the standard normal density and cumulative distribution functions respectively.

Remark 1. If the estimator $\widehat{\kappa}_h(t)$ has Hermite rank q for some $q \in \{2, 3, \ldots\}$ then the asymptotic distribution depends on the size of the bandwidth relative to $q\alpha$. Firstly, if $n^{1-q\alpha_x} h^7 L^{2q}(n) \to \infty$ then it can be shown using a similar argument used in the Proof of Theorem 4 with the result of Theorem 2 of Avram and Taqqu (1987) that the normed process $n^{q\alpha_x/2} L^{-q}(n) \left(\widehat{\kappa}_h(t) - \kappa_h(t) \right) \xrightarrow{\mathscr{D}} \mathcal{H}_q(t) \mathscr{H}_q$ where,

$$\mathcal{H}_q(t) = \frac{\kappa_h(t)}{s_X^2 \sigma_\eta \phi \left(\Phi^{-1}(t) \right)} \int_{\mathbb{R}} \phi \left(\frac{\Phi^{-1}(t) - u}{s_X} \right) H_q \left(\frac{\Phi^{-1}(t) - u}{s_X} \right) \phi \left(\frac{u}{\sigma_\eta} \right) du$$

and $H_q(x)$ is the Hermite polynomial of degree q and $\mathscr{H}_q$ is the Hermite-Rosenblatt process,

$$\mathscr{H}_q = C_{\mathscr{H}} \int_{-\infty < x_1 < x_2 \ldots < x_q < 1} \left\{ \int_0^1 \prod_{i=1}^q \left((y - x_i)^+ \right)^{-\frac{\alpha+1}{2}} dy \right\} dB(x_1) \ldots dB(x_q)$$

where $C_{\mathscr{H}} = (q!(1 - q\alpha)(2 - q\alpha)/(2 \left(\int_0^\infty (x^2 + x)^{-(1+\alpha)/2} dx \right)^q))^{1/2}$ and B denotes a standard Brownian motion. In Avram and Taqqu (1987), they considered Appell polynomials for a generalised sequence of stationary LRD random variables. In our case the LRD variables are Gaussian and consequently the Appell polynomials reduce to the Hermite polynomials. On the other hand, if the bandwidth satisfies $n^{1-q\alpha_x} h^7 L^{2q}(n) \to 0$ then (21) holds.

As will be seen in Section 4.5, some large deviations results are needed to be able to distinguish between the signal generated by the $\kappa_h(t)$ term and the stochastic bias and noise contributions. Unfortunately, a slightly weaker large deviations result is proved under Assumption (A) in Theorem 5. In particular

we assume that the scale function, $\sigma(\cdot) \equiv \sigma$, is constant however this restriction could possibly be relaxed by using a different method. The large deviations result for Assumption (B) in Theorem 6 does not carry this restriction and the scale function need not be constant.

Theorem 5. *Let $K \in \mathcal{K}_{s \wedge r}$ and the design and error variables satisfy Assumption (A). Further assume that the bandwidth $h = h(n)$ also satisfies,*

$$\frac{|\log h|^3}{nh^3} + \frac{L^2(n) |\log h|^2}{n^{\alpha_\varepsilon} h^{\frac{4}{3}}} \to 0 \quad \text{as } n \to \infty. \tag{22}$$

Then define,

$$S_n^A(t) := \sum_{i=1}^n \Psi_i(t) = \frac{1}{\upsilon(t)\sqrt{nh}} \sum_{i=1}^n K_3 \left(\frac{F(X_i) - t}{h} \right) (\mu_F(t) + \sigma \varepsilon_i).$$

where $\sigma > 0$ is constant. Also define,

$$B_n(x) = \sqrt{2 \log n} + \frac{x}{\sqrt{2 \log n}} - \frac{1}{\sqrt{2 \log n}} \left(\frac{1}{2} \log \log n + \log \left(2\sqrt{\pi} \right) \right) \tag{23}$$

and a partition of $[0, 1]$,

$$T_n = \{ t_j = 2hj, j = 1, \ldots, m_n - 1 \} \tag{24}$$

where $m_n = \lceil \frac{1}{2h} \rceil$. Then,

$$\lim_{n \to \infty} P \left(\sup_{t \in T_n} \left| S_n^A(t) \right| \leq B_{m_n}(x) \right) = e^{-2e^{-x}},$$

for all $x \in \mathbb{R}$.

Theorem 6. *Let $K \in \mathcal{K}_{s \wedge r}$, $\sigma \in \mathcal{G}_r$ with $s \wedge r \geq 3$ and the design and error variables satisfy Assumption (B) and assume that the bandwidth $h = h(n)$ also satisfies,*

$$\frac{1}{nh^3} + \frac{L^2(n)}{n^{\alpha_x} h^{\frac{8}{3}}} + \frac{h^{2(s \wedge r)+1} n}{|\log h|} \to 0 \quad \text{as } n \to \infty. \tag{25}$$

Then define,

$$S_n^B(t) := \sum_{i=1}^n \Xi_i(t) = \frac{1}{\upsilon(t)\sqrt{nh}} \sum_{i=1}^n K_3 \left(\frac{F(X_i) - t}{h} \right) (\mu_F(t) + \sigma(X_i) \varepsilon_i),$$

then for T_n defined in (24) with $m_n = \left\lceil \frac{1}{2h} \right\rceil$,

$$\lim_{n \to \infty} P \left(\sup_{t \in T_n} \left| S_n^B(t) \right| \leq \sqrt{2 \log m_n} \right) = 0.$$

4.4. Localisation step

Recall from (12), that the probe function given by $\kappa_h(t)$ gives a signal from the localisation term, $L_h(t)$ with some approximation error and the estimator adds a stochastic bias and error term,

$$\widehat{\kappa}_h(t) = L_h(t) + \mathcal{O}(h^{s-3}) + Z_h(t) + b_h(t). \tag{26}$$

Clearly, $h^{-2} > h^{s-3}$, since $s \geq 3$. So to be able to discern the signal generated from $L_h(t) = \mathcal{O}(h^{-2})$, it is required that $L_h(t)$ dominates the stochastic terms, $Z_h(t)$ and $b_h(t)$.

By construction of the kernel function, (cf. Cheng and Raimondo (2008)), $K_1(\cdot)$ has two unique extrema in the form of a unique global minimum and maximum in the interval $[-1, 1]$. This implies that $K_1(\cdot/h)$ has the same unique extrema in an interval of a length $\mathcal{O}(h)$. Consequently, $L_h(\cdot)$ has two unique global extrema near $t^* = \lambda + \mathcal{O}(h)$ and $t_* = \lambda - \mathcal{O}(h)$. As in the fixed design scenario considered by Cheng and Raimondo (2008); Wishart (2009) define,

$$t_* := \underset{t \in (0,1)}{\arg\min} \, L_h(t) \,, \qquad t^* := \underset{t \in (0,1)}{\arg\max} \, L_h(t).$$

However, in practice the location of t_* and t^* are not known and estimated using $\widehat{\kappa}_h(t)$ with,

$$\widehat{t}_* = \underset{t \in (0,1)}{\arg\min} \, \widehat{\kappa}_h(t) \,, \qquad \widehat{t}^* = \underset{t \in (0,1)}{\arg\max} \, \widehat{\kappa}_h(t).$$

If $\mu_F \in \mathscr{F}_s([0,1], \lambda)$ then,

$$|L_h(t^*) + L_h(t_*)| \geq Ch^{-2}. \tag{27}$$

There are two respective bandwidth restrictions, ((A1), (A2); (B1), (B2)) for the asymptotic behaviour of the estimator under Assumption (A) and Assumption (B) respectively. Starting with (A1) and (B1), to have a well defined signal, it is required that, $h^{-2} \geq Cn^{-\frac{1}{2}} h^{-\frac{7}{2}} \Rightarrow h \geq Cn^{-\frac{1}{3}}$. Furthermore, since it is assumed that $s \wedge r \geq 3$, to ensure that (20) and (21) always hold it suffices to choose h such that $h \leq Cn^{-\frac{1}{7} + (\alpha_x \vee \alpha_\varepsilon)/7 - \delta}$, for some $\delta > 0$ or,

$$Cn^{-\frac{1}{3} + \delta} < h < Cn^{-\frac{1}{7} - \delta} \tag{28}$$

for some $\delta > 0$. With this choice, the bandwidth restrictions given by (A1) and (B1) will always hold.

It is worth noting that under this choice, the order of the stochastic terms does not involve α_x or α_ε, the level of dependence. Note that h is chosen in a very similar manner if ε_i and X_i, $i \geq 1$, were i.i.d. Consequently, there will be no influence of the (long range) dependence on the change point estimation. The influence of the long range dependence will only affect testing purposes of the threshold used to determine if a signal is genuine and this will be discussed in the next subsection.

4.5. Kink detection step

For simplicity in notation, assume that $[\mu_F]^{(1)}(\lambda) > 0$, which means, $t_* < t^*$ (a similar argument follows if $[\mu_F]^{(1)}(\lambda) < 0 \Rightarrow t_* > t^*$.) To detect a kink, first standardise the statistic $\widehat{\kappa}_h(t)$ to have unit variance. This will allow us to appropriately notice if there is a change-point present when the observed extrema of $\widehat{\kappa}_h(t)$ exceed the threshold for the noise process. Define this standardised process as,

$$\mathcal{T}_{\widehat{\kappa}}(t) := \frac{\sqrt{nh^7}\widehat{\kappa}_h(t)}{\upsilon(t)}.$$

Then by (26) the $\mathcal{T}_{\widehat{\kappa}}(t)$ process has expansion,

$$\mathcal{T}_{\widehat{\kappa}}(t) = \frac{n^{\frac{1}{2}}h^{\frac{7}{2}}}{\upsilon(t)}L_h(t) + \mathcal{O}(n^{\frac{1}{2}}h^{s+\frac{1}{2}}) + \frac{n^{\frac{1}{2}}h^{\frac{7}{2}}}{\upsilon(t)}\left(Z_h(t) + b_h(t)\right). \qquad (29)$$

As seen earlier, the information regarding a kink is generated by the $L_h(t)$ process. A thresholding regime will be considered to be able to distinguish between the signal generated by $L_h(t)$ against the noise signal generated by the $Z_h(t)$ and $b_h(t)$ terms. This thresholding will be split into the two scenarios for Assumption (A) and (B).

Begin by giving a general decomposition of the estimator for both cases by using, $\gamma_i^*(t) := \left(\mu(X_i) - \mu_F(t)\right)K_3\left(\frac{F(X_i)-t}{h}\right) = \gamma_i(t) - \mu_F(t)K_3\left(\frac{F(X_i)-t}{h}\right)$ and using (18) and (19). So,

$$\begin{aligned}
\mathcal{T}_{\widehat{\kappa}}(t) &= \frac{\sqrt{nh^7}}{\upsilon(t)}\kappa_h(t) + \frac{1}{\upsilon(t)\sqrt{nh}}\sum_{i=1}^{n}\left(\gamma_i(t) - \mathbb{E}\gamma_1(t) + \zeta_i(t)\varepsilon_i\right) \\
&= \frac{\sqrt{nh^7}}{\upsilon(t)}\kappa_h(t) + \frac{1}{\upsilon(t)\sqrt{nh}}\sum_{i=1}^{n}\left(\gamma_i^*(t) - \mathbb{E}\gamma_1^*(t)\right. \\
&\qquad\qquad\qquad\qquad\left. + \mu_F(t)K_3\left(\frac{F(X_i)-t}{h}\right) + \zeta_i(t)\varepsilon_i\right). \qquad (30)
\end{aligned}$$

Focus on Assumption (A) and assume $\sigma(\cdot) \equiv C$, constant. The assumption that the scale function is constant is required in the proof of the maximal deviation result in Theorem 5. It may be possible to relax this condition and have the same result for $\sigma \in \mathscr{G}_r$ under Assumption (A) but remains a conjecture at this stage. Nevertheless, to control the stochastic terms in (30) first apply Lemma 2 and use (10),

$$\mathcal{T}_{\widehat{\kappa}}(t) = \frac{\sqrt{nh^7}}{\upsilon(t)}\kappa_h(t) + o_{\text{a.s.}}\left(\sqrt{|\log h|}\right) + S_n^A(t).$$

Then consider the values of t on the initial coarse grid T_n (see (24)) where the increments are of size $2h$. The grid values will be refined later in Section 4.6.

From Theorem 5, it is known that $\sup_{t \in T_n} \left| S_n^A(t) \right|$ will diverge to infinity no faster than $\sqrt{2 \left| \log 2h \right|}$. Also, if $\mu \in \mathscr{G}_s$, then from (14), $\kappa_h(t) = \mathcal{O}(h^{s-3})$ and

$$\lim_{n \to \infty} P \left(\sup_{t \in T_n} |\mathcal{T}_{\widehat{\kappa}}(t)| \geq \sqrt{2 \left| \log 2h \right|} \right) = 0. \tag{31}$$

However, if $\mu \in \mathscr{F}_s(\theta)$, then (27) holds and by (29), $\max_{t \in (t_*, t^*)} \mathcal{T}_{\widehat{\kappa}}(t) \geq Cn^{\frac{1}{2}} h^{\frac{3}{2}} > \sqrt{2 \left| \log 2h \right|}$ and a kink is detected when,

$$\max_{t \in T_n} |\mathcal{T}_{\widehat{\kappa}}(t)| \geq \sqrt{2 \left| \log 2h \right|}. \tag{32}$$

A very similar argument holds for Assumption (B). In this case assume that the scale function $\sigma \in \mathscr{G}_r$ with $r \geq 3$ and proceed as before. In conjunction with (30) and (10) apply Lemma 2,

$$\mathcal{T}_{\widehat{\kappa}}(t) = \frac{\sqrt{nh^7}}{\upsilon(t)} \kappa_h(t) + S_n^B(t) + D_n(t) + \mathcal{O}_p \left(h\sqrt{|\log h|} \right)$$

where the extra term $D_n(t)$ is defined,

$$D_n(t) := \frac{1}{\upsilon(t)\sqrt{nh}} \sum_{i=1}^{n} \left\{ \mathbb{E}\left[\gamma_i^*(t) \middle| \mathcal{F}_{i-1} \right] - \mathbb{E}\gamma_i^*(t) \right\}.$$

Using Lemma 3 (see Appendix in Section 5), with the bandwidth condition (25), $\sup_{t \in (h, 1-h)} |D_n(t)| = o_p \left(\sqrt{|\log h|} \right)$. Also, the bandwidth restriction (28) guarantees that (25) and consequently Theorem 6 holds. Then for Assumption (B) the same argument applies that was used to show (31) for Assumption (A).

This thresholding technique does raise some restrictions that could possibly be removed by another technique. Recall from (28), that $h > Cn^{-\frac{1}{3}+\delta}$ for some $\delta > 0$ is required to be able to distinguish the signal from the stochastic terms. Also, (22) and (25) are required to be able to apply Theorem 5 and Theorem 6 respectively and obtain a large deviation result for the process. Therefore to ensure both conditions are satisfied, it is sufficient to consider $\alpha_x > \frac{8}{9}$ or $\alpha_\varepsilon > \frac{4}{9}$.

4.6. Zero-crossing technique

If a kink is detected (when (32) is satisfied) then the method can proceed to the zero-crossing step. This step considers the interval $\widehat{A}_h := \left[\widehat{t}_*, \widehat{t}^* \right]$, which will contain λ and $\left| \widehat{t}^* - \widehat{t}_* \right| = \mathcal{O}(h)$. The main idea behind the zero-crossing technique is that for $t \in \widehat{A}_h$, $\widehat{\kappa}_h(t) \approx \kappa_h(t)$. Using Lemma 1 we can locate the zero-crossing-time of $\kappa_h(t)$ which occurs at $t = \lambda$ with an accuracy of order δ, $\delta < h$. This is done by minimising $|\widehat{\kappa}_h(t)|$ within the interval $\widehat{A}_h$:

$$\widehat{\lambda} = \arg\min_{t \in \widehat{A}_h} |\widehat{\kappa}_h(t)| = \arg\min_{t \in \widehat{A}_h} |\mathcal{T}_{\widehat{\kappa}}(t)|.$$

By comparing (12) with the bounds in Lemma 1 we see that the minimum is well defined if,

$$\delta h^{-3} \geq C h^{s-3} \quad \text{and} \quad \delta h^{-3} \geq C n^{-\frac{1}{2}} h^{-\frac{7}{2}}. \tag{33}$$

We will obtain the best possible accuracy if we choose δ as small as possible, as long as both inequalities of (33) still hold. The left hand expression of (33) implies that $\delta \asymp h^s$ and substituting this into the right hand expression of (33) we derive the order of the smallest possible bandwidth

$$h_* \asymp n^{-\frac{1}{2s+1}}.$$

We now apply Lemma 1 with $\delta_* = h_*^s$ to locate the change point λ in μ_F with an accuracy of order,

$$\left| \widehat{\lambda} - \lambda \right| = \delta_* = h_*^s \asymp n^{-\frac{s}{2s+1}}.$$

Remark 2. There are some limitations to the procedure presented thus far in terms of detection. More specifically, dependent on the location of λ relative to the grid values in T_n, the detection phase may fail. Indeed, define the closest grid value $\lambda^* := \arg\min_{t \in T_n} |\lambda - t|$. If λ_* is too close to λ, that is, $|\lambda - \lambda^*| < \delta$ then the procedure will not detect a kink since $L_h(t_i) = \mathcal{O}\left(h^{s-3}\right)$ for $t_i \in \{\lambda^* - 2h, \lambda^*, \lambda^* + 2h\}$ and consequently $\kappa_h(t_i) = \mathcal{O}\left(\sqrt{|\log h|}\right)$ and (32) will not hold. However, if $\delta < |\lambda - \lambda^*| < h$ then a kink will be detected since (32) holds and the aforementioned procedures in Sections 4.5 - 4.6 will follow. Furthermore, the limitations imposed by the coarse T_n grid affect only the kink detection step and will not influence on the zero-crossing step.

4.7. Modified estimator of kink

Recall that $\theta = Q(\lambda)$. In practice the true distribution function F is unknown, so it is estimated in the usual manner by the empirical distribution function $F_n(x) = n^{-1} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}}$ and consequently can obtain an estimator of Q via the empirical quantile function $Q_n(\cdot)$. Estimate θ by, $\widehat{\theta} = Q_n(\widehat{\lambda})$. The rate of convergence of this estimator is evaluated below,

$$
\begin{aligned}
|\widehat{\theta} - \theta| &= |Q_n(\widehat{\lambda}) - Q(\lambda)| \\
&\leq |Q_n(\widehat{\lambda}) - Q(\widehat{\lambda})| + |Q(\widehat{\lambda}) - Q(\lambda)| \\
&\leq |Q_n(\widehat{\lambda}) - Q(\widehat{\lambda})| + L_Q|\widehat{\lambda} - \lambda| \\
&\leq |Q_n(\widehat{\lambda}) - Q(\widehat{\lambda})| + \mathcal{O}_p(n^{-\frac{s}{2s+1}}).
\end{aligned}
\tag{34}
$$

The rate of convergence in (34) is therefore contingent on the maximum of the rate from the generalised quantile process for the design variables or the rate from the initial unscaled kink estimator. Under Assumption (A), the quantile

process involves independent and identically distributed design variables and for all $t \in (0, 1)$,

$$|Q_n(t) - Q(t)| = \mathcal{O}_p(n^{-\frac{1}{2}}) \tag{35}$$

(see Csörgő (1983) and references therein for a detailed treatment). For Assumption (B), the rate is dependent on α_x and for all $t \in (0, 1)$,

$$|Q_n(t) - Q(t)| = \mathcal{O}_p(n^{-\frac{\alpha_x}{2}} L(n)) \tag{36}$$

(see Theorem 5.1 of Ho and Hsing (1996)). Therefore, using (35) and (36) in (34),

$$|\widehat{\theta} - \theta| = \begin{cases} \mathcal{O}_p(n^{-\frac{s}{2s+1}}) & \text{under Assumption (A) with } \sigma \equiv C, \\ \\ \mathcal{O}_p\left(n^{-\frac{s}{2s+1}} \vee \left(n^{-\frac{\alpha_x}{2}} L(n)\right)\right) & \text{under Assumption (B) with } \sigma \in \mathscr{G}_r, \end{cases}$$

where $s \wedge r \geq 3$ which proves Theorem 1.

Remark 3. The method can be extended to the multiple kink scenario by observing multiple instances of (32). For each instance of (32) there is a corresponding interval $\widehat{A}_h$ and the localisation and zero-crossing-time steps are executed on each of those intervals to produce an estimate for each kink location. The interested reader is referred to Cheng and Raimondo (2008); Wishart (2009) for a more detailed treatment of the method in the multiple kink scenario with numerical examples. However, it is worth pointing out that there are some limitations to the accuracy of this method in this situation. Problems will arise if the multiple change-points are not well spaced apart in the sense that they are within order h of each other. To see this, let λ_1 and λ_2 be two such change-points. When t is within order h of both the change points, the localisation term, $L_h(t)$ will not produce two unique disjoint signals for the kinks. Instead the signals generated by $K_1\left(\frac{\lambda_i - t}{h}\right)$ for $i = 1, 2$ will interact and be confounded in one overlapping signal.

5. Mathematical Appendix

Before giving the proofs, some notation is described. Let X denote a random variable and denote the L_p-norm $\|X\|_p^p = \mathbb{E}|X|^p$ and $\|\cdot\| = \|\cdot\|_2$. For a function $f : \mathcal{X} \longrightarrow \mathbb{R}$ denote the sup-norm $|f|_\infty = \sup_{x \in \mathcal{X}} |f(x)|$. Throughout this Section a Taylor expansion of composite functions will be used to exploit the vanishing moment condition of K_3. For the Taylor expansion to be well defined, the derivatives of the composite functions need to exist. A generalised chain rule for composite functions exists (see the Faà di Bruno formula from Hernández Encinas, Martín del Rey and Muñoz Masqué (2005) and references therein), and these are of the form,

$$\frac{d^n}{dx^n} f(g(x)) := \frac{d^n}{dx^n} (f \circ g)(x) = \sum_{\mathbf{k} \in \mathcal{K}_n} \frac{n!}{k_1! k_2! \dots k_n!} (f^{(k)} \circ g)(x) \prod_{i=1}^n \left(\frac{g^{(i)}(x)}{i!}\right)^{k_i} \tag{37}$$

where $\mathcal{K}_n = \{k_i \in \{\mathbb{Z}^+ \cup 0\} : k_1 + 2k_2 + \ldots + nk_n = n\}$ and $k = \sum_{i=1}^{n} k_i$. Also, through tedious but elementary calculus it can be shown that, the nth derivative of $Q = F^{-1}$ will exist, and the Taylor expansions of μ_F and σ_F up to order n will exist if $f^{(n)}$ exists.

Proof of Lemma 2. Begin with the proof of the first claim under Assumption (A). Since $\gamma_i^*(t)$ will be non-zero only if $F(X_i) \in (t - h, t + h)$, there exists a $\tau_i \in (-1, 1)$ that depends on X_i such that,

$$\gamma_i^*(t) = K_3 \left(\frac{F(X_i) - t}{h} \right) (\mu_F(t + \tau_i h) - \mu_F(t))$$

$$= h\tau_i K_3 \left(\frac{F(X_i) - t}{h} \right) \mu_F^{(1)}(t + h\xi_i |\tau_i| h) =: h\nu_i(t),$$

and ξ_i depends on τ_i. The $\nu_i(t)$ terms are independent random variables, each of which have variance that is of order h. Therefore by the Law of Iterated Logarithm (see Bingham (1986)) we have the following result,

$$\limsup_{n \to \infty} \frac{1}{\sqrt{nh \log \log n}} \sum_{i=1}^{n} \left(\nu_i(t) - \mathbb{E}\nu_i(t) \right) = C \qquad \text{a.s.}$$

$$-\liminf_{n \to \infty} \frac{1}{\sqrt{nh \log \log n}} \sum_{i=1}^{n} \left(\nu_i(t) - \mathbb{E}\nu_i(t) \right) = C \qquad \text{a.s.}$$

Therefore we have,

$$\sum_{i=1}^{n} \left(\gamma_i^*(t) - \mathbb{E}\gamma_i^*(t) \right) = h \sum_{i=1}^{n} \left(\nu_i(t) - \mathbb{E}\nu_i(t) \right) = \mathcal{O}_{\text{a.s.}} \left(\sqrt{nh^3 \log \log n} \right)$$

$$= o_{\text{a.s.}} \left(\sqrt{nh^3 |\log h|} \right)$$

which proves the first claim of the Lemma. Now to concentrate on the claim for Assumption (B), a proof of a similar claim in Lemma 4 of Zhao and Wu (2006) is adapted to our framework. This technique bounds the martingale difference sequence $\gamma_i^*(t) - \mathbb{E}\left[\gamma_i^*(t) | \mathcal{F}_{i-1} \right]$ above and below by two discretised martingale difference sequences and uses an exponential martingale inequality to gain the required probabilistic bounds. To do this, again exploit the Taylor expansion of μ in Definition 1 and use the fact that $Support(K_3) = [-1, 1]$, which means that there exists a τ_i dependent on X_i with $|\tau_i| \leq 1$ such that $F(X_i) = t + \tau_i h$ and,

$$\gamma_i^*(t) = K_3 \left(\frac{F(X_i) - t}{h} \right) (\mu_F(t + \tau_i h) - \mu_F(t)) \mathbb{1}_{(t-h,t+h)} (F(X_i))$$

$$= \tau_i h K_3 \left(\frac{F(X_i) - t}{h} \right) \mu_F^{(1)}(t + \xi |\tau_i| h) \mathbb{1}_{(t-h,t+h)} (F(X_i)), \qquad (38)$$

where $|\xi| \leq 1$. Then split the function in (38) into its positive and negative parts by defining $\xi_i := t + \xi |\tau_i| h$ and $\tau_i \mu_F^{(1)}(\xi_i) = \left(\tau_i \mu_F^{(1)}(\xi_i) \right)^+ - \left(\tau_i \mu_F^{(1)}(\xi_i) \right)^- =:$

$\mu_{F,1}^+(\xi_i) - \mu_{F,1}^-(\xi_i)$ where $f^+ = (f \vee 0)$, $f^- = (-f \wedge 0)$ denote the respective positive and negative parts of f. Then,

$$
\begin{aligned}
\gamma_i^*(t) &= h\left(\mu_{F,1}^+(\xi_i) - \mu_{F,1}^-(\xi_i)\right)\left(K_3^+\left(\frac{F(X_i)-t}{h}\right) - K_3^-\left(\frac{F(X_i)-t}{h}\right)\right) \\
&= h\left(\mu_{F,1}^+(\xi_i)K_3^+\left(\frac{F(X_i)-t}{h}\right) - \mu_{F,1}^+(\xi_i)K_3^-\left(\frac{F(X_i)-t}{h}\right)\right. \\
&\qquad\left. -\mu_{F,1}^-(\xi_i)K_3^+\left(\frac{F(X_i)-t}{h}\right) + \mu_{F,1}^-(\xi_i)K_3^-\left(\frac{F(X_i)-t}{h}\right)\right) \\
&:= \left(\varsigma_i^{++}(t) - \varsigma_i^{+-}(t) - \varsigma_i^{-+}(t) + \varsigma_i^{--}(t)\right),
\end{aligned}
\tag{39}
$$

By the linearity of the conditional expectation operator and (39) we can decompose the martingale difference sequence into parts,

$$
\begin{aligned}
\gamma_i^*(t) &- \mathbb{E}\left[\gamma_i^*(t)\middle|\mathcal{F}_{i-1}\right] \\
&= \varsigma_i^{++}(t) - \mathbb{E}\left[\varsigma_i^{++}(t)\middle|\mathcal{F}_{i-1}\right] - \left(\varsigma_i^{+-}(t) - \mathbb{E}\left[\varsigma_i^{+-}(t)\middle|\mathcal{F}_{i-1}\right]\right) \\
&\qquad - \left(\varsigma_i^{-+}(t) - \mathbb{E}\left[\varsigma_i^{-+}(t)\middle|\mathcal{F}_{i-1}\right]\right) + \varsigma_i^{--}(t) - \mathbb{E}\left[\varsigma_i^{--}(t)\middle|\mathcal{F}_{i-1}\right]
\end{aligned}
\tag{40}
$$

To begin with we will concentrate on the first martingale difference term on the RHS of (40) and bound it above and below by a discretised version that does not depend on t directly. For this discretization let $N = \lceil(nh^{-3})^{\frac{1}{2}}\rceil$ and $t_j = \frac{j}{N}$ where $0 \leq j \leq N$. Then for any $t \in [0,1]$ there exists a j such that $t \in [t_j, t_{j+1})$ and the distance $|t_{j+1} - t_j| = \mathcal{O}(N^{-1})$. Define the two new tweaked martingale difference sequences versions of $\varsigma_i^{++}(t)$,

$$
\overline{\varsigma}_{i,j}^{++} = h\mu_{F,1}^+(\xi_{i,j})\frac{1}{2}\left\{K_3^+\left(\frac{F(X_i)-t_j}{h}\right) + K_3^+\left(\frac{F(X_i)-t_{j+1}}{h}\right)\right\}\mathbb{1}_{\chi_j}(F(X_i))
$$

$$
\underline{\varsigma}_{i,j}^{++} = h\mu_{F,1}^+(\xi_{i,j})K_3^+\left(\frac{F(X_i)-t_j}{h}\right)\mathbb{1}_{(t_{j+1}-h,t_j+h)}(F(X_i))
$$

where $\xi_{i,j} = \xi_i - (t - t_j)$ and $\chi_j = (t_j - h, t_{j+1} + h)$. It can be shown that, the martingale difference sequence $\varsigma_i^{++}(t) - \mathbb{E}\left[\varsigma_i^{++}(t)\middle|\mathcal{F}_{i-1}\right]$ can be bounded uniformly in t above and below by,

$$
\varsigma_i^{++}(t) - \mathbb{E}\left[\varsigma_i^{++}(t)\middle|\mathcal{F}_{i-1}\right] \leq \overline{\varsigma}_{i,j}^{++} - \mathbb{E}\left[\overline{\varsigma}_{i,j}^{++}\middle|\mathcal{F}_{i-1}\right] + CN^{-1}
$$

$$
\varsigma_i^{++}(t) - \mathbb{E}\left[\varsigma_i^{++}(t)\middle|\mathcal{F}_{i-1}\right] \geq \underline{\varsigma}_{i,j}^{++} - \mathbb{E}\left[\underline{\varsigma}_{i,j}^{++}\middle|\mathcal{F}_{i-1}\right] - CN^{-1}.
$$

We have the following result,

$$
\begin{aligned}
\sup_{t\in(0,1)}\left|\sum_{i=1}^n\left(\varsigma_i^{++}(t) - \mathbb{E}\left[\varsigma_i^{++}(t)\middle|\mathcal{F}_{i-1}\right]\right)\right| \\
\leq \max_{0\leq j\leq N-1}\left(\left|\overline{S_n}(j)\right| + \left|\underline{S_n}(j)\right|\right) + CnN^{-1} \\
= \max_{0\leq j\leq N-1}\left(\left|\overline{S_n}(j)\right| + \left|\underline{S_n}(j)\right|\right) + o\left(\sqrt{nh^3\left|\log h\right|}\right),
\end{aligned}
$$

where for each fixed j, $\underline{S_n}(j)$ and $\overline{S_n}(j)$ are martingales with respect to the filtration $\mathcal{F}_n$ and are defined,

$$\underline{S_n}(j) = \sum_{i=1}^{n} \left(\underline{\varsigma}_{i,j}^{++} - \mathbb{E}\left[\underline{\varsigma}_{i,j}^{++} \,\middle|\, \mathcal{F}_{i-1} \right] \right)$$

$$\overline{S_n}(j) = \sum_{i=1}^{n} \left(\overline{\varsigma}_{i,j}^{++} - \mathbb{E}\left[\overline{\varsigma}_{i,j}^{++} \,\middle|\, \mathcal{F}_{i-1} \right] \right).$$

These martingales will be bounded by an exponential martingale inequality. Consider firstly the martingale $\underline{S_n}(j)$, its martingale differences are bounded $\left| \underline{\varsigma}_{i,j}^{++} - \mathbb{E}\left[\underline{\varsigma}_{i,j}^{++} \,\middle|\, \mathcal{F}_{i-1} \right] \right| \leq 2h \left| \mu^{(1)} \right|_\infty |K_3|_\infty =: C_b h$. Also using the Lipschitz property of Q and the bounded domain of K_3,

$$\mathbb{E}\left[\left(\underline{\varsigma}_{i,j}^{++} - \mathbb{E}\left[\underline{\varsigma}_{i,j}^{++} \,\middle|\, \mathcal{F}_{i-1} \right] \right)^2 \,\middle|\, \mathcal{F}_{i-1} \right]$$

$$\leq \int_{\mathbb{R}} \left(h\mu_{F,1}^+(u) K_3^+ \left(\frac{F(u) - t_j}{h} \right) \right)^2 f_X(u|\mathcal{F}_{i-1})\, du$$

$$\leq 2h^3 L_Q \left| K_3 \right|_\infty^2 \left| \mu_{F,1}^+ \right|_\infty |f_\eta|_\infty =: C_{cv} h^3.$$

Then, a martingale inequality for bounded differences given by Theorem 1.5A of de la Peña (1999) can be used to yield,

$$P\left(\underline{S_n}(j) \geq x \right) \leq \exp\left\{ -\frac{x}{2a} \sinh^{-1}\left(\frac{ax}{2y} \right) \right\}, \tag{41}$$

where $a = C_b h$ and $y = C_{cv} n h^3$. Furthermore if $ax/2y = o(1)$ then using a Taylor expansion of $\sinh^{-1}$,

$$\sinh^{-1}\left(\frac{ax}{2y} \right) = \frac{ax}{2y} + o\left(\left(\frac{ax}{2y} \right)^2 \right). \tag{42}$$

Now consider the chance that $\max_{1 \leq j \leq n} \underline{S_n}(j)$ exceeds the threshold with order $x = C_T \sqrt{nh^3 \left| \log h \right|}$ for some $C_T > 0$ which combined with $a = C_b h$ and $y = C_{cv} n h^3$ implies, $ax/2y = \mathcal{O}\left(\sqrt{\left| \log h \right| / nh} \right) = o(1)$ and by (41) and (42),

$$P\left(\underline{S_n}(j) \geq C_T \sqrt{nh^3 \left| \log h \right|} \right) \leq \exp\left\{ -\frac{C_T^2}{4C_{cv}} \left| \log h \right| + o(1) \right\} \tag{43}$$

So, fix $\epsilon > 0$ and use (43),

$$P\left(\max_{0 \leq j \leq N-1} \underline{S_n}(j) \geq C_T \sqrt{nh^3 \left| \log h \right|} \right) \leq \sum_{j=0}^{N-1} P\left(\underline{S_n}(j) \geq C_T \sqrt{nh^3 \left| \log h \right|} \right)$$

$$\leq N \exp\left\{ -\frac{C_T^2}{4C_{cv}} \left| \log h \right| \right\} \exp\left\{ o(1) \right\}$$

$$\leq C n^{\frac{1}{2}} h^{C_T^2/4C_{cv} - \frac{3}{2}}. \tag{44}$$

By choosing C_T large enough will ensure that $Cn^{\frac{1}{2}}h^{C_T^2/4C_{cv}-\frac{3}{2}} < \epsilon$. The similar conclusion can be reached that for any $\epsilon > 0$ there exists a finite constant C such that,

$$P\left(-\max_{0\le j\le N-1}\underline{S_n}(j) \ge C\sqrt{nh^3\left|\log h\right|}\right) < \epsilon. \tag{45}$$

Therefore, (44) and (45) ensure that,

$$\max_{0\le j\le N-1}\underline{S_n}(j) = \mathcal{O}_p(\sqrt{nh^3\left|\log h\right|}).$$

Using a comparable argument, the same conclusion can be reached for the $\overline{S_n}(j)$,

$$\max_{0\le j\le N-1}\left|\overline{S_n}(j)\right| = \mathcal{O}_p(\sqrt{nh^3\left|\log h\right|}).$$

Also, a similar technique can be used to bound the other martingale difference terms given in (40), details omitted. □

Proof of Theorem 2. To prove the Theorem we appeal to similar results that were shown by Kulik (2008); Wu and Mielniczuk (2002) by decomposing the stochastic terms into two parts, a martingale part and a LRD part. This is done by defining,

$$\chi_i(t) = \frac{(\zeta_i(t) - \mathbb{E}\zeta_1(t))\,\varepsilon_i + \gamma_i(t) - \mathbb{E}\gamma_1(t)}{\sqrt{n\,(\mathbb{V}\mathrm{ar}\zeta_1(t) + \mathbb{V}\mathrm{ar}\gamma_1(t))}}$$

and then decomposing the standardised estimator $\widehat{\kappa}_h(t)$ into two terms,

$$\begin{aligned}
\sqrt{nh^7}\left(\widehat{\kappa}_h(t) - \kappa_h(t)\right) &= \sqrt{nh^7}\left(Z_h(t) + b_h(t)\right) \\
&= \frac{1}{\sqrt{nh}}\left(\sum_{i=1}^n \zeta_i(t)\varepsilon_i + \sum_{i=1}^n(\gamma_i(t) - \mathbb{E}\gamma_1(t))\right) \\
&= \sqrt{h^{-1}\left(\mathbb{V}\mathrm{ar}\zeta_1(t) + \mathbb{V}\mathrm{ar}\gamma_1(t)\right)}\sum_{i=1}^n\chi_i(t) + \frac{\mathbb{E}\zeta_1(t)}{\sqrt{nh}}\sum_{i=1}^n\varepsilon_i.
\end{aligned} \tag{46}$$

The Theorem will follow by showing that either the first or last term on the RHS of (46) dominates under the bandwidth conditions (A1) or (A2) respectively. More specifically, it will be shown that the dominating term will follow a CLT and the other term converges to zero in probability; then Slutsky's Theorem completes the proof. Firstly consider the case where (A1) holds, then apply the martingale CLT of Brown (1971) to show,

$$\sum_{i=1}^n\chi_i(t) \xrightarrow{\mathscr{D}} \mathcal{N}(0,1). \tag{47}$$

Note that $\{\chi_i(t), \mathcal{G}_i\}$ form a martingale difference sequence. So it remains to check that the sum of the conditional variances converge in probability to the

unconditional sum and the Lindeberg condition holds. Before we prove the Lindeberg condition note that for $t \in (h, 1-h)$,

$$\mathbb{E}\zeta_1^2(t) = \int_{\mathbb{R}} \sigma^2(x) K_3^2 \left(\frac{F(x) - t}{h} \right) dF(x) = h \int_{-1}^{1} \sigma_F^2(t + hu) K_3^2(u) \ du. \quad (48)$$

Exploiting (10) and the assumption that $\sigma \in \mathscr{G}_r$,

$$\mathbb{E}\zeta_i(t) = h \int_{-1}^{1} \sigma_F(t + hu) K_3(u) \ du$$

$$= \frac{h^{(s \wedge r)+1}}{(s \wedge r)!} \int_{-1}^{1} \sigma_F^{(s \wedge r)}(t + \tau hu) u^{s \wedge r} K_3(u) \ du = h^{(s \wedge r)+1} \upsilon_*(t), \quad (49)$$

where $\tau \in (0, 1)$. Therefore, using (48) and (49),

$$\mathbb{V}\mathrm{ar}\zeta_1(t) = h \int_{-1}^{1} \sigma_F^2(t + hu) K_3^2(u) \ du$$

$$- \frac{h^{2(s \wedge r)+2}}{((s \wedge r)!)^2} \left(\int_{-1}^{1} \sigma_F^{(s \wedge r)}(t + \tau hu) u^{s \wedge r} K_3(u) \ du \right)^2.$$

Due to the fact that the bandwidth is assumed to follow $h \in (0, 1)$, there exists a h_0 such that for all $0 < h \le h_0 < 1$,

$$\mathbb{V}\mathrm{ar}\zeta_1(t) \ge \frac{h \inf_{x \in \mathbb{R}} |\sigma^2(x)|}{2} \int_{-1}^{1} K_3^2(u) \ du. \quad (50)$$

From (48), it follows $h^{-1}\mathbb{E}\zeta_1^2(t) \to \sigma_F^2(t) \int_{-1}^{1} K_3^2(u) \ du$ and similarly from (49), $h^{-\frac{1}{2}}\mathbb{E}\zeta_1(t) = o(1)$. Therefore, $h^{-1}\mathbb{V}\mathrm{ar}\zeta_1(t) \to \sigma_F^2(t) \int_{-1}^{1} K_3^2(u) \ du$. Also, the same argument applies for the $\gamma_i(t)$ term to yield,

$$h^{-1} \left(\mathbb{V}\mathrm{ar}\zeta_1(t) + \mathbb{V}\mathrm{ar}\gamma_1(t) \right) \xrightarrow{h \to 0} \upsilon^2(t).$$

Now the Lindeberg condition is shown to hold. Let $\epsilon > 0$ be arbitrary,

$$\sum_{i=1}^{n} \mathbb{E}\chi_i^2(t) \mathbb{1}_{\{|\chi_i(t)| > \epsilon\}} = n\mathbb{E}\chi_1^2(t) \mathbb{1}_{\{|\chi_1(t)| > \epsilon\}}$$

$$= \frac{\mathbb{E}\left[\left(\varepsilon_1 \left(\zeta_1(t) - \mathbb{E}\zeta_1(t) \right) + \gamma_1(t) - \mathbb{E}\gamma_1(t) \right)^2 \mathbb{1}_{A_n} \right]}{\mathbb{V}\mathrm{ar}\zeta_1(t) + \mathbb{V}\mathrm{ar}\gamma_1(t)}. \quad (51)$$

where $A_n = \{|\chi_1(t)| > \epsilon\}$. The size of this set can be maximised using (50),

$$A_n \subset \left\{ 2 |K_3|_{\infty} |\varepsilon_1| \left(|\sigma|_{\infty} + |\mu|_{\infty} \right) > \epsilon \sqrt{n\mathbb{V}\mathrm{ar}\zeta_1(t)} \right\}$$

$$\subset \left\{ 2 |K_3|_{\infty} |\varepsilon_1| \left(|\sigma|_{\infty} + |\mu|_{\infty} \right) > \epsilon \sqrt{nh \frac{\inf_{x \in \mathbb{R}} |\sigma^2(x)|}{2} \int_{-1}^{1} K_3^2(u) \ du} \right\}. \quad (52)$$

Using the fact that $nh \to \infty$ and $h \to 0$ as $n \to \infty$ we see that $A_n \to \emptyset$, the empty set. Consequently with (51), (52) and $n\mathbb{E}\chi_1^2(t) < \infty$ imply that,

$$\sum_{i=1}^{n} \mathbb{E}\chi_i^2(t)\mathbb{1}_{\{|\chi_i(t)|>\epsilon\}} \stackrel{n\to\infty}{\longrightarrow} 0,$$

and the Lindeberg condition holds. By a consequence of (11), let $\epsilon > 0$ be arbitrary,

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\right| > \epsilon\right) \le \frac{1}{n^2\epsilon^2}\mathbb{V}\mathrm{ar}\left(\sum_{i=1}^{n}\varepsilon_i\right) \le \frac{C_1^2 n^{-\alpha}L^2(n)}{\epsilon^2},$$

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2 - 1\right| > \epsilon\right) \le \frac{1}{n^2\epsilon^2}\mathbb{V}\mathrm{ar}\left(\sum_{i=1}^{n}\varepsilon_i^2\right) \le \frac{\left(C_2^2 n^{-1} \vee C_3^2 n^{-2\alpha}L^2(n)\right)}{\epsilon^2},$$

with both of the above equations being $o(1)$. So, the sum of the conditional variances to converge in probability to one:

$$\sum_{i=1}^{n}\mathbb{E}\left[\chi_i^2(t)\big|\mathcal{G}_{i-1}\right] = \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2\frac{\mathbb{V}\mathrm{ar}\zeta_1(t)}{\mathbb{V}\mathrm{ar}\zeta_1(t) + \mathbb{V}\mathrm{ar}\gamma_1(t)} + \frac{\mathbb{V}\mathrm{ar}\gamma_1(t)}{\mathbb{V}\mathrm{ar}\zeta_1(t) + \mathbb{V}\mathrm{ar}\gamma_1(t)}$$

$$+ \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\frac{2\mathbb{C}\mathrm{ov}\left(\zeta_1(t),\gamma_1(t)\right)}{\mathbb{V}\mathrm{ar}\zeta_1(t) + \mathbb{V}\mathrm{ar}\gamma_1(t)}$$

$$\stackrel{p}{\longrightarrow} 1,$$

and by the martingale CLT, (47) follows.

Now we show that the last term on the RHS of (46) converges in probability to zero. Consider an arbitrary $\epsilon > 0$, then using (49) and (11),

$$P\left(\left|\frac{\mathbb{E}\zeta_1(t)}{\sqrt{nh}}\sum_{i=1}^{n}\varepsilon_i\right| > \epsilon\right) \le \frac{\left(\mathbb{E}\zeta_1(t)\right)^2}{\epsilon^2 nh}\mathbb{V}\mathrm{ar}\left(\sum_{i=1}^{n}\varepsilon_i\right)$$

$$\le Ch^{2(s\wedge r)+1}n^{1-\alpha}L^2(n)$$

$$= o(1),$$

and the last line follows by the bandwidth restriction given in (A1). Thus, the proof of the first claim under the 'small' bandwidth scenario holds.

Consider now the 'large' bandwidth scenario. Using (46), (47) and (49),

$$\widehat{\kappa}_h(t) - \kappa_h(t) = \mathcal{O}_p\left(n^{-\frac{1}{2}}h^{-\frac{7}{2}}\right) + \frac{v_*(t)}{nh^{3-(s\wedge r)}}\sum_{i=1}^{n}\varepsilon_i. \tag{53}$$

Also, from Ho and Hsing (1997), it is known that

$$\frac{1}{n^{1-\frac{\alpha}{2}}L(n)}\sum_{i=1}^{n}\varepsilon_i \stackrel{\mathscr{D}}{\longrightarrow} \mathcal{N}(0, C_1^2). \tag{54}$$

Therefore, normalising the expression on (53),

$$\frac{n^{\frac{\alpha}{2}} h^{3-(s \wedge r)}}{L(n)} \left(\widehat{\kappa}_h(t) - \kappa_h(t)\right) = \mathcal{O}_p \left(\frac{h^{-\frac{1}{2}-(s \wedge r)} n^{-\frac{1-\alpha}{2}}}{L(n)} \right) + \frac{\upsilon_*(t)}{n^{1-\frac{\alpha}{2}} L(n)} \sum_{i=1}^{n} \varepsilon_i,$$

and the result follows from (A2) and (54) with Slutsky's Theorem $\qquad\square$

Proof of Theorem 3. First break down the estimator into its separate martingale and LRD part in a similar fashion to the method employed in the proof of Theorem 2. Using (30), apply Lemma 2,

$$\begin{aligned}
\widehat{\kappa}_h(t) &- \kappa_h(t) \\
&= \frac{1}{nh^4} \sum_{i=1}^{n} \left(\gamma_i(t) - \mathbb{E}\gamma_i(t) + \zeta_i(t)\varepsilon_i \right) \\
&= \mathcal{O}_p \left(\sqrt{\frac{|\log h|}{nh^5}} \right) + \frac{1}{nh^4} \sum_{i=1}^{n} \left(\mathbb{E}\left[\gamma_i^*(t)|\, \mathcal{F}_{i-1}\right] - \mathbb{E}\gamma_i^*(t) \right) \\
&\quad + \frac{1}{nh^4} \sum_{i=1}^{n} \left(\mu_F(t) K_3 \left(\frac{F(X_i)-t}{h} \right) + \zeta_i(t)\varepsilon_i \right) \\
&= \frac{\mu_F(t)}{nh^4} \sum_{i=1}^{n} \left(K_3 \left(\frac{F(X_i)-t}{h} \right) - \mathbb{E}\left[K_3 \left(\frac{F(X_i)-t}{h} \right) \middle| \mathcal{F}_{i-1} \right] \right) \\
&\quad + \frac{1}{nh^4} \sum_{i=1}^{n} \zeta_i(t)\varepsilon_i + \frac{1}{nh^4} \sum_{i=1}^{n} \left(\mathbb{E}\left[\gamma_i(t)|\, \mathcal{F}_{i-1}\right] - \mathbb{E}\gamma_i(t) \right) + \mathcal{O}_p \left(\sqrt{\frac{|\log h|}{nh^5}} \right)
\end{aligned}$$
$$\tag{55}$$

Define the standardised stochastic terms,

$$\Delta_i(t) := \frac{\zeta_i(t)\varepsilon_i + \mu_F(t) \left(K_3 \left(\frac{F(X_i)-t}{h} \right) - \mathbb{E}\left[K_3 \left(\frac{F(X_i)-t}{h} \right) \middle| \mathcal{F}_{i-1} \right] \right)}{\upsilon(t)\sqrt{nh}}.$$

Then in a similar fashion to the Proof of Theorem 2 it will be shown by the martingale CLT of Brown (1971) that,

$$\sum_{i=1}^{n} \Delta_i(t) \xrightarrow{\mathscr{D}} \mathcal{N}(0,1). \tag{56}$$

Indeed, $\Delta_i(t)$ is a martingale difference sequence with respect to the σ-fields $\{\mathcal{F}_i\}$. Thus we need to check that the Lindeberg condition holds and that the sum of the conditional variances converge in probability to 1. First, focus on the convergence of the conditional variances. The conditional variances can be

broken into two parts,

$$
\sum_{i=1}^{n} \mathbb{E}\left[\Delta_i^2(t)\middle|\mathcal{F}_{i-1}\right]
$$

$$
= \frac{\mu_F^2(t)}{nh\,\upsilon^2(t)} \sum_{i=1}^{n} \mathbb{E}\left[\left(K_3\left(\frac{F(X_i)-t}{h}\right) - \mathbb{E}\left[K_3\left(\frac{F(X_i)-t}{h}\right)\middle|\mathcal{F}_{i-1}\right]\right)^2\middle|\mathcal{F}_{i-1}\right]
$$

$$
+ \sum_{i=1}^{n} \frac{\mathbb{E}\left[\zeta_i^2(t)\middle|\mathcal{F}_{i-1}\right]}{nh\,\upsilon^2(t)}. \tag{57}
$$

Dealing with the second term on the RHS of (57), use Lemma 1 of Zhao and Wu (2008),

$$
\frac{1}{nh} \sum_{i=1}^{n} \mathbb{E}\left[\zeta_i^2(t)\middle|\mathcal{F}_{i-1}\right] = \frac{1}{nh} \sum_{i=1}^{n} \mathbb{E}\zeta_i^2(t) + \frac{1}{nh} \sum_{i=1}^{n} \left(\mathbb{E}\left[\zeta_i^2(t)\middle|\mathcal{F}_{i-1}\right] - \mathbb{E}\zeta_i^2(t)\right)
$$

$$
= \int_{-\frac{t}{h}}^{\frac{1-t}{h}} \sigma_F^2(t+hx)K_3^2(x)\,dx + \mathcal{O}_p(n^{-\frac{\alpha}{2}}L(n))
$$

$$
= \sigma_F^2(t)\int_{-1}^{1} K_3^2(x)\,dx + \mathcal{O}(h^2) + \mathcal{O}_p(n^{-\frac{\alpha}{2}}L(n)) \tag{58}
$$

A bound is required for $\mathbb{E}\left[K_3\left(\frac{F(X_i)-t}{h}\right)\middle|\mathcal{F}_{i-1}\right]^2$ to deal with the first term of (57). Define $X_{i,i-1} := X_i - \eta_i = \mu_X + \sum_{j=1}^{\infty} c_j\eta_{i-j}$ and $Z_i := s_X^{-1}(X_{i,i-1} - \mu_X)$ and define $\widetilde{f}_\eta(x) := f_X\left(x\middle|\mathcal{F}_{i-1}\right) = f_\eta(x - X_{i,i-1})$ and $g(x) = 1/x$. Then $X_{i,i-1}$ and Z_i are $\mathcal{F}_{i-1}$-measurable and for all $t \in (h, 1-h)$ the conditional expectation can be evaluated as follows.

$$
\mathbb{E}\left[K_3\left(\frac{F(X_i)-t}{h}\right)\middle|\mathcal{F}_{i-1}\right]
$$

$$
= \int_{\mathbb{R}} K_3\left(\frac{F(v)-t}{h}\right) f_X\left(v\middle|\mathcal{F}_{i-1}\right)\,dv
$$

$$
= h\int_{-1}^{1} K_3\left(x\right)\left(\widetilde{f}_\eta \circ Q\right)(t+hx)\left(g \circ f_X \circ Q\right)(t+hx)\,dx. \tag{59}
$$

Use a Taylor expansion of the composite functions, $p(t) := \left(\widetilde{f}_\eta \circ Q\right)(t)$ and $q(t) := \left(g \circ f_X \circ Q\right)(t)$ by using the Faà di Bruno chain rule given in (37); starting with the latter Taylor expansion,

$$
(g \circ f_X \circ Q)(t+hx)
$$

$$
= \sum_{j=0}^{s\wedge r-1} \frac{h^j x^j\,(g \circ f_X \circ Q)^{(j)}(t)}{j!} + \frac{h^{s\wedge r}x^{s\wedge r}\,(g \circ f_X \circ Q)^{(s\wedge r)}(t+\tau hx)}{(s\wedge r)!},
$$

$$
\tag{60}
$$

where $|\tau| < 1$. The intermediate derivatives for $j = 0, 1, \ldots, s \wedge r$ are given by

$$
(g \circ f_X \circ Q)^{(j)}(t) = \sum_{\mathbf{k} \in \mathcal{K}_j} (-1)^k k! \left((f_X \circ Q)(t)\right)^{-(k+1)} \prod_{\ell=1}^{j} \left(\frac{(f_X \circ Q)^{(\ell)}(t)}{j!} \right)^{k_\ell}
$$
$$
= \mathcal{O}(1)
$$

due to restrictions imposed in Assumption (B). Similarly,

$$
\left(\widetilde{f}_\eta \circ Q\right)(t+hx) = \sum_{j=0}^{s \wedge r-1} \frac{h^j x^j}{j!} \left(\widetilde{f}_\eta \circ Q\right)^{(j)}(t) + \frac{h^{s \wedge r} x^{s \wedge r}}{(s \wedge r)!} \left(\widetilde{f}_\eta \circ Q\right)^{(s \wedge r)}(t+\delta hx)
$$
(61)

where $|\delta| \leq 1$. Therefore, using (61) and (60) in (59) with the vanishing moment condition (10) implies that,

$$
\mathbb{E}\left[K_3\left(\frac{F(X_i)-t}{h}\right)\Big|\mathcal{F}_{i-1}\right]
$$
$$
= h^{s \wedge r+1}\left\{ \sum_{\substack{j=0 \\ j+\ell \geq s \wedge r}}^{s \wedge r-1} \sum_{\ell=0}^{s \wedge r-1} h^{\ell+j-s \wedge r} \frac{p^{(j)}(t)q^{(\ell)}(t)}{j!\ell!} \int_{-1}^{1} x^{\ell+j} K_3(x)\,dx \right.
$$
$$
+ \sum_{j=0}^{s \wedge r-1} \frac{p^{(j)}(t)h^j}{(s \wedge r)!j!} \int_{-1}^{1} x^{s \wedge r+j} K_3(x) q^{(s \wedge r)}(t+\tau hx)\,dx
$$
$$
+ \sum_{\ell=0}^{s \wedge r-1} \frac{q^{(\ell)}(t)h^\ell}{(s \wedge r)!\ell!} \int_{-1}^{1} x^{s \wedge r+\ell} K_3(x) p^{(s \wedge r)}(t+\delta hx)\,dx
$$
$$
\left. + \frac{h^{s \wedge r}}{((s \wedge r)!)^2} \int_{-1}^{1} x^{2(s \wedge r)} K_3(x) q^{(s \wedge r)}(t+\tau hx) p^{(s \wedge r)}(t+\delta hx)\,dx \right\}.
$$

However, by Assumption (B), $f_\eta^{(j)}$ and Q are Lipschitz continuous for $j = 0, \ldots, s$ and therefore bounded. Consequently $p^{(j)}$ and $q^{(j)}$ are also bounded which means that uniformly in t,

$$
\widetilde{K_3}(X_{i,i-1}, t) := \mathbb{E}\left[K_3\left(\frac{F(X_i)-t}{h}\right)\Big|\mathcal{F}_{i-1}\right] < Ch^{s \wedge r+1} \quad \text{a.s.} \qquad (62)
$$

Define, $g(X_{i,i-1}, t) := \widetilde{K_3}^2(X_{i,i-1}, t) - \mathbb{E}\widetilde{K_3}^2(X_{i,i-1}, t)$, then $\mathbb{E}g(X_{i,i-1}, t) = 0$ and by Jensen's Inequality $\mathbb{E}\widetilde{K_3}(X_{i,i-1}, t)^2 < \infty$. It will be shown by an application of Theorem 1 of Wu (2007) that $\sum_{i=1}^{n} g(X_{i,i-1}, t) = \mathcal{O}_p\left(h^{s \wedge r+2} n^{1-\frac{\alpha}{2}} L(n)\right)$. Define, $\vartheta_i = \sup_{t \in (h, 1-h)} \|\mathbb{E}\left[g(X_{i,i-1}, t)|\mathcal{F}_0\right] - \mathbb{E}\left[g(X_{i,i-1}, t)|\mathcal{F}_{-1}\right]\|$ to measure the physical dependence. To bound ϑ_i, let η_0' be an i.i.d. copy of η_0 and define $X_{i,i-1}^* = X_{i,i-1} - c_i\eta_0 + c_i\eta_0'$ with the associated sigma field $\mathcal{F}_i^* = \sigma\left(\eta_i, \eta_{i-1}, \ldots, \eta_1, \eta_0', \eta_1, \ldots; \varepsilon_1, \ldots, \varepsilon_i\right)$. Then by Theorem 1 of Wu (2005) it was shown that there is a bound $\vartheta_i \leq \sup_{t \in (h, 1-h)} \|g(X_{i,i-1}, t) - g(X_{i,i-1}^*, t)\|$.

Using this, (62) and the Lipschitz property of f_η it will be shown that $\vartheta_i < Ch^{s\wedge r+2}i^{-\beta}L(i)$,

$$
\begin{aligned}
\vartheta_i &\leq \sup_{t\in(h,1-h)} \left\| g(X_{i,i-1},t) - g(X_{i,i-1}^*,t) \right\| \\
&\leq Ch^{s\wedge r+1} \sup_{t\in(h,1-h)} \left\| \widetilde{K_3}\left(X_{i,i-1},t\right) - \widetilde{K_3}\left(X_{i,i-1}^*,t\right) \right\| \\
&\leq Ch^{s\wedge r+1} \sup_{t\in(h,1-h)} \int_{\mathbb{R}} \left| K_3\left(\frac{F(u)-t}{h}\right) \right| du \left\| X_{i,i-1} - X_{i,i-1}^* \right\| \\
&\leq Ch^{s\wedge r+2} \left\| \eta_0 - \eta_0' \right\| c_i = Ch^{s\wedge r+2} i^{-\beta} L(i),
\end{aligned}
$$

where the last line follows due to the Lipschitz property of Q and the bounded domain of K_3. Then by Theorem 1 of Wu (2007) and Karamata's Theorem, $\left\| \sum_{i=1}^n g(X_{i,i-1},t) \right\|^2 = \mathcal{O}\left(h^{2(s\wedge r)+4}n^{2-\alpha}L^2(n)\right)$. Using this and (62),

$$
\begin{aligned}
\frac{1}{nh}\sum_{i=1}^n \widetilde{K_3}^2(X_{i,i-1},t) &= \frac{1}{nh}\sum_{i=1}^n g(X_{i,i-1},t) + \frac{1}{nh}\sum_{i=1}^n \mathbb{E}\widetilde{K_3}^2(X_{i,i-1},t) \\
&= \mathcal{O}_p\left(h^{s\wedge r+1}n^{-\frac{\alpha}{2}}L(n)\right) + \mathcal{O}\left(h^{2(s\wedge r)+2}\right) \\
&= o_p(1)
\end{aligned}
\tag{63}
$$

Then the first term on the RHS of (57) can be bounded by (63) and a similar application of Lemma 1 of Zhao and Wu (2008),

$$
\begin{aligned}
\frac{\mu_F^2(t)}{nh}\sum_{i=1}^n &\mathbb{E}\left[\left(K_3\left(\frac{F(X_i)-t}{h}\right) - \mathbb{E}\left[K_3\left(\frac{F(X_i)-t}{h}\right)\Big|\mathcal{F}_{i-1}\right]\right)^2\Big|\mathcal{F}_{i-1}\right] \\
&= \frac{\mu_F^2(t)}{nh}\sum_{i=1}^n \left\{ \mathbb{E}K_3^2\left(\frac{F(X_i)-t}{h}\right) + \mathbb{E}\left[K_3\left(\frac{F(X_i)-t}{h}\right)\Big|\mathcal{F}_{i-1}\right]^2 \right. \\
&\qquad \left. + \left(\mathbb{E}\left[K_3^2\left(\frac{F(X_i)-t}{h}\right)\Big|\mathcal{F}_{i-1}\right] - \mathbb{E}K_3^2\left(\frac{F(X_i)-t}{h}\right)\right) \right\} \\
&= \mu_F^2(t)\int_{-1}^1 K_3^2(x)\,dx + \mathcal{O}_p(n^{-\frac{\alpha}{2}}L(n)) + \mathcal{O}\left(h^2\right)
\end{aligned}
\tag{64}
$$

Substituting (64) and (58) into (57) implies that,

$$
\sum_{i=1}^n \mathbb{E}\left[\Delta_i^2(t)\big|\mathcal{F}_{i-1}\right] \xrightarrow{p} 1.
$$

For the Lindeberg condition, let $\epsilon > 0$ and define $A_n = \{|\Delta_1(t)| > \epsilon\}$, then similar to the procedure used in the Proof of Theorem 2, it can be shown that $A_n \to \emptyset$ and the Lindeberg condition holds. Thus by the martingale CLT, (56) holds and by using (B1) in the decomposition given in (55) the result follows by Slutsky's Theorem. □

Proof of Theorem 4. Again, use the decomposition (55) used in the Proof of Theorem 3. Then, define the standardised process,

$$\Upsilon_i(t) := \frac{\mathbb{E}\left[\gamma_i(t)\mid \mathcal{F}_{i-1}\right] - \mathbb{E}\gamma_i(t)}{h^4 n^{1-\frac{\alpha}{2}} L(n)\mathcal{H}_1(t)}.$$

It will be shown via use of a Hermite expansion of the LRD variables that,

$$\sum_{i=1}^n \Upsilon_i(t) \xrightarrow{\mathscr{D}} \mathcal{N}(0, C_1^2). \tag{65}$$

To do this, split the LRD variable X_i into two parts, $X_i = \eta_i + X_{i,i-1}$. Define the standardised version of $X_{i,i-1}$, $Z_i = s_X^{-1}(X_{i,i-1} - \mu_X)$, $Z_i \sim \mathcal{N}(0,1)$. Notice that $\Upsilon_i(t)$ and Z_i are both $\mathcal{F}_{i-1}$-measurable and define $G(Z_i, t) := \mathbb{E}\left[\gamma_i(t)\mid \mathcal{F}_{i-1}\right] - \mathbb{E}\gamma_i(t)$. Then clearly, $\mathbb{E}G(Z_i, t) = 0$ and by Jensen's inequality, $\mathbb{E}G(Z_i, t)^2 < \infty$. So by Taqqu (1975), $G(Z_i, t)$ can be re-expressed by its Hermite expansion,

$$G(Z_i, t) = \sum_{m=1}^\infty \frac{a_m}{m!} H_m(Z_i)$$

where $a_m = \mathbb{E}\left[H_m(Z_1)G(Z_1, t)\right]$ is the mth Hermite coefficient. For our case it is assumed that $a_1 \neq 0$. Evaluating a_1,

$$\begin{aligned}
a_1 &= \mathbb{E}\left[Z_1 G(Z_1, t)\right] \\
&= \mathbb{E}\left[Z_1 \frac{1}{\sigma_\eta} \int_{\mathbb{R}} \mu(u + \mu_X + s_X Z_1) K_3\left(\frac{\Phi(u + s_X Z_1) - t}{h}\right) \phi\left(\frac{u}{\sigma_\eta}\right) du\right] \\
&= \frac{1}{\sigma_\eta} \int_{\mathbb{R}} \int_{\mathbb{R}} z\mu(u + \mu_X + s_X z) K_3\left(\frac{\Phi(u + s_X z) - t}{h}\right) \phi(z)\phi\left(\frac{u}{\sigma_\eta}\right) dz\, du \\
&= \frac{h}{s_X^2 \sigma_\eta} \int_{\mathbb{R}} \int_{-\frac{t}{h}}^{\frac{1-t}{h}} \frac{\Phi^{-1}(t + hw) - u}{\phi(\Phi^{-1}(t + hw))} \mu_F(t + hw) K_3(w) \\
&\qquad\qquad \times \phi\left(\frac{\Phi^{-1}(t + hw) - u}{s_X}\right) \phi\left(\frac{u}{\sigma_\eta}\right) dw\, du.
\end{aligned}$$

By exploiting the Faà di Bruno formula further, it can be shown via Taylor expansions that the asymptotic behaviour of a_1 satisfies,

$$a_1 \sim \frac{h^4 \kappa_h(t)}{s_X^3 \sigma_\eta \phi\left(\Phi^{-1}(t)\right)} \int_{\mathbb{R}} \phi\left(\frac{\Phi^{-1}(t) - u}{s_X}\right)\left(\Phi^{-1}(t) - u\right)\phi\left(\frac{u}{\sigma_\eta}\right) du = h^4 \mathcal{H}_1(t)$$

From Corollary 5.1 of Taqqu (1975),

$$\sum_{i=1}^n \Upsilon_i(t) \sim \frac{1}{n^{1-\frac{\alpha}{2}} L(n)} \sum_{i=1}^n Z_i \xrightarrow{\mathscr{D}} \mathcal{N}(0, C_1^2)$$

Therefore (65) holds by Slutsky's Theorem in the decomposition given in (55) in conjuction with (56), (65) and (B2). $\qquad\square$

Proof of Theorem 5. First, choose distinct integers $0 \leq j_1, j_2, \ldots, j_k \leq m_n$ for some fixed $k \in \mathbb{N}$. We adapt the proof of Theorem 5 of Zhao and Wu (2008) to our case. The proof of their result was reliant on another result given by Theorem 1 of Grama and Haeusler (2006) which requires a martingale difference sequence that has third order moments. We obtain such a sequence below. Define,

$$S_{n,k}^A(\mathbf{t}) = \sum_{i=1}^n \Psi_i(\mathbf{t}) = \left[S_n^A(t_{j_1}), S_n^A(t_{j_2}), \ldots, S_n^A(t_{j_k}) \right]^T.$$

and the associated sigma field $\mathcal{A}_i = \sigma(X_i, \ldots, X_1; \eta_{i+1}, \eta_i, \ldots)$. Then the sequence $\left\{ S_{n,k}^A(\mathbf{t}), \mathcal{A}_n \right\}_{n \in \mathbb{Z}}$ is a martingale since $\mathbb{E} K_3 \left(\frac{F(X_i)-t}{h} \right) = 0$ for all $t \in (h, 1-h)$. Let $\mathcal{Q}$ be the quadratic characteristic matrix of $S_{n,k}^A$, that is,

$$\mathcal{Q} = \sum_{i=1}^n \mathbb{E} \left[\Psi_i(\mathbf{t}) \Psi_i(\mathbf{t})^T \big| \mathcal{A}_{i-1} \right] := (\mathcal{Q}_{rr'})_{1 \leq r, r' \leq k}.$$

$$\mathcal{Q}_{rr'} = \frac{\mathbb{E} \left(K_3 \left(\frac{F(X_1)-t_{j_r}}{h} \right) K_3 \left(\frac{F(X_1)-t_{j_{r'}}}{h} \right) \right)}{nh \sqrt{\upsilon(t_{j_r}) \upsilon(t_{j_{r'}})}} \left\{ \sigma^2 \sum_{i=1}^n \varepsilon_i^2 \right.$$
$$\left. + \left(\mu_F(t_{j_r}) + \mu_F(t_{j_{r'}}) \right) \sigma \sum_{i=1}^n \varepsilon_i + n \mu_F(t_{j_r}) \mu_F(t_{j_{r'}}) \right\}.$$

However, by construction, if $r \neq r'$, then $|t_{j_r} - t_{j_{r'}}| \geq 2h$ and the kernel function $K_3 : [-1,1] \longrightarrow \mathbb{R}$ which implies that the set $\left\{ h^{-1} |F(x) - t_{j_r}| \leq 1 \right\} \cap \left\{ h^{-1} |F(x) - t_{j_{r'}}| \leq 1 \right\}$ is empty for all $x \in \mathbb{R}$. Therefore when $r \neq r'$, $\mathcal{Q}_{rr'} = 0$. If $r = r'$, then by (11),

$$\mathcal{Q}_{rr} = \frac{1}{n \left(\sigma^2 + \mu_F^2(t_{j_r}) \right)} \sum_{i=1}^n \left(\sigma^2 \varepsilon_i^2 + 2\mu_F(t_{j_r}) \sigma \varepsilon_i + \mu_F^2(t_{j_r}) \right).$$

$$\| \mathcal{Q}_{rr} - 1 \| \leq \frac{1}{n \left(\sigma^2 + \mu_F^2(t_{j_r}) \right)} \left(\sigma^2 \left\| \sum_{i=1}^n \left(\varepsilon_i^2 - 1 \right) \right\| + 2 |\mu_F(t_{j_r})| \left\| \sum_{i=1}^n \sigma \varepsilon_i \right\| \right)$$
$$= \mathcal{O} \left(n^{-\frac{\alpha}{2}} L(n) \right).$$

Let $(u_{rr'})_{1 \leq r, r' \leq k}$ be the $k \times k$ identity matrix. Then by the above argument $\mathbb{E} |\mathcal{Q}_{rr'} - u_{rr'}|^{\frac{3}{2}} = \mathcal{O} \left(n^{-\frac{3\alpha}{4}} L^{\frac{3}{2}}(n) \right)$ uniformly over $1 \leq r, r' \leq k$. Additionally, $\sum_{i=1}^n \mathbb{E} |\Psi_i(t)|^3 = \mathcal{O}(n^{-\frac{1}{2}} h^{-\frac{1}{2}})$. Combining the two yields, $\sum_{i=1}^n \mathbb{E} |\Psi_i(\mathbf{t})|^3 + \mathbb{E} |\mathcal{Q}_{rr'} - u_{rr'}|^{\frac{3}{2}} = \mathcal{O}((nh)^{-\frac{1}{2}} + n^{-\frac{3\alpha}{4}} L^{3/2}(n))$. Considering the asymptotic behaviour of (23), $(1 + B_{m_n}(x))^4 \exp \left\{ \frac{B_{m_n}^2(x)}{2} \right\} = \mathcal{O} \left(h^{-1} |\log h|^{\frac{3}{2}} \right)$ and using (22) it follows that $(1 + B_{m_n}(x))^4 \exp \left\{ \frac{B_{m_n}^2(x)}{2} \right\} \Lambda_n \to 0$. Therefore the same framework and argument applies that was used in the proof of Theorem 5 of Zhao and Wu (2008) and the result follows. $\square$

Proof of Theorem 6. The proof of the Theorem uses similar moderate deviation inequalities from Grama and Haeusler (2006) that were used in the proof of Theorem 5. However, slight modification is needed. Firstly fix $k \in \mathbb{N}$ and choose distinct integers $1 \le j_1, j_2, \ldots, j_k < m_n$ then modify $S_n^B(t)$ to obtain a martingale by adding and subtracting the conditional expectation. Define,

$$S_n^{B^*}(t) := \sum_{i=1}^{n} \Xi_i^*(t) = \sum_{i=1}^{n} \left(\Xi_i(t) - \frac{\mu_F(t)}{\upsilon(t)\sqrt{nh}} \mathbb{E}\left[K_3 \left(\frac{F(X_i) - t}{h} \right) \middle| \mathcal{F}_{i-1} \right] \right).$$

With this definition, $\left\{ S_n^{B^*}(t), \mathcal{F}_n \right\}_{n \in \mathbb{Z}^+}$ is a martingale and

$$S_n^B(t) = S_n^{B^*}(t) + \frac{\mu_F(t)}{\upsilon(t)\sqrt{nh}} \sum_{i=1}^{n} \mathbb{E}\left[K_3 \left(\frac{F(X_i) - t}{h} \right) \middle| \mathcal{F}_{i-1} \right]. \qquad (66)$$

Define $R_n^{B^*}(t) := \mu_F(t)/(\upsilon(t)\sqrt{nh}) \sum_{i=1}^{n} \mathbb{E}\left[K_3 \left(\frac{F(X_i)-t}{h} \right) \middle| \mathcal{F}_{i-1} \right]$. The proof of the result will follow if it can be shown that the supremum over T_n for both terms of (66) are $o_p \left(\sqrt{|\log h|} \right)$. Starting with the latter term, from (62) and (25) it follows that $n^{-\frac{1}{2}} h^{-\frac{1}{2}} \sum_{i=1}^{n} \mathbb{E}\left[K_3 \left(\frac{F(X_i)-t}{h} \right) \middle| \mathcal{F}_{i-1} \right] = \mathcal{O}_{\text{a.s.}} \left(n^{\frac{1}{2}} h^{s \wedge r + \frac{1}{2}} \right)$. Also, since k is fixed $\mathbb{E}\left[\sup_{t_{j_r}} \left| R_n^{B^*}(t) \right| \right]^2 = \mathcal{O}\left(n h^{2(s \wedge r)+1} \right) = o(1)$ so by the Chebyshev inequality,

$$P \left(\sup_{t_{j_r}} \left| R_n^{B^*}(t_{j_r}) \right| \ge \sqrt{2 |\log h|} \right) = o(1). \qquad (67)$$

Now turn attention to the first term on the RHS of (66) and apply a moderate deviation martingale result from Corollary 2 of Grama and Haeusler (2006).

To be able to use their Corollary a bound is needed on the trace norm of the quadratic characteristic matrix of the martingale and a bound on the Euclidean norm of the martingale difference sequence. These will be investigated, starting with the former. For a symmetric $k \times k$ matrix U define the trace norm $\|U\|_{\text{tr}} := \sum_{i=1}^{p} |e_i|$ where e_i are the eigenvalues of U. For a sequence $\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$ define the usual Euclidean norm of a sequence $|\mathbf{x}|_2 := \left(\sum_i x_i^2 \right)^{\frac{1}{2}}$. Now let $\mathcal{Q}$ be

the quadratic characteristic matrix of $S_{n,k}^{B^*}(\mathbf{t})$, that is,

$$\mathcal{Q} = \sum_{i=1}^{n} \mathbb{E}\left[\Xi_i^*(\mathbf{t})\Xi_i^*(\mathbf{t})^T \middle| \mathcal{F}_{i-1}\right] := (\mathcal{Q}_{rr'})_{1 \le r, r' \le k}.$$

$$\mathcal{Q}_{rr'} = \sum_{i=1}^{n} \mathbb{E}\left[\Xi_i^*(t_{j_r})\Xi_i^*(t_{j_{r'}}) \middle| \mathcal{F}_{i-1}\right]$$

$$= \sum_{i=1}^{n} \frac{\mathbb{E}\left[\sigma^2(X_i)K_3\left(\frac{F(X_i)-t_{j_r}}{h}\right)K_3\left(\frac{F(X_i)-t_{j_{r'}}}{h}\right)\middle| \mathcal{F}_{i-1}\right]}{nhv(t_{j_r})v(t_{j_{r'}})}$$

$$+ \frac{\mu_F\left(t_{j_r}\right)\mu_F\left(t_{j_{r'}}\right)}{nhv(t_{j_r})v(t_{j_{r'}})}\sum_{i=1}^{n}\mathbb{E}\left[K_3\left(\frac{F(X_i)-t_{j_r}}{h}\right)K_3\left(\frac{F(X_i)-t_{j_{r'}}}{h}\right)\middle| \mathcal{F}_{i-1}\right]$$

$$- \frac{\mu_F\left(t_{j_r}\right)\mu_F\left(t_{j_{r'}}\right)}{nhv(t_{j_r})v(t_{j_{r'}})}\sum_{i=1}^{n}\widetilde{K_3}^2(X_{i,i-1}, t). \tag{68}$$

By a similar domain argument that was presented in the proof of Theorem 5, if $r \ne r'$, the first and second terms on the RHS of (68) are zero. Using this fact with (62) it follows that,

$$\mathcal{Q}_{rr'} = \mathcal{O}_{\text{a.s.}}\left(h^{2(s \wedge r)+1}\right), \tag{69}$$

for $r \ne r'$. On the other hand, if $r = r'$, then,

$$\mathcal{Q}_{rr} = \frac{1}{nhv^2(t_{j_r})}\sum_{i=1}^{n}\left\{\mathbb{E}\left[\sigma^2(X_i)K_3^2\left(\frac{F(X_i)-t_{j_r}}{h}\right)\middle| \mathcal{F}_{i-1}\right]\right.$$

$$+ \mu_F^2\left(t_{j_r}\right)\mathbb{E}\left[K_3^2\left(\frac{F(X_i)-t_{j_r}}{h}\right)\middle| \mathcal{F}_{i-1}\right]$$

$$\left. - \mu_F^2\left(t_{j_r}\right)\mathbb{E}\left[K_3\left(\frac{F(X_i)-t_{j_r}}{h}\right)\middle| \mathcal{F}_{i-1}\right]^2\right\}. \tag{70}$$

Define the eigenvalues of $\mathcal{Q}$ to be $e_1 \le e_2 \le \ldots \le e_k$. To evaluate the trace norm of $\mathcal{Q}$ consider,

$$\|\mathcal{Q} - I\|_{\text{tr}}^2 \le \sum_{r=1}^{k}(e_r - 1)^2 = \sum_{r=1}^{k}\left(\mathcal{Q}^2\right)_{rr} - 2\sum_{r=1}^{k}\mathcal{Q}_{rr} + k$$

$$= \sum_{r=1}^{k}\left(\left(\mathcal{Q}^2\right)_{rr} - 1\right) - 2\sum_{r=1}^{k}\left(\mathcal{Q}_{rr} - 1\right) \tag{71}$$

Using (58) and (64) in (70),

$$\mathbb{E}\left|\mathcal{Q}_{rr} - 1\right|^2 = \mathcal{O}(n^{-\alpha}L^2(n) + h^4) \tag{72}$$

The diagonal terms of $\mathcal{Q}^2$ can be evaluated by virtue of (69),

$$\left(\mathcal{Q}^2\right)_{rr} = \sum_{j=1}^{k} \mathcal{Q}_{rj}\mathcal{Q}_{jr} = \left(\mathcal{Q}_{rr}\right)^2 + \sum_{\substack{j=1 \\ j\neq r}}^{k} \mathcal{Q}_{rj}\mathcal{Q}_{jr} = \left(\mathcal{Q}_{rr}\right)^2 + \mathcal{O}_{\text{a.s.}}\left(h^{4(s\wedge r)+2}\right)$$

(73)

Then with (73) and (72) we have,

$$\left(\mathcal{Q}^2\right)_{rr} - 1 = \left(\mathcal{Q}_{rr}\right)^2 - 1 + \mathcal{O}_{\text{a.s.}}\left(h^{4(s\wedge r)+2}\right)$$
$$= \mathcal{O}_p\left(n^{-\frac{\alpha}{2}}L(n) + h^2\right)$$

(74)

Using the bounds in (74) and (72) in (71) with the Lyapunov inequality we find,

$$\mathbb{E}\,\|\mathcal{Q} - I\|_{\text{tr}}^{\frac{3}{2}} \leq \left(\mathbb{E}\,\|\mathcal{Q} - I\|_{\text{tr}}^2\right)^{\frac{3}{4}} = \mathcal{O}\left(n^{-\frac{3\alpha}{8}}L^{\frac{3}{4}}(n) + h^{\frac{3}{2}}\right)$$

(75)

Turn attention now to the Euclidean norm of the martingale differences.

$$|\Xi_i^*(\mathbf{t})|_2 = \left(\sum_{r=1}^{k}\Xi_i^{*2}(t_{j_r})\right)^{\frac{1}{2}} \leq \sqrt{k}\sup_{1\leq r\leq k}|\Xi_i^*(t_{j_r})|$$

$$\leq Cn^{-1}h^{-1}\left(|\mu|_\infty \sup_{1\leq r\leq k}\left|K_3\left(\frac{F(X_i)-t_{j_r}}{h}\right) - \widetilde{K}_3(X_{i,i-1},t_{j_r})\right|\right.$$

$$\left. + |\varepsilon_i|\sup_{1\leq r\leq k}|\zeta_i(t_{j_r})|\right)$$

$$= Cn^{-1}h^{-1}\left(\sup_{1\leq r\leq k}\left|K_3\left(\frac{F(X_i)-t_{j_r}}{h}\right)\right| + |\varepsilon_i|\sup_{1\leq r\leq k}|\zeta_i(t_{j_r})|\right)$$

$$+ \mathcal{O}_{\text{a.s.}}\left(n^{-\frac{1}{2}}h^{(s\wedge r)+\frac{1}{2}}\right).$$

Then consider the third power of the above statement and expand,

$$|\Xi_i^*(\mathbf{t})|_2^3 \leq Cn^{-\frac{3}{2}}h^{-\frac{3}{2}}\left(\left(\sup_{1\leq r\leq k}\left|K_3\left(\frac{F(X_i)-t_{j_r}}{h}\right)\right|\right)^3\right.$$

$$\left. + |\varepsilon_i|^3\left(\sup_{1\leq r\leq k}|\zeta_i(t_{j_r})|\right)^3\right) + \mathcal{O}_{\text{a.s.}}\left(n^{-\frac{3}{2}}h^{3(s\wedge r)+\frac{3}{2}}\right). \quad (76)$$

To give a bound on the expectation of the above term, it is sufficient to look at the higher order expectation terms and apply the Lyapunov inequality. Starting with the expected value of the suprema of the kernel function, using the fact that the kernel function has support on $[-1,1]$ and $\left|t_{j_r} - t_{j_{r'}}\right| \geq 2h$ for $r \neq r'$,

$$\mathbb{E}\left(\sup_{1\leq r\leq k}\left|K_3\left(\frac{F(X_i)-t_{j_r}}{h}\right)\right|\right)^3 = \sum_{r=1}^{k}\int_{\mathbb{R}}\left|K_3\left(\frac{F(x)-t_{j_r}}{h}\right)\right|^3 dF(x)$$
$$= \mathcal{O}(h)$$

(77)

since there is at most one value of t_{j_r} where $K_3(h^{-1}(F(x) - t_{j_r}))$ is non-zero for any given $x \in \mathbb{R}$ and fixed $k \in \mathbb{N}$. The scale function σ is also bounded, so by a similar procedure it can be shown that,

$$\mathbb{E}\left(\sup_{1 \le r \le k} |\zeta_i(t_{j_r})|\right)^3 = \mathcal{O}(h). \tag{78}$$

Using the bounds given in (77) and (78) in (76) and applying the expectation operator yields,

$$\sum_{i=1}^n \mathbb{E}\,|\Xi_i^*(\mathbf{t})|_2^3 = \mathcal{O}\left(n^{-\frac{1}{2}}h^{-\frac{1}{2}} + n^{-\frac{1}{2}}h^{3(s \wedge r) + \frac{1}{2}}\right) = \mathcal{O}\left(n^{-\frac{1}{2}}h^{-\frac{1}{2}}\right). \tag{79}$$

Using (25) with (75) and (79) ensure that

$$\sum_{i=1}^n \mathbb{E}\,|\Xi_i^*(\mathbf{t})|_2^3 + \mathbb{E}\,\|Q - I\|_{\mathrm{tr}}^{\frac{3}{2}} = o(h),$$

which in turn ensures that the required condition

$$\frac{1}{2h} \le C\left(\sum_{i=1}^n \mathbb{E}\,|\Xi_i^*(\mathbf{t})|_2^3 + \mathbb{E}\,\|Q - I\|_{\mathrm{tr}}^{\frac{3}{2}}\right)^{-1}$$

holds. Therefore the conditions are satisfied, so apply Corollary 2 of Grama and Haeusler (2006) and there exists a $q \in (0, 1]$ such that,

$$
\begin{aligned}
P\left(\sup_{t \in T_n} \left|S_n^{B^*}(t)\right| \ge \sqrt{2 \log m_n}\right) &\le P\left(\left|S_n^{B^*}(\mathbf{t})\right|_2 \ge \sqrt{2 \log m_n}\right) \\
&\le \frac{1}{2^{\frac{k}{2} - 1}\Gamma\left(\frac{k}{2}\right)} m_n^{-q} (2q \log m_n)^{\frac{k}{2} - 1} (1 + o(1)) \\
&= \mathcal{O}\left(h^q |\log h|^{\frac{k}{2} - 1}\right) = o(1). \tag{80}
\end{aligned}
$$

The proof then follows as a direct result of (66), (67) and (80). $\qquad\square$

Lemma 3. *Assume that $\mu \in \mathscr{G}_s$, $K \in \mathscr{K}_{s \wedge r}$ with $s \wedge r \ge 3$ and Assumption (B) holds. Then,*

$$\sup_{t \in (h, 1-h)} |D_n(t)| = \mathcal{O}_p\left(h^{s \wedge r + \frac{1}{2}} n^{\frac{1-\alpha}{2}} L(n)\right).$$

Proof of Lemma 3. Using Assumption (B), the higher order derivatives of the densities f_η exist and are bounded. Combining this with (Zhao and Wu, 2008, Lemma 1) it can be shown that for $m = 0, 1, \ldots, s \wedge r$;

$$\left\|\sup_{t \in (h, 1-h)} \left|\frac{d^m}{dt^m} I^{[n]}(t)\right|\right\| = \mathcal{O}\left(n^{1 - \frac{\alpha}{2}} L(n)\right),$$

where $I^{[n]}(t) := \sum_{i=1}^n \{f_X(x|\mathcal{F}_{i-1}) - f_X(x)\}$. The result then follows by using the vanishing moment property of K in conjuction with the Faà di Bruno formula in a similar manner to the procedure that proved (62). $\qquad\square$

Acknowledgements

The first author would like to thank his Ph.D. supervisor N. C. Weber for his guidance and support; Wei Biao Wu for comments which helped in reaching the final result and an anonymous referee that helped improve the presentation of the paper.

References

AVRAM, F. and TAQQU, M. S. (1987). Noncentral limit theorems and Appell polynomials. *Ann. Probab.* **15** 767–775. MR885142 (88i:60058)

BINGHAM, N. H. (1986). Variants on the law of the iterated logarithm. *Bull. London Math. Soc.* **18** 433–467. MR847984 (87k:60087)

BROWN, B. (1971). Martingale Central Limit Theorems. *The Annals of Mathematical Statistics* **42** 59–66.

BROWN, L. D. and LOW, M. G. (1996). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.* **24** 2384–2398. MR1425958 (98a:62042)

CAVALIER, L. (2004). Estimation in a problem of fractional integration. *Inverse Problems* **20** 1445–1454. MR2109128 (2005h:62225)

CHENG, M.-Y. and RAIMONDO, M. (2008). kernel methods for optimal change-points estimation in derivatives. *J. Comput. Graph. Statist.* **17** 56–75. MR2424795

CSÖRGŐ, M. (1983). *Quantile processes with statistical applications. CBMS-NSF Regional Conference Series in Applied Mathematics* **42**. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA. MR745130 (86g:60045)

CSÖRGŐ, S. and MIELNICZUK, J. (1995). Nonparametric regression under long-range dependent normal errors. *Ann. Statist.* **23** 1000–1014. MR1345211 (96k:62091)

CSÖRGŐ, S. and MIELNICZUK, J. (1999). Random-design regression under long-range dependent errors. *Bernoulli* **5** 209–224. MR1681695 (2000i:62045)

DE LA PEÑA, V. H. (1999). A general class of exponential inequalities for martingales and ratios. *Ann. Probab.* **27** 537–564. MR1681153 (2000c:60020)

GIJBELS, I., HALL, P. and KNEIP, A. (1999). On the estimation of jump points in smooth curves. *Ann. Inst. Statist. Math.* **51** 231–251. MR1707773 (2001c:62046)

GOLDENSHLUGER, A., TSYBAKOV, A. and ZEEVI, A. (2006). Optimal change-point estimation from indirect observations. *Ann. Statist.* **34** 350–372. MR2275245 (2008b:62022)

GRAMA, I. G. and HAEUSLER, E. (2006). An asymptotic expansion for probabilities of moderate deviations for multivariate martingales. *J. Theoret. Probab.* **19** 1–44. MR2256478 (2007m:60113)

GUO, H. and KOUL, H. L. (2008). Asymptotic inference in some heteroscedastic regression models with long memory design and errors. *Ann. Statist.* **36** 458–487. MR2387980 (2009e:62350)

HERNÁNDEZ ENCINAS, L., MARTÍN DEL REY, A. and MUÑOZ MASQUÉ, J. (2005). Faà di Bruno's formula, lattices, and partitions. *Discrete Appl. Math.* **148** 246–255. MR2147794 (2006c:11026)

HO, H.-C. and HSING, T. (1996). On the asymptotic expansion of the empirical process of long-memory moving averages. *Ann. Statist.* **24** 992–1024. MR1401834 (97g:60032)

HO, H.-C. and HSING, T. (1997). Limit theorems for functionals of moving averages. *Ann. Probab.* **25** 1636–1669. MR1487431 (98m:60027)

HUH, J. and PARK, B. U. (2004). Detection of a change point with local polynomial fits for the random design case. *Aust. N. Z. J. Stat.* **46** 425–441. MR2088981 (2006b:62061)

JOHNSTONE, I. M. (1999). Wavelet shrinkage for correlated data and inverse problems: adaptivity results. *Statist. Sinica* **9** 51–83. MR1678881 (2000g:62087)

JOHNSTONE, I. M. and SILVERMAN, B. W. (1997). Wavelet threshold estimators for data with correlated noise. *J. Roy. Statist. Soc. Ser. B* **59** 319–351. MR1440585 (98h:62054)

KOROSTELËV, A. P. and TSYBAKOV, A. B. (1993). *Minimax theory of image reconstruction. Lecture Notes in Statistics* **82**. Springer-Verlag, New York. MR1226450 (95a:62028)

KULIK, R. (2008). Nonparametric deconvolution problem for dependent sequences. *Electron. J. Stat.* **2** 722–740. MR2430252

KULIK, R. and RAIMONDO, M. (2009a). L^p−Wavelet regression with correlated errors and Inverse Problems. *Statist. Sinica* **19** 1479-1489.

KULIK, R. and RAIMONDO, M. (2009b). Wavelet regression in random design with heteroscedastic dependent errors. *Ann. Statist.* **37** 3396–3430. MR2549564

LIN, Z., LI, D. and CHEN, J. (2008). Change point estimators by local polynomial fits under a dependence assumption. *J. Multivariate Anal.* **99** 2339–2355. MR2463393 (2010a:62110)

MIELNICZUK, J. and WU, W. B. (2004). On random-design model with dependent errors. *Statist. Sinica* **14** 1105–1126. MR2126343 (2006f:62043)

REISS, M. (2008). Asymptotic equivalence for nonparametric regression with multivariate and random design. *Ann. Statist.* **36** 1957–1982. MR2435461 (2009m:62119)

ROBINSON, P. M. and HIDALGO, F. J. (1997). Time series regression with long-range dependence. *Ann. Statist.* **25** 77–104. MR1429918 (98g:62176)

TAQQU, M. S. (1975). Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **31** 287–302. MR0400329 (53 ##4164)

WANG, Y. (1996). Function estimation via wavelet shrinkage for long-memory data. *Ann. Statist.* **24** 466–484. MR1394972 (98d:62072)

WANG, Y. (1999). Change-points via wavelets for indirect data. *Statist. Sinica* **9** 103–117. MR1678883 (2000a:62075)

WANG, L. (2008). Change-Point Estimation in Long Memory Nonparametric Models with Applications. *Communications in Statistics-Simulation and*

Computation **37** 48–61.

WISHART, J. (2009). Kink estimation with correlated noise. *Journal of the Korean Statistical Society* **38** 131–143.

WU, W. B. (2005). Nonlinear system theory: another look at dependence. *Proc. Natl. Acad. Sci. USA* **102** 14150–14154 (electronic). MR2172215

WU, W. B. (2007). Strong invariance principles for dependent random variables. *Ann. Probab.* **35** 2294–2320. MR2353389 (2008j:60086)

WU, W. B. and MIELNICZUK, J. (2002). kernel density estimation for linear processes. *Ann. Statist.* **30** 1441–1459. MR1936325 (2003j:62064)

YANG, Y. (2001). Nonparametric regression with dependent errors. *Bernoulli* **7** 633–655. MR1849372 (2002g:62078)

ZHAO, Z. and WU, W. (2006). kernel quantile regression for nonlinear stochastic models Technical Report, Department of Statistics, University of Chicago.

ZHAO, Z. and WU, W. B. (2008). Confidence bands in nonparametric time series regression. *Ann. Statist.* **36** 1854–1878. MR2435458