NONPARAMETRIC ESTIMATION OF CHANGE-POINTS IN DERIVATIVES

JUSTIN RORY WISHART

A thesis submitted in fulfillment of
the requirements for the degree of
Doctor of Philosophy

THE UNIVERSITY OF SYDNEY

School of Mathematics and Statistics

Faculty of Science

May 2011

ABSTRACT

In this thesis, the main concern is to analyse change-points in a non-parametric regression model. More specifically, the analysis is focussed on the estimation of the location of jumps in the first derivative of the regression function. These change-points will be referred to as kinks.

The estimation method is closely based on the zero-crossing technique (ZCT) introduced by Goldenshluger, Tsybakov and Zeevi (2006). The work of Goldenshluger et al. (2006) was aimed at estimating jumps in the regression function in the indirect non-parametric regression model and shown to be optimal in the minimax sense. Their analysis was applied in practice by Cheng and Raimondo (2008) whereby a class of kernel functions is constructed to use ZCT with a kernel smoothing implementation. Moreover, Cheng and Raimondo (2008) adapted the technique to estimating kinks from a fixed design model with i.i.d. errors.

The thesis extends the aforementioned kink estimation technique in two ways. The first extension is to include a long-range dependent (LRD) error structure in the fixed design scenario. The rate of convergence of the resultant LRD method is shown to be reliant on the level of dependence and the smoothness of the underlying regression function. This rate of convergence is shown to be optimal in the sense of the minimax rate.

The second extension is to include a regression model with random design and LRD structures. The random design regression models considered include an i.i.d. random design with LRD errors and a separate model with a LRD design with i.i.d. errors.

For the case of LRD design variables, the rate of convergence for the estimator is again reliant on the level of dependence and the smoothness of the regression function. However, interestingly for the case of i.i.d. design and LRD errors, the rate of convergence is shown to not rely on the level of dependence but only rely on the smoothness of the regression function and in fact agrees with the minimax rate for fixed design with i.i.d. errors.

To conclude, it is summarised where original work occurs in this thesis. Firstly, the extension of the ZCT to the fixed design framework with LRD noise arose with discussions with my initial Ph.D. supervisor Dr Marc Raimondo before his passing. The method is based on the technique proposed by Cheng and Raimondo (2008) but the mathematical analysis and development of the extension to the LRD framework and its minimax optimality is my own work. For the second extension which covers the random design regression framework, the main idea and premise arose through discussions with Assistant Professor Rafał Kulik. I wish it to be known that although the published versions of the work are in joint names with Assistant Professor Kulik, the great bulk of the mathematical analysis and development presented in this thesis is my own. Finally my current supervisor's contribution, Professor N. C. Weber, has been to provide direction in terms of checking the accuracy, clarity and style of the work.

PUBLICATIONS

Wishart, J. (2009). Kink estimation with correlated noise. *Journal of the Korean Statistical Society* **38** 131–143.
URL http://dx.doi.org/10.1016/j.jkss.2008.08.001

Wishart, J and Kulik, R (2010). Kink estimation in stochastic regression with dependent errors and predictors. *Electron. J. Statist.* **4** 875–913.
URL http://dx.doi.org/10.1214/10-EJS571

It is very certain that,
when it is not in our power to determine what is true,
we ought to act according to what is most probable.

–René Descartes.

ACKNOWLEDGEMENTS

It is my pleasure to sincerely thank and acknowledge many people who have kindly assisted me both directly and indirectly in the last few years while I wrote this thesis.

Firstly, I would like to thank my honours supervisor Dr Michael Stewart for developing my interest in the field on mathematical research. Also, I would like to acknowledge the contribution of my initial Ph.D. supervisor, the late Dr Marc Raimondo. He provided much assistance, encouragement and support in my first year of my Ph.D. studies, which is in a topic I was not familiar with at the time. Furthermore, he proposed the fundamental underlying idea behind this whole thesis before he passed away at the end of the first year of my studies. These fundamental ideas are evident in Chapter 2 which is the foundation for the entire work in the thesis. For this reason I would like to dedicate the whole of Chapter 2 to his memory.

The largest thanks without a doubt would have to be given to my supervisor, Professor N.C. Weber. Even since my undergraduate studies Neville has provided me with guidance and support, both mathematical and administrative. I would like to thank him for taking me under his wing as a Ph.D. student after Marc passed away when I could have easily given up and pursued other interests. He encouraged to do the mathematics with rigour and motivated me to do more than I thought capable. This thesis would have not been possible without his tireless efforts, organisation and support over the past few years.

I would like to thank Assistant Professor Rafał Kulik, who like Neville, took me under his wing and provided guidance after the passing of Marc. His discussions and support both during my visit to BIRS in Banff, my stay in Ottawa and the email dialogue afterwards gave me another direction of research to pursue and motivation to continue.

I would like to thank many of my fellow students in the school who I have built friendships with over the years and made the whole Ph.D. experience enjoyable, to name a few: Patrick, Garth, Duncan, Erwin, Jo, James Kelly, Emi, Anna and Jenny. I would also like to thank staff members Dr Samuel Müller and Dr John Ormerod who provided valuable comments and insight while I presented my work.

I would also like to thank the University of Sydney and the School of Mathematics & Statistics for their financial support over the years which allowed me to more readily focus on my studies.

Finally I would like to thank the members of my family: my parents and my brother Nathan. I would like to thank them for supporting my decisions and giving their wisdom and guidance in all facets of my life over the years.

CONTENTS

LIST OF FIGURES

ACRONYMS

fARIMA fractional Autoregressive Integrated Moving Average

ARIMA Autoregressive Integrated Moving Average

fBm fractional Brownian motion

i.i.d. independent and identically distributed

LRD Long-Range Dependent

RMSE Root Mean-Square Error

SRD Short-Range Dependent

ZCT Zero-Crossing Technique

NOMENCLATURE

θ	Location of a change-point	4
$\asymp$	Equal up to a constant	13
$\xrightarrow{\mathscr{D}}$	Converges in distributon	25
$x \vee y$	Maximum of x and y	24
$x \wedge y$	Minimum of x and y	75
$\lfloor x \rfloor$	Largest integer before x	25
$\lceil x \rceil$	Smallest integer following x	100
x_+	Positive part of x	72
$[\mu]$	Change-point function	4
$f \circ g$	Composite of two functions, f and g	96
$\mathbb{1}$	Indicator function	12
$K*\mu$	Convolution of the functions K and μ	9
$\widehat{f}_n$	An estimator of f, given n observations	12
$\widetilde{f}$	The Fourier tranform of the function f	17
P_n	Legendre polynomial of degree n	30
$\mathbb{R}$	Field of real numbers	2
$\mathbb{R}^+$	Field of non-negative real numbers	7
$\mathbb{Z}$	Field of integers	8
$\mathbb{Z}^+$	Field of non-negative integers	35
$\mathscr{B}_p^{(s,q)}$	Besov space of functions	15
$\mathscr{C}_s^m$	Class of functions with kinks	27
$\mathscr{H}_\alpha^s$	Hölder space of functions	14
$\mathcal{H}$	Class of fractionally Wiener integrable functions	74
$\mathscr{K}_s$	Class of high order kernel functions	35
$\mathscr{L}_p$	Space of p-integrable functions	14
$\mathscr{M}_s$	Kink functional class for minimax result	78
$\mathscr{T}_s$	Kink functional class in the Fourier domain	77
$\mathscr{T}_s^m$	Multiple kink functional class in the Fourier domain	78
$\mathscr{W}_p^s$	Sobolev space of functions	15
$\mathcal{F}$	Fourier transform operator	17
$\mathcal{D}^n$	Differential operator of order n	31
$\mathcal{D}^{-\nu}$	Fractional integral operator of order ν	70
$\mathcal{D}_*^{-\nu}$	Adjoint fractional integral operator of order ν	70
$\langle \cdot, \cdot \rangle$	Inner product on the $\mathscr{L}_2$ space of functions	8
$\langle \cdot, \cdot \rangle_{\mathcal{H}}$	Inner product for the functional class $\mathcal{H}$	43
$\lvert \cdot \rvert_p$	Norm of the $\mathscr{L}_p$ space.	14
$\lVert X \rVert_p^p$	The p^{th} moment of a random variable, X.	16
$\frac{d\mathbb{P}_1}{d\mathbb{P}_0}$	Radon-Nikodym derivative of two probability measures	76
$\mathcal{K}$	Kullback-Leibler divergence of two probability measures	79
$\mathcal{R}^*$	Minimax risk	13
$\mathcal{S}_n$	Faá di Bruno summation set of order n	96
H	Self-similarity or Hurst parameter	23

B	Standard Brownian motion	24
B_H	Fractional Brownian motion	24
F_X	Cumulative distribution function for the random variable X	87
F_n	Empirical cumulative distribution function	87
Q	Quantile function	87
Q_n	Empirical quantile function	108

Part I

INTRODUCTION

1

NONPARAMETRIC REGRESSION

A classical topic in statistics is regression analysis. It examines the relationship between two variables of interest where the influence of an independent variable X is measured against a dependent variable Y. Some simple examples that could fall under this framework include the relation of the level of rainfall (X) on tree growth (Y); the relation of a child's age (X) on their height (Y).

Clearly there is rarely a perfect causal relationship, (children of the same age can have different heights). Therefore the conditional mean of the variable Y on X is usually pursued in estimation and the deviations from this conditional mean are modelled by an error term. The regression model for this general framework would be written,

$$Y = \mathbb{E}\left[Y|\, X\right] + \mathcal{E},$$

where $\mathcal{E}$ is the error term that models the deviations of Y from its conditional mean on X and it is assumed that $\mathbb{E}\mathcal{E} = 0$. A very basic approach would be to assume that the conditional mean has a linear structure with $\mathbb{E}\left[Y|\, X\right] = a + bX$ for some constants $a, b \in \mathbb{R}$ which will be unknown in practice. A model can be constructed by estimating a and b with $\widehat{a}$ and $\widehat{b}$ by various methods, the most straightforward approach being least squares estimation methods.

The benefits of using an underlying estimated regression model are two-fold. Firstly, some insight into the underlying behaviour and relationship between X and Y can be gained by using the estimated model as a model based data analysis tool. This would be revealed in this case by interpreting the structure of the straight line generated by $\widehat{a} + \widehat{b}X$. Secondly, the estimated regression model could potentially be used for prediction purposes as to the behaviour of Y for unobserved values of X. The emphasis for this thesis will be on the former, using a regression model as data analysis tool.

A linear structure is very limited in its scope, a natural extension would be to assume that the conditional mean has the structure of a polynomial. That is, consider the class of all polynomials of degree d in $\mathbb{R}$ and denote them by

$$\pi_d = \left\{ p : \mathbb{R} \longrightarrow \mathbb{R} : p(x) = a_0 + a_1 x + \ldots + a_d x^d, a_i \in \mathbb{R}, i = 0, 1, \ldots, d \right\}.$$

Then assume that $\mathbb{E}\left[Y|\, X\right] = g(X) \in \pi_d$ for some $d \in \mathbb{N}$. This would be a much richer framework and the estimated coefficients inherent in π_d could be estimated in various ways.

Other structures could be possible such as assuming $\mathbb{E}\left[Y|\, X\right]$ has a logarithmic or exponential behaviour. All the regression models introduced so far are specific instances of a broader class of models called *parametric* models. These parametric models have a well established theory and estimators with 'good' performance have been constructed that exploit the parametric nature of the underlying model. The exact meaning of 'good' performance will be clarified later in Section 1.4 but the 'good' performance is gained by tailoring the estimator to utilise the restrictions imposed by the parametric framework. However, there is an inherent weakness to this approach and that is the parametric class of models impose a very rigid structure on the data and there is a danger of mis-specifying the model. This can have severe implications in both the interpretation of the regression model as a data analysis tool and in prediction of other values.

An alternative to this parametric approach is the *nonparametric* approach. The nonparametric approach is to relax the rigid assumptions imposed by the parametric framework and only impose mild assumptions. A basic general nonparametric approach is to assume that the mean of the dependant variable of interest Y depends on $X \in \mathcal{X} \subseteq \mathbb{R}$ through a so called regression function $\mu : \mathcal{X} \longrightarrow \mathbb{R}$. Moreover, the regression function μ is assumed to be derived from a broad (usually very large) functional class $\mathscr{F}$ that has mild restrictions. A simple class is when $\mathscr{F} = \mathrm{Lip}(\mathcal{X}, \mathrm{L})$, the class of Lipschitz continuous functions, that is, $\mu \in \mathrm{Lip}(\mathcal{X}, \mathrm{L})$ when $\mu : \mathcal{X} \longrightarrow \mathbb{R}$ such that $|\mu(t) - \mu(s)| < L\,|t - s|$ for all $t, s \in \mathcal{X}$ and some constant $L > 0$. In the nonparametric framework the regression model would be of the form,

$$Y = \mu(X) + \mathcal{E}. \tag{1.1}$$

The nonparametric approach has the benefit of not forcing a strict structure on the data and in a sense allowing the data to 'speak for itself' which reduces or avoids the problem of mis-specification. However, the nonparametric approach does have one disadvantage, namely when a parametric model has been correctly specified the nonparametric estimators in general have poor efficiency in estimation in comparison to their parametric equivalents. Therefore, one can see that the two approaches are opposing trade-offs between performance and mis-specification. Throughout this thesis the focus will be solely on the nonparametric approach.

In the context of the nonparametric approach some notation and nomenclature is required. The independent variables (X) are called the *design*. The design can be a set of random variables $X_1, X_2, \ldots, X_n$ or an equally spaced grid of points and $X_i = x_i = \frac{i}{n}$, for $i = 1, 2, \ldots, n$. The former is referred to as *random design* and the latter *fixed design*. In keeping with the usual notation in the literature, upper case will denote random variables and lower case deterministic variables.

The analysis of nonparametric regression function estimators has been a popular field of study and some important contributions to the literature in the fixed design scenario include Speckman (1985); Khas'minskiĭ (1992); Wand and Jones (1995); Härdle, Kerkyacharian, Picard and Tsybakov (1998); Efromovich (1999); Fan and Yao (2003);

Wasserman (2006); Tsybakov (2009). For the random design case, some of the most recent work includes Baraud (2002); Zhang, Wong and Zheng (2002); Bertin (2004); Birgé (2004); Kerkyacharian and Picard (2004); Chesneau (2007); Kohler (2008); Kulik and Raimondo (2009).

1.1 CHANGE-POINT ANALYSIS

A sub-area of regression analysis that is very active in research is change-point analysis. Change-point analysis is concerned with identifying sharp transitions or structural changes in the underlying regression function. These are characterised by a jump in the regression function at a particular point. More specifically, there is a point $\theta \in \mathbb{R}$ such that,

$$[\mu]\,(\theta) := \mu(\theta^{+}) - \mu(\theta^{-}) = a.$$

where $\mu(\theta^{-}) = \lim_{x\uparrow\theta} \mu(x)$ and $\mu(\theta^{+}) = \lim_{x\downarrow\theta} \mu(x)$ and $a \in \mathbb{R}$ is non-zero.

Knowledge of these locations could explain the change in qualitative or quantitative behaviour of an underlying process. For example, in a situation of quality control, μ might represent a measure of quality in a manufacturing process and a change-point would represent a sudden increase or decrease in quality. It would be of considerable value to the manufacturer to not only know that manufacturing quality has changed but also when the change occurred. Another example arises in image analysis where it might be desirable to separate two objects on a single image. The detection and location for sharp changes in colour would be of interest for that application. Some notable references on the material are Korostelëv (1987); Korostelëv and Tsybakov (1993); Raimondo (1998); Gijbels, Hall and Kneip (1999); Goldenshluger, Tsybakov and Zeevi (2006); Goldenshluger, Juditsky, Tsybakov and Zeevi (2008b,a).

Another area that has received less attention is the detection and estimation of change-points in the slope (and higher order derivatives) of the regression function μ. We describe this jump in the first derivative of μ as a kink and will 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 nonparametric model. This can be related back to the previous two examples of quality control and image analysis. Starting with quality control, recall that μ represents the quality of a manufacturing process, if μ changes from a constant to a monotone decreasing function, a kink would represent when the manufacturing quality started to decline (possibly wear and tear on a part of the process becoming more apparent). In the context of image analysis, consider the scenario where the image is the electrocardiogram (ECG) for a person, a kink would represent the rapid acceleration and deceleration that is intrinsic in a heartbeat. Some notable references on the material are Müller (1992); Wang (1995); Gijbels and Goderniaux (2004); Cheng and Raimondo (2008).

A function $\mu : \mathcal{X} \longrightarrow \mathbb{R}$ is said to have a kink if there is a jump in the first derivative $\mu^{(1)}$. More formally a function μ has a kink at θ with intensity $|a| > 0$ if

$$\left[\mu^{(1)}\right](\theta) := \mu^{(1)}(\theta^{+}) - \mu^{(1)}(\theta^{-}) = a.$$

For the purposes of regression analysis, the knowledge of kink locations can reveal a change in the trend of the model or reveal periods where the function experiences sharp changes in curvature. This was demonstrated in Cheng and Raimondo (2008) when they analysed the motorcycle dataset which is a common dataset that is used in the nonparametric literature (see (Härdle, 1990, Table 1, Appendix 2) for the dataset). The dataset itself records the experienced accelerations of the head on a test dummy during a simulated motorcycle crash. The initial analysis for this context has been performed by Cheng and Raimondo (2008) where they identified three kinks at $\theta_1 = 0.21$, $\theta_2 = 0.33$ and $\theta_3 = 0.54$ which corresponds to the moments that the motorcycle crashed and the head experienced whiplash due to the rapid deceleration. An underlying function of the accelerations has been estimated using a segmented regression with the kink locations represented by the solid line in Figure 1. In other cases, kink analysis

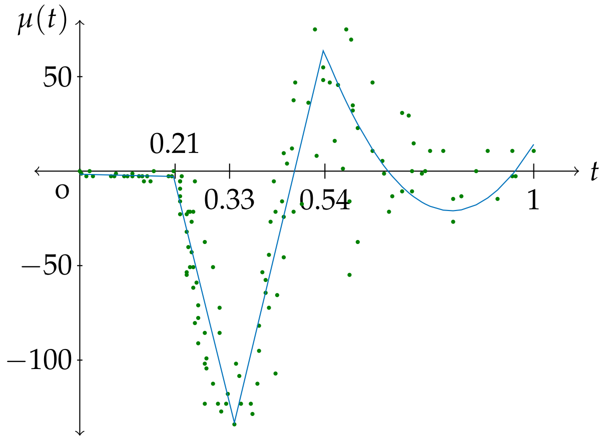


Figure 1: Scatterplot of the motorcycle dataset with an estimate of the true underlying function with three kinks.

can also locate qualitative changes in behaviour. Another common dataset used in the literature is the nursing time of a beluga whale calf dataset in Cheng and Raimondo (2008). After analysis a kink was located corresponding to the time of a bacterial infection. The reader is referred to Cheng and Raimondo (2008) for further analysis and discussion of both the motorcycle and whale datasets.

For the change-point estimation problem, parametric or nonparametric methods can be pursued. Similar to the regression estimation problem discussed earlier, the same dichotomous trade-off in the approaches arise with respect to misspecification and estimation efficiency. Consider firstly the parametric approach and then the nonparametric approach.

In the parametric approach, a crucial problem is to determine the number of kinks present in the dataset. Once this is known, the location of each kink can be estimated to best fit the data using the parametric family of distributions that the data are assumed to follow.

A naïve approach is to estimate and locate a kink by inspection of a scatterplot as was used in Kunst, Looman and Mackenbach (1993) but this method fails when the kink locations are not easily discernible by the naked eye. Another body of literature uses search algorithms such as a grid search algorithms (see Lerman (1980)). There is also a body of literature on Bayesian methods for parametric change point analysis and the interested reader is referred to Chen, Chan, Gerlach and Hsieh (2010) who compare some Bayesian methods against the grid search algorithms. Another parametric method uses a maximum likelihood approach. One such method has been suggested by Hawkins (2001) that assumes each data segment comes from a possibly different exponential family of distributions. Another similar recent treatment of the topic has been covered in the context of a simple linear regression model with a change point by Liu and Qian (2010) who use a empirical likelihood ratio statistic.

In the nonparametric approach of estimating kink locations there is no such restriction on the data segments. The method does not require the number of kinks to be estimated or to be specified beforehand. Work has been done for nonparametrically estimating kink locations in certain situations by Müller (1992); Korostelëv and Tsybakov (1993); Raimondo (1998); Gijbels, Hall and Kneip (1999); Luan and Xie (2001); Goldenshluger et al. (2006); Cheng and Raimondo (2008); Menéndez, Ghosh and Beran (2010). Some contributions to the area based on the work in this thesis have appeared in Wishart (2009); Wishart and Kulik (2010). The nonparametric approach is the method we will pursue for analysis since it is more lenient to the data and does not enforce any restrictive assumptions.

Nonparametric estimation of sharp cusps has been investigated by Wang (1995); Raimondo (1998); Wang (1999). A sharp cusp can have the interpretation of a kink with unbounded intensity. More rigorously, for $c \in (0,1)$, a function μ has an c-level cusp at θ if for $h > 0$ tending to zero there exists a constant $C > 0$ such that,

$$|\mu(\theta + h) - \mu(\theta - h)| \geq 2C\,|h|^c\,.$$

It is worth noting that a c-level cusp can be thought of as an intermediate case between the previous two change-points cases. Indeed, notice that when when $c = 0$, a cusp is the same as a jump in μ and when $c = 1$, a cusp is the same as a kink in μ.

1.2 NONPARAMETRIC ESTIMATION METHODS

In what follows various techniques will be explored that have been used previously in the literature to solve nonparametric regression and change-point problems.

1.2.1 *Kernel smoothing*

There exist various different types of kernel estimators of the regression function μ. The most applauded is the Nadaraya-Watson estimator constructed independently by Nadaraya (1964) and Watson (1964). Given a kernel function $K : \mathbb{R} \longrightarrow \mathbb{R}$ and a bivariate set of observations (X, Y) the random design version of their estimator is of the form,

$$\widehat{\mu}(x) = \frac{\sum_{i=1}^{n} Y_i K\left(\frac{X_i - x}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right)}, \tag{1.2}$$

where h is the bandwidth of the estimator. The kernel smoothing approach is essentially a weighted averaging approach across the values of Y. It is usually assumed that the weights are all positive and normalised to have a total value of 1, that is $K : \mathbb{R} \longrightarrow \mathbb{R}^+$ with $\int_{\mathbb{R}} K(x)\, dx = 1$ although this is not always necessary. For each fixed x, the weights for each Y_i in the estimate (1.2) are controlled by the kernel function K and the bandwidth h and the distance the corresponding X_i value is from x. A fixed design variant is the estimator given by,

$$\widehat{\mu}(x) = \frac{1}{nh} \sum_{i=1}^{n} Y_i K\left(\frac{x_i - x}{h}\right).$$

In terms of performance, the choice of K has little impact on the final estimate produced. The main contributor to the weighting scheme is the bandwidth h. A large bandwidth will produce a smoother estimate with low variability at the cost of high bias. In contrast, an estimate produced using a smaller bandwidth will have a lower bias at the cost of higher variance. The kernel smoothing approach methods usually revolve around selecting the bandwidth in an 'optimal' way to balance the bias and variance effects to produce a reliable estimate.

There are some other variants of the kernel smoothing approach in the form of the *Gasser-Müller* estimator introduced by Gasser and Müller (1979) and the *Priestley-Chao* estimator introduced by Priestley and Chao (1972). These are very similar estimators but have a slightly different weighting scheme to the observations in the averaging approach. More specifically, the weights are a convolution of the kernel function and the step function with design point boundaries. There are some slight advantages and disadvantages between using the convolution weighted estimators (Gasser-Müller and Priestley-Chao) or the evaluation weighted estimator of Nadaraya-Watson, however, they are not directly relevant to this thesis and the interested reader is referred to Chu and Marron (1991) for a detailed discussion on the topic.

Some classical references in this field are Wand and Jones (1995) and Fan and Gijbels (1996) which cover both density estimation and nonparametric regression. Some notable papers for our purposes are Wu and Mielniczuk (2002); Zhao and Wu (2008); Liu and Wu (2010). The thesis uses a kernel smoothing approach and estimators are constructed in Part II for the fixed design case and in Part IV and Part V for the random design case.

1.2.2 *Orthogonal projection estimators*

A different approach is to express the regression function μ in terms of an orthonormal basis on the space $\mathscr{F}$. Moreover, this can be defined in the framework of Hilbert spaces using the inner product

$$\langle \mu, \varphi_\lambda \rangle := \int_{\mathbb{R}} \mu(x)\overline{\varphi}_\lambda(x)\, dx,$$

where $\overline{\varphi}$ denotes the complex conjugate of φ. Indeed, if a function $\mu \in \mathscr{F}$, then a set of orthonormal basis functions $\{\varphi_\lambda\}_{\lambda\in\Lambda}$ of $\mathscr{F}$ is determined. Then the function μ can be expressed in terms of this basis:

$$\mu(x) = \sum_{\lambda\in\Lambda} \langle \mu, \varphi_\lambda \rangle \varphi_\lambda(x). \tag{1.3}$$

A simple orthonormal basis is the Fourier basis where $\Lambda = \mathbb{N}$ and $\varphi_\lambda(x) = e_{2\pi i\lambda}(x) := e^{2\pi i\lambda x}$. Then (1.3) reduces down to the celebrated Fourier series decomposition of μ.

A main pitfall of the orthogonal projection method occurs when an infinite number of projections are required to reconstruct μ (that is, $\{\#\lambda : \langle \mu, \varphi_\lambda \rangle \neq 0\} = \infty$). This can cause an issue since, in practice, the summation set in (1.3) needs to be truncated at some finite level.

A projection method that is common in change point analysis that is not pursued in the thesis is the wavelet expansion. Assuming $\mu \in \mathscr{F}$ then the homogeneous wavelet expansion is given by,

$$\mu(x) = \sum_{j\in\mathbb{Z}} \sum_{k\in\mathbb{Z}} \beta_{j,k}\psi_{j,k}(x)$$

where

$$\psi_{j,k}(x) = 2^{\frac{j}{2}}\psi(2^j x - k)$$

is the 'mother' wavelet function at scale j and location k and $\beta_{j,k} = \langle \mu, \psi_{j,k} \rangle$ are the wavelet (or 'detail') coefficients. This can be further expressed in terms of the scale or 'father' wavelet function in the inhomogeneous expansion at scale level j_0,

$$\mu(x) = \sum_{k\in\mathbb{Z}} \alpha_{j_0,k}\phi_{j_0,k}(x) + \sum_{j=j_0}^{\infty} \sum_{k\in\mathbb{Z}} \beta_{j,k}\psi_{j,k}(x)$$

where ϕ is the scale or 'father' wavelet function with $\phi_{j,k}(x) = 2^{\frac{j}{2}}\phi(2^j x - k)$ and $\alpha_{j,k} = \langle \mu, \phi_{j,k} \rangle$ are the scale coefficients. The inhomogeneous expansion condenses the homogeneous expansion into a simpler format. It does this by starting with the orthogonal projection of μ from the father wavelet $\phi_{j_0,k}$ which captures the general shape of μ at scale level j_0. The remainder of the *details* for the function μ are explained by the mother wavelet projections. The interested reader for wavelet methodology is referred to Mallat (1999) for a thorough theoretical treatment and Härdle et al. (1998) for treatment in the context of statistical analysis.

The wavelet expansion is a popular method in change-point analysis since a jump is a localised phenomena. As such, the jump can be captured by examining the higher level 'detail' coefficients. Some notable contributions of the wavelet expansion for change point analysis are Wang (1995); Raimondo (1998); Wang (1999); Antoniadis and Gijbels (2002); Cavalier (2004); Wang (2008).

1.2.3 *Other methods*

There are alternative methods that have been used in the nonparametric framework that include (but are not limited to) splines, neural networks and maximum likelihood estimators. The reader is referred to Efromovich (1999); Eubank (1999); Wasserman (2006) and references therein for treatment on other methods.

1.3 MODEL ASSUMPTIONS

Up until this point, the assumptions on the underlying nonparametric model given in (1.1) have received very little attention. The only one being that in the nonparametric framework $\mathbb{E}\left[Y|\, X\right] = \mu(X)$ for some regression function $\mu \in \mathscr{F}$. However, this is not always the case. In some applications it is assumed that $\mathbb{E}\left[Y|\, X\right] = K \circ \mu(X)$ where K is a functional operator on the function μ. A relevant example of this being appropriate would be in image analysis where an image is recorded using a blurry lens. The functional operator K could be a convolution operator where its corresponding kernel function K represents the effect of the blurring process on the lens and the underlying image would be the regression function μ. This generalised regression model is given by,

$$Y = \mathrm{K}{*}\mu(X) + \sigma(X)\mathcal{E}, \tag{1.4}$$

where $\mathrm{K}{*}\mu$ denotes the convolution of $\mathrm{K}: \mathcal{X} \longrightarrow \mathbb{R}$ and μ and $\sigma : \mathcal{X} \longrightarrow \mathbb{R}^+$ is a scale function that allows possible heteroskedastic variance in the errors. The model (1.4) is not pursued here in its full generality. Special cases of model (1.4) are pursued and will be introduced as the need arises.

The structure of both the design and error variables has not yet been definitively stated. There is already a wide body of literature in the context of nonparametric regression for independent and identically distributed (i.i.d.) random variables. This thesis will attempt to extend the analysis to the more general assumption of dependence in either the design or error variables. In particular, the assumption will be on second-order stationary Long-Range Dependent (LRD) processes, starting with a fixed design model with LRD errors in Part II. Then the analysis is continued to the random design case with i.i.d. design variables and LRD errors variables in Part IV. The analysis concludes in Part V where a random design model is considered with LRD design variables and i.i.d. error variables. This leads to the specific definitions of what is meant by a second-order stationary LRD process. Starting with the concept of second-order stationary.

Definition 1.1. *A random process $\{X_t\}_{t\in\mathbb{R}}$ is said to be second-order stationary if, for all $t \in \mathbb{R}$ and $k \in \mathbb{R}$,*

$$\mathbb{E}X_t = \mu \qquad \text{and} \qquad \mathrm{Cov}\left(X_t, X_{t+k}\right) = \gamma(|k|)$$

where $\gamma : \mathbb{R} \longrightarrow \mathbb{R}$ is the autocovariance function.

This is known as second-order stationarity, weakly stationary or covariance stationary. In the literature, there are other definitions of stationarity that have stronger assumptions (referred to as strong stationarity or strictly stationary). For purposes here, a process will be called stationary if it is covariance stationary, that is, it satisfies Definition 1.1. This leads to the definitions of stationary Short-Range Dependent (SRD) and LRD processes.

Definition 1.2. *A stationary random process $\{X_t\}_{t\in\mathbb{R}}$ is said to be SRD if,*

$$\sum_{k=0}^{\infty} \gamma(k) < \infty$$

where γ is the autocovariance function.

Definition 1.3. *A stationary random process $\{X_t\}_{t\in\mathbb{R}}$ is said to be LRD if, there exists a real number $\alpha \in (0,1)$ such that,*

$$\gamma(k) \sim L^2(k)k^{-\alpha}.$$

where $L : \mathbb{R}^+ \longrightarrow \mathbb{R}^+$ is a slowly varying function. That is for all $k > 0$,

$$\frac{L(tk)}{L(t)} \to 1 \quad \text{as } t \to \infty.$$

Long-range dependence is also known as strong dependence or long memory in other parts of the literature.

There is a very large body of literature on LRD random variables and some consider the topic controversial due to possible ambiguity as to whether a process is non-stationary or has long range dependence. This is beyond the scope of this thesis but it is worth stating the merits of the analysis based on long range dependent variables. Two important seminal works in the area include Hurst (1951) and Mandelbrot and Van Ness (1968). It was Hurst (1951) who noticed that the annual minimum water levels of the Nile river throughout recorded history exhibited similar behaviour or persistence over time. Then the work of Mandelbrot and Van Ness (1968) took this idea and formalised it mathematically. Since then, LRD has been applied in many areas to describe phenomena. Some notable applications of LRD analysis include economics with possible LRD in financial returns, volatility and stock trading volumes; hydrology in rainfall and temperature data; and computer science with data network traffic data. There are many more applications of LRD analysis and the interested reader is referred to Beran (1992, 1994); Doukhan, Oppenheim and Taqqu (2003) for more details.

There are a number of popular processes that have been introduced that exhibit this property. Four common processes that satisfy Definition 1.3 that are used in practice are the parametric fractional Autoregressive Integrated Moving Average (fARIMA) models, fractionally Integrated Generalised Autoregressive Conditionally Heteroskedastic (fIGARCH) models, the fractional Brownian motion (fBm) model and causal linear processes. The two processes that will be used throughout the thesis to model LRD will be fBm and causal linear processes. Their specific definitions will be delayed until later with fBm in Part II and causal linear processes in Part IV.

The fARIMA models were proposed by Granger and Joyeux (1980); Hosking (1981) to generalise the already well established theory of Autoregressive Integrated Moving Average (ARIMA) models by Box and Jenkins (1970) to satisfy LRD. The fractionally Integrated Generalised Autoregressive Conditionally Heteroskedastic (fIGARCH) model was proposed by Baillie, Bollerslev and Mikkelsen (1996); Andersen and Bollerslev (1997) to generalise the well established theory of GARCH models by Engle (1982) to satisfy LRD. The fIGARCH model has more applications to financial data due to its ability to satisfy heteroskedastic conditional variance and LRD. Both models are flexible, however, these two processes require many parameters for all of the components to be fully specified.

The fBm was first formally introduced explicitly by Mandelbrot and Van Ness (1968) as an extension of a standard Brownian motion. The increments of fBm are called fractional Gaussian noise and they are used in practice to model errors that exhibit LRD behaviour. The other process of interest for this thesis is the causal linear process. Roughly speaking, the causal linear process is an infinite linear combination of latent random variables. A detailed account of applications and references of all the aforementioned LRD processes can be found in Beran (1994).

1.4 MINIMAX RISK

To measure the performance of an estimator the notion of risk is introduced. In particular, the minimax risk measure is used throughout the thesis to compare the performance of constructed estimators.

In this chapter some of the fundamental ideas behind the minimax theory are explored for two reasons. Firstly, a short review of methods and their optimal performance rates are given for comparison with our later analysis. Secondly, the theory and methods will be used to check that the technique introduced in Part II is optimal in the minimax sense.

Loosely speaking, the key idea behind minimax estimation is to construct an estimator that minimises the risk of the estimator across the worst possible scenario in the functional class $\mathscr{F}$. Let $f \in \mathscr{F}$ where f is a desired quantity of interest such as a regression function or a change point and $\mathscr{F}$ is its corresponding functional class. Let $\widehat{f}_n$ be an estimator of f given a sample of n observations.

Risk is to be defined in terms of a pseudo-metric and loss function that measures the distance of the estimator to the true value. The framework and notation is borrowed from Tsybakov (2009). Let us start by defining a pseudo-metric.

Definition 1.4. *Let $f, f_0, f_1 \in \mathscr{F}$, then $d : \mathscr{F} \times \mathscr{F} \longrightarrow \mathbb{R}^+$ is a pseudo-metric if it satisfies,*

1. $d(f_0, f_1) \geq 0$
2. $d(f_0, f_0) = 0$
3. $d(f_0, f_1) \leq d(f_0, f) + d(f, f_1)$.

A loss function $w : \mathbb{R}^+ \longrightarrow \mathbb{R}^+$ is a monotone increasing function with $w(0) = 0$ and is used in combination with a pseudo-metric to give a standard measure of the distance between f and $\widehat{f}$. Some common loss functions include,

$$w(u) = \mathbb{1}_{\{|u| \geq A\}} \quad \text{for some constant } A > 0, \qquad \text{(Indicator Loss)}$$
$$w(u) = |u|, \qquad \text{(Absolute Loss)}$$
$$w(u) = u^2. \qquad \text{(Square Loss)}$$

Let $\widehat{f}_n := \widehat{f}((X_1, Y_1), \ldots, (X_n, Y_n))$ be an estimator of f given a sample of n observed bivariate observations. Given the definitions of loss and a pseudo-metric the risk of the estimator $\widehat{f}_n$ of f is defined,

$$R\left(\widehat{f}_n, f, \rho_n\right) := R\left(\widehat{f}_n, f, \rho_n, d, w\right) = \mathbb{E}_f\left\{w\left(\rho_n^{-1} d\left(\widehat{f}_n, f\right)\right)\right\}$$

where ρ_n is a positive sequence that is referred to as the rate of convergence. Define the maximal risk by maximising the risk over the functional class $\mathscr{F}$,

$$R^*_{\mathscr{F}}\left(\widehat{f}_n, \rho_n\right) := R^*\left(\widehat{f}_n, \mathscr{F}, \rho_n, d, w,\right) = \sup_{f\in\mathscr{F}} \mathbb{E}_f\left\{w\left(\rho_n^{-1} d\left(\widehat{f}_n, f\right)\right)\right\}.$$

The minimax risk is the smallest possible value of the maximal risk and defined to be,

$$\begin{aligned}\mathcal{R}^*_{\mathscr{F}}\left(\rho_n\right) &:= \inf_{\widehat{f}_n} R^*_{\mathscr{F}}\left(\widehat{f}_n, \rho_n\right)\\ &= \inf_{\widehat{f}_n}\sup_{f\in\mathscr{F}} \mathbb{E}_f\left\{w\left(\rho_n^{-1} d\left(\widehat{f}_n, f\right)\right)\right\}.\end{aligned}$$

When the indicator loss function is used, the risk function becomes,

$$R\left(\widehat{f}_n, f, \rho_n\right) = P_f\left(d\left(\widehat{f}_n, f\right) \geq A\rho_n\right), \qquad \text{(Probabilistic loss)}$$

where A is some fixed level. Then the maximal risk $R^*_{\mathscr{F}}\left(\widehat{f}_n, \rho_n\right)$ becomes,

$$R^*_{\mathscr{F}}\left(\widehat{f}_n, \rho_n\right) = \sup_{f\in\mathscr{F}} P_f\left(\rho_n^{-1} d\left(\widehat{f}_n, f\right) \geq A\right).$$

This reduces the maximal risk to a probability statement concerning the pseudo-metric distance of the estimator and the rate of convergence ρ_n and has important consequences that will be discussed in Section 3.2.1. The most common risk function used in practice is the squared-error loss which uses,

$$R\left(\widehat{f}_n, f, \rho_n\right) = \rho_n^{-2}\mathbb{E}_f\left|\widehat{f}_n - f\right|^2. \qquad \text{(Squared-error loss)}$$

In practice it is usually very difficult to determine the precise minimax risk of an estimator over a functional class. It is easier to determine the asymptotic behaviour of the minimax risk in terms of the sample size, n. In particular, the weaker statement will be pursued that an estimator $\widehat{f}_n$ will be said to be optimal in the sense of the minimax rate if $R^*_{\mathscr{F}}\left(\widehat{f}_n, \rho_n\right) \asymp \mathcal{R}^*_{\mathscr{F}}\left(\rho_n\right)$ where $a_n \asymp b_n$ if there exist constants $0 < c < C < \infty$ such that $c < a_n/b_n < C$. The rate of convergence ρ_n that satisfies this property is said to be the *minimax rate* of convergence and the corresponding estimator $\widehat{f}_n$ is said to be *rate optimal*.

The minimax rate can be determined in terms of the *upper rate of convergence* and *lower rate of convergence* of the minimax risk.

Definition 1.5. *We say that a sequence ρ_n is a lower rate of convergence for the functional class $\mathscr{F}$ in the pseudo-metric d if,*

$$\liminf_{n\to\infty}\inf_{\widehat{f}_n} R^*_{\mathscr{F}}\left(\widehat{f}_n, \rho_n\right) \geq c_*$$

for some positive constant $c_* < \infty$.

Definition 1.6. *We say that a sequence* ρ_n *is an upper rate of convergence for the functional class* $\mathscr{F}$ *in the pseudo-metric* d *if,*

$$\limsup_{n\to\infty} \inf_{\widehat{f}_n} R^*_{\mathscr{F}}\left(\widehat{f}_n, \rho_n\right) \leq C^*$$

for some positive constant $C^* < \infty$.

Clearly, if $R^*_{\mathscr{F}}(\widehat{f}_n, \rho_n)$ satisfies both Definition 1.5 and Definition 1.6 then ρ_n is the minimax rate for estimators $\widehat{f}_n$ of f over the functional class $\mathscr{F}$.

Until now it has been assumed that the minimax framework has an underlying functional class $\mathscr{F}$ that contains f and there is good reason for this. In the early stages of investigation in the literature it became evident that if f was assumed to be an arbitrary function and not restricted to a functional space $\mathscr{F}$, then for any estimator $\widehat{f}_n$, there is a function f such that the non-rate normalised maximal risk $R^*(\widehat{f}_n, 1) = \sup_f \mathbb{E}_f\{w(d(\widehat{f}_n, f))\}$ does not converge to zero. Devroye and Györfi (1990) is a relatively recent treatment of this fact in the context of density estimation.

1.4.1 *Common functional classes*

For comparison purposes for the estimators constructed in Part II, Part IV and Part V, the already established minimax rates will be reviewed for a number of different models along with their respective assumptions on the functional class $\mathscr{F}$. Before proceeding some of the common functional classes will be defined.

Definition 1.7 ($\mathscr{L}_p$-space)**.** *Let* $p > 0$*, then the* $\mathscr{L}_p$*-space is defined,*

$$\mathscr{L}_p(\mathcal{X}, \mathbb{R}) = \left\{ \mu : \mathcal{X} \longrightarrow \mathbb{R} \,\middle|\, |\mu|_p := \left(\int_{\mathcal{X}} |\mu(x)|^p \, dx \right)^{1/p} < \infty \right\}.$$

Furthermore in the case of $p = \infty$,

$$\mathscr{L}_\infty(\mathcal{X}, \mathbb{R}) = \left\{ \mu : \mathcal{X} \longrightarrow \mathbb{R} \,\middle|\, |\mu|_\infty := \sup_{x \in \mathcal{X}} |\mu(x)| < \infty \right\}.$$

Definition 1.8 (Hölder-space)**.** *Let* $\alpha > 0$, $s > 0$ *and* $L > 0$ *constant, then* $\mathscr{H}^s_\alpha$ *is defined,*

$$\mathscr{H}^s_\alpha = \left\{ \mu : \mathcal{X} \longrightarrow \mathbb{R} \,\middle|\, \left|\mu^{(s)}(x) - \mu^{(s)}(y)\right| \leq L\,|x-y|^\alpha \textit{ for all } x, y \in \mathcal{X} \textit{ and some } L > 0 \right\},$$

where $\mu^{(s)}(x) = d^s\mu(x)/dx^s$ *is the order* s *derivative of* μ.

The Sobolev and Besov spaces require some auxiliary definitions. Start with the definition of a weak derivative of a function f.

Definition 1.9. *A function f is said to have a weak derivative $g = f'_w$ if for all $s < t$, $\int_s^t g(x)\, dx = f(t) - f(s)$.*

The weak derivative of order s will be denoted $f_w^{(s)}$.

Definition 1.10 (Sobolev-space). *Let $s > 0$ and $p > 0$ then a function $\mu \in \mathscr{W}_p^s(\mathcal{X}, \mathbb{R})$ if $\mu_w^{(s)} \in \mathscr{L}_p(\mathcal{X}, \mathbb{R})$ and $\mu \in \mathscr{L}_p(\mathcal{X}, \mathbb{R})$, that is,*

$$\mathscr{W}_p^s = \left\{ \mu : \mathcal{X} \longrightarrow \mathbb{R} \,\middle|\, |\mu|_p < \infty \quad and \quad \left|\mu_w^{(s)}\right|_p < \infty \right\}.$$

The Besov space is sometimes used as a functional class which contains the Sobolev and Hölder spaces as special cases. The Besov space relies on a quantity called the *Moduli of continuity*. First define the translation operator $T_{-h}f(x) := f(x-h)$.

Definition 1.11 (Moduli of continuity). *Let $f \in \mathscr{L}_p(\mathbb{R})$ with $1 \le p < \infty$, $\Delta_h f := T_{-h}f - f$ and for $m \in \mathbb{N}$, $\Delta_h^m f = \underbrace{\Delta_h \ldots \Delta_h}_{m\ times} f$. Then for $t > 0$, the moduli of continuity are defined,*

$$\omega_p^m(f, t) = \sup_{|h| \le t} |\Delta_h^m f|_p .$$

Definition 1.12 (Besov-space). *Let $1 \le p, q \le \infty$ and $s = n + \delta$ where $n \in \{0, 1, \ldots\}$ and $0 < \delta \le 1$. Then the Besov space is defined*

$$\mathscr{B}_p^{(s,q)} = \left\{ \mu : \mathcal{X} \longrightarrow \mathbb{R} \,\middle|\, \mu \in \mathscr{W}_p^n(\mathcal{X}, \mathbb{R}) \quad and \quad \left| t^{-\delta} \omega_p^2\left(\mu^{(n)}, t\right) \right|_p^* < \infty \right\},$$

where the modified $\mathscr{L}_p$ norm $|\cdot|_p^*$ is defined,

$$|g|_p^* = \begin{cases} \left(\int_0^\infty \frac{|g(t)|^p}{t}\, dt \right)^{1/p}, & if\ 1 \le p < \infty, \\ \operatorname*{ess\,sup}_{t \in (0,\infty)} |g(t)|, & if\ p = \infty. \end{cases}$$

1.4.2 *Minimax rates for regression function estimation*

In this section, minimax rates for estimators of μ covered in the literature will be considered along with their respective functional classes. Start with the simplest model in the fixed design context given by,

$$Y_i = \mu(x_i) + \varepsilon_i. \tag{1.5}$$

It was shown by Nussbaum (1985) that when ε_i are uncorrelated random variables and $\mathscr{F} = \mathscr{W}_2^s$ then the minimax rate using the squared-error loss is,

$$\rho_n = n^{-s/(2s+1)}. \tag{1.6}$$

Similar results were shown by Speckman (1985).

Now consider the case where (1.5) holds but the error variables are a dependent sequence, measured either by a LRD sequence with $\alpha \in (0, 1)$ or a SRD sequence. Then it has been shown in numerous scenarios (two of which are discussed below) that the minimax rate is,

$$\rho_n = n^{-\alpha s/(2s+\alpha)}. \tag{1.7}$$

The two cases where the rate (1.7) was shown to hold are by Hall and Hart (1990); Wang (1996). The former covered the case of kernel estimation when $\mathscr{F} = \mathscr{H}_0^2$ and the risk function is the probabilistic loss with the pseudo-metric that measures the pointwise distance about some $x_0 \in \mathcal{X}$. The latter using a wavelet estimator where $\mathscr{F} = \mathscr{B}_p^{(s,q)}$ (and $s > \alpha\,(2 - p)\,/\,(2p)$) and the risk function is the squared-error loss. An interesting result is that when the SRD case is considered it is possible to attain the same minimax rate for the uncorrelated noise scenario in (1.6). This is discussed further in Hall and Hart (1990).

Shift the focus now to estimators of μ in the random design model given by,

$$Y_i = \mu\,(X_i) + \varepsilon_i, \tag{1.8}$$

where X_i, ε_i are both random variables. The next result requires the definition of the p^{th} moment of a random variable. Let ξ be a random variable, then the p^{th} moment of ξ is,

$$\|\xi\|_p^p := \mathbb{E}\,|\xi|^p\,,$$

with the special case $\|\cdot\| := \|\cdot\|_2$. Then, in the simplest random design case when X_i and ε_i are both i.i.d. it has been shown by Stone (1982) that the minimax rate is,

$$\rho_n = n^{-s/(2s+1)}$$

where $\mathscr{F} = \mathscr{H}_\alpha^s$ for some $0 < \alpha \leq 1$ and the risk measure uses the probabilistic loss with the pseudo-metric $d = \|\cdot\|_p^p$ for some $p > 0$. This is the same as the minimax rate in the fixed design scenario with i.i.d. errors. However, if X_i are i.i.d. and ε_i are LRD then Yang (2001) has shown that the minimax rate using the squared-error loss is

$$\rho_n = n^{-min(s/(2s+1),\alpha/2)}$$

when $\mathscr{F} = \mathscr{B}_s^{(p,q)}$ and $s > 1/q - 1/2$. This result has recently been extended by Kulik and Raimondo (2009) to include the scenario of heteroskedastic noise. That is for the model,

$$Y_i = \mu(X_i) + \sigma(X_i)\,\varepsilon_i. \tag{1.9}$$

Their result only requires the added assumption that $\inf_{x \in \mathcal{X}} \sigma(x) > 0$.

As of yet, the minimax rate has not been established for (1.9) with LRD design variables. However, work on asymptotic theory for it has been developed by Zhao and Wu (2008) and Liu and Wu (2010) for the case of LRD design variables and i.i.d. error variables. There is also some literature for both LRD design and error variables by Guo and Koul (2008) and Kulik and Lorek (2011).

Consider now the indirect model or inverse problem when the observations are assumed to be derived from the fixed design model,

$$Y_i = K * \mu(x_i) + \varepsilon_i. \tag{1.10}$$

Unsurprisingly, the minimax rates of convergence for estimation under model (1.10) depend on the behaviour of the convolution kernel K. The level of severity of the blurring effect of K is measured by the so called *degree of ill-posedness* which is defined in the Fourier domain. Define the Fourier transform operator $\mathcal{F}$.

Definition 1.13. *The Fourier transform of a function f is,*

$$\mathcal{F}f(\omega) := \widetilde{f}(\omega) = \int_{\mathbb{R}} f(x)\, e^{-2\pi i \omega x}\, dx.$$

This leads to the definition of the degree of ill-posedness.

Definition 1.14. *Assume that $K \in \mathscr{L}_1(\mathbb{R}, \mathbb{R})$ and there exist constants $\beta > 0$ and $0 < c < C < \infty$ such that,*

$$c\left(1 + |\omega|^2\right)^{-\beta/2} \le |\mathcal{F}K(\omega)| \le C\left(1 + |\omega|^2\right)^{-\beta/2}$$

for all $\omega \in \mathbb{R}$. Then the function K is said to have degree of ill-posedness of level β.

It was shown by Donoho (1995) that when using the squared-error loss the minimax rate for functional estimation is

$$\rho_n = n^{-s/(2s+2\beta+1)}.$$

1.4.3 *Minimax rates for change point estimation*

In this section, the focus is on estimation of change-points. As such, the functional classes for the regression function $\mu : \mathcal{X} \longrightarrow \mathbb{R}$ need to not only satisfy similar smoothness properties to the ones considered in Section 1.4.2 but also need to have a change-point at some point $\theta \in \mathcal{X}$. The change-points considered are either a jump in the function or a jump in one of its derivatives. Similar to the Section 1.4.2, the minimax rates of estimators of the change-points θ will be considered and their respective functional classes with additional change-point properties will be introduced as the need arises.

Definition 1.15. *Let $\theta \in \mathcal{X}$, $0 \leq c < \delta \leq 1$ and $L > 0$ be constants. Then the Lipschitz class of functions that have a c-level cusp at θ is denoted $\mathscr{J}_c^\delta(\mathcal{X}, \theta)$ and $\mu \in \mathscr{J}_c^\delta(\mathcal{X}, \theta)$ if $\mu : \mathcal{X} \longrightarrow \mathbb{R}$ and,*

i) as h tends to zero, there exists a constant $C > 0$ such that,

$$|\mu(\theta + h) - \mu(\theta - h)| \geq 2C |h|^c;$$

ii) if $c = 0$, for all $x, y \in \mathcal{X}$ such that $x < y$ and $\theta \notin (x, y)$

$$|\mu(x) - \mu(y)| \leq L |x - y|^\delta; \quad \text{and}$$

iii) if $0 < c < 1$, then μ is differentiable everywhere on $\mathcal{X}$ except at the point θ.

Assuming the direct model (1.5) with i.i.d. error variables, it was claimed by Raimondo (1998) that the minimax rate is,

$$\rho_n = n^{-1/(2c+1)} \tag{1.11}$$

when $\mathscr{F} = \mathscr{J}_c^\delta(\mathcal{X}, \theta)$ and the indicator loss is used. However, this is not entirely correct, in fact it has been shown by Neumann (1997); Goldenshluger et al. (2006),

$$\rho_n = \min\left\{n^{-1/(2c+1)}, n^{-2/(2c+3)}\right\},$$

which has an elbow in the rate when $c = \frac{1}{2}$, is the true minimax rate for this problem. Discussion of the specific assumptions and proofs are given in Goldenshluger et al. (2006).

For comparison recall that when $c = 0$, a c-cusp is the same as a jump and the minimax rate is $\rho_n = n^{-1}$ which is consistent with the theory already established by Korostelëv (1987).

Also, an estimator has been constructed by Cheng and Raimondo (2008) to estimate a jump in $\mu^{(1)}$, the first derivative of μ among a smoother functional class $\mathscr{F} = \mathscr{C}_s^1(\mathcal{X}, \theta)$ defined below.

Definition 1.16. *Let $s \geq 2$ be an integer and $a \in \mathbb{R}$ with $a \neq 0$. Then, we say that $\mu \in \mathscr{C}_s^1(\mathcal{X}, \theta)$ if,*

1. *$\mu : \mathcal{X} \longrightarrow \mathbb{R}$ has a kink, that is there exists $\theta \in \mathcal{X}$ such that,*

$$\left[\mu^{(1)}\right](\theta) = \mu^{(1)}\left(\theta^{+}\right) - \mu^{(1)}\left(\theta^{-}\right) = a.$$

2. *For all $x \in \mathbb{R}, x > 0$,*

$$\mu^{(1)}(\theta - x) = \sum_{j=0}^{s-1} \frac{(-x)^j}{j!} \mu^{(j+1)}\left(\theta^{-}\right) + \mathcal{O}\left(x^s\right).$$

3. *For all $x \in \mathbb{R}, x > 0$,*

$$\mu^{(1)}(\theta + x) = \sum_{j=0}^{s-1} \frac{x^j}{j!} \mu^{(j+1)}\left(\theta^{+}\right) + \mathcal{O}\left(x^s\right).$$

This is the functional class that will be used in Part II and a more generalised version is defined therein. The functional classes considered thus far for minimax rates only satisfy a Lipschitz condition away from the change-point, the class $\mathscr{C}_s^1(\mathcal{X}, \theta)$ considers the possibility of smoother functions indexed by the parameter s. The estimator constructed by Cheng and Raimondo (2008) was shown to achieve the rate,

$$\rho_n = n^{-s/(2s+1)}.$$

However, their proof requires a slight modification of their functional class. This is discussed further in Part II and Part III.

For the random design model in (1.8) an estimator for a sharp change point has been constructed by Park and Kim (2006) when both the design and errors are i.i.d. random variables. Their estimator achieves nearly the same rate as (1.11) (up to a logarithmic term) which suggest near minimax optimality. However, the minimax rates for the random design framework have not been established.

Similarly, for the heteroskedastic random design model given in (1.9), Huh and Park (2004) have constructed an estimator for sharp change points when the design and error variables are i.i.d. which achieves the same rate as (1.11).

Finally, for the indirect model in (1.10), minimax rates were established by Goldenshluger et al. (2006) who proved that the minimax rate is,

$$\rho_n = n^{-(s+1)/(2s+2\beta+1)}$$

where the degree of ill-posedness $\beta \in (1/2, 1)$ and they assume that $\mathscr{F} = \mathscr{G}_s(\theta)$ which is defined below.

Definition 1.17. *Let $s > 1$ be an integer. Then a function $\mu \in \mathscr{G}_s(\theta)$ if $\mu \in \mathscr{L}_2(\mathbb{R}, \mathbb{R})$ and μ has a change-point at $\theta \in [0, 1]$ such that,*

1. *The left and right first derivatives are equal, that is,* $\mu^{(1)}(\theta_-) = \mu^{(1)}(\theta_+)$ *and the function* $\mu^{(1)}$ *is continuous.*
2. *The function* $\mu^{(1)} \in \mathscr{L}_2(\mathbb{R}, \mathbb{R})$ *and*

$$\int_{\mathbb{R}} \left| \mathcal{F}\mu^{(1)}(\omega) \right| |\omega|^{s-1} \, d\omega < \infty.$$

For the remainder of the thesis, the symbol θ will refer to a *kink* in a given regression function μ.

1.5 THESIS OUTLINE

To begin the body of the thesis, the minimax framework is introduced as it will be the main tool of measurement to compare the performance of estimators. This is covered in Section 1.4 along with a literature review of regression (and change point) estimators with their respective performance and functional classes.

The main kink estimation technique is covered Part II and is a kernel smoothing approach to estimate kink locations. The technique assumes a fixed design framework where the $\mathcal{E}$ variables are LRD. It is the crucial method that underpins all the subsequent analysis through the thesis. The performance and minimax optimality of this method is then clarified in Part III. An extension to the random design framework is given in Part IV and Part V. In this context the driving LRD process is assumed to be a causal linear process. The assumption in Part IV is that X are i.i.d. and $\mathcal{E}$ are LRD. In Part V we assume that X are LRD and $\mathcal{E}$ are i.i.d.. The assumption of both LRD errors and design variables is a topic for future investigation.

Part II

FIXED DESIGN

2

KINK ESTIMATION UNDER LONG RANGE DEPENDENCE

In this chapter the primary problem is to find a kink location from the fixed design framework in the presence of LRD noise. Aspects of this work has already been published in Wishart (2009), although full treatment is given here with tightened results and increased generality. Furthermore, in Part III the result is shown to be rate optimal in the minimax sense for a particular class of functions to be introduced later.

The method discussed here is called the Zero-Crossing Technique (ZCT) which is a technique that was pioneered by Goldenshluger et al. (2006). The ZCT is crucial to the entire thesis since it is used both here in the fixed design framework and is extended for use later in the random design cases given in Part IV and Part V. The ZCT was pursued for our purposes since it has already been established by Goldenshluger et al. (2006) as a optimal method in the minimax sense for the indirect model in the fixed design setting with i.i.d. errors. As will be shown, the 'optimal' ZCT approach has been modified to allow for a LRD error structure in the direct fixed design model.

The chapter is broken into three Sections. The assumptions on the regression model are outlined in Section 2.1. There are two regression models that are under consideration, the sampling model used in practice and the theoretical asymptotic model that is used in the analysis. The ZCT estimation method is constructed and analysed in Section 2.2. As alluded to earlier, a kernel smoothing implementation of the ZCT is used and the class of kernel functions is constructed along with the methodology of the ZCT. A numerical study of the efficiency of the modified ZCT is given in Section 2.3 to give weight to the results on the minimax optimality of the method presented in Part III. Finally a discussion of the results is presented in Section 2.4. The discussion includes the extension of the technique to other situations of multiple kink locations and change points in higher order derivatives. Lastly, the reason behind the ZCT methodology is explored with respect to its minimax optimality.

2.1 MODEL ASSUMPTIONS

There will be two main models considered in this chapter. The *sampling model* that will be used in practice and is introduced in Section 2.1.1 and the *asymptotic model* that is used for the theoretical analysis and is introduced in Section 2.1.2. The sampling model is the discrete model that will be observable in practice for an appropriate dataset. The asymptotic model is essentially the passage of the sampling model to the limit as the sample size n approaches infinity. The last sentence should be interpreted with caution however, since the dynamics between each model is much more subtle as will become evident in Section 2.1.3. Lastly, the relevant functional class $\mathscr{F} = \mathscr{C}_s^m(\mathcal{X}, \boldsymbol{\theta})$ for the regression function in our model will be described in Section 2.1.4.

2.1.1 *Sampling model*

In practice, it is assumed that the function μ is not directly observable. It is only observed at a discrete set of points in the presence of noise. That is, the pairs $\{(x_i, y_i)\}_{i=1}^n$ are observed such that,

$$y_i = \mu(x_i) + \sigma\xi_i \tag{2.1}$$

where $\{x_i\}_{i=1}^n$ are a uniform grid of points contained in a compact interval $I \subset \mathbb{R}$. For our case, the error variables $\{\xi_i\}_{i=1}^n$ are assumed to be a Gaussian LRD sequence with mean zero and unit variance and $\sigma > 0$ is constant. Without loss of generality set $I = [0,1]$ since any compact interval can be rescaled to $[0,1]$ by an affine transformation. By doing this, the design variables can be defined as $x_i = i/n$ for $i = 1, 2, \ldots, n$. This simplification eases the notation and it allows the analysis to be conducted and compared with other methods without worrying about the scaling of the design points.

2.1.2 *Asymptotic model*

As will become evident later, for theoretical purposes and mathematical convenience it is easier to consider the regression model and the kink estimator for the model that is in some sense least the asymptotic proxy of (2.1). This asymptotic model based on the fractional Brownian motion (fBm). This is referred to as the fractional white noise model and has nice mathematical properties in comparison to the sampling model (2.1). Before the fractional white noise model is formally introduced some preliminary definitions and concepts are covered.

Definition 2.1. *A process $\{Y(t)\}_{t\in\mathbb{R}^+}$ is said to be self-similar of order $H \in (0,1)$ if for all $t \in \mathbb{R}^+$ and any constant $a > 0$,*

$$Y(t) \overset{\mathscr{D}}{=} a^{-H} Y(at)$$

where the equality means equality in distribution.

The H parameter is referred to in the literature as the *Hurst* parameter in honour of Harold Hurst and his seminal work on the presence of long-range dependent noise in hydrology given in Hurst (1951). The fBm concept is an extension of Brownian motion that can exhibit dependence among its increments which is typically controlled by the Hurst parameter, H. In contrast, the increments of a standard Brownian motion are independent. The most common definition of fBm in the modern literature from a probabilistic point of view is given by

Definition 2.2. *The fBm process* $\{B_H(t)\}_{t\in\mathbb{R}^+}$ *is a Gaussian process which has mean zero and covariance structure,*

$$\mathbb{E}B_H(t)B_H(s) = \frac{1}{2}(t^{2H} + s^{2H} - |t-s|^{2H}).$$

The standard Brownian motion is recovered in Definition 2.2 with the choice $H = 1/2$. The first formal definition of fBm was given by Mandelbrot and Van Ness (1968) as a particular moving average representation of a regular Brownian motion $B(t)$.

Definition 2.3. *The fBm* B_H *defined over the whole real line has representation,*

$$B_H(t) = \frac{C_{H,1}}{\Gamma\left(H+\frac{1}{2}\right)} \int_{\mathbb{R}} (t-s)_+^{H-1/2} - (-s)_+^{H-1/2} dB(s).$$

where $x \vee y := \max\{x, y\}$ *and* $x_+ = 0 \vee x$,

$$C_{H,1} := \sqrt{2H\sin(\pi H)\,\Gamma(2H)}.$$

The constant $C_{H,1}$ *is a normalisation constant to ensure that Definition 2.2 is satisfied.*

There is another definition in the literature that defines the fBm over a compact interval. This is the definition we will use since we are only interested in the specific compact interval $[0,1]$ for most of our purposes.

Definition 2.4. *The fBm* B_H *defined over the interval* $[0,1]$ *can be represented by,*

$$B_H(t) = \frac{C_H}{\Gamma\left(H+\frac{1}{2}\right)} \int_0^t (t-s)^{H-1/2} dB(s),$$

for $t \in [0,1]$ *and* $H \in (1/2, 1]$ *where*

$$C_H = \sqrt{\frac{\pi H}{\cos(\pi H)\,\Gamma(1-2H)}}.$$

The above definitions ensure that the fBm $B_H(t)$ has been normalised such that $\mathbb{V}\mathrm{ar}B_H(t) = t^{2H}$. As previously mentioned the level of dependence in the fBm is controlled by H. The regular Brownian motion is recovered with the choice $H = 1/2$ giving independence among increments. If $0 < H < 1/2$, the increments are negatively correlated and are SRD. If $H > 1/2$, then the increments are positively correlated and are LRD. Now the preliminaries for the asymptotic model have been covered, it can now be defined. A continuous proxy model of the sampling model in (2.1), or *fractional white noise model*, assumes the structure,

$$dY(t) = \mu(t)dt + \varepsilon^{\alpha} dB_H(t), \tag{2.2}$$

where $B_H(t)$ is a normalised fractional Brownian motion (fBm) on I with Hurst parameter $H \in [1/2, 1)$ and the noise level is controlled by ε and $\alpha := 2 - 2H$. More specifically, there is a particular choice for ε that connects the two models in practice. Then the choice $\varepsilon \asymp n^{-1/2}$ ensures that there is a close relationship between the two models (2.1) and (2.2). The asymptotic model (2.2) was first considered by Wang (1996) to study the effect of LRD on contemporary nonparametric methods. Under this notation we see that $0 < \alpha \leq 1$ so that the level of dependence is scaled between 0 and 1. The white noise model is included at $\alpha = 1$ and there is an increased level of dependence in the increments as α tends to zero.

2.1.3 *Connection between the models*

Ideally, a proof of asymptotic equivalence between models (2.1) and (2.2) would show that all results obtained under (2.2) will have the same asymptotic behaviour under (2.1). Asymptotic equivalence has been shown to hold between (2.1) and (2.2) when the error variables are i.i.d. and $\alpha = 1$ by Brown and Low (1996). Unfortunately, asymptotic equivalence has not yet been proven between (2.1) and (2.2) in general. It has been shown by Carter (2010) that for asymptotic equivalence to hold it is sufficient to impose a stringent covariance structure for the errors under (2.1). It is unknown whether it is necessary at this stage.

As will become evident in Section 2.3, the estimator that is developed assuming model (2.2) performs well in (2.1), the discretised model that will hold in practice. So for our purposes the relationship between the two models will be pursued to justify the use of (2.2). Wang (1996) was interested in the estimation of μ itself and proposed (although never fully proved) that the sampling model (2.1) has the same asymptotic risk as the fractional white noise model, (2.2). Their approach makes use of a result by (Taqqu, 1975, Corollary 5.1) that shows that a discrete LRD sequence, $\{\xi_i\}_{i=1}^n$, under appropriate normalisation, converges to a fBm. The corollary by Taqqu is stated below for easy reference.

Corollary 5.1 (Taqqu, 1975). *Let $\{\xi_i\}_{i=1}^n$ be a sequence of mean zero, unit variance, stationary random variables with autocorrelation $\mathbb{E}\xi_0\xi_k \sim k^{-\alpha}$. Then for a fixed $t \in (0,1)$ the following convergence result holds,*

$$\frac{C_\alpha}{n^{1-\alpha/2}} \sum_{j=1}^{\lfloor nt \rfloor} \xi_j \xrightarrow{\mathscr{D}} B_H(t) \quad \text{as } n \to \infty,$$

where $C_\alpha = \sqrt{(2-\alpha)(1-\alpha)/2}$ and the ceiling function is defined

$$\lfloor x \rfloor := \max\{i \in \mathbb{Z} | i \leq x\}.$$

To compare models (2.1) and (2.2), first look at the finite discrete analogue of Y given a set of observations $\{(x_i, y_i)\}_{i=1}^n$ that follow (2.1). Define the discrete version of μ with,

$$\mu_n(x) = \sum_{i=1}^{n} \mathbb{1}_{(x_{i-1}, x_i]}(x)\mu(x_i),$$

where $x_0 \equiv 0$. Use a similar technique to the one suggested by Wang (1996, 1997) by defining the cumulative process $\{Y_n(x_i) | x_i = i/n, i \in \{1, 2, \ldots, n\}\}$ with,

$$Y_n(0) := 0, \qquad Y_n(x_i) := \frac{1}{n} \sum_{j=1}^{i} y_j.$$

Then for any fixed $t \in (0, 1)$, there exists a $i \in \{1, 2, \ldots, n\}$ such that $i = \lfloor nt \rfloor$. Moreover, to exploit Corollary 5.1 (Taqqu, 1975), for this fixed value of t, further define the cumulative process Y_n,

$$\begin{aligned}
Y_n(t) &:= Y_n(i/n) \\
&= Y_n\left(\frac{\lfloor nt \rfloor}{n}\right) \\
&= n^{-1} \sum_{j=1}^{\lfloor nt \rfloor} \mu(x_j) + \frac{\sigma}{n} \sum_{j=1}^{\lfloor nt \rfloor} \xi_j \\
&= \frac{1}{n} \sum_{j=1}^{\lfloor nt \rfloor} \mu(x_j) + \frac{n^{-\frac{\alpha}{2}}\sigma}{C_\alpha} \left\{ \frac{C_\alpha}{n^{1-\frac{\alpha}{2}}} \sum_{j=1}^{\lfloor nt \rfloor} \xi_j \right\}. \qquad (2.3)
\end{aligned}$$

By the result of Corollary 5.1 (Taqqu, 1975) (see page 25) there exists a sequence $\delta_{t,n}$ such that,

$$\frac{C_\alpha}{n^{1-\alpha/2}} \sum_{j=1}^{\lfloor nt \rfloor} \xi_j = B_H(t) + \delta_{t,n}, \quad \text{where } \delta_{t,n} = o_p(1). \qquad (2.4)$$

Let $\varepsilon_n^\alpha := n^{-\alpha/2}\sigma/C_\alpha$. Combining, (2.4) and (2.3) and noting that the μ function is only evaluated on its discrete points we get,

$$Y_n(t) = \int_0^t \mu_n(s)\, ds + \varepsilon_n^\alpha B_H(t) + o_p\left(n^{-\alpha/2}\right). \qquad (2.5)$$

Therefore, using the calibration $\varepsilon_n \asymp n^{-1/2}$ in (2.2) with (2.5) we have that for any $t \in (0, 1)$,

$$|Y_n(t) - Y(t)| = \left| \int_0^t (\mu_n(s) - \mu(s))\, ds \right| + o_p\left(n^{-\alpha/2}\right).$$

That is, Y and its discrete equivalent Y_n differ by at most the $\mathscr{L}_1$ distance between μ_n and μ and a term that is $o_p\left(n^{-\alpha/2}\right)$. Moreover, if μ is at least Lipschitz continuous with Lipschitz constant L_μ and then for any $s \in [0,1]$, $|\mu_n(s) - \mu(s)| = \mathcal{O}\left(n^{-1}\right)$. Indeed,

$$|\mu(s) - \mu_n(s)| = \left|\sum_{i=1}^{n} \mathbb{1}_{(x_{i-1},x_i]}(s)\left(\mu(s) - \mu(x_i)\right)\right| \leq L_\mu \max_{2\leq i\leq n} |x_i - x_{i-1}| = L_\mu n^{-1} = \mathcal{O}\left(n^{-1}\right).$$

Therefore, we can see that the sampling model (2.1) is in some sense close to the asymptotic fractional white noise (2.2) since,

$$|Y_n(x) - Y(x)| = o_p\left(n^{-\alpha/2}\right).$$

It is worth noting that the fractional white noise model seems like a reasonable choice for analysis of any estimator that will in practice use the sampling model (2.1), as long as the rate of convergence for the estimator dominates the stochastic difference between the two models. This stochastic difference is given by a term that is of order $o_p\left(n^{-\alpha/2}\right)$. As will be shown in Part III, the minimax rate for estimation of a kink location θ in a particular functional class that we consider is $o(n^{-\alpha/2})$. This implies that the fractional white noise model might be a viable theoretical alternative to the sampling model used in practice.

2.1.4 *Functional class*

We assume that we are in a regression model framework where the objective function μ is a smooth function with finitely many kinks. More specifically we assume that μ comes from a particular smoothness class $\mathscr{C}_s^m(\mathcal{X}, \boldsymbol{\theta})$. This class is defined below.

Definition 2.5. *Let $m \geq 1$ and $s \geq 2$ be integers and $\mathbf{a} = (a_1, a_2, \ldots, a_m) \in \mathbb{R}^m$ be fixed such that, $a_i \in \mathbb{R} \setminus \{0\}$. Then, we say that $\mu \in \mathscr{C}_s^m(\mathcal{X}, \boldsymbol{\theta})$ if,*

1. *The function $\mu : \mathcal{X} \longrightarrow \mathbb{R}$ has m kinks, that is, there exists a $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_m)$ such that $\theta_i \neq \theta_j$ for $i \neq j$ and $\theta_i \in \mathcal{X}$ for $i = 1, 2, \ldots, m$; and*

$$\left[\mu^{(1)}\right](\theta_i) = \mu^{(1)}(\theta_i^+) - \mu^{(1)}(\theta_i^-) = a_i \neq 0.$$

2. *The higher order derivatives $\mu^{(j)}$ exist for $j = 2, 3, \ldots, s-1$; are finite everywhere and satisfy,*

$$\mu^{(j)}(\theta_i^+) = \mu^{(j)}(\theta_i^-) \qquad \textit{for } i = 1, 2, \ldots, m. \tag{2.6}$$

3. *For all $x > 0$ such that $(\theta_i - x) \in \mathcal{X}$ and $\theta_j \notin (\theta_i - x, \theta_i)$ for $j \neq i$,*

$$\mu^{(1)}(\theta_i - x) = \sum_{k=0}^{s-2} \frac{(-x)^k}{k!} \mu^{(k+1)}(\theta_i^-) + \mathcal{O}(x^{s-1}). \tag{2.7}$$

4. *For all $x > 0$ such that $(\theta_i + x) \in \mathcal{X}$ and $\theta_j \notin (\theta_i, \theta_i + x)$ for $j \neq i$,*

$$\mu^{(1)}(\theta_i + x) = \sum_{k=0}^{s-2} \frac{x^k}{k!} \mu^{(k+1)}(\theta_i^+) + \mathcal{O}(x^{s-1}). \tag{2.8}$$

This definition is used to ensure that the regression function μ is sufficiently smooth to be used for analysis. The second condition of Definition 2.5 might seem overly restrictive but it is a technical condition that is used to exploit the smoothness of the function with the yet to be introduced class of kernel functions used in the estimator $\widehat{\theta}_n$. The third and fourth conditions of Definition 2.5 ensure that μ has, at least, order s derivatives at all points away from the change-points $\boldsymbol{\theta}$ and that $\mu'(x)$ can be approximated from the left hand or right hand side of each kink, θ_i, by using Taylor's formula. An example of a function $\mu \in \mathscr{C}_3^1\left([0,1], \frac{2}{3}\right)$ is given in Figure 2. In particular the function is defined by,

$$\mu(x) := \begin{cases} 4 - \frac{85}{9}x + \frac{97}{12}x^2 - x^3, & \text{if } 0 \leq x < \frac{2}{3}; \\ \frac{82}{9} - \frac{508}{9}x + \frac{1159}{12}x^2 - \frac{181}{4}x^3 & \text{if } \frac{2}{3} < x \leq 1; \\ 0 & \text{otherwise.} \end{cases} \tag{2.9}$$

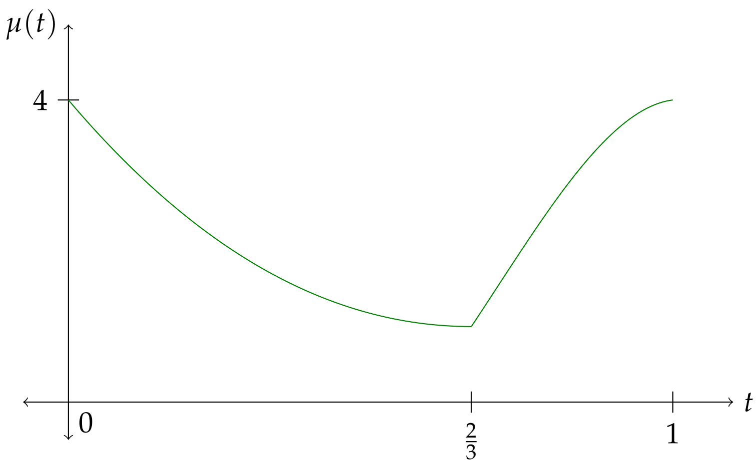


Figure 2: The example function μ with a single kink at $\theta = \frac{2}{3}$ and $s = 3$.

For completeness and comparison purposes introduce another smoothness class $\mathscr{S}_s$ to denote the class of smooth functions that do not have a kink (or any change-points).

Definition 2.6. *Let $s \geq 2$ be an integer. Then, $\mu \in \mathscr{S}_s(\mathcal{X})$ if $\mu : \mathcal{X} \longrightarrow \mathbb{R}$ and*

$$\mu(t + x) = \sum_{j=0}^{s-1} \frac{x^j}{j!} \mu^{(j+1)}(t) + \mathcal{O}(x^s)$$

for all $t, x \in \mathcal{X}$.

2.2 KINK ESTIMATION METHOD

The kink estimation procedure that will be pursued employs the Zero-Crossing Technique (ZCT). The ZCT was pioneered by Goldenshluger et al. (2006) and implemented for the i.i.d. noise scenario by Cheng and Raimondo (2008). The method considered by Goldenshluger et al. (2006) considers change point estimation from the perspective of indirect independent noisy observations. That is, they assume that observations $x \in I$ are derived from the model,

$$dY(x) = (\mathrm{K}{*}\mu)(x)dx + \xi dB(x)$$

where $\mathrm{K}{*}\mu$ is the convolution of μ and K, $B(\cdot)$ is a standard Brownian motion and $\xi \in (0,1)$ is the noise level. In this scenario the regression function μ is not observed directly but $(\mathrm{K}{*}\mu)$ is instead. Intuitively this can be interpreted as observing a blurred signal of μ where the level of blurring is controlled by the function K. Furthermore, the degree of difficulty of estimation or severity of the blurring effect in K is measured by a quantity called the degree of ill-posedness that will be described later in Section 3.2. The ZCT described by Goldenshluger et al. (2006) is a general theoretical construction of an estimator for a change point in μ. They also proved that the method is optimal in the minimax sense.

The ZCT was adapted by Cheng and Raimondo (2008) to the fixed design model in the direct setting. The ZCT will be extended in this chapter to the fixed design model in the direct setting with LRD errors. The key results have been published in Wishart (2009). It is extended further later in the thesis to the random design setting in Parts IV and V. The key results for the random design treatment have been published in Wishart and Kulik (2010).

The work of Cheng and Raimondo (2008) and Goldenshluger et al. (2006) will be reviewed to give the reader a detailed context of the method. In Cheng and Raimondo (2008) it was assumed that the observations are derived from the asymptotic white noise model, where for each $x \in I$,

$$dY(x) = \mu(x)dx + \varepsilon dB(x). \tag{2.10}$$

The estimation technique considered in Cheng and Raimondo (2008) is concerned with kink estimation and they followed a kernel smoothing approach that closely follows the ZCT. The kernel function they employed for their construction of an optimal estimator is a high order kernel function built from a Legendre basis. In contrast to Goldenshluger et al. (2006), it was built specifically to deal with observations in the direct setting. It is defined in the next section.

2.2.1 *Kernel construction*

The kernel function that is needed for the ZCT requires a few technical conditions that are given below. These conditions are required to create an estimator that can locate θ and additionally exploit the smoothness of the functional class $\mathscr{C}_s^1(\mathcal{X},\theta)$. As will become evident in the description of estimation method in Section 2.2, high order derivatives of the kernel function K are used. To ease the notation, define $K_i(x) := d^iK(x)/dx^i$. The technical conditions for the kernel function, K, are:

$(K:1)$ Support(K) = [-1,1] and K_3 is at least Lipschitz continuous.

$(K:2)$ $K_1(x) = -K_1(-x)$.

$(K:3)$ $K_1(0) = K_1(-1) = K_1(1) = 0$, $K_2(-1) = K_2(1) = 0$.

$(K:4)$ If $s \geq 2$ then,

$$m_j = \int_{-1}^{1} x^j K_2(x)dx = 0, \qquad \text{for } j = 0, 1, \dots, s-2.$$

$(K:5)$ $|K_1(x)| \geq C|x|$, for all $x \in [-q, q]$, for some constants $q \in (0, 1)$, $C > 0$.

$(K:6)$ $K_1(x) > 0$ for $x \in [-b, 0]$ for some $0 < b < 1$ and there exists a unique global maximum at $-q^* \in (-b, 0)$. Since K_1 is odd it follows that $K_1(x) < 0$ for $x \in [0, b]$ and there exists a unique global minimum at $q^* \in (0, b)$.

These conditions ensure that the kernel K has bounded support and that the smoothed derivatives are able to exploit the smoothness conditions (2.7) and (2.8) by the vanishing moments property in (K:3). Condition (K:6) ensures that the smoothed first derivative has a unique global maximum and minimum near the zero-crossing times. This is a critical property that the kernel function needs to possess for the ZCT to work. This will become evident in the description of the method in Section 2.2.1.1. The crucial points of this class of kernel functions are shown in Figure 3. The first positive root of K_1, denoted by q_0, is shown as well. This point is used in a key Lemma in Part III which concerns the minimax optimality of this method.

The specific construction of the general class of high order kernels is now pursued. In particular, the Legendre polynomial basis is used and the coefficients for each Legendre component are set to ensure that conditions (K:1) - (K:6) hold.

Start by introducing the Legendre polynomials that will be used as the basis for construction. The Legendre polynomials can be defined succinctly on $[-1, 1]$ by the Roderiguéz formula,

$$P_n(x) := \frac{1}{2^n n!} \frac{d^n}{dx^n}(x^2 - 1)^n.$$

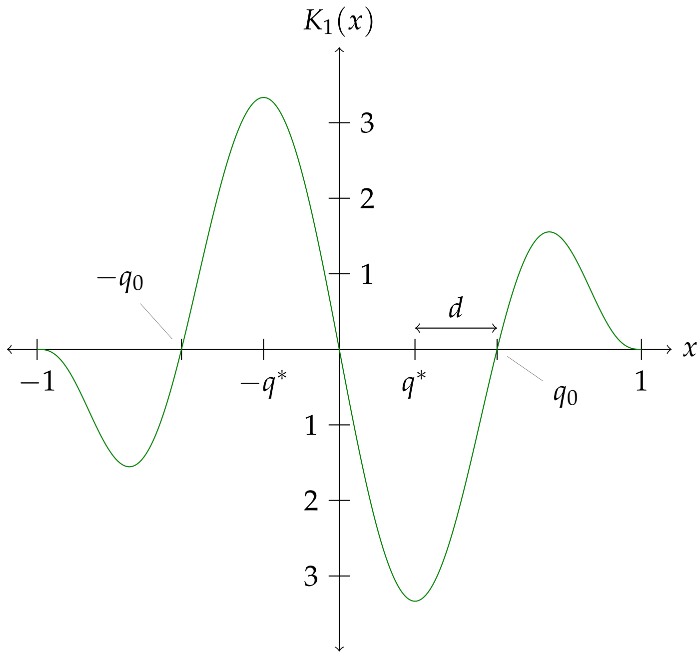


Figure 3: Plot of the K_1 function showing the crucial points used in the analysis.

To save on notation define the differential operator $\mathcal{D}^n := d^n/dx^n$. Using the definition with the binomial expansion and Leibniz rule of differentiation it can be shown that,

$$P_n(x) = 2^{-n} \sum_{i=0}^{\lfloor \frac{n}{2} \rfloor} (-1)^i \frac{(2n-2i)!}{i!(n-i)!(n-2i)!} x^{n-2i}. \tag{2.11}$$

The kernel function K is constructed by considering K_2, the second order derivative, to be an even function and an order $s+3$ polynomial where $s \geq 3$ is an odd integer. To guarantee that s is an odd number let $s = 2k+1$ where $k \in \mathbb{N}$. It will be shown that the required polynomial that satisfies (K:1)–(K:6) has,

$$K_2(x) = \gamma_s \sum_{j=\lfloor s/2 \rfloor}^{s+1} \alpha_{j,s} x^{2j-s+1}, \tag{2.12}$$

$$\text{where } \gamma_s = \frac{(2s+3)!}{2^{2s+3}(s-1)!(s+1)!}, \quad \alpha_{j,s} = \frac{(-1)^{\lfloor s/2 \rfloor + j+1}(2j)!}{j!(s-j+1)!(2j-s+1)!}.$$

$$K_3(x) = \gamma_s \sum_{j=\lfloor s/2 \rfloor + 1}^{s+1} \beta_{j,s} x^{2j-s},$$

$$\text{where } \beta_{j,s} = \frac{(-1)^{\lfloor s/2 \rfloor + j+1}(2j)!}{j!(s-j+1)!(2j-s)!}.$$

A simple consequence of (2.11) is that $P_n(-x) = (-1)^n P_n(x)$, implying that P_n will be even (or odd) if the index n is even (or odd). It is required by (K:4) that the moment condition is satisfied,

$$m_j = \int_{-1}^{1} x^j K_2(x)dx = \langle x^j, K_2\rangle = 0 \qquad \text{for } j = 0, 1, \ldots, 2k-1. \tag{2.13}$$

Notice that by forcing K_2 to be an even function it follows that $m_j = 0$ for all odd j. Using (2.11) and (2.13),

$$\begin{aligned}\langle P_{2\ell}, K_2\rangle &= 2^{-2\ell}\sum_{j=0}^{\ell}(-1)^j \frac{(4\ell-2j)!}{j!(2\ell-j)!(2\ell-2j)!}\langle x^{2\ell-2j}, K_2\rangle \\ &= \sum_{j=0}^{\ell}(-1)^j a(2\ell, j) m_{2\ell-2j} \\ &\qquad \text{where } a(2\ell, j) = \frac{2^{-2\ell}(4\ell-2j)!}{j!(2\ell-j)!(2\ell-2j)!}.\end{aligned} \tag{2.14}$$

Combining (2.13) and (2.14) it is easy to see that,

$$\langle P_i, K_2\rangle = \begin{cases} a(2k,0)m_{2k}, & \text{if } i = 2k \\ a(2k+2,0)m_{2k+2} - a(2k+2,1)m_{2k}, & \text{if } i = 2k+2 \\ a(2k+4,0)m_{2k+4} - a(2k+4,1)m_{2k+2} + a(2k+4,2)m_{2k}, & \text{if } i = 2k+4 \\ 0, & \text{otherwise.} \end{cases} \tag{2.15}$$

Let s be given. Since K_2 is a polynomial of degree $s+3 = 2k+4$, the function K_2 can be constructed via three even Legendre polynomials,

$$K_2(x) = \sum_{i=k}^{k+2} \frac{\langle P_{2i}, K_2\rangle}{\langle P_{2i}, P_{2i}\rangle} P_{2i}(x), \tag{2.16}$$

where the coefficients are given by (2.14). Using (2.15) and (2.16) and the fact that $\langle P_{2k}, P_{2k}\rangle = 2/(4k+1)$ yields,

$$
\begin{aligned}
K_2(x) &= \sum_{i=k}^{k+2} \frac{\langle P_{2i}, K_2\rangle}{\langle P_{2i}, P_{2i}\rangle} P_{2i}(x) \\
&= \sum_{i=k}^{k+2} \tfrac{4i+1}{2} \left\{ \sum_{j=0}^{i} (-1)^j a(2i,j) m_{2i-2j} \right\} P_{2i}(x) \\
&= \tfrac{4k+1}{2} a(2k,0) m_{2k} P_{2k}(x) + \tfrac{4k+5}{2} P_{2k+2}(x) (a(2k+2,0) m_{2k+2} - a(2k+2,1) m_{2k}) \\
&\quad + \tfrac{4k+9}{2} P_{2k+4}(x)\, (a(2k+4,0) m_{2k+4} - a(2k+4,1) m_{2k+2} + a(2k+4,2) m_{2k}) \\
&= m_{2k} \left(\tfrac{4k+1}{2} a(2k,0) P_{2k}(x) - \tfrac{4k+5}{2} a(2k+2,1) P_{2k+2}(x) + \tfrac{4k+9}{2} a(2k+4,2) P_{2k+4}(x) \right) \\
&\quad + m_{2k+2} \left(\tfrac{4k+5}{2} a(2k+2,0) P_{2k+2}(x) - \tfrac{4k+9}{2} a(2k+4,1) P_{2k+4}(x) \right) \\
&\qquad + m_{2k+4} \tfrac{4k+9}{2} a(2k+4,0) P_{2k+4}(x).
\end{aligned} \tag{2.17}
$$

We also require that the boundary points are zero, that is, $K_2(1) = K_2(-1) = K_3(1) = K_3(-1) = 0$. We only consider the condition that $K_2(1) = K_3(1) = 0$. The other two conditions follow by symmetry since $K_2(-x) = K_2(x)$. We first consider the value of the m_i coefficients at $x = 1$. It will be shown that $P_n(1) = 1$ for all $n \in \mathbb{N}$ via the Roderiguéz representation and Liebniz rule of differentiation,

$$
\begin{aligned}
P_n(1) &= \frac{1}{2^n n!} \mathcal{D}^n \left\{ (x-1)^n (x+1)^n \right\} \big|_{x=1} \\
&= \frac{1}{2^n n!} \sum_{i=0}^{n} \binom{n}{i} \mathcal{D}^i (x+1)^n \mathcal{D}^{n-i} (x-1)^n \Bigg|_{x=1} \\
&= \frac{n! 2^n}{2^n n!} \\
&= 1.
\end{aligned}
$$

The coefficients of each of the m components are now evaluated when $x = 1$, starting with m_{2k},

$$
\begin{aligned}
&\frac{4k+1}{2} a(2k,0) - \frac{4k+5}{2} a(2k+2,1) + \frac{4k+9}{2} a(2k+4,2) \\
&\quad = \frac{4k+1}{2^{2k+1}} \frac{(4k)!}{((2k)!)^2} - \frac{4k+5}{2^{2k+3}} \frac{(4k+2)!}{(2k+1)!(2k)!} + \frac{4k+9}{2^{2k+6}} \frac{(4k+4)!}{(2k+2)!(2k)!} \\
&\quad = \frac{(4k+1)!}{2^{2k+1}((2k)!)^2} \left\{ 1 - \frac{(4k+5)}{2} + \frac{(4k+9)(4k+3)}{2^3} \right\} \\
&\quad = \frac{(4k+1)!(4k+3)(4k+5)}{2^{2k+4}((2k)!)^2}.
\end{aligned} \tag{2.18}
$$

Evaluating the coefficient for m_{2k+2},

$$\begin{aligned}
&\frac{4k+5}{2}a(2k+2,0) - \frac{4k+5}{2}a(2k+4,1) \\
&\qquad = \frac{(4k+5)!}{2^{2k+3}((2k+2)!)^2} - \frac{(4k+9)(4k+6)!}{2^{2k+5}(2k+3)!(2k+2)!} \\
&\qquad = -\frac{(4k+5)!(4k+7)}{2^{2k+4}((2k+2)!)^2}.
\end{aligned} \tag{2.19}$$

The coefficient of the last term m_{2k+4} is $(4k+9)!/(2^{2k+5}((2k+4)!)^2)$. Using (2.19), (2.18) and the requirement that $K_2(1) = 0$ implies, after simplification, that

$$m_{2k} = \frac{4(4k+7)}{(2k+2)(2k+1)} \left\{ m_{2k+2} - \frac{2(4k+9)}{(2k+4)(2k+3)} m_{2k+4} \right\}. \tag{2.20}$$

It is also needed that $K_3(1) = 0$ which by (2.17) requires the knowledge of $P_{2k}^{(1)}(1)$ which is evaluated below by again using the Liebniz rule of differentiation with the Roderiguéz formula.

$$\begin{aligned}
\mathcal{D}P_{2k}(1) &= \frac{1}{2^{2k}(2k)!}\mathcal{D}^{2k+1}\left\{(x+1)^{2k}(x-1)^{2k}\right\}\Big|_{x=1} \\
&= \frac{1}{2^{2k}(2k)!}\sum_{j=1}^{2k+1}\binom{2k+1}{j}\mathcal{D}^{j}\left\{(x+1)^{2k}\mathcal{D}^{2k+1-j}(x-1)^{2k}\right\}\Big|_{x=1} \\
&= \frac{1}{2^{2k}(2k)!}\sum_{j=1}^{2k+1}\binom{2k+1}{j}\frac{(2k)!}{(2k-i)!}(x+1)^{2k-j}\frac{(2k)!}{(j-1)!}(x-1)^{j-1}\Big|_{x=1} \\
&= \frac{2k(2k+1)}{2}.
\end{aligned}$$

Using the above with (2.17) and $K_3(1) = 0$ yields another linear combination of the three m_{2k}, m_{2k+2} and m_{2k+4} coefficients. Evaluating the m_{2k} coefficient,

$$\begin{aligned}
&\frac{(4k+1)!}{2^{2k+1}((2k)!)^2}P_{2k}^{(1)}(1) - \frac{(4k+2)!(4k+5)}{2^{2k+3}(2k)!(2k+1)!}P_{2k+1}^{(1)}(1) + \frac{(4k+4)!(4k+9)}{2^{2k+6}(2k)!(2k+2)!}P_{2k+4}^{(1)}(1) \\
&\qquad = \frac{(4k+1)!(4k+3)(4k+5)(2k^2+9k+14)}{2^{2k+4}((2k)!)^2}.
\end{aligned} \tag{2.21}$$

Evaluating the coefficient for m_{2k+2},

$$\begin{aligned}
&\frac{(4k+5)!}{2^{2k+3}((2k+2)!)^2}P_{2k+2}^{(1)}(1) - \frac{(4k+9)(4k+6)!}{2^{2k+5}(2k+3)!(2k+2)!}P_{2k+4}^{(1)}(1) \\
&\qquad = -\frac{(4k+5)!(4k+7)(2k^2+9k+12)}{2^{2k+4}((2k+2)!)^2}.
\end{aligned} \tag{2.22}$$

The coefficient of the last term m_{2k+4} is $(4k+9)!(2k+4)(2k+5)/(2^{2k+6}((2k+4)!)^2)$. Using (2.22), (2.21) and the requirement that $K_3(1) = 0$ implies, after simplification, that

$$m_{2k} = \frac{4(4k+7)}{(2k+1)(2k+2)}\left\{\frac{2k^2+9k+12}{2k^2+9k+14}m_{2k+2} - \frac{(2k+5)(4k+9)}{(2k+3)(2k^2+9k+14)}m_{2k+4}\right\}. \quad (2.23)$$

Assume without loss of generality that m_{2k} is assumed to be known. Then (2.20) and (2.23) can be used to determine m_{2k+2} and m_{2k+4} in terms of m_{2k}. Indeed,

$$\begin{bmatrix} 1 & -\frac{2(4k+9)}{(2k+4)(2k+3)} \\ \frac{2k^2+9k+12}{2k^2+9k+14} & -\frac{(2k+5)(4k+9)}{(2k+3)(2k^2+9k+14)} \end{bmatrix}\begin{bmatrix} m_{2k+2} \\ m_{2k+4}\end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}\frac{(2k+1)(2k+2)}{4(4k+7)}m_{2k}$$

which after inversion yields,

$$\begin{bmatrix} m_{2k+2} \\ m_{2k+4}\end{bmatrix} = m_{2k}\frac{(2k+1)(2k+2)}{2(4k+7)}\begin{bmatrix} 1 \\ \frac{(2k+1)(2k+2)}{4(4k+9)}\end{bmatrix}. \quad (2.24)$$

Substituting (2.24) into the expansion given in (2.17) and simplifying,

$$K_2(x) = \frac{m_{2k}(4k-1)!}{2^{2k+1}(4k+7)((2k)!)^2}\left\{(4k+7)P_{2k}(x) - 2(4k+5)P_{2k+2}(x) + (4k+3)P_{2k+4}(x)\right\}. \quad (2.25)$$

Recall that $P_n(x)$ is a sum of even (or odd) powers of x if n is an even (or odd) number. Therefore, $P^{(1)}_{2n}(x)$ will be a sum of odd powers and therefore an odd function. Also, $P^{(2)}_{2n}(x)$ will be an even functions by the same reasoning. Using the above and (2.25), it follows that $K_3(x)$ is an odd function and consequently $K_3(0) = 0$. Note, as stated in the construction given in Cheng and Raimondo (2008), K_2 has a unique global minimum at zero if $m_{2k} = (-1)^{k+1}$. This in turn ensures that (K:5) and (K:6) hold. The kernel function K_2 was constructed using three even degree polynomials which guarantee that $K_2 : [-1,1] \longrightarrow \mathbb{R}$ is a polynomial of order $2k+4 = s+3$. This ensures that (K:1) and (K:2) are satisfied. Also, the boundary conditions (K:3) and vanishing moments (K:4) were explicitly built into the construction ensuring they are satisfied.

By substituting in the definition of P_n in terms of the polynomial power basis (see (2.11)) into (2.25), one can obtain the result (2.12) after simplification. This leads to the definition of the class of high order kernels that will be used throughout the thesis.

Definition 2.7. *Let $s, k \in \mathbb{Z}^+$ with $k = \lfloor s/2 \rfloor$. Then the polynomial function $K : [-1,1] \longrightarrow \mathbb{R}$ will be said to belong to the functional class $\mathscr{K}_s$ if,*

$$K(x) = K(k,x) = a_k \sum_{j=k-1}^{2k+2} b_{j,k} x^{2j-2k+2} \mathbb{1}_{[-1,1]}(x)\,, \quad (2.26)$$

where the polynomial coefficients are defined by

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

The class $\mathscr{K}_s$ satisfies all the required properties (K:1) – (K:6) allowing the ZCT to be described. The reader should note that the kernel construction assumed that $s = 2k+1$ for some $k \in \mathbb{Z}^+$, namely s is a positive odd number. The kernel functions defined by (2.26) still satisfy (K:1) – (K:6) regardless if $s = 2k$ or $s = 2k+1$ for some fixed $k \in \mathbb{Z}^+$.

2.2.1.1 *Use of the kernels for the Zero-crossing technique*

As proposed earlier, the estimation method for kink locations will be described in the fractional white noise model. In particular, it will focus on the probe functional described 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 $\mu \in \mathscr{C}_s^1([0,1],\theta)$, the kernel function $K \in \mathscr{K}_s$ and where $h = h(n)$ is the bandwidth that depends on n. This is the technique developed by Cheng and Raimondo (2008) which is based on the pioneering work of Goldenshluger et al. (2006). It is assumed at the very least that bandwidth satisfies, $h + \frac{1}{nh} \to 0$, as $n \to \infty$. These are standard regularity conditions for kernel smoothing techniques and additional conditions on the bandwidth will be stated as needed.

To the reader familiar with kernel smoothing, the probe functional $\kappa_h(t)$ bears a strong resemblance to a kernel smoothing estimate of the third derivative of μ (see for example Mack and Müller, 1989). The essential idea with using this probe functional is when t is 'close' to the change point at θ, the probe functional $\kappa_h(t)$ will have large extreme values of opposite sign when t is larger or smaller than θ respectively. Furthermore, due to the extrema of opposing signs and the fact that the smoothing approach will, by construction, yield a smooth continuous curve in $\kappa_h(t)$, there will be a zero crossing time between the two extrema near the kink location at θ. This idea is explored in more detail with the following argument. Start by integration by substitution and exploit the bounded domain of K_3 in (K:1),

$$\begin{aligned}\kappa_h(t) &= h^{-4} \int_0^1 K_3\left(\frac{x-t}{h}\right) \mu(x)\, dx \\ &= h^{-3} \int_{-t/h}^{(1-t)/h} K_3(u)\, \mu(t+hu)\, du.\end{aligned}$$

Notice that a two-sided kernel function is being used, that is, the domain of K includes both positive and negative values. This implies that values above and below the point of interest t are being used in the weighting scheme of K to produce an estimate. This is fine until t reaches the boundaries near $t = \pm 1$ at which point difficulties will arise since no observations will exist for $\mu(t)$ when $t > 1$ or $t < 0$. Therefore, the points near the boundaries need to be discarded to avoid edge bias. In particular, consider only $t \in (h, 1-h)$ and use repeated integration by parts with the vanishing boundaries of (K:3),

$$\begin{aligned}
\kappa_h(t) &= -h^{-2} \int_{-1}^{1} K_2\,(u)\, \mu^{(1)}(t+hu)\, du \\
&= \quad \vdots \\
&= -\int_{-1}^{1} K\,(u)\, \mu^{(3)}(t+hu)\, du \\
&= -\mu^{(3)}(t) + \mathcal{O}(h),
\end{aligned}$$

as long as $\int K(x)\, dx = 1$. However, this condition is not assumed to hold here. In fact, the properties of the class of kernel functions $K \in \mathscr{K}_s$ are exploited to obtain an estimate of θ in a different manner. Recall the re-expressed $\kappa_h(t)$,

$$\kappa_h(t) = -h^{-2} \int_{-1}^{1} K_2\,(x)\, \mu^{(1)}(t+hx)\, dx. \tag{2.27}$$

Define the interval

$$L_\theta := \{t\,|\,|\theta - t| < h\}\,, \tag{2.28}$$

and define $\tau := (\theta - t)/h$. Then $|\tau| < 1$ for all $t \in L_\theta$. Split (2.27) 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 the Taylor expansion inherent in $\mathscr{C}_s^1([0,1], \theta)$ define,

$$\begin{aligned}
J_h(t) := -h^{-2} \Bigg(&\int_{-1}^{\tau} K_2\,(x) \left(\mu^{(1)}(t+hx) - \mu^{(1)}(\theta^-)\right) dx \\
&+ \int_{\tau}^{1} K_2\,(x) \left(\mu^{(1)}(t+hx) - \mu^{(1)}(\theta^+)\right) dx \Bigg).
\end{aligned}$$

Then, using (K:4) with (2.7), (2.8) in combination with (2.6), gives the order bound,

$$J_h(t) = \mathcal{O}\left(h^{s-3}\right).$$

Therefore, this allows us to express $\kappa_h(t)$ in the following way,

$$\begin{aligned}
\kappa_h(t) &= -h^{-2} \int_{-1}^{\tau} K_2\,(x)\, \mu^{(1)}(\theta^-)\, dx - h^{-2} \int_{\tau}^{1} K_2\,(x)\, \mu^{(1)}(\theta^+)\, dx + J_h(t) \\
&= h^{-2} K_1(\tau)[\mu^{(1)}](\theta) + J_h(t) =: L_h(t) + J_h(t).
\end{aligned} \tag{2.29}$$

The so called localisation term $L_h(t)$ captures the magnitude of the kink around the value $K_1(\tau)$. It is also of order h^{-2} which is large for small h. Then the expansion in the above equation ensures that $\kappa_h(\cdot) = \mathcal{O}(h^{-2})$. More specifically we have the following,

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

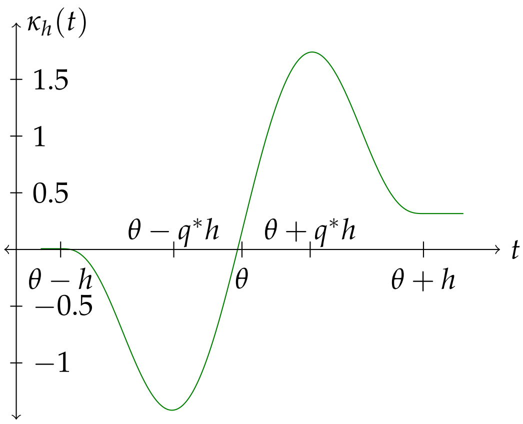
The above is more of a heuristic description of why $\kappa_h(t)$ can be used as an effective tool to locate the position of a kink. As seen in Goldenshluger et al. (2006) and Cheng and Raimondo (2008), a more technical result in the $\delta-$separation rate Lemma shown below gives a more rigorous, albeit subtle, explanation as to why the above representation is effective.

Lemma 2.1 (δ-separation rate). *Let $K \in \mathscr{K}_s$ and $\mu \in \mathscr{C}_s^1([0,1],\theta)$. In what follows the constant $0 < q < 1$ is given in (K:5). Let $h > 0$ and $\delta > 0$ satisfy $\delta < qh$. Let $A_{\delta,h} = \{t : \delta < |t-\theta| < qh\}$. Then for $\kappa_h(t) = \kappa_h(t,\mu)$:*

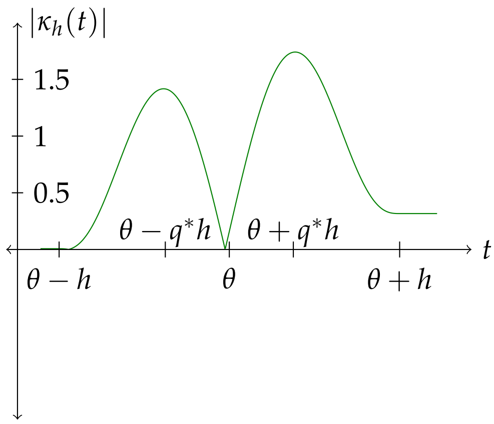
(a) $|\kappa_h(\theta)| \leq Ch^{s-3}$,

(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 $|\theta - t| > h$, $|\kappa_h(t)| \leq Ch^{s-3}$.



(a) Plot of the probe functional $\kappa_h(t)$.



(b) Plot of the probe functional $|\kappa_h(t)|$.

Figure 4: Demonstration of the $\delta-$separation methodology.

The specific behaviour of $\kappa_h(t)$ can be seen in Figure 4 where in this instance a function μ was chosen such that $\left[\mu^{(1)}\right](\theta) > 0$. In Figure 4 (a), it can be seen that there is a maximum and minimum located a distance q^*h above and below θ respectively. In Figure 4 (b), the zero crossing time or time when $|\kappa_h(t)|$ is near zero can be seen to occur very close to $t = \theta$. The discrepancy between the zero crossing time and θ is encapsulated by $\delta \in (h^s, h)$.

The proof of this Lemma is given in Cheng and Raimondo (2008). However, their proof requires a minor correction as the extra regularity condition 2. is needed in Definition 2.5 for the smoothness class $\mathscr{C}_s^1([0,1],\theta)$. The corrected proof is given below.

Proof of Lemma 2.1 . First re-express $\kappa_h(t)$ by exploiting the conditions of K. Recall from (2.27) that $\kappa_h(t)$ can be re-written as,

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

Proof of (a): Let $\tau = (\theta - t)/h$. Then $|\tau| < 1$ for all $t \in L_\theta$ (see (2.28)). We now split $\kappa_h(t)$ into two integrals,

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

Define the deterministic bias term,

$$\begin{aligned} J_h(t) := -h^{-2}\Bigg(&\int_{-1}^{\tau} K_2\,(x)\left(\mu^{(1)}(t+hx) - \mu^{(1)}(\theta^-)\right)\,dx \\ &+ \int_{\tau}^{1} K_2\,(x)\left(\mu^{(1)}(t+hx) - \mu^{(1)}(\theta^+)\right)\,dx\Bigg). \end{aligned}$$

Then $\kappa_h(t)$ can be expressed in the following way,

$$\begin{aligned} \kappa_h(t) &= -h^{-2}\int_{-1}^{\tau} K_2\,(x)\,\mu^{(1)}(\theta^-)\,dx - h^{-2}\int_{\tau}^{1} K_2\,(x)\,\mu^{(1)}(\theta^+)\,dx + J_h(t) \\ &= -h^{-2}\mu^{(1)}(\theta^-)K_1\,(x)\,\Big|_{x=-1}^{\tau} - h^{-2}\mu^{(1)}(\theta^+)K_1\,(x)\,\Big|_{x=\tau}^{1} + J_h(t) \\ &= h^{-2}K_1(\tau)[\mu^{(1)}](\theta) + J_h(t). \end{aligned} \tag{2.30}$$

Noting that $\tau = 0$ when $t = \theta$ so $K_1(\tau) = K_1(0) = 0$ by (K:3). Thus, (2.30) implies that $\kappa_h(\theta) = J_h(\theta)$ and we focus on $J_h(t)$ for our upper bound. Using $\tau = h^{-1}(\theta - t)$ we write,

$$\begin{aligned} J_h(t) = -h^{-2}\Bigg\{&\int_{-1}^{\tau} K_2\,(x)\left(\mu^{(1)}(\theta - h(\tau - x)) - \mu^{(1)}(\theta^-)\right)\,dx \\ &+ \int_{\tau}^{1} K_2\,(x)\left(\mu^{(1)}(\theta + h(x-\tau)) - \mu^{(1)}(\theta^+)\right)\,dx\Bigg\}. \end{aligned} \tag{2.31}$$

First, we use (2.6) and (2.7) to obtain,

$$\begin{aligned}
J_h(t) &= -h^{-2}\left\{\sum_{i=1}^{s-2}\frac{h^i\mu^{(i+1)}(\theta_-)}{i!}\int_{-1}^{\tau}(x-\tau)^iK_2(x)\,dx\right.\\
&\qquad\left.\sum_{i=1}^{s-2}\frac{h^i\mu^{(i+1)}(\theta_+)}{i!}\int_{\tau}^{1}(x-\tau)^iK_2(x)\,dx+\mathcal{O}(h^{s-1})\right\}\\
&= -h^{-2}\left\{\sum_{i=1}^{s-2}\sum_{j=0}^{i}\binom{i}{j}\frac{h^i(-\tau)^{i-j}\mu^{(i+1)}(\theta_\pm)}{i!}\int_{-1}^{1}x^iK_2(x)\,dx+\mathcal{O}(h^{s-1})\right\}\\
&= \mathcal{O}(h^{s-3}).
\end{aligned}\tag{2.32}$$

This finishes the proof of *(a)*.

Proof of (b): If $t \in A_{\delta,h} = \{t : \delta < |\theta - t| < qh\}$ then $|\tau| \in (\delta/h, q)$. Then recall from (2.30) that,

$$\kappa_h(t) = L_h(t) + J_h(t).$$

The first term of (2.30) satisfies,

$$|h^{-2}K_1(\tau)[\mu^{(1)}](\theta)| \geq h^{-2}|\tau||a| \geq \delta h^{-3}|a|. \tag{2.33}$$

From *(a)* it is known that for all $t \in L_\theta$, $|J_h(t)| \leq Ch^{s-3}$ then we can apply the reverse triangle inequality to obtain the desired result. Indeed,

$$\begin{aligned}
|\kappa_h(t)| &\geq \left||h^{-2}K_1(\tau)[\mu^{(1)}](\theta)| - |J_h(t)|\right| && \text{by (2.30)}\\
&\geq \left|\delta h^{-3}|a| - c_3h^{s-3}\right| && \text{by (2.33) and (2.32)}\\
&\geq C\delta h^{-3},
\end{aligned}$$

as long as $\delta \geq Ch^s$.

Proof of (c): Using (K:4) and the re-expressed $\kappa_h(t)$ (see *Proof of (a)*), we have the equivalent expression,

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

Since we are considering $\{t : |\theta - t| > q^*h\}$, t and $t + hx$ are well separated from the change point θ such that if $t \leq \theta$ then $t \pm h \leq \theta$ and if $t \geq \theta$ then $t \pm h \geq \theta$. This allows us to be able to apply (2.7) directly,

$$\begin{aligned}\kappa_h(t) &= -h^{-2}\int_{-1}^{1} K_2(x)\left(\mu^{(1)}(t+hx)-\mu^{(1)}(t)\right)\,dx\\ &= -h^{-2}\left\{\sum_{i=1}^{s-2}\frac{h^i\mu^{(i+1)}(\theta_\pm)}{i!}\int_{-1}^{1}x^iK_2(x)\,dx+\mathcal{O}(h^{s-1})\right\}\\ &= \mathcal{O}(h^{s-3})\end{aligned}$$

which completes the proof. □

2.2.2 *Estimation procedure*

The preliminaries have been covered and in this section the estimation procedure is covered that includes both a kink detection procedure and an estimation of the change point. Without loss of generality, to ease the presentation, when $\mu \in \mathscr{C}_s^1([0,1],\theta)$ also assume that $[\mu^{(1)}](\theta) > 0$, similar arguments will apply for the scenario when $\mu \in \mathscr{C}_s^1([0,1],\theta)$ and $[\mu^{(1)}](\theta) < 0$. As indicated, the estimation procedure will be considered under the fractional white noise. The functional that will be used for estimation is,

$$\widehat{\kappa}_h(t) = \int_0^1 K_3\left(\frac{x-t}{h}\right)\,dY(x).$$

Using stochastic integration we see that,

$$\begin{aligned}\widehat{\kappa}_h(t) &= h^{-4}\int_0^1 K_3\left(\frac{x-t}{h}\right)\mu(x)\,dx+\varepsilon_n^{\alpha}h^{-4}\int_0^1 K_3\left(\frac{x-t}{h}\right)\,dB_H(x)\\ &= L_h(t)+J_h(t)+Z_h(t),\end{aligned}$$

where $L_h(t)$, $J_h(t)$ are the aforementioned *localisation* and *deterministic bias* terms defined by (2.29) in Section 2.2.1.1. The stochastic error term is

$$Z_h(t) := h^{-4}\varepsilon_n^{\alpha}\int_0^1 K_3\left(\frac{x-t}{h}\right)\,dB_H(x).$$

The estimation technique is divided into three main steps:

I) The *localisation step*, where the interval $(h, 1-h)$ is truncated into a shorter interval that is suspected to contain θ;

II) The *kink detection step*, whereby it is determined if suspicions of a present kink are genuine;

III) The *zero-crossing step* which uses the ZCT to hone in a finer estimate of θ.

To illustrate the technique, each of the procedural steps are demonstrated a set of observations $\{(x_i, y_i)\}_{i=1}^n$ that satisfy the sampling model (2.1). It has the specific regression function $\mu \in \mathscr{C}^1_{s=3}([0,1], \theta = 2/3)$ introduced in (2.9) (and shown in Figure 2) with LRD noise generated in the software `R` (see R Development Core Team, 2010) with the R-package `fracdiff` (see Fraley, Leisch, Maechler, Reisen and Lemonte, 2009). Start with the *localisation step* below.

Localisation step

Recall that it was also shown by (2.29) that the probe functional $\kappa_h(t)$ has the localisation term,

$$L_h(t) = \begin{cases} h^{-2}K_1\left(\frac{\theta - t}{h}\right)[\mu^{(1)}](\theta), & \text{if } \mu \in \mathscr{C}^1_s([0,1],\theta) \text{ and } t \in L_\theta = \{t \mid |\theta - t| < h\} \\ 0, & \text{if } \mu \in \mathscr{S}_s \text{ or } \left\{\mu \in \mathscr{C}^1_s([0,1],\theta) \text{ and } t \notin L_\theta\right\}. \end{cases} \tag{2.34}$$

By (K:6), it is known that the kernel function K_1 has two unique extrema in the form of a unique global minimum at $q^* \in (0,1)$ and unique global maximum at $-q^*$. In light of (2.34), $L_h(t)$ has the same unique extrema with a unique global minimum at the point

$$t_* := \arg\min_{t\in(h,1-h)} L_h(t) = \theta - q^* h$$

and a unique global maximum at the point

$$t^* := \arg\max_{t\in(h,1-h)} L_h(t) = \theta + q^* h.$$

It is required that the localisation term, $L_h(t)$ dominates the deterministic bias term $J_h(t) = \mathcal{O}\left(h^{s-3}\right)$. Clearly, when $|\theta - t| < h$, $L_h(t) \geq Ch^{-2} > Ch^{s-3}$ and $J_h(t) = \mathcal{O}(h^{s-3})$, so $L_h(t)$ dominates the deterministic signal of $\kappa_h(t)$. We now need to investigate the distribution of the noise process, $Z_h(t)$. This is needed to ensure that $Z_h(t)$ isn't too large and does not interfere with the signal generated by $L_h(t)$. Consider the probabilistic bound given in the Lemma below.

Lemma 2.2. *Let $t \in (h, 1-h)$, $K_3 \in \mathscr{L}_1([-1,1],\mathbb{R}) \cup \mathscr{L}_2([-1,1],\mathbb{R})$, then $Z_h(t)$ is a Gaussian process with zero mean and variance,*

$$\sigma_Z^2 := \mathbb{V}ar\, Z_h(t) = h^{-(\alpha+6)}\varepsilon^{2\alpha}\tau_1^2,$$

where the constant τ_1^2 is given by,

$$\tau_1^2 := (2-\alpha)(1-\alpha)/2 \int_{-1}^{1}\int_{-1}^{1} K_3(x)K_3(y)|x-y|^{-\alpha} dx dy.$$

Proof of Lemma 2.2 . Firstly define, $V_\alpha := (2-\alpha)(1-\alpha)/2$ then a result of Gripenberg and Norros (1996) gives a measure of the variability of a stochastic integral with respect to fBm.

Proposition 2.2 (Gripenberg and Norros, 1996). *If $f, g \in \mathscr{L}_1(x, \mathbb{R}) \cap \mathscr{L}_2(x, \mathbb{R})$, $H \in (\frac{1}{2}, 1)$ and B_H is a fBm then,*

$$\langle f, g\rangle_{\mathcal{H}} := \mathbb{E}\left\{\int_x f(x)dB_H(x)\int_x g(y)dB_H(y)\right\} = V_\alpha \int_x\int_x f(x)g(y)|x-y|^{-\alpha}dxdy.$$

From the previous section it is known that K_3 is a finite degree polynomial constructed via three Legendre polynomials. Therefore, $K_3 \in \mathscr{L}^2([-1,1],\mathbb{R}) \cap \mathscr{L}^1([-1,1],\mathbb{R})$ so apply Proposition 2.2 (Gripenberg and Norros, 1996),

$$\begin{aligned}\mathbb{E}Z_h^2(t) &= h^{-8}\varepsilon^{2\alpha}\mathbb{E}\left\{\int_0^1 K_3\left(\frac{x-t}{h}\right)dB_H(x)\right\}^2\\ &= h^{-8}\varepsilon^{2\alpha}V_\alpha\int_0^1\int_0^1 K_3\left(\frac{x-t}{h}\right)K_3\left(\frac{y-t}{h}\right)|x-y|^{-\alpha}\,dx\,dy.\end{aligned}$$

By changing variable and recalling that $t \in (h, 1-h)$ we obtain the result:

$$\sigma_Z^2 = \mathbb{E}Z_h^2(t) = h^{-(\alpha+6)}\varepsilon^{2\alpha}V_\alpha\int_{-1}^1\int_{-1}^1 K_3(x)K_3(y)|x-y|^{-\alpha}\,dx\,dy = h^{-(\alpha+6)}\varepsilon^{2\alpha}\tau_1^2. \quad \square$$

Using the probabilistic bound given by Lemma 2.2 it follows that,

$$\widehat{\kappa}_h(t) = L_h(t) + J_h(t) + \mathcal{O}_p\left(n^{-\alpha/2}h^{-\alpha/2-3}\right).$$

To ensure that $L_h(t)$ dominates $Z_h(t)$ as well $J_h(t)$ it is required that $h^{-2} > Cn^{-\alpha/2}h^{-\alpha/2-3}$ or equivalently $h \geq Cn^{-\alpha/(\alpha+2)}$. This is guaranteed to hold if h is chosen such that,

$$h \asymp n^{-\alpha/(\alpha+2)+\eta}, \text{ for some } \eta > 0. \tag{2.35}$$

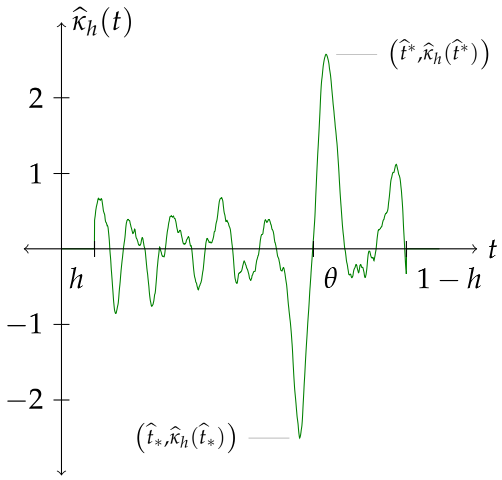
If the bandwidth is chosen such that (2.35) holds, then the points t_* and t^* can be estimated with,

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

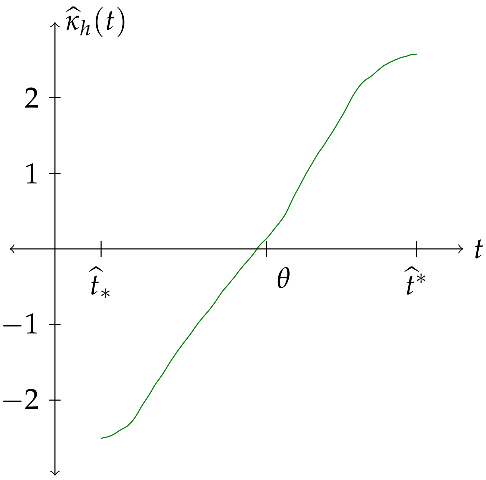
By this construction, the interval $\widehat{A}_h = (\widehat{t}_*, \widehat{t}^*)$ has a length which is order h and contains θ with high probability. The estimator $\widehat{\kappa}_h(t)$ for this choice of μ with a simulated realisation of LRD errors is shown in Figure 5. Notice the extrema points used for the truncation to the interval $\widehat{A}_h$.

Kink detection step

To ensure that the signal generated by $\kappa_h(t)$ was actually influenced by $L_h(t)$ and a



(a) Initial probe functional with extrema near θ.



(b) Truncation of the interval after identifying extrema.

Figure 5: Plot of a the noisy probe functional $\widehat{\kappa}_h(t)$ where μ has a kink at $\theta = 2/3$.

genuine kink is present, the procedure needs to be able to distinguish between $L_h(t)$ and any large deviations generated by the noise process $Z_h(t)$. Begin by standardising the statistic $\widehat{\kappa}_h(t)$ with the standard deviation of the noise $Z_h(t)$. Define this standardised statistic as,

$$\mathcal{T}_{\widehat{\kappa}}(t) := \frac{\varepsilon_n^{-\alpha}}{\tau_1} n^{\alpha/2} h^{\alpha/2+3} \widehat{\kappa}_h(t) = \frac{\varepsilon_n^{-\alpha}}{\tau_1} n^{\alpha/2} h^{\alpha/2+3} \left(L_h(t) + J_h(t) + Z_h(t) \right). \tag{2.36}$$

The extrema of $L_h(t)$ and $J_h(t)$ have already been considered in previous sections. A large deviations result needed for the process $Z_h(t)$ and is given in the Lemma below.

Lemma 2.3. *Let $I \subseteq (h, 1-h)$, then for all $\lambda \geq 2\sigma_Z = \mathbb{V}\mathrm{ar} Z_h(t)$ we have,*

$$P\left(\sup_{t \in I} |Z_h(t)| \geq \lambda \right) \leq C h^{\frac{\alpha+4}{2}} \varepsilon^{-\alpha} |I| \lambda \exp\left(-\frac{\lambda^2 h^{\alpha+6}}{2\tau_1^2 \varepsilon^{2\alpha}} \right)$$

where $C > 0$ is a constant and $|\cdot|$ denotes the Lebesgue measure.

Proof of Lemma 2.3 . Modifying the equivalent result from Goldenshluger et al. (2006), the proof requires a result from van der Vaart and Wellner (1996). Some auxiliary definitions are required. Let $r \in \mathbb{R}^+$, $B \subset [0,1]$ and $\rho : \mathbb{R}^+ \times \mathbb{R}^+ \longrightarrow \mathbb{R}^+$ be a semi-metric. Then define $N(r, B, \rho)$ to be the covering number of B with respect to the semi-metric ρ. That is, $N(r, B, \rho)$ is the number of balls of radius r in the semi-norm ρ that cover B. Also define $\varsigma^2(X) := \sup_{t \in B} \mathbb{V}\mathrm{ar} X_t$. The needed result is (van der Vaart and Wellner, 1996, Proposition A.2.7) and is stated below.

Proposition A.2.7 (Van der Vaart and Wellner, 1996). *Let X be a separable, mean-zero Gaussian process such that for some $K > \varsigma(X)$, some $V > 0$, and some $0 < \varepsilon_0 \leq \varsigma(X)$,*

$$N(\varepsilon, B, \rho) \leq \left(\frac{K}{\varepsilon}\right)^V, \qquad 0 < \varepsilon \leq \varepsilon_0,$$

then there exists a universal constant D such that, for all $\lambda \geq \varsigma^2(X)(1+\sqrt{V})/\varepsilon_0$,

$$P\left(\sup_{t\in B} X_t \geq \lambda\right) \leq \left(\frac{DK\lambda}{\sqrt{V}\varsigma^2(X)}\right)^V \overline{\Phi}\left(\frac{\lambda}{\varsigma(X)}\right),$$

where $\overline{\Phi}(x) = P(Z \geq x)$ and $Z \sim \mathcal{N}(0,1)$.

Consider the semi-metric σ^2 which gives a distance measure between two stochastic processes,

$$\begin{aligned}
\sigma^2(t,s) &:= \mathbb{E}|Z_h(t) - Z_h(s)|^2 \\
&= \varepsilon^{2\alpha} h^{-8} \mathbb{E}\left|\int_0^1 K_3\left(\frac{x-t}{h}\right) - K_3\left(\frac{x-s}{h}\right) dB_H(x)\right|^2 \\
&= \varepsilon^{2\alpha} h^{-8} V_\alpha \int_0^1 \int_0^1 \left\{K_3\left(\frac{x-t}{h}\right) - K_3\left(\frac{x-s}{h}\right)\right\} \\
&\qquad \left\{K_3\left(\frac{y-t}{h}\right) - K_3\left(\frac{y-s}{h}\right)\right\} |x-y|^{-\alpha} dxdy \\
&= \varepsilon^{2\alpha} h^{-(\alpha+6)} V_\alpha \int_{-\frac{t}{h}}^{\frac{1-t}{h}} \int_{-\frac{t}{h}}^{\frac{1-t}{h}} \left\{K_3(u) - K_3\left(u + \frac{t-s}{h}\right)\right\} \\
&\qquad \left\{K_3(v) - K_3\left(v + \frac{t-s}{h}\right)\right\} |u-v|^{-\alpha} dudv. \qquad (2.37)
\end{aligned}$$

Without loss of generality let $0 < h < s < t < 1-h$. First consider the case when $t - s < 2h$. The domain of integration $D = (-t/h, (1-t)/h)$ needs to be considered with respect to the support of the integrand $K_3(x) - K_3(x + (t-s)/h)$ which will be defined by the set $S = [-1 - (t-s)/h, 1]$. Due to the restriction that $0 < h < s < t$, $S \subset D$, the set S can be split into three disjoint intervals, $S = [-1-(t-s)/h, -1) \cup [-1, 1-(t-s)/h] \cup (1-(t-s)/h, 1] =: S_1 \cup S_2 \cup S_3$. Then apply the Lipschitz condition (K:1) over each of the domains S_i. Start with S_1 which is outside the domain of $K_3(x)$. Thus,

$$\begin{aligned}
\left|K_3(x) - K_3\left(x + \frac{t-s}{h}\right)\right| \mathbb{1}_{[-1-\frac{t-s}{h}, -1)}(x) &= \left|K_3\left(x + \frac{t-s}{h}\right) - K_3(-1)\right| \mathbb{1}_{[-1-\frac{t-s}{h}, -1)}(x) \\
&\leq L_{K_3} \left|x + \frac{t-s}{h} + 1\right| \mathbb{1}_{[-1-\frac{t-s}{h}, -1)}(x) \\
&\leq L_{K_3} h^{-1} |t-s|. \qquad (2.38)
\end{aligned}$$

Similarly, the set S_3 is outside the domain of the shifted version of K_3 so,

$$\left|K_3(x) - K_3\left(x + \frac{t-s}{h}\right)\right| \mathbb{1}_{[1-\frac{t-s}{h},1)}(x) \le L_{K_3}|x-1| \mathbb{1}_{[1-\frac{t-s}{h},1)}(x) \le L_{K_3}h^{-1}|t-s|. \tag{2.39}$$

The set S_2 covers both the domains of $K_3(x)$ and the shifted version so by a direct application of the Lipschitz inequality,

$$\left|K_3(x) - K_3\left(x + \frac{t-s}{h}\right)\right| \mathbb{1}_{[-1,1-\frac{t-s}{h})}(x) \le L_{K_3}h^{-1}|t-s|. \tag{2.40}$$

Thus, all three equations, (2.38), (2.40) and (2.39) imply that for $t-s < 2h$,

$$\left|K_3(x) - K_3\left(x + \frac{t-s}{h}\right)\right| \mathbb{1}_S(x) \le L_{K_3}h^{-1}|t-s|. \tag{2.41}$$

Consequently, using (2.41) inside (2.37), it follows that for $t-s < 2h$,

$$\sigma^2(t,s) \le \varepsilon^{2\alpha}h^{-(\alpha+8)}V_\alpha L_{K_3}^2|t-s|^2\int_{-1}^{1}\int_{-1}^{1}|u-v|^{-\alpha}\,dudv = \mathcal{O}\left(\varepsilon^{2\alpha}h^{-(\alpha+8)}|t-s|^2\right). \tag{2.42}$$

On the other hand, consider the case when $t-s > 2h$. The support of the integrand $K_3(x) - K_3(x+(t-s)/h)$ is over two disjoint intervals $S = [-1-(t-s)/h, 1-(t-s)/h] \cup [-1,1] := S_1 \cup S_2$ where $S_1 \cap S_2 = \emptyset$. To ease the presentation define $I_{K_3}(x) := K_3(x) - K_3(x+(t-s)/h)$. Then, over the set S_1, only the shifted version is non-zero, namely,

$$I_{K_3}(x)\,\mathbb{1}_{S_1}(x) = -K_3(x+(t-s)/h)\,\mathbb{1}_{S_1}(x).$$

Conversely, $I_{K_3}(x)\,\mathbb{1}_{S_2}(x) = K_3(x)\mathbb{1}_{S_2}(x)$. Using a slight abuse of notation, the double integral needs to be considered over all possible combinations of the values $u, v \in S_1$ or S_2,

$$\begin{aligned}\int_{-\frac{t}{h}}^{\frac{1-t}{h}}\int_{-\frac{t}{h}}^{\frac{1-t}{h}} I_{K_3}(u)\,I_{K_3}(v)\,|u-v|^{-\alpha}\,du\,dv \\ = \int_{S_1}\int_{S_2} + \int_{S_1}\int_{S_1} + \int_{S_2}\int_{S_2} + \int_{S_2}\int_{S_1} I_{K_3}(u)\,I_{K_3}(v)\,|u-v|^{-\alpha}\,du\,dv.\end{aligned} \tag{2.43}$$

Starting with the first term of (2.43),

$$\int_{S_1}\int_{S_2} I_{K_3}(u)\, I_{K_3}(v)\, |u-v|^{-\alpha} du dv = \int_{-1-\frac{t-s}{h}}^{1-\frac{t-s}{h}} \int_{-1}^{1} K_3\left(u+\frac{t-s}{h}\right) K_3(v)\, |u-v|^{-\alpha}\, du\, dv$$
$$= \int_{-1}^{1}\int_{-1}^{1} K_3(x)\, K_3(v) \left|x-v-\frac{t-s}{h}\right|^{-\alpha} dx\, dv. \tag{2.44}$$

Now it was assumed that $t-s>2h$, meaning that over the set $U=\{(x,v): x,v\in[-1,1]\}$, there is no discontinuity in $|x-v-(t-s)/h|^{-\alpha}$ so $|x-v-(t-s)/h|^{-\alpha} = ((t-s)/h+v-x)^{-\alpha}$. Consider a Taylor expansion of the function $x^{-\alpha}$ where $x+y>0$ with $x,y\neq 0$ and $\alpha\in(0,1]$,

$$(x+y)^{-\alpha} = x^{-\alpha} - \alpha y x^{-(1+\alpha)} + \frac{\alpha(1+\alpha)}{2} y^2 (x+\xi y)^{-(2+\alpha)},$$

where $\xi=\xi(x,y)\in(0,1)$. Apply this Taylor expansion in a piecewise manner to the function $((t-s)/h+v-x)^{-\alpha}$. So for $(x,v)\in U$ with $x\neq v$, there exists a $\xi\in(0,1)$ such that,

$$\left(\frac{t-s}{h}+v-x\right)^{-\alpha} = |x-v|^{-\alpha} - \alpha h^{-1}(t-s)\,|x-v|^{-(1+\alpha)} + \frac{\alpha(1+\alpha)}{2} h^{-2}(t-s)^2 \left|x-v+\xi\frac{t-s}{h}\right|^{-(2+\alpha)}.$$

Using this expansion in (2.44),

$$\begin{aligned}
&V_\alpha \int_{-1}^{1}\int_{-1}^{1} K_3(x)\, K_3(v) \left|x-v-\frac{t-s}{h}\right|^{-\alpha} dx\, dv \\
&\quad= V_\alpha \int_{-1}^{1}\int_{-1}^{1} K_3(x)\, K_3(v) \left\{ |x-v|^{-\alpha} - \alpha h^{-1}(t-s)\,|x-v|^{-(1+\alpha)} \right. \\
&\qquad\qquad \left. + \frac{\alpha(1+\alpha)}{2} h^{-2}(t-s)^2 \left|x-v+\xi\frac{t-s}{h}\right|^{-(2+\alpha)} \right\} dx\, dv \\
&\quad= \tau_1^2 - \alpha h^{-1}(t-s) V_\alpha \int_{-1}^{1}\int_{-1}^{1} K_3(x)\, K_3(v)\, |x-v|^{-(1+\alpha)}\, dx\, dv \\
&\qquad + \frac{\alpha(1+\alpha)}{2} h^{-2}(t-s)^2 V_\alpha \int_{-1}^{1}\int_{-1}^{1} K_3(x)\, K_3(v) \left|x-v+\xi(x,v)\frac{t-s}{h}\right|^{-(2+\alpha)} dx\, dv. \\
&\quad= \mathcal{O}\left(h^{-2}(t-s)^2\right)
\end{aligned}$$

By symmetry it can also be shown that,

$$V_\alpha \int_{S_2}\int_{S_1} I_{K_3}(u)\, I_{K_3}(v)\, |u-v|^{-\alpha} du dv = \mathcal{O}\left(h^{-2}(t-s)^2\right).$$

The other two cases in (2.43) are over the same sets and are in fact equal. Indeed,

$$
\begin{aligned}
V_\alpha & \int_{S_1} \int_{S_1} I_{K_3}(u)\, I_{K_3}(v)\, |u-v|^{-\alpha} du dv \\
&= V_\alpha \int_{-1-\frac{t-s}{h}}^{1-\frac{t-s}{h}} \int_{-1-\frac{t-s}{h}}^{1-\frac{t-s}{h}} K_3\left(u + \frac{t-s}{h}\right) K_3\left(v + \frac{t-s}{h}\right) |u-v|^{-\alpha} du dv \\
&= V_\alpha \int_{-1}^{1} \int_{-1}^{1} K_3(x)\, K_3(y)\, |x-y|^{-\alpha} du dv \\
&= V_\alpha \int_{S_2} \int_{S_2} I_{K_3}(x)\, I_{K_3}(y)\, |x-y|^{-\alpha} du dv \\
&= \tau_1^2. \qquad (2.45)
\end{aligned}
$$

So by equations (2.43) – (2.45), it follows for $t-s > 2h$, $\sigma^2(t,s) = \mathcal{O}\left(\varepsilon^{2\alpha} h^{-(\alpha+8)} |t-s|^2\right)$. Then with (2.42), for general $t-s > 0$,

$$
\sigma^2(t,s) = \mathcal{O}\left(\varepsilon^{2\alpha} h^{-(\alpha+8)} |t-s|^2\right).
$$

Thus, the the number of balls of radius r using the semi-norm $\sigma(t,s)$ that cover the interval I does not exceed $Cr^{-1}h^{-(\alpha+8)/2}|I|$. That is,

$$
N(r, I, \sigma) \le r^{-1} \sup_{t,s} \sigma(t,s) \le Cr^{-1}\varepsilon^\alpha h^{-(\alpha+8)/2}|I|. \qquad (2.46)
$$

In the notation of Proposition A.2.7 (Van der Vaart and Wellner, 1996) (see page 44), set $K = C\varepsilon^\alpha h^{-(\alpha+8)/2}|I|$, $r = \varepsilon^\alpha$, $\varepsilon_0 = \sigma_Z$, $\varsigma(X) = \sigma_Z$, $V = 1$ and $X = |Z_h|$. Apply (2.46) and Proposition A.2.7 (Van der Vaart and Wellner, 1996) implies that for all $\lambda \ge 2\sigma_Z$, there exists constant $C > 0$ such that,

$$
P\left(\sup_{t \in I} |Z_h(t)| \ge \lambda\right) \le C\varepsilon^\alpha h^{-(\alpha+8)/2}|I| \frac{\lambda}{\sigma_Z^2} \overline{\Phi}\left(\frac{\lambda}{\sigma_Z}\right). \qquad (2.47)
$$

Using Lemma 2.2 with (2.47) there exists a constant $C > 0$ such that,

$$
\begin{aligned}
P\left(\sup_{t \in I} |Z_h(t)| \ge \lambda\right) &\le Ch^{(\alpha+4)/2}\varepsilon^{-\alpha}|I|\lambda\overline{\Phi}\left(\frac{\lambda}{\sigma_Z}\right) \\
&\le Ch^{(\alpha+4)/2}\varepsilon^{-\alpha}|I|\sigma_Z\phi\left(\frac{\lambda}{\sigma_Z}\right) \\
&\le Ch^{(\alpha+4)/2}\varepsilon^{-\alpha}|I|\lambda \exp\left(-\frac{\lambda^2 h^{\alpha+6}}{2\tau_1^2\varepsilon^{2\alpha}}\right).
\end{aligned}
$$

The second last line follows due to the inequality $\overline{\Phi}(z) < z^{-1}\phi(z)$ (see Feller, 1968, p175). The last line follows since $\lambda \ge 2\sigma_Z$. □

Consider the threshold at $\lambda = \lambda_\epsilon = \sigma_Z\sqrt{(2+\epsilon)\,|\log h|}$ for some $\epsilon > 0$. The bandwidth $h \to 0$ as $n \to \infty$, so for a sufficiently large n, we have $\lambda_\epsilon \geq 2\sigma_Z$. Apply Lemma 2.3 with this choice of λ_ϵ,

$$P\left(\sup_{t\in(h,1-h)} |Z_h(t)| \geq \sigma_Z\sqrt{(2+\epsilon)\,|\log h|}\right) \leq Ch^{-1}\sqrt{|\log h|}\exp\{-(1+\epsilon)\,|\log h|\}$$
$$= Ch^{\epsilon}\sqrt{|\log h|} = o(1). \tag{2.48}$$

Checking the large deviations of the estimator $\widehat{\kappa}_h(t)$, consider first the scenario when $\mu \in \mathscr{S}_s$, then there is no kink in μ and $L_h(t) \equiv 0$. Recall the standardised estimator of $\kappa_h(t)$ given in (2.36) and by (2.48) it follows that for an arbitrary choice of $\epsilon > 0$,

$$\lim_{n\to\infty} P\left(\sup_{t\in(h,1-h)} |\mathcal{T}_{\widehat{\kappa}}(t)| \geq \sqrt{(2+\epsilon)\,|\log h|}\right) = 0.$$

On the other hand, if $\mu \in \mathscr{C}^1_s([0,1],\theta)$, then for $t \in (t_*, t^*)$, $L_h(t) \geq Ch^{-2}$ and

$$\lim_{n\to\infty} P\left(\sup_{t\in(h,1-h)} |\mathcal{T}_{\widehat{\kappa}}(t)| \geq \sqrt{2\,|\log h|}\right) = 1.$$

That is, whenever $\mu \in \mathscr{S}_s$, $\sup_{t\in(h,1-h)} |\mathcal{T}_{\widehat{\kappa}}(t)|$ will diverge to infinity at a rate no faster than $\sqrt{2\,|\log h|}$. With this in mind, a kink is detected when the condition,

$$\sup_{t\in(h,1-h)} |\mathcal{T}_{\widehat{\kappa}}(t)| \geq \sqrt{2\,|\log h|}, \tag{2.49}$$

is observed in practice. For our example function the boundary is exceeded and a kink is flagged and shown in Figure 6. If a kink is detected through this procedure, the method can proceed to the final *zero-crossing step*.

Zero-crossing step

The idea behind the zero-crossing step is to use the aforementioned ZCT to refine the interval $\widehat{A}_n = (\widehat{t}_*, \widehat{t}^*)$ down to a singular point $\widehat{\theta}_n$, the estimate of θ. Indeed, by using Lemma 2.1, the zero-crossing time of $\kappa_h(t)$ within the interval $A_h = (t_*, t^*)$ can be located with an accuracy of order δ where $\delta < h$. The estimate of θ is constructed with,

$$\widehat{\theta}_n = \arg\min_{t\in\widehat{A}_n} |\widehat{\kappa}_h(t)| = \arg\min_{t\in\widehat{A}_n} |\mathcal{T}_{\widehat{\kappa}}(t)|\,. \tag{2.50}$$

This is demonstrated for noisy regression model in Figure 7. Recall again that $\widehat{\kappa}_h(t) = L_h(t) + J_h(t) + Z_h(t)$ and by comparing the bounds given in (2.32), (2.34) and Lemma 2.2

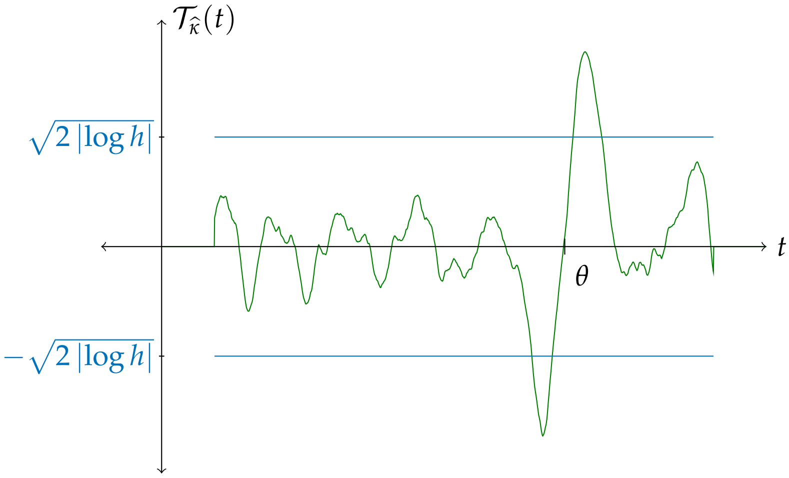


Figure 6: Illustration of the kink detection step.

with the bounds given in Lemma 2.1 implies that the minimum given in (2.50) is well defined if,

$$\delta h^{-3} \geq C h^{s-3} \quad \text{and} \quad \delta h^{-3} \geq C n^{-\alpha/2} h^{-\alpha/2-3}. \tag{2.51}$$

The best possible accuracy of the estimator given in (2.50) will be obtained if $\delta > 0$ is chosen as small as possible while still satisfying the inequalities in (2.51). The first inequality of (2.51) implies that $\delta \asymp h^s$ which when substituted into the second inequality of (2.51) implies that the best possible bandwidth is given by,

$$h_* \asymp n^{-\alpha/(2s+\alpha)}. \tag{2.52}$$

Now apply Lemma 2.1 with this optimal choice of bandwidth and $\delta = \delta_* \asymp h_*^s$ allows the method to obtain an accuracy of order $n^{-\alpha s/(2s+\alpha)}$. More specifically, this proves that the estimator satisfies the probabilistic bound,

$$\left|\widehat{\theta}_n - \theta\right| = \mathcal{O}_p\left(n^{-\alpha s/(2s+\alpha)}\right).$$

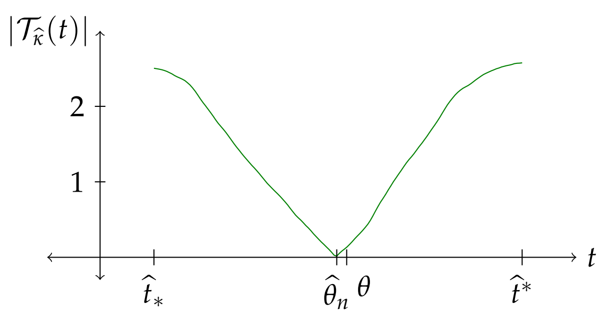


Figure 7: Illustration of the zero-crossing step when μ has a kink.

2.3 NUMERICAL PERFORMANCE

A numerical study is investigated here to confirm that the observed behaviour in practice agrees with the asymptotic theory that the level of dependence is detrimental to the rate of convergence. More specifically, it was shown in Section 2.2.2 that the rate of convergence for the estimator $\widehat{\theta}_n$ follows the rate,

$$\rho_n = n^{-\alpha s/(2s+\alpha)}.$$

The numerical analysis that follows in this subsection will reflect this behaviour via Monte-Carlo simulations. A discretely sampled regression model will be repeatedly simulated with a known regression function, μ, known level of noise, σ and known level of dependence in the error variables, α. The performance measure that will be used as a proxy for ρ_n will be ascertained by finding the bandwidth h that minimises the Root Mean-Square Error (RMSE) of the estimator $\widehat{\theta}_n$.

To start with, the specific structural assumptions for the simulated regression model is covered in Section 2.3.1, the performance measure of the RMSE along with the justification for the bandwidth as a proxy for the rate of convergence is covered in Section 2.3.2. Finally in Section 2.3.3 the numerical results are shown along with a brief discussion.

2.3.1 *Framework for the numerical study*

The numerical study that is performed in this Section is implemented in `R` (see R Development Core Team, 2010)). The sampling model is used where a discrete set of observations $\{(x_i, y_i)\}_{i=1}^n$ are generated using the grid $x_i = i/n$ where $y_i = \mu(x_i) + \sigma\varepsilon_i$. The same regression function that was considered in Section 2.2 is used here and is defined,

$$\mu(x) := \begin{cases} 4 - \frac{85}{9}x + \frac{97}{12}x^2 - x^3, & \text{if } 0 \le x < \frac{2}{3}; \\ \frac{82}{9} - \frac{508}{9}x + \frac{1159}{12}x^2 - \frac{181}{4}x^3 & \text{if } \frac{2}{3} < x \le 1; \\ 0 & \text{otherwise.} \end{cases} \tag{2.53}$$

The regression function has a clear kink at $\theta = 2/3$ (refer to Figure 2). Furthermore, it satisfies the additional smoothness and matching derivative assumptions required by the functional class $\mathscr{C}_3^1([0,1], 2/3)$.

The error variables are generated using the `R`-package `fracdiff` by Fraley, Leisch, Maechler, Reisen and Lemonte (2009). The `fracdiff` package generates observations from the fARIMA model where the innovations are assumed to be Gaussian. The level of dependence is controlled by $d = (1-\alpha)/2$ which is done for aesthetic reasons since it is more intuitive in the plots to have a dependence measure that is at zero for the case of independence and increases as the level of dependence increases.

2.3.2 *Performance measure*

It was shown in the theoretical exposition in Section 2.2 that there is a balance between choosing the bandwidth h such that the signal generated by $L_h(t) = \mathcal{O}\left(h^{-2}\right)$ dominates the stochastic error and deterministic bias terms, $Z_h(t)$ and $b_h(t)$, inside the functional $\kappa_h(t)$. More importantly, the stochastic term, $Z_h(t) = \mathcal{O}_p\left(n^{-\alpha/2}h^{-\alpha/2-3}\right)$ needs to be controlled. It was shown that the optimal choice of bandwidth to control the error and bias terms was given by $h_* \asymp n^{-\alpha/(2s+\alpha)}$ (see (2.52) see page 50). With this equation in mind, it is easy to see that as α decreases, the bandwidth h increases as is demonstrated in the graphic below.

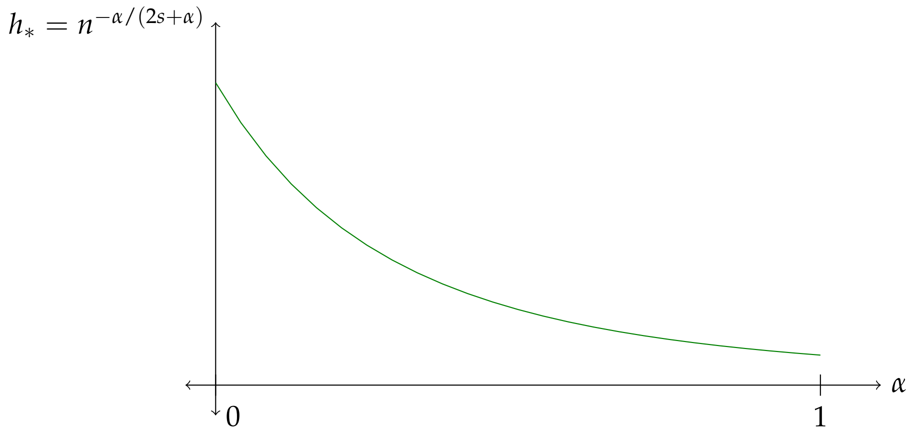


Figure 8: Theoretical optimal bandwidth h_* for a given level of n.

In our simulations we seek to find the best possible h for different levels of dependence and compare it with the expected theoretical result in Figure 8. This is done by using the kink estimation technique on repeated simulations of the same regression function given by (2.53) over different levels of dependent noisy errors. Then the best h is found using the 'oracle strategy'. This is done by using minimising the RMSE measure over the argument h with the following,

$$h^* = \arg\min_h \sqrt{\mathbb{E}\left|\widehat{\theta}_n - \theta\right|^2}.$$

The value h^* is the oracle bandwidth for the problem. The value h^* will be evaluated numerically using repeated Monte-Carlo iterations of the sample regression model and calculating the sample proxy of the RMSE which is the square root of the sample mean of the squared deviations from the known true kink location at $\theta = 2/3$.

To do this, first define the set of candidate bandwidths from the set $\mathcal{H}_n = \{0.15, \dots, 0.4\}$. These possible bandwidths are relatively large in comparison to the bandwidths used in usual kernel smoothing of regression functions. In the case of kernel smoothing estimation of regression functions, assuming LRD noise is present, then a typical kernel smoothing estimator of μ would be of the form,

$$\widehat{\mu}_n(t) = \frac{1}{nh} \sum_{i=1}^{n} Y_i K\left(\frac{x_i - t}{h}\right).$$

This estimator would have variance $\mathbb{V}\mathrm{ar}\left(\widehat{\mu}_n(t)\right) \asymp n^{-\alpha} h^{-\alpha}$. However, we are interested in the estimation of the smoothed third derivative of μ in $\widehat{\kappa}_h(t)$ and as such have an estimator of the form,

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

Consequently, this estimator satisfies, $\mathbb{V}\mathrm{ar}\left(\kappa_h(t)\right) \asymp n^{-\alpha} h^{-\alpha-6}$ and one can see that the variability has increased by a factor of h^{-6}. Thus, a larger bandwidth needs to be chosen to reduce this inflation factor to a reasonable level and obtain a reliable estimate of $\kappa_h(t)$.

The procedure for calculating the oracle bandwidth is as follows. First, fix a level of Monte-Carlo iterations M and define a sequence of k dependence levels, $\alpha_i \in \mathcal{A}$ where $1 \le i \le k$ and a sequence of ℓ candidate bandwidths, $h_j \in \mathcal{H}_\ell$ where $1 \le j \le \ell$. Then, at each level (α_i, h_j), generate M regression models each of which have n observations and obtain M corresponding kink location estimates $\left\{\widehat{\theta}_1, \widehat{\theta}_2, \dots, \widehat{\theta}_M\right\} =: \Theta_M$. The sample proxy for RMSE is then calculated for this choice of α_i and h_j via,

$$\widehat{\mathrm{RMSE}}_{\alpha_i, h_j} = \sqrt{\frac{1}{M} \sum_{\widehat{\theta}_i \in \Theta_M} \left(\widehat{\theta}_i - \theta\right)^2},$$

where θ is known to be $2/3$. Then the sample oracle h_i^* at dependence level α_i is the value that minimises the sample RMSE across all the candidate bandwidths $h_j \in \mathcal{H}_\ell$,

$$h_{i,n}^* = \underset{h_j \in \mathcal{H}_\ell}{\arg\min}\, \widehat{\mathrm{RMSE}}_{\alpha_i, h_j}.$$

The above procedure is repeated for all dependence levels $\alpha_i \in \mathcal{A}$ to gain a numerical equivalent of Figure 8, this is discussed further in Section 2.3.3 below.

2.3.3 *Numerical results*

The numerical results are displayed in Figure 9 and show the numerically evaluated oracle bandwidths, h_n^*, evaluated above across different sample sizes, n, and the number of Monte-Carlo iterations to numerically calculate the RMSE is chosen to be $M = 2048$.

Overall, it is evident that the sample oracle bandwidths are in agreement with the theoretical levels given by Figure 8 since in both Figure 9 (a) and (b) there is a clear increasing trend in the sample oracle bandwidth as the level of dependence increases (As the dependence increases, α decreases meaning d increases).

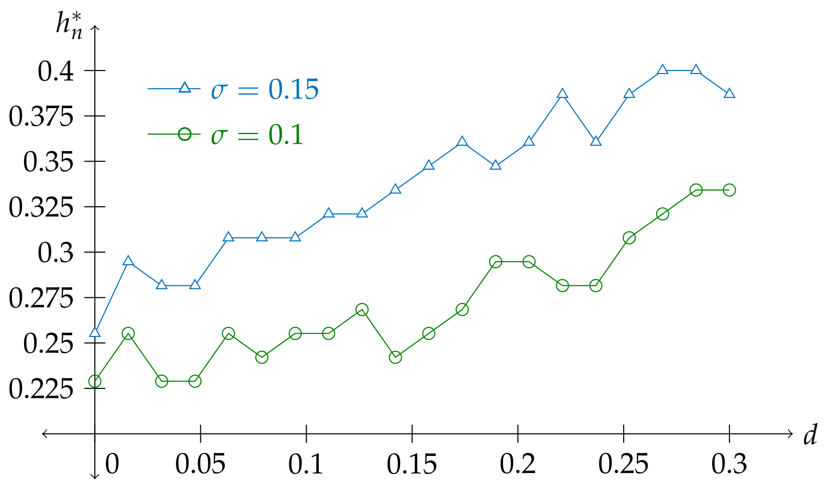
Moreover, as expected, with a larger σ coefficient amplifying the effects of the noise, a larger bandwidth is needed to reduce the effect of the noise. This is shown in both Figure 9 (a) and (b) since the higher noise level requires a larger oracle bandwidth at a given dependence level than a lower noise level. Also, one can see that estimation efficiency increases with n since larger noise levels are given in Figure 9 (b) (where $n = 1024$) yet the oracle bandwidths are very similar to those in Figure 9 (a) (where $n = 512$).

2.4 DISCUSSION

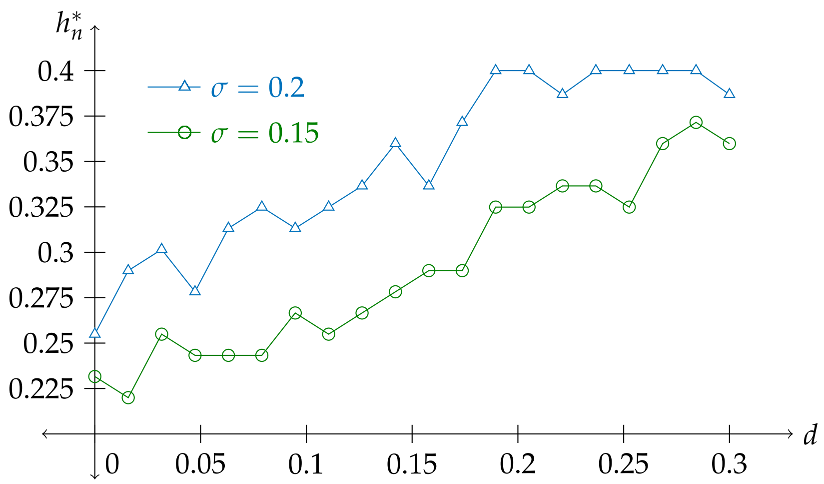
2.4.1 *Comparison for kink detection*

To demonstrate the estimation technique further, the analysis is conducted on two similar smooth functions, one with a kink and the other without a kink. In the same vein as the previous sections in this chapter, the example regression function with a kink, $\mu \in \mathscr{C}_3^1([0,1], \frac{2}{3})$ is used and a similar equivalent function without a kink is considered that belongs to the null functional class, $\mu \in \mathscr{S}_s$ (see Definition 2.6 see page 28 for details). In the smooth version of μ, the kink point has been removed and the function is completely smooth. This is demonstrated in Figure 10 where the analysis is conducted side-by-side for easy comparison.

In both instances, the regression models used the same sequence of error variables. The analysis that involves the function with a kink is given in Figure 10 (a) and (b). The analysis involving the completely smooth regression function without a kink is given in Figure 10 (c) and (d). Notice that, when there is no kink present, the detection bounds are satisfied and the method correctly doesn't identify a kink. Conversely, when a kink is present the detection bounds are violated and the technique proceeds to the zero crossing step discussed in the last part of Section 2.2.2 located see page 49.

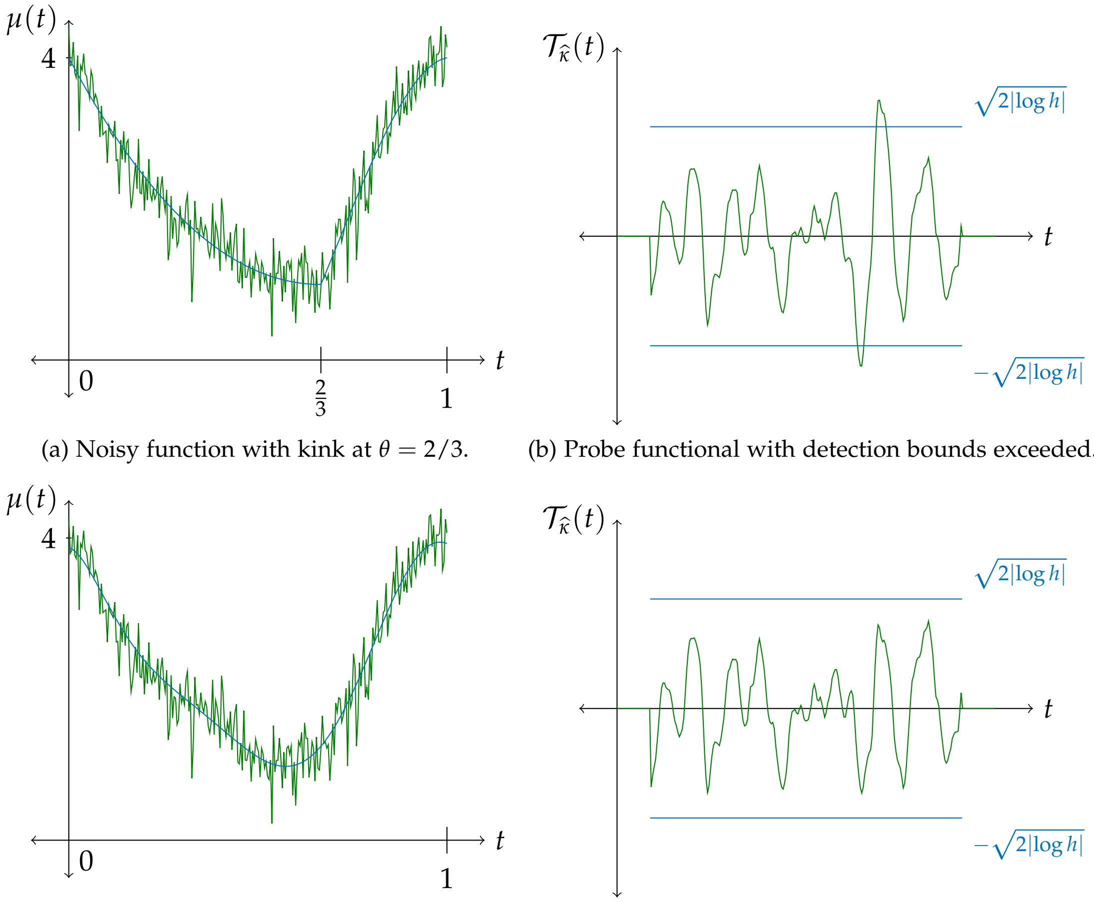


(a) Sample size $n = 512$ with $M = 2048$ Monte-Carlo iterations of RMSE.



(b) Sample size $n = 1024$ with $M = 2048$ Monte-Carlo iterations of RMSE.

Figure 9: Plots of the numerical oracle bandwidths against the dependence measure $d = \frac{1-\alpha}{2}$.



(a) Noisy function with kink at $\theta = 2/3$.

(b) Probe functional with detection bounds exceeded.

(c) Noisy function with no kink.

(d) Probe functional with detection bounds satisfied.

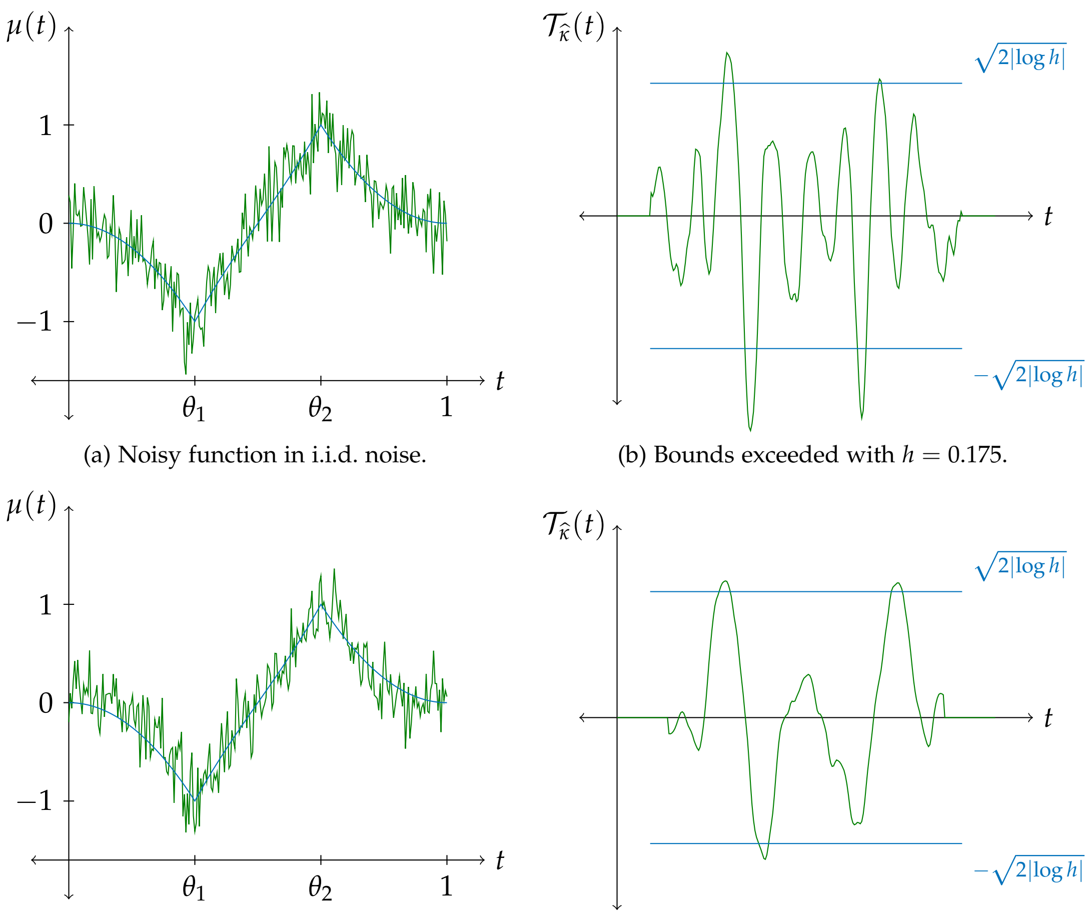
Figure 10: Comparison of the method when μ has a kink or is entirely smooth.

2.4.2 *Multiple change points*

The procedure can be extended to include the class of functions $\mu \in \mathscr{C}_s^m([0,1], \boldsymbol{\theta})$ by observing multiple instances where (2.49) is observed. For each instance of (2.49) there is a corresponding pair $\widehat{t}_*(i)$ and $\widehat{t}^*(i)$ that correspond to the localisation term $L_h(t)$ for the i^{th} kink with $i = 1, 2, \ldots, m$. With each pair, the interval $\widehat{A}_n$ is considered and the localisation and zero-crossing-time steps are executed on each of those intervals to produce an estimate for each kink location.

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, consider the toy example where θ_1 and θ_2 be two such change-points where $|\theta_1 - \theta_2| = h/2$. 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 $K_1\left((\lambda_i - t)/h\right)$ for $i = 1, 2$ will interact and be confounded in one overlapping signal. Therefore, the method will work in the multiple kink scenario in practice only if the kink locations are spaced apart such that $|\theta_i - \theta_j| \geq h$ for $i \neq j$.

The multiple kink scenario with analysis is demonstrated in Figure 11. It is demonstrated for both the case where the error variables are i.i.d. in with the observed regression model in Figure 11 (a) and the normalised probe functional $\mathcal{T}_{\widehat{\kappa}}(t)$ in Figure 11 (b). In each instance of the two kinks, θ_1, and θ_2, there are a pair of local extrema with a local minimum and local maximum within order h of the true kink locations. The complete analysis would then involve the truncation of the interval into two subintervals and conducting the zero crossing step on each subinterval to gain the estimator $\widehat{\boldsymbol{\theta}}_n = \left(\widehat{\theta}_{1,n}, \widehat{\theta}_{2,n}\right)$.



(a) Noisy function in i.i.d. noise.

(b) Bounds exceeded with $h = 0.175$.

(c) Noisy function in LRD noise ($\alpha = 0.6$).

(d) Bounds exceeded with $h = 0.27$.

Figure 11: Illustration of the method when μ has two kinks, one at $\theta_1 = 1/3$ and $\theta_2 = 2/3$.

2.4.3 *Higher-order change-points*

There was a claim earlier in Cheng and Raimondo (2008) that their method could be easily generalised to higher order derivatives. The claim in particular presented the idea that to identify a change point in the ν^{th} derivative of μ for some $\nu \geq 2$ then the term $\kappa_h^{[\nu]}(t) = h^{-(\nu+3)} \int_0^1 K_{\nu+2}\left(\frac{x-t}{h}\right) \mu(x)\, dx$ can be used instead of $\kappa_h(t)$ and all the analogous steps in the argument in the previous section apply. However, this is not actually the case. The reason being is that for $\nu \geq 2$ the deterministic bias term, $J_h^{[\nu]}(t)$ will dominate the signal generated by the localisation term $L_h^{[\nu]}(t)$. Indeed, assume $[\mu^{(\nu)}](\theta) \neq 0$ and the analogous conditions of Definition 2.5 apply for this case. Then consider the expansion on $\kappa_h^{[\nu]}(t)$,

$$\kappa_h^{[\nu]}(t) = h^{-(\nu+3)} \int_0^1 K_{\nu+2}\left(\frac{x-t}{h}\right) \mu(x)\, dx = h^{-(\nu+2)} \int_{-1}^1 K_{\nu+2}(x)\, \mu(t+hx)\, dx.$$

Assume that the boundary points of K_i are all zero for $i = \nu+2, \nu+1, \ldots, 2$. Then a similar iterative procedure of integration by parts yields,

$$\begin{aligned}
\kappa_h^{[\nu]}(t) &= h^{-(\nu+2)} \int_{-1}^1 K_{\nu+2}(x)\, \mu(t+hx)\, dx. \\
&= h^{-(\nu+2)} \left\{ K_{\nu+1}(x)\mu(t+hx)\Big|_{-1}^1 - h \int_{-1}^1 K_{\nu+1}(x)\mu^{(1)}(t+hx)dx \right\} \\
&= -h^{-(\nu+1)} \int_{-1}^1 K_{\nu+1}(x)\mu^{(1)}(t+hx)dx \\
&= \quad \vdots \\
&= (-1)^\nu h^{-2} \int_{-1}^1 K_2(x)\mu^{(\nu)}(t+hx)dx.
\end{aligned}$$

Similar to the argument in Section 2.2, define $\tau = (\theta - t)/h$ and consider all $\{t : |\theta - t| < h\}$, then $|\tau| < 1$ and it can be shown that,

$$\kappa_h^{[\nu]}(t) = L_h^{[\nu]}(t) + J_h^{[\nu]}(t)$$

where $L_h^{[\nu]}(t) = (-1)^{\nu+1} h^{-2} K_1(\tau)[\mu^{(\nu)}](\theta)$ and,

$$\begin{aligned}
J_h^{[\nu]}(t) = (-1)^\nu h^{-2} \Bigg\{ &\int_{-1}^\tau K_2(x) \left(\mu^{(\nu)}(\theta + h(x-\tau)) - \mu^{(\nu)}(\theta_-)\right) dx \\
&+ \int_\tau^1 K_2(x) \left(\mu^{(\nu)}(\theta + h(x-\tau)) - \mu^{(\nu)}(\theta_+)\right) dx \Bigg\}.
\end{aligned}$$

Also, if it is assumed that $\mu^{(s)}$ is the highest order derivative that exists and is finite where $s \geq \nu + 2$, then the vanishing moment property of K_2 implies that the deterministic bias term, $J_h^{[\nu]}(t) = \mathcal{O}\left(h^{s-\nu-2}\right)$. Now focus on part (b) of Lemma 2.1. As seen in Goldenshluger et al. (2006), the equivalent minimax optimal rate for change point estimation of order ν in the indirect model is $n^{-s/(2s+2\nu-1)}$. For the Lemma to obtain this best possible accuracy it is needed that δ is chosen as small as possible, namely $\delta \asymp h^s$. If this level of δ is again chosen here, then it is required that the lower bound of the $L_h^{[\nu]}(t)$ remains larger than the deterministic bias $J_h^{[\nu]}(t)$, that is, $\delta h^{-3} \geq C h^{s-\nu-2}$ with $\delta \geq C h^s$. This requires $h^{s-3} \geq C h^{s-\nu-2}$ which is impossible for $\nu \geq 2$.

For the sake of argument, allow δ to be chosen differently such that $\delta h^{-3} \geq C h^{s-\nu-2}$ holds, namely $\delta \geq C h^{s-\nu+1}$ (see the proof of (b) in Lemma 2.1). After considering the variance of the error process $\mathbb{V}\mathrm{ar}\, Z_h^{[\nu]}(t) = \mathcal{O}(n^{-1}h^{-\nu-5/2})$ it can be shown that the method will attain the rate, $n^{-(s-\nu+1)/(2s+1)}$ which is slower than the minimax rate of $n^{-s/(2s+2\nu-1)}$ stated by Goldenshluger et al. (2006).

2.4.4 *Why focus on the third derivative?*

As discussed in Goldenshluger et al. (2006), it is a natural question to ask why the method need focus on the third derivative? Perhaps the method could be just as reliable if only the second derivative was used? The answer to this question is that the achieved rates are not optimal. The rate relies on focusing on the three main components, the $\delta-$separation, the stochastic variance and the deterministic bias. To ease the presentation here, consider the scenario with i.i.d. errors instead of LRD errors which have been discussed so far in Part II. As seen in Section 2.2 when the third derivative is used, the $\delta-$separation and the bias are of order δh^{-3} and h^{s-3} respectively. For the i.i.d. scenario, the stochastic variance is of order $n^{-1}h^{-7}$. By considering the optimal balance of these three terms, it was shown that the optimal rate can be achieved with $\rho_n = n^{-s/(2s+1)}$.

If the alternative method was considered where the probe functional is

$$\kappa_{h,2}(t) = h^{-3} \int_0^1 K_2\left(\frac{x-t}{h}\right) \mu(x)\, dx,$$

that is, the second derivative is considered. The method would be required to locate the maximum of $\kappa_{h,2}(t)$ in the presence of noise. It can be shown via very similar methods to the ones already seen that, $\kappa_{h,2}(t) = L_{h,2}(t) + J_{h,2}(t)$ where the localisation term $L_{h,2}(t) = -h^{-1}K(\tau)[\mu^{(1)}](\theta)$ and the deterministic bias $J_{h,2}(t) = \mathcal{O}\left(h^{s-2}\right)$.

The function K is chosen such that the maximum located around $K(\theta)$ is bounded by a parabolic function. In particular choose K such that for some constants $\delta, q \in (0,1)$ such that $\delta < q$ then for all $x \in (\delta, q)$, $K(x) \geq K(0) - Cx^2 =: Q(x)$ (see for example, Figure 12). Indeed, recall (K:5) stated that $|K_1(x)| \geq C|x|$ for all $x \in (-q,q)$ using this with the boundary conditions and symmetry relations in (K:2), (K:3) and (K:6) implies that, $-K_1(x) \geq Cx$ for $x \in (0, q^*)$. Then we have,

$$\begin{aligned} K(x) - K(0) &= \int_0^x K_1(t)\, dt \\ &= -\int_0^x (-K_1(t))\, dt \\ &\geq -Cx^2. \end{aligned}$$

The fact that K_1 is odd implies that K is even and the result follows for $x \in (-q,q)$. Then it can be shown that the $\delta-$separation is of order $\delta^2 h^{-3}$. Consider the $\delta-$separation in

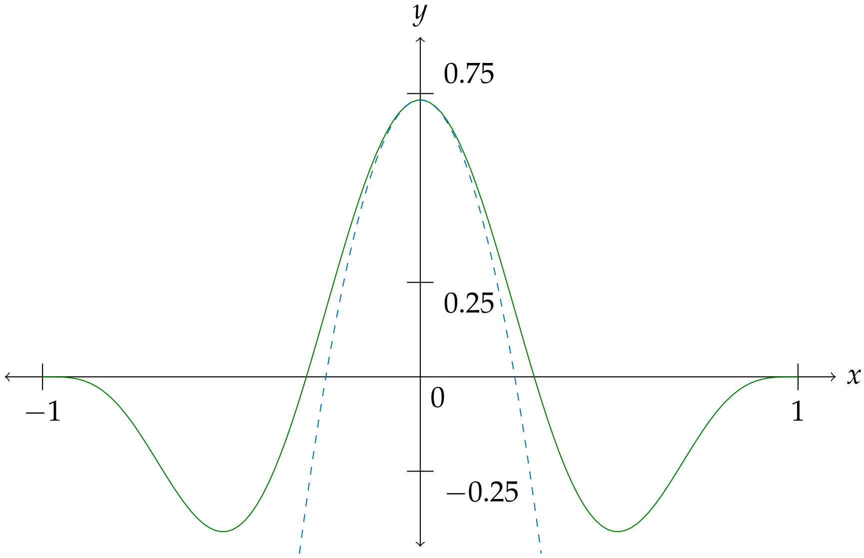


Figure 12: Plot showing the K function and its dominance over a quadratic function Q for the choice $k = 1$. The solid line represents $y = K(x)$ while the dashed line represents the quadratic function $y = Q(x)$.

more depth. This is given by, $\inf_{t \in \widehat{A}_{h,2}} \{\kappa_{h,2}(t) - \kappa_{h,2}(\theta)\}$ where $\widehat{A}_{h,2}$ is modified to be the interval with endpoints as the pair of values that exceed the known large deviation boundaries. Beforehand, $\kappa_h(\theta) = J_h(\theta)$ since the localisation term $L_h(\theta) \equiv 0$. However, here $L_{h,2}(\theta) \neq 0$ since K is maximised at zero. Considering the whole infimum further recall the value $\tau = (\theta - t)/h$ and consider $|\theta - t| \in (\delta, qh)$,

$$\begin{aligned}
\inf_{t \in \widehat{A}_{h,2}} \{\kappa_{h,2}(t) - \kappa_{h,2}(\theta)\} &= \inf_{t \in \widehat{A}_{h,2}} \{L_{h,2}(t) - L_{h,2}(\theta) + J_{h,2}(t) - J_{h,2}(\theta)\} \\
&\geq \inf_{t \in \widehat{A}_{h,2}} \{L_{h,2}(t) - L_{h,2}(\theta)\} - 2 \sup_{t \in (h,1-h)} J_{h,2}(t) \\
&= \inf_{t \in \widehat{A}_{h,2}} \left\{-h^{-1}[\mu^{(1)}](\theta)\,(K(\tau) - K(0))\right\} + \mathcal{O}\left(h^{s-2}\right) \\
&\geq Ch^{-1}\tau^2 + \mathcal{O}\left(h^{s-2}\right) \\
&\geq C\delta^2 h^{-3} + \mathcal{O}\left(h^{s-2}\right) \\
&\geq C\delta^2 h^{-3}.
\end{aligned}$$

Furthermore, in the i.i.d. framework, the stochastic variance in this scenario is of order $n^{-1}h^{-5}$. Considering the balance between the three terms achieves the rate $n^{-(s+1)/(4s+2)}$ which is slower than the optimal rate $\rho_n = n^{-s/(2s+1)}$.

Part III

OPTIMALITY OF FIXED DESIGN RESULT

3 MINIMAX OPTIMALITY

This chapter is concerned with showing the minimax optimality of the estimation method constructed in Section 2.2 for the *fractional white noise* model. The minimax optimality is not shown over the entire class $\mathscr{C}_s^1([0,1],\theta)$ but a subset of the class. As per the definitions given in Section 1.4, the minimax optimality requires checking both the upper and lower rate of convergence for the estimator.

The upper bound will be pursued first and is proved in Section 3.1 over the functional class $\mathscr{C}_s^1([0,1],\theta)$. The lower bound is proved over a slightly different functional class denoted by $\mathscr{T}_s(\theta)$. This class is characterised in the Fourier domain and is not directly equivalent to the class $\mathscr{C}_s^1([0,1],\theta)$. The definition of $\mathscr{T}_s(\theta)$ along with the lower bound result and proof is given in Section 3.2.

The upper and lower bound results given in Sections 3.1 – 3.2 will then prove the minimax optimality of the method by Theorem 3.1 below.

Theorem 3.1. *Suppose that $\mu \in \mathscr{C}_s^1([0,1],\theta) \cap \mathscr{T}_s(\theta)$ is observed in the fractional white noise model (2.2) with the calibration $\varepsilon \asymp n^{-\frac{1}{2}}$. Then the minimax rate of estimating θ is,*

$$\rho_n \asymp n^{-\alpha s/(2s+\alpha)}.$$

Therefore, the minimax optimality of the result is proven over the intersection of the two classes $\mathscr{C}_s^1([0,1],\theta) \cap \mathscr{T}_s(\theta)$. The intersection of these two classes is non-empty and infact only a slightly smaller subset than $\mathscr{C}_s^1([0,1],\theta)$. This is discussed in more depth when the classes are compared at the start of Section 3.2.

Corollary 3.1. *Suppose that $\mu \in \mathscr{C}_s^m([0,1],\boldsymbol{\theta}) \cap \mathscr{T}_s^m(\boldsymbol{\theta})$ where $\mathscr{T}_s^m(\boldsymbol{\theta})$ is given by Definition 3.9 and μ is observed in the fractional white noise model (2.2) with the calibration $\varepsilon \asymp n^{-\frac{1}{2}}$. Then the minimax rate of estimating $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_m)$ is also $\rho_n \asymp n^{-\alpha s/(2s+\alpha)}$.*

Proof of Corollary 3.1 . As was discussed in Section 2.4, the estimation argument for multiple kinks can be generalised to $m \geq 2$ change-points easily. The corollary will follow immediately from the proof of Theorem 3.1. □

3.1 UPPER BOUND

Before we state the theorem for the upper rate of convergence for the estimator used in the kink estimation technique, we state a key Lemma that is used in the proof. The Lemma is concerned with the estimated localisation values $\widehat{t}_*, \widehat{t}^*$ and gives them a probabilistic bound which is crucial for the *localisation step* shown in Section 2.2.

Lemma 3.1. *Let $h^{(\alpha+2)/2}\varepsilon^{-\alpha} \geq C$. Then, for $\mu \in \mathscr{C}^1_s([0,1],\theta)$ and for h sufficiently small,*

$$P\left(\left|\widehat{t}_* - t_*\right| > \frac{hd}{2}\right) \vee P\left(\left|\widehat{t}^* - t^*\right| > \frac{hd}{2}\right) \leq Ch^{\frac{\alpha}{2}}\varepsilon^{-\alpha}\exp\left\{-C\frac{h^{(\alpha+2/2}}{\varepsilon^{2\alpha}}\right\}$$

where $d = |q^* - q_0| > 0$ and q_0 is the first zero of K_1 to the right of q^*, the abscissa of the global minimum.

Proof of Lemma 3.1 . We first consider the inequality for $P\left(\left|\widehat{t}_* - t_*\right| > \frac{hd}{2}\right)$. Let

$$\Delta := \left\{t \in [0,1] : |t - t_*| > \frac{hd}{2}\right\},$$

and recall that $t_* = \underset{t\in(h,1-h)}{\arg\min}\, L_h(t) = \theta - q^*h$ then,

$$\begin{aligned}
P\left(\left|\widehat{t}_* - t_*\right| > \frac{hd}{2}\right) &\leq P\left(\exists\, t \in \Delta : \widehat{\kappa}_h\left(t_*\right) > \widehat{\kappa}_h(t)\right) \\
&= P\left(\exists\, t \in \Delta : \widehat{\kappa}_h\left(t_*\right) - \kappa_h\left(t_*\right) + \kappa_h(t) - \widehat{\kappa}_h(t) > \kappa_h(t) - \kappa_h\left(t_*\right)\right) \\
&\leq P\left(2\sup_{t\in(0,1)}\left|\widehat{\kappa}_h(t) - \kappa_h(t)\right| > \inf_{t\in\Delta}\kappa_h(t) - \kappa_h\left(t_*\right)\right) \\
&= P\left(2\sup_{t\in(0,1)}\left|Z_h(t)\right| > \inf_{t\in\Delta}\kappa_h(t) - \kappa_h\left(t_*\right)\right). \qquad (3.1)
\end{aligned}$$

Consider the infimum by expanding $\kappa_h(t)$ in terms of the localisation term $L_h(t)$ and the deterministic bias $J_h(t)$,

$$\inf_{t\in\Delta}\left\{\kappa_h(t) - \kappa_h\left(t_*\right)\right\} \geq h^{-2}[\mu^{(1)}](\theta)\inf_{t\in\Delta}\left\{K_1\left(\frac{\theta - t}{h}\right) - K_1\left(\frac{\theta - t_*}{h}\right)\right\} - 2\sup_{t\in(0,1)} J_h(t).$$

Change variable $x = h^{-1}\left(\theta - t\right)$ so $t \in \Delta$ is equivalent to $x \in \{u : |u - q^*| > d/2\}$,

$$\begin{aligned}
\inf_{t\in\Delta}\left\{\kappa_h(t) - \kappa_h\left(t_*\right)\right\} &\geq h^{-2}[\mu^{(1)}](\theta)\inf_{\left\{x:|x-q^*|>\frac{d}{2}\right\}}\left\{K_1\left(x\right) - K_1\left(q^*\right)\right\} - 2\sup_{t\in(0,1)}\left|J_h(t)\right| \\
&\geq Ch^{-2} - 2Ch^{s-3} \\
&\geq Ch^{-2}. \qquad (3.2)
\end{aligned}$$

The last inequality follows since h^{-2} dominates h^{s-3} as h tends to zero. Notice that the restriction on h such that $h^{(\alpha+2)/2}\varepsilon^{-\alpha} \geq C$ with the choice $\lambda = 1/2Ch^{-2}$ satisfies $\lambda \geq 2\sigma_Z \asymp h^{-\frac{\alpha}{2}-3}\varepsilon^{\alpha}$. Using (3.1) and (3.2), apply Lemma 2.3,

$$\begin{aligned} P\left(\left|\widehat{t}_* - t_*\right| > \frac{hd}{2}\right) &\leq P\left(2\sup_{t\in(0,1)}|Z_h(t)| > \inf_{t\in\Delta}\kappa_h(t) - \kappa_h(t_*)\right) \\ &\leq P\left(\sup_{t\in(0,1)}|Z_h(t)| > \frac{Ch^{-2}}{2}\right) \\ &\leq Ch^{(\alpha+4)/2}\varepsilon^{-\alpha}h^{-2}\exp\left\{-C\frac{h^{-4}h^{\alpha+6}}{\varepsilon^{2\alpha}}\right\} \\ &= Ch^{\alpha/2}\varepsilon^{-\alpha}\exp\left\{-C\frac{h^{\alpha+2}}{\varepsilon^{2\alpha}}\right\}. \end{aligned} \tag{3.3}$$

The proof for the t^* case follows by a similar argument. Define, $\Delta = \left\{t \in (0,1) : |t - t^*| > \frac{hd}{2}\right\}$ and recall that $t^* = \arg\max_t L_h(t) = \theta + q^*h$ then,

$$P\left(\left|\widehat{t}^* - t^*\right| > \frac{hd}{2}\right) \leq Ch^{\alpha/2}\varepsilon^{-\alpha}\exp\left\{-C\frac{h^{\alpha+2}}{\varepsilon^{2\alpha}}\right\}. \tag{3.4}$$

Using both (3.3) and (3.4),

$$\begin{aligned} P\left(\left|\widehat{t}_* - t_*\right| > \frac{hd}{2}\right) \vee P\left(\left|\widehat{t}^* - t^*\right| > \frac{hd}{2}\right) &\leq Ch^{\alpha/2}\varepsilon^{-\alpha}\exp\left\{-C\frac{h^{\alpha+2}}{\varepsilon^{2\alpha}}\right\} \\ &= Ch^{\alpha/2}\varepsilon^{-\alpha}\exp\left\{-C\frac{h^{\alpha+2}}{\varepsilon^{2\alpha}}\right\}. \end{aligned}$$

□

Theorem 3.2. *Suppose $\mu \in \mathscr{C}_s^1([0,1],\theta)$ and $K \in \mathscr{K}_s$ and $\widehat{A}_n = \left(\widehat{t}_*, \widehat{t}^*\right)$. Then let*

$$\widehat{\theta}_n = \arg\min_{t\in\widehat{A}_n}|\widehat{\kappa}_h(t)|$$

be an estimator for the change-point θ with the bandwidth $h = h_*$ which is of the order,

$$h_* \asymp n^{-\alpha/(2s+\alpha)}. \tag{3.5}$$

Then, there exists a constant $C^* < \infty$ that does not depend on n such that under the square loss,

$$\limsup_{n\to\infty}\inf_{\widehat{\theta}_n} R^*_{\mathscr{C}_s^1}(\widehat{\theta}_n, n^{-\alpha s/(2s+\alpha)}) \leq C^*.$$

Proof of Theorem 3.2 . Notice that using our choice of $h = h_* \asymp n^{-\alpha/(2s+\alpha)}$ then,

$$\begin{aligned}
\varepsilon^{-\alpha} h_*^{(\alpha+2)/2} &\asymp n^{\alpha/2-(\alpha(\alpha+2))/(2(2s+\alpha)}\\
&= n^{\alpha(s-1)/(2s+\alpha)}\\
&= n^{\beta} \qquad \text{for some } \beta > 0, \qquad (3.6)\\
&\geq C,
\end{aligned}$$

for some $C > 0$ independent of n. This will allow us to apply Lemma 3.1 with the bandwidth choice $h = h_*$. Now let $E_n = \left\{|\widehat{t}_* - t_*| \leq hd/2\right\} \cap \left\{|\widehat{t}^* - t^*| \leq hd/2\right\}$.

$$\begin{aligned}
\mathbb{E}\left|\widehat{\theta}_n - \theta\right|^2 &= \mathbb{E}\left|\widehat{\theta}_n - \theta\right|^2 \mathbb{1}_{E_n} + \mathbb{E}\left|\widehat{\theta}_n - \theta\right|^2 \mathbb{1}_{E_n^c}\\
&\leq \mathbb{E}\left\{\left|\widehat{\theta}_n - \theta\right|^2 \mathbb{1}_{E_n}\right\} + P\left(E_n^c\right). \qquad (3.7)
\end{aligned}$$

Apply Lemma 3.1 to bound the probability $P\left(E_n^c\right)$,

$$\begin{aligned}
P(E_n^c) &= P\left(\left\{\left|\widehat{t}_* - t_*\right| > \frac{hd}{2}\right\} \cup \left\{\left|\widehat{t}^* - t^*\right| > \frac{hd}{2}\right\}\right)\\
&\leq C h^{\alpha/2} \varepsilon^{-\alpha} \exp\left\{-C \frac{h^{\alpha+2}}{\varepsilon^{2\alpha}}\right\}\\
&\leq C n^{\alpha s/(2s+\alpha)} \exp\left\{-C n^{2\alpha(s-1)/(2s+\alpha)}\right\} \overset{n\to\infty}{\to} 0 \qquad \text{by (3.6).} \qquad (3.8)
\end{aligned}$$

Recall that, $|t_* - \theta| = |t^* - \theta| = q^* h$. Thus, on the set E_n

$$\begin{aligned}
\left|\widehat{t}_* - \theta\right| &= \left|\widehat{t}_* - t_* + t_* - \theta\right|\\
&\leq \left|\widehat{t}_* - t_*\right| + |t_* - \theta|\\
&\leq \frac{hd}{2} + q^* h =: rh. \qquad (3.9)
\end{aligned}$$

By a similar argument, the other estimator $\widehat{t}^*$ can be bounded,

$$\left|\widehat{t}^* - \theta\right| \leq rh. \qquad (3.10)$$

Recall $\widehat{\theta}_n \in \left(\widehat{t}_*, \widehat{t}^*\right)$ and combining this with (3.9) and (3.10) implies,

$$\left|\widehat{\theta}_n - \theta\right| \leq rh. \qquad (3.11)$$

Now let $0 < \delta < rh$ and $\Delta_j = \left[\delta 2^{j-1}, \delta 2^j\right]$ and $J = \min\left\{j : \delta 2^j > rh\right\}$ then by (3.11),

$$\mathbb{E}\left\{\left|\widehat{\theta}_n - \theta\right|^2 \mathbb{1}_{E_n}\right\} \leq \delta^2 + \sum_{j=1}^{J} \delta^2 2^{2j} P\left(\left\{\left|\widehat{\theta}_n - \theta\right| \in \Delta_j\right\} \cap E_n\right). \qquad (3.12)$$

Define, $T_j = \left\{t : |t-\theta| \in \Delta_j\right\}$ then $|T_j| = \delta 2^{j-1}$ and

$$\begin{aligned} P\left(\left\{\left|\widehat{\theta}_n - \theta\right| \in \Delta_j\right\} \cap E_n\right) &\leq P\left(\exists\, t \in T_j : |\widehat{\kappa}_h(\theta)| \geq |\widehat{\kappa}_h(t)|\right) \\ &= P\left(\exists\, t \in T_j : |\widehat{\kappa}_h(\theta)| - |\kappa_h(\theta)| + |\kappa_h(t)| - |\widehat{\kappa}_h(t)| \geq |\kappa_h(t)| - |\kappa_h(\theta)|\right) \\ &\leq P\left(2\sup_{t\in T_j} |\widehat{\kappa}_h(t) - \kappa_h(t)| \geq \inf_{t\in T_j} |\kappa_h(t)| - |\kappa_h(\theta)|\right). \end{aligned} \tag{3.13}$$

Now apply Lemma 2.1 on $\inf_{t\in\Delta_j} |\kappa_h(t)| - |\kappa_h(\theta)|$. This can be done by noting that we can substitute the set T_j with $A_{\delta 2^{j-1},h}$ provided that $\delta 2^{j-1} \geq Ch^s$. Then part (b) of Lemma 2.1 implies that,

$$\begin{aligned} \inf_{t\in T_j} |\kappa_h(t)| &\geq \inf_{A_{\delta 2^{j-1},h}} |\kappa_h(t)| \\ &\geq C\delta 2^{j-1} h^{-3}. \end{aligned} \tag{3.14}$$

Apply part (a) of Lemma 2.1 along with (3.14),

$$\begin{aligned} \inf_{t\in T_j} |\kappa_h(t)| - |\kappa_h(\theta)| &\geq \inf_{t:\delta 2^{j-1}<|t-\theta|<rh} |\kappa_h(t)| - |\kappa_h(\theta)| \\ &\geq C\delta 2^{j-1}h^{-3} - Ch^{s-3} \\ &\geq C\delta 2^{j-1} h^{-3}. \end{aligned} \tag{3.15}$$

Now if we choose $\delta \asymp h^s$ with $h = h_*$ then $\delta \asymp n^{-\frac{\alpha s}{2s+\alpha}}$ and for a sufficiently large C, $C\delta 2^{j-2}h^{-3} \geq 2\sigma_Z = 2\tau_1 \varepsilon^{\alpha} h^{-(\alpha+6)/2}$. Then using (3.13) and (3.15) with the choice $\lambda = C\delta 2^{j-1}h^{-3}$ apply Lemma 2.3,

$$\begin{aligned} P\left(\left\{\left|\widehat{\theta}_n - \theta\right| \in \Delta_j\right\} \cap E_n\right) &\leq P\left(2\sup_{t\in T_j} |\widehat{\kappa}_h(t) - \kappa_h(t)| \geq \inf_{t\in T_j} |\kappa_h(t)| - |\kappa_h(\theta)|\right) \\ &\leq P\left(\sup_{t\in T_j} |Z_h(t)| \geq C\delta 2^{j-2} h^{-3}\right) \\ &\leq C\delta 2^{j+1} h^{\alpha/2-1} \varepsilon^{-\alpha} |T_j| \exp\left(-C\frac{\delta^2 2^{2j} h^{\alpha}}{\varepsilon^{2\alpha}}\right) \\ &\leq C\delta^2 2^{2j} h^{(\alpha-2)/2} \varepsilon^{-\alpha} \exp\left(-C2^{2j}\right). \end{aligned} \tag{3.16}$$

The last line uses the fact that under the choice of optimal bandwidth (3.5) and $\delta \asymp h^s$ then $\delta^2 h^\alpha \varepsilon^{-2\alpha} = \mathcal{O}(1)$. Furthermore, notice that,

$$\begin{aligned}
\delta^2 2^{2j} h^{(\alpha-2)/2} \varepsilon^{-\alpha} &\asymp h^{(4s+\alpha-2)/2} \varepsilon^{-\alpha} \\
&= \varepsilon^{2\alpha(s-1)/(2s+\alpha)} \\
&\asymp n^{-\alpha(s-1)/(2s+\alpha)} \\
&\overset{n\to\infty}{\to} 0 \quad \text{since } s \geq 3 \text{ and } \alpha \in (0,1].
\end{aligned} \tag{3.17}$$

Thus, by (3.12), (3.16) and (3.17) we have,

$$\begin{aligned}
\mathbb{E}\left\{\left|\widehat{\theta}_n - \theta\right|^2 \mathbb{1}_{E_n}\right\} &\leq \delta^2 + \sum_{j=1}^{J} \delta^2 2^{2j} P\left(\left\{\left|\widehat{\theta}_n - \theta\right| \in \Delta_j\right\} \cap E_n\right) \\
&\leq \delta^2 \left(1 + o(1)\right).
\end{aligned} \tag{3.18}$$

Finally, using (3.7), (3.8) and (3.18) we have,

$$\mathbb{E}\left|\widehat{\theta}_n - \theta\right|^2 \leq \mathbb{E}\left\{\left|\widehat{\theta}_n - \theta\right|^2 \mathbb{1}_{E_n}\right\} + P(E_n^c) = \delta^2 \left(1 + o(1)\right) \asymp n^{-\alpha s/(2s+\alpha)} \left(1 + o(1)\right). \tag{3.19}$$

Therefore, by (3.19) there exists a positive constant $C^* < \infty$ that doesn't depend on n such that under the square loss function $w(u) = u^2$ we have,

$$\begin{aligned}
\limsup_{n\to\infty} \inf_{\widehat{\theta}_n} R^*_{\mathscr{C}_s^1}\left(\widehat{\theta}_n, n^{-\alpha s/(2s+\alpha)}\right) &= \limsup_{n\to\infty} \inf_{\widehat{\theta}_n} \mathbb{E}\left|n^{\alpha s/(2s+\alpha)}\left(\widehat{\theta}_n - \theta\right)\right|^2 \\
&\leq \limsup_{n\to\infty} \inf_{\widehat{\theta}_n} n^{2\alpha s/(2s+\alpha)} \mathbb{E}\left|\widehat{\theta}_n - \theta\right|^2 \\
&\leq C^*. \qquad \square
\end{aligned}$$

3.2 LOWER BOUND

The lower bound result is spread over three subsections. In Section 3.2.1, the generalised framework for constructing a minimax lower bound is outlined. Due to the nature of the problem, there is a close link between fBm and fractional calculus. Some key aspects of fractional calculus and fractional stochastic calculus that are crucial to obtaining the lower bound result are given in Sections 3.2.2 – 3.2.3. The main lower bound result along with the definitions and discussion of the functional class $\mathscr{T}_s(\theta)$ are given in Section 3.2.4.

3.2.1 *General reduction scheme*

A general outline of obtaining a reduction of the general minimax risk was obtained by Tsybakov (2009) and is explained here for completeness and easy reference. It is based on the following three points.

(i) Consider a first simplification by an application of the Markov Inequality. Recall from Section 1.4 that the loss function w is positive and monotone, then let $A > 0$ and,

$$\mathbb{E}\left\{w\left(\rho_n^{-1}d\left(\hat{f}_n, f\right)\right)\right\} \geq w\left(A\right)P\left(\rho_n^{-1}d\left(\hat{f}_n, f\right) \geq A\right) = w\left(A\right)P\left(d\left(\hat{f}_n, f\right) \geq \delta\right). \tag{3.20}$$

So, by (3.20) we can reduce the general minimax risk, $\mathcal{R}^*_{\mathscr{F}}\left(\rho_n\right)$, to the minimax probability as follows,

$$\begin{aligned}\mathcal{R}^*_{\mathscr{F}}\left(\rho_n\right) &= \inf_{\hat{f}_n}\sup_{f\in\mathscr{F}}\mathbb{E}\left\{w\left(\rho_n^{-1}d\left(\hat{f}_n, f\right)\right)\right\}\\ &\geq \inf_{\hat{f}_n}\sup_{f\in\mathscr{F}} w\left(A\right)P\left(d\left(\hat{f}_n, f\right) \geq A\rho_n\right).\end{aligned}$$

Hence, to obtain a lower bound on the minimax risk, it is sufficient to consider a lower bound on the minimax probability given by,

$$\inf_{\hat{f}_n}\sup_{f\in\mathscr{F}} P\left(d\left(\hat{f}_n, f\right) \geq A\rho_n\right).$$

(ii) Reduce the problem further by considering two elements $f_0, f_1 \in \mathscr{F}$. Obviously if $f_0, f_1 \in \mathscr{F}$, then,

$$\inf_{\hat{f}_n}\sup_{f\in\mathscr{F}} P\left(d\left(\hat{f}_n, f\right) \geq A\rho_n\right) \geq \inf_{\hat{f}_n}\max_{f\in\{f_0,f_1\}} P\left(d\left(\hat{f}_n, f\right) \geq A\rho_n\right) \tag{3.21}$$

whenever we have a finite subset $\{f_0, f_1\} \subset \mathscr{F}$. In practice, $\{f_0, f_1\}$ will be suitably chosen for the problem.

(iii) Choose functions f_0, f_1 separated by a distance of at least $2A\rho_n$. So by the triangle inequality,

$$d\left(\hat{f}_n, f_0\right) + d\left(\hat{f}_n, f_1\right) \geq d\left(f_0, f_1\right) \geq 2A\rho_n. \tag{3.22}$$

Then, for any estimator $\hat{f}_n$,

$$P\left(d\left(\hat{f}_n, f_j\right) \geq A\rho_n\right) \geq P\left(\zeta^* \neq j\right), \quad j = 0, 1; \tag{3.23}$$

where ζ^* is the test of the minimum distance defined by,

$$\zeta^* = \arg\min_{j\in\{0,1\}} d\left(\widehat{f}_n, f_j\right).$$

Indeed, consider the following argument. From (3.22) it implies that $d(\widehat{f}_n, f_0)$ and $d(\widehat{f}_n, f_1)$ cannot both be simultaneously be less than $A\rho_n$. Say for example that $d(\widehat{f}_n, f_{j'}) < A\rho_n$, then $d(\widehat{f}_n, f_j) > A\rho_n$. This implies that,

$$\begin{aligned} P\left(d\left(\widehat{f}_n, f_j\right) \geq A\rho_n\right) &\geq P\left(d\left(\widehat{f}_n, f_{j'}\right) < A\rho_n\right) \\ &= P\left(d\left(\widehat{f}_n, f_{j'}\right) < d\left(\widehat{f}_n, f_j\right)\right) \\ &= P\left(\zeta^* \neq j\right). \end{aligned}$$

If we now combine (3.23) and (3.21), we can construct two functions f_0, f_1 that satisfy (3.22) such that,

$$\begin{aligned} \inf_{\widehat{f}_n} \sup_{f\in\mathscr{F}} P\left(d\left(\widehat{f}_n, f\right) \geq A\rho_n\right) &\geq \inf_{\widehat{f}_n} \max_{f\in\{f_0,f_1\}} P\left(d\left(\widehat{f}_n, f\right) \geq A\rho_n\right) \\ &\geq \inf_{\widehat{f}_n} \max_{j\in\{0,1\}} P\left(\zeta^* \neq j\right) \\ &= \inf_{\widehat{f}_n} \max_{j\in\{0,1\}} P\left(\arg\min_{k=0,1} d\left(\widehat{f}_n, f_k\right) \neq j\right) \\ &\geq \inf_{\zeta} \max_{j\in\{0,1\}} P\left(\zeta \neq j\right) \end{aligned}$$

where $\inf_\zeta$ should be interpreted as the infimum over all test functions, ζ.

Therefore, to establish the minimax lower bound, we just need to check that

$$\inf_{\zeta} \max_{j=0,1} P\left(\zeta \neq j\right) \geq C,$$

where $0 < C < \infty$ is a positive constant that does not depend on n and that f_0 and f_1 satisfy (3.22). The value $\inf_\zeta \max_{j=0,1} P\left(\zeta \neq j\right)$ is called the minimax probability of error for the problem of testing the two hypotheses f_0, f_1.

3.2.2 *Fractional calculus*

In this section, the machinery and analysis of some relevant fractional calculus are outlined with the use of some fractional functional operators. These are defined on both the real domain and the Fourier domain. There is good reason for this, as will be seen in Section 3.2.4, the functional class $\mathscr{T}_s(\theta)$, is defined in the Fourier domain and consequently requires this machinery to be defined in the Fourier domain as well. To be sure that the results of Parseval and Plancherel can be used, assume at the very least that $\mu \in \mathscr{L}_2([0,1],\mathbb{R})$. Start with the definition of the fractional integral operator and its adjoint.

Definition 3.1. *Let $\nu > 0$ and $\mu \in \mathscr{L}_1([0,1],\mathbb{R})$ then for $x \in [0,1]$ the fractional integral operator $\mathcal{D}^{-\nu}\mu(x)$ exists almost everywhere and is given by,*

$$\begin{aligned}\mathcal{D}^{-\nu}\mu(x) &:= \begin{cases} \frac{1}{\Gamma(\nu)}\int_0^1 (x-t)^{\nu-1}\mu(t)dt, & \text{if } x \geq 1 \\ \frac{1}{\Gamma(\nu)}\int_0^x (x-t)^{\nu-1}\mu(t)dt, & \text{if } 0 < x < 1 \\ 0, & \text{if } x \leq 0. \end{cases} \\ &= \mathrm{D}^{-\nu} * \mu(x)\end{aligned}$$

where $\mathrm{D}^{-\nu}(x) := \frac{1}{\Gamma(\nu)} x^{\nu-1}\mathbb{1}_{(0,\infty)}(x)$.

Proposition 3.1. *Let $\nu > 0$ and $\mu \in \mathscr{L}_1([0,1],\mathbb{R})$, then the fractional integral operator $\mathcal{D}^{-\nu}$ given in Definition 3.1 has an adjoint $\mathcal{D}_*^{-\nu}$ that satisfies,*

$$\begin{aligned}\mathcal{D}_*^{-\nu}\mu(x) &:= \begin{cases} 0, & \text{if } x \geq 1 \\ \frac{1}{\Gamma(\nu)}\int_x^1 (t-x)^{\nu-1}\mu(t)dt, & \text{if } 0 < x < 1 \\ \frac{1}{\Gamma(\nu)}\int_0^1 (t-x)^{\nu-1}\mu(t)dt, & \text{if } x \leq 0. \end{cases} \\ &= \mathrm{D}_*^{-\nu} * \mu(x),\end{aligned}$$

where $\mathrm{D}_*^{-\nu}(x) := \mathrm{D}^{-\nu}(-x) = \dfrac{1}{\Gamma(\nu)}(-x)^{\nu-1}\mathbb{1}_{(0,\infty)}(-x)$.

Proof of Proposition 3.1 . If $\mu, g \in \mathscr{L}_1([0,1],\mathbb{R})$, then use Definition 3.1 and apply Fubini's Theorem,

$$\begin{aligned}\langle \mathcal{D}^{-\nu}\mu, g\rangle &= \int_0^1 (\mathcal{D}^{-\nu}\mu)(x)g(x)dx \\ &= \frac{1}{\Gamma(\nu)}\int_0^1\int_0^x (x-t)^{\nu-1}\mu(t)dt\, g(x)dx \\ &= \frac{1}{\Gamma(\nu)}\int_0^1\int_t^1 (x-t)^{\nu-1}g(x)dx\, \mu(t)dt \\ &= \int_0^1 \mu(t)\mathcal{D}_*^{-\nu}g(t)dt \\ &= \langle \mu, \mathcal{D}_*^{-\nu}g\rangle.\end{aligned}$$

□

As stated earlier, the operators have a representation in the Fourier domain and these will be useful for our purposes in the lower bound result in Section 3.2.4. The next proposition determines this Fourier representation.

Proposition 3.2. *Let $\nu > 0$ with $\mathcal{D}^{-\nu}$ and $\mathcal{D}_*^{-\nu}$ be the fractional integral operator and its adjoint given by Definition 3.1 and Proposition 3.1 respectively. Then these operators have the Fourier domain representation satisfying,*

$$\mathcal{F}\mathcal{D}^{-\nu}\mu(\omega) = (2\pi i\omega)^{-\nu}\widetilde{\mu}(\omega) \qquad \text{and} \qquad \mathcal{F}\mathcal{D}_*^{-\nu}\mu(\omega) = (-2\pi i\omega)^{-\nu}\widetilde{\mu}(\omega).$$

Proof of Proposition 3.2 .

$$\begin{aligned}
\mathcal{F}\mathrm{D}^{-\nu}(\omega) &= \int_{\mathbb{R}} \mathrm{D}^{-\nu}(x)e^{-2\pi i\omega x}dx \\
&= \int_{\mathbb{R}} \frac{x^{\nu-1}\mathbb{1}_{\{x\geq 0\}}}{\Gamma(\nu)}e^{-2\pi i\omega x}dx \\
&= \int_0^\infty \frac{x^{\nu-1}}{\Gamma(\nu)}e^{-2\pi i\omega x}dx \\
&= (2\pi i\omega)^{-\nu}\int_0^\infty \frac{u^{\nu-1}}{\Gamma(\nu)}e^{-u}du \\
&= (2\pi i\omega)^{-\nu}.
\end{aligned}$$

Thus, by the Convolution theorem,

$$\begin{aligned}
\mathcal{F}\mathcal{D}^{-\nu}\mu(\omega) &= \mathcal{F}\left(\mathrm{D}^{-\nu}*\mu\right)(\omega) \\
&= \mathcal{F}\mathrm{D}^{-\nu}(\omega)\mathcal{F}\mu(\omega) \\
&= (2\pi i\omega)^{-\nu}\widetilde{\mu}(\omega).
\end{aligned}$$

□

The fractional integral operators have an inverse operator or *fractional derivatives* operator given by,

Definition 3.2. *Let $0 < \nu < 1$ then the fractional derivative operator $\mathcal{D}^\nu$ is given by,*

$$\mathcal{D}^\nu\mu(x) := \mathcal{D}\mathcal{D}^{-(1-\nu)}\mu(x),$$

where $\mathcal{D} = \frac{d}{dx}$ is the regular differential operator and $\mathcal{D}^{-(1-\nu)}$ is the fractional integral operator given by Definition 3.1.

Similarly there is a corresponding adjoint operator defined below.

Definition 3.3. *Let $0 < \nu < 1$ then the adjoint fractional derivative operator $\mathcal{D}^\nu$ is given by,*

$$\mathcal{D}_*^\nu\mu(x) := \mathcal{D}\mathcal{D}_*^{-(1-\nu)}\mu(x),$$

where $\mathcal{D} = \frac{d}{dx}$ is the regular differential operator and $\mathcal{D}_*^{-(1-\nu)}$ is the adjoint fractional integral operator given by Proposition 3.1.

The fractional derivative operators given by Definition 3.2 and Definition 3.3 are inverses for $\mathcal{D}^{-\nu}$ and $\mathcal{D}_*^{-\nu}$ since the fractional integral operators are linear and $\mathcal{D}$ is the inverse of $\mathcal{D}^{-1}$, the regular integral operator.

3.2.3 *Fractional stochastic calculus*

To be able to bound the error process $Z_h(t)$, some extra results from fractional stochastic calculus are needed. This is done by re-expressing the fBm in a particular way in terms of a fractional derivative. Recall from Definition 2.4 that fBm has the representation,

$$B_H(t) = \frac{C_H}{\Gamma\left(H+\frac{1}{2}\right)} \int_0^t (t-s)^{H-\frac{1}{2}}\, dB(s).$$

This quantity can be linked to fractional calculus with the next proposition.

Proposition 3.3. *Assume $t \in [0,1]$, $x \in [0,1]$ and $\nu > 0$, then the following holds,*

$$\mathcal{D}_*^{-\nu} \mathbb{1}_{(0,t)}(x) = \frac{1}{\Gamma(\nu+1)} (t-x)_+^{\nu},$$

where

$$x_+ := \begin{cases} x, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0. \end{cases} \tag{3.24}$$

Proof of Proposition 3.3 . Note that,

$$(t-x)_+ = \begin{cases} (t-x), & \text{if } 0 < x < t, \\ 0, & \text{if } 0 < t < x. \end{cases}$$

The proof is split into these two cases.

- For $0 \leq x < t \leq 1$,

$$\begin{aligned} \mathcal{D}_*^{-\nu} \mathbb{1}_{(0,t)}(x) &= \frac{1}{\Gamma(\nu)} \int_x^1 (u-x)^{\nu-1} \mathbb{1}_{(0,t)}(u) du \\ &= \frac{1}{\Gamma(\nu)} \int_x^t (u-x)^{\nu-1} du \\ &= \frac{(t-x)^{\nu}}{\Gamma(\nu+1)}. \end{aligned}$$

- For $0 \leq t < x \leq 1$,

$$\mathcal{D}_*^{-\nu} \mathbb{1}_{(0,t)}(x) = \frac{1}{\Gamma(\nu)} \int_x^1 (u-x)^{\nu-1} \mathbb{1}_{(0,t)}(u) du = 0. \qquad \square$$

We are now in a position to be able to show that $B_H(t)$ can be considered as the fractional integral of standard Brownian in the following sense.

Proposition 3.4. *Let $H \in \left(\frac{1}{2}, 1\right)$ and B be a standard Brownian motion. Then the fBm B_H can be expressed by,*

$$B_H(t) = C_H \mathcal{D}^{-(H-\frac{1}{2})} B(t),$$

where C_H is given in Definition 2.4.

Proof. First, an auxiliary result will be shown that,

$$\frac{d}{dt} \int_0^t \frac{(t-s)^{H-\frac{1}{2}}}{\Gamma\left(H+\frac{1}{2}\right)} B(s) ds = \int_0^t \frac{(t-s)^{H-\frac{1}{2}}}{\Gamma\left(H+\frac{1}{2}\right)} dB(s). \tag{3.25}$$

This result uses Leibniz Integral rule and Lebesgue-Stieltjes integration by parts. Apply the Leibniz integral rule first, note that the derivative of the lower integral limit is trivially zero and the integrand evaluated at the upper limit at t is zero. Therefore, by the Leibniz integral rule, the differential operator passes into the integrand,

$$\begin{aligned}
\frac{d}{dt} \int_0^t \frac{(t-s)^{H-\frac{1}{2}}}{\Gamma(H+\frac{1}{2})} B(s) ds &= \int_0^t \frac{d}{dt} \left(\frac{(t-s)^{H-\frac{1}{2}}}{\Gamma(H+\frac{1}{2})} \right) B(s) ds \\
&= \int_0^t \frac{d}{ds} \left(-\frac{(t-s)^{H-\frac{1}{2}}}{\Gamma(H+\frac{1}{2})} \right) B(s) ds, \\
&= \int_0^t B(s) dF_t(s), \qquad \text{where } F_t(s) = -\frac{(t-s)^{H-\frac{1}{2}}}{\Gamma(H+\frac{1}{2})} \\
&= -\int_0^t F_t(s)\, dB(s) + B(t^+) F_t(t^+) - B(0^-) F_t(0^-) \\
&= \int_0^t \frac{(t-s)^{H-\frac{1}{2}}}{\Gamma(H+\frac{1}{2})} dB(s),
\end{aligned}$$

which proves (3.25). Apply the fractional integral operator to $B(t)$ and use (3.25),

$$\begin{aligned}
C_H \mathcal{D}^{-(H-\frac{1}{2})} B(t) &= C_H \int_0^t \frac{(t-s)^{H-\frac{3}{2}}}{\Gamma(H-\frac{1}{2})} B(s) ds \\
&= C_H \int_0^t \frac{d}{dt} \left(\frac{(t-s)^{H-\frac{1}{2}}}{\Gamma(H+\frac{1}{2})} \right) B(s) ds \\
&= C_H \int_0^t \frac{(t-s)^{H-\frac{1}{2}}}{\Gamma(H+\frac{1}{2})} dB(s) \\
&= B_H(t).
\end{aligned}$$

□

The next idea is to consider the set of functions μ such that the stochastic integral with respect to fBm is well defined. That is, the process,

$$\int_0^t \mu(x)\, dB_H(x),$$

is a well defined random variable with finite variance. The class of functions μ that satisfy this property will be denoted by $\mathcal{H}$ and is defined below.

Definition 3.4. *Let $H \in \left(\frac{1}{2}, 1\right)$ be constant, then the class $\mathcal{H}$ is defined,*

$$\mathcal{H} = \left\{ \mu : [0,1] \longrightarrow \mathbb{R} \,\middle|\, |\mu|_{\mathcal{H}} = \langle \mu, \mu \rangle_{\mathcal{H}} = V_\alpha \int_0^1 \int_0^1 \mu(s)\mu(t)\, |t-s|^{-\alpha}\, ds\, dt < \infty \right\}.$$

A useful embedding of the square integrable functions exists with $\mathscr{L}_2\left([0,1], \mathbb{R}\right) \subset \mathcal{H}$ and is verified with an easy argument in (Biagini, Hu, Øksendal and Zhang, 2008, p33).

The next proposition involves stochastic integrals with respect to fBm and shows that they are equivalent to a stochastic integral with respect to a standard Brownian motion. As is standard with stochastic integrals, they are equal in the mean-square limit sense. This is explained further after the following Proposition.

Proposition 3.5. *Assume $H \in (\frac{1}{2}, 1)$ and $\mathcal{D}_*^{-(H-\frac{1}{2})}\mu \in \mathscr{L}_2\left([0,1], \mathbb{R}\right)$ then the integral of μ with respect to fBm can be re-expressed as the following integral with respect to a standard Brownian motion,*

$$\int_0^1 \mu(x)\, dB_H(x) = C_H \int_0^1 \mathcal{D}_*^{-(H-\frac{1}{2})}\mu(x)\, dB(x). \tag{3.26}$$

Proof of Proposition 3.5 . Firstly, there is an isometry between the inner products, $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ and $\langle \cdot, \cdot \rangle$. This is given by,

Lemma 3.1.2 (Biagini, Hu, Øksendal and Zhang, 2008). *If $\langle f, g \rangle_{\mathcal{H}} < \infty$ then the following isometry holds,*

$$\langle f, g \rangle_{\mathcal{H}} = C^2(H) \langle \mathcal{D}_*^{-(H-\frac{1}{2})} f, \mathcal{D}_*^{-(H-\frac{1}{2})} g \rangle,$$

where the constant $C(H)$ is given by,

$$C^2(H) = \frac{H(2H-1)\Gamma\left(\frac{3}{2}-H\right)}{\Gamma\left(H-\frac{1}{2}\right)\Gamma\left(2-2H\right)}.$$

This is important as it justifies that both sides of the equality in (3.26) are well defined random variables. Start the proof by considering simple functions μ_n of the form

$$\mu_n(x) = \sum_{i=1}^{n} c_i \mathbb{1}_{[x_{i-1}, x_i)}(x)$$

where $\{0 = x_0 < x_1 < \ldots < x_n = 1\}$ is a finite partition of $[0,1]$ and $\{c_i : 1 \leq i \leq n\}$ are constants. We also consider simple functions such that,

$$|\mu_n(x)| \leq |\mu(x)| \quad \text{and} \quad \lim_{n\to\infty} \mu_n(x) = \mu(x)$$

for all $x \in [0,1]$. The left hand side of (3.26) becomes,

$$\begin{aligned} LHS &= \int_0^1 \mu_n(x)\, dB_H(x) \\ &= \int_0^1 \sum_{i=0}^n c_i \mathbb{1}_{[x_{i-1},x_i)}(x)\, dB_H(x) \\ &= \sum_{i=0}^n c_i (B_H(x_i) - B_H(x_{i-1})). \end{aligned} \tag{3.27}$$

Using Proposition 3.3 and the linearity of $\mathcal{D}_*^{-(H-\frac{1}{2})}$ the right hand side of (3.26) becomes,

$$\begin{aligned} RHS &= C_H \int_0^1 \mathcal{D}_*^{-(H-\frac{1}{2})} \mu_n(x)\, dB(x) \\ &= C_H \int_0^1 \mathcal{D}_*^{-(H-\frac{1}{2})} \sum_{i=0}^n c_i \mathbb{1}_{[x_{i-1},x_i)}(x)\, dB(x) \\ &= C_H \sum_{i=0}^n c_i \int_0^1 \mathcal{D}_*^{-(H-\frac{1}{2})} \mathbb{1}_{[x_{i-1},x_i)}(x)\, dB(x) \\ &= C_H \sum_{i=0}^n c_i \int_0^1 \mathcal{D}_*^{-(H-\frac{1}{2})} (\mathbb{1}_{(0,x_i)}(x) - \mathbb{1}_{(0,x_{i-1})}(x))\, dB(x) \\ &= C_H \sum_{i=0}^n c_i \left\{ \int_0^1 \mathcal{D}_*^{-(H-\frac{1}{2})} \mathbb{1}_{(0,x_i)}(x)\, dB(x) - \int_0^1 \mathcal{D}_*^{-(H-\frac{1}{2})} \mathbb{1}_{(0,x_{i-1})}(x)\, dB(x) \right\} \\ &= C_H \sum_{i=0}^n \frac{c_i}{\Gamma(H+\frac{1}{2})} \left\{ \int_0^1 (x_i - x)_+^{H-\frac{1}{2}}\, dB(x) - \int_0^1 (x_{i-1} - x)_+^{H-\frac{1}{2}}\, dB(x) \right\} \\ &= \sum_{i=0}^n c_i (B_H(x_i) - B_H(x_{i-1})). \end{aligned} \tag{3.28}$$

The last line follows using Definition 2.4. Now decompose the function μ into its positive and negative parts: $\mu(x) = \mu^+(x) - \mu^-(x)$ where $\mu^+(x) = \mu(x) \vee 0$ and $\mu^-(x) = -(\mu(x) \wedge 0)$ and $x \wedge y := \min\{x,y\}$. Do a similar procedure for $\mu_n(x) = \mu_n^+(x) - \mu_n^-(x)$, then apply the Monotone Convergence theorem with (3.27) and (3.28),

$$\begin{aligned} \lim_{n\to\infty} \int_0^1 \mu_n(x)\, dB_H(x) &= C_H \lim_{n\to\infty} \int_0^1 \mathcal{D}_*^{-(H-\frac{1}{2})} \mu_n(x)\, dB(x) \\ &= C_H \lim_{n\to\infty} \int_0^1 \mathcal{D}_*^{-(H-\frac{1}{2})} \mu_n^+(x)\, dB(x) - C_H \lim_{n\to\infty} \int_0^1 \mathcal{D}_*^{-(H-\frac{1}{2})} \mu_n^-(x)\, dB(x) \\ &= C_H \int_0^1 \mathcal{D}_*^{-(H-\frac{1}{2})} \mu^+(x)\, dB(x) - C_H \int_0^1 \mathcal{D}_*^{-(H-\frac{1}{2})} \mu^-(x)\, dB(x) \\ &= C_H \int_0^1 \mathcal{D}_*^{-(H-\frac{1}{2})} \mu(x)\, dB(x). \end{aligned} \tag{3.29}$$

With (3.29) in mind, the proof will be complete if it can be shown that,

$$\lim_{n\to\infty} \int_0^1 \mu_n(x)\, dB_H(x) = \int_0^1 \mu(x)\, dB_H(x). \tag{3.30}$$

To show (3.30) holds, apply Proposition 2.2 (Gripenberg and Norros, 1996) (see page 43) and consider the mean-square convergence directly,

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{E}\left|\int_0^1 \mu_n(x)\, dB_H(x) - \int_0^1 \mu(x)\, dB_H(x)\right|^2 &= \lim_{n\to\infty} \mathbb{E}\left|\int_0^1 (\mu_n(x) - \mu(x))\ dB_H(x)\right|^2 \\
&= \lim_{n\to\infty} \langle \mu_n - \mu, \mu_n - \mu\rangle_{\mathcal{H}} \\
&= \lim_{n\to\infty} \left\{\langle \mu_n, \mu_n\rangle_{\mathcal{H}} - 2\langle \mu_n, \mu\rangle_{\mathcal{H}} + \langle \mu, \mu\rangle_{\mathcal{H}}\right\} \\
&= 0. \qquad \square
\end{aligned}$$

The final piece of machinery that is useful for proving the lower bound result is a Girsanov type theorem for fBm. The theorem of Girsanov is concerned with the behaviour of probability measures for Brownian motions with separate drift functions. As was shown in Section 3.2.1, the lower bound can be constructed by comparing the behaviour of two suitably chosen functions f_0 and f_1 from a functional class $\mathscr{F}$. As such, a Girsanov type theorem for fBm is a valuable tool for this and has been found by Biagini, Hu, Øksendal and Zhang (2008, Theorem 3.2.4). This theorem needs some auxiliary definitions which are given first.

Definition 3.5. *Let $(\mathbb{R}, \mathscr{R})$ be a measurable space and $\mathbb{P}_0$ and $\mathbb{P}_1$ be two probability measures on $(\mathbb{R}, \mathscr{R})$. Then we say that $\mathbb{P}_0$ is absolutely continuous with respect to $\mathbb{P}_1$ if whenever $\mathbb{P}_1(A) = 0$ then $\mathbb{P}_0(A) = 0$. We write $\mathbb{P}_0 \ll \mathbb{P}_1$.*

Definition 3.6. *If $\mathbb{P}_0 \ll \mathbb{P}_1$, then there is non-negative function, f, such that,*

$$\mathbb{P}_0(A) = \int_A f\, d\mathbb{P}_1 \quad \text{where } A \in \mathscr{R}.$$

The function f is called the Radon-Nikodymn derivative of $\mathbb{P}_0$ with respect to $\mathbb{P}_1$ and is sometimes denoted,

$$\frac{d\mathbb{P}_0}{d\mathbb{P}_1} := f.$$

Theorem 3.2.4 (Biagini, Hu, Øksendal and Zhang, 2008). *Let $\mathbb{P}_0$ be the probability measure associated with B_H and define the functions $\Delta : [0,1] \longrightarrow \mathbb{R}$ and $\underline{\Delta} : [0,1] \longrightarrow \mathbb{R}$ such that,*

$$\Delta(t) = \int_0^1 \underline{\Delta}(s)\phi(s,t)ds$$

where,

$$\phi(t,s) = \frac{(2-\alpha)(1-\alpha)}{2}|t-s|^{-\alpha}.$$

Then define

$$\mathring{B}_H(t) = B_H(t) + \int_0^t \Delta(s)ds$$

which has an associated probability measure $\mathbb{P}_1$*, then the Radon-Nikodym derivative is,*

$$\begin{aligned}\frac{d\mathbb{P}_1}{d\mathbb{P}_0} &= \exp\left\{-\int_0^1 \underline{\Delta}(x)\,dB_H(x) - \frac{1}{2}\int_0^1\int_0^1 \underline{\Delta}(s)\underline{\Delta}(t)\phi(s,t)dsdt\right\}\\ &= \exp\left\{-\int_0^1 \underline{\Delta}(x)\,dB_H(x) - \frac{1}{2}\langle\underline{\Delta},\underline{\Delta}\rangle_{\mathcal{H}}\right\}.\end{aligned}$$

3.2.4 *Lower bound result*

As alluded to earlier, a slightly different class of functions is used in the lower bound in comparison to the upper bound. This new class of functions is close to the class $\mathscr{C}_s^1([0,1],\theta)$ and is characterised in the Fourier domain and denoted $\mathscr{T}_s(\theta)$. To prove the lower bound result the technique used in Goldenshluger et al. (2006) is modified to the case of direct observations with LRD noise by considering the risk in the Fourier domain. Define this new functional class, $\mathscr{T}_s(\theta)$.

Definition 3.7. *A function* $\mu \in \mathscr{T}_s(\theta)$ *if* $\mu : [0,1] \longrightarrow \mathbb{R}$ *and*

1. $\left|\left[\mu^{(1)}\right](\theta)\right| = a > 0.$
2. *The left and right second derivatives are equal at the change point, that is,*

$$\mu^{(2)}\left(\theta^-\right) = \mu^{(2)}\left(\theta^+\right).$$

3. *The function* $\mu \in \mathscr{L}_2([0,1],\mathbb{R})$ *and satisfies the following condition,*

$$\int_{\mathbb{R}} |\mathcal{F}\mu(\omega)||\omega|^s\,d\omega < \infty.$$

The functional classes $\mathscr{T}_s(\theta)$ and $\mathscr{C}_s^1([0,1],\theta)$ need to be compared to gain insight to the minimax rate result in Theorem 3.1 (see page 18 for the definition of $\mathscr{C}_s^1([0,1],\theta)$). In the forthcoming argument, refer to the defining properties of both classes in Definition 1.16 and Definition 3.7. Clearly, for general $s \geq 2$, property 1. is the same for both functional classes. However, property 3. of $\mathscr{T}_s(\theta)$ is slightly stronger than property 3. and 4. of $\mathscr{C}_s^1([0,1],\theta)$. Indeed, properties 3. and 4. of $\mathscr{C}_2([0,1],\theta)$ are the Taylor expansions of the function around its kink location θ. These require uniformly bounded derivatives up to order s. To this end, note that,

$$\begin{aligned}\sup_{x\in[0,1]}\left|\mu^{(s)}(x)\right| &= \sup_{x\in[0,1]}\left|\mathcal{F}^{-1}\mathcal{F}\mu^{(s)}(x)\right| \\ &= \sup_{x\in[0,1]}\left|\int_{\mathbb{R}}\mathcal{F}\mu^{(s)}(\omega)e^{2\pi ix\omega}\ d\omega\right| \\ &= \sup_{x\in[0,1]}\left|\int_{\mathbb{R}}(2\pi i\omega)^s\,\widetilde{\mu}(\omega)e^{2\pi ix\omega}\ d\omega\right| \\ &\leq (2\pi)^s\int_{\mathbb{R}}|\omega|^s\,|\widetilde{\mu}(\omega)|\ d\omega. \end{aligned} \tag{3.31}$$

This proves the claim that property 3. of $\mathscr{T}_s(\theta)$ is slightly stronger than the uniformly bounded derivative condition in the Taylor expansions in $\mathscr{C}_s^1([0,1],\theta)$.

On the other hand, property 2. of $\mathscr{C}_s^1([0,1],\theta)$ requires that all the intermediate derivatives, $\mu^{(j)}$, have equal left and right limits for $j=2,3,\ldots,s-1$ while property 2. of $\mathscr{T}_s(\theta)$ only requires that the second derivative satisfies that condition.

Therefore, one functional class is not contained in the other functional class and the minimax result will apply to functions that are in both classes. These functions, $\mu\in\mathscr{M}_s(\theta):=\mathscr{C}_s^1([0,1],\theta)\cap\mathscr{T}_s(\theta)$ and are defined as follows.

Definition 3.8. *A function $\mu\in\mathscr{M}_s(\theta)$ if,*

1. $\left|\left[\mu^{(1)}\right](\theta)\right| = a > 0.$
2. *The left and right higher order derivatives are equal at the change point, that is,*

$$\mu^{(j)}\left(\theta^-\right) = \mu^{(j)}\left(\theta^+\right) \quad \text{for } j=2,3,\ldots,s-1.$$

3. *The function $\mu\in\mathscr{L}_2([0,1],\mathbb{R})$ and satisfies the following condition,*

$$\int_{\mathbb{R}}|\mathcal{F}\mu(\omega)||\omega|^s\,d\omega<\infty.$$

For Corollary 3.1, the class of functions that have m kinks need to be given with a Fourier representation. These will be given by

Definition 3.9. *Let $m\geq 1$ and $s\geq 2$ be integers and $\mathbf{a}=(a_1,a_2,\ldots,a_m)\in\mathbb{R}^m$ be fixed such that, $a_i\in\mathbb{R}\setminus\{0\}$ and $\boldsymbol{\theta}=(\theta_1,\theta_2,\ldots,\theta_m)$ with $\theta_i\neq\theta_j$ for $i\neq j$ and $\theta_i\in(0,1)$ for all $i=1,2,\ldots,m$. Then, we say that $\mu\in\mathscr{T}_s^m(\boldsymbol{\theta})$ if $\mu:[0,1]\longrightarrow\mathbb{R}$ and*

1. *μ has m kinks, that is, for all $i=1,2,\ldots,m$ we have,*

$$\left[\mu^{(1)}\right](\theta_i) = \mu^{(1)}(\theta_i^+) - \mu^{(1)}(\theta_i^-) = a_i \neq 0.$$

2. *The left and right second derivatives are equal at all change points, that is,*

$$\mu^{(2)}\left(\theta_i^-\right) = \mu^{(2)}\left(\theta_i^+\right) \qquad \text{for } i = 1, 2, \ldots, m \text{ and } j = 2, 3, \ldots, s-1.$$

3. *The function $\mu \in \mathscr{L}_2\left([0,1], \mathbb{R}\right)$ and satisfies the following condition,*

$$\int_{\mathbb{R}} |\mathcal{F}\mu(\omega)||\omega|^s \, d\omega < \infty.$$

The lower bound result is now stated and proved for the functional class $\mathscr{T}_s\left(\theta\right)$. Which combined with the above argument will prove the minimax result for the class $\mathscr{M}_s\left(\theta\right)$.

Theorem 3.3. *Suppose $\mu \in \mathscr{T}_s\left(\theta\right)$ is observed from the model (2.2) and also assume that $\alpha > 0$ is bounded away from zero such that there exists a k with $0 < k < \alpha \leq 1$, then, there exists a positive constant $c_* < \infty$ that does not depend on n such that the lower rate of convergence for an estimator for the kink location θ with the square loss is of the form,*

$$\liminf_{n\to\infty} \inf_{\widehat{\theta}_n} R^*_{\mathscr{T}_s}\left(\widehat{\theta}_n, n^{-\alpha s/(2s+\alpha)}\right) \geq c^*.$$

Proof of Theorem 3.3 . The proof uses the methodology outlined in Section 3.2.1 by considering the Kullback-Leibler divergence between two particular choices of functions $\mu_0, \mu_1 \in \mathscr{T}_s\left(\theta\right)$. The Kullback-Leibler divergence is defined in terms of absolutely continuous probability measures below.

Definition 3.10. *If two probability measures are absolutely continuous with respect to each other, that is, $\mathbb{P}_0 \ll \mathbb{P}_1$ and $\mathbb{P}_1 \ll \mathbb{P}_0$ (see Definition 3.5), the Kullback-Leibler divergence of two probability measures is defined to be,*

$$\begin{aligned}\mathcal{K}\left(\mathbb{P}_0, \mathbb{P}_1\right) &:= \int \ln \frac{d\mathbb{P}_0}{d\mathbb{P}_1} d\mathbb{P}_0 \\ &= \mathbb{E}_{\mathbb{P}_0}\left\{\ln \frac{d\mathbb{P}_0}{d\mathbb{P}_1}\right\},\end{aligned}$$

where $\frac{d\mathbb{P}_0}{d\mathbb{P}_1}$ is the Radon-Nikodym derivative of measure $\mathbb{P}_0$ with respect to measure $\mathbb{P}_1$.

This is then linked to the lower rate of convergence by

Theorem 2.2 (iii). (Tsybakov, 2009). *If $\mathbb{P}_0, \mathbb{P}_1$ are two probability measures on $(\mathbb{R}, \mathscr{B})$ and the Kullback-Leibler divergence $\mathcal{K}(\mathbb{P}_1, \mathbb{P}_0) \leq C < \infty$ for some $C > 0$ then,*

$$\inf_{\zeta} \max_{j=0,1} P\left(\zeta \neq j\right) \geq \max\left(\tfrac{1}{4}\exp(-C), \tfrac{1-\sqrt{\frac{C}{2}}}{2}\right).$$

Consider a function $\mu_0 \in \mathscr{T}_s(\theta_0)$ where $\theta_0 \in (0, \frac{1}{2})$. Then define, $\theta_1 = \theta_0 + \delta$ where $\delta \in (\theta_0, 1-\theta_0)$. There is no loss of generality imposed by this, a symmetric argument can be setup to accommodate the case when $\theta_0 \in [\frac{1}{2}, 1)$. Define the functions $v : [0,1] \longrightarrow \mathbb{R}$ and $v_N : [0,1] \longrightarrow \mathbb{R}$ such that

$$v(x) := a((\theta_1 \wedge x) - \theta_0)\mathbb{1}_{(\theta_0,1]}(x), \qquad v_N(x) := \int_{-N}^{N} \tilde{v}(\omega)e^{2\pi i x\omega}\, d\omega,$$

where a is the size of the jump given in Definition 3.7. Note that, $v_N(x)$ is close to $v(x)$ in the sense that it is the inverse Fourier transform of $\tilde{v}(\omega)\mathbb{1}_{|\omega|\leq N}$. Also, note with these defintions that the derivative,

$$v^{(1)}(x) = a\mathbb{1}_{[\theta_0,\theta_1]}(x).$$

The function $(\mu_0 - v)$ has a single kink at θ_1 since $[\mu_0^{(1)} - v^{(1)}](\theta_1) = -a$. Then define $\mu_1 := \mu_0 - (v - v_N)$. The function v_N is infinitely differentiable across the whole real line and smooth for finite N, which implies that $\mu_1 = \mu_0 - (v - v_N)$ has a single change-point in the derivative at θ_1 and $[\mu_1^{(1)}](\theta_1) = -a$. Consider the behaviour of v in the Fourier domain,

$$\begin{aligned}
\tilde{v}\,(\omega) &= \mathcal{F}\mathcal{D}^{-1}\mathcal{D}v\,(\omega) \\
&= \mathcal{F}\mathcal{D}^{-1}v^{(1)}\,(\omega) \\
&= (2\pi i\omega)^{-1}\,\mathcal{F}v^{(1)}\,(\omega) \\
&= (2\pi i\omega)^{-1}\,a\int_{\mathbb{R}} \mathbb{1}_{[\theta_0,\theta_1]}\,(x)\,e^{-2\pi i\omega x}dx \\
&= (2\pi i\omega)^{-1}\,a\int_{\theta_0}^{\theta_1} e^{-2\pi i\omega x}dx \\
&= a\,(2\pi i\omega)^{-2}\left(e^{-2\pi i\omega\theta_0} - e^{-2\pi i\omega\theta_1}\right) \\
&= a\,(2\pi i\omega)^{-2}\,e^{-2\pi i\omega\theta_0}\left(1 - e^{-2\pi i\omega\delta}\right).
\end{aligned}$$

So the modulus of $\tilde{v}$ is,

$$\begin{aligned}
|\tilde{v}(\omega)| &= \frac{a}{(2\pi)^2\,|\omega|^2}\left|1 - e^{-2\pi i\omega\delta}\right| \\
&\leq \frac{a}{(2\pi)^2\,|\omega|^2}\,|2\pi\omega\delta| \\
&= \frac{a\delta}{2\pi\,|\omega|}. \qquad (3.32)
\end{aligned}$$

We now investigate the required conditions on N and consequently v_N to ensure that $\mu_1 \in \mathscr{T}_s(\theta_1)$. Using (3.32) we see that,

$$\begin{aligned}
\int_{\mathbb{R}} |\mathcal{F}v_N(\omega)|\, |\omega|^s\, d\omega &= \int_{-N}^{N} |\widetilde{v}(\omega)|\, |\omega|^s\, d\omega \\
&\leq \frac{a\delta}{2\pi} \int_{-N}^{N} |\omega|^{s-1}\, d\omega \\
&= \frac{a\delta}{\pi} \int_{0}^{N} \omega^{s-1}\, d\omega \\
&= \frac{a\delta}{\pi s} N^s,
\end{aligned}$$

which is bounded if N is chosen to be $N = \left(\frac{s\pi C}{a\delta}\right)^{1/s}$ where C is a finite positive constant. Then by (3.31), the same choice of N will ensure that v_N has order s bounded derivatives. By definition $v^{(2)}(x) = 0$ almost everywhere which implies that $\mu_1^{(2)} = \mu_0^{(2)} + v_N^{(2)}$ almost everywhere as well. Consequently it follows that, $\mu_1 \in \mathscr{T}_s^1(\theta)$ since,

$$\begin{aligned}
\int_{\mathbb{R}} |\mathcal{F}\mu_1(\omega)|\, |\omega|^s\, d\omega &\leq \int_{\mathbb{R}} |\mathcal{F}\mu_0(\omega)|\, |\omega|^s\, d\omega + \int_{\mathbb{R}} |\mathcal{F}v_N(\omega)|\, |\omega|^s\, d\omega \\
&\leq C < \infty,
\end{aligned}$$

for some positive constant C. To be able to exploit the fractional Girsanov Theorem, define $\Delta = \mu_0 - \mu_1 = v - v_N$ and it can be seen that $\Delta\colon [0,1] \longrightarrow \mathbb{R}$. The fractional Girsanov Theorem also needs a paired function $\underline{\Delta}\colon [0,1] \longrightarrow \mathbb{R}$. Define such a function,

$$\underline{\Delta}(t) = \varepsilon^{-\alpha} C^{-2}(H) \mathcal{D}_*^{H-1/2} \mathcal{D}^{H-1/2} \Delta(t),$$

where $C(H)$ is given in Lemma 3.1.2 (Biagini, Hu, Øksendal and Zhang, 2008) (see page 74). This definition will ensure that the following equality holds, at least in the distributional sense,

$$\Delta(t) = \varepsilon^{\alpha} \int_0^1 \underline{\Delta}(s)\phi(s,t)ds, \tag{3.33}$$

where $\phi(t,s) = 1/2(2-\alpha)(1-\alpha)|t-s|^{-\alpha}$. The justification for this claim follows. Apply the fractional integral operators to the definition of $\underline{\Delta}$,

$$\Delta(t) = \varepsilon^{\alpha} C^2(H) \mathcal{D}^{-(H-1/2)} \mathcal{D}_*^{-(H-1/2)} \underline{\Delta}(t).$$

Then consider the inner product $\langle \underline{\Delta}, \Delta\rangle$ and apply Lemma 3.1.2 (Biagini, Hu, Øksendal and Zhang, 2008) (see page 74),

$$\begin{aligned}
\varepsilon^{-\alpha}\langle \underline{\Delta}, \Delta\rangle &= C^2(H) \langle \underline{\Delta}, \mathcal{D}^{-(H-1/2)} \mathcal{D}_*^{-(H-1/2)} \underline{\Delta}\rangle \\
&= C^2(H) \langle \mathcal{D}_*^{-(H-1/2)} \underline{\Delta}, \mathcal{D}_*^{-(H-1/2)} \underline{\Delta}\rangle \\
&= \langle \underline{\Delta}, \underline{\Delta}\rangle_{\mathcal{H}}.
\end{aligned}$$

So in the distributional sense, (3.33) is valid.

To evaluate the behaviour of $\underline{\Delta}$ in the Fourier domain use Definition 3.1 and Proposition 3.1, so $\underline{\Delta}$ can be written in terms of the fractional integral operator and its adjoint,

$$\begin{aligned}
\varepsilon^{-\alpha}\Delta(t) &= \int_0^1 \underline{\Delta}(s)\phi(s,t)ds \\
&= \frac{(2-\alpha)(1-\alpha)\Gamma(1-\alpha)}{2}\frac{1}{\Gamma(1-\alpha)}\int_0^1 \underline{\Delta}(s)|t-s|^{-\alpha}ds \\
&= \frac{\Gamma(3-\alpha)}{2\Gamma(1-\alpha)}\left\{\int_0^t \underline{\Delta}(s)(t-s)^{-\alpha}ds + \int_t^1 \underline{\Delta}(s)(s-t)^{-\alpha}ds\right\} \\
&= \frac{\Gamma(3-\alpha)}{2}\left\{\mathcal{D}^{-(1-\alpha)}\underline{\Delta}(t) + \mathcal{D}_*^{-(1-\alpha)}\underline{\Delta}(t)\right\}.
\end{aligned}$$

Change into the Fourier domain and apply Proposition 3.2,

$$\begin{aligned}
\varepsilon^{-\alpha}\mathcal{F}\Delta(\omega) &= \frac{\Gamma(3-\alpha)}{2}\left\{\mathcal{F}\mathcal{D}^{-(1-\alpha)}\underline{\Delta}(\omega) + \mathcal{F}\mathcal{D}_*^{-(1-\alpha)}\underline{\Delta}(\omega)\right\} \\
&= \frac{\Gamma(3-\alpha)}{2}\left\{(2\pi i\omega)^{-(1-\alpha)}\widetilde{\underline{\Delta}}(\omega) + (-2\pi i\omega)^{-(1-\alpha)}\widetilde{\underline{\Delta}}(\omega)\right\} \\
&= \frac{\Gamma(3-\alpha)}{2}\widetilde{\underline{\Delta}}(\omega)\left\{(2\pi i\omega)^{-(1-\alpha)} + (-2\pi i\omega)^{-(1-\alpha)}\right\}.
\end{aligned}$$

Recall, by assumption in Theorem 3.3, there exists a positive value k such that $0 < k < \alpha$. Therefore, there is a bounded positive constant C_k that depends on k such that $\left|\widetilde{\underline{\Delta}}(\omega)\right|$ is bounded,

$$\begin{aligned}
\widetilde{\underline{\Delta}}(\omega) &= \frac{C\varepsilon^{-\alpha}\widetilde{\Delta}(\omega)}{(2\pi i\omega)^{-(1-\alpha)} + (-2\pi i\omega)^{-(1-\alpha)}} \\
|\widetilde{\underline{\Delta}}(\omega)| &= \frac{C\varepsilon^{-\alpha}|\widetilde{\Delta}(\omega)|}{|\omega^{-(1-\alpha)}(1+(-1)^{-(1-\alpha)})|} \\
|\widetilde{\underline{\Delta}}(\omega)| &\leq \frac{C_k\varepsilon^{-\alpha}|\widetilde{\Delta}(\omega)|}{|\omega|^{-(1-\alpha)}} \\
|\widetilde{\underline{\Delta}}(\omega)|^2 &\leq C_k^2\varepsilon^{-2\alpha}|\widetilde{\Delta}(\omega)|^2|\omega|^{2-2\alpha}.
\end{aligned} \tag{3.34}$$

Also, there is a simple bound for $\widetilde{\Delta}$ given by,

$$\begin{aligned}
\widetilde{\Delta}(\omega) &= \mathcal{F}(v - v_N)(\omega) \\
&= \widetilde{v}(\omega) - \widetilde{v}(\omega)\mathbb{1}_{\{|\omega|\leq N\}} \\
&= \widetilde{v}(\omega)\mathbb{1}_{\{|\omega|\geq N\}}.
\end{aligned} \tag{3.35}$$

So, by (3.34) and (3.35) the bound can be further simplified,

$$\begin{aligned} |\widetilde{\underline{\Delta}}(\omega)|^2 &\leq C_k \varepsilon^{-2\alpha} |\widetilde{v}(\omega)|^2 \mathbb{1}_{\{|\omega|\geq N\}} |\omega|^{2-2\alpha} \\ &\leq C_k a^2 \delta^2 \varepsilon^{-2\alpha} |\omega|^{-2\alpha} \mathbb{1}_{\{|\omega|\geq N\}}. \end{aligned} \tag{3.36}$$

The last inequality follows by (3.32). Now let $\mathbb{P}_0$ and $\mathbb{P}_1$ be the probability measures associated with model under μ_0 and μ_1 respectively. That is, $\mathbb{P}_0$ is the measure associated with,

$$dY_0(x) = \mu_0(x)dx + \varepsilon^\alpha \, dB_H(x)$$

and $\mathbb{P}_1$ is the measure associated with,

$$dY_1(x) = \mu_1(x)dx + \varepsilon^\alpha \, dB_H(x).$$

Define, $\mathring{B}_H(x) := \varepsilon^{-\alpha}\mathcal{D}^{-1}\Delta(x) + B_H(x)$ and then under the $\mathbb{P}_0$ measure,

$$\begin{aligned} dY_0(x) &= \mu_0(x)dx + \varepsilon^\alpha \, dB_H(x), \\ Y_0(x) &= \mathcal{D}^{-1}\mu_0(x) + \varepsilon^\alpha B_H(x) \\ &= \mathcal{D}^{-1}\mu_1(x) + \varepsilon^\alpha \left(\frac{\mathcal{D}^{-1}\Delta(x)}{\varepsilon^\alpha} + B_H(x) \right) \\ &= \mathcal{D}^{-1}\mu_1(x) + \varepsilon^\alpha \mathring{B}_H(x) \\ dY_0 &= \mu_1(x)dx + \varepsilon^\alpha \, d\mathring{B}_H(x). \end{aligned}$$

Thus the model Y_0 can be modified to be the same as the model Y_1 if the measure is used under the $\mathring{B}_H(x)$ process. Apply Theorem 3.2.4 (Biagini, Hu, Øksendal and Zhang, 2008). (see page 76) with Proposition 2.2 (Gripenberg and Norros, 1996) (see page 43). It follows that the Radon-Nikodym derivative is as follows,

$$\frac{d\mathbb{P}_1}{d\mathbb{P}_0} = \exp\left\{ -\int_0^1 \underline{\Delta}(x)\, dB_H(x) - \frac{1}{2}\mathbb{E}_{\mathbb{P}_0}\left(\int_0^1 \underline{\Delta}(x)\, dB_H(x) \right)^2 \right\}. \tag{3.37}$$

Recall $H = 1 - \alpha/2$ and apply Proposition 3.5 to (3.37),

$$\begin{aligned} \ln \frac{d\mathbb{P}_0}{d\mathbb{P}_1} &= C_H \int_0^1 \left(\mathcal{D}_*^{-(H-1/2)} \underline{\Delta} \right)(x)\, dB(x) + \frac{C_H^2}{2} \mathbb{E}_{\mathbb{P}_0} \left(\int_0^1 \left(\mathcal{D}_*^{-(H-1/2)} \underline{\Delta} \right)(x)\, dB(x) \right)^2 \\ &= C_H \int_0^1 \left(\mathcal{D}_*^{-(1-\alpha)/2} \underline{\Delta} \right)(x)\, dB(x) + \frac{C_H^2}{2} \int_0^1 \left(\left(\mathcal{D}_*^{-(1-\alpha)/2} \underline{\Delta} \right)(x) \right)^2 dx. \end{aligned} \tag{3.38}$$

The above integral is well defined. Indeed, by virtue of Lemma 3.1.2 (Biagini, Hu, Øksendal and Zhang, 2008) (see page 74),

$$C^2(H)\langle \mathcal{D}_*^{-(H-1/2)}\underline{\Delta}, \mathcal{D}_*^{-(H-1/2)}\underline{\Delta}\rangle = \langle \underline{\Delta}, \underline{\Delta}\rangle_{\mathcal{H}} < \infty.$$

So the Kullback-Leibler divergence can be evaluated using (3.38) with the Plancherel identity, Proposition 3.2 and (3.36),

$$\begin{aligned}
\mathcal{K}(\mathbb{P}_0, \mathbb{P}_1) &= \mathbb{E}_{\mathbb{P}_0} \ln \frac{d\mathbb{P}_0}{d\mathbb{P}_1} \\
&= \frac{C_H^2}{2} \int_0^1 \left| \left(\mathcal{D}_*^{-(1-\alpha)/2}\underline{\Delta}\right)(x)\right|^2 dx \\
&= \frac{C_H^2}{2} \int_{\mathbb{R}} \left| \mathcal{F}\left(\mathcal{D}_*^{-(1-\alpha)/2}\underline{\Delta}\right)(\omega)\right|^2 d\omega \\
&= \frac{C_H^2}{2} \int_{\mathbb{R}} \left|(-2\pi i\omega)^{-(1-\alpha)/2}\right|^2 \left|\widetilde{\underline{\Delta}}(\omega)\right|^2 d\omega \\
&\le C_k \int_{\mathbb{R}} |\omega|^{\alpha-1} \left|\widetilde{\underline{\Delta}}(\omega)\right|^2 d\omega \\
&\le C_k a^2 \delta^2 \varepsilon^{-2\alpha} \int_{|\omega|\ge N} |\omega|^{-\alpha-1}\, d\omega \\
&= C_k \delta^2 \varepsilon^{-2\alpha} \int_{\omega\ge N} \omega^{-\alpha-1}\, d\omega \\
&= C_k \delta^2 \varepsilon^{-2\alpha} N^{-\alpha} \\
&= C_k \delta^2 \varepsilon^{-2\alpha} \left(\frac{s\pi C}{a\delta}\right)^{-\alpha/s} \\
&= C_k \delta^{(2s+\alpha)/s} \varepsilon^{-2\alpha}.
\end{aligned}$$

Now choose $\delta \asymp \varepsilon^{2\alpha s/(2s+\alpha)}$ which guarantees that $\mathcal{K}(\mathbb{P}_0, \mathbb{P}_1) \le C_k < \infty$ for some finite positive constant C_k. Then by Theorem 2.2 (iii). (Tsybakov, 2009) (see page 79) combined with the fact that $\varepsilon \asymp n^{-1/2}$ it follows that the lower rate of convergence for the minimax risk is $\varepsilon^{2\alpha s/(2s+\alpha)} \asymp n^{-\alpha s/(2s+\alpha)}$. □

Part IV

RANDOM DESIGN WITH LRD ERROR VARIABLES

4 RANDOM DESIGN WITH LRD ERRORS

In this chapter the primary problem is to extend the kink estimation method that was constructed for the fixed design model in Part II to the random design model. In particular, for the random design framework with i.i.d. design variables and LRD error variables. Similar to Part II, work on this topic has already been published in Wishart and Kulik (2010), although full treatment is given here with tightened results and more generality.

Similar to Part II, the ZCT is described and adapted further to deal with the random design framework. The random design kink estimation method in this chapter is described over three sections. A review of literature on nonparametric regression in the random design context is given in Section 4.1. The particular assumptions on the model are in Section 4.2. The adaptation of the ZCT to estimate kink locations for the random design model is given in Section 4.3. To help with the presentation, a demonstration of the random design extension is given in Section 4.3.3.

4.1 RANDOM DESIGN REVIEW

Let $\{X_i\}_{i=1}^n$ be i.i.d. random variables independent of $\{\varepsilon_i\}_{i=1}^n$. As in Part II, $\{\varepsilon_i\}_{i=1}^n$ are assumed to be a sequence of LRD error variables. Assume that a bivariate dataset $\{(X_i, Y_i)\}_{i=1}^n$ is observed that follows the regression model,

$$Y_i = \mu(X_i) + \sigma(X_i)\varepsilon_i, \tag{4.1}$$

where the design variables X_i are supported on $\mathcal{X} \subseteq \mathbb{R}$ with regression function, $\mu : \mathcal{X} \longrightarrow \mathbb{R}$ and scale function $\sigma : \mathcal{X} \longrightarrow \mathbb{R}^+$.

It is worth noting that this might not be the most relevant model in practice. Instead, it might be more feasible to consider a model with dependence in the design variables. However, model (4.1) with i.i.d. design variables and LRD errors is a good exercise and first step before analysing the opposite scenario with LRD design variables and i.i.d. errors which is a more difficult framework and is considered separately in the final part of the thesis in Part V.

For the kink estimation procedure, the ZCT described in Part II will again be pursued here for the random design model (4.1). However, before proceeding further, the differences between model (4.1) and other nonparametric and parametric models and their link to the theory given in Part II are examined.

Firstly, (4.1) can be thought of as an extension to the fixed design models given in Part II in the sense that the design points are no longer restricted to a grid of points and the scale function $\sigma(\cdot)$ allows heteroskedasticity 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 is balanced on the behaviour of the scale function and on the level of dependence in the errors. This delicate asymptotic behaviour will become evident and discussed in the analysis section in Section 4.3.4. Before we reach that stage, a review of the existing literature for this topic is covered.

There is an extensive treatment in the literature on both parametric and nonparametric methods for regression models with a random design framework that assume i.i.d. design variables and error variables. The methodologies used include but are not limited to kernel smoothing, wavelet decompositions and orthogonal series. Some of these articles have established optimal methods in the minimax sense (see Fan (1992); Antoniadis, Grégoire and Vial (1997); Müller (1997); Baraud (2002); Zhang, Wong and Zheng (2002); Bertin (2004); Birgé (2004); Kerkyacharian and Picard (2004); Chesneau (2007); Kohler (2008)). The methods of change point estimation have also been considered in the i.i.d. random design framework by Korostelëv (1987); Korostelëv and Tsybakov (1993); Gijbels, Hall and Kneip (1999); Huh and Park (2004); Park and Kim (2006).

4.2 MODEL ASSUMPTIONS

The specific assumptions on the structure of the random design model defined (4.1) will be outlined in this section. There will be some common definitions and notation that overlap for both this part and Part V.

4.2.1 *Common random design definitions and notation*

Denote $F = F_X$ to be the cumulative distribution function of X and denote the empirical distribution function of X by $F_n(x) := n^{-1}\sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}}$. Also let $Q = F^{-1}$ and $Q_n = F_n^{-1}$ be the quantile and empirical quantile functions respectively. Finally, we need to impose some mild restrictions on σ. For technical reasons that will be revealed in the proofs of the forthcoming convergence results in Section 4.3.4, assume σ is bounded away from zero and infinity in the sense that,

$$0 < \inf_{t \in \mathcal{X}} \sigma(t) < \sup_{t \in \mathcal{X}} \sigma(t) < \infty \tag{4.2}$$

and that $\sigma \in \mathscr{S}_r$ where $r \geq 3$. In both Parts a specific type of LRD random variable will be assumed to hold. This type of random variable is called a *causal LRD linear process* and is defined below.

Definition 4.1. *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 (see Appendix A) and $\alpha \in (0, 1]$. 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 $\mathbb{E}\eta_t^2 = \left(\sum_{j=0}^{\infty} c_j^2\right)^{-1} =: \sigma_\eta^2$.

It is worth mentioning, that due to the structure of c_i coefficients in Definition 4.1, a causal LRD process does not have bounded support, it is supported across the whole real line. As a side issue, it is often claimed in the literature that the asymptotic covariance structure of a causal LRD linear process satisfying Definition 4.1 can be evaluated by 'simple routine' calculations based on the theorem of Karamata (see Theorem A.1 of Appendix A). Namely, it is written that given a causal LRD linear process, $\{\xi_i\}_{i=1}^n$, the asymptotic covariance structure is

$$\operatorname{Cov}(\xi_i, \xi_{i+n}) \sim C_\alpha n^{-\alpha} L^2(n).$$

This result is true but does not follow from routine applications of the theorem of Karamata. In fact the asymptotic structure is much more delicate and a proof is given in Appendix B along with other similar results.

4.2.2 *LRD error model assumptions*

The specific details of the i.i.d. design and LRD error variables that are used in this chapter are now stated and encapsulated by,

- *Assumption (E)*

  The error variables $\{\varepsilon_i\}_{i=1}^n$ are assumed to be derived from a causal LRD linear process satisfying Definition 4.1 with parameter $\alpha_\varepsilon \in (0, 1]$, unit variance and mean zero.

  The design variables, $\{X_i\}_{i=1}^n$ are i.i.d. random variables with support $\mathcal{X} \subseteq \mathbb{R}$ and common density f_X such that $f_X(x) > 0$ for all $x \in \mathcal{X}$ and $\sup_{x \in \mathcal{X}} |f_X^{(s \wedge r)}(x)| < \infty$. Furthermore, the random variables $\{\varepsilon_i\}_{i=1}^n$ are assumed to be independent of

$\{X_i\}_{i=1}^n$. Also assume that $f_X(x) > 0$ for all $x \in \mathbb{R}$ with the added condition that there exists an $a, b \in (0,1)$ such that $0 < a < 1 - b < 1$ and

$$\inf_{Q(a)<x<Q(1-b)} f_X(x) > 0.$$

For Assumption (E), define the associated set of σ-fields,

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

The strictly positive constraint on f_X given in Assumption (E) implies that F is strictly increasing as well.

4.3 RANDOM DESIGN KINK ESTIMATION METHOD

The main result of this section is concerned with the construction and analysis of an estimator, $\widehat{\theta}_n$, of the kink location θ. The theoretical result with a brief discussion is given in Section 4.3.1. The extension of the ZCT used in Part II to the random design framework is given in Section 4.3.2. To ease the representation of the method, a visual demonstration of the random design extension is given in Section 4.3.3. The asymptotic convergence results and large deviation results are given in Section 4.3.4. The final method and proof of the result is given in Section 4.3.5.

In Part II the stochastic analysis of the method was substantially easier since an asymptotic model was considered. There have been papers that show that there is an asymptotic equivalence between random design models and the white noise model. For example, it has been shown by Brown, Cai, Low and Zhang (2002) and Reiß (2008) that there exists an asymptotic equivalence between model (4.1) and (2.10) (the *white noise model*) when $\sigma(\cdot) \equiv C$, the design variables are i.i.d. uniform random variables and the error variables are i.i.d. and and independent of the design variables. However, this is not the case for general i.i.d. design points and $\sigma : \mathcal{X} \longrightarrow \mathbb{R}$.

Therefore, since there is no argument that can justify the use of an asymptotic model in the general random design framework considered here, the finite random design model (4.1) is used for all the analysis in this part (and in Part V). This makes the analysis more difficult.

4.3.1 *Convergence result for LRD errors*

The main result of this chapter is concerned with the construction and analysis of an estimator, $\widehat{\theta}_n$, of the kink location θ. The estimator $\widehat{\theta}_n$ is constructed such that it convergences to the true kink location θ in the sense that it satisfies the probabilistic bound in the theorem below.

Theorem 4.1. *Suppose a bivariate sequence of observations* $\{X_i, Y_i\}_{i=1}^n$ *that follow model* (4.1) *and satisfy Assumption (E) are observed such that* $\mu \in \mathscr{C}_s^1(\mathcal{X}, \theta)$ *and* $\sigma \in \mathscr{S}_r$ *where* $s \wedge r \geq 3$. *Then an estimator,* $\widehat{\theta}_n$ *of the change point,* θ, *can be constructed such that,*

$$\left|\widehat{\theta}_n - \theta\right| = \mathcal{O}_p\left(n^{-s/(2s+1)}\right).$$

The minimax optimality of this result is not pursued in this thesis. However, parallels can be drawn between this result and the minimax results in the literature for the fixed design case. Recall from Part III that the minimax rate for kink estimation from the *fractional white noise* model (2.2) is $\rho_n = n^{-\alpha s/(2s+\alpha)}$. In particular, the case of fixed design with i.i.d. errors occurs when $\alpha = 1$ and results in the rate $\rho_n = n^{-s/(2s+1)}$. This rate coincides with the rate in Theorem 4.1 which suggests that it could be reasonable to conjecture that the rate is in fact optimal in the minimax sense.

4.3.2 *Random design extension of* ZCT

To be able to use the ZCT, a proxy is needed for the probe functional given in the fixed design setting by

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

In the fixed design framework it was assumed without loss of generality that the support of the design variables is [0,1]. This assumption has been relaxed here and it is assumed that $\mu \in \mathscr{C}_s^1(\mathcal{X}, \theta)$ with $\mathcal{X} \subseteq \mathbb{R}$. Then $\kappa_h(t)$ is estimated in the random design setting by considering,

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

This quantity is an unbiased estimator of $\mathcal{K}_h(t) := \kappa_h(t, \mu_F)$ where μ_F is the regression function of μ with a rescaled domain.

$$\begin{aligned}
\mathbb{E}\widehat{\mathcal{K}}_h(t) &= h^{-4}\mathbb{E}\mu(X_1)K_3\left(\frac{F(X_1) - t}{h}\right) \\
&= h^{-4} \int_{\mathbb{R}} \mu(u) K_3\left(\frac{F(u) - t}{h}\right) dF(u) \\
&= h^{-4} \int_0^1 \mu_F(x) K_3\left(\frac{x-t}{h}\right) dx = \mathcal{K}_h(t) \qquad (4.3)
\end{aligned}$$

where $\mu_F(\cdot) = \mu(F^{-1}(\cdot))$. Therefore, if $\mu \in \mathscr{C}_s^1(\mathcal{X}, \theta)$, then $\mu_F \in \mathscr{C}_s^1([0,1], \lambda)$ where $\theta = F^{-1}(\lambda)$. In (4.3), the observed quantity is the smoothed third derivative of μ_F and the method applied in Section 2.2 to the regression function μ can be applied here to the modified regression function μ_F.

With the above argument in mind an estimator of $\widehat{\theta}_n$ is constructed for the kink location of μ in the random design setting that is approximately the same as the estimator for kink location λ of μ_F in the regular design setting. This is done by estimating the value of λ by $\widehat{\lambda}_n$ using the established ZCT in the regular design setting. Then $\widehat{\lambda}_n$ is rescaled by the quantile function to obtain an estimate of θ. This extra step will be referred to as the *quantile rescaling step*.

Thus to assess the performance of our estimator we need to check that the convergence of $\widehat{\mathcal{K}_h}(t)$ to $\mathcal{K}_h(t)$ is sufficiently fast. Consider the two following processes,

$$\begin{aligned}\gamma_i(t) &= \mu(X_i)K_3\left(\frac{F(X_i)-t}{h}\right),\\ \zeta_i(t) &= \sigma(X_i)K_3\left(\frac{F(X_i)-t}{h}\right).\end{aligned} \tag{4.4}$$

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

$$\widehat{\mathcal{K}_h}(t) = \mathcal{K}_h(t) + \mathcal{Z}_h(t) + \mathcal{b}_h(t), \tag{4.5}$$

where $\mathcal{Z}_h(t)$ and $\mathcal{b}_h(t)$ represent the respective stochastic error and stochastic bias contributions to the estimator and are given by,

$$\begin{aligned}\mathcal{Z}_h(t) &= n^{-1}h^{-4}\sum_{i=1}^{n}\zeta_i(t)\varepsilon_i.\\ \mathcal{b}_h(t) &= n^{-1}h^{-4}\sum_{i=1}^{n}\left(\gamma_i(t)-\mathbb{E}\gamma_1(t)\right).\end{aligned} \tag{4.6}$$

The analysis of the above terms are given in the Section 4.3.4. However before proceeding with the in-depth analysis, the random design extension is demonstrated in Section 4.3.3 on a simulated dataset to illustrate the procedure with the quantile scaling extension.

4.3.3 *Demonstration of the random design method with LRD errors*

Before proceeding into the asymptotic analysis of all the stochastic terms inherent in this random design framework, an illustration of the method in this setting is shown in Figure 13 (see page 94) and discussed to give the reader a heuristic outline of the approach. Similar to the numerical study in Section 2.3, the R software (see R Development Core Team, 2010), was used to conduct the simulation with the added R-package `fracdiff` to simulate the LRD variables. The actual variables that are simulated using the `fracdiff` package are a fARIMA process which is slightly different to our assumption of a causal LRD linear process given in Assumption (E). In particular, the variables simulated in our use of `fracdiff` are a fARIMA$(0, d, 0)$ process where $d = \frac{1-\alpha}{2}$. To describe this process and justify its use, first define the Gamma function denoted by,

$$\Gamma(z) := \int_0^\infty x^{z-1} e^{-x}\, dx.$$

Then by (Hosking, 1981, Theorem 1 (a)) a fARIMA$(0, d, 0)$ process, $\{e_i\}_{i=1}^n$ has a linear process representation,

$$e_i = \sum_{j=0}^{\infty} c_j \eta_{i-j},$$

where

$$c_j = \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(d)} \sim C_d j^{d-1} = C_\alpha j^{-(1+\alpha)/2}$$

and $\{\eta_i\}_{i=0}^\infty$ are a sequence of i.i.d. latent random variables. Notice that the definition of e_i is very close to the definition of a causal LRD linear process given in Definition 4.1. The only minor discrepancy arises in the linear coefficients which are $c_j = j^{-(1+\alpha)/2} L(j)$ in Definition 4.1 while $c_j \sim j^{-(1+\alpha)/2}$ in the fARIMA$(0, d, 0)$ representation given above. Therefore, the fARIMA process generated by `fracdiff` seems a very worthy proxy for purposes of this demonstration.

A random design regression model was simulated with $n = 1024$ observations. The regression function used is similar to (2.9), however the kink location has been moved to occur at $\theta = 1/2$ instead of $\theta = 2/3$.

The LRD error variables were simulated using the R-package `fracdiff` (see Fraley et al., 2009) with the choice of dependence parameter $\alpha = 1/2$. The realisation of the error variables is shown in Figure 13 (a). The design variables are i.i.d. and are simulated from a beta distribution. This type of random variable is defined in terms of its the probability density function below,

$$f_X(x; \beta_1, \beta_2) = \begin{cases} \frac{\Gamma(\beta_1+\beta_2)}{\Gamma(\beta_1)\Gamma(\beta_2)} x^{\beta_1-1}(1-x)^{\beta_2-1}, & \text{if } x \in (0,1) \\ 0, & \text{otherwise,} \end{cases} \tag{4.7}$$

where β_1 and β_2 are parameters that affect the shape and scale of the distribution. The beta distribution was chosen as the design distribution with parameters, $\beta_1 = 1.75$ and $\beta_2 = 1.25$. This choice of parameters gives slightly more weight to the right side of the $(0, 1)$ interval and is seen in Figure 13 (b). A scatterplot of the simulated random design regression model for Assumption (E) is shown in Figure 13 (c) and it can be seen that due to the density of the design, there are fewer observations where $X_i \in (0, 1/2)$ and more observations for $X_i \in (1/2, 1)$.

The initial kink estimator of λ is calculated using the ZCT on the estimator $\mathcal{K}_h(t)$ after it has been normalised to the quantity $\mathcal{T}_{\widehat{\mathcal{K}}}(t)$ (defined later in Section 4.3.5). This is demonstrated in Figure 13 (d). The analysis determined a value of $\widehat{\lambda}_n = 0.401$ for the zero-crossing time. After rescaling by the order statistics of the design variables an estimate of $\widehat{\theta}_n = 0.497$ is obtained which is close to the true location at $\theta = 0.5$. Thus, one can see that the ZCT can be successfully adapted to the random design framework in practice with the added quantile rescaling step.

4.3.4 *Asymptotic results for LRD errors*

In this section the analysis of the stochastic bias and stochastic error contributions given by (4.6) for the random design extension are analysed. These stochastic terms need to be considered before proceeding to the next stage of the ZCT to ensure that the stochastic contributions do not overwhelm the signal generated by the $\mathcal{K}_h(t)$ term.

The ZCT is a technique that was constructed for use in the fixed design framework. It was shown in Part II that the only stochastic term that needed to be considered was the stochastic integral process, $Z_h(t) = \varepsilon_n^{\alpha_\varepsilon} h^{-4} \int_0^1 K_3\left(\frac{x-t}{h}\right) dB_H(x)$. Here, the method is being extended to the random design framework and the equivalent process of $Z_h(t)$ in the random design framework is given by $\mathcal{Z}_h(t)$, the discrete analogue of the stochastic integral which also includes the scale function $\sigma \in \mathscr{S}_r(\mathcal{X})$. Furthermore, there is an extra stochastic bias term $\mathcal{b}_h(t)$ which measures the discrepancy between the kernel smoothing approach in both the fixed design and random design frameworks.

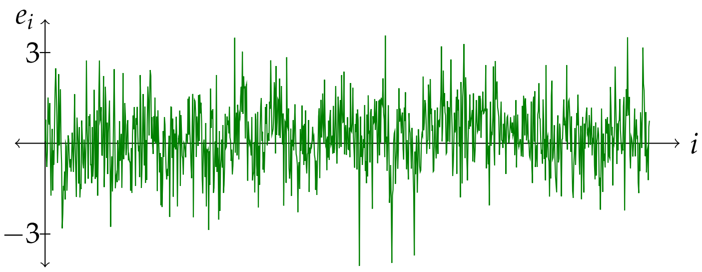
We now state some central limit theorems for the estimator, $\widehat{\mathcal{K}_h}(t)$. The convergence of the estimator $\widehat{\mathcal{K}_h}(t)$ is reliant on a balance between the size of the bandwidth relative to the level of dependence α_ε. The specific details of this relationship between h and n^{α_ε} will be shown in detail inside the Theorems. Roughly speaking, if the bandwidth is too 'large' compared to α_ε, more observations are included for the weighted averages around each point of interest t. This has a knock-on effect that the dependence structure of the errors dominate for each point and the estimator converges to a process that needs to be normed by a sequence that relies on α_ε. Conversely, if the bandwidth is 'small' compared to α_ε then the dependence of the random variables is negligible and the asymptotic behaviour of $\widehat{\mathcal{K}_h}(t)$ behaves similar to the independent scenario and a regular central limit theorem holds with a norming sequence that is not reliant on α_ε. This is shown below in the following Theorem.

Theorem 4.2. *Let $K \in \mathscr{K}_{s\wedge r}$, $\mu \in \mathscr{C}_s^1(\mathcal{X}, \theta)$, $\sigma \in \mathscr{S}_r$ with $s \wedge r \geq 3$ and $t \in (h, 1-h)$. Also if the design variables and error random variables follow Assumption (E) and the bandwidth $h = h(n)$ also satisfies,*

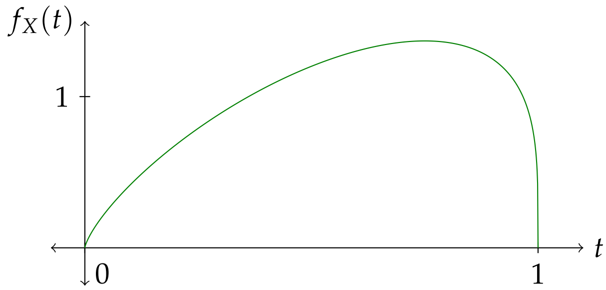
$$h^{2(s\wedge r)+1} n^{1-\alpha_\varepsilon} L^2(n) \to 0 \quad \text{as } n \to \infty, \tag{4.8}$$

then the following convergence result holds,

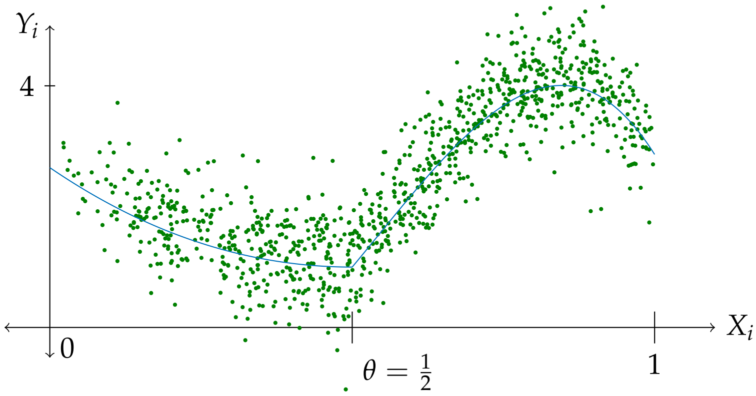
$$\sqrt{nh^7}\left(\widehat{\mathcal{K}_h}(t) - \mathcal{K}_h(t)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, v^2(t)\right), \tag{4.9}$$



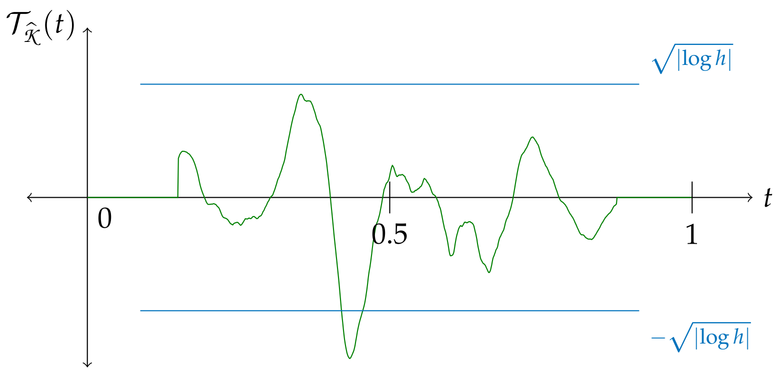
(a) Plot of LRD errors with Gaussian innovations and $\alpha = 1/2$.



(b) Density of the i.i.d. design X_i; a beta distribution with shape parameters $1.75, 1.25$.



(c) Scatterplot of simulated random design regression model.



(d) Bandwidth choice $h = 0.25$, $\widehat{\lambda}_n = 0.401$, with $\widehat{\theta}_n = 0.497$

Figure 13: Demonstration of the kink estimation technique under Assumption (E).

where

$$v^2(t) := \left(\sigma_F^2(t) + \mu_F^2(t)\right) \int_{-1}^{1} K_3^2(x)\, dx.$$

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

$$h^{2(s\wedge r)+1} n^{1-\alpha_\varepsilon} L^2(n) \to \infty \quad \text{as } n \to \infty, \tag{4.10}$$

then,

$$\frac{n^{\frac{\alpha_\varepsilon}{2}} h^{3-(s\wedge r)}}{L(n)} \left(\widehat{\mathcal{K}}_h(t) - \mathcal{K}_h(t)\right) \xrightarrow{\mathscr{D}} \mathcal{N}\left(0, C_1^2 v_*^2(t)\right), \tag{4.11}$$

where

$$v_*(t) := \frac{\sigma_F^{(s\wedge r)}(t)}{(s\wedge r)!} \int_{-1}^{1} x^{s\wedge r} K_3(x)\, dx \quad \text{and} \quad C_1^2 := \frac{2\sigma_\eta^2}{(1+\alpha_\varepsilon)(2+\alpha_\varepsilon)} \int_0^\infty \left(x^2 + x\right)^{-\frac{1+\alpha_\varepsilon}{2}} dx.$$

Recall from Part II that a large deviations result is needed to be able to distinguish between the stochastic contributions and the signal generated by $\mathcal{K}_h(t)$. Beforehand, the assumed structure was the fractional white noise model that included a Gaussian process as the stochastic error term. This was dealt with using the well established classical extreme value theory for Gaussian distributions. However, this is not the case here since neither the design variables nor error variables are assumed to be Gaussian.

Theorem 4.3. *Consider the functions $\mu \in \mathscr{S}_s(\mathcal{X}), \sigma \in \mathscr{S}_r(\mathcal{X})$ and assume $K \in \mathscr{K}_{s\wedge r}$ for some $s \wedge r \geq 3$. If the design and error variables follow Assumption (E) and the bandwidth is chosen such that (4.8) holds. Then the following large deviations result holds,*

$$\sqrt{nh^7} \sup_{t\in(h,1-h)} \left|\widehat{\mathcal{K}}_h(t) - \mathcal{K}_h(t)\right| = \mathcal{O}_p\left(\sqrt{|\log h|}\right).$$

Proof of Theorem 4.2 . To prove the Theorem we appeal to similar results that were shown by Wu and Mielniczuk (2002); Kulik (2008) by decomposing the stochastic terms into two parts, a martingale part and a LRD part. This is done by defining,

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

(see (4.4) for definitions of $\gamma_i(t)$ and $\zeta_i(t)$). Then decompose the standardised probe functional $\widehat{\kappa}_h(t)$ into two terms,

$$\begin{aligned}
\sqrt{nh^7}\left(\widehat{\mathcal{K}}_h(t) - \mathcal{K}_h(t)\right) &= \sqrt{nh^7}\left(\mathcal{Z}_h(t) + \mathscr{b}_h(t)\right) \\
&= \frac{1}{\sqrt{nh}}\left(\sum_{i=1}^n \zeta_i(t)\varepsilon_i + \sum_{i=1}^n (\gamma_i(t) - \mathbb{E}\gamma_1(t))\right) \\
&= \sqrt{h^{-1}\,(\mathbb{V}\mathrm{ar}\zeta_1(t) + \mathbb{V}\mathrm{ar}\gamma_1(t))} \sum_{i=1}^n \chi_i(t) + \frac{\mathbb{E}\zeta_1(t)}{\sqrt{nh}} \sum_{i=1}^n \varepsilon_i.
\end{aligned} \tag{4.12}$$

The Theorem will follow by showing that either the first or last term on the RHS of (4.12) dominates under the bandwidth conditions (4.8) or (4.10) respectively. More specifically, it will be shown that the dominating term will follow a central limit theorem and the other term converges to zero in probability; then Slutsky's Theorem completes the proof.

Firstly consider the case where (4.8) holds and use a martingale central limit theorem. In particular, (Brown, 1971, Theorem 2) will be used and is stated for completeness below.

Theorem 2 (Brown, 1971). *Let $\{M_n, \mathcal{F}_n\}$ be a martingale with $d_n = M_n - M_{n-1}$. Suppose further that:*

1. *The conditional variances of the martingale converge in probability to the unconditional variance, that is,*

$$\frac{1}{\mathbb{E}M_n^2}\sum_{i=1}^n \mathbb{E}\left[d_i^2 \middle| \mathcal{F}_{i-1}\right] \xrightarrow{p} 1;$$

2. *The Lindeberg condition holds. That is, for any $\epsilon > 0$,*

$$\frac{1}{\mathbb{E}M_n^2}\sum_{i=1}^n \mathbb{E}d_i^2 \mathbb{1}_{\left\{|d_i| \geq \epsilon\sqrt{\mathbb{E}M_n^2}\right\}} \xrightarrow{n\to\infty} 0.$$

Then the central limit theorem follows,

$$\frac{1}{\sqrt{\mathbb{E}M_n^2}}\sum_{i=1}^n d_i \xrightarrow{\mathscr{D}} \mathcal{N}(0,1),$$

as $n \to \infty$.

The above Theorem will be applied to the martingale difference sequence $\{\chi_i(t), \mathcal{E}_i\}$. 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 $t \in (h, 1-h)$,

$$\mathbb{E}\zeta_1^2(t) = \int_{\mathbb{R}} \sigma^2(x) K_3^2\left(\frac{F(x)-t}{h}\right) dF(x) = h\int_{-1}^1 \sigma_F^2(t+hu) K_3^2(u)\, du. \tag{4.13}$$

To appropriately bound the above term, a Taylor expansion of σ_F is obtained to exploit the vanishing moment property of (K:4). This requires a generalised chain rule to find the derivatives of the composite function, $\sigma_F(\cdot) = \sigma(F^{-1}(\cdot))$. A generalised chain rule for composite functions exists (see the Faà di Bruno formula from Hernández Encinas et al. (2005) and references therein). Define the composite function $f \circ g(x) := f(g(x))$, then the derivatives are of the form,

$$\frac{d^n}{dx^n} f(g(x)) := \frac{d^n}{dx^n}(f\circ g)(x) = \sum_{\mathbf{k}\in\mathcal{S}_n} \frac{n!}{k_1!k_2!\ldots k_n!}(f^{(k)}\circ g)(x)\prod_{i=1}^n\left(\frac{g^{(i)}(x)}{i!}\right)^{k_i} \tag{4.14}$$

where $\mathcal{S}_n = \{k_i \in \{\mathbb{Z}^+ \cup 0\} : k_1 + 2k_2 + \ldots + nk_n = n\}$ and $k = \sum_{i=1}^n k_i$. Also, through tedious but elementary calculus it can be shown that, the n^{th} derivative of $Q = F^{-1}$ will exist if $f^{(n)}$ exists. Using the assumption that $\sigma \in \mathscr{S}_r$ and the bounded derivatives of the density f_X in Assumption (E), the Taylor expansion can be obtained and most terms vanish due to the vanishing moment property of (K:4),

$$\begin{aligned}\mathbb{E}\zeta_i(t) &= h \int_{-1}^{1} \sigma_F(t+hu) K_3(u)\, du \\ &= \frac{h^{(s\wedge r)+1}}{(s \wedge r)!} \int_{-1}^{1} \sigma_F^{(s\wedge r)}(t+\tau hu) u^{s\wedge r} K_3(u)\, du = h^{(s\wedge r)+1} v_*(t), \end{aligned} \tag{4.15}$$

where $\tau \in (0,1)$. Therefore, using (4.13) and (4.15),

$$\mathbb{V}\text{ar}\zeta_1(t) = h \int_{-1}^{1} \sigma_F^2(t+hu) K_3^2(u)\, du - \frac{h^{2(s\wedge r)+2}}{((s\wedge r)!)^2} \left(\int_{-1}^{1} \sigma_F^{(s\wedge r)}(t+\tau hu) u^{s\wedge r} K_3(u)\, du \right)^2.$$

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

$$\mathbb{V}\text{ar}\zeta_1(t) \geq \frac{h \inf_{x\in\mathbb{R}} \left|\sigma^2(x)\right|}{2} \int_{-1}^{1} K_3^2(u)\, du. \tag{4.16}$$

From (4.15), it follows, $h^{-\frac{1}{2}}\mathbb{E}\zeta_1(t) = o(1)$ and from (4.13), $h^{-1}\mathbb{E}\zeta_1^2(t) \to \sigma_F^2(t) \int_{-1}^{1} K_3^2(u)\, du$. Therefore, $h^{-1}\mathbb{V}\text{ar}\zeta_1(t) = h^{-1}\left(\mathbb{E}\zeta_1^2(t) - (\mathbb{E}\zeta_1(t))^2\right) \to \sigma_F^2(t) \int_{-1}^{1} K_3^2(u)\, du$. Also, the same argument applies for the $\gamma_i(t)$ term to yield,

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

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

$$\begin{aligned}\sum_{i=1}^{n} \mathbb{E}\chi_i^2(t) \mathbb{1}_{\{|\chi_i(t)|>\epsilon\}} &= n\mathbb{E}\chi_1^2(t)\mathbb{1}_{\{|\chi_1(t)|>\epsilon\}} \\ &= \frac{\mathbb{E}\left[\left(\varepsilon_1\left(\zeta_1(t) - \mathbb{E}\zeta_1(t)\right) + \gamma_1(t) - \mathbb{E}\gamma_1(t)\right)^2 \mathbb{1}_{A_n(\epsilon)}\right]}{\mathbb{V}\text{ar}\zeta_1(t) + \mathbb{V}\text{ar}\gamma_1(t)}. \end{aligned} \tag{4.17}$$

where $A_n(\epsilon) = \left\{\left|\varepsilon_1\left(\zeta_1(t) - \mathbb{E}\zeta_1(t)\right) + \gamma_1(t) - \mathbb{E}\gamma_1(t)\right| > \epsilon\sqrt{n\left(\mathbb{V}\text{ar}\zeta_1(t) + \mathbb{V}\text{ar}\gamma_1(t)\right)}\right\}$. The size of this set can be maximised using (4.16),

$$\begin{aligned} A_n(\epsilon) &\subset \left\{2\left|K_3\right|_\infty \left|\varepsilon_1\right| \left(|\sigma|_\infty + |\mu|_\infty\right) > \epsilon\sqrt{n\mathbb{V}\text{ar}\zeta_1(t)}\right\} \\ &\subset \left\{2\left|K_3\right|_\infty \left|\varepsilon_1\right| \left(|\sigma|_\infty + |\mu|_\infty\right) > \epsilon\sqrt{nh\frac{\inf_{x\in\mathbb{R}}\left|\sigma^2(x)\right|}{2} \int_{-1}^{1} K_3^2(u)\, du}\right\}. \end{aligned} \tag{4.18}$$

Using the fact that $nh \to \infty$ and $h \to 0$ as $n \to \infty$ we see that $A_n(\epsilon) \to \emptyset$, the empty set. Consequently with (4.17), (4.18) 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 Lemma B.2 and Lemma B.4 (on pages 151 and 151 respectively), let $\epsilon > 0$ be arbitrary,

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\right| > \epsilon\right) \leq \frac{1}{n^2\epsilon^2}\mathbb{V}\mathrm{ar}\left(\sum_{i=1}^{n}\varepsilon_i\right) \leq \frac{C_1^2 n^{-\alpha_\varepsilon}L^2(n)}{\epsilon^2} = o(1),$$
$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2 - 1\right| > \epsilon\right) \leq \frac{1}{n^2\epsilon^2}\mathbb{V}\mathrm{ar}\left(\sum_{i=1}^{n}\varepsilon_i^2\right) \leq \frac{(C_2^2 n^{-1} \vee C_3^2 n^{-2\alpha_\varepsilon}L^2(n))}{\epsilon^2} = o(1).$$

Then by the above, the sum of the conditional variances converge in probability to one:

$$\sum_{i=1}^{n}\mathbb{E}\left[\chi_i^2(t)\middle|\,\mathcal{E}_{i-1}\right] = \frac{\sum_{i=1}^{n}\varepsilon_i^2\,\mathbb{V}\mathrm{ar}\zeta_1(t) + n\,\mathbb{V}\mathrm{ar}\gamma_1(t) + 2\mathbb{C}\mathrm{ov}\left(\zeta_1(t),\gamma_1(t)\right)\sum_{i=1}^{n}\varepsilon_i}{n\,\left(\mathbb{V}\mathrm{ar}\zeta_1(t) + \mathbb{V}\mathrm{ar}\gamma_1(t)\right)} \xrightarrow{p} 1.$$

Therefore all the conditions of Theorem 2 (Brown, 1971) hold giving,

$$\sum_{i=1}^{n}\chi_i(t) \xrightarrow{\mathscr{D}} \mathcal{N}(0,1). \tag{4.19}$$

Now we show that the last term on the RHS of (4.12) converges in probability to zero. Consider an arbitrary $\epsilon > 0$, then using (4.15) and Lemma B.2 (see page 151),

$$\begin{aligned} P\left(\left|\frac{\mathbb{E}\zeta_1(t)}{\sqrt{nh}}\sum_{i=1}^{n}\varepsilon_i\right| > \epsilon\right) &\leq \frac{\left(\mathbb{E}\zeta_1(t)\right)^2}{\epsilon^2 nh}\mathbb{V}\mathrm{ar}\left(\sum_{i=1}^{n}\varepsilon_i\right) \\ &\leq Ch^{2(s\wedge r)+1}n^{1-\alpha_\varepsilon}L^2(n) \\ &= o(1), \end{aligned}$$

and the last line follows by the bandwidth restriction given in (4.8). Thus, the proof of (4.9) holds with (4.8) and Slutsky's Theorem.

Consider now the claim of (4.11). Using (4.12), (4.19) and (4.15),

$$\widehat{\mathcal{K}_h}(t) - \mathcal{K}_h(t) = \mathcal{O}_p\left(n^{-\frac{1}{2}}h^{-\frac{7}{2}}\right) + \frac{v_*(t)}{nh^{3-(s\wedge r)}}\sum_{i=1}^{n}\varepsilon_i. \tag{4.20}$$

Also, from (Ho and Hsing, 1997, Corollary 3.3), it is known that

$$\frac{1}{n^{1-\frac{\alpha_\varepsilon}{2}}L(n)}\sum_{i=1}^{n}\varepsilon_i \xrightarrow{\mathscr{D}} \mathcal{N}(0, C_1^2). \tag{4.21}$$

Therefore, normalising the expression in (4.20),

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

and the result follows from (4.10) and (4.21) with Slutsky's Theorem. □

Proof of Theorem 4.3 . Start the proof by concentrating on the first term in the summation involving the stochastic bias $\gamma_i(t)$. The $\gamma_i(t)$ terms are independent random variables, each of which have variance that is of order h. Indeed for $t \in (h, 1-h)$,

$$\begin{aligned}
\mathbb{V}\mathrm{ar}\gamma_1(t) &\le \mathbb{E}\gamma_1^2(t) \\
&= \int_{\mathcal{X}} \mu^2(x) K_3^2\left(\frac{F(x)-t}{h}\right) dF(x) \\
&= h \int_{-1}^{1} \mu_F^2(t+hx) K_3^2(x)\, dx = \mathcal{O}(h).
\end{aligned}$$

Therefore by the Law of the Iterated Logarithm for the i.i.d. random variables (see Bingham (1986)) we have the following result,

$$\begin{aligned}
\limsup_{n\to\infty} \frac{1}{\sqrt{nh\log\log n}} \sum_{i=1}^n \left(\gamma_i(t) - \mathbb{E}\gamma_i(t)\right) &= C_1 \qquad \text{a.s.} \\
\liminf_{n\to\infty} \frac{1}{\sqrt{nh\log\log n}} \sum_{i=1}^n \left(\gamma_i(t) - \mathbb{E}\gamma_i(t)\right) &= -C_2 \qquad \text{a.s.}
\end{aligned}$$

where C_1, C_2 are positive constants or equivalently,

$$\frac{1}{\sqrt{nh}} \sup_{t\in(h,1-h)} \left|\sum_{i=1}^n (\gamma_i(t) - \mathbb{E}\gamma_1(t))\right| = \mathcal{O}_{\text{a.s.}}\left(\sqrt{\log\log n}\right) = o_{\text{a.s.}}\left(\sqrt{|\log h|}\right).$$

Shift attention now to the LRD stochastic error contribution $\zeta_i(t)\varepsilon_i$. Begin with a mean corrected version,

$$\frac{1}{\sqrt{nh}} \sup_{t\in(h,1-h)} \left|\sum_{i=1}^n \zeta_i(t)\varepsilon_i\right| \le \frac{1}{\sqrt{nh}} \sup_{t\in(h,1-h)} \left|\sum_{i=1}^n (\zeta_i(t) - \mathbb{E}\zeta_1(t))\,\varepsilon_i\right| + \sup_{t\in(h,1-h)} |\mathbb{E}\zeta_1(t)| \frac{1}{\sqrt{nh}} \left|\sum_{i=1}^n \varepsilon_i\right|$$

Use a similar Taylor expansion argument that was used to show (4.15) to obtain the following bound,

$$\begin{aligned}
\sup_{t\in(h,1-h)} |\mathbb{E}\zeta_1(t)| &= \sup_{t\in(h,1-h)} \left| \int_{\mathbb{R}} \sigma(x) K_3 \left(\frac{F(x)-t}{h} \right) dF(x) \right| \\
&= \sup_{t\in(h,1-h)} \left| h \int_{-\frac{t}{h}}^{\frac{1-t}{h}} \sigma_F(t+hx) K_3(x)\, dx \right| \\
&= \sup_{t\in(h,1-h)} \left| \frac{h^{(s\wedge r)+1}}{(s\wedge r)!} \int_{-1}^{1} \sigma_F^{(s\wedge r)}(t+qhx) x^{(s\wedge r)} K_3(x)\, dx \right| \quad \text{for some } q\in(0,1), \\
&\le \frac{2h^{(s\wedge r)+1}}{(s\wedge r)!} \left| \sigma^{(s\wedge r)} \right|_\infty \int_0^1 x^{(s\wedge r)} |K_3(x)|\, dx.
\end{aligned}$$

By the above, Lemma B.2 (see page 151) and (4.8) imply,

$$\begin{aligned}
\frac{1}{\sqrt{nh}} \sup_{t\in(h,1-h)} \left| \sum_{i=1}^n \zeta_i(t)\varepsilon_i \right| &\le \frac{1}{\sqrt{nh}} \sup_{t\in(h,1-h)} \left| \sum_{i=1}^n \left(\zeta_i(t) - \mathbb{E}\zeta_1(t)\right) \varepsilon_i \right| + \mathcal{O}_p\left(h^{(s\wedge r)+1/2} n^{(1-\alpha_\varepsilon)/2} L(n) \right) \\
&= \frac{1}{\sqrt{nh}} \sup_{t\in(h,1-h)} \left| \sum_{i=1}^n \left(\zeta_i(t) - \mathbb{E}\zeta_1(t)\right) \varepsilon_i \right| + o_p(1) \qquad (4.22)
\end{aligned}$$

To deal with the other LRD term in (4.22), define $d_i(t) := \left(\zeta_i(t) - \mathbb{E}\zeta_1(t)\right)\varepsilon_i$, $M_n(t) = \sum_{i=1}^n d_i(t)$ and the sigma field $\mathcal{A}_i = \sigma(\ldots, \eta_i, \eta_{i+1}; X_1, X_2, \ldots, X_i)$. Then $\{M_n(t), \mathcal{A}_n\}$ is a martingale. To deal with the supremum across t for this martingale, approximate the martingale difference sequence $d_i(t)$ with a discretised version $d_i(t_j)$ for some partition t_j of the interval $(h, 1-h)$. Then the discretised martingale is bounded using an exponential martingale inequality. For the discretisation, define the uniform grid $t_j = j/N$ where $N = n^2$ and $j \in G_N = \{\lceil hN \rceil, \lceil hN \rceil + 1, \ldots, \lfloor (1-h)N \rfloor\}$ (where $\lceil x \rceil := \min\{i \in \mathbb{Z} | i \ge x\}$) so for any $t \in (h, 1-h)$ there exists a $j \in G_N$ such that $t \in [t_j, t_{j+1})$. Then the LRD term of (4.22) can be bounded with,

$$\begin{aligned}
\frac{1}{\sqrt{nh}} \sup_{t\in(h,1-h)} \left| \sum_{i=1}^n \left(\zeta_i(t) - \mathbb{E}\zeta_1(t)\right) \varepsilon_i \right| &\le \frac{1}{\sqrt{nh}} \max_{j\in G_N} \sup_{t\in[t_j,t_{j+1})} \left| \sum_{i=1}^n \left(d_i(t) - d_i(t_j) \right) \right| \\
&\quad + \frac{1}{\sqrt{nh}} \max_{j\in G_N} \left| \sum_{i=1}^n d_i(t_j) \right| \qquad (4.23)
\end{aligned}$$

Starting with the first term on the RHS of (4.23). For any $t \in [t_j, t_{j+1})$ it can be shown that $\left|\zeta_i(t) - \zeta_i(t_j)\right| \le |\sigma|_\infty L_{K_3} h^{-1} n^{-2}$. To show this, a careful consideration of the location of t and t_j are considered with respect to the support of K_3. Define the indicator sets,

$$I_{ij} = \mathbb{1}\left\{\left|\frac{F(X_i)-t_j}{h}-1\right| \leq Cn^{-2}h^{-1}\right\} \cup \left\{\left|\frac{F(X_i)-t_{j+1}}{h}+1\right| \leq Cn^{-2}h^{-1}\right\}$$
$$\mathring{I}_{ij} = \mathbb{1}\left\{\left|\frac{F(X_i)-t_j}{h}\right| \leq 1\right\} \cap \left\{\left|\frac{F(X_i)-t_{j+1}}{h}\right| \leq 1\right\}.$$

The set $\mathring{I}_{ij}$ represents the situations when both $h^{-1}(F(X_i)-t_j)$ and $h^{-1}(F(X_i)-t_{j+1})$ lie in the domain of K_3. The set I_{ij} represents the edge situation when one of the values $\{t_j, t_{j+1}\}$ is near the boundaries inside the interval $[-1,1]$ while the other value is outside the interval $[-1,1]$. Then consider the $\zeta_i(t)$ terms under each of the new sets I_{ij} and $\mathring{I}_{ij}$,

$$\begin{aligned}
&\left|\sigma(X_i)K_3\left(\frac{F(X_i)-t}{h}\right) - \sigma(X_i)K_3\left(\frac{F(X_i)-t_j}{h}\right)\right| \\
&\qquad \leq |\sigma|_\infty \left|\left(K_3\left(\frac{F(X_i)-t}{h}\right) - K_3\left(\frac{F(X_i)-t_j}{h}\right)\right)(I_{ij}+\mathring{I}_{ij})\right| \\
&\qquad \leq |\sigma|_\infty I_{ij}\left|K_3\left(\frac{F(X_i)-t}{h}\right) - K_3\left(\frac{F(X_i)-t_j}{h}\right)\right| \\
&\qquad\qquad + |\sigma|_\infty \mathring{I}_{ij}\left|K_3\left(\frac{F(X_i)-t}{h}\right) - K_3\left(\frac{F(X_i)-t_j}{h}\right)\right|.
\end{aligned} \tag{4.24}$$

Start with the second term on the RHS of (4.24) which concerns the situation when $\mathring{I}_{ij}$ is non-zero. For any $t \in [t_j, t_{j+1})$ there exists a $0 \leq \delta_j < n^{-2}$ such that $t = t_j + \delta_j$ and by the Lipschitz property of K_3 in (K:1),

$$\left|K_3\left(\frac{F(X_i)-t}{h}\right) - K_3\left(\frac{F(X_i)-t_j}{h}\right)\right| \mathring{I}_{ij} \leq L_{K_3}h^{-1}\left|t-t_j\right| \leq Ch^{-1}n^{-2}. \tag{4.25}$$

Shift attention now to the first term on the RHS of (4.24) and consider initially the situation when $0 < \left(1-\frac{F(X_i)-t_j}{h}\right) \leq Cn^{-2}h^{-1}$. A similar argument can be reached for values of $\{t_j, t_{j+1}\}$ near the opposite boundary and consequently over the whole range for I_{ij}. They are omitted for ease of readability. Consider the situation when $h^{-1}(F(X_i)-t) < 1$, then the same argument holds from (4.25). So it remains to consider when $h^{-1}(F(X_i)-t) > 1$ which is outside the support of K_3. Substituting $K_3\left(\frac{F(X_i)-t}{h}\right) = 0 = K_3(1)$ and using a similar argument to before, apply the Lipschitz property of K_3,

$$\left|K_3\left(\frac{F(X_i)-t}{h}\right) - K_3\left(\frac{F(X_i)-t_j}{h}\right)\right| I_{ij} = \left|K_3(1) - K_3\left(\frac{F(X_i)-t_j}{h}\right)\right| \leq Ch^{-1}n^{-2}.$$

Therefore it has been shown that $|\sigma|_\infty I_{ij} \left|K_3\left(\frac{F(X_i)-t}{h}\right) - K_3\left(\frac{F(X_i)-t_j}{h}\right)\right| \le Ch^{-1}n^{-2}$ which combined with (4.25) implies,

$$\begin{aligned}\left|d_i(t) - d_i(t_j)\right| &= |\varepsilon_i| \left|\zeta_i(t) - \zeta_i(t_j) + \mathbb{E}\left[\zeta_1(t_j) - \zeta_1(t)\right]\right| \\ &\le Ch^{-1}n^{-2}\,|\varepsilon_i| \end{aligned} \tag{4.26}$$

To deal with the LRD term $|\varepsilon_i|$, define the two truncated random variables,

$$\breve{\varepsilon}_i := \varepsilon_i \mathbb{1}_{|\varepsilon_i|\ge\sqrt{nh}} - \mathbb{E}\varepsilon_1\mathbb{1}_{|\varepsilon_1|\ge\sqrt{nh}} \quad \text{and} \quad \hat{\varepsilon}_i := \varepsilon_i - \breve{\varepsilon}_i.$$

Now $\mathbb{E}\varepsilon_1 = 0$ and $nh \to \infty$ which implies that $\mathbb{E}\varepsilon_1\mathbb{1}_{|\varepsilon_1|\ge\sqrt{nh}} = o(1)$ and $|\hat{\varepsilon}_1| \le C\sqrt{nh}$. Define as well $\breve{d}_i(t) := \breve{\varepsilon}_i\left(\zeta_i(t) - \mathbb{E}\zeta_1(t)\right)$ and $\hat{d}_i(t) := \hat{\varepsilon}_i\left(\zeta_i(t) - \mathbb{E}\zeta_1(t)\right)$. Then use (4.26) and consider the following moment bounds,

$$\frac{1}{\sqrt{nh}}\mathbb{E}\sup_{t\in[t_j,t_{j+1})}\left|\sum_{i=1}^{n}\left(\breve{d}_i(t) - \breve{d}_i(t_j)\right)\right| \le \frac{C}{\sqrt{n^3h^3}}\mathbb{E}\left|\breve{\varepsilon}_1\right| = \mathcal{O}\left(\frac{1}{\sqrt{n^3h^3}}\right). \tag{4.27}$$

Similarly, using the additional fact that $|\hat{\varepsilon}_i| < C\sqrt{nh}$ it can be shown that,

$$\frac{1}{\sqrt{nh}}\mathbb{E}\sup_{t\in[t_j,t_{j+1})}\left|\sum_{i=1}^{n}\left(\hat{d}_i(t) - \hat{d}_i(t_j)\right)\right| = \mathcal{O}\left(\frac{1}{nh}\right). \tag{4.28}$$

By (4.27), (4.28) and the Markov inequality it is easy to see that,

$$\begin{aligned}\frac{1}{\sqrt{nh\,|\log h|}}\sup_{t\in[t_j,t_{j+1})}\left|\sum_{i=1}^{n}\left(\breve{d}_i(t) - \breve{d}_i(t_j)\right)\right| &= o_p(1) \\ \frac{1}{\sqrt{nh\,|\log h|}}\sup_{t\in[t_j,t_{j+1})}\left|\sum_{i=1}^{n}\left(\hat{d}_i(t) - \hat{d}_i(t_j)\right)\right| &= o_p(1).\end{aligned} \tag{4.29}$$

The result (4.29) ensures that,

$$\frac{1}{\sqrt{nh\,|\log h|}}\sup_{t\in[t_j,t_{j+1})}\left|\sum_{i=1}^{n}\left(d_i(t) - d_i(t_j)\right)\right| = o_p(1).$$

The last term that needs to be dealt with is the second term on the RHS of (4.23). Recall that $\{M_n(t_j), \mathcal{A}_n\}$ form a martingale. The exponential martingale inequality of Freedman (1975) is appealed to.

Theorem (Freedman, 1975). *Let $M_n = \sum_{i=1}^n d_i$ be a martingale with respect to the sigma field $\mathcal{F}_n$ with $Y_n = \sum_{i=1}^n \mathbb{E}\left[d_i^2\middle|\mathcal{F}_{i-1}\right]$ then for all $x, y > 0$,*

$$P\left(M_n \ge x, Y_n \le y \textit{ for some } n\right) \le \exp\left\{-\frac{x^2}{2(x+y)}\right\}.$$

The Theorem of Freedman requires knowledge of the sum of conditional variances. Using Lemma B.4 (see page 151),

$$\begin{aligned}
\sup_{t\in(h,1-h)} \sum_{i=1}^{n} \mathbb{E}\left[d_i^2(t)\middle|\mathcal{A}_{i-1}\right] &= \sup_{t\in(h,1-h)} \mathbb{V}\mathrm{ar}\left(\zeta_1(t)\right)\sum_{i=1}^{n}\varepsilon_i^2 \\
&\le Ch\sum_{i=1}^{n}\varepsilon_i^2 \\
&= \begin{cases} \mathcal{O}_p\left(hn^{1/2}\right), & \text{if } \frac{1}{2}<\alpha_\varepsilon\le 1 \\ \mathcal{O}_p\left(hn^{1/2}\sqrt{L^*(n)}\right), & \text{if } \alpha_\varepsilon=\frac{1}{2} \\ \mathcal{O}_p\left(hn^{1-\alpha_\varepsilon}L^2(n)\right), & \text{if } 0<\alpha_\varepsilon<\frac{1}{2}. \end{cases}
\end{aligned} \tag{4.30}$$

This will lead to three separate applications of Freedman's martingale inequality. Starting with the case $\frac{1}{2}<\alpha_\varepsilon\le 1$. Indeed, let $V_n(t_j)=\sum_{i=1}^{n}\mathbb{E}\left[d_i^2(t_j)\middle|\mathcal{A}_{i-1}\right]$, consider some $C_T>0$, $C_V>0$ and apply Section 4.3.4,

$$\begin{aligned}
&P\left(\max_{j\in G_N} M_n(t_j)\ge C_T\sqrt{nh\left|\log h\right|}, V_n(t_j)\le C_v nh\right) \\
&\le P\left(\bigcup_{j\in G_N}\left\{M_n(t_j)\ge C_T\sqrt{nh\left|\log h\right|}, V_n(t_j)\le C_v nh\right\}\right) \\
&\le \sum_{j\in G_N} P\left(M_n(t_j)\ge C_T\sqrt{nh\left|\log h\right|}, V_n(t_j)\le C_v nh\right) \\
&\le N\exp\left\{-\frac{C_T^2 nh\left|\log h\right|}{2\left(C_T\sqrt{nh\left|\log h\right|}+C_v nh\right)}\right\} \\
&= N\exp\left\{-\frac{C_T^2 nh\left|\log h\right|}{2C_v nh\left(1+\frac{C_T}{C_v}\sqrt{\frac{\left|\log h\right|}{nh}}\right)}\right\} \\
&= N\exp\left\{-\frac{C_T^2}{2C_v}\left|\log h\right|\right\}(1+o(1)) \\
&\le Cn^2h^{C_T^2/(2C_v)}
\end{aligned} \tag{4.31}$$

which can be made arbitrarily small for an appropriately large C_T. Also by (4.30), $V_n(t_j)/(nh)=\mathcal{O}_p\left(n^{-1/2}\right)=o_p(1)$ which in combination with (4.31) implies,

$$P\left(\max_{j\in G_N} M_n(t)\ge C_T\sqrt{nh\left|\log h\right|}\right)\le Cn^2h^{C_T^2/(2C_v)}+o(1),$$

which again can be made arbitrarily small with an appropriately large $C_T > 0$. For the other two cases, notice that $nh > hn^{1/2}\sqrt{L^*(n)}$ and $nh > hn^{1-\alpha_\varepsilon}L(n)$. Therefore the argument will apply to both the other cases giving,

$$\frac{1}{\sqrt{nh}}\max_{j\in G_N}\left|\sum_{i=1}^{n} d_i(t_j)\right| = \mathcal{O}_p\left(\sqrt{|\log h|}\right), \tag{4.32}$$

for $\alpha_\varepsilon \in (0,1]$. The combination of (4.22), (4.23), (4.31) and (4.32) complete the proof. □

4.3.5 *Estimation method for LRD errors*

In this section the specific analysis of the kink estimation method is pursued. A similar three step procedure that was used in the fixed design framework in Section 2.2 is adapted to the random design framework along with the additional quantile rescaling step. As before, the probabilistic bounds and asymptotic behaviour are considered carefully to ensure that the method will obtain a reliable estimate of θ. As was shown in (4.5), the random design estimator has decomposition,

$$\widehat{\mathcal{K}_h}(t) = \mathcal{K}_h(t) + \mathcal{Z}_h(t) + \mathcal{b}_h(t).$$

Using (4.3) along with the familiar expansion of the localisation term and deterministic bias term from Section 2.2.1.1, the probe functional, $\mathcal{K}_h(t) = \kappa_h(t, \mu_F)$ has expansion,

$$\begin{aligned}\mathcal{K}_h(t) &= h^{-2}\left[\mu_F^{(1)}\right](\lambda)K_1\left(\frac{\lambda - t}{h}\right) + \mathcal{J}_h(t) + \mathcal{Z}_h(t) + \mathcal{b}_h(t)\\ &=: \mathcal{L}_h(t) + \mathcal{J}_h(t) + \mathcal{Z}_h(t) + \mathcal{b}_h(t),\end{aligned}$$

where $\mathcal{J}_h(t) = \mathcal{O}\left(h^{s-3}\right)$ takes a similar definition to the fixed design version given in (2.31), namely, μ is substituted for μ_F.

Furthermore, again to avoid trivial complications of the location of the global minimum and maximum generated by $\mathcal{L}_h(t)$, assume that $[\mu_F^{(1)}](\lambda) > 0$. Similar results will hold in the opposite scenario when $[\mu_F^{(1)}](\lambda) < 0$ but are omitted for brevity. The same three step procedure used in Section 2.2 is again applied to first estimate the kink-point λ with estimator $\widehat{\lambda}_n$ and then an additional fourth quantile scaling step is used at the end to map $\widehat{\lambda}_n$ back to $\widehat{\theta}_n$, the estimate of the true change point at θ. Start the procedure with the localisation step.

Localisation step

By using a similar argument to (2.34) for the set $L_\lambda = \{t : |\lambda - t| < h\}$,

$$\mathcal{L}_h(t) = \begin{cases} h^{-2}K_1\left(\frac{\lambda-t}{h}\right)[\mu_F^{(1)}](\lambda), & \text{if } \mu \in \mathscr{C}_s^1([0,1],\lambda) \text{ and } t \in L_\lambda \\ 0, & \text{if } \mu \in \mathscr{S}_s \text{ or } \left\{\mu \in \mathscr{C}_s^1([0,1],\lambda) \text{ and } t \notin L_\lambda\right\}. \end{cases}$$

Also, $\mathcal{L}_h(t)$ has the same unique extrema with a unique global minimum at the point

$$t_* := \arg\min_{t\in(h,1-h)} \mathcal{L}_h(t) = \lambda - q^*h$$

and a unique global maximum at the point

$$t^* := \arg\max_{t\in(h,1-h)} \mathcal{L}_h(t) = \lambda + q^*h.$$

When $|\lambda - t| < h$, $\mathcal{L}_h(t) \geq Ch^{-2} > Ch^{s-3} \geq \mathcal{I}_h(t)$, and $\mathcal{L}_h(t)$ dominates the deterministic signal of $\mathcal{K}_h(t)$.

To construct estimates of the unique global extrema, t_* and t^*, the localisation term $\mathcal{L}_h(t)$ also needs to dominate the stochastic terms, $\mathcal{B}_h(t)$ and $\mathcal{Z}_h(t)$. By virtue of Theorem 4.2, there are two respective bandwidth restrictions to consider, ((4.8) and (4.10)) for the asymptotic behaviour of the estimator under Assumption (E) that correspond to the 'small' and 'large' bandwidth scenarios respectively. The stochastic terms $\mathcal{Z}_h(t) + \mathcal{B}_h(t) = \widehat{\mathcal{K}_h}(t) - \mathcal{K}_h(t)$, so under the assumption of (4.8), apply Theorem 4.2 and

$$\mathcal{Z}_h(t) + \mathcal{B}_h(t) = \mathcal{O}_p\left(n^{-1/2}h^{-7/2}\right). \tag{4.33}$$

Therefore, to have a well defined signal that is dominated by $\mathcal{L}_h(t)$, it is required that $\mathcal{L}_h(t)$ dominates (4.33), $h^{-2} \geq Cn^{-1/2}h^{-7/2}$ or equivalently,

$$h \geq Cn^{-1/3}.$$

Furthermore, since it is assumed that $s \wedge r \geq 3$, to ensure that (4.9) in Theorem 4.2 always holds it is sufficient to ensure that the bandwidth (4.8) always holds. This can be guaranteed if h is chosen such that $h \leq Cn^{-1/7+\alpha_\varepsilon/7-\delta}$, for some $\delta > 0$ or,

$$Cn^{-1/3+\delta} < h < Cn^{-1/7-\delta} \tag{4.34}$$

for some $\delta > 0$. With this optimal bandwidth choice in (4.34), the upper bound ensures that the bandwidth is small enough such that the restriction given by (4.8) holds which in turn ensures that (4.33) holds and the stochastic contribution doesn't depend on α_ε. The lower bound in (4.34) ensures that the bandwidth is not too small and the $\mathcal{L}_h(t)$ signal still dominates the stochastic contribution.

The points t_* and t^* can be estimated with,

$$\widehat{t}_* = \arg\min_{t\in(h,1-h)} \widehat{\mathcal{K}_h}(t) \qquad \text{and} \qquad \widehat{t}^* = \arg\max_{t\in(h,1-h)} \widehat{\mathcal{K}_h}(t).$$

By this construction, the interval $\widehat{A}_h = (\widehat{t}_*, \widehat{t}^*)$ has a length which is order h and contains λ with high probability.

It is worth noting that under this choice, the order of the stochastic terms does not involve α_ε, the level of dependence. Note that h is chosen in a very similar manner if ε_i and X_i, $i \geq 1$, are i.i.d. Consequently, there will be no influence of the LRD on the change point estimation (at least asymptotically).

Kink detection step

As before, to ensure the signal generated by $\mathcal{K}_h(t)$ is genuine and not an artefact of the noise, some large deviation results concerning the stochastic terms $\mathcal{Z}_h(t) + \mathcal{b}_h(t)$ are needed. First standardise the statistic $\widehat{\mathcal{K}_h}(t)$ to have unit variance with,

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

where $v(t)$ is given in Theorem 4.2. Now consider the large deviations when no kink is present in μ, that is, assume $\mu \in \mathscr{S}_s$. Since (4.8) is guaranteed to hold from (4.34), apply the large deviations result Theorem 4.3 and for any $\epsilon > 0$ there exists a $C_T > 0$ large enough such that,

$$\begin{aligned} &P\left(\sup_{t\in(h,1-h)} \left|\mathcal{T}_{\widehat{\mathcal{K}}}(t)\right| \geq C_T\sqrt{2\left|\log h\right|}\right) \\ &\quad = P\left(\sqrt{nh^7} \sup_{t\in(h,1-h)} \left|\widehat{\mathcal{K}_h}(t) - \mathcal{K}_h(t)\right| \geq \inf_{t\in(h,1-h)} v(t) C_T\sqrt{2\left|\log h\right|}\right) \leq \epsilon \qquad (4.35) \end{aligned}$$

The above statement holds since $\inf_t v(t) > 0$ due to added restriction that $\inf_{x\in X} \sigma(x) > 0$ (see (4.2)). On the other hand, if $\mu_F \in \mathscr{C}_s^1([0,1],\lambda)$, then for $t \in (t_*, t^*)$, $\mathcal{L}_h(t) \geq Ch^{-2}$ and for the same large $C_T > 0$ used to ensure (4.35),

$$\lim_{n\to\infty} P\left(\sup_{t\in(h,1-h)} \left|\mathcal{T}_{\widehat{\mathcal{K}}}(t)\right| \geq C_T\sqrt{2\left|\log h\right|}\right) = 1.$$

That is, whenever $\mu \in \mathscr{S}_s$, $\sup_{t\in(h,1-h)} \left|\mathcal{T}_{\widehat{\mathcal{K}}}(t)\right|$ will diverge to infinity at a rate no faster than $\sqrt{2\left|\log h\right|}$ and a kink is detected when

$$\sup_{t\in(h,1-h)} \left|\mathcal{T}_{\widehat{\mathcal{K}}}(t)\right| \geq \sqrt{2\left|\log h\right|}$$

is observed in practice. If a kink is detected through this procedure, the method proceeds to the zero-crossing step.

Zero-crossing step

The same ZCT is used to refine the interval $\widehat{A}_n = (\widehat{t}_*, \widehat{t}^*)$ down to a singular point $\widehat{\lambda}_n$, the estimate of λ. To avoid repetition the reader is referred to the latter part of Section 2.2 for the complete treatment of the ZCT and just the main results are stated here. The estimate of λ is constructed with,

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

Recall again that $\widehat{\mathcal{K}_h}(t) = \mathcal{L}_h(t) + \mathcal{I}_h(t) + \mathcal{Z}_h(t) + \mathcal{b}_h(t)$ and the best possible accuracy occurs if $\delta > 0$ can be chosen as small as possible such that the following inequalities hold,

$$\delta h^{-3} \geq Ch^{s-3} \quad \text{and} \quad \delta h^{-3} \geq Cn^{-1/2}h^{-7/2}. \tag{4.36}$$

The best possible bandwidth that ensures (4.36) holds along with the required bandwidth condition (4.34) is given by,

$$h_* \asymp n^{-1/(2s+1)}.$$

Now apply Lemma 2.1 with this optimal choice of bandwidth and $\delta = \delta_* \asymp h_*^s$ allows the method to obtain an accuracy of order $n^{-s/(2s+1)}$. More specifically, this proves that the estimator satisfies the probabilistic bound,

$$\left|\widehat{\lambda}_n - \lambda\right| = \mathcal{O}_p\left(n^{-s/(2s+1)}\right). \tag{4.37}$$

One final step remains to rescale $\widehat{\lambda}_n$ to an estimate of the true kink location at θ with the quantile rescaling step.

Quantile rescaling step

Recall that $\theta = F^{-1}(\lambda)$. 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 similarly, one can obtain an estimator of Q using the sample quantile function with order statistics. Namely, given the design, $X_1, X_2, \ldots, X_n$; consider the order statistics, $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$, then the sample quantile function is defined,

$$Q_n(x) := \inf\{t | F_n(t) \leq x\} = X_{(i)} \quad \text{if } \tfrac{i-1}{n} < x \leq \tfrac{i}{n}, \quad i = 1, 2, \ldots, n.$$

Estimate θ by, $\widehat{\theta}_n = Q_n(\widehat{\lambda})$. The rate of convergence of this estimator is evaluated below,

$$\begin{aligned} |\widehat{\theta}_n - \theta| &= |Q_n(\widehat{\lambda}_n) - Q(\lambda)| \\ &\leq |Q_n(\widehat{\lambda}_n) - Q(\widehat{\lambda}_n)| + |Q(\widehat{\lambda}_n) - Q(\lambda)| \end{aligned} \tag{4.38}$$

Under Assumption (E), it follows that the derivative of the Quantile function, $Q(t)$, is uniformly bounded on $t \in (a, 1-b)$ for some constants a, b. Indeed, by assumption, there exists an a, b such that $0 < a < 1 - b < 1$ and

$$\inf_{Q(a) < x < Q(1-b)} f_X(x) > 0.$$

Consequently, the derivative of Q given by $Q^{(1)}(\cdot) = 1/(f(Q(\cdot)))$ is uniformly bounded. Therefore, there exists a $\xi \in (0,1)$ such that,

$$Q(\widehat{\lambda}_n) - Q(\lambda) = (\widehat{\lambda}_n - \lambda)Q^{(1)}(\lambda + \xi(\widehat{\lambda}_n - \lambda)) \leq C(\widehat{\lambda}_n - \lambda)$$

The rate of convergence in (4.38) 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 (E), the quantile process involves independent and identically distributed design variables and for all $t \in (0,1)$,

$$|Q_n(t) - Q(t)| = \mathcal{O}_p(n^{-1/2}) \tag{4.39}$$

(see Csörgő (1983) and references therein for a detailed treatment). Therefore, the proof of Theorem 4.1 follows using (4.39) and (4.37) in (4.38) to yield,

$$|\widehat{\theta}_n - \theta| = \mathcal{O}_p(n^{-s/(2s+1)}).$$

Part V

RANDOM DESIGN WITH LRD DESIGN VARIABLES

5 RANDOM DESIGN WITH LRD DESIGN VARIABLES

In a complimentary situation to Chapter 4, the focus in this chapter is to estimate a kink location from a regression function in the random design framework with LRD design variables and i.i.d. noise. In a similar vein to the previous chapter, the method is an extension of the ZCT described in Part II. Work on this topic has already been published in Wishart and Kulik (2010) and forms the basis for analysis discussed here but a more in depth comprehensive treatment is given with increased generality.

A review of some literature that is concerned with LRD random design regression models specifically with LRD design variables is given in Section 5.1. The specific assumptions on the model are outlined in Section 5.2 with the estimation method finally considered in Section 5.3.

5.1 REVIEW OF STOCHASTIC REGRESSION WITH LRD DESIGN

In this random design case it is assumed that $\{X_i\}_{i=1}^n$ are a sequence of LRD random variables and $\{\varepsilon_i\}_{i=1}^n$ are a sequence of i.i.d. random variables. Furthermore, it is assumed that the sequences $\{X_i\}_{i=1}^n$ and $\{\varepsilon_i\}_{i=1}^n$ are independent of each other and a bivariate dataset $\{(X_i, Y_i)\}_{i=1}^n$ is observed that follows the regression model,

$$Y_i = \mu(X_i) + \sigma(X_i)\varepsilon_i, \tag{5.1}$$

with regression function, $\mu : \mathbb{R} \longrightarrow \mathbb{R}$ and scale function $\sigma : \mathbb{R} \longrightarrow \mathbb{R}^+$. Unless otherwise stated, in this chapter, the observations $\{(X_i, Y_i)\}_{i=1}^n$ are assumed to follow (5.1).

A model with this structure is more applicable in reality in comparison to having the i.i.d. design and LRD error structure considered in Chapter 4. Consider for example, a financial state space model where the design variables, X, represent interest rates and Y represents a financial index such as the ASX 200. Clearly, any current level of interest rates is dependent on previous levels of interest rates implying dependence in the X variables. Also, the unexplained variation in Y from the regression function μ can be modelled by the heteroskedastic scale function σ with i.i.d. errors.

Again, for the kink estimation procedure, the ZCT will be used. However, before proceeding further a review of the literature is conducted on random design regression models with LRD design variables.

As discussed in Section 4.1 there is an extensive treatment in the literature on both parametric and nonparametric methods for regression models with a random design framework that assume i.i.d. design variables and error variables. The reader is referred to that chapter and references within for a general treatment of nonparametric regression estimation in random design models. Here, a review of regression models with LRD design variables is given.

A contemporary analysis on least-squares estimation for a linear regression model with LRD design variables has been conducted by Guo and Koul (2008). A treatment of estimators in a nonparametric regression model with LRD design variables have been considered by Zhao and Wu (2008); Liu and Wu (2010); Kulik and Lorek (2011). The methods of change point estimation have also been considered in the LRD random design framework by Wang (2008) and Wang and Cai (2010).

5.2 LRD DESIGN MODEL ASSUMPTIONS

The model is assumed to still satisfy the common assumptions described in Section 4.2.1. The specific details of the LRD design and i.i.d. error variables that are used in this chapter are stated below.

- *Assumption (X)*

  The error variables $\{\varepsilon_i\}_{i=1}^n$ are assumed to be i.i.d. random variables with mean zero, unit variance and $\mathbb{E}\varepsilon_1^4 < \infty$.

  The design variables $\{X_i\}_{i=1}^n$ are assumed to be derived from a causal LRD linear process satisfying Definition 4.1 with parameter $\alpha_x \in (0,1]$ and independent of $\{\varepsilon_i\}_{i=1}^n$. Furthermore, assume that $f_\eta^{(j)}$ is Lipschitz continuous for $j = 0, 1, \ldots, s$ where f_η is the density of the η_i variables, $\mathbb{E}\eta_1^4 < \infty$. Also assume that $f_X(x) > 0$ for all $x \in \mathbb{R}$ with the added condition that there exists an $a, b \in (0,1)$ such that $0 < a < 1 - b < 1$ and

$$\inf_{Q(a)<x<Q(1-b)} f_X(x) > 0, \tag{5.2}$$

  where $Q = F_X^{-1}$, the quantile function of the design variables.

For Assumption (X), define the associated set of σ-fields,

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

The strictly positive constraint on f_X given in Assumption (X) implies that F is strictly increasing as well.

5.3 ESTIMATION METHOD UNDER LRD DESIGN

The general estimation method for random design with LRD design variables closely parallels the procedure covered in Section 4.3. As such, to avoid needless repetition, the reader is referred to the relevant sections when prompted for a comprehensive treatment. The main discrepancies and extensions between Assumption (E) and Assumption (X) and the effects on the convergence rates in kink estimation will be the focus.

The convergence result is stated first in Section 5.3.1. Similar to Part IV, the method is demonstrated on a simulated random design regression model example that satisfies Assumption (X) in Section 5.3.2. The asymptotic results and large deviations results are given in Section 5.3.3. The estimation method and analysis is given in Section 5.3.4.

5.3.1 *Convergence result for LRD design*

The main result of this section is concerned with the construction and analysis of an estimator, $\widehat{\theta}_n$, of the kink location θ. The estimator $\widehat{\theta}_n$ is constructed such that it convergences to the true kink location θ in the sense that it satisfies the probabilistic bound in the theorem below.

Theorem 5.1. *Suppose a sequence of bivariate observations $\{X_i, Y_i\}_{i=1}^n$ that follow model (5.1) and satisfy Assumption (X) are observed such that $\mu \in \mathscr{C}_s^1(\mathbb{R}, \theta)$ and $\sigma \in \mathscr{S}_r(\mathbb{R})$ where $s \wedge r \geq 3$. Then an estimator, $\widehat{\theta}_n$ of the change point, θ, can be constructed such that,*

$$\left|\widehat{\theta}_n - \theta\right| = \mathcal{O}_p\left(n^{-s/(2s+1)} \vee \left(n^{-\alpha_x/2} L(n)\right)\right).$$

Recall that the convergence result under Assumption (E) given in Theorem 4.1 had a rate $\rho_n = n^{-s/(2s+1)}$. In contrast, the convergence result in Theorem 5.1 relies on α, the level of LRD in the design. Therefore, one can see that, at least in the results displayed here, the effect of LRD is more detrimental to the rate of convergence when present in the design variables as opposed to the error variables.

The estimation method used to obtain the rate in Theorem 5.1 follows the same idea proposed in Section 4.3.2. Namely, consider the estimator,

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

As was shown in (4.5), this random design estimator has decomposition,

$$\begin{aligned}
\widehat{\mathcal{K}_h}(t) &= \mathcal{K}_h(t) + \mathcal{Z}_h(t) + \mathscr{b}_h(t) \\
&= h^{-2}\left[\mu_F^{(1)}\right](\lambda) K_1\left(\frac{\lambda - t}{h}\right) + \mathcal{I}_h(t) + \mathcal{Z}_h(t) + \mathscr{b}_h(t) \\
&:= \mathcal{L}_h(t) + \mathcal{I}_h(t) + \mathcal{Z}_h(t) + \mathscr{b}_h(t).
\end{aligned}$$

The above quantity includes the localisation term, $\mathcal{L}_h(t)$, the deterministic bias term $\mathcal{I}_h(t) = \mathcal{O}\left(h^{s-3}\right)$ and $\mathcal{Z}_h(t)$ and $\mathscr{b}_h(t)$ which are the respective stochastic error and stochastic bias contributions given by (4.6). So to ensure that the estimator $\widehat{\mathcal{K}_h}(t)$ converges sufficiently fast to $\mathcal{L}_h(t)$, the asymptotic behaviours of the stochastic terms, $\mathcal{Z}_h(t)$ and $\mathcal{B}_h(t)$ need to be considered for the case of Assumption (X). Before considering the analysis of these terms, the demonstration of the method in this context is given in the next section.

5.3.2 *Demonstration of the random design method with LRD design*

An illustration of the method in the LRD design context is shown in Figure 14 (see page 115). A similar implementation was used with the R software (see R Development Core Team, 2010), with the added R-package `fracdiff` (see Fraley, Leisch, Maechler, Reisen and Lemonte, 2009) to simulate the LRD variables. The package `fracdiff` actually simulates fARIMA random variables instead of the assumed causal LRD linear process. However, as justified in Section 4.3.3 they are sufficiently close for purposes of demonstration.

A random design regression model was simulated with $n = 1024$ observations. The regression function μ is given by (2.9) (see Figure 2) which has a kink at $\theta = 2/3$.

The LRD design variables were simulated using the R-package `fracdiff` with the choice of dependence parameter $\alpha = 2/3$ with the underlying latent variables generated with a beta distribution (see (4.7)) with parameters $\beta_1 = 2$ and $\beta_2 = 1$ and shifted to have a mean at zero. More specifically, the density of the latent variables is given by $f_\eta(x) = g(x - 2/3; \beta_1 = 2, \beta_2 = 1)$ where g is the probability density function defined in (4.7). The realisation of the error variables is shown in Figure 14 (a).

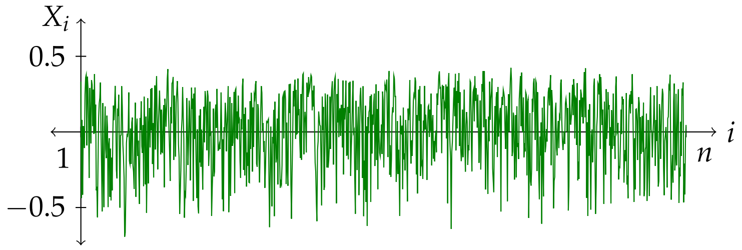
The error variables are i.i.d. and are simulated from a reflected gamma distribution with shape and scale parameters both being 2. Namely, they were generated from the density,

$$f_\epsilon(x) = \tfrac{1}{4}|x|e^{-|x|/2}, \qquad \text{for } x \in \mathbb{R}.$$

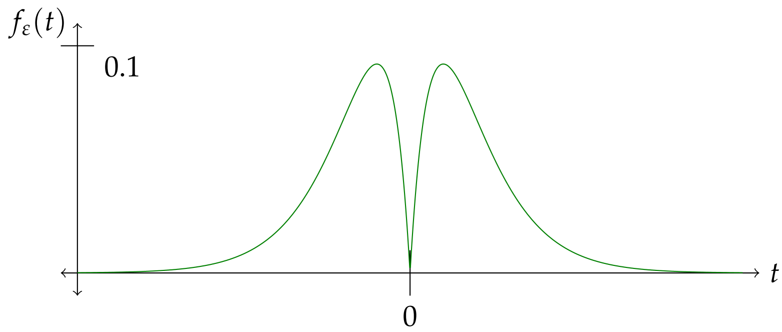
The reflected gamma distribution was chosen since it is non-standard and has more kurtosis than the normal and less clustering around zero.

A scatterplot of the simulated random design regression model for Assumption (X) is shown in Figure 14 (c). For comparison with previous analyses, let $\{\mathring{X}_i\}_{i=1}^n$ be the sequence of rescaled design points such that they all lie in the $[0,1]$ interval. It can be seen that due to the density of the design, there are fewer observations where $\mathring{X}_i$ is near zero and the majority of the design variables lie in the upper half of the interval where $\mathring{X}_i \in (1/2, 1)$.

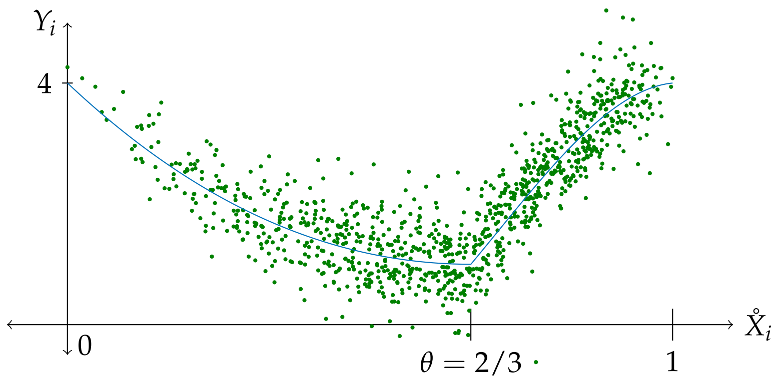
The initial kink estimator of λ is calculated using the ZCT on the estimator $\mathcal{K}_h(t)$ after it has been normalised to the quantity $\mathcal{T}_{\widehat{\mathcal{K}}}(t)$ (defined later in Section 4.3.5). This is demonstrated in Figure 14 (d). The analysis determined a value of $\widehat{\lambda}_n = 0.501$ for the zero-crossing time. After rescaling by the order statistics of the design variables an estimate for $\theta = 2/3$ was obtained with $\widehat{\theta}_n = 0.653$. Thus, again one can see that the ZCT can be successfully adapted to the random design framework with the added quantile rescaling step.



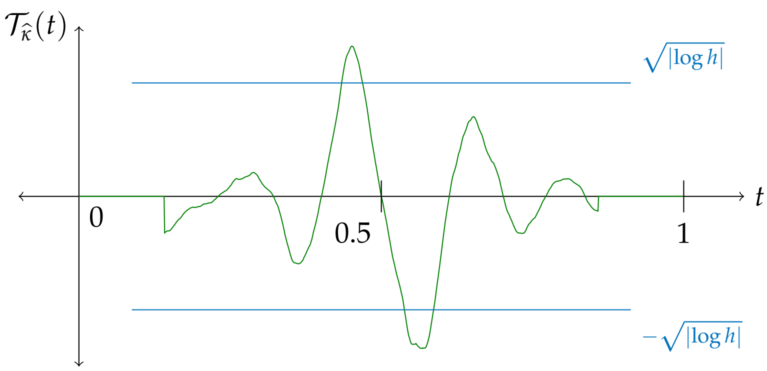
(a) Plot of the simulated LRD design variables, X_i, with $\alpha = 2/3$.



(b) A plot of the density of the error variables, ε_i.



(c) Scatterplot of random design regression model with rescaled design points.



(d) Bandwidth choice $h = 0.28$, $\widehat{\lambda}_n = 0.501$, with $\widehat{\theta}_n = 0.653$

Figure 14: Demonstration of the kink estimation technique under Assumption (X).

5.3.3 *Asymptotic results for LRD design*

In this section the analysis of the stochastic bias and stochastic error contributions given by (4.6) for the random design extension are analysed. These stochastic terms need to be considered before proceeding to the next stage of the ZCT to ensure that the stochastic contributions do not overwhelm the signal generated by the $\mathcal{K}_h(t)$ term.

We now state some central and non-central limit theorems for the estimator, $\widehat{\mathcal{K}_h}(t)$. Similar to Part IV, the convergence of the estimator $\widehat{\mathcal{K}_h}(t)$ is reliant on a balance between the size of the bandwidth relative to the level of dependence α_x where a 'large' bandwidth will cause the asymptotic distribution to rely on α_x. Conversely, if the bandwidth is 'small' compared to α_x then the dependence of the random variables is negligible and the asymptotic behaviour of $\widehat{\mathcal{K}_h}(t)$ behaves similar to the independent scenario and a regular central limit theorem holds with a norming sequence that is not reliant on α_x. This is shown below in the following Theorems.

Theorem 5.2 and Theorem 5.3 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 process.

Theorem 5.2. *Let $K \in \mathscr{K}_{s\wedge r}$, $\mu \in \mathscr{C}_s^1(\mathbb{R}, \theta)$, $\sigma \in \mathscr{S}_r(\mathbb{R})$ with $s \wedge r \geq 3$. If the design points and error random variables follow Assumption (X) and the bandwidth $h = h(n)$ satisfies,*

$$h^7 n^{1-\alpha_x} L^2(n) \to 0 \quad \text{as } n \to \infty, \tag{5.3}$$

then for $t \in (h, 1-h)$ the estimator obeys the following law,

$$\sqrt{nh^7}\left(\widehat{\mathcal{K}_h}(t) - \mathcal{K}_h(t)\right) \xrightarrow{\mathscr{D}} \mathcal{N}\left(0, v^2(t)\right). \tag{5.4}$$

See Theorem 4.2 for the definition of $v(t)$.

In the case of 'large' bandwidth behaviour, (when (5.3) does not hold), the asymptotic behaviour is dominated by the LRD variables. Under Assumption (X), the LRD behaviour is mainly prevalent in the stochastic bias term, $\mathscr{b}_h(t)$, since the stochastic error term, $\mathscr{Z}_h(t)$ involves the i.i.d. error variables ε. The behaviour under this scenario is much more subtle and complex. In the next theorem, the framework is 'simplified' somewhat by assuming the design variables are a causal LRD *Gaussian* process. Then the theory of a Hermite expansion is used to determine the asymptotic behaviour of $\widehat{\mathcal{K}_h}(t)$.

The Hermite expansion is defined in terms of the *Hermite polynomials* which have a close relationship with the Gaussian density and therefore are a ripe candidate to determine the asymptotic behaviour of a Gaussian random variable.

Definition 5.1. *The common definition of the Hermite polynomials uses the differential operator and for an $n \in \{0, 1, 2, \ldots\}$ the n^{th}-order Hermite polynomial, $H_n(x)$, is defined by,*

$$\phi(x) H_n(x) = (-\mathcal{D})^n \phi(x),$$

where ϕ is the standard normal density.

Even under the added simplification that the design variables are a *Gaussian causal LRD linear process*, the asymptotic distribution depends on the so called *Hermite rank* (see Taqqu, 1975) which is defined below.

Definition 5.2. *Let $Z \sim \mathcal{N}(0, 1)$ and define the set of functions*

$$\mathcal{H} := \left\{ f : \mathbb{R} \longrightarrow \mathbb{R} \middle| \mathbb{E} f(Z) = 0, \mathbb{E} f^2(Z) < \infty \right\}.$$

Then a function $f \in \mathcal{H}$ is said to have Hermite rank $q \geq 1$ if,

$$q = \min_{r \in \mathbb{Z}^+ \setminus \{0\}} \{ r : \mathbb{E} f(Z) H_r(Z) \neq 0 \}.$$

Theorem 5.3. *Let $K \in \mathscr{K}_{s \wedge r}$, $\mu \in \mathscr{C}_s^1(\mathbb{R}, \theta)$, $\sigma \in \mathscr{S}_r(\mathbb{R})$ with $s \wedge r \geq 3$. Assume the design points and error random variables follow Assumption (X) and that the design variables are a causal LRD Gaussian process. If $t \in (h, 1-h)$ and the bandwidth $h = h(n)$ satisfies,*

$$h^7 n^{1-\alpha_x} L^2(n) \to \infty \quad \text{as } n \to \infty, \tag{5.5}$$

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

$$\frac{n^{\frac{\alpha_X}{2}}}{L(n)} \left(\widehat{\mathcal{K}_h}(t) - \mathcal{K}_h(t) \right) \xrightarrow{\mathscr{D}} \mathcal{N}(0, \mathcal{C}_1^2(t))$$

where

$$\mathcal{C}_1(t) = \frac{C_1 \mathcal{K}_h(t)}{s_X^2 \sigma_\eta \phi(\Phi^{-1}(t))} \int_{\mathbb{R}} \phi\left(\frac{\Phi^{-1}(t) - u}{s_X} \right) \left(\Phi^{-1}(t) - u \right) \phi\left(\frac{u}{\sigma_\eta} \right) du,$$

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

Remark 5.1. If the estimator $\widehat{\mathcal{K}_h}(t)$ 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 5.3 with the result of (Avram and Taqqu, 1987, Theorem 2) that the normed process

$$n^{q\alpha_x/2} L^{-q}(n) \left(\widehat{\mathcal{K}_h}(t) - \mathcal{K}_h(t) \right) \xrightarrow{\mathscr{D}} \mathcal{C}_q(t) \mathcal{R}_q$$

where the constant $\mathcal{C}_q(t)$ is given by,

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

and $\mathcal{R}_q$ is the Hermite-Rosenblatt process. That is,

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

where B denotes a standard Brownian motion and the constant $C_{q,\alpha}$ is given by,

$$C_{q,\alpha} = \sqrt{\frac{q!(1 - q\alpha)(2 - q\alpha)}{2C_1^{2q}}}.$$

In Avram and Taqqu (1987), they considered Appell polynomials of generalised stationary LRD random variables. In our case the LRD variables are Gaussian and the Appell polynomials reduce to the Hermite polynomials. Otherwise, if $n^{1-q\alpha_x} h^7 L^{2q}(n) \to 0$ then (5.4) holds.

Thus, one can see that depending on the choice of bandwidth, there are different asymptotic behaviours of the estimator $\widehat{\mathcal{K}_h}(t)$. For our purposes, it is desirable that the asymptotic behaviour does not rely on the level of dependence. Recall from Section 4.3 that the random design estimator has decomposition,

$$\begin{aligned}\widehat{\mathcal{K}_h}(t) &= h^{-2} \left[\mu_F^{(1)}\right](\lambda) K_1\left(\frac{\lambda - t}{h}\right) + \mathcal{J}_h(t) + \mathcal{Z}_h(t) + \mathscr{b}_h(t) \\ &:= \mathcal{L}_h(t) + \mathcal{J}_h(t) + \mathcal{Z}_h(t) + \mathscr{b}_h(t),\end{aligned}$$

where the deterministic bias, $\mathcal{J}_h(t) = \mathcal{O}\left(h^{s-3}\right)$. Ideally, it would be beneficial if Theorem 5.2 were to apply to ensure that $\widehat{\mathcal{K}_h}(t) - \mathcal{K}_h(t) = \mathcal{Z}_h(t) + \mathscr{b}_h(t) = \mathcal{O}\left(\sqrt{nh^7}\right)$. This will occur if the bandwidth is chosen such that,

$$Cn^{-1/3+\delta} < h < Cn^{-1/7-\delta}, \qquad \text{for some } \delta > 0. \tag{5.6}$$

Indeed, this guarantees that (5.3) holds and consequently, Theorem 5.2 is to hold as well.

The final asymptotic result that is required is a large deviations result to be able to distinguish between the stochastic contributions and the signal generated by $\mathcal{K}_h(t)$. The large deviations results are given in the same fashion as in Section 4.3.4.

Theorem 5.4. *Assume that $s \wedge r \geq 3$, $\mu \in \mathscr{S}_s(\mathbb{R})$, $\sigma \in \mathscr{S}_r(\mathbb{R})$ with $K \in \mathscr{K}_{s\wedge r}$. Let Assumption (X) hold along with bandwidth condition, (5.6). Then the large deviations result holds,*

$$\sqrt{nh^7} \sup_{t\in(h+a,1-h-b)} \left| \widehat{\mathcal{K}_h}(t) - \mathcal{K}_h(t) \right| = \mathcal{O}_p\left(\sqrt{|\log h|}\right),$$

where the constants a and b depend on the distribution of the design variables and satisfies (5.2).

The dependence structure is more complicated for the case of the design variables. The first complication arises due to the support of the causal LRD random variables being across the whole real line. This will become evident in the proof of the result. To ensure the bound holds it is required that the density of the design variables, is bounded away from zero on some interval compact interval and is shown by condition (5.2). The proof of the result is spread over several Lemmas for ease of presentation. Some of the Lemmas consider a 'mean corrected' version of the stochastic processes. These processes are defined by,

$$\begin{aligned} \gamma_i^*(t) &= (\mu(X_i) - \mu_F(t))\, K_3\left(\frac{F(X_i)-t}{h}\right), \\ \zeta_i^*(t) &= (\sigma(X_i) - \sigma_F(t))\, K_3\left(\frac{F(X_i)-t}{h}\right). \end{aligned} \tag{5.7}$$

Proof of Theorem 5.4 . The proof follows by carefully expressing $\widehat{\mathcal{K}_h}(t) - \mathcal{K}_h(t)$ in a particular decomposition,

$$\begin{aligned} &\sqrt{nh^7}\left(\widehat{\mathcal{K}_h}(t) - \mathcal{K}_h(t)\right) \\ &= \frac{1}{\sqrt{nh}} \sum_{i=1}^n \left(\gamma_i(t) - \mathbb{E}\gamma_1(t) + \zeta_i(t)\varepsilon_i\right) \\ &= \frac{1}{\sqrt{nh}} \sum_{i=1}^n \left(\gamma_i^*(t) - \mathbb{E}\gamma_1(t) + \zeta_i(t)\varepsilon_i + \mu_F(t) K_3\left(\frac{F(X_i)-t}{h}\right)\right) \\ &= \frac{1}{\sqrt{nh}} \sum_{i=1}^n \Bigg(\gamma_i^*(t) - \mathbb{E}\left[\gamma_i^*(t) \middle| \mathcal{F}_{i-1}\right] + \mathbb{E}\left[\gamma_i(t) \middle| \mathcal{F}_{i-1}\right] - \mathbb{E}\gamma_i(t) \\ &\qquad + \mu_F(t)\left(K_3\left(\frac{F(X_i)-t}{h}\right) - \mathbb{E}\left[K_3\left(\frac{F(X_i)-t}{h}\right)\middle| \mathcal{F}_{i-1}\right]\right) \\ &\qquad + \zeta_i^*(t)\varepsilon_i + \sigma_F(t) K_3\left(\frac{F(X_i)-t}{h}\right)\varepsilon_i\Bigg). \end{aligned}$$

Each of these five terms in the decomposition are dealt with separately in Lemmas 5.1 – 5.5. The proof follows by appealing to each of those Lemmas which are stated below and each of their respective proofs given thereafter. □

Lemma 5.1. *Assume that $\sigma \in \mathscr{S}_r(\mathbb{R})$, K_3 is a Lipschitz continuous function and Assumption (X) holds along with (5.6). Then,*

$$\sup_{t\in(h+a,1-h-b)} \left|\sum_{i=1}^{n} \zeta_i^*(t)\varepsilon_i\right| = \mathcal{O}_p\left(\sqrt{nh^3\left|\log h\right|}\right).$$

Lemma 5.2. *Assume that K_3 is a Lipschitz continuous function and Assumption (X) holds along with (5.6). Then,*

$$\sup_{t\in(h+a,1-h-b)} \left|\sum_{i=1}^{n} K_3\left(\frac{F(X_i)-t}{h}\right)\varepsilon_i\right| = \mathcal{O}_p\left(\sqrt{nh\left|\log h\right|}\right).$$

Lemma 5.3. *Assume that $\mu:\mathbb{R}\longrightarrow\mathbb{R}$ with $\left|\mu^{(1)}\right|_\infty < \infty$, K_3 is a Lipschitz continuous function, Assumption (X) holds along with the bandwidth condition (5.6). Then,*

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

Lemma 5.4. *Assume that $\mu \in \mathscr{S}_s(\mathbb{R})$, $K \in \mathscr{K}_{s\wedge r}$ where $s\wedge r \geq 3$ and Assumption (X) holds. Then,*

$$\sup_{t\in(h,1-h)} \left|\sum_{i=1}^{n} \left(\mathbb{E}\left[\gamma_i(t)\middle|\,\mathcal{X}_{i-1}\right] - \mathbb{E}\gamma_1(t)\right)\right| = \mathcal{O}_p\left(h^{s\wedge r+1}n^{1-\alpha_x/2}L(n)\right).$$

Lemma 5.5. *Assume that $\mu \in \mathscr{S}_r(\mathbb{R})$, K_3 is a Lipschitz continuous function and Assumption (X) holds along with (5.6). Then,*

$$\sup_{t\in(h+a,1-h-b)} \left|\sum_{i=1}^{n} \left(K_3\left(\frac{F(X_i)-t}{h}\right) - \mathbb{E}\left[K_3\left(\frac{F(X_i)-t}{h}\right)\middle|\,\mathcal{X}_{i-1}\right]\right)\right| = \mathcal{O}_p\left(\sqrt{nh\left|\log h\right|}\right).$$

The proofs of all the large deviation type results in this section (Lemmas 5.1 – 5.5) are based heavily on the methods employed by Liu and Wu (2010) and modified for our purposes. They involve considering the large deviations of discretised versions of the processes and comparing the distance between the continuous and discrete versions of the processes (using truncated random variables where needed).

Due to their frequent use in the forthcoming proofs of Lemmas 5.1 – 5.5, define the truncated random variables,

$$\begin{aligned} \check{\varepsilon}_i &:= \varepsilon_i \mathbb{1}_{\{|\varepsilon_i|\geq\sqrt{nh}\}} - \mathbb{E}\varepsilon_i \mathbb{1}_{\{|\varepsilon_i|\geq\sqrt{nh}\}}, \\ \widehat{\varepsilon}_i &:= \varepsilon_i - \check{\varepsilon}_i. \end{aligned} \tag{5.8}$$

Proof of Lemma 5.1 . Using the truncated error variables defined in (5.8) split the summation in the Lemma into two parts, $\sum_{i=1}^{n} \zeta_i^*(t)\varepsilon_i = \sum_{i=1}^{n} \zeta_i^*(t)\left(\widehat{\varepsilon}_i + \breve{\varepsilon}_i\right)$, with the large deviations of both being considered separately. Starting with the latter,

$$\sup_{t\in(h,1-h)} \left|\sum_{i=1}^{n} \zeta_i^*(t)\breve{\varepsilon}_i\right| = o_p\left(\sqrt{nh^3 \left|\log h\right|}\right).$$

Indeed, first consider a Taylor expansion of $\sigma \in \mathscr{S}_r\left(\mathbb{R}\right)$ about $\sigma_F(t)$,

$$\begin{aligned}
\zeta_i^*(t) &= K_3\left(\frac{F(X_i)-t}{h}\right)\left(\sigma(X_i) - \sigma_F(t)\right) \\
&= K_3\left(\frac{F(X_i)-t}{h}\right)\left(\sigma_F(h(F(X_i)-t)/h - t) - \sigma_F(t)\right) \\
&= K_3\left(\frac{F(X_i)-t}{h}\right)\left(h\xi_{i,t}\sigma_F^{(1)}(t + q_{i,t}(F(X_i)-t))\right)
\end{aligned}$$

where $q_{i,t} \in (0,1)$ and $\xi_{i,t} = (F(X_i)-t)/h$ with $|\xi_{i,t}| < 1$ since *Support* $K_3 = [-1,1]$. Therefore,

$$\begin{aligned}
\sup_{t\in(h,1-h)} |\zeta_i^*(t)| &= \sup_{t\in(h,1-h)} \left|h\xi_{i,t}K_3\left(\frac{F(X_i)-t}{h}\right)\sigma_F^{(1)}(t + q_{i,t}(F(X_i)-t))\right| \\
&\le h\left|K\right|_\infty \left|\sigma^{(1)}\right|_\infty = Ch.
\end{aligned}$$

Using the above with $\mathbb{E}\varepsilon_1^4 < \infty$ an upper bound can be obtained for the expectation since,

$$\begin{aligned}
\mathbb{E}\sup_{t\in(h,1-h)} \left|\sum_{i=1}^{n} \zeta_i^*(t)\breve{\varepsilon}_i\right| &\le Cnh\mathbb{E}\left|\breve{\varepsilon}_1\right| \\
&= Cnh\mathbb{E}\left|\varepsilon_1 \mathbb{1}_{|\varepsilon_1|\ge\sqrt{nh}} - \mathbb{E}\varepsilon_1\mathbb{1}_{|\varepsilon_1|\ge\sqrt{nh}}\right| \\
&\le Cnh\mathbb{E}\left|\varepsilon_1\right|\mathbb{1}_{|\varepsilon_1|\ge\sqrt{nh}} \\
&\le Cnh\sqrt{P\left(|\varepsilon_1| \ge \sqrt{nh}\right)}, \qquad \text{as } \mathbb{E}\varepsilon_1^4 < \infty, \\
&\le C.
\end{aligned}$$

where the last three lines follow from the Hölder and Chebyshev inequalities. A simple application of the Markov inequality yields,

$$P\left(\sup_{t\in(h,1-h)} \left|\sum_{i=1}^{n} \zeta_i^*(t)\breve{\varepsilon}_i\right| \ge \sqrt{nh^3\left|\log h\right|}\right) = \mathcal{O}\left(\frac{1}{\sqrt{nh^3\left|\log h\right|}}\right) = o(1),$$

where the last statement follows due to (5.6). Now the other half of the decomposition needs to be dealt with. Consider a partition of $t \in (h, 1-h)$ into a grid by defining $t_j := j/N$ where $N = n^2$ and $j \in G_N(a,b)$ where,

$$G_N(a,b) = \{\lceil (h+a)N \rceil, \lceil (h+a)N \rceil + 1, \dots, \lfloor (1-h-b)N \rfloor\}. \tag{5.9}$$

Then for any $t \in (h, 1-h)$, there exists a t_j, t_{j+1} such that, $t \in [t_j, t_{j+1})$ and the decomposition is used,

$$\begin{aligned} \sup_{t\in(h,1-h)} \left| \sum_{i=1}^{n} \zeta_i^*(t)\widehat{\varepsilon}_i \right| \leq \max_{j\in G_N(a,b)} \sup_{t_j \leq t < t_{j+1}} \left| \sum_{i=1}^{n} \left(\zeta_i^*(t) - \zeta_i^*(t_j)\right) \widehat{\varepsilon}_i \right| \\ + \max_{j\in G_N(a,b)} \left| \sum_{i=1}^{n} \zeta_i^*(t_j)\widehat{\varepsilon}_i \right|. \end{aligned} \tag{5.10}$$

To deal with the latter term of (5.10), the exponential martingale inequality of Freedman (1975) is used with the martingale difference sequence constructed below.

Notice that $\{\zeta_i^*(t_j)\widehat{\varepsilon}_i, \mathcal{X}_i\}$ form a martingale difference sequence. Define,

$$X_{i,i-1} := X_i - \eta_i = \mu_X + \sum_{j=1}^{\infty} c_j \eta_{i-j}$$

Then, $X_{i,i-1}$ is $\mathcal{X}_{i-1}$-measurable and the conditional density $f_X(x|\mathcal{X}_{i-1})$ can take the following form,

$$f_X(x|\mathcal{X}_{i-1}) = f_\eta\left(x - X_{i,i-1}\right).$$

(see Lemma C.1 for details). Also, since $Support\, K_3 = [-1, 1]$, at the very least $x \in (Q(t-h), Q(t+h))$ for $K_3\left(\frac{F(x)-t}{h}\right)$ to be non-zero. Then the using the appropriate

a, b from the definition of Assumption (X), the sum of the conditional variances are bounded using (5.2),

$$\begin{aligned}
&\sup_{t\in(h+a,1-h-b)} \sum_{i=1}^{n} \mathbb{E}\left[\left(\zeta_i^*(t)\widehat{\varepsilon}_i\right)^2 \middle| \mathcal{X}_{i-1}\right] \\
&= \mathbb{E}\widehat{\varepsilon}_1^2 \sup_{t\in(h+a,1-h-b)} \sum_{i=1}^{n} \int_{\mathbb{R}} K_3^2\left(\frac{F(u)-t}{h}\right)\left(\sigma(u)-\sigma_F(t)\right)^2 f_X(u|\mathcal{X}_{i-1})\,du \\
&= \mathbb{E}\widehat{\varepsilon}_1^2 \sup_{t\in(h+a,1-h-b)} \sum_{i=1}^{n} \int_{\mathbb{R}} K_3^2\left(\frac{F(u)-t}{h}\right)\left(\sigma_F(F(u)-t+t)-\sigma_F(t)\right)^2 f_\eta(u-X_{i,i-1})\,du \\
&= h\mathbb{E}\widehat{\varepsilon}_1^2 \sup_{t\in(h+a,1-h-b)} \sum_{i=1}^{n} \int_{-t/h}^{(1-t)/h} K_3^2(x)\left(\sigma_F(t+hx)-\sigma_F(t)\right)^2 \frac{f_\eta(Q(t+hx)-X_{i,i-1})}{f_X(Q(t+hx))}\,dx \\
&\le h^3\mathbb{E}\widehat{\varepsilon}_1^2 \sup_{t\in(h+a,1-h-b)} \sum_{i=1}^{n} \int_{-t/h}^{(1-t)/h} x^2K_3^2(x)\left(\sigma_F^{(1)}(t+qhx)\right)^2 \frac{f_\eta(Q(t+hx)-X_{i,i-1})}{f_X(Q(t+hx))}\,dx \\
&\le nh^3 \frac{\mathbb{E}\widehat{\varepsilon}_1^2 \left|\sigma^{(1)}\right|_\infty^2 |f_\eta|_\infty \int_{-1}^{1} K_3^2(x)\,dx}{\inf_{Q(a)<x<Q(1-b)} f_X(x)} =: C_V nh^3. \qquad (5.11)
\end{aligned}$$

Thus apply Theorem (Freedman, 1975) (see page 102) with the choice $y = C_V nh^3$ and $x = C_T\sqrt{nh^3|\log h|}$ for some $C_T > 0$,

$$\begin{aligned}
&P\left(\max_{j\in G_N(a,b)}\left|\sum_{i=1}^{n}\zeta_i^*(t_j)\widehat{\varepsilon}_i\right| \ge C_T\sqrt{nh^3\,|\log h|}\right) \\
&\le P\left(\bigcup_{j\in G_N(a,b)}\left\{\left|\sum_{i=1}^{n}\zeta_i^*(t_j)\widehat{\varepsilon}_i\right| \ge C_T\sqrt{nh^3\,|\log h|}\right\}\right) \\
&\le \sum_{j\in G_N(a,b)} P\left(\left\{\left|\sum_{i=1}^{n}\zeta_i^*(t_j)\widehat{\varepsilon}_i\right| \ge C_T\sqrt{nh^3\,|\log h|}\right\}\right) \\
&\le N\exp\left\{-\frac{C_T^2 nh^3\,|\log h|}{2\left(\sqrt{nh^3\,|\log h|}+C_V nh^3\right)}\right\} \\
&= N\exp\left\{-\frac{C_T^2}{2C_V}\,|\log h|\right\}(1+o(1)) \\
&\le Cn^2h^{C_T^2/(2C_V)}.
\end{aligned}$$

The above bound can be made arbitrarily small for a sufficiently large enough choice for C_T so therefore,

$$\max_{j\in G_N(a,b)}\left|\sum_{i=1}^{n}\zeta_i^*(t_j)\widehat{\varepsilon}_i\right| = \mathcal{O}_p\left(\sqrt{nh^3\,|\log h|}\right). \qquad (5.12)$$

To deal with the first term on the RHS of (5.10), the location of t and t_j are considered with respect to the support of K_3. Define the indicator sets,

$$I_{ij} := \mathbb{1}\left\{\left|\frac{F(X_i)-t_j}{h}-1\right| \le Cn^{-2}h^{-1}\right\} \cup \left\{\left|\frac{F(X_i)-t_{j+1}}{h}+1\right| \le Cn^{-2}h^{-1}\right\}$$
$$\breve{I}_{ij} := \mathbb{1}\left\{\left|\frac{F(X_i)-t_j}{h}\right| \le 1\right\} \cap \left\{\left|\frac{F(X_i)-t_{j+1}}{h}\right| \le 1\right\}. \tag{5.13}$$

The set $\breve{I}_{ij}$ represents the situation when both $h^{-1}(F(X_i)-t_j)$ and $h^{-1}(F(X_i)-t_{j+1})$ lie in $Support\, K_3 = [-1,1]$. The set I_{ij} represents the other possible non-zero situation when one of the values $\{t_j, t_{j+1}\}$ is near the boundaries at 1 or -1 but the other value is outside the domain of K_3. Notice that,

$$\begin{aligned}|\widehat{\varepsilon}_i| &= |\varepsilon_i - \breve{\varepsilon}_i| \\ &\le |\varepsilon_i|\, \mathbb{1}_{\{|\varepsilon_i|\le\sqrt{nh}\}} + \left|\mathbb{E}\varepsilon_i \mathbb{1}_{\{|\varepsilon_i|\ge\sqrt{nh}\}}\right| \\ &\le C\sqrt{nh}. \end{aligned} \tag{5.14}$$

Then consider the term under each of the sets I_{ij} and $\breve{I}_{ij}$,

$$\begin{aligned}\sup_{t_j\le t<t_{j+1}}\left|\sum_{i=1}^n\left(\zeta_i^*(t)-\zeta_i^*(t_j)\right)\widehat{\varepsilon}_i\right| &= \sup_{t_j\le t<t_{j+1}}\left|\sum_{i=1}^n\left(\zeta_i^*(t)-\zeta_i^*(t_j)\right)\left(I_{ij}+\breve{I}_{ij}\right)\widehat{\varepsilon}_i\right| \\ &\le \sup_{t_j\le t<t_{j+1}}\left|\sum_{i=1}^n\left(\zeta_i^*(t)-\zeta_i^*(t_j)\right)I_{ij}\widehat{\varepsilon}_i\right| \\ &\quad + \sup_{t_j\le t<t_{j+1}}\left|\sum_{i=1}^n\left(\zeta_i^*(t)-\zeta_i^*(t_j)\right)\breve{I}_{ij}\widehat{\varepsilon}_i\right|. \end{aligned} \tag{5.15}$$

Start with the latter term of (5.15) which concerns the situation when $\breve{I}_{ij}$ is nonzero. For any $t \in [t_j, t_{j+1})$ there exists a $\delta_j \geq 0$ such that $t = t_j + \delta_j$ and

$$\begin{aligned}
\left|\zeta_i^*(t) - \zeta_i^*(t_j)\right| \breve{I}_{ij} = \breve{I}_{ij} \Bigg| K_3\left(\frac{F(X_i)-t}{h}\right)(\sigma(X_i) - \sigma_F(t)) \\
- K_3\left(\frac{F(X_i)-t_j}{h}\right)(\sigma(X_i) - \sigma_F(t_j)) \Bigg| \\
\leq |\sigma|_\infty \left| K_3\left(\frac{F(X_i)-t}{h}\right) - K_3\left(\frac{F(X_i)-t_j}{h}\right)\right| \\
+ \left|\sigma_F(t_j) K_3\left(\frac{F(X_i)-t_j}{h}\right) - \sigma_F(t_j+\delta_j) K_3\left(\frac{F(X_i)-t}{h}\right)\right| \\
\leq L_{K_3} |\sigma|_\infty h^{-1} \left|t - t_j\right| \\
+ \left|\sigma_F(t_j) K_3\left(\frac{F(X_i)-t_j}{h}\right) - \sigma_F(t_j+\delta_j) K_3\left(\frac{F(X_i)-t}{h}\right)\right| \\
\leq L_{K_3} |\sigma|_\infty h^{-1} n^{-2} + |\sigma|_\infty \left| K_3\left(\frac{F(X_i)-t_j}{h}\right) - K_3\left(\frac{F(X_i)-t}{h}\right)\right| \\
+ |\delta_j| \left|\sigma_F^{(1)}(t_j + q_j\delta_j) K_3\left(\frac{F(X_i)-t_j}{h}\right)\right| \\
\leq Ch^{-1}n^{-2}.
\end{aligned} \tag{5.16}$$

Using the above with (5.14) in the latter term of (5.15),

$$\sup_{t_j \leq t < t_{j+1}} \left| \sum_{i=1}^n \left(\zeta_i^*(t) - \zeta_i^*(t_j)\right) \varepsilon_i \breve{I}_{ij} \right| \leq \frac{C}{\sqrt{nh}} \tag{5.17}$$

where C is independent of j. Shift attention now to the first term on the RHS of (5.15) and consider initially the situation where $\left|\frac{F(X_i)-t_j}{h} - 1\right| \leq Cn^{-2}h^{-1}$. A similar argument can be reached for values of t_j, t_{j+1} near the opposite boundary and consequently over the whole range for I_{ij}. Consider the situation where $h^{-1}(F(X_i) - t) < 1$, then the same argument holds from (5.16). So it remains to consider $h^{-1}(F(X_i) - t) > 1$ which is outside the support of K_3. Substituting $K_3\left(\frac{F(X_i)-t}{h}\right) = 0 = K_3(1)$ and using a similar argument to before,

$$\begin{aligned}
\left|\zeta_i^*(t) - \zeta_i^*(t_j)\right| I_{ij} = I_{ij} \Bigg| K_3\left(\frac{F(X_i)-t}{h}\right)(\sigma(X_i) - \sigma_F(t)) \\
- K_3\left(\frac{F(X_i)-t_j}{h}\right)(\sigma(X_i) - \sigma_F(t_j)) \Bigg| \\
= I_{ij} \left|\sigma(X_i) - \sigma_F(t)\right| \left| K_3(1) - K_3\left(\frac{F(X_i)-t_j}{h}\right)\right| \\
\leq Ch^{-1}n^{-2}.
\end{aligned}$$

So we also have,

$$\sup_{t_j \le t < t_{j+1}} \left| \sum_{i=1}^{n} \left(\zeta_i^*(t) - \zeta_i^*(t_j) \right) \hat{\varepsilon}_i I_{ij} \right| \le \frac{C}{\sqrt{nh}}. \tag{5.18}$$

So by using (5.17),

$$\max_{j \in G_N(a,b)} \sup_{t_j \le t < t_{j+1}} \left| \sum_{i=1}^{n} \left(\zeta_i^*(t) - \zeta_i^*(t_j) \right) \hat{\varepsilon}_i \right| \le \frac{C}{\sqrt{nh}}.$$

Then an application of the Markov inequality with bandwidth condition (5.6) implies that,

$$\max_{j \in G_N(a,b)} \sup_{t_j \le t < t_{j+1}} \left| \sum_{i=1}^{n} \left(\zeta_i^*(t) - \zeta_i^*(t_j) \right) \hat{\varepsilon}_i \right| = o_p \left(\sqrt{nh^3 \left| \log h \right|} \right).$$

This combined with (5.12) in (5.10) completes the proof. □

Proof of Lemma 5.2 . The proof of this Lemma is based on the argument in Proof of Lemma 5.1. Split the summation into two parts,

$$\sum_{i=1}^{n} K_3 \left(\frac{F(X_i) - t}{h} \right) \varepsilon_i = \sum_{i=1}^{n} K_3 \left(\frac{F(X_i) - t}{h} \right) \left(\hat{\varepsilon}_i + \breve{\varepsilon}_i \right),$$

where the variables $\hat{\varepsilon}_i$ and $\breve{\varepsilon}_i$ are defined by (5.8). Then the large deviations of the two summations are considered separately. First it is shown that

$$\sup_{t \in (h, 1-h)} \left| \sum_{i=1}^{n} K_3 \left(\frac{F(X_i) - t}{h} \right) \breve{\varepsilon}_i \right| = \mathcal{O}_p \left(\sqrt{nh \left| \log h \right|} \right).$$

Evaluating the expectation, recall that $\mathbb{E}\varepsilon_1^4 < \infty$ and apply the inequalities of Hölder and Chebyshev,

$$\begin{aligned}
\mathbb{E} \sup_{t \in (h, 1-h)} \left| \sum_{i=1}^{n} K_3 \left(\frac{F(X_i) - t}{h} \right) \breve{\varepsilon}_i \right| &\le |K_3|_\infty n \mathbb{E} \left| \breve{\varepsilon}_1 \right| \\
&= Cn\mathbb{E} \left| \varepsilon_1 \mathbb{1}_{|\varepsilon_1| \ge \sqrt{nh}} - \mathbb{E} \varepsilon_1 \mathbb{1}_{|\varepsilon_1| \ge \sqrt{nh}} \right| \\
&\le Cn\mathbb{E} \left| \varepsilon_1 \right| \mathbb{1}_{|\varepsilon_1| \ge \sqrt{nh}} \\
&\le Cn \sqrt{P \left(|\varepsilon_1| \ge \sqrt{nh} \right)} \\
&\le Ch^{-1}.
\end{aligned}$$

A simple application of the Markov inequality and (5.6) yields,

$$P\left(\sup_{t\in(h,1-h)}\left|\sum_{i=1}^{n}K_3\left(\frac{F(X_i)-t}{h}\right)\breve{\varepsilon}_i\right|\geq\sqrt{nh\left|\log h\right|}\right)=\mathcal{O}\left(\frac{1}{\sqrt{nh^3\left|\log h\right|}}\right)=o(1).$$

Now the other half of the decomposition needs to be dealt with. Consider a partition of $t\in(h,1-h)$ into a grid by defining $t_j:=j/N$ where $N=n^2$ and $j\in G_N(a,b)$ (see (5.9)). Then for any $t\in(h+a,1-h-b)$, there exists a t_j,t_{j+1} such that, $t\in[t_j,t_{j+1})$ and the decomposition is used,

$$\begin{aligned}\sup_{t\in(h,1-h)}&\left|\sum_{i=1}^{n}K_3\left(\frac{F(X_i)-t}{h}\right)\widehat{\varepsilon}_i\right|\\&\leq\max_{j\in G_N(a,b)}\sup_{t_j\leq t<t_{j+1}}\left|\sum_{i=1}^{n}\left(K_3\left(\frac{F(X_i)-t}{h}\right)-K_3\left(\frac{F(X_i)-t_j}{h}\right)\right)\widehat{\varepsilon}_i\right|\\&\qquad+\max_{j\in G_N(a,b)}\left|\sum_{i=1}^{n}K_3\left(\frac{F(X_i)-t_j}{h}\right)\widehat{\varepsilon}_i\right|.\end{aligned}\tag{5.19}$$

Dealing with the latter term of (5.19), apply the exponential martingale inequality in Theorem (Freedman, 1975) (see page 102). Notice that $\left\{K_3\left(\frac{F(X_i)-t_j}{h}\right)\widehat{\varepsilon}_i,\mathcal{X}_i\right\}$ form a martingale difference sequence. Again, by a similar argument to (5.11) the sum of the conditional variances are bounded by virtue of (5.2),

$$\begin{aligned}\sup_{t\in(h+a,1-h-b)}&\sum_{i=1}^{n}\mathbb{E}\left[\left(K_3\left(\frac{F(X_i)-t}{h}\right)\widehat{\varepsilon}_i\right)^2\middle|\mathcal{X}_{i-1}\right]\\&=\mathbb{E}\widehat{\varepsilon}_1^2\sup_{t\in(h+a,1-h-b)}\sum_{i=1}^{n}\int_{\mathbb{R}}K_3^2\left(\frac{F(u)-t}{h}\right)f_X(u|\mathcal{X}_{i-1})\,du\\&=\mathbb{E}\widehat{\varepsilon}_1^2\sup_{t\in(h+a,1-h-b)}\sum_{i=1}^{n}h\int_{-1}^{1}K_3^2\left(x\right)\frac{f_\eta(x-X_{i,i-1})}{f_X\left(Q(t+hx)\right)}\,dx\\&\leq 2nh\frac{\mathbb{E}\widehat{\varepsilon}_1^2\left|f_\eta\right|_\infty\int_{-1}^{1}K_3^2(x)\,dx}{\inf\limits_{Q(a)<x<Q(1-b)}f_X(x)}=:C_Vnh.\end{aligned}$$

Thus apply Theorem (Freedman, 1975) (see page 102) with $y = C_V nh$ and consider $x = C_T$ for some $C_T > 0$,

$$
\begin{aligned}
P&\left(\max_{j\in G_N(a,b)}\left|\sum_{i=1}^{n} K_3\left(\frac{F(X_i)-t_j}{h}\right)\widehat{\varepsilon}_i\right| \geq C_T\sqrt{nh\left|\log h\right|}\right)\\
&\leq P\left(\bigcup_{j\in G_N(a,b)}\left\{\left|\sum_{i=1}^{n} K_3\left(\frac{F(X_i)-t_j}{h}\right)\widehat{\varepsilon}_i\right| \geq C_T\sqrt{nh\left|\log h\right|}\right\}\right)\\
&\leq \sum_{j\in G_N(a,b)} P\left(\left\{\left|\sum_{i=1}^{n} K_3\left(\frac{F(X_i)-t_j}{h}\right)\widehat{\varepsilon}_i\right| \geq C_T\sqrt{nh\left|\log h\right|}\right\}\right)\\
&\leq N\exp\left\{-\frac{C_T^2 nh\left|\log h\right|}{2\left(\sqrt{nh\left|\log h\right|}+C_V nh\right)}\right\}\\
&= N\exp\left\{-\frac{C_T^2}{2C_V}\left|\log h\right|\right\}(1+o(1))\\
&\leq C n^2 h^{C_T^2/(2C_V)}.
\end{aligned}
$$

The above statement can be made arbitrarily small for a sufficiently large enough choice for C_T so therefore,

$$
\max_{j\in G_N(a,b)}\left|\sum_{i=1}^{n} K_3\left(\frac{F(X_i)-t}{h}\right)\widehat{\varepsilon}_i\right| = \mathcal{O}_p\left(\sqrt{nh\left|\log h\right|}\right). \tag{5.20}
$$

To deal with the first term on the RHS of (5.19), again t and t_j are considered with respect to the support of K_3 by using the indicator sets I_{ij} and $\check{I}_{ij}$ (see (5.13)).

$$
\begin{aligned}
\sup_{t_j\leq t<t_{j+1}}&\left|\sum_{i=1}^{n}\left(K_3\left(\frac{F(X_i)-t}{h}\right)-K_3\left(\frac{F(X_i)-t_j}{h}\right)\right)\widehat{\varepsilon}_i\right|\\
&\leq \sup_{t_j\leq t<t_{j+1}}\left|\sum_{i=1}^{n}\left(K_3\left(\frac{F(X_i)-t}{h}\right)-K_3\left(\frac{F(X_i)-t_j}{h}\right)\right)I_{ij}\widehat{\varepsilon}_i\right|\\
&\quad+\sup_{t_j\leq t<t_{j+1}}\left|\sum_{i=1}^{n}\left(K_3\left(\frac{F(X_i)-t}{h}\right)-K_3\left(\frac{F(X_i)-t_j}{h}\right)\right)\check{I}_{ij}\widehat{\varepsilon}_i\right|.
\end{aligned} \tag{5.21}
$$

Start with the situation when $\check{I}_{ij}$ is nonzero. For any $t \in [t_j, t_{j+1})$ there exists a $\delta_j \geq 0$ such that $t = t_j + \delta_j$ and

$$
\left|K_3\left(\frac{F(X_i)-t}{h}\right)-K_3\left(\frac{F(X_i)-t_j}{h}\right)\right|\check{I}_{ij} \leq L_{K_3}h^{-1}\left|t-t_j\right| \leq Ch^{-1}n^{-2}.
$$

Using the above with (5.14) in the latter term of (5.21),

$$\sup_{t_j \le t < t_{j+1}} \left| \sum_{i=1}^{n} \left(K_3 \left(\frac{F(X_i) - t}{h} \right) - K_3 \left(\frac{F(X_i) - t_j}{h} \right) \right) \hat{\varepsilon}_i \check{I}_{ij} \right| \le \frac{C}{\sqrt{nh}} \tag{5.22}$$

where C is independent of j. Shift attention now to the first term of (5.21) and consider initially the situation when $\left| \frac{F(X_i)-t_j}{h} - 1 \right| \le Cn^{-2}h^{-1}$. A similar argument can be reached for values of t_j, t_{j+1} near the opposite boundary and consequently over the whole range for I_{ij}. Consider the situation when $h^{-1}(F(X_i) - t) < 1$, then the same argument holds from (4.25). So it remains to consider when $h^{-1}(F(X_i) - t) > 1$ which is outside the support of K_3. Substituting $K_3\left(\frac{F(X_i)-t}{h}\right) = 0 = K_3(1)$ and using a similar argument to before,

$$\left| K_3 \left(\frac{F(X_i) - t}{h} \right) - K_3 \left(\frac{F(X_i) - t_j}{h} \right) \right| I_{ij} = \left| K_3(1) - K_3 \left(\frac{F(X_i) - t_j}{h} \right) \right| \le Ch^{-1}n^{-2}.$$

So the similar result can be shown that,

$$\sup_{t_j \le t < t_{j+1}} \left| \sum_{i=1}^{n} \left(K_3 \left(\frac{F(X_i) - t}{h} \right) - K_3 \left(\frac{F(X_i) - t_j}{h} \right) \right) \hat{\varepsilon}_i I_{ij} \right| \le \frac{C}{\sqrt{nh}}. \tag{5.23}$$

So by using (5.22) and (5.23) gives,

$$\max_{j \in G_N(a,b)} \sup_{t_j \le t < t_{j+1}} \left| \sum_{i=1}^{n} \left(K_3 \left(\frac{F(X_i) - t}{h} \right) - K_3 \left(\frac{F(X_i) - t_j}{h} \right) \right) \hat{\varepsilon}_i \right| \le \frac{C}{\sqrt{nh}},$$

Using the Markov inequality with the restricted bandwidth in (5.6) implies that,

$$\max_{j \in G_N(a,b)} \sup_{t_j \le t < t_{j+1}} \left| \sum_{i=1}^{n} \left(K_3 \left(\frac{F(X_i) - t}{h} \right) - K_3 \left(\frac{F(X_i) - t_j}{h} \right) \right) \hat{\varepsilon}_i \right| = o_p \left(\sqrt{nh^3 |\log h|} \right).$$

This combined with (5.20) in (5.19) completes the proof. □

Proof of Lemma 5.3 . Again, the proof is based on the Proof of Lemma 5.1. However, this time there is no nuisance variable ε_i to deal with. Define the terms, $d_i(t) = \gamma_i^*(t) - \mathbb{E}\left[\gamma_i^*(t) \middle| \mathcal{X}_{i-1}\right]$ ensuring that $\{\sum_{i=1}^n d_i(t), \mathcal{X}_n\}$ is a martingale. Then consider a partition of $t \in (h+a, 1-h-b)$ into a grid by defining $t_j := j/N$ where $N = n^2$ and $j \in G_N(a,b)$ where $G_N(a,b)$ is defined by (5.9). Then for any $t \in (h+a, 1-h-b)$, there exists a t_j, t_{j+1} such that, $t \in [t_j, t_{j+1})$ and the decomposition is used,

$$\sup_{t \in (h,1-h)} \left| \sum_{i=1}^{n} d_i(t) \right| \le \left\{ \max_{j \in G_N(a,b)} \sup_{t_j \le t < t_{j+1}} \left| \sum_{i=1}^{n} (d_i(t) - d_i(t_j)) \right| + \max_{j \in G_N(a,b)} \left| \sum_{i=1}^{n} d_i(t_j) \right| \right\}. \tag{5.24}$$

As seen in a similar light earlier, to deal with the second term of (5.24), apply the exponential martingale inequality Theorem (Freedman, 1975) (see page 102). First, exploit the Taylor expansion of μ and use the fact that $Support\,(K_3) = [-1,1]$, meaning there exists a τ_i dependent on X_i with $|\tau_i| \leq 1$ such that $F(X_i) = t + \tau_i h$ and,

$$\begin{aligned} \gamma_i^*(t) &= K_3\left(\frac{F(X_i)-t}{h}\right)\left(\mu_F(t+\tau_i h) - \mu_F(t)\right)\mathbb{1}_{(t-h,t+h)}\left(F(X_i)\right) \\ &= \tau_i h K_3\left(\frac{F(X_i)-t}{h}\right)\mu_F^{(1)}(t+\xi\,|\tau_i|\,h)\mathbb{1}_{(t-h,t+h)}\left(F(X_i)\right), \end{aligned} \tag{5.25}$$

where $|\xi| \leq 1$. Then consider the sum of the conditional variances and use (5.25) and (5.2),

$$\begin{aligned} &\sup_{t\in(h+a,1-h-b)}\sum_{i=1}^{n}\mathbb{E}\left[d_i^2(t)\middle|\,\mathcal{X}_{i-1}\right] \\ &= \sup_{t\in(h+a,1-h-b)}\sum_{i=1}^{n}\left\{\mathbb{E}\left[\left(\gamma_i^*(t)\right)^2\middle|\,\mathcal{X}_{i-1}\right] - \mathbb{E}\left[\gamma_i^*(t)\middle|\,\mathcal{X}_{i-1}\right]^2\right\} \\ &\leq \sup_{t\in(h+a,1-h-b)}\sum_{i=1}^{n}\mathbb{E}\left[\left(\gamma_i^*(t)\right)^2\middle|\,\mathcal{X}_{i-1}\right] \\ &= h^2\sup_{t\in(h+a,1-h-b)}\sum_{i=1}^{n}\mathbb{E}\left[\tau_i^2\left(\mu_F^{(1)}(t+\xi\,|\tau_i|\,h)\right)^2K_3^2\left(\frac{F(X_i)-t}{h}\right)\middle|\,\mathcal{X}_{i-1}\right] \\ &\leq h^2\left|\mu_F^{(1)}\right|_\infty^2\sup_{t\in(h+a,1-h-b)}\sum_{i=1}^{n}\mathbb{E}\left[K_3^2\left(\frac{F(X_i)-t}{h}\right)\middle|\,\mathcal{X}_{i-1}\right] \\ &= h^2\left|\mu_F^{(1)}\right|_\infty^2\sup_{t\in(h+a,1-h-b)}\sum_{i=1}^{n}\int_{\mathbb{R}}K_3^2\left(\frac{F(u)-t}{h}\right)f_\eta(u-X_{i,i-1})\,du \\ &= h^3\left|\mu_F^{(1)}\right|_\infty^2\sup_{t\in(h+a,1-h-b)}\sum_{i=1}^{n}\int_{-1}^{1}K_3^2(x)\,\frac{f_\eta(Q(t+hx)-X_{i,i-1})}{f(Q(t+hx))}\,dx \\ &\leq nh^3\frac{\left|\mu_F^{(1)}\right|_\infty^2|f_\eta|_\infty}{\inf\limits_{Q(a)\leq t\leq Q(1-b)}f_X(t)}\int_{-1}^{1}K_3^2(x)\,dx =: C_V nh^3. \end{aligned} \tag{5.26}$$

Thus apply Theorem (Freedman, 1975) (see page 102) with $y = C_V nh^3$ and $x = C_T$ for some $C_T > 0$,

$$\begin{aligned}
&P\left(\max_{j\in G_N(a,b)}\left|\sum_{i=1}^{n} d_i(t_j)\right| \geq C_T\sqrt{nh^3\left|\log h\right|}\right)\\
&\leq P\left(\bigcup_{j\in G_N(a,b)}\left\{\left|\sum_{i=1}^{n} d_i(t_j)\right| \geq C_T\sqrt{nh^3\left|\log h\right|}\right\}\right)\\
&\leq \sum_{j\in G_N(a,b)} P\left(\left\{\left|\sum_{i=1}^{n} d_i(t_j)\right| \geq C_T\sqrt{nh^3\left|\log h\right|}\right\}\right)\\
&\leq N\exp\left\{-\frac{C_T^2 nh^3\left|\log h\right|}{2\left(\sqrt{nh^3\left|\log h\right|}+C_V nh^3\right)}\right\}\\
&= N\exp\left\{-\frac{C_T^2}{2C_V}\left|\log h\right|\right\}(1+o(1))\\
&\leq Cn^2h^{C_T^2/(2C_V)}.
\end{aligned}$$

The above statement can be made arbitrarily small for a sufficiently large enough choice for C_T so therefore,

$$\max_{j\in G_N(a,b)}\left|\sum_{i=1}^{n} d_i(t_j)\right| = \mathcal{O}_p\left(\sqrt{nh^3\left|\log h\right|}\right). \tag{5.27}$$

To deal with the first term of (5.10), a careful consideration of the location of t and t_j with respect to the support of K_3 and the indicator sets I_{ij} and $\check{I}_{ij}$ (see (5.13)) is required.

$$\begin{aligned}
\sup_{t_j\leq t<t_{j+1}}\left|\sum_{i=1}^{n}\left(d_i(t)-d_i(t_j)\right)\right| \leq &\sup_{t_j\leq t<t_{j+1}}\left|\sum_{i=1}^{n}\left(d_i(t)-d_i(t_j)\right)I_{ij}\right|\\
&+\sup_{t_j\leq t<t_{j+1}}\left|\sum_{i=1}^{n}\left(d_i(t)-d_i(t_j)\right)\check{I}_{ij}\right|.
\end{aligned} \tag{5.28}$$

Start with the latter term of (5.28) which concerns the situation when $\check{I}_{ij}$ is nonzero.

$$\left|d_i(t)-d_i(t_j)\right|\check{I}_{ij} \leq \left|\gamma_i^*(t)-\gamma_i^*(t_j)\right|\check{I}_{ij} + \left|\mathbb{E}\left[\gamma_i^*(t)\middle|\mathcal{X}_{i-1}\right]-\mathbb{E}\left[\gamma_i^*(t_j)\middle|\mathcal{X}_{i-1}\right]\right|\check{I}_{ij} \tag{5.29}$$

By the same argument for (5.16),

$$\left|\gamma_i^*(t)-\gamma_i^*(t_j)\right|\check{I}_{ij} \leq Cn^{-2}h^{-1}. \tag{5.30}$$

So (5.29) and (5.30) imply that,

$$\left|d_i(t)-d_i(t_j)\right|\check{I}_{ij} \leq Cn^{-2}h^{-1}. \tag{5.31}$$

By an equivalent argument to the one that showed (5.18) it can be shown that,

$$\left|d_i(t) - d_i(t_j)\right| I_{ij} \leq Cn^{-2}h^{-1}. \tag{5.32}$$

So (5.31) and (5.32) imply that,

$$\max_{j \in G_N(a,b)} \sup_{t_j \leq t < t_{j+1}} \left| \sum_{i=1}^{n} \left(d_i(t) - d_i(t_j)\right) \right| \leq \frac{C}{nh}.$$

Since the above bound holds, using Markov's inequality and the restricted bandwidth (5.6) we have,

$$\max_{j \in G_N(a,b)} \sup_{t_j \leq t < t_{j+1}} \left| \sum_{i=1}^{n} \left(d_i(t) - d_i(t_j)\right) \right| = o_p\left(\sqrt{nh^3 \left|\log h\right|}\right).$$

This combined with (5.27) in (5.24) completes the proof. □

Proof of Lemma 5.4. Consider the value of $\sum_{i=1}^{n} \left(\mathbb{E}\left[\gamma_i(t) \middle| \mathcal{X}_{i-1}\right] - \mathbb{E}\gamma_1(t)\right)$ when $t \in (h, 1-h)$,

$$\begin{aligned}\sum_{i=1}^{n} \left(\mathbb{E}\left[\gamma_i(t) \middle| \mathcal{X}_{i-1}\right] - \mathbb{E}\gamma_1(t)\right) &= \int_{\mathbb{R}} K_3\left(\frac{F(u)-t}{h}\right) \mu(u) I^{[n]}(u)\, du \\ &= h \int_{-1}^{1} K_3(x)\, \mu_F(t+hx) I_F^{[n]}(t+hx)\, dx \end{aligned} \tag{5.33}$$

where $I^{[n]}(x) = \sum_{i=1}^{n} \left(f_X(x|\mathcal{X}_{i-1}) - f_X(x)\right)$. To bound the above process consider a Taylor expansion of both μ_F and $I_F^{[n]}$ and exploit the vanishing moments of $\mathscr{K}_{s\wedge r}$ in (K:4). Consider first the expansion of $I_F^{[n]}$ which requires the higher order derivatives of $I^{[n]}$. If Assumption (X) holds, then $f_\eta^{(s)}$ exists and $I_F^{[n]}$ has a Taylor expansion and for some $p \in [0,1]$,

$$I_F^{[n]}(t+hx) = \sum_{j=0}^{s\wedge r-1} \frac{(hx)^j \mathcal{D}^j I_F^{[n]}(t)}{j!} + \frac{(hx)^{(s\wedge r)}\, \mathcal{D}^{(s\wedge r)} I_F^{[n]}(t+phx)}{(s\wedge r)!}.$$

Indeed, calculating the intermediate derivatives using the Faà di Bruno formula (see (4.14) on page 96),

$$\mathcal{D}^j I_F^{[n]}(t) = \mathcal{D}^j I^{[n]} \circ Q(t) = \sum_{\mathbf{k} \in \mathcal{S}_j} \frac{j!}{k_1! k_2! \ldots k_j!} (\mathcal{D}^k I^{[n]} \circ Q)(t) \prod_{i=1}^{j} \left(\frac{Q^{(i)}(t)}{i!}\right)^{k_i}.$$

By a similar argument a Taylor expansion of μ_F exists. Then exploit the vanishing moment property of $K \in \mathscr{K}_{s\wedge r}$ in (K:4) by using the above expansion and the extra smoothness of μ_F in (5.33). Thus, there exists a $p, q \in [0,1]$ such that,

$$
\begin{aligned}
&\int_{-1}^{1} K_3(x)\, \mu_F(t+hx) I_F^{[n]}(t+hx)\, dx \\
&= \sum_{i=0}^{s\wedge r-1} \sum_{j=0}^{s\wedge r-1} \frac{h^{i+j} \mu_F^{(i)}(t) \mathcal{D}^j I_F^{[n]}(t)}{i!j!} \int_{-1}^{1} x^{i+j} K_3(x)\, dx \\
&\quad + \sum_{i=0}^{s\wedge r-1} \frac{h^{i+s\wedge r} \mu_F^{(i)}(t)}{i!(s\wedge r)!} \int_{-1}^{1} \mathcal{D}^{(s\wedge r)} I_F^{[n]}(t+phx) x^{i+s\wedge r} K_3(x)\, dx \\
&\quad + \sum_{j=0}^{s\wedge r-1} \frac{h^{j+s} \mathcal{D}^j I_F^{[n]}(t)}{j!(s\wedge r)!} \int_{-1}^{1} \mu_F^{(s)}(t+qhx) x^{j+s\wedge r} K_3(x)\, dx \\
&\quad + \frac{h^{2(s\wedge r)}}{((s\wedge r)!)^2} \int_{-1}^{1} \mathcal{D}^{(s\wedge r)} I_F^{[n]}(t+phx) \mu_F^{(s)}(t+qhx) x^{2(s\wedge r)} K_3(x)\, dx \\
&= h^{(s\wedge r)} \Bigg(\sum_{\substack{i=0 \\ i+j\geq s\wedge r}}^{s\wedge r-1} \sum_{j=0}^{s\wedge r-1} \frac{h^{i+j-(s\wedge r)} \mu_F^{(i)}(t) \mathcal{D}^j I_F^{[n]}(t)}{i!j!} \int_{-1}^{1} x^{i+j} K_3(x)\, dx \\
&\quad + \sum_{i=0}^{s\wedge r-1} \frac{h^{i} \mu_F^{(i)}(t)}{i!(s\wedge r)!} \int_{-1}^{1} \mathcal{D}^{(s\wedge r)} I_F^{[n]}(t+phx) x^{i+(s\wedge r)} K_3(x)\, dx \\
&\quad + \sum_{j=0}^{s\wedge r-1} \frac{h^{j} \mathcal{D}^j I_F^{[n]}(t)}{j!(s\wedge r)!} \int_{-1}^{1} \mu_F^{(s)}(t+qhx) x^{j+s\wedge r} K_3(x)\, dx \\
&\quad + \frac{h^{s\wedge r}}{((s\wedge r)!)^2} \int_{-1}^{1} \mathcal{D}^{(s\wedge r)} I_F^{[n]}(t+phx) \mu_F^{(s)}(t+qhx) x^{2(s\wedge r)} K_3(x)\, dx \Bigg).
\end{aligned}
\tag{5.34}
$$

Therefore it follows by (5.34) and Remark C.1 (see page 170) that,

$$
\sup_{t\in(h,1-h)} \left| \sum_{i=1}^{n} \left(\mathbb{E}\left[\gamma_i(t) \middle| \mathcal{X}_{i-1}\right] - \mathbb{E}\gamma_1(t) \right) \right| = \mathcal{O}_P\left(h^{s\wedge r+1} n^{1-\alpha_x/2} L(n) \right). \qquad \square
$$

Proof of Lemma 5.5 . Define, $d_i(t) = K_3\left(\frac{F(X_i)-t}{h}\right) - \mathbb{E}\left[K_3\left(\frac{F(X_i)-t}{h}\right) \middle| \mathcal{X}_{i-1}\right]$. The proof follows from a similar argument to Proof of Lemma 5.1. Define the grid $N = n^2$ and $t_j = j/N$ for $j \in G_N(a,b)$. Then, similar to the Proof of Lemma 5.1, the two terms that need to be dealt with are,

$$
\frac{1}{\sqrt{nh}} \left\{ \max_{j\in G_N(a,b)} \left| \sum_{i=1}^{n} d_i(t_j) \right| + \max_{j\in G_N(a,b)} \sup_{t_j\leq t<t_{j+1}} \left| \sum_{i=1}^{n} \left(d_i(t) - d_i(t_j) \right) \right| \right\}. \tag{5.35}
$$

The former is dealt with the inequality in Theorem (Freedman, 1975) (see page 102). Clearly $\{d_i(t), \mathcal{X}_i\}$ form a martingale difference sequence. Similar to (5.11) the sum of the conditional variances are bounded,

$$\begin{aligned}\sup_{t\in(h+a,1-h-b)} \sum_{i=1}^{n} \mathbb{E}\left[d_i(t)^2 \middle| \mathcal{X}_{i-1}\right] &\leq \sup_{t\in(h+a,1-h-b)} \sum_{i=1}^{n} \mathbb{E}\left[K_3^2\left(\frac{F(X_i)-t}{h}\right) \middle| \mathcal{X}_{i-1}\right] \\ &\leq 2nh \frac{|f_\eta|_\infty \int_{-1}^{1} K_3^2(x)\,dx}{\inf\limits_{Q(a)<x<Q(1-b)} f_X(x)} =: C_V nh.\end{aligned}$$

Apply Theorem (Freedman, 1975) (see page 102),

$$\begin{aligned}&P\left(\max_{j\in G_N(a,b)} \left|\sum_{i=1}^{n} d_i(t_j)\right| \geq C_T\sqrt{nh\,|\log h|}\right) \\ &\leq P\left(\bigcup_{j\in G_N(a,b)} \left\{\left|\sum_{i=1}^{n} d_i(t_j)\right| \geq C_T\sqrt{nh\,|\log h|}\right\}\right) \\ &\leq \sum_{j\in G_N(a,b)} P\left(\left|\sum_{i=1}^{n} d_i(t_j)\right| \geq C_T\sqrt{nh\,|\log h|}\right) \\ &\leq N \exp\left\{-\frac{C_T^2 nh\,|\log h|}{2\left(C_T\sqrt{nh\,|\log h|} + C_V nh\right)}\right\} \\ &\leq n^2 \exp\left\{-\frac{C_T^2}{2C_V}\,|\log h|\right\}(1+o(1)) \\ &\leq Cn^2 h^{C_T^2/(2C_V)} = Cn^2 h^{C_T^2/(2C_V)},\end{aligned}$$

which can be made arbitrarily small by an appropriately large C_T. Therefore,

$$\max_{j\in G_N(a,b)} \left|\sum_{i=1}^{n} d_i(t_j)\right| = \mathcal{O}_p\left(\sqrt{nh\,|\log h|}\right). \tag{5.36}$$

To deal with the other term, by construction, $\left|K_3\left(\frac{F(X_i)-t}{h}\right) - K_3\left(\frac{F(X_i)-t_j}{h}\right)\right| \leq Cn^{-2}h^{-1}$. This consequently implies that, $\left|d_i(t) - d_i(t_j)\right| \leq Cn^{-2}h^{-1}$. Therefore,

$$\max_{j\in G_N(a,b)} \sup_{t_j\leq t<t_{j+1}} \left|\sum_{i=1}^{n} \left(d_i(t) - d_i(t_j)\right)\right| \leq \frac{C}{nh}.$$

So by the above and the Markov inequality and (5.6),

$$P\left(\max_{j\in G_N(a,b)} \sup_{t_j\leq t<t_{j+1}} \left|\sum_{i=1}^{n} \left(d_i(t) - d_i(t_j)\right)\right| \geq C\sqrt{nh\,|\log h|}\right) \leq \frac{C}{\sqrt{n^3h^3\,|\log h|}} = o(1). \tag{5.37}$$

The result follows from (5.37), (5.36) and (5.35). □

Proof of Theorem 5.2 . 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 4.2. Recall the modified processes in (5.7) and by the same argument used to show (5.26), it can be shown that,

$$\left\| \sum_{i=1}^{n} \left(\gamma_i^*(t) - \mathbb{E}\left[\gamma_i^*(t) | \mathcal{X}_{i-1}\right] \right) \right\| = \mathcal{O}\left(\sqrt{nh^3}\right).$$

Apply the above bound and Lemma 5.4 to yield,

$$\begin{aligned}
&\widehat{\mathcal{K}}_h(t) - \mathcal{K}_h(t) \\
&= \frac{1}{nh^4} \sum_{i=1}^{n} \Big(\gamma_i(t) - \mathbb{E}\gamma_i(t) + \zeta_i(t)\varepsilon_i \Big) \\
&= \frac{1}{nh^4} \sum_{i=1}^{n} \Bigg(\gamma_i(t) - \mathbb{E}\left[\gamma_i(t) | \mathcal{X}_{i-1}\right] + \mathbb{E}\left[\gamma_i(t) | \mathcal{X}_{i-1}\right] - \mathbb{E}\gamma_i(t) + \zeta_i(t)\varepsilon_i \Bigg) \\
&= \frac{1}{nh^4} \sum_{i=1}^{n} \left(\gamma_i^*(t) - \mathbb{E}\left[\gamma_i^*(t) | \mathcal{X}_{i-1}\right] \right) \\
&\qquad + \frac{\mu_F(t)}{nh^4} \sum_{i=1}^{n} \left(K_3\left(\frac{F(X_i) - t}{h}\right) - \mathbb{E}\left[K_3\left(\frac{F(X_i) - t}{h}\right) \middle| \mathcal{X}_{i-1} \right] \right) \\
&\qquad\quad + \frac{1}{nh^4} \sum_{i=1}^{n} \left(\mathbb{E}\left[\gamma_i(t) | \mathcal{X}_{i-1}\right] - \mathbb{E}\gamma_i(t) + \zeta_i(t)\varepsilon_i \right) \\
&= \frac{1}{nh^4} \sum_{i=1}^{n} \left(\mu_F(t) \left(K_3\left(\frac{F(X_i) - t}{h}\right) - \mathbb{E}\left[K_3\left(\frac{F(X_i) - t}{h}\right) \middle| \mathcal{X}_{i-1} \right] \right) + \zeta_i(t)\varepsilon_i \right) \\
&\qquad\quad + \mathcal{O}\left(h^{s\wedge r - 3} n^{\frac{1-\alpha_X}{2}} L(n) \right) + \mathcal{O}_p\left(\frac{1}{nh^5} \right).
\end{aligned} \tag{5.38}$$

Recall we are considering the term, $\sqrt{nh^7}\left(\widehat{\mathcal{K}}_h(t) - \mathcal{K}_h(t)\right)$ so define the standardised stochastic terms,

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

Then in a similar fashion to the Proof of Theorem 4.2 it will be shown by the martingale central limit theorem of Theorem 2 (Brown, 1971) (see page 96) that,

$$\sum_{i=1}^{n} \Delta_i(t) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1). \tag{5.39}$$

Indeed, $\{\Delta_i(t), \mathcal{X}_i\}_{i=1}^{n}$ is a martingale difference sequence. 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,

$$\begin{aligned}
&\sum_{i=1}^{n} \mathbb{E}\left[\Delta_i^2(t) \middle| \mathcal{X}_{i-1}\right] \\
&\quad = \frac{\mu_F^2(t)}{nh\, v^2(t)} \sum_{i=1}^{n} \mathbb{E}\left[\left(K_3\left(\frac{F(X_i)-t}{h}\right) - \mathbb{E}\left[K_3\left(\frac{F(X_i)-t}{h}\right)\middle| \mathcal{X}_{i-1}\right]\right)^2 \middle| \mathcal{X}_{i-1}\right] \\
&\qquad + \sum_{i=1}^{n} \frac{\mathbb{E}\left[\zeta_i^2(t)\middle| \mathcal{X}_{i-1}\right]}{nh\, v^2(t)}.
\end{aligned} \tag{5.40}$$

Dealing with the second term on the RHS of (5.40), use Proposition C.1 from Appendix C,

$$\begin{aligned}
&\frac{1}{nh}\sum_{i=1}^{n} \mathbb{E}\left[\zeta_i^2(t)\middle|\mathcal{X}_{i-1}\right] \\
&\quad = \frac{1}{nh}\sum_{i=1}^{n} \mathbb{E}\zeta_i^2(t) + \frac{1}{nh}\sum_{i=1}^{n}\left(\mathbb{E}\left[\zeta_i^2(t)\middle| \mathcal{X}_{i-1}\right] - \mathbb{E}\zeta_i^2(t)\right) \\
&\quad = \int_{-\frac{t}{h}}^{\frac{1-t}{h}} \sigma_F^2(t+hx)K_3^2(x)\,dx + \frac{1}{n}\int_{-1}^{1} \sigma_F^2(t+hx)K_3^2\,(x)\; I_F^{[n]}(t+hx)\,dx \\
&\quad = \int_{-1}^{1} \sigma_F^2(t+hx)K_3^2(x)\,dx + \mathcal{O}_p(n^{-\alpha_x/2}L(n)) \\
&\quad = \sigma_F^2(t)\int_{-1}^{1} K_3^2(x)\,dx + \mathcal{O}(h^2) + \mathcal{O}_p(n^{-\alpha_x/2}L(n)).
\end{aligned} \tag{5.41}$$

To bound the first term of (5.40), a bound is required for $\mathbb{E}\left[K_3\left(\frac{F(X_i)-t}{h}\right)\middle| \mathcal{X}_{i-1}\right]^2$. This is achieved using a Taylor expansion of composite functions to exploit the vanishing moment condition of K_3. Recall from (4.14) in the Proof of Lemma 5.4 the Faà di Bruno formula for composite functions. 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 $\mathring{f}_\eta(x) := f_X\left(x\middle|\mathcal{X}_{i-1}\right) = f_\eta(x - X_{i,i-1})$ and $g(x) = 1/x$. Then $X_{i,i-1}$ and Z_i are $\mathcal{X}_{i-1}$-measurable and for all $t \in (h, 1-h)$ the conditional expectation can be evaluated as follows.

$$\begin{aligned}
\mathbb{E}\left[K_3\left(\frac{F(X_i)-t}{h}\right)\middle|\mathcal{X}_{i-1}\right] &= \int_{\mathbb{R}} K_3\left(\frac{F(v)-t}{h}\right) f_X\left(v\middle|\mathcal{X}_{i-1}\right)\,dv \\
&= h\int_{-1}^{1} K_3\,(x)\,\frac{f_X(Q(t+hx)|\mathcal{X}_{i-1})}{f_X(Q(t+hx))}\,dx \\
&= h\int_{-1}^{1} K_3\,(x)\left(\mathring{f}_\eta \circ Q\right)(t+hx)\,(g\circ f_X\circ Q)\,(t+hx)\,dx.
\end{aligned} \tag{5.42}$$

Use the Faà di Bruno chain rule on the composite functions, $p(t) := \left(\mathring{f}_\eta \circ Q\right)(t)$ and $q(t) := (g \circ f_X \circ Q)(t)$ to obtain the Taylor expansions. Start with $q(t)$ below,

$$(g \circ f_X \circ Q)(t+hx) = \sum_{j=0}^{s\wedge r-1} \frac{h^j x^j (g \circ f_X \circ Q)^{(j)}(t)}{j!} + \frac{h^{s\wedge r} x^{s\wedge r} (g \circ f_X \circ Q)^{(s\wedge r)}(t+\tau hx)}{(s\wedge r)!}, \tag{5.43}$$

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

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

due to restrictions imposed in Assumption (X). Similarly for $p(t)$,

$$\left(\mathring{f}_\eta \circ Q\right)(t+hx) = \sum_{j=0}^{s\wedge r-1} \frac{h^j x^j \left(\mathring{f}_\eta \circ Q\right)^{(j)}(t)}{j!} + \frac{h^{s\wedge r} x^{s\wedge r} \left(\mathring{f}_\eta \circ Q\right)^{(s\wedge r)}(t+\delta hx)}{(s\wedge r)!} \tag{5.44}$$

where $|\delta| \le 1$. Therefore, using (5.44) and (5.43) in (5.42) with the vanishing moment condition (K:4) implies that,

$$\begin{aligned}
&\mathbb{E}\left[K_3\left(\frac{F(X_i)-t}{h}\right)\middle|\mathcal{X}_{i-1}\right] \\
&\quad = h^{s\wedge r+1}\Bigg\{\sum_{j=0}^{s\wedge r-1}\sum_{\substack{\ell=0\\ j+\ell\ge s\wedge r}}^{s\wedge r-1} h^{\ell+j-s\wedge r}\frac{p^{(j)}(t)q^{(\ell)}(t)}{j!\ell!}\int_{-1}^{1} x^{\ell+j}K_3(x)\,dx \\
&\qquad + \sum_{j=0}^{s\wedge r-1}\frac{p^{(j)}(t)h^j}{(s\wedge r)!j!}\int_{-1}^{1} x^{s\wedge r+j}K_3(x)q^{(s\wedge r)}(t+\tau hx)\,dx \\
&\qquad + \sum_{\ell=0}^{s\wedge r-1}\frac{q^{(\ell)}(t)h^\ell}{(s\wedge r)!\ell!}\int_{-1}^{1} x^{s\wedge r+\ell}K_3(x)p^{(s\wedge r)}(t+\delta hx)\,dx \\
&\qquad + \frac{h^{s\wedge r}}{((s\wedge r)!)^2}\int_{-1}^{1} x^{2(s\wedge r)}K_3(x)q^{(s\wedge r)}(t+\tau hx)p^{(s\wedge r)}(t+\delta hx)\,dx\Bigg\}.
\end{aligned}$$

However, by Assumption (X), $f_\eta^{(j)}$ 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,

$$\mathbb{E}\left[K_3\left(\frac{F(X_i)-t}{h}\right)\middle|\mathcal{X}_{i-1}\right] < Ch^{s\wedge r+1} \quad \text{a.s.} \tag{5.45}$$

Define,

$$\mathring{K}_3(X_{i,i-1},t) := \mathbb{E}\left[K_3\left(\frac{F(X_i)-t}{h}\right)\middle|\,\mathcal{X}_{i-1}\right] \quad \text{and}$$
$$g(X_{i,i-1},t) := \mathring{K}_3^{\,2}(X_{i,i-1},t) - \mathbb{E}\mathring{K}_3^{\,2}(X_{i,i-1},t),$$

then $\mathbb{E}g(X_{i,i-1},t)=0$ and by Jensen's Inequality $\mathbb{E}g(X_{i,i-1},t)^2<\infty$. It will be shown by an application of (Wu, 2007, Theorem 1) that $\sum_{i=1}^n g(X_{i,i-1},t)=\mathcal{O}_p\left(h^{s\wedge r+2}n^{1-\alpha_x/2}L(n)\right)$.

Using a similar technique to the one used in the Proof of Proposition C.1 (see page 170), define the physical dependence measure,

$$v_i := \sup_{t\in(h,1-h)} \left\|\mathbb{E}\left[g(X_{i,i-1},t)\middle|\,\mathcal{X}_0\right]-\mathbb{E}\left[g(X_{i,i-1},t)\middle|\,\mathcal{X}_{-1}\right]\right\|.$$

To bound v_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 $\mathcal{X}_i^*=\sigma\left(\eta_i,\eta_{i-1},\ldots,\eta_1,\eta_0',\eta_1,\ldots;\varepsilon_1,\ldots,\varepsilon_i\right)$. In the same fashion define $\mathring{K}_3(X^*_{i,i-1},t):=\mathbb{E}\left[K_3\left(\frac{F(X_i^*)-t}{h}\right)\middle|\,\mathcal{X}^*_{i-1}\right]$. By Theorem 1 of Wu (2005) it was shown that $v_i\le\sup_{t\in(h,1-h)}\left\|g(X_{i,i-1},t)-g(X^*_{i,i-1},t)\right\|$. Using this, (5.45) and the Lipschitz property of f_η it will be shown that $v_i=\mathcal{O}\left(h^{s\wedge r+2}i^{-\beta}L(i)\right)$,

$$\begin{aligned}
v_i &\le \sup_{t\in(h,1-h)}\left\|g(X_{i,i-1},t)-g(X^*_{i,i-1},t)\right\|\\
&= \sup_{t\in(h,1-h)}\left\|\left(\mathring{K}_3\left(X_{i,i-1},t\right)+\mathring{K}_3\left(X^*_{i,i-1},t\right)\right)\left(\mathring{K}_3\left(X_{i,i-1},t\right)-\mathring{K}_3\left(X^*_{i,i-1},t\right)\right)\right\|\\
&\le Ch^{s\wedge r+1}\sup_{t\in(h,1-h)}\left\|\mathring{K}_3\left(X_{i,i-1},t\right)-\mathring{K}_3\left(X^*_{i,i-1},t\right)\right\|\\
&= Ch^{s\wedge r+1}\sup_{t\in(h,1-h)}\left\|\int_{\mathbb{R}}K_3\left(\frac{F(u)-t}{h}\right)\left(f_\eta\left(u-X_{i,i-1}\right)-f_\eta\left(u-X^*_{i,i-1}\right)\right)du\right\|\\
&\le Ch^{s\wedge r+1}\sup_{t\in(h,1-h)}\int_{\mathbb{R}}\left|K_3\left(\frac{F(u)-t}{h}\right)\right|du\left\|X_{i,i-1}-X^*_{i,i-1}\right\|\\
&\le Ch^{s\wedge r+2}\left\|\eta_0-\eta_0'\right\|c_i = Ch^{s\wedge r+2}i^{-\beta}L(i),
\end{aligned}$$

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, $\left\|\sum_{i=1}^n g(X_{i,i-1},t)\right\|^2=\mathcal{O}\left(h^{2(s\wedge r)+4}n^{2-\alpha_x}L^2(n)\right)$. Using this and (5.45),

$$\begin{aligned}
\frac{1}{nh}\sum_{i=1}^n\mathbb{E}\left[K_3\left(\frac{F(X_i)-t}{h}\right)\middle|\,\mathcal{X}_{i-1}\right]^2 &= \frac{1}{nh}\sum_{i=1}^n g(X_{i,i-1},t)+\frac{1}{nh}\sum_{i=1}^n\mathbb{E}\mathring{K}_3^2(X_{i,i-1},t)\\
&= \mathcal{O}_p\left(h^{s\wedge r+1}n^{-\alpha_x/2}L(n)\right)+\mathcal{O}\left(h^{2(s\wedge r)+2}\right)\\
&= o_p(1). \qquad (5.46)
\end{aligned}$$

Then the first term on the RHS of (5.40) can be bounded by (5.46) and a similar application of Lemma 1 of Zhao and Wu (2008),

$$\begin{aligned}
&\frac{\mu_F^2(t)}{nh}\sum_{i=1}^{n}\mathbb{E}\left[\left(K_3\left(\frac{F(X_i)-t}{h}\right)-\mathbb{E}\left[K_3\left(\frac{F(X_i)-t}{h}\right)\Big|\,\mathcal{X}_{i-1}\right]\right)^2\Big|\,\mathcal{X}_{i-1}\right]\\
&\qquad=\frac{\mu_F^2(t)}{nh}\sum_{i=1}^{n}\left\{\mathbb{E}K_3^2\left(\frac{F(X_i)-t}{h}\right)+\mathbb{E}\left[K_3\left(\frac{F(X_i)-t}{h}\right)\Big|\,\mathcal{X}_{i-1}\right]^2\right.\\
&\qquad\qquad\left.+\left(\mathbb{E}\left[K_3^2\left(\frac{F(X_i)-t}{h}\right)\Big|\,\mathcal{X}_{i-1}\right]-\mathbb{E}K_3^2\left(\frac{F(X_i)-t}{h}\right)\right)\right\}\\
&\qquad=\mu_F^2(t)\int_{-1}^{1}K_3^2(x)\,dx+\mathcal{O}_p(n^{-\alpha_x/2}L(n))+\mathcal{O}\left(h^2\right). \qquad (5.47)
\end{aligned}$$

Substituting (5.47) and (5.41) into (5.40) implies that,

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

For the Lindeberg condition, let $\epsilon>0$ and define $A_n(\epsilon)=\{|\Delta_1(t)|>\epsilon\}$, then similar to the procedure used in the Proof of Theorem 4.2, it can be shown that $A_n(\epsilon)\to\emptyset$ for any $\epsilon>0$ so the Lindeberg condition holds. Thus apply the martingale central limit theorem of Theorem 2 (Brown, 1971) (see page 96) ensuring that (5.39) holds. Then, using (5.3) in the decomposition given in (5.38), the result follows by Slutsky's Theorem. □

Proof of Theorem 5.3 . Again, use the decomposition (5.38) from the Proof of Theorem 5.2,

$$\begin{aligned}
&\widehat{\mathcal{K}_h}(t)-\mathcal{K}_h(t)\\
&=\frac{1}{nh^4}\sum_{i=1}^{n}\Big(\gamma_i(t)-\mathbb{E}\left[\gamma_i(t)|\,\mathcal{X}_{i-1}\right]+\mathbb{E}\left[\gamma_i(t)|\,\mathcal{X}_{i-1}\right]-\mathbb{E}\gamma_i(t)+\zeta_i(t)\varepsilon_i\Big)\\
&=\mathcal{O}_p\left(n^{-1/2}h^{-7/2}\right)+\frac{1}{nh^4}\sum_{i=1}^{n}\Big(\mathbb{E}\left[\gamma_i(t)|\,\mathcal{X}_{i-1}\right]-\mathbb{E}\gamma_1(t)\Big). \qquad (5.48)
\end{aligned}$$

Standardising the process,

$$\begin{aligned}
\frac{n^{\alpha_x/2}}{L(n)}\left(\widehat{\mathcal{K}_h}(t)-\mathcal{K}_h(t)\right)&=\frac{1}{n^{1-\alpha_x/2}L(n)h^4}\sum_{i=1}^{n}\Big(\mathbb{E}\left[\gamma_i(t)|\,\mathcal{X}_{i-1}\right]-\mathbb{E}\gamma_1(t)\Big)\\
&\quad+\mathcal{O}_p\left(n^{-(1-\alpha_x)/2}L^{-1}(n)h^{-7/2}\right).
\end{aligned}$$

By the assumed bandwidth condition (5.5), $n^{-(1-\alpha_x)/2}L^{-1}(n)h^{-7/2} = o(1)$, so it follows that the last term is negligible and the asymptotic behaviour is dominated by the LRD conditional expectation terms. As such, define the standardised process,

$$Y_i(t) := \frac{\mathbb{E}\left[\gamma_i(t)|\,\mathcal{X}_{i-1}\right] - \mathbb{E}\gamma_i(t)}{h^4 n^{1-\alpha_x/2}L(n)\mathcal{C}_1(t)}.$$

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

$$\sum_{i=1}^{n} Y_i(t) \xrightarrow{\mathscr{D}} \mathcal{N}(0,1). \tag{5.49}$$

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 $X_{i,i-1}$, $Z_i := s_X^{-1}\left(X_{i,i-1} - \mu_X\right)$, $Z_i \sim \mathcal{N}(0,1)$. Notice that $Y_i(t)$ and Z_i are both $\mathcal{X}_{i-1}$-measurable and define $G(Z_i, t) := \mathbb{E}\left[\gamma_i(t)|\,\mathcal{X}_{i-1}\right] - \mathbb{E}\gamma_i(t)$. 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,

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

where $a_m(t) = \mathbb{E}\left[H_m(Z_1)G(Z_1,t)\right]$ is the m^{th} Hermite coefficient. For our case it is assumed that $a_1 \neq 0$. To evaluate a_1, re-express the conditional expectation using the standardised version of X_i,

$$\begin{aligned}
\mathbb{E}\left[\gamma_i(t)|\,\mathcal{X}_{i-1}\right] &= \mathbb{E}\left[\mu(X_i)K_3\left(\frac{\Phi\left(X_i - \mu_X\right) - t}{h}\right)\middle|\,\mathcal{X}_{i-1}\right] \\
&= \mathbb{E}\left[\mu(\eta_i + \mu_X + s_X Z_i)K_3\left(\frac{\Phi\left(\eta_i + s_X Z_i\right) - t}{h}\right)\middle|\,\mathcal{X}_{i-1}\right] \\
&= \mathbb{E}_\eta\left[\mu(\eta_i + \mu_X + s_X Z_i)K_3\left(\frac{\Phi\left(\eta_i + s_X Z_i\right) - t}{h}\right)\right] \\
&= \frac{1}{\sigma_\eta}\int_{\mathbb{R}} \mu(u + \mu_X + s_X Z_i)K_3\left(\frac{\Phi\left(u + s_X Z_i\right) - t}{h}\right)\phi\left(\frac{u}{\sigma_\eta}\right) du.
\end{aligned}$$

Then substitute the above into a_1,

$$\begin{aligned}
a_1(t) &= \mathbb{E}\left[Z_1 G(Z_1,t)\right] = \mathbb{E}\left[Z_1\mathbb{E}\left[\gamma_1(t)|\,\mathcal{X}_0\right]\right] - \mathbb{E}Z_1\mathbb{E}\gamma_1(t) \\
&= \mathbb{E}\left[Z_1\frac{1}{\sigma_\eta}\int_{\mathbb{R}} \mu(u + \mu_X + s_X Z_1)K_3\left(\frac{\Phi(u + s_X Z_1) - t}{h}\right)\phi\left(\frac{u}{\sigma_\eta}\right) du\right] \\
&= \frac{1}{\sigma_\eta}\int_{\mathbb{R}}\int_{\mathbb{R}} z\mu(u + \mu_X + s_X z)K_3\left(\frac{\Phi(u + s_X z) - t}{h}\right)\phi(z)\phi\left(\frac{u}{\sigma_\eta}\right) dz\, du \\
&= \frac{h}{s_X^2\sigma_\eta}\int_{\mathbb{R}}\int_{-\frac{t}{h}}^{\frac{1-t}{h}} \frac{\Phi^{-1}(t+hw) - u}{\phi(\Phi^{-1}(t+hw))}\mu_F(t+hw)K_3(w)\,\phi\left(\frac{\Phi^{-1}(t+hw) - u}{s_X}\right)\phi\left(\frac{u}{\sigma_\eta}\right) dw\, du.
\end{aligned}$$

Recall from (5.43) (with $g(x) = 1/x$) in combination with Assumption (X) that the density in the denominator has a Taylor expansion up to order s. By exploiting the Faà di Bruno formula again, it can be shown very easily that for $w \in (-1, 1)$,

$$\phi\left(\frac{\Phi^{-1}(t+hw)-u}{s_X}\right) = \phi\left(\frac{\Phi^{-1}(t)-u}{s_X}\right) + \mathcal{O}(h).$$

Lastly, by definition, the smoothed third derivative, $h^3\mathcal{K}_h(t) = \int_{-1}^{1}\mu_F(t+hw)K_3(w)\,dw$. This all leads to the conclusion that,

$$a_1(t) \sim \frac{h^4\mathcal{K}_h(t)}{s_X^2\sigma_\eta\phi\left(\Phi^{-1}(t)\right)}\int_{\mathbb{R}}\phi\left(\frac{\Phi^{-1}(t)-u}{s_X}\right)\left(\Phi^{-1}(t)-u\right)\phi\left(\frac{u}{\sigma_\eta}\right)du = \frac{h^4\mathcal{C}_1(t)}{C_1}.$$

From Corollary 5.1 (Taqqu, 1975) (see page 25),

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

Therefore (5.49) holds by Slutsky's Theorem in the decomposition given in (5.48) in conjunction with (5.39), (5.49) and (5.5). □

5.3.4 *Estimation method under LRD design*

Recall from the general random design estimation method in Section 4.3.5 that $\widehat{\mathcal{K}_h}(t)$, has decomposition,

$$\begin{aligned}\widehat{\mathcal{K}_h}(t) &= h^{-2}\left[\mu_F^{(1)}\right](\lambda)K_1\left(\frac{\lambda-t}{h}\right) + \mathcal{J}_h(t) + \mathcal{Z}_h(t) + \mathcal{b}_h(t)\\ &:= \mathcal{L}_h(t) + \mathcal{J}_h(t) + \mathcal{Z}_h(t) + \mathcal{b}_h(t),\end{aligned}$$

where $\mathcal{L}_h(t)$ is the localisation term, $\mathcal{J}_h(t) = \mathcal{O}\left(h^{s-3}\right)$, is the deterministic bias and $\mathcal{Z}_h(t) + \mathcal{b}_h(t)$ are the stochastic bias and stochastic error processes analysed in Section 5.3.3.

As before, to avoid trivial complications of the location of the global minimum and maximum generated by $\mathcal{L}_h(t)$, assume that $\left[\mu_F^{(1)}\right](\lambda) > 0$. Apply the same random design procedure outlined in Section 4.3.5. This involves firstly using the ZCT to estimate the shifted change point λ with $\widehat{\lambda}_n$. Then rescale the $\widehat{\lambda}_n$ back to an estimate of θ using the empirical quantile function. Begin with the *Localisation step* of the ZCT.

Localisation step

By using a similar argument to (2.34) for the set $L_\lambda = \{t : |\lambda - t| < h\}$,

$$\mathcal{L}_h(t) = \begin{cases} h^{-2} K_1\left(\frac{\lambda - t}{h}\right)\left[\mu^{(1)}\right](\lambda), & \text{if } \mu \in \mathscr{C}_s^1([0,1],\lambda) \text{ and } t \in L_\lambda \\ 0, & \text{if } \mu \in \mathscr{S}_s \text{ or } \left\{\mu \in \mathscr{C}_s^1([0,1],\lambda) \text{ and } t \notin L_\lambda\right\}. \end{cases}$$

Also, $\mathcal{L}_h(t)$ has the same unique extrema with a unique global minimum at the point

$$t_* := \arg\min_{t \in (h, 1-h)} \mathcal{L}_h(t) = \lambda - q^* h$$

and a unique global maximum at the point

$$t^* := \arg\max_{t \in (h, 1-h)} \mathcal{L}_h(t) = \lambda + q^* h.$$

Also, when $|\lambda - t| < h$, $\mathcal{L}_h(t) \geq Ch^{-2} > Ch^{s-3}$ and $\mathcal{J}_h(t) = \mathcal{O}(h^{s-3})$, so $\mathcal{L}_h(t)$ dominates the deterministic signal of $\mathcal{K}_h(t)$.

To construct estimates of the unique global extrema, t_* and t^*, the localisation term $\mathcal{L}_h(t)$ also needs to dominate the stochastic terms, $\mathcal{b}_h(t)$ and $\mathcal{Z}_h(t)$. If the bandwidth is chosen such that,

$$Cn^{-1/3+\delta} < h < Cn^{-1/7-\delta}, \qquad \text{for some } \delta > 0,$$

then it guarantees that (5.3) holds. Then apply Theorem 5.2 to yield,

$$\mathcal{Z}_h(t) + \mathcal{b}_h(t) = \mathcal{O}_p\left(n^{-1/2} h^{-7/2}\right). \tag{5.50}$$

So to have a well defined signal where $\mathcal{L}_h(t)$ dominates the stochastic contributions as well, it is required that $\mathcal{L}_h(t)$ dominates (5.50), $h^{-2} \geq Cn^{-1/2} h^{-7/2}$ or equivalently,

$$h \geq Cn^{-1/3},$$

which holds by construction.

So again, with the optimal choice of bandwidth, the stochastic contributions asymptotically don't depend on α_x. The points t_* and t^* can be estimated with,

$$\hat{t}_* = \arg\min_{t \in (h, 1-h)} \widehat{\mathcal{K}}_h(t) \qquad \hat{t}^* = \arg\max_{t \in (h, 1-h)} \widehat{\mathcal{K}}_h(t).$$

By this construction, the interval $\hat{A}_n = (\hat{t}_*, \hat{t}^*)$ has a length which is order h and contains λ with high probability.

Kink detection step

As before, to ensure the signal generated by $\mathcal{K}_h(t)$ is genuine and not an artefact of the noise, some large deviations results concerning $\mathcal{Z}_h(t) + \mathcal{b}_h(t)$ are needed. First standardise the statistic $\widehat{\mathcal{K}}_h(t)$ to have unit variance with,

$$\mathcal{T}_{\widehat{\mathcal{K}}}(t) := \frac{\sqrt{nh^7}}{v(t)} \left(\widehat{\mathcal{K}}_h(t) - \mathcal{K}_h(t) \right).$$

where $v(t)$ is defined in Theorem 4.2 on page 93. Consider the large deviations when μ is smooth without a kink, that is, $\mu \in \mathscr{S}_s$. Since (5.3) is guaranteed to hold from (5.6), apply the large deviations result Theorem 5.4 and for any $\epsilon > 0$ there exists a $C_T > 0$ large enough such that,

$$\begin{aligned} &P\left(\sup_{t\in(h,1-h)} \left| \mathcal{T}_{\widehat{\mathcal{K}}}(t) \right| \geq C_T \sqrt{2\left|\log h\right|} \right) \\ &\quad = P\left(\sqrt{nh^7} \sup_{t\in(h,1-h)} \left| \widehat{\mathcal{K}}_h(t) - \mathcal{K}_h(t) \right| \geq \inf_{t\in(h,1-h)} v(t) C_T \sqrt{2\left|\log h\right|} \right) \leq \epsilon. \end{aligned} \tag{5.51}$$

Again, on the other hand, if $\mu_F \in \mathscr{C}_s^1([0,1],\lambda)$, then for $t \in (t_*, t^*)$, $\mathcal{L}_h(t) \geq Ch^{-2}$ and for the same large $C_T > 0$ used to ensure (5.51),

$$\lim_{n\to\infty} P\left(\sup_{t\in(h,1-h)} \left| \mathcal{T}_{\widehat{\mathcal{K}}}(t) \right| \geq C_T \sqrt{2\left|\log h\right|} \right) = 1.$$

That is, whenever $\mu \in \mathscr{S}_s$, $\sup_{t\in(h,1-h)} \left|\mathcal{T}_{\mathcal{K}}(t)\right|$ will diverge to infinity at a rate no faster than $\sqrt{2\left|\log h\right|}$ and a kink is detected when the condition,

$$\sup_{t\in(h,1-h)} \left| \mathcal{T}_{\widehat{\mathcal{K}}}(t) \right| \geq \sqrt{2\left|\log h\right|},$$

is observed in practice. If a kink is detected through this procedure, the method proceeds to the zero-crossing step.

Zero-crossing step

The same ZCT is used to refine the interval $\widehat{A}_n = (\widehat{t}_*, \widehat{t}^*)$ down to a singular point $\widehat{\lambda}_n$, the estimate of λ. To avoid repetition the reader is referred to the latter part of Section 2.2 for the complete treatment of the ZCT and just the main results are stated here. The estimate of λ is constructed with,

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

Recall again that $\widehat{\mathcal{K}}_h(t) = \mathcal{L}_h(t) + \mathcal{J}_h(t) + \mathcal{Z}_h(t) + \mathcal{b}_h(t)$ and the best possible accuracy occurs if $\delta > 0$ can be chosen as small as possible such that the following inequalities hold,

$$\delta h^{-3} \geq Ch^{s-3} \quad \text{and} \quad \delta h^{-3} \geq Cn^{-1/2}h^{-7/2}. \tag{5.52}$$

The best possible bandwidth that ensures (5.52) holds along with the required bandwidth condition (5.6) is given by,

$$h_* \asymp n^{-1/(2s+1)}.$$

Now apply Lemma 2.1 with this optimal choice of bandwidth and $\delta = \delta_* \asymp h_*^s$ allows the method to obtain an accuracy of order $n^{-s/(2s+1)}$. More specifically, this proves that the estimator satisfies the probabilistic bound,

$$\left|\widehat{\lambda}_n - \lambda\right| = \mathcal{O}_p\left(n^{-s/(2s+1)}\right).$$

One final step remains to rescale $\widehat{\lambda}_n$ to an estimate of the true kink location at θ.

Quantile rescaling step

Recall that $\theta = F^{-1}(\lambda)$. 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 we can obtain an estimator of Q via the empirical quantile function $Q_n = F_n^{-1}$. Estimate θ by, $\widehat{\theta}_n = Q_n(\widehat{\lambda}_n)$. The rate of convergence of this estimator is evaluated below,

$$\begin{aligned}\left|\widehat{\theta}_n - \theta\right| &= \left|Q_n(\widehat{\lambda}_n) - Q(\lambda)\right| \\ &\leq \left|Q_n(\widehat{\lambda}_n) - Q\left(\widehat{\lambda}_n\right)\right| + \left|Q(\widehat{\lambda}_n) - Q(\lambda)\right|.\end{aligned} \tag{5.53}$$

Under Assumption (X), it follows that the derivative of the Quantile function, $Q(t)$, is uniformly bounded on $t \in (a, 1-b)$ for some constants a, b. Indeed, by assumption, there exists an $a, b \in (0,1)$ such that $0 < a < 1-b < 1$ and

$$\inf_{Q(a)<x<Q(1-b)} f_X(x) > 0.$$

Consequently, the derivative of Q given by, $Q^{(1)} = 1/(f \circ Q)$ is uniformly bounded. Therefore, there exists a $\xi \in (0,1)$ such that,

$$Q(\widehat{\lambda}_n) - Q(\lambda) = \left(\widehat{\lambda}_n - \lambda\right) Q^{(1)}\left(\lambda + \xi\left(\widehat{\lambda}_n - \lambda\right)\right) \leq C\left(\widehat{\lambda}_n - \lambda\right) \tag{5.54}$$

The rate of convergence in (5.53) 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 (X), the quantile process involves LRD variables. The rate of convergence of the empirical quantile function has been considered by (Ho and Hsing, 1996, Theorem 5.1) which states,

Theorem 5.1 (Ho and Hsing, 1996). *Let $X_i = \mu_X + \sum_{j=1}^{n} c_j \eta_{i-j}$ be a linear process with coefficients $c_i = i^{-\beta} L(i)$ with $\beta \in \left(\frac{1}{2}, 1\right)$ such that, $\mathbb{E}\eta_1 = 0$ and $\mathbb{E}\eta_1^4 < \infty$. Also assume that F_η is four times differentiable and there exists an $a, b \in (0,1)$ such that $0 < a < 1 - b < 1$ and*

$$\inf_{Q(a) < t < Q(1-b)} f_X(t) > 0.$$

Then for all $t \in (a, 1-b)$,

$$\frac{n^{\alpha_x/2}}{L(n)} \left(Q_n(t) - Q(t)\right) \xrightarrow{\mathscr{D}} \mathcal{N}(0,1).$$

The conditions of the Theorem are satisfied by Assumption (X) so apply Theorem 5.1 (Ho and Hsing, 1996). Then for all $t \in (h + a, 1 - b - h)$,

$$|Q_n(t) - Q(t)| = \mathcal{O}_p\left(n^{-\alpha_x/2} L(n)\right). \tag{5.55}$$

Therefore, using (5.55) and (5.54) in (5.53),

$$\left|\widehat{\theta}_n - \theta\right| = \mathcal{O}_p\left(n^{-s/(2s+1)} \vee \left(n^{-\alpha_x/2} L(n)\right)\right),$$

which proves Theorem 5.1.

Part VI

APPENDIX

A

REGULARLY VARYING FUNCTIONS

The reader is directed to Bingham, Goldie and Teugels (1989); Seneta (1976) for a detailed treatment of regularly varying functions. Some literature exists on regularly varying sequences, although the treatment is less extensive and the interested reader is referred to Bojanic and Seneta (1973); Galambos and Seneta (1973).

Definition A.1. *A function L is said to be slowly varying at infinity if $L : \mathbb{R}^+ \setminus \{0\} \longrightarrow \mathbb{R}^+$ such that for all $t > 0$,*

$$\lim_{x\to\infty} \frac{L(tx)}{L(x)} = 1. \tag{A.1}$$

The above definition leads to the concept of a regularly varying function.

Definition A.2. *A function f is said to be regularly varying at infinity with index p if $f(x) = x^p L(x)$ and L is a slowly varying function.*

Firstly, the famous Theorem of Karamata is given below. First define the two integral expressions,

$$\begin{aligned} L_p(x) &= \int_0^x y^p L(y)\, dy \\ L_p^*(x) &= \int_x^\infty y^p L(y)\, dy \end{aligned} \tag{A.2}$$

The two Lemmas and Theorem are taken from Feller (1971) and we refer the reader to the text for their proof,

Lemma A.1. *If $L > 0$ varies slowly. The integrals defined in (A.2) will converge at $x = \infty$ for $p < -1$ and diverge for $p > -1$.*
If $p > 1$ then L_p varies regularly with exponent $p+1$. If $p < 1$ then L_p^* varies regularly with exponent $p+1$, and this remains true for $p+1 = 0$ if L_{-1}^* exists.

Lemma A.2. *Let L be a slowly varying function. Then, for any $\epsilon > 0$ and all sufficiently large $x > 0$,*

$$x^{-\epsilon} < L(x) < x^{\epsilon}.$$

Theorem A.1. *Let L be a slowly varying function,*

(a) If $p < -1$ then,

$$\frac{L_p^*(n)}{n^{1+p}L(n)} \overset{n\to\infty}{\longrightarrow} -\frac{1}{p+1}.$$

(b) If $p > -1$ then,

$$\frac{L_p(n)}{n^{1+p}L(n)} \xrightarrow{n\to\infty} \frac{1}{1+p}.$$

The key useful properties of Slowly varying functions for our purposes are listed below, beginning with the so called Uniform Convergence Theorem.

Theorem A.2 (Uniform Convergence Theorem)**.** *If L is a slowly varying function, then for every fixed $a, b \in \mathbb{R}^+$ such that $0 < a < b < \infty$, then (A.1) holds uniformly for $t \in [a, b]$.*

A result that will be used extensively is taken from (Seneta, 1976, p18).

Lemma A.3. *For every $p > 0$ and function L that is slowly varying,*

$$x^p L(x) \xrightarrow{x\to\infty} \infty \qquad \text{and} \qquad x^{-p}L(x) \xrightarrow{x\to\infty} 0$$

The proof of the above result is an easy consequence of Lemma A.2. The next Theorem is useful for bounded the regularly varying sequences at the boundaries near zero and infinity and appears in Bojanic and Seneta (1973) and (Seneta, 1976, p20.) respectively,

Theorem A.3. *For every $p > 0$ and $L(\cdot)$ slowly varying,*

$$\sup_{i\geq n}\{i^{-p}L(i)\} \sim n^{-p}L(n) \qquad \text{and} \qquad \sup_{0<i\leq n}\{i^p L(i)\} \sim n^p L(n)$$

A last useful property is the convergence result for Slowly varying sequences and is taken from Bojanic and Seneta (1973)

Theorem A.4. *A positive sequence $\{L(n)\}_{n=1}^{\infty}$ is slowly varying if and only if,*

$$\lim_{n\to\infty}\frac{1}{n^{1+p}}\sum_{i=1}^{n} i^p L(i) = \frac{1}{1+p},$$

for some $p > -1$.

Theorem A.5. *Given a positive slowly varying sequence $\{L(n)\}_{n=1}^{\infty}$,*

$$\lim_{n\to\infty}\sum_{i=1}^{n} i^p L(i) = C,$$

when $p < -1$.

Proof. The result follows from Lemma A.2. Indeed, choose $\epsilon > 0$ such that $p + \epsilon < -1$ and then by Lemma A.2 there exists an $n_0 \in \mathbb{N}$ such that for all $i \geq n_0$, $L(i) < i^{\epsilon}$ which implies,

$$\sum_{i=n_0}^{\infty} i^p L(i) < \sum_{i=n_0}^{\infty} i^{p+\epsilon} = C.$$

□

B

LONG RANGE DEPENDENT RANDOM VARIABLES

In this section, the technical details and analysis of the properties of causal LRD linear processes is studied. In particular, the focus will be on the asymptotic behaviour of the covariance structure of the process and other functionals.

Assume throughout this chapter that $\{X_i\}_{i=1}^n$ is a causal LRD linear process (see Definition 4.1) with parameter $\alpha \in (0,1]$. As stated in Section 4.2.1, the asymptotic covariance structure of X_i satisfies,

$$\mathbb{C}\mathrm{ov}\left(X_i, X_{i+n}\right) \sim C_\alpha n^{-\alpha} L^2(n). \tag{B.1}$$

More importantly, it is commonly claimed in the literature that (B.1) can be proven by routine applications of Karamata's Theorem (see Theorem A.1) which for our purposes states,

$$\int_0^x y^p L(y)\, dy \sim \frac{x^{p+1}L(x)}{p+1} \quad \text{for } p > -1. \tag{B.2}$$

The structure is more delicate than an application of Karamata's Theorem in (B.2). To see this, consider the ratio below,

$$\begin{aligned} \frac{\mathbb{E}\left(X_0 - \mu_X\right)\left(X_n - \mu_X\right)}{n^{-\alpha}L^2(n)} &= \frac{\sigma_\eta^2 n^\alpha}{L^2(n)} \sum_{i=0}^{\infty} c_i c_{i+n} \\ &= \frac{\sigma_\eta^2 n^\alpha}{L^2(n)} \left(c_n + \sum_{i=1}^{\infty} c_i c_{i+n} \right) \\ &= \frac{\sigma_\eta^2 n^\alpha}{L^2(n)} \left(n^{-(1+\alpha)/2} L(n) + \sum_{i=1}^{\infty} (i^2 + ni)^{-(1+\alpha)/2} L(i) L(i+n) \right) \\ &= \sigma_\eta^2 \frac{n^{(\alpha-1)/2}}{L(n)} + \sigma_\eta^2 \frac{n^\alpha}{L^2(n)} \sum_{i=1}^{\infty} (i^2 + ni)^{-(1+\alpha)/2} L(i) L(i+n). \end{aligned} \tag{B.3}$$

The first term in (B.3) is asymptotically negligible (see proof of Lemma B.1). So the asymptotic behaviour as n approaches infinity, is determined by the summation term. The theorem of Karamata involves an integral of a slowly varying function and the asymptotic relation involves the upper limit of that integral (see (B.2)). In a case of direct application it would involve n being the upper limit in the integral.

On the other hand, the summation term in (B.3) involves an infinite summation of a slowly varying sequence where the asymptotics of n are focused inside the summand $(i^2 + ni)^{-\beta} L(i) L(i+n)$. Therefore, the theorem of Karamata cannot be applied in the direct sense and the proof requires looking at the asymptotic behaviour of a summation involving regularly varying sequences. Moreover, as shown in Bojanic and Seneta (1973) the analysis of regularly varying sequences share analogous results to regularly varying functions. However, the development and proofs of the results for regularly varying sequences are in general *not* a simple extension of arguments from the regularly varying functions. This is shown to be the case in the proofs of the forthcoming Lemmas.

For purposes of convenience, in this chapter, the dependence parameter β will be defined with $\beta := (1+\alpha)/2$. The first Lemma gives the asymptotic behaviour of the covariance structure of a LRD causal linear process.

Lemma B.1. *Let X_i be a LRD causal linear process that satisfies Definition 4.1. Then the covariance structure of X_i satisfies,*

$$\gamma_n = \mathbb{C}\text{ov}\,(X_0, X_n) \sim \sigma_\eta^2 \left(\int_0^\infty (x^2+x)^{-\beta}\, dx \right)\, n^{-\alpha} L^2(n).$$

The above Lemma is a powerful tool for considering the asymptotic covariance structure of some other functionals of LRD processes such as the ones given in Lemmas B.2 – B.4.

Lemma B.2. *Let X_i be a LRD causal linear process that satisfies Definition 4.1. Then the variance of the sum of a LRD causal linear process satisfies,*

$$\mathbb{V}\text{ar}\left(\sum_{i=1}^n X_i \right) \sim \frac{2\sigma_\eta^2}{(1-\alpha)(2-\alpha)} \left(\int_0^\infty (x^2+x)^{-\beta}\, dx \right)\, n^{2-\alpha} L^2(n).$$

Lemma B.3. *Let X_i be a LRD causal linear process that satisfies Definition 4.1. Then square of the process, X_i^2 is second-order stationary and satisfies,*

$$\mathbb{C}\text{ov}(X_0^2, X_k^2) = 2\mathbb{C}\text{ov}^2\,(X_0, X_k) - 2\sigma_\eta^2 \sum_{j=0}^\infty c_j^2 c_{j+k}^2 \sim 2\sigma_\eta^4 \left(\int_0^\infty (x^2+x)^{-\beta}\, dx \right)^2 k^{-2\alpha} L^4(k).$$

Lemma B.4. *Let X_i be a LRD causal linear process that satisfies Definition 4.1. Then the variance of the sum of squares of the process satisfies,*

$$\mathbb{V}\text{ar}\left(\sum_{i=1}^n X_i^2 \right) \sim \begin{cases} \dfrac{4\sigma_\eta^4}{(1-2\alpha)(2-2\alpha)} \left(\displaystyle\int_0^\infty (x^2+x)^{-\beta}\, dx \right)^2 n^{2-2\alpha} L^4(n), & \textit{if } 0 < \alpha < \frac{1}{2} \\ 4\sigma_\eta^4 \left(\displaystyle\int_0^\infty (x^2+x)^{-\beta}\, dx \right)^2 n L^*(n), & \textit{if } \alpha = \frac{1}{2} \\ n \left(\mathbb{V}\text{ar} X_0^2 + 2 \displaystyle\sum_{i=0}^\infty \mathbb{C}\text{ov}\,(X_0^2, X_i^2) \right), & \textit{if } \frac{1}{2} < \alpha < 1. \end{cases}$$

where L^* is defined,

$$L^*(n) := \sum_{i=1}^{n} \left(1 - \frac{i}{n}\right) i^{-1} L^4(i).$$

Proof of Lemma B.1 . The proof will follow by showing that the ratio of $\mathbb{C}\mathrm{ov}\,(X_0, X_n)$ and $n^{-\alpha} L^2(n)$ is asymptotically equal to the integral constant,

$$\sigma_\eta^2 \int_0^\infty (x^2 + x)^{-\beta}\, dx.$$

Recall from (B.3) that

$$\frac{\mathbb{C}\mathrm{ov}(X_0, X_n)}{n^{-\alpha} L^2(n)} = \sigma_\eta^2 \frac{n^{\beta-1}}{L(n)} + \sigma_\eta^2 \frac{n^\alpha}{L^2(n)} \sum_{i=1}^{\infty} (i^2 + ni)^{-\beta} L(i) L(i+n)$$

By construction, $1 - \beta < 0$ which in combination with Lemma A.3 and the inverse of a slowly varying function is still a slowly varying function implies that the first term of (B.3) converges to zero. The second term of (B.3) involves an infinite sum and for later use define the summation set,

$$S(n, a, b) = S(a, b) := \left\{ a, a + \frac{1}{n}, a + \frac{2}{n}, \ldots, b \right\}, \tag{B.4}$$

where $a, b \in \mathbb{R}^+$ with $a < b$. Then the quantity in (B.3) can be re-expressed by rescaling the dummy variable of summation with the choice $i = nj$. Doing this will allow us to appeal to some of the relevant theorems and properties of slowly varying functions,

$$\begin{aligned}
\frac{\gamma_n}{n^{-\alpha} L^2(n)} &= o(1) + \sigma_\eta^2 \frac{n^\alpha}{L^2(n)} \sum_{nj=1}^{\infty} ((nj)^2 + n(nj))^{-\beta} L(nj) L(nj+n) \\
&= o(1) + \sigma_\eta^2 \frac{n^{\alpha-2\beta}}{L^2(n)} \sum_{j \in S(n^{-1}, \infty)} (j^2 + j)^{-\beta} L(nj) L(n(j+1)) \\
&= o(1) + \sigma_\eta^2 \frac{n^{-1}}{L^2(n)} \sum_{j \in S(n^{-1}, \infty)} (j^2 + j)^{-\beta} L(nj) L(n(j+1)).
\end{aligned} \tag{B.5}$$

To show that ratio converges to the integral, $\int_0^\infty (x^2 + x)^{-\beta}\, dx$, it will be shown that the asymptotic distance of the ratio to the integral is zero. However, the uniform convergence property of slowly varying functions requires that the argument of the slowly varying function is separated away from zero and infinity. Therefore, the distance needs to be carefully split up into two separate cases. Then the methodology applied is similar but slightly more complex than the continuous integral case given in (Seneta, 1976, Theorem 2.6 and Theorem 2.7). The analogous integral case was proved by Seneta (2009) and the adaptation to the discrete summation case is explored here.

To be able to adapt the proof, first choose an $\epsilon > 0$ such that $1 - 2\beta + \epsilon < 0 < 1 - \beta - \epsilon$. With this choice of ϵ, then for any $\delta > 0$ the following integrals are finite,

$$\int_0^\delta x^{-\epsilon}(x^2+x)^{-\beta}\,dx < \infty \qquad \text{and} \qquad \int_\delta^\infty x^{\epsilon}(x^2+x)^{-\beta}\,dx < \infty. \tag{B.6}$$

Indeed, for the LHS of (B.6), the only danger of the integral becoming infinite is at the origin and the integral is bounded as follows,

$$\begin{aligned}\int_0^\delta x^{-\epsilon}(x^2+x)^{-\beta}\,dx &= \int_0^\delta x^{-\epsilon-\beta}(x+1)^{-\beta}\,dx \\ &< \int_0^\delta x^{-\epsilon-\beta}\,dx \\ &< \frac{\delta^{1-\epsilon-\beta}}{1-\epsilon-\beta} < \infty.\end{aligned}$$

The last line is finite due to the choice of ϵ. Similarly, for the second term of (B.6), the only danger of the integral becoming infinite is at right limit point at infinity. The integral is similarly bounded as follows,

$$\begin{aligned}\int_\delta^\infty x^{\epsilon}(x^2+x)^{-\beta}\,dx &= \int_\delta^\infty x^{\epsilon-2\beta}\left(1+\frac{1}{x}\right)^{-\beta}\,dx \\ &< \int_\delta^\infty x^{\epsilon-2\beta}\,dx \\ &< \frac{\delta^{1+\epsilon-2\beta}}{2\beta-1-\epsilon} < \infty.\end{aligned}$$

Where again, the last line is finite due to the choice of ϵ. Using (B.5), the ratio can then be split into three parts,

$$\begin{aligned}\frac{\gamma_n}{n^{-\alpha}L^2(n)} = o(1) &+ \sigma_\eta^2\frac{n^{-1}}{L^2(n)}\sum_{j\in S(n^{-1},\delta)}(j^2+j)^{-\beta}L(nj)L(n(j+1)) \\ &+ \sigma_\eta^2\frac{n^{-1}}{L^2(n)}\sum_{j\in S(\delta+n^{-1},\infty)}(j^2+j)^{-\beta}L(nj)L(n(j+1)).\end{aligned}$$

Separately it will be shown that,

$$\lim_{n\to\infty} n^{-1}\sum_{j\in S(n^{-1},\delta)}(j^2+j)^{-\beta}\frac{L(nj)L(n(j+1))}{L^2(n)} - \int_0^\delta (x^2+x)^{-\beta}\,dx = 0 \quad \text{and} \tag{B.7}$$

$$\lim_{n\to\infty} n^{-1}\sum_{j\in S(\delta+n^{-1},\infty)}(j^2+j)^{-\beta}\frac{L(nj)L(n(j+1))}{L^2(n)} - \int_\delta^\infty (x^2+x)^{-\beta}\,dx = 0. \tag{B.8}$$

In both cases given in (B.7) and (B.8) a term involving a ratio of slowly varying functions appears. The following bound will be useful,

$$
\begin{aligned}
&\left|\frac{L(nj)L(n(j+1))}{L^2(n)} - 1\right| \\
&= \left|\left(\frac{L(nj)}{L(n)} - 1\right)\left(\frac{L(n(j+1))}{L(n)} - 1\right) + \frac{L(nj)}{L(n)} - 1 + \frac{L(n(j+1))}{L(n)} - 1\right| \\
&\leq \left|\frac{L(nj)}{L(n)} - 1\right| \times \left|\frac{L(n(j+1))}{L(n)} - 1\right| + \left|\frac{L(nj)}{L(n)} - 1\right| + \left|\frac{L(n(j+1))}{L(n)} - 1\right|. \qquad \text{(B.9)}
\end{aligned}
$$

Consider $0 < j < \infty$ with $j \in \mathbb{R}$, then take the lim sup of (B.9) with respect to n and by Theorem A.2 (Uniform Convergence Theorem),

$$
\begin{aligned}
\limsup_{n\to\infty}\left|\frac{L(nj)L(n(j+1))}{L^2(n)} - 1\right| &\leq \limsup_{n\to\infty}\left|\frac{L(nj)}{L(n)} - 1\right| \times \limsup_{n\to\infty}\left|\frac{L(n(j+1))}{L(n)} - 1\right| \\
&\quad + \limsup_{n\to\infty}\left|\frac{L(nj)}{L(n)} - 1\right| + \limsup_{n\to\infty}\left|\frac{L(n(j+1))}{L(n)} - 1\right| \\
&= 0. \qquad \text{(B.10)}
\end{aligned}
$$

Furthermore, the convergence in (B.10) is uniform. Now the proof of (B.7) is given. To be able to exploit (B.10), consider a value $m \in \mathbb{R}$ such that $0 < m < \delta$ and then split (B.7) further as follows,

$$
\begin{aligned}
&\left|\sum_{j\in S(n^{-1},\delta)} n^{-1}(j^2+j)^{-\beta}\frac{L(nj)L(n(j+1))}{L^2(n)} - \int_0^\delta (x^2+x)^{-\beta}\,dx\right| \\
&\qquad \leq \left|\sum_{j\in S(m+n^{-1},\delta)} n^{-1}(j^2+j)^{-\beta}\left(\frac{L(nj)L(n(j+1))}{L^2(n)} - 1\right)\right| \\
&\qquad\quad + \left|\sum_{j\in S(n^{-1},m)} n^{-1}(j^2+j)^{-\beta}\frac{L(nj)L(n(j+1))}{L^2(n)}\right| \\
&\qquad\quad + \left|\sum_{j\in S(m+n^{-1},\delta)} n^{-1}(j^2+j)^{-\beta} - \int_0^\delta (x^2+x)^{-\beta}\,dx\right|. \qquad \text{(B.11)}
\end{aligned}
$$

The last term on the RHS of (B.11), is the distance between the integral and its finite analog, therefore it follows,

$$
\begin{aligned}
&\lim_{m\downarrow 0}\lim_{n\to\infty}\left|\sum_{j\in S(m+n^{-1},\delta)} n^{-1}(j^2+j)^{-\beta} - \int_0^\delta (x^2+x)^{-\beta}\,dx\right| \\
&\quad = \lim_{m\downarrow 0}\left|\int_m^\delta (x^2+x)^{-\beta}\,dx - \int_0^\delta (x^2+x)^{-\beta}\,dx\right| = 0. \qquad \text{(B.12)}
\end{aligned}
$$

Consider now the first term on the RHS of (B.11). The summand terms are positive and the arguments in the summand of this term are well separated from zero and infinity, meaning that (B.10) can be used. This means, that for any $\tau > 0$ there exists a $N \in \mathbb{N}$ such that for all $n \geq N$,

$$\begin{aligned}
&\left| \sum_{j \in S(m+n^{-1},\delta)} n^{-1}(j^2+j)^{-\beta} \left(\frac{L(nj)L(n(j+1))}{L^2(n)} - 1 \right) \right| \\
&= \sum_{j \in S(m+n^{-1},\delta)} n^{-1}(j^2+j)^{-\beta} \left| \frac{L(nj)L(n(j+1))}{L^2(n)} - 1 \right| \\
&< \tau \sum_{j \in S(m+n^{-1},\delta)} n^{-1}(j^2+j)^{-\beta}.
\end{aligned} \tag{B.13}$$

However, the summation term in (B.13) converges to the integral,

$$\int_0^\delta (x^2+x)^{-\beta}\, dx,$$

which is finite due to (B.6) and this implies that the summation in (B.13) is finite as well. Thus, for an arbitrary $\tau > 0$ there exists a finite $N \in \mathbb{R}^+$ such that for all $n \geq N$,

$$\left| \sum_{j \in S(m+n^{-1},\delta)} n^{-1}(j^2+j)^{-\beta} \left(\frac{L(nj)L(n(j+1))}{L^2(n)} - 1 \right) \right| \leq C\tau, \tag{B.14}$$

where C is independent of n, m and δ. Now to deal with the second term of (B.11) the slowly varying functions need to be bounded. By Theorem A.3 there exists a constant $C > 0$ that is independent of n such that,

$$\sup_{0<i\leq n} \left\{ i^{\epsilon/2} L(i) \right\} \leq C n^{\epsilon/2} L(n). \tag{B.15}$$

Thus the second term of (B.11) can be bounded. Indeed, apply (B.15) and (B.6) as follows,

$$\begin{aligned}
&\sum_{j\in S(n^{-1},m)} n^{-1}(j^2+j)^{-\beta}\frac{L(nj)L(n(j+1))}{L^2(n)} \\
&= \frac{n^{-\epsilon}}{L^2(n)}\sum_{j\in S(n^{-1},m)} n^{-1}j^{-\epsilon}(j^2+j)^{-\beta}(nj)^{\epsilon/2}L(nj)\,(n(j+1))^{\epsilon/2}\,L(n(j+1))\left(\frac{n(j+1)}{nj}\right)^{-\epsilon/2} \\
&\le \frac{n^{-\epsilon}}{L^2(n)}\sum_{j\in S(n^{-1},m)} n^{-1}j^{-\epsilon}(j^2+j)^{-\beta}\left\{\sup_{0<j\le m}(nj)^{\epsilon/2}L(nj)\right\}\left\{\sup_{0<j\le m}(n(j+1))^{\epsilon/2}\,L(n(j+1))\right\} \\
&\qquad\qquad \times\left\{\sup_{0<j\le m}\left(1+\frac{1}{j}\right)^{-\epsilon/2}\right\} \\
&\le \frac{n^{-\epsilon}}{L^2(n)}\sum_{j\in S(n^{-1},m)} n^{-1}j^{-\epsilon}(j^2+j)^{-\beta}\left\{\sup_{0<nj\le nm}(nj)^{\epsilon/2}L(nj)\right\}\left\{\sup_{0<\frac{y}{n}-1\le m}y^{\epsilon/2}L(y)\right\} \\
&\le C\frac{n^{-\epsilon}}{L^2(n)}(nm)^{\epsilon/2}L(nm)\sum_{j\in S(n^{-1},m)} n^{-1}j^{-\epsilon}(j^2+j)^{-\beta}\left\{\sup_{0<y\le nm}y^{\epsilon/2}L(y)\right\} \\
&\le C\frac{n^{-\epsilon}}{L^2(n)}(nm)^{\epsilon}L^2(nm)\sum_{j\in S(n^{-1},m)} n^{-1}j^{-\epsilon}(j^2+j)^{-\beta} \\
&= C\frac{L^2(nm)}{L^2(n)}m^{\epsilon}\sum_{j\in S(n^{-1},m)} n^{-1}j^{-\epsilon}(j^2+j)^{-\beta}. \qquad \text{(B.16)}
\end{aligned}$$

Taking the lim sup over the index n in (B.16) and use (B.6) and Theorem A.2,

$$\begin{aligned}
&\limsup_{n\to\infty}\sum_{j\in S(n^{-1},m)} n^{-1}(j^2+j)^{-\beta}\frac{L(nj)L(n(j+1))}{L^2(n)} \\
&\qquad\le \limsup_{n\to\infty} C\frac{L^2(nm)}{L^2(n)}m^{\epsilon}\sum_{j\in S(n^{-1},m)} n^{-1}j^{-\epsilon}(j^2+j)^{-\beta} \\
&\qquad\le Cm^{\epsilon}. \qquad \text{(B.17)}
\end{aligned}$$

However, by (B.6) the integral on the RHS of (B.17) is finite and then take the lim sup of (B.17) over the index m,

$$\limsup_{m\downarrow 0} Cm^{\epsilon}\int_0^m x^{-\epsilon}(x^2+x)^{-\beta}\,dx = 0. \qquad \text{(B.18)}$$

Therefore, (B.16), (B.17) and (B.18) ensure that the second term of (B.11) converges to zero. This combined with (B.14) and (B.12) ensure that all terms in (B.11) converge to zero and consequently complete the proof of (B.7). To prove the result, it remains to prove (B.8) which is very similar but has a few subtle variations and the full proof is given below for completeness. Again, to be able to exploit (B.10), consider a value $m \in \mathbb{R}$ such that $\delta < m < \infty$ and then split (B.8) as follows,

$$\begin{aligned}
&\left| \sum_{j \in S(\delta+n^{-1},\infty)} n^{-1}(j^2+j)^{-\beta} \frac{L(nj)L(n(j+1))}{L^2(n)} - \int_\delta^\infty (x^2+x)^{-\beta}\,dx \right| \\
&\quad \leq \left| \sum_{j \in S(\delta+n^{-1},m)} n^{-1}(j^2+j)^{-\beta} \left(\frac{L(nj)L(n(j+1))}{L^2(n)} - 1 \right) \right| \\
&\qquad + \left| \sum_{j \in S(m+n^{-1},\infty)} n^{-1}(j^2+j)^{-\beta} \frac{L(nj)L(n(j+1))}{L^2(n)} \right| \\
&\qquad\quad + \left| \sum_{j \in S(\delta+n^{-1},m)} n^{-1}(j^2+j)^{-\beta} - \int_\delta^\infty (x^2+x)^{-\beta}\,dx \right|.
\end{aligned} \tag{B.19}$$

The last term on the RHS of (B.19), is the distance between the integral and its finite analog, therefore it follows,

$$\begin{aligned}
&\lim_{m\to\infty} \lim_{n\to\infty} \left| \sum_{j \in S(\delta+n^{-1},m)} n^{-1}(j^2+j)^{-\beta} - \int_\delta^\infty (x^2+x)^{-\beta}\,dx \right| \\
&\quad = \lim_{m\to\infty} \left| \int_\delta^n (x^2+x)^{-\beta}\,dx - \int_\delta^\infty (x^2+x)^{-\beta}\,dx \right| = 0.
\end{aligned} \tag{B.20}$$

Consider now the first term on the RHS of (B.19). Again, the summand terms are positive and the arguments in the summand of this term are well separated from zero and infinity, meaning that (B.10) can be used. So for any $\tau > 0$ there exists a $N \in \mathbb{R}^+$ such that for all $n \geq N$,

$$\begin{aligned}
&\left| \sum_{j \in S(\delta+n^{-1},m)} n^{-1}(j^2+j)^{-\beta} \left(\frac{L(nj)L(n(j+1))}{L^2(n)} - 1 \right) \right| \\
&\leq \sum_{j \in S(\delta+n^{-1},m)} n^{-1}(j^2+j)^{-\beta} \left| \frac{L(nj)L(n(j+1))}{L^2(n)} - 1 \right| \\
&< \tau \sum_{j \in S(\delta+n^{-1},m)} n^{-1}(j^2+j)^{-\beta}.
\end{aligned} \tag{B.21}$$

Again, the summation in (B.21) is bounded by a finite integral, so it follows, for any $\tau > 0$ there exists a $n \in \mathbb{R}^+$ such that for all $n \geq N$,

$$\left| \sum_{j \in S(m+n^{-1},\delta)} n^{-1}(j^2+j)^{-\beta} \left(\frac{L(nj)L(n(j+1))}{L^2(n)} - 1 \right) \right| \leq C\tau, \tag{B.22}$$

where C is independent of m, n and δ. As before, to deal with the second term of (B.19) the slowly varying functions need to be bounded and by Theorem A.3 there exists a constant $C > 0$ that is independent of n such that,

$$\sup_{i \geq n} \left\{ i^{-\epsilon/2} L(i) \right\} \leq C n^{-\epsilon/2} L(n). \tag{B.23}$$

Thus the second term of (B.19) can be bounded by applying (B.23) and (B.6),

$$\begin{aligned}
&\sum_{j \in S(m+n^{-1},\infty)} n^{-1}(j^2+j)^{-\beta} \frac{L(nj)L(n(j+1))}{L^2(n)} \\
&= \frac{n^\epsilon}{L^2(n)} \sum_{j \in S(m+n^{-1},\infty)} n^{-1} j^\epsilon (j^2+j)^{-\beta} (nj)^{-\epsilon/2} L(nj) \left(n(j+1)\right)^{-\epsilon/2} L(n(j+1)) \left(\frac{n(j+1)}{nj}\right)^{\epsilon/2} \\
&\leq \frac{n^\epsilon}{L^2(n)} \sum_{j \in S(m+n^{-1},\infty)} n^{-1} j^\epsilon (j^2+j)^{-\beta} \left\{ \sup_{m \leq j < \infty} (nj)^{-\epsilon/2} L(nj) \right\} \left\{ \sup_{m \leq j < \infty} \left(n(j+1)\right)^{-\epsilon/2} L(n(j+1)) \right\} \\
&\qquad \times \left\{ \sup_{m \leq j < \infty} \left(1 + \frac{1}{j}\right)^{\epsilon/2} \right\} \\
&= \frac{n^\epsilon}{L^2(n)} \left(1 + \frac{1}{m}\right)^{\epsilon/2} \sum_{j \in S(m+n^{-1},\infty)} n^{-1} j^\epsilon (j^2+j)^{-\beta} \left\{ \sup_{nm \leq nj < \infty} (nj)^{-\epsilon/2} L(nj) \right\} \left\{ \sup_{m \leq \frac{y}{n} - 1 < \infty} y^{-\epsilon/2} L(y) \right\} \\
&\leq C \frac{n^\epsilon}{L^2(n)} (nm)^{-\epsilon/2} L(nm) \left(1 + \frac{1}{m}\right)^{\epsilon/2} \sum_{j \in S(m+n^{-1},\infty)} n^{-1} j^\epsilon (j^2+j)^{-\beta} \left\{ \sup_{nm \leq y < \infty} y^{-\epsilon/2} L(y) \right\} \\
&\leq C \frac{L^2(nm)}{L^2(n)} n^\epsilon (nm)^{-\epsilon} \left(1 + \frac{1}{m}\right)^{\epsilon/2} \sum_{j \in S(n+n^{-1},\infty)} n^{-1} j^\epsilon (j^2+j)^{-\beta} \\
&= C \frac{L^2(nm)}{L^2(n)} m^{-\epsilon} \left(1 + \frac{1}{m}\right)^{\epsilon/2} \sum_{j \in S(m+n^{-1},\infty)} n^{-1} j^\epsilon (j^2+j)^{-\beta}.
\end{aligned} \tag{B.24}$$

Taking the $\limsup$ over the index n in (B.24) and use (B.6) and Theorem A.2,

$$\begin{aligned}
\limsup_{n \to \infty} &\sum_{j \in S(m+n^{-1},\infty)} n^{-1}(j^2+j)^{-\beta} \frac{L(nj)L(n(j+1))}{L^2(n)} \\
&\leq \limsup_{n \to \infty} C \frac{L^2(nm)}{L^2(n)} m^{-\epsilon} \left(1 + \frac{1}{m}\right)^{\epsilon/2} \sum_{j \in S(m+n^{-1},\infty)} n^{-1} j^\epsilon (j^2+j)^{-\beta} \\
&\leq C m^{-\epsilon},
\end{aligned} \tag{B.25}$$

for all $m \geq m_0$ where $m_0 \in \mathbb{R}^+$ is sufficiently large. Therefore then take the $\limsup$ of (B.25) over the index m,

$$\limsup_{m \to \infty} \limsup_{n \to \infty} \sum_{j \in S(m+n^{-1},\infty)} n^{-1}(j^2+j)^{-\beta} \frac{L(nj)L(n(j+1))}{L^2(n)} \leq C \limsup_{m \to \infty} m^{-\epsilon} = 0. \tag{B.26}$$

Therefore, (B.24), (B.25) and (B.26) ensure that the second term of (B.19) converges to zero. This combined with (B.22) and (B.20) ensure that (B.19) converges to zero and consequently complete the proof of (B.8) and Lemma B.1 is proven. □

Proof of Lemma B.2 . From the proof of Lemma B.1 it was shown that, $\mathbb{C}\text{ov}\,(X_0, X_k) = \sigma_\eta^2 \sum_{j=0}^{\infty} c_j c_{j+k}$, which means that X_i is second-order stationary. Since the variance operator is location invariant, without loss of generality assume $\mu_X = 0$ then,

$$\begin{aligned}
\mathbb{V}\text{ar}\left(\sum_{i=1}^{n} X_i\right) &= \mathbb{E}\left(\sum_{i=1}^{n} X_i\right)^2 \\
&= \sum_{i=1}^{n} \mathbb{E}X_i^2 + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} \mathbb{E}X_i X_j \\
&= n\mathbb{E}X_0^2 + 2\sum_{i=1}^{n-1} (n-i)\mathbb{E}X_0 X_i \\
&= n + 2\sum_{i=1}^{m-1} (n-i)\mathbb{E}X_0 X_i + 2\sum_{i=m}^{n-1} (n-i)\mathbb{E}X_0 X_i, \qquad \text{(B.27)}
\end{aligned}$$

for some $m \in \mathbb{N}^+$ such that $0 < m < n$. In particular, to be able to use Lemma B.1, consider a value of m such that $m \to \infty$ and $m = o(n^{1-\alpha-\epsilon})$ where $\epsilon > 0$ is chosen such that $\alpha + \epsilon < 1$. Then, it will be shown that asymptotically, the last term on the RHS of (B.27) dominates and satisfies,

$$\begin{aligned}
\sum_{i=m}^{n-1} (n-i)\mathbb{E}X_0 X_i &\sim \sigma_\eta^2 \left(\int_0^1 (1-x)x^{-\alpha}\,dx\right)\left(\int_0^\infty (x^2+x)^{-\beta}\,dx\right) n^{2-\alpha} L^2(n) \\
&= \frac{\sigma_\eta^2}{(1-\alpha)(2-\alpha)} \left(\int_0^\infty (x^2+x)^{-\beta}\,dx\right) n^{2-\alpha} L^2(n). \qquad \text{(B.28)}
\end{aligned}$$

Indeed, consider the ratio of the terms in (B.28), apply Lemma B.1 and rescale the dummy variable of summation $i = nj$,

$$\begin{aligned}
&\frac{n^{\alpha-2}}{\sigma_\eta^2 L^2(n)} \left(\int_0^\infty (x^2+x)^{-\beta}\,dx\right)^{-1} \sum_{i=m}^{n-1} (n-i)\mathbb{E}X_0 X_i \\
&\quad\sim \frac{n^{\alpha-2}}{L^2(n)} \sum_{i=m}^{n-1} (n-i) i^{-\alpha} L^2(i) \\
&\quad= \frac{n^{\alpha-2}}{L^2(n)} \sum_{j\in S(\frac{m}{n}, 1-n^{-1})} (n-nj)\,(nj)^{-\alpha} L^2(jn) \\
&\quad= \sum_{j\in S(\frac{m}{n}, 1-n^{-1})} n^{-1}\,(1-j)\,j^{-\alpha} \frac{L^2(jn)}{L^2(n)}, \qquad \text{(B.29)}
\end{aligned}$$

where S was defined as the summation set given in (B.4). The result will follow by the same method used in the proof of Lemma B.1, some regularity conditions just need to be checked. Therefore, consider a value of $\frac{m}{n} < \delta < 1$ and choose an $\epsilon > 0$ such that $\alpha + \epsilon < 1$, then the following integral is finite,

$$\begin{aligned}\int_0^\delta x^{-\epsilon}(1-x)x^{-\alpha}\,dx &= \frac{x^{1-\alpha-\epsilon}}{1-\alpha-\epsilon} - \frac{x^{2-\alpha-\epsilon}}{2-\alpha-\epsilon}\Big|_0^\delta \\ &= \frac{\delta^{1-\alpha-\epsilon}}{1-\alpha-\epsilon} - \frac{\delta^{2-\alpha-\epsilon}}{2-\alpha-\epsilon} < \infty. \qquad \text{(B.30)}\end{aligned}$$

Therefore, by (B.29) and (B.30), the argument that was used in the proof of Lemma B.1 applies and ensure that (B.28) holds. The remainder of the proof shows that this term dominates the two remaining terms of (B.27). Due to the fact that L is slowly varying ensures that $1/L$ is also slowly varying since the asymptotic property is preserved. Therefore, Starting with the first term of (B.27), Lemma A.3 implies,

$$\frac{n}{n^{2-\alpha}L^2(n)} = C\frac{n^{\alpha-1}}{L^2(n)} = o(1). \qquad \text{(B.31)}$$

Using Lemma A.3 and the fact that $m = o(n^{1-\alpha-\epsilon})$,

$$\frac{1}{n^{2-\alpha}L^2(n)}\sum_{i=1}^{m-1}(n-i)\mathbb{E}X_0X_i \leq \frac{n^{\alpha-1}m}{L^2(n)} = o(1), \qquad \text{(B.32)}$$

Therefore, (B.28), (B.31) and (B.32) ensure that,

$$\mathbb{V}\text{ar}\left(\sum_{i=1}^n X_i\right) \sim \frac{2\sigma_\eta^2}{(1-\alpha)(2-\alpha)}\left(\int_0^\infty (x^2+x)^{-\beta}\,dx\right) n^{2-\alpha}L^2(n). \qquad \square$$

Proof of Lemma B.3 . To prove the first claim, start with the expectation of the product of two squared processes at lag k,

$$
\begin{aligned}
\mathbb{E}X_0^2X_k^2 &= \mathbb{E}\left(\sum_{j=0}^{\infty} c_j\eta_{-j}\right)^2\left(\sum_{i=0}^{\infty} c_i\eta_{k-i}\right)^2 \\
&= \mathbb{E}\left(\sum_{j=0}^{\infty} c_j^2\eta_{-j}^2 + \sum_{j=0}^{\infty}\sum_{\substack{j'=0\\ j'\neq j}}^{\infty} c_jc_{j'}\eta_{-j}\eta_{-j'}\right)\left(\sum_{i=0}^{\infty} c_i^2\eta_{k-i}^2 + \sum_{i=0}^{\infty}\sum_{\substack{i'=0\\ i'\neq i}}^{\infty} c_ic_{i'}\eta_{k-i}\eta_{k-i'}\right) \\
&= \sum_{j=0}^{\infty}\sum_{i=0}^{\infty} c_j^2\mathbb{E}\eta_{-j}^2\eta_{k-i}^2 + \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{\substack{j'=0\\ j'\neq j}}^{\infty} c_jc_{j'}c_i^2\mathbb{E}\eta_{k-i}^2\eta_{-j}\eta_{-j'} \\
&\quad + \sum_{j=0}^{\infty}\sum_{i=0}^{\infty}\sum_{\substack{i'=0\\ i'\neq i}}^{\infty} c_j^2c_ic_{i'}\mathbb{E}\eta_{-j}^2\eta_{k-i}\eta_{k-i'} + \sum_{j=0}^{\infty}\sum_{\substack{j'=0\\ j'\neq j}}^{\infty}\sum_{i=0}^{\infty}\sum_{\substack{i'=0\\ i'\neq i}}^{\infty} c_jc_{j'}c_ic_{i'}\mathbb{E}\eta_{-j}\eta_{-j'}\eta_{k-i}\eta_{k-i'} \\
&= \sum_{j=0}^{\infty} c_j^2c_{j+k}^2\mathbb{E}\eta_{-j}^4 + \sum_{j=0}^{\infty}\sum_{\substack{i=0\\ i\neq k+j}}^{\infty} c_j^2c_i^2\mathbb{E}\eta_{-j}^2\eta_{k-i}^2 + 2\sum_{j=0}^{\infty}\sum_{\substack{j'=0\\ j'\neq j}}^{\infty} c_jc_{j'}c_{j+k}c_{j'+k}\mathbb{E}\eta_{-j}^2\eta_{-j'}^2 \\
&= \mathbb{E}\eta_0^4\sum_{j=0}^{\infty} c_j^2c_{j+k}^2 + \left(\sigma_\eta^2\sum_{j=0}^{\infty} c_j^2\right)\left(\sigma_\eta^2\sum_{i=0}^{\infty} c_i^2\right) - \mathbb{E}\eta_0^4\sum_{j=0}^{\infty} c_j^2c_{j+k}^2 + 2\sigma_\eta^4\sum_{j=0}^{\infty}\sum_{\substack{j'=0\\ j'\neq j}}^{\infty} c_jc_{j'}c_{j+k}c_{j'+k} \\
&= 1 + 2\sigma_\eta^4\sum_{j=0}^{\infty}\sum_{\substack{j'=0\\ j'\neq j}}^{\infty} c_jc_{j'}c_{j+k}c_{j'+k}.
\end{aligned}
$$

$$
\mathrm{Cov}(X_0^2, X_k^2) = 2\sigma_\eta^4\sum_{j=0}^{\infty}\sum_{\substack{j'=0\\ j'\neq j}}^{\infty} c_jc_{j'}c_{j+k}c_{j'+k}. \tag{B.33}
$$

Compare this to the square of the covariance of the regular LRD causal process given by,

$$
\begin{aligned}
\mathrm{Cov}^2\left(X_0, X_k\right) &= \left(\sigma_\eta^2\sum_{i=0}^{\infty} c_ic_{i+k}\right)^2 \\
&= \sigma_\eta^4\sum_{i=0}^{\infty} c_i^2c_{i+k}^2 + \sigma_\eta^4\sum_{i=0}^{\infty}\sum_{\substack{i'=0\\ i'\neq i}}^{\infty} c_ic_{i'}c_{i+k}c_{i'+k}.
\end{aligned} \tag{B.34}
$$

Substituting (B.34) into (B.33) yields the result for the equality of the first assertion,

$$
\mathrm{Cov}(X_0^2, X_k^2) = 2\mathrm{Cov}^2\left(X_0, X_k\right) - 2\sigma_\eta^2\sum_{j=0}^{\infty} c_j^2c_{j+k}^2. \tag{B.35}
$$

By an application of Lemma B.1, it is clear that,

$$2\mathrm{Cov}^2\left(X_0, X_k\right) \sim 2\sigma_\eta^4 \left(\int_0^\infty (x^2+x)^{-\beta}\, dx\right)^2 k^{-2\alpha} L^4(k).$$

Therefore, the result follows if it can be shown that the last term of (B.35) is asymptotically negligible. Therefore, consider the ratio of the summation and $k^{-2\alpha}L^4(k)$,

$$\begin{aligned}\frac{1}{k^{-2\alpha}L^4(k)}\sum_{i=0}^{\infty} c_i^2 c_{i+k}^2 &= \frac{k^{2\alpha}c_k^2}{L^4(k)} + \frac{k^{2\alpha}}{L^4(k)}\sum_{i=1}^{\infty} c_i^2 c_{i+k}^2 \\ &= \frac{k^{2\alpha}c_k^2}{L^4(k)} + \frac{k^{2\alpha}}{L^4(k)}\sum_{i=1}^{k} c_i^2 c_{i+k}^2 + \frac{k^{2\alpha}}{L^4(k)}\sum_{i=k+1}^{\infty} c_i^2 c_{i+k}^2. \end{aligned} \tag{B.36}$$

The first term on the RHS of (B.36) converges to zero. Indeed, by Lemma A.3,

$$0 \le \frac{k^{2\alpha}c_k^2}{L^4(k)} = \frac{k^{2\beta-2}}{L^2(k)} = o(1).$$

Then deal with the second term on the RHS of (B.36), consider $\epsilon > 0$ such that $\alpha + \epsilon < 1$ then apply Lemma A.2, choosing k large enough such that for all $y \ge k$, $L^2(i+k) < (i+k)^\epsilon$ which implies,

$$\begin{aligned}\frac{k^{2\alpha}}{L^4(k)}\sum_{i=1}^{k} c_i^2 c_{i+k}^2 &= \frac{k^{2\alpha}}{L^4(k)}\sum_{i=1}^{k}(i^2+ki)^{-2\beta}L^2(i)L^2(i+k) \\ &\le \frac{k^{2\alpha-2\beta}}{L^4(k)}\sum_{i=1}^{k} i^{-2\beta}L^2(i)L^2(i+k) \\ &< \frac{k^{\alpha-1}}{L^4(k)}\sum_{i=1}^{k} i^{-2\beta}(i+k)^{\epsilon}L^2(i) \\ &< 2^{\epsilon}\frac{k^{\alpha-1+\epsilon}}{L^4(k)}\sum_{i=1}^{k} i^{-2\beta}L^2(i) \\ &= o(1). \end{aligned}$$

The last line follows due to Theorem A.5 and Lemma A.3. To deal with the last term of (B.36), rescale the dummy variable of summation by using the substitution $j = ik$ and recall the summation set, S, defined in (B.4), then the last term becomes,

$$\begin{aligned}\frac{1}{k^{-2\alpha}L^4(k)}\sum_{i=k+1}^{\infty} c_i^2 c_{i+k}^2 &= \frac{k^{2\alpha}}{L^4(k)}\sum_{j\in S(1+k^{-1},\infty)}((jk)^2+k(jk))^{-2\beta}L^2(jk)L^2(j(k+1)) \\ &= k^{-1}\sum_{j\in S(1+k^{-1},\infty)} k^{-1}(j^2+j)^{-2\beta}\frac{L^2(jk)L^2(j(k+1))}{L^4(k)}. \end{aligned} \tag{B.37}$$

The summation term is very similar to the one that was dealt with in the proof of Lemma B.1. By comparison to the sum in the proof of Lemma B.1, it will converge to the integral,

$$\int_1^\infty (x^2+x)^{-2\beta}\,dx < \infty.$$

Indeed, the proof of this claim will follow by an adapted proof of Lemma B.1. The only conditions that need to be checked are that there exists an $\epsilon > 0$ such that,

$$\int_1^\infty x^\epsilon (x^2+x)^{-2\beta}\,dx < \infty.$$

Indeed, choose an $\epsilon > 0$ such that $1 - 4\beta + \epsilon < 0$ then,

$$\begin{aligned}\int_1^\infty x^\epsilon (x^2+x)^{-2\beta}\,dx &= \int_1^\infty x^{\epsilon-4\beta}\left(1+\frac{1}{x}\right)^{-2\beta} dx \\ &< \int_1^\infty x^{\epsilon-4\beta}\,dx \\ &< \frac{1}{4\beta-1-\epsilon} < \infty.\end{aligned}$$

The last line is finite due to the choice of ϵ. Therefore by the same argument that was used in the proof of Lemma B.1, we have,

$$\lim_{k\to\infty} k^{-1} \sum_{j\in S(1+k^{-1},\infty)} (j^2+j)^{-2\beta}\frac{L^2(jk)L^2(j(k+1))}{L^2(k)} = \int_1^\infty (x^2+x)^{-2\beta}\,dx < \infty.$$

Therefore using the above equation, (B.37) becomes,

$$\frac{1}{k^{-2\alpha}L^4(k)}\sum_{i=k+1}^{\infty} c_i^2 c_{i+k}^2 = \mathcal{O}\left(k^{-1}\right) = o(1). \qquad \square$$

Proof of Lemma B.4 . To prove the claim, first check that $\mathbb{E}X_i^4 < \infty$ to ensure that $\mathbb{V}\mathrm{ar}X_i^2$ is finite.

$$
\begin{aligned}
\mathbb{E}X_0^4 &= \mathbb{E}\left(\sum_{j=0}^{\infty} c_j\eta_{-j}\right)^4 \\
&= \mathbb{E}\left(\left(\sum_{j=0}^{\infty} c_j\eta_{-j}\right)^2\right)^2 \\
&= \mathbb{E}\left(\sum_{j=0}^{\infty} c_j^2\eta_{-j}^2 + \sum_{j\neq j'} c_jc_{j'}\eta_{-j}\eta_{-j'}\right)^2 \\
&= \sum_{j=0}^{\infty} c_j^4\mathbb{E}\eta_{-j}^4 + \sum_{j=0}^{\infty}\sum_{\substack{j'=0\\ j'\neq j}}^{\infty} c_j^2c_{j'}^2\mathbb{E}\eta_{-j}^2\mathbb{E}\eta_{-j'}^2 + 2\sum_{j=0}^{\infty}\sum_{\substack{j'=0\\ j'\neq j}}^{\infty} c_j^3c_{j'}\mathbb{E}\eta_{-j}^3\eta_{-j'} + \mathbb{E}\left(\sum_{j=0}^{\infty}\sum_{\substack{j'=0\\ j'\neq j}}^{\infty} c_jc_{j'}\eta_{-j}\eta_{-j'}\right)^2 \\
&= \mathbb{E}\eta_0^4\sum_{j=0}^{\infty} c_j^4 + \sum_{j=0}^{\infty}\sum_{\substack{j'=0\\ j'\neq j}}^{\infty} c_j^2c_{j'}^2\mathbb{E}\eta_{-j}^2\eta_{-j'}^2 + \sum_{j=0}^{\infty}\sum_{\substack{j'=0\\ j'\neq j}}^{\infty}\sum_{k=0}^{\infty}\sum_{\substack{k'=0\\ k'\neq k}}^{\infty} c_jc_{j'}c_kc_{k'}\mathbb{E}\eta_{-j}\eta_{-j'}\eta_{-k}\eta_{-k'} \\
&= \mathbb{E}\eta_0^4\sum_{j=0}^{\infty} c_j^4 + 3\sigma_\eta^4\sum_{j=0}^{\infty}\sum_{\substack{j'=0\\ j'\neq j}}^{\infty} c_j^2c_{j'}^2 \\
&= \mathbb{E}\eta_0^4\sum_{j=0}^{\infty} j^{-4\beta}L^4(j) + 3\sigma_\eta^4\sum_{j=0}^{\infty}\sum_{\substack{j'=0\\ j'\neq j}}^{\infty} j^{-2\beta}j'^{-2\beta}L^2(j)L^2(j') \\
&= \mathbb{E}\eta_0^4\sum_{j=0}^{\infty} j^{-4\beta}L^4(j) + 6\sigma_\eta^4\sum_{j=0}^{\infty} j^{-2\beta}L^2(j)\sum_{j'=j+1}^{\infty} j'^{-2\beta}L^2(j')
\end{aligned}
\tag{B.38}
$$

To show that the above sums are finite appeal to Theorem A.5. The first summation is finite since, $-4 < 4\beta < -2$, so by Theorem A.5,

$$
\sum_{j=0}^{\infty} j^{-4\beta}L^4(j) = C < \infty. \tag{B.39}
$$

Also, $-2 < 2\beta < -1$, so for the second double summation bound from above by removing the restriction on the inner summation and apply Theorem A.5,

$$
\begin{aligned}
\sum_{j=0}^{\infty} j^{-2\beta}L^2(j)\sum_{\substack{j'=0\\ j'\neq j}}^{\infty} j'^{-2\beta}L^2(j') &< \left(\sum_{j=0}^{\infty} j^{-2\beta}L^2(j)\right)\left(\sum_{j'=0}^{\infty} j'^{-2\beta}L^2(j')\right) \\
&= C^2 < \infty.
\end{aligned}
\tag{B.40}
$$

Therefore, (B.40), (B.39) and (B.38) imply that,

$$
\mathbb{V}\mathrm{ar}X_0^2 < \mathbb{E}X_0^4 < \infty \tag{B.41}
$$

It is now possible to evaluate the variance of the sum of squares. Indeed, earlier in Lemma B.3 it was proven that,

$$\mathbb{C}\text{ov}(X_0^2, X_k^2) = 2\mathbb{C}\text{ov}^2\left(X_0, X_k\right) - 2\sigma_\eta^2 \sum_{j=0}^{\infty} c_j^2 c_{j+k}^2,$$

which in combination with $\mathbb{E}X_0X_k = \sigma_\eta^2 \sum_{j=0}^{\infty} c_j c_{j+k}$ proves that the square process, X_i^2 is second-order stationary. Using this and (B.41) yields,

$$\begin{aligned}\mathbb{V}\text{ar}\left(\sum_{i=1}^{n} X_i^2\right) &= \sum_{i=1}^{n} \mathbb{V}\text{ar}X_i^2 + \sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} \mathbb{C}\text{ov}\left(X_i^2, X_j^2\right) \\ &= n\mathbb{V}\text{ar}X_0^2 + 2\sum_{i=1}^{n-1}(n-i)\mathbb{C}\text{ov}\left(X_0^2, X_i^2\right) \\ &= n\mathbb{V}\text{ar}X_0^2 + 2\sum_{i=1}^{m-1}(n-i)\mathbb{C}\text{ov}\left(X_0^2, X_i^2\right) + 2\sum_{i=m}^{n-1}(n-i)\mathbb{C}\text{ov}\left(X_0^2, X_i^2\right),\end{aligned} \tag{B.42}$$

where $0 < m < n$. For any finite m, the first two terms of (B.42) are $\mathcal{O}(n)$. However, by Lemma B.3, $\mathbb{C}\text{ov}\left(X_0^2, X_i^2\right) \sim Ci^{-2\alpha}L^4(i)$ which in combination with Theorem A.3 implies that,

$$\sum_{i=m}^{n-1}(n-i)\mathbb{C}\text{ov}\left(X_0^2, X_i^2\right) \sim Cn^{2-2\alpha}L^4(n),$$

which is larger or smaller order than n if α is smaller or larger than $1/2$. More specifically it will first be shown that in general when $\alpha \neq \frac{1}{2}$,

$$\begin{aligned}\sum_{i=m}^{n-1}(n-i)\mathbb{C}\text{ov}\left(X_0^2, X_i^2\right) &\sim 2\sigma_\eta^2\left(\int_0^1 (1-x)x^{-2\alpha}\,dx\right)\left(\int_0^\infty (x^2+x)^{-\beta}\,dx\right)^2 n^{2-2\alpha}L^4(n) \\ &= \frac{2\sigma_\eta^2}{(1-2\alpha)(2-2\alpha)}\left(\int_0^\infty (x^2+x)^{-\beta}\,dx\right)^2 n^{2-2\alpha}L^4(n).\end{aligned} \tag{B.43}$$

By Lemma B.3 for any $\delta > 0$ there exists a finite $m < n$ such that for all $i \geq m$,

$$\left|\frac{\mathbb{C}\text{ov}\left(X_0^2, X_i^2\right)}{2\sigma_\eta^4\left(\int_0^\infty (x^2+x)^{-\beta}\,dx\right)^2 i^{-2\alpha}L^4(i)} - 1\right| < \delta \tag{B.44}$$

Using (B.44) an upper bound can be constructed such that,

$$\begin{aligned}
&\left(\int_0^\infty (x^2+x)^{-\beta}\, dx \right)^2 \frac{n^{2\alpha-2}}{2\sigma_\eta^4 L^4(n)} \sum_{i=m}^{n-1} (n-i)\mathbb{C}\mathrm{ov}\left(X_0^2, X_i^2\right) \\
&\quad \leq \frac{n^{2\alpha-2}(1+\delta)}{L^4(n)} \sum_{i=m}^{n-1} (n-i) i^{-2\alpha} L^4(i) \\
&\quad = (1+\delta) \sum_{S \in \left(\frac{m}{n}, 1-n^{-1}\right)} n^{-1}(1-j) j^{-2\alpha} \frac{L^4(nj)}{L^4(n)} \qquad \text{(B.45)}
\end{aligned}$$

where S was defined as the summation set given in (B.4). This is again very similar to the result proved earlier, some regularity conditions just need to be checked. Therefore, consider a value of $\frac{m}{n} < \delta < 1$ and choose an $\epsilon > 0$ such that $2\alpha + \epsilon < 1$, then the following integral is finite,

$$\begin{aligned}
\int_0^\delta x^{-\epsilon}(1-x)x^{-2\alpha}\, dx &= \frac{x^{1-2\alpha-\epsilon}}{1-2\alpha-\epsilon} - \frac{x^{2-2\alpha-\epsilon}}{2-2\alpha-\epsilon}\Bigg|_0^\delta \\
&= \frac{\delta^{1-\alpha-\epsilon}}{1-\alpha-\epsilon} - \frac{\delta^{2-\alpha-\epsilon}}{2-\alpha-\epsilon} < \infty.
\end{aligned}$$

Also, by (B.44), a similar procedure can be used to produce a lower bound such that for any $\delta > 0$ there exists a finite $m < n$ such that for all $i \geq m$,

$$\left(\int_0^\infty (x^2+x)^{-\beta}\, dx \right)^{-2} \frac{n^{2\alpha-2}}{2\sigma_\eta^4 L^4(n)} \sum_{i=m}^{n-1} (n-i)\mathbb{C}\mathrm{ov}\left(X_0^2, X_i^2\right) \geq (1-\delta) \sum_{S \in \left(\frac{m}{n}, 1-n^{-1}\right)} n^{-1}(1-j) j^{-2\alpha} \frac{L^4(nj)}{L^4(n)}. \tag{B.46}$$

Thus by (B.46), (B.45) and the fact $\delta > 0$ was arbitrary, the regularity conditions for the similar result in Lemma B.1 hold and ensure that (B.43) is satisfied. Now consider the three cases for α around the value $\frac{1}{2}$.

Case 1: $0 < \alpha < \frac{1}{2}$.

In this case, $n = o(n^{2-2\alpha} L^4(n))$ since there exists an $0 < \epsilon < 2\alpha$ and by Lemma A.2 there is a finite value m such that for all $n \geq m$, $L^4(n) > n^{-\epsilon}$ and consequently,

$$\frac{n}{n^{2-2\alpha} L^4(n)} < \frac{1}{n^{1-2\alpha+\epsilon}} = o(1)$$

Therefore, the last term of (B.42) dominates and ensure the first part of the proof holds.

Case 2: $\alpha = \frac{1}{2}$.

In this case, the third term of (B.42) dominates again. Indeed, by (B.44), for any $\delta > 0$, there exists a finite $m < n$ such that all $i \geq m$,

$$\left| \frac{\sum_{i=m}^{n-1} (n-i)\mathbb{C}\text{ov}\left(X_0^2, X_i^2\right)}{2\sigma_\eta^4 \left(\int_0^\infty (x^2+x)^{-\beta}\, dx \right)^2 nL^*(n)} - 1 \right| < \delta.$$

The other two terms of (B.42) are neglible since they are $\mathcal{O}(n)$ it will be shown that $L^*(n)$ diverges. It will be shown that $L^*(n)$ can be bounded below by an arbitrarily large constant. First choose an $\epsilon > 0$, then by Lemma A.2 there exists an $n_0 \in \mathbb{N}$ such that for all $i \geq n_0$ and $n \geq n_0$,

$$\begin{aligned} \sum_{i=n_0}^{n} \left(1 - \frac{i}{n}\right) i^{-1} L^4(i) &> \sum_{i=n_0}^{n} \left(1 - \frac{i}{n}\right) i^{-1-\epsilon} \\ &= \sum_{i=n_0}^{n} i^{-1-\epsilon} - n^{-1} \sum_{i=n_0}^{n} i^{-\epsilon} \\ &= C_\epsilon - n^{-\epsilon} n^{-1+\epsilon} \sum_{i=n_0}^{n} i^{-\epsilon}, \end{aligned} \tag{B.47}$$

where $C_\epsilon > 0$ diverges as $\epsilon \to 0$. The second term on the RHS of (B.47) converges to zero, by changing variable of summation $i = nj$ it follows,

$$n^{-1+\epsilon} \sum_{i=1}^{n} i^{-\epsilon} = \sum_{j \in S(n^{-1},1)} n^{-1} j^{-\epsilon} \sim \int_0^1 x^{-\epsilon}\, dx = \frac{1}{1-\epsilon} < \infty.$$

Therefore, by the above and (B.47) it follows,

$$\liminf_{n\to\infty} \sum_{i=n_0}^{n} \left(1 - \frac{i}{n}\right) i^{-1} L^4(i) \geq C_\epsilon$$

and consequently for any constant $M \in \mathbb{N}$ there exists an $\epsilon > 0$ such that $M < C_\epsilon < L^*(n)$. So $L^*(n)$ is unbounded and must diverge and completes the proof of the second case.

Case 3: $\frac{1}{2} < \alpha < 1$.

Finally, in this scenario consider the ratio of (B.42) and n,

$$\begin{aligned} \frac{1}{n}\mathbb{V}\text{ar}\left(\sum_{i=1}^{n} X_i^2\right) &= \mathbb{V}\text{ar}X_0^2 + 2\sum_{i=1}^{n-1} \left(1 - \frac{i}{n}\right) \mathbb{C}\text{ov}\left(X_0^2, X_i^2\right) \\ &= \mathbb{V}\text{ar}X_0^2 + 2\sum_{i=1}^{n-1} \mathbb{C}\text{ov}\left(X_0^2, X_i^2\right) - 2n^{-1}\sum_{i=1}^{n-1} i\mathbb{C}\text{ov}\left(X_0^2, X_i^2\right). \end{aligned}$$

The result will follow if it can be shown that when $\frac{1}{2} < \alpha < 1$ the two series in the above equation satisfy,

$$\sum_{i=1}^{n-1} \mathbb{C}\text{ov}\left(X_0^2, X_i^2\right) \overset{n\to\infty}{\longrightarrow} C, \tag{B.48}$$

$$\sum_{i=1}^{n-1} i\mathbb{C}\text{ov}\left(X_0^2, X_i^2\right) = o(n). \tag{B.49}$$

Starting with (B.48), the tail sums will be shown to converge to zero as $n \to \infty$. Let $\delta > 0$ be arbitrary, then by Lemma B.3, there exists an $m \in \mathbb{N}$ such that for all $i \geq m$,

$$\left| \frac{\mathbb{C}\text{ov}\left(X_0^2, X_i^2\right)}{2\sigma_\eta^4 \left(\int_0^\infty (x^2 + x)^{-\beta}\, dx\right) i^{-2\alpha} L^4(i)} - 1 \right| \leq \delta. \tag{B.50}$$

Also, consider the following series and apply Theorem A.3 and Lemma A.3

$$\sum_{i=1}^{n} i^{-2\alpha} L^4(i) \overset{n\to\infty}{\longrightarrow} n^{1-2\alpha} L^4(n) = o(1), \tag{B.51}$$

when $\frac{1}{2} < \alpha < 1$. Therefore, (B.50) and (B.51) ensure that the tail series of (B.48) converge to zero which implies that (B.48) must hold. Applying a similar argument will show (B.49). Consider the following series and apply Theorem A.3 and Lemma A.3,

$$\sum_{i=1}^{n} i^{1-2\alpha} L^4(i) \sim n^{2-2\alpha} L^4(n) = o(n), \tag{B.52}$$

since $\frac{1}{2} < \alpha < 1$. Therefore, (B.52) in combination with (B.50) ensure that (B.49) holds since $\delta > 0$ was arbitrary. So, consider the ratio of (B.42) and n and use (B.48) and (B.49) implies,

$$\begin{aligned} \frac{1}{n}\mathbb{V}\text{ar}\left(\sum_{i=1}^{n} X_i^2\right) &= \mathbb{V}\text{ar}X_0^2 + 2\sum_{i=1}^{n-1}\left(1 - \frac{i}{n}\right)\mathbb{C}\text{ov}\left(X_0^2, X_i^2\right) \\ &\overset{n\to\infty}{\longrightarrow} \mathbb{V}\text{ar}X_0^2 + 2\sum_{i=1}^{\infty}\mathbb{C}\text{ov}\left(X_0^2, X_i^2\right) \end{aligned}$$

□

C

LEVEL OF DEPENDENCE

In this section we consider the level of dependence for use in obtaining probabilistic bounds on the LRD processes and use the techniques of Wu (2007). We use a particular process obtain bounds for general LRD processes and the applications will be seen in the succeeding sections. Firstly, some extra smoothness conditions on the density of the η_i process are given to ensure some regularity conditions for the density and conditional density of X_i.

Lemma C.1. *Suppose X_i is a LRD causal process satisfying Definition 4.1 with parameter $\alpha \in (0,1)$ with the associated set of sigma fields $\mathcal{F}_i = \sigma(\eta_i, \eta_{i-1}, \ldots)$ and the density of the η_i variables, $f_\eta(t)$, is Lipschitz continuous. Then,*

1. *The conditional density function of X, $f_X\left(t|\, \mathcal{F}_{i-1}\right)$, exists. Furthermore, it is Lipschitz continuous and therefore bounded.*
2. *The density function of X_i, f_X, exists and is Lipschitz continuous. Furthermore it satisfies,*

$$f_X(t) = \mathbb{E}\, f_X(t|\, \mathcal{F}_{i-1}) = \mathbb{E} f_\eta(t - X_{i,i-1}).$$

where

$$X_{i,i-1} = \mathbb{E}\left[X_i|\, \mathcal{F}_{i-1}\right] = \sum_{j=1}^{\infty} c_j \eta_{i-j}$$

3. *If we additionally assume that for some positive integer k, and $f_\eta^{(i)}$ is Lipschitz continuous for $0 \le i \le k$ then $f_X^{(i)}$ exists for $0 \le i \le k$ and is also Lipschitz continuous and satisfies,*

$$f_X^{(i)}(t) = \mathbb{E} f_X^{(i)}(t|\mathcal{F}_{i-1}) = \mathbb{E} f_\eta^{(i)}(t - X_{i,i-1}),$$

for all $0 \le i \le k$.

Lemma C.1 and its proof are based heavily on (Wu and Mielniczuk, 2002, Lemma 1). Now a generalised process can be pursued to obtain bounds for the dependence structure of LRD processes and the applications will be seen in the succeeding sections. This process is defined,

$$I^{[n]}(x) := \sum_{i=1}^{n} \left\{ f_X\left(x|\, \mathcal{F}_{i-1}\right) - f_X\left(x\right) \right\}.$$

and a bound for the above process is given in the next Proposition which crucially relies on techniques used in Wu (2007).

Proposition C.1. *Let X_i be a LRD causal process that satisfies Definition 4.1 with parameter $\alpha \in (0,1)$ and sigma-field $\mathcal{X}_i = \sigma(\eta_i, \eta_{i-1}, \ldots)$. Furthermore, assume that for $i = 0,1,2$; $f_\eta^{(i)}$ is Lipschitz continuous. The process,*

$$I^{[n]}(x) = \sum_{i=1}^{n} \left\{ f_X\left(x | \mathcal{F}_{i-1}\right) - f_X\left(x\right) \right\}.$$

satisfies the following asymptotic bound,

$$\sup_{x\in\mathbb{R}} \left\| I^{[n]}(x) \right\|^2 = \mathcal{O}(n^{2-\alpha}L^2(n))$$

Proposition C.2. *Let $T > 0$ be fixed in $\mathbb{R}$ and $f_\eta^{(i)}$ be Lipschitz continuous for $i = 0,1,2,3$. Then,*

$$\left\| \sup_{|x|\leq T} \left| I^{[n]}(x) \right| \right\| = \mathcal{O}\left(n^{1-\frac{\alpha}{2}} L(n) \right).$$

Remark C.1. A more generalised result of Proposition C.2 can be reached that bounds the j^{th} order derivative of the dependence process,

$$\mathcal{D}^j I^{[n]}(x) = \sum_{i=1}^{n} \left\{ f_X^{(j)}\left(x | \mathcal{F}_{i-1}\right) - f_X^{(j)}(x) \right\}.$$

Indeed, if $f_\eta^{(i)}$ is Lipschitz continuous for $i = j, j+1, j+2, j+3$ and $j \geq 1$ then for some fixed $T > 0$ with $t \in \mathbb{R}$,

$$\left\| \sup_{|x|\leq T} \left| \mathcal{D}^j I^{[n]}(x) \right| \right\| = \mathcal{O}\left(n^{1-\frac{\alpha}{2}} L(n) \right).$$

The proof of the above statement follows by Proposition C.1 and Proposition C.2 where in the statements and proofs the levels f_η are changed to $f_\eta^{(j)}$.

Proof of Lemma C.1 . We begin by proving 1. Use the decomposition,

$$X_i = \sum_{i=1}^{\infty} c_j \eta_{i-j} = \eta_i + \sum_{j=1}^{\infty} c_j \eta_{i-j} =: \eta_i + X_{i,i-1}.$$

Expanding the conditional cumulative distribution function,

$$
\begin{aligned}
F_X\left(t \middle| \mathcal{F}_{i-1}\right) &= P\left(X_i \leq t \middle| \mathcal{F}_{i-1}\right) \\
&= P\left(\sum_{j=0}^{\infty} c_j \eta_{i-j} \leq t \middle| \mathcal{F}_{i-1}\right) \\
&= P\left(\eta_i \leq t - X_{i,i-1}\right) \\
&= F_\eta\left(t - X_{i,i-1}\right).
\end{aligned}
$$

Thus the conditional density exists and is given by,

$$
f_X(t|\mathcal{F}_{i-1}) = f_\eta(t - X_{i,i-1}). \tag{C.1}
$$

Furthermore, the conditional density is Lipschitz continuous and therefore bounded due to the Lipschitz property of f_η,

$$
\begin{aligned}
|f_X(t|\mathcal{F}_{i-1}) - f_X(s|\mathcal{F}_{i-1})| &= \left|f_\eta(t - X_{i,i-1}) - f_\eta(s - X_{i,i-1})\right| \\
&\leq L_\eta \left|t - s\right|.
\end{aligned}
$$

This completes the proof of 1. Let G be the cumulative distribution function of $X_{i,i-1}$, then we show that the density of X exists, firstly, by Fubini's Theorem,

$$
\begin{aligned}
F_X(x) &= P(X_i \leq x) = P(\eta_i + X_{i,i-1} \leq x) \\
&= \mathbb{E}\mathbb{1}_{\{\eta_i + X_{i,i-1} \leq x\}} \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{1}_{\{z+y \leq x\}} dF_\eta(z) dG(y) \\
&= \int_{\mathbb{R}} F_\eta(x - y) dG(y)
\end{aligned} \tag{C.2}
$$

Let $\Delta > 0$ and for any $x, y \in \mathbb{R}$ using the Mean Value Theorem,

$$
\frac{F_\eta\left(x - y + \Delta\right) - F_\eta(x - y)}{\Delta} = f_\eta(x - y + \tau\Delta) \tag{C.3}
$$

where $|\tau| \leq 1$. Using (C.2), (C.2) and (C.3) we obtain,

$$\begin{aligned}
&\left| \frac{F_X(x+\Delta) - F_X(x)}{\Delta} - \mathbb{E} f_\eta(x - X_{i,i-1}) \right| \\
&= \left| \int_{\mathbb{R}} \frac{F_\eta(x-y+\Delta) - F_\eta(x-y)}{\Delta} - \mathbb{E} f_\eta(x - X_{i,i-1}) dG(y) \right| \\
&= \left| \int_{\mathbb{R}} \frac{F_\eta(x-y+\Delta) - F_\eta(x-y)}{\Delta} - f_\eta(x-y) dG(y) \right| \\
&\leq \int_{\mathbb{R}} \left| \frac{F_\eta(x-y+\Delta) - F_\eta(x-y)}{\Delta} - f_\eta(x-y) \right| dG(y) \\
&\leq \int_{\mathbb{R}} \left| f_\eta(x-y+\tau\Delta) - f_\eta(x-y) \right| dG(y) \\
&\leq L_\eta \Delta \int_{\mathbb{R}} dG(y) \\
&= L_\eta \Delta. \qquad \text{(C.4)}
\end{aligned}$$

Thus, taking limits of (C.4) as $\Delta \to 0$ and using the squeeze law implies,

$$\lim_{\Delta \to 0} \left| \frac{F_X(x+\Delta) - F_X(x)}{\Delta} - \mathbb{E} f_\eta(x - X_{i,i-1}) \right| = 0. \qquad \text{(C.5)}$$

Consequently, by the formal definition of a limit in conjunction with (C.5) yields,

$$f_X(x) = \lim_{\Delta \to 0} \frac{F_X(x+\Delta) - F_X(x)}{\Delta} = \mathbb{E} f_\eta(x - X_{i,i-1}).$$

The last thing that remains to be checked is the Lipschitz continuity of f_X,

$$\begin{aligned}
|f_X(t) - f_X(s)| &= \left| \mathbb{E} f_\eta(t - X_{i,i-1}) - \mathbb{E} f_\eta(s - X_{i,i-1}) \right| \\
&\leq \mathbb{E} \left| f_\eta(t - X_{i,i-1}) - f_\eta(s - X_{i,i-1}) \right| \\
&\leq L_\eta |t - s|
\end{aligned}$$

Finally the proof of 3 follows in an identical manner from the one used in 2. First consider the finite analog of the derivative. For some $\Delta > 0$,

$$\begin{aligned}
&\left| \frac{f_X(x+\Delta) - f_X(x)}{\Delta} - \mathbb{E} f_\eta^{(1)}(x - X_{i,i-1}) \right| \\
&= \left| \int_{\mathbb{R}} \frac{f_\eta(x-y+\Delta) - f_\eta(x-y)}{\Delta} - \mathbb{E} f_\eta^{(1)}(x - X_{i,i-1}) dG(y) \right| \\
&= \left| \int_{\mathbb{R}} \frac{F_\eta(x-y+\Delta) - F_\eta(x-y)}{\Delta} - f_\eta(x-y) dG(y) \right| \\
&\leq \int_{\mathbb{R}} \left| \frac{f_\eta(x-y+\Delta) - f_\eta(x-y)}{\Delta} - f_\eta^{(1)}(x-y) \right| dG(y) \\
&\leq \int_{\mathbb{R}} \left| f_\eta^{(1)}(x-y+\tau\Delta) - f_\eta^{(1)}(x-y) \right| dG(y) \\
&\leq L_\eta^{[1]} \Delta \int_{\mathbb{R}} dG(y) \\
&= L_\eta^{[1]} \Delta.
\end{aligned}$$

Then taking the limits as $\Delta \to 0$ and applying the squeeze law implies,

$$f_X^{(1)}(x) = \lim_{\Delta \to 0} \frac{f_X(x+\Delta) - f_X(x)}{\Delta} = \mathbb{E} f_\eta^{(1)}(x - X_{i,i-1}).$$

Again, for the Lipschitz continuity,

$$\begin{aligned}
\left| f_X^{(1)}(t) - f_X^{(1)}(s) \right| &= \left| \mathbb{E} f_\eta^{(1)}(t - X_{i,i-1}) - \mathbb{E} f_\eta^{(1)}(s - X_{i,i-1}) \right| \\
&\leq \mathbb{E} \left| f_\eta^{(1)}(t - X_{i,i-1}) - f_\eta^{(1)}(s - X_{i,i-1}) \right| \\
&\leq L_\eta^{[1]} |t-s|
\end{aligned}$$

□

Proof of Proposition C.1 . As stated earlier, the proof of this Proposition is heavily reliant on the following theorem, taken from (Wu, 2007, Theorem 1) and slightly modified for our purposes,

Theorem 1 (Wu, 2007). *If* $\mathbb{E} g(x, \xi_i) = 0$ *and* $\|g(x, \xi_i)\| < \infty$ *then,*

$$\left\| \sum_{i=1}^{n} g(x, \xi_i) \right\|^2 \leq \sum_{i=-n}^{\infty} (Y_{i+n} - Y_i)^2,$$

where $v_i := \sup_{x \in \mathbb{R}} \|\mathcal{P}_0 g(x, \xi_i)\| = \sup_{x \in \mathbb{R}} \|\mathbb{E}\left[g(x, \xi_i) | \mathcal{F}_0\right] - \mathbb{E}\left[g(x, \xi_i) | \mathcal{F}_{-1}\right]\|$ *and*

$$Y_n = \begin{cases} \sum_{i=1}^{n} v_i & \text{for } n \geq 1, \\ 0 & \text{otherwise.} \end{cases}$$

To be able to use the above theorem on the $I^{[n]}$ process, the conditional density needs to be expressed in that framework whereby the contribution of an individual

η_0 is measured. Without loss of generality, assume $\mathbb{E}X = 0$ and define $g(t, X_{i,i-1}) := f_{X_i}(t|\mathcal{F}_{i-1}) - \mathbb{E} f_{X_i}(t|\mathcal{F}_{i-1})$, then from Lemma C.1,

$$\begin{aligned} g(t, X_{i,i-1}) &= f_{X_i}(t|\mathcal{F}_{i-1}) - \mathbb{E} f_{X_i}(t|\mathcal{F}_{i-1}) \\ &= f_\eta(t - X_{i,i-1}) - f_{X_i}(t). \end{aligned}$$

where $X_{i,i-1} := X_i - \eta_0 = \sum_{j=1}^n c_j \eta_{i-j}$ and f_X is the density of the linear process X. By construction, $\mathbb{E} g(t, X_{i,i-1}) = 0$ for any t and i. From the Proof of Lemma C.1 it was shown that, $f_X(t|\,\mathcal{F}_{i-1}) = f_\eta(t - X_{i,i-1})$. By assumption the Taylor expansion exists,

$$f_\eta(t + x) = f_\eta(t) + x f_\eta^{(1)}(t) + x^2 f_\eta^{(2)}(t + \tau x) \tag{C.6}$$

where $0 \le \tau \le 1$. Consider the projection $\mathcal{P}_0$ of the random variable $g(t, X_{i,i-1})$,

$$\begin{aligned} \mathcal{P}_0 g(t, X_{i,i-1}) &= \mathbb{E}\left[g(t, X_{i,i-1})|\,\mathcal{F}_0\right] - \mathbb{E}\left[g(t, X_{i,i-1})|\,\mathcal{F}_{-1}\right] \\ &= \mathbb{E}\left[f_{X_i}(t|\,\mathcal{F}_{i-1}) - f_{X_i}(t)|\,\mathcal{F}_0\right] - \mathbb{E}\left[f_{X_i}(t|\,\mathcal{F}_{i-1}) - f_{X_i}(t)|\,\mathcal{F}_{-1}\right] \\ &= \mathbb{E}\left[f_{X_i}(t|\,\mathcal{F}_{i-1})|\,\mathcal{F}_0\right] - \mathbb{E}\left[f_{X_i}(t|\,\mathcal{F}_{i-1})|\,\mathcal{F}_{-1}\right] \\ &= \mathbb{E}\left[f_\eta(t - X_{i,i-1})\big|\,\mathcal{F}_0\right] - \mathbb{E}\left[f_\eta(t - X_{i,i-1})\big|\,\mathcal{F}_{-1}\right] \qquad \text{by (C.1)} \\ &= \mathbb{E}\left[f_\eta(t) - X_{i,i-1} f_\eta^{(1)}(t) - X_{i,i-1}^2 f_\eta^{(2)}(\tau_1)\Big|\,\mathcal{F}_0\right] \\ &\qquad - \mathbb{E}\left[f_\eta(t) - X_{i,i-1} f_\eta^{(1)}(t) - X_{i,i-1}^2 f_\eta^{(2)}(\tau_1)\Big|\,\mathcal{F}_{-1}\right] \qquad \text{by (C.6)} \\ &= -f_\eta^{(1)}(t) c_i \eta_0 + \mathbb{E}\left[\left(\sum_{j=1}^{\infty} c_j \eta_{i-j}\right)^2 f_\eta^{(2)}(\tau_1)\,\middle|\,\mathcal{F}_0\right] \\ &\qquad\qquad - \mathbb{E}\left[\left(\sum_{j=1}^{\infty} c_j \eta_{i-j}\right)^2 f_\eta^{(2)}(\tau_1)\,\middle|\,\mathcal{F}_{-1}\right] \\ &= -f_\eta^{(1)}(t) c_i \eta_0 + D_i(t). \end{aligned} \tag{C.7}$$

Applying the $\|\cdot\|$ norm to (C.7),

$$\begin{aligned} \|\mathcal{P}_0 g(t, X_{i,i-1})\| &= \left\| -f_\eta^{(1)}(t) c_i \eta_0 + D_i(t) \right\| \\ &\le c_i \left| f_\eta^{(1)} \right|_\infty \|\eta_0\| + \|D_i(t)\|. \end{aligned} \tag{C.8}$$

Focus attention on the $D_i(t)$ term,

$$
\begin{aligned}
|D_i(t)| \leq \left|f_\eta^{(2)}\right|_\infty \Bigg| \mathbb{E}\left[\left(\sum_{j=1}^{i-1} c_j\eta_{i-j} + \sum_{j=i}^{\infty} c_j\eta_{i-j}\right)^2 \middle| \mathcal{F}_0\right] \\
- \mathbb{E}\left[\left(\sum_{j=1}^{i-1} c_j\eta_{i-j} + \sum_{j=i}^{\infty} c_j\eta_{i-j}\right)^2 \middle| \mathcal{F}_{-1}\right]\Bigg| \\
= \left|f_\eta^{(2)}\right|_\infty \Bigg| \mathbb{E}\left(\sum_{j=1}^{i-1} c_j\eta_{i-j}\right)^2 + \left(\sum_{j=i}^{\infty} c_j\eta_{i-j}\right)^2 \\
- \mathbb{E}\left(\sum_{j=1}^{i} c_j\eta_{i-j}\right)^2 - \left(\sum_{j=i+1}^{\infty} c_j\eta_{i-j}\right)^2 \Bigg| \\
= \left|f_\eta^{(2)}\right|_\infty \left| c_i^2\left(\eta_0^2 - \mathbb{E}\eta_0^2\right) + 2c_i\eta_0 \sum_{j=i+1}^{\infty} c_j\eta_{i-j}\right| \\
\leq \left|f_\eta^{(2)}\right|_\infty \left(c_i^2\left|\eta_0^2 - \mathbb{E}\eta_0^2\right| + 2c_i\left|\eta_0 \sum_{j=i+1}^{\infty} c_j\eta_{i-j}\right|\right).
\end{aligned} \tag{C.9}
$$

We just need to deal with the last term of (C.9) and we can bound the $D_i(t)$ process,

$$
\begin{aligned}
\left\|\eta_0 \sum_{j=i+1}^{\infty} c_j\eta_{i-j}\right\|^2 &= \|\eta_0\|^2 \left\|\sum_{j=i+1}^{\infty} c_j\eta_{i-j}\right\|^2 \\
&= \|\eta_0\|^2 \sum_{j=i+1}^{\infty} c_j^2 \|\eta_{i-j}\|^2 \\
&\leq \|\eta_0\|^4 \sum_{j=0}^{\infty} c_j^2 < \infty.
\end{aligned} \tag{C.10}
$$

Combined (C.9) and (C.10) yields,

$$
\|D_i(t)\| \leq \left|f_\eta^{(2)}\right|_\infty \left(c_i^2\left\|\eta_0^2 - \mathbb{E}\eta_0^2\right\| + 2c_i\|\eta_0\|^4 \sum_{j=0}^{\infty} c_j^2\right). \tag{C.11}
$$

So using (C.11) and (C.8),

$$
\begin{aligned}
v_i &= \sup_{t\in(0,1)} \|\mathcal{P}_0 g(t, X_{i,i-1})\| \\
&\leq \sup_{t\in(0,1)} \left(c_i \left|f_\eta^{(1)}\right|_\infty \|\eta_0\| + \left|f_\eta^{(2)}\right|_\infty \left(c_i^2\left\|\eta_0^2 - \mathbb{E}\eta_0^2\right\| + 2c_i\|\eta_0\|^4 \sum_{j=0}^{\infty} c_j^2\right)\right) \\
&= A_1 c_i + A_2 c_i^2
\end{aligned} \tag{C.12}
$$

where $A_1 = \left|f_\eta^{(1)}\right|_\infty \|\eta_0\| + 2\left|f_\eta^{(2)}\right|_\infty \|\eta_0\|^4 \sum_{j=0}^\infty c_j^2$ and $A_2 = \left|f_\eta^{(2)}\right|_\infty \left\|\eta_0^2 - \mathbb{E}\eta_0^2\right\|$. Define, $\beta := (1-\alpha)/2$, then recall that, $c_i = i^{-\beta}L(i)$ and define $\ell(\cdot) = L(\cdot) \vee L^2(\cdot)$ and $A_* = 2(A_1 \vee A_2)$ then by (C.12),

$$\begin{aligned}
v_i &\le \frac{A_*}{2}\ell(i)\left(i^{-\beta} + i^{-2\beta}\right) \\
&\le C\ell(i)i^{-\beta} \\
&= C\frac{L(i) \vee L^2(i)}{L(i)} i^{-\beta}L(i) \\
&= C\left(1 \vee L(i)\right) i^{-\beta}L(i) \\
&\sim Ci^{-\beta}L^2(i) \\
&= c_i^*.
\end{aligned}$$

where c_i^* is a regularly varying sequence. Thus we can see that the contribution of each η to the overall dependence structure of the LRD process can be bounded by a regularly varying sequence with parameter β. Therefore, apply Theorem 1 (Wu, 2007) (see page 173),

$$\begin{aligned}
\left\|I^{[n]}(x)\right\|^2 &\le \sum_{i=-n}^{\infty} (Y_{i+n} - Y_i)^2 \\
&= \sum_{i=1}^{n} Y_i^2 + \sum_{i=1}^{n} (Y_{i+n} - Y_i)^2 + \sum_{i=n+1}^{\infty} (Y_{i+n} - Y_i)^2 \\
&\le nY_n^2 + \sum_{i=1}^{n} Y_{i+n}^2 + \sum_{i=n+1}^{\infty} (Y_{i+n} - Y_i)^2 \\
&\le nY_n^2 + nY_{2n}^2 + \sum_{i=n+1}^{\infty} (Y_{i+n} - Y_i)^2 \\
&\le 2nY_{2n}^2 + \sum_{i=n+1}^{\infty} (Y_{i+n} - Y_i)^2
\end{aligned} \tag{C.13}$$

We now deal with each term of the RHS of (C.13) separately. We begin with the first term and apply Theorem A.4,

$$\begin{aligned}
2nY_{2n}^2 &= 2n\left(\sum_{i=1}^{2n} v_i\right)^2 \\
&\le Cn\left(\sum_{i=1}^{2n} i^{-\beta}L(i)\right)^2 \\
&\sim Cn^{3-2\beta}L^2(n).
\end{aligned} \tag{C.14}$$

To bound the second term, apply Theorem A.3,

$$
\begin{aligned}
Y_{i+n} - Y_i &= \sum_{j=i+1}^{i+n} v_j \\
&\leq n \left\{ \sup_{j \in \{i+1,\dots,i+n\}} v_j \right\} \\
&\leq n \left\{ \sup_{j \geq i} v_j \right\} \\
&\leq Cn \left\{ \sup_{j \geq i} j^{-\beta} L(j) \right\} \\
&\sim Cni^{-\beta} L(i). \qquad \text{(C.15)}
\end{aligned}
$$

The bound in (C.15) can then be used to bound the sum,

$$
\begin{aligned}
\sum_{i=n+1}^{\infty} (Y_{i+n} - Y_i)^2 &\leq Cn^2 \sum_{i=n+1}^{\infty} v_i^2 \\
&\leq Cn^2 \int_n^{\infty} x^{-2\beta} L^2(x)\, dx \\
&\leq Cn^{3-2\beta} L^2(n) \quad \text{by Lemma A.1}. \qquad \text{(C.16)}
\end{aligned}
$$

Substituting (C.16) and (C.14) into (C.13) yields,

$$
\sup_{x \in \mathbb{R}} \left\| I^{[n]}(x) \right\|^2 \leq Cn^{3-2\beta} L^2(n) = Cn^{2-\alpha} L^2(n). \qquad \square
$$

Proof of Proposition C.2 . Notice that,

$$
\left| I^{[n]}(x) - I^{[n]}(-T) \right| = \left| \int_{-T}^{x} \mathcal{D}I^{[n]}(t)\, dt \right| \leq \int_{-T}^{T} \left| \mathcal{D}I^{[n]}(t) \right| \, dt
$$

Therefore, $\left| I^{[n]}(x) \right| \leq \left| I^{[n]}(-T) \right| + \int_{-T}^{T} \left| \mathcal{D}I^{[n]}(t) \right| \, dt$ which gives,

$$
\begin{aligned}
\left\| \sup_{|x| \leq T} \left| I^{[n]}(x) \right| \right\| &\leq \left\| I^{[n]}(-T) \right\| + \left\| \int_{-T}^{T} \left| \mathcal{D}I^{[n]}(t) \right| \, dt \right\| \\
&\leq \sup_{x \in \mathbb{R}} \left\| I^{[n]}(x) \right\| + 2T \sup_{x \in \mathbb{R}} \left\| \mathcal{D}I^{[n]}(x) \right\|
\end{aligned}
$$

However, it is assumed that $f_\eta^{(i)}$ are Lipschitz continuous for $i = 0, 1, 2, 3$; so two applications of Proposition C.1 yield

$$
\sup_{x \in \mathbb{R}} \left\| I^{[n]}(x) \right\| = \mathcal{O}\left(n^{1-\frac{\alpha}{2}} L(n) \right) \quad \text{and} \quad \sup_{x \in \mathbb{R}} \left\| \mathcal{D}I^{[n]}(x) \right\| = \mathcal{O}\left(n^{1-\frac{\alpha}{2}} L(n) \right). \qquad \square
$$

BIBLIOGRAPHY

Andersen, T. and Bollerslev, T. (1997). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance* **4** 115–158.
URL http://dx.doi.org/10.1016/S0927-5398(97)00004-2 (Cited on page 11.)

Antoniadis, A. and Gijbels, I. (2002). Detecting abrupt changes by wavelet methods. *J. Nonparametr. Stat.* **14** 7–29. Statistical models and methods for discontinuous phenomena (Oslo, 1998).
URL http://dx.doi.org/10.1080/10485250211396 (Cited on page 9.)

Antoniadis, A., Grégoire, G. and Vial, P. (1997). Random design wavelet curve smoothing. *Statist. Probab. Lett.* **35** 225–232.
URL http://dx.doi.org/10.1016/S0167-7152(97)00017-5 (Cited on page 87.)

Avram, F. and Taqqu, M. S. (1987). Noncentral limit theorems and Appell polynomials. *Ann. Probab.* **15** 767–775.
URL http://dx.doi.org/10.1214/aop/1176992170 (Cited on pages 117 and 118.)

Baillie, R. T., Bollerslev, T. and Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *J. Econometrics* **74** 3–30.
URL http://dx.doi.org/10.1016/S0304-4076(95)01749-6 (Cited on page 11.)

Baraud, Y. (2002). Model selection for regression on a random design. *ESAIM Probab. Statist.* **6** 127–146 (electronic).
URL http://dx.doi.org/10.1051/ps:2002007 (Cited on pages 4 and 87.)

Beran, J. (1992). Statistical Methods for Data with Long-Range Dependence. *Statistical Science* **7** 404–416.
URL http://dx.doi.org/10.1214/ss/1177011127 (Cited on page 11.)

Beran, J. (1994). *Statistics for long-memory processes*, vol. 61 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, New York. (Cited on page 11.)

Bertin, K. (2004). Minimax exact constant in sup-norm for nonparametric regression with random design. *J. Statist. Plann. Inference* **123** 225–242.
URL http://dx.doi.org/10.1016/S0378-3758(03)00154-X (Cited on pages 4 and 87.)

Biagini, F., Hu, Y., Øksendal, B. and Zhang, T. (2008). *Stochastic Calculus for Fractional Brownian Motion and Applications*. Probability and its Applications (New York), Springer-Verlag London Ltd., London. (Cited on pages 74 and 76.)

Bingham, N. H. (1986). Variants on the law of the iterated logarithm. *Bull. London Math. Soc.* **18** 433–467.
URL http://dx.doi.org/10.1112/blms/18.5.433 (Cited on page 99.)

Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1989). *Regular variation*, vol. 27 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge. (Cited on page 147.)

BIRGÉ, L. (2004). Model selection for Gaussian regression with random design. *Bernoulli* **10** 1039–1051.
URL http://dx.doi.org/10.3150/bj/1106314849 (Cited on pages 4 and 87.)

BOJANIC, R. and SENETA, E. (1973). A unified theory of regularly varying sequences. *Math. Z.* **134** 91–106.
URL http://dx.doi.org/10.1007/BF01214468 (Cited on pages 147, 148, and 151.)

BOX, G. E. P. and JENKINS, G. M. (1970). *Times series analysis. Forecasting and control*. Holden-Day, San Francisco, Calif. (Cited on page 11.)

BROWN, B. (1971). Martingale Central Limit Theorems. *The Annals of Mathematical Statistics* **42** 59–66.
URL http://dx.doi.org/10.1214/aoms/1177693494 (Cited on page 96.)

BROWN, L. D., CAI, T. T., LOW, M. G. and ZHANG, C.-H. (2002). Asymptotic equivalence theory for nonparametric regression with random design. *Ann. Statist.* **30** 688–707. Dedicated to the memory of Lucien Le Cam.
URL http://dx.doi.org/10.1214/aos/1028674838 (Cited on page 89.)

BROWN, L. D. and LOW, M. G. (1996). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.* **24** 2384–2398.
URL http://dx.doi.org/10.1214/aos/1032181159 (Cited on page 25.)

CARTER, A. V. (2010). Asymptotically sufficient statistics in nonparametric regression experiments with correlated noise. *Journal of Probability and Statistics* **2009**.
URL http://dx.doi.org/10.1155/2009/275308 (Cited on page 25.)

CAVALIER, L. (2004). Estimation in a problem of fractional integration. *Inverse Problems* **20** 1445–1454.
URL http://dx.doi.org/10.1088/0266-5611/20/5/007 (Cited on page 9.)

CHEN, C., CHAN, J., GERLACH, R. and HSIEH, W. (2010). A comparison of estimators for regression models with change points. *Statistics and Computing* 1–2010.1007/s11222-010-9177-0.
URL http://dx.doi.org/10.1007/s11222-010-9177-0 (Cited on page 6.)

CHENG, M.-Y. and RAIMONDO, M. (2008). Kernel methods for optimal change-points estimation in derivatives. *J. Comput. Graph. Statist.* **17** 56–75.
URL http://dx.doi.org/10.1198/106186008X289164 (Cited on pages iii, 4, 5, 6, 18, 19, 29, 35, 36, 38, 39, and 57.)

CHESNEAU, C. (2007). Regression with random design: a minimax study. *Statist. Probab. Lett.* **77** 40–53.
URL http://dx.doi.org/10.1016/j.spl.2006.05.010 (Cited on pages 4 and 87.)

CHU, C.-K. and MARRON, J. S. (1991). Choosing a kernel regression estimator. *Statist. Sci.* **6** 404–436. With comments and a rejoinder by the authors.
URL http://dx.doi.org/10.1214/ss/1177011586 (Cited on page 7.)

CSÖRGŐ, M. (1983). *Quantile processes with statistical applications*, vol. 42 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA. (Cited on page 109.)

DEVROYE, L. and GYÖRFI, L. (1990). No empirical probability measure can converge in the total variation sense for all distributions. *Ann. Statist.* **18** 1496–1499. URL http://dx.doi.org/10.1214/aos/1176347765 (Cited on page 14.)

DONOHO, D. L. (1995). Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. *Appl. Comput. Harmon. Anal.* **2** 101–126. URL http://dx.doi.org/10.1006/acha.1995.1008 (Cited on page 17.)

DOUKHAN, P., OPPENHEIM, G. and TAQQU, M. S. (eds.) (2003). *Theory and applications of long-range dependence*. Birkhäuser Boston Inc., Boston, MA. (Cited on page 11.)

EFROMOVICH, S. (1999). *Nonparametric curve estimation*. Springer Series in Statistics, Springer-Verlag, New York. Methods, theory, and applications. (Cited on pages 3 and 9.)

ENGLE, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50** 987–1007. URL http://dx.doi.org/10.2307/1912773 (Cited on page 11.)

EUBANK, R. L. (1999). *Nonparametric regression and spline smoothing*, vol. 157 of *Statistics: Textbooks and Monographs*. 2nd ed. Marcel Dekker Inc., New York. (Cited on page 9.)

FAN, J. (1992). Design-adaptive nonparametric regression. *J. Amer. Statist. Assoc.* **87** 998–1004. URL http://dx.doi.org/10.2307/2290637 (Cited on page 87.)

FAN, J. and GIJBELS, I. (1996). *Local polynomial modelling and its applications*, vol. 66 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London. (Cited on page 7.)

FAN, J. and YAO, Q. (2003). *Nonlinear time series*. Springer Series in Statistics, Springer-Verlag, New York. Nonparametric and parametric methods. URL http://dx.doi.org/10.1007/b97702 (Cited on page 3.)

FELLER, W. (1968). *An introduction to probability theory and its applications. Vol. I*. Third edition, John Wiley & Sons Inc., New York. (Cited on page 48.)

FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications. Vol. II*. Second edition, John Wiley & Sons Inc., New York. (Cited on page 147.)

FRALEY, C., LEISCH, F., MAECHLER, M., REISEN, V. and LEMONTE, A. (2009). *fracdiff: Fractionally differenced ARIMA aka ARFIMA(p,d,q) models*. R package version 1.3-2. URL http://CRAN.R-project.org/package=fracdiff (Cited on pages 42, 51, 92, and 114.)

FREEDMAN, D. A. (1975). On tail probabilities for martingales. *Ann. Probability* **3** 100–118. URL http://dx.doi.org/10.1214/aop/1176996452 (Cited on pages 102 and 122.)

GALAMBOS, J. and SENETA, E. (1973). Regularly varying sequences. *Proc. Amer. Math. Soc.* **41** 110–116. URL http://dx.doi.org/10.1090/S0002-9939-1973-0323963-5 (Cited on page 147.)

Gasser, T. and Müller, H.-G. (1979). Kernel estimation of regression functions. In *Smoothing techniques for curve estimation (Proc. Workshop, Heidelberg, 1979)*, vol. 757 of *Lecture Notes in Math.* Springer, Berlin, 23–68.
URL http://dx.doi.org/10.1007/BFb0098489 (Cited on page 7.)

Gijbels, I. and Goderniaux, A.-C. (2004). Data-driven discontinuity detection in derivatives of a regression function. *Comm. Statist. Theory Methods* **33** 851–871.
URL http://dx.doi.org/10.1081/STA-120028730 (Cited on page 4.)

Gijbels, I., Hall, P. and Kneip, A. (1999). On the estimation of jump points in smooth curves. *Ann. Inst. Statist. Math.* **51** 231–251.
URL http://dx.doi.org/10.1023/A:1003802007064 (Cited on pages 4, 6, and 87.)

Goldenshluger, A., Juditsky, A., Tsybakov, A. and Zeevi, A. (2008a). Change-point estimation from indirect observations. II. Adaptation. *Ann. Inst. Henri Poincaré Probab. Stat.* **44** 819–836.
URL http://dx.doi.org/10.1214/07-AIHP144 (Cited on page 4.)

Goldenshluger, A., Juditsky, A., Tsybakov, A. B. and Zeevi, A. (2008b). Change-point estimation from indirect observations. I. Minimax complexity. *Ann. Inst. Henri Poincaré Probab. Stat.* **44** 787–818.
URL http://dx.doi.org/10.1214/07-AIHP110 (Cited on page 4.)

Goldenshluger, A., Tsybakov, A. and Zeevi, A. (2006). Optimal change-point estimation from indirect observations. *Ann. Statist.* **34** 350–372.
URL http://dx.doi.org/10.1214/009053605000000750 (Cited on pages iii, 4, 6, 18, 19, 22, 29, 36, 38, 44, 58, 59, and 77.)

Granger, C. W. J. and Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *J. Time Ser. Anal.* **1** 15–29.
URL http://dx.doi.org/10.1111/j.1467-9892.1980.tb00297.x (Cited on page 11.)

Gripenberg, G. and Norros, I. (1996). On the prediction of fractional Brownian motion. *J. Appl. Probab.* **33** 400–410.
URL http://dx.doi.org/10.2307/3215063 (Cited on page 43.)

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.
URL http://dx.doi.org/10.1214/009053607000000686 (Cited on pages 17 and 112.)

Hall, P. and Hart, J. D. (1990). Nonparametric regression with long-range dependence. *Stochastic Process. Appl.* **36** 339–351.
URL http://dx.doi.org/10.1016/0304-4149(90)90100-7 (Cited on page 16.)

Härdle, W. (1990). *Applied nonparametric regression*, vol. 19 of *Econometric Society Monographs*. Cambridge University Press, Cambridge. (Cited on page 5.)

Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (1998). *Wavelets, approximation, and statistical applications*, vol. 129 of *Lecture Notes in Statistics*. Springer-Verlag, New York. (Cited on pages 3 and 9.)

Hawkins, D. M. (2001). Fitting multiple change-point models to data. *Comput. Statist. Data Anal.* **37** 323–341.
URL http://dx.doi.org/10.1016/S0167-9473(00)00068-2 (Cited on page 6.)

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.
URL http://dx.doi.org/10.1016/j.dam.2005.02.009 (Cited on page 96.)

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.
URL http://dx.doi.org/10.1214/aos/1032526953 (Cited on page 145.)

Ho, H.-C. and Hsing, T. (1997). Limit theorems for functionals of moving averages. *Ann. Probab.* **25** 1636–1669.
URL http://dx.doi.org/10.1214/aop/1023481106 (Cited on page 98.)

Hosking, J. R. M. (1981). Fractional differencing. *Biometrika* **68** 165–176.
URL http://dx.doi.org/10.1093/biomet/68.1.165 (Cited on pages 11 and 92.)

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.
URL http://dx.doi.org/10.1111/j.1467-842X.2004.00340.x (Cited on pages 19 and 87.)

Hurst, H. (1951). Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers* **116** 35. (Cited on pages 11 and 23.)

Kerkyacharian, G. and Picard, D. (2004). Regression in random design and warped wavelets. *Bernoulli* **10** 1053–1105.
URL http://dx.doi.org/10.3150/bj/1106314850 (Cited on pages 4 and 87.)

Khas′minskiĭ, R. Z. (ed.) (1992). *Topics in nonparametric estimation*, vol. 12 of *Advances in Soviet Mathematics*. American Mathematical Society, Providence, RI. (Cited on page 3.)

Kohler, M. (2008). Multivariate orthogonal series estimates for random design regression. *J. Statist. Plann. Inference* **138** 3217–3237.
URL http://dx.doi.org/10.1016/j.jspi.2008.01.011 (Cited on pages 4 and 87.)

Korostelëv, A. P. (1987). On minimax estimation of a discontinuous signal. *Theory Probab. Appl.* **32** 727–730.
URL http://dx.doi.org/10.1137/1132110 (Cited on pages 4, 18, and 87.)

Korostelëv, A. P. and Tsybakov, A. B. (1993). *Minimax theory of image reconstruction*, vol. 82 of *Lecture Notes in Statistics*. Springer-Verlag, New York. (Cited on pages 4, 6, and 87.)

Kulik, R. (2008). Nonparametric deconvolution problem for dependent sequences. *Electron. J. Stat.* **2** 722–740.
URL http://dx.doi.org/10.1214/07-EJS154 (Cited on page 95.)

Kulik, R. and Lorek, P. (2011). Some results on random design regression with long memory errors and predictors. *Journal of Statistical Planning and Inference* **141** 508–523.

URL http://dx.doi.org/10.1016/j.jspi.2010.06.030 (Cited on pages 17 and 112.)

Kulik, R. and Raimondo, M. (2009). Wavelet regression in random design with heteroscedastic dependent errors. *Ann. Statist.* **37** 3396–3430.
URL http://dx.doi.org/10.1214/09-AOS684 (Cited on pages 4 and 17.)

Kunst, A., Looman, C. and Mackenbach, J. (1993). Outdoor Air Temperature and Mortality in the Netherlands: A Time-Series Analysis. *American Journal of Epidemiology* **137** 331–341. (Cited on page 6.)

Lerman, P. (1980). Fitting segmented regression models by grid search. *Applied Statistics* **29** 77–84.
URL http://dx.doi.org/10.2307/2346413 (Cited on page 6.)

Liu, W. and Wu, W. B. (2010). Simultaneous nonparametric inference of time series. *Ann. Statist.* **38** 2388–2421.
URL http://dx.doi.org/10.1214/09-AOS789 (Cited on pages 7, 17, 112, and 120.)

Liu, Z. and Qian, L. (2010). Changepoint Estimation in a Segmented Linear Regression via Empirical Likelihood. *Commun. Stat.-Simul. Comput.* **39** 85–100.
URL http://dx.doi.org/10.1080/03610910903312193 (Cited on page 6.)

Luan, Y. and Xie, Z. (2001). The wavelet identification for jump points of derivative in regression model. *Statistics & Probability Letters* **53** 167 – 180.
URL http://dx.doi.org/10.1016/S0167-7152(01)00070-0 (Cited on page 6.)

Mack, Y. P. and Müller, H.-G. (1989). Derivative estimation in nonparametric regression with random predictor variable. *Sankhy?: The Indian Journal of Statistics, Series A* **51** pp. 59–72.
URL http://www.jstor.org/stable/25050723 (Cited on page 36.)

Mallat, S. (1999). *A wavelet tour of signal processing*. Academic Press Inc., San Diego, CA. (Cited on page 9.)

Mandelbrot, B. B. and Van Ness, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Rev.* **10** 422–437.
URL http://dx.doi.org/10.1137/1010093 (Cited on pages 11 and 24.)

Menéndez, P., Ghosh, S. and Beran, J. (2010). On rapid change points under long memory. *Journal of Statistical Planning and Inference* **140** 3343 – 3354.
URL http://dx.doi.org/10.1016/j.jspi.2010.04.051 (Cited on page 6.)

Müller, H.-G. (1992). Change-points in nonparametric regression analysis. *Ann. Statist.* **20** 737–761.
URL http://dx.doi.org/10.1214/aos/1176348654 (Cited on pages 4 and 6.)

Müller, H.-G. (1997). Density adjusted kernel smoothers for random design nonparametric regression. *Statist. Probab. Lett.* **36** 161–172.
URL http://dx.doi.org/10.1016/S0167-7152(97)00059-X (Cited on page 87.)

Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability and its Applications* **9** 141–142.
URL http://link.aip.org/link/?TPR/9/141/1 (Cited on page 7.)

Neumann, M. H. (1997). Optimal change-point estimation in inverse problems. *Scand. J. Statist.* **24** 503–521.
URL http://dx.doi.org/10.1111/1467-9469.00077 (Cited on page 18.)

Nussbaum, M. (1985). Spline smoothing in regression models and asymptotic efficiency in L_2. *Ann. Statist.* **13** 984–997.
URL http://dx.doi.org/10.1214/aos/1176349651 (Cited on page 16.)

Park, C. and Kim, W.-C. (2006). Wavelet estimation of a regression function with a sharp change point in a random design. *J. Statist. Plann. Inference* **136** 2381–2394.
URL http://dx.doi.org/10.1016/j.jspi.2005.08.017 (Cited on pages 19 and 87.)

Priestley, M. B. and Chao, M. T. (1972). Non-parametric function fitting. *J. Roy. Statist. Soc. Ser. B* **34** 385–392. (Cited on page 7.)

R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
URL http://www.R-project.org (Cited on pages 42, 51, 91, and 114.)

Raimondo, M. (1998). Minimax estimation of sharp change points. *Ann. Statist.* **26** 1379–1397.
URL http://dx.doi.org/10.1214/aos/1024691247 (Cited on pages 4, 6, 9, and 18.)

Reiss, M. (2008). Asymptotic equivalence for nonparametric regression with multivariate and random design. *Ann. Statist.* **36** 1957–1982.
URL http://dx.doi.org/10.1214/07-AOS525 (Cited on page 89.)

Seneta, E. (1976). *Regularly varying functions*. Lecture Notes in Mathematics, Vol. 508, Springer-Verlag, Berlin. (Cited on pages 147, 148, and 152.)

Seneta, E. (2009). Personal Communication. (Cited on page 152.)

Speckman, P. (1985). Spline smoothing and optimal rates of convergence in nonparametric regression models. *Ann. Statist.* **13** 970–983.
URL http://dx.doi.org/10.1214/aos/1176349650 (Cited on pages 3 and 16.)

Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10** 1040–1053.
URL http://links.jstor.org/sici?sici=0090-5364(198212)10:4<1040:OGROCF>2.0.CO;2-2&origin=MSN (Cited on page 16.)

Taqqu, M. S. (1975). Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **31** 287–302.
URL http://dx.doi.org/10.1007/BF00532868 (Cited on pages 25, 117, and 140.)

Tsybakov, A. B. (2009). *Introduction to Nonparametric Estimation*. Springer Publishing Company, Incorporated. (Cited on pages 4, 12, and 68.)

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes*. Springer Series in Statistics, Springer-Verlag, New York. With applications to statistics. (Cited on page 44.)

Wand, M. P. and Jones, M. C. (1995). *Kernel smoothing*, vol. 60 of *Monographs on Statistics and Applied Probability*. Chapman and Hall Ltd., London. (Cited on pages 3 and 7.)

WANG, L. (2008). Change-Point Estimation in Long Memory Nonparametric Models with Applications. *Commun. Stat.-Simul. Comput.* **37** 48–61.
URL http://dx.doi.org/10.1080/03610910701723583 (Cited on pages 9 and 112.)

WANG, L. and CAI, H. (2010). Wavelet change-point estimation for long memory non-parametric random design models. *J. Time Series Anal.* **31** 86–97.
URL http://dx.doi.org/10.1111/j.1467-9892.2009.00646.x (Cited on page 112.)

WANG, Y. (1995). Jump and sharp cusp detection by wavelets. *Biometrika* **82** 385–397.
URL http://dx.doi.org/10.1093/biomet/82.2.385 (Cited on pages 4, 6, and 9.)

WANG, Y. (1996). Function estimation via wavelet shrinkage for long-memory data. *Ann. Statist.* **24** 466–484.
URL http://dx.doi.org/10.1214/aos/1032894449 (Cited on pages 16, 25, and 26.)

WANG, Y. (1997). Minimax estimation via wavelets for indirect long-memory data. *J. Statist. Plann. Inference* **64** 45–55.
URL http://dx.doi.org/10.1016/S0378-3758(96)00205-4 (Cited on page 26.)

WANG, Y. (1999). Change-points via wavelets for indirect data. *Statist. Sinica* **9** 103–117. (Cited on pages 6 and 9.)

WASSERMAN, L. (2006). *All of nonparametric statistics*. Springer Texts in Statistics, Springer, New York. (Cited on pages 4 and 9.)

WATSON, G. S. (1964). Smooth regression analysis. *Sankhyā Ser. A* **26** 359–372. (Cited on page 7.)

WISHART, J. (2009). Kink estimation with correlated noise. *Journal of the Korean Statistical Society* **38** 131–143.
URL http://dx.doi.org/10.1016/j.jkss.2008.08.001 (Cited on pages 6, 22, and 29.)

WISHART, J. and KULIK, R. (2010). Kink estimation in stochastic regression with dependent errors and predictors. *Electron. J. Statist.* **4** 875–913.
URL http://dx.doi.org/10.1214/10-EJS571 (Cited on pages 6, 29, 86, and 111.)

WU, W. B. (2005). Nonlinear system theory: another look at dependence. *Proc. Natl. Acad. Sci. USA* **102** 14150–14154 (electronic).
URL http://dx.doi.org/10.1073/pnas.0506715102 (Cited on page 138.)

WU, W. B. (2007). Strong invariance principles for dependent random variables. *Ann. Probab.* **35** 2294–2320.
URL http://dx.doi.org/10.1214/009117907000000060 (Cited on pages 138, 169, 170, and 173.)

WU, W. B. and MIELNICZUK, J. (2002). Kernel density estimation for linear processes. *Ann. Statist.* **30** 1441–1459.
URL http://dx.doi.org/10.1214/aos/1035844982 (Cited on pages 7, 95, and 169.)

YANG, Y. (2001). Nonparametric regression with dependent errors. *Bernoulli* **7** 633–655.
URL http://dx.doi.org/10.2307/3318730 (Cited on page 16.)

ZHANG, S., WONG, M.-Y. and ZHENG, Z. (2002). Wavelet threshold estimation of a regression function with random design. *J. Multivariate Anal.* **80** 256–284.
URL http://dx.doi.org/10.1006/jmva.2000.1980 (Cited on pages 4 and 87.)

ZHAO, Z. and WU, W. B. (2008). Confidence bands in nonparametric time series regression. *Ann. Statist.* **36** 1854–1878.
URL http://dx.doi.org/10.1214/07-AOS533 (Cited on pages 7, 17, 112, and 139.)