Non-Standard Statistical Inference Under Short and Long Range Dependence

by

Pramita Bagchi

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Statistics)
in The University of Michigan
2015

Doctoral Committee:

Professor Moulinath Banerjee, Co-Chair
Associate Professor Stilian A. Stoev, Co-Chair
Professor Xuming He
Professor Bin Nan

To Ma, Thampa and Didun

ACKNOWLEDGEMENTS

First of all I would like to thank my advisors Prof. Moulinath Banerjee and Prof. Stilian Stoev, for being incredible mentors throughout these last five years. Their constant support and guidance have enriched me as a student and motivated me to be a better researcher.

I am extremely thankful to my department for providing a friendly and dynamic environment to work and to my officemates Naveen, Juan, Emmy, Jing, Can, Sandipan for making my stay here enjoyable. Naveen has been a dear friend from my ISI days and I am grateful to have him by my side for last ten years. I would also like to thank Prof. Vijay Nair and Prof. Xuming He for guiding me during my stressful job search process.

It was only when I was lonely and away from my family I realized how lucky I am to have so many friends. Thank you Bharamardi for making all of us a family. I thank Rito and Santy for helping to make this faraway land home in the early years. I am grateful to Shrijita, Sreyoshidi, Sumantada, Pamda, Bhramardi for sharing so many wonderful times, making my life at Ann Arbor memorable and my ISI and school friends Somak, Bana, Gourab, Jhaba, Ankita for patiently tolerating me for so many years. I am extremely lucky to have my best friend Sebanti as a part of my life for last ten years. From the first day at ISI she has been a big part of this journey. Thank you Sebanti for listening to all my frustrations, being patient with me at my worst times and sharing all the wonderful memories.

Finally the biggest acknowledgment goes out to my parents for always supporting and believing in me. Nothing would have been possible without their constant love, care and sacrifice. Thank you Ma & Baba.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF APPENDICES

Appendix

LIST OF ABBREVIATIONS

CLT Central Limit Theorem

CI Confidence Interval

fBm fractional Brownian Motion

GCM Greatest Convex Minorant

IRE Isotonic Regression Estimator

LRD Long Range Dependence

SRD Short Range Dependence

ABSTRACT

Non Standard Problems Under Short and Long Range Dependence

by

Pramita Bagchi

Chair: Moulinath Banerjee and Stilian Stoev

The work discusses different non-standard problems under different types of short and long range dependence.

In the first part we introduce new point-wise confidence interval estimates for monotone functions observed with additive and dependent noise. Existence of such monotone trend is quite common in time series data. We study both short- and long-range dependence regimes for the errors. The interval estimates are obtained via the method of inversion of certain discrepancy statistics. This approach avoids the estimation of nuisance parameters such as the derivative of the unknown function, which other methods are forced to deal with. The resulting estimates are therefore more accurate, stable, and widely applicable in practice under mild assumptions on the trend and error structure. While motivated by earlier work in the independent context, the dependence of the errors, especially long-range dependence leads to new phenomena and new universal limits based on convex minorant functionals of drifted fractional Brownian motion.

In the second part we investigate the problem of M-estimation, the technique of extracting a parameter estimate by minimizing a loss function is used in almost every

statistical problems. We focus on the general theory of such estimators in the presence of dependence in data, a very common feature in time series or econometric applications. Unlike the case of independent and identically distributed observations, there is a lack of an overarching asymptotic theory for M-estimation under dependence. In order to develop a general theory, we have proved a new triangular version of functional central limit theorem for dependent observations, which is useful for broader applications beyond our current paper. We use this general CLT along with standard empirical process techniques to provide the rate and asymptotic distribution of minimizer of a general empirical process. We have used our theory to make inferences for many important problems like change point problems, excess-mass-baseline-inverse problem, different regression settings including maximum score estimator, least absolute deviation regression and censored regression among others.

CHAPTER I

Introduction

Dependence is a natural phenomenon appearing in time series data. We have studied some important non-standard statistical problems under general dependence structure.

In the first part we have considered the estimation of a trend function observed with additive noise. This is a canonical problem of substantial interest and have been widely studied in the statistical literature (see e.g. Clifford et. al.(2005)[37], Fan and Yao (2003)[7], Robinson (2009)[58], Wu and Zhao (2007) [69]). Most existing methods are based upon smoothness conditions on the trend (e.g. higher order differentiability, or curvature). They do not incorporate shape constraints like monotonicity or convexity, even in the presence of such information. Monotonicity, in particular, is naturally associated with trend functions arising in many disciplines like climatology (e.g. global warming), environmental and air pollution (e.g. ground-level Ozone or fine particulate matter (PM_{10} or $PM_{2.5}$) as a function of temperature or humidity), engineering (diurnal trends in network traffic loads), among many others. One motivating application, illustrated in Figure 2.4 below, involves the annual global temperature anomalies data available in the NASA website [2]. The data comprises of annual temperature records, measured relative to a baseline mean temperature, during the period 1850–1999; see also Jones and Mann (2014)[39] for a study of the paleoclimatic temperature and Steig et. al. (2009)[60] for evidence of warming trends at Antarctic locations. In the context of environmental pollution, monotone trends have been observed, for example, in the monitoring of water quality (Meal, 2001[43]), and mercury concentration of edible fish (Hussian et. al. 2005 [38]). In air pollution monitoring it is often the case that important factors (temperature, humidity, elevation) have isotonic effect on the concentration of pollutants[46]. In many such scenarios, a natural model for the response as a function of time (or another natural covariate) is to write it as in (2.1) as the sum of an unobserved *monotone trend* function and *dependent noise*. A fundamental problem of interest is then to provide accurate confidence intervals for the underlying parameter that work well in practice under a variety of trend and dependence conditions of the data.

In the context of independent observations, the study of isotonic inference dates back to Rao (1969)[53]. Since then, the field has amassed a large body of research (see e.g. Banerjee and Wellner (2001, 2005)[14, 15], Banerjee (2007,2009) [12, 13], Brunk (1970) [19], Groeneboom (1985) [30], Groeneboom and Wellner (1992) [33], Sun and Woodroofe (1993, 1996) [67, 62], to name a few). *Yet*, isotonic inference in the presence of *dependence* is relatively less developed, despite a clear need for it. Recent breakthrough was achieved thanks to the important work of Anevski and Hossjer (2006) [9] and Zhao and Woodroofe (2012)[71]. [9] develops a general asymptotic scheme for inference under order restrictions that applies, in principle, to arbitrary dependence in the model. A number of practical, as well as, theoretical challenges, however, remain open. Most notably, deriving confidence intervals (based on the work [9] and [71]) requires estimation of the derivative $m'(t_0)$ of the unknown function. This is known to be a difficult problem in the context of shape restricted inference and often leads to biased confidence intervals and substantial under-coverage in practice, as will be demonstrated later. Here we have discussed new methodology based on discrepancy type statistics and the corresponding theory, purely within the isotonic regression framework, in the signal plus noise model for making point wise inference on a monotone trend function that largely circumvent the nuisance parameter estimation problems above and are substantially more robust to functions ill-behaved around the point of interest. Our approach should be contrasted with ones that *combine* isotonization with smoothing; see, for example Mammen (1991) [41], Mukherjee (1988) [45], Pal and Woodroofe (2007) [49], Ramsay (1998) [52] where a variety of methods of this type have been developed in the i.i.d. framework, but typically under higher order smoothness assumptions. Here, our goal is to work under minimal smoothness assumptions and to provide estimates that apply to a broad variety of both weakly and strongly dependent (stationary) error structures.

The second part deals with general problem of M-estimation under dependence. Statistical estimators obtained by maximizing or minimizing objective functions are called M-estimators and can be viewed as a broad generalization of the ubiquitous least squares or maximum likelihood estimates. The literature on M–estimation for i.i.d. data is very well developed; see, for example, Huber (1967)[36], Newey & McFadden (1994)[47], Fan, Hu & Tourong (1994)[28], Arcones (2000) [10], Kim & Pollard (1990)[40] and Van Der Vaart & Wellner (1996)[65]. Among these, [40] and [65] provide the most generals frameworks that use modern empirical process techniques to handle problems even with non-standard rates of convergence. In contrast, considerably less attention has been paid to M–estimation under dependence. However, dependence is a natural phenomenon arising in diverse fields like economics, finance, climate studies and geology. While M-estimation under dependence has been studied in specific contexts like nonparametric regression ([22]), linear models ([68]), local M-estimation ([20]), a general treatment that handles both standard and non-standard problems as in [65] (for the i.i.d. case) is missing. Caner [21] extends the set-up of [65] for mixing-type errors and establishes certain results on convergence rates but does not provide asymptotic distributions. In this chapter, we extend the treatment of [65] to short-range dependent errors – specifically, those coming from absolutely regular mixing sequences – and provide unified results on the consistency, rate of convergence and asymptotic distributions of M-estimators in this setting.

A key challenge in extending the results of [65] under dependence is the relative paucity of suitable empirical process techniques for dependent data; see, for example, [8] for a reasonably comprehensive summary of available empirical process theory under weak dependence. While there are a number of structures available to model weak dependence [mixing as well as the Woodroofe type conditions etc etc], empirical process type results are only available for absolutely regular mixing or β-mixing data. Based on the seminal work of Rio ([55]), Doukhan, Massart and Rio developed functional central limit theorems under β-mixing in [25] and [26] which were subsequently used in [21] in the study of convergence rates of M-estimators under β-mixing. Our work also builds on the the rich framework of [26] but necessitates some interesting extensions. In particular, to obtain the asymptotic distribution of the (appropriately normalized) M-estimator, we need to investigate the limit behavior of certain local (in a neighborhood of the true parameter) processes arising from the objective functions, and to that end, we develop a double–array version of functional central limit theorem in [26]. Thanks to the extension of the Lindeberg Central Limit Theorem for β mxing by [56], the only missing, though non–trivial, ingredient for the functional CLT is establishing tightness, which we accomplish via a generalization of the maximal inequality provided in [26] to double–arrays. This requires an extension of the chaining type arguments of [26] and is expected to be of independent interest.

CHAPTER II

Inference for Monotone Functions Under Short and Long Range Dependence

2.1 Problem formulation and Preliminaries

Consider the isotonic regression model

$$Y_i = m(t_i) + \epsilon_i, \quad i = 1, \cdots, n, \tag{2.1}$$

where $m : [0,1] \to \mathbb{R}$ is an unknown *monotone non-decreasing* function, $t_i = i/n,\ i = 1, \cdots, n$ is a fixed uniform design and where the errors ϵ_i have zero means and variance $\mathrm{Var}(\epsilon_i) = \sigma^2$. We are interested in the case where the noise is a stationary time series $\{\epsilon_k\}_{k\in\mathbb{Z}}$ with non-trivial dependence structure. As indicated, we will consider both the short– and long–range dependent regimes, described in detail in Section 2.1.1. The trend m will be assumed to satisfy the following general condition.

Assumption C. *The regression function $m(t)$ is continuously differentiable in a neighborhood of t_0 with $m'(t_0) > 0$.*

Our ultimate goal is to construct an asymptotic confidence interval for $m(t_0)$, $(0 < t_0 < 1)$, which is largely robust to the dependence structure of the errors. To this end, we consider the testing problem: $H_0 : m(t_0) = \theta_0$ vs. $H_1 : m(t_0) \neq \theta_0$.

Confidence intervals for $m(t_0)$ will be obtained by inversion of acceptance regions of tests for the above problem. Consider the usual isotonic regression estimate (IRE) of m (cf. [57]), obtained as

$$(\hat{m}_n(t_i),\ i = 1, \cdots, n) = \operatorname*{Argmin}_{m_1 \leq \cdots \leq m_n} \sum_{i=1}^{n} (Y_i - m_i)^2. \tag{2.2}$$

To address the above testing problem, we also consider the following *constrained* isotonic estimate $\hat{m}_n^0$. Let $l = \lfloor nt_0 \rfloor$, so that $t_l \leq t_0 < t_{l+1}$ and define

$$(\hat{m}_n^0(t_i),\ i = 1, \cdots, n) = \operatorname*{Argmin}_{m_1 \leq \cdots \leq m_l \leq \theta_0 \leq m_{l+1} \leq \cdots \leq m_n} \sum_{i=1}^{n} (Y_i - m_i)^2. \tag{2.3}$$

Note that both functions $\hat{m}_n$ and $\hat{m}_n^0$ are identified only at the grid points. By convention, we extend them as left-continuous piecewise constant functions defined on the entire interval $(0, 1]$.

Our hypothesis tests will be based on the following *discrepancy statistics* which are scaled versions of $\mathbb{L}_n$ and $\mathbb{T}_n$ introduced in the previous section. Namely,

$$\begin{aligned} L_n &= \frac{n}{\sigma_n^2} \left(\sum_{i=1}^{n} (Y_i - \hat{m}_n^0(t_i))^2 - \sum_{i=1}^{n} (Y_i - \hat{m}_n(t_i))^2 \right) \equiv \frac{n}{\sigma_n^2} \mathbb{L}_n \\ T_n &= \frac{n}{\sigma_n^2} \sum_{i=1}^{n} \left(\hat{m}_n(t_i) - \hat{m}_n^0(t_i) \right)^2 \equiv \frac{n}{\sigma_n^2} \mathbb{T}_n, \end{aligned} \tag{2.4}$$

where $\sigma_n^2 = \mathrm{Var}(\sum_{k=1}^{n} \epsilon_k)$. A third statistic which will prove particularly useful in the long-range dependence case is the 'ratio statistic' $R_n := L_n / T_n$. Its asymptotic properties will be derived from the joint asymptotic behavior of L_n and T_n.

2.1.1 Dependence Structure

In this section, we introduce and discuss our formal assumptions on the dependence structure of the errors ϵ_i's in (2.1). These assumptions will be tacitly adopted for the rest of the paper and will require some technicalities for a precise description.

We suppose that the errors have zero means, finite variances and form a strictly stationary time series $\{\epsilon_k\}_{k\in\mathbb{Z}}$. Let $S_n = \sum_{k=1}^n \epsilon_k$, and consider the piece-wise linear cumulative sum diagram

$$w_n(t) = \frac{1}{\sigma_n}\Big(\sum_{i=1}^{\lfloor nt\rfloor} \epsilon_i + (nt - \lfloor nt\rfloor)\epsilon_{\lfloor nt\rfloor+1}\Big), \quad \text{where} \quad \sigma_n^2 = \mathrm{Var}(S_n). \tag{2.5}$$

The asymptotic behavior of the process $\{w_n(t)\}_{t\geq 0}$ is generally determined by the degree of dependence of the errors in addition to their tail behavior. If the ϵ_i's are weakly dependent, then as in the usual Donsker theorem, the limit is the Brownian motion and the corresponding statistical results are similar to the situation of independent errors. On the other hand, as noted in the introduction, strong dependence of the ϵ_i's leads to different types of limits and new statistical theory. We shall consider two different regimes: **[i]** short-range dependent errors, and, **[ii]** long-range dependent errors. Let $\mathrm{Cov}(k) = \mathrm{Cov}(\epsilon_1, \epsilon_{1+k})$. A stationary finite variance time series $\{\epsilon_k\}_{k\in\mathbb{Z}}$ is said to be *short-range dependent* if $\sum_k |\mathrm{Cov}(k)| < \infty$. Otherwise, if $\sum_k |\mathrm{Cov}(k)| = \infty$, the time series is referred to as *long-range dependent*. As indicated above, the dependence structure of the errors plays a critical role in determining the type of the limit process.

2.1.1.1 Short Range Dependence

To formalize weak dependence, let $\|\cdot\|$ denote the L^2 norm on the probability space and introduce the discrete filtration $\mathcal{F}_n = \sigma\{\epsilon_m,\ m \leq n\},\ n \in \mathbb{Z}$, i.e. $\mathcal{F}_n$ is the σ-algebra generated by all errors up to and including 'time' n. In the short range dependent case, following [71], we shall assume that

$$\sum_{n=1}^{\infty} n^{-\frac{3}{2}} \|\mathbb{E}(S_n|\mathcal{F}_0)\| < \infty. \tag{2.6}$$

It is shown in [50] that if (2.6) is satisfied then,

$$\Gamma := \sum_{k=0}^{\infty} 2^{-\frac{1}{2}k} \|\mathbb{E}(S_{2^k}|\mathcal{F}_0)\| < \infty \quad \textit{and} \quad \mathbb{E}\left[\max_{k \le n} S_k^2\right] \le 6\left[\mathbb{E}(\epsilon_1^2) + \Gamma\right] n. \tag{2.7}$$

Furthermore, the limit

$$\tau^2 = \lim_{n \to \infty} \frac{1}{n} \mathbb{E}(S_n^2) < \infty \tag{2.8}$$

exists and the process $\{w_n(t)\}_{0 \le t \le 1}$ converges in distribution to the Brownian motion $\mathbb{B}$ in the space $D[0,1]$ equipped with the usual J_1-Skorohod topology. This result and a careful continuous mapping argument will be used in the sequel to establish the asymptotic behavior of our statistics under short range dependence.

Remark II.1. In [9], weak dependence was quantified in terms of mixing conditions. Here, we use an alternative condition (2.6) from [71], implied by the strong mixing Assumption (A9) of [9], and therefore weaker.

2.1.1.2 Long Range Dependence

A great variety of models exhibit long-range dependence. We focus here on a special but important case when $\epsilon_k = g(\xi_k),\ k \in \mathbb{Z}$, where $\{\xi_i\}_{i \in \mathbb{Z}}$ is a stationary Gaussian time series with zero mean. The function g is deterministic and from $L^2(\phi)$ where ϕ denotes standard normal density, i.e., $\mathbb{E}(g(Z))^2 < \infty$ where $Z \sim N(0,1)$. In this setting, an elegant theory characterizing the possible limits of the cumulative sums in (2.5) was developed in the seminal work of Taqqu ([63], [64]).

Following [9], let $\mathrm{Cov}(k) = \mathbb{E}(\xi_i \xi_{i+k})$ be such that $\mathrm{Cov}(0) = 1$ and $\mathrm{Cov}(k) = k^{-d} l_0(k)$, where $0 < d < 1$ is fixed and l_0 is a function slowly varying at infinity, i.e., for all $a > 0$, $l_0(ax)/l_0(x) \to 1$, as $x \to \infty$.

Observe that $\mathbb{E}(\epsilon_i^2) = \int_{\mathbb{R}} g(z)^2 \phi(z) dz < \infty$. Thus, using the Hermite polynomial expansion of the function g, we have

$$\epsilon_i := g(\xi_i) = \sum_{k=r}^{\infty} \frac{\eta_k}{k!} H_k(\xi_i),$$

where the series converges in $L^2(\mathbb{P})$, and where $\eta_k = \mathbb{E}(g(\xi_i)H_k(\xi_i)),\ k \geq r$. Here the H_k's are the Hermite polynomials of order k and the summation starts from $r \geq 1$ – the index of the first nonzero coefficient in the expansion. The index r is referred to as the Hermite rank of the function g. For the rest of the paper we will restrict our discussion to the case $r = 1$

The results of Taqqu ([63], [64]) show that if $0 < d < 1$, the sequence $\{\epsilon_i\}$ also exhibits long range dependence and, in fact,

$$\left\{\sigma_n^{-1} \sum_{i=1}^{\lfloor nt \rfloor} \epsilon_i\right\} \Longrightarrow \left\{B_H(t)\right\}_{t\in[0,1]}, \tag{2.9}$$

in $D[0,1]$ equipped with Skorohod topology, where the limit process B_H is in $C[0,1]$ a.s. It can be shown that

$$\sigma_n^2 = \eta_1^2 n^{2-d} l_1(n)(1 + o(1)), \tag{2.10}$$

where l_1 is another slowly varying function: $l_1(k) = 2l_0(k)/(1-d)(2-d))$.
The limit process B_H and is known as the fractional Brownian motion (fBm) with self-similarity parameter $H = 1-d/2$ also known as the Hurst index. Recall that self-similarity of B_H means that for all $c > 0$, the processes $\{B_H(ct)\}_{t\in\mathbb{R}}$ and $\{c^H B_H(t)\}_{t\in\mathbb{R}}$ are equal in distribution. The stationarity of the increments and self-similarity imply that

$$\mathrm{Cov}(B_H(t), B_H(s)) = \frac{\sigma^2}{2}\left(|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right), \quad t, s \in \mathbb{R}. \tag{2.11}$$

For more details on the properties of the fractional Brownian Motion (fBm), see e.g. the review chapter by Taqqu in [27]. In this paper, we focus only on the long-range dependence regime with Hermite rank $r = 1$, i.e. for strongly dependent errors whose cumulative sums, appropriately normalized, converge to the fractional Brownian motion.

2.2 Main results

We discuss in this section the joint asymptotic behavior of the statistics L_n and T_n. To derive their asymptotic distribution, we focus on a shrinking neighborhood of t_0 at rate $d_n \downarrow 0$, which will determined by the type of dependence structure of the error sequence, since the the constrained and unconstrained isotonic estimates of m, namely $\hat{m}_n$ and $\hat{m}_n^0$ are equal outside of neighborhoods of this order of magnitude. For example, under independence or short-range dependence $d_n \sim n^{-1/3}$, while under long-range dependence the rate will involve the Hurst index. More formally, let $z := d_n^{-1}(t - t_0)$ and define

$$X_n(z) = \frac{1}{d_n}\left(\hat{m}_n\left(t_0 + zd_n\right) - \theta_0\right) \text{ and } Y_n(z) = \frac{1}{d_n}\left(\hat{m}_n^0\left(t_0 + zd_n\right) - \theta_0\right), \tag{2.12}$$

for $z \in (a_n, b_n] := (-d_n^{-1}t_0, d_n^{-1}(1 - t_0)]$. Here $\theta_0 = m(t_0)$. It turns out that the statistics L_n and T_n can be represented asymptotically as fairly simple integrals involving X_n and Y_n and also that the set of z's on which they differ is contained, with high probability, in a compact set. These are the contents of the below propositions.

Proposition II.2. *For L_n and T_n as in* (2.4)*, we have*

$$\begin{aligned} L_n &= \frac{n^2 d_n^3}{\sigma_n^2}\Big(\int_{(a_n, b_n]} \left(X_n^2(z) - Y_n^2(z)\right) dz + o_P(1)\Big) \qquad (2.13) \\ T_n &= \frac{n^2 d_n^3}{\sigma_n^2}\Big(\int_{(a_n, b_n]} \left(X_n(z) - Y_n(z)\right)^2 dz + o_P(1)\Big). \end{aligned}$$

For a proof of this result, see Section 2.5.2.3.

Lemma II.3. *Let* $D_n := \{z \in \mathbb{R} : X_n(z) \neq Y_n(z)\}$. *For any* $\epsilon > 0$, *there exist* $M_\epsilon > 0$ *and* $n_\epsilon > 0$, *such that*

$$\mathbb{P}\Big(D_n \subset [-M_\epsilon, M_\epsilon]\Big) \geq 1 - \epsilon,$$

for all $n \geq n_\epsilon$.

The proof of this lemma is given in Section 2.5.2.2. It is then clear that fathoming the asymptotic behavior of L_n, T_n requires an understanding of the asymptotic behavior of the processes (X_n, Y_n) on compact sets, since with high probability, the difference set D_n is contained in a compact set. If we could show that (X_n, Y_n) converge to limit processes (X_∞, Y_∞) with increasing sample size on every compact set (in a strong-enough metric under which integral type functionals are continuous), then, roughly speaking, up to adequate normalizations our limits for L_n and T_n should have forms:

$$\int (X_\infty^2(z) - Y_\infty^2(z))\, dz \ \text{ and } \ \int (X_\infty(z) - Y_\infty(z))^2\, dz\,,$$

respectively. It turns out that the topology of L^2 convergence on compact sets is adequate for this purpose. To properly state the limiting distribution of (L_n, T_n), we now introduce the greatest convex minorant (GCM) functionals.

Greatest convex minorants: Let $\mathcal{T}_I(f)$ denote the GCM of a real-valued function f, defined on an interval $I \subseteq \mathbb{R}$. For an interval $J \subset I$, we denote the GCM of the *restriction of* f *to* J by $\mathcal{T}_J(f)$. When f is defined on $\mathbb{R}$, we sometimes write $\mathcal{T}(f)$ for $\mathcal{T}_{\mathbb{R}}(f)$ and $\mathcal{T}_c(f)$ for $\mathcal{T}_{[-c,c]}(f)$. Also, let $\mathcal{L}(f)$ denote the left derivative functional of a convex function f, which is a well-defined, non-decreasing and left-continuous function (cf. Theorem 24.1 of [59]).

The processes X_n and Y_n can be represented as greatest convex minorant functionals of a normalized version of the the process U_n, the linear interpolation of the cumulative sum process of the Y_i's, namely:

$$U_n(t) = \frac{Y_1 + Y_2 + \cdots + Y_{\lfloor nt \rfloor}}{n} + \frac{(nt - \lfloor nt \rfloor)}{n} Y_{\lfloor nt \rfloor + 1}, \; t \in [0, 1]\,. \tag{2.14}$$

More specifically, defining:

$$\mathbb{V}_n(z) := d_n^{-2} \Big(U_n(t_0 + d_n z) - U_n(t_0) - m(t_0) d_n z \Big), \qquad z \in (a_n, b_n]\,, \tag{2.15}$$

we can write:f

$$\begin{aligned} X_n(z) =& \mathcal{L} \circ \mathcal{T}_{(a_n, b_n]} \left(\mathbb{V}_n \right)(z) \qquad (2.16) \\ Y_n(z) =& \left(\mathcal{L} \circ \mathcal{T}_{(a_n, l_n]} \left(\mathbb{V}_n \right)(z) \wedge 0 \right) \mathbf{1}_{(a_n, l_n]}(z) + 0 \times \mathbf{1}_{(l_n, 0]}(z) \\ &+ \left(\mathcal{L} \circ \mathcal{T}_{(l_n, b_n]} \left(\mathbb{V}_n \right)(z) \vee 0 \right) \mathbf{1}_{(0, b_n]}(z), \end{aligned}$$

where $l_n = d_n^{-1}(t_l - t_0)$. This is a direct consequence of the well-known representation of $\hat{m}_n$ and $\hat{m}_n^0$ in terms of U_n (See (II.23) in Section 2.5.2) followed by an appropriate renormalization. The limiting properties of (X_n, Y_n) are therefore driven by those of $\mathbb{V}_n$, which is fortunately well-studied ([9]). More concretely, we know:

Theorem 1. *Consider the processes $\mathbb{V}_n$ in the space $C(\mathbb{R})$ equipped with the topology of uniform convergence on compact sets. Then, as $n \to \infty$,*

$$\mathbb{V}_n \Longrightarrow \{\mathbb{G}(z)\}_{z \in \mathbb{R}} \equiv \{\mathbb{G}_{a,b}(z)\}_{z \in \mathbb{R}} := \{a \mathbb{W}(z) + b z^2\}_{z \in \mathbb{R}}, \tag{2.17}$$

where $b = \frac{1}{2} m'(t_0)$ and (i) (under weak dependence) $d_n = n^{-\frac{1}{3}}$, $\mathbb{W}$ is a two-sided Brownian motion on $\mathbb{R}$, and $a := \tau$ given in (2.8).
(ii) (under strong dependence) $d_n = l_2(n) n^{-\frac{d}{2+d}}$, $\mathbb{W}$ is the fBm process B_H and $a := |\eta_1|$. (Here l_2 is a slowly varying function related to l_1 as shown in the proof of

the theorem, provided in Section 2.5.2.1.)

We therefore expect the limits of (X_n, Y_n) to be given by replacing $a_n, b_n, l_n, \mathbb{V}_n$ by their corresponding limits in (2.16). Thus, the limits, (X_∞, Y_∞) above, should have the form:

$$\begin{aligned} \mathcal{S}_{a,b}(z) &= \mathcal{L} \circ \mathcal{T}(\mathbb{G})(z) \\ \mathcal{S}^h_{a,b}(z) &= \begin{cases} \mathcal{L} \circ \mathcal{T}_{(-\infty,0)}(\mathbb{G})(z) \wedge h & , \ z \in (-\infty, 0) \\ \lim_{u \uparrow 0} \mathcal{L} \circ \mathcal{T}_{(-\infty,0)}(\mathbb{G})(u) \wedge h & , \ z = 0 \\ \mathcal{L} \circ \mathcal{T}_{(0,\infty)}(\mathbb{G})(z) \vee h & , \ z \in (0, \infty) \end{cases} \end{aligned} \tag{2.18}$$

We next define the space L^2_{loc} which appears in the statement of convergence of (X_n, Y_n). This is the space of all functions which are square integrable on compact sets. The convergence in this space is accordingly defined, that is, a sequence of functions $f_n \to f$ as $n \to \infty$ in L^2_{loc} if $\int_I (f_n - f)^2 \to 0$ as $n \to \infty$ for every compact interval I. In fact with this convergence the space is metrizable. The next theorem formalizes the limit behavior of (X_n, Y_n).

Theorem 2. *As $n \to \infty$, we have*

$$\{(X_n(z), Y_n(z))\}_{z \in \mathbb{R}} \Longrightarrow \{(\mathcal{S}_{a,b}(z), \mathcal{S}^0_{a,b}(z))\}_{z \in \mathbb{R}} \quad \text{in } L^2_{loc} \times L^2_{loc}, \tag{2.19}$$

where the components of the limit process are defined in (2.18).

The formal proof of this theorem is highly technical and provided in Section 2.5.2.2.

Remark II.4. Under short–range dependence, the convergence in (2.19) can be shown to hold in the sense of finite–dimensional distributions, which together with monotonicity implies convergence in L^2_{loc}. In the long-range dependence case, this remains an open problem since extensive knowledge about the GCM of fBm plus quadratic drift is lacking. The weaker form of L^2_{loc}-convergence, however, suffices to deduce the limit behavior of L_n and T_n as shown below.

We are now ready to state the limit distributions of the statistics L_n and T_n in terms of $\mathcal{S}_{a,b}(z)$ and $\mathcal{S}^0_{a,b}(z)$. Define:

$$\mathbb{L}_{a,b} = \int\limits_{\mathbb{R}} \left((\mathcal{S}_{a,b}(z))^2 - (\mathcal{S}^0_{a,b}(z))^2 \right) dz \quad \text{and} \quad \mathbb{T}_{a,b} = \int\limits_{\mathbb{R}} \left(\mathcal{S}_{a,b}(z) - \mathcal{S}^0_{a,b}(z) \right)^2 dz. \tag{2.20}$$

Remark II.5. The processes $\mathcal{S}_{a,b}(z)$ and $\mathcal{S}^0_{a,b}(z)$ differ on a compact interval. This is suggested by the asymptotic behavior of D_n and rigorously established in Theorem 18 of Appendix B *showing that the statistics in* (2.20) *are proper random variables.*

Remark II.6. In the long–range dependent case where $\mathbb{L}_{a,b}$ and $\mathbb{T}_{a,b}$ depend on the Hurst index H, we denote them by $\mathbb{L}^{(H)}_{a,b}$ and $\mathbb{T}^{(H)}_{a,b}$. When $a = b = 1$, we drop the subscripts and write $\mathbb{L}$ and $\mathbb{T}$ in the short–range dependent case, and $\mathbb{L}^{(H)}$ and $\mathbb{T}^{(H)}$ in the long–range dependent case. **In the following sections we will, often, drop H and just use $\mathbb{L}$ and $\mathbb{T}$ for both short and long range dependence when there is no chance of confusion.**

Theorem 3. *For short–range dependent errors, $(L_n, T_n) \Rightarrow (\mathbb{L}, \mathbb{T})$, as $n \to \infty$.*

Theorem 4. *For long–range dependent errors, as $n \to \infty$,*

$$\frac{\sigma_n^2}{n^2 d_n^3}(L_n, T_n) \Longrightarrow a^2 \left(\frac{a}{b}\right)^{\frac{2H-1}{2-H}} (\mathbb{L}^{(H)}, \mathbb{T}^{(H)}), \tag{2.21}$$

where $a = |\eta_1|, b = \frac{1}{2}m'(t_0)$, σ_n^2 is the variance of partial sum of ϵ_i's and d_n is as in Theorem 1 (ii).

The proof for the long-range case is given in Section 2.5.2.3. The proof in the simpler, short-range case is similar and omitted for brevity.

Remark II.7. It is natural to ask whether our results can be extended to the case of higher order long–range dependence, i.e. Hermite ranks $r \geq 2$. This is an interesting open challenge. In this case, non-Gaussian limits such as the Rosenblatt process ($r = 2$) arise in place of the fractional Brownian motion and new probabilistic tools need to be developed. One key difficulty in this context would be to study the GCM functionals of drifted self–similar processes with stationary increments represented as iterated stochastic integrals. Another challenge is showing that the processes $\mathcal{S}_{a,b}$ and $\mathcal{S}^0_{a,b}$ coincide outside a compact set. Our Theorem 18 in Appendix B establishes this in the case $r = 1$ by critically using the underlying normality (see the proof of Lemma B.5). Further, the behavior of the slope processes X_n and Y_n (Theorem 2) depends on certain path properties of quadratically drifted (fractional) Brownian motion, which we establish in the appendix. We expect that such results will be possible but rather technical when $r \geq 2$. Last but not least, methodology involving higher order Hermite rank r should also account for its statistical estimation. This is to the best of our knowledge a largely unexplored problem of independent interest.

2.3 Methodology

2.3.1 An asymptotically pivotal ratio statistic

Recall (2.4). To be able to use these statistics in the SRD case one needs a suitable 'plug–in' estimate for σ_n^2. This, however, is not difficult as $\sigma_n^2 \sim n\tau^2$, where the parameter τ^2 in (2.8) can typically be estimated well in practice, as in (2.23). The Wald type confidence intervals of AH and ZW, however, require the estimation of $m'(t_0)$ in addition to τ^2. The estimation of the latter is a much harder problem and typically leads to biased estimates in practice (More details are provided in our simulation results in Section 6).

The use of the statistics L_n and T_n in the Long Range Dependence (LRD) case, however, is much more challenging, because by (2.10), σ_n^2 (and in turn d_n) involve an unknown slowly varying function and Hurst parameter, as well as the derivative

$m'(t_0)$. In practice, ignoring slowly varying functions, one can use established estimators for the Hurst parameter, which is a challenging problem in its own right, but the dependence on $m'(t_0)$ remains. An elegant way to eliminate the need for a plug-in estimate of σ_n^2 as well as the constants a and b in the LRD case is to consider the ratio statistic, introduced next.

Note that L_n and T_n are always non-negative by definition. By (2.4), if $T_n = 0$ we have $L_n = 0$. Also as shown in Lemma II.28 from Section 2.5.3, we have $L_n \geq T_n$. Therefore L_n and T_n are either both equal to 0 or both strictly positive. Similarly (2.20) implies that $\mathbb{L} = 0$ iff $\mathbb{T} = 0$ and by Theorems 3 and 4 and the Portmanteau theorem, we obtain that $\mathbb{L} \geq \mathbb{T}$ almost surely.

Now, define the ratio statistic $R_n = L_n/T_n$, where $0/0$ is interpreted as 1. By the discussion in the above paragraph, $\mathbb{P}(R_n < \infty) = 1$.

Theorem 5. *For both short– and long–range dependent errors, we have*

$$R_n \Longrightarrow \mathcal{R} := \frac{\mathbb{L}}{\mathbb{T}}, \qquad \text{as } n \to \infty,$$

where the limit has a proper probability distribution.

Proof. The convergence follows from Theorems 3, 4 and the Continuous Mapping Theorem, provided that $\mathbb{P}(\mathbb{T} = 0) = 0$. The latter is true thanks to Theorem 19 in Appendix B. □ □

Remark II.8. *The limit distribution of R_n is essentially pivotal, involving only the Hurst parameter!*

2.3.2 Construction of Confidence Intervals

Let $L_n(\theta)$ and $T_n(\theta)$ denote the residual sum of squares and L_2 statistics, respectively, for testing $H_0 : m(t_0) = \theta$ against $H_a : m(t_0) \neq \theta$. Letting θ_0 denote the true value of $m(t_0)$, an asymptotic level $1 - \alpha$ confidence set for θ_0, using inversion of L_n, is given by $\{\theta : L_n(\theta) \leq F_{\mathbb{L}}^{\leftarrow}(1-\alpha)\}$, where $F_{\mathbb{L}}^{\leftarrow}$ denotes the left–continuous quantile function of $F_{\mathbb{L}}$, the distribution function of $\mathbb{L}$. The statistics T_n can be used similarly to obtain confidence interval. The shape of $L_n(\theta)$ (or $T_n(\theta)$) is described in Lemma II.29 of Section 2.5.3. For convenience, we state that result below.

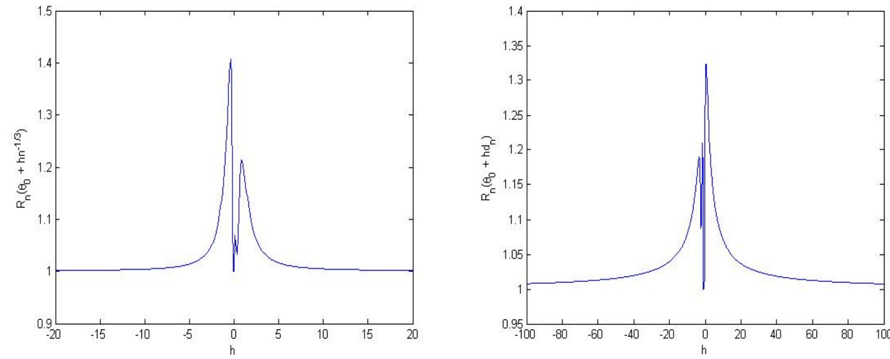
Proposition II.9. *Both $L_n(\theta)$ and $T_n(\theta)$ are continuous in θ, monotone non-increasing on $(-\infty, \hat{\theta}_n]$, monotone non-decreasing on $(\hat{\theta}_n, \infty)$ and $L_n(\hat{\theta}_n) = T_n(\hat{\theta}_n) = 0$. Also, both $L_n(\theta)$ and $T_n(\theta)$ diverge to infinity as $|\theta| \to \infty$.*

Next, let

$$C_L(\alpha) := \inf\{\theta : L_n(\theta) < F_{\mathbb{L}}^{\leftarrow}(1-\alpha)\} \ \text{ and } \ C_U(\alpha) := \sup\{\theta : L_n(\theta) < F_{\mathbb{L}}^{\leftarrow}(1-\alpha)\}.$$

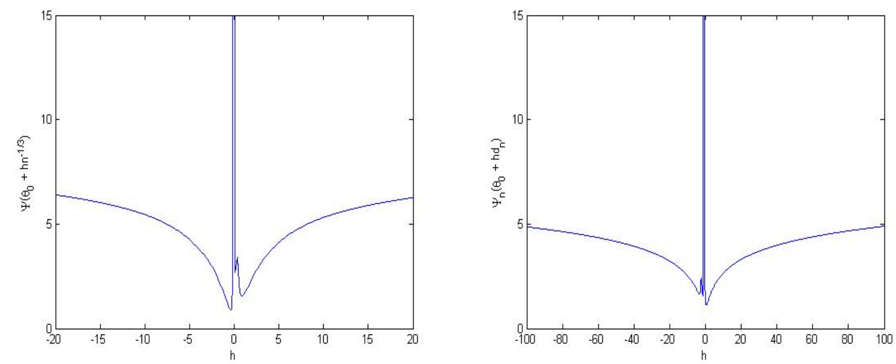
Then, by the above proposition, $[C_L(\alpha), C_U(\alpha)]$ is precisely the set $\{\theta : L_n(\theta) \leq F_{\mathbb{L}}^{\leftarrow}(1-\alpha)\}$, giving us a $100(1-\alpha)\%$ confidence interval for θ_0. The simulated quantiles for $\mathbb{L}$ for both short– and long–range dependent errors with different Hurst parameters H can be found in Table B.3 and B.4 in Appendix B. A confidence interval based on T_n may be similarly obtained. As noted above, under long range dependence, L_n and T_n are not as useful as in the short range case, while R_n still manages to eliminate the key nuisance parameter $m'(t_0)$ and it is to this that we turn our attention next.

Consider first, the shape of $R_n(\theta)$ as a function of θ. It assumes the value 1 at $\theta = \hat{m}_n(t_0)$, converges to 1 as $|\theta| \to \infty$ and displays irregular humps in between. Figure 2.1 illustrates the behavior of this statistic as a function of h, where $h = n^{1/3}(\theta - \theta_0)$ under SRD and $h = d_n^{-1}(\theta - \theta_0)$ under LRD. As a sensible inversion of R_n should avoid values away from $\hat{m}_n(t_0)$, an asymptotic confidence set should look like: $\{\theta : R_n(\theta) > \zeta\}$, where ζ is an appropriate quantile (depending on the level of confidence desired) of $\mathcal{R}$, the limiting random variable in Theorem 5. This will, however, not yield a confidence interval but a rather irregular confidence set, and, in particular, may miss values of θ close to θ_0.



(a) Short Range Dependence (b) Long Range Dependence

Figure 2.1: Shape of Ratio Statistic as a function of h



(a) Short Range Dependence (b) Long Range Dependence

Figure 2.2: Shape of Ψ-Statistic as a function of h

Another issue with using R_n is that the quantiles of $\mathcal{R}$ grow extremely slowly from 1 and are hard to represent in a table. For matters of practical convenience, we therefore make a monotone transformation of R_n, namely,

$$\Psi_n(\theta) = \begin{cases} -\log(R_n(\theta) - 1), & \text{if } R_n(\theta) > 1 \\ \infty, & \text{if } R_n(\theta) = 1. \end{cases}$$

Then, the following Proposition follows easily from Theorem 5 and the Continuous Mapping Theorem.

Proposition II.10. *Under the assumptions of Theorem 5, we have*

$$\Psi_n(\theta_0) \xrightarrow{d} \Psi := -\log(\mathcal{R} - 1) \text{ as } n \to \infty. \tag{2.22}$$

where $\mathbb{P}(\Psi = \infty) = \mathbb{P}(\mathcal{R} = 1)$.

As Ψ_n is a monotone decreasing transformation of $R_n = L_n/T_n$, it exhibits the same irregularities; see Figure 2.2, and therefore, in terms of Ψ_n, our confidence set $\{\theta : \Psi_n(\theta) < -\log(\zeta - 1)\}$, is still irregular. To avoid this, we propose a confidence interval of the form $[\widetilde{C}_L(\alpha), \widetilde{C}_u(\alpha)]$, where $\widetilde{C}_L$ and $\widetilde{C}_U$ are defined thus:

$$\widetilde{C}_L(\alpha) := \inf\{\theta : \Psi_n(\theta) < F_\Psi^{\leftarrow}(1-\alpha)\}, \;\; \widetilde{C}_U(\alpha) := \sup\{\theta : \Psi_n(\theta) < F_\Psi^{\leftarrow}(1-\alpha)\}.$$

Note that this gives us a conservative $100(1-\alpha)\%$ C.I. for θ_0.
While our knowledge of the behavior of $R_n(\theta)$ is limited, we do have the following result.

Proposition II.11. *Let* $\theta \neq \theta_0$ *and* $R_n(\theta)$ *be the ratio statistic calculated under the null hypothesis* $H_{0,\theta} : m(t_0) = \theta$*. Then,* $R_n(\theta) \xrightarrow{P} 1$ *as* $n \to \infty$.

Therefore the probability that any $\theta \neq \theta_0$ is outside our proposed honest confidence interval converges to 1. The proof of this lemma is available in Section 2.5.3.

Remark II.12. It is unclear that Ψ is a proper random variable, i.e., $\mathbb{P}(\Psi < \infty) = 1$. Extensive simulations suggest that this should be the case, and also that the distribution function is continuous and strictly increasing. It is possible that the distribution of $\mathcal{R}$ may harbor a small mass at the point 1 (and therefore Ψ a mass at ∞), undetectable by simulations. But Proposition II.11 implies that confidence intervals (at level $100(1-\alpha)\%$) based on Ψ (or equivalently on $\mathcal{R}$) would be consistent provided that $\alpha > \mathbb{P}(\mathcal{R} = 1)$, since the $(1-\alpha)$ quantile of Ψ would then be finite. Based on our simulations, if such an α does exist it would have be orders

Table 2.1: Quantiles of Ψ

p	SRD	H = 0.7	H = 0.8	H = 0.9	H = 0.95
0.50	2.21 (0.021)	2.19 (0.006)	2.11 (0.012)	2.20 (0.051)	2.66 (0.025)
0.80	24.25 (0.020)	23.79 (0.019)	10.89 (0.494)	5.72 (0.132)	5.77 (0.122)
0.85	24.67 (0.022)	24.51 (0.036)	24.14 (0.023)	8.43 (0.539)	8.30 (0.096)
0.90	25.00 (0.041)	25.12 (0.031)	25.28 (0.054)	26.43 (0.165)	27.05 (0.248)
0.95	25.21 (0.023)	25.92 (0.017)	26.32 (0.026)	28.02 (0.489)	33.13 (0.188)

of magnitude smaller than .01, so this would have no bearing on the construction of usual confidence intervals.

Remark II.13. Note that, by Proposition II.9, the intervals produced by inverting L_n and T_n are always of finite length. In contrast, the quantile $F_{\Psi}^{\leftarrow}(1-\alpha)$ may lie entirely below the graph of the statistic $\Psi_n(\theta)$ with some positive probability. In particular, this corresponds to the case where $L_n(\theta) = T_n(\theta)$ for all θ. As shown in Proposition II.30 in Section 2.5.3, this happens at the points where isotonic regression estimator jumps. (Note however for a pre-fixed point of interest the probability of it being a jump point is zero.) In this case, the confidence interval from inversion of the Ψ–statistic is the empty set. Also, note that with non-zero probability, the confidence interval based on Ψ_n can be the entire range of the function m, though this probability, by the observation following Proposition II.11, goes to 0 as n increases.

Selected quantiles of Ψ are presented in Table 2.3.2. See Tables B.1 and B.2 in Appendix **??** for a detailed presentation of the quantiles of Ψ.
Finally to construct confidence interval using Ψ-statistic for long-range dependence case needs an estimate of the Hurst index H. To this end we state the following proposition about the effect of a plug–in estimator of H.

Proposition II.14. *If the quantiles of Ψ are continuous as a function of H, any consistent plug-in estimate of H will give confidence interval with appropriate coverage.*

Remark II.15. Note that in Remark B.9 in Appendix B we conjecture that the quantiles of Ψ are continuous as a function of H. Therefore by Proposition II.14 any plug–in consistent estimate of H should give reasonable confidence interval.

Remark II.16. The order of the confidence intervals based on L_n and T_n is shown to be d_n. (See Theorem 17 from Appendix A.) Based on extensive simulations we conjecture that intervals based on Ψ_n are of the same order. But analysis for Ψ_n is difficult due to irregular nature of its sample path as it is ratio of two convex functions.

2.3.3 Construction of Confidence Band

Our methodology is applicable for pointwise confidence intervals. Construction of uniform confidence band is a challenging problem even in the case with iid errors. Construction of a good confidence band for our problem is an open question and beyond the scope of this paper. However here we propose a conservative method for constructing a simultaneous confidence intervals under short range dependence for the function m relying on its monotonicity property.

To this end define $L_n(\theta, t)$ to be the L_n test statistic for testing $H_0 : m(t) = \theta$. We can define $\Psi_n(\theta, t)$ and $T_n(\theta, t)$ similarly. First consider the problem of construction of simultaneous confidence intervals for the function m at k fixed points $t_1 \leq t_2 \leq \cdots \leq t_k$. By Theorem 7 from Section 2.5.3, $\{L_n(\theta, t_i)\}_\theta$ for $i = 1, \ldots, k$ are asymptotically independent. Therefore we look at $\{L_n(\theta, t_i)\}$ for $i = 1, \ldots, k$ as function of θ and use method of inversion to construct $(1-\alpha)^{1/k}$-level confidence intervals (l_i, u_i) at each of these points, which ensures $\mathbb{P}((l_i, u_i) \ni m(t_i), \forall i) \sim \prod_{i=1}^{k} \mathbb{P}((l_i, u_i) \ni m(t_i)) \geq (1-\alpha)$ for sufficiently large n as l_i and u_i are functions of the process $\{L_n(\theta, t_i)\}_\theta$.

Next we extend this approach to construct a confidence band for the function m. The first step is to monotonize the sequences l_i and u_i, i.e., define $\tilde{l}_1 = l_1$ and

$\tilde{l}_i = \max(\tilde{l}_{i-1}, l_i)$ for $i \geq 2$ and similarly $\tilde{u}_1 = u_1$ and $\tilde{u}_i = \max(\tilde{u}_{i-1}, u_i)$ for $i \geq 2$. Now define the functions l and u as $l(t) = \tilde{l}_i$ if $t \in [t_i, t_{i+1})$ and $u(t) = \tilde{u}_{i+1}$ if $t \in (t_i, t_{i+1}]$.

Proposition II.17. *Let $l(.)$ and $u(.)$ be functions constructed as stated above. Then $(l(t), u(t))$ gives an honest asymptotic $100(1-\alpha)\%$ confidence band for the function m, i.e., $\mathbb{P}\left((l(t), u(t)) \ni m(t),\ \forall t\right) \geq (1-\alpha)$.*

Proof. Let (l_i, u_i) be such that $l_i \leq m(t_i) \leq u_i$, then by construction we have $\tilde{l}_i \leq l_i \leq m(t_i) \leq u_i \leq \tilde{u}_i$. Therefore $l(t_i) \leq m(t_i) \leq u(t_i)$ for $i = 1, \ldots, k$. For any $t \in (t_i, t_{(i+1)})$ we have $m(t) \geq m(t_i) \geq \tilde{l}_i = l(t)$ and $m(t) \leq m(t_{i+1}) \leq \tilde{u}_{(i+1)} = u(t)$. Therefore $\mathbb{P}((l(t), u(t)) \ni m(t),\ \forall t) \geq P((l_i, u_i) \ni m(t_i),\ \forall i = 1, \ldots, k)$ and the last probability is at least $(1-\alpha)$ by the discussion above. $\square$

Remark II.18. Note that theoretically this method works for long range dependent errors using either L_n or Ψ_n. But due to the plug in estimator of $m'(t_0)$ the confidence bands constructed using this method significantly undercover. On the other hand, due to lack of structure in shape of the Ψ-statistic, the confidence bands constructed using this method with Ψ-statistic are generally too conservative. Construction of appropriate confidence intervals under LRD errors is a challenging problem and beyond the scope of this work.

The number of points k should depend on the number of data–points we have. The size of the flat stretches of the isotonic regression estimator is of the order d_n, therefore to ensure the independence of (l_i, u_i) across i we should choose $k \approx 1/d_n$.

2.4 Simulation and Data Analysis

2.4.1 Performance of Point-wise Confidence Intervals

To study the performance of our confidence intervals we consider two choices for $m(t)$, namely:

$$m_1(t) = e^t \quad \text{and} \quad m_2(t) = \begin{cases} t, & t \in (0, 1/4] \\ 1/4 + 20000(t - 1/4)^2, & t \in (1/4, 1/4 + 1/200] \\ t + 3/4, & t \in (1/4 + 1/200, 1]. \end{cases}$$

Observe the capricious behavior of m_2 in the interval $(1/4, 1/4 + 1/200]$, where the function grows rapidly. We choose the midpoint $t_0 = 1/4 + 1/400$ from this interval. For m_1 we choose $t_0 = 1/2$.

In the following sections we demonstrate that our confidence intervals outperform existing methods for both conventional and challenging trend functions such as m_1 and m_2 respectively. We also show that the intervals perform well under both short and long range dependent errors.

Data were generated from the models $y_i = m_j(i/n) + \epsilon_i$, for $i = 1, 2, \ldots, n$, and $j = 1, 2$. The errors were generated from different ARMA processes, fractional Gaussian noise for different Hurst indices and a FARIMA process. The marginal variance of the errors was 0.2 in all cases. Three statistics: R_n, L_n and the Isotonic Regression Estimator (IRE) (defined in (2.2)) were used to construct confidence intervals for $m_1(0.5)$ in the first case, and $m_2(0.25 + 1/400)$ in the second. To use IRE, we constructed Wald-type confidence intervals based on the results of [9] and [71]. The required quantiles for this method can be found in [31] for the weak dependence case. For long range dependent errors, we simulated (approximations to) the quantiles for some specific values of H. The average length and coverage of

90% confidence intervals based on 1000 repetitions were reported for various sample sizes (n). The (Binomial) standard error of the coverage was calculated to be 0.3%. Constructing confidence intervals using R_n is straightforward and follows the method outlined in the previous section. In order to use L_n and the IRE, estimates of τ^2 and $m'(t_0)$ (only for the IRE) were needed under short range dependence, while estimates of $m'(t_0), \sigma_n^2$ and η_1 were required for long range dependence. Note that η_1 is simply the common standard deviation of the errors. For weakly dependent errors τ^2 was estimated as

$$\hat{\tau}^2 = \hat{\gamma}_n(0) + \sum_{k \leq \sqrt{n}} \left(1 - \frac{k}{\sqrt{n}}\right) \hat{\gamma}_n(k) \tag{2.23}$$

where $\hat{\gamma}_n(k)$ is the auto-covariance of $\hat{e}_i := Y_i - \hat{m}_n(t_i)$ at lag k (see [70]). This is a consistent estimator under the presence of a monotone trend as argued in [70]. Estimation of $m'(t_0)$ is the most challenging part. Even for i.i.d. data, principled estimation in the monotone function setting is challenging – see Section 3.1 of [15] for a discussion – and the difficulties are only exacerbated under dependence. Kernel based estimation, as in [15] was used; thus,

$$\hat{m}'(t_0) = \frac{1}{h} \int K\left(\frac{t_0 - t}{h}\right) d\hat{m}_n(t)$$

where h is the bandwidth and K, a Gaussian kernel. The bandwidth was chosen by the method of cross-validation. For this we divide the dataset in two parts randomly. Each data points were assigned to one of the two sets with probability 0.5 using an auxiliary Bernoulli(1/2) random variable. Let D_i denote the set of indices for i-th subset, for $i = 1, 2$. Then for a given bandwidth h we calculate $\hat{m}'_{h,D_i}(t)$, the estimate of $m'(t)$ based on set D_i as

$$\hat{m}'_{h,D_i}(t) = \frac{1}{h} \int K\left(\frac{t - s}{h}\right) d\hat{m}_{D_i}(s)$$

where $\hat{m}_{D_i}(t)$ is the MLE of $m(t)$ based on set D_i. We then numerically integrate $\hat{m}'_{h,D_i}(t)$ to obtain $\hat{m}_{h,D_i}(t)$. Then we calculate

$$CV(h) = \sum_{i \in D_1} (y_i - \hat{m}_{h,D_2}(t_i))^2 + \sum_{i \in D_2} (y_i - \hat{m}_{h,D_1}(t_i))^2.$$

Note that in calculating CV we use the estimate based on one group to calculate the residual sum of square of the other group of the data set. We choose the value of h that minimizes $CV(h)$ as optimal bandwidth.

For short range dependent data oversmoothing with respect to order of the spacing of the jumps ($n^{-1/3}$) we also used theoretically optimal bandwidth $n^{-1/7}$ for derivative estimation of a monotone function (see [32]) to compare the results. For long range dependent (FGN) errors, σ_n^2 was replaced by $n^{2H}\sigma^2$ (using (2.10), because for $r = 1$, $\eta_r^2 = \sigma^2$ and $H = 1 - d/2$), and σ^2 estimated by the empirical variance of the Y_i's. ln the case of FARIMA, σ^2 was estimated using the approximate maximum likelihood method discussed in [35].

Remark II.19. In our simulations we used models with trivial slowly varying function components (recall (2.10)). For simplicity, we also used the actual value of the Hurst index H in the calculation of d_n and σ_n^2. The effect of plug-in estimates of H is discussed in Section 2.4.3.

Discussion of the simulation results: From Tables 2.2 to 2.3, we observe that for short range dependent errors, L_n and Ψ_n are performing much better, in terms of coverage, than the Wald-type confidence intervals based on the IRE. Note that the IRE based Confidence Interval (CI)s show systematic under-coverage, especially for m_2, as the derivative estimation procedure is highly unstable in this situation. The L_n and Ψ_n based intervals both exhibit coverage much closer to the nominal, though the Ψ_n based ones tend to over-cover, which can be attributed to the manner of their construction; see the comments following Proposition II.10. The average lengths of the CIs using Ψ_n are also substantially larger than their L_n based counterparts. As estimation of τ^2 is, often, not terribly difficult, we recommend using L_n whenever possible, i.e. unless we have very little information about the dependence structure of the errors, or if the dependence structure involves estimating too many parameters compared to sample size.

Under long range dependence, the Ψ_n based method outperforms both L_n and the IRE based methods, in terms of coverage, as is evident from Tables 2.4 to 2.8, with the latter intervals showing systematic under-coverage, especially at higher values of H and under FARIMA errors. While L_n was seen to be reliable in the short range case, its performance suffers under long range dependence because the derivative $m_2'(t_0)$ now needs to be estimated for its construction. Under m_2, the coverage of the IRE based CIs worsens significantly, owing to reasons similar to the short range case. The average lengths of the intervals using Ψ_n are consistently larger than those from the the other methods, showing that the lengths of asymptotically pivotal R_n–based CIs adapt nicely to the underlying variability in order to maintain close–to–nominal coverage. Additional simulations (not reported here) were run to assess the performance of *oracle* L_n-based CIs, constructed using the *true* values of the nuisance parameters. It was seen that such oracle CIs are substantially better: close–to–nominal coverage was restored and the average lengths were now less than the Ψ_n–based CIs. *Of course, the oracle CIs are not available in practice, but the experiments underscore the importance of (asymptotic) pivotality.*

Finally, in view of our discussion, we recommend using L_n or T_n under short range dependence unless the dependence structure is unknown or the covariance is difficult to estimate. For long range dependent data and short range dependent data where the covariance is difficult to estimate, we recommend using Ψ_n to construct confidence intervals.

Table 2.2: Confidence Intervals for ARMA(2,2) with AR coeffs 0.8, -0.5 and MA coeffs -0.2,0.3

n	$m_1(t)$								$m_2(t)$							
	L_n		Ψ_n		IRE1[1]		IRE2[2]		L_n		Ψ_n		IRE1		IRE2	
	Cov.	Len.	Cov.	Len.	Cov.	Len.	Cov.	Len.	Cov.	Len.	Cov.	Len.	Cov.	Len.	Cov.	Len.
100	89.1	0.443	90.8	0.537	86.5	0.504	87.3	0.511	88.3	0.407	89.9	0.523	81.8	0.485	80.1	0.478
200	89.7	0.352	92	0.392	85.8	0.413	84.9	0.433	87.6	0.312	89.8	0.437	78.6	0.399	79.2	0.401
500	89.3	0.261	91.4	0.307	84.9	0.312	85.7	0.332	88.9	0.230	90.4	0.312	79.9	0.298	73.4	0.225
1000	90.2	0.208	90.7	0.262	85.9	0.257	86.8	0.239	85.9	0.180	90.2	0.279	76.9	0.215	80.2	0.219
2000	91	0.163	90.9	0.205	86.8	0.209	85.5	0.211	90.1	0.141	90.8	0.211	80.1	0.199	75.5	0.186
5000	89.1	0.121	91.7	0.169	89.9	0.169	88.7	0.178	90.2	0.105	90.5	0.134	81.5	0.114	81.4	0.112

Table 2.3: Confidence Intervals for AR(2) with AR coeffs 0.95, 0.8

n	$m_1(t)$								$m_2(t)$							
	L_n		Ψ_n		IRE1[3]		IRE2[4]		L_n		Ψ_n		IRE1		IRE2	
	Cov.	Len.	Cov.	Len.	Cov.	Len.	Cov.	Len.	Cov.	Len.	Cov.	Len.	Cov.	Len.	Cov.	Len.
100	87.4	1.006	90.8	1.783	84.2	1.558	85.7	1.679	83.3	0.879	89.7	1.003	68.9	0.901	73.5	1.001
200	88.9	0.814	90.5	1.498	82.7	1.293	85.5	1.311	82.4	0.775	89.8	0.892	77.5	0.865	74.7	0.828
500	89.2	0.573	91.1	1.134	83.8	1.020	81.2	1.001	88.5	0.499	90.3	0.687	70.3	0.662	73.1	0.698
1000	89.1	0.458	91.6	0.827	85.9	0.809	83.6	0.798	89.9	0.387	91.6	0.568	72.4	0.525	71.9	0.517
2000	90.3	0.357	92.3	0.689	86.5	0.715	81.5	0.687	87.8	0.296	90.8	0.499	75.6	0.468	72.3	0.455
5000	89.3	0.283	91.4	0.514	83.5	0.592	82.9	0.527	89.5	0.198	91.3	0.401	76.9	0.379	77.8	0.400

Table 2.4: Confidence Intervals for fractional Gaussian noise with H=0.7

n	$m_1(t)$						$m_2(t)$					
	L_n		Ψ_n		IRE		L_n		Ψ_n		IRE	
	Cov.	Len.	Cov.	Len.	Cov.	Len.	Cov.	Len.	Cov.	Len.	Cov.	Len.
100	87.7	0.497	90.2	0.977	86.9	0.511	88.7	0.487	88.2	0.965	76.1	0.678
200	85.8	0.419	90.4	0.783	82.5	0.431	89.5	0.402	89.1	0.762	77.5	0.559
500	82.1	0.346	90.5	0.636	83.8	0.359	89.8	0.385	90.5	0.658	72.9	0.425
1000	85.7	0.298	91.9	0.529	84.4	0.300	89.7	0.311	89.3	0.527	78.3	0.369
2000	90	0.260	92.3	0.438	82.7	0.272	89.9	0.286	90.2	0.451	78.9	0.297
5000	84.9	0.199	91.8	0.332	87.6	0.200	88.4	0.178	90.7	0.348	80	0.235

Table 2.5: Confidence Intervals for fractional Gaussian noise with H=0.8

n	$m_1(t)$						$m_2(t)$					
	L_n		Ψ_n		IRE		L_n		Ψ_n		IRE	
	Cov.	Len.	Cov.	Len.	Cov.	Len.	Cov.	Len.	Cov.	Len.	Cov.	Len.
100	75.9	0.600	90.7	0.882	89.4	0.609	79.9	0.621	85.8	0.892	70	0.712
200	86.6	0.529	89.6	0.779	71.2	0.532	82.5	0.539	88.7	0.793	63.6	0.645
500	71.6	0.454	92.4	0.663	72.7	0.469	83.8	0.478	90.2	0.778	71.1	0.573
1000	72.2	0.401	91.1	0.594	81.2	0.416	74.5	0.425	89	0.601	62.8	0.481
2000	84.1	0.370	90.5	0.512	77.8	0.372	84.6	0.397	89.7	0.577	61.5	0.419
5000	84.9	0.306	90.9	0.448	82.7	0.320	88.1	0.324	90.4	0.463	73.4	0.395

Table 2.6: Confidence Intervals for fractional Gaussian noise with H=0.9

n	$m_1(t)$						$m_2(t)$					
	L_n		Ψ_n		IRE		L_n		Ψ_n		IRE	
	Cov.	Len.	Cov.	Len.	Cov.	Len.	Cov.	Len.	Cov.	Len.	Cov.	Len.
100	84.5	0.662	90.4	0.892	89.1	0.659	81.7	0.688	87.6	0.902	71.2	0.723
200	81.5	0.602	89.6	0.804	73.2	0.600	77.8	0.615	83.9	0.823	72.3	0.667
500	71.6	0.558	90.4	0.741	81.7	0.550	78.5	0.587	88.7	0.774	66.3	0.597
1000	72.2	0.503	92.1	0.691	75.8	0.500	73.9	0.514	89.5	0.712	71.8	0.561
2000	83.1	0.478	91.1	0.658	79.2	0.470	80.1	0.495	89.9	0.675	73.5	0.518
5000	81.7	0.434	91.8	0.597	76.7	0.432	81.2	0.462	89.7	0.613	74.9	0.499

Table 2.7: Confidence Intervals for fractional Gaussian noise with H=0.99

n	$m_1(t)$						$m_2(t)$					
	L_n		Ψ_n		IRE		L_n		Ψ_n		IRE	
	Cov.	Len.	Cov.	Len.	Cov.	Len.	Cov.	Len.	Cov.	Len.	Cov.	Len.
100	80.2	0.795	87.8	0.912	71.2	0.801	75.4	0.812	84.7	0.921	70.2	0.872
200	71.3	0.742	88.4	0.864	81.9	0.751	72.1	0.798	83.1	0.901	68.4	0.805
500	70.6	0.697	89.1	0.831	70.2	0.701	70.8	0.745	85.3	0.858	65.5	0.763
1000	81.5	0.654	88.9	0.814	70.8	0.660	71.1	0.687	86.2	0.832	63.9	0.705
2000	68.8	0.631	89.1	0.803	81.8	0.638	70.5	0.651	85.8	0.816	71.4	0.674
5000	73.7	0.593	89.9	0.785	71.2	0.601	72.7	0.612	88.9	0.773	72.6	0.655

Table 2.8: Confidence Intervals for FARIMA(2,1,1) with AR coeffs 0.5,-0.5; MA coeff 0.6 and d=0.2, i.e. H = 0.7

n	$m_1(t)$						$m_2(t)$					
	L_n		Ψ_n		IRE		L_n		Ψ_n		IRE	
	Cov.	Len.	Cov.	Len.	Cov.	Len.	Cov.	Len.	Cov.	Len.	Cov.	Len.
100	82.1	0.599	89.1	0.986	86.3	0.601	77.5	0.667	89.2	0.979	60.1	0.754
200	81.9	0.520	89.9	0.888	84.8	0.521	78.2	0.589	88.7	0.876	62.5	0.682
500	79.8	0.432	90.3	0.747	81.2	0.439	71.6	0.495	91.1	0.735	74.8	0.595
1000	80.4	0.362	91.2	0.632	79.8	0.370	72.8	0.401	90.1	0.609	71.1	0.502
2000	77.6	0.311	90.9	0.546	89.2	0.318	74.1	0.362	90.4	0.515	72.9	0.468
5000	81.7	0.262	91.5	0.446	81.2	0.273	77.5	0.298	90.8	0.407	72.6	0.375

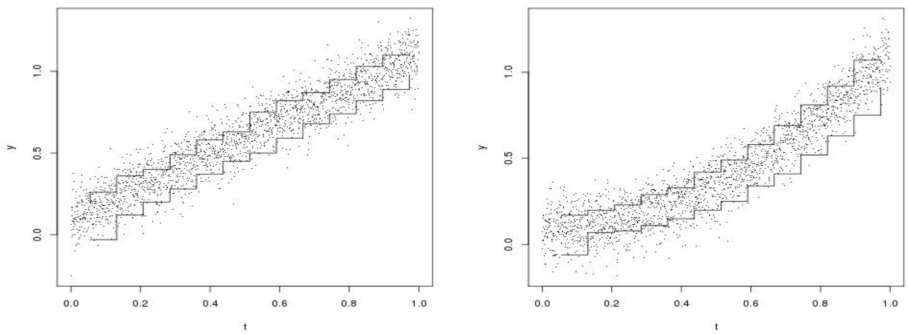
Table 2.9: Coverage of 90% Uniform Confidence Intervals

Function	Errors	n	Coverage
$m(t) = t$	AR(2) coeff (0.7, -0.6)	2000	92.1%
$m(t) = t$	AR(2) coeff (0.7, -0.6)	5000	91.9%
$m(t) = t$	ARMA(1,1) coeff (0.8, 0.4)	2000	93.0%
$m(t) = t$	ARMA(1,1) coeff (0.8, 0.4)	5000	92.6%
$m(t) = t^2$	AR(2) coeff (0.7, -0.6)	2000	93.5%
$m(t) = t^2$	AR(2) coeff (0.7, -0.6)	5000	91.7%
$m(t) = t^2$	ARMA(1,1) coeff (0.8, 0.4)	2000	92.3%
$m(t) = t^2$	ARMA(1,1) coeff (0.8, 0.4)	5000	92.5%

2.4.2 Performance of Uniform Confidence Intervals Under SRD

To study the performance of confidence band proposed in Section 2.3.3 we used two choices for the trend function namely $m_1(t) = t$ and $m_2(t) = t^2$. We used two different dependence structures for errors. The first one is a AR(2) model with AR coefficients 0.7 and -0.6, for the second dependence structure we used a ARMA(1,1) model with AR coefficient 0.8 and MA coefficient 0.4. The marginal variance for both the structures were taken to be 0.2. Table 2.9 presents simulated coverage of 90% confidence bands calculated from 1000 iterations for sample sizes $n = 2000$ and $n = 5000$. We chose 13 and 17 equidistant points starting from the 10-th data-point respectively for sample sizes 2000 and 5000 to construct the confidence band. As we can see from the table the bands constructed in this method gives reasonable coverage.

Figure 2.3 shows confidence band for a simulated dataset from the isotonic regression model with trend functions m_1 and m_2 and the errors come from a AR(2) model with AR coefficients 0.7 and -0.6 and variance 0.2. Both the data-set have $n = 2000$ data-points and we used 13 equally spaced points to construct the confidence bands.



(a) $m_1(t) = t$ (b) $m_2(t) = t^2$

Figure 2.3: 90% confidence band

2.4.3 Estimation of H

In order to construct confidence intervals, using either Ψ_n or L_n, we need to estimate the Hurst index H. As discussed in Remark II.15 any consistent plug-in estimate of H should give right coverage. For our data analysis, we used a wavelet-based method e.g. as in [5] and [61]; for other methods, see [29] and [27]. This methods are robust to presence of a monotone trend because the wavelet functions kills polynomial trend of certain order in the data. As all the reasonably well-behaved functions can be well-approximated by polynomials, by increasing order of wavelets appropriately we can kill the effect of trend functions in our data. Table 2.10 presents coverage and average length (based on 1000 iterations) for 90% confidence interval using Ψ_n under different LRD models and different sample size. We used wavelets of order 7 to estimate H. It is evident from the table the plug-in estimate preserves right coverage without making the interval too large.

Also, the dependence of Ψ_n–based inference on H is minimal in the sense that H is only required to determine the cut-off value for inversion and does not enter into the computation of Ψ_n itself (unlike what happens with the IRE or L_n). Hence, if there were a general nesting of quantiles of Ψ with respect to H, one could have built conservative confidence intervals at any given level without estimating H! Such type

Table 2.10: Confidence Intervals with estimated H

n	fGn, H= 0.6		fGn, H=0.8		fGn, H = 0.9		fARIMA, H=0.7	
	Cov.	Len.	Cov.	Len.	Cov.	Len.	Cov.	Len.
100	90.1	0.643	90.8	0.737	90.5	0.904	89.3	0.607
200	89.7	0.552	92	0.592	89.8	0.813	88.6	0.512
500	91.3	0.461	91.4	0.407	88.9	0.712	89.9	0.430
1000	90.2	0.408	90.7	0.362	89.9	0.657	90.9	0.380
2000	91	0.363	90.9	0.305	90.1	0.509	90.1	0.341
5000	89.1	0.221	91.7	0.269	89.9	0.369	90.2	0.205

of robustness to long-range dependence is too much to hope for. Nevertheless, while the nesting property is absent in general, at both 90% and 95% levels, our estimated quantiles increase as a function of H for $0.5 \leq H \leq 0.95$, as a quick inspection of Table I (and more extensive simulations not reported here) reveals. This empirical observation can, therefore, be used to construct conservative Ψ_n-based confidence intervals at these two levels, by using the quantiles corresponding to $H = 0.95$. Values of H greater than 0.95 indicate extreme levels of long-range dependence, which should be dealt with care, but are rarely encountered in practice. Note that such conservative CI's are *completely agnostic* as to whether the underlying dependence is short– or long–range, exemplifying the robustness of our method. The bottom–line here is that if little is qualitatively known about the extent of dependence, it is better to go with the conservative intervals above, whereas if reasonably reliable information about the error structure is available, the best distributional approximation to the Ψ–statistic (generally at the expense of estimating H) should be used.

2.4.4 Analysis of Global Temperature Anomaly and Internet Usage Data

Here we apply our methodology to a short-range dependent and a long-range dependent dataset. In view of the discussion at the end of Section 5, in the former case we use the L_n statistic and the the ratio based statistic Ψ_n in the more challenging long-range dependent case.

Global Temperature Anomaly Data (Short Range Dependence) We consider the global warming data used in [71], which consists of global annual temperature anomalies, measured in degrees celsius from 1850 to 2009. These anomalies are, simply, temperature deviations measured with respect to the base period 1961-1990. The autocorrelation plot of these data suggests that the dependence can be well accounted for using an autoregressive model of order two (AR(2)), see [71]. The short range dependence condition (2.6) applies to AR(2) time series. Figure 2.4 represents the data along with its isotonic regression estimates and point-wise confidence intervals obtained by using L_n. The estimate of the asymptotic standard deviation τ was taken to be 0.1248 from [70]. Note that the point-wise confidence intervals form a rather smooth band, which mimics, in shape, the isotonic regression curve. Note that our analysis does not use win any way the AR(2) model suggested in [71]. Apart from the value of τ, which is often easy to estimate, our methodology is completely agnostic to the nature of short-range dependence and regularity of the trend. It is remarkable that this near-universality does not come at the expense of overly wide intervals. As it can be seen from Figure 2.4, our confidence intervals clearly indicate sufficient evidence for systematic growth of the global temperature anomalies. Such results are to be expected, given the compelling evidence from numerical climate model simulations of the Intergovernmental Panel of Climate Change IPCC (2007)[3]. Nevertheless, our statistical methodology is based exclusively on data rather than numerical prediction. It can provide an alternative, statistically justified approach in complex situations, where the nature of dependence and the underlying structure of the trend are largely unknown.

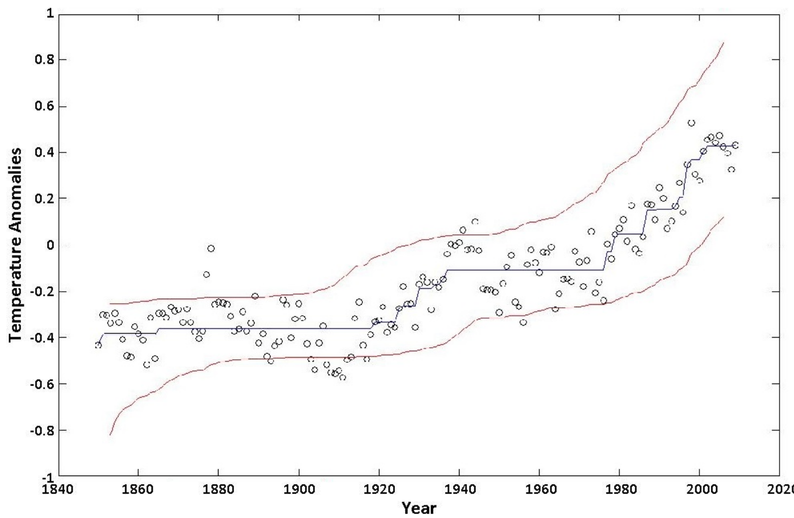


Figure 2.4: 90% point-wise confidence intervals for global temperature anomaly data

Internet Traffic Data (Long Range Dependence) This example involves computer network traffic data obtained from the Internet2 network [1]. The data consists of number of bytes per 100 millisecond time-intervals over a fast backbone link measured on 17th March, 2009. Such traffic traces exhibit typical diurnal patterns with clearly defined periods of monotone non-decreasing or non-increasing trends throughout the day. This is associated with usual growth/decay of the number of active users in the beginning/middle of the day. Further, it is very well known and documented that Internet traffic traces exhibit long-range dependence (see e.g. [66] and [61]). Such data provide an ideal test-bed for the performance of our confidence intervals based on the ratio statistic. To be able to provision network capacity as well as detect anomalous network activity, it is important to have accurate estimates of confidence intervals that are robust to the presence of long–range dependence and account for natural traffic trends, without imposing stringent parametric/smoothness assumptions. This is particularly important in the network traffic context, where unusual changes in the regularity of the trend may occur and methods that involve estimation of derivatives require great care to implement and, in fact, can lead to non-robust interval estimates.

We focused on the time period 11 : 06 to 14 : 36 GMT. This period corresponds to 6 : 06 AM to 11 : 36 AM in the local EST time, where there is a typical monotone non-decreasing diurnal trend due to systematic increase of the number of active users in the beginning of the day. The Hurst parameter was estimated to be $\hat{H} = 0.9491$ using wavelet methods [61].

Figure 2.5 shows 90% confidence intervals at 100 time points based on the Ψ_n statistic with $H = 0.95$. Observe that the confidence intervals track closely the monotone trend and can be used to detect the onset of anomalous activity in the network. The p-values of the associated test can be further used to track significance of changes in the underlying traffic trends. This is one natural further application of our methodology.

Note that in contrast to the intervals based on the L_n and T_n statistics, the intervals based on Ψ_n are not smooth over time. Since we do not perform any type of smoothing, nor aim to produce a confidence band, this feature is not alarming. In fact, it shows that our statistic is rather adaptive and sensitive to changes in the variability! That is, as seen in extensive simulations (not reported) and also from Figure 2.5, the confidence intervals automatically expand when the variability is large relative to the slope of the trend and rapidly shrink otherwise. This adaptivity property may be attributed to the fact our interval estimates provide accurate coverage and are at the same time nearly dependence–universal. That is, the same statistic Ψ_n is used under both short- and long–range dependence without having to estimate nuisance parameters. The only input necessary to calibrate the critical values of the test is the Hurst long–range dependence parameter, which may be estimated or constrained to obtain conservative interval estimates. This unusual adaptivity/irregularity behavior of the proposed confidence intervals is yet to be fully understood. It will be the subject of a future work.

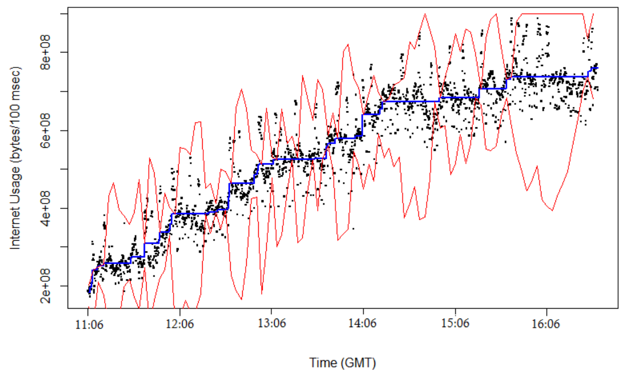


Figure 2.5: 90% point–wise confidence intervals for internet traffic data

2.5 Technical Complements

In this section we provide proofs of the main theorems discussed in the paper.

2.5.1 Some Auxiliary Lemmas

In this section we state some Lemmas useful for our purpose.

Lemma II.20. *Let f_n, f be convex functions, defined on an open interval $I \subset \mathbb{R}$. If $\lim_{n\to\infty} \sup_{x\in I} |f_n(x) - f(x)| = 0$, then, for all $x \in I$,*

$$\partial_\ell f(x) \le \liminf_{n\to\infty} \partial_\ell f_n(x) \le \limsup_{n\to\infty} \partial_r f_n(x) \le \partial_r f(x),$$

where $\partial_\ell f$ and $\partial_r f$ denote the left and right derivatives of f. If, moreover, the function f is differentiable at a point x with derivative $f'(x)$, then both $\partial_\ell f_n(x)$ and $\partial_r f_n(x)$ converge to $f'(x)$, as $n \to \infty$.

For the proof, see e.g. p. 330 in [57].

Lemma II.21. *Let f_n, f be convex functions, defined on an open interval $I \subset \mathbb{R}$. If $f_n \to f$, as $n \to \infty$ uniformly on all compact subsets of I, then $\partial_\ell f_n \to \partial_\ell f$ in L^2_{loc}.*

Proof. Since f is convex it is a.e. differentiable and by Lemma II.20, we have $\partial_\ell f_n(z) \to \partial_\ell f(z) \equiv \partial_r f(z)$, as $n \to \infty$, for almost all $z \in I$.

Now, for any given $[c,d] \subset I$, one can find $a \le c < d \le b$, such that $[a,b] \subset I$ and f is differentiable at both a and b. Thus, by Lemma II.20 $\partial_\ell f_n(x) \to \partial_\ell f(x)$, $x \in \{a,b\}$ as $n \to \infty$. Since $\partial_\ell f_n : [a,b] \to \mathbb{R}$, is non–decreasing, we have $\partial_\ell f_n(a) \le \partial_\ell f_n(z) \le \partial_\ell f_n(b)$, $z \in [a,b]$, and by the fact that the last lower and upper bounds converge, we have

$$\sup_{z\in[a,b]} |\partial_\ell f_n(z)| \le 1 + \max\{|\partial_\ell f(a)|, |\partial_\ell f(b)|\} < \infty,$$

for all sufficiently large n. Therefore, by the dominated convergence theorem

$$\int_{[a,b]} (\partial_\ell f_n(z) - \partial_\ell f(z))^2 \, dz \to 0, \quad \text{as } n \to \infty,$$

which completes the proof. $\square$

Lemma II.22. *Let $M(I)$ denote the set of monotone non-decreasing and left–continuous functions defined on the interval I equipped with the L^2_{loc} convergence. Define the concatenation map $C_h : M(-\infty,0) \times M(0,\infty) \to M(-\infty,\infty)$, where*

$$C_h(f,g)(x) := \begin{cases} f(x) \wedge h & , \text{ if } x \in (-\infty,0) \\ \lim_{u\uparrow 0} f(u) \wedge h & , \text{ if } x = 0 \\ g(x) \vee h & , \text{ if } x \in (0,\infty) \end{cases} \tag{2.24}$$

Then, $C_h : (M(-\infty,0) \times M(0,\infty), L^2_{loc} \times L^2_{loc}) \to (M(-\infty,\infty), L^2_{loc})$ is continuous.

Proof. Let $f_n \to f$ and $g_n \to g$ in $(M(-\infty,0), L^2_{loc})$ and $(M(0,\infty), L^2_{loc})$ respectively and let $a < 0 < b$. It is enough to show that $\int_{[a,0]}(f_n(x)\wedge\theta - f(x)\wedge\theta)^2 dx \to 0$ and that $\int_{[0,b]}(g_n(x)\vee\theta - g(x)\vee\theta)^2 dx \to 0,\ n \to \infty$. We only focus on the first integral since the second one can be treated similarly.

Observe that $(f_n(x) \wedge \theta - f(x) \wedge \theta)^2 \leq (f_n(x) - f(x))^2$. Therefore by the fact that $f_n \to f$ in $L^2_{loc}(-\infty, 0)$, it follows that it is enough to show that

$$\lim_{\epsilon \to 0} \lim_{n \to \infty} \int\limits_{[-\epsilon,0]} (f_n(x) \wedge \theta - f(x) \wedge \theta)^2 dx = 0. \tag{2.25}$$

Observe that by the monotonicity of f_n, we have

$$|f_n(x) \wedge \theta| \leq \max\{|\theta|, |f_n(x-1)|\}, \quad \text{for all } x \in [-1, 0].$$

Therefore, using the inequality $(u - v)^2 \leq 2u^2 + 2v^2$, we get

$$\int\limits_{[-\epsilon,0]} (f_n(x) \wedge \theta - f(x) \wedge \theta)^2 dx \leq 2 \int\limits_{[-\epsilon-1,-1]} (f_n^2(x) + f^2(x))dx + 4|\theta|^2\epsilon \tag{2.26}$$

Since $f_n \to f$ in $L^2_{loc}(-\infty, 0)$, we get $\int_{[-\epsilon-1,-1]} f_n^2(x)dx \to \int_{[-\epsilon-1,-1]} f^2(x)dx$, and the latter vanishes as $\epsilon \downarrow 0$. Therefore, the right-hand side of (2.26) vanishes as $n \to \infty$ and as $\epsilon \downarrow 0$, which implies (2.25). □

2.5.2 Proof of Results of Section 2.2

In this section we present the proofs of the Results discussed in Section 2.2. First note that the isotonic regression estimator can be represented in terms of the partial sum process U_n defined as in (2.14).

Proposition II.23. *We have*

$$\begin{aligned} \hat{m}_n(t) =& \mathcal{L} \circ \mathcal{T}_{(0,1]}(U_n)(t) \\ \hat{m}_n^0(t) =& (\mathcal{L} \circ \mathcal{T}_{(0,t_l]}(U_n)(t) \wedge \theta_0)\mathbf{1}_{(0,t_l]}(t) \\ &+ \theta_0 \mathbf{1}_{(t_l,t_0]}(t) + (\mathcal{L} \circ \mathcal{T}_{(t_l,1]}(U_n)(t) \vee \theta_0)\mathbf{1}_{(t_0,1]}(t). \end{aligned} \tag{2.27}$$

This representation follows from Chapter 2 of [57] in the case of $\hat{m}_n$, and from Section 2 of [14], in the case of $\hat{m}_n^0$.

2.5.2.1 The Process $\mathbb{V}_n$

Proof of Theorem 1: It is enough to show that for all $c > 0$, we have $\mathbb{V}_n|_{[-c,c]} \Rightarrow \mathbb{G}|_{[-c,c]}$ in $C([-c, c])$ equipped with the uniform norm. Fix $c > 0$ and note that since $(a_n, b_n] \uparrow \mathbb{R}$, as $n \to \infty$, without loss of generality we may assume that $[-c, c] \subset (a_n, b_n]$. Write $\mathbb{V}_n(z) = \mathbb{W}_n(z) + \Lambda_n(z)$, $z \in [-c, c]$, where

$$\mathbb{W}_n(z) := d_n^{-2} n^{-1} \left(\upsilon_n(t_0 + z d_n) - \upsilon_n(t_0) \right),$$

with $\upsilon_n(t) = \sum_{i=1}^{\lfloor nt \rfloor} \epsilon_i + (nt - \lfloor nt \rfloor)\epsilon_{\lfloor nt \rfloor + 1}$. Then,

$$\Lambda_n(z) = \Upsilon_n(z) + \mathbf{R}_n(z), \tag{2.28}$$

where $\Upsilon_n(z) = d_n^{-2}\big[M(t_0 + d_n z) - M(t_0) - m(t_0) d_n z\big]$, and $\mathbf{R}_n(z) = d_n^{-2}\big[(M_n - M)(t_0 + d_n z) - (M_n - M)(t_0)\big]$. Hence we have

$$\sup_{z \in (a_n, b_n]} |\mathbf{R}_n(z)| \le 2 d_n^{-2} \sup_{0 \le t \le 1} |M_n(t) - M(t)| = O(d_n^{-2} n^{-1}).$$

The latter vanishes as $n \to \infty$ because,

$$d_n^{-2} n^{-1} = \begin{cases} n^{-\frac{1}{3}} & \text{under weak dependence} \\ n^{-\frac{2-d}{2+d}} & \text{under strong dependence.} \end{cases}$$

Thus, the remainder term $\mathbf{R}_n$ in (2.28) can be neglected and a Taylor series expansion of the deterministic function M at t_0 in the term Υ_n yields,

$$\Lambda_n(z) \to \frac{1}{2} m'(t_0) z^2 \tag{2.29}$$

as $n \to \infty$ uniformly on $[-c, c]$.

Now, we deal with the term $\mathbb{W}_n$. By the stationarity of $\{\epsilon_i\}_{i\in\mathbb{Z}}$, we have

$$\{v_n(t_0 + zd_n) - v_n(t_0)\}_{z\in\mathbb{R}} \stackrel{d}{=} \{v_n(zd_n)\}_{z\in\mathbb{R}} = \{v_{\hat{n}}(z)\}_{z\in\mathbb{R}},$$

where $\widehat{n} := nd_n$. Note that $\widehat{n}$ may not be an integer. The definition of v_n makes sense even if n is not an integer. For rest of the proof we will use other sequences indexed by $\widehat{n}$ and we define $\sigma_{\widehat{n}} = \sigma_{\lfloor \widehat{n} \rfloor}$ and $w_{\widehat{n}}(t) = v_{\widehat{n}}(t)/\sigma_{\widehat{n}}$ (recall (2.5)). With this convention our following arguments remain valid at least asymptotically if $\widehat{n} \to \infty$ as $n \to \infty$.

Next we can write,

$$\{\mathbb{W}_n(z)\}_{z\in\mathbb{R}} \stackrel{d}{=} \{d_n^{-2} n^{-1} v_{\widehat{n}}(z)\}_{z\in\mathbb{R}} = \{d_n^{-2} n^{-1} \sigma_{\widehat{n}} w_{\widehat{n}}(z)\}_{z\in\mathbb{R}}, \tag{2.30}$$

where w_n is as in (2.5). Observe that under both short- and long-range dependence assumptions, we have $w_{\widehat{n}}|_{[-c,c]} \Rightarrow \mathbb{W}|_{[-c,c]}$, as $\widehat{n} \to \infty$ in the Skorokhod J_1-topology, where $\mathbb{W}$ denotes either the two–sided Brownian motion or the process $B_{r,H}$ (recall Section 2.1.1). Since the limit processes (in both cases) have versions with continuous paths, the J_1-convergence implies also convergence in the uniform topology. To complete the proof, it remains to show that $\widehat{n} \to \infty$ in both cases with the appropriate choice of d_n and constants.

(i) Under short–range dependence, with $d_n = n^{-\frac{1}{3}}$, we have $\hat{n} \equiv nd_n \to \infty$ and, by (2.8), $d_n^{-2} n^{-1} \sigma_{\hat{n}} \to \tau$ as $n \to \infty$, which yields $a = \tau$.

(ii) Under long–range dependence, we want d_n such that $d_n^{-2} n^{-1} \sigma_{\hat{n}} \to |\eta_1|$ as $n \to \infty$

where η_1 is the Hermite rank. By relation (2.10) this is equivalent to

$$\begin{aligned} |\eta_1| &= d_n^{-2} n^{-1} |\eta_1| (nd_n)^{1-\frac{d}{2}} l_1(nd_n)^{\frac{1}{2}} \\ &\Longleftrightarrow d_n^{1+\frac{d}{2}} = n^{-\frac{d}{2}} l_1(nd_n)^{\frac{1}{2}} \\ &\Longleftrightarrow d_n = n^{-\frac{d}{2+d}} l_2(n), \end{aligned}$$

where l_2 is another slowly varying function at infinity. This choice of d_n ensures that $\hat{n} \equiv nd_n \to \infty$ as $n \to \infty$ and $a = |\eta_1|$, by (2.9). This completes the proof.

2.5.2.2 The Processes X_n and Y_n

The processes $\mathbb{V}_n(z)$, $X_n(z)$ and $Y_n(z)$ are only defined for $z \in (a_n, b_n]$. Ultimately, we have that $(a_n, b_n] \uparrow \mathbb{R}$. For technical convenience, however, we shall extend the definitions of these processes to the entire real line. This is best done by extending $\mathbb{V}_n$ in such a way that Relations (2.16) continue to hold for all $z \in (-\infty, \infty)$. To this end, let

$$\mathbb{V}_n(z) := \begin{cases} \mathbb{V}_n(z) & , \ z \in (a_n, b_n] \\ \lambda_\ell(z - a_n) + \mathbb{V}_n(a_n+) & , \ z \in (-\infty, a_n] \\ \lambda_r(z - b_n) + \mathbb{V}_n(b_n) & , \ z \in (b_n, \infty), \end{cases} \tag{2.31}$$

where $\lambda_\ell = \lim_{z \downarrow a_n} \mathcal{L} \circ \mathcal{T}_{(a_n,b_n]}(\mathbb{V}_n)(z)$ and $\lambda_r = \mathcal{L} \circ \mathcal{T}_{(a_n,b_n]}(\mathbb{V}_n)(b_n)$. That is, λ_ℓ and λ_r may be viewed as the smallest and largest left slopes of the GCM of $\mathbb{V}_n$ over the interval $(a_n, b_n]$.

The so-defined extension of $\mathbb{V}_n$ has the following important property:

$$\mathcal{T}_{(-\infty,\infty)}(\mathbb{V}_n)(z) = \mathcal{T}_{(a_n,b_n]}(\mathbb{V}_n)(z), \quad \text{for all } z \in (a_n, b_n],$$

and in fact $\mathcal{T}_{(-\infty,c]}(\mathbb{V}_n)(z) = \mathcal{T}_{(a_n,c]}(\mathbb{V}_n)(z)$, $z \in (a_n, c]$ and

$\mathcal{T}_{(c,\infty)}(\mathbb{V}_n)(z) = \mathcal{T}_{(c,b_n]}(\mathbb{V}_n)(z)$, $z \in (c, b_n]$. This shows that Relations (2.16) continue to hold and in fact a_n and b_n therein can be replaced by $-\infty$ and ∞, respectively. Therefore, from now on, we shall consider the processes $X_n = \{X_n(z)\}_{z\in\mathbb{R}}$ and $Y_n = \{Y_n(z)\}_{z\in\mathbb{R}}$, defined as follows

$$\begin{aligned} X_n(z) =& \mathcal{L} \circ \mathcal{T}_{(-\infty,\infty)}(\mathbb{V}_n)(z) \\ Y_n(z) =& \big(\mathcal{L} \circ \mathcal{T}_{(-\infty,l_n]}(\mathbb{V}_n)(z) \wedge 0\big)\,\mathbf{1}_{(-\infty,l_n]}(z) + 0 \times \mathbf{1}_{(l_n,0]}(z) \\ &+ \big(\mathcal{L} \circ \mathcal{T}_{(l_n,\infty]}(\mathbb{V}_n)(z) \vee 0\big)\,\mathbf{1}_{(0,\infty)}(z). \end{aligned} \tag{2.32}$$

The paths of the processes X_n and Y_n are left–continuous non–decreasing step–functions, which are constant on $(-\infty, a_n]$ and (b_n, ∞). As argued above, over $(a_n, b_n]$ they are given by (2.12).

For the next step we need the following result from [9]:

Theorem 6 (Adapted from [9])**.** *Consider a sequence of stochastic processes* $\{V_n(z)\}_{z\in\mathbb{R}}$, $n = 1, 2, \cdots$ *with paths in* $C(\mathbb{R})$*. Assume that*

(1) (Compact boundedness) For every compact set K *and* $\delta > 0$*, there is a finite* $M = M(K, \delta)$ *such that*

$$\limsup_{n\to\infty} \mathbb{P}\left(\sup_{z\in K} |V_n(z)| > M\right) < \delta \tag{2.33}$$

(2) (Lower bound) For every $\delta > 0$*, there are finite* $0 < \tau = \tau(\delta)$ *and* $0 < \kappa = \kappa(\delta)$ *such that*

$$\liminf_{n\to\infty} \mathbb{P}\left(\inf_{|z|\geq\tau} (V_n(z) - \kappa|z|) > 0\right) > 1 - \delta \tag{2.34}$$

(3) (Small downdippings) Given $\epsilon, \delta, \tilde{\tau} > 0$,

$$\limsup_{n\to\infty} \mathbb{P}\left(\inf_{\tilde{\tau}\le z\le c} \frac{V_n(z)}{z} - \inf_{\tilde{\tau}\le z} \frac{V_n(z)}{z} > \epsilon\right) < \delta \tag{2.35}$$

$$\limsup_{n\to\infty} \mathbb{P}\left(\inf_{z\le -\tilde{\tau}} \frac{V_n(z)}{z} - \inf_{-c\le z\le -\tilde{\tau}} \frac{V_n(z)}{z} < -\epsilon\right) < \delta \tag{2.36}$$

for all large enough $c > 0$

Then for any finite interval I in $\mathbb{R}$ and $\epsilon > 0$,

$$\lim_{c\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(\sup_I |\mathcal{T}_{[-c,c]}(V_n)(.) - \mathcal{T}(V_n)(.)| > \epsilon\right) = 0. \tag{2.37}$$

This also holds true if we replace $\mathcal{T}$ by $\mathcal{T}_O$ for any interval $O \subseteq \mathbb{R}$ or $\mathcal{T}_{O_n}$ where O_n is a sequence of intervals such that $O_n \uparrow O$, with $O \subseteq \mathbb{R}$. In these cases $\mathcal{T}_{[-c,c]}$ in (2.37) *is replaced by $\mathcal{T}_{K_c}$, for some sequence of compact intervals K_c such that $K_c \uparrow O$ as $c \to \infty$.*

Proposition II.24. *The processes $V_n := \mathbb{V}_n$ in* (2.15) *satisfy the conditions of Theorem 6.*

This result will be used to “localize" certain continuous mapping arguments to a compact interval.

Proof of Lemma II.3: Observe that if for some $M > 0$, $X_n(M) = Y_n(M)$, then $X_n(z) = Y_n(z)$, for all $z \ge M$. This is because of Lemma B.2 from Appendix B and the fact that X_n are step functions where the jump points are precisely the points where the GCM of $\mathbb{V}_n$ touches the curve. Similarly, $X_n(-M) = Y_n(-M)$ implies $X_n(z) = Y_n(z)$, for $z \le -M$. Therefore, it is enough to show that $\limsup_{n\to\infty} \mathbb{P}(X_n(M) \ne Y_n(M)) \to 0$, as $M \to \infty$. The case when $M \to -\infty$ can be treated similarly.

We claim that if $X_n(M) \ne Y_n(M)$, then either $\hat{m}_n^0(t_0 + Md_n) = \theta_0$ or $\hat{m}_n(t_0 + Md_n) = \hat{m}_n(t_0)$ (see also page 159, [11]). The proof of this claim will be given at the end of this proof. This, since

$\{X_n(0) = X_n(M)\} = \{\hat{m}_n(t_0) = \hat{m}_n(t_0 + Md_n)\}$ and

$\{Y_n(M) = 0\} = \{\hat{m}_n^0(t_0 + Md_n) = \theta_0\}$, implies

$$\{X_n(M) \neq Y_n(M)\} \subset \{Y_n(M) = 0\} \cup \{X_n(0) = X_n(M)\}.$$

Now as Y_n is a non-decreasing step function and $Y_n(0) = 0$,

$$\begin{aligned}
\limsup_{n\to\infty} \mathbb{P}(Y_n(M) = 0) &\leq \limsup_{n\to\infty} \mathbb{P}\left(\int_0^M Y_n^2(z)dz = 0\right) \\
&\leq \mathbb{P}\left(\int_0^M (\mathcal{S}_{a,b}^0(z))^2 dz = 0\right), \qquad (2.38)
\end{aligned}$$

where the last inequality follows from (2.19) and the Portmanteau Theorem (see e.g. page 16 of [17]). Similarly, since X_n is non–decreasing

$$\begin{aligned}
\limsup_{n\to\infty} \mathbb{P}(X_n(0) = X_n(M)) &\leq \limsup_{n\to\infty} \mathbb{P}\Big(\int_0^1 X_n^2(z)dz - \int_{M-1}^M X_n^2(z)dz = 0\Big) \\
&\leq \mathbb{P}\Big(\int_0^1 \mathcal{S}_{a,b}(z)^2(z)dz - \int_{M-1}^M \mathcal{S}_{a,b}(z)^2 dz = 0\Big), \qquad (2.39)
\end{aligned}$$

by (2.19) and the Portmanteau Theorem. Now, observe that by Lemma B.8 from Appendix B, we have that the right–hand sides of (2.38) and (2.39) vanish, as $M \to \infty$. This implies the desired inequality.

Now to prove the claim that $X_n(M) \neq Y_n(M)$ implies either $\hat{m}_n^0(t_0 + Md_n) = \theta_0$ or $\hat{m}_n(t_0 + Md_n) = \hat{m}_n(t_0)$ recall (2.12). Suppose that $\hat{m}_n^0(t_0 + Md_n) \neq \theta_0$ and $\hat{m}_n(t_0 + Md_n) \neq \hat{m}_n(t_0)$. Note that $\hat{m}_n(t) = \mathcal{L} \circ T_{[0,1]}(U_n)(t)$ is a step–function which changes only at points t, where the GCM $T_{[0,1]}(U_n)(t)$ of U_n equals the function value $U_n(t),\ t \in (0, 1]$. Therefore, the fact that $\hat{m}_n(t_0) \neq \hat{m}_n(t_0 + Md_n)$, implies that for some $t^* \in (t_0, t_0 + Md_n]$, we have $T_{(0,1]}(U_n)(t^*) = U_n(t^*)$. Note, however, that the constrained GCM $T_{(t_l,1]}(U_n)(t),\ t \in (t_l, 1]$ lies between the unconstrained one and the function, i.e.

$$T_{(0,1]}(U_n)(t) \le T_{(t_l,1]}(U_n)(t) \le U_n(t), \quad t \in (t_l, 1].$$

This implies that $T_{(t_l,1]}(U_n)(t^*) = T_{(0,1]}(U_n)(t^*) = U_n(t^*)$ and as in the proof of Lemma B.2 (2) from Appendix B, the two GCMs coincide over the interval $[t^*, 1]$, and so do their slopes

$$\mathcal{L} \circ T_{(t_l,1]}(U_n)(t) \equiv \mathcal{L} \circ T_{(0,1]}(U_n)(t),\ t \in (t^*, 1]. \tag{2.40}$$

On there other hand, since $\hat{m}_n^0(t_0 + Md_n) = \max\{\theta_0, \mathcal{L} \circ T_{(t_l,1]}(U_n)(t_0 + Md_n)\} \neq \theta_0$, we have that $\hat{m}_n^0(t_0 + Md_n) = \mathcal{L} \circ T_{(t_l,1]}(U_n)(t_0 + Md_n)$, which by (2.40) implies that $\hat{m}_n^0(t_0 + Md_n) = \hat{m}_n(t_0 + Md_n)$, since $t^* < t_0 + Md_n$. This completes our proof. □

Remark II.25. Since we do not have convergence of finite dimensional distributions of $\{X_n(z), Y_n(z)\}_{z\in\mathbb{R}}$ here we cannot use the techniques used to prove the same version of this Lemma in the iid case (see Page 159 of [11]).

Proof of Theorem 2: We will show that

$$\begin{aligned} \mathbb{GCM}_n \;\; &:= \;\; \Big(\mathcal{T}(\mathbb{V}_n), \mathcal{T}_{(-\infty,l_n]}(\mathbb{V}_n)|^{(-\infty,0)}, \mathcal{T}_{(l_n,\infty)}(\mathbb{V}_n)|_{(0,\infty)}\Big) \\ &\Longrightarrow \Big(\mathcal{T}(\mathbb{G}), \mathcal{T}_{(-\infty,0)}(\mathbb{G}), \mathcal{T}_{(0,\infty)}(\mathbb{G})\Big), \end{aligned} \tag{2.41}$$

where $\mathcal{T}_{(-\infty,l_n]}(\mathbb{V}_n)|^{(-\infty,0)}$ denotes the extension of the process $\mathcal{T}_{(-\infty,l_n]}(\mathbb{V}_n)$ to $(-\infty, 0)$. This extension is defined as in (2.31), i.e., we extend the convex function $\mathcal{T}_{(-\infty,l_n]}(\mathbb{V}_n)$ linearly in $[l_n, 0)$ to maintain convexity. The weak convergence (2.41) is in the space $\mathcal{C}(\mathbb{R}) \times \mathcal{C}(-\infty, 0) \times \mathcal{C}(0, \infty)$ equipped with the product topology of local uniform convergence on compacta.

If (2.41) holds, then the result follows from a continuous mapping argument. Indeed, consider the map

$$J : \mathcal{C}(\mathbb{R}) \times \mathcal{C}(-\infty, 0) \times \mathcal{C}(0, \infty) \to M(\mathbb{R}) \times M(-\infty, 0) \times M(0, \infty),$$

defined as $J(f, f_-, f_+) := (\mathcal{L}f, \mathcal{L}f_-, \mathcal{L}f_+)$, where $M(I)$ denotes the space of monotone real valued functions on an interval I equipped with the topology of L^2 convergence on compact sets. Observe that with the concatenation map
$C_0 : M(-\infty, 0) \times M(0, \infty) \to M(\mathbb{R})$ defined in (2.24) with $h = 0$.
we have

$$(X_n, Y_n) = ((\mathrm{id}, C_0) \circ J)(\mathbb{GCM}_n),$$

where $\mathrm{id} : M(\mathbb{R}) \to M(\mathbb{R})$ denotes the identity. By Lemmas II.21 and II.22 from Section 2.5.1, the maps J and C_0 are continuous and so is the composition $((\mathrm{id}, C_0) \circ J)$. This in view of (2.41) yields (2.19).
Now to complete the proof, we will use Theorem 6 along with the standard converging together Lemma II.26 as well as the continuity Lemma B.4 from Appendix B to establish (2.41). Before proceeding further, first we mention a result that we will use later in the proof. Note that for any interval I, not necessarily compact we have $\{\mathcal{T}_I(\mathbb{V}_n)(z)\}_{z\in I}$ converges in distribution to $\{\mathcal{T}_I(\mathbb{G})(z)\}_{z\in I}$ uniformly on compacta. Indeed by Theorem 1, $\{\mathbb{V}_n(z)\}_{z\in\mathbb{R}}$ converges in distribution to $\{\mathbb{G}(z)\}_{z\in\mathbb{R}}$ as a process uniformly on compact sets. The map $\mathcal{T}_K : C(K) \mapsto \mathcal{C}(K)$ is continuous for any compact set K, where both the spaces are equipped with topology of uniform convergence. So an application of the Continuous Mapping Theorem gives us the result for any compact interval I. If I is not compact we prove the result useing converging together lemma (Lemma II.26) and approximating I by some compact interval. The conditions of the Lemma can be verified using continuous mapping (as argued earlier) and Theorem 6. We adopt a similar method

to establish joint convergence though it is technically more challenging and involved. It is enough to show that for any fixed compact intervals $I \subset (-\infty, \infty)$, $I_- \subset (-\infty, 0)$ and $I_+ \subset (0, \infty)$, we have that (2.41) holds restricted to $\mathcal{C}(I) \times \mathcal{C}(I_-) \times \mathcal{C}(I_+)$, equipped with the uniform topology.

Let us fix such intervals and given $\delta > 0$ small and $c > 0$ large enough so that $I \subset [-c, c]$, $I_- \subset [-c, -1/c]$ and $I_+ \subset [-\delta, c]$, define,

$$\xi_{\delta,c,n} := \Big(\mathcal{T}_{[-c,c]}(\mathbb{V}_n)|_I,\, \mathcal{T}_{[-c,-1/c]}(\mathbb{V}_n)|_{I_-},\, \mathcal{T}_{[-\delta,c]}(\mathbb{V}_n)|_{I_+}\Big).$$

Let also

$$\xi_{\delta,c} := \Big(\mathcal{T}_{[-c,c]}(\mathbb{G})|_I,\, \mathcal{T}_{[-c,-1/c]}(\mathbb{G})|_{I_-},\, \mathcal{T}_{[-\delta,c]}(\mathbb{G})|_{I_+}\Big),$$

define $\xi := (\mathcal{T}_{(-\infty,\infty)}(\mathbb{G})|_I,\, \mathcal{T}_{(-\infty,0)}(\mathbb{G})|_{I_-},\, \mathcal{T}_{(0,\infty)}(\mathbb{G})|_{I_+}\Big)$, and finally

$$\eta_n := \Big(\mathcal{T}_{(-\infty,\infty)}(\mathbb{V}_n)|_I,\, \mathcal{T}_{(-\infty,l_n]}(\mathbb{V}_n)1_{(-\infty,0)}|_{I_-},\, \mathcal{T}_{(l_n,\infty)}(\mathbb{V}_n)1_{(0,\infty)}|_{I_+}\Big).$$

We will verify that $\xi_{\delta,c,n},\ \xi_{\delta,c},\ \xi$ and η_n satisfy the conditions of Lemma II.26. The GCM maps $\mathcal{T}_{[-c,c]},\ \mathcal{T}_{[-c,-1/c]}$ and $\mathcal{T}_{[-\delta,c]}$ are continuous on the spaces $C([-c,c])$, $C([-c,-1/c])$ and $C([-\delta,c])$ equipped with the the uniform norm. Therefore, by Theorem 1 and the Continuous Mapping Theorem, we obtain $\xi_{\delta,c,n} \Rightarrow \xi_{\delta,c},\ n \to \infty$, which verifies condition *(i)* of Lemma II.26.

Since $E = C(I) \times C(I_-) \times C(I_+)$ equipped with the uniform topology, it is enough to verify condition *(iii)* of Lemma II.26 for each of the three coordinates separately where d is the uniform metric on the corresponding interval (I, I_- or I_+). Recall that the processes $V_n := \mathbb{V}_n$ satisfy the conditions of Theorem 6 and hence (2.37) implies the condition *(iii)* for the first coordinate. To verify the condition for the second coordinate we apply Theorem 6 with $O_n := (-\infty, l_n]$, $O = (-\infty, 0)$ and $K_c = [-c, -1/c]$. Dealing with the third coordinate is more involved owing to the fact that in Theorem 6 the sequence of sets O_n increases to O, whereas the the intervals $[l_n, \infty) \downarrow (0, \infty)$. So the Theorem does not directly apply. To take care of the third coordinate, we will use Theorem 6 along with Lemma B.4 from Appendix B. Given $\epsilon > 0$, we have,

$$
\begin{aligned}
&\lim_{\delta \to 0} \lim_{c \to \infty} \limsup_{n \geq 1} \mathbb{P}\left(\sup_{z \in I_+} |\mathcal{T}_{[-\delta,c]}(\mathbb{V}_n)(z) - \mathcal{T}_{(l_n,\infty)}(\mathbb{V}_n)(z)| \geq \epsilon\right) \\
\leq &\lim_{\delta \to 0} \lim_{c \to \infty} \limsup_{n \geq 1} \mathbb{P}\left(\sup_{z \in I_+} |\mathcal{T}_{[-\delta,c]}(\mathbb{V}_n)(z) - \mathcal{T}_{[-\delta,\infty)}(\mathbb{V}_n)(z)| \geq \epsilon/2\right) \\
&+ \lim_{\delta \to 0} \lim_{c \to \infty} \limsup_{n \geq 1} \mathbb{P}\left(\sup_{z \in I_+} |\mathcal{T}_{[-\delta,\infty)}(\mathbb{V}_n)(z) - \mathcal{T}_{(l_n,\infty)}(\mathbb{V}_n)(z)| \geq \epsilon/2\right) \qquad (2.42)
\end{aligned}
$$

The first term in the right hand side is 0 by Theorem 6. Note that $l_n \uparrow 0$ as $n \to \infty$, so for given $\delta > 0$, for large enough n, we have $-\delta < l_n \leq 0$. Therefore the GCM function $\mathcal{T}_{(l_n,\infty)}(\mathbb{V}_n)(t)$ lies in between the GCMs $\mathcal{T}_{[-\delta,\infty)}(\mathbb{V}_n)(t)$ and $\mathcal{T}_{[0,\infty)}(\mathbb{V}_n)(t)$ for all $t \in I_+$. So the second term in (2.42) is bounded above by

$$
\lim_{\delta \to 0} \limsup_{n \geq 1} \mathbb{P}\left(\sup_{z \in I_+} |\mathcal{T}_{[-\delta,\infty)}(\mathbb{V}_n)(z) - \mathcal{T}_{[0,\infty)}(\mathbb{V}_n)(z)| \geq \epsilon/2\right).
$$

One can show that (will be proved at the end)

$$
\sup_{z \in I_+} |\mathcal{T}_{[0,\infty)}(\mathbb{V}_n)(z) - \mathcal{T}_{[-\delta,\infty)}(\mathbb{V}_n)(z)| \leq |\mathcal{T}_{[0,\infty)}(\mathbb{V}_n)(0) - \mathcal{T}_{[-\delta,\infty)}(\mathbb{V}_n)(0)|. \qquad (2.43)
$$

Now using the fact $\mathcal{T}_{[0,\infty)}(\mathbb{V}_n)(0) = 0$ and (2.43), the second term in (2.42) can be bounded above by:

$$
\lim_{\delta \to 0} \limsup_{n \geq 1} \mathbb{P}\left(|\mathcal{T}_{[-\delta,\infty)}(\mathbb{V}_n)(0)| \geq \epsilon/2\right) \leq \lim_{\delta \to 0} \mathbb{P}\left(|\mathcal{T}_{[-\delta,\infty)}(\mathbb{G})(0)| \geq \epsilon/2\right), \qquad (2.44)
$$

where the last inequality follows from the Portmanteau Theorem and the fact that $\mathcal{T}_{[-\delta,\infty)}(\mathbb{V}_n)$ converges in distribution to $\mathcal{T}_{[-\delta,\infty)}(\mathbb{G})$ uniformly on compact set as mentioned earlier. The last quantity in (2.44) is zero by Lemma B.4 from Appendix B (since the sample paths of $\mathbb{G}$ satisfy the conditions of that lemma with probability 1), which completes the proof of condition *(iii)* of Lemma II.26.

It was shown in Theorem 1 of [9] that Theorem 6 applies to the processes $V_n := \mathbb{G}$. Thus using similar arguments as above applying Relation (2.37) and Lemma B.4 of Appendix B we can show that $\xi_{\delta,c} \Rightarrow \xi$, as $c \to \infty$ and $\delta \uparrow 0$ (in fact the convergence is in probability).

Now it remains to prove (2.43) to complete the proof. To prove this first notice that, $\mathcal{T}_{[-\delta,\infty)}(\mathbb{V}_n)(z) \leq \mathcal{T}_{[0,\infty)}(\mathbb{V}_n)(z)$ for $z \in [0,\infty)$ and if for some $z_* \geq 0$ we have $\mathcal{T}_{[-\delta,\infty)}(\mathbb{V}_n)(z_*) = \mathcal{T}_{[0,\infty)}(\mathbb{V}_n)(z_*)$ then the two GCMs coincide on $[z_*,\infty)$. Let, $z_* = \inf(z \geq 0 : \mathcal{T}_{[-\delta,\infty)}(\mathbb{V}_n)(z) = \mathcal{T}_{[0,\infty)}(\mathbb{V}_n)(z))$. If $z_* = 0$, (2.43) is trivial, otherwise as argued in Lemma A.1 of [9], the GCM $\mathcal{T}_{[-\delta,\infty)}(\mathbb{V}_n)(z)$ is a linear function for $z \in [0, z_*]$. Therefore the left slope $\mathcal{L} \circ \mathcal{T}_{[-\delta,\infty)}(\mathbb{V}_n)(z) \equiv \mathcal{L} \circ \mathcal{T}_{[-\delta,\infty)}(\mathbb{V}_n)(z_*) \equiv$ *const.*, for all $z \in [0, z_*)$. Moreover by the fact that $\mathcal{T}_{[-\delta,\infty)}(\mathbb{V}_n)(z_*) = \mathcal{T}_{[0,\infty)}(\mathbb{V}_n)(z_*)$, and domination we get $\mathcal{L} \circ \mathcal{T}_{[0,\infty)}(\mathbb{V}_n)(z_*) \leq \mathcal{L} \circ \mathcal{T}_{[-\delta,\infty)}(\mathbb{V}_n)(z_*)$. This since $z \mapsto \mathcal{T}_{[0,\infty)}(\mathbb{V}_n)(z)$ is a non-decreasing function while $z \mapsto \mathcal{T}_{[-\delta,\infty)}(\mathbb{V}_n)(z_*)$ is constant on $(0, z_*)$, implies that the slope $\mathcal{L} \circ (\mathcal{T}_{[0,\infty)}(\mathbb{V}_n)(z) - \mathcal{T}_{[-\delta,\infty)}(\mathbb{V}_n)(z)) \leq 0$ for all $z \in (0, z_*)$. This shows that the function $z \mapsto \mathcal{T}_{[0,\infty)}(\mathbb{V}_n)(z) - \mathcal{T}_{[-\delta,\infty)}(\mathbb{V}_n)(z)$ is monotone non-increasing on $[0, z_*]$ and (2.43) holds. This completes the proof. $\square$

2.5.2.3 The Statistics L_n and T_n

Finally we are ready to prove the asymptotic properties of the statistics L_n and T_n. We first state a version of the converging together lemma which is used later, and is an adaptation of Theorem 8.6.2 in [54].

Lemma II.26. *Let* $\xi, \xi_{\delta,c,n},\ \xi_{\delta,c},\ \eta_n,\ n \in \mathbb{N}, \delta,\ c > 0$ *be random elements taking values in a metric space* (E,d). *If (i)* $\xi_{\delta,c,n} \Rightarrow \xi_{\delta,c}$, *as* $n \to \infty$, *(ii)* $\xi_{\delta,c} \Rightarrow \xi$, *as*

$c \to \infty$ and $\delta \uparrow 0$ and (iii) for all $\epsilon > 0$,

$$\lim_{\delta \uparrow 0} \lim_{c \to \infty} \limsup_{n \geq 1} \mathbb{P}(d(\xi_{\delta,c,n}, \eta_n) > \epsilon) = 0, \tag{2.45}$$

then $\eta_n \Rightarrow \xi$, as $n \to \infty$.

Proof of Proposition II.2: By adding and subtracting θ_0 and expanding the squares in the two sums in Relation (2.4), we obtain

$$\begin{aligned} L_n &= \frac{n}{\sigma_n^2}\Big(\underbrace{-2\sum_{i=1}^{n}(Y_i - \theta_0)(\hat{m}_n^0(t_i) - \theta_0)}_{=:A_0} + \underbrace{\sum_{i=1}^{n}(\hat{m}_n^0(t_i) - \theta_0)^2}_{=:B_0}\Big) \\ &\quad - \frac{n}{\sigma_n^2}\Big(\underbrace{-2\sum_{i=1}^{n}(Y_i - \theta_0)(\hat{m}_n(t_i) - \theta_0)}_{=:A} + \underbrace{\sum_{i=1}^{n}(\hat{m}_n(t_i) - \theta_0)^2}_{=:B}\Big). \end{aligned}$$

It is known by the so–called pooled adjacent violators (PAV) characterization of isotonic regression that $\hat{m}_n(t_i)$s are sample averages of Y_js over non–overlapping blocks of indices j. (see [19]). This is also true for $\hat{m}_n^0(t_i)$s whenever $\hat{m}_n^0(t_i) \neq \theta_0$. Therefore, by grouping together the terms in the sum A that correspond to the same $\hat{m}_n(t_i)$s, we obtain that $A = 2B$. Similarly, we have $A_0 = 2B_0$ and therefore,

$$L_n = \frac{n}{\sigma_n^2}(-B_0 + B) = \frac{n}{\sigma_n^2}\Big(\sum_{i=1}^{n}(\hat{m}_n(t_i) - \theta_0)^2 - \sum_{i=1}^{n}(\hat{m}_n^0(t_i) - \theta_0)^2\Big).$$

Recall now that $X_n(z) = d_n^{-1}(\hat{m}_n(t_0 + d_n z) - \theta_0)$, and
$Y_n(z) = d_n^{-1}(\hat{m}_n^0(t_0 + d_n z) - \theta_0)$, for $z \in (-d_n t_0, (1 - t_0)d_n] =: (a_n, b_n]$. Further, by definition, we have that $X_n(z) \equiv Y_n(z)$, for all $z \notin (a_n, b_n]$ and therefore the integrals in (2.13) are finite.

By the charecterization $\hat{m}_n(t)$ is constant over
$(t_{i-1}, t_i] \equiv ((i-1)/n, i/n],\ i = 1, \cdots, n$, and $\hat{m}_n^0(t)$ is constant over all $(t_{i-1}, t_i] \not\ni t_0$.

Thus,

$$L_n = \frac{n^2}{\sigma_n^2}\Big(\int_0^1 (\hat{m}_n(t)-\theta_0)^2 dt - \int_0^1 (\hat{m}_n^0(t)-\theta_0)^2 dt\Big) + R_n$$
$$= \frac{n^2 d_n^3}{\sigma_n^2} \int_{(a_n,b_n]} (X_n^2(z) - Y_n^2(z))dz + R_n, \qquad (2.46)$$

where R_n is given below and where the last relation follows by the change of variables to local coordinates $z = d_n^{-1}(t-t_0)$.

Since the only interval $(t_{i-1}, t_i],\ i = 1, \cdots, n$ where $\hat{m}_n^0(t)$ is potentially non constant is the one containing t_0, i.e. $i = [nt_0]+1 = l+1$, we get

$$R_n = \frac{n^2}{\sigma_n^2}\Big(\int_{t_l}^{t_{l+1}} (\hat{m}_n^0(t)-\theta_0)^2 dt - \frac{1}{n}(\hat{m}_n^0(t_{l+1})-\theta_0)^2\Big).$$

By the monotonicity of $\hat{m}_n^0(t)$, we have
$(\hat{m}_n^0(t_{l+1})-\theta_0)^2 \le (\hat{m}_n^0(t)-\theta_0)^2 + (\hat{m}_n^0(s)-\theta_0)^2$, for all $t \le t_{l+1} \le s$, which implies

$$R_n \le \frac{2n^2}{\sigma_n^2} \int_{t_l}^{t_{l+2}} (\hat{m}_n^0(t)-\theta_0)^2 dt = \frac{2n^2 d_n^3}{\sigma_n^2} \int_{\Delta_n} Y_n^2(z)dz, \qquad (2.47)$$

where $\Delta_n := d_n^{-1}([nt_0]/n - t_0,\ [nt_0]/n + 2/n - t_0) \subset [-1/nd_n, 3/nd_n]$.

By Theorem 2, we have that $Y_n \Rightarrow \mathcal{S}_{a,b}^0$, and since Δ_n is a shrinking interval around 0 the Portmanteau Theorem implies that for all $\epsilon > 0$ and $\delta > 0$,

$$\begin{aligned}\limsup_{n\to\infty} \mathbb{P}\Big(\int_{\Delta_n} Y_n^2(z)dz \ge \epsilon\Big) &\le \limsup_{n\to\infty} \mathbb{P}\Big(\int_{-\delta}^{\delta} Y_n^2(z)dz \ge \epsilon\Big) \\ &\le \mathbb{P}\Big(\int_{-\delta}^{\delta} (\mathcal{S}_{a,b}^0(z))^2 dz \ge \epsilon\Big) \qquad (2.48)\end{aligned}$$

As shown in the proof of Theorem 19 of Appendix B, $\mathcal{S}_{a,b}^0(z)$ is zero in a neighborhood of 0, and therefore $\int_{-\delta}^{\delta}(\mathcal{S}_{a,b}^0(z))^2dz \to 0$, as $\delta \downarrow 0$, in probability. Therefore the right–hand side of (2.48) can be made arbitrarily small. This implies $\int_{\Delta_n} Y_n^2(z)dz \to 0$, in probability, as $n \to \infty$, which which in view of (2.46) and (2.47) yields (2.13). The argument for the statistic T_n is similar. □

Proof of Theorem 4 By Lemma II.3 and Theorem 18 (Appendix B), for every $\epsilon > 0$ there exists an interval $K_\epsilon := [-M_\epsilon, M_\epsilon]$ such that, for all large n,

$$\mathbb{P}\Big[D_n \subset [-M_\epsilon, M_\epsilon]\Big] > 1-\epsilon \text{ and } \mathbb{P}\Big[D_{a,b} \subset [-M_\epsilon, M_\epsilon]\Big] > 1-\epsilon.$$

Let now

$$\xi_{\epsilon,n} = \Big(\int\limits_{K_\epsilon} \Big(X_n^2(z) - Y_n^2(z)\Big)dz, \int\limits_{K_\epsilon} \Big(X_n(z) - Y_n(z)\Big)^2 dz\Big),$$

$$\xi_\epsilon = \Big(\int\limits_{K_\epsilon} \Big((\mathcal{S}_{a,b}(z))^2 - (\mathcal{S}_{a,b}^0(z))^2\Big)dz, \int\limits_{K_\epsilon} \Big(\mathcal{S}_{a,b}(z) - \mathcal{S}_{a,b}^0(z)\Big)^2 dz\Big).$$

Also, let

$$\eta_n = \Big(\int\limits_{(a_n,b_n]} \Big(X_n^2(z) - Y_n^2(z)\Big)dz, \int\limits_{\mathbb{R}} \Big(X_n(z) - Y_n(z)\Big)^2 dz\Big),$$

$$\xi = \Big(\int\limits_{D_{a,b}} \Big((\mathcal{S}_{a,b}(z))^2 - (\mathcal{S}_{a,b}^0(z))^2\Big)dz, \int\limits_{D_{a,b}} \Big(\mathcal{S}_{a,b}(z) - \mathcal{S}_{a,b}^0(z)\Big)^2 dz\Big).$$

Since K_ϵ contains $D_n := \{z : X_n(z) \neq Y_n(z)\}$ with probability greater than $1-\epsilon$ and $(a_n, b_n]$ grows up to $\mathbb{R}$, for large n, we have $\lim_{\epsilon\downarrow 0}\limsup_{n\geq 1}\mathbb{P}(\xi_{\epsilon,n} \neq \eta_n) = 0$. We similarly have that $\lim_{\epsilon\downarrow 0}\mathbb{P}(\xi_\epsilon \neq \xi) = 0$. Finally, by Theorem 2 and the continuous mapping Theorem, for all fixed $\epsilon > 0$, we have $\xi_{\epsilon,n} \Rightarrow \xi_\epsilon$, as $n \to \infty$. Thus, all conditions of the converging together lemma (cf Lemma II.26) hold, where in this simple case there is no dependence on $\delta > 0$. Hence $\eta_n \Rightarrow \xi,\ n \to \infty$, which, in view of Proposition II.2, yields

$$\frac{\sigma_n^2}{n^2 d_n^3}(L_n, T_n) \Longrightarrow \left(\mathbb{L}_{a,b}^{(H)}, \mathbb{T}_{a,b}^{(H)}\right)$$

as $n \to \infty$.

To complete the proof, it remains to show that

$$\left(\mathbb{L}_{a,b}^{(H)}, \mathbb{T}_{a,b}^{(H)}\right) \stackrel{d}{=} a^2 \left(\frac{a}{b}\right)^{\frac{2H-1}{2-H}} \left(\mathbb{L}^{(H)}, \mathbb{T}^{(H)}\right). \tag{2.49}$$

This follows from a scaling argument. Indeed, by the H–self–similarity of B_H, for $\mathbb{G}(z) \equiv \mathbb{G}_{a,b}^H(z) = aB_H(z) + bz^2$, we have

$$\{\mathbb{G}_{a,b}^H(z)\}_{z\in\mathbb{R}} \stackrel{d}{=} a(a/b)^{\frac{H}{2-H}} \left\{ \left(\mathbb{G}_{1,1}^H((b/a)^{\frac{1}{2-H}} z)\right) \right\}_{z\in\mathbb{R}}. \tag{2.50}$$

Thus, the process $\{(\mathcal{S}_{a,b}(z), \mathcal{S}_{a,b}^0(z))\}_{z\in\mathbb{R}}$ equals in distribution

$$a\,(b/a)^{\frac{1-H}{2-H}} \left\{ \left(\mathcal{S}_{1,1}((b/a)^{\frac{1}{2-H}}\, z),\, \mathcal{S}_{1,1}^0((b/a)^{\frac{1}{2-H}}\, z)\right) \right\}_{z\in\mathbb{R}}, \tag{2.51}$$

which by substituting in (2.20) and making a change of variables yields (2.49). □

Remark II.27. The result of Theorem 3 can be formally recovered from the statement of Theorem 4 by letting $H = 1/2$, $a = \tau$, $d_n = n^{-1/3}$, using the fact that $\sigma_n^2/n \to \tau^2$ and noting that $\mathbb{L}^{(1/2)}$ and $\mathbb{T}^{(1/2)}$ are precisely the $\mathbb{L}$ and $\mathbb{T}$ of Theorem 3 respectively.

2.5.3 Behavior of the Statistics $L_n(\theta)$, $T_n(\theta)$ and $R_n(\theta)$

The following Lemmas describe the shape of the statistics we have discussed before.

Lemma II.28. *Define L_n and T_n as in* (2.4). *We have $L_n \geq T_n$.*

Proof. Note that $L_n \geq T_n$ is equivalent to

$$
\begin{aligned}
& \sum_{i=1}^{n}(Y_i - \hat{m}_n^0(t_i))^2 - \sum_{i=1}^{n}(Y_i - \hat{m}_n(t_i))^2 \geq \sum_{i=1}^{n}\left(\hat{m}_n(t_i) - \hat{m}_n^0(t_i)\right)^2 \\
\iff & \sum_{i=1}^{n}(Y_i - \hat{m}_n(t_i))^2 + \sum_{i=1}^{n}(Y_i - \hat{m}_n^0(t_i))(Y_i - \hat{m}_n(t_i)) \leq 0 \\
\iff & \sum_{i=1}^{n}(Y_i - \hat{m}_n(t_i))(\hat{m}_n^0(t_i) - \hat{m}_n(t_i)) \leq 0. \qquad (2.52)
\end{aligned}
$$

Notice that by the definition of isotonic regression (recall (2.2)) the vector $\overrightarrow{\hat{m}} := (\hat{m}_n(t_i))_{i=1}^n$ is the projection of the vector $\overrightarrow{y} = (y_i)_{i=1}^n$ onto the convex set $V := \{\overrightarrow{x} = (x_i)_{i=1}^n : x_1 \leq x_2 \leq \cdots \leq x_n\}$. The vector $\overrightarrow{\hat{m}^0} := (\hat{m}_n^0(t_i))_{i=1}^n$ is in V. So by the well-known characterization of projections onto closed convex sets we have $(\overrightarrow{y} - \overrightarrow{\hat{m}})^T(\overrightarrow{\hat{m}^0} - \overrightarrow{\hat{m}}) \leq 0$. The last inequality is equivalent to (2.52). Hence we have the result. $\square$

Lemma II.29. *Both $L_n(\theta)$ and $T_n(\theta)$ are continuous in θ, $L_n(\hat{\theta}_n) = T_n(\hat{\theta}_n) = 0$ and monotone non-increasing on $(-\infty, \hat{\theta}_n]$ and monotone non-decreasing on $(\hat{\theta}_n, \infty)$. Also $L_n(\theta)$ and $T_n(\theta)$ diverge to ∞ as θ goes to ∞ or $-\infty$.*

Proof. Let $\tilde{m}_n(t)$ be the left derivative of greatest convex minorant of $U_n(t)$ fitted separately for left and right side of t_l. Then the (constrained) estimate of m under the constraint $m(t_0) = \theta$ is given by (2.54). Also, if $\hat{\theta}_n$ is the isotonic regression estimate of $m(t_0)$, then $\hat{m}_n(t) = \hat{m}_n^{\hat{\theta}_n}(t)$.

Let, $a_1 < a_2 < \cdots < a_m$ be the distinct values of $\tilde{m}_n(t)$ and the corresponding design points are $s_1 < s_2 < \cdots < s_m$. Also, let, $s_k < t_0 < s_{k+1}$. Then $L_n(\theta)$ can be written as

$$\begin{aligned} L_n(\theta) =& \frac{n}{\sigma_n^2}\sum_{i=1}^{n}\left[\left((\hat{m}_n(t_i)-\theta)^2-(\hat{m}_n^{\theta}(t_i)-\theta)^2\right)\right] \\ =& \frac{n}{\sigma_n^2}\sum_{i=1}^{k}\left(\left(a_i\wedge\hat{\theta}_n-\theta\right)^2-(a_i\wedge\theta-\theta)^2\right) \\ &+\frac{n}{\sigma_n^2}\sum_{i=k+1}^{m}\left(\left(a_i\vee\hat{\theta}_n-\theta\right)^2-(a_i\vee\theta-\theta)^2\right) \end{aligned} \tag{2.53}$$

From (2.53) it follows that $L_n(\hat{\theta}_n)=0$, L_n is continuous in θ and it diverges to ∞ as $|\theta|\to\infty$. The fact that $L_n(\theta)$ is monotone non-increasing on $(\infty,\hat{\theta}_n]$ and monotone non-decreasing on $(\hat{\theta}_n,\infty)$ is argued considering θ in different intervals and using simple algebra.

We also have

$$T_n(\theta)=\frac{n}{\sigma_n^2}\left[\sum_{i=1}^{k}\left(a_i\wedge\hat{\theta}_n-a_i\wedge\theta\right)^2+\sum_{i=k+1}^{m}\left(a_i\vee\hat{\theta}_n-a_i\vee\theta\right)^2\right],$$

and similar arguments will show the results for $T_n(\theta)$. □

Proposition II.30. *At the jump points of isotonic regression estimator $L_n(\theta)=T_n(\theta)$ for all values of θ.*

Proof. Note that with the formulation as in the proof of Lemma II.29, at the jump points we have $a_i\vee\hat{\theta}_n=a_i$ and $a_i\wedge\hat{\theta}_n=a_i$ for all i. Now without loss of generality assume that $a_{l-1}<\theta\le a_l$ and $l\le k$. The other cases can be handled similarly. From the representation (2.53) we can write

$$L_n(\theta)=\frac{n}{\sigma_n^2}\sum_{i=l+1}^{k}(a_i-\theta)^2.$$

Similar calculations yield the same form for $T_n(\theta)$. Hence the result. □

To prove the next result, we need one preliminary result first.

Lemma II.31. *Assume that a sequence of stochastic process* $\{m_n(t)\}_{t\in[a,b]}$ *converges to* $\{m(t)\}_{t\in[a,b]}$ *in finite dimensional distributions. The functions* m_n *are monotone non-decreasing and* m *is continuous monotone non-decreasing and non-random. Then* $\{m_n(t)\}_{t\in[a,b]}$ *converges to* $\{m(t)\}_{t\in[a,b]}$ *in distribution uniformly.*

Proof. Consider a grid $a = t_1 < t_2 < \cdots < t_k = b$ such that $\|m(t_i) - m(t_{(i+1)})\| < \epsilon$ for some given $\epsilon > 0$. Then by monotonicity of m_n and m we have,

$$
\begin{aligned}
\sup_{t\in[a,b]} |m_n(t) - m(t)| = &\max_{i=1,2,\ldots n} (|m_n(t_i) - m(t_{(i+1)})| \vee |m_n(t_{(i+1)}) - m(t_i)|) \\
\leq &\max_{i=1,2,\ldots k} (|m_n(t_i) - m(t_i)| + |m(t_i) - m(t_{(i+1)})| \\
&+ |m_n(t_{(i+1)}) - m(t_{(i+1)})| + |m(t_{i+1}) - m(t_i)|) \\
< &2\epsilon + \max_{i=1,2,\ldots k} (|m_n(t_i) - m(t_i)| + |m_n(t_{(i+1)}) - m(t_{(i+1)})|)
\end{aligned}
$$

The second term converges to zero in probability because of the finite dimensional convergence of m_n to m. As $\epsilon > 0$ is arbitrary this implies that $\{m_n(t)\}_{t\in[a,b]}$ converges to $\{m(t)\}_{t\in[a,b]}$ in probability uniformly and hence in distribution.

□

Proposition II.32. *Let* $\theta \neq \theta_0$ *and* $R_n(\theta)$ *be the ratio statistic calculated under the restriction* $m(t_0) = \theta$*. Then, under* $H_0 : m(t_0) = \theta_0$, $R_n(\theta) \xrightarrow{P} 1$ *as* $n \to \infty$*.*

Proof. Assume $\theta > \theta_0$. By [9], $\hat{m}_n(t) \xrightarrow{P} m(t)$ for all $t \in (0, 1)$. Let $\tilde{m}_n(t)$ be the slope of the GCM of $U_n(t)$ (see (2.14)) where the GCM is fitted separately at left and right of t_l, the nearest design point at the left of t_0. The isotonic regression estimate of $m(t)$ under the constraint $m(t_0) = \theta$ is given by

$$\hat{m}_n^\theta(t) = \begin{cases} \tilde{m}_n(t) \wedge \theta, & \text{if } t \leq t_l \\ \theta, & \text{if } t_l < t \leq t_0 \\ \tilde{m}_n(t) \vee \theta, & \text{if } t > t_0. \end{cases} \tag{2.54}$$

Considering the regression problem on the intervals $[0, t_l]$ and $[t_l, 1]$ separately and by the fact $t_l \to t_0$ as $n \to \infty$, we have for $t \in (0, t_0)$, $\hat{m}_n^\theta(t) \xrightarrow{P} m(t) \wedge \theta = m(t)$ and for $t \in (t_0, 1)$, $\hat{m}_n^\theta(t) \xrightarrow{P} m(t) \vee \theta$. So,
$\hat{m}_n^\theta(t) \xrightarrow{P} m(t)\mathbf{1}_{(t<t_0)} + (m(t) \vee \theta)\mathbf{1}_{(t\geq t_0)} := m^\theta(t)$ for all $t \in (0, 1)$.
Now as $\hat{m}_n(t) \xrightarrow{P} m(t)$, for any $0 < a < b < 1$ $\{\hat{m}_n(t)\}_{t\in[a,b]} \to \{m(t)\}_{t\in[a,b]}$ in finite dimensional distribution. Also, $\hat{m}_n(t)$ is increasing and $m(t)$ is continuous, increasing and non random. So Lemma II.31 implies that
$\{\hat{m}_n(t)\}_{t\in[a,b]} \Rightarrow \{m(t)\}_{t\in[a,b]}$ uniformly on $D[a, b]$. As $m(t)$ is non-random this implies that $\{\hat{m}_n(t)\}_{t\in[a,b]}$ converges in probability to $\{m(t)\}_{t\in[a,b]}$ in $D[a, b]$. Similar arguments can be applied to establish the convergence of $\{\hat{m}_n^\theta(t)\}_{t\in[a,b]}$ to $\{m^\theta(t)\}_{t\in[a,b]}$ in probability as a process in $D[a, b]$. So,$\{\hat{m}_n(t), \hat{m}_n^\theta(t)\}_{t\in[a,b]}$ converges jointly to $\{m(t), m^\theta(t)\}_{t\in[a,b]}$ in probability.

Now look at the statistic $R_n(\theta)$:

$$
\begin{aligned}
R_n(\theta) =& \frac{L_n(\theta)}{T_n(\theta)} = \frac{\sum_{i=1}^{n} \left((\hat{m}_n(t_i)-\theta)^2 - (\hat{m}_n^\theta(t_i)-\theta)^2\right)}{\sum_{i=1}^{n} \left(\left(\hat{m}_n(t_i)-\theta\right) - \left(\hat{m}_n^\theta(t_i)-\theta\right)\right)^2} \\
=& \frac{\int_0^1 ((\hat{m}_n(t)-\theta)^2 - (\hat{m}_n^\theta(t)-\theta)^2)dt}{\int_0^1 ((\hat{m}_n(t)-\theta) - (\hat{m}_n^\theta(t)-\theta))^2 dt} \\
=& \frac{1}{1-2\bar{R}_n(\theta)}
\end{aligned}
$$

where,

$$
\bar{R}_n(\theta) = \frac{\int_0^1 (\hat{m}_n^\theta(t)-\theta)(\hat{m}_n(t)-\hat{m}_n^\theta(t))dt}{\int_0^1 (\hat{m}_n(t)+\hat{m}_n^\theta(t)-2\theta)(\hat{m}_n(t)-\hat{m}_n^\theta(t))dt} \tag{2.55}
$$

By Lemma II.3 given $\epsilon > 0$ we can find $0 < a < b < 1$ such that $P(\tilde{m}_n \neq \hat{m}_n \subset [a,b]) > 1-\epsilon$ for sufficiently large n. Pick a and b such that $m(a) < \theta_0$ and $m(b) > \theta$. As $\hat{m}_n(t)$ converges in probability to $m(t)$ for $t \in (0,1)$ we have $P(\hat{m}_n(b) > \theta) > 1-\epsilon$ and $P(\hat{m}_n(a) < \theta_0) > 1-\epsilon$ for large enough n. Consider the event

$$
A = \{\tilde{m}_n \neq \hat{m}_n \subset [a,b]\} \cap \{\hat{m}_n(b) > \theta\} \cap \{\hat{m}_n(a) < \theta_0\}.
$$

From above discussion we have $P(A) > 1-3\epsilon$. For all $\omega \in A$, if $t \notin [a,b]$, $\hat{m}_n(t) \equiv \tilde{m}_n(t) > \theta$, and therefore for $t > b$, $\hat{m}_n^\theta(t) = \tilde{m}_n(t) \vee \theta = \tilde{m}_n(t) = \hat{m}_n(t)$ and for $t < a$, $\hat{m}_n^\theta(t) = \tilde{m}_n(t) \wedge \theta = \tilde{m}_n(t) = \hat{m}_n(t)$.

So, $\exists [a,b] \subset (0,1)$ such that $P(\hat{m}_n^\theta \neq \hat{m}_n \subset [a,b]) \to 1$ as $n \to \infty$.

So all the integrals in $\bar{R}_n(\theta)$ can be considered as integral over $[a,b]$. The integrand of the denominator converges in probability to

$(m^\theta(t) + m(t) - 2\theta)(m(t) - m^\theta(t)) = (m(t) - \theta)\mathbf{1}_{(t \geq t_0, m(t) < \theta)}$. As m is continuous and increasing and $m(t_0) = \theta_0 < \theta$, so this limiting function is positive on the interval $[t_0, m^{-1}(\theta) \wedge 1]$. And the integrand of the numerator converges in probability to $(m^\theta(t) - \theta)(m(t) - m^\theta(t)) = 0$. As the integrals in both numerator and denominator are continuous functional of $\hat{m}_n$ and $\hat{m}_n^\theta$ in $L^2_{[0,1]}$ we have $\bar{R}_n(\theta) \xrightarrow{P} 0$ as $n \to \infty$. This in turn implies that $R_n(\theta) \xrightarrow{P} 1$ as $n \to \infty$. $\square$

Theorem 7. *Let $L_n(\theta, t)$ be the test statistic for testing $H_0 : m(t) = \theta$. If $t_1 \neq t_2$ then $L_n(\theta, t_1)$ and $L_n(\theta, t_2)$ are asymptotically independent.*

Proof. We will prove the result for simpler case where the errors are Gaussian. The general case is similar but more technical.

Note that by Proposition A.2 and (2.16) the process $L_n(\theta, t_1)$ is a functional of the process $V_n^1(z) := d_n^{-2}\Big(U_n(t_1 + d_n z) - U_n(t_1) - m(t_1) d_n z\Big)$ and following the proof of Theorem 1 we can write $V_n^1(z) = \sum_{i=\lfloor nt_1 \rfloor}^{\lfloor nt_1 + nd_n z \rfloor} \epsilon_i + \Lambda_n(z) + o(1)$, where $\Lambda_n(z)$ is a deterministic term. By direct calculation we have $\mathrm{Cov}(V_n^1(z), V_n^2(z)) \to 0$ for all z as $n \to \infty$ under our assumptions of both SRD and LRD. As for Gaussian errors zero correlation implies independence we have our result. $\square$

CHAPTER III

M-Estimation Under Dependence

3.1 Problem Formulation and Notations

We have data $\{X_{ni}\}_{i=1}^{n}$ from a double array of row-wise stationary χ valued random variables, i.e., for each n, the random variables $X_{n1}, X_{n2}, \ldots, X_{nn}$ come from a stationary process. For most practical situations χ is a Borel subset of $\mathbb{R}^m$. Consider the parameter space Θ_n to be a metric space with metric d_n. For all practical purposes d_n is the natural metric associated with the parameter space Θ_n. Furthermore, let $\mathfrak{M}_n = \{m_{n,\theta} : \theta \in \Theta_n\}$ be the class of real–valued functions defined on χ. We want to study the convergence of maxima of the following criterion function:

$$\theta \mapsto \mathbb{P}_n(m_{n,\theta}) = \frac{1}{n}\sum_{i=1}^{n} m_{n,\theta}(X_{ni})$$

to the maxima of $P_n(m_{n,\theta}) = \mathbb{E}\mathbb{P}_n(m_{n,\theta})$. We will work under the assumption that the function $\theta \mapsto P_n(m_{n,\theta})$ has a well separated maxmimum, i.e.,

Assumption (A): There exists a point θ_n such that

$$P_n(m_{n,\theta_n}) > \sup_{\theta \notin \Theta^0} P_n(m_{n,\theta}) + \xi \tag{3.1}$$

for all $\xi > 0$ and every open set Θ^0 containing θ_n.

Definition 1. Define the M-estimator to be the near maximizer of $\theta \mapsto \mathbb{P}_n(m_{n,\theta})$ in the sense that $\hat{\theta}_n$ is the M-estimator if for all n,

$$\mathbb{P}_n(m_{n,\hat{\theta}_n}) \geq \sup_{\theta \in \Theta_n} \mathbb{P}_n(m_{n,\theta}) - o_p(1).$$

To establish the asymptotic properties of the M-estimator we need to investigate the convergence of the associated empirical processes uniformly over θ. Following basic empirical process notation we define the process $\mathbb{G}_n$ as

$$\mathbb{G}_n(\theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (m_{n,\theta}(X_{ni}) - \mathbb{E}m_{n,\theta}(X_{ni})) \tag{3.2}$$

for all $\theta \in \Theta_n$. For any class of functions $\mathcal{M}$, we define the L_{2r} norm as $\|m_{n,\theta}\|_{2r} := (\mathbb{E}|m_{n,\theta}(X_{n1})|^{2r})^{1/2r}$ for $r \geq 1$. We work under the assumption that the L_2 norm is uniformly continuous with respect to the metric d_n, i.e.

Assumption D: $\|m_{n,\theta} - m_{n,\theta_n}\|_2 \to 0$ if and only if $d_n(\theta_n, \theta) \to 0$.

In other words the metric induced by L_2 norm on the class of functions $\mathfrak{M}_n$ is equivalent to the natural metric associated with the parameter space.

The uniform convergence of the process over θ, and hence the properties of $\hat{\theta}_n$, depends on the behavior of the class $\mathfrak{M}_n$. The behavior of this class of functions here is phrased in terms of bracketing numbers and bracketing entropy.

Given two real valued functions l and u, the bracket $[l, u]$ is the set of all functions f satisfying $l \leq f \leq u$. An ϵ–bracket is a bracket $[l, u]$ with $u - l \leq \epsilon$. The bracketing number for a class of functions $\mathcal{M}$ with respect to the norm $\|.\|$, denoted by $N_{[]}(\epsilon, \mathcal{M}, \|.\|)$ is the minimum number of ϵâĹŠ-brackets needed to cover $\mathcal{M}$. The entropy with bracketing is the logarithm of the bracketing number. Note that the upper and lower bounds u and l of the brackets need not belong to $\mathcal{M}$ themselves but are assumed to have finite norms. Finally, we define a quantity which will appear repeatedly in our analysis. Consider the class of functions

$$\mathcal{M}_{n,\delta} = \{m_{n,\theta} - m_{n,\theta_n} : d_n(\theta, \theta_n) < \delta\} \tag{3.3}$$

and define the function

$$\phi_n(\delta) = \int_0^\infty [\log N_{[]}(\epsilon, \mathcal{M}_{n,\delta}, \|.\|_{2r})]^{1/2} d\epsilon \tag{3.4}$$

Note that the class $\mathcal{M}_{n,\delta}$ and hence ϕ_n depends on the choice of the metric d_n on Θ_n. When d_n is the Euclidean metric on Θ_n, i.e., $d_n(\theta_1, \theta_2) = \|\theta_1 - \theta_2\|$, we write $\psi_n(\delta) := \phi_{n,\|.\|}(\delta)$, i.e., ψ_n is the entropy integral of the class of functions $\mathcal{M}_{n,\delta} = \{m_{n,\theta} - m_{n,\theta_n} : \|\theta - \theta_n\| < \delta\}$.

3.1.1 Dependence Structure

The dependence structure of the stochastic process plays an important role in our analysis. Formally we consider data generated from an absolutely regular mixing stochastic process.

Definition 2. Define β-mixing coefficients between two sigma-algebras $\mathcal{A}$ and $\mathcal{B}$ as

$$\beta(\mathcal{A}, \mathcal{B}) = \sup_{\substack{A_1, A_2, \ldots A_m \in \mathcal{A} \\ B_1, B_2, \ldots, B_n \in \mathcal{B} \\ \uplus A_i = \uplus B_j = \Omega}} \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{n} |P(A_i \cap B_j) - P(A_i)P(B_j)|.$$

For a sequence of random variables $\{X_i\}_{i \in \mathbb{Z}}$, the β-mixing coefficients are defined as

$$\beta_k = \sup_{l \in \mathbb{Z}} \beta(\sigma(X_i : i \leq l - k), \sigma(X_i : i \geq l)) \tag{3.5}$$

and the sequence is said to be a β-mixing or absolutely regular mixing sequence if $\beta_k \to 0$ as $k \to \infty$. Here we list some important properties of β mixing that will be used in this paper.

Lemma III.1. *If $\{X_i\}_{i\in\mathbb{Z}}$ is a β-mixing sequence then so is $\{f(X_i)\}_{i\in\mathbb{Z}}$ for any measurable function f. The β-mixing coefficients of the sequence $f(X_i)$ are bounded above by those of the sequence X_i.*

Proof. The proof follows trivially from the definition of β-mixing and the fact that for any random variable X and measurable function f, we have
$\sigma(f(X)) \subset \sigma(X)$. □

Lemma III.2. *(Adapted from [16]) Let X and Y be two random variables taking their values in Borel spaces S_1 and S_2 respectively and U be a Uniform(0,1) random variable which is independent of (X, Y). Then there exists a random variable $Y^* = f(X, Y, U)$ where f is a measurable function from $S_1 \times S_2 \times [0, 1]$ into S_2 such that*

(i) Y^* is independent of X and has the same distribution as Y and

(ii) $P(Y \neq Y^*) = \beta(\sigma(X), \sigma(Y))$.

The assumption of β-mixing is quite common in the literature and covers a variety of commonly encountered situations. Davydov ([24]) showed that Markov chains under Harris recurrence conditions are geometrically β-mixing. For example, if $(X_t, \sigma_t)_{t\in\mathbb{Z}}$ is a stationary GARCH(p,q) process such that the noise sequence is absolutely continuous with Lebesgue density being strictly positive in a neighborhood of zero and $\mathbb{E}|\epsilon_0|^s < \infty$ for some $s > 0$ then both the sequences $(X_t)_{t\in\mathbb{Z}}$ and $(X_t^2)_{t\in\mathbb{Z}}$ are geometrically β-mixing (see [18]). Mokkadem ([44]) showed that stationary vector valued ARMA processes with innovations from an absolutely continuous distribution with respect to Lebesgue measure are geometrically β-mixing.

For our purposes, we need to work with a double array of random variables. For a double array $\{X_{ni}\}$ define β-mixing coefficients as

$$\beta_{n,k} = \sup_{l \in \mathbb{Z}} \beta(\mathcal{F}_{n,l-k}, \mathcal{G}_{n,l}) \tag{3.6}$$

where $\mathcal{F}_{n,j} = \sigma(X_{ni} : i \leq j)$ and $\mathcal{G}_{n,j} = \sigma(X_{ni} : i \geq j)$ and the mixing rate function $\beta_{(n)}(u) = \beta_{n,[u]}$ for $u \in \mathbb{R}^+$. In the sequel, we use two different conditions on the mixing rate.

(M1) (Polynomial decay) $\beta_{(n)}(u) \leq Cu^{-L}$ with $L > 1$, C and L do not depend on n.

(M2) (Exponential decay) $\beta_{(n)}(u) \leq Cb^u$ for some $b \in (0, 1)$, c and b do not depend on n.

Note that assumption (M1) imposes a weaker condition on the dependence structure than (M2), but it requires existence of higher moments. The rate of convergence of the M-estimator is derived under either of the assumptions but to derive asymptotic distribution of the M-estimator we require (M2).

3.2 Asymptotic Properties of the M-estimator

In this section we establish the asymptotic properties of the M-estimator. To this end, we first state a functional maximal inequality for a double array of β-mixing random variables. The framework for this maximal inequality and a subsequent central limit theorem stated in Section 3.2.1 is more general than our M-estimation set-up. Recall that the M-estimator can be viewed as the maximizer of the process $\{\mathbb{P}_n(m_{n,\theta})\}_{\theta \in \Theta_n}$. The general functional maximal inequality established in Section 3.2.1 provides an upper bound for the oscillation of this process, which determines the rate of convergence. The same result is also used to establish the tightness of a properly normalized version of the process $\{\mathbb{P}_n(m_{n,\theta})\}_{\theta \in \Theta_n}$, which is essential to obtain its asymptotic distribution. An application of the argmax continuous mapping theorem then gives the asymptotic distribution of normalized M-estimator.

3.2.1 A Maximal Inequality and Central Limit Theorem for Double Array of Dependent Random Variables:

Before discussing the main results we state a general maximal inequality which will be proved in Appendix C. Consider a double array of real valued stochastic processes $\{Z_{n,i}\} := \{Z_{n,i}(f)\}_{f \in \mathfrak{F}}$, indexed by a class of functions $\mathfrak{F}$, where $n \in \mathbb{N}$ and $i = 1, 2, \ldots, n$. The class $\mathfrak{F}$ is equipped with a metric ρ. Note that the class of functions $\mathfrak{F}$ and the associated metric ρ can depend on n. But for notational convenience we drop the index n here. For fixed n the processes $Z_{n,i}$ will have the same distribution, i.e., $Z_{n,i} \stackrel{d}{=} Z_{n,1}$, but they can be dependent in i. We shall quantify their dependence via β mixing coefficients $\beta_{n,k}$ and mixing rate function $\beta_{(n)}(t)$ defined as in (3.6).

We establish a maximal inequality for the empirical processes corresponding to the double array $Z_{n,i}$, $i = 1, 2, \ldots, n$. The precise rate $\varphi_n(\delta)$, the upper bound in the maximal inequality to be established below, depends on the mixing rate function as well as the tail behavior of the random variables. To make things precise, define $Q_{Z_{n,1}(f)}$, the quantile function of the random variable $Z_{n,1}(f), f \in \mathcal{F}$ to be the inverse function of $t \mapsto \mathbb{P}(|Z_{n,1}(f)| > t)$. Note that for any non-increasing cadlag function $H : \mathbb{R}^+ \to \mathbb{R}^+$, the inverse is defined as $H^{-1}(u) = \sup\{t \in \mathbb{R}^+ : H(t) > u\}$, with the convention that $\sup \emptyset = 0$. Also define the $\|.\|_{2,\beta_{(n)}}$-norm on $\mathfrak{F}$ as

$$\|f\|_{2,\beta_{(n)}} := \sqrt{n \int_0^1 \beta_{(n)}^{-1}(u)(Q_{Z_{n1}(f)}(u))^2 du}. \tag{3.7}$$

This norm is the key ingredient of our analysis as it combines the mixing rate function and tail behavior of the random variables. This norm is motivated by [55] and a very similar version of this was used by [26]. This norm also provides an upper bound on the variance the sum of β-mixing random variables.

Lemma III.3. *If* $\|f\|_{2,\beta_{(n)}} < \infty$, *then* $\mathrm{Var}\Big(\sum_{i=1}^{n} Z_{ni}(f)\Big) \leq 4\|f\|^2_{2,\beta_{(n)}}$.

Proof. By Theorem 1.2 of [55] we have

$$\mathrm{Var}\Big(n^{-1/2}\sum_{i=1}^{n} Z_{ni}(f)\Big) \leq \int_0^1 \beta^{-1}(u)(Q_{Z_{n1}(f)}(u))^2 du. \tag{3.8}$$

Hence the result follows. □

Proposition III.8, which we state later, provides sufficient conditions on $Q_{Z_{n1}}(f)$ and $\beta_{(n)}(.)$ that ensure the $\|.\|_{2,\beta_{(n)}}$-norm is finite.

Remark III.4. For the case of independence the $\|.\|_{2,\beta_{(n)}}$-norm is reduced to essentially the L_2-norm, i.e.,

$$\|f\|_{2,\beta_{(n)}} := \sqrt{n\mathbb{E}(Z^2_{n1}(f))} = \sqrt{\sum_{i=1}^{n}\mathbb{E}(Z^2_{ni}(f))}.$$

Theorem 8. *Assume that the following four conditions hold:*

(A1)For all $f \in \mathfrak{F}$, $\sup_n \|f\|_{2,\beta_{(n)}} < \infty$

(A2) The mixing coefficients satisfy the summability condition, i.e., $\sum_k \beta_{n,k} < \infty$ *for all* n.

(A3) For any sequence $\delta_n \to 0$ *as* $n \to \infty$, *we have*

$$\lim_{n\to\infty} \sup_{\rho(f,g)<\delta_n} \|f-g\|_{2,\beta_{(n)}} = 0.$$

(A4)Let $\mathfrak{F}_{n,\delta} := \{f - g \in \mathfrak{F} : f, g \in \mathfrak{F}, \|f-g\|_{2,\beta_{(n)}} < \delta\}$ *and the bracketing entropy integral*

$$\varphi_n(\delta) := \int_0^\infty \sqrt{\log N_{[]}(\epsilon, \mathfrak{F}_{n,\delta}, \|.\|_{2,\beta_{(n)}})d\epsilon} < \infty$$

for all n *and for all* $\delta > 0$.

Then, for sufficiently large n, *there exists a constant* $A > 0$ *(not depending on* n*)*

such that

$$\mathbb{E} \sup_{h \in \mathfrak{F}_{n,\delta}} \left| \sum_{i=1}^{n} (Z_{ni}(h) - \mathbb{E} Z_{ni}(h)) \right| < 2A\varphi_n(\delta).$$

Remark III.5. This is an extension of Theorem 1 of [26] to the double array set-up. The proof relies on Barbee's Lemma (Lemma III.2) and chaining arguments.

Remark III.6. Note that the distribution of Z_{ni} has to depend on n in order to satisfy assumption (A1). For example, for a sequence of i.i.d. random variables Z_i, if we define $Z_{ni} := Z_i/\sqrt{n}$ then (A1) is satisfied.

It is difficult to compute the $2, \beta_{(n)}$ norm directly. The following propositions allow us to establish connections between the L_{2r} or L_2 norm with the $2, \beta_{(n)}$ norm.

Proposition III.7. *The L_{2r} norm and $2, \beta_{(n)}$ norm are related as below:*

(i) If the mixing coefficients are summable, i.e., they satisfy (A2) and $\|f\|_{2,\beta_{(n)}} < \infty$ then $\|\sqrt{n} Z_{n1}(f)\|_2 \leq \|f\|_{2,\beta_{(n)}}$.

(ii) If the mixing coefficients satisfy (M1) (polynomial decay) then $\|f\|_{2,\beta_n} \leq C_1 \|\sqrt{n} Z_{n1}(f)\|_{2r}$ for a constant $C_1 > 0$ and L and r are related as $L > r/(r-1)$.

(iii) If the mixing coefficients satisfy (M2) (exponential decay) then $\|f\|_{2,\beta_{(n)}} \leq C_2 \|\sqrt{n} Z_{n1}(f)\|_2$ for some constant $C_2 > 0$.

Proof. The first assertion is true by Lemma 1 of [26] and (ii) and (iii) follow from Holder's inequality. □

Corollary III.8. *Assumption (A1) is satisfied if either of the following holds:*

(a) For all $f \in \mathcal{F}$, we have $\sup_n \|\sqrt{n} Z_{n1}(f)\|_{2r} < \infty$ and $\beta_{(n)}(s) \leq C s^{-L}$, where $L > r/(r-1)$ for some $r > 1$.

(b) For all $f \in \mathcal{F}$, we have $\sup_n \|\sqrt{n} Z_{n1}(f)\|_2 < \infty$ and $\beta_{(n)}(s) \leq b^s$ for some $b \in (0, 1)$.

Corollary III.9. *If either (a) or (b) of Corollary III.8 holds then* $\|f\|_{2,\beta_{(n)}} \to 0$ *as* $n \to \infty$ *if and only if* $\|\sqrt{n}Z_{n1}(f)\|_2 \to 0$ *as* $n \to \infty$*.*

Proof. The "only if" part follows from Proposition III.7 (i).

If (b) holds the "if" part follows from Proposition III.7 (iii).

To prove the "if" direction for (a), we shall show that under our assumption there exists some $r' < r$ such that $\|f\|_{2,\beta_{(n)}} \le K\|\sqrt{n}Z_{n1}(f)\|_{2r'}$. This along with $\|\sqrt{n}Z_{n1}(f)\|_2 \to 0$ as $n \to \infty$ imply that $\|\sqrt{n}Z_{n1}(f)\|_{2r'} \to 0$ as $n \to \infty$. To this end, note that for $r' < r$ by Holder inequality we can write

$$\int_0^1 \beta_{(n)}^{-1}(u)Q^2(u)du \le \left(\int_0^1 Q^{2r'}(u)\right)^{1/r'} \left(\int_0^1 \left(\beta_{(n)}^{-1}(u)\right)^{r'/r'-1}\right)^{r'-1/r'}.$$

Choose $r' < r$ such that $L > r'/(r'-1) > r/(r-1)$. The first integral above is just $[\mathbb{E}(Z_{n1}(f)^{2r'})]^{1/r'}$ and the second integral is finite because $\beta_{(n)}(s) \le Cs^{-L}$ implies $\beta_{(n)}^{-1}(u) \le c^{1/L}u^{-1/L}$. Therefore we have $\|f\|_{2,\beta_{(n)}} \le K\|\sqrt{n}Z_{n1}(f)\|_{2r'}$. Also $\sup_n \|\sqrt{n}Z_{n1}(f)\|_{2r} < \infty$ implies that $\{(\sqrt{n}Z_{n1}(f))^{2r'}\}_n$ is uniformly integrable and $\|\sqrt{n}Z_{n1}(f)\|_2 \to 0$ as $n \to \infty$ implies that $\sqrt{n}Z_{n1}(f) \xrightarrow{p} 0$ as $n \to \infty$. Therefore we have $\|\sqrt{n}Z_{n1}(f)\|_{2r'} \to 0$ as $n \to \infty$. This completes the proof. □

Remark III.10. By Proposition III.7 and Corollary III.9 if the mixing coefficients satisfy (M1) or (M2) the $\|.\|_{2,\beta_{(n)}}$-norm is asymptotically equivalent to $\sqrt{n}\|.\|_2$. Therefore the the class of functions $\mathcal{F}_{n,\delta}$ and entropy integral $\varphi_n(.)$ can be expressed in terms of $\|.\|_2$ norm which is easier to work with. Therefore, with a small change of variable, we restate Assumption (A4) as

(A4') For $\mathcal{F}_{n,\delta} := \sqrt{n}\{f - g : f, g \in \mathfrak{F}, \|f-g\|_2 < \delta\}$, the bracketing entropy integral

$$\varphi_n'(\delta) := \int_0^\infty \sqrt{\log N_{[]}(\epsilon, \mathcal{F}_{n,\delta}, \|.\|_2)}d\epsilon < \infty$$

for all $\delta > 0$.

Under (A1)-(A3) and (A4') we have for large enough n,

$$\mathbb{E} \sup_{h \in \mathcal{F}_{n,\delta}} \left| \sum_{i=1}^{n} (Z_{ni}(h) - \mathbb{E}Z_{ni}(h)) \right| < 2A\varphi_n'(\delta). \tag{3.9}$$

Note that by dominated convergence theorem $\varphi_n(\delta) \to 0$ as $\delta \to 0$. Therefore, Theorem 8 provides equicontinuity for the process $\{\sum_{i=1}^{n}(Z_{ni}(f) - \mathbb{E}(Z_{ni}(f)))\}_{f\in\mathfrak{F}}$ and hence we have the following Corollary:

Corollary III.11. *Under the assumptions of Theorem 8 $\{\sum_{i=1}^{n}(Z_{ni}(f) - \mathbb{E}(Z_{ni}(f)))\}_{f\in\mathfrak{F}}$ is uniformly asymptotically equicontinuous and converges as a process provided the finite dimensional distributions converge.*

To obtain the convergence of finite dimensional distributions we state a generalized version of the Lindeberg Theorem which is Corollary 1 from [56].

Theorem 9. *Let $\{X_{ni}\}_{n\in\mathbb{Z},1\le i\le n}$ be a double array of real valued random variables with finite variance and mean zero. Let $\beta_{n,k}$ be the sequence of beta mixing coefficients of the sequence $\{X_{ni}\}_{i=1}^{n}$ and $\beta_{(n)}^{-1}$ be the inverse function of the associated mixing rate function defined as $\beta_{(n)}(t) = \beta_{n,\lfloor t\rfloor}$. We set*

$$S_{ni} = X_{n1} + X_{n2} + \cdots + X_{ni} \textit{ and } V_{ni} = \mathrm{Var(S_{ni})}.$$

Suppose furthermore that

$$\limsup_{n\to\infty} \max_{i\in\{1,2,\ldots n\}} (V_{ni}/V_{nn}) < \infty \tag{3.10}$$

and let $Q_{ni} := Q_{X_{ni}}$ denote the quantile function of X_{ni}. Then, $S_{nn}/\sqrt{V_{nn}}$ converges to a standard normal distribution if

$$V_{nn}^{-3/2} \sum_{i=1}^{n} \int_0^1 \beta_{(n)}^{-1}(x) Q_{ni}^2(x) \min(\beta_{(n)}^{-1}(x) Q_{ni}(x), \sqrt{V_{nn}}) dx \to 0 \tag{3.11}$$

as n tends to ∞.

3.2.2 Consistency

First we show the M-estimator $\hat{\theta}_n$ is consistent for the true parameter θ_n.

Theorem 10. *Assume that $\hat{\theta}_n$ is tight and either of the following two conditions holds:*

(C1) $\sup_n \|m_{n,\theta}(X_{n1})\|_{2r} < \infty$ for all $\theta \in \Theta_n$ and some $r > 1$ and the β mixing rates satisfy condition (M1) (i.e, polynomial rate of decay) with $L > r/(r-1)$.

(C2) $\sup_n \|m_{n,\theta}(X_{n1})\|_2 < \infty$ for all $\theta \in \Theta_n$ and the β mixing rates satisfy condition (M2) (i.e, exponential rate of decay).

Also let $\phi_n(\delta)$ as defined in (3.4) be finite for all $\delta > 0$ for all n. Then $d_n(\hat{\theta}_n, \theta_n)$ converges to zero in probability as $n \to \infty$.

Proof. Define $Z_{ni}(\theta) = \frac{1}{\sqrt{n}}(m_{n,\theta}(X_{ni}) - \mathbb{E}m_{n,\theta}(X_{ni})$. First note that for a fixed θ we can apply Theorem 9 to conclude that the finite dimensional distributions $\sum_{i=1}^n Z_{ni}(\theta)$ converges to a normal distribution. Indeed (C1) (or (C2)) implies $V_{nn} = \mathrm{Var}(\sum_{i=1}^n Z_{ni}(\theta))$ is bounded uniformly for all n and by (3.8)

$$\frac{V_{ni}}{V_{nn}} = \frac{i}{nV_{nn}} \int_0^1 \beta_{(n)}^{-1}(u) Q^2_{m_{n,\theta}(X_{n1})}(u)du.$$

The last integral is finite by Holder inequality and hence (3.10) is satisfied. To check (3.11) note that in this case, $\sum_{i=1}^n \int_0^1 \beta_{(n)}^{-1}(x) Q_{ni}^2(x) = \|\theta\|_{2,\beta_{(n)}} < \infty$ by Proposition III.8 and $\beta_{(n)}^{-1}(x) Q_{ni}(x) \to 0$ for all $x \in (0,1)$ as $n \to \infty$. Therefore an application of the Dominated Convergence Theorem gives us assumption (3.11).

Next we apply Theorem 8 to obtain a uniform central limit theorem for $Z_{ni}(\theta)$ for $\theta \in K$, where K is any compact subset of Θ_n. Observe that by Proposition III.8

(A1) is satisfied. Assumptions (A2)-(A4) follow trivially from the assumptions of this theorem. Hence we have our desired Central Limit Theorem (CLT), which in addition with Slutsky's Theorem gives a uniform weak law of large numbers, i.e., we have

$$\sup_{\theta \in K} |\mathbb{P}_n m_{n,\theta} - P_n m_{n,\theta}| \xrightarrow{p} 0 \tag{3.12}$$

for all compact subsets K of Θ_n as $n \to \infty$. Now fix a large compact set K such that $\hat{\theta}_n \in K$ with very high probability, let $\delta > 0$ and write

$$\begin{aligned}
&\mathbb{P}\left(\hat{\theta}_n \notin B(\theta_n, \delta)\right) \\
&\leq \mathbb{P}\left(\sup_{\theta \in B^c(\theta_n,\delta) \cap K} \mathbb{P}_n m_{n,\theta} \geq \sup_{\theta \in B(\theta_n,\delta)} \mathbb{P}_n m_{n,\theta} - o_p(1)\right) \\
&\leq \mathbb{P}\left(\sup_{\theta \in B^c(\theta_n,\delta) \cap K} P_n m_{n,\theta} + \epsilon \geq \sup_{\theta \in B(\theta_n,\delta)} P_n m_{n,\theta} - \epsilon - o_p(1)\right)
\end{aligned}$$

for all small enough $\epsilon > 0$ and $n > N_\epsilon$ by (3.12). But as $\epsilon > 0$ is arbitrary, the last quantity converges to zero by definition of θ_n. Hence $P(d_n(\hat{\theta}_n, \theta_n) > \delta) \to 0$ for all $\delta > 0$ as $n \to \infty$. □

3.2.3 Rate of Convergence

The next result gives the rate of convergence of the M-estimator. Note that rate of convergence is generally defined in terms of d_n, the natural metric on parameter space and d_n is equivalent to the metric induced by L_2 norm by Assumption (D). Therefore the following proposition holds trivially.

Proposition III.12. *Let $\mathcal{M}_{n,\delta}$ be defined as in* (3.3) *and $\mathcal{F}_{n,\delta}$ be defined as in Assumption (A4'). We have $\mathcal{F}_{n,\delta} = \mathcal{M}_{n,\gamma(\delta)}$ for some onto function $\gamma : (0,1) \mapsto (0,1)$ and hence $\varphi_n(\delta) = \phi'_n(\gamma(\delta))$.*

Theorem 11. *Assume that either (C1) or (C2) holds and $P_n m_{n,\theta} - P_n m_{n,\theta_n} \leq -C d_n^2(\theta, \theta_n)$ for all θ in a d_n–neighborhood of θ_n. Let $\phi_n(\delta)$, as* *defined in* (3.4) *with* $r = 1$, *be finite for* $\delta > 0$ *for all* n *and* $\delta \mapsto \phi_n(\delta)/\delta^\alpha$ *be decreasing for some* $\alpha < 2$ *and small* δ. *Then, if* r_n *is such that* $r_n^2\phi_n(1/r_n) \leq \sqrt{n}$ *for sufficiently large* n *we have* $r_n d_n(\hat{\theta}_n, \theta_n) = O_p(1)$.

Proof. We will prove this theorem by checking all the assumptions of Theorem 3.2.5 from [65] under the conditions imposed above. To check the maximal inequality assumption of Theorem 3.2.5 we apply our Theorem 8 with $Z_{ni}(\theta) = \frac{1}{\sqrt{n}}(m_{n,\theta}(X_{ni}) - m_{n,\theta_n}(X_{ni}))$ and $\mathfrak{F} = \Theta_n$. Assumptions (A2) and (A3) are immediate from the statement of this Theorem and (A1) follows from Proposition III.8. By Proposition III.12 and Remark III.10, Assumption (A4) is satisfied and we have our desired maximal inequality with upper bound $2A\phi_n(\delta)/\sqrt{n}$. The consistency of $\hat{\theta}_n$ is guaranteed by Theorem 10. The other assumptions of Theorem 3.2.5 follow trivially from the assumptions of this Theorem and hence Theorem 11 follows. □

Remark III.13. The same result was proved in Theorem 1 from [21] under a slightly less general setting.

For the calculation of the rate of convergence it is essential to calculate the quantity $\phi_n(\delta)$. As there is no direct method of obtaining this quantity, we use the following result from [8] to get a tight bound on the bracketing number and consequently the function ϕ.

Theorem 12. *(Adapted from [8] page 201) If* Θ_n *is a subset of* $\mathbb{R}^d$ *and for all* $\theta \in \Theta_n$

$$\sup_n \left[\mathbb{E} \sup_{\theta_1 \in \Theta_n : \rho_n(\theta_1,\theta) < \delta} |m_{n,\theta}(X_{n1}) - m_{n,\theta_1}(X_{n1})|^p\right]^{1/p} \leq C\delta^\nu \tag{3.13}$$

for all $\delta > 0$ *small, some* $\nu > 0$, $p > 1$ *and* $C > 0$ *not depending on* θ, *then*

$$N_{[]}(\epsilon, \mathcal{M}_{n,\delta}, \|.\|_p) \leq C^*(1/\epsilon)^{d/\nu} \vee 1.$$

Corollary III.14. *Under the conditions of Theorem 12, with $\nu \in (0,1]$, we have $\phi_n(\delta) \leq C_1\delta^\nu$ and consequently the rate of convergence $r_n = n^{1/(4-2\nu)}$.*

Proof. Note that by (3.13) if $\rho_n(\theta_1, \theta_2) < \delta$ then $\|m_{n,\theta_1} - m_{n,\theta_2}\|_p \leq \delta^\nu$ and hence $N_{[]}(\epsilon, \mathcal{M}_{n,\delta}, \|.\|_p) = 1$ for $\epsilon \geq 2\delta^\nu$. Now the first part of the corollary follows by direct integration and the inequality $\log x \leq (x-1)$ for $x \geq 1$. The second part then follows from Theorem 11. □

Remark III.15. This result is also proved in [21] under a different set-up.

3.2.4 Asymptotic Distribution

Our next goal is to derive the limit distribution of $l_n(\hat{\theta}_n - \theta_0)$, where l_n is such that $l_n(\hat{\theta}_n - \theta_0) = O_p(1)$. Note that in the previous section we have derived the rate of convergence in terms of d_n, a general metric on the parameter space. This metric is usually a function of the Euclidean distance of the parameters, therefore the rate of convergence we obtained in terms of d_n translates easily to some rate of convergence with respect to the Euclidean norm. In most of the examples we have discussed below $d_n(\theta_1, \theta_2)$ is either a polynomial function of $|\theta_1 - \theta_2|$ or a scaled version of $|\theta_1 - \theta_2|$. For such simple functions it is easy to calculate l_n such that
$l_n(\hat{\theta}_n - \theta_0) = O_p(1)$.
The derivation of the asymptotic distribution of $l_n(\hat{\theta}_n - \theta_0)$ will be similar to the proof of Theorem 3.2.10 of [65] and hence we will need a dependent version of a triangular functional Central Limit Theorem. The two necessary ingredients for this version of CLT are a dependent version of Lindeberg CLT and extension of the tightness type result, Theorem 2.11.9 from [65], under dependence. The first extension, Theorem 9 is due to [56] and the tightness follows from Theorem 8.
For simplicity of notation in the next theorem we will denote coordinate-wise multiplication and division of two vectors a and b as ab and a/b respectively.

Theorem 13. *Suppose that the β-mixing coefficients of $\{X_{ni}\}_{n\in\mathbb{N},i\le n}$ satisfy (M2). Assume that $l_n(\hat\theta_n-\theta_n)$ is tight and the quantity $\psi_n(1/l_n)$ is bounded uniformly for all n. Define $\tilde m_{n,\theta} := (m_{n,\theta}-m_{n,\theta_n})/\psi_n(1/l_n)$. Furthermore, assume that all functions in the the class $\tilde{\mathcal{M}}_{n,K} := \{\tilde m_{n,\theta} : |\theta-\theta_n| < K/l_n\}$ have finite second moment for all $K>0$ and*

$$\lim_{n\to\infty}\{\sqrt{n}P_n(\tilde m_{n,\theta_n+h/l_n}-\tilde m_{n,\theta_n})\}_{\|h\|\le K}\to\{A(h)\}_{\|h\|\le K}, \tag{3.14}$$

$$\lim_{\epsilon\downarrow 0}\limsup_n \sup_{\substack{\|h-g\|<\epsilon\\ \|h\|\vee\|g\|\le K}} P_n(\tilde m_{\theta_n+g/l_n}-\tilde m_{\theta_n+h/l_n})^2 = 0, \tag{3.15}$$

$$\lim_n P_n(\tilde m_{\theta_n+g/l_n}-\tilde m_{\theta_n+h/l_n})^2 = H(g,h), \tag{3.16}$$

for all positive constants K. Let B be a zero-mean Gaussian process such that $B(g)=B(h)$ almost surely only if $g=h$ and with covariance kernel H and bounded continuous sample paths on compacta. If $B(h)+A(h)\to-\infty$ as $|h|\to\infty$ we have $l_n(\hat\theta_n-\theta_n)$ converges in distribution to the unique maximizer of the process $h\mapsto B(h)+A(h)$.

Proof. Note that $l_n(\hat\theta_n-\theta_n)$ is the maximizer of $h\mapsto\mathbb{P}_n(m_{\theta_n+h/l_n})$ and hence also maximizer of $h\mapsto\sqrt{n}(\mathbb{P}_n(m_{n,\theta_n+h/l_n})-\mathbb{P}_n(m_{n,\theta_n}))/\psi_n(1/l_n)$. We will obtain the limit distribution of the process $\{\sqrt{n}(\mathbb{P}_n(m_{n,\theta_n+h/l_n})-\mathbb{P}_n(m_{n,\theta_n})/\psi_n(1/l_n))\}_{h:\|h\|\le K}$ and apply argmax continuous mapping theorem to obtain the limit distribution of $l_n(\hat\theta_n-\theta_n)$.

In order to do that write

$$\sqrt{n}(\mathbb{P}_n(m_{n,\theta_n+h/l_n})-\mathbb{P}_n(m_{n,\theta_n}))/\psi_n(1/l_n) = I(h)+II(h)$$

where the first term is the centered version of the above process and can be written as

$$
\begin{aligned}
I(h) =& \frac{\sqrt{n}}{\psi_n(1/l_n)} \left((\mathbb{P}_n(m_{n,\theta_n+h/l_n}) - \mathbb{P}_n(m_{n,\theta_n})) - (P_n(m_{n,\theta_n+h/l_n}) - P_n(m_{n,\theta_n})) \right) \\
=& \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left((\tilde{m}_{n,\theta_n+h/l_n}(X_{ni}) - \tilde{m}_{n,\theta_n}(X_{ni})) - \mathbb{E}(\tilde{m}_{n,\theta_n+h/l_n}(X_{ni}) - \tilde{m}_{n,\theta_n}(X_{ni})) \right),
\end{aligned}
$$

and the second term is the following non-random function of h

$$
\begin{aligned}
II(h) =& \frac{\sqrt{n}}{\psi_n(1/l_n)} \left(P_n(m_{n,\theta_n+h/l_n}) - P_n(m_{n,\theta_n}) \right) \\
=& \frac{1}{\sqrt{n}} \left(P_n(\tilde{m}_{n,\theta_n+h/l_n}) - P_n(\tilde{m}_{n,\theta_n}) \right).
\end{aligned}
$$

Note that by (3.14) the second term $II(h)$ converges in distribution to $A(h)$ as a process in h uniformly on compacta. Next we will show that the first term is asymptotically tight by applying Theorem 8 with $Z_{ni}(h) = (\tilde{m}_{\theta_n+h/l_n}(X_{ni}) - \tilde{m}_{\theta_n}(X_{ni}))/\sqrt{n}$. Note that by Proposition III.1 the β-mixing coefficients of $\{Z_{ni}\}_{i=1}^{n}$ for all $n \in \mathbb{N}$ are dominated by that of $\{X_{ni}\}_{i=1}^{n}$. So Assumption (A2) is satisfied. Next we check that the norm defined in (3.7) is finite in this situation for $\|h\| \leq K$. Note that here the class of functions $\mathfrak{F} = 1/\sqrt{n}\tilde{\mathcal{M}}_{n,K}$ and therefore (A1) is satisfied by Proposition III.7 (iii) and finite second moment assumption for the class $\tilde{\mathcal{M}}_{n,K}$. Next we check assumption (A4'). To do that we use a change of variable technique. Following the notations of assumption (A4') here we have $\mathfrak{F}_{n,\delta} = \tilde{\mathcal{M}}_{n,\delta} = [1/\psi_n(l_n)]\mathcal{M}_{n,\delta/l_n}$. Recall that $\mathcal{M}_{n,\delta} := \{m_{n,\theta} - m_{n,\theta_n} : |\theta - \theta_n| \leq \delta\}$ and $\psi_n(\delta) := \int_0^\infty \left[\log N\left(\epsilon, \mathcal{M}_{n,\delta}, \|.\|_2\right)\right]^{1/2} d\epsilon$.

Indeed we have

$$\begin{aligned}
\phi_n(\delta) &= \int_0^\infty \left[\log N\left(\epsilon, \frac{1}{\psi_n(1/l_n)}\mathcal{M}_{n,\delta/l_n}, \|.\|_2\right)\right]^{1/2} d\epsilon \\
&= \int_0^\infty \left[\log N\left(\psi_n(1/l_n)\epsilon, \mathcal{M}_{n,\delta/l_n}, \|.\|_2\right)\right]^{1/2} d\epsilon \\
&= \frac{1}{\psi_n(1/l_n)} \int_0^\infty \left[\log N\left(\epsilon, \mathcal{M}_{n,\delta/l_n}, \|.\|_2\right)\right]^{1/2} d\epsilon \\
&= \frac{\psi_n(\delta/l_n)}{\psi_n(1/l_n)} \leq 1
\end{aligned}$$

Finally (3.15) implies Assumption (A3). So, by Theorem 8, we have our desired asymptotic equicontinuity.

Finally we check the conditions of Theorem 9 to get finite dimensional convergence of the process $\{I(h)\}_{h:\|h\|\leq K}$. Note that in this case, V_{nn} is bounded by the $2, \beta_{(n)}$ norm defined in (3.7) and that is finite by our earlier discussion. The condition (3.11) is translated here as

$$\int_0^1 \beta_{(n)}^{-1}(u) Q^2_{f_n(X_{ni})}(u) \inf_n \{n^{-1/2}\beta^{-1}(u) Q_{f_n(X_{ni})}(u), 1\} du \to 0$$

as $n \to \infty$, where f_n is any function of the class $\mathfrak{F} = \mathcal{M}_{n,K}/\sqrt{n}$. Now for each fixed $u \in (0,1)$ the quantity $n^{-1/2}\beta^{-1}(u)Q_{f_n(X_ni)}(u) \to 0$ as $n \to \infty$ and $\int_0^1 \beta_{(n)}^{-1}(u)Q^2_{f_n(X_{ni})}(u) < \infty$ for each n. Therefore the Dominated Convergence Theorem implies (3.11). Hence the process $\{I(h)\}_h$ converges to a zero mean Gaussian process uniformly on compacta.

So the process $\{\sqrt{n}(\mathbb{P}_n(m_{n,\theta_n+h/l_n}) - \mathbb{P}_n(m_{n,\theta_n})/\psi_n(1/l_n))\}_{h:\|h\|\leq K}$ converges in distribution to a Gaussian process and by (3.14) and (3.16), the mean and covariance of the limiting process are $A(\dot{})$ and $H(\dot{,})$ respectively. Therefore the limit process can be written as $B(h) + A(h)$ where B is as defined in the statement of the

Theorem. Lemma 2.6 from [40] implies that the limit process in fact has a unique maximum. Now, we apply the Argmax Continuous Mapping (Theorem 3.2.2 from [65]) to complete the proof. $\square$

Remark III.16. The assumption that $B(h) + A(h) \to -\infty$ as $|h| \to \infty$ is satisfied if

$$P\left(\lim \sup_{|h|\to\infty} \frac{B(h)}{A(h)} > \epsilon\right) = 0.$$

For most of the applications $A(h)$ is generally a polynomial in $|h|$ and the last equality holds for all such $A(h)$. The proof is similar to the proof of Lemma 2.5 from [40].

Remark III.17. The kernel H of (3.16) is calculated as the limit of the covariance kernel of the process $\{\sqrt{n}(\mathbb{P}_n(m_{n,\theta_n+h/l_n}) - \mathbb{P}_n(m_{n,\theta_n}))/\psi_n(1/l_n)\}_h$.

3.3 Applications

In this section we apply our theory to some specific models to obtain the asymptotic properties of corresponding M-estimators. Though our theory is applicable for triangular array data, for notational convenience in this section we focus on simpler models with data coming from a single array of random variables. In most of the examples (if nothing is mentioned) we have assumed the covariates U_i are independent and the errors ϵ_i come from a dependent time series. We also assume that U_i and ϵ_i are independent of each other. Our theory allows us to analyze several other scenario like U_i dependent on ϵ_i or (U_i, ϵ_i) coming from a 2 dimensional dependent time series etc. The calculation of covariance kernel will differ in these different cases giving different limiting distribution. All the other details will be similar and hence we will have same rate of convergence in all these scenarios for all the examples illustrated below.

The rate of convergence calculation is valid under either one of the following assumptions:

(i) $\mathbb{E}|\epsilon_1|^{2r} < \infty$ for some $r > 1$ and the β-mixing coefficients satisfy (M1) (polynomial mixing rate) with $L > r/(r-1)$.

(ii) $\mathbb{E}|\epsilon_1|^2 < \infty$ and the β-mixing coefficients satisfy (M2) (exponential mixing rate).

The limiting distribution of M-estimator is calculated under assumption (ii).

3.3.1 Inverse of a Monotone Function

We consider the problem of estimating inverse of a monotone function (See [15]). We have data (U_i, Y_i) for $i = 1, 2, \ldots n$ with $Y_i = \mu(U_i) + \epsilon_i$, where μ is a smooth monotone increasing function on $[0, 1]$ with $\int_0^1 \mu(u)du < \infty$. We want to estimate $\theta_0 = \mu^{-1}(\tau_0)$ for some fixed level τ_0. We assume the design points U_i's are i.i.d. Uniform(0,1). The M-estimator $\hat{\theta}_n$ can be obtained by minimizing the criterion function

$$\mathbb{M}_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \tau_0)\,\mathbf{1}(U_i \le \theta).$$

Let

$$\mathbb{M}(\theta) = \mathbb{E}\mathbb{M}_n(\theta) = \int_0^\theta (\mu(u) - \tau_0)dx.$$

Note that $\mathbb{M}'(\theta) = \mu(\theta) - \tau_0$, so $\mathbb{M}'(\theta_0) = 0$, and due to the monotonicity of μ, we have $\mathbb{M}'(\theta) < 0$ for $\theta < \theta_0$ and $\mathbb{M}'(\theta) > 0$ for $\theta > \theta_0$. Therefore $\mathbb{M}$ has a unique minimum at θ_0. We assume for some integer $k > 1$, $\mu^{(j)}(\theta_0) = 0$ for all $j = 1, 2, \ldots, k-1$ and $\mu^{(k)}(\theta_0) > 0$.

Rate of Convergence: To obtain the rate of convergence of the estimator, we apply Theorem 11 with $m_{n,\theta}(U_i, \epsilon_i) := (\mu(U_i) + \epsilon_i - \tau_0)\,\mathbf{1}(U_i \le \theta)$ and $\theta_n \equiv \theta_0$. Define the metric $d(\theta_1, \theta_2) := |\theta_1 - \theta_2|^{(k+1)/2}$ on $\Theta_n \equiv \Theta := [0, 1]$. Note that $\mathbb{M}^{(l)}(\theta) = \mu^{(l-1)}(\theta)$, so the first k derivatives of $\mathbb{M}$ vanish at θ_0 and $(k+1)$-th derivative is positive at θ_0. So we have for every θ in a neighbourhood of θ_0,

$$\mathbb{M}(\theta) - \mathbb{M}(\theta_0) = \frac{\mu^{(k)}(\tilde{\theta})}{(k+1)!}|\theta - \theta_0|^{(k+1)} \geq Kd^2(\theta, \theta_0)$$

for some $\tilde{\theta}$ in between θ_0 and θ and an appropriate constant $K > 0$. Next, we use Corollary III.14 to calculate $\phi_n(\delta)$ under this setup. Without loss of generality assume $\theta_2 < \theta_1$. For $0 < \delta < 1$ and $\eta = \delta^{2/(k+1)}$ we can write

$$\begin{aligned}
&\mathbb{E} \sup_{\theta_1 : d(\theta_1, \theta_2) < \delta} \left(m_{n,\theta_1}(U_1, \epsilon_1) - m_{n,\theta_2}(U_1, \epsilon_1)\right)^2 \\
\leq &\mathbb{E} \sup_{\theta_1 : |\theta_1 - \theta_2| \leq \eta} (\mu(U_1) + \epsilon_1 - \tau_0)^2 \mathbf{1}(\theta_2 < U_1 < \theta_1) \\
\leq &\left[\sup_{u \in [0,1]} (\mu(u) - \tau_0)^2 + \sigma^2\right] \eta = C\delta^{2/(k+1)},
\end{aligned} \tag{3.17}$$

where $C = \sup_{u \in [0,1]} (\mu(u) - \tau_0)^2 + \sigma^2$. Therefore by Corollary III.14 for all sufficiently small $\delta > 0$ we have $\phi_n(\delta) \leq K\delta^{1/(k+1)}$ for some constant K not depending on δ. Furthermore we have $r_n d(\hat{\theta}_n, \theta_0) = O_p(1)$, where $r_n = n^{1/(4-2/(k+1))} = n^{(k+1)/2(2k+1)}$. In terms of the usual Euclidean metric, this translates to we obtain $n^{1/(2k+1)}|\hat{\theta}_n - \theta_0| = O_p(1)$.

Asymptotic Distribution: To investigate the asymptotic distribution of $\Delta_n := n^{1/(2k+1)}(\hat{\theta}_n - \theta_0)$ we use Theorem 13 with $l_n = n^{1/(2k+1)}$. Calculations similar to (3.17) gives $\psi_n(1/l_n) \sim 1/\sqrt{l_n} = n^{-1/2(2k+1)}$. Note that $\mathbb{E}(m_{n,\theta} - m_{n,\theta_0})^2 \leq C|\theta - \theta_0|$ and therefore the functions of the class $\mathcal{M}_n$ defined in Theorem 13 have bounded second moments. Taylor expansion yields that (3.14) holds with

$$A(h) = \frac{1}{(k+1)!}\mu^{(k)}(\theta_0)|h|^{k+1}.$$

Checking (3.15) is again similar to (3.17). Finally by Remark III.17, to check (3.16) we calculate limiting covariance of the process $\{\sum_{i=1}^n Q_{ni}(h)\}_h$ where

$$Q_{ni}(h) := \frac{n^{1/2(2k+1)}}{\sqrt{n}} (\mu(U_i) + \epsilon_i - \tau_0) \left[\mathbf{1}\left(U_i \leq \theta_0 + \frac{h}{n^{1/(2k+1)}} \right) - \mathbf{1}\left(U_i \leq \theta_0 \right) \right].$$

First note that if $h_1 < 0 < h_2$, we have $\mathrm{Cov}(\sum_i Q_{ni}(h_1), \sum_i Q_{ni}(h_2)) = 0$. Now without loss of generality assume $h_1, h_2 > 0$ and using the fact $\mathbb{E}(\epsilon_i) = 0$ and independence of U_i and ϵ_i we can write the covariance kernel as

$$\begin{aligned}
&\mathrm{Cov}\left(\sum_{i=1}^n Q_{ni}(h_1), \sum_{i=1}^n Q_{ni}(h_2) \right) \\
&= \left(\frac{n^{1/2(2k+1)}}{n} \right)^2 \left[\sum_i \mathbb{E}(\mu(U_i) - \tau_0)^2 \mathbf{1}\left(\theta_0 \leq U_i \leq \theta_0 + \frac{\min(h_1, h_2)}{n^{1/(2k+1)}} \right) \right. \\
&\quad - \sum_i \left[\mathbb{E}(\mu(U_i) - \tau_0) \mathbf{1}\left(\theta_0 \leq U_i \leq \theta_0 + \frac{h_1}{n^{1/(2k+1)}} \right) \right]^2 \\
&\quad + \sum_i \mathbb{E}\epsilon_i^2 \mathbf{1}\left(\theta_0 \leq U_i \leq \theta_0 + \frac{\min(h_1, h_2)}{n^{1/(2k+1)}} \right) \\
&\left. + \sum_{i \neq j} \mathrm{Cov}\left(\epsilon_i \mathbf{1}\left(\theta_0 \leq U_i \leq \theta_0 + \frac{h_1}{n^{1/(2k+1)}} \right), \epsilon_j \mathbf{1}\left(\theta_0 \leq U_j \leq \theta_0 + \frac{h_2}{n^{1/(2k+1)}} \right) \right) \right] \\
&= n^{-2k/(2k+1)} \left[n \int_{\theta_0}^{\theta_0 + \frac{\min(h_1,h_2)}{n^{1/(2k+1)}}} (\mu(u) - \tau_0)^2 du - n \left(\int_{\theta_0}^{\theta_0 + \frac{\min(h_1,h_2)}{n^{1/(2k+1)}}} (\mu(u) - \tau_0) du \right)^2 \right. \\
&\quad \left. + n\sigma^2 \frac{\min(h_1, h_2)}{n^{1/(2k+1)}} + n \left(\frac{1}{n} \sum_{i \neq j} \mathrm{Cov}(\epsilon_i, \epsilon_j) \right) \frac{h_1 h_2}{n^{2/(2k+1)}} \right] \\
&= n^{1/(2k+1)} \left[\left(\frac{\mu^{(k)}(\theta_0)}{k!} \frac{\min(h_1^k, h_2^k)}{n^{k/(2k+1)}} + o(1/n^{k/(2k+1)}) \right)^2 [n^{-1/(2k+1)} - n^{-2/(2k+1)}] \right. \\
&\quad \left. + \sigma^2 \frac{\min(h_1, h_2)}{n^{1/(2k+1)}} + O(1/n^{2/(2k+1)}) \right] \\
&= \sigma^2 \min(h_1, h_2) + o(1). \qquad (3.18)
\end{aligned}$$

So the covariance kernel can be written as $H(h_1, h_2) = \sigma^2 \min(|h_1|, |h_2|)$ if $\mathrm{sign}(h_1) = \mathrm{sign}(h_2)$ and it is 0 otherwise. This yields $B(h) := \sigma^2 W(h)$ and hence by

Theorem 13 we have

$$n^{1/(2k+1)}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \arg\min_h \left[\sigma^2 W(h) + \mu^{(k)}(\theta_0)|h|^{k+1}\right] \tag{3.19}$$

as $n \to \infty$.

Remark III.18. Note that in this example $\mathbb{E}((\sqrt{n}Q_{ni}(h)^2) = O(1)$, but for any $r > 1$, the expectation $\mathbb{E}((\sqrt{n}Q_{ni}(h))^{2r}) \to infty$ as $n \to \infty$. Therefore the moment assumption of Lemma III.7(ii) does not hold in this case. In fact this is true for deriving the asymptotic distribution for all the examples discussed here. Therefore we derive the asymptotic distribution of M-estimator only under (M2), i.e., exponential rate of β-mixing.

We also discuss an interesting variant of this model called minimum effective dose identification problem.

Minimum Effective Dose Model: Here the data (U_i, Y_i) for $i = 1, 2, \ldots, n$ comes from the model $Y_i = \mu(U_i) + \epsilon_i$ with U_i's i.i.d. uniform(0,1). The function μ is continuous on $[0, 1]$ with $\mu(x) = \tau_0$ for $x \leq \theta_0$ and $\mu(x) > \tau_0$ for $x > \theta_0$. The function is assumed to be monotone increasing in $[\theta_0, \theta_0 + \epsilon_0]$ and can behave erratically otherwise. Let $\mu^{(j)}(\theta_0+) = 0$ for $j = 1, 2, \ldots, k-1$, $\mu^{(k)}(\theta_0+) > 0$ and $\mu^{(k)}$ be continuous on $[\theta_0, \theta_0 + \epsilon_0]$. Our goal is to estimate θ_0. We assume that τ_0 is known and we define $\hat{\theta}_n$ to be the minimizer of

$$\mathbb{M}_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \tau_n)\mathbf{1}(U_i \leq \theta)$$

where $\tau_n = \tau_0 + c_0 n^{-\gamma}$, with $\gamma \geq k/(2k+1)$. Now let

$$M_n(\theta) = \mathbb{E}\mathbb{M}_n(\theta) = \int_0^\theta (\mu(x) - \tau_n)dx.$$

Then we have $M_n'(\theta) = \mu(x) - \tau_n$. Since τ_n converges to τ_0 , by the monotonicity of μ on $[\theta_0, \theta_0 + \epsilon_0]$ for all sufficiently large n, there is a unique solution θ_n to $M_n'(\theta) = 0$ that lies in $[\theta_0, \theta_0 + \epsilon_0]$ and $\theta_n \to \theta_0$ as $n \to \infty$. Taylor expansion shows that $n^{\gamma/k}(\theta_n - \theta_0) \to (k!c_0/\mu^{(k)}(\theta_0))^{1/k}$ as $n \to \infty$. So to study the asymptotics of $\hat{\theta}_n$, we will investigate the behavior of $\hat{\theta}_n - \theta_n$. The analysis is similar to the previous example.

Rate of Convergence: Define a metric ρ_n on $\Theta_n \equiv \Theta = [0, 1]$ as follows:

$$\rho_n^2(\theta_1, \theta_2) = \begin{cases} n^{-\gamma}|\theta_1 - \theta_2|, & \text{if } \theta_1, \theta_2 < \theta_0 \\ |\theta_1 - \theta_2|^{k+1}, & \text{if } \theta_1, \theta_2 \geq \theta_0 \\ \rho_n(\theta_1, \theta_0) + \rho_n(\theta_0, \theta_2), & \text{if } \theta_1 < \theta_0 \leq \theta_2. \end{cases}$$

Note that as in Example 1, $M_n^{(l)}(\theta) = \mu^{(l-1)}(\theta)$, so we have for every $\theta > \theta_0$ in a neighbourhood of θ_n,

$$M_n(\theta) - M_n(\theta_n) = \int_{\theta_n}^{\theta} (\mu(x) - \tau_n)dx = \sum_{l=1}^{k} \frac{\mu^{(l)}(\theta_n)}{l!}(\theta_n - \theta)^{(l+1)}.$$

In fact as μ is increasing at θ_0, the first $(k-1)$ derivatives at θ_n are positive. Also $\mu^{(k)}(\theta_n) \to \mu^{(k)}(\theta_0)$ as $n \to \infty$, we finally have $M_n(\theta) - M_n(\theta_n) \geq const.|\theta_n - \theta|^{(k+1)}$. If $\theta < \theta_0$, we can write

$$M_n(\theta) - M_n(\theta_n) = \int_{\theta}^{\theta_0} c_0 n^{-\gamma} dx + \int_{\theta_0}^{\theta_n} (\mu(x) - \tau_n)dx \geq const.[n^{-\gamma}(\theta_0 - \theta) + |\theta_0 - \theta_n|^{(k+1)}],$$

as $\theta_n > \theta_0$ and is very close to θ_0 for large n. So for all θ in a neighbourhood of θ_n, we have $M_n(\theta_n) - M_n(\theta) \geq \rho_n^2(\theta, \theta_n)$. Next we calculate a bound for $\phi_n(\delta)$ via Theorem 12 by calculating the quantity $\mathbb{E} \sup_{\theta_1:\rho_n(\theta_1,\theta_2)\leq\delta} (m_{n,\theta_1}(U, \epsilon) - m_{n,\theta_2}(U, \epsilon))^2$, where

$$m_{n,\theta_1}(U,\epsilon) := (\mu(U) + \epsilon - \tau_n)\mathbf{1}(U \leq \theta).$$

Note that for $\theta_1, \theta_2 > \theta_0$ with $\rho_n(\theta_1, \theta_2) \leq \delta$, we have

$$\begin{aligned} &\mathbb{E} \sup_{\theta_1 : \rho_n(\theta_1,\theta_2) \leq \delta} (m_{n,\theta_1}(U,\epsilon) - m_{n,\theta_2}(U,\epsilon))^2 \\ &\leq \left[\sup_{u \in [0,1]} (\mu(u) - \tau_0)^2 + \sigma^2 \right] \mathbb{E}\mathbf{1}(\theta < U < \theta + \delta^{2/(k+1)}) = C\delta^{2/(k+1)} \end{aligned} \tag{3.20}$$

and for $\theta_1, \theta_2 < \theta_0$ with $\rho_n(\theta_1, \theta_2) \leq \delta$,

$$\begin{aligned} &\mathbb{E} \sup_{\theta_1 : \rho_n(\theta_1,\theta_2) \leq \delta} (m_{n,\theta_1}(U,\epsilon) - m_{n,\theta_2}(U,\epsilon))^2 \\ =&[(\tau_0 - \tau_n)^2 + \sigma^2]\mathbb{E}\mathbf{1}(\theta < U < \theta + n^{\gamma}\delta^2) \leq C(n^{-\gamma}\delta^2 + n^{\gamma}\delta^2) \end{aligned} \tag{3.21}$$

for some constant C. The case of $\theta_1 < \theta_0 < \theta_2$ can be treated by dividing the integral in two parts and which finally gives us

$$\phi_n(\delta) \sim A\sqrt{\delta^{2/(k+1)} + n^{-\gamma}\delta^2 + n^{\gamma}\delta^2}.$$

Note that $\phi_n(\delta)/\delta^{\alpha}$ is decreasing for $1 < \alpha < 2$. Hence using Theorem 11 we have if $r_n^2\phi_n(1/r_n) \leq \sqrt{n}$, then $r_n\rho_n(\hat{d}_n, d_n) = O_p(1)$. This given

$$r_n^2 n^{-\gamma} + r_n^2 n^{\gamma} + r_n^{4-2/(k+1)} \leq n.$$

Simplifying the last expression we obtain
$r_n^2 = n^{(k+1)/(2k+1)} \wedge n^{1-\gamma} \wedge n^{1+\gamma} = n^{(k+1)/(2k+1)} \wedge n^{1-\gamma}$, as $\gamma > 0$. The choice of $\gamma \geq k/(2k+1)$, gives $r_n^2 = n^{(k+1)/(2k+1)}$. Therefore if $\hat{\theta}_n > \theta_0$, we have $n^{1/(2k+1)}|\hat{\theta}_n - \theta_n| = O_p(1).$, If $\hat{\theta}_n < \theta_0$, this implies that, $n^{(k+1)/(2k+1)-\gamma}|\hat{\theta_n} - \theta_0| = O_p(1)$ and $n^{1/(2k+1)}|\theta_n - \theta_0| = O_p(1)$. If we choose

$\gamma = k/(2k+1)$, this gives us $n^{1/(2k+1)}|\hat{\theta_n} - \theta_n| = O_p(1)$.

Asymptotic Distribution: We take $\gamma = k/(2k+1)$ and want to obtain the asymptotic distribution of $l_n(\hat{\theta}_n - \theta_n)$, with $l_n = n^{1/(2k+1)}$. Recall that by earlier discussion $K_n := l_n(\theta_n - \theta_0) \to c := (k!c_0/\mu^{(k)}(\theta_0))^{1/k}$. Calculations similar to (3.20) and (3.21) give $\psi_n(\delta) \leq C\sqrt{(\delta + n^{2k/(2k+1)}\delta)}$ and hence $\psi_n(1/l_n)\ n^{-1/2(2k+1)}$. The finiteness of second moments of functions from the class $\mathcal{M}_n$ defined in Theorem 13 can be checked by direct calculation. To check (3.14) note that

$$P_n(\tilde{m}_{n,\theta_n+h/l_n} - \tilde{m}_{n,\theta_n})/\sqrt{n} = n^{(k+1)/(2k+1)}(M_n(\theta_n + h/r_n) - M_n(\theta_n)).$$

For $h \geq -c$ by Taylor expansion, we have

$$n^{(k+1)/(2k+1)}(M_n(\theta_n + h/r_n) - M_n(\theta_n)) \xrightarrow{d} \frac{\mu^{(k)}(\theta_0+)|h|^{k+1}}{(k+1)!}$$

as $n \to \infty$ and for $h < -c$,

$$\begin{aligned} n^{(k+1)/(2k+1)}(M_n(\theta_n + h/r_n) - M_n(\theta_n)) &\xrightarrow{d} c_0|h| - c_0 c + \frac{\mu^{(k)}(\theta_0+)|c|^{k+1}}{(k+1)!} \\ &= c_0(|h| - ck/(k+1)) \end{aligned}$$

, as $n \to \infty$. Note that the drift $A(h)$ is continuous. The (3.15) can be checked again by easy calculations. Finally to check (3.16), we calculate the limiting covariance similar to (3.18) and obtain $B(h) = \sigma^2 W(h)$ in this case. Finally we get $n^{1/(2k+1)}(\hat{\theta}_n - \theta_n)$ converges in distribution to $\arg\min_h(\sigma^2 W(h) + \mu^{(k)}(\theta_0+)|h|^{k+1}/(k+1)!\mathbf{1}(h \geq -c) + c_0(|h| - ck/(k+1))\mathbf{1}(h < -c))$.

3.3.2 Mode Estimation Problem

Consider data (Y_i, U_i) coming from the model $Y_i = f(|U_i - \theta_0|) + \epsilon_i$, where $f : [0, \infty) \mapsto \mathbb{R}$ is a monotone decreasing differentiable function with $f(0) < \infty$ and

U_i are i.i.d. uniform. Note that the regression function is unimodal and symmetric around the parameter θ_0.

Estimation of θ_0 with varying bandwidth: We want to estimate the mode θ_0 and it is estimated by

$$\hat{\theta}_n := \arg\max_\theta M_n(\theta) = \arg\max_\theta \frac{1}{n} \sum_{i=1}^{n} (Y_i \mathbf{1}(U_i \in [\theta - h_n, \theta + h_n]))$$

where $h_n \to 0$ as $n \to \infty$. The true parameter θ_0 can be written as

$$\theta_0 = \arg\max_\theta \mathbb{E} M_n(\theta) = \arg\max_\theta \int_{\theta - h_n}^{\theta + h_n} f(|x - \theta_0|)dx =: \arg\max_\theta \mathbb{M}_n(\theta).$$

Indeed $M_n'(\theta_0) = 0$ and $M_n''(\theta_0) = 2f'(h_n) < 0$ since f is monotone decreasing.

Rate of Convergence: As before we apply Theorem 11 to obtain the rate of convergence. In this case $m_{n,\theta}(Y_i, U_i) := Y_i \mathbf{1}(U_i \in [\theta - b, \theta + b])$ and the metric associated is $d_n(\theta_1, \theta_2) := \sqrt{f'(h_n)}|\theta_1 - \theta_2|$. Existence of second derivative at θ_0 implies that $|M(\theta) - M(\theta_0)| \leq d_n^2(\theta, \theta_0)$ for all θ in a neighbourhood of θ_0. To check the remaining assumptions we use Corollary III.14 and write

$$\begin{aligned}
&\mathbb{E} \sup_{\theta_1 : d_n(\theta_1, \theta_2) < \delta} |m_{n,\theta_1}(Y_1, U_1) - m_{n,\theta_2}(Y_2, U_2)|^2 \\
=&\mathbb{E} \sup_{\theta_1 : |\theta_1 - \theta_2| < \delta / \sqrt{f'(h_n)}} |Y_1 \mathbf{1}(U_1 \in [\theta_1 - h_n, \theta_1 + h_n]) - Y_1 \mathbf{1}(U_1 \in [\theta_2 - h_n, \theta_2 + h_n])|^2 \\
=&\mathbb{E} \sup_{\theta_1 : |\theta_1 - \theta_2| < \delta / \sqrt{f'(h_n)}} |Y_1 (\mathbf{1}(U_1 \in [\theta_1 - h_n, \theta_2 - h_n]) - \mathbf{1}(U_1 \in [\theta_1 + h_n, \theta_2 + h_n]))|^2 \\
=&\mathbb{E} \sup_{\theta_1 : |\theta_1 - \theta_2| < \delta / \sqrt{f'(h_n)}} Y_1^2 (\mathbf{1}(U_1 \in [\theta_1 - h_n, \theta_2 - h_n]) + \mathbf{1}(U_1 \in [\theta_1 + h_n, \theta_2 + h_n])) \\
=&\mathbb{E} \sup_{\theta_1 : |\theta_1 - \theta_2| < \delta / \sqrt{f'(h_n)}} (f^2 |U_1 - \theta_0| + \epsilon^2)(\mathbf{1}(U_1 \in [\theta_1 - h_n, \theta_2 - h_n] \cup [\theta_1 + h_n, \theta_2 + h_n])) \\
\leq& 2(f^2(0) + \sigma^2) \frac{\delta}{f'(h_n)}.
\end{aligned} \tag{3.22}$$

Hence by Corollary III.14 we have $\phi_n(\delta) \leq \frac{\sqrt{\delta}}{(f'(h_n))^{1/4}}$ and using the equation $r_n^2\phi_n(1/r_n) = \sqrt{n}$ we obtain $r_n = n^{1/3}/(f'(h_n))^{1/6}$. Using the form of d_n we finally have $(nf'(h_n))^{1/3}(\hat{\theta}_n - \theta_0) = O_p(1)$.

Asymptotic Distribution: We apply Theorem 13 with $l_n = (nf'(h_n))^{1/3}$. From the previous calculations, we have $\psi_n(1/l_n) = 1/\sqrt{l_n}$. Note that by calculations similar to (3.22) we have $\mathbb{E}(m_{n,\theta}(Y_1, U_1) - m_{n,\theta_0}(Y_1, U_1))^2 \leq C|\theta - \theta_0|$ which implies that the functions of the class $\mathcal{M}_n$ defined in the Theorem have finite second moment. By Taylor expansion, with the notation of Theorem 13 we can write

$$\begin{aligned}
\sqrt{n}\mathbb{E}(\tilde{m}_{n,\theta_0+h/l_n}(Y_i) - \tilde{m}_{n,\theta_0}(Y_i)) =& \sqrt{nl_n}[\mathbb{M}_n(\theta_0 + h/l_n) - \mathbb{M}_n(\theta_0)] \\
=& \sqrt{nl_n}[\mathbb{M}''(\theta_0) + o(1)]\frac{h^2}{2l_n^2} \\
=& n^{2/3}(f'(h_n))^{1/3}\frac{(2f'(h_n) + o(1)}{n^{2/3}(f'(h_n))^{2/3}}h^2 \\
\to& \sqrt{f'(0)}h^2
\end{aligned}$$

as $nto\infty$. Therefore condition (3.14) is satisfied with $A(h) = \sqrt{f'(0)}h^2$. Condition (3.15) can be checked by direct calculations similar to (3.22). To check (3.16) we calculate the limit of the covariance kernel of the process $\{\sum_{i=1}^n Q_{ni}(h)\}_h$ with

$$\begin{aligned}
Q_{ni}(h) :=& (f'(h_n))^{1/6}n^{-1/3}(f(|U_i - \theta_0|) + \epsilon_i)\,[\mathbf{1}(U_i \in [\theta_0 + h/l_n - h_n, \theta_0 + h/l_n + h_n]) \\
&- \mathbf{1}(U_i \in [\theta_0 - h_n, \theta_0 + h_n])\Big]
\end{aligned}$$

as suggested by Remark III.17. The calculations are similar to (3.18) and the limiting covariance kernel turns out to be $H(g, h) = 2(f'(0))^{1/3}(f^2(0) + \sigma^2)\min(g, h)$ if both g and h are positive or negative and $H(g, h) = 0$ if g and h have different signs. Therefore we finally have

$$(nf'(h_n))^{1/3}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \arg\max_h[\sqrt{2(f^2(0) + \sigma^2)}W(h) + \sqrt{f'(0)}h^2]$$

as $n \to \infty$, where W is the two sided Brownian motion.

Estimation of θ_0 with fixed bandwidth: Alternatively we can choose a fixed bandwidth b in place of $h_n \to 0$. The calculation of rate of convergence and asymptotic distribution are similar in this situation, but the rate of convergence does not involve $f'(h_n)$ and we obtain simple cube-root convergence in this scenario. The asymptotic result turns out to be

$$n^{1/3}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \arg\max_h [\sqrt{2(f^2(b) + \sigma^2)} W(h) + 2(f'(b))h^2].$$

Therefore choosing a fixed bandwidth vs. shrinking bandwidth depend on the slope of the function f near 0. If the function is flat near 0, so $|f'(h_n)|$ is small we should use a fixed bandwidth to get better rate of convergence, on the other hand for a steep slope around 0 a shrinking bandwidth will perform better.

3.3.3 Change Point Estimation With Diminishing Signal

The next example we consider is the change point problem with diminishing signal, i.e, we have data (U_i, Y_i) for $i = 1, 2, \ldots, n$ such that

$$Y_i = \alpha_{n,0}\mathbf{1}(U_i \leq t_0) + \beta_{n,0}\mathbf{1}(U_i > t_0) + \epsilon_i,$$

where U_is are i.i.d. uniform(0,1). We assume $\alpha_{n,0} - \beta_{n,0} = c_0 n^{-\gamma}$ for some $\gamma < 1/2$ and our goal is to estimate the change point t_0. We define the parameter space $\Theta_n := \{\theta_n := (\alpha_n, \beta_n, t) \in \mathbb{R}^2 \times [0,1] : \alpha_n - \beta_n = c_0 n^{-\gamma}\}$ and the true parameter $\theta_{n,0} := (\alpha_{n,0}, \beta_{n,0}, t_0)$ is estimated to be $\arg\min_{\theta \in \Theta_n} \mathbb{M}_n(\theta)$, where

$$\begin{aligned}
\mathbb{M}_n(\theta_n) =& \frac{1}{n}\sum_{i=1}^n \left(Y_i - \alpha_n \mathbf{1}(U_i \leq t) - \beta_n \mathbf{1}(U_i > t)\right)^2 \\
=& \frac{1}{n}\sum_{i=1}^n \left[(Y_i - \alpha_n)^2 \mathbf{1}(U_i \leq t) + (Y_i - \beta_n)^2 \mathbf{1}(U_i > t)\right].
\end{aligned}$$

It is easy to see that $\theta_{n,0}$ is the minimizer of the process

$$M_n(\theta_n) := \mathbb{E}\mathbb{M}_n(\theta_n) = \mathbb{E}\left(Y_1 - \alpha_n\right)^2 \mathbf{1}(U_1 \le d) + \mathbb{E}\left(Y_1 - \beta_n\right)^2 \mathbf{1}(U_1 > d).$$

We will study the asymptotic properties of $\hat{\theta}_n$.

Rate of Convergence: We define the distance metric ρ_n on Θ_n to be

$$\rho_n(\theta_{n,1}, \theta_{n,2}) = |\alpha_{n,1} - \alpha_{n,2}| + |\beta_{n,1} - \beta_{n,2}| + n^{-\gamma}\sqrt{|t_1 - t_2|}.$$

It is easy to check that for all $\theta_n \in \Theta_n$, we have $M_n(\theta_n) - M_n(\theta_{n,0}) \le \rho_n^2(\theta_n, \theta_{n,0})$. To check the next assumption of Theorem 11, we need a tight bound for $\phi_n(\delta)$. We employ Theorem 12 with $m_{n,\theta_n}(U,Y) := (Y - \alpha_n)^2\mathbf{1}(U \le t) + (Y - \beta_n)^2\mathbf{1}(U > t)$. To this end note that $m_{n,\theta_n}(U,Y) - m_{n,\theta_{n,0}}(U,Y) = (2\epsilon + g_{\theta_n}(U))g_{\theta_n}(U)$, where

$$g_{\theta_n}(U) = \alpha_{n,0}\mathbf{1}(U \le t_0) + \beta_{n,0}\mathbf{1}(U > t_0) - \alpha_n\mathbf{1}(U \le t) - \beta_n\mathbf{1}(U > t) \tag{3.23}$$

with

$$\mathbb{E}(g_{\theta_n}^2(X)) = (\alpha_n - \alpha_{n,0})^2 d_0 + (\beta_n - \beta_{n,0})^2(1 - d_0) + c_0^2 n^{-2\gamma}|d - d_0| + o(1) \tag{3.24}$$

and

$$\mathbb{E}(g_{\theta_n}^4(X)) = (\alpha_n - \alpha_{n,0})^4 d_0 + (\beta_n - \beta_{n,0})^4(1 - d_0) + c_0^4 n^{-4\gamma}|d - d_0| + o(1). \tag{3.25}$$

Hence we have $n\|f\|_2^2 \le (4\sigma^2(1 + c_0^2) + K^2 + c_0^4)K^2$ Using this formulation for $\delta < 1$, we have

$$\begin{aligned}
&\mathbb{E} \sup_{\theta_{n,1}:\rho_n(\theta_{n,1},\theta_{n,2})<\delta} (m_{n,\theta_{n,1}}(U,Y) - m_{n,\theta_{n,2}}(U,Y))^2 \\
=&4\sigma^2\mathbb{E} \sup_{\theta_{n,1}:\rho_n(\theta_{n,1},\theta_{n,2})<\delta} (g_{\theta_{n,1}}(U) - g_{\theta_{n,2}}(U))^2 + \mathbb{E} \sup_{\rho_n(\theta_{n,1},\theta_{n,2})<\delta} (g_{\theta_{n,1}}(U) - g_{\theta_{n,2}}(U))^4 \\
\leq&(2 + c_0^2)\delta^2 + 2\delta^4 + c_0^4 n^{-2\gamma}\delta^2 \leq (4 + c_0^2 + c_0^4)\delta^2
\end{aligned} \tag{3.26}$$

Therefore by Remark III.14 we have $r_n = n^{1/(4-2)} = n^{1/2}$ and hence $\sqrt{n}\rho_n(\hat{\theta}_n, \theta_n) = O_p(1)$. This finally gives $\left(\sqrt{n}(\hat{\alpha}_n - \alpha_n), \sqrt{n}(\hat{\beta}_n - \beta_n), n^{1-2\gamma}(\hat{t}_n - t_0)\right) = O_p(1)$.

Asymptotic Distribution: We use Theorem 13 to obtain the asymptotic distribution of $l_n(\hat{\theta}_n - \theta_{n,0})$ with $l_n = (\sqrt{n}, \sqrt{n}, n^{1-2\gamma})$. Repeating the calculation as above we get

$$\psi_n(\delta_1, \delta_2, \delta_3) \sim \sqrt{\delta_1^2 + \delta_2^2 + n^{-2\gamma}\delta_3}$$

which gives $\psi_n(1/l_n) = \sqrt{n}$. The class of functions $\mathcal{M}_n$ can be written as $\{(\tilde{m}_{n,\theta_n}(U,Y) : |\alpha_n - \alpha_{n,0}| \leq 1/\sqrt{n}, |\beta_n - \beta_{n,0}| \leq 1/\sqrt{n}, |t - t_0| \leq n^{2\gamma-1}\}$ where

$$\tilde{m}_{n,\theta}(U,Y) = 2\epsilon + g_{\theta_n}(U))g_{\theta_n}(U)/\sqrt{n}$$

and some algebra shows that the functions of this class have finite second moment. Now let $\theta_{n,h} = \theta_{n,0} + (h_1/\sqrt{n}, h_2/\sqrt{n}, h_3/n^{1-2\gamma})$ and note that $P_n[(\tilde{m}_{n,\theta_{n,h}}(U,Y) - \tilde{m}_{n,\theta_{n,0}}(U,Y))/\sqrt{n}] = \mathbb{E}(Q_n(h))$ where

$$Q_n(h) = \sum_{i=1}^n (2\epsilon_i + g_{\theta_{n,h}}(U_i))g_{\theta_{n,h}}(U_i) = 2\sum_{i=1}^n \epsilon_i g_{\theta_{n,h}}(U_i) + \sum_{i=1}^n g^2_{\theta_{n,h}}(U_i) = 2A_n(h) + B_n(h).$$

Note that $\mathbb{E}(A_n(h)) = 0$ and we expand $B_n(h)$ as

$$B_n(h) = c_0 n^{-2\gamma} \sum_{i=1}^{n} \mathbf{1}(U_i \in (t_0, t_0 + \frac{h_3}{n^{1-2\gamma}}) + \frac{h_1^2}{n} \sum_{i=1}^{n} \mathbf{1}(U_i \leq t_0) + \frac{h_2^2}{n} \sum_{i=1}^{n} \mathbf{1}(U_i > t_0) + o(1).$$

From this it is easy to see that $\mathbb{E}(B_n(h)) \to c_0|h_3| + h_1^2 t_0 + h_2^2(1-t_0)$ as $n \to \infty$ uniformly on h as h ranges over bounded sets, which finally gives us (3.14) with $A(h) = c_0|h_3| + h_1^2 t_0 + h_2^2(1-t_0)$. Checking (3.15) is straightforward and similar to (3.26). Next by Remark III.17 we calculate limiting covariance kernel to check (3.16). USing the independence of ϵ_i and U_i, the covariance kernel can be written as

$$\begin{aligned}
&\mathrm{Cov}(Q_n(h_1), Q_n(h_2)) \\
=&4 \sum_{i=1}^{n} \mathbb{E}\epsilon_i^2 \mathbb{E}(g_{\theta_{n,h_1}}(U_i) g_{\theta_{n,h_2}}(U_i)) + 4 \sum_{i \neq j} \epsilon_i \epsilon_j \mathbb{E}(g_{\theta_{n,h_1}}(U_i)) \mathbb{E}(g_{\theta_{n,h_2}}(U_j)) \\
&+ \sum_{i=1}^{n} \mathbb{E}(g^2_{\theta_{n,h_1}}(U_i) g^2_{\theta_{n,h_2}}(U_i)) - \sum_{i=1}^{n} \mathbb{E}^2(g^2_{\theta_{n,h_1}}(U_i))
\end{aligned}$$

The expressions are easy to calculate using (3.23) and (3.24). All the terms except the first term are $o(1)$ and the first term yields the limiting covariance kernel is $H((h_{11}, h_{12}, h_{13}), (h_{21}, h_{22}, h_{23})) = \sigma^2 t_0 h_{11} h_{21} + \sigma^2 (1-t_0) h_{12} h_{22} + c_0^2 \sigma^2 \min(|h_{31}|, |h_{33}|)$, if h_{31} and h_{32} are both positive or both negative, otherwise the last term is just replaced by zero. Hence $B(h) \xrightarrow{d} h_1 N(0, \sigma^2 t_0) + h_2 N(0, \sigma^2(1-t_0) + c_0 \sigma W(h)$ in this example. Finally combining all the results we have

$$\sqrt{n}(\hat{\alpha}_n - \alpha_{n0}) \xrightarrow{d} N\left(0, \frac{\sigma^2}{t_0}\right)$$

$$\sqrt{n}(\hat{\beta}_n - \beta_{n0}) \xrightarrow{d} N\left(0, \frac{\sigma^2}{1-t_0}\right)$$

$$n^{1-2\gamma}(\hat{d}_n - d_0) \xrightarrow{d} \arg\min_h c_0 \sigma W(h) + c_0|h|$$

and these three components are asymptotically independent.

3.3.4 Maximum Score Estimator and Similar Problems

Maximum Score Estimator is a semiparametric estimate for discrete choice regression model in Econometrics (See [42]). We observe data (Y_i, U_i) for $i = 1, 2, \ldots, n$ such that $Y_i = \mathbf{1}(U_i'\theta_0 + \epsilon_i > 0)$. Unlike the general dependence structure we are considering here we assume that (U_i, ϵ_i) are jointly β mixing with the mixing coefficients satisfying (M2). The parameter θ_0 is estimated by maximizing the objective function

$$M_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}(2Y_i - 1)\mathbf{1}(U_i'\theta \geq 0)$$

over a suitably normalized parameter space. Generally θ is normalized to have unit norm in order to grantee the identifiability of the model. Here we use the formulation from [4] and write the model as

$$Y_i = \mathbf{1}(U_{di} + \tilde{U}_i'\theta_0 + \epsilon_i).$$

The random covariate U_i is divided in two parts: the component $\tilde{U}_i \in \mathbb{R}^{d-1}$ and a single continuous covariate U_{di}. We restrict the coefficient of U_{di} to be 1 and the parameter θ_0 is in $\mathbb{R}^{d-1}$. We maximize the objective function

$$M_n(\theta) := \frac{1}{n}\sum_{i=1}^{n}(2Y_i - 1)[\mathbf{1}(U_{di} + \tilde{U}_i'\theta \geq 0) - \mathbf{1}(U_{di} + \tilde{U}_i'\theta_0 \geq 0)].$$

We assume the following:

(Q1) The support of the distribution of U_1 is not contained in any proper linear subspace of $\mathbb{R}^d$,

(Q2) $0 < Pr(Y_1 = 1|U_1 = u) < 1$ for almost every u,

(Q3) The distribution of U_{d1} conditional on $\tilde{U}_1$ has everywhere positive density with respect to Lebesgue measure with density $g(u_d|\tilde{U}_1 = u)$.

(Q4) $\text{Median}(\epsilon_i|U_i) = 0$.

(Q5) The components of $\tilde{U}_1$ and $\tilde{U}_1\tilde{U}_1'$ have finite absolute first moment.

(Q6) The function $g'(u_d|\tilde{U}_1)$, derivative of the density function $g(.|.)$ exists and for some $M > 0$, $|g'(u_d|\tilde{U}_1 = u)| \leq M$ for all u_d and almost every u.

(Q7) For all ϵ in a neighborhood around 0 and for all u_d in a neighborhood around $-\tilde{u}'\theta_0$, for elmost every $\tilde{u}$, the conditional density of ϵ_1 given U_{d1} and $\tilde{U}_1$, $(\epsilon|u_d, \tilde{u})$ exists and bounded above by M.

(Q8) Let $F(\epsilon|u_d, \tilde{u})$ be the conditional distribution function of ϵ_1 given U_{d1} and $\tilde{U}_1$. For all ϵ in a neighborhood around 0 and for all u_d in a neighborhood around $-\tilde{u}'\theta_0$, for elmost every $\tilde{u}$ the quantity $\partial F(\epsilon|u_d, \tilde{u})/\partial x_d$ exists and $|\partial F(\epsilon|u_d, \tilde{u})/\partial x_d| \leq M$.

(Q9) The parameter θ_0 is an interior point of the parameter space $\Theta \subset \mathbb{R}^{d-1}$.

(Q10) The matrix

$$V := \mathbb{E}[2f(0|\tilde{u}, -\tilde{u}'\theta_0)g(-\tilde{u}'\theta_0|\tilde{u})\tilde{u}\tilde{u}']$$

is positive definite.

[4] shows that $M(\theta) := \mathbb{E}(M_n(\theta)) \leq 0$ and $M(\theta_0) = 0$. Therefore θ_0 is a maximizer of $M(\theta)$.

Rate of convergence: To obtain the rate of convergence we apply Theorem 11 with $m_\theta(Y_i, U_i) := (2Y_i - 1)[\mathbf{1}(U_{di} + \tilde{U}_i'\theta \geq 0) - \mathbf{1}(U_{di} + \tilde{U}_i'\theta_0 \geq 0)]$. By[4] we have $\nabla_{\theta\theta}M(\theta_0) = -V$ and hence

$$M(\theta) - M(\theta_0) \geq -C|\theta - \theta_0|^2$$

for all θ in a neighborhood of θ_0. Next to check the entropy condition, write

$$\begin{aligned}
&\mathbb{E} \sup_{\theta_1:|\theta_1-\theta_2|<\delta} \|m_{\theta_1}(Y_1,U_1) - m_{\theta_2}(Y_1,U_1)\|^2 \\
\leq &\mathbb{P} \sup_{\theta_1:|\theta_1-\theta_2|<\delta} \{\tilde{U}'\theta_1 \geq -U_d > \tilde{U}'\theta_2 \text{ or } \tilde{U}'\theta_2 \geq -U_d > \tilde{U}'\theta_1\}.
\end{aligned}$$

By (Q1), (Q2), (Q3) and (Q5) the right hand side is of order δ . So from our assumption of angular component, the right hand side is $O(\delta)$. So as the earlier examples by Remark III.14 we have $r_n = n^{1/3}$.

Asymptotic Distribution: As before we use Theorem 13 to obtain the asymptotic distribution of $n^{1/3}(\hat{\theta}_n - \theta_0)$. As discussed earlier, in this case $\psi_n(\delta) = \delta^{1/2}$ which gives $f(l_n) = n^{2/3} = l_n^2$. By the existence of the second derivative, we have

$$l_n^2(M(\theta_0 + h/l_n) - M(\theta_0)) \xrightarrow{d} \frac{1}{2}h'Vh$$

as $n \to \infty$. Note that the matrix V is negative definite by (Q10). For all functions $f \in \mathcal{M}_n$ as defined in Theorem 13 we have

$$\begin{aligned}
\mathbb{E}(f^2) =& l_n \mathbb{P}\left(\tilde{U}'\theta \geq -U_d > \tilde{U}'\theta_0 \text{ or } \tilde{U}'\theta < -U_d \leq \tilde{U}'\theta_0\right) \\
=& l_n O(|\theta - \theta_0|) = O(1).
\end{aligned}$$

Assumption (3.15) can be checked similarly. Finally by Remark III.17 to check assumption (3.16) we have to calculate the limiting covariance kernel. As shown in [40], the covariance Kernel can be written as
$H(s,t) = 1/2\,(L(s,0) + L(t,0) - L(s,t))$, where

$$L(s,t) = \lim_{n\to\infty} n^{1/3}\mathrm{Var}\sum_{i=1}^{n}\left(m_{\theta_0+n^{-1/3}s}(U_i,\epsilon_i) - m_{\theta_0+n^{-1/3}t}(U_i,\epsilon_i)\right).$$

For fixed s and t define $g_{n,i} = m_{\theta_0+n^{-1/3}s}(U_i, \epsilon_i) - m_{\theta_0+n^{-1/3}t}(U_i, \epsilon_i)$. The limit $n^{1/3}\mathrm{Var}(g_{n,i})$ is given in [4]. We have to obtain the limit of $n^{1/3}\sum_{m=1}^{\infty}\mathrm{Cov}(g_{n,i}, g_{n,i+m})$. Note that g_{ni} are functions of indicator functions and can take values -1, 0 and 1. We can write, for all $n, m > 1$ and $j, k = -1, 0, 1$,

$$
\begin{aligned}
&|\mathbb{P}\{g_{n,i} = j; g_{n,i+m} = k\} - \mathbb{P}\{g_{n,i} = j\}\mathbb{P}\{g_{n,i+m} = k\}| \\
=&|\mathbb{P}\{h(U_i, \epsilon_i) \in A_1, h(U_{i+m}, \epsilon_{i+m}) \in A_2|B(U_i)B(U_{i+m})\}\mathbb{P}(B(U_i)B(U_{i+m})) \\
&- \mathbb{P}\{h(U_i, \epsilon_i) \in A_1|B(U_i)\}\mathbb{P}(B(U_i))\mathbb{P}\{h(U_i, \epsilon_{i+m}) \in A_2|B(U_{i+m})\}\mathbb{P}(B(U_{i+m}))|
\end{aligned}
$$

where A_1 and A_2 are some subsets of $\{-1, 0, 1\}$ depending on j and k and $B(U_i)$ is the event corresponding to the possible values of $\mathbf{1}(\tilde{U}_i'(\theta_0 + sn^{-1/3} \geq -U_{di}) - \mathbf{1}(\tilde{U}_i'(\theta_0 + tn^{-1/3} \geq -U_d)$. Note that $\mathbb{P}(B(U_i)) \leq Cn^{-1/3}$ and $\mathbb{P}(B(U_i)B(U_{i+m})) \leq C_1 n^{-2/3}$ and the mixing assumption implies that $\mathbb{P}\{h(U_i, \epsilon_i) \in A_1, h(U_{i+m}, \epsilon_{i+m}) \in A_2|B(U_i)B(U_{i+m})\} - \mathbb{P}\{h(U_i, \epsilon_i) \in A_1|B(U_i)\}\mathbb{P}\{h(U_i, \epsilon_{i+m}) \in A_2|B(U_{i+m})\} \leq \beta_m/2$. Thus, $\{g_{n,i}\}$ is an α-mixing array whose mixing coefficients are bounded by $Kn^{-2/3}\beta_m$. By applying the α-mixing inequality from [23], for $p, q, r > 1$, such that $1/p + 1/q + 1/r = 1$ we have

$$
\mathrm{Cov}(g_{n,i}, g_{n,i+m}) \leq Cn^{-2/3p}\beta_m^{1/p}\|g_{n,i}\|_q\|g_{n,i+m}\|_r.
$$

Note that

$$
\|g_{n,i}\|_q = \left[\mathbb{E}|\mathbf{1}(\tilde{U}_i'(\theta_0 + s_1 n^{-1/3}) > -U_d) - \mathbf{1}(\tilde{U}_i'(\theta_0 + s_2 n^{-1/3}) > -U_d)|\right]^{1/q} = O\left(n^{1/(3q)}\right).
$$

Combining these results, choosing $p < 1.5$, $n^{1/3}\sum_{m=1}^{\infty}\mathrm{Cov}(g_{n,i}, g_{n,i+m}) \to 0$ as $n \to \infty$. Therefore, the covariance kernel is same as the independent case in [4] (page 1199). So the asymptotic distribution of $n^{1/3}(\hat{\theta}_n - \theta_0)$ is the same as the i.i.d. case.

Remark III.19. A number of similar interesting problems discussed in [40] can be solved by our method.

3.3.5 Linear and Penalized Linear Regression

M-estimation in Linear Models: We consider the linear model $Y_i = U_i'\theta_0 + \epsilon_i$ for $i = 1, 2, \ldots, n$, where θ_0 is a $p \times 1$ unknown regression coefficient vector, which we want to estimate. The design vectors U_is are $p \times 1$ and bounded in each component with $\mathbb{E}(U_1) = \mu$ and $\mathbb{E}(U_1'U_1) = \Sigma$. We estimate θ by

$$\hat{\theta}_n = \arg\min_{\theta} \frac{1}{n}\sum_{i=1}^{n} \rho(Y_i - U_i'\theta) =: \arg\min_{\theta} M_n(\theta),$$

where ρ is a convex function. For several choices of ρ we get well known models like Huber's estimate, $\mathcal{L}^q$ regression or LAD. Note that this set-up is similar to [68], but we consider slightly general loss function and different dependence structure. We assume the following properties of ρ:

(P1)$\mathbb{E}(\rho(\epsilon_1 + t))$ is differentiable as a function of t with derivative φ.

(P2)$\varphi(0) = 0$ and φ has a strictly positive derivative at $t = 0$.

(P3)$\|\rho(\epsilon_1 + t) - \rho(\epsilon_1)\|_2$ is differentiable as a function of t.

Note that as ρ is a loss function it is non-negative and $\rho(t) > 0$ for $t > 0$. By convexity of ρ, we have $\mathbb{E}(\rho(\epsilon_1 + t)) - \mathbb{E}(\rho(\epsilon_1)) \geq \mathbb{E}(\rho(t)) \geq 0$ and is equal to 0 iff $t = 0$. Hence we can write the true parameter as

$$\theta_0 = \arg\min_{\theta} \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(\rho(Y_i - U_i'\theta)) = \frac{1}{n}\sum_{i=1}^{n} \arg\min_{\theta} \mathbb{E}(\rho(\epsilon_1 + U_i'(\theta_0 - \theta))) =: \arg\min_{\theta} M(\theta)$$

Rate of Convergence: We take d_n to be the Euclidian metric on $\Theta := \mathbb{R}^p$. First note that by (P2), if t is in a small neighbourhood of 0, $U_1'y \leq M'y$, where $M := (M_1, M_2, \ldots, M_p)'$ is such that M_i is the lower (or upper) bound for U_1 if y_i is negative (or positive). Also convexity of ρ implies convexity of $\mathbb{E}(\rho(\epsilon_1 + .))$, which in turn implies that φ is non-decreasing. So for θ in a small neighborhood of θ_0, we can write,

$$\begin{aligned}
M(\theta) - M(\theta_0) &= \mathbb{E}(\rho(\epsilon_1 + U_1'(\theta_0 - \theta)) - \rho(\epsilon_1)) \\
&= \mathbb{E}(\mathbb{E}(\rho(\epsilon_1 + U_1'(\theta_0 - \theta)) - \rho(\epsilon_1)|U_1)) \\
&= \mathbb{E}(U_1'(\theta_0 - \theta)\varphi(\epsilon_1 + t_u)) \\
&\leq M'(\theta_0 - \theta)\varphi(\epsilon_1 + M'(\theta_0 - \theta))) \\
&\leq M'(\theta_0 - \theta)[KM'(\theta_0 - \theta)] \\
&\leq const\|\theta_0 - \theta\|^2
\end{aligned}$$

The last inequality is by Cauchy-Schwarz. This verifies the first condition of Theorem 11. To check the second condition of the Theorem, we have to obtain a tight bound for $\phi_n(\delta)$. Define $m_{n,\theta}(U, \epsilon) := \rho(U_i'\theta_0 - U_i'\theta + \epsilon)$. It is easy to check that (similar calculations as before)

$$\mathbb{E} \sup_{\|\theta_1 - \theta_2\| \leq \delta} (m_{n,\theta_1}(U, \epsilon) - m_{n,\theta_2}(U, \epsilon))^2 \leq A\delta^2.$$

Therefore by Corollary III.14 we obtain $\sqrt{n}(\hat{\theta}_n - \theta_0) = O_p(1)$.

Asymptotic Distribution: To obtain the asymptotic distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ note that in this case $\psi_n(\delta) = \phi_n(\delta)$ and therefore $\psi_n(1/l_n) = 1/\sqrt{n}$. Therefore for all functions g in the class $\mathcal{M}_n := \{\tilde{m}_{n,\theta} : \|\theta - \theta_0\| < 1/l_n\}$ we have,

$$\mathbb{E}g^2 = \mathbb{E}\left(\rho(\epsilon_i - x_i'(\theta - \theta_0)) - \rho(\epsilon_i)\right)^2/n = nO(1/l_n^2) = O(1),$$

by (P3). To check (3.14) note that

$$\sqrt{n}P_n(\tilde{m}_{n,\theta_n + h/l_n} - \tilde{m}_{n,\theta_n}) = n(M_n(\theta_n + h/\sqrt{n}) - M_n(\theta_n))$$

and by Taylor expansion we have,

$$n(M(\theta_0 + h/\sqrt{n}) - M(\theta_0)) \xrightarrow{d} \frac{1}{2}\varphi'(0)h'\Sigma h$$

as $n \to \infty$. Also Assumption (3.15) follows directly from (P3). To check (3.16) note that the quantity $P_n(m_{\theta_n+g/l_n} - m_{\theta_n+h/l_n})^2/\psi_n^2(1/l_n)$ has a limit as $n \to \infty$ by Law of Large Number and (P3), but the limit is difficult to calculate for general ρ. In fact if ρ is differentiable with derivative ψ, it is easy to calculate the covariance kernel of the limit process $B(h)$ as the limit of the covariance kernel for the process $Q_n(h) := \sum_{i=1}^n(\rho(\epsilon_i - x_i'h/\sqrt{n}) - \rho(\epsilon_i))$. It can be written as

$$\mathrm{Cov}(Q_n(h_1), Q_n(h_2)) = \mathrm{Cov}\left(\sum_{i=1}^n \psi(\epsilon_i)U_i'h_1/\sqrt{n}, \sum_{i=1}^n \psi(\epsilon_i)U_i'h_2/\sqrt{n}\right) + o(1)$$

as $n \to \infty$. Now $\psi(\epsilon_i)$ has β-mixing coefficients bounded above by the β-mixing coefficients of ϵ_i. Therefore let $\tau_\psi^2 := \mathbb{E}\sum_{k\in\mathbb{Z}}(\psi(\epsilon_0), \psi(\epsilon_k))$. Then the right hand side converges to $h_1'\Delta h_2$, where $\Delta := \|\psi(\epsilon_1)\|^2\Sigma + \tau_\psi^2\mu\mu'$. Finally we have $B(h) = N(0,\Delta)h$. So $\sqrt{n}(\hat{\theta}_n - \theta_0)$ converges in distribution to $\arg\min_h N(0,\Delta)h + \frac{1}{2}h'\Sigma h$ as $n \to \infty$. Computing the minimizer we get,

$$\varphi'(0)\Sigma^{1/2}\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0,\Delta)$$

as $n \to \infty$. For non-differentiable ρ however the covariance kernel does not have a general form. For example, if $\rho(t) = |t(\tau - \mathbf{1}(t<0))|$ with $\tau \in (0,1)$, the loss function for quantile regression for τ-th quantile gives limiting kernel to be $H(h_1, h_2) = (1-\tau)h_1'\mathbb{E}(f_y(U'\theta_0)UU')h_2$.

Remark III.20. If (U_i, ϵ_i) are jointly β-mixing, the covariance kernel is different. For example, if ρ is differentiable with derivative ψ then the limiting covariance kernel we obtain is $h_1'\Delta_1 h_2$ where $\Delta_1 := \lim_{n\to\infty}\mathrm{Var}(\sum_{i=1}^n \psi(\epsilon_i)U_i/n)$ and the last limit exists because of the property of β-mixing. Therefore in this case $\sqrt{n}(\hat{\theta}_n - \theta_0)$ converges in distribution to $\arg\min_h N(0, \Delta_1)h + \frac{1}{2}h'\Sigma h$ as $n \to \infty$.

Remark III.21. In fact assumption (P1) can be relaxed. As the analysis above deals with the local behavior of φ, we need the function $t \mapsto \mathbb{E}(\rho(\epsilon_1 + t)$ to be differentiable in a neighborhood of 0 only.

M-Estimation for Penalized Linear Models: Here we consider a penalized least square regression model, i.e., we consider a linear model as above and estimate θ_0 by minimizing

$$\tilde{M}_n(\theta) := \frac{1}{n}\sum_{i=1}^{n}(Y_i - U_i'\theta)^2 + \lambda_n \sum_{j=1}^{p}|\theta_j|$$

where $\sqrt{n}\lambda_n \to \lambda_0$ for some λ_0 as $n \to \infty$ Suppose θ_n be the minimizer of

$$\mathbb{E}(\tilde{M}_n(\theta)) = \sigma^2 + (\theta - \theta_0)'\Sigma(\theta - \theta_0) + \lambda_n \sum_{j=1}^{p}|\theta_j|$$

and $\Delta_n = \theta_n - \theta_0$. As $\theta_n = \theta_0 + \Delta_n$ minimizes $\mathbb{E}(\tilde{M}_n(\theta))$ we can write,

$$\Delta_n'\Sigma\Delta_n \leq \lambda_n \sum_{j=1}^{p}[|\theta_{0j}| - |\theta_{0j} + \Delta_{nj}|]. \tag{3.27}$$

As $\Sigma = \mathbb{E}(XX')$ is a positive definite matrix this implies the right hand side of the above equation must be positive. Hence we have $\|\theta_0\|_1 \geq \|\theta_0 + \Delta_n\|_1$ and therefore $\|\Delta_n\|_1 \leq 2\|\theta_0\|_1$ and this implies $\|\theta_0\|_1 - \|\theta_0 + \Delta_n\|_1$. Substituting this last inequality in (3.27) we obtain

$$\Delta_n'\Sigma\Delta_n \leq \lambda_n\|\Delta_n\|_1.$$

As $\lambda_n = O(1/\sqrt{n})$ this implies $\Delta_n = O(1/\sqrt{n})$. Checking the assumptions of Theorem 11 is similar to the linear model case owing to the fact that for θ in a neighborhood of θ_0, the penalization term $\lambda_n(\|\theta\|_1 - \|\theta_0\|_1) \to 0$ as $n \to \infty$.

Therefore in this case we have $\sqrt{n}(\hat{\theta}_n - \theta_0) = O_p(1)$.

To obtain the asymptotic distribution note that $\psi_n(1/l_n) = 1/\sqrt{n}$ as before. Following the calculations for general linear model case, for all functions g in the call of functions $\mathcal{M}_n$ defined in Theorem 13 we have

$$\mathbb{E}g^2 \leq 2(nO(1/l_n^2) + n\lambda_n^2 O(1/l_n^2)) = O(1).$$

Assumption (3.15) can also be checked similarly to the general linear model case using the fact $\lambda_n = O(1/\sqrt{n})$. To check the other conditions first assume $\lim_n \sqrt{n}(\theta_n - \theta_0) = \Delta$. Now write

$$\begin{aligned}
&n[\mathbb{E}(\tilde{M}_n(\theta_n + h/\sqrt{n}) - \mathbb{E}(\tilde{M}_n(\theta_n)] \\
=&h'\Sigma h + 2h'\Sigma[\sqrt{n}(\theta_n - \theta_0)] + n\lambda_n(\|\theta_n + h/\sqrt{n}\|_1 - \|\theta_n\|_1)
\end{aligned}$$

Now as shown in [34] page 18, the last term converges to

$$P(h) := \lambda_0 \sum_{j=1}^{p} (h_j \text{sign}(\theta_j)\mathbf{1}(\theta_j \neq 0) + |h_j|\mathbf{1}(\theta_j = 0))$$

as $n \to \infty$. Therefore (3.14) is satisfied with $A(h) := h'\Sigma h + 2h'\Sigma\Delta + P(h)$. To check (3.16) we calculate the limiting covariance kernel which is not affected by introduction of the penalization term. So we get the covariance kernel using $\rho(t) = t^2$ for the general case and we get $B(h) = h'N(0, \sigma^2\Sigma + \tau^2\mu\mu')$. Finally we calculate the quantity Δ. Note that Δ is minimizer of $n(\mathbb{E}(\tilde{M}(\theta_0 + h/\sqrt{n})) - \mathbb{E}(\tilde{M}(\theta_0))$ and by similar argument as above this quantity converges to $h'\Sigma h + P(h)$ which is minimized uniquely at 0. Therefore we have $\Delta = 0$ giving $A(h) = h'\Sigma h + P(h)$. Hence we finally have

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \arg\min_h h'\Sigma h + P(h) + h'N(0, \sigma^2\Sigma + \tau^2\mu\mu')$$

as $n \to \infty$.

APPENDICES

APPENDIX A

Order of The Length of Likelihood Based Confidence Intervals for Isotonic Regression Problem

A.1 Properties of Confidence Intervals Using L_n and T_n

A.1.1 Local Asymptotic Behavior of L_n and T_n

To study the properties of the confidence interval we need to study the behavior of the statistics $L_n(\theta)$ and $T_n(\theta)$ as a process in θ. In order to do that we change the time scale to local scale $h = (t - t_0)/d_n$, where $1/d_n$ is the rate for convergence of $\hat{m}_n(t_0)$ to $m(t_0)$. By [14], [71] and [9], we have

$$d_n = \begin{cases} n^{-1/3} & \text{under independence or weak dependence} \\ n^{-\frac{1-H}{2-H}} & \text{under strong dependence} \end{cases}$$

We redefine all the processes of interest in this local scale. To this end define $\theta_{n,h} = \theta_0 + hd_n$ and for $z \in (-t_0/d_n, (1-t_0)/d_n] := (a_n, b_n]$ define the process

$$\mathbb{V}_n(z, \theta_{n,h}) = d_n^{-2}(U_n(t_0 + zd_n) - U_n(t_0) - \theta_{n,h} z d_n).$$

Define the version of Isotonic regression estimator and its constrained counterpart in the new time scale to be

$$
\begin{aligned}
X_n(z,\theta_{n,h}) =& d_n^{-1}(\hat{m}_n(t_0+zd_n)-\theta_{n,h}) \\
=& \mathcal{L}(\mathcal{T}(\mathbb{V}_n(.,\theta_{n,h})),(a_n,b_n]) \\
=& \mathcal{L}(\mathcal{T}((\mathbb{V}_n(.,\theta_0),(a_n,b_n]))-h
\end{aligned}
\tag{A.1}
$$

and

$$
\begin{aligned}
Y_n(z,\theta_{n,h}) =& d_n^{-1}(\hat{m}_n(t_0+zd_n)-\theta_{n,h}) \\
=& (\mathcal{L}(\mathcal{T}(\mathbb{V}_n(.,\theta_{n,h}),(a_n,l_n]))\wedge 0)\mathbf{1}_{(a_n,l_n]}+0\times\mathbf{1}_{(l_n,0]} \\
& +(\mathcal{L}(\mathcal{T}(\mathbb{V}_n(.,\theta_{n,h}),(l_n,b_n]))\vee 0)\mathbf{1}_{(0,b_n]} \\
=& (\mathcal{L}(\mathcal{T}(\mathbb{V}_n(.,\theta_0),(a_n,l_n]))\wedge h)\mathbf{1}_{(a_n,l_n]}+h\times\mathbf{1}_{(l_n,0]} \\
& +(\mathcal{L}(\mathcal{T}(\mathbb{V}_n(.,\theta_0),(l_n,b_n]))\vee h)\mathbf{1}_{(0,b_n]}-h
\end{aligned}
\tag{A.2}
$$

The relation (A.2) follows because $(a-h)\wedge 0 = a\wedge h - h$ and $(a-h)\vee 0 = a\vee h - h$. We extend $X_n(z,\theta_{n,h})$ and $Y_n(z,\theta_{n,h})$ to be defined for all $z\in\mathbb{R}$ as left-continuous step functions constant outside the interval $(a_n,b_n]$. Before stating the results involving X_n and Y_n, we introduce the processes that will appear in the limit. For $z\in\mathbb{R}$ and $1/2\le H<1$, define $\mathbb{G}(z)\equiv\mathbb{G}_{a,b}(z):=aW_H(z)+bz^2$, where $W_{1/2}$ is the two sided version of the Brownian motion $\mathbb{B}$ and corresponds to the short range dependent or independent case and for $H>1/2$, the process W_H is the two sided version of the fractional Brownian motion $\mathbb{B}_H$. Note that W_H is a self similar Gaussian process with self similarity parameter parameter H. Now define the following functionals of $\mathbb{G}$,

$$\begin{aligned}\mathcal{S}_{a,b}(z) &= \mathcal{L} \circ \mathcal{T}(\mathbb{G})(z) \\ \mathcal{S}^h_{a,b}(z) &= \begin{cases} \mathcal{L} \circ \mathcal{T}_{(-\infty,0)}(\mathbb{G})(z) \wedge h & , \ z \in (-\infty, 0) \\ \lim_{u \uparrow 0} \mathcal{L} \circ \mathcal{T}_{(-\infty,0)}(\mathbb{G})(u) \wedge h & , \ z = 0 \\ \mathcal{L} \circ \mathcal{T}_{(0,\infty)}(\mathbb{G})(z) \vee h & , \ z \in (0, \infty) \end{cases}\end{aligned} \tag{A.3}$$

At this stage we will establish the limit distribution of X_n and Y_n and will show that $\mathbb{L}_n$ and $\mathbb{T}_n$ can be expressed as functionals of the processes X_n and Y_n. Then some compactification argument along with the continuous mapping theorem will be used to obtain the limiting finite dimensional distributions of $\mathbb{L}_n$ and $\mathbb{T}_n$. This along with the tightness of the local processes $\{\mathbb{L}_n(\theta_0 + hd_n)\}_{h\in\mathbb{R}}$ and $\{\mathbb{T}_n(\theta_0 + hd_n)\}_{h\in\mathbb{R}}$ will give a process convergence result.

We start with stating a theorem on large sample behavior of X_n and Y_n.

Theorem 14. *The process $\{(X_n(z, \theta_{n,h}), Y_n(z, \theta_{n,h}))\}_{z\in\mathbb{R}}$ converges in distribution to $\{(\mathcal{S}_{a,b}(z) - h, \mathcal{S}^h_{a,b}(z) - h)\}_{z\in\mathbb{R}}$ in $L^2_{loc} \times L^2_{loc}$, where $\mathcal{S}_{a,b}$ and $\mathcal{S}^h_{a,b}$ defined in* (A.3).

Proof. By Theorem 1, we have $\mathbb{V}_n \Rightarrow \mathbb{G}$ uniformly on compacta as a process. Therefore as argued in Theorem 4.2 therein we have:

$$\begin{aligned}\mathbb{GCM}_n &:= \Big(\mathcal{T}(\mathbb{V}_n), \mathcal{T}_{(-\infty,l_n]}(\mathbb{V}_n)1_{(-\infty,0)}, \mathcal{T}_{(l_n,\infty)}(\mathbb{V}_n)1_{(0,\infty)}\Big) \\ &\Longrightarrow \Big(\mathcal{T}(\mathbb{G}), \mathcal{T}_{(-\infty,0)}(\mathbb{G}), \mathcal{T}_{(0,\infty)}(\mathbb{G})\Big),\end{aligned}$$

in the space $\mathcal{C}(\mathbb{R}) \times \mathcal{C}(-\infty, 0) \times \mathcal{C}(0, \infty)$ equipped with the local uniform convergence on compact sets.

Now as in Theorem 2 consider

$$J : \mathcal{C}(\mathbb{R}) \times \mathcal{C}(-\infty, 0) \times \mathcal{C}(0, \infty) \to M(\mathbb{R}) \times M(-\infty, 0) \times M(0, \infty),$$

defined as $J(f, f_-, f_+) := (\mathcal{L}f, \mathcal{L}f_-, \mathcal{L}f_+)$. Define map $\tilde{C}_h : M(-\infty, 0) \times M(0, \infty) \to M(\mathbb{R})$ as $\tilde{C}_h(f_1, f_2) = C_h(f_1, f_2) - h$, where C_h is the concatenation map in Lemma II.22. We have

$$(X_n, Y_n) = ((\mathrm{id}, \tilde{C}_h) \circ J)(\mathbb{GCM}_n),$$

where $\mathrm{id} : M(\mathbb{R}) \to M(\mathbb{R})$ denotes the identity. By Lemmas II.21 and II.22, the maps J and C_h are continuous and so is $\tilde{C}_h$ and the composition $((\mathrm{id}, \tilde{C}_h) \circ J)$. Hence an application of the Continuous Mapping Theorem gives us the result. $\square$

The following corollary will be useful for sequel.

Corollary A.1. *For every fixed $h \in \mathbb{R}$, we have that $(X_n(0), \{X_n(z, \theta_0 + hd_n), Y_n(z, \theta_0 + hd_n)\}_{z\in\mathbb{R}})$ converges in distribution to $(\mathcal{S}_{a,b}(0), \{\mathcal{S}_{a,b}(z) - h, \mathcal{S}^h_{a,b}(z) - h\}_{z\in\mathbb{R}})$ in $\mathbb{R} \times L^2_{loc} \times L^2_{loc}$.*

Proof. As $f \mapsto \mathcal{T}_c(f)$ is a continuous mapping , Theorem 1 $\mathcal{T}_K(\mathbb{V}_n) \Longrightarrow \mathcal{T}_K(\mathbb{G})$ as $n \to \infty$ uniformly on compact sets. So by the arguments in the proof of Theorem 14 we have (44). Now Lemma II.21 implies that the map $L : f \mapsto \partial_\ell f$ is a continuous map from the space of convex function to L^2_{loc}. By [9] (page 1890-1891) $\mathcal{S}_{a,b}$ is continuous at 0 almost surely. So Lemma II.20 and the Continuous Mapping Theorem imply the result. $\square$

We shall now obtain expressions for $L_n(\theta_{n,h})$ and $T_n(\theta_{n,h})$ through the processes X_n and Y_n. First let D^h_n be the set on which $\hat{m}_n$ and $\hat{m}^{\theta_{n,h}}_n$ differ. Then, for any $\epsilon > 0$, we can find $M_\epsilon > 0$ such that with probability greater than $1 - \epsilon$, $D^h_n \subset [t_0 - M_\epsilon d_n, t_0 + M_\epsilon d_n]$, eventually. The proof is similar to Lemma II.3. Let $\tilde{D}^h_n = d^{-1}_n(D^h_n - t_0)$.

Proposition A.2. *The quantities $L_n(\theta_{n,h})$ and $T_n(\theta_{n,h})$ can be expressed as*

$$L_n(\theta_{n,h}) = \frac{n^2 d_n^3}{\sigma_n^2}\Big(\int_{\mathbb{R}} \big(X_n^2(z,\theta_{n,h}) - Y_n^2(z,\theta_{n,h})\big)\,dz + o_P(1)\Big)$$
$$T_n(\theta_{n,h}) = \frac{n^2 d_n^3}{\sigma_n^2}\Big(\int_{\mathbb{R}} \left(X_n(z,\theta_{n,h}) - Y_n(z,\theta_{n,h})\right)^2 dz + o_P(1)\Big) \tag{A.4}$$

Proof. The proof is basically similar to the proof of Proposition II.2.

$$\begin{aligned}
L_n(\theta_{n,h}) =& \frac{n}{\sigma_n^2}\Big[-\sum_{i=1}^n (Y_i - \hat{m}_n(t_i))^2 + \sum_{i=1}^n (Y_i - \hat{m}_n^{\theta_{n,h}}(t_i))^2\Big] \\
=& \frac{n}{\sigma_n^2}\Big[2\sum_{i=1}^n (Y_i - \theta_{n,h})(\hat{m}_n(t_i) - \theta_{n,h}) - 2\sum_{i=1}^n (Y_i - \theta_{n,h})(\hat{m}_n^{\theta_{n,h}}(t_i) - \theta_{n,h})\Big] \\
&- \frac{n}{\sigma_n^2}\Big[\sum_{i=1}^n (\hat{m}_n(t_i) - \theta_{n,h})^2 - \sum_{i=1}^n (\hat{m}_n^{\theta_{n,h}}(t_i) - \theta_{n,h})^2\Big] \\
=& \frac{n^2}{\sigma_n^2}\sum_{i=1}^n \Big[\frac{1}{n}\big((\hat{m}_n(t_i) - \theta_{n,h})^2 - (\hat{m}_n^{\theta_{n,h}}(t_i) - \theta_{n,h})^2\big)\Big] \\
=& \frac{n^2}{\sigma_n^2}\Big(\int_0^1 (\hat{m}_n(t) - \theta_0)^2 dt - \int_0^1 (\hat{m}_n^{\theta_{n,h}}(t) - \theta_{n,h})^2 dt\Big) + R_n \\
=& \frac{n^2 d_n^3}{\sigma_n^2}\int_{\mathbb{R}} (X_n^2(z,\theta_{n,h}) - Y_n^2(z,\theta_{n,h}))dz + R_n
\end{aligned}$$

In view of Theorem 14 with similar arguments as in the proof of Proposition II.2 we can bound the remainder R_n and obtain:

$$L_n(\theta_{n,h}) = \frac{n^2 d_n^3}{\sigma_n^2}\Big(\int_{\mathbb{R}} \big(X_n^2(z,\theta_{n,h}) - Y_n^2(z,\theta_{n,h})\big)\,dz + o_P(1)\Big).$$

Similarly we have,

$$T_n(\theta_{n,h}) = \frac{n^2 d_n^3}{\sigma_n^2}\Big(\int_{\mathbb{R}} \left(X_n(z,\theta_{n,h}) - Y_n(z,\theta_{n,h})\right)^2 dz + o_P(1)\Big).$$

$\square$

Let $D^h_{a,b}$ be the set where $\mathcal{S}_{a,b}(z)$ and $\mathcal{S}^h_{a,b}(z)$ differ. Define,

$$\begin{aligned} L_\infty(h) =& \frac{1}{a^2} \int\limits_{D^h_{a,b}} \Big((\mathcal{S}_{a,b}(z))^2 - (\mathcal{S}^h_{a,b}(z))^2 \Big) dz \\ & - \frac{2h}{a^2} \int\limits_{D^h_{a,b}} \Big(\mathcal{S}_{a,b}(z) - \mathcal{S}^h_{a,b}(z) \Big) dz, \\ T_\infty(h) =& \frac{1}{a^2} \int\limits_{D^h_{a,b}} \Big(\mathcal{S}_{a,b}(z) - \mathcal{S}^h_{a,b}(z) \Big)^2 dz. \end{aligned} \tag{A.5}$$

For long range dependent errors $L_\infty(h)$ and $T_\infty(h)$ depend on H.

Remark A.3. Note that if $a = b = 1$, we have $L_\infty(0) = \mathbb{L}$ and $T_\infty(0) = \mathbb{T}$ where $\mathbb{L}$ and $\mathbb{T}$ appear in the limit distribution of L_n and T_n.

Theorem 15. *As $n \to \infty$ the process $\{f(n)(L_n(\theta_0 + hd_n), T_n(\theta_0 + hd_n))\}_{h\in\mathbb{R}}$ converges to $\{(L_\infty(h), T_\infty(h))\}_{h\in\mathbb{R}}$ in finite dimensional distributions, where*

$$f(n) = \begin{cases} 1 & \textit{under weak dependence} \\ \frac{n^2 d_n^3}{\sigma_n^2} & \textit{under strong dependence.} \end{cases} \tag{A.6}$$

Proof. Consider $\theta_{n,h_1}, \theta_{n,h_2}, \dots, \theta_{n,h_k}$ for $1 \leq i \leq k$. The arguments presented in the proof of Theorem 14 can be easily extended to show that
$(X_n(., \theta_{n,h_i}), Y_n(., \theta_{n,h_i}) : 1 \leq i \leq k)$ converges in distribution to
$(S_{a,b}(.) - h_i, S^{h_i}_{a,b}(.) - h_i : 1 \leq i \leq k)$ under $(L^2_{loc})^{2k}$ metric.
The rest of the proof involves a localization argument similar to the one in the proof of Theorem 4. Given $\epsilon > 0$, we can get a compact set K_ϵ of the form $[-M_\epsilon, M_\epsilon]$ such that for $1 \leq i \leq k$,

$$\mathbb{P}\Big[\tilde{D}^{h_i}_n \subset [-M_\epsilon, M_\epsilon]\Big] > 1 - \frac{\epsilon}{2k} \text{ and } \mathbb{P}\Big[D^{h_i}_{a,b} \subset [-M_\epsilon, M_\epsilon]\Big] > 1 - \frac{\epsilon}{2k}. \tag{A.7}$$

Define,

$V_{n,\epsilon,i} = \frac{1}{a^2}\int_{K_\epsilon}\Big(X_n^2(z,\theta_{n,h_i}) - Y_n^2(z,\theta_{n,h_i})\Big)dz$
$W_{\epsilon,i} = \frac{1}{a^2}\int_{K_\epsilon}\Big((\mathcal{S}_{a,b}(z))^2 - (\mathcal{S}_{a,b}^{h_i}(z))^2\Big)dz - \frac{2h_i}{a^2}\int_{K_\epsilon}\Big((\mathcal{S}_{a,b}(z)) - (\mathcal{S}_{a,b}^{h_i}(z))\Big)dz$
$V_{n,\epsilon,(i+k)} = \frac{1}{a^2}\int_{K_\epsilon}\Big(X_n(z,\theta_{n,h_i}) - Y_n(z,\theta_{n,h_i})\Big)^2 dz$
$W_{\epsilon,(i+k)} = \frac{1}{a^2}\int_{K_\epsilon}\Big(\mathcal{S}_{a,b}(z) - \mathcal{S}_{a,b}^{h_i}(z)\Big)^2 dz.$

Also, introduce the quantities

$\xi_{ni} = \frac{1}{a^2}\int_{\tilde{D}_n^{h_i}}\Big(X_n^2(z,\theta_{n,h_i}) - Y_n^2(z,\theta_{n,h_i})\Big)dz,$
$\xi_i = \frac{1}{a^2}\int_{D_{a,b}^{h_i}}\Big((\mathcal{S}_{a,b}(z))^2 - (\mathcal{S}_{a,b}^{h_i}(z))^2\Big)dz - \frac{2h_i}{a^2}\int_{D_{a,b}^{i}}\Big((\mathcal{S}_{a,b}(z)) - (\mathcal{S}_{a,b}^{h_i}(z))\Big)dz$
$\xi_{n(i+k)} = \frac{1}{a^2}\int_{\tilde{D}_n^{h_i}}\Big(X_n(z,\theta_{n,h_i}) - Y_n(z,\theta_{n,h_i})\Big)^2 dz$
$\xi_{i+k} = \frac{1}{a^2}\int_{D_{a,b}^{h_i}}\Big(\mathcal{S}_{a,b}(z) - \mathcal{S}_{a,b}^{h_i}(z)\Big)^2 dz.$

Define the vectors

$V_{n\epsilon} = (V_{n,\epsilon,1},\ldots,V_{n,\epsilon,2k}), W_\epsilon = (W_{\epsilon,1},\ldots,W_{\epsilon,2k}), \xi_n = (\xi_{n1},\ldots,\xi_{n2k})$ and $\xi = (\xi_1,\ldots,\xi_{2k})$. Since by (A.7) with probability at least $(1-\epsilon)$, K_ϵ contains all $\tilde{D}_n^{h_i}$ eventually, we have $\mathbb{P}[V_{n\epsilon} \neq \xi_n] < \epsilon$ for all sufficiently large n. By (A.7) we similarly have $\mathbb{P}[W_\epsilon \neq \xi] < \epsilon$. Note, however, that by the continuous mapping theorem, for each fixed $\epsilon > 0$, we have $V_{n\epsilon} \Rightarrow W_\epsilon$, as $n \to \infty$, because by Theorem 14, the process $(X_n(.,\theta_i), Y_n(.,\theta_i) : 1 \le i \le k)$ converges in distribution to $(\mathcal{S}_{a,b}(.) - h_i, \mathcal{S}_{a,b}^{h_i}(.) - h_i : 1 \le i \le k)$ in $(L^2_{loc})^{2k}$. We have thus, verified all the conditions of the converging together Lemma II.26 and hence $\xi_n \Rightarrow \xi$, as $n \to \infty$. Finally in view of Proposition A.2, by Slutsky's Theorem we have $(L_n(\theta_0 + hd_n), T_n(\theta_0 + hd_n))$ converges to $(L_\infty(h), T_\infty(h))$ in finite dimensional distribution. □

The tightness of the processes $\{L_n(\theta_0 + hd_n)\}_{h\in\mathbb{R}}$ and $\{T_n(\theta_0 + hd_n)\}_{h\in\mathbb{R}}$ rely on the convexity of the sample paths of L_n, T_n, L_∞ and T_∞. So we establish some results on shape of the processes.

Lemma A.4. *With probability one, the processes L_∞ and T_∞ as defined in* (A.5) *have continuous sample paths as functions of h.*

Proof. We shall focus on L_∞. The argument is similar for T_∞. Define

$$\widetilde{\mathcal{S}}_{a,b}(z) = \mathcal{L} \circ \mathcal{T}_{(\infty,0]}(\mathbb{G})\mathbf{1}_{(\infty,0]}(z) \tag{A.8}$$

$$+ \mathcal{L} \circ \mathcal{T}_{(0,\infty)}(\mathbb{G})\mathbf{1}_{(0,\infty)}(z), \tag{A.9}$$

that is, on $(\infty, 0]$ ($(0,\infty)$ respectively) $\widetilde{\mathcal{S}}_{a,b}(z)$ is the left derivative of the GCM of $\mathbb{G}$ restricted to $(\infty, 0]$ ($(0,\infty)$, resp.).

For any given $\epsilon > 0$ we shall investigate the quantity

$$\begin{aligned} L_\infty(h+\epsilon) - L_\infty(h) =& \frac{1}{a^2}\int \Big((\mathcal{S}^h_{a,b}(z))^2 - (\mathcal{S}^{(h+\epsilon)}_{a,b}(z))^2\Big)dz \\ &- \frac{2h}{a^2}\int \Big(\mathcal{S}^h_{a,b}(z) - \mathcal{S}^{(h+\epsilon)}_{a,b}(z)\Big)dz \\ &- \frac{2\epsilon}{a^2}\int_{D^{h+\epsilon}_{a,b}} \Big(\mathcal{S}_{a,b}(z) - \mathcal{S}^{(h+\epsilon)}_{a,b}(z)\Big)dz. \end{aligned} \tag{A.10}$$

Consider the following sets:

$$\begin{aligned} A^- =& \{z \in (-\infty, 0] : \widetilde{\mathcal{S}}_{a,b}(z) < h\} \\ B^- =& \{z \in (-\infty, 0] : h \le \widetilde{\mathcal{S}}_{a,b}(z) < h+\epsilon\} \\ C^- =& \{z \in (-\infty, 0] : h+\epsilon \le \widetilde{\mathcal{S}}_{a,b}(z)\} \\ A^+ =& \{z \in (0,\infty) : \widetilde{\mathcal{S}}_{a,b}(z) < h\} \\ B^+ =& \{z \in (0,\infty) : h \le \widetilde{\mathcal{S}}_{a,b}(z) < h+\epsilon\} \\ C^+ =& \{z \in (0,\infty) : h+\epsilon \le \widetilde{\mathcal{S}}_{a,b}(z)\} \end{aligned}$$

Now note that on A^- and C^+ we have $\mathcal{S}^h_{a,b}(z) = \mathcal{S}^{h+\epsilon}_{a,b}(z) = \mathcal{S}^c_{a,b}(z)$. So the first two integrals of (A.10) are zero on these two sets. On C^- and A^+, we have $\mathcal{S}^h_{a,b}(z) = h$ and $\mathcal{S}^{h+\epsilon}_{a,b}(z) = (h+\epsilon)$. On B^-, $\mathcal{S}^h_{a,b}(z) = h$ and $\mathcal{S}^{h+\epsilon}_{a,b}(z) = \widetilde{\mathcal{S}}_{a,b}(z) < h+\epsilon$ and on B^+, $\mathcal{S}^h_{a,b}(z) = \widetilde{\mathcal{S}}_{a,b}(z) > h$ and $\mathcal{S}^{h+\epsilon}_{a,b}(z) = h+\epsilon$. So on all these four sets $\mathcal{S}^h_{a,b}(z)$ and

$\mathcal{S}_{a,b}^{h+\epsilon}(z)$ differ by at most ϵ. As a result on these sets the integrand of the first integral in (A.10) is less than $\epsilon(\epsilon - h)$ and that of the second integral is less than ϵ. Also notice that $(B^- \cup C^- \cup A^+ \cup B^+) \subseteq (D_{a,b}^h \cup D_{a,b}^{h+\epsilon})$ and by Theorem 18 for each h, $D_{a,b}^h$ is contained in a random compact interval with probability one. By monotonicity consideration it can be seen that, as h increases the set $D_{a,b}^h$ shifts to the right. Therefore if $[L_0, U_0]$ and $[L_1, U_1]$ are compact intervals containing $D_{a,b}^h$ and $D_{a,b}^{h+1}$, respectively, for all $\epsilon \in (0,1)$, the set $D_{a,b}^h \cup D_{a,b}^{h+\epsilon}$ is contained in $[L_0, U_1]$. Hence we can make the first two integrals arbitrarily small by choosing small ϵ. The third integral is also finite as again $D_{a,b}^{h+\epsilon}$ is subset of the compact interval $[L_0, U_1]$ and the integrand is bounded on the set they are being integrated on. So the difference $|L_\infty(h+\epsilon) - L_\infty(h)| \to 0$ as $\epsilon \to 0$. So $L_\infty(h)$ is continuous in h. $\square$

Lemma A.5. *The shape of L_∞ and T_∞ can be described as follows:*

(i) $L_\infty(S_{a,b}(0)) = T_\infty(S_{a,b}(0)) = 0$.

(ii) L_∞ *and* T_∞ *have strictly decreasing sample paths on* $(-\infty, S_{a,b}(0)]$ *and strictly increasing sample paths on* $(S_{a,b}(0), \infty)$.

Proof. Note that $S_{a,b}(z) = S_{a,b}^{S_{a,b}(0)}(z)$ for all $z \in \mathbb{R}$. This in view of (A.5) immediately proves (i). We shall prove (ii) only for L_∞ and $h > S_{a,b}(0)$. The other cases can be treated similarly. Consider the process $\widetilde{\mathcal{S}}_{a,b}$ in (A.8). By definition (A.5), we can write

$$\begin{aligned}
a^2 L_\infty(h) =& \int_{\mathbb{R}} \left((S_{a,b}(z) - h)^2 - \left(S_{a,b}^h(z) - h\right)^2 \right) dz \\
=& \int_{(-\infty,0]} \left((S_{a,b}(z) - h)^2 - \left(S_{a,b}^h(z) - h\right)^2 \right) dz \\
&+ \int_{(0,\infty)} \left((S_{a,b}(z) - h)^2 - \left(S_{a,b}^h(z) - h\right)^2 \right) dz \\
=& I_1(h) + I_2(h)
\end{aligned}$$

Let $z_1 < 0 < z_2$ be respectively the largest negative and smallest positive touch points of the process $\mathbb{G}$ with its GCM. Indeed by Lemma B.1 these points exist. Observe that $\mathcal{T}_{(\infty,0]}(\mathbb{G})(z) = \mathcal{T}(\mathbb{G})(z)$, for all $z \leq z_1$, that is the constrained and unconstrained GCMs coincide to the left of z_1. Therefore $\widetilde{\mathcal{S}}_{a,b}(z) \equiv \mathcal{S}_{a,b}(z), z \leq z_1$. Similarly $\widetilde{\mathcal{S}}_{a,b}(z) \equiv \mathcal{S}_{a,b}(z)$ for all $z \geq z_2$. Also by characterization of GCM it is easy to see that $\widetilde{\mathcal{S}}_{a,b}(z) \geq \mathcal{S}_{a,b}(z)$ for $z \in (z_1, 0]$ and $\widetilde{\mathcal{S}}_{a,b}(z) \leq \mathcal{S}_{a,b}(z)$ for $z \in (0, z_2)$. Finally, note that by Lemma B.3, $\mathcal{S}_{a,b}(z) \equiv \mathcal{S}_{a,b}(0) = \text{const}$ on $z \in (z_1, z_2]$.
Now consider the first integral $I_1(h)$. As we have $h > \mathcal{S}_{a,b}(0)$, the integrand of $I_1(h)$ is zero outside the interval $(z_1, 0]$. Let $h_1 > h_2$, then

$$
\begin{aligned}
& I_1(h_1) - I_1(h_2) \\
= & \int_{(z_1,0]} \left((\mathcal{S}_{a,b}(0) - h_1)^2 - (\mathcal{S}_{a,b}(0) - h_2)^2\right) dz \\
& - \int_{(z_1,0]} \left(\left(\widetilde{\mathcal{S}}_{a,b}(z) \wedge h_1 - h_1\right)^2 - (\left(\widetilde{\mathcal{S}}_{a,b}(z) \wedge h_2 - h_2\right)^2\right) dz \\
= & \int_{(z_1,0]} (h_1 + h_2 - 2\mathcal{S}_{a,b}(0))(h_1 - H_2) dz - \\
& \int_{(z_1,0]} \left(h_1 - h_2 - \widetilde{\mathcal{S}}_{a,b}(z) \wedge h_1 + \widetilde{\mathcal{S}}_{a,b}(z) \wedge h_2\right) \\
& \left(h_1 + h_2 - \widetilde{\mathcal{S}}_{a,b}(z) \wedge h_1 - \widetilde{\mathcal{S}}_{a,b}(z) \wedge h_2\right) dz.
\end{aligned}
$$

As $h_1 > h_2$, we have $\widetilde{\mathcal{S}}_{a,b}(z) \wedge h_1 - \widetilde{\mathcal{S}}_{a,b}(z) \wedge h_2 \geq 0$. Therefore the second integrand is bounded above by $\left(h_1 + h_2 \widetilde{\mathcal{S}}_{a,b}(z) \wedge h_1 - \widetilde{\mathcal{S}}_{a,b}(z) \wedge h_2\right)(h_1 - h_2)$. Hence we have,

$$
I_1(h_1) - I_1(h_2) \geq \int_{(z_1,0]} (h_1 - h_2) \left(\widetilde{\mathcal{S}}_{a,b}(z) \wedge h_1 + \widetilde{\mathcal{S}}_{a,b}(z) \wedge h_2 2\mathcal{S}_{a,b}(0)\right) dz,
$$

which is non-negative because $h_1 > h_2 > \mathcal{S}_{a,b}(0)$ and $\widetilde{\mathcal{S}}_{a,b}(z) > \mathcal{S}_{a,b}(0)$ for $z \in (z_1, 0]$. To show the monotonicity second integral $I_2(h)$, notice that if $\widetilde{\mathcal{S}}_{a,b}(z) > h$, then

$z > z_1$ and $\mathcal{S}_{a,b}$ coincides with $\widetilde{\mathcal{S}}_{a,b}$, making the integrand of the second integral $I_2(h)$ zero. Moreover if $\widetilde{\mathcal{S}}_{a,b}(z) \leq h$, we have $\mathcal{S}^h_{a,b}(z) = h$. Therefore $I_2(h)$ can be written as $\int_{\widetilde{\mathcal{S}}_{a,b}(z)\leq h}(\mathcal{S}_{a,b}(0)h)^2 dz$. As $h > \mathcal{S}_{a,b}(0)$, the second integral $I_2(h)$ is a strictly increasing function of h.

Finally as $I_1(h)$ is non-decreasing and $I_2(h)$ is strictly increasing in h, we have the strict monotonicity of $L_\infty(h)$ for $h > \mathcal{S}_{a,b}(0)$. □

Now we are ready to establish the required tightness result.

Lemma A.6. $\{f(n)(L_n(\theta_0 + hd_n), T_n(\theta_0 + hd_n))\}_{h\in\mathbb{R}}$ *as a process in* h *is tight on* $C(I) \times C(I)$ *for any compact interval* I*, where* $f(n)$ *is defined in* (A.*6).*

Proof. Note that by Theorem 15 and the Portmanteau Theorem we have $\limsup_{n\to\infty} \mathbb{P}(|f(n)L_n(\theta_0)| \geq a) \leq \mathbb{P}(|L_\infty(0)| \geq a) \to 0$ as $a \to \infty$. Now, let $I = [A, B]$ and $w_x(\delta) = \sup_{|s-t|\leq\delta} |x(s) - x(t)|$. By Theorem 3 of [9] and Corollary 1 of [71], we have $d_n^{-1}(\hat{\theta}_n - \theta_0) := \Delta_n$ converges in distribution to some random variable (say, Δ) and $L_\infty(\Delta) = 0$. Also by Lemma A.1, we have joint convergence of Δ_n and $f(n)L_n$.

Let $\epsilon > 0$ and $\eta > 0$ be given. By Lemma A.4,

$$\lim_{\delta\to 0} \mathbb{P}(w_{L_\infty}(\delta) > \epsilon/2) = 0.$$

So, choose $\delta_1 \in (0, 1/2)$ s.t., for all $0 < \delta \leq \delta_1$, $\mathbb{P}(w_{L_\infty}(\delta) > \epsilon/2) < \eta$. Now, take $\delta = 2\delta_1$. Take $A = h_0 < h_1 < \cdots < h_k = B$ with $h_j - h_{j-1} = \frac{\delta}{2}$ for $j = 1, 2, \ldots, k$. By the monotonicity property of $L_n(\theta_0 + hd_n)$ to the left and right of Δ_n, we have

$$\begin{aligned}
&\limsup_{n\to\infty} \mathbb{P}(w_{f(n)L_n}(\delta) \geq \epsilon) \\
\leq &\limsup_{n\to\infty} \mathbb{P}(\max_{1\leq j\leq k} f(n)|L_n(\theta_0 + h_j d_n) - L_n(\theta_0 + h_{j-1}d_n)|\mathbf{1}_{\Delta_n\in[t_{j-1},t_j)^c} \\
&+ \max\{f(n)|L_n(\theta_0 + h_j d_n)|, f(n)|L_n(\theta_0 + h_{j-1}d_n)|\}\mathbf{1}_{\Delta_n\in[t_{j-1},t_j)} \geq \epsilon/2)
\end{aligned}$$

Now, since $(\Delta_n, \{f(n)L_n(\theta_0 + hd_n)\}_{h\in\mathbb{R}}) \Longrightarrow (\Delta, \{L_\infty(h)\}_{h\in\mathbb{R}})$ the Portmanteau Theorem implies that the last lim sup is bounded above by

$$\begin{aligned}
&\mathbb{P}(\max_{1\le j\le k}(|L_\infty(h_j) - L_\infty(h_{j-1})|\mathbf{1}_{\Delta\in[t_{j-1},t_j)^c} \\
&\quad + \max\{|L_\infty(h_j)|, |L_\infty(h_{j-1})|\}\mathbf{1}_{\Delta\in[t_{j-1},t_j)}) \ge \epsilon/2) \\
\le&\mathbb{P}(w_{L_\infty}(\delta/2) \ge \epsilon/2) < \eta
\end{aligned}$$

As $\eta > 0$ is arbitrary, we have

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\mathbb{P}(w_{f(n)L_n}(\delta) \ge \epsilon) = 0.$$

So, by Theorem 7.3 from [17], we have $\{L_n(\theta_0 + hd_n)\}_{h\in\mathbb{R}}$ is tight. Similar arguments yield the tightness of $\{T_n(\theta_0 + hd_n)\}_{h\in\mathbb{R}}$ and the proof is complete. □

Theorem 16. *$\{f(n)(L_n(\theta_0 + hd_n), T_n(\theta_0 + hd_n))\}_{h\in\mathbb{R}}$ converge in distribution to $\{(L_\infty(h), T_\infty(h))\}_{h\in\mathbb{R}}$ uniformly in $C(I) \times C(I)$ where $f(n)$ is defined as in (A.6).*

Proof. Theorem 15 and Lemma A.6 together imply the result. □

Remark A.7. If $\theta_n = \theta_0 + a_n$ where $a_n = o(d_n)$, then $(L_n(\theta_n), T_n(\theta_n) \Rightarrow (\mathbb{L}, \mathbb{T})$.

A.1.2 Length of the Confidence Interval Constructed Via L_n and T_n

Theorem 17. *The length of the confidence intervals based on L_n and T_n decreases at a rate d_n, that is, if (a_n, b_n) is the confidence interval, $b_n - a_n = O_p(d_n)$ as $n \to \infty$.*

Proof. By Theorem 16 we have

$$f(n)L_n(\theta_0 + hd_n) \Rightarrow L_\infty(h) \tag{A.11}$$

uniformly on $C(I)$ for every compact interval I. Also by Lemma **??** the function $h \mapsto f(n)L_n(\theta_0 + hd_n)$ is non-negative and it is minimized at $h = H_n = d_n^{-1}(\hat{\theta}_n - \theta_0)$. Let (a_n, b_n) be the L_n-based confidence interval and c_0 be the cut-off. So we have

$$d_n^{-1}(a_n - \theta_0) = \inf\{h : L_n(\theta_0 + hd_n) \leq c_0\}$$

$$d_n^{-1}(b_n - \theta_0) = \sup\{h : L_n(\theta_0 + hd_n) \leq c_0\}$$

Let $\mathcal{F}$ be the set of continuous real-valued U-shaped functions f with minimum value 0 and $K : \mathcal{F} \to \mathbb{R}$ be the functional $K(g) = \inf\{h : g(h) \leq c_0\}$, with the convention $\inf \emptyset = -\infty$. Let $\{f_n\}_{n=1}^{\infty} \in \mathcal{F}$ be such that $f_n \to f$ uniformly on compact sets as $n \to \infty$. Also assume that $f \in \mathcal{F}$ is strictly decreasing before reaching the minimum and strictly increasing after that.Furthermore suppose that $f(x) \to \infty$ as $|x| \to \infty$. Note that by the continuity of f_n and f we have, $f(K(f)) = f_n(K(f_n)) = c_0$. First let, x_0 be such that, $f(x_0) = 0$ and $K(f) = x_1$. Let $0 < \delta < x_0 - x_1$. As $f_n \to f$ uniformly on $[x_0 - \delta, x_0 + \delta]$ we have $f_n(x_0 - \delta) \to f(x_0 - \delta) > 0$, $f_n(x_0) \to f(x_0) = 0$ and $f_n(x_0 + \delta) \to f(x_0 + \delta) > 0$ as $n \to \infty$. So, there exists N such that for all $n > N$, $f_n(x_0 - \delta) > 0$ and $f_n(x_0 + \delta) > 0$. So f_n assumes value 0 at some point inside $(x_0 - \delta, x_0 + \delta)$ for all $n > N$. Now pick some $0 < \epsilon < x_0 - x_1 - \delta$. As $f_n \to f$ uniformly on $[x_1 - \epsilon, x_1 + \epsilon]$ there exists $N_1 > N$ such that for all $n > N_1$ we have, $f_n(x_1 - \epsilon) < f(x_1) = c_0 < f_n(x_1 + \epsilon)$. So $K(f_n) \in (x_1 - \epsilon, x_1 + \epsilon)$ eventually. As $\epsilon > 0$ can be chosen arbitrarily small we have $K(f_n) \to K(f)$ as $n \to \infty$.
Now, by Lemma **??** the function $h \mapsto L_\infty(h)$ is continuous, U-shaped, strictly decreasing on $(-\infty, S_{a,b}(0)]$ and strictly increasing on $[S_{a,b}(0), \infty)$. Therefore the functional K is continuous at L_∞ almost surely. So by continuous mapping theorem applied to (A.11) we have $K(L_n(\theta_0 + \dot{d}_n)) = d_n^{-1}(a_n - \theta_0) \Rightarrow K(L_\infty)$ as $n \to \infty$. So, $d_n^{-1}(a_n - \theta_0) = O_p(1)$. Similarly we can show that $d_n^{-1}(b_n - \theta_0) = O_p(1)$. So the

length of the confidence interval $(b_n - a_n) = O_p(d_n)$. $\square$

A.1.3 Length of Confidence Intervals Constructed Via Ψ_n

Obtaining a theoretical result about length of confidence intervals constructed using Ψ_n is more challenging. Unlike $L_n(\theta)$ and $T_n(\theta)$, sample paths of $\Psi_n(\theta)$ are not convex as a function of θ. In fact $\Psi_n(\theta)$ being a function of ratio of two convex function functions do not have nice regular sample paths. (See Figure 2.2) Due to this it is difficult to prove tightness results and consequently process convergence of $\Psi_n(\theta)$.

Extensive simulation suggests that the order of the confidence intervals for Ψ_n is also d_n.

APPENDIX B

Properties of fBm with a Quadratic Drift and the random variables $\mathbb{L}$ and $\mathbb{T}$

B.1 Properties of GCM of a Function

Before discussing the properties of functionals of Brownian and fractional Brownian motion, we provide some technical lemmas about GCMs of deterministic functions. For a real-valued function f on $\mathbb{R}$, introduce the constrained GCM:

$$\mathcal{T}_0(f)(z) = \begin{cases} \mathcal{T}_{(0,\infty)}(f)(z) & , \text{ if } z \in (0,\infty) \\ \mathcal{T}_{(-\infty,0]}(f)(z) & , \text{ if } z \in (-\infty,0]. \end{cases} \tag{B.1}$$

Lemma B.1. *Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function such that $f(t)/|t| \to \infty$ as $|t| \to \infty$. Then, there exist points $t_1 < 0$ and $t_2 > 0$, such that $\mathcal{T}(f)(t_i) = f(t_i),\ i = 1, 2.$*

Proof. Observe that if ϕ is a convex minorant of f such that $f(t) = \phi(t)$, then $\mathcal{T}(f)(t) = f(t)$. We will construct such a ϕ with $f(t_2) = \phi(t_2),\ t_2 > 0$. The existence of a negative touch point can be treated similarly.

Note that f is bounded below since $f(t) \to \infty,\ |t| \to \infty$ and let $a := \inf_{t\in\mathbb{R}} f(t) > -\infty$. If this infimum is attained at $t_2 > 0$, then $\phi(t) := a,\ t \in \mathbb{R}$ is the desired convex minorant. Otherwise, let $\ell(t) := \lambda t + a$ be a line with positive slope $\lambda > 0$, that passes through the points $(0, a)$ and $(x, f(x))$, for some $x > 0$. Such a line exists since $f(x) \to \infty$. We consider two cases:

(i) If $\ell(t) \leq f(t)$, for all $t \geq 0$, then the convex minorant $\phi(t) := a1_{(-\infty,0]}(t) + \ell(t)1_{(0,\infty)}(t),\ t \in \mathbb{R}$ has the desired property with $t_2 := x$.

(ii) If $\ell(t) > f(t)$, for some $t \geq 0$, then the infimum $b := \inf_{t\geq 0}(f(t) - \ell(t)) < 0$ is negative, and it is attained at some $t_2 > 0$, because f is continuous and $f(t)/t \to \infty,\ t \to \infty$. In this case, $\phi(t) := a1_{(-\infty,|b|/\lambda]}(t) + (\ell(t) + b)1_{(|b|/\lambda,\infty)}(t),\ t \in \mathbb{R}$ is a convex minorant for f with $\phi(t_2) = f(t_2)$. This completes the proof. □

Lemma B.2. *For any continuous real-valued function f on $\mathbb{R}$ and $\mathcal{T}_0$ as in* (B.1)*, we have:*

(i) $\mathcal{T}(f)(t) \leq \mathcal{T}_0(f)(t) \leq f(t),\ t \in \mathbb{R}$.

(ii) If $f(t_0) = \mathcal{T}(f)(t_0)$ for some $t_0 > 0$ ($t_0 < 0$, respectively) then $\mathcal{T}(f)(t) = \mathcal{T}_0(f)(t)$ for all $t \geq t_0$ ($t \leq t_0$, respectively).

Proof. The first assertion is immediate by definition since the constrained GCM is no greater than the unconstrained one. To prove the second statement, suppose $\mathcal{T}_0(f)(t_0) = \mathcal{T}(f)(t_0)$ for some $t_0 > 0$. Let $\phi(t) = \mathcal{T}(f)(t)\mathbf{1}_{(-\infty,t_0]}(t) + \mathcal{T}_0(f)(t)\mathbf{1}_{(t_0,\infty)}(t)$. By the already established part *(i)*, we have that $\mathcal{T}(f)(t) \leq \mathcal{T}_0(f)(t) \leq f(t)$, and hence one can show that the function ϕ is convex and it is also a minorant of f. Therefore, $\phi(t) \equiv \mathcal{T}_0(f)(t) \leq \mathcal{T}(f)(t)$ for all $t \geq t_0$, which in view of part *(i)* yields $\mathcal{T}(f)(t) = \mathcal{T}_0(f)(t),\ t \geq t_0$. The case $t_0 < 0$ can be treated similarly. □

Lemma B.3. *Let f be a continuous function with $f(0) = 0,\ f(t)/|t| \to \infty$ as $|t| \to \infty$ and such that both $\inf_{(-\infty,0)} f(t)$ and $\inf_{(0,\infty)} f(t)$ are negative. Then we* *have the following:*

(i) There exist $x_* < 0 < x^*$, such that the GCM of f coincides with the line joining $(x_*, f(x_*))$ and $(x^*, f(x^*))$ on the interval $[x_*, x^*]$.

(ii) Moreover, $\mathcal{L} \circ \mathcal{T}(f)(0) = 0$ implies that $\inf\limits_{t\in(-\infty,0]} f(t) = \inf\limits_{t\in[0,\infty)} f(t)$.

Proof. Let $C = \{x \in \mathbb{R} : f(x) = \mathcal{T}(f)(x)\}$. First note that C is a closed set since f and $\mathcal{T}(f)$ are both continuous. By Lemma B.1 there exist a positive point and a negative point in C. Observe that by the convexity of $\mathcal{T}(f)$ and the fact $\inf_{(-\infty,0)} f(t)$ and $\inf_{(0,\infty)} f(t)$ are negative, the point 0 does not belong to C. Indeed, no convex function passing through the origin could take strictly negative values on both $(-\infty, 0)$ and $(0, \infty)$.

Define $x_* := \sup\{x \in C, x < 0\}$ and $x^* := \inf\{x \in C, x > 0\}$. Since C is closed we have that $x_*, x^* \in C$, also $(x_*, x^*) \cap C = \emptyset$ and $x_* < 0 < x^*$, since $0 \notin C$. To prove *(i)*, consider the function

$$\widetilde{\mathcal{T}}(x) := \begin{cases} \mathcal{T}(f)(x) & , \text{ if } x \notin [x_*, x^*] \\ \ell(x) & , \text{ if } x \in [x_*, x^*], \end{cases}$$

where $\ell(x) = (f(x^*) - f(x_*))(x - x_*)/(x^* - x_*) + f(x_*)$ is the equation of the line joining the points $(x_*, f(x_*))$ and $(x^*, f(x^*))$. We will show that $\widetilde{\mathcal{T}}$ coincides with $\mathcal{T}(f)$. Since $\mathcal{T}(f)$ is convex and its graph passes through the points $(x_*, f(x_*))$ and $(x^*, f(x^*))$, it follows that $\widetilde{\mathcal{T}}$ is convex, and in fact $\widetilde{\mathcal{T}}(x) \geq \mathcal{T}(f)(x),\ x \in \mathbb{R}$. Therefore, it remains to show that $\widetilde{\mathcal{T}}$ is a minorant of f, which amounts to proving $\widetilde{\mathcal{T}}(x) \equiv \ell(x) \leq f(x),\ x \in [x_*, x^*]$.

Suppose that $\min_{x\in[x_*,x^*]}(f(x) - \ell(x)) < 0$, which by continuity is attained at some $x_0 \in (x_*, x^*)$. Define the linear function $\widetilde{\ell}(x) := \ell(x) + (f(x_0) - \ell(x_0))$ and observe that $\widetilde{\ell}(x) < \ell(x),\ x \in \mathbb{R}$ and hence $\widetilde{\ell}$ is a (trivial) convex minorant to f. We also have that $\widetilde{\ell}(x_0) = f(x_0)$, which implies that $x_0 \in C$. This is however a contracticiton since $x_0 \in (x_*, x^*)$, and as argued above $(x_*, x^*) \cap C = \emptyset$.

We now prove *(ii)*. Suppose that the slope of GCM on $[x_*, x^*]$ is 0. Then, by part *(i)*, the horizontal line passing through the points $(x_*, f(x_*))$ and $(x^*, f(x^*))$ is a convex minorant of f over the entire real line, which touches the graph of the function at x_* and x^*. This shows that the infima of f on the intervals $(-\infty, 0)$ and $(0, \infty)$ are equal and are attained at x_* and x^*, respectively. $\square$

Lemma B.4. *Assume that f is continuous, $f(0) = 0$, $\inf_{t\in(0,\infty)} f(t) < 0$, and $f(t) \to \infty$ as $t \to \infty$. Then $\lim_{\delta\downarrow 0} \mathcal{T}_{[-\delta,\infty)}(f)(0) = 0$.*

Proof. First note that for each fixed t, $\mathcal{T}_{[-\delta,\infty)}(f)(t)$ is a non-increasing function of $\delta > 0$ and define

$$\varphi(t) := \lim_{\delta\downarrow 0} \mathcal{T}_{[-\delta,\infty)}(f)(t) \equiv \sup_{\delta>0} \mathcal{T}_{[-\delta,\infty)}(f)(t).$$

The function φ is a convex minorant of f on $[0, \infty)$. We will prove the result by contradiction. Indeed, assume that $\lim_{\delta\downarrow 0} \mathcal{T}_{[-\delta,\infty)}(f)(0) = \varphi(0) =: -a < 0$. Since f is continuous and $f(t) \to \infty$, as $t \to \infty$, we have $\inf_{[0,\infty)} f(t) = f(t^*)$, for some point $t^* \geq 0$. Since also $f(0) = 0 > \inf_{t\in(0,\infty)} f(t) = f(t^*)$, it follows that $t^* > 0$. The continuity of f and the fact that $f(0) = 0$, imply that $\exists\ \delta_0 > 0$ small enough such that $f(t) \geq f(t_*)$ for all $t \in [-\delta_0, \infty)$. Therefore, the horizontal line passing through $(t_*, f(t_*))$ is a minorant of f in the interval $[-\delta_0, \infty)$. This shows that t_* is a touch point of $\mathcal{T}_{[-\delta,\infty)}(f)$ and f for $0 \leq \delta \leq \delta_0$ and hence it is a touch point of φ and f, too. Also, $\mathcal{T}_{[-\delta,\infty)}(f)$, for $0 \leq \delta \leq \delta_0$ and consequently φ lie above the horizontal line mentioned above. As a result both convex minorants $\mathcal{T}_{[0,\infty)}(f)$ and φ are non-increasing in the interval $[0, t_*]$.

Consider now the convex set
$C = \{\vec{x} = (x_1, x_2) \in \mathbb{R}^2 : 0 \leq x_1 \leq t_*, \mathcal{T}_{[0,\infty)}(f)(x_1) \leq x_2 \leq 0\}$. Let $d = \inf_{\vec{x}\in C} \|\vec{x} - \vec{v}\|$ be the distance from the point $\vec{v} := (0, -a)$ to the set C. Define the d-neighborhood of C, $C_d = \{\vec{y} \in \mathbb{R}^2 : \inf_{\vec{x}\in C} \|\vec{y} - \vec{x}\| \leq d\}$. Note that C_d is also a

closed convex set and $\vec{v}$ is a point on its boundary. Thus, by the characterization of convex sets there exists a vector to $\vec{\Delta} = (\Delta_1, \Delta_2)$, such that $\langle \vec{u} - \vec{v}, \vec{\Delta} \rangle > 0$, for all interior points $\vec{u}$ of C_d.

We will show that both components Δ_1 and Δ_2 of $\vec{\Delta}$ are positive. Indeed, since $\vec{0}$ is an interior point of C_d, we have $\langle \vec{0} - \vec{v}, \vec{\Delta} \rangle = \Delta_2 a > 0$, which implies $\Delta_2 > 0$. Also, considering the interior point $(t_*, f(t_*))$ of C_d, we get $\langle (t_*, f(t_*)) - (0, \varphi(0)), \vec{\Delta} \rangle = t_*\Delta_1 + (f(t_*) - \varphi(0))\Delta_2 > 0$, which in turn implies $t_*\Delta_1 > (\varphi(0) - f(t_*))\Delta_2$. The last quantity is positive since $f(t_*) = \varphi(t_*) < \varphi(0)$. This yields $\Delta_1 > 0$.

Note that the compact set $K := \{(t, f(t)) : t \in [0, t_*]\}$ is contained in the interior of C_d. Therefore,

$$\inf_{t \in [0,t_*]} \langle \vec{\Delta},\, (t, f(t)) - (0, \varphi(0)) \rangle = \inf_{t \in [0,t_*]} \Big(\Delta_1 t + \Delta_2 (f(t) - \varphi(0)) \Big) := c > 0.$$

This means that the linear function $\ell(t) := -(\Delta_1/\Delta_2)t + c/2\Delta_2 + \varphi(0)$ is a minorant of the function f on the interval $[0, t_*]$. Since also the minimum of f on $[0, \infty)$ is attained at t_* and the slope of this aforementioned line is negative, the function $\ell(t)$ is also a minorant of f on $[0, \infty)$.

By the continuity of f at 0, the above line is a minorant of f on $[-\delta, \infty)$ for some $\delta > 0$ and it passes through the point $(0, \varphi(0) + c/2\Delta_2)$. This, however, contradicts the fact that any such minorant should be no greater than $\mathcal{T}_{[-\delta,\infty)}(f)(0) \le \varphi(0)$ at 0. $\square$

B.2 Properties of the GCM-type functionals of the fractional Brownian motion plus quadratic drift

Let now $\{B_H(t)\}_{t \in \mathbb{R}}$ denote the two-sided fractional Brownian motion with Hurst index $H \in [1/2, 1)$. These processes have versions with continuous paths and in the sequel we shall work with such versions. Note that if $H = 1/2$, then B_H is the usual Brownian motion involved in the limit of under short range dependence. For $1/2 < H < 1$ it is two-sided version of fBm with Hurst index H. Define

$$\mathbb{G}_{a,b}^H(t) = aB_H(t) + bt^2, \quad a, b > 0.$$

By using the self similarity of the fBm one can show that

$$\{\mathbb{G}_{a,b}^H(z)\}_{z\in\mathbb{R}} \stackrel{d}{=} a(a/b)^{\frac{H}{2-H}} \left\{ \left(\mathbb{G}_{1,1}^H((b/a)^{\frac{1}{2-H}} z) \right) \right\}_{z\in\mathbb{R}}. \tag{B.2}$$

Thus in what follows it suffices to focus on the standardized process

$$\xi_H(t) := \mathbb{G}_{1,1}(t) = B_H(t) + t^2, \ t \in \mathbb{R}.$$

We shall establish certain properties of the process ξ_H, its GCM $\mathcal{T}(\xi_H)$ and its constrained counterpart $\mathcal{T}_0(\xi_H)$. All results established below for ξ_H and its functionals extend immediately to $\mathbb{G}_{a,b}^H$ (and its functionals) in light of (B.2).

Lemma B.5. *With probability one, we have $B_H(t)/t \to 0$, as $|t| \to \infty$.*

Proof. Define $X_k = \sup_{t\in[k-1,k]} |B_H(t)|$. By self-similarity we then have $X_k \stackrel{d}{=} k^H \sup_{t\in[\frac{k-1}{k},1]} |B_H(t)|$. By the Borell inequality ([6]), for some $c > 0$ and all $\epsilon > 0$,

$$\mathbb{P}(X_k > \epsilon(k-1)) \le \mathbb{P}\Big(\sup_{t\in[\frac{k-1}{k},1]} |B_H(t)| > \frac{\epsilon(k-1)}{k^H} \Big) \le e^{-c\frac{(k-1)^2}{k^{2H}}}.$$

This implies that, $\sum_{k=1}^{\infty} \mathbb{P}\left(X_k/(k-1) > \epsilon\right) < \infty$, and by the Borel-Cantelli Lemma, $\mathbb{P}(X_k/(k-1) > \epsilon \ i.o.) = 0$. Consequently, $\sup_{j\ge k} X_j/(j-1) \to 0$ a.s. as $k \to \infty$. Thus, $\sup_{t>k} |B_H(t)/t| \le \sup_{j\ge k} X_j/(j-1)$ implies that, with probability one, $|B_H(t)/t| \to 0$, as $t \to \infty$. The case $t \to -\infty$ is treated similarly. $\square$

Corollary B.6. *With probability one, we have $\inf_{t\in\mathbb{R}}(B_H(t) + t^2) > -\infty$.*

Proof. By Lemma B.5, with probability one, we have $B_H(t)/t \to 0$, as $|t| \to \infty$. Thus, with probability one, given $C > 0, \exists T_c$ such that $|B_H(t)/t| \leq C, \forall t > T_c$. Hence $\inf_{t\in[-T_c,T_c]}(B_H(t) + t^2) > -\infty$ and consequently $|B_H(t) + t^2| \geq t^2 - Ct > -\infty$ for $|t| > T_c$. This proves the result. $\square$

Proposition B.7. *Let*

$$M^- := \inf_{t\in(-\infty,0]} \xi_H(t) \quad \textit{and} \quad M^+ := \inf_{t\in[0,\infty)} \xi_H(t)$$

be the values of the infima of ξ_H over the negative and positive half-lines, respectively. We then have that:

(i) $\mathbb{P}\{M^- < 0\} = \mathbb{P}\{M^+ < 0\} = 1$.

(ii) $\mathbb{P}\{M^- = M^+\} = 0$.

Proof. *(i)* Note that $\mathbb{P}\{M^+ > 0\} = 1$ is equivalent to $\mathbb{P}\{B_H(t) + t^2 \geq 0,\ \forall t > 0\} = 0$. By the self-similarity of B_H, for all $c > 0$, we have that

$$\begin{aligned}
\mathbb{P}\{B_H(t) + t^2 > 0,\ \forall t > 0\} &= \mathbb{P}\{c^{-H} B_H(ct) + t^2 \geq 0,\ \forall t > 0\} \\
&= \mathbb{P}\{B_H(\tau) \geq -c^{H-2}\tau^2,\ \forall \tau > 0\} \\
&=: \mathbb{P}(A_c) \equiv \text{const.}
\end{aligned}$$

Note that $A_c \downarrow \{B_H(\tau) \geq 0,\ \forall \tau > 0\}$ as $c \to \infty$ and since the above probability does not depend on $c > 0$, it is enough to show that $\mathbb{P}\{B_H(\tau) \geq 0,\ \forall \tau > 0\} = 0$. This follows from law of iterated logarithm for fractional Brownian motion (see [48]).

(ii) In view of *(i)*, part *(ii)* readily follows from Lemma 2.6 in [40] applied to the process $\{\xi_H(t)\}_{t\in\mathbb{R}}$. $\square$

Finally we state the main results of this section.

Theorem 18. *The set where $\mathcal{T}(\xi_H)$ and $\mathcal{T}_0(\xi_H)$ differ is contained in a compact set almost surely.*

Proof. As ξ_H has a continuous path almost surely by Lemma B.2 the set where $\mathcal{T}(\xi_H)$ and $\mathcal{T}_0(\xi_H)$ differ is contained in the closed interval $[x_*, x^*]$, where x_* and x^* are as in the Lemma B.3. It remains to prove that x_* and x^* are finite almost surely. Lemma B.5 implies that $\xi_H(t)/t \overset{a.s.}{\to} \infty$ as $t \to \infty$. Thus, Lemma B.1 applied to the sample paths of ξ_H yields that x_* and x^* are finite with probability one. □

Theorem 19. *We have* $\mathbb{P}(\mathbb{T} = 0) = 0$ *under both short- and long-range dependence.*

Proof. Focus on the one-sided minorant $\mathcal{T}_{(-\infty,0)}(\xi_H)(t)$ for $t \in (-\infty, 0)$. By Proposition B.7, we have that $M^- = \inf_{t\in(-\infty,0)} \xi_H(t)$ is almost surely negative and since $\xi_H(t) \to \infty$ as $t \to -\infty$, the infimum is attained at some point $t_* < 0$. Therefore, the function $f(t) = M^-$ is a convex minorant to $\xi_H(t)$ on $(-\infty, 0]$ which has touch-point at $t = t_*$. Thus,

$$f(t_*) \leq \mathcal{T}_{(-\infty,0)}(\xi_H)(t_*) \leq \xi_H(t_*) = f(t_*)$$

and the one sided GCM $\mathcal{T}_{(-\infty,0)}(\xi_H)(t_*) = \xi_H(t_*)$. By convexity, we have $\mathcal{L} \circ \mathcal{T}_{(-\infty,0)}(\xi_H)(t) \geq 0 = \mathcal{L}(f)(t)$ for all $t \in (t_*, 0)$. This implies that the constrained slope (defining $\mathcal{S}^0_{1,1}(t)$ in (2.18)) in $(t_*, 0)$ is equal to zero. One can similarly show that $\mathcal{S}^0_{1,1}(t)$ vanishes on an interval to the right of zero, almost surely.
By Lemma B.3 *(i)*, we can find an interval around 0 in which the slope of $\mathcal{T}(\xi_H)(t)$ is constant. By the second part of the same lemma, if this slope is zero, then we have $M^- = M^+$, where M^+ and M^- are as defined in Proposition B.7, which has probability zero by the second assertion of Proposition B.7. This in view of (2.18) implies that the slope $\mathcal{S}_{1,1}(t)$ is constant and non-zero in a neighborhood of zero, almost surely.
We have thus shown that $\mathcal{S}^0_{1,1}(t)$ vanishes while $\mathcal{S}_{1,1}(t) \neq 0$ in a neighborhood of 0 almost surely. This by (2.20) implies $\mathbb{P}(\mathbb{T} \neq 0) = 1$. □

Lemma B.8. *With probability one, both $\mathcal{S}_{1,1}(t)$ and $\mathcal{S}^0_{1,1}(t)$ diverge to infinity as $t \to \infty$.*

Proof. Since by Theorem 18, $\mathcal{S}_{1,1}(t)$ and $\mathcal{S}^0_{1,1}(t)$ coincide eventually as $|t| \to \infty$, it is enough to prove the result for $\mathcal{S}_{1,1}(t)$.

Recall that by (2.18), we have $\mathcal{S}_{1,1}(t) = \mathcal{L} \circ \mathcal{T}(\xi_H)(t)$. We shall show that for any given $C > 0$, there exists t_0 s.t. $\mathcal{S}_{1,1}(t) = \mathcal{L} \circ \mathcal{T}(\xi_H)(t) \geq C$ for all $t > t_0$. To do so, we will construct a convex minorant of ξ_H, which touches ξ_H at some point and has slope C at that point. Then the GCM of ξ_H will lie above this minorant and hence it will have a slope at least as large as C to the right of that touch–point.

To construct such a minorant, let $c = \inf_{t \in \mathbb{R}} \xi_H(t)$. By Proposition B.7 we have that $c < 0$ with probability 1. Since $\xi_H(t) \to \infty$ as $|t| \to \infty$, there exists a $t_* \in \mathbb{R}$ where the infimum of $\xi_H(t)$ is attained.

Consider the curve $f(t) = c + C(t - t_*)1_{(t \geq t_*)}$. Then $f(t)$ is convex, t_* is a common point of $\xi_H(t)$ and $f(t)$, and $f(t)$ is below $\xi_H(t)$ for $t < t_*$.

By Lemma B.5 we have $\xi_H(t)/t \to \infty$ as $t \to \infty$. So one can find a $t^* > t_*$ such that $\xi_H(t) > f(t)$ for all $t > t^*$.

If $f(t)$ and $\xi_H(t)$ have no more common points in $(t_*, t^*]$, then $f(t)$ is our desired minorant. Otherwise, consider a class of functions $f_l(t) = c + C(t - l)1_{(t \geq l)}$. Note that if $l = t^*$, $f_l(t) < \xi_H(t)$ for all $t > t_*$. Define a function $d(l) = \min_{t \in (t_*, \infty)}(\xi_H(t) - f_l(t))$. It is easy to see that d is a continuous function of l. From our discussion above, $d(t_*) \leq 0$ and $d(t^*) > 0$. Therefore there exists a l_* in $[t_*, t^*)$ such that $d(l_*) = 0$. This $f_{l_*}(t)$ is our desired minorant. $\square$

B.3 Simulation of Quantiles of $\mathbb{L}$ and Ψ

In this section we discuss simulation of quantiles of $\mathbb{L}$ and $\Psi := -\log(\mathbb{L}/\mathbb{T} - 1)$. For simulating the quantiles of Ψ and $\mathbb{L}$ we used two methods (MD1) and (MD2).

(MD1) The method (MD1) relies on approximating the limit distributions by empirical distributions for large sample sizes. To elaborate, we simulated data from a simple model $y_i = i/n + \epsilon_i$, for $i = 1, 2, , \ldots, n$ with $n = 10^6$. The errors $\{\epsilon_i\}$ were therefore taken to i.i.d. normals (with variance 0.2) for simulating approximates to the limiting quantiles in the short range dependent case, and from fractional Gaussian noise with Hurst index $H = 0.7, 0.8, 0.9, 0.95, 0.99$ and variance 0.2, for the long range dependent case. The statistics L_n, T_n (not tabulated) and Ψ_n from this simulated dataset were then calculated, and the procedure repeated $M = 10^4$ times. The estimates of the quantiles were obtained from the sorted values of these statistics. Note that, as the limiting distributions are the same in the i.i.d. and short range cases, using the i.i.d. errors above suffices.

(MD2) For simulating the quantiles of Ψ and $\mathbb{L}$ by this method, we used discrete approximations of the limit processes. To this end, we simulated a sequence Z_j of standard normal variables (for the SRD case) and fractional Gaussian noise with Hurst index $H = 0.7, 0.8, 0.9, 0.95, 0.99$ (for the LRD case) of length $2n$ where $n = 10^6$ and constructed the partial sum process on $-2 \leq t \leq 2$ as follows:

$$W_n(t) = \frac{1}{n^H}\Big\{\mathbf{1}_{(t\geq 0)} \sum_{j=n+1}^{n+[nt/2]} Z_j - \mathbf{1}_{(t<0)} \sum_{j=n-[n(-t)/2]}^{n} Z_j\Big\}.$$

We then generated the process $Y_n(t) = W_n(t) + t^2$ on the grid with step size $\Delta = 2 \times 10^{-6}$ and computed the greatest convex minorant G_m and the constrained greatest convex minorant G_m^0 using the R implementation of PAVA in R package fdrtool(Klaus and Strimmer, 2013[**?**]). Corresponding approximations $\hat{\mathcal{S}}_{a,b}$ and $\hat{\mathcal{S}}^0_{a,b}$ of the slope processes $\mathcal{S}_{a,b}$ and $\mathcal{S}^0_{a,b}$ were obtained and numerical approximations of the random variables $\mathbb{L}$ and $\mathbb{T}$ computed using (2.20) where $\mathcal{S}_{a,b}$ and $\mathcal{S}^0_{a,b}$ were replaced by $\hat{\mathcal{S}}_{a,b}$ and $\hat{\mathcal{S}}^0_{a,b}$. Quantiles of the resulting statistic $\Psi = -\log(\mathbb{L}/\mathbb{T} - 1)$ were calculated based on $M = 10^4$ independent replications.

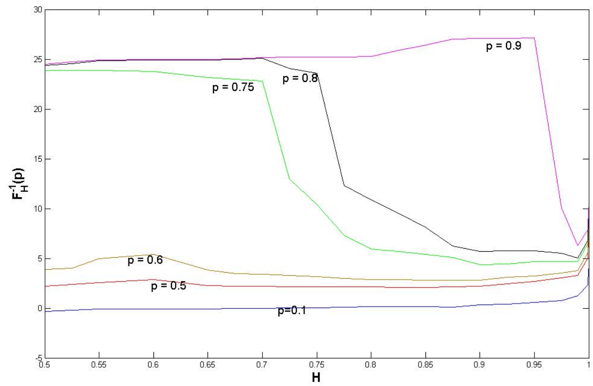


Figure B.1: Quantiles of Ψ as a function of H

The standard errors of these quantiles are based on the observations that for large n, F_n, the distribution function of the statistic L_n (Ψ_n, respectively) will be a good approximation to the corresponding limit distribution, and furthermore, if $F_{n,M}$ is the corresponding empirical CDF, based on M independent Monte-Carlo samples, then

$$\sqrt{M}(F_{n,M}^{\leftarrow}(\alpha) - F_n^{\leftarrow}(\alpha)) \approx_d N\left(0, \frac{\alpha(1-\alpha)}{f_n(F_n^{\leftarrow}(1-\alpha))}\right) \tag{B.3}$$

as $M \to \infty$. Here, $F_n^{\leftarrow}$ denotes the left continuous inverse of F_n and f_n denotes the density of F_n, which is assumed to exist at the point $F_n^{\leftarrow}(1-\alpha)$. The quality of the approximation is confirmed by the fact that (MD1) and (MD2) provide similar quantile estimates relative to the Monte-Carlo estimation error. The denominator of the variance in (B.3) was estimated by $\widehat{f}_n(F_{n,M}^{\leftarrow}(1-\alpha))$, where $\widehat{f}_n$ is obtained by differencing $F_{n,M}$.

Quantiles of $\mathbb{L}$ and Ψ calculated using both the methods are presented in the Tables B.1 – B.3. Figure B.1 shows the quantiles of Ψ as a function of h.

Remark B.9. Based on the simulated quantiles of Ψ, we make the following conjectures about its distribution function, F_H, for values of the Hurst index H between 0.5 (short-range dependence case) and 1.

Table B.1: Simulated Quantiles of Ψ: Method 1

p	SRD	H=0.7	H=0.8	H = 0.9	H = 0.95
.10	-0.35 (0.011)	-0.02 (0.011)	0.17 (0.015)	0.33 (0.052)	0.58 (0.062)
.15	0.10 (0.024)	0.26 (0.019)	0.47 (0.040)	0.63 (0.001)	0.90 (0.002)
.20	0.39 (0.013)	0.51 (0.018)	0.71 (0.015)	0.86 (0.012)	1.14 (0.014)
.25	0.55 (0.009)	0.73 (0.023)	0.92 (0.002)	1.09 (0.008)	1.36 (0.008)
.30	0.83 (0.027)	0.96 (0.026)	1.13 (0.036)	1.29 (0.020)	1.58 (0.030)
.35	1.04 (0.012)	1.20 (0.011)	1.36 (0.022)	1.50 (0.012)	1.99 (0.012)
.40	1.25 (0.020)	1.48 (0.027)	1.59 (0.019)	1.71 (0.008)	2.21 (0.011)
.45	1.69 (0.053)	1.79 (0.016)	1.85 (0.003)	1.94 (0.003)	2.41 (0.005)
.50	2.21 (0.021)	2.19 (0.006)	2.11 (0.012)	2.20 (0.051)	2.66 (0.025)
.55	2.88 (0.047)	2.69 (0.053)	2.43 (0.010)	2.49 (0.041)	2.93 (0.042)
.60	3.87 (0.068)	3.41 (0.030)	2.88 (0.028)	2.80 (0.015)	3.24 (0.020)
.65	6.28 (0.103)	4.69 (0.310)	3.47 (0.067)	3.18 (0.038)	3.59 (0.072)
.70	20.03 (0.022)	7.92 (0.349)	4.31 (0.173)	3.70 (0.011)	4.03 (0.097)
.75	23.91 (0.032)	22.82 (0.060)	5.95 (0.286)	4.40 (0.035)	4.68 (0.060)
.80	24.25 (0.020)	23.79 (0.019)	10.89 (0.494)	5.72 (0.132)	5.77 (0.122)
.85	24.67 (0.022)	24.51 (0.036)	24.14 (0.023)	8.43 (0.539)	8.30 (0.096)
.90	25.00 (0.041)	25.12 (0.031)	25.28 (0.054)	26.43 (0.165)	27.05 (0.248)
.95	25.21 (0.023)	25.92 (0.017)	26.32 (0.026)	28.02 (0.489)	33.13 (0.188)

Table B.2: Simulated Quantiles of Ψ: Method 2

p	SRD	H=0.7	H=0.8	H = 0.9	H = 0.95
.10	-0.34 (0.011)	-0.02 (0.011)	0.15 (0.015)	0.33 (0.052)	0.59 (0.062)
.15	0.09 (0.024)	0.24 (0.019)	0.44 (0.040)	0.63 (0.001)	0.91 (0.002)
.20	0.38 (0.013)	0.50 (0.018)	0.69 (0.015)	0.85 (0.012)	1.13 (0.014)
.25	0.54 (0.009)	0.77 (0.023)	0.93 (0.002)	1.10 (0.008)	1.36 (0.008)
.30	0.85 (0.027)	0.91 (0.026)	1.11 (0.036)	1.30 (0.020)	1.58 (0.030)
.35	1.02 (0.012)	1.19 (0.011)	1.39 (0.022)	1.52 (0.012)	1.99 (0.012)
.40	1.25 (0.020)	1.44 (0.027)	1.60 (0.019)	1.72 (0.008)	2.22 (0.011)
.45	1.70 (0.053)	1.78 (0.016)	1.84 (0.003)	1.95 (0.003)	2.40 (0.005)
.50	2.23 (0.021)	2.17 (0.006)	2.10 (0.012)	2.21 (0.051)	2.70 (0.025)
.55	2.85 (0.047)	2.70 (0.053)	2.45 (0.010)	2.51 (0.041)	2.92 (0.042)
.60	3.90 (0.068)	3.38 (0.030)	2.85 (0.028)	2.83 (0.015)	3.27 (0.020)
.65	6.30 (0.103)	4.48 (0.310)	3.50 (0.067)	3.21 (0.038)	3.59 (0.072)
.70	20.02 (0.022)	7.80 (0.349)	4.20 (0.173)	3.71 (0.011)	4.04 (0.097)
.75	23.89 (0.032)	22.80 (0.060)	6.12 (0.286)	4.40 (0.035)	4.66 (0.060)
.80	24.22 (0.020)	23.76 (0.019)	10.80 (0.494)	5.83 (0.132)	5.73 (0.122)
.85	24.61 (0.022)	24.49 (0.036)	24.10 (0.023)	8.51 (0.539)	8.24 (0.096)
.90	25.01 (0.041)	25.16 (0.031)	25.29 (0.054)	26.51 (0.165)	27.10 (0.248)
.95	25.27 (0.023)	25.90 (0.017)	26.30 (0.026)	28.13 (0.489)	33.15 (0.188)

Table B.3: Simulated Quantiles of $\mathbb{L}$: Method 1

p	SRD	H=0.7	H=0.8	H = 0.9	H = 0.95
.10	0.01 (0.001)	0.02 (0.001)	0.03 (0.000)	0.04 (0.000)	0.05 (0.000)
.15	0.02 (0.000)	0.05 (0.000)	0.08 (0.004)	0.09 (0.005)	0.10 (0.004)
.20	0.04 (0.001)	0.10 (0.000)	0.14 (0.018)	0.18 (0.018)	0.19 (0.018)
.25	0.06 (0.001)	0.16 (0.000)	0.23 (0.003)	0.29 (0.004)	0.31 (0.004)
.30	0.10 (0.002)	0.24 (0.001)	0.34 (0.002)	0.36 (0.002)	0.39 (0.003)
.35	0.13 (0.004)	0.34 (0.010)	0.48 (0.028)	0.50 (0.029)	0.51 (0.025)
.40	0.17 (0.005)	0.44 (0.018)	0.66 (0.005)	0.69 (0.005)	0.72 (0.005)
.45	0.22 (0.010)	0.57 (0.012)	0.87 (0.011)	0.90 (0.011)	0.92 (0.011)
.50	0.28 (0.009)	0.71 (0.010)	1.12 (0.039)	1.47 (0.041)	1.61 (0.033)
.55	0.35 (0.013)	0.91 (0.042)	1.43 (0.034)	1.61 (0.036)	1.75 (0.035)
.60	0.43 (0.020)	1.14 (0.068)	1.79 (0.028)	1.88 (0.029)	1.92 (0.031)
.65	0.54 (0.005)	1.42 (0.009)	2.21 (0.024)	2.49 (0.026)	2.62 (0.027)
.70	0.66 (0.002)	1.76 (0.037)	2.79 (0.041)	3.25 (0.043)	3.87 (0.042)
.75	0.82 (0.004)	2.18 (0.000)	3.52 (0.019)	4.37 (0.020)	4.99 (0.020)
.80	1.00 (0.002)	2.77 (0.017)	4.48 (0.044)	5.01 (0.047)	5.36 (0.045)
.85	1.23 (0.003)	3.56 (0.045)	5.85 (0.018)	6.66 (0.019)	7.11 (0.018)
.90	1.62 (0.002)	4.63 (0.035)	7.74 (0.060)	9.64 (0.063)	10.21 (0.063)
.95	2.26 (0.006)	6.64 (0.120)	11.23 (0.029)	15.61 (0.031)	19.91 (0.033)

Table B.4: Simulated Quantiles of $\mathbb{L}$: Method 2

p	SRD	H=0.7	H=0.8	H = 0.9	H = 0.95
.10	0.01 (0.001)	0.03 (0.001)	0.03 (0.000)	0.04 (0.000)	0.05 (0.000)
.15	0.02 (0.000)	0.05 (0.000)	0.08 (0.004)	0.08 (0.005)	0.10 (0.004)
.20	0.04 (0.001)	0.10 (0.000)	0.15 (0.018)	0.18 (0.018)	0.19 (0.018)
.25	0.07 (0.001)	0.16 (0.000)	0.23 (0.003)	0.29 (0.004)	0.31 (0.004)
.30	0.09 (0.002)	0.24 (0.001)	0.34 (0.002)	0.36 (0.002)	0.39 (0.003)
.35	0.13 (0.004)	0.35 (0.010)	0.50 (0.028)	0.48 (0.029)	0.52 (0.025)
.40	0.17 (0.005)	0.43 (0.018)	0.66 (0.005)	0.69 (0.005)	0.71 (0.005)
.45	0.24 (0.010)	0.57 (0.012)	0.85 (0.011)	0.90 (0.011)	0.93 (0.011)
.50	0.28 (0.009)	0.71 (0.010)	1.10 (0.039)	1.48 (0.041)	1.61 (0.033)
.55	0.35 (0.013)	0.97 (0.042)	1.39 (0.034)	1.62 (0.036)	1.76 (0.035)
.60	0.44 (0.020)	1.13 (0.068)	1.76 (0.028)	1.90 (0.029)	1.92 (0.031)
.65	0.55 (0.005)	1.42 (0.009)	2.25 (0.024)	2.51 (0.026)	2.61 (0.027)
.70	0.65 (0.002)	1.78 (0.037)	2.82 (0.041)	3.28 (0.043)	3.87 (0.042)
.75	0.80 (0.004)	2.18 (0.000)	3.51 (0.019)	4.38 (0.020)	5.01 (0.020)
.80	1.00 (0.002)	2.74 (0.017)	4.42 (0.044)	5.06 (0.047)	5.36 (0.045)
.85	1.23 (0.003)	3.52 (0.045)	5.84 (0.018)	6.67 (0.019)	7.13 (0.018)
.90	1.62 (0.002)	4.67 (0.035)	7.79 (0.060)	9.66 (0.063)	10.24 (0.063)
.95	2.25 (0.006)	6.68 (0.120)	11.19 (0.029)	15.63 (0.031)	20.00 (0.033)

C1 For every $0 < p < 1$, $F_H^{-1}(p) \to \infty$ as $H \to 1$ though not monotonically.

C2 For a fixed p, $H \mapsto F_H^{-1}(p)$ increases at first with H followed by a decrease and increase again.

These observations, in particular, the presence of partial nesting, are important since they provide a way of avoiding the estimation of the Hurst parameter H while constructing C.I.s in certain situations.

Remark B.10. In the boundary case $H = 1$, the fractional Brownian motion is degenerate and equals $\mathbb{B}_H(t) = Zt,\ t \in \mathbb{R}$, for some centered Normal random variable Z (recall (2.11)). So the sample paths of $\mathbb{G}_{a,b}(t) = a\mathbb{B}_H(t) + bt^2$ are convex and therefore coincide with their GCM. Similarly the one sided GCMs also coincide with the original curve. With this explicit characterization of GCM by (**??**) we get $\mathbb{L}_{a,b} = \mathbb{T}_{a,b} = \int_{D_0} \mathcal{S}_{a,b}(z)^2 dz$, where $D_0 = \{z : \mathcal{S}^0_{a,b}(z) = 0\}$. This gives $\mathcal{R} = 1$ and $\Psi = \infty$ with probability 1, leading us to conjecture that the distribution of $\Psi = \Psi_H$ converges to the degenrate distribution at ∞, as $H \uparrow 1$. This, however, appears hard to verify numerically since when H very close to 1, the simulation methods for fBm break down.

APPENDIX C

Triangular Array Functional Central Limit Theorem Under Dependence

C.1 Derivation of a Functional Central Limit Theorem for Triangular array of β-mixing Variables

Let $\{Z_{ni}\}_{n>0,i\in[1,n]}$ be stochastic processes with finite second moment indexed by a totally bounded metric space $(\mathfrak{F},\rho)$. For each n, $Z_{n1},Z_{n2},\ldots,Z_{nn}$ are identically distributed. We will assume that the underlying probability space is rich enough to equip a random variable U distributed as $unif[0,1]$ and independent of $\{Z_{ni}\}_{n>0,i\in[1,n]}$. Let Q_f^n denote the quantile function of $f(Z_{n1})$.
Define strong mixing coefficients

$$\beta_{n,k} = \sup_{l\in\mathbb{Z}} \beta(\mathcal{F}_{n,l-k},\mathcal{G}_{n,l}) \tag{C.1}$$

where $\mathcal{F}_{n,j}=\sigma(Z_{ni}:i\leq j)$ and $\mathcal{G}_{n,j}=\sigma(Z_{ni}:i\geq j)$. Also define the mixing rate function $\beta_{(n)}(u)=\beta_{n,[u]}$ for $u\in\mathbb{R}^+$.

We want to investigate the convergence of

$$G_n = \sum_{i=1}^{n} (Z_{ni} - E(Z_{ni})). \tag{C.2}$$

Define norm $\|.\|_{2,\beta_{(n)}}$ on $\mathfrak{F}$ as

$$\|f\|_{2,\beta_{(n)}} = \sqrt{n \int_0^1 \beta_{(n)}^{-1}(u)(Q_f^n(u))^2 du} \tag{C.3}$$

and class $\mathcal{L}_{2,\beta_{(n)}}(P)$ as the class of real-valued function f such that $\|f\|_{2,\beta_{(n)}} < \infty$.

We make the following assumptions:

Assumption B1: The beta-mixing coefficients of $\{Z_{ni}\}_{n>0, i\in[1,n]}$ satisfy the summability condition for all n, that is, for all n

$$\sum_{k=0}^{\infty} \beta_{n,k} < \infty \tag{C.4}$$

Also the sequence $\tau_n = \left(\sum_{k\geq 0} \beta_{n,k}\right)^{1/2}$ is uniformly bounded by some constant τ.

Assumption B2: For every sequence $\delta_n \downarrow 0$ we have

$$\lim_{n\to\infty} \sup_{\rho(f,g)<\delta_n} \|f - g\|_{2,\beta_{(n)}} = 0. \tag{C.5}$$

Assumption B3: For every $\delta_n \downarrow 0$ we have

$$\lim_{n\to\infty} \int_0^{\delta_n} \sqrt{\log N(\epsilon, \mathfrak{F}, \|.\|_{2,\beta_{(n)}}) d\epsilon} = 0. \tag{C.6}$$

Theorem 20. *Under assumptions B1, B2 and B3, the sequence $\sum_{i=1}^{n} (Z_{ni} - \mathbb{E}Z_{ni})$ is asymptotically tight in $l^\infty(\mathfrak{F})$ and converges in distribution provided it converges marginally.*

To Prove Theorem 20 we need to establish some maximal inequalities (Following [26]). The first result states one such inequality for a finite class of bounded functions.

Lemma C.1. *Let $\mathfrak{G}$ be any finite subclass of $\mathcal{L}_{2,\beta_{(n)}}(P)$ such that for all $g \in \mathfrak{G}$, $\mathbb{E}(g(Z_{n1})) = 0$ and for some positive real numbers a and δ, $\|g\|_\infty \leq a$ and $\|g\|_{2,\beta_{(n)}} \leq \delta$. Let $L(\mathfrak{G}) = \max(1, \log|\mathfrak{G}|)$, where $|\mathfrak{G}|$ denotes the cardinality of $\mathfrak{G}$. Then for any integer $1 \leq q \leq n$, there exists some positive constant C such that*

$$\mathbb{E}\left(\sup_{g\in\mathfrak{G}}\Big|\sum_{i=1}^{n} Z_{ni}(g)\Big|\right) \leq C\left(\delta\sqrt{L(\mathfrak{G})} + aqL(\mathfrak{G}) + an\beta_{n,q}\right).$$

Proof. By Proposition 2 of [26] (which is basically an extension of Berbee's Lemma) we can construct a sequence $\{Z^*_{ni}\}_{i\in[1,n]}$ of real valued random variables such that the random vectors $Y_k = \left(Z_{n,qk+1}, \ldots, Z_{n,q(k+1)}\right)$ and $Y^*_k = \left(Z^*_{n,qk+1}, \ldots, Z^*_{n,q(k+1)}\right)$ satisfy the following conditions:

(i) For any $k > 0$, Y_k and Y^*_k have the same distribution and $\mathbb{P}(Y_k \neq Y^*_k) \leq \beta_{n,q}$.

(ii) The random vectors $(Y^*_{2k})_{k>0}$ are independent and random vectors $(Y^*_{2k-1})_{k>0}$ are independent.

Now we can write

$$\mathbb{E}\left(\sup_{g\in\mathfrak{G}}\Big|\sum_{i=1}^{n} Z_{ni}(g)\Big|\right) \leq \mathbb{E}\left(\sup_{g\in\mathfrak{G}}\Big|\sum_{i=1}^{n} Z^*_{ni}(g)\Big|\right) + \mathbb{E}\left(\sup_{g\in\mathfrak{G}}\Big|\sum_{i=1}^{n}(Z_{ni}(g) - Z^*_{ni}(g))\Big|\right). \tag{C.7}$$

Note that an upper bound on the second term can be given by $2\|g\|_\infty \sum_{i=1}^{n} \mathbf{1}_{(Z_{ni}\neq Z^*_{ni})}$. This combined with the fact that $\mathbb{P}(Z_{ni} \neq Z^*_{ni}) \leq \beta_{n,q}$ yields that

$$\mathbb{E}\left(\sup_{g\in\mathfrak{G}}\Big|\sum_{i=1}^{n}(Z_{ni}(g) - Z^*_{ni}(g))\Big|\right) \leq 2an\beta_{n,q}. \tag{C.8}$$

To control the first term, consider

$$Y_k^*(g) = \sum_{qk+1}^{q(k+1)\wedge n} g\left(Z_{ni}^*\right).$$

These random variables $Y_k^*(g)$ are centered and bounded by qa. By Theorem 1.2 of [55] for any g in $\mathfrak{G}$ we have

$$q^{-1}\mathrm{Var}\mathrm{Y}_\mathrm{k}^*(\mathrm{g}) \leq 4\|\mathrm{g}\|_{2,\beta_{(\mathrm{n})}}^2/\mathrm{n} \leq 4\delta^2/\mathrm{n}.$$

Recall also that $(Y_{2k}^*)_{k>0}$ are independent and so are $(Y_{2k+1}^*)_{k>0}$. Therefore applying Bernstein's inequality ([51] page 193) we get that for any $\lambda > 0$, there exists a positive constant c such that

$$\mathbb{P}\left(\Big|\sum_{i=1}^n Z_{ni}^*(g)\Big| > \lambda\right) \leq 4\exp\left(-c\inf\left((\lambda/\delta)^2, \lambda/(qa)\right)\right). \tag{C.9}$$

It now follows that

$$\begin{aligned}
\mathbb{P}\left(\sup_{g\in\mathfrak{G}}|\sum_{i=1}^n Z_{ni}^*(g)\Big| > \lambda\right) &\leq 4\exp\left(-\left(c\inf\left((\lambda/\delta)^2, \lambda/(qa)\right) - L(\mathfrak{G})\right)^+\right) \\
&\leq 4\exp\left(-\left(c(\lambda/\delta)^2 - L(\mathfrak{G})\right)^+\right) + \\
&\quad 4\exp\left(-\left(c\lambda/(qa) - L(\mathfrak{G})\right)^+\right).
\end{aligned}$$

Let λ_0 and λ_1 be positive number defined by the equations

$$c(\lambda_0/\delta)^2 = L(\mathfrak{G}) \quad \text{and} \quad \mathrm{c}\lambda_1 = \mathrm{aqL}(\mathfrak{G}). \tag{C.10}$$

Hence integrating the above inequality we obtain

$$\begin{aligned}
&\mathbb{E}\left(\sup_{g\in\mathfrak{G}}\Big|\sum_{i=1}^{n} Z_{ni}^{*}(g)\Big|\right) \qquad (C.11)\\
&\leq 4(\lambda_0+\lambda_1)+4\int_0^\infty \exp\left(-\frac{c\lambda^2}{\delta^2}\right)d\lambda+4\int_0^\infty \exp\left(-\frac{c\lambda}{qa}\right)d\lambda\\
&=4(\lambda_0+\lambda_1)+4\delta\sqrt{\frac{\pi}{c}}+\frac{4}{cqa}.
\end{aligned}$$

The equations (C.10) and (C.11) together imply that

$$\mathbb{E}\left(\sup_{g\in\mathfrak{G}}\Big|\sum_{i=1}^{n} Z_{ni}^{*}(g)\Big|\right)\leq C\left(\delta\sqrt{L(\mathfrak{G})}+aqL(\mathfrak{G})\right) \qquad (C.12)$$

for some positive constant C which in turn together with (C.8) implies the result. □

As the next step we extend our result for any bounded class of functions. For any class of functions $\mathfrak{F}$, define $H_{\beta_{(n)}}(.,\mathfrak{F})=\log N(.,\mathfrak{F},\|.\|_{2,\beta_{(n)}})$ and for $\sigma>0$,

$$\varphi_n(\sigma)=\int_0^\sigma \sqrt{H_{\beta_{(n)}}(\epsilon,\mathfrak{F})}d\epsilon. \qquad (C.13)$$

Theorem 21. *Given any $\sigma>0$, let $\mathfrak{F}:=\mathfrak{F}_{n,\sigma}\subset\mathcal{L}_{2,\beta_{(n)}}(P)$ be a class of functions f with $\|f\|_{2,\beta_{(n)}}\leq\sigma$ and $\varphi_n(\sigma)<\infty$. Also assume that for all $f\in\mathfrak{F}_{n,\sigma}$, $|f|\leq M$ for some $M\geq 1$. Then there exists a constant K depending on τ, such that for all positive integer q,*

$$\mathbb{E}\left(\sup_{f\in\mathfrak{F}_{n,\sigma}}\Big|\sum_{i=1}^{n}(Z_{ni}-\mathbb{E}(Z_{ni})(f)\Big|\right)\leq K\left(\varphi_n(\sigma)+\frac{Mq\varphi_n^2(\sigma)}{\sigma^2}+nM\beta_{n,q}\right). \qquad (C.14)$$

Proof. To prove this we use chaining argument similar to proof of Theorem 2 from [26]. And for notational convenience we replace the centered random variable

$Z_{ni} - \mathbb{E}Z_{ni}$ by just Z_{ni}.

First by Lemma 6 of [26] we can find a non-increasing function $H_n(.)$ majorizing $H_{\beta_{(n)}}(., \mathfrak{F}_\sigma)$ such that $x \mapsto x^4 H_n(x)$ in nondecreasing and

$$\int_0^\sigma \sqrt{H_n(t)}dt \leq 4\varphi_n(\sigma). \tag{C.15}$$

We shall assume that $\varphi(\sigma)/\sigma < 2^{-6}\sqrt{n}$ as the other case is trivial. Define $\delta_0 = \sigma$ and $\delta_k = 2^{-k}\delta_0$ for all integers $k \geq 1$. For simplicity we suppress n from the notation. For each nonnegative integer k, we can choose a covering of $\mathfrak{F}_{n,\sigma}$ by brackets $B_{k,j} = [g_{k,j}, h_{k,j}]$ for $i \leq j \leq J_k$ where $\|h_{k,j} - g_{k,j}\|_{2,\beta_{(n)}} \leq \delta_k$ and $J_k \leq \exp(H_n(\delta_k))$. In each bracket $B_{k,j}$ fix a point $f_{k,j}$ from the bracket. Define a mapping $\psi_k : \mathfrak{F}_{n,\sigma} \to \{1, 2, \ldots, J_k\}$ by

$$\psi_k(f) = \min\{j \in \{1, 2, \ldots, J_k\} : f \in B_{k,j}\}$$

and set

$$\Pi_k f = f_{k,\psi_k f} \quad \text{and} \quad \Delta_k f = h_{k,\psi_k f} - g_{k,\psi_k f}.$$

It now follows that

$$|f - \Pi_k f| \leq \Delta_k f \quad \text{and} \|\Delta_k f\| \leq \delta_k.$$

For any $\delta > 0$, define $\mathbb{H}(\delta) = \sum_{\delta_k \geq \delta} H_n(\delta_k)$.

We need to define some parameters to use Lemma C.1. First we define the parameters

$$q(\delta) = \min\{s \in \mathbb{N}^* : \beta_{(n)}(s)/s \leq \mathbb{H}(\delta)/n\}$$

and

$$\epsilon(\delta) = ((q(\delta) - 1) \vee 1)\mathbb{H}(\delta)/n.$$

Let $q_k = q(\delta_{k+1})$, $\epsilon_k = \epsilon(\delta_{k+1})$ and $b_k = 2\delta_k(\mathbb{H}(\delta_{k+1}))^{-1/2}$. Note that both $(q_k)_k$ and

$(b_k)_k$ are non-increasing. The following inequalities which can be easily derived from the definitions will be useful later:

$$\epsilon_k \leq \beta_{(n)}^{-1}(\epsilon_k)\mathbb{H}(\delta_{k+1})/n, \tag{C.16}$$

and

$$q_k \leq \frac{2n\epsilon_k}{\mathbb{H}(\delta_{k+1})}. \tag{C.17}$$

Finally let $N = \min\{k \geq 0 : \delta_k \leq 2^6\varphi(\sigma)/\sqrt{n}\}$ and

$\nu(f) = [\min\{k \geq 0 : q_k\Delta_k f > b_k\}] \wedge N$.

Let I denote the identity operator, then it can be decomposed as

$$I = \Pi_0 + \sum_{k=0}^{N}(I - \Pi_k)\mathbf{1}_{\nu=k} + \sum_{k=1}^{N}(\Pi_k - \Pi_{k-1})\mathbf{1}_{\nu\geq k}.$$

Since $b_{k-1} \geq b_k$, we can write

$$\{\nu(f) \geq k, q_k\Delta_k f > b_{k-1}\} = \{\nu(f) = k, q_k\Delta_k f > b_{k-1}\}.$$

Plugging this relation in the decomposition above we get

$$\begin{aligned} I =& \Pi_0 + (I - \Pi_0)\mathbf{1}_{\nu=0} \\ &+ \sum_{k=1}^{N-1}(I - \Pi_{k-1})\mathbf{1}_{\nu=k, q_k\Delta_k > b_{k-1}} + (I - \Pi_{N-1})\mathbf{1}_{\nu\geq N} \\ &+ \sum_{k=1}^{N-1}(I - \Pi_k)\mathbf{1}_{\nu=k, q_k\Delta_k \leq b_{k-1}} \\ &+ \sum_{k=1}^{N-1}(\Pi_k - \Pi_{k-1})\mathbf{1}_{\nu\geq k, q_k\Delta_k \leq b_{k-1}}. \end{aligned} \tag{C.18}$$

Applying this decomposition we get

$$\mathbb{E}\left(\sup_{f\in\mathfrak{F}_{n,\sigma}}\Big|\sum_{i=1}^{n} Z_{ni}(f)\Big|\right) \leq E_1 + E_2 + E_3 + E_4 + E_5 + E_6.$$

where

$$\begin{aligned}
E_1 &= \mathbb{E}\Big(\sup_{f\in\mathfrak{F}_{n,\sigma}}\Big|\sum_{i=1}^{n} Z_{ni}(\Pi_0 f)\Big|\Big)\\
E_2 &= \mathbb{E}\Big(\sup_{f\in\mathfrak{F}_{n,\sigma}}\Big|\sum_{i=1}^{n} Z_{ni}\left[(f-\Pi_0 f)\mathbf{1}_{\nu(f)=0}\right]\Big|\Big)\\
E_3 &= \sum_{k=1}^{N-1}\mathbb{E}\Big(\sup_{f\in\mathfrak{F}_{n,\sigma}}\Big|\sum_{i=1}^{n} Z_{ni}\left[(f-\Pi_{k-1} f)\mathbf{1}_{\nu(f)=k,q_k\Delta_k f>b_{k-1}}\right]\Big|\Big)\\
E_4 &= \sum_{k=1}^{N-1}\mathbb{E}\Big(\sup_{f\in\mathfrak{F}_{n,\sigma}}\Big|\sum_{i=1}^{n} Z_{ni}\left[(f-\Pi_{k-1} f)\mathbf{1}_{\nu(f)=k,q_k\Delta_k f\leq b_{k-1}}\right]\Big|\Big)\\
E_5 &= \mathbb{E}\Big(\sup_{f\in\mathfrak{F}_{n,\sigma}}\Big|\sum_{i=1}^{n} Z_{ni}\left[(f-\Pi_{N-1} f)\mathbf{1}_{\nu(f)\geq N}\right]\Big|\Big)\\
E_4 &= \sum_{k=1}^{N-1}\mathbb{E}\Big(\sup_{f\in\mathfrak{F}_{n,\sigma}}\Big|\sum_{i=1}^{n} Z_{ni}\left[(\Pi_k f-\Pi_{k-1} f)\mathbf{1}_{\nu(f)\geq k,q_k\Delta_k f\leq b_{k-1}}\right]\Big|\Big).
\end{aligned}$$

Now we shall focus on each of these 6 terms separately and control them

Note that Π_0 ranges over a finite set of functions with cardinality J_0. So by direct application of Lemma C.1 and (C.15) we get for any integer q,

$$\begin{aligned}
E_1 &\leq C\left(\sigma(H_n(\sigma))^{1/2} + 2MqH_n(\sigma) + 2nM\beta_{n,q}\right)\\
&\leq C\left(\sigma(H_n(\sigma))^{1/2} + 2^5\frac{Mq\varphi_n^2(\sigma)}{\sigma^2} + 2nM\beta_{n,q}\right) \qquad \text{(C.19)}
\end{aligned}$$

To control the other terms we shall need some inequalities. First we state a claim which is modified version of Claim 3 from [26].

Claim C.2. *For any integer $k \le N-1$, we have*

$$\sup_{f\in\mathfrak{F}_{n,\sigma}} \|\Delta_k f \mathbf{1}_{q_k\Delta_k f > b_k}\|_1 \le \frac{2\delta_k}{n}(\mathbb{H}(\delta_{k+1}))^{1/2}.$$

Proof of Claim C.2: We will apply Lemma 4 of [26] with $a = b_k/q_k$, $\delta = \delta_k$ and $\epsilon = \epsilon_k$. To check the conditions of the Lemma is satisfied, note that by definition of b_k and (C.17) we have

$$a\sqrt{\epsilon\beta_{(n)}^{-1}(\epsilon)} \ge \frac{\sqrt{\beta_{(n)}^{-1}(\epsilon_k)\mathbb{H}(\delta_{k+1})}}{n\sqrt{\epsilon_k}}\delta_k,$$

and the right hand side is bigger than δ_k by (C.16). Therefore by Lemma 4 of [26] we have,

$$\|\Delta_k f \mathbf{1}_{q_k\Delta_k f > b_k}\|_1 \le 2\delta_k\sqrt{\frac{\epsilon_k}{\beta_{(n)}^{-1}(\epsilon_k)}},$$

which in turn with (C.16) proves that

$$\|\Delta_k f \mathbf{1}_{q_k\Delta_k f > b_k}\|_1 \le \frac{2\delta_k}{n}(\mathbb{H}(\delta_{k+1}))^{1/2}$$

which proves our claim.

Also notice that if $|g| \le h$, then

$$|\sum_{i=1}^{n} Z_{ni}(g)| \le |\sum_{i=1}^{n} Z_{ni}(h)| + 2n\|h\|_1. \tag{C.20}$$

Now to control E_2, note that $|f - \Pi_0 f| \le \Delta_0 f$ and $\{\nu(f) = 0\} = \{q_0\Delta_0 f > b_0\}$. These together with (C.20) imply that

$$E_2 \le \mathbb{E}\Big(\sup_{f\in\mathfrak{F}_{n,\sigma}} \Big|\sum_{i=1}^{n} Z_{ni}\left[\Delta_0 f \mathbf{1}_{q_0\Delta_0 f > b_0}\right]\Big|\Big) + 2n \sup_{f\in\mathfrak{F}_{n,\sigma}} \|\Delta_0 f \mathbf{1}_{q_0\Delta_0 f > b_0}\|_1. \tag{C.21}$$

By Claim 1 from [26] we can write

$$\|\Delta_0 f \mathbf{1}_{q_0\Delta_0 f > b_0} - \mathbb{E}[\Delta_0 f \mathbf{1}_{q_0\Delta_0 f > b_0}]\| \leq (1+\tau)\sigma,$$

so we can apply Lemma C.1 with $a = 2M$ and $\delta = (1+\tau)\sigma$, and thus we obtain for any integer q,

$$\begin{aligned}
&\mathbb{E}\Big(\sup_{f\in\mathfrak{F}_{n,\sigma}}\Big|\sum_{i=1}^{n} Z_{ni}\left[\Delta_0 f \mathbf{1}_{q_0\Delta_0 f > b_0}\right]\Big|\Big) \\
&\leq C\left((1+\tau)\sigma(H_n(\sigma))^{1/2} + 2MqH_n(\sigma) + 2nM\beta_{n,q}\right)
\end{aligned}$$

To control the second term of (C.21) we use Claim C.2 with $k = 0$, and conclude that for any integer q we have

$$E_2 \leq C\left((1+\tau)\sigma(H_n(\sigma))^{1/2} + 2MqH_n(\sigma) + 2nM\beta_{n,q}\right) + 4\sigma(\mathbb{H}(\delta_1))^{1/2}. \qquad \text{(C.22)}$$

Next by the fact $|f - \Pi_k f| \leq \Delta_k$ and (C.20) we can decompose E_3 as below:

$$\begin{aligned}
E_3 \leq &\sum_{k=1}^{N-1}\mathbb{E}\Big(\sup_{f\in\mathfrak{F}_{n,\sigma}}\Big|\sum_{i=1}^{n} Z_{ni}\left[\Delta_{k-1} f \mathbf{1}_{\nu(f)=k, q_k\Delta_k f > b_{k-1}}\right]\Big|\Big) \\
&+ \sum_{k=1}^{N-1} 2n \sup_{f\in\mathfrak{F}_{n,\sigma}} \|\Delta_{k-1} f \mathbf{1}_{\nu(f)=k, q_k\Delta_k f > b_{k-1}}\|_1. \qquad \text{(C.23)}
\end{aligned}$$

Now by Claim 1 from [26] we have,

$$\|\Delta_{k-1} f \mathbf{1}_{\nu(f)=k, q_k\Delta_k f > b_{k-1}} - \mathbb{E}\Delta_{k-1} f \mathbf{1}_{\nu(f)=k, q_k\Delta_k f > b_{k-1}}\|_{2,\beta_{(n)}} \leq (1+\tau)\delta_{k-1},$$

and $\nu(f) = k$ implies that $q_{k-1}\Delta_{k-1} f \leq b_{k-1}$. So to we apply Lemma C.1 with $q = q_{k-1}$, $a = b_{k-1}/q_{k-1}$ and $\delta = (1+\tau)\delta_{k-1}$ and the definitions of the parameters, an upper bound of the first term of (C.23) is given by:

$$C \sum_{k=1}^{N-1} \left((1+\tau)\delta_{k-1} (\mathbb{H}(\delta_k))^{1/2} + 2b_{k-1}\mathbb{H}(\delta_k) \right),$$

which in turn is bounded by $C(5+\tau)\sum_{k=1}^{N-1} \delta_{k-1}(\mathbb{H}(\delta_k))^{1/2}$. For the second term, first note that

$$\{\nu(f) = k, q_k \Delta_k f > b_{k-1}\} \subset \{q_{k-1}\Delta_{k-1} f \leq b_{k-1} < q_k \Delta_k f\}.$$

Hence we have,

$$\|\Delta_{k-1} f \mathbf{1}_{\nu(f)=k, q_k \Delta_k f > b_{k-1}}\|_1 \leq \|\Delta_k f \mathbf{1}_{q_k \Delta_k f > b_k}\|_1.$$

Now application of Claim C.2 provides the following bound for E_3,

$$E_3 \leq C(5+\tau) \sum_{k=1}^{N-1} \delta_{k-1} (\mathbb{H}(\delta_k))^{1/2} + 4 \sum_{k=1}^{N-1} \delta_k (\mathbb{H}(\delta_{k+1}))^{1/2}. \tag{C.24}$$

Now similar arguments as above gives,

$$\begin{aligned} E_4 \leq & \sum_{k=1}^{N-1} \mathbb{E}\Big(\sup_{f \in \mathfrak{F}_{n,\sigma}} \Big| \sum_{i=1}^{n} Z_{ni} \left[\Delta_k f \mathbf{1}_{\nu(f)=k, q_k \Delta_k f \leq b_{k-1}} \right] \Big| \Big) \\ & + \sum_{k=1}^{N-1} 2n \sup_{f \in \mathfrak{F}_{n,\sigma}} \|\Delta_k f \mathbf{1}_{\nu(f)=k, q_k \Delta_k f \leq b_{k-1}}\|_1. \end{aligned} \tag{C.25}$$

As argued above the $(2, \beta_{(n)})$-norm of the desired quantity is bounded above by $(1+\tau)\delta_k$ and hence Lemma C.1 can be applied with $q = q_k$, $a = b_{k-1}/q_k$ and $\delta = (1+\tau)\delta_k$ which gives an upper bound of the first term as:

$$C \sum_{k=1}^{N-1} \left((1+\tau)\delta_k (\mathbb{H}(\delta_k))^{1/2} + b_{k-1}\mathbb{H}(\delta_k) + b_{k-1}\mathbb{H}(\delta_{k+1}) \right).$$

As $x^4 H_n(x)$ is nondecreasing, we have $H_n(\delta_{k+1}) \leq 16 H_n(\delta_k)$ and therefore $\mathbb{H}(\delta_{k+1}) \leq 17\mathbb{H}(\delta_k)$. So finally substituting the definition of b_k in the above quantity an upper bound for the first term becomes $C(37+\tau)\sum_{k=1}^{N-1}\delta_k(\mathbb{H}(\delta_k))^{1/2}$. Noting that

$$\{\nu(f) = k, q_k\Delta_k f \leq b_{k-1}\} \subset \{b_k < q_k\Delta_k f\},$$

we can bound the second term here by the same quantity used for the second term of E_3. Hence,

$$E_4 \leq C(37+\tau)\sum_{k=1}^{N-1}\delta_k(\mathbb{H}(\delta_k))^{1/2} + 4\sum_{k=1}^{N-1}\delta_k(\mathbb{H}(\delta_{k+1}))^{1/2}. \tag{C.26}$$

For the next term we can write

$$E_5 \leq \mathbb{E}\Big(\sup_{f\in\mathfrak{F}_{n,\sigma}}\Big|\sum_{i=1}^{n} Z_{ni}\left[\Delta_{N-1} f\mathbf{1}_{\nu(f)\geq N}\right]\Big|\Big) + 2n\sup_{f\in\mathfrak{F}_{n,\sigma}}\|\Delta_{N-1}f\|_1. \tag{C.27}$$

As $\nu(f)\geq N$ implies that $q_{N-1}\Delta_{N-1}f \leq b_{N-1}$, Lemma C.1 can be applied with $q = q_{N-1}$, $a = b_{N-1}/q_{N-1}$ and $\delta = (1+\tau)\delta_{N-1}$ and thus the first term of the above inequality can be majorized by

$$C(1+\tau)\delta_{N-1}(\mathbb{H}(\delta_{N-1}))^{1/2} + 2b_{N-1}\mathbb{H}(\delta_N)$$

which in turn is bounded above by $C(5+\tau)\delta_{N-1}(\mathbb{H}(\delta_N))^{1/2}$. Since $\|\Delta_{N-1}f\|_1 \leq \|\Delta_{N-1}f\|_2 \leq \|\Delta_{N-1}f\|_{2,\beta_n}/\sqrt{n} \leq \delta_{N-1}/sqrtn$, we get

$$E_5 \leq C(5+\tau)\delta_{N-1}(\mathbb{H}(\delta_N))^{1/2} + 16\varphi_n(\sigma). \tag{C.28}$$

The term E_6 can be treated exactly similarly to the first term of E_4. We apply Lemma C.1 here with $q = q_k$, $a = 4b_{k-1}/q_k$ and $\delta = 3(1+\tau)\delta_k$, and obtain the final upper bound for this term

$$E_6 \leq C(307 + 3\tau) \sum_{k=1}^{N-1} \delta_k (\mathbb{H}(\delta_k))^{1/2}. \tag{C.29}$$

Finally summing over (C.19), (C.22), (C.24), (C.26), (C.28) and (C.29) we obtain the final upper bound of the expectation as

$$K' \sum_{k \geq 0} \delta_k (\mathbb{H}(\delta_k))^{1/2} + 16\varphi_n(\sigma) + 2^6 CM \left(\frac{q\varphi_n^2(\sigma)}{\sigma^2} + n\beta_{n,q} \right).$$

And finally to complete the proof the first term can be bounded above by $16K'\varphi_n(\sigma)$ as shown in page 418 of [26]. □

Now we shall state the final result towards our proof to asymptotic equicontinuity.

Theorem 22. *Let $\sigma > 0$ be a positive number and define $\mathcal{F}_{n,\sigma}$ be defined as in Theorem 21 which has an envelope F_n. Let B_n be a function on $\mathbb{R}^+$ defined as $B_n(x) = \int_0^x \beta_{(n)}^{-1}(t)dt$. For any measurable function h, define*

$$\delta_h^n(\epsilon) = \sup_{t \leq \epsilon} Q_h^n(t) \sqrt{B_n(t)}.$$

Assume that if $\epsilon_n \to 0$ as $n \to \infty$, $\lim_{n\to\infty} \delta_{F_n}^n(\epsilon_n) = 0$. Then there exists some constant depending on τ^2 such that for any positive integer n, we have

$$\mathbb{E}\left(\sup_{f \in \mathfrak{F}_{n,\sigma}} \left| \sum_{i=1}^{n} (Z_{ni} - \mathbb{E}Z_{ni}(f)) \right| \right) \leq A\varphi_n(\sigma) \left[1 + \frac{\delta_{F_n}^n(1 \wedge \epsilon(\sigma, n))}{\sigma} \right], \tag{C.30}$$

where $\epsilon(\sigma, n)$ is the unique solution of the equation

$$x^2 / B_n(x) = \varphi_n^2(\sigma)/(n\sigma)^2.$$

Proof. For notational simplicity denote $\epsilon(\delta, n)$ by ϵ and set $q = \beta_{(n)}^{-1}(\epsilon)$. Then by

(C.17) we have

$$q \leq \frac{\epsilon n \sigma^2}{\varphi_n^2(\sigma)}. \tag{C.31}$$

With the notations $M = Q^n_{F_n}(\epsilon)$ and $\mathfrak{F}_M = \{f\mathbf{1}_{F_n \leq M} : f \in \mathfrak{F}\}$ we can write

$$\mathbb{E}\left(\sup_{f\in\mathfrak{F}_{n,\sigma}}\Big|\sum_{i=1}^{n} Z_{ni}(f)\Big|\right) \leq \mathbb{E}\left(\sup_{f\in\mathfrak{F}_{n,\sigma,M}}\Big|\sum_{i=1}^{n} Z_{ni}(f)\Big|\right) + 2n\|F_n\mathbf{1}_{F_n>M}\|_1. \tag{C.32}$$

To bound the first term of the right hand side of the above inequality we use Theorem 21. Using relation (C.31) with the inequality (C.14), we get that the first term is bounded above by $K(\varphi_n(\sigma) + 2n\epsilon Q^n_{F_n}(\epsilon))$. Now by the definition of ϵ, we have,

$$2n\epsilon Q^n_{F_n}(\epsilon) \leq \frac{2n\epsilon}{\sqrt{B_n(\epsilon)}}\delta^n_{F_n}(\epsilon) \leq 2\delta^n_{F_n}(\epsilon)\frac{\varphi_n(\sigma)}{\sigma}.$$

For the second term, we apply Lemma 4 form [26] to obtain

$$\|F_n\mathbf{1}_{F_n>M}\|_1 \leq \frac{2\delta^n_{F_n}(\epsilon)\epsilon}{\delta^n_{F_n}(\epsilon)} \leq 2\delta^n_{F_n}(\epsilon)\frac{\varphi_n(\sigma)}{n\sigma}.$$

Hence,

$$\mathbb{E}\left(\sup_{f\in\mathfrak{F}_{n,\sigma}}\Big|\sum_{i=1}^{n} Z_{ni}(f)\Big|\right) \leq K\varphi_n(\sigma) + 2(K+2)\frac{\delta^n_{F_n}(\epsilon)\varphi_n(\sigma)}{\sigma}$$

and this completes the proof. □

Proof. Proof of Theorem 20 We will apply Theorem 22 on $\mathfrak{F}_{n,\sigma} = \{f - g : f, g \in \mathfrak{F}, \|f-g\|_{2,\beta_{(n)}} \leq \sigma\}$. Under assumption B3, the class of functions $\mathcal{F}_{n,\sigma}$ admits an envelope F_n where $\{F_n\}_{n=1}^{\infty}$ is uniformly integrable. That implies $\delta^n_{F_n}(\epsilon) \to 0$ as $\epsilon \to 0$. Also $\epsilon(\sigma, n) \to 0$ as $n \to \infty$ by definition. So for large enough n we have

$$\mathbb{E}\left(\sup_{f\in\mathfrak{F}_{n,\sigma}}\Big|\sum_{i=1}^{n}(Z_{ni} - \mathbb{E}Z_{ni}(f))\Big|\right) \leq 2A\varphi_n(\sigma).$$

This along with Assumption B3 proves Theorem 20. $\square$

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] Backbone link traces from the internet2 network.

[2] Nasa giss surface temperature analysis.

[3] I. (2007). Intergovernmental panel on climate change - facts, June 2014.

[4] J. Abrevaya and J. Huang. On the bootstrap of the maximum score estimator. *Econometrica*, 73(4):1175–1204, 2005.

[5] P. Abry and D. Veitch. Wavelet analysis of long-range-dependent traffic. *Institute of Electrical and Electronics Engineers. Transactions on Information Theory*, 44(1):2–15, 1998.

[6] R. J. Adler and J. E. Taylor. *Random Fields and Geometry*. Springer, New York, 2005.

[7] J. an and Q. Yao. *Nonlinear Time Series. Nonparametric and Parametric Methods*. Springer, 2003.

[8] D. W. Andrews. An introduction to econometric applications of empirical process theory for dependent random variables. *Econometric Reviews*, 12(2):183–216, 1993.

[9] D. Anevski and O. Hössjer. A general asymptotic scheme for inference under order restrictions. *The Annals of Statistics*, 34(4):1874–1930, 2006.

[10] M. A. Arcones. M-estimators converging to a stable limit. *Journal of multivariate analysis*, 74(2):193–221, 2000.

[11] M. Banerjee. *Likelihood Ratio Inference in Regular and Non-regular Problems.* PhD thesis, University of Washington, 2000.

[12] M. Banerjee. Likelihood based inference for monotone response models. *The Annals of Statistics*, 35(3):931–956, 2007.

[13] M. Banerjee. Inference in exponential family regression models under certain shape constraints. In *In Advances in Multivariate Statistical Methods, Statistical Science and Interdisciplinary Research*, volume 4, pages 249–272. Citeseer, 2009.

[14] M. Banerjee and J. A. Wellner. Likelihood ratio tests for monotone functions. *The Annals of Statistics*, 29(6):1699–1731, 2001.

[15] M. Banerjee and J. A. Wellner. Confidence intervals for current status data. *Scandinavian Journal of Statistics*, 32(3):405–424, 2005.

[16] H. C. Berbee. Random walks with stationary increments and renewal theory. *MC Tracts*, 112:1–223, 1979.

[17] P. Billingsley. *Convergence of probability measures.* Wiley, New York, 1999.

[18] F. Boussama. Ergodicity, mixing and estimation in garch models. *Unpublished Ph. D. Dissertation, University of Paris*, 7, 1998.

[19] H. D. Brunk. Estimation of isotonic regression. In M. L. Puri, editor, *Nonparametric Techniques in Statistical Inference*, pages 177–197. Cambridge University Press, London, 1970.

[20] Z. Cai and E. Ould-SaıÌĹd. Local m-estimator for nonparametric time series. *Statistics & probability letters*, 65(4):433–449, 2003.

[21] M. Caner. M-estimators with non-standard rates of convergence and weakly dependent data. *Journal of statistical planning and inference*, 136(4):1207–1219, 2006.

[22] Y. Cheng and J. G. De Gooijer. Bahadur representation for the nonparametric m-estimator under alpha-mixing dependence. *Tinbergen Institute Discussion Papers No. TI*, 67(4), 2005.

[23] Y. A. Davydov. Convergence of distributions generated by stationary stochastic processes. *Theory of Probability & Its Applications*, 13(4):691–696, 1968.

[24] Y. A. Davydov. Mixing conditions for markov chains. *Teoriya Veroyatnostei i ee Primeneniya*, 18(2):321–338, 1973.

[25] P. Doukhan, P. Massart, and E. Rio. The functional central limit theorem for strongly mixing processes. In *Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques*, volume 30, pages 63–82. Gauthier-Villars, 1994.

[26] P. Doukhan, P. Massart, and E. Rio. Invariance principles for absolutely regular empirical processes. In *Annales de l'IHP Probabilités et statistiques*, volume 31, pages 393–427. Elsevier, 1995.

[27] P. Doukhan, G. Oppenheim, and M. S. Taqqu. *Theory and applications of long-range dependence.* Springer, 2003.

[28] J. Fan, T.-C. Hu, and Y. K. Truong. Robust non-parametric function estimation. *Scandinavian Journal of Statistics*, pages 433–446, 1994.

[29] G. Faÿ, E. Moulines, F. Roueff, and M. S. Taqqu. Estimators of long-memory: Fourier versus wavelets. *Journal of Econometrics*, 151(2):159–177, 2009.

[30] P. Groeneboom. Estimating a monotone density. In M. L. Lucien and R. A. Olshen, editors, *Proceeding of the Berkeley Conference in Honor of Jezry Neyman and Jack Kiefer*, volume II, 1985.

[31] P. Groeneboom. Brownian motion with a parabolic drift and airy functions. *Probability theory and related fields*, 81(1):79–109, 1989.

[32] P. Groeneboom and G. Jongbloed. Density estimation in the uniform deconvolution model. *Statistica Neerlandica*, 57(1):136–157, 2003.

[33] P. Groeneboom and J. A. Wellner. *Information bounds and nonparametric maximum likelihood estimation*, volume 19. Springer, 1992.

[34] S. Gupta. A note on the asymptotic distribution of lasso estimator for correlated data. *Sankhya A*, 74(1):10–28, 2012.

[35] J. Haslett and A. E. Raftery. Space-time modelling with long-memory dependence: Assessing ireland's wind power resource. *Applied Statistics*, 38(1):1–50, 1989.

[36] P. J. Huber. The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 221–233, 1967.

[37] C. Hurvich, G. Lang, and P. Soulier. Estimation of long memory in the presence of a smooth nonparametric trend. *Journal of the American Statistical Association*, 100(471):853–871, 2005.

[38] M. Hussian, A. Grimvall, O. Burdakov, and O. Sysoev. Monotonic regression for the detection of temporal trends in environmental quality data. *MATCH Commun. Math. Comput. Chem*, 54:535–550, 2005.

[39] P. D. Jones and M. E. Mann. Climate over past millennia. *Reviews of Geophysics*, 42(2), 2004.

[40] J. Kim and D. Pollard. Cube root asymptotics. *The Annals of Statistics*, 18(1):191–219, 1990.

[41] E. Mammen. Estimating a smooth monotone regression function. *The Annals of Statistics*, pages 724–740, 1991.

[42] C. F. Manski. Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator. *Journal of econometrics*, 27(3):313–333, 1985.

[43] D. Meal. Statistical analysis for monotone trend. *The NCSU Water Quality Group Newsletter*, pages 1–11, 2011.

[44] A. Mokkadem. Mixing properties of arma processes. *Stochastic processes and their applications*, 29(2):309–315, 1988.

[45] H. Mukerjee. Monotone nonparametric regression. *The Annals of Statistics*, pages 741–750, 1988.

[46] N. A. A. Q. S. (NAAQS). Amendment to the "clean air act", 1990.

[47] W. K. Newey and D. McFadden. Large sample estimation and hypothesis testing. *Handbook of econometrics*, 4:2111–2245, 1994.

[48] H. Oodaira. On strassen's version of the law of the iterated logarithm for gaussian processes. *Probability Theory and Related Fields*, 21(4):289–299, 1972.

[49] J. K. Pal and M. Woodroofe. Large sample properties of shape restricted regression estimators with smoothness adjustments. *Statistica Sinica*, 17(4):1601, 2007.

[50] M. Peligrad and S. Utev. A new maximal inequality and invariance principle for stationary sequences. *The Annals of Probability*, 33(2):798–815, 2005.

[51] D. Pollard. *Convergence of stochastic procresses*. Springer, Berlin, 1984.

[52] J. Ramsay. Estimating smooth monotone functions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(2):365–375, 1998.

[53] B. P. Rao. Estimation of unimodal density. *Sankhyā - Series A*, 31(1):23–36, 1969.

[54] S. I. Resnick. *A Probability Path*. Birkhäuser, Boston, 1999.

[55] E. Rio. Covariance inequalities for strongly mixing processes. In *Annales de l'institut Henri Poincaré (B) Probabilités et Statistiques*, volume 29, pages 587–597. Gauthier-Villars, 1993.

[56] E. Rio. About the lindeberg method for strongly mixing sequences. *ESAIM, Probabability and Statistics*, 1:35–61, 1995.

[57] T. Robertson, F. Wright, and R. L. Dykstra. *Order Restricted Statistical Inference*. Wiley, New York, 1988.

[58] P. M. Robinson. Inference on nonparametrically trending time series with fractional errors. *Econometric Theory*, 25(6):1716–1733, 2009.

[59] R. T. Rockafellar. *Convex Analysis*, volume 28 of *Princeton Mathematics Series*. Princeton University Press, Princeton, 1970.

[60] E. J. Steig, D. P. Schneider, S. D. Rutherford, M. E. Mann, J. C. Comiso, and D. T. Shindell. Warming of the antarctic ice-sheet surface since the 1957 international geophysical year. *Nature*, 457(7228):459–462, 2009.

[61] S. Stoev, M. S. Taqqu, C. Park, and J. Marron. On the wavelet spectrum diagnostic for hurst parameter estimation in the analysis of internet traffic. *Computer Networks*, 48(3):423–445, 2005.

[62] J. Sun and M. Woodroofe. Adaptive smoothing for a penalized npmle of a non-increasing density. *Journal of statistical planning and inference*, 52(2):143–159, 1996.

[63] M. S. Taqqu. Weak convergence to fractional brownian motion and to the rosenblatt process. *Probability Theory and Related Fields*, 31(4):287–302, 1975.

[64] M. S. Taqqu. Convergence of integrated processes of arbitrary hermite rank. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 50(1):53–83, 1979.

[65] A. W. van der Vaart and J. A. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer, New York, 1996.

[66] W. Willinger, M. S. Taqqu, W. E. Leland, and D. V. Wilson. Self-similarity in high-speed packet traffic: analysis and modeling of ethernet traffic measurements. *Statistical science*, pages 67–85, 1995.

[67] M. Woodroofe and J. Sun. A penalized maximum likelihood estimate of f (0+) when f is non-increasing. *Statistica Sinica*, 3(2):501–515, 1993.

[68] W. B. Wu et al. M-estimation of linear models with dependent errors. *The Annals of Statistics*, 35(2):495–521, 2007.

[69] W. B. Wu and Z. Zhao. Inference of trends in time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(3):391–410, 2007.

[70] W. M. Wu, Wei Biao and G. Mentz. Isotonic regression: Another look at the changepoint problem. *Biometrika*, 88(3):793–804, 2001.

[71] O. Zhao and M. Woodroofe. Estimating a monotone trend. *Statistica Sinica*, 22(1):359–378, 2012.