This is to certify that the
dissertation entitled

Inference on Long Memory Processes

presented by

Hongwen Guo

has been accepted towards fulfillment
of the requirements for the

Doctoral degree in Department of Statistics and Probability

Major Professor's Signature

May 10, 2006

Date

Inference on long memory processes

By

Hongwen Guo

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Statistics and Probability

2006

ABSTRACT

Inference on long memory processes

by

Hongwen Guo

This dissertation discusses regression models with a long memory heteroscedastic error process with long memory parameter H. When the regression function is formulated nonparametrically and uniformly on the unit interval, the consistency and the finite dimensional weak convergence of the regression function and variance function estimators are established. For the regression function estimators, the asymptotic normality is established for the values of the long memory parameter $1/2 < H < 1$; while for the heteroscedastic function estimators, the asymptotic normality is established for $1/2 < H < 3/4$, non-normality for $3/4 < H < 1$. We also establishes the uniform convergence rate of the regression function estimators for a large class of innovations, including bounded and Gaussian innovations. Additionally, the local Whittle estimator of H based on the standardized nonparametric residuals is shown to be $\log(n)$-consistent and the finite dimensional distributions of the studentized versions of the regression function estimators are shown to be asymptotically normal.

While when the regression function is linear, the design is long memory Gaussian with the long memory parameter h, in some circumstance, the first order asymptotic distribution of the least square estimator of the slope parameter is observed to be degenerate. Under some additional mild conditions, the second order asymptotic distribution of

this estimator is shown to be normal whenever $h+H < 3/2$; non-normal otherwise. The asymptotic distribution of the kernel type estimators of the heteroscedasticity function is found to be normal whenever $H < (1+h)/2$, and non-normal otherwise. In addition, an estimator of H based on pseudo residuals in a more general heteroscedastic regression model is shown to be $\log(n)$-consistent. We also discuss the consistency of a cross validation type estimator of the heteroscedasticity function in a more general regression model under the assumed long memory set up.

All of these findings are then used to propose a lack-of-fit test of a parametric regression model. Some simulations and an application to currency exchange rate data sets are included in this study.

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my advisor, Professor Hira L. Koul, for his guidance, encouragement and patience throughout my graduate studies at the Michigan State University. It is not often to find an advisor who is willing to help with every detail of research and academic problems. His advice and encouragement were essential to the completion of this dissertation.

My thanks also go to Professors Richard Baillie for the insight of economics and finance application; Yimin Xiao for helpful suggestions and friendly discussion during the course of my study; Lijian Yang for his lectures and his suggestion of my future career; and Dennis Gilliland for serving on my dissertation committee.

I also wish to thank the Department of Statistics and Probability of Michigan State University, faculty, staff, particularly, Professor Stapleton, and graduate students, for providing a good working atmosphere,

The Laha Travel Awards, the graduate travel grant from MSU and the Department of Statistics and Probability enabled me to attend the JSM 2005 where the first part of this dissertation was presented. The grant from the 9th New Researcher Meeting and from the Department of Statistics and Probability of MSU enabled me to attend the 9th New Researcher Meeting of Statistics and Probability and JSM2006, where the second part was presented.

Last, but not the least, I would like to thank my husband Zhongjun and my daughter Lan for their understanding and love during the past four years. This dissertation would never have been begun, nonetheless finished, without their support.

TABLE OF CONTENTS

List of Tables

List of Figures

Chapter 1

Introduction

The aim of inference is to recover a relationship between variables disturbed by random noise. When the noise is an independent sequence, a large number of classical results about large sample theory can be applied for inference of nonparametric or parametric models. In particular, the Central Limit Theorem can be used to derive the asymptotic distribution of the concerned statistic. Even with weak dependence (including m-dependence, mixing) in the noise, some of the above results still apply. Whence the degree the dependence exceeds certain point, new phenomena arise. Long memory processes fall in this category. In such processes the autocorrelations decay to zero so slowly that their sum diverges.

The work of Hurst (1951, 1956) on the Nile river data aroused the wide interest of mathematicians, statisticians, econometricians, physicists and others in long memory study. In 1968, Mandelbrot, Wallis, and van Ness published a series of papers, providing a solid mathematical model for a long memory process, which is called fractional Brownian motion. Later, Granger and Joyeux (1980), and Hosking (1981) independently proposed another model from the economics point of view, named fractional autoregressive integrated moving average (FARIMA) model. Meanwhile, more and more scientists have found the presence of long memory in their data in economics, finance, hydrology, physics, telecommunication, and other sciences. Taqqu (1986) cited more than 250 theoretical papers on long memory. Beran (1992, 1994) gives numerous examples from a variety of scientific areas. The survey paper of Baillie (1996) includes 138 papers on long memory processes and their applications, particularly in economics and finance. Since long memory sequences have properties rather different form those of classical independent sequences, even mixing sequences, it has been under intensive studies, and a large variety of applications make them more exciting, cf. Dehling, Mikosch, and Sørensoen (2002), Doukhan, Oppenheim and Taqqu (2003) and Robinson (2003) .

1.1 Long memory processes

1.1.1 Mathematical models

There are two typical mathematical models for long memory stochastic processes: fractional Brown motion (FBM) and fractional autoregressive integrated moving average (FARIMA).

In 1951-1956, Hurst studied the minimum water level of the Nile river from 622 to 1284 AD. He found there were long period behavior of the river: long time periods of dryness were followed by long time periods of yearly returning floods. Floods had the effect of fertilizing the soil so that in flood years the yield of crop was particularly abundant, as described in the Bible: "Seven years of great abundance are coming throughout the land of Egypt, but seven years of famine will follow them." (ref. Beran 1994). This is the so called "Hurst effect" or long memory effect.

Mandelbrot and Wallis (1968, 1969), and Mandelbrot and van Ness (1968b) formulated fractional Brownian motion (FBM) to model the Hurst effect, which is a self-similar Gaussian process with mean zero. Let $\{Y_n, n = 1, 2, \cdots,\}$ be a FBM,

$X_n = Y_{n+1} - Y_n$ be the difference process. Then X_n is a stationary Gaussian process (called fractional Gaussian noise). Its covariances satisfy the following condition: $\forall\ t, k > 0$,

$$\gamma_k =: \operatorname{cov}(X_t, X_{t+k}) = \frac{\sigma^2}{2}|(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}|, \quad 0 < H < 1,$$

where $\sigma^2 = EX_1^2$. The parameter H is called the Hurst index or long memory index. The covariance γ_k and the spectral density function $f(\lambda)$ of this process satisfy

$$\begin{aligned} \gamma_k &\sim Ck^{2H-2}, \qquad k \to \infty; \\ f(\lambda) &\sim C\lambda^{1-2H}, \qquad \lambda \to 0. \end{aligned} \tag{1.1.1}$$

Note that for $1/2 < H < 1$, $\sum_k \gamma_k = \infty$. In this case, the process X_n is said to have long memory (or long range dependence),

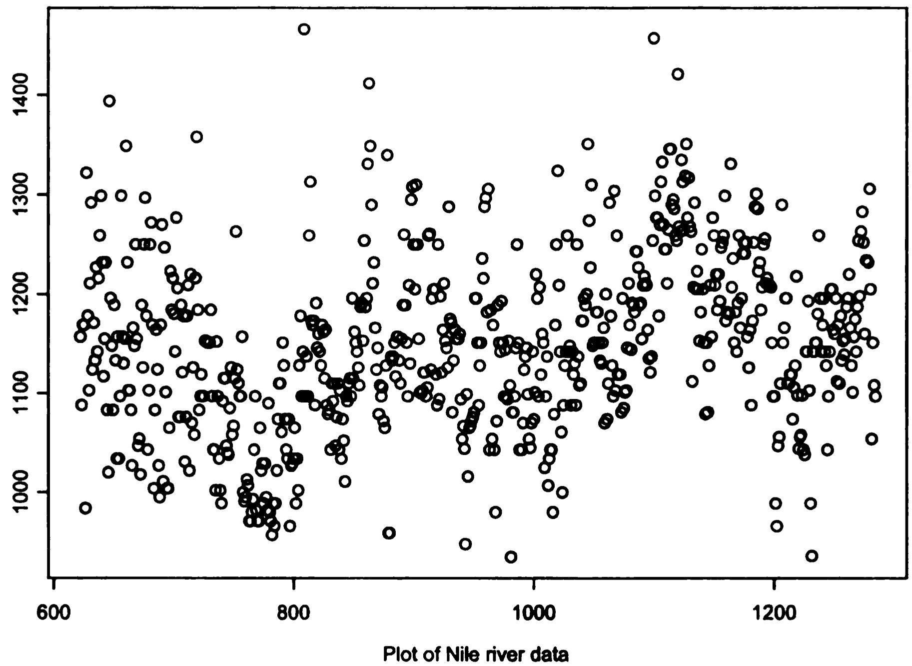
Another model is proposed by Granger and Joyeux (1980), Hosking (1981) independently. Let X_t be a stationary process generated by the following dynamic system

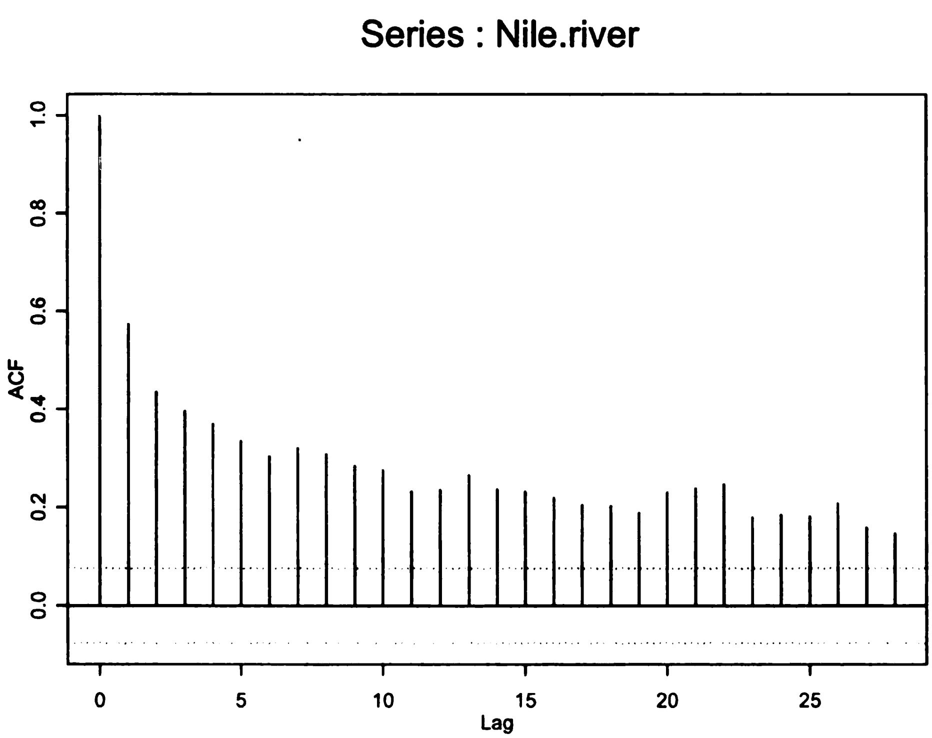
$$\phi(B)(1-B)^d X_t = \psi(B)\epsilon_t, \tag{1.1.2}$$

where $-1/2 < d := H - 1/2 < 1/2$, ϵ_t are i.i.d. random variables, B is the back shift operator, and ϕ and ψ are $p, q-$polynomials with roots outside unit circles. The process X_t is called an $FARIMA(p, d, q)$ process. When $0 < d < 1/2$, It satisfies (1.1.1), hence has long memory. By definition, we can see it is generated from the classical $ARIMA$ by allowing the difference order to be fractions. In the case $f(0)$ is finite and positive, we say the process has short memory.

The Nile river data posses these characteristics, see Figure 1.1, where ACF is the sample auto-covariance of the data.

Figure 1.1: Scatter plot and ACF of the Nile river data.





1.1.2 Importance

In statistics and probability theory, a large number of limit theorems are based on the assumption that the sequences are independent random variables. When the dependence of observations are weak, where the covariances decay to zero exponentially, these results may continue hold. Once the rate of decay is slow, hyperbolically, say, the long memory phenomena arise. In this case, the covariances are not summable, which leads to different results of large sample behaviors of various statistics. We give a simple example to illustrate this point.

Let X_n be an i.i.d. sequence with mean $\mu = EX_1$ and $0 < EX_1^2 < \infty$. By the classical Central Limit Theorem, we have

$$n^{1/2}(\bar{X} - \mu) \rightarrow_d N(0, \sigma_1),$$

where $\bar{X}$ is the sample mean, $\sigma_1 > 0$ is some constant. We can see $\bar{X}$ converges to μ at a rate of $\sqrt{n}$.

Now, if we assume X_n is a stationary long memory process, under some additional conditions, an available result from Avram and Taqqu (1987), Taqqu (1995) or Davydov (1971) is

$$n^{1-H}(\bar{X} - \mu) \rightarrow_d N(0, \sigma_2), \quad \text{for } 1/2 < H < 1, \text{ and some } \sigma_2 > 0.$$

Here, the convergence rate is slower than $\sqrt{n}$. In some cases, the limit distribution is not normal. What makes things more difficult is that because H is unknown in practice, in order to conduct inference about μ, we now have to find an $\log(n)$-consistent estimator of H.

1.2 Applications of long memory processes

Taqqu (1968), Beran (1994), Baillie (1996), Robinson (2003), and Doukhan et. al. (2003), among others, have cited many applications of long memory processes. Here we quote a few examples from the literature.

Geophysics: Besides the above Nile river data, Hurst (1951), Mandelbrot and Wallis (1968) had two more data of FBM: Temperature data and Tree Ring series. Figure 1.2 provides 5405 annual tree ring measurements observed at Mount Campito from 3436 BC to 1969 AD. We can observe the slow decay of autocovariances from the ACF plot, thus it also posses the Hurst effect. Hipel and Mcleod (1978) and Noakes et al. (1988) provide analysis of other tree ring series, mud varvs on river floors, high tides.

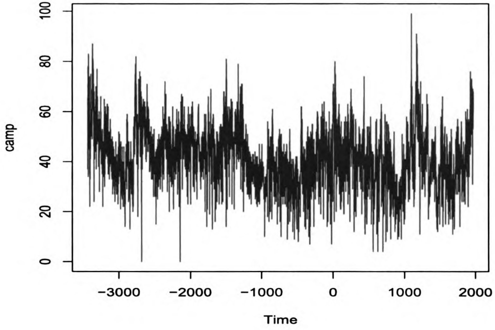
Economics and finance: In economics and finance, data sets, such as GNP, Asset price, stock return, exchange rates, etc. have been analyzed by many researchers. Some of these data are found to have long memory phenomena, see Baillie (1996) and Robinson (2003). The most noted long memory phenomena lie in the absolute returns and the volatility (square return) process of high frequency finance data.

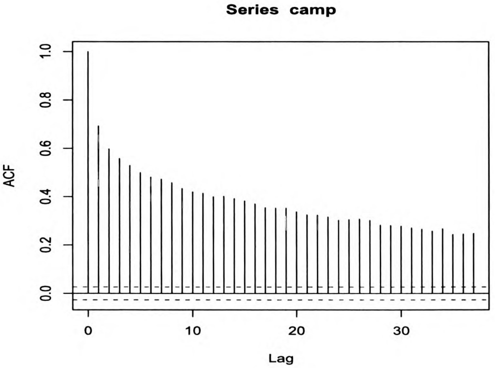
Physiology: Among the models used in research for complex physiologic signals, long memory time series models have been used to evaluate the heart beat rate, and other physiological time series. A typical realization of fractional Brownian motion is a fractal. Many physiological time series can be thought as fractals or multifractals. See www.physionet.org/ and reference there for more on this.

Physics: Either for systems which are far from equilibrium (e.g., turbulent flows), or in equilibrium but very close to a critical point (e.g., the transition from a solid to liquid phase, or from a non-magnetic phase to a magnetized one), phase transitions may have fluctuations which decay like power laws, and so do many non-equilibrium systems. In physics, it is called $1/f$ noise.

Telecommunication: Long memory also finds its application in computer networks, see Willinger et. al. (1998). The web site www.cs.bu.edu/pub/barford/ss_lrd.html provides links and very brief summaries for some of the work done in this area.

Figure 1.2: Plot and ACF of Mount Campito annual tree ring measurements





1.3 Main results in this research

In addition to the long memory phenomena in economics and finance data there is another well known fact– "volatility smile": the conditional variance function of the time series varies upon time. To capture this phenomenon, Engle (1983) proposed the parametric ARCH model, later it is generalized to GARCH by others. It is part of the reason that Engle, together with Ganger, won the Nobel prize for economics in 2003.

Because of the mixture of long memory processes with ARMA-type or ARCH-type models, there is a need to develop a variety of consistent estimators and test procedures to obtain statistical inference from some real data.

Many authors consider the regression models with long memory and homoscedastic errors under fixed or random design, and established various asymptotic results for the regression function estimators. In order to apply statistical methodology to economic or financial data, especially to capture the changing conditional variance, the model with heteroscedasticity shall be considered and developed. In addition, without enough knowledge of the models, nonparametric regression methods shall be investigated for estimating regression function or conditional variance function. The study in this dissertation is aimed at solving the above problems and applying the theoretical results to real data.

In this dissertation, we consider the heteroscedastic regression model,

$$Y_t = r(X_t) + \sigma(X_t)u_t, \qquad t = 1, 2, \cdots, \tag{1.3.1}$$

where u_t is the long memory moving average process with the long memory parameter H, and where X_t are either uniform design on $[0, 1]$ or a long memory Gaussian process and independent of u_t. Chapter 2 discusses the first case while Chapter 3 discusses the later case.

In Chapter 2, the consistency and the finite dimensional weak convergence of the regression function and variance function estimators are established. For the regression function estimators, the asymptotic normality is established for the values of the long memory parameter $1/2 < H < 1$; while for the heteroscedastic function estimators, the asymptotic normality is established for $1/2 < H < 3/4$, non-normality for $3/4 < H < 1$. This chapter also establishes the uniform convergence rate of the regression function estimators to be $(nb)^{1-H}/\log^2 n$ for $1/2 < H < 1$ and for a large class of innovations, including bounded and Gaussian innovations, where n is series size and b is the bandwidth used in estimating regression function. Additionally, the local Whittle estimator of H based on the standardized nonparametric residuals is shown to be $\log(n)$-consistent and the finite dimensional distributions of the studentized versions of the regression function estimators are shown to be asymptotically normal. These results thus generalize some of the results of Robinson (1997) to heteroscedastic regression models with long memory moving average errors.

In Chapter 3, we discuss the asymptotic inference of the model (1.3.1) when $r(x)$ is some linear function and X_t is a long memory (LM) Gaussian design with long memory index $1/2 < h < 1$ and $\sigma(x)$ is still a nonparametric function. The first order asymptotic distribution of the least square estimator of the slope parameter is observed to be degenerate in some cases. Under some additional mild conditions, the second order asymptotic distribution of this estimator is shown to be normal whenever $h + H < 3/2$; non-normal otherwise. The asymptotic distribution of the kernel type estimators of the heteroscedasticity function is found to be normal whenever $H < (1 + h)/2$, and non-normal otherwise. In addition, an estimator of H based on pseudo residuals in a more general heteroscedastic regression model is shown to be $\log(n)$-consistent. A small simulation study included in this chapter shows that the above estimators of the regression parameters and the variance function are more stable for the values of h, H in the range $0.6 - 0.85$ compared to the values of h, H larger than 0.85. We also discuss the consistency of a cross validation type estimator of the heteroscedasticity function in a more general regression model under the assumed LM set up. All of these findings are then used to propose a lack-of-fit test of a parametric regression model, with an application to currency exchange rate data sets that exhibit LM.

1.4 Main issues and technical problems related in this research

In this section, we introduce some background material needed in this research, then discuss some issues and main difficulties tackled in this dissertation.

1.4.1 Non-Gaussian distribution

Let Z_t be a stationary Gaussian process with mean zero and variance one. Nonlinear functionals of Gaussian processes with power asymptotic correlation function were considered by Dobrushin and Major (1979), Taqqu (1975, 1979), Gorodetskii (1977), etc. Beautiful theorems are proved based on the Hermite expansion of

$$g(x) = \sum \alpha_k H_k(x)$$

where $g \in L^2 \equiv \{g : Eg(Z) = 0, Eg^2(Z) < \infty\}$. The functions $H_k(x)$ are the Hermite Polynomial defined by :

$$H_k(x) = (-1)^k \frac{d^k}{dx^k} e^{-x^2/2},$$

and $\alpha_k = Eg(Z)H_k(Z)$, cf. Sansone (1959). Note that $\{H_k(x), k = 0, 1, \cdots\}$ is a complete orthogonal basis of L^2, and satisfies $EH_l(Z)H_q(Z) = q!$ for $l = q$, $= 0$

otherwise. The Hermite rank of the function $g(x)$, is defined to be

$$m = min_{k=1,2,\cdots,}\{k, \alpha_k \neq 0\}.$$

In many cases, for a function $g(\cdot)$ of the Hermite rank m, the partial sum of $g(Z_i)$ is dominated by the Hermite polynomials $H_m(Z_i), i = 1, \cdots, n$. Because of the Gaussianity, moment bounds and a variety of inequalities can be used for the analysis of statistical problems related to these processes.

When one generalizes the above results to linear processes, the counterpart of the Hermite polynomials are the Appell polynomials, cf. Szegö (1959), Surgailis (1980, 1982), or Avram and Taqqu (1982). Define

$$e_k = E\Big(\frac{d^k}{dx^k}G(x)|_{x=X_0}\Big), \quad k = 1, 2, \cdots .$$

Then $m = \min\{k, e_k \neq 0\}$ is called the Appell rank of G w.r.t. X. Since Appell polynomials are not orthogonal, special care has to be taken when dealing with long memory linear processes such as FARIMA.

1.4.2 Strong dependence and asymptotic properties

Consider the model

$$Y_i = r(X_i) + u_i, \quad i = 1, \cdots, n. \tag{1.4.1}$$

When the errors u_i are i.i.d. with $Eu_1 = 0$ and finite variance, the following theorem (cf. Theorem 4.2.1, Hardle 1990) states the asymptotic distribution of the Nadaraya-Watson kernel estimator for one-dimensional predictor variables. Let K be a kernel function, b be the bandwidth, and f be the density function of X.

Theorem 1.4.1 *Suppose*

(A1). $\int |K(u)|^{2+\eta} du < \infty$ *for some* $\eta > 0$;

(A2). $b \sim n^{-1/5}$;

(A3). r and f are twice differentiable;

(A4) the distinct points $x_1, x_2, \cdots, x_k$ are continuity points of $\sigma^2(x)$ and $E(|Y|^{2+\eta}|X = x)$ and $f(x_j) > 0, j = 1, 2, \cdots, k$.

Then, the suitably normalized Nadaraya-Watson kernel smoother $\hat{r}(x_j)$ at the k different locations $x_1, \cdots, x_k$ converges in distribution to a multivariate normal random vector with some mean vector B and identity covariance matrix,

$$(nb)^{1/2}\left\{\frac{\hat{m}(x_j) - m(x_j)}{(\sigma^2(x_j)c_K/f(x_j))^{1/2}}\right\} \Longrightarrow N(B, I).$$

Hall and Hart (1990) have shown that, in the model (1.4.1) with $X_t = t/n, t = 1, \cdots, n$, and long memory Gaussian moving average error $\{u_t\}$ with covariance $\gamma_j \sim cj^{-\alpha}, 0 < \alpha := 2 - 2H < 1$, the optimal convergence rate of $\hat{m}(x)$ is $n^{-2\alpha/(4+\alpha)}$ when the mean function has 2 derivatives. Thus the optimal bandwidth is $b \sim n^{-\frac{\alpha}{4+\alpha}}$, which is slower than that of the i.i.d. setup, and depends on the long memory parameter H.

In our model (1.3.1), the error is a non-Gaussian and heteroscedastic long memory moving average process, we use the technique of decomposition and truncation of a linear process to achieve this optimal bandwidth, see Sections 2.2.4-5 below for details.

In addition to the estimation of the regression function, we also consider estimation of conditional variance function. The derivation of the consistency and asymptotic distribution of this estimator need weak convergence of some functional of long memory moving average processes. The corresponding results appear in Sections 2.2.3-4, 3.3.3. below.

1.4.3 Estimation of long memory parameter

As we have mentioned above, when one works with models with long memory errors, there is a crucial hurdle to overcome for inference - the $\log(n)$-consistency estimation of the long memory index H. We use the local Whittle estimator method to solve the problem.

Consider a stationary process X_t with mean μ and lag-j-autocovariance γ_j and spectrum $f(\lambda) = (2\pi)^{-1} \sum \gamma_j \cos(j\lambda)$ which satisfies the following condition.

"For some $H \in (\frac{1}{2}, 1)$,

$$f(\lambda) \sim L(\frac{1}{\lambda})\lambda^{1-2H} \quad \lambda \to 0+. \tag{1.4.2}$$

where $L(x)$ is a so called slowly varying function satisfying $L(tx)/L(x) \to 1$, as $x \to \infty$, for any $t \in R$". Condition (1.4.2) forces some restriction on α_j in (2.1.2) below (cf. Surgailis 1982).

The spectral density function of fractional Brownian motion is (cf. Sinai 1976)

$$f(\lambda) = \frac{\sigma^2 \sin(H\pi)}{\pi}\Gamma(2H+1)(1-\cos\lambda)\sum|\lambda + 2\pi j|^{-2H-1},$$

and that of fractional ARIMA is (cf. Brockwell and Davis 1987 and Hosking 1981)

$$f(\lambda) = \frac{\sigma^2}{2\pi}|1 - e^{i\lambda}|^{1-2H}\frac{|\psi(e^{i\lambda})|^2}{|\phi(e^{i\lambda})|^2}.$$

Both of them satisfy Condition (1.4.2).

Parametric models for $f(\lambda)$, $\lambda \in (-\pi, \pi]$ have been considered by many authors. The asymptotic distributional properties of parameter estimates have been derived by Fox and Taqqu (1986), Dahlhaus (1989) for the Gaussian process, and by Giraitis and Surgailis (1990) for the linear process, in the case $H \in (1/2, 1)$, and under some regularity conditions. These properties are highly desirable ones: $n^{1/2}$-consistency and asymptotic normality. However, these properties also depend on the correct specification of $f(\lambda)$, $\lambda \in (-\pi, \pi]$. In the event of any mis-specification, estimates will be inconsistent. To overcome this problem, semi-parametric estimates of H have been proposed. Robinson (1995b) showed that the two leading semi-parametric estimates of H have desirable asymptotic properties in a broad setting, these are the log-periodogram estimate which originated in Geweke and Porter-Hudak (1983), and the semiparametric Gaussian or local Whittle estimate which originated in Künsch (1987). Both estimates depend on a smooth parameter m, the number of low-frequency periodogram ordinates employed in the estimation.

The other methods for estimating the index H include the R/S, Variogram, wavelet method, cf. Beran (1994) and Abry et. al. (2002). Because the frequency-domain approach seems much more elegant than the time-domain one in this semi-parametric setting, and because simulation results of Taqqu and Teverovsky (1997) show that the local Whittle estimator is more robust, we use the local Whittle estimator to estimate H in the present setup. In order to obtain $\log(n)$-consistency of the estimator, Robinson (1997) provided some sufficient conditions under the model (1.4.1) with uniform design and long memory moving average errors.

The new challenge in our study is that the estimator has to be based on residuals, in which the conditional variance function estimator is involved. Derivation of these results require uniform consistent estimation of regression and variance functions with certain rates. Details are in Sections 2.2.6 and 3.3.4. In addition, in chapter 3, we show that $\log(n)$-consistency of the local Whittle estimator holds true even when the design of the parametric regression model is a long memory Gaussian process.

Chapter 2

Nonparametric regression with heteroscedastic long memory errors

2.1 Introduction

A stochastic process is said to have long memory, or to be long range dependent, if its auto-covariances decay at a hyperbolic rate in the lag. Long memory processes have been found to arise in a variety of physical and social sciences, see, e.g. Beran (1992, 1994), Baillie (1996), Dehling, Mikosch, and Sørensoen (2002), Doukhan, Oppenheim and Taqqu (2003), Robinson (2003), and the references therein. On the other hand nonparametric heteroscedastic regression models are also found to be very useful in practice.

The focus of this chapter is to analyze the asymptotic behavior of some inference procedures in these models with uniform non-random design and long memory errors. More precisely, consider the model

(2.1.1) $$Y_t = r(\frac{t}{n}) + \sigma(\frac{t}{n})u_t, \qquad t = 1, 2, \cdots, n,$$

where r is a real valued function and σ a positive function, both defined on $[0, 1]$, and where the errors u_t form the moving average process, i.e., for some $1/2 < H < 1$,

$$(2.1.2) \quad u_t := \sum_{j=0}^{\infty} \alpha_j \varepsilon_{t-j}, \quad \alpha_j \sim c\, j^{-(\frac{3}{2}-H)}, \text{ for } j \text{ large and for some } |c| < \infty.$$

The innovations ε_t are assumed to be i.i.d random variables (not necessary Gaussian) with mean 0 and unit variance. Then the spectral density of u_t's satisfies

$$(2.1.3) \qquad f(\lambda) \sim G\lambda^{1-2H} \quad \text{as } \lambda \to 0+$$

where G is a positive constant. In this paper $a_n \sim b_n$ means that $\lim_{n\to\infty} a_n/b_n = 1$.

Robinson (1997) considered the homoscedastic version of the above model where $\sigma(x) \equiv \sigma$, a positive constant, but with the errors more general in that the innovations are stationary martingale differences. Among many interesting results of his paper, he gave a central limit theorem for certain weighted partial sums of a covariance stationary process, and applied it to prove the asymptotic normality of the kernel type regression function estimator $\hat{r}$ and its studentized version. In the presence of long memory of the errors, an important element of this studentization is to have a $\log(n)$-consistent estimator of the parameter H. Such estimators are usually based on the residuals $\{Y_t - \hat{r}(t/n),\ 1 \le t \le n\}$. But, Robinson showed that in the case of long-range dependence the raw data $\{Y_t,\ 1 \le t \le n\}$ can also be used to provide such an estimator of H in homoscedastic models. One possible intuitive reason for this is that under the homscedastic set up where $\sigma(i/n) = \sigma$, a constant, the processes Y_t and u_t have the same covariance structure.

This chapter analyzes the asymptotic distribution of the kernel type estimators of both r, σ and the local Whittle estimator of H based on the estimated residuals under the above heteroscedastic setup. To proceed further, let K and W be density kernel functions and $b \equiv b_n$ and $c \equiv c_n$ be bandwidth sequences. The kernel estimators of

r and σ to be investigated here are

$$\text{(2.1.4)} \qquad \hat{r}(x) = \frac{1}{nb}\sum_t K(\frac{nx-t}{nb})Y_t,$$

$$\hat{\sigma}^2(x) = \frac{1}{nc}\sum_{t=1}^{n} W(\frac{nx-t}{nc})[Y_t - \hat{r}(\frac{t}{n})]^2.$$

Fan and Yao (1998) proved the finite dimensional asymptotic normality of the estimators $\hat{\sigma}$ for the stationary and absolutely regular errors. Under the above long memory setup, we establish their finite dimensional asymptotic normality for $1/2 < H < 3/4$. For $3/4 < H < 1$, these distributions are non-normal.

Since the parameters H and G appear in the standardization of these estimators in such a way that it is necessary to have $\log(n)$-consistent and consistent estimators of H and G, respectively, in order to use $\hat{r}$, $\hat{\sigma}$ for the large sample inference about r, σ. Using the method of Robinson (1997), it is proved that the local Whittle estimators $\hat{H}$, $\hat{G}$ of H, G based on the residuals $(Y_t - \hat{r}(t/n))/\hat{\sigma}(t/n)$ have these properties, which in turn make the studentization of $\hat{r}$ and $\hat{\sigma}^2$ feasible.

This chapter has seven sections. Section 2.2 verifies the asymptotic properties of $\hat{r}$, sections 2.3 and 2.4 discuss the asymptotic properties of $\hat{\sigma}$, while section 2.5 discusses the uniform convergence rate of $\hat{r}$. Section 2.6 discusses the estimation of H, G, and the asymptotic distributions of the studentized versions of $\hat{r}$ and $\hat{\sigma}$. Section 2.7 is an application, and Section 2.8 is the Appendix containing some preliminary results from Robinson (1997), some other miscellaneous results and some proofs. In the sequel, $=_d$ means equivalence in distribution.

2.2 Asymptotic normality of $\hat{r}$

Consider the Nadaraya-Watson estimate $\hat{r}(x)$ based on the model (2.1.1) and (2.1.2). We give additional needed conditions on the kernel, and the regression and variance functions.

Assumption 2.1 Let K be an even positive differentiable density with support $[-1, 1]$, and with a bounded derivative.

Assumption 2.2 Either the function r satisfies a Lipschitz condition of degree τ, $0 < \tau \leq 1$, or r is differentiable with derivative satisfying a Lipschitz condition of degree $\tau - 1, 1 < \tau \leq 2$.

Assumption 2.3 The function σ is continuous on $[0, 1]$ and bounded away from 0.

Assumption 2.4 The function σ satisfies Assumption 2.3 and is continuously twice differentiable on $[0, 1]$.

To discuss the bias in $\hat{r}$, introduce the continuity modulus of a function g on $[0, 1]$:

$$\omega(\delta; g) = \sup_{x, |b|<\delta} |g(x + b) - g(x)|, \qquad \delta > 0.$$

Lemma 9.1 in Härdle, Kerkyacharian, Picard, and Tsybakov (1997) shows that

$$(2.2.1) \qquad \omega(a\delta; g) \leq (a + 1)\omega(\delta; g), \qquad a > 0,\ \delta > 0.$$

We also need to recall, say from Beran (Theorem 2.1; 1998) or Zygmund (Chapter V.2, 1968), that for any long range dependent stationary process u_t with the spectral density f, (2.1.3) holds if and only if, with $\theta(H) := 2\Gamma(2 - 2H)\cos\{\pi(1 - H)\}$,

$$(2.2.2) \quad \gamma_j := \mathrm{Cov}(u_0, u_j) = \int_{-\pi}^{\pi} f(\lambda) e^{i\lambda j} d\lambda \sim G\theta(H) j^{-2(1-H)}, \quad j \to \infty.$$

Let $k_t(x) = K(\frac{nx-t}{nb}),\ t \geq 1,\ 0 < x < 1$.

$$(2.2.3) \qquad \hat{r}(x) = \frac{1}{nb}\sum_{t=1}^{n} k_t(x) r(\frac{t}{n}) + \frac{1}{nb}\sum_{t=1}^{n} k_t(x)\sigma(\frac{t}{n}) u_t, \quad \forall x.$$

Under (2.1.1) and Assumptions 2.1-3, a routine calculation shows that

$$(2.2.4) \qquad \begin{aligned} \sup_{0<x<1} |E(\hat{r}(x) - r(x))| &= O(b^\tau), && 0 < \tau \leq 1 \\ &= O(b^\tau + \frac{1}{nb}), && 1 < \tau \leq 2, \end{aligned}$$

Because for x close to 0 or 1, $\int_0^\infty |K(u)bu\{r'(x+\theta bu)\}|du = O(b)$, under Assumptions 2.1-3 with $\tau = 2$, one obtains that

(2.2.5)
$$\sup_{0\le x\le 1} |E(\hat{r}(x) - r(x))| = O(b + \frac{1}{nb}).$$

In order that the bias is small enough to permit centering at $r(x)$ in the central limit theorem, we impose the following additional assumption on the bandwidth.

Assumption 2.5: With τ as in Assumption 2.2,

$$(nb)^{-1} + (nb)^{1-H} b^\tau \to 0, \quad \text{as } n \to \infty.$$

The following lemma gives the asymptotic behavior of the covariance structure of $\hat{r}$. Let

$$\rho(H) \;:=\; \theta(H) \int\int K(v)K(w)|v-w|^{-2(1-H)} dwdv, \qquad D = G\theta(H).$$

Lemma 2.2.1 *Under (2.1.1), (2.1.2), the Assumptions 2.1-3 and 2.5, we have*

$$(nb)^{2-2H} Cov(\hat{r}(x), \hat{r}(y)) \to G\rho(H)\sigma(x)\sigma(y)1_{\{x=y\}}, \qquad \forall x, y \in (0,1).$$

Proof. Using (2.2.3), we obtain, for $x \neq y$,

$$\begin{aligned}
&(nb)^{2-2H} E[(\hat{r}(x) - r(x))(\hat{r}(y) - r(y))] \\
&= (nb)^{2-2H} \Big\{ \Big(\frac{1}{n^2 b^2} \sum_{t=1}^{n} k_t(x) r(\frac{t}{n}) - r(x) \Big) \Big(\frac{1}{nb} \sum_{t=1}^{n} k_t(y) r(\frac{t}{n}) - r(y) \Big) \\
&\quad + \frac{1}{n^2 b^2} \sum_{s\neq t} k_t(x) k_s(y) \sigma(\frac{t}{n}) \sigma(\frac{s}{n}) E u_t u_s \Big\} \\
&:= I + II
\end{aligned}$$

By (2.2.4) and the Assumption 2.5, $I = O((nb)^{2-2H} b^{2\tau}) \to 0$.

Next, using (2.2.2), we obtain,

$$II \sim D \frac{(nb)^{2-2H}}{(nb)^2} \int_1^n \int_1^n k_t(x) k_s(y) \sigma(\frac{t}{n}) \sigma(\frac{s}{n}) \, |t-s|^{-2(1-H)} dtds$$

$$
\begin{aligned}
&\sim D\,(nb)^{2-2H}\int\int K(z)K(w)\sigma(x-bz)\sigma(y-bw) \\
&\qquad\qquad\qquad \times|n(x-y)-nb(z-w)|^{-2(1-H)}dzdw \\
&\sim D\,\frac{b^{2-2H}}{|x-y|^{2-2H}}\int\int K(z)K(w)\sigma(x-bz)\sigma(y-bw) \\
&\qquad \times|1-b\frac{z-w}{x-y}|^{-2(1-H)}dzdw \\
&\leq D\,\frac{b^{2-2H}}{|x-y|^{2-2H}}\sigma(x)\sigma(y) \\
&\quad +D\,\frac{b^{2-2H}}{|x-y|^{2-2H}}\int\int K(z)K(w)\sigma(x-bz)\sigma(y-bw) \\
&\qquad \times|(2-2H)b\frac{z-w}{x-y}|dzdw \\
&= O(b^{2-2H})
\end{aligned}
$$

where the last but one inequality follows from the fact that for any $|a|<1/2$ and $1/2<H<1$, $|(1+a)^{-2(1-H)}-1|\leq|(2H-2)a|$.

Assuming $x=y$,

$$
\begin{aligned}
\mathrm{Var}\,(\hat{r}(x)) &= (nb)^{-2}\sum_{s,t}k_s(x)k_t(x)\sigma(\frac{t}{n})\sigma(\frac{s}{n})r(s-t) \\
&\sim D\,\frac{1}{(nb)^2}\int_1^n\int_1^n k_t(x)k_s(x)\sigma(\frac{t}{n})\sigma(\frac{s}{n})|t-s|^{-2(1-H)}dsdt \\
&= D\,\frac{1}{(nb)^2}\int_{\frac{x-\frac{1}{n}}{b}}^{\frac{x-1}{b}}\int_{\frac{x-\frac{1}{n}}{b}}^{\frac{x-1}{b}} K(z)K(w)|nb(z-w)|^{-2(1-H)}(nb)^2 \\
&\qquad \times\sigma(x-bz)\sigma(x-bw)dzdw \\
&\sim (nb)^{-2(1-H)}G\sigma^2(x)\rho(H).
\end{aligned}
$$

This proves the lemma. □

The proof of the asymptotic normality of $\hat{r}$ is facilitated by the following result from Robinson (1997), reproduced here for the sake of completeness and easy reference. Let $\mathcal{N}_k(\mu,\Sigma)$ stand for a k variate normal distribution with mean vector μ and covariance matrix Σ, k a positive integer. Write $\mathcal{N}$ for $\mathcal{N}_1$. Consider the

Assumption RI. Let β_j, $j\in\mathbb{Z}:=\{0,\pm1,\pm2,\cdots\}$ be a set of square summable real numbers and $\{\varepsilon_t, t \in \mathbb{Z}\}$ be a sequence of r.v.'s such that $E(\varepsilon_t|\mathcal{F}_{t-1}) = 0$, $E(\epsilon_t^2|\mathcal{F}_{t-1}) = 1$, a.s., $t \in \mathbb{Z}$, where $\mathcal{F}_t := \sigma\text{-field}\{\varepsilon_s, s \le t\}$, $t \in \mathbb{Z}$.

Let w_{tn}, $1 \le t \le n$, $n \ge 1$, be an array of constants, $v_{jn} := \sum_{t=1}^{n} w_{tn}\beta_{t-j}$, $u_t = \sum_{j=-\infty}^{\infty} \beta_j \varepsilon_{t-j}$, $t \in \mathbb{Z}$, and $S_n := \sum_{t=1}^{n} w_{tn} u_t$. Robinson (1997), we obtain

Lemma 2.2.2 *Suppose Assumption RI holds. In addition, assume that there exists a positive sequence $a = a_n$ such that*

$$\sum_{j=-\infty}^{\infty} v_{jn}^2 = 1, \qquad \forall n \ge 1, \tag{2.2.6}$$

$$\Big(\sum_{t=1}^{n} w_{tn}^2 \sum_{|j|>a} \beta_j^2\Big)^{1/2} + \max_{1\le t\le n} |w_{tn}| \sum_{|j|\le a} |\beta_j| \to 0. \tag{2.2.7}$$

Then, $S_n \Longrightarrow \mathcal{N}(0,1)$, as $n \to \infty$.

The next result gives the limiting finite dimensional distribution of $\hat{r}_n$, where $\beta = (G\rho(H))^{1/2}$.

Theorem 2.2.1 *Let (2.1.1) and the Assumptions 2.1-4 hold. Then for any integer $\ell \ge 1$ and for any distinct x_i, $i = 1, \cdots, \ell$, in $(0,1)$, the joint distribution of $(nb)^{1-H}(\beta\sigma(x_i))^{-1}\{\hat{r}(x_i) - r(x_i)\}, i = 1, \cdots, \ell$, converge weakly to $\mathcal{N}_\ell(0, \mathcal{I})$, as $n \to \infty$.*

Proof. Fix an integer $\ell \ge 1$ and real numbers $h_1, \cdots, h_\ell$, not all zero. Let $\sigma_i = \sigma(x_i)$. By the Cramer-Wold device, it suffices to prove

$$T_n := \sum_{i=1}^{\ell} h_i \frac{(nb)^{1-H}}{\beta\sigma_i}\{\hat{r}(x_i) - r(x_i)\} \Longrightarrow \mathcal{N}(0, \sum_{i=1}^{\ell} h_i^2).$$

Let

$$\begin{aligned}
\tilde{S}_n &:= \frac{1}{(nb)^H} \sum_{i=1}^{\ell} \frac{h_i}{\beta\sigma_i}\Big\{\sum_{t=1}^{n} k_t(x_i)\sigma(\frac{t}{n})u_t\Big\}, \\
V_n^2 &:= \mathrm{Var}(\tilde{S}_n), \qquad S_n := \frac{1}{V_n}\tilde{S}_n.
\end{aligned}$$

In view of (2.2.3), (2.2.4), and the Assumption 2.5, $T_n - \tilde{S}_n = o_P(1)$. It thus suffices to prove the claimed result for $\tilde{S}_n$. If we show that

$$(2.2.8) \qquad S_n \Longrightarrow \mathcal{N}(0,1), \qquad \text{and} \qquad V_n^2 \to \sum_{i=1}^{\ell} h_i^2,$$

then, by Slutsky's theorem, the claim will follow. But, by Lemma 2.2.1,

$$V_n^2 \sim (nb)^{2-2H} \sum_{i=1}^{\ell} \frac{h_i^2}{\beta^2\sigma_i^2} \operatorname{Var}\Big(\frac{1}{nb}\sum_{t=1}^{n} k_t(x_i)\sigma(\frac{t}{n})u_t\Big) \to \sum_{i=1}^{\ell} h_i^2.$$

Now consider the claim about S_n. Rewrite $S_n = \sum_{t=1}^{n} w_{tn} u_t$, where

$$\omega_{tn} := \frac{1}{(nb)^H V_n} \sum_{i=1}^{\ell} \frac{h_i}{\beta\sigma_i} \sigma(\frac{t}{n}) k_t(x_i).$$

In view of Lemma 2.2.2 applied with these ω_{tn}, $a = 1$, and with $\beta_j = \alpha_j$, $j \geq 0$, $\beta_j = 0$, $j \leq -1$, to prove the first part of (2.2.8), it suffices to verify that

$$(\sum_{t=1}^{n} \omega_{tn}^2 \sum_{j>1} \alpha_j^2)^{1/2} + \max_{1\leq t\leq n} |\omega_{tn}| \sum_{j\leq 1} |\alpha_j| \to 0.$$

But, because α_j's are square summable and the function σ is bounded from below, the left hand side of this expression is bounded above by a positive and finite constant times

$$\frac{1}{(nb)^H V_n}\Big\{\sum_{t=1}^{n} \Big(k_t(x_i)\sigma(\frac{t}{n})\Big)^2\Big\}^{1/2} + \frac{1}{(nb)^H V_n} \max_{1\leq t\leq n, 1\leq i\leq \ell} k_t(x_i) \sum_{j\leq 1} |\alpha_j|$$
$$= O((nb)^{1/2-H}) + O((nb)^{-H}) \to 0,$$

because $1/2 < H < 1$ and $nb \to \infty$, thereby completing the proof of (2.2.8). □

Remark 2.2.1 Hall and Hart (1990) show that under the long memory Gaussian errors setup the optimal bandwidth b for estimating $r(x)$ is of the order $n^{-(1-H)/(3-H)}$ and it is achieved by kernel smoothing method. If one uses this optimal b then the bias in Theorem 2.2.1 is not negligible. However, if b is chosen proportional to $n^{-(1-H)/(3-H)}$ times a sequence that tends slowly to zero, say $1/\log n$, then the bias vanishes asymptotically.

2.3 Estimates of $\sigma^2(x)$

This section discusses the asymptotic behavior of bias and variance of the estimator $\hat{\sigma}^2$. With c and W as in (2.1.4), let $I(b) := \{t : nb \le t \le n - nb\}$, and

$$w_t(x) := \frac{1}{nc}W(\frac{nx-t}{nc}), \quad \tilde{\sigma}^2(x) := \frac{1}{nc}\sum_{t=1}^{n} w_t(x)\sigma^2(\frac{t}{n})u_t^2, \; x \in [0,1].$$

The following decomposition of $\hat{\sigma}^2$ is often used in the sequel.

$$\begin{aligned}
(2.3.1) \qquad \hat{\sigma}^2(x) &= \frac{1}{nc}\Big\{\sum_{t\in I(b)} + \sum_{t\notin I(b)}\Big\} w_t(x)\Big[Y_t - \hat{r}(\frac{t}{n})\Big]^2 \\
&= \frac{1}{nc}\Big\{\sum_{t\in I(b)} + \sum_{t\notin I(b)}\Big\} w_t(x)\Big[r(\frac{t}{n}) - \hat{r}(\frac{t}{n})\Big]^2 \\
&\quad + \frac{2}{nc}\sum_{t=1}^{n} w_t(x)\,\sigma(\frac{t}{n})\,u_t\Big[r(\frac{t}{n}) - \hat{r}(\frac{t}{n})\Big] + \tilde{\sigma}^2(x).
\end{aligned}$$

To proceed further, we need to introduce additional assumption:

Assumption 2.6. The kernel W is another even positive differentiable density with support $[-1,1]$, and with a bounded derivative, and the bandwidth c satisfies $c \to 0$ and $nc \to \infty$.

We shall first analyze the bias and variance of $\tilde{\sigma}^2$. Recall that $\gamma_0 = Eu_0^2 = \sum_{j=0}^{\infty}\alpha_j^2$. From now on we shall assume that the u_t's are standardized, i.e., $\gamma_0 = 1$. Consider the bias of $\tilde{\sigma}^2$. Similar to (2.2.4), using Assumptions 2.4, 2.6, and the Taylor expansion of σ^2, we obtain, uniformly in $0 < x < 1$,

$$\begin{aligned}
E\tilde{\sigma}^2(x) &= \frac{1}{nc}\sum_{t=1}^{n} w_t(x)\sigma^2(\frac{t}{n}) \\
&= \int W(z)\Big(\sigma^2(x) - cz(\sigma^2(x))^{(1)} + \frac{c^2z^2}{2}(\sigma^2(x-czv))^{(2)}\Big)dz \\
&\qquad + O(\frac{1}{nc}) \\
(2.3.2) \qquad &= \sigma^2(x) + C_x c^2 + O(\frac{1}{nc}),
\end{aligned}$$

where $\sigma^2(x)^{(k)}$ stands for the kth derivative of $\sigma^2(x)$ and $C_x = \sigma^2(x)^{(2)} \times \int W(z)z^2dz/2$.

Next, in order to compute the variance of $\tilde{\sigma}^2(x)$, we need the following lemma about the covariance structure of the stationary process u_t^2.

Lemma 2.3.1 *With $\{u_t\}$ is as in* (2.1.2) *and under the assumption that $E\varepsilon_0^4 < \infty$, $\forall\ 1/2 < H < 1$,*

$$
\begin{aligned}
(2.3.3) \quad E(u_0^2 - 1)(u_t^2 - 1) &= 2\gamma_t^2 + \sum_s \alpha_s^2 \alpha_{t+s}^2 [E\varepsilon_0^4 - 3] \\
&= 2D^2 t^{2(2H-2)} + o(t^{2(2H-2)}), \quad \text{as } t \to \infty.
\end{aligned}
$$

Proof. Using the fact that $\{\varepsilon_t\}$ are standardized i.i.d. r.v.'s, we obtain

$$
\begin{aligned}
& E(u_0^2 - 1)(u_t^2 - 1) \\
&= E(\sum_{s,l,u,v} \alpha_s \alpha_l \alpha_u \alpha_v \varepsilon_{-s} \varepsilon_{-l} \varepsilon_{t-u} \varepsilon_{t-v}) - 1 \\
&= E(\sum_{s,u,s \neq u} \alpha_s^2 \varepsilon_s^2 \alpha_u^2 \varepsilon_{t-u}^2) + 2E(\sum_{s,l,s \neq l} \alpha_s \alpha_{s+t} \varepsilon_s^2 \alpha_l \alpha_{l+t} \varepsilon_l^2) \\
&\qquad + E(\sum_s \alpha_s^2 \alpha_{s+t}^2 \varepsilon_s^2 \varepsilon_{s+t}^2) - 1 \\
&= \sum_{s,u,s \neq u} \alpha_s^2 \alpha_u^2 + 2(\sum_{s,l,s \neq l} \alpha_s \alpha_{s+t} \alpha_l \alpha_{l+t}) + E\varepsilon_0^4 \sum_s \alpha_s^2 \alpha_{s+t}^2 - 1 \\
&= [(\sum_s \alpha_s^2)^2 - \sum_s \alpha_s^2 \alpha_{t+s}^2] + 2[(\sum_s \alpha_s \alpha_{s+t})^2 - \sum_s \alpha_s^2 \alpha_{t+s}^2] \\
&\qquad + E\varepsilon_0^4 \sum_s \alpha_s^2 \alpha_{s+t}^2 - 1 \\
&= 2\gamma_t^2 + \sum_s \alpha_s^2 \alpha_{t+s}^2 [E\varepsilon_0^4 - 3] \\
&\sim 2D^2 t^{2(2H-2)}, \qquad \text{as } t \to \infty.
\end{aligned}
$$

The last statement follows because $\alpha_j \sim c j^{H-3/2}$ as $j \to \infty$, and because

$$
\frac{1}{t^{4H-4}} \sum_s \alpha_s^2 \alpha_{s+t}^2 = O(\frac{1}{t} + \frac{1}{t^{2-2H}}) \to 0, \qquad \text{as } t \to \infty. \qquad \square
$$

Remark 2.3.1 The term $[E\varepsilon_0^4 - 3]$ is exactly zero in the case $\{\varepsilon_i\}$ are standard normal random variables.

Now, using (2.3.2), we obtain

$$\mathrm{Var}(\tilde{\sigma}^2(x)) = \frac{1}{(nc)^2}\sum_{s,t} w_t(x)w_s(x)\sigma^2(\frac{s}{n})\sigma^2(\frac{t}{n})E\{(u_t^2-1)(u_s^2-1)\}.$$

Let $\mathcal{D}_n(x)$ denote the leading term on the right hand side of the above equation. To analyze it further, it is necessary to consider the two ranges of H, v.i.z., $1/2 < H < 3/4$ and $3/4 < H < 1$ separately. We define

$$\|W\|^2 := \int W^2(v)dv, \qquad \tilde{C}_1(x,H) := E(u_0^4-1)\sigma^4(x)\|W\|^2,$$

$$C_2(x,H) := \sigma^4(x)\int\int W(\omega)W(\nu)|\omega-\nu|^{4H-4}d\omega d\nu, \quad \tilde{C}_2(x,H) := 2D^2C_1(x,H).$$

For $1/2 < H < 3/4$, the process $u_t^2 - 1$ is short-range dependent. Moreover, for an $0 < x < 1$,

$$\begin{aligned}
nc\mathcal{D}_n(x) &= \frac{1}{nc}\sum_{s,t} w_t(x)w_s(x)\sigma^2(\frac{j}{n})\sigma^2(\frac{s}{n})E(u_s^2-1)(u_t^2-1)\\
&= \frac{1}{(nc)}\sum_{s=t,t=1}^{n} w_t^2(x)\sigma^4(\frac{t}{n})E(u_t^4-1)\\
&\quad +\frac{1}{(nc)}\sum_{s,t,s\neq t} w_t(x)w_s(x)\sigma^2(\frac{j}{n})\sigma^2(\frac{s}{n})E(u_s^2-1)(u_t^2-1)
\end{aligned}$$

$$\begin{aligned}
&= E(u_0^4-1)\int W^2(\omega)\sigma^4(x-c\omega)d\omega\\
&+ \frac{1}{nc}\int W(\omega)W(\nu)\sigma^2(x-c\omega)\sigma^2(x-c\nu)E(u_0^2u_{nc(\omega-\nu)}^2)d\omega d\nu + O(\frac{1}{nc})\\
&\sim \tilde{C}_1(x,H) + (nc)^{2(2H-2)-1}\tilde{C}_2(x,H) + O(\frac{1}{nc}) + O(\frac{1}{nc})\\
&= \tilde{C}_1(x,H) + o(1).
\end{aligned}$$

For $3/4 < H < 1$, the case of very long-range dependence, the sequence $u_t^2 - 1$ is still long-range dependent. Proceeding as above and using the continuity of σ and Lemma 2.3.1, we obtain,

$$\begin{aligned}
\mathcal{D}_n(x) &\sim C(nc)^{-1} + 2D^2\frac{1}{(nc)^2}\int_1^n\int_1^n w_s(x)w_t(x)\sigma^2(\frac{t}{n})\sigma^2(\frac{s}{n})|t-s|^{4H-4}dsdt\\
&\sim (nc)^{4H-4}\tilde{C}_2(x,H), \qquad \forall\, 0 < x < 1.
\end{aligned}$$

The above discussion is summarized in the following

Lemma 2.3.2 *Suppose (2.1.1) and (2.1.2) hold. In addition, suppose* $E\varepsilon_0^4 < \infty$ *and the Assumption 2.4 holds. Then, for each* $x \in (0,1)$, $\tilde{\sigma}^2(x) \to \sigma^2(x)$ *in probability,* $E\{\tilde{\sigma}^2(x) - \sigma^2(x)\} = O(\frac{1}{n} + c^2)$, *and*

$$
\begin{aligned}
Var(\tilde{\sigma}^2(x)) &= \tilde{C}_1(x,H)(nc)^{-1}, & 1/2 < H < 3/4; \\
&= \tilde{C}_2(x,H)\,(nc)^{-4(1-H)}, & 3/4 < H < 1.
\end{aligned}
$$

Now we are ready to analyze the bias and variance behavior of $\hat{\sigma}^2(x)$. Recall the decomposition (2.3.1). By the proof of Lemma 2.2.1, $E(\hat{r}(x) - r(x))^2 = O((nb)^{-2(1-H)} + b^2)$ uniformly $x \in [0,1]$. Also, the continuity of W and Assumption 2.6 imply

$$
\sup_{0 \le x \le 1} \left| \frac{1}{nc} \sum_{t \notin I(b)} w_t(x) \right| = o(1).
$$

Hence,

$$
\begin{aligned}
E\frac{1}{nc} \sum_{t \notin I(b)} w_t(x)[r(\frac{t}{n}) - \hat{r}(\frac{t}{n})]^2 &= O((nb)^{-2(1-H)})\frac{1}{nc} \sum_{t \notin I(b)} w_t(x) \\
&= o((nb)^{-2(1-H)} + b^2),
\end{aligned}
$$

so that the expected value of the first term in the right hand side of (2.3.1) is asymptotically equivalent to

$$
\begin{aligned}
&\frac{1}{nc} \sum_{t \in I(b)} w_t(x)[\beta^2 \sigma^2(\frac{t}{n})(nb)^{-2(1-H)}] \\
\sim\ & \beta^2 (nb)^{-2(1-H)} \frac{1}{nc} \sum_{t \in I(b)} w_t(x)\sigma^2(\frac{t}{nb}) \\
=\ & \beta^2 (nb)^{-2(1-H)} \{\int W(v)\sigma^2(x)dv + O(\frac{1}{n})\} \\
\sim\ & \beta^2 (nb)^{-2(1-H)} \sigma^2(x), \quad \forall x \in (0,1).
\end{aligned}
$$

By Lemma 2.3.2, the expected value of the third term in the right hand side of (2.3.1) is $\sigma^2(x) + C_x c^2 + O(\frac{1}{n})$. Now consider the second term of (2.3.1). By (2.2.3),

$$
\begin{aligned}
(2.3.4) \qquad & E\Big[\frac{1}{nc}\sum_{t=1}^{n} w_t(x)\sigma(\frac{t}{n})u_t(r(\frac{t}{n}) - \hat{r}(\frac{t}{n}))\Big] \\
&= -E\Big[\frac{1}{nc}\sum_{t=1}^{n} w_t(x)\sigma(\frac{t}{n})u_t\{\frac{1}{nb}\sum_{s=1}^{n} k_s(\frac{t}{n})\sigma(\frac{s}{n})u_s\}\Big] \\
&= \frac{-1}{nc}\sum_{t=1}^{n} w_t(x)\sigma(\frac{t}{n})\frac{1}{nb}\sum_{s=1}^{n} k_s(\frac{t}{n})\sigma(\frac{s}{n})Eu_s u_t.
\end{aligned}
$$

Since for any fix $N < n$,

$$
\Big|\frac{1}{nc}\sum_{t=1}^{n} w_t(x)\sigma(\frac{t}{n})\frac{1}{nb}\sum_{s:|s-t|<N} k_s(\frac{t}{n})\sigma(\frac{s}{n})Eu_s u_t\Big| = O(\frac{1}{nb}),
$$

hence (2.3.4) is asymptotically

$$
\begin{aligned}
&\sim \frac{-D}{nc}\Big\{\sum_{t\in I(b)} + \sum_{t\notin I(b)}\Big\}w_t(x)\sigma(\frac{t}{n})\frac{1}{nb}\sum_{s=1}^{n} k_s(\frac{t}{n})\sigma(\frac{s}{n})|s-t|^{-2(1-H)} \\
&\sim \frac{-D}{(nc)}\sum_{t\in I(b)} w_t(x)\sigma(\frac{t}{n})\int K(v)\sigma(\frac{t}{n})(nb)^{-2(1-H)}|v|^{-2(1-H)}dv \\
&\sim -D(nb)^{-2(1-H)}\sigma^2(x)\int K(v)|v|^{-2(1-H)}dv,
\end{aligned}
$$

since, similar to the first term of (2.3.1), $\sum_{t\notin I(b)}$-part is negligible compared to $\sum_{t\in I(b)}$-part, as $n \to \infty$.

Combining the above approximations, we obtain that for each $x \in (0,1)$, the bias of $\hat{\sigma}^2(x)$ satisfies

$$
(2.3.5) \qquad E\Big(\hat{\sigma}^2(x) - \sigma^2(x)\Big) \sim C_{bias}(x,H)(nb)^{-2(1-H)} + C_x c^2,
$$

where C_x is defined in (2.3.2),

$$
\begin{aligned}
C_{bias}(x,H) = D\sigma^2(x)\Big\{&\int\int K(\omega)K(\nu)|\omega-\nu|^{-2(1-H)}d\omega d\nu \\
&-2\int K(v)|v|^{-2(1-H)}dv\Big\}.
\end{aligned}
$$

The following additional assumption is needed to obtain the asymptotic distribution of $\hat{\sigma}^2(x)$.

Assumption 2.7 Assume the bandwidth $c = o(b)$ and $(nc)^{2(1-H)}c^2 \to 0$.
Note that under this assumption, the above asymptotic bias of $\hat{\sigma}^2(x)$ depends only on the bandwidth b used for estimating r, and not on the bandwidth c.

Lemma 2.3.3 *Under (2.1.1), (2.1.2), and the Assumptions 2.1-3 with $\tau = 2$, suppose that $c = o(b)$. Then, for every $x \in (0,1)$,*

$$(2.3.6) \qquad (nc)^{-1}\, Var\Big(\hat{\sigma}^2(x) - \tilde{\sigma}^2(x)\Big) = o(1), \qquad \frac{1}{2} < H < \frac{3}{4};$$

$$(2.3.7) \qquad (nc)^{4(1-H)}\, Var\Big(\hat{\sigma}^2(x) - \tilde{\sigma}^2(x)\Big) = o(1), \qquad \frac{3}{4} < H < 1.$$

The proof of this lemma is computationally involved and deferred to the Appendix.

Remark 2.3.2 In view of the above results, the mean square error of $\hat{\sigma}^2(x)$ satisfies

$$(2.3.8) \quad E\Big(\hat{\sigma}^2(x) - \sigma^2(x)\Big)^2 = O\Big((nc)^{-1} + c^4\Big), \qquad 1/2 < H < 3/4,$$

$$(2.3.9) \qquad = O\Big((nc)^{4(H-1)} + c^4\Big), \qquad 3/4 < H < 1.$$

From this one sees that if $c = O(n^{-(1-H)/(2-H)})$, then this mean square error tends to zero and at the same time this c is of the smaller order than the optimal bandwidth b selected in Remark 2.2.1. This is one reason for introducing the Assumption 2.7. Another reason is that if $c = b$, then the result (2.3.7) of Lemma 2.3.3 for $\frac{3}{4} < H < 1$ is not true because in this case $\mathrm{Var}(\hat{\sigma}^2(x) - \tilde{\sigma}^2(x)) = O((nb)^{4(H-1)})$.

The requirement that $(nc)^{2-2H}c^2 \to 0$ will lead to centered asymptotic distribution for $\hat{\sigma}^2(x)$ discussed in the next section.

2.4 Asymptotic distribution of $\hat{\sigma}^2(x)$

This sections discusses the finite dimensional asymptotic distribution of the variance estimator $\hat{\sigma}^2(x)$. This is facilitated by using some results about the Appell polynomials. Let A_m denote the m-th Appell polynomial associated with the distribution of u_0, $1/2 < \beta := (3/2) - H < (m+1)/2m$, and assume that $E\varepsilon_0^{2m} < \infty$. Theorem 2 of Avram and Taqqu (1987) shows that, as $n \to \infty$,

$$(2.4.1) \quad \frac{1}{n^{1+m(H-1)}} \sum_{k=1}^{[n\cdot]} A_m(u_k) \Longrightarrow Z_m(\cdot), \qquad \text{(in } D[0,T]\text{), uniform metric),}$$

for any constant $T \in (0, \infty)$, where, for $t \geq 0$,

$$\begin{aligned} Z_m(t) &= m! \int\int_{-\infty<\omega_1<\omega_2<\cdots<\omega_m} \\ &\times \int \Big\{ \int_0^t \prod_{j=1}^m (\nu - \omega_j)_+^{-\beta} \Big\} dB(\omega_1) \cdots dB(\omega_m) \end{aligned}$$

is the Hermite process, where B is the standard Brownian motion on $[0, \infty)$. Some properties of this process are discussed in Hosking (1996) and Taqqu (1975, 1978). In particular this is well defined for $t \in [0, T]$, for every $T < \infty$.

Recall, say from Surgailis (1982), that 2-nd Appell polynomial associated with u_0 is $A_2(x) = x^2 - 1$. Since $Eu_0 = 0, Eu_0^2 = 1$, upon applying (2.4.1) with $m = 2$, and noting that in this case $1/2 < \beta < 3/4$, or equivalently, $3/4 < H < 1$, we obtain

$$(2.4.2) \qquad \frac{1}{n^{2H-1}} \sum_{k=1}^{[nt]} (u_k^2 - 1) \Longrightarrow Z_2(t), \qquad 3/4 < H < 1.$$

The next theorem discusses the asymptotic distribution of $\tilde{\sigma}^2(x)$. For that purpose we need

$$Y_2 = \int_0^2 W'(1-s) Z_2(s) ds.$$

For completeness, we also recall the formula for the summation by parts which is often used in the sequel. For any two sequences $\{a_j, j = 1, 2, \cdots\}$ and $\{b_j, j = 1, 2, \cdots\}$, and for any integers $1 \leq m < n$,

$$(2.4.3) \quad \sum_{j=m}^n a_j b_j = b_n \sum_{j=1}^n a_j - b_m \sum_{j=1}^{m-1} a_j + \sum_{j=m}^{n-1} \sum_{k=1}^j a_k (b_j - b_{j+1}).$$

Lemma 2.4.1 *Suppose (2.1.1), (2.1.2) and the Assumptions 2.1-4 and 2.6 hold. Then, for every* $x \in (0,1)$,

$$(2.4.4) \qquad (nc)^{2-2H}\{\tilde{\sigma}^2(x) - \sigma^2(x)\} \Longrightarrow \sigma^2(x)Y_2, \qquad 3/4 < H < 1.$$

Proof. Define $S_n(v) = \sum_{k=0}^{[nv]}(u_k^2 - 1)$ for $0 \le v \le 1$, $S_n = S_n(1)$. Fix an $x \in (0,1)$. Then, $w_n(x) \equiv W((x-1)/c) = 0$, for all $c < 1-x$. Hence, (2.4.3) yields that the left hand side of (2.4.4) equals

$$\begin{aligned}
&(nc)^{1-2H}\sum_{t=1}^{n} W\Big(\frac{nx-t}{nc}\Big)\sigma^2\Big(\frac{t}{n}\Big)(u_t^2 - 1) \\
&= (nc)^{1-2H}\Big\{S_n w_n(x)\sigma^2\Big(\frac{n}{n}\Big) - S_0 w_1(x)\sigma^2\Big(\frac{1}{n}\Big) \\
&\qquad + \sum_{t=1}^{n-1} S_n\Big(\frac{t}{n}\Big)\Big[w_t(x)\sigma^2\Big(\frac{t}{n}\Big) - w_{t+1}(x)\sigma^2\Big(\frac{t+1}{n}\Big)\Big]\Big\} \\
&= (nc)^{1-2H}\Big\{\sum_{t=1}^{n-1} S_n\Big(\frac{t}{n}\Big)\Big[w_t(x)\sigma^2\Big(\frac{t}{n}\Big) - w_{t+1}(x)\sigma^2\Big(\frac{t+1}{n}\Big)\Big]\Big\} + o_p(1) \\
&= (nc)^{1-2H}\int_{-1}^{1} S_n(x - c\nu)\Big[W(\nu)\sigma^2(x - c\nu) \\
&\qquad\qquad - W\Big(\nu - \frac{1}{nc}\Big)\sigma^2\Big(x - c\nu + \frac{1}{n}\Big)\Big]d\nu + o_p(1) \\
&= (nc)^{1-2H}\Big\{\int_{-1}^{1} S_n(x - c\nu)W'(\nu)\sigma^2(x)d\nu\Big\} + o_p(1).
\end{aligned}$$

This fact, the stationarity of $\{u_t\}$, combined with the fact that $\int_{-1}^{1} W'(\nu)d\nu = 0$, yields that the leading term in the right hand side above is equal to

$$\begin{aligned}
&(nc)^{1-2H}\sigma^2(x)\int_{-1}^{1}\Big(S_n(x - c\nu) - S_n(x-c)\Big)W'(\nu)d\nu \\
&= (nc)^{1-2H}\sigma^2(x)\int_{-1}^{1} W'(\nu)\sum_{k=[n(x-c)]}^{[n(x-c\nu)]}(u_k^2 - 1)d\nu \\
&=_d (nc)^{1-2H}\sigma^2(x)\int_{-1}^{1} W'(\nu)S_{r_x(\nu)}(1)d\nu \\
&= \sigma^2(x)\int_0^2 W'(1-t)\frac{S_{[nc]}(t)}{[nc]^{2H-1}}dt \\
&\qquad\qquad \Longrightarrow \sigma^2(x)Y_2,
\end{aligned}$$

where $r_x(\nu) = [n(x - c\nu)] - [n(x - c)] + 1$. The last claim follows from (2.4.2) and the continuous mapping theorem (Billingsley, 1968, Theorem 5.1), since $W'(\cdot)$ is bounded and the functional $T(f) = \int_0^2 W'(1-t)f(t)dt$ for $f \in D[0,2]$ is continuous. □

In view of Lemmas 2.3.3 and 2.4.1, the following theorem is immediate.

Theorem 2.4.1 *Under (2.1.1), (2.1.2), and the Assumptions 2.1-7, for each $x \in (0,1)$,*

$$(nc)^{2(1-H)}\{\hat{\sigma}^2(x) - \sigma^2(x)\} \Longrightarrow \sigma^2(x)Y_2, \qquad 3/4 < H < 1.$$

Now consider the case $\frac{1}{2} < H < \frac{3}{4}$. By Lemma 2.2.1, in this case the covariances of the process $u_t^2 - 1$ are absolutely summable. Hence this process is short range dependent and suitably standardized $\hat{\sigma}^2(x)$ will be asymptotically normal. To prove this, we use a result of Wu (2002), which in turn is based on some results of Maxwell and Woodroofe (2000). For the sake of completeness, we state Corollary 1 in Wu (2002) as Lemma 2.4.2 below. Let $||\cdot||$ denote L_2-norm, and let

$$u_{t,+} = \sum_{i=0}^{t-1} \alpha_i \varepsilon_{t-i}, \quad u_{t,-} := \sum_{i \geq t} \alpha_i \varepsilon_{t-i}, \quad u_t = u_{t,+} + u_{t,-}, \quad t \geq 1.$$

Let φ be a function from $\mathbb{R}$ to $\mathbb{R}$, $\varphi_n(x) := E\varphi(x + u_{n,+})$, $S_n(\varphi) = \sum_{t=1}^n \varphi(u_t)$.

Lemma 2.4.2 *If $E[|\varphi_n(u_1)|]^p < \infty$ and $||\varphi_n(u_{n,-})|| = O(n^{\kappa-1})$ for some $p > 2$ and $\kappa < \frac{1}{2}$, then*

$$\{\sqrt{n}S_{[nt]}(\varphi), 0 \leq t \leq 1\} \Longrightarrow \{\sigma_0 B(t), 0 \leq t \leq 1\},$$

in $D[0,1]$, with respect to the uniform metric.

Our needed result is given in

Lemma 2.4.3 *Suppose (2.1.2) holds with $E|\varepsilon_0|^p < \infty$, for some $p > 4$. Then, for $\frac{1}{2} < H < \frac{3}{4}$,*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{[nt]}(u_i^2 - 1) \Longrightarrow \sigma_0 B(t), \qquad (\textit{in } D[0,2], \textit{ uniform metric}),$$

where B is the Browninan motion on $[0,2]$ *and*

$$\sigma_0^2 := E(u_0^2-1)^2 + 2\sum_{i=1}^{\infty} E[(u_0^2-1)(u_i^2-1)]$$

$$(2.4.5) \qquad = 2\sum_{t=0}^{\infty}\gamma_t^2 + (E\varepsilon_0^4 - 3)\sum_{t=0}^{\infty}\sum_{s=1}^{\infty}\alpha_s^2\alpha_{t+s}^2, \qquad \gamma_t = (\sum_i \alpha_i\alpha_{t+i})^2.$$

Proof. The proof follows from lemma 2.4.2 applied to $\varphi(x) = x^2 - 1$. Note that for this φ,

$$\varphi_t(x) := E\varphi(x+u_{t,+}) = E[(x+u_{t,+})^2 - 1] = x^2 - \sum_{i\ge t}\alpha_i^2.$$

According to Lemma 2.4.2, it suffices to verify that for some $\kappa < \frac{1}{2}$,

$$(2.4.6) \qquad ||\varphi_t(u_{t,-})|| = O(t^{\kappa-1}).$$

But this is satisfied with $\kappa = 2H - 1$, because

$$\begin{aligned} ||\varphi_t(u_{t,-})||^2 &= E\Big[(\sum_{i\ge t}\alpha_i\varepsilon_i)^2 - \sum_{i\ge t}\alpha_i^2\Big]^2 \\ &= E\Big[\Big(\sum_{i\ge t}\alpha_i\varepsilon_i\Big)^4 - 2\sum_{i\ge t}\alpha_i^2 E\Big(\sum_{i\ge t}\alpha_i\varepsilon_i\Big)^2 + \Big(\sum_{i\ge t}\alpha_i^2\Big)^2\Big] \\ &= E\varepsilon_0^4\sum_{i\ge t}\alpha_i^4 + 3\sum_{i\ne j, i\ge t, j\ge t}\alpha_i^2\alpha_j^2 - \sum_{i\ge t}\alpha_i^2 \\ &= (E\varepsilon_0^4 - 3)\sum_{i\ge t}\alpha_i^4 + 2\sum_{i\ge t}\alpha_i^2 = O(t^{4H-4}). \end{aligned}$$

The claim about σ_0^2 follows from (3.5.2). □

Theorem 2.4.2 *Suppose (2.1.1), (2.1.2) with* $E|\varepsilon_0|^p < \infty$*, for some* $p > 4$*, and the Assumptions 2.1-7 hold. Then, for every* $0 < x < 1$,

$$\sqrt{nc}\{\hat\sigma^2(x) - \sigma^2(x))\} \Longrightarrow \sigma^2(x)\,\sigma_0\,Z, \qquad \frac{1}{2} < H < \frac{3}{4},$$

where $Z := -\int_0^2 B(t)dW(1-t)$ *has the* $\mathcal{N}(0, 2\|W\|^2)$ *distribution.*

Proof. Arguing as in the proof of Theorem 2.4.1,

$$(nc)^{-1/2}\sum_{t=1}^{n} W(\frac{nx-t}{nc})\sigma^2(\frac{t}{n})(u_t^2-1)$$

$$= (nc)^{-1/2}\Big\{S_n w_n(x)\sigma^2(\frac{n}{n}) - S_0 w_1(x)\sigma^2(\frac{1}{n})$$

$$+\sum_{t=1}^{n-1} S_n(\frac{t}{n})\Big[w_t(x)\sigma^2(\frac{t}{n}) - w_{t+1}(x)\sigma^2(\frac{t+1}{n})\Big]\Big\}$$

$$= (nc)^{-1/2}\Big\{\sum_{t=1}^{n-1} S_n(\frac{t}{n})\Big[w_t(x)\sigma^2(\frac{t}{n}) - w_{t+1}(x)\sigma^2(\frac{t+1}{n})\Big]\Big\}$$

$$+o_P(1)$$

$$= (nc)^{-1/2}\Big\{\int_{-1}^{1} S_n(x-c\nu)\Big[W(\nu)\sigma^2(x-c\nu)$$

$$-W(\nu-\frac{1}{nc})\sigma^2(x-c\nu+\frac{1}{n})\Big]d\nu + O_P(\frac{1}{n})\Big\} + o_P(1)$$

$$= (nc)^{-1/2}\Big\{\int_{-1}^{1} S_n(x-c\nu)W'(\nu)\sigma^2(x)d\nu)\Big\} + o_P(1),$$

Since $\int_{-1}^{1} W'(\nu)d\nu = 0$, the leading term in the right hand side above is equal to

$$(nc)^{-1/2}\sigma^2(x)\int_{-1}^{1} W'(\nu)\{\sum_{k=[n(x-c)]}^{[n(x-c\nu)]}(u_k^2-1)\}d\nu$$

$$= (nc)^{-1/2}\sigma^2(x)\int_{-1}^{1} W'(\nu)S_{r_x(\nu)}(1)d\nu =_d \sigma^2(x)\int_{-1}^{1} W'(\nu)\frac{S_{[nc]}(\nu)}{[nc]^{1/2}}d\nu.$$

The theorem follows from this fact, Lemma 2.4.3 and the continuous mapping theorem. □

2.5 Uniform convergence rate of $\hat{r}$

In this section, we consider the uniform convergence rate of $\hat{r}$ and uniform consistency of $\hat{\sigma}$. As is well known, under i.i.d. error setup, the point wise consistency rate of the kernel type regression estimator $\hat{r}$ is $1/\sqrt{nb}$, and the uniform convergence rate is nearly the same: $\log n/\sqrt{nb}$. As seen in Theorem 2.2.1, when errors are long range dependent, the point wise convergence rate is $(nb)^{-(1-H)}$. We shall next show that if the innovations ε_i satisfy the Cramer's Condition, then the uniform consistency rate of $\hat{r}$ is $(nb)^{-(1-H)}\log n$. Towards this goal we first recall the Bernstein inequality from Doukan (1994).

Lemma 2.5.1 *Let $X_i, i = 1, \cdots, n$, be mean zero finite variance independent random variables. Assume, additionally, that they satisfy the Cramer's Condition: For some $C < \infty$,*

$$E|X_i|^k \leq C^{k-2} k! EX_i^2 \quad k = 2, 3, \cdots, \quad i = 1, 2, \cdots, n. \tag{2.5.1}$$

Let $S_n = \sum_{i=1}^n X_i$, $s_n^2 = \sum_{i=1}^n Var(X_i)$. Then, for any $\epsilon > 0$,

$$P(|S_n| > \epsilon) \leq 2\exp\{\frac{-\epsilon^2}{4\, s_n^2 + 2C\epsilon}\}.$$

Remark 2.5.1 A large class of random variables including bounded, Gaussian and Gamma r.v.'s, satisfy this condition.

Rewrite the moving average process as $u_t = \sum_{j=-\infty}^{\infty} \alpha_{t-j}\varepsilon_j$ by defining $\alpha_j = 0$ if $j < 0$. Then for some integer $L > 0$,

$$\frac{1}{nb}\sum_{t=1}^{n} k_t(x)\sigma(\frac{t}{n})u_t = S_L + S_L^- + S_L^+, \qquad S_L = \sum_{j=-L+1}^{L} v_{nj}\varepsilon_j,$$

$$S_L^- = \sum_{j=-\infty}^{-L} v_{nj}\varepsilon_j, \quad S_L^+ = \sum_{j=L+1}^{\infty} v_{nj}\varepsilon_j, \quad v_{nj} = \frac{1}{nb}\sum_{t=1}^{n} k_t(x)\sigma(\frac{t}{n})\alpha_{t-j}$$

Let $\sigma^* := \sup_{x\in[0,1]}\sigma(x)$. Observe that for large L, $\mathrm{Var}S_L^+ = 0$,

$$\begin{aligned} \mathrm{Var}S_L^- &\leq \sum_{j=-\infty}^{-L} \alpha_j^2 \Big(\sum_{t=1}^{n} \frac{1}{nb} k_t(x)\sigma(\frac{t}{n})\Big)^2 \\ &= \Big(\frac{(\sigma^*)^2}{nb}\int K(\frac{nx-t}{nb})dt + O(\frac{1}{n})\Big)^2 L^{2H-2} = O(L^{2H-2}). \end{aligned}$$

Consider S_L. Let $X_j := v_{nj}\varepsilon_j$. If ε_j's satisfy the Cramer's Condition, then

$$E|X_j|^k \le |v_{nj}|^k C^{k-2}(k!)E|\varepsilon_j|^2 \le |v_{nj}|^{k-2}C^{k-2}(k!)EX_j^2.$$

Moreover, using the fact σ and K are bounded above and the Cauchy-Schwarz inequality, we obtain that for any $\beta > 0$,

$$\begin{aligned}
|v_{nj}| &= \frac{1}{nb}\Big|\Big(\sum_{|t-j|>\beta} + \sum_{|t-j|\le\beta}\Big)k_t(x)\sigma(\frac{t}{n})\alpha_{t-j}\Big| \\
&\le C\frac{1}{nb}\Big(\sum_{t=1}^{n} k_t^2(x)\Big)^{1/2}\Big(\sum_{|j|>\beta}\alpha_j^2\Big)^{1/2} + C\max_{1\le t\le n}|\frac{1}{nb}k_t(x)|\sum_{|j|\le\beta}|\alpha_j| \\
&\le C\Big[\frac{1}{\sqrt{nb}}\beta^{H-1} + \frac{1}{nb}\sum_{|j|\le\beta}|\alpha_j|\Big].
\end{aligned}$$

Choosing $\beta = nb$ in this bound yields that for all sufficiently large n,

$$\max_{1\le j\le n}|v_{nj}| \le C\,(nb)^{H-3/2}. \tag{2.5.2}$$

Thus these X_j's satisfy the Cramer's condition for all sufficiently large n with C there replaced by $C_2 = C\,(nb)^{H-3/2}$.

To apply the Bernstein's inequality, we need to analyze the variance of these r.v.'s. Towards that we have

$$\sigma_j^2 := \mathrm{Var}X_j \le \frac{(\sigma^*)^2}{(nb)^2}\sum_{s,t}k_t(x)k_s(x)|\alpha_{t-j}\alpha_{s-j}|,$$

$$\begin{aligned}
s_n^2 = \sum_{j=1}^{n}\sigma_j^2 &= \frac{(\sigma^*)^2}{(nb)^2}\sum_{s,t}k_t(x)k_s(x)\sum_{j=-L}^{L}|\alpha_{t-j}\alpha_{s-j}| \\
&\le \frac{(\sigma^*)^2}{(nb)^2}\sum_{s,t}k_t(x)k_s(x)\sum_{j=-\infty}^{\infty}|\alpha_{t-j}\alpha_{s-j}| \\
&= \frac{(\sigma^*)^2}{(nb)^2}\sum_{s,t}k_t(x)k_s(x)\gamma_{|t-s|} = O((nb)^{2H-2}).
\end{aligned}$$

Applying Bernstein's Inequality with $\epsilon_n = (nb)^{H-1}\log n$, we get for n sufficiently large,

$$\begin{aligned} P(|S_L| > \epsilon_n) &\leq 2\exp\{\frac{-\epsilon_n^2}{4C(nb)^{2H-2} + 2C(nb)^{H-3/2}\epsilon_n}\} \\ &= 2\exp\{\frac{-\log^2 n}{4C + O((nb)^{-1/2})}\}. \end{aligned} \tag{2.5.3}$$

This inequality will be used to prove the following

Lemma 2.5.2 *In addition to the Assumptions 2.1-3 with $\tau = 2$, suppose that $\{\varepsilon_i\}$ satisfy the Cramer's Condition. Then, for any $0 < a_0 \leq a_1 < 1$,*

$$\sup_{x\in[a_0,a_1]} |\hat{r}(x) - r(x)| = O_p((nb)^{(H-1)}\log n),$$

where the constant in O_p does not depend on a_0 and a_1.

Proof. Let $N = N_n$ be a sequence of positive integers such that $N = O(n)$. Let $a_0 = x_0 < x_1 < \cdots < x_{N-1} < x_N = a_1$ denote a partition of the interval $[a,b]$ such that $x_j - x_{j-1} \leq 1/N$, for all $1 \leq j \leq N$. By (2.2.4),

$$\begin{aligned} \sup_{x\in[a,b]} |\hat{r}(x) - r(x)| &\leq \sup_{x\in[a,b]} \Big(|\frac{1}{nb}\sum_{t=1}^{n} k_t(x)\sigma(\frac{t}{n})u_t| \\ &\quad + |\frac{1}{nb}\sum_{t=1}^{n} k_t(x)r(\frac{t}{n}) - r(x)|\Big) \\ &\leq \sup_{x_i}\Big|\frac{1}{nb}\sum_{t=1}^{n} k_t(x_i)\sigma(\frac{t}{n})u_t\Big| \\ &\quad + \max_i \sup_{x_{i-1} < x \leq x_i} |\frac{1}{nb}\sum_{t=1}^{n}(k_t(x_i) - k_t(x))\sigma(\frac{t}{n})u_t| + O(b^2) \\ &\leq \sup_{x_i}\Big(|S_L(x_i)| + |S_L^+(x_i)| + |S_L^-(x_i)|\Big) \\ &\quad + \max_i \sup_{x_{i-1} < x \leq x_i} |\frac{1}{nb}\sum_{t=1}^{n}(k_t(x_i) - k_t(x))\sigma(\frac{t}{n})u_t| \\ &\quad + O(b^2) \\ &= I + II + O(b^2), \qquad \text{say.} \end{aligned}$$

Consider the term II. Using Lipschitz condition of the kernel function K, and the Ergodic Theorem, we have

$$\begin{aligned} II &= \max_{0\le j\le N} \sup_{x_{j-1} < x \le x_j} \left|\frac{1}{nb}\sum_{t=1}^{n}\Big(k_t(x) - k_t(x_j)\Big)\sigma(\frac{t}{n})u_t\right| \\ &\le \max_{0\le j\le N} \sup_{x_{j-1} < x \le x_j} \frac{C}{nb}\frac{|x_j - x_{j-1}|}{b}\sum_{t=nx-nb}^{nx+nb}|u_t| \\ &\le \max_{0\le j\le N} \sup_{x_{j-1} < x \le x_j} \frac{C}{Nb}\frac{1}{nb}\sum_{t=nx-nb}^{nx+nb}|u_t| \\ &= O_p(\frac{1}{Nb}) = o_p((nb)^{H-1}). \end{aligned}$$

Consider the first term I. By choosing large L, say $L = n^{2/(2-2H)}$, the terms involving S_L^+ and S_L^- are of smaller order than $(nb)^{H-1}\log n$. In view of the assumed Cramer's condition for ε_i, the lemma thus follows from (2.5.3) in a routine fashion. □

2.6 Estimation of H

In practice, the parameters G and H appearing in the spectral density (2.1.3) are unknown. In order to be able to use the studendized versions of $\hat{r}(x)$ and $\hat{\sigma}(x)$ to make the large sample inference about $r(x)$ and $\sigma(x)$, a $\ln(n)$ consistent estimator of H and a consistent estimator of G under the model (2.1.1), (2.1.2) and (2.1.3) are needed.

In the parametric case, i.e., when the spectral density is specified up to a finite dimensional parameter, the Whittle estimators are known to be $n^{1/2}$-consistent and asymptotically normal, proved by Fox and Taqqu (1986) for the Gaussian processes, and by Giraitis and Surgailis (1990) for linear processes. For semi-parametric models where the spectral density is like in (2.1.3), the popular estimates are the local Whittle estimates as in Robinson (1995a, 1995b, 1997).

In a finite sample simulation study, Taqqu and Teverovsky (1995, 1997) observed that the Whittle estimator of H based on a stationary observable process, where there are no nuisance parameters except the mean, is by far the most accurate when the correct parametric model is chosen, and robust against certain departures from the correct model. Otherwise, judging from the simulation data, the local Whittle estimator with an appropriate m is preferable over other estimators when the form of the spectral density is not completely known, and in this case, its consistency rate is $m^{1/2}$ with $m = o(n)$. Taqqu and Teverovsky (1997) suggest a value of $m \approx n/32$ if the length n of the time series is 10,000. Dalla, Giraitis and Hidalgo (2005) recently give the convergence rate of the estimator $\hat{H}$ for some general stationary processes satisfying (2.1.3) and "ergodicity" conditions, including linear processes and functional Gaussian processes. It is thus desirable to develop their analogs that will be useful for making inference in the model (2.1.1) and (2.1.2).

To define these estimates, for an arbitrary stationary process $\xi_t, t = 1, 2, \cdots, n$, define the discrete Fourier transform and periodogram

$$c_{\xi,n}(\lambda) = (2\pi n)^{-1/2} \sum_{t=1}^{n} \xi_t e^{it\lambda}, \qquad I_{\xi,n}(\lambda) = |c_{\xi,n}(\lambda)|^2.$$

Let $\hat{e}_t := y_t - \hat{r}(t/n)$. Fix $1/2 < \Delta_1 < \Delta_2 < 1$. With $\lambda_j := 2\pi j/n$ and an integer $m \in [1, n/2)$, define $\Delta_1 \leq \psi \leq \Delta_2$,

$$\bar{Q}(\psi) := \frac{1}{m} \sum_{j=1}^{m} \lambda_j^{2\psi-1} I_{\hat{e},n}, \quad \bar{R}(\psi) = \log \bar{Q}(\psi) - (2\psi - 1) \sum_{j=1}^{m} \log \lambda_j.$$

The local Whittle estimates of G and H are defined to be

$$\bar{G} = \bar{Q}(\bar{H}), \quad \bar{H} = \arg \min_{\psi \in [\Delta_1, \Delta_2]} \bar{R}(\psi).$$

Under some regularity conditions including the assumption that $\sigma(x)$ is constant in x, Robinson (1997) proved that

$$\ln(n)(\bar{H} - H) \to_p 0, \qquad \bar{G} - G \longrightarrow_p 0. \tag{2.6.1}$$

Among the conditions required by Robinson (1997) to prove (2.6.1) are the following two conditions.

(2.6.2) There exist $\beta \in (0,2]$ and $G_1 \neq 0$ such that the spectral density f satisfies

$$f(\lambda) = \lambda^{1-2H}(G + G_1\lambda^{\beta} + o(\lambda^{\beta})), \quad \text{as } \lambda \to 0^+.$$

(2.6.3) $\alpha(\lambda) = \sum_{j=1}^{\infty} \alpha_j e^{ij\lambda}$ is differentiable for all λ in a neighborhood of 0, and

$$(d/d\lambda)\alpha(\lambda) = O(|\alpha(\lambda)|/\lambda), \text{ as } \lambda \to 0^+.$$

In view of (2.1.2) and (2.3.11) of Zygmund (1968, page 70), (2.6.2) is satisfied in our case with $\beta = 2H - 1$, while Li (2004) has shown that (2.6.3) is also satisfied here.

We shall now construct the analogs of $\hat{G}$ and $\hat{H}$ under the heteroscedastic regression set up (2.1.1) and (2.1.2), where $\sigma(x)$ may depend on x, that will satisfy (2.6.1).

Note that in the above estimators, the periodograms are based on the entire sample. Here we shall construct analogs of these estimators that use periodograms based on only a subset of the residuals. One of the reasons for doing this is to eliminate the boundary effects of $\hat{r}(x)$ and $\hat{\sigma}(x)$ on $\hat{H}$. More precisely, we first use all observations $Y_1, \cdots, Y_n$ to estimate $r(x)$ and $\sigma(x)$, but with bandwidths d_n and e_n possibly different from those in sections 2.2-4. Then we use periodograms based on a subset of residuals, say, $\hat{u}_{k_0+1}, \cdots, \hat{u}_{k_0+M}$, to construct $\hat{H}$ as follows, where k_0 and M are some positive integers, with $k_0 + M < n$, and where $\hat{u}_t := (y_t - \hat{r}(t/n))/\hat{\sigma}(t/n)$, $t \geq 1$. Accordingly, for any process ξ_t, and $1/2 < \psi < 1$, let

$$\omega_{\xi,M}(\lambda) = (2\pi M)^{-1/2} \sum_{t=k_0+1}^{k_0+M} \xi_t e^{it\lambda}, \quad I_{\xi,M}(\lambda) = |\omega_{\xi,M}(\lambda)|^2,$$

$$Q_{\xi}(\psi) := \frac{1}{m} \sum_{j=1}^{m} \beta_j^{2\psi-1} I_{\xi,M}.$$

With $\beta_j := 2\pi j/M$ and an integer $m \in [1, M/2)$, $\Delta_1 \le \psi \le \Delta_2$, define

$$\hat{R}(\psi) := \log Q_{\hat{u}}(\psi) - (2\psi - 1)\sum_{j=1}^{m} \log \beta_j, \tag{2.6.4}$$

$$\hat{G} := Q_{\hat{u}}(\hat{H}), \qquad \hat{H} = \mathrm{argmin}_{\psi \in [\Delta_1, \Delta_2]} \hat{R}(\psi). \tag{2.6.5}$$

We shall show that

$$\ln(M)(\hat{H} - H) \to_p 0, \qquad \hat{G} - G \to_p 0. \tag{2.6.6}$$

As will be seen later one can take $M = n^a$ for some $a < 1$. Then the first claim of (2.6.6) is equivalent to $\log(n)(\hat{H} - H) \longrightarrow_p 0$.

One of the assumptions needed for proving (2.6.6) is

Assumption 2.8. For $H \ge \Delta_1 > 1/2$, as $n \ge M \to \infty$, $m = o(M)$ and

$$(\ln M)^4\Big(\frac{1}{m^2 d^2}\frac{M}{n} + \frac{(\log m)^2}{m} + d + Md^4 + \frac{1}{nd^2}\Big) \to 0,$$

where $d \equiv d_n$ is the bandwidth used to estimate $r(x)$.

We shall use the following decomposition of $\hat{u}_t$ in the sequel.

$$\begin{aligned} \hat{u}_t &= u_t + V_t, \quad V_t := \xi_t + v_t + \zeta_t, \quad \xi_t := u_t\Big(\frac{\sigma(t/n)}{\hat{\sigma}(t/n)} - 1\Big), \\ v_t : &= \frac{r(\frac{t}{n}) - \hat{r}(\frac{t}{n})}{\sigma(\frac{t}{n})}, \quad \zeta_t = v_t\Big(\frac{\sigma(t/n)}{\hat{\sigma}(t/n)} - 1\Big), \quad \text{and} \\ \tilde{u}_t &= u_t + v_t = \frac{Y_t - \hat{r}(\frac{t}{n})}{\sigma(\frac{t}{n})}. \end{aligned} \tag{2.6.7}$$

Let

$$\begin{aligned} \tilde{R}(\psi) &:= \log Q_{\tilde{u}}(\psi) - (2\psi - 1)\sum_{j=1}^{m} \log \beta_j. \\ \widetilde{G} &:= Q_{\tilde{u}}(\tilde{H}), \qquad \tilde{H} = \mathrm{argmin}_{\psi \in [\Delta_1, \Delta_2]} \tilde{R}(\psi). \end{aligned}$$

The following proposition is helpful in proving (2.6.6).

Proposition 2.6.1 *Suppose the Assumptions 2.1-3 with $\tau = 2$ and 2.8 hold. Then,*

$$\log(M)(\tilde{H} - H) \to_p 0, \tag{2.6.8}$$

$$\tilde{G} - G \to_p 0. \tag{2.6.9}$$

The proof of this proposition is facilitated by the following two lemmas and the inequalities

$$\begin{aligned} |I_{\hat{u},M} - I_{u,M}| &\leq 2|I_{u,M}I_{V,M}|^{1/2} + I_{V,M}, \\ |I_{\tilde{u},M} - I_{u,M}| &\leq 2|I_{u,M}I_{v,M}|^{1/2} + I_{v,M} \\ I_{V,M} &\leq 5(I_{\xi,M} + I_{v,M} + I_{\zeta,M}), \quad \forall M \geq 1. \end{aligned} \tag{2.6.10}$$

In the first lemma,

$$f_j = G\beta_j^{1-2H}, \ 1 \leq j \leq m. \qquad G(\psi) = \frac{G}{m}\sum_{j=1}^{m} \beta_j^{2(\psi-H)}.$$

Lemma 2.6.1 *Under conditions (2.1.2), (2.1.3) and $m = o(M)$, we obtain that uniformly in $0 < t \leq 1$,*

$$[tm]^{-1}\sum_{j=1}^{[tm]} \frac{I_{u,M}(\beta_j)}{f_j} \to_p 1, \qquad a.s. \text{ for } 1 \leq j \leq m, \quad m \to \infty. \tag{2.6.11}$$

$$\sup_{1/2<\psi<1} \left|\frac{Qu(\psi) - G(\psi)}{G(\psi)}\right| = o_p(1). \tag{2.6.12}$$

Proof. By assumption, u_t, $t \in \mathbb{Z}$, is a stationary and ergodic process. Hence the claim (2.6.11) follows directly from the Proposition 2.4 in Dalla, Giraitis and Hidalgo (2004). To prove (2.6.12), using the first claim, the facts that for $|\delta| < 1$ and for all large m,

$$(m-1)^{\delta+1} - m^{\delta+1} = m^{\delta+1}[(\frac{m-1}{m})^{\delta+1} - 1] \sim m^{\delta}, \ \ G(\psi) = O((\frac{m}{M})^{2(\psi-H)}),$$

and (2.4.3), we obtain

$$Q_u(\psi) - G(\psi) = \frac{G}{m}\sum_{j=1}^{m} \beta_j^{2(\psi-H)}[\frac{I_{u,M}(\beta_j)}{f_j} - 1]$$

$$= \frac{G}{m}\sum_{j=1}^{m}[\frac{I_{u,M}(\beta_j)}{f_j} - 1](\frac{2\pi m}{M})^{2(\psi-H)}$$

$$+\frac{G}{m}\sum_{k=1}^{m-1}\sum_{i=1}^{k}[\frac{I_{u,M}(\lambda_i)}{f_i} - 1]\Big((\frac{2k\pi}{M})^{2(\psi-H)} - (\frac{2(k+1)\pi}{M})^{2(\psi-H)}\Big)$$

$$= o_p((\frac{m}{M})^{2(\psi-H)}). \qquad \square$$

Lemma 2.6.2 *For the moving average process u_t of (2.1.2),*

$$\sum_{j=1}^{m} j^{\psi}\frac{I_{u,M}(\beta_j)}{f_j} = O_p(m^{\psi+1}), \qquad \psi > -1.$$

$$= O_p(1), \qquad \psi < -1.$$

Proof. Formula (2.4.3) and (2.6.11) give

$$\sum_{j=1}^{m} j^{\psi}\frac{I_{u,M}(\beta_j)}{f_j} \sim (\frac{1}{m}\sum_{j=1}^{m}\frac{I_{u,M}(\beta_j)}{f_j})m^{\psi+1} + \sum_{j=1}^{m-1}(\frac{1}{j}\sum_{i=1}^{j}\frac{I_{u,M}(\lambda_i)}{f_i})j^{\psi}$$

$$= O_p(m^{\psi+1} + \sum_{j=1}^{m-1} j^{\psi}). \qquad \square$$

Proof of Proposition 2.6.1. The main idea of the proof here is similar to that of Theorem 3 of Robinson (1997). The proof is given in several steps. We shall first prove (2.6.8).

For an $\epsilon \in (0, \frac{1}{2}\ln(M))$, $\delta \in (\epsilon/\ln(M), 1/2)$, let $M_\epsilon = \{\psi : |(\ln M)(\psi - H)| < \epsilon\}$, $N_\delta = \{\psi : |\psi - H| < \delta\}$, $\Theta = [\Delta_1, \Delta_2]$ with $\Delta_1 > 1/2$, and $S(\psi) := \widetilde{R}(\psi) - \widetilde{R}(H)$. For any subset $A \subset \mathbb{R}$, let $\overline{A}$ denote its complement. We have

$$P\Big(|\ln(M)(\tilde{H} - H)| \geq \epsilon\Big) = P\Big(\tilde{H} \in \Theta\bigcap\overline{M}_\epsilon\Big)$$

$$= P\Big(\inf_{\overline{M}_\epsilon\bigcap\Theta} \tilde{R}(\psi) \leq \inf_{M_\epsilon\bigcap\Theta} \tilde{R}(\psi)\Big) \leq P\Big(\inf_{\overline{M}_\epsilon\bigcap\Theta} S(\psi) \leq 0)$$

$$\leq P\Big(\inf_{\overline{N}_\delta\bigcap\Theta} S(\psi) \leq 0\Big) + P\Big(\inf_{\overline{M}_\epsilon\bigcap N_\delta\bigcap\Theta} S(\psi) \leq 0\Big),$$

where the third inequality above holds because $H \in M_\epsilon \bigcap \Theta$. As in the proof of Theorem 3 of Robinson (1997), the right hand side of the above expression tends to 0 if

$$\sup_{\Theta} |\frac{Q_{\tilde{u}}(\psi) - G(\psi)}{G(\psi)}| + (\ln M)^2 \sup_{\Theta \bigcap N_\delta} |\frac{Q_{\tilde{u}}(\psi) - G(\psi)}{G(\psi)}| \to_p 0,$$

which in turn, in view of the triangle inequality, is implied by

$$\sup_{\Theta} |\frac{Q_u(\psi) - G(\psi)}{G(\psi)}| + (\ln M)^2 \sup_{\Theta \bigcap N_\delta} |\frac{Q_u(\psi) - G(\psi)}{G(\psi)}| \to_p 0, \tag{2.6.13}$$

$$\sup_{\Theta} |\frac{Q_{\tilde{u}}(\psi) - Q_u(\psi)}{G(\psi)}| + (\ln M)^2 \sup_{\Theta \bigcap N_\delta} |\frac{Q_{\tilde{u}}(\psi) - Q_u(\psi)}{G(\psi)}| \to_p 0. \tag{2.6.14}$$

Let

$$\mathcal{D}_j =: \frac{I_{\tilde{u},M}(\beta_j) - I_{u,M}(\beta_j)}{f_j}, \ \hat{\mathcal{D}}_j =: \frac{I_{\hat{u},M}(\beta_j) - I_{u,M}(\beta_j)}{f_j}. \tag{2.6.15}$$

2. *Sufficient conditions* According to the proof of Theorem 3 in Robinson (1997), to prove (2.6.13) and (2.6.14), it suffices to show that

$$\sum_{i=1}^{m-1} (\frac{i}{m})^{2(\Delta_1 - H)+1} \frac{1}{i^2} | \sum_{j=1}^{i} \mathcal{D}_j | \to_p 0, \tag{2.6.16}$$

$$(\log M)^2 \sum_{i=1}^{m-1} (\frac{i}{m})^{1-2\delta} \frac{1}{i^2} | \sum_{j=1}^{i} \mathcal{D}_j | \to_p 0 \tag{2.6.17}$$

for some arbitrarily small $\delta > 0$, and

$$\frac{(\log M)^2}{m} \sum_{j=1}^{m} \mathcal{D}_j \to_p 0. \tag{2.6.18}$$

Towards verifying the above conditions, we first need to obtain a bound on

$I_{v,M}(\beta_j)/f_j$. Recall that $f_j = G\beta_j^{1-2H}$, $v_t := (r(\frac{t}{n}) - \hat{r}(\frac{t}{n}))/\sigma(\frac{t}{n})$. We have

$$
\begin{aligned}
EI_{v,M}(\lambda) &= \frac{1}{2\pi M} E \sum_{s,t=1}^{M} \frac{\hat{r}(\frac{t}{n}) - r(\frac{t}{n})}{\sigma(\frac{t}{n})} \frac{r(\frac{s}{n}) - \hat{r}(\frac{s}{n})}{\sigma(\frac{s}{n})} e^{i(t-s)\lambda} \\
&= \frac{1}{2\pi M} \Big| \sum_{t=1}^{M} \frac{\left(\frac{1}{nd}\sum_{l=1}^{n} k_l(\frac{t}{n}) r(\frac{l}{n}) - r(\frac{t}{n})\right)}{\sigma(\frac{t}{n})} e^{it\lambda} \Big|^2 \\
&\quad + \frac{1}{2\pi M} \sum_{s,t=1}^{M} \frac{1}{\sigma(\frac{t}{n})\sigma(\frac{s}{n}) n^2 d^2} \sum_{l,p=1}^{n} k_l(\frac{t}{n}) k_p(\frac{s}{n}) \sigma(\frac{l}{n}) \sigma(\frac{p}{n}) \\
&\quad \times E u_l u_p e^{i(t-s)\lambda}
\end{aligned}
$$

$$
(2.6.19) \qquad := A(\lambda) + B(\lambda).
$$

Consider the term $B(\lambda)$ first. By a change of variable, we obtain

$$
\begin{aligned}
J_\lambda(\omega) &:= \sum_{t=1}^{M} \sum_{l=1}^{n} K(\frac{t-l}{nd}) \frac{\sigma(\frac{l}{n})}{\sigma(\frac{t}{n})} e^{it\lambda - il\omega} \\
&= \sum_{|s| \le nd} K(\frac{s}{nd}) e^{is\omega} \sum_{t=1}^{M} \frac{\sigma(\frac{t-s}{n})}{\sigma(\frac{t}{n})} e^{it(\lambda-\omega)}.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
B(\lambda) &= (2\pi M n^2 d^2)^{-1} \int_{-\pi}^{\pi} f(\omega) \sum_{t,s=1}^{M} \frac{1}{\sigma(\frac{t}{n})\sigma(\frac{s}{n})} \sum_{l,p=1}^{n} k_l(\frac{t}{n}) k_p(\frac{s}{n}) \\
&\quad \times \sigma(\frac{l}{n}) \sigma(\frac{p}{n}) e^{i(p-l)\omega} e^{i(t-s)\lambda} d\omega \\
&= (2\pi M n^2 d^2)^{-1} \int_{-\pi}^{\pi} f(\omega) |J_\lambda(\omega)|^2 d\omega \\
&\le (2\pi M n^2 d^2)^{-1} [B_1(\lambda) + B_2(\lambda) + B_3(\lambda)],
\end{aligned}
$$

where

$$
\begin{aligned}
B_1(\lambda) &:= \int_{\lambda/2}^{3\lambda/2} |f(\omega) - f(\lambda)| |J_\lambda(\omega)|^2 d\omega, \\
B_2(\lambda) &:= \{ \int_{3\lambda/2}^{\pi} + \int_{-\pi}^{\lambda/2} \} f(\omega) |J_\lambda(\omega)|^2 d\omega, \\
B_3(\lambda) &:= f(\lambda) \int_{-\pi}^{\pi} |J_\lambda(\omega)|^2 d\omega, \quad \lambda \in [-\pi, \pi].
\end{aligned}
$$

To bound B_j's we first obtain some bounds on J_λ. Let $D_k(\lambda) = \sum_{j=1}^k e^{ij\lambda}$, $s_k = D_k(\lambda - \omega)$. Recall from Zygmund (1968, page 51) that

$$|D_k(\lambda)| \le C/\lambda, \quad \forall\, \lambda \in [-\pi, \pi],\ k \ge 1. \tag{2.6.20}$$

For $M \le n$, if we take $a_j = e^{ij(\lambda-\omega)}$ and $b_k = \sigma((t-s)/n)/\sigma(t/n)$, (2.4.3) yield

$$\begin{aligned}
|\sum_{t=1}^{M} \frac{\sigma(\frac{t-s}{n})}{\sigma(\frac{t}{n})} e^{it(\lambda-\omega)}| &\le |s_M \frac{\sigma(\frac{M-s}{n})}{\sigma(\frac{M}{n})}| + |\sum_{k=1}^{M-1} s_k \Big(\frac{\sigma(\frac{k-s}{n})}{\sigma(\frac{k}{n})} - \frac{\sigma(\frac{k+1-s}{n})}{\sigma(\frac{k+1}{n})},\Big)| \\
&\le C\,|D_M(\lambda-\omega)| + \frac{C}{n}\sum_{k=1}^{M-1} |D_k(\lambda-\omega)|.
\end{aligned}$$

Hence, using the fact $|K(s/nd)e^{is\omega}| \le C$, for all s, ω, and by (2.6.20), we obtain

$$\begin{aligned}
|J_\lambda(\omega)| &\le C\,nd|D_M(\lambda-\omega)| + Cd\sum_{k=1}^{M-1} |D_k(\lambda-\omega)| \\
&\le C\frac{nd}{|\lambda-\omega|}, \qquad \forall\, \lambda, \omega \in [-\pi, \pi].
\end{aligned} \tag{2.6.21}$$

Similarly, (2.4.3) applied with $a_j = e^{ij\lambda}$, $b_j = K((t-s)/(nd))\sigma(s/n)/\sigma(t/n)$, and Lipschitz properties of K and σ yield

$$\Big|\sum_{t=1}^{M} K(\frac{t-s}{nd})\frac{\sigma(\frac{s}{n})}{\sigma(\frac{t}{n})} e^{it\lambda}\Big| \le \frac{C}{\lambda}, \qquad \forall\, \lambda \in [-\pi, \pi].$$

Since $\int_{-\pi}^{\pi} e^{i(k-l)\omega} d\omega = 0$ of $k \ne l$, we obtain that for all $\lambda \in [-\pi, \pi]$ and for all $n \ge 1$,

$$\int_{-\pi}^{\pi} |J_\lambda(\omega)|^2 d\omega = 2\pi \sum_{s=1}^{n} \Big|\sum_{t=1}^{M} K(\frac{t-s}{nd})\frac{\sigma(\frac{s}{n})}{\sigma(\frac{t}{n})} e^{it\lambda}\Big|^2 \le C\frac{n}{\lambda^2}. \tag{2.6.22}$$

Now, we are ready to give bounds on B_j's. Clearly, by (2.6.22),

$$B_3(\lambda) \le C\, f(\lambda) \min\{1, (Mnd^2\lambda^2)^{-1}\}), \qquad \forall\, \lambda.$$

Next, consider B_2. Note that $3\lambda/2 \le \omega \le \pi$, implies $\lambda/2 \le \omega - \lambda \le \pi - \lambda$ and $-\pi \le \omega < \lambda/2$ implies $\lambda/2 \le \lambda - \omega \le \pi + \lambda$. Hence, by (2.6.21),

$$B_2(\lambda) \le C\frac{n^2 d^2}{\lambda^2}.$$

Finally, since $\sup_{\lambda/2\le\omega\le 3\lambda/2}\{|f(\lambda)-f(\omega)|/|\lambda-\omega|\} = O(f(\lambda)/\lambda)$, as $\lambda \to 0^+$, from (2.6.20) and (2.6.21), $B_1(\lambda)$ is bounded above by

$$\frac{Cf(\lambda)}{M\lambda}\int_{\lambda/2}^{3\lambda/2}|D(\lambda-\omega)|d\omega \le \frac{Cf(\lambda)}{M\lambda}\int_0^{\lambda/2}|D(\omega)|d\omega.$$

This bound and Lemma 5 of Robinson (1994b) imply that uniformly in $j = 1,\cdots,m$, $B_1(\beta_j) = O(f_j(1+\log j)/j)$, as $M\to\infty$. This fact in turn together with the above bounds on B_3 and B_2 readily imply that uniformly in $j = 1,\cdots,m$,

$$B(\beta_j) = O_P\Big(f_j\{\min(1,\frac{M}{n}\frac{1}{d^2j^2}) + \frac{1+\log j}{j}\}\Big),\ M\to\infty. \tag{2.6.23}$$

Next, consider the term $A(\lambda)$ in (2.6.19). Arguing as in Robinson (1997), we have $A(\lambda) \le 2A_1^2(\lambda) + 2A_2^2(\lambda)$, where

$$\begin{aligned} A_1(\lambda) &:= \frac{1}{\sqrt{2\pi M}}\Big|\sum_{t=1}^{M}\frac{\frac{1}{nd}\sum_{l=1}^{n}k_l(\frac{t}{n})(r(\frac{l}{n})-r(\frac{t}{n}))}{\sigma(\frac{t}{n})}e^{it\lambda}\Big|, \\ A_2(\lambda) &:= \frac{1}{\sqrt{2\pi M}}\Big|\sum_{t=1}^{M}\frac{\Big(\frac{1}{nd}\sum_{l=1}^{n}k_l(\frac{t}{n})-1\Big)r(\frac{t}{n})}{\sigma(\frac{t}{n})}e^{it\lambda}\Big|. \end{aligned}$$

By the two term Taylor expansion and the Assumption 2.3, we obtain

$$\begin{aligned} A_1(\lambda) \;&\le\; \frac{C}{\sqrt{M}nd}\sum_{t=1}^{M}|r'(\frac{t}{n})|\Big|\sum_{l=1}^{n}k_l(\frac{t}{n})(\frac{l-t}{n})\Big| \\ &\quad + \frac{C}{\sqrt{M}nd}\sum_{t=1}^{M}\sum_{l=1}^{n}(\frac{l-t}{n})^2|k_l(\frac{t}{n})|. \end{aligned} \tag{2.6.24}$$

The last term in the above bound is $O(\sqrt{M}d^2)$, since

$$\begin{aligned} \sum_{t=1}^{M}\sum_{l=1}^{n}(\frac{l-t}{n})^2|k_l(\frac{t}{n})| \;&\sim\; \sum_{t=1}^{M}\int(\frac{l-t}{n})^2|k_l(\frac{t}{n})|dl \\ &\sim\; \sum_{t=1}^{M}nd\int d^2\omega^2|K(\omega)|d\omega = O(Mnd^3). \end{aligned}$$

Using $\int vK(v)dv = 0$ and the boundedness of r', the first term on the right of (2.6.24) is seen to be bounded above by

$$\frac{C}{\sqrt{M}nd}\sum_{t=1}^{M}\Big|\sum_{l=1}^{n}k_l(\frac{t}{n})(\frac{l-t}{n}) - nd^2\int_{-1}^{1}vK(v)dv\Big|$$
$$\leq \frac{Cd}{\sqrt{M}}\sum_{t=1}^{M}\Big|\sum_{l=1}^{n}\int_{(t-l)/nd}^{(t+1-l)/nd}\Big\{(\frac{t-l}{nd})K(\frac{t-l}{nd}) - vK(v)\Big\}dv\Big|$$

for large n. By the Mean Value Theorem this is bounded by

$$\frac{Cd}{\sqrt{M}}\sum_{t=1}^{M}\Big\{\frac{1}{n^2d^2}\sum_{l=1}^{n}K(\frac{t-l}{nd}) + \frac{1}{nd}\int_{-1}^{1}|v|dv\Big\} = O(\frac{\sqrt{M}}{n}).$$

Thus $A_1^2(\lambda) = O(Md^4 + M/n^2)$. Similarly $A_2^2(\lambda) = O(M/(n^2d^2))$.

These bounds together with (2.6.19) and (2.6.23) imply that uniformly in $1 \leq j \leq m$,

$$\frac{I_{v,M}(\beta_j)}{f_j} = O_P\Big(\min(1, \frac{M}{n}\frac{1}{d^2j^2}) + \frac{1+\log j}{j} + \frac{Md^4}{f_j} + \frac{M}{n^2d^2f_j}\Big). \tag{2.6.25}$$

We are now ready to verify the sufficient conditions (2.6.16) - (2.6.18). By changing the order of summation, the left-hand side of (2.6.16) is seen to be bounded above by

$$A_1 := Cm^{2(H-\Delta_1)-1}\sum_{j=1}^{m}j^{2(\Delta_1-H)}|\mathcal{D}_j|.$$

By (2.6.7) and the Cauchy-Schwarz inequality,

$$\sum_{j=1}^{m}j^{2(\Delta_1-H)}|\mathcal{D}_j| \leq C\sum_{j=1}^{m}j^{2(\Delta_1-H)}\frac{|I_{u,M}(\beta_j)\,I_{v,M}(\beta_j)|^{1/2} + I_{v,M}(\beta_j)}{f_j}$$
$$\leq C\Big(\sum_{j=1}^{m}j^{2(\Delta_1-H)}\frac{|I_{u,M}(\beta_j)|}{f_j}\Big)^{1/2}\Big(\sum_{j=1}^{m}j^{2(\Delta_1-H)}\frac{|I_{v,M}(\beta_j)|}{f_j}\Big)^{1/2}$$
$$+\sum_{j=1}^{m}j^{2(\Delta_1-H)}\frac{|I_{v,M}(\beta_j)|}{f_j}.$$

This bound together with Lemmas 2.6.1 and 2.6.2, (2.6.25), Assumption 2.8 and the following calculations imply that $A_1 = o_p(1)$.

$$m^{2(H-\Delta_1)-1}\sum_{j=1}^{m} j^{2(\Delta_1-H)}\frac{M}{n}\frac{1}{d^2 j^2} = \frac{M}{n}\frac{1}{d^2 m^2},$$

$$m^{2(H-\Delta_1)-1}\sum_{j=1}^{m} j^{2(\Delta_1-H)}\frac{Md^4}{f_j} = M^{2-2H}m^{2H-1}d^4 \leq Md^4,$$

$$m^{2(H-\Delta_1)-1}\sum_{j=1}^{m} j^{2(\Delta_1-H)}\frac{M}{n^2 d^2 f_j} = \frac{M^{2-2H}m^{2H-1}}{n^2 d^2} \leq \frac{1}{nd^2}.$$

The left-hand side of (2.6.17) is bounded above by

$$\begin{aligned} & C(\log M)^2 m^{2\delta-1}\sum_{j=1}^{m} j^{-2\delta}|\mathcal{D}_j| \\ \leq \;& C(\log M)^2 m^{2\delta-1}\sum_{j=1}^{m} j^{-2\delta}\{\frac{(I_{u,M}(\beta_j)I_{v,M}(\beta_j))^{1/2}}{f_j} + \frac{I_{v,M}(\beta_j)}{f_j}\} \end{aligned}$$

which tends to zero in probability by applying Lemmas 2.6.1, 2.6.2, Assumption 2.8 and the similar calculations for $0 < \delta < 1/2$:

$$\begin{aligned} & (\log M)^2 m^{2\delta-1}\sum_{j=1}^{m} j^{-2\delta}\frac{I_{v,M}(\beta_j)}{f_j} \\ \leq \;& C(\log M)^2\{\frac{M}{nm^2 d^2} + \frac{\log^2 m}{m} + Md^4\frac{m^{2H-1}}{M^{2H-1}} + \frac{1}{nd^2}\frac{M^{2-2H}}{nm^{2H-1}}\}. \end{aligned}$$

And the same argument implies that (2.6.18) tends to zero in probability. □

We shall next show that (2.6.8) implies (2.6.9). To that effect, we have

$$\begin{aligned} |\tilde{G} - G| &\leq |Q_{\tilde{u}}(\tilde{H}) - Q_{\tilde{u}}(H)| + |Q_{\tilde{u}}(H) - G_u(H)| + |G_u(H) - G| \\ &= |\frac{1}{m}\sum_{j=1}^{m}(\beta_j^{2\tilde{H}-1} - \beta_j^{2H-1})I_{\tilde{u},M}(\beta_j)| \\ &\quad + |\frac{1}{m}\sum_{j=1}^{m}\beta_j^{2H-1}(I_{\tilde{u},M}(\beta_j) - I_{u,M}(\beta_j))| \\ &\quad + |\frac{1}{m}\sum_{j=1}^{m}\beta_j^{2H-1}I_{u,M}(\beta_j) - G|. \end{aligned}$$

By (2.6.11), the last term in this bound is $o_P(1)$. The middle term is exactly equal to $|Gm^{-1}\sum_{j=1}^{m}\mathcal{D}_j|$, which is $o_P(1)$ by (2.6.18). The first term is bounded above by

$$|\frac{1}{m}\sum_{j=1}^{m}\beta_j^{2H-1}(1-\beta_j^{2(\tilde{H}-H)})I_{\tilde{u},M}(\beta_j)| \leq \max_{1\leq j\leq m}|1-\beta_j^{2(H-\tilde{H})}|Q_{\tilde{u}}(H).$$

Moreover,

$$|\beta_j^{2(H-\tilde{H})}-1| = |\exp 2(H-\tilde{H})\ln\beta_j - 1| \leq 4|(H-\tilde{H})|\ln(M).$$

This bound together with (2.6.11) and the fact that $Q_{\tilde{u}}(H)\to_P Q_u(H)$ implied by (2.6.14) thus shows that (2.6.8) implies (2.6.9). This also completes the proof of the Proposition 2.6.1.

Remark 2.6.1 If we let $M = n^a$, $m = n^\eta$ and $d = n^{-\delta}$, Assumption 2.8 holds when $a/4 \leq \delta < 1/2$ and $0 < \delta < \eta \leq a \leq 1$.

Now, let $k_0 = k_{0n}$ be a sequence of positive integers such that $k_0 \to \infty$, $k_0/n \to x_0 \in (0,1)$ and $(k_0+M)/n \to x_1 \in (0,1)$, as $n\to\infty$. Let $\omega_k(\beta_j) = \sum_{t=k_0}^{k_0+k} u_t e^{it\beta_j}$. By Zygmund (1968, V.2), for any integer $M \geq k \to \infty$, if $m = o(k)$,

$$\frac{|\omega_k(\beta_j)|^2}{kf_j} = O_P(1), \quad 1\leq j\leq m \tag{2.6.26}$$

Let $\Psi_t = \sigma(t/n)/\hat{\sigma}(t/n)-1$. Using Lemma 2.8.3 and Lemma 2.8.4 in the Appendix here, we obtain that uniformly for $k_0 \leq k \leq k_0 + M$,

$$\Psi_k \to_P 0, \quad \sup_k|\Psi_k - \Psi_{k+1}| = O_P(\tau_n), \quad \tau_n = \frac{1}{nd}\vee\frac{1}{nh}, \tag{2.6.27}$$

where $h = h_n$ is a bandwidth sequence for estimating $\sigma(t)$. We need the following

Assumption 2.9 Assume that $m^4\log^4 M/M^3 + \tau_n M \to 0$.

Now we are ready to prove the following main theorem.

Theorem 2.6.1 *Suppose Assumptions 2.1-4, 2.8 and 2.9 hold with $\tau = 2$. Then (2.6.6) holds.*

Proof. For $t = k_0 + 1, k_0 + 2, \cdots, k_0 + M$, recall the residuals can be written as $\hat{u}_t = u_t + \xi_t + v_t + \eta_t$ as in (2.6.7) and (2.6.8). From the proof of Proposition 2.6.1, we know the sufficient conditions are analogs of (2.6.16)-(2.6.17) with $\mathcal{D}_j$ replaced by $\hat{\mathcal{D}}_j$. Using inequalities (2.6.10) and (2.6.11), we only need to show the following claim: uniformly in $1 \le j \le m$,

$$\frac{I_{\xi,M}(\beta_j)}{f_j} = o_P(1), \quad \frac{I_{\eta,M}(\beta_j)}{f_j} = O_P(\frac{I_{v,M}(\beta_j)}{f_j}). \tag{2.6.28}$$

By (2.4.3), we obtain that uniformly for $1 \le j \le m$,

$$\frac{I_{\xi,M}(\beta_j)}{f_j} = \frac{1}{2\pi M f_j}\Big|\sum_{t=k_0}^{k_0+M} u_t \Psi_t e^{it\beta_j}\Big|^2$$

$$= \frac{1}{2\pi M}\Big|\frac{\omega_M(\beta_j)}{f_j^{1/2}}\Psi_{k_0+M} + \sum_{k=k_0}^{k_0+M-1}\frac{\omega_k(\beta_j)}{f_j^{1/2}}(\Psi_k - \Psi_{k+1})\Big|^2$$

$$\le C\Psi_{k_0+M}^2 \frac{I_{u,M}(\beta_j)}{f_j} + \frac{C}{M}\sup_{k_0 \le k \le k_0+M}(\Psi_k - \Psi_{k+1})^2\Big(\sum_{k=1}^{M-1}\frac{|\omega_k(\beta_j)|}{f_j^{1/2}}\Big)^2$$

$$= I + II.$$

But, by (2.6.26) and (2.6.27), $I = o_P(I_{u,M}(\beta_j)/f_j) = o_P(1)$.

Next, consider the term II. For any positive integer $N < M - 1$, we have the inequality

$$\frac{(\sum_{k=1}^{M-1}|\omega_k(\beta_j)|)^2}{f_j} \le 2\frac{(\sum_{k=1}^{N}|\omega_k(\beta_j)|)^2}{f_j} + 2\frac{(\sum_{k=N+1}^{M-1}|\omega_k(\beta_j)|)^2}{f_j}.$$

By the Ergodic Theorem, $\sup_{k \le N}|\omega_k(\beta_j)| = O_P(N)$, for any sequence of positive integers $N \to \infty$. Using this fact we obtain that the first term in this upper bound is $O_P(N^4)$, for all $1 \le j \le m$. For the second term we choose $N = O(m \log M)$. Then by by (2.6.26), it is $O_P(M^3)$, for all $1 \le j \le m$. Hence, in view of (2.6.27), $II = O_P(N^4\tau_n^2/M) + O_P(\tau_n^2 M^2) = O_P(\tau_n^2 M^2) = o_P(1)$, by Assumption 2.9. Therefore the first claim in (2.6.28) holds.

Let $S_k(\beta_j) = \sum_{t=k_0+1}^{k_0+k} v_t e^{it\beta_j}$. Notice that in the proof of (2.6.25), the only requirement is $m = o(M)$. Therefore we can similarly obtain that for $N \le k \le M$, uniformly for $1 \le j \le m$,

$$\frac{|S_k(\beta_j)|^2}{M f_j} = O_P\Big(\frac{I_{v,M}(\beta_j)}{f_j}\Big). \tag{2.6.29}$$

Similarly to the proof of the first claim in (2.6.28), using the relation $\zeta_t = v_t \Psi_t$, see (2.6.8), we have

$$\begin{aligned}
\frac{I_{\zeta,M}(\beta_j)}{f_j} &= \frac{1}{2\pi M f_j}\Big|\sum_{t=k_0}^{k_0+M} v_t \Psi_t e^{it\beta_j}\Big|^2 \\
&= \frac{1}{2\pi M}\Big|\frac{S_M(\beta_j)}{f_j^{1/2}}\Psi_{k_0+M} + \sum_{k=k_0}^{k_0+M-1}\frac{S_k(\beta_j)}{f_j^{1/2}}(\Psi_k - \Psi_{k+1})\Big|^2 \\
&\le C\Psi_{k_0+M}^2 \frac{I_{v,M}(\beta_j)}{f_j} + \frac{C}{M}\sup_k(\Psi_k - \Psi_{k+1})^2\Big(\sum_{k=1}^{M-1}\frac{|S_k(\beta_j)|}{f_j^{1/2}}\Big)^2 \\
&= III + IV.
\end{aligned}$$

The term $III = o_P(I_{v,M}(\beta_j)/f_j)$. By (2.6.27), $|S_k(\beta_j)| = O_P(N)$, uniformly for all $k \le N$. Hence

$$\begin{aligned}
\frac{(\sum_{k=1}^{M-1}|S_k(\beta_j)|)^2}{f_j} &\le 2\frac{(\sum_{k=1}^{N}|S_k(\beta_j)|)^2}{f_j} + 2\frac{(\sum_{k=N+1}^{M-1}|S_k(\beta_j)|)^2}{f_j} \\
&= O_P(N^4) + O_P(M^3\frac{I_{v,M}(\beta_j)}{f_j}),
\end{aligned}$$

so that the second claim in (2.6.28) also holds. □

Remark 2.6.2 Let $h_n \le d_n$. Then as defined in Remark 2.6.1, Assumption 2.8 and 2.9 hold simultaneously if $a < 4\delta \wedge (1-\delta)$, $\delta < 1/2$ and $\delta < \eta < 3a/4$. If we take $\delta = 1/5$, then $a = 4/5$, we still have a large proportion of residuals to estimate H from.

Based on $\hat{H}$, we are now ready to state and prove the following corollary about the studentized versions of $\hat{r}(x)$ and $\hat{\sigma}^2(x)$ for $1/2 < \Delta_1 < H < \Delta_2$.

Corollary 2.6.1 *Let (2.1.1), (2.1.2) and the Assumptions 2.1-9 hold. Let $\hat{H}$ and $\hat{G}$ be as in (2.6.4). Then, for every fixed integer $k \geq 1$, and for any distinct $x_1, \cdots, x_k$ in $(0,1)$,*

$$\frac{(nb)^{1-\hat{H}}}{\hat{G}\rho(\hat{H})\hat{\sigma}(x_i)}\{\hat{r}(x_i) - r(x_i)\},\ i = 1, \cdots, k, \Longrightarrow N_k(0, \mathcal{I}), \tag{2.6.30}$$

Proof. Let $Z_i = (nb)^{1-H}(\hat{r}(x_i) - r(x_i))/\sigma_i$, $i = 1, \cdots, k$. Then, the left hand side of (2.6.30) can be rewritten as

$$(nb)^{H-\hat{H}}\frac{G\rho(H)\sigma_i}{\hat{G}\rho(\hat{H})\hat{\sigma}_i}Z_i,\ i = 1, \cdots k.$$

By Lemma 5 in Robinson (1997), $\rho(H)$ is continuous on $(0,1)$. Hence, by (2.6.1), $\rho(\hat{H}) \to_p \rho(H)$. In view of the Slutsky's theorem, the consistency of $\hat{G}$ and $\hat{\sigma}^2(x)$ and Theorem 2.2.1, the result (2.6.30) follows from (2.6.1) and that the fact that $\ln(nb) \leq \ln(n)$, for all sufficiently large n. □

Similarly, from Theorem 2.4.1, we obtain

Corollary 2.6.2 *Let (2.1.1), (2.1.2) and the Assumptions 2.1-9, and $\hat{H}$ and $\hat{G}$ be as in (2.6.4). Then,*

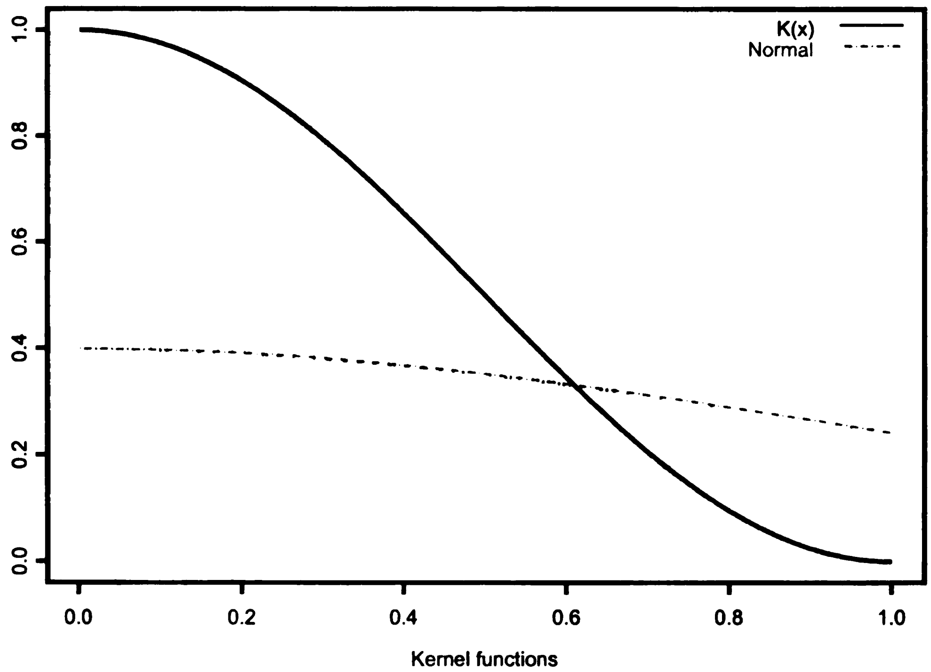
$$(nc)^{2(1-\hat{H})}\hat{\sigma}^{-2}(x)\{\hat{\sigma}^2(x) - \sigma^2(x)\} \Longrightarrow Y_2, \quad x \in (0,1),\ \frac{3}{4} < H < 1.$$

In the case $\frac{1}{2} < H < \frac{3}{4}$, from Theorem 2.4.2 we obtain that the asymptotic distribution of $(nc)^{1/2}(\hat{\sigma}^2(x) - \sigma^2(x))$ is normal with mean zero and standard deviaton $\sigma_2(x) = \sigma^2(x)\sigma_0 2^{1/2}\|W\|$, where σ_0 is given at (2.4.5). Thus to studentize $\hat{\sigma}^2(x)$ in this case, there is no need to estimate H, but one must estimate σ_0. Let $\hat{\sigma}_0$ be a consistent estimator of σ_0, and set $\hat{\sigma}_2(x) = \hat{\sigma}^2(x)\hat{\sigma}_0 2^{1/2}\|W\|$. Then by Theorem 2.4.2, we readily obtain $\sqrt{nb}\hat{\sigma}_2^{-1}(x)\{\hat{\sigma}^2(x) - \sigma^2(x)\} \Longrightarrow \mathcal{N}(0,1)$.

2.7 Application to the Nile river data

Now we apply the obtained results to analyze the Nile river data. In Figure 2.1, we plot Robinson's kernel (solid line) $K(x) = .5(1 + \cos(x\pi))I(|x| \leq 1)$ and the normal kernel (dashed line). Figure 2.2 is the kernel estimation of the regression function.

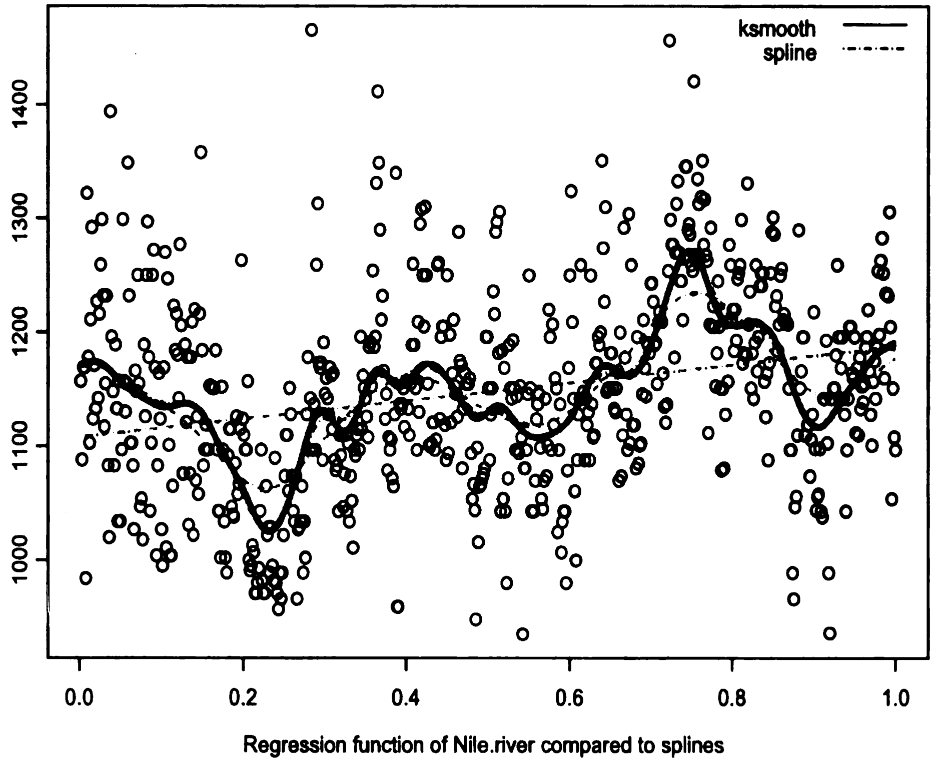
Figure 2.1: Kernel functions .



There is no difference in the estimation of regression function by using Robinson's kernel or the normal kernel, which is depicted by the solid line, and as a comparison we also try spline methods with different numbers of knots, which are dashed lines in this figure. When the bandwidth $b = .05obse$, we observed that Robinson's estimation and normal kernel smoothing and smooth spline with df=20 are the same.

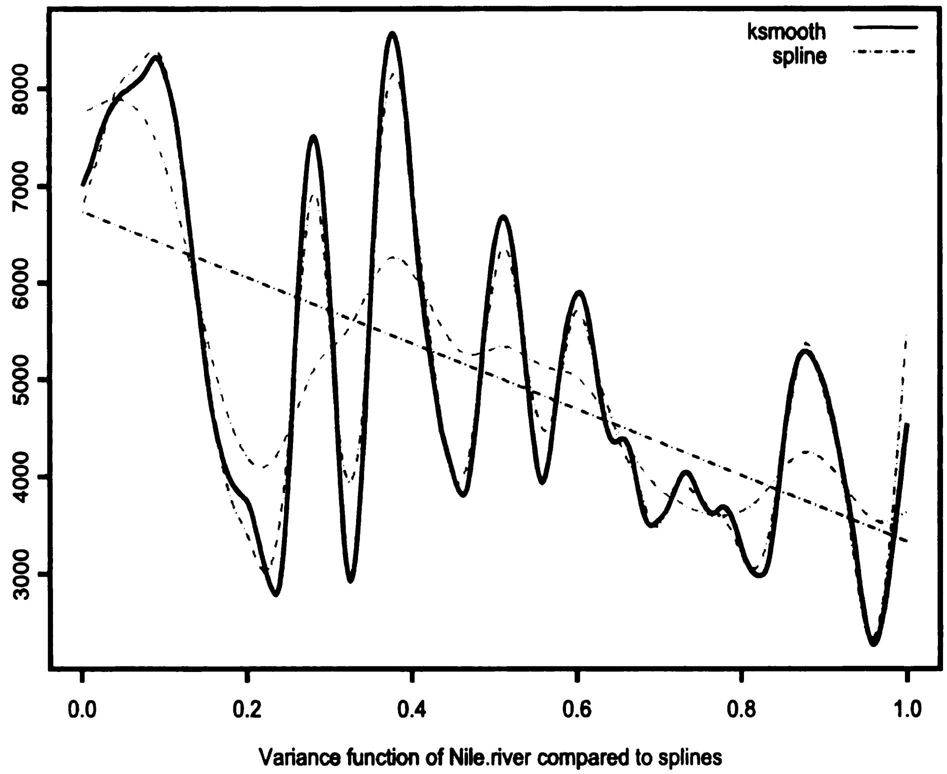
The evidence of Heteroscedasticity is presented in the Figure 2.3. The solid curve is the kernel estimation of variance function with kernel $K(x)$, which is similar to

Figure 2.2: Estimation of regression function.



that of smooth spline with df=20. For this small bandwidth $b = c = .05$, the local

Figure 2.3: Estimation of conditional variance function.



Whittle estimators are less than a half for both residuals $\hat{u} = (Y - \hat{m}(X))/\hat{\sigma}(X)$ and $\varepsilon = Y - \hat{m}(X)$, $\hat{H}_{\hat{u}} = 0.3945635$ and $\hat{H}_{\varepsilon} = 0.3803844$.

When we choose larger bandwidth, $b = c = .1$, say, we observe that all behaviors of regression function and variance function estimations are similar, except $\hat{H}_{\hat{u}} = 0.6497344$ and $\hat{H}_{\varepsilon} = 0.6570367$. Note that $\hat{H}_{Nileriver} \geq .9$. The findings may suggest that the long memory effect of the Nile river data may be partially caused by the non stationary trend and variation with time, instead of the error process u_t. The long /short memory effect are caused by different choices of bandwidth, which agrees with the results in Robinson (1997). In either case (with many other choices of b and c), it is suggested that $Y_t = r(t/n) + \sigma(t/n)u_t$ with heteroscedasticity may be a reasonable model for the Nile river data.

2.8 Appendix

This section contains some needed proofs of some of the results stated in the previous sections. To prepare for the proof of Lemma 2.3.3, we compute the fourth order moments of u_t as follows, where $\gamma_j := \sum_{k=0}^{\infty} \alpha_k \alpha_{j+k}$, with α_j's in (2.1.2).

$$\begin{aligned} Eu_0u_t^3 &= E\sum_{i,j,k,l} \alpha_i\alpha_j\alpha_k\alpha_l\varepsilon_{-i}\varepsilon_{t-j}\varepsilon_{t-k}\varepsilon_{t-l} \\ &= 3\sum_{-i=t-j,\,t-k=t-l}\ \sum \alpha_i\alpha_{t+i}\alpha_k^2 + 3\sum_i \alpha_i\alpha_{t+i}^3(E\varepsilon_0^4 - 1) \\ &\sim 3\gamma_t + o(\gamma_t), \end{aligned} \tag{2.8.1}$$

as $t \to \infty$. Similarly, for $1 \le s \le t$,

$$\begin{aligned} Eu_0u_su_t^2 &= E\sum_{i,j,k,l} \alpha_i\alpha_j\alpha_k\alpha_l\varepsilon_{-i}\varepsilon_{s-j}\varepsilon_{t-k}\varepsilon_{t-l} \\ &= \sum_{-i=s-j,\,t-k=t-l}\ \sum \alpha_i\alpha_{s+i}\alpha_k^2 + 2\sum_{-i=t-k,\,s-j=t-l}\ \sum \alpha_i\alpha_{t+i}\alpha_j\alpha_{t-s+j} \\ &\qquad + \sum_i \alpha_i\alpha_{s+i}\alpha_{t+i}^2E\varepsilon_0^4 \\ &= \gamma_s + 2\gamma_t\gamma_{t-s} + \sum_i \alpha_i\alpha_{s+i}\alpha_{t+i}^2(E\varepsilon_0^4 - 3) \\ &\sim \gamma_s + 2\gamma_t\gamma_{t-s} \qquad \text{as } t - s \to \infty, \end{aligned} \tag{2.8.2}$$

since the third term in the last but one equality above is bounded above by

$$\begin{aligned} & s^{H-3/2}t^{2H-3}\Big\{\int_0^s x^{H-3/2}(1+\frac{x}{s})^{H-3/2}(1+\frac{x}{t})^{2H-3}dx\Big\} \\ &\quad + \int_s^t x^{H-3/2}(1+\frac{x}{s})^{H-3/2}dx \\ &\quad + \int_t^{\infty} x^{H-3/2}(1+\frac{x}{s})^{H-3/2}(1+\frac{x}{t})^{2H-3}dx \\ &= O(s^{-2(1-H)}t^{2H-3} + t^{4H-5} + t^{4H-5}) = O(t^{4H-5}). \end{aligned}$$

Hence we also obtain that

$$Eu_0 u_s(u_t^2 - 1) \sim 2\gamma_t \gamma_{t-s}, \qquad \text{as } t - s \to \infty. \tag{2.8.3}$$

For s, t, r such that $|t-s|, |r-t|$ and $|s-r|$ are all large enough, we have

$$\begin{aligned} Eu_0 u_s u_t u_r &= E\sum \alpha_i \alpha_j \alpha_k \alpha_l \varepsilon_{-i} \varepsilon_{s-j} \varepsilon_{t-k} \varepsilon_{r-l} \\ &= \sum_i \alpha_i \alpha_{s+i} \alpha_{t+i} \alpha_{r+i} E\varepsilon_0^4 \\ &\quad + \sum_{-i=s-j} \sum_{t-k=r-k} \alpha_i \alpha_{s+i} \alpha_k \alpha_{r-t+k} \\ &\quad + \sum_{-i=t-k} \sum_{s-j=r-l} \alpha_i \alpha_{t+i} \alpha_j \alpha_{t-s+j} \\ &\quad + \sum_{-i=r-l,} \sum_{s-j=t-k} \alpha_i \alpha_{r+i} \alpha_j \alpha_{t-s+j} \\ &\sim \gamma_s \gamma_{r-t} + \gamma_t \gamma_{r-s} + \gamma_r \gamma_{t-s}. \end{aligned} \tag{2.8.4}$$

Proof of Lemma 2.3.3. We begin with the following decomposition:

$$\begin{aligned} &\mathrm{Var}(\hat\sigma^2(x) - \tilde\sigma^2(x)) \\ &= \mathrm{Var}\Big\{\frac{1}{nc}\sum_t w_t(x)\Big[\Big\{r_t + \sigma_t u_t - \hat r_t\Big\}^2 - \sigma_t^2 u_t^2\Big]\Big\} \\ &= E\Big\{\frac{1}{nc}\sum_t w_t(x)\Big[\Big(r_t - \hat r_t\Big)^2 + 2\sigma_t u_t (r_t - \hat r_t)\Big]\Big\}^2 \\ &\quad - E^2\Big[\hat\sigma^2(x) - \tilde\sigma^2(x)\Big] \\ &= E\Big\{\frac{1}{(nc)^2}\sum_{s=t=1}^{n} w_t^2(x)\Big([r_t - \hat r_t]^2 + 2\sigma_t u_t [r_t - \hat r_t]\Big)^2\Big\} \\ &\quad + E\Big\{\sum_{s \neq t} \frac{1}{(nc)^2} w_t(x) w_s(x)\Big([r_t - \hat r_t]^2 + 2\sigma_t u_t [r_t - \hat r_t]\Big) \\ &\qquad \times \Big([r_s - \hat r_s]^2 + 2\sigma_s u_s [r_s - \hat r_s]\Big)\Big\} - E^2[\hat\sigma^2(x) - \tilde\sigma^2(x)] \\ &:= I + II - III, \end{aligned}$$

First note that, by (2.3.5), we obtain

$$(nb)^{4-4H} III = C_{bias}^2(x, H).$$

Next, consider the term I. For this, we consider the cross-product terms needed in the calculations and use (2.8.1)-(2.8.4), (2.2.4) frequently. Also, recall (2.2.3) and let

$$C_{nt} := \frac{1}{nb}\sum_{j=1}^{n} k_j(t/n)r_j - r_t, \qquad Z_{nt} := \frac{1}{nb}\sum_{j=1}^{n} k_j(t/n)\sigma_j u_j, \quad \forall t.$$

Then $\hat{r}_t - r_t = C_{nt} + Z_{nt}$ and

$$Eu_t^2[r_t - \hat{r}_t]^2 = C_{nt}^2 Eu_t^2 - 2C_{nt}Eu_t^2 Z_{nt} + Eu_t^2 Z_{nt}^2.$$

But, by (2.2.4), $\max_{1\le t\le n}|C_{nt}| = O(b + \frac{1}{n})$. Moreover,

$$\begin{aligned} Eu_t^2 Z_{nt}^2 &= \frac{1}{n^2b^2}\sum_{i,j} k_i(\frac{t}{n})k_j(\frac{t}{n})\sigma_i\sigma_j Eu_iu_ju_t^2 \\ &= \frac{1}{n^2b^2}\sum_{i} k_i^2(\frac{t}{n})\sigma_i^2[Eu_i^2u_t^2 - 1] + \frac{1}{n^2b^2}\sum_{i} k_i^2(\frac{t}{n})\sigma_i^2 \\ &\quad + \frac{1}{n^2b^2}\sum_{i\neq j} k_i(\frac{t}{n})k_j(\frac{t}{n})\sigma_i\sigma_j Eu_iu_ju_t^2. \end{aligned}$$

By Lemma 2.3.1, the first term on the r.h.s. of the above expression is seen to be of the order $o(1/nb)$, while by expanding σ_i^2 using Assumption 2.4, one can verify that the second term is approximately equal to $\sigma_t^2\int K^2(v)dv/nb$, both holding for all $1 \le t \le n$. Using (2.8.2), for all $1 \le t \le n$, the third term is approximately equal to

$$\begin{aligned} &\frac{1}{n^2b^2}\sum_{i\neq j} k_i(\frac{t}{n})k_j(\frac{t}{n})\sigma_i\sigma_j[\gamma_{j-i} + 2\gamma_{t-i}\gamma_{t-j}] \\ &= \frac{\sigma_t^2 G\rho(H)}{(nb)^{2-2H}}(1 + o(1)) + \frac{2D^2\sigma_t^2}{(nb)^{4-4H}}\Big(\int K(u)u^{-2(1-H)}du\Big)^2(1 + o(1)), \\ &\qquad\qquad \frac{3}{4} < H < 1; \\ &= \frac{\sigma_t^2 G\rho(H)}{(nb)^{2-2H}} + o(\frac{1}{nb}), \qquad \frac{1}{2} < H < \frac{3}{4}, \end{aligned}$$

Note also that $EZ_{nt}^2 = Var(\hat{r}_t)$. Thus, by the Cauchy-Schwarz inequality and by Lemma 2.2.1 $|E(u_t^2 Z_{nt})| \leq Cn^{2H-2}$. Therefore, for all $1 \leq t \leq n$,

$$\begin{aligned} Eu_t^2[r_t - \hat{r}_t]^2 &= \frac{\sigma_t^2 \int K^2(v)dv}{nb} + \frac{\sigma_t^2 G\rho(H)}{(nb)^{2-2H}} + o(\frac{1}{nb}) + o(\frac{1}{(nb)^{2-2H}}) \\ &\quad + O(b + \frac{1}{n}). \end{aligned}$$

Next, we need to evaluate EZ_{nt}^4. For this purpose, let

$$K_{\omega vpq} := K(\omega)K(v)K(p)K(q), \quad d\omega vpq := d\omega dv dp dq$$

, and $\int_j$ denote the product integral j-times, $j = 2, 3, 4$, and $D = G\theta(H)$. Then,

$$\begin{aligned} EZ_{nt}^4 &= \frac{1}{n^4 b^4} \sum_{i,j,m,l} k_i(\frac{t}{n})k_j(\frac{t}{n})k_m(\frac{t}{n})k_l(\frac{t}{n})\sigma_i\sigma_j\sigma_m\sigma_l Eu_iu_ju_mu_l \\ &= O(\frac{1}{n^3b^3}) + \frac{1}{n^4b^4}\Big\{4 \sum_{i=j=m\neq l} k_i^3(\frac{t}{n})k_l(\frac{t}{n})\sigma_i^3\sigma_l Eu_i^3 u_l \\ &\quad +6 \sum_{i=j\neq k=l} k_i^3(\frac{t}{n})k_m(\frac{t}{n})\sigma_i^2\sigma_m^2[(Eu_i^2u_m^2 - 1) + 1] \\ &\quad +4 \sum_{i=j\neq m\neq l} k_i^2(\frac{t}{n})k_m(\frac{t}{n})k_l(\frac{t}{n})\sigma_i^2\sigma_m\sigma_l Eu_i^2 u_m u_l \\ &\quad + \sum_{i\neq j\neq m\neq l} k_i(\frac{t}{n})k_j(\frac{t}{n})k_m(\frac{t}{n})k_l(\frac{t}{n})\sigma_i\sigma_j\sigma_m\sigma_l Eu_iu_ju_mu_l\Big\} \\ &= O(\frac{1}{n^3b^3}) + \frac{4\sigma_t^4(nb)^{-2(1-H)}}{n^2b^2}\int_2 K^3(\omega)K(v)|\omega - v|^{-2(1-H)}d\omega dv \\ &\quad + \frac{6D^2\sigma_t^4(nb)^{4H-4}}{n^2b^2}\int_2 K^2(\omega)K^2(v)|\omega - v|^{-2(1-H)}d\omega dv \\ &\quad + \frac{6\sigma_t^4}{n^2b^2}\int_2 K^2(\omega)K^2(v)d\omega dv \\ &\quad + \frac{4\sigma_t^4}{nb}\int_3 K^2(\omega)K(v)K(q)\Big[D(nb)^{-2(1-H)}|v - q|^{-2(1-H)} \\ &\quad +2D^2(nb)^{4H-4}|\omega - q|^{-2(1-H)}|u - v|^{-2(1-H)}\Big]d\omega dv dq \\ &\quad + \frac{\sigma_t^4 D^2}{(nb)^{4-4H}}\int_4 K_{\omega vpq}\Big[(|\omega - v|\cdot|q - p|)^{-2(1-H)} \\ &\quad +(|p - \omega|\cdot|v - q|)^{-2(1-H)} + (|q - \omega|\cdot|v - p|)^{-2(1-H)}\Big]d\omega vpq. \end{aligned}$$

Therefore, using the invariance of the product $K_{\omega v p q}$ under permutation, we obtain

(2.8.5)
$$
\begin{aligned}
& E\Big(r_t - \hat{r}_t\Big)^4 \\
= \;& \frac{3\sigma_t^4 D^2}{(nb)^{4-4H}} \Big(\int_2 K(\omega)K(v)|\omega - v|^{-2(1-H)} d\omega dv \Big)^2 \\
& + O(b + \frac{1}{n}) + o(\frac{1}{nb} + \frac{1}{(nb)^{4-4H}}).
\end{aligned}
$$

Similarly, we can obtain that, for all t,

$$
\begin{aligned}
& Eu_t\Big(r_t - \hat{r}_t\Big)^3 \\
= \;& \frac{-3D^2\sigma_t^3}{(nb)^{4-4H}} \int K(\omega)|\omega|^{-2(1-H)} d\omega \int_2 K(v)K(p)|v-p|^{-2(1-H)} dv dp \\
& + o(\frac{1}{(nb)^{4-4H}}).
\end{aligned}
$$

Hence, for all $x \in (0,1)$, we obtain that

$$
\begin{aligned}
I \;&=\; \frac{1}{n^2c^2} \sum_t w_t^2(x) E\Big\{ (r_t - \hat{r}_t)^4 + 4\sigma_t^2 u_t^2 (r_t - \hat{r}_t)^2 + 4\sigma_t u_t (\sigma_t^2 u_t)^3 \Big\} \\
&=\; O(\frac{1}{(nc)(nb)^{2-2H}} + \frac{1}{(nb)(nc)} + \frac{b}{nc}).
\end{aligned}
$$

Now consider the second term II.

$$
\begin{aligned}
II = \;& \frac{1}{n^2c^2} \sum_{s \neq t} w_t(x) w_s(x) \\
& \times \Big\{ \frac{1}{n^4b^4} \sum_{i,j,m,l} k_i(\frac{t}{n}) k_j(\frac{t}{n}) k_m(\frac{s}{n}) k_l(\frac{s}{n}) \sigma_i \sigma_j \sigma_m \sigma_l E u_i u_j u_m u_l \\
& -2\sigma_t \frac{1}{n^3b^3} \sum_{i,m,l} k_i(\frac{t}{n}) k_m(\frac{s}{n}) k_l(\frac{s}{n}) \sigma_i \sigma_m \sigma_l E u_i u_m u_l u_t \\
& -2\sigma_s \frac{1}{n^3b^3} \sum_{i,j,m} k_i(\frac{t}{n}) k_j(\frac{t}{n}) k_m(\frac{s}{n}) \sigma_i \sigma_j \sigma_m E u_i u_j u_m u_t \\
& +4\sigma_t \sigma_s \frac{1}{n^2b^2} \sum_{i,m} k_i(\frac{t}{n}) k_m(\frac{s}{n}) \sigma_i \sigma_m E u_i u_m u_t u_s \Big\} \\
= \;& II_1 - II_2 - II_3 + II_4.
\end{aligned}
$$

Consider II_1. Decompose the second sum in II_1 into five parts: $i = j = k = l$, $i = j = k \neq l$, $i = j \neq k = l$, $i = j \neq k \neq l$, and $i \neq j \neq k \neq l$. Then the corresponding first four parts of II_1 are readily seen to be of the order $O((nb)^{-3})$, $O(\frac{1}{(nb)^{4-2H}})$, $O(\frac{1}{n^2b^2})$, and $O(\frac{1}{(nb)^{3-2H}})$, respectively. To analyze the last part, let $\ell(s,t,p,\omega) := |s-t+nb(p-\omega)|^{-2(1-H)}$, for any positive integers s,t,p,ω. This part is then approximately equal to

$$\frac{1}{n^2c^2}\sum_{t\neq s} w_t(x)w_s(x)\sigma_t^2\sigma_s^2 \int_4 K_{\omega vpq}\Big[D^2(nb)^{4H-4}(|\omega-v|\cdot|p-q|)^{-2(1-H)}$$
$$+\ell(s,t,p,\omega)\,\ell(s,t,q,v)+\ell(s,t,q,\omega)\,\ell(s,t,p,v)\Big]d\omega vpq$$
$$\sim \frac{D^2}{c^2}\int_1^n\int_1^n w_s(x)w_t(x)\sigma_t^2\sigma_s^2$$
$$\Big\{\int_4 K_{\omega vpq}\Big[(nb)^{4H-4}(|\omega-v|\cdot|p-q|)^{-2(1-H)}$$
$$+\ell(s,t,p,\omega)\,\ell(s,t,q,v)+\ell(s,t,q,\omega)\,\ell(s,t,p,v)\Big]d\omega vpq\Big\}dsdt$$
$$\sim \sigma^4(x)D^2\Big\{\frac{1}{(nb)^{4-4H}}\Big[\int_2 K(p)K(q)|p-q|^{-2(1-H)}dpdq\Big]^2$$
$$+2\int_2 W(z_1)W(z_2)\Big[\int_2 K(p)K(q)|nc(z_1-z_2)$$
$$+nb(p-q)|^{-2(1-H)}dpdq\Big]^2dz_1dz_2\Big\}$$
$$\sim \frac{3\sigma^4(x)D^2}{(nb)^{4-4H}}\Big[\int_2 K(p)K(q)|p-q|^{-2(1-H)}dpdq\Big]^2.$$

The last step follows from the Assumption 2.7: $c/b \to 0$.

In II_2, the part for which $i = j = k$ is $O(\frac{1}{(nc)^{4-2H}})$; the part for which $i = k \neq l$ is $O(\frac{1}{(nc)^{3-2H}})$; the remainder is when $i \neq j \neq l$, this part is approximately equal to

$$\frac{1}{n^2c^2}\sum_{t\neq s} w_s(x)w_t(x)[2\sigma_s^2\sigma_t^2]\int_3 K(u)K(v)K(p)D^2$$
$$\times\Big[(nb)^{4H-4}(|\omega|\cdot|v-p|)^{-2(1-H)}$$

$$
\begin{aligned}
&+\ell(s,t,p,0)[\ell(s,t,p,\omega)+\ell(s,t,v,\omega)]\Big]d\omega dv dp \\
\sim\ & \sigma^4(x)D^2\Big\{\frac{2}{(nb)^{4-4H}}\int K(v)|v|^{-2(1-H)}dv \\
&\times\int K(p)K(q)|p-q|^{-2(1-H)}dpdq \\
&+4\int\int W(z_1)W(z_2)[\int K(v)|nc(z_1-z_2)+nbv|^{-2(1-H)}dv \\
&\times\int K(p)K(q)|nc(z_1-z_2)+nb(p-q)|^{-2(1-H)}dpdq] \\
&\times dz_1dz_2\Big\} \\
\sim\ & \frac{6\sigma^4(x)D^2}{(nb)^{4-4H}}\int K(v)|v|^{-2(1-H)}dv\int K(p)K(q)|p-q|^{-2(1-H)}dpdq.
\end{aligned}
$$

The last step above again uses the Assumption 2.7. The approximation to the term II_3 is the same as that for the term II_2.

For the term II_4, we have the following: the part for which $i=k$ is $O(\frac{1}{(nc)^{3-2H}})$; the part for which $i\neq k$ is approximately equal to

$$
\begin{aligned}
&\frac{1}{n^2c^2}\sum_{t\neq s}w_t(x)w_s(x)4\sigma_t^2\sigma_s^2\int_2 K(\omega)K(v)D^2\Big[(|t-s|^{-2(1-H)}\ell(s,t,\omega,v) \\
&+\ell(s,t,\omega,0)\ell(s,t,v,0)+(|nbv|\cdot|nbu|)^{-2(1-H)}\Big]d\omega dv \\
\sim\ & 4\sigma^4(x)D^2\Big\{\int_2 W(z_1)W(z_2)|nc(z_1-z_2)|^{-2(1-H)} \\
&\times\Big[\int_2 K(\omega)K(v)|nc(z_1-z_2)+nb(v-\omega)|^{-2(1-H)}d\omega dv\Big]dz_1dz_2 \\
&+\int_2 W(z_1)W(z_2)dz_1dz_2 \\
&\times\int_2 K(\omega)K(v)\Big(|nc(z_1-z_2)nb\omega||nc(z_1-z_2)+nbv|\Big)^{-2+2H}d\omega dv \\
&+\frac{1}{(nb)^{4-4H}}\Big(\int K(v)|v|^{-2(1-H)}dv\Big)^2\Big\} \\
\sim\ & \frac{4\sigma^4(x)D^2}{(nc)^{2-2H}(nb)^{2(1-H)}}\int_2 W(z_1)W(z_2)|(z_1-z_2)|^{-2+2H} \\
&\times\Big[\int_2 K(\omega)K(v)|v-\omega|^{-2+2H}d\omega dv\Big]dz_1dz_2
\end{aligned}
$$

$$+\frac{4D^2}{(nb)^{4-4H}}\Big(\int K(\omega)|nb\omega|^{-2(1-H)}d\omega\Big)^2$$
$$+\frac{4\sigma^4(x)D^2}{(nb)^{4-4H}}\Big(\int K(v)|v|^{-2(1-H)}dv\Big)^2.$$

Consequently, we obtain that for each $x \in (0,1)$,

$$\mathrm{Var}(\hat{\sigma}^2(x) - \tilde{\sigma}^2(x)) = o((nc)^{-4(1-H)} + (nc)^{-1}),$$

which in turn completes the proof of the Lemma 2.3.3. □

The following two lemmas are needed for proving some uniform convergence results about $\hat{\sigma}_t^2$ and $\hat{r}_t$ given in the Lemmas 2.8.3 and 2.8.4 below.

Lemma 2.8.1 *Let u_t be defined in (2.1.2). Suppose Assumption 2.10 holds, then*

$$\max_{1\le t\le N} |u_t| = O_P(\log N), \text{ as } N \to \infty.$$

Proof. The proof uses the Bernstein inequality given in Lemma 2.5.1 and the truncation technique. Rewrite the moving average process as $u_t = \sum_{j=-\infty}^{\infty} \alpha_{t-j}\varepsilon_j$ by defining $\alpha_j = 0$ if $j < 0$. Then for some integer $L > 0$,

$$u_t = T_{t,L} + T_{t,L}^- + T_{t,L}^+, \qquad T_{t,L} = \sum_{j=-L+1}^{L} \alpha_{t-j}\varepsilon_j,$$
$$T_{t,L}^- = \sum_{j=-\infty}^{-L} \alpha_{t-j}\varepsilon_j, \quad T_{t,L}^+ = \sum_{j=L+1}^{\infty} \alpha_{t-j}\varepsilon_j,$$

Note that $T_{t,L}^+ = 0$, if $L > N$. To use Bernstein's inequality, we calculate the variance of $T_{t,L}$:

$$\mathrm{Var}(T_{t,L}) = \sum_{j=-L+1}^{L} \alpha_{t-j}^2 \le \sum_{j=-\infty}^{\infty} \alpha_j^2 < \infty, \text{ for any } L,\ t,$$

Hence by Bernstein's inequality, we obtain that there exists a $\infty > C_3 > 0$, such that

$$\begin{aligned} P(|T_{t,L}| > C_3 \log N) &\le 2\exp\Big\{\frac{-C_3^2 \log^2 N}{C + C\log N}\Big\} \le C\frac{1}{N^2}, \\ P(\max_{1\le t\le N} |T_{t,L}| > C_3 \log N) &\le \sum_{1\le t\le N} P(|T_{t,L}| > C_3 \log N) \to 0, \end{aligned}$$

for any positive integer L. Next, because $E(T_{t,L}^-)^2 = \sum_{j=-\infty}^{-L} \alpha_j^2 \le CL^{2H-2}$, by choosing $L = N^{1/(2-2H)}$, we obtain that

$$P\Big(\max_{1\le t\le N} |T_{t,L}^-| > C\log N\Big) \le CN\frac{L^{2H-2}}{\log^2 N} \to 0.$$

The lemma follows. □

Lemma 2.8.2 *Let u_t be defined in (2.1.2). Suppose Assumption 2.10 holds, and nd, $nh \to \infty$. Then for any finite positive integer N_0,*

$$\sup_{1\le k\le N_0 nd} |\frac{1}{N_0 nd}\sum_{l=1}^{k} u_l| \to_P 0, \qquad \sup_{1\le k\le N_0 nh} |\frac{1}{N_0 nh}\sum_{l=1}^{k}(u_l^2-1)| \to_P 0.$$

Proof. By the Ergodic Theorem, we have

$$\frac{1}{k}\sum_{l=1}^{k} u_l \to_{a.s.} 0 \quad \text{as } k\to\infty,$$

that is , $\forall\, \delta > 0$,

$$\begin{aligned} 0 &= P\Big(|\frac{1}{k}\sum_{l=1}^{k} u_l| \ge \delta \text{ i.o.}\Big) = P\Big(\bigcap_{N=1}^{\infty}\bigcup_{k=N}^{\infty}\big(|\frac{1}{k}\sum_{l=1}^{k} u_l| \ge \delta\big)\Big) \\ &= = \lim_{N\to\infty} P\Big(\bigcup_{k=N}^{\infty}\big(|\frac{1}{k}\sum_{l=1}^{k} u_l| \ge \delta\big)\Big). \end{aligned}$$

By choosing $N = \log(N_0 nd)$, we obtain that

$$\sup_{k\ge N} |\frac{1}{k}\sum_{l=1}^{k} u_l| \to_P 0, \quad \text{which implies} \quad \sup_{N\le k\le N_0 nb} |\frac{1}{nd}\sum_{l=1}^{k} u_l| \to_P 0,$$

as $n\to\infty$. On the other hand, by Lemma 2.8.1,

$$\sup_{1\le k\le N} |\frac{1}{nd}\sum_{l=1}^{k} u_l| \le \frac{N}{nd}\sup_{1\le k\le N} |u_l| = O_P(\frac{N\log(N)}{nd}) = o_P(1), \text{ as } n\to\infty.$$

Hence the first claim in the lemma follows.

The second claim is obtained by a similar argument and the fact that $\sup_{1\le k\le N} \sum_{l=1}^{k} |u_l^2 - 1| = O_P(N + N\log^2 N)$. □

Let d and h be the bandwidth for estimating $r(x)$ and $\sigma(x)$ respectively, and $d, h \to 0$, $nd, nh \to \infty$ in the following. Let $A_0 = \{k_0 + 1, \cdots, k_0 + M\}$, where $k_0/n \to x_0$, $(k_0 + M)/n \to x_1 \geq x_0$, with $x_0, x_1 \in (0,1)$, and $A_1 := \{j \in \mathbb{Z}; k_0 + 1 - nh \leq j \leq k_0 + M + nh\}$.

Lemma 2.8.3 *Suppose Assumptions 2.1-3 and 2.6, 2.10 hold. Then, under (2.1.1), (2.1.2),*

$$\sup_{t \in A_0} |\hat{\sigma}_t^2 - \sigma_t^2| = o_P(1).$$

Proof. In view of the decomposition (2.3.1) of $\hat{\sigma}^2(x)$ with b and c replaced by d and h, using Lemma 2.5.2, the first term in (2.3.1) is bounded above by

$$\sup_{i \in A_1} [\hat{r}_i - r_i]^2 \frac{C}{nh} \sum_{i=1}^{n} W(\frac{i-t}{nh}) \to_P 0.$$

Consider the third term in (2.3.1). Let $S_k(t) = \sum_{l=t-nh}^{t-nh+k} (u_l^2 - 1)$ for $t \in A_0$ and $1 \leq k \leq 2nh$. Since

$$\begin{aligned}
\frac{1}{nh}|S_k(t) - \sum_{l=k_0-2nh}^{k_0+2nh} (u_l^2 - 1)| \;&\leq\; \frac{1}{nh}|\sum_{l=k_0-2nh}^{t-nh-1} (u_l^2 - 1)| \\
&+ \frac{1}{nh}|\sum_{l=t-nh+k+1}^{k_0+2nh} (u_l^2 - 1)|,
\end{aligned}$$

by Lemma 2.8.2, we obtain that $\sup_{t \in A_0} \sup_{1 \leq k \leq 2nh} |\frac{S_k(t)}{nh}| \to_P 0$, which implies that the third term in (2.3.1) is $o_P(1)$ uniformly in $t \in A_0$, since

$$\begin{aligned}
\frac{1}{nh}\sum_{l=1}^{n} \quad W(\frac{t-l}{nh})\sigma_l(u_l^2 - 1) &\equiv \frac{S_{2nh}(t)}{nh} W(1) \\
&+ \frac{1}{nh} \sum_{k=t-nh}^{t+nh-1} S_k(t) \Big(W(\frac{t-k}{nh})\sigma_k - W(\frac{t-k-1}{nh})\sigma_{k+1} \Big)
\end{aligned}$$

by (2.4.3). This also shows that the second term in (2.3.1) is $o_P(1)$ uniformly in $t \in A_0$. □

The following "Lipschitz" properties are needed for estimating H. Let $\tau_n = (1/(nd) \vee 1/(nh))$.

Lemma 2.8.4 *Suppose Assumptions 2.1-4, 2.6, 2.10 hold. Then*

$$\sup_{j\in A_1} |\hat{r}_{j+1} - \hat{r}_j| = O_P(\frac{\log n}{nd}). \quad \sup_{j\in A_0} |\hat{\sigma}^2_{j+1} - \hat{\sigma}^2_j| = O_P(\tau_n \log^2 n).$$

Proof. By the decomposition of $\hat{r}$ in (2.2.3), we consider

$$\begin{aligned}
&\sup_{j\in A_1} \Big|\frac{1}{nd}\sum_{l=1}^{n} K(\frac{j-l}{nd})\Big[\sigma_{l+1}u_{l+1} - \sigma_l u_l\Big]\Big| \\
&\quad \le \sup_{j\in A_1} \Big|\frac{1}{nd}\sum_{l=1}^{n} K(\frac{j-l}{nd})[\sigma_{l+1} - \sigma_l]u_{l+1}\Big| \\
&\quad + \sup_{j\in A_1} \Big|\frac{1}{nd}\sum_{l=1}^{n} K(\frac{j-l}{nd})\sigma_l[u_{l+1} - u_l]\Big|.
\end{aligned}$$

By the Ergodic Theorem and the Lipschitz property of σ, the first term in this bound is $O_P(1/(nd))$. To deal with the second term, let $j_0 = j - nh, j^1 = j + nh$. Note that $\sum_{l=j_0}^{k}(u_{l+1} - u_l) = u_{k+1} - u_{j_0}$. The second term is bounded above by

$$\begin{aligned}
(2.8.6) \qquad &\sup_{j\in A_1} \frac{1}{nd}\Big|(u_{j^1} - u_{j_0})K(-1)\sigma_{j^1}\Big| + \sup_{j\in A_1} \frac{1}{nd}\sum_{l=j_0}^{j^1-1} |u_{l+1} - u_{j_0}| \\
&\times\Big|K(\frac{j-l}{nd})\sigma_l - K(\frac{j-l-1}{nd})\sigma_{l+1}\Big| \\
&= O_P\Big(\frac{C\log n}{nd} + \frac{C}{(nd)^2}\sum_{k_0-3nd}^{k_0+3nd} \log n\Big) \to_p 0,
\end{aligned}$$

since $\max_{1\le j\le n} |u_j| = O_P(\log n)$ by Lemma 2.8.1. Hence the first claim follows.

Now consider the second claim. By definition and the first claim,

$$\begin{aligned}
|\hat{\sigma}^2_{j+1} - \hat{\sigma}^2_j| = \Big|\frac{1}{nh}\sum_{l=1}^{n-1} W(\frac{j-l}{nh})\Big\{&\Big[(r_{l+1} - \hat{r}_{l+1})^2 - (r_l - \hat{r}_l)^2\Big] \\
&+\Big[\sigma^2_{l+1}u^2_{l+1} - \sigma^2_l u^2_l\Big]
\end{aligned}$$

$$+2\Big[(r_{l+1}-\hat r_{l+1})-(r_l-\hat r_l)\Big]\times\Big[\sigma_{l+1}u_{l+1}-\sigma_l u_l\Big]\Big\}\Big|$$

$$= O_P(\frac{\log n}{nd}) + C\Big|\frac{1}{nh}\sum_{l=1}^{n-1} W(\frac{j-l}{nh})\Big[\sigma_{l+1}^2-\sigma_l^2\Big]|u_{l+1}^2|$$

$$+\Big|\frac{1}{nh}\sum_{l=1}^{n-1} W(\frac{j-l}{nh})\sigma_l^2\Big(u_{l+1}^2-u_l^2\Big)\Big| := I + II + III.$$

By the Lipschitz property of σ and the Ergodic Theroem,

$$\sup_{j\in A_0} II \le C(nnh)^{-1}\sum_{l=1}^{n-1}|u_{l+1}^2| = O_P((nh)^{-1}).$$

Similar to (2.8.6), the term III is bounded above by

$$\frac{1}{nh}\Big|(u_{j1}^2-u_{j0}^2)W(-1)\sigma_{j1}^2\Big|$$

$$+\frac{1}{nh}\sum_{l=j_0}^{j^1-1}|u_{l+1}^2-u_{j0}^2|\Big|W(\frac{j-l}{nh})\sigma_l^2 - W(\frac{j-l-1}{nh})\sigma_{l+1}^2\Big|.$$

the above bound is $o_P(1)$ uniformly in j by observing that

$$\max_{1\le j\le n} u_j^2 \le (\max_{1\le j\le n}|u_j|)^2 = O_P(\log^2 n)$$

by Lemma 2.8.1, thus proves the second claim. □

Chapter 3

Asymptotic inference for some regression models under heteroscedasticity and long memory design and errors

3.1 Introduction

It has been of great interest for statisticians to analyze the statistical models with long memory errors, from simple linear to nonparametric regression models, cf., Csörgö and Mielniczuk (2000), Dahlhaus (1995), Ho and Hsing (1995), Koul (1992), Koul and Mukherjee (1993), Robinson (1997) and Yajima (1988, 1991) among others. The asymptotic distributional properties of numerous well known estimators of the underlying parameters in these models are different from those when independent or weakly dependent errors are assumed.

Baillie and Bollerslev (1994, 2000) , Cheung (1993) and Maynard and Phillips (2001) noted that some times when regressing spot exchange rate returns on the lagged forward premium, both the error and the covariate processes may have long memory. We also notice in section 6 below that some monthly currency exchange rate data exhibit long memory. It is thus of interest to analyze statistical behavior of some inference procedures for the underlying parameters in regression models with long memory errors and long memory designs. To begin with we focus on a simple linear regression model with nonparametric heteroscedastic errors.

Consider the model where one observes a strictly stationary bivariate process (X_t, Y_t), $t \in \mathbb{Z} := \{0, \pm 1, \cdots\}$, both having finite and positive variances and obeying the model

$$(3.1.1)\quad Y_t = \beta_0 + \beta_1 X_t + \sigma(X_t) u_t, \quad \text{for some } (\beta_0, \beta_1) \in \mathbb{R}^2,$$

$$(3.1.2)\quad u_t := \sum_{j=0}^{\infty} b_j \varepsilon_{t-j}, \quad b_j \sim C j^{-(3/2-H)}, \text{as } j \to \infty, \text{ for some } \tfrac{1}{2} < H < 1.$$

Here, ε_t are standardized i.i.d. innovations, independent of the X_t-process. The sequence b_j is also assumed to be non-increasing in j and C is a constant such that $\sum_{j=0}^{\infty} b_j^2 = 1$. Throughout this chapter, for any two sequences, $a_n \sim b_n$ means that $a_n/b_n \to 1$, as $n \to \infty$. Note that under this set up, $\sigma^2(x) = \text{Var}(Y|X = x)$, $x \in \mathbb{R}$, where (X, Y) denotes a copy of (X_0, Y_0). We further assume that $\{X_t\}$ is a Gaussian process with zero mean, standard deviation one, and an auto-covariance function $\gamma_X(k)$ that is non-increasing in k and satisfies

$$(3.1.3)\qquad \gamma_X(k) \sim G_X \theta(h) k^{-2(1-h)}, \quad \text{as } k \to \infty, \text{ for some } 1/2 < h < 1,$$

where $\theta(h) = 2\Gamma(2-2h)\cos(\pi(1-h))$ and $G_X > 0$ is a constant.

For any stationary second order process ξ_t, $t \in \mathbb{Z}$, let f_ξ and γ_ξ denote its spectral density and auto-covariance functions, respectively. Let $\alpha(\lambda) = \sum_{j=0}^{\infty} b_j e^{ij\lambda}$. From (3.1.2) and Corollary 4.10.2 of Bingham, Goldie and Teugels (1987), we observe that $f_u(\lambda) = |\alpha(\lambda)|^2/(2\pi)$ and that for some positive constant G_u,

$$(3.1.4)\quad f_u(\lambda) \sim G_u \lambda^{1-2H}, \ \lambda \to 0+; \qquad \gamma_u(k) \sim G_u \theta(H) k^{-2(1-H)}, \ k \to \infty,$$
$$f_X(\lambda) \sim G_X \lambda^{1-2h}, \quad \lambda \to 0+.$$

Several authors have discussed regression models with long memory (LM) errors when $\sigma(x) \equiv c$, a constant. The asymptotic distributions of the least squares estimator (LSE) and M- and R- estimators in non-random design linear regression models with LM errors are established in Giraitis, Koul and Surgailis (1996), Koul and Mukherjee (1993), and Yajima (1988, 1991). The latter two papers observed that a large class of M- and R- estimators are asymptotically equivalent to the LSE in the first order. Similar results were obtained for nonlinear regression models with LM subordinate Gaussian errors in Koul (1996). Dahlhaus (1995) discussed the asymptotic efficiency of the generalized LSE in linear regression models with certain polynomial type designs and LM Gaussian errors. Robinson and Hidalgo (1997) discussed the linear regression model with LM moving average (LMMA) processes in both design and errors with the error spectral density known up to an unknown Euclidean parameter vector. They showed that a class of generalized LSE's are $n^{1/2}$-consistent for $\hat{\beta}$ under some conditions. In addition, for the same model, Hidalgo and Robinson (2002) removed the assumption of the error spectral density being parametric and used semi-parametric methods to obtain similar results. Ho and Hsu (2005) obtained asymptotic normality of a class of generalized LSE in polynomial trend regression models with errors being subordinated to a LMMA process. All these papers, however, are dealing with homoscedastic errors only.

Here we analyze the asymptotic distribution of the LSE $(\hat{\beta}_0, \hat{\beta}_1)'$ of $(\beta_0, \beta_1)'$ in the model (3.1.1) and (3.1.2) with heteroscedastic errors. In addition, we analyze the asymptotic distributional properties of the kernel type nonparametric estimators of $\sigma^2(x)$ and the $\log(n)$-consistency of the local Whittle estimator H based on the least square residuals, assuming the above model holds. We also provide an asymptotically distribution free test for testing the lack-of-fit of a linear heteroscedastic regression model under the assumed long memory set up. This chapter also contains a simulation study and an application to some foreign currency exchange rate data.

This chapter is organized as follows. Section 3.2 discusses the asymptotic distribution of $(\hat\beta_0, \hat\beta_1)'$. It turns out that $n^{1-H}(\hat\beta_0 - \beta_0) \to_d N(0, a^2)$ for all $1/2 < H, h < 1$, while if $EX\sigma(X) = 0$, then the first order asymptotic distribution of $\hat\beta_1 - \beta_1$ is degenerate at 0, where a is given in Lemma 3.2.1 below. In this case we then consider the second order properties of $\hat\beta_1$. We obtain that in the case $H + h < 3/2$, $n^{1/2}(\hat\beta_1 - \beta_1)$ is asymptotically normal with mean zero and some positive variance. On the other hand even when $H \wedge h > 3/4$ and both u_t and X_t are Gaussian, $\hat\beta_1$ has non-normal limit distribution with normalization n^{2-H-h}.

Section 3.3 contains a discussion about the asymptotic distribution of the kernel type estimators of $\sigma^2(x)$ with long memory design. It is observed that for an appropriately chosen bandwidth sequence, when $H < (1 + h)/2$, $n^{1-h}(\hat\sigma^2(x) - \sigma^2(x))$ is asymptotically normal with mean zero and some positive variance and when $H > (1 + h)/2$, the asymptotic distribution of $n^{2-2H}(\hat\sigma^2(x) - \sigma^2(x))$ is non-normal.

As is evident from the above discussion, to carry out the inference about β_0, β_1 and $\sigma^2(x)$, we also need a $\ln(n)$-consistent estimators of H. We address this issue in a more general model $Y_t = \beta' m(X_t) + \sigma(X_t)u_t$, where β is now a $q \times 1$ parameter vector, $m(x)$ is a vector of some known q functions, and where $\sigma(X_t)$ and u_t are as before. Section 3.4 adopts the approach in Robinson (1997) to obtain a $\log(n)$-consistent estimator of H based on the pseudo residuals $Y_t - \hat\beta' m(X_t)$ in this model, where $\hat\beta$ is the LSE. This is unlike the case of nonparametric heteroscedastic regression model with non-random uniform design on $[0, 1]$ and LMMA errors, where it is necessary to base estimators of H on the standardized residuals that need a uniformly consistent estimator of $\sigma(x)$, see Section 2.2.6 in Chapter 2.

Section 3.5 constructs a test of lack-of-fit of a parametric regression model. Let $\mu(x) := E(Y|X = x)$ and consider the problem of testing $H_0 : \mu(x) = \beta' m(x)$, for some $\beta \in \mathbb{R}^q$ and all $x \in \mathbb{R}$, against the alternatives that H_0 is not true, where $m(x)$ is a $q \times 1$ vector of known functions such that $A := Em(X)m(X)'$ is positive definite. In the presence of long memory in design and/or errors and when $\sigma(x) \equiv c$, Koul, Baillie and Sugailis (2004) (KBS) proposed a test for H_0 based on the marked empirical process

$$\hat{\mathcal{V}}_n(x) = \sum_{t=1}^{n} \Big(Y_t - \hat{\beta}' m(X_t)\Big) I(X_t \leq x), \quad x \in \bar{\mathbb{R}} := [-\infty, \infty],$$

where $\hat{\beta}$ is the least squares estimator of β under H_0. They showed that $n^{-H}\hat{\mathcal{V}}_n$ converges weakly to a process degenerated in x, thus making the implementation of the tests based on this process relatively easier, compared to when the errors and/or the design processes are i.i.d.

Under the current set up and some conditions, Theorem 3.5.1 below proves that under H_0, $n^{-H}\hat{\mathcal{V}}_n(x)$ converges weakly to $J_\sigma(x)\tau(H)Z$ in $D(\bar{\mathbb{R}})$ and uniform metric, where Z is a $N(0,1)$ r.v., $\tau^2(H) := G\theta(H)(2H^2 - H)^{-1}$, and

$$\begin{aligned} (3.1.5) \qquad J_\sigma(x) &:= F_\sigma(x) - \kappa_\sigma' A^{-1}\alpha(x), \quad F_\sigma(x) := E(\sigma(X)I(X \leq x)), \\ \alpha(x) &:= Em(X)I(X \leq x), \quad x \in \bar{\mathbb{R}}; \quad \kappa_\sigma := E\sigma(X)m(X). \end{aligned}$$

To use this process for testing H_0, we thus need a uniformly consistent estimator of $J_\sigma(x)$ and a consistent estimator of $\tau(H)$. Section 3.5 constructs uniformly consistent estimator of J_σ, under H_0, based on the leave-one-observation-out or a cross validation estimator of $\sigma(x)$. The regular kernel type estimator is not useful here because the behavior of $\hat{\sigma}^2(X_t)$ is not stable. The estimators of G and H constructed in section 4 are used to provide a consistent estimator of $\tau(H)$ under H_0.

Section 3.6 includes a finite sample simulation and an application to some monthly currency exchange rate data that exhibits long memory. Section 3.7 is the Appendix consisting of some needed lemmas and proofs.

In the sequel, $\to_d$ stands for the convergence in distribution of a sequence of r.v.'s while $\Longrightarrow$ denotes the weak convergence of a sequence of stochastic processes, and $u_P(1)$ denotes a sequence of stochastic processes that tends to zero uniformly over its time domain, in probability. All limits are taken as $n \to \infty$, unless specified otherwise.

3.2 Asymptotics of the LSE

In this section, we consider the asymptotic distribution of the LSE $(\hat{\beta}_0, \hat{\beta}_1)$ in the model (3.1.1) - (3.1.3). For this, we need the following assumption.

Assumption 3.1. X_t and u_t are independent.

Let $\sigma_t := \sigma(X_t)$, $e_t := \sigma_t u_t$, $\bar{X} := \sum_{t=1}^n X_t/n$, $\bar{u} := \sum_{t=1}^n u_t/n$, $\bar{Y} := \sum_{t=1}^n Y_t/n$, $\bar{e} := \frac{1}{n}\sum_{t=1}^n e_t$, $s_X^2 := \sum_{t=1}^n (X_t - \bar{X})^2/n$. Then the least squares estimators are

$$(3.2.1) \qquad \hat{\beta}_1 := \frac{\sum_{t=1}^n (X_t - \bar{X})(Y_t - \bar{Y})}{\sum_{t=1}^n (X_t - \bar{X})^2} = \beta_1 + \frac{1}{s_X^2}\Big(\frac{1}{n}\sum_{t=1}^n X_t e_t - \bar{X}\bar{e}\Big).$$

$$(3.2.2) \qquad \hat{\beta}_0 := \bar{Y} - \bar{X}\hat{\beta}_1 = \beta_0 + \bar{e} - \frac{1}{s_X^2}\Big(\bar{X}\frac{1}{n}\sum_{t=1}^n X_t e_t - (\bar{X})^2\bar{e}\Big).$$

Note that $s_X^2 \to_{a.s.} \gamma_X(0)$ by the Ergodic Theorem.

To proceed further we need the following result. Let $\nu(x)$ be a function on $\mathbb{R}$ with $E\nu^2(X) < \infty$. Let $\nu_0 = E\nu(X)$. By Assumption 3.1 and (3.7.3) in the Appendix, there is a $C < \infty$ such that $E[(\nu(X_0) - \nu_0)u_0(\nu(X_t) - \nu_0)u_t] \le Ct^{-2(1-H)}t^{2h-2}$, for all sufficiently large and positive t. Hence,

$$(3.2.3) \qquad n^{-H}\sum_{i=1}^n \nu(X_i)u_i = \nu_0 n^{-H}\sum_{i=1}^n u_i + n^{-H}\sum_{i=1}^n (\nu(X_i) - \nu_0)u_i$$
$$= \nu_0 n^{-H}\sum_{i=1}^n u_i + o_P(1).$$

Next, let Z_1, Z_2, be two independent r.v.'s, Z_j having $N(0, \psi_j^2)$ distribution, $j = 1, 2$, where $\psi_1^2 = G_u\theta(H)/H(2H-1)$, $\psi_2^2 = G_X\theta(h)/h(2h-1)$. From Davydov (1970), we obtain

$$(3.2.4) \qquad n^{-H}\sum_{i=1}^{n} u_i \to_d Z_1, \qquad n^{-h}\sum_{i=1}^{n} X_i \to_d Z_2.$$

Now apply (3.2.3) and (3.2.4) to $\nu(x) \equiv \sigma(x)$ to obtain that $\bar{e} = O_p(n^{1-H})$. We also have $\bar{X} = O_p(n^{1-h})$ from (3.2.4). By (3.2.1)-(3.2.4), and Slutsky's Theorem, we thus obtain the following result where $J_0 = E\sigma(X)$ and $J = E\sigma(X)X$.

Lemma 3.2.1 *Suppose (3.1.1)-(3.1.3) and the Assumption 3.1 hold. Then,*

$$n^{1-H}\Big(\hat{\beta}_0 - \beta_0, \hat{\beta}_1 - \beta_1\Big) \to_d \gamma_X(0)^{-1}\Big(J_0, J\Big)Z_1.$$

Note that by the Cauchy-Schwarz inequality $J^2 \leq E\sigma^2(X)\,EX^2 < \infty$. The result about $\hat{\beta}_1$ is useful only if $J \neq 0$. But if σ is an even function then $J = 0$. For example, in financial econometrics, the volatility (the conditional variance function) is often assumed to have the form $a + bx^2$. In these cases the limit distribution of $n^{1-H}(\hat{\beta}_1 - \beta_1)$ is degenerate. It is thus important to investigate the higher order approximation to the distribution of a suitably standardized $\hat{\beta}_1$ under the following

Assumption 3.2. $\sigma(x)$ is an even function of $x \in \mathbb{R}$ and $\gamma_X(0) = 1$.

Let H_j, $j \geq 1$ denote the Hermite polynomials, see, e.g. Taqqu (1975). The Hermite expansion of the function $x\sigma(x)$ is

$$(3.2.5) \qquad x\sigma(x) = \sum_{j=0}^{\infty} \frac{c_j}{j!}H_j(x), \qquad c_j := E(X\sigma(X)H_j(X)),\ j \geq 0.$$

The Hermite rank of $x\sigma(x)$ is defined to be $\min\{j \geq 1; c_j \neq 0\}$. Since $c_1 = EX^2\sigma(X) \neq 0$, the Hermite rank of $x\sigma(x)$ is 1. Let $\tilde{\beta}_1 := \frac{1}{n}\sum_{t=1}^{n} X_t e_t$ and

$$\mathcal{Z}_2 \;:=\; \tilde{C}\int\int_0^1 \Big[(s-x_1)^{-(3-2H)/2}(s-x_2)^{-(3-2h)/2}\Big]$$

$$
\begin{aligned}
& \times I(x_1 < s, x_2 < s) ds\, dB_1(x_1) dB_2(x_2), \\
\tilde{C} &:= \frac{\sqrt{G_u G_X}}{\sqrt{a(H)a(h)}}, \\
a(z) &:= \int_0^\infty v^{-(3-2z)/2}(1+v)^{-(3-2z)/2} dv, \; 1/2 < z < 1,
\end{aligned}
$$

where B_1 and B_2 are two Wiener random measures. We are now ready to prove

Lemma 3.2.2 *Suppose (3.1.1)-(3.1.3) and the Assumptions 3.1-2 hold, with the innovations ε_j being standard Gaussian r.v.'s. Then, for $h \wedge H > 3/4$, $n^{2-(H+h)}\tilde{\beta}_1 \to_d c_1 \mathcal{Z}_2$, where B_1 and B_2 in $\mathcal{Z}_2$ of (3.2.6) are now two independent Wiener random measures.*

Proof. Using the above Hermite expansion, we obtain

$$
\begin{aligned}
\frac{1}{n}\sum_{t=1}^{n} X_t \sigma_t u_t &= \frac{c_1}{n}\sum_{t=1}^{n} X_t u_t + \frac{1}{n}\sum_{t=1}^{n} u_t \sum_{j=2}^{\infty} \frac{c_j}{j!} H_j(X_t) \\
&=: S_n + T_n. \qquad (3.2.6)
\end{aligned}
$$

Under Assumption 3.1, $\mathrm{Var}S_n = O(n^{-4+2H+2h})$. Because of the orthogonality of the Hermite polynomials,

$$
\begin{aligned}
\mathrm{Var} \quad (T_n) &= n^{-2}\sum_{s=1}^{n}\sum_{t=1}^{n} E u_s u_t E\Big(\sum_{j=2}^{\infty}\frac{c_j}{j!}H_j(X_s)\sum_{k=2}^{\infty}\frac{c_k}{k!}H_k(X_t)\Big) \\
&\le Cn^{-2}\sum_{s=1}^{n}\sum_{t=1}^{n} |s-t|^{2H-2}|s-t|^{4h-4} \\
&\le Cn^{-4+2H+2h-(2-2h)}\ln n = o(\mathrm{Var}S_n).
\end{aligned}
$$

An application of Theorem 6.1 in Fox and Taqqu (1987) to the leading term S_n gives the desired result.

The following lemma directly follows from (3.2.4) and the Hermite expansion.

Lemma 3.2.3 *Suppose (3.1.1)-(3.1.3) and the Assumptions 3.1-2 hold. Then,*

$$
n^{2-H-h}\bar{X}\bar{e} \to_d Z_1 Z_2. \qquad (3.2.7)
$$

Theorem 3.2.1 *Under the conditions of Lemma 3.2.2,*

$$n^{2-H-h}(\hat{\beta}_1 - \beta_1) \to_d c_1 Z_2 - Z_1 Z_2. \tag{3.2.8}$$

Moreover, the $Correl(Z_2, Z_1 Z_2)$ *equals*

$$\frac{\sqrt{2(2H+2h-3)(2H+2h-2)}}{(2H+2h-1)} \sqrt{\frac{Hh}{(2H-1)(2h-1)}}. \tag{3.2.9}$$

Proof. In view of (3.2.1), the claim (3.2.8) follows from Lemmas 3.2.2 and 3.2.3. To prove (3.2.9), proceed as follows. Let $\kappa_1 = G_X G_u \theta(H)\theta(h)$. By (3.2.3) and (3.2.6),

$$\begin{aligned}
\mathrm{Var}(\frac{1}{n}\sum_{t=1}^{n} X_t e_t) &\sim c_1^2 \mathrm{Var}(\frac{1}{n}\sum_{t=1}^{n} X_t u_t) \sim \frac{2c_1^2 \kappa_1 n^{2H+2h-4}}{(2H+2h-3)(2H+2h-2)}, \\
\mathrm{Var}(\bar{X}\bar{e}) &\sim J_0^2 \mathrm{Var}(\bar{X}\bar{u}) \sim \frac{J_0^2 \kappa_1 n^{2H+2h-4}}{(2H-1)(2h-1)Hh}.
\end{aligned}$$

Next, by the Hermite expansion of $x\sigma(x)$ and $\sigma(x)$, we obtain

$$\begin{aligned}
& E\Big(\frac{1}{n}\sum_{t=1}^{n} X_t e_t \bar{X}\bar{e}\Big) \\
\sim\ & c_1 J_0 n^{-3} \sum_{t=1}^{n}\sum_{s=1}^{n}\sum_{k=1}^{n} E(X_t X_s) E(u_t u_k) \\
\sim\ & c_1 J_0 \kappa_1 n^{-3+2H+2h-4} \sum_{t=1}^{n}\sum_{s=1}^{n}\sum_{k=1}^{n} |\frac{t}{n} - \frac{s}{n}|^{2H-2} |\frac{t}{n} - \frac{k}{n}|^{2h-2} \\
\sim\ & \frac{4\kappa_1 c_1 J_0 n^{2H+2h-4}}{(2H-1)(2h-1)(2H+2h-1)}.
\end{aligned}$$

This proves (3.2.9). □

The above theorem considered the case of u_t's being Gaussian and $3/4 < H, h < 1$. We shall now discuss the asymptotic distribution of the LSE's when u_t's form the moving average (3.1.2) and when $H + h < 3/2$. This in turn is facilitated by the following lemma where $\nu_t := \nu(X_t)$ and $U_n := n^{-1/2}\sum_{t=1}^{n} \nu_t u_t$.

Lemma 3.2.4 *Suppose (3.1.1)-(3.1.3) and the Assumptions 3.1-2 hold. In addition, suppose* ν *is a measurable function such that* $E\nu(X) = 0$, $E\nu^2(X) < \infty$,

$\max_{\{0\le x\le \ln n\}} |\nu(x)|/n^{1/2-\eta} \to 0$ *for some* $0<\eta<1/2$, *and the Hermite rank of* $\nu(X)$ *is* 1. *Then, for* $H+h<3/2$, $U_n \to_d N(0,\kappa_2)$, *where* $\kappa_2^2 = \lim EU_n^2 < \infty$.

Proof. The proof uses the truncation method similar to the one used by Robinson and Hidalgo (1997). The main idea here is to approximate U_n by a weighted partial sum of the independent innovations $\{\varepsilon_i\}$. Fix H, h such that $H+h<3/2$. Let $M = M_n > n^{(2h-1)/(2-2H)}$, and define

$$U_{n,M} := n^{-1/2} \sum_{t=1}^{n} \nu_t \sum_{j=-M}^{n} b_{t-j}\varepsilon_j, \qquad T_{n,M} := U_n - U_{n,M}.$$

Because the Hermite rank of ν_t is 1 and by the Assumption 3.1 and (3.7.3),

$$(3.2.10)\quad ET_{n,M}^2 \;=\; \frac{1}{n} \sum_{j=-\infty}^{-M-1} E(\sum_{t=1}^{n} \nu_t b_{t-j})^2 \le C n^{2h-1} M^{-2+2H} \to 0.$$

Hence it suffices to show that the weak limit of $U_{n,M}$ is a normal distribution. We prove this by showing that the conditional distribution of $U_{n,M}$, given $\mathcal{F} := \sigma$-$\{X_t, t=1,2,\cdots\}$, converges weakly to $N(0,a)$, for some nonrandom $a>0$. Let $d_{n,j} := \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \nu_t b_{t-j}$. By the Lindeberg-Feller Theorem, this is equivalent to showing that conditional on $\mathcal{F}$,

$$(3.2.11)\qquad \sum_{j=-M}^{n} d_{n,j}^2 \to_p \text{ a positive constant,}$$

$$(3.2.12)\qquad P(\max_{-M\le j\le n} |d_{n,j}| > \delta) \to 0 \text{ for all } \delta > 0.$$

Consider (3.2.11). Let $\gamma_\nu(k) = E\nu_0\nu_k$, recall $\gamma_u(k) = Eu_0u_k = \sum_{j=0}^{\infty} b_j b_{j+k}$. Let

$$s^2 := E(U_{n,M}^2|\mathcal{F}) = \frac{1}{n} \sum_{j=-M}^{n} (\sum_{t=1}^{n} \nu_t b_{t-j})^2, \qquad \Gamma_j := \frac{1}{n} \sum_{t=1}^{n-j} \nu_t \nu_{t+j}.$$

Thus proving (3.2.11) is equivalent to showing that

$$(3.2.13)\qquad s^2 \to_p \kappa_2^2.$$

Rewrite $s^2 = A + 2B$, where

$$A = \frac{1}{n} \sum_{j=-M}^{n} \sum_{t=1}^{n} \nu_t^2 b_{t-j}^2, \quad B = \frac{1}{n} \sum_{j=-M}^{n} \sum_{s=1}^{n-1} \sum_{t=s+1}^{n} \nu_t \nu_s b_{t-j} b_{s-j}.$$

But $A = A_1 + A_2$, where

$$\begin{aligned} A_1 &= \frac{1}{n} \sum_{j=-M}^{n} \sum_{t=1}^{n} (\nu_t^2 - \gamma_\nu(0)) b_{t-j}^2 \\ &= \frac{1}{n} \sum_{t=1}^{n} \left(\nu_t^2 - \gamma_\nu(0)\right)\left(\sum_{k=0}^{\infty} b_k^2 - \sum_{k=t+M+1}^{\infty} b_k^2 \right) \to_p 0, \\ A_2 &:= \gamma_\nu(0) \frac{1}{n} \sum_{j=-M}^{n} \sum_{t=1}^{n} b_{t-j}^2 \\ &= \gamma_\nu(0) \frac{1}{n} \sum_{t=1}^{n} \left(\sum_{k=0}^{\infty} b_k^2 - \sum_{k=t+M+1}^{\infty} b_k^2 \right) \to \gamma_\nu(0)\gamma_u(0). \end{aligned}$$

since $\sum_{k=M}^{\infty} b_k^2 \to 0$, $\frac{1}{n} \sum_{t=1}^{n} (\nu_t^2 - \gamma_\nu(0)) \to_{a.s.} 0$ and $\frac{1}{n} \sum_{t=1}^{n} |\nu_t^2 - \gamma_\nu(0)|$ $\to_{a.s.} C < \infty$ by the Ergodic Theorem. Hence

$$A \to_p \gamma_\nu(0)\gamma_u(0). \tag{3.2.14}$$

Also, $B = \sum_{k=1}^{n-1} \left(\frac{1}{n} \sum_{s=1}^{n-k} \nu_s \nu_{k+s} \sum_{j=0}^{s+M} b_{k+j} b_j \right) =: B_1 - B_2$, where

$$B_1 = \sum_{k=1}^{n-1} \left(\frac{1}{n} \sum_{s=1}^{n-k} \nu_s \nu_{k+s} \gamma_u(k) \right), \quad B_2 = \sum_{k=1}^{n-1} \left(\frac{1}{n} \sum_{s=1}^{n-k} \nu_s \nu_{k+s} \sum_{j=s+M+1}^{\infty} b_{k+j} b_j \right).$$

By (3.2.10), $B_2 \to_p 0$. For the term B_1, we have

$$B_1 = \sum_{k=1}^{n-1} \left(\Gamma_k - \gamma_\nu(k) \right) \gamma_u(k) + \sum_{k=1}^{n-1} \gamma_\nu(k) \gamma_u(k) =: B_{11} + B_{12}.$$

By applying Theorem 6 of Arcones (1994) and the fact that the Hermite rank of the bivariate function $\nu_t \nu_{t+k} - \gamma_\nu(k)$ is 2, we obtain $\sup_k E|\Gamma_k - \gamma_\nu(k)| \le C n^{2h-2}$ for $1/2 < h < 1$. Then

$$E|B_{11}| \le \sum_{k=1}^{n-1} \gamma_u(k) E|\Gamma_k - \gamma_\nu(k)| \to 0. \tag{3.2.15}$$

Thus (3.2.13) follows from (3.2.14) and (3.2.15) and the fact that $\lim_n B_{12}$ exists for $H + h < 3/2$.

Next, in order to prove (3.2.12), we recall that $\max_{1 \le t \le n} |X_t| = O_P(\ln n)$ from Berman (1992). For some integer $l > 0$, by the Cauchy-Schwarz inequality, we obtain that

$$
\begin{aligned}
\max_{-M \le j \le n} |d_{n,j}| &= \max_{-M \le j \le n} n^{-1/2} \sum_{t=1}^{n} \nu_t b_{t-j} \Big(I(|t-j| > l) + I(|t-j| \le l) \Big) \\
&\le n^{-1/2} \Big(\sum_{t=1}^{n} \nu_t^2 \Big)^{1/2} \max_{-M \le j \le n} \Big(\sum_{t=1}^{n} b_{t-j}^2 I(|t-j| > l) \Big)^{1/2} \\
&\quad + n^{-1/2} \max_{1 \le t \le n} |\nu_t| \max_{-M \le j \le n} \sum_{t=1}^{n} |b_{t-j}| I(|t-j| \le l) \\
&= O_P\Big(l^{-1+H} + n^{-1/2} \Big(\max_{\{0 \le x \le \ln n\}} |\nu(x)| \Big) l^{H-1/2} \Big).
\end{aligned}
$$

In view of the assumption $\max_{\{0 \le x \le \ln n\}} |\nu(x)|/n^{1/2-\eta} \to 0$, the above upper bound tends to zero in probability for any $l = O(n^{2\eta/(2H-1)})$. Hence (3.2.12) follows, thereby completing the proof of the lemma. □

Now, take $\nu(x) = x\sigma(x)$ in the above lemma. Assuming that $EX^2\sigma^2(X) < \infty$, under Assumption 3.2, $EX\sigma(X) = 0$, and the Hermite rank of this function is 1. Also, since $\max_{1 \le t \le n} |X_t| = O_P(\ln n)$, and by Lemma 3.2.3, $\bar{X}\bar{e} = o_P(n^{-1/2})$ for $H + h < 3/2$, we readily obtain

Theorem 3.2.2 *Suppose (3.1.1) - (3.1.3), and the Assumptions 3.1-2 hold. In addition, suppose* $EX^2\sigma^2(X) < \infty$, *and* $\max_{\{0 \le x \le \ln n\}} \sigma(x)/n^{1/2-\eta} \to 0$, *for some* $0 < \eta < 1/2$. *Then, for* $H + h < 3/2$, $n^{1/2}(\hat{\beta}_1 - \beta_1) \to_d N(0, \kappa_2)$.

3.3 Asymptotic distribution of the variance function

In this section we investigate the asymptotic distribution of the kernel type estimator of the conditional variance function $\sigma^2(x)$ for every fixed $x \in \mathbb{R}$. To introduce this estimator, let K be a density function on $[-1,1]$, $b = b_n$ be sequence of positive numbers and ϕ denote the density of the $N(0,1)$ r.v. Let $K_b(x) \equiv K(x/b)/b$. The kernel type estimator of $\sigma^2(x)$ corresponding to the given K and b is defined to be

$$\hat{\sigma}^2(x) := \frac{1}{n\phi(x)} \sum_{t=1}^{n} K_b(x - X_t)\, \hat{\varepsilon}_t^2, \qquad \hat{\varepsilon}_t := Y_t - \hat{\beta}_0 - \hat{\beta}_1 X_t.$$

Let $\tilde{\sigma}^2(x) = (n\phi(x))^{-1} \sum_{t=1}^{n} K_b(x - X_t)(\sigma_t u_t + \bar{e})^2$. Note that

$$\hat{\sigma}^2(x) - \tilde{\sigma}^2(x) = \frac{1}{n\phi(x)} \sum_{t=1}^{n} K_b(x - X_t)[2(\sigma_t u_t + \bar{e}) \times (X_t - \bar{X})(\beta_1 - \hat{\beta}_1) + (X_t - \bar{X})^2(\beta_1 - \hat{\beta}_1)^2], \tag{3.3.1}$$

In view of (3.2.1), (3.2.2), and the facts $s_X^2 \to_p \gamma_X(0)$, and

$$\frac{1}{n}\sum_{t=1}^{n} e_t X_t - \bar{X}\bar{e} = O_p(n^{H+h-2}),$$
$$\frac{1}{n}\sum_{t=1}^{n} K_b(x - X_t)(X_t - \bar{X})^2 = O_p(1),$$
$$\frac{1}{n}\sum_{t=1}^{n} K_b(x - X_t)(\sigma_t u_t + \bar{e})(X_t - \bar{X}) = O_p(1),$$

by Theorems 3.2.1 and 3.2.2, we obtain

$$\hat{\sigma}^2(x) - \tilde{\sigma}^2(x) = O_p(n^{H+h-2}), \qquad \forall\, x \in \mathbb{R}. \tag{3.3.2}$$

As will be seen in the sequel $n^a(\tilde{\sigma}^2(x) - \sigma^2(x)) = O_p(1)$, with $a = 1-h$ or $a = 2-2H$, depending on whether $H < (1+h)/2$ or $H > (1+h)/2$. In either case from (3.3.2) it then follows that $n^a(\hat{\sigma}^2(x) - \tilde{\sigma}^2(x)) = o_p(1)$. Thus, it suffices to obtain the asymptotic distribution of $\tilde{\sigma}^2(x)$, for a given x. Now fix an $x \in \mathbb{R}$ and consider the following additional assumptions.

Assumption 3.3. Kernel function K is a smooth density symmetric around zero and with a compact support $[-1, 1]$.

Assumption 3.4. The variance function σ^2 is twice continuously differentiable in a neighborhood of x, and satisfies $\inf_y \sigma(y) \geq c > 0$.

We claim that if $b \to 0$, $nb \to \infty$ and Assumptions 3.1-4 hold, then

$$E\tilde{\sigma}^2(x) - \sigma^2(x) = O(b^2 + n^{2H-2}). \tag{3.3.3}$$

To see this, rewrite $\tilde{\sigma}^2(x) - \sigma^2(x) := A_n + B_n + C_n$, where

$$\begin{aligned}
A_n &= \frac{1}{n\phi(x)} \sum_{t=1}^{n} K_b(x - X_t)\sigma_t^2 u_t^2 - \sigma^2(x) \\
B_n &= \frac{1}{n\phi(x)} \sum_{t=1}^{n} K_b(x - X_t)(\bar{e})^2, \\
C_n &= 2\frac{1}{n\phi(x)} \sum_{t=1}^{n} K_b(x - X_t)\sigma_t u_t \bar{e}.
\end{aligned}$$

First, by the Assumptions 3.3-4 and a routine argument, it can be seen that $EA_n = O(b^2)$.

Next, let $\phi_{t,s,k}$ denote the joint density function of (X_t, X_s, X_k). By the Assumptions 3.1 and 3.3,

$$\begin{aligned}
EB_n &= \frac{1}{n^3\phi(x)} \sum_{t,s,k} E\{K_b(x - X_t)\sigma_s\sigma_k\} E(u_k u_s) \\
&= \frac{1}{n^3\phi(x)} \sum_{t,s,k} \int\int\int K_b(x - y)\sigma(z)\sigma(w)\phi_{t,s,k}(y, z, w)dydzdw\, E(u_k u_s) \\
&= \frac{1}{n^3\phi(x)} \sum_{t,s,k} \int\int\int K(v)\sigma(z)\sigma(w)\phi_{t,s,k}(x - vb, z, w)dvdzdw\, E(u_k u_s) \\
&\sim \frac{1}{n^3\phi(x)} \sum_{t,s,k} \int\int \sigma(z)\sigma(w)\phi_{t,s,k}(x, z, w)dzdw\, E(u_k u_s)
\end{aligned}$$

since $\int K(v)dv = 1$. By Lemma 3.7.4, the double integral in the above expression is uniformly bounded from above for sufficiently large $|t-s|$ and $|t-k|$. Hence $EB_n \le Cn^{-2}\sum_{s=1}^{n}\sum_{k=1}^{n} Eu_k u_s = O(n^{2H-2})$. A similar calculation implies $C_n = O(n^{2H-2})$. These facts readily yield (3.3.3).

Next, to obtain the asymptotic distribution $\tilde{\sigma}^2(x) - \sigma^2(x)$, we need to analyze the order of magnitude of each term in the following decomposition:

$$\tilde{\sigma}^2(x) - \sigma^2(x) = I + II + III + 2IV, \tag{3.3.4}$$

where

$$
\begin{aligned}
I &= \frac{1}{n\phi(x)}\sum_{t=1}^{n} K_b(x-X_t)\sigma_t^2(u_t^2-1),\\
II &= \frac{1}{n\phi(x)}\sum_{t=1}^{n} K_b(x-X_t)\sigma_t^2 - \sigma^2(x) = II_1 + II_2,\\
II_1 &= \frac{1}{n\phi(x)}\sum_{t=1}^{n}\Big(K_b(x-X_t)\sigma_t^2 - \tau_b\Big), \quad \tau_b := EK_b(x-X)\sigma^2(X),\\
II_2 &= \frac{\tau_b}{\phi(x)} - \sigma^2(x),\\
III &= \frac{(\bar{e})^2}{n\phi(x)}\sum_{t=1}^{n} K_b(x-X_t), \quad IV = \frac{\bar{e}}{n\phi(x)}\sum_{t=1}^{n} K_b(x-X_t)e_t.
\end{aligned}
$$

First consider the term II. Note that $II_2 \le Cb^2$. To analyze the term II_1, we need one more assumption on the bandwidth.

Assumption 3.5. The bandwidth b satisfies $n^{2-2h}b \to \infty$ for $3/4 < h < 1$, and $n^{2h-1}(\ln n)^{-1}b \to \infty$ for $1/2 < h \le 3/4$.

Note that the Assumption 3.5 implies that $n^{2-2h} = o(nb)$. Let Z be the standard normal r.v.. Using the reduction principle of the kernel estimation presented in Lemma 3.7.1 in Appendix, we obtain

Lemma 3.3.1 *Suppose the Assumptions 3.1-5 hold. Then*

$$n^{1-h}II_1 \to_d x\,\sigma^2(x)\psi_1 Z.$$

To deal with the term I, using Lemma 3.5.1, (3.7.2) and that $EI = 0$, we obtain that $EI^2 = O((nb)^{-1} + n^{4H-4})$, for $3/4 < H < 1$, and $EI^2 = O((nb)^{-1})$, for $1/2 < H < 3/4$. From the proof of Lemma 3.7.1 one also obtains that $I = O_p((nb)^{-1/2}\ln^{1/2}(n))$, for $H = 3/4$. From (3.2.4), we can see $III = O_p(n^{2H-2})$. And an argument similar to (3.7.1) yields that $IV = O_p(n^{2H-2})$. We summarize these results here for a later use: Under the Assumption 3.5,

$$\begin{aligned} I &= O_p\Big(\frac{1}{\sqrt{nb}} + n^{2H-2} I(H > 3/4) + \frac{\sqrt{\ln(n)}}{\sqrt{nb}} I(H = 3/4)\Big) \\ II &= O_p(n^{h-1} + b^2), \quad III = IV = O_p(n^{2H-2}). \end{aligned} \tag{3.3.5}$$

The following theorem gives the asymptotic distributions of $\hat{\sigma}^2(x)$, where $Z^*_{n2} := n^{2-2H} I$, $Z_n = n^{2-2H}(III + IV)$, and $\kappa_3 := J_0^2 + 2J_0\sigma^2(x)$.

Theorem 3.3.1 *Suppose (3.1.1), (3.1.2), (3.1.3), and the Assumptions 3.1-5 hold.*
(a). In addition, suppose

$$H < (1+h)/2, \qquad n^{1-h} b^2 \to 0. \tag{3.3.6}$$

Then, $n^{1-h}\Big(\hat{\sigma}^2(x) - \sigma^2(x)\Big) \to_d \sigma^2(x) x\, \psi_1\, Z$.
(b). In addition, suppose

$$H > (1+h)/2, \qquad n^{1-H} b \to 0. \tag{3.3.7}$$

Then,

$$n^{2-2H}(\hat{\sigma}^2(x) - \sigma^2(x)) = \sigma^2(x) Z^*_{n2} + \kappa_3 \psi_1^2 Z_n^2 + o_p(1). \tag{3.3.8}$$

Moreover, $Z^*_{n2} \to_d \mathcal{Z}^*_2$ *and* $Z_n \to_d Z$, *where* $\mathcal{Z}^*_2$ *is the* $\mathcal{Z}_2$ *of (3.2.6) with* $B_1 = B_2$. *Furthermore,*

$$Correl(Z^*_{n2}, Z_n^2) \to \frac{2H}{4H-1}\sqrt{\frac{4H-3}{2H-1}}. \tag{3.3.9}$$

Proof. In view of (3.3.2), it suffices to prove the above claims with $\hat{\sigma}^2$ replaced by $\tilde{\sigma}^2$.

PROOF OF (a). In this case, (3.3.5) implies that the term II is the dominating term in the decomposition (3.3.4). Thus this claim follows from Lemma 3.3.1 and (3.3.3).

PROOF OF (b). Note that because $h > 1/2$, $H > (1+h)/2$ implies that $H > 3/4$. Hence here the terms I, III and IV are the dominating terms in the decomposition (3.3.4). By (3.7.2), we obtain $n^{2-2H} I \to_d \sigma^2(x) \mathcal{Z}_2^*$.

Next, because $\frac{1}{n}\sum_{t=1}^n K_b(x - X_t) - \phi(x) = o_P(1)$, and $\phi(x) > 0$, we obtain, in view of Lemma 3.2.1, that $n^{2-2H}\Big((n\phi(x))^{-1}\sum_{t=1}^n K_b(x-X_t) - 1\Big)(\bar{e})^2 = o_P(1)$. By (3.7.1), $\frac{1}{n\phi(x)}\Big(\sum_{t=1}^n K_b(x-X_t)e_t - \tau_b\sum_{t=1}^n u_t\Big) = o_P(n^{-1+H})$. These fact and $\tau_b \to \phi(x)\sigma^2(x)$ imply that $n^{2-2H}(III + IV) \to_d \kappa_3\psi_1^2 Z^2$, thereby completing the proof of (3.3.8).

To prove (3.3.9), note that $E(n^{1-H}\bar{u})^2 \to \psi_1^2$, and by (3.5.4),

$$\begin{aligned}
E(\bar{u})^4 &= \frac{4}{n^4}\sum_{t=1}^{n-1}\sum_{s_1=1}^{n-t}\sum_{s_2=1}^{n-t}\sum_{s_3=1}^{n-t} E u_0 u_{s_1} u_{s_2} u_{s_3} \\
&\sim \frac{4}{n^4}\sum_{t=1}^{n-1}\sum_{s_1=1}^{n-t}\sum_{s_2=1}^{n-t}\sum_{s_3=1}^{n-t} \Big\{\gamma_X(s_1)\gamma_X(s_3 - s_2) \\
&\qquad\qquad + \gamma_X(s_2)\gamma_X(s_1 - s_3) + \gamma_X(s_3)\gamma_X(s_2 - s_1)\Big\} \\
&\sim \frac{12 G_X \theta(H)}{n^4}\sum_{t=1}^{n-1}\sum_{s_1=1}^{n-t}\gamma_X(s_1)\frac{(n-t)^{2H}}{H(2H-1)} \sim 3\psi_1^2 n^{4H-4}.
\end{aligned}$$

Hence $\mathrm{Var}(\bar{u}^2) \sim 2\psi_1^2 n^{-4+4H}$. By (3.5.3), similar calculations yield

$$E\Big(\bar{u}^2\frac{1}{n}\sum_{t=1}^n (u_t^2 - 1)\Big) \sim \frac{4G^2\theta^2(H)}{(2H-1)^2(4H-1)} n^{-4+4H}.$$

In addition,

$$\mathrm{Var}\Big(\frac{1}{n}\sum_{t=1}^n (u_t^2 - 1)\Big) \sim \frac{2G^2\theta^2(H)}{(4H-3)(2H-1)} n^{-4+4H}.$$

This completes the proof of the theorem. □

Remark 3.3.1 Suppose we choose $b = O(n^{-\delta})$. Then the Assumption 3.5 and (3.3.6) hold, for all δ in the range $\frac{1-h}{2} < \delta < 2(1-h)$ whenever $h > \frac{3}{4}$; and for all δ in the range $\frac{1-h}{2} < \delta < 2h - 1$, whenever $h \leq \frac{3}{4}$ in case (a). Similarly, in case (b), Assumption 3.5 and (3.3.7) hold for $1 - H < \delta < 2h - 1$ whenever $h < \frac{3}{4}$; and for $1 - H < \delta < 2 - 2h$ whenever $h > \frac{3}{4}$.

Remark 3.3.2 We remark here that by using the truncation method as in Andrews (1995), the above Theorem 3.3.1 continues to hold for some symmetric density kernel function K (normal density function, say) with infinite support and finite variance.

3.4 Estimation of the LM parameter

As is seen from the above results in order to carry out inference about the underlying parameters in the above regression model or in order to carry out a lack-of-fit test as is done in the next section, we need a $\log(n)$-consistent estimators of H. In this section we consider the regression model

$$(3.4.1) \qquad Y_t = \beta' m(X_t) + \sigma(X_t) u_t, \qquad t \in \mathbb{Z},$$

where $m(x) = (m_0(x), m_1(x), \cdots, m_{q-1}(x))'$ with $m_0(x) \equiv 1$ and $m_1, \cdots, m_{q-1}$ some known functions. The process X_t, the function $\sigma(x)$ and the errors u_t are as in (3.1.2) and (3.1.3). In what follows we prove the $\log(n)$-consistency of the local Whittle estimator of H based on the least square residuals $\hat{\varepsilon}_t := Y_t - \hat{\beta}' m(X_t)$, $t = 1, \cdots, n$. To proceed further, we need the

Assumption 3.6. $Em_i(X) = 0$ and $Em_i^2(X) < \infty$ for $i = 1, \cdots, q-1$ and $A := Em(X)m(X)'$ is positive definite.

For a process $\xi_t, t = 1, 2, \cdots$, let

$$\omega_\xi(\lambda) = (2\pi n)^{-1/2} \sum_{t=1}^{n} \xi_t e^{it\lambda}, \qquad I_\xi(\lambda) = |\omega_\xi(\lambda)|^2, \qquad \lambda \in [-\pi, \pi],$$

denote its discrete Fourier transform and periodogram, respectively.

Fix $1/2 < \Delta_1 < \Delta_2 < 1$. With $\lambda_j := 2\pi j/n$ and an integer $m \in [1, n/2)$, for $\Delta_1 \le \psi \le \Delta_2$, define

$$Q(\psi) := \frac{1}{m}\sum_{j=1}^{m} \lambda_j^{2\psi-1} I_{\hat{\varepsilon}}(\lambda_j), \quad R(\psi) = \log Q(\psi) - (2\psi - 1)\sum_{j=1}^{m} \log \lambda_j.$$

Then the local Whittle estimates of G and H in the model (3.1.1) based on the residuals $\hat{\varepsilon}_t$ are defined to be

$$\hat{G} = Q(\hat{H}), \qquad \hat{H} = \text{argmin}_{\psi\in[\Delta_1,\Delta_2]} R(\psi).$$

Under some regularity conditions including the assumption that the regression model is nonparametric with non-random uniform design on $[0, 1]$ and homoscedastic ($\sigma(x) \equiv$ a constant) long memory moving average errors, Robinson (1997) proved the $\log(n)$-consistency of the analog of $\hat{H}$ and the consistency of the analog of $\hat{G}$. The following theorem shows that these results continue to hold in the regression model (3.4.1) under much simpler restrictions on the smoothing parameter m than those required for non-random design.

Theorem 3.4.1 *Suppose (3.1.2), (3.1.3), (3.4.1), and the Assumptions 3.1-2 and 3.6 hold. If, in addition,*

$$(3.4.2) \qquad (\ln n)^4\Big((\frac{m}{n})^{2H-1} + \frac{m^{2(H-h)}}{n^{1+H-2h}}\Big) \to 0.$$

Then

$$(3.4.3) \qquad \ln(n)(\hat{H} - H) \to_p 0, \qquad \hat{G} - G_u \longrightarrow_p 0.$$

Proof. The basic proof is the same as in Robinson (1997), with some difference in technical details. So we shall be brief, indicating only the main differences. With $J_0 = E\sigma(X)$, let $\eta_t := e_t - J_0 u_t$, $\xi_t := \beta_0 - \hat{\beta}_0$ and $\zeta_t := \sum_{j=1}^{q-1}(\beta_j - \hat{\beta}_j)m_j(X_t)$. Then $\hat{\varepsilon}_t$ can be rewritten as $\hat{\varepsilon}_t = \xi_t + \zeta_t + \eta_t + J_0 u_t$. Let $f_j = \lambda_j^{1-2H}$, and $\mathcal{D}_j =:$

$[I_{\hat{\varepsilon}}(\lambda_j) - J_0^2 I_u(\lambda_j)]/f_j$. According to the proof of Theorem 3 in Robinson (1997), to prove (3.4.3) for $1/2 < H < 1$, it suffices to verify the following three claims:

$$(3.4.4) \qquad \sum_{i=1}^{m-1} (\frac{i}{m})^{2(\Delta_1 - H)+1} \frac{1}{i^2} | \sum_{j=1}^{i} \mathcal{D}_j | \to_p 0,$$

$$(3.4.5) \qquad (\log n)^2 \sum_{i=1}^{m-1} (\frac{i}{m})^{1-2\delta} \frac{1}{i^2} | \sum_{j=1}^{i} \mathcal{D}_j | \to_p 0, \quad \text{for some small } \delta > 0,$$

$$(3.4.6) \qquad \frac{(\log n)^2}{m} \sum_{j=1}^{m} \mathcal{D}_j \to_p 0.$$

To verify these conditions, we use the following elementary inequalities,

$$(3.4.7) \qquad \begin{aligned} |I_{\hat{\varepsilon}}(\lambda) - J_0^2 I_u(\lambda)| &\le 2J_0 |I_u(\lambda) I_V(\lambda)|^{1/2} + I_V(\lambda), \\ I_V(\lambda) &\le 3(I_\xi(\lambda) + I_\zeta(\lambda) + I_\eta(\lambda)), \quad \forall \lambda \in [-\pi, \pi], \end{aligned}$$

where $V_t := \xi_t + \zeta_t + \eta_t$. In view of (3.4.7), it suffices to obtain upper bounds on I_ξ, I_ζ and I_η.

Recall that for the Dirichlet kernel $D_k(\lambda) := \sum_{t=1}^{k} e^{it\lambda}$, $|D_k(\lambda)| \le C/\lambda$, for all $\lambda \in [-\pi, \pi]$, $k \ge 1$. Also, note that from (3.2.3) applied q-times to $\nu(x) = m_j(x)\sigma(x)$, $j = 0, 1, \cdots, q-1$, we obtain that

$$(3.4.8) \qquad n^{1-H} \|\hat{\beta} - \beta\| = O_P(1).$$

These bounds imply that

$$(3.4.9) \qquad \frac{I_\xi(\lambda_j)}{f_j} = O_P(\frac{1}{n^{3-2H} \lambda_j^{3-2H}}), \quad \text{uniformly for } 1 \le j \le m.$$

Next, by Assumption 3.2, the function $\sigma(X) - J_0$ has the Hermite rank $r \ge 1$, and hence by Lemma 3.7.2 , we obtain, uniformly for $1 \le j \le m$,

$$(3.4.10) \qquad \begin{aligned} \frac{I_\eta(\lambda_j)}{f_j} &= O_P(\lambda_j^{r(2-2h)}), \quad 0 < r(2-2h) + (2-2H) < 1; \\ &= O_P(\lambda_j^{2H-1} \log n), \quad r(2-2h) + (2-2H) \ge 1. \end{aligned}$$

For the terms ζ_t, Assumption 3.6 implies that the Hermite ranks of $m_j(X), j = 1, 2, \cdots, q-1$, are at least one, therefore, by (3.4.8) and Corollary 4.10.2 of Bingham et al (1987), using (3.4.7), we obtain that uniformly for $1 \leq j \leq m$,

$$\frac{I_\zeta(\lambda_j)}{f_j} = O_P(\lambda_j^{2(H-h)} n^{H-1}). \tag{3.4.11}$$

Now we are ready to verify (3.4.4)-(3.4.6). By changing the order of summation, the l.h.s. of (3.4.4) is bounded above by

$$Cm^{2(H-\Delta_1)-1} \sum_{j=1}^{m} j^{2(\Delta_1 - H)} |\mathcal{D}_j|,$$

for $H > \Delta_1$, and by $Cm^{-1} \log m \sum_{j=1}^{m} |\mathcal{D}_j|$, for $H = \Delta_1$. But, by (3.4.7) and the Cauchy-Schwarz inequality,

$$\begin{aligned}
\sum_{j=1}^{m} j^{2(\Delta_1-H)} |\mathcal{D}_j| &\leq C \sum_{j=1}^{m} j^{2(\Delta_1-H)} \frac{J_0 |I_u(\lambda_j) I_V(\lambda_j)|^{1/2} + I_V(\lambda_j)}{f_j} \\
&\leq C \Big(\sum_{j=1}^{m} j^{2(\Delta_1-H)} \frac{|I_u(\lambda_j)|}{f_j} \Big)^{1/2} \Big(\sum_{j=1}^{m} j^{2(\Delta_1-H)} \frac{|I_V(\lambda_j)|}{f_j} \Big)^{1/2} \\
&\quad + C \sum_{j=1}^{m} j^{2(\Delta_1-H)} \frac{|I_V(\lambda_j)|}{f_j}.
\end{aligned}$$

This bound together with Lemma 3.7.3, (3.4.9)-(3.4.11) and the following calculations imply (3.4.4) for $H > \Delta_1$:

$$m^{2(H-\Delta_1)-1} \sum_{j=1}^{m} j^{2(\Delta_1-H)} \frac{I_\xi(\lambda_j)}{f_j} = o_P(1).$$

$$m^{2(H-\Delta_1)-1} \sum_{j=1}^{m} j^{2(\Delta_1-H)} \frac{I_\eta(\lambda_j)}{f_j} = o_P(1).$$

$$m^{2(H-\Delta_1)-1} \sum_{j=1}^{m} j^{2(\Delta_1-H)} \frac{I_\zeta(\lambda_j)}{f_j} = o_P(1).$$

The proof of (3.4.4) for $H = \Delta_1$ is similar. One can also verify other conditions using the same method. □

Remark 3.4.1 In Theorem 3.4.1, suppose $m = Cn^a$ for $0 < a < 1$. Then, the condition (3.4.2) holds for $H \geq h$. In the case $H < h$, it holds for any $a > (2h - H - 1)/(2h - 2H)$. In particular, as discussed in Henry and Robinson (1995), when the spectral density function $f_u(\lambda)$ of a Gaussian error process u_t satisfies some mild conditions, the optimal bandwidth $m = Cn^a$, with $a = 4/5$. This choice of a always satisfies (3.4.2).

3.5 Regression model diagnostics

In this section we investigate the weak convergence of $\hat{\mathcal{V}}_n$ under H_0 and the assumptions (3.1.2) and (3.1.3), where now $Y_t - \beta' m(X_t) = \sigma(X_t)u_t$. Assume $m(x)' = (m_0(x), m_1(x), \cdots, m_{q-1}(x))$ with $m_0 \equiv 1$.

Recall the definition (3.1.5). Also, let $\bar{A}_n := \frac{1}{n}\sum_{t=1}^n m(X_t)m(X_t)'$. By the Ergodic Theorem, $\bar{A}_n \to_{a.s.} A$. Let $\hat{\beta}$ be the LSE of β in this model and $Z_n = \sum_{t=1}^n m(X_t)\sigma_t u_t$, and for an $x \in \bar{\mathbb{R}}$,

$$\begin{aligned}
\bar{\alpha}_n(x) &:= \frac{1}{n}\sum_{t=1}^n m(X_t)I(X_t \leq x), \\
\mathcal{V}_n(x) &:= \sum_{t=1}^n \sigma_t u_t I(X_t \leq x), \quad \hat{\mathcal{V}}_n(x) := \sum_{t=1}^n \Big(Y_t - \hat{\beta}' m(X_t)\Big) I(X_t \leq x),
\end{aligned}$$

Next, we state a Glivenko-Cantelli type result that is used repeatedly in the sequel: for a measurable real valued function g, with $E|g(X)| < \infty$,

$$\sup_{x \in \bar{\mathbb{R}}} \left| \frac{1}{n}\sum_{t=1}^n g(X_t)I(X_t \leq x) - Eg(X)I(X \leq x) \right| \longrightarrow_{a.s.} 0. \tag{3.5.1}$$

For a non-negative g this follows from the Ergodic Theorem and the classical Glivenko-Cantelli argument where $\bar{\mathbb{R}}$ is partitioned such that the oscillation of the measure $Eg(X)I(X \leq x)$ is small. The result for a general g is obtained from this result applied to $g^{\pm}$ and the triangle inequality. We shall be also using the Ergodic Theorem repeatedly without mentioning.

We also need to recall that $\hat{\beta} - \beta = \bar{A}_n^{-1} \frac{1}{n} \sum_{t=1}^{n} m(X_t)\sigma_t u_t$. We are now ready to state and prove

Theorem 3.5.1 *Suppose (3.1.1)-(3.1.3), Assumptions 3.1 and 3.6 hold. Then,*

$$n^{-H} \hat{\mathcal{V}}_n(x) \Longrightarrow J_\sigma(x)\tau(H)N(0,1), \qquad \text{in } D(\bar{\mathbb{R}}), \text{ uniform metric.}$$

Proof. First note $\hat{\mathcal{V}}_n(x) = \mathcal{V}_n(x) - Z_n' \bar{A}_n \bar{\alpha}_n(x)$. Moreover,

$$\begin{aligned} \mathcal{V}_n(x) &= F_\sigma(x) \sum_{t=1}^{n} u_t + \sum_{t=1}^{n} u_t \Big\{ \sigma_t I(X_t \le x) - F_\sigma(x) \Big\}, \\ Z_n &= \kappa_\sigma \sum_{t=1}^{n} u_t + \sum_{t=1}^{n} u_t \Big\{ m(X_t)\sigma_t - \kappa_\sigma) \Big\}. \end{aligned}$$

Since $\bar{\alpha}_n(x) \to_{a.s.} \alpha(x)$, by (3.2.3), (3.2.4), and (3.5.1), we obtain that

$$n^{-H} \hat{\mathcal{V}}(x) = n^{-H} \sum_{t=1}^{n} u_t \Big(F_\sigma(x) - \kappa_\sigma' A^{-1} \alpha(x) \Big) + u_p(1).$$

□

The above result is useful only if $J_\sigma(x) \not\equiv 0$. Observe that $J_\sigma(x) = \int_{-\infty}^{x} \Big(\sigma(y) - \mu_\sigma' A^{-1} m(y) \Big) \phi(y) dy$. Hence, the condition $J_\sigma \equiv 0$ implies that the functions $\{\sigma(x), 1, m_1(x), \cdots, m_q(x)\}$ are linearly dependent, for almost all $x \in \mathbb{R}$.

In the simple linear regression model, when $m(x) = (1, x)'$, $J_\sigma(x) \equiv 0$ if and only if $\sigma(x) \equiv c$, a constant. In this case,

$$J_\sigma(x) = E\sigma(X)I(X \le x) - E\sigma(X)\Phi(x) + EXI(X \le x)EX\sigma(X).$$

Clearly, $\sigma(x) \equiv c$ implies $J_\sigma(x) \equiv 0$. On the other hand, if $J_\sigma(x) = 0$ for all x, then upon differentiating the equation $J_\sigma(x) = 0$ with respect to x, we have $\sigma(x) = E\sigma(X) + xEX\sigma(X)$, which contradicts the assumption that $\sigma(x) > 0$ for all x if $EX\sigma(X) \ne 0$, and which in turn implies that $\sigma(x)$ must be a constant.

In order to implement the above result, we shall require that $J_\sigma(x) \ne 0$ for some $x \in \mathbb{R}$. We also need a uniformly consistent estimator of $J_\sigma(x)$ in order to apply tests based on the process $\hat{\mathcal{V}}_n$. The following condition is needed for this purpose:

Assumption 3.7. $\sigma(x) \geq c > 0$ and has a continuous first derivative function of $x \in \mathbb{R}$.

Observe that the only unknown entity in J_σ is $\sigma(X)$. For technical reasons the theory is much harder if we use the previous estimator of σ^2 in this process. We shall use an alternate estimator based on the ideas of cross validation method that leaves one observation out each time. Let

$$\hat{\Lambda}^2_{-i}(x) := \frac{1}{n-1}\sum_{t \neq i}^{n} K_b(x - X_t)\hat{\varepsilon}^2_t, \qquad i = 1, \cdots, n.$$

Then, $V_i(x) := \hat{\Lambda}_{-i}(x)\phi^{-1/2}(x)$ is an estimator of $\sigma(x)$, and

$$\hat{J}_n(x) = \frac{1}{n}\sum_{t=1}^{n} V_t(X_t)I(X_t \leq x) - \frac{1}{n}\sum_{t=1}^{n} m(X_t)V_t(X_t)\bar{A}_n^{-1}\bar{\alpha}_n(x)$$

is an estimator of $J_\sigma(x)$. Its uniform consistency is assessed by the following theorem.

Theorem 3.5.2 *Suppose (3.1.1)-(3.1.3), and Assumptions 3.1, 3.6, and 3.7 hold. In addition, suppose* $b \to 0$, $b^{-1}n^{2h-2} = O(1)$, $E\sigma^2(X)\phi^{1/2}(X) < \infty$, $E\sigma^2(X)m_j^4(X) < \infty$, *and* $E|m_j(2X)|^2 < \infty$, *for* $j = 1, \cdots, q-1$. *Then, under* H_0, $\sup_{x \in \bar{\mathbb{R}}} |\hat{J}_n(x) - J_\sigma(x)| = o_P(1)$.

The proof of this theorem follows from several lemmas proved below. We begin with stating some preliminary facts about some moments of the LMMA process u_t. Let $d_t := E(u_0^2 - 1)(u_t^2 - 1)$, $t \in \mathbb{Z}$. The following lemma is proved in Chapter 2.2.3.

Lemma 3.5.1 *Suppose* $\{u_t\}$ *is as in* (3.1.2) *with* $E\varepsilon_0^4 < \infty$. *Then, for all* $1/2 < H < 1$,

$$d_t = 2D^2 t^{2(2H-2)} + o(t^{2(2H-2)}), \quad \text{as } t \to \infty, \tag{3.5.2}$$

$$Eu_0 u_s(u_t^2 - 1) \sim 2\gamma_t\gamma_{t-s}, \quad \textit{as } |t - s| \to \infty, \tag{3.5.3}$$

and for s, t, r *such that* $|t - s|, |r - t|$ *and* $|s - r|$ *all tending to infinity, we have*

$$Eu_0 u_s u_t u_r \sim \gamma_s\gamma_{r-t} + \gamma_t\gamma_{r-s} + \gamma_r\gamma_{t-s}. \tag{3.5.4}$$

Next, let $\tilde{\Lambda}^2_{-t}(x) = (n-1)^{-1} \sum_{i \neq t}^n K_b(x - X_i)\sigma_i^2 u_i^2$.

Lemma 3.5.2 *Suppose the conditions of Theorem 3.5.2 hold. Then*

$$\max_{1 \leq t \leq n} E\Big(\tilde{\Lambda}^2_{-t}(X_t) - \sigma_t^2 \phi(X_t)\Big)^2 \to 0.$$

Proof. It suffices to show

$$\max_{1 \leq t \leq n} E\Big(\frac{1}{n-1} \sum_{j \neq t}^{n} K_b(X_t - X_j)\sigma_j^2(u_j^2 - 1)\Big)^2 \to 0, \tag{3.5.5}$$

$$\max_{1 \leq t \leq n} E\Big(\frac{1}{n-1} \sum_{j \neq t}^{n} K_b(X_t - X_j)\sigma_j^2 - \sigma_t^2\phi(X_t)\Big)^2 \to 0. \tag{3.5.6}$$

To prove (3.5.5), the expectation in the l.h.s. of (3.5.5) equals $A_{n,t} + B_{n,t}$, where

$$A_{n,t} =: \frac{1}{(n-1)^2} \sum_{j \neq t} E\{K_b^2(X_t - X_j)\sigma_j^4\} E(u_j^2 - 1)^2,$$

$$B_{n,t} =: \frac{1}{(n-1)^2} \sum_{j \neq t}^{n} \sum_{k \neq t, k \neq j}^{n} E\{K_b(X_t - X_j)K_b(X_t - X_k)\sigma_j^2\sigma_k^2\} \times E(u_j^2 - 1)(u_k^2 - 1).$$

Upon applying (3.7.7) below with $g(X_t, X_j, X_k) = EK_b(X_t - X_j)K_b(X_t - X_k)\sigma_j^2\sigma_k^2$ and using the fact that for this g, $||g||^2 \leq Cb^{-1}$, we obtain that, uniformly in t,

$$\begin{aligned} B_{n,t} &\leq \frac{C}{(n-1)^2} \sum_{j \neq t}^{n} \sum_{k \neq t, k \neq j}^{n} \Big\{E^0 K_b(X_t - X_j)K_b(X_t - X_k)\sigma_j^2\sigma_k^2 \\ &\qquad + Cb^{-1}\lambda^{1/2}_{|t-j|,|t-k|}\Big\} E(u_j^2 - 1)(u_k^2 - 1) \\ &\leq Cn^{4H-4} + Cb^{-1/2}n^{h-1+4H-4} \to 0, \end{aligned} \tag{3.5.7}$$

by Lemma 3.5.1. Similarly, by (3.7.8) we obtain that uniformly in t, $A_{n,t} \leq C(nb)^{-1}$. Hence (3.5.5) holds.

To prove (3.5.6), rewrite the l.h.s of (3.5.6) as

$$
\begin{aligned}
&E\frac{1}{(n-1)^2}\sum_{j,k\neq t}^{n} K_b(X_t - X_j)K_b(X_t - X_k)\sigma_j^2\sigma_k^2 \\
&\qquad -2E\frac{1}{n-1}\sigma^2(X_t)\phi(X_t)\sum_{j\neq t}^{n} K_b(X_t - X_j)\sigma_j^2 + E\sigma^4(X)\phi^2(X) \\
&=: C_{n,t} - 2D_{n,t} + E\sigma^4(X)\phi^2(X), \qquad \text{say.}
\end{aligned}
$$

Similar to the argument in (3.5.7), by (3.7.7) and (3.7.8), the terms $C_{n,t}$ and $D_{n,t}$ tend to $E\sigma^4(X)\phi^2(X)$ uniformly in t, thereby proving (3.5.6). □

By applying the simple inequality

(3.5.8) $$|a^{1/2} - b^{1/2}|^2 \leq |a - b|, \quad a, b \geq 0,$$

we obtain the following:

Corollary 3.5.1 *Suppose the conditions of Theorem 3.5.2 hold. Then,*

(3.5.9) $$\max_{1\leq t\leq n} E|\tilde{\Lambda}_{-t}(X_t) - \sigma_t\phi^{1/2}(X_t)|^4 \to 0.$$

Lemma 3.5.3 *Under the conditions of Theorem 3.5.2,*

$$\frac{1}{n}\sum_{t=1}^{n}\left|\hat{\Lambda}_{-t}(X_t) - \tilde{\Lambda}_{-t}(X_t)\right|^2\phi^{-1/2}(X_t) \ \to_p \ 0.$$

Proof. Applying (3.5.8) again, it suffices to show that

(3.5.10) $$\frac{1}{n}\sum_{t=1}^{n}\left|\hat{\Lambda}^2_{-t}(X_t) - \tilde{\Lambda}^2_{-t}(X_t)\right|\phi^{-1/2}(X_t) \ \to_p \ 0.$$

But $\left|\hat{\Lambda}^2_{-t}(X_t) - \tilde{\Lambda}^2_{-t}(X_t)\right|$ is bounded above by

$$\frac{1}{n-1}\sum_{i\neq t}^{n} K_b(X_t - X_i)\left|[(\beta - \hat{\beta})'m(X_i)]^2 + 2(\beta - \hat{\beta})'m(X_i)\sigma_i u_i + \sigma_i^2 u_i^2\right|.$$

Moreover, (3.7.8) in the Appendix implies that, for $k = 0, 1, 2$, $0 \le j \le q-1$,

$$
\begin{aligned}
& E\Big|\frac{1}{n(n-1)}\sum_{t=1}^{n}\sum_{i\neq t}^{n} K_b(X_t - X_i)\sigma_i u_i m_j^k(X_i)\phi^{-1/2}(X_t)\Big| \\
\le\ & \frac{C}{n(n-1)}\sum_{t=1}^{n}\sum_{i\neq t}^{n}\int\int K_b(x-y)\sigma(y)|m_j(y)|^k\phi^{-1/2}(x)\phi_{i,t}(x,y) \\
& \qquad\qquad\qquad\qquad\qquad \times dxdy \\
\le\ & \frac{C}{n(n-1)}\sum_{t=1}^{n}\sum_{i\neq t}^{n}\int\int K_b(x-y)\sigma(y)|m_j(y)|^k\phi^{1/2}(x)\phi(y)dxdy \\
& +\frac{C}{n^{1-h}b^{1/2}} = O(1). \\
& E\frac{1}{n(n-1)}\sum_{t=1}^{n}\sum_{i\neq t}^{n} K_b(X_t - X_i)|m_j(X_i)|^k\,\phi^{-1/2}(X_t) \\
\le\ & \frac{C}{n(n-1)}\sum_{t=1}^{n}\sum_{i\neq t}^{n}\int\int K_b(x-y)|m_j(y)|^k\phi^{1/2}(x)\phi(y)dxdy \\
& +\frac{C}{n^{1-h}b^{1/2}} = O(1).
\end{aligned}
$$

In the above we used the fact that $Em_j^4(X) < \infty$, which is implied by the assumption $E\sigma^2(X)m_j^4(X) < \infty$ and the Assumption 3.7, for $j = 1, \cdots, q-1$. Therefore, (3.5.10) holds because $\beta_1 - \hat{\beta}_1 \to_p 0$, $\bar{e} \to_p 0$ and $\bar{X} \to_p 0$. □

In the sequel, for a $y \in \mathbb{R}^q$, $|y|$ and $\|y\|$ stand, respectively, for the q-vector of the absolute values of the coordinates of y and the Euclidean norm of y. For a finite dimensional square matrix A, $\|A\|$ stands for its Euclidean norm.

Corollary 3.5.2 *Under the conditions of Theorem 3.5.2,*

$$
\frac{1}{n}\sum_{t=1}^{n}\Big|\Big(V_t(X_t) - \sigma_t\Big)m(X_t)\Big| \to_p 0.
$$

Proof. We only prove the claim for $m_1(x)$, the proof for the other terms is similar.

For this, it suffices to show that

$$(3.5.11)\qquad \frac{1}{n}\sum_{t=1}^{n}\left|\hat{\Lambda}_{-t}(X_t)-\tilde{\Lambda}_{-t}(X_t)\right||m_1(X_t)|\phi^{-1/2}(X_t) = o_P(1),$$

$$(3.5.12)\qquad \frac{1}{n}\sum_{t=1}^{n}\left|\tilde{\Lambda}_{-t}(X_t)\phi^{-1/2}(X_t)-\sigma(X_t)\right||m_1(X_t)| = o_P(1).$$

By the Hölder inequality, the expectation of the l.h.s. of (3.5.12) is bounded above by

$$\frac{1}{n}\sum_{t=1}^{n}E^{1/3}\Big(\tilde{\Lambda}_{-t}(X_t)-\sigma(X_t)\phi^{1/2}(X_t)\Big)^3E^{2/3}\Big(m_1^{3/2}(X_t)\phi^{-3/4}(X_t)\Big).$$

Since $E|m_1(X)|^{3/2}/\phi^{3/4}(X) = E|m_1(2X)|^{3/2} < \infty$, (3.5.12) follows from (3.5.9).

Next, by the Cauchy-Schwarz inequality, the l.h.s. of (3.5.11) is bounded above by

$$\Big\{\frac{1}{n}\sum_{t=1}^{n}\left|\hat{\Lambda}_{-t}(X_t)-\tilde{\Lambda}_{-t}(X_t)\right|^2\phi^{-1/2}(X_t)\times\frac{1}{n}\sum_{t=1}^{n}m_1^2(X_t)\phi^{-1/2}(X_t)\Big\}^{1/2}.$$

But because $\frac{1}{n}\sum_{t=1}^{n}m_1^2(X_t)(\phi(X_t))^{-1/2}\rightarrow_{a.s.} Em_1^2(X)(\phi(X)^{-1/2}\le C\,Em_1^2(2X) < \infty$, (3.5.11) follows from this bound and Lemma 3.5.3. □

PROOF OF THEOREM 3.5.2. The proof consists of the following two parts. First,

$$\begin{aligned}
&\sup_x\frac{1}{n}\Big|\sum_{t=1}^{n}V_t(X_t)I(X_t\le x)-F_\sigma(x)\Big|\\
&\le \sup_x\frac{1}{n}\Big|\sum_{t=1}^{n}V_t(X_t)I(X_t\le x)-\sigma_tI(X_t\le x)\Big|\\
&\qquad+\sup_x\frac{1}{n}\Big|\sum_{t=1}^{n}\sigma_tI(X_t\le x)-F_\sigma(x)\Big|\\
&\le \frac{1}{n}\sum_{t=1}^{n}\Big|V_t(X_t)-\sigma_t\Big|+\sup_x\frac{1}{n}\Big|\sum_{t=1}^{n}\sigma_tI(X_t\le x)-F_\sigma(x)\Big|.
\end{aligned}$$

The first term in this bound tends to zero in probability by Corollary 3.5.2. The second term tends to zero almost surely by (3.5.1). Secondly,

$$
\begin{aligned}
&\sup_x \Big|\frac{1}{n}\sum_{t=1}^{n}\Big(V_t(X_t)m(X_t)'\bar{A}_{\frac{1}{n}}\bar{\alpha}_n(x) - E\sigma(X)m(X)'A^{-1}\alpha(x)\Big)\Big| \\
\le\ & \Big\|\frac{1}{n}\sum_{t=1}^{n}\Big(V_t(X_t) - \sigma_t\Big)m(X_t)\Big\|\,\|\bar{A}_{\frac{1}{n}}\|\,\sup_x\|\bar{\alpha}_n(x)\| \\
&+\frac{1}{n}\sum_{t=1}^{n}\sigma_t\|m(X_t)\|\,\|\bar{A}_{\frac{1}{n}} - A^{-1}\|\sup_x\|\bar{\alpha}_n(x)\| \\
&+\frac{1}{n}\sum_{t=1}^{n}\sigma_t\|m(X_t)\|\,\|A^{-1}\|\,\sup_x\|\bar{\alpha}_n(x) - \alpha(x)\| \\
&+\Big\|\frac{1}{n}\sum_{t=1}^{n}(\sigma_t m(X_t) - \kappa_\sigma)\Big\|\,\|A^{-1}\|\,\sup_x\|\alpha(x)\|,
\end{aligned}
$$

which tends to zero in probability by the facts

$$\sup_x|\alpha(x)| \le \sqrt{E\|m(X)\|^2} < \infty,$$

$\sup_x|\bar{\alpha}_n(x) - \alpha(x)| \to_{a.s.} 0$ by (3.5.1), and Corollary 3.5.2. □

An immediate consequence of the above results is that in the case $m(x) = (1, x)'$, the test that rejects H_0, whenever

$$D_n := \frac{1}{n^{\hat{H}}\tau(\hat{H})\sup_x|\hat{J}_n(x)|}\sup_x|\hat{\mathcal{V}}_n(x)| \ge z_{\alpha/2}, \tag{3.5.13}$$

is of the asymptotic size α, where z_α is $100(1-\alpha)$th percentile of the standard normal distribution.

Now we consider the consistency of the proposed test under any fixed alternative $\mu(x) = \ell(x)$, where $\ell(x)$ is not a linear function, $E\ell^2(X) < \infty$ and $\sup_{x\in\bar{\mathbb{R}}}|E\Big(\beta^T m(X) - \ell(X)\Big)I(X \le x)| \ne 0$. Then, under this alternative,

$$
\begin{aligned}
&n^{-H}\mathcal{V}_n(x) \\
=\ & n^{-H}\sum_{t=1}^{n}\sigma(X_t)u_t I(X_t \le x) + n^{-H}\sum_{t=1}^{n}\Big(\ell(X_t) - \beta^T m(X_t)\Big)I(X_t \le x) \\
=\ & I(x) + II(x), \qquad \text{say.}
\end{aligned}
$$

By (3.5.1), $n^{H-1}II(x) = E\{\beta^T m(X) - \ell(X)\}I(X \leq x) + u_p(1)$. Also, by (3.2.3) and (3.2.4), $\sup_x |I(x)| = O_p(1)$. Thus the power of the tests will tend to one against any fixed alternative $\ell(x)$ of the above type.

3.6 Empirical results

3.6.1 A small simulation study

In this section we report the findings of a finite sample simulation study. In this simulation, for simplicity, we take $m(x) = (1, x)'$, $\beta_0 = 0$, $\beta_1 = 2$ and $\sigma^2(x) = 1+x^2$. The processes $\{u_t\}$ is taken to be ARIMA$(0, H-1/2, 0)$ with standardized Gaussian innovations and $\{X_t\}$ is taken to be fractional Gaussian noise with long memory parameter h. The values of H, h range in the interval $[.6, .95]$ with increments of .05. In order that the generated processes are stationary, we trim off the first 500 generated observations of both $\{u_t\}$ and $\{X_t\}$ processes. These processes were generated using the codes given in Beran (1994, Ch. 12).

We first concentrate on the properties of $\hat{\beta}_1$ and $\hat{H}$. Table 3.1 provides the root mean square errors (RMSE) of the LSE $\hat{\beta}_1$ with sample size 500 and 2000 replications. As can be seen from this table, when $H + h$ increases, so does the RMSE of $\hat{\beta}_1$. Typically, when $H + h < 3/2$, the RMSE is small.

Table 3.2 provides the RMSE's of the local Whittle estimator $\hat{H}$ of H based on the samples of size 500 with 1000 replications. The calculation of $\hat{H}$ is based on pseudo residuals $\hat{\varepsilon}_t = Y_t - \hat{\beta}_1 X_t$ without estimating the variance function $\sigma^2(x)$. From this table, we observe that for $H \leq 0.85$, the overall RMSE is less than 0.072 and stable regardless of the values of h.

Next, to asses the finite sample behavior of $\hat{\sigma}^2$, we simulated the function estimator $\hat{\sigma}^2(x)$ for the values of x in the grid $x_1 = -1.50$, $x_2 = -1.49, \cdots, x_{301} = 1.50$, and

Table 3.1: RMSE of the LSE $\hat{\beta}_1$ for sample size $n = 500$.

$H \setminus h$	.60	.65	.70	.75	.80	.85	.90	.95
.60	.0087	.0086	.0088	.0098	.0104	.0115	.0135	.0192
.65	.0084	.0095	.0107	.0117	.0123	.0134	.0176	.0247
.70	.0104	.0101	.0114	.0135	.0146	.0176	.0215	.0341
.75	.0108	.0121	.0135	.0154	.0194	.0227	.0304	.0465
.80	.0123	.0139	.0176	.0192	.0244	.0333	.0479	.0735
.85	.0141	.0177	.0218	.0283	.0362	.0487	.0704	.1254
.90	.0186	.0237	.0310	.0398	.0540	.0834	.1201	.2087
.95	.0257	.0340	.0519	.0647	.1137	.1762	.2738	.4962

Table 3.2: RMSE of $\hat{H}$ based on $Y_t - \hat{\beta}_1 X_t$ for sample size $n = 500$.

$H \setminus h$	.60	.65	.70	.75	.80	.85	.90	.95
.60	.0396	.0386	.0393	.0387	.0377	.0394	.0398	.0394
.65	.0428	.0427	.0440	.0426	.0433	.0451	.0416	.0432
.70	.0481	.0475	.0477	.0482	.0504	.0486	.0475	.0469
.75	.0536	.0548	.0554	.0547	.0538	.0507	.0529	.0494
.80	.0623	.0630	.0623	.0592	.0582	.0597	.0581	.0551
.85	.0708	.0720	.0707	.0670	.0658	.0631	.0622	.0577
.90	.0833	.0832	.0808	.0780	.07688	.0724	.0678	.0651
.95	.1129	.1109	.1089	.1030	.0946	.08520	.0782	.0672

Table 3.3: Ranges for δ of the bandwidths for estimation σ.

$H \setminus h$	.65	.75	.85	.95
.65	(**a**) (.175, .3)	(**a**) (.125, .5)	(**a**) (.075, .3)	(**a**) (.025, .1)
.75	(**a**) (.175, .3)	(**a**) (.125, .5)	(**a**) (.075, .3)	(**a**) (.025, .1)
.85	(**b**) (.15, .3)	(**a**) (.125, .5)	(**a**) (.075, .3)	(**a**) (.025, .1)
.95	(**b**) (.05, .3)	(**b**) (.05, .5)	(**b**) (.05, .3)	(**a**) (.025, .1)

Table 3.4: Summary of ASE($\hat{\sigma}^2$) for $H = .65$

$h \setminus Summary$	bandwidth	Q1	Median	Mean	Q3
.65	$3n^{-.2}$	.0261	.0369	.0424	.0512
.75	$3.5n^{-.2}$	.0256	.0383	.0432	.0557
.85	$4n^{-.2}$	.0273	.0417	.0595	.0617
.95	$1.5n^{-.099}$	.0366	.0663	.1138	.1058

Table 3.5: Summary of ASE($\hat{\sigma}^2$) for $H = .75$

$h \setminus Summary$	bandwidth	Q1	Median	Mean	Q3
.65	$4n^{-.2}$	.0442	.0711	.0887	.1127
.75	$4n^{-.2}$	.0465	.0652	.0888	.1076
.85	$4n^{-.2}$	.04667	.0774	.1043	.1252
.95	$2n^{-.099}$	.0627	.0995	.2190	.1902

Table 3.6: Summary of ASE($\hat{\sigma}^2$) for H = .85

$h \setminus Summary$	bandwidth	Q1	Median	Mean	Q3
.65	$4.5n^{-.2}$	.1562	.2724	.5402	.5584
.75	$6n^{-.2}$	.1594	.3113	.5449	.6330
.85	$5n^{-.2}$	.1625	.3252	.5475	.6103
.95	$2.5n^{-.099}$	.1704	.3155	.7092	.6235

Table 3.7: Summary of ASE($\hat{\sigma}^2$) for H = .95

$h \setminus Summary$	bandwidth	Q1	Median	Mean	Q3
.65	$6n^{-.2}$	1.153	3.214	16.24	11.83
.75	$7n^{-.2}$	1.137	3.078	14.75	11.25
.85	$7.5n^{-.2}$	1.018	2.611	12.77	11.59
.95	$4.5n^{-.099}$	1.136	3.374	12.57	11.85

for $0.65 \leq H, h \leq 0.95$. We used the built-in smoothing function of R program with the normal kernel and sample size 500 repeated 500 times. The ranges for δ in the bandwidths $b = Cn^{-\delta}$ are given in Table 3.3 according to the Remark 3.3.1. The symbols (**a**) and (**b**) indicate 'Case a' and 'Case b' in Theorem 3.3.1, respectively. Based on Table 3.3, for convenience we used $\delta = 0.2$, $b = Cn^{-.2}$ in our simulations for all cases of H and h considered except when $h = .95$. In the case $h = .95$, we used $\delta = 0.099$. The constant C is adjusted for different values of H and h according to the average squared errors: $ASE := \sum_{k=1}^{301}(\hat{\sigma}^2(x_k)/\sigma^2(x_k) - 1)^2/301$. We record those C values which possibly make ASE the smallest. Some summary statistics of ASE are reported in Tables 3.4-3.7. It can be seen that the estimator $\hat{\sigma}^2(x)$ is relatively stable for the values of $H, h \leq .85$. Similar results are observed when we replace the normal kernel by the kernel function $K(x) = .5(1+\cos(x\pi))I(|x| \leq 1)$ or the uniform kernel.

3.6.2 Application to a foreign exchange data set

In this section we shall apply the above proposed regression model diagnostic test to fit a simple linear regression model with heteroscedastic errors to some currency exchange rate data obtained from

www.federalreserve.gov/releases/H10/hist/.

The data are noon buying rates in New York for cable transfers payable in foreign currencies.

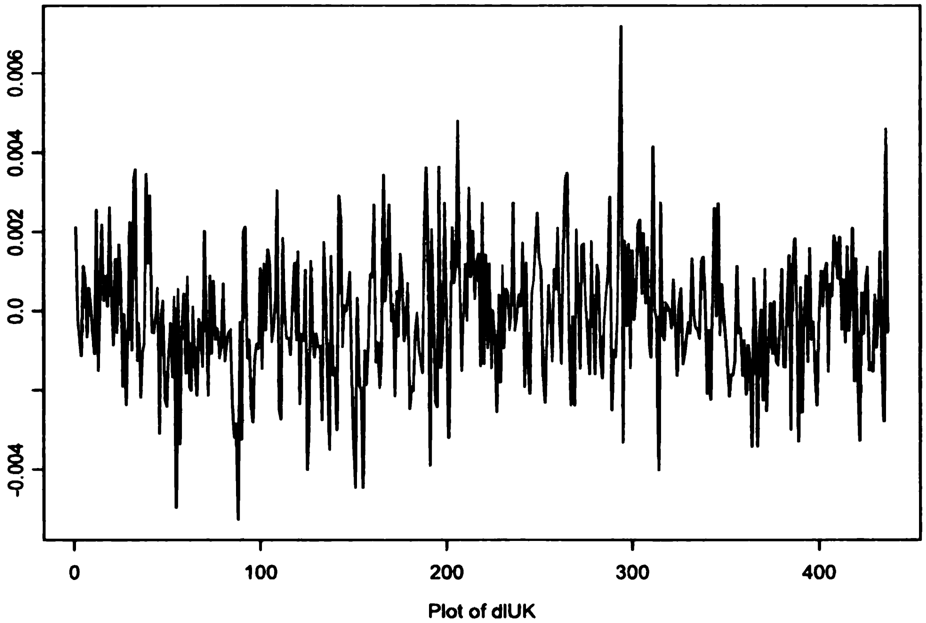
In this example, we use the currency exchange rates of the United Kingdom Pounds (UK£) vs. US$ and the Switzerland Franc (SZF) vs. US$ from January 4, 1971 to December 2, 2005. We first delete missing values and obtain about 437 monthly observations. The symbols dlUK and dlSZ stand for differenced log exchange rate of UK£ vs. US$ and SZF vs. US$, respectively. From the figure 3.1, we observed that these two sequences appear to be stationary. Also,

$$\text{mean(dlUK)} = -0.0001775461, \ \text{Stdev(dlUK)}=0.001701488,$$
$$\text{mean(dlSZ)}=-0.00004525129, \ \text{Stdev(dlSZ)}=0.001246904.$$

The local Whittle estimated values of the LM parameters of dlUK and dlSZ processes, respectively, are 0.6610273 and .7147475. In computing local Whittle estimator the choice of the smoothing parameter m is crucial. Taqqu and Teverovsky (1997) recommend $m = n/4$ and $m = n/32$ for sample sizes $n = 100$ and $n = 10,000$, respectively. Our sample size being in between these two, we chose $m = n/8$ in obtaining the above local Whittle estimates.

Let Y = dlSZ and X =dlUK. Comparing the X-process with a simulated fractional Gaussian noise with $\hat{H}_X = 0.6610273$ and $n = 437$, Figure 3.2 suggests that the marginal distribution of X is Gaussian.

Figure 3.1: The time series plots of dlUK and dlSZ.



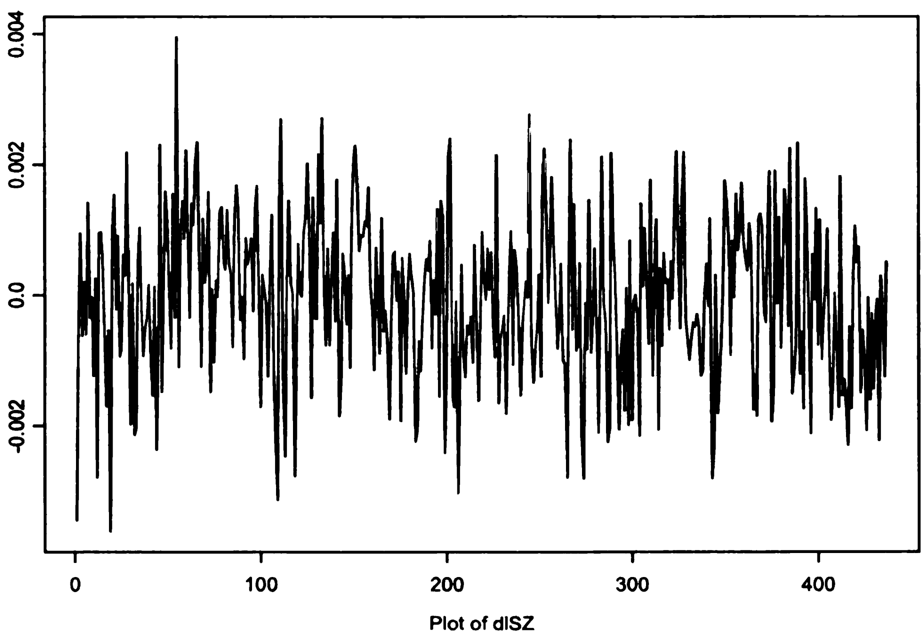


Figure 3.2: QQ-plot of dlUK.

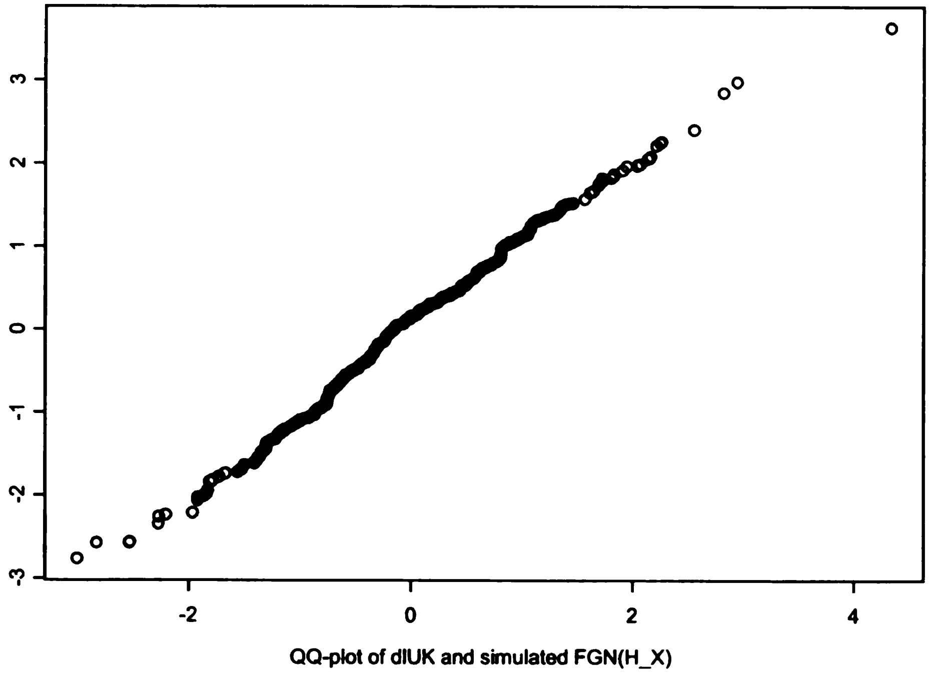


Figure 3.3: Regression estimation of $r(x)$.

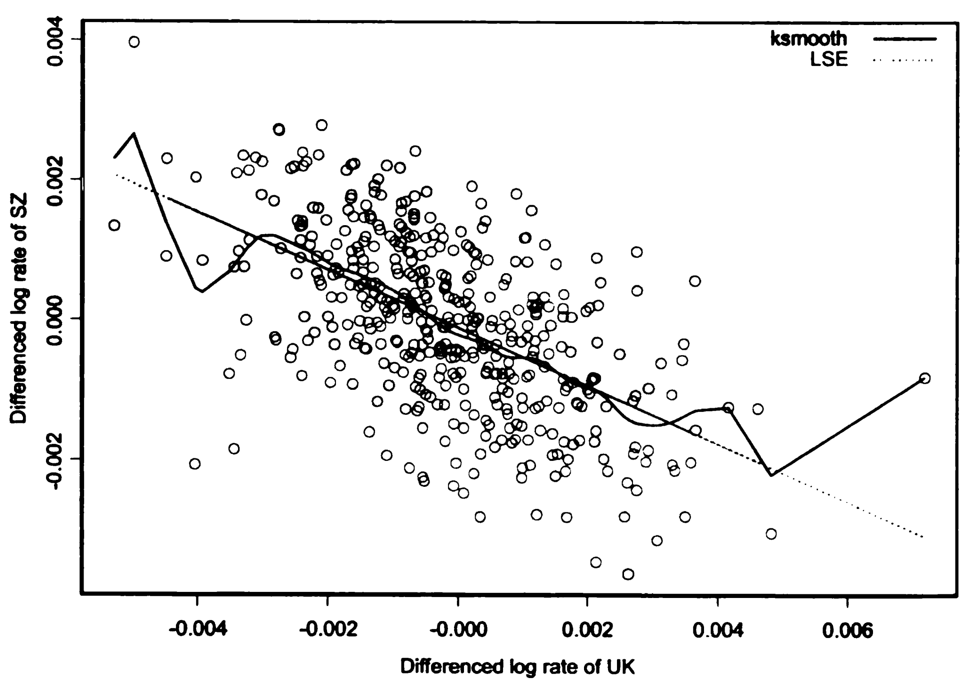
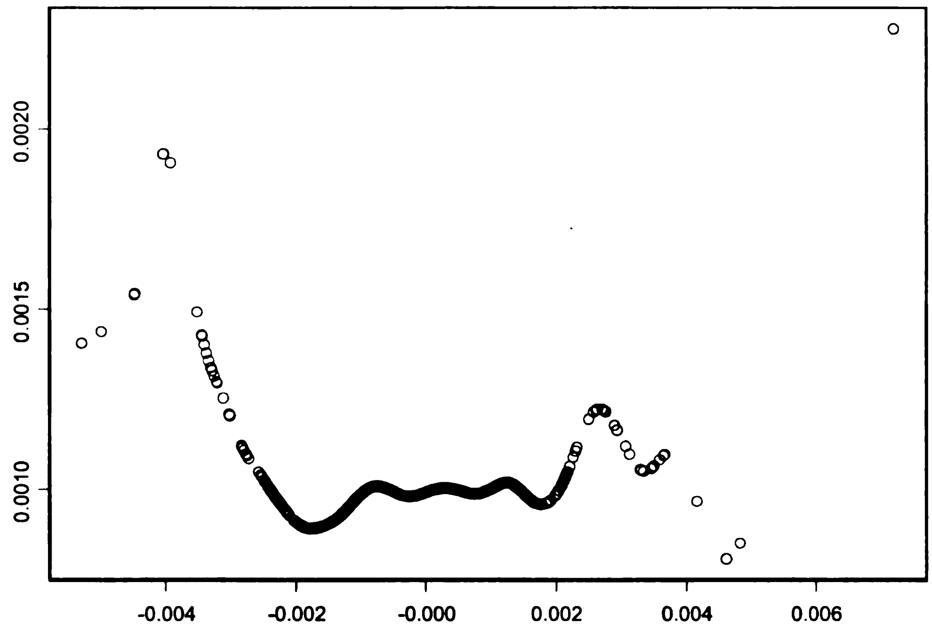


Figure 3.4: Kernel estimation of $\sigma(x)$.



Next, we regressed Y on X, using non-parametric kernel regression estimator and parametric simple linear regression model of Y upon X. Both of these estimates are depicted in Figure 3.3. They display a negative association between X and Y. The estimated linear equation is $\hat{Y} = -0.000118775 - 0.4141107X$, with a residual standard error of 0.00102992. Figure 3.4 provides the nonparametric kernel estimator of $\sigma(x)$ when regressing Y on X. The kernel function $K(x) = .5(1+\cos(x\pi))I(|x| \leq 1)$.

The estimators of the parameter H based on the residuals $\hat{\varepsilon} = Y - \hat{\beta}X$ and $\hat{u} = (Y - \hat{b}X)/\hat{\sigma}(X)$ are equal to 0.6046235 and 0.6246576, respectively. This again suggest the presence of long memory in the error process.

Finally, to check if the regression of Y on X is linear, we obtained $D_n = 0.4137897$ with the asymptotic p-value 66%. As expected, this test fails to reject the null hypothesis that there exists a linear relationship between these two processes.

3.7 Appendix

We first state a result similar to the reduction principle of Taqqu (1975), but with the kernel function $K_b(\cdot)$ involved. Let $\mathcal{G} := \{\nu : E\nu^2(X) < \infty\}$. Recall the Hermite expansion from (3.2.5) above. We have

Lemma 3.7.1 *Let X_t be a stationary Gaussian process. Then,*

$$\begin{aligned} & \frac{1}{n}\sum_{t=1}^{n}\Big(K_b(x-X_t)\sigma^2(X_t)-\tau_b\Big) - \frac{\sigma^2(x)x\phi(x)}{n}\sum_{t=1}^{n}X_t \\ = \ & O_p\Big(\frac{1}{\sqrt{nb}}+\frac{1}{\sqrt{bn^{4-4h}}}\Big), \qquad \forall x \in \mathbb{R}. \end{aligned}$$

Proof. Fix an $x \in \mathbb{R}$. Let $\nu_n(X) := \sqrt{b}[K_b(x-X)\sigma^2(X)-\tau_n]$ with $\tau_n = EK_b(x-X)\sigma^2(X)$. Note that $E\nu_n(X) \equiv 0$, for each $n \geq 1$. Moreover, by the change of variable formula,

$$E\nu_n^2(X) = \int K^2(w)\sigma^2(x-bw)\phi(x-bw)dw.$$

From this one sees that $\sup_{n\geq 1} E\nu_n^2(X) < \infty$, so that $\nu_n(X) \in \mathcal{G}$, for each $n \geq 1$. This in turn implies that $\sup_{n\geq 1}\sum_{j=1}^{\infty} c_{nj}^2/j! < \infty$, where

$$\begin{aligned} c_{nj} &:= \sqrt{b}\int K_b(x-y)\sigma^2(y)H_j(y)\phi(y)dy \\ &= \sqrt{b}\int K(w)\sigma^2(x-wb)H_j(x-wb)\phi(x-wb)dw \\ &= \sqrt{b}\{H_j(x)\sigma^2(x)\phi(x)+o(1)\}, \qquad \forall j \geq 1. \end{aligned}$$

Hence,

$$E\Big(\frac{1}{n}\sum_{t=1}^{n}\nu_n(X_t) - c_1\frac{1}{n}\sum_{t=1}^{n}X_t\Big)^2$$

$$= n^{-2}\sum_{s=1}^{n}\sum_{t=1}^{n}\sum_{j=2}^{\infty}\frac{c_j^2}{(j!)^2}EH_j(X_s)H_j(X_t)$$

$$\le \sum_{j=2}^{\infty}\frac{c_j^2}{j!}\Big\{\frac{1}{n}+n^{-2}\sum_{t=1}^{n}\sum_{s\neq t}\gamma_X^2(|t-s|)\Big\} \le C(\frac{1}{n}+n^{4h-4}).$$

This proves the lemma. □

Under the Assumption 3.5, a similar argument yields the following.

$$E\Big\{\frac{1}{n}\sum_{t=1}^{n}\Big(K_b(x-X_t)\sigma^2(X_t)-\tau_b\Big)u_t\Big\}^2 = O(\frac{1}{nb})+o(\frac{1}{n^{2-2H}}). \tag{3.7.1}$$

$$E\Big\{\frac{1}{n}\sum_{t=1}^{n}\Big(K_b(x-X_t)\sigma^2(X_t)-\tau_b\Big)(u_t^2-1)\Big\}^2 = O(\frac{1}{nb})+o(\frac{1}{n^{4-4H}}). \tag{3.7.2}$$

Now let ν be an arbitrary function such that $E\nu(X) = 0$, $E\nu^2(X) = 1$. Let $\xi_t := \nu(X_t)$. For simplicity of the exposition, let ρ_k now stand for $\gamma_X(k)$ of (3.1.3), and $r \ge 1$ be the Hermite rank of $\nu(X)$. Then, the auto-covariance function of the process ξ is

$$\gamma_\xi(k) = \rho_k^r\Big(\frac{c_r^2}{r!}+\sum_{j=r}^{\infty}\frac{c_{j+1}^2}{(j+1)!}\rho_k^{j+1-r}\Big).$$

Because ρ_k is decreasing in k, we readily obtain that $\gamma_\xi(k)$ is also decreasing in k. Moreover, the second factor in $\gamma_\xi(k)$ is bounded above by $\sum_{j\ge1}c_j^2/j! = 1$. Therefore, there exists a constant $C = C(r, G)$ free of k, such that

$$\gamma_\xi(k) \sim C\rho_k^r. \tag{3.7.3}$$

Lemma 3.7.2 *Let $\xi_t = \nu(X_t)$, u_t be defined as in (3.1.2), and $I_{\xi u}$ denote the periodogram of $\xi_t u_t$. Then, as $\lambda \to 0$,*

$$\begin{aligned} EI_{\xi u}(\lambda) &= O(\lambda^{r(2-2h)+1-2H}), && 0 < r(2-2h) + (2-2H) < 1; \\ &= O(\log n), && r(2-2h) + (2-2H) \geq 1. \end{aligned}$$

Proof. For the first case we use the Corollary 4.10.2 of Bingham et al (1987) which says that for any $0 < \alpha < 1$,

$$\sum_{t=1}^{n} t^{-\alpha} e^{it\lambda} \to \lambda^{\alpha-1}\Gamma(1-\alpha)\Big(\sin\frac{\pi}{2}\alpha + i\cos\frac{\pi}{2}\alpha\Big), \quad \lambda \to 0. \tag{3.7.4}$$

Using the fact that $\gamma_u(k) \sim Ck^{2H-2}$, (see (3.1.4)), (3.7.3), (3.7.4) with $0 < \alpha = r(2-2h) + (2-2H) < 1$, and the independence of ξ_t and u_t, we obtain

$$\begin{aligned} EI_{\xi u}(\lambda) &= \frac{1}{2\pi n}\sum_{j=1}^{n}\sum_{k=1}^{n} \gamma_\xi(j-k)\gamma_u(j-k)e^{i(j-k)\lambda} \\ &\sim C\frac{1}{n}\sum_{k=1}^{n}\sum_{t=k-n}^{k-1} t^{-[r(2-2h)+(2-2H)]}e^{it\lambda} \\ &\leq C\lambda^{r(2-2h)+1-2H}, \quad \lambda \to 0. \end{aligned}$$

Otherwise, when $r(2-2h) + (2-2H) \geq 1$, $EI_{\xi u}(\lambda) \leq C\log n$ for all $\lambda \in [-\pi, \pi]$. □

The following lemma in Dalla, Giraitis and Hidalgo (2006) is needed for the proof of Theorem 3.4.1:

Lemma 3.7.3 *Under (3.1.2) and when $m = o(n)$, then uniformly in $0 < v \leq 1$,*

$$[vm]^{-1}\sum_{j=1}^{[vm]} \frac{I_u(\lambda_j)}{f_j} \to 1, \quad a.s., \quad m \to \infty. \tag{3.7.5}$$

The next lemma approximates the averages of certain covariances of a square integrable function of a Gaussian vector by the corresponding average where the components of the Gaussian r.v. are i.i.d. Accordingly, let E^0 denote the expectation when a

Gaussian random vector is standard normal random vector. Let $A_{0,s,t}$ be the covariance matrix of X_0, X_s, X_t, $0 \le s \le t$, and $B_{0,s,t} = A_{0,s,t} - I_3 = ((b_{i,j}(s,t)))$, where I_3 is the 3×3 identity matrix. Let $\varrho_{s,t}$ denote the largest eigne value of $B_{0,s,t}$. From Luenberger (1979) (Ch. 6.2, page 194) we obtain that $\varrho_{s,t} \le \max_i \sum_{j=1}^{3} |b_{i,j}(s,t)|$. This in turn implies that

$$\varrho_{s,t} \le 3\gamma_X(t-s) \vee \gamma_X(s) \vee \gamma_X(t). \tag{3.7.6}$$

We are now ready to state and prove the

Lemma 3.7.4 *Let g be a function defined on $\mathbb{R}^3$ such that $Eg^2(X_0, X_1, X_2)$ and $E^0 g^2(X_0, X_1, X_2)$ are finite. Then, uniformly in $i = 1, \cdots, n$,*

$$\frac{1}{(n-1)^2} \sum_{t \neq i} \sum_{s \neq i} \Big(Eg(X_i, X_s, X_t) - E^0 g(X_i, X_s, X_t) \Big) \to 0.$$

Proof. We use Theorem 2.1 of Soulier (2001). By the definition, the function $g(x,y,z) - E^0 g(X_i, X_s, X_t)$ has the Hermite rank $r \ge 1$. For sufficiently large $|s-i|, |t-i|$ and $|t-s|$, by Theorem 2.1 of Soulier (2001), there is a constant $C < \infty$ free of i, t, s, such that

$$\Big| Eg(X_i, X_s, X_t) - E^0 g(X_i, X_s, X_t) \Big| \le C ||g|| \varrho_{|s-i|,|t-i|}^{r/2}. \tag{3.7.7}$$

Then the lemma follows from (3.7.6) and (3.1.3). □

By a similar argument, we can obtain that there exists C free of s, t, such that

$$|Eg(X_s, X_t) - E^0 g(X_s, X_t)| \le C ||g|| \gamma_X^{r/2}(t-s). \tag{3.7.8}$$

BIBILIOGRAPHY

Abry, P., Flandrin, P., Taqqu, M.S. and Veitch, D. (2002). Self-similarity and long-range dependence through the wavelet lens, *Theory and Applications of Long-Range Dependence*, Doukhan, P., Oppenheim, G. and Taqqu, M.S. (Eds.) Birkhäuser. 527-556.

Andrews, D. (1995). Nonparametric kernel estimation for semiparametric models. *Econometric Theory* **11** 560-596.

Arcones, M. (1994). Limit Theorems for nonlinear functionals of a stationary Gaussian sequence of vectors. *Ann. Probab.* **22** 2242-2274.

Avram, F. and Taqqu, M. (1987). Noncentral limit theorems and Appell polynomials, *Ann. Probab.* **15** No.2,767-775.

Baillie, R.T. (1996). Long memory processes and fractional integration in econometrics. *Journal of Econometrics,* **73**, 5-59.

Baillie, R.T. and T. Bollerslev.(1994). The long memory of the forward premium. *Journal of International Money and Finance* **13** 309-324.

Baillie, R.T. and T. Bollerslev. (2000). The forward premium anomaly is not as bad as you think. *Journal of International Money and Finance* **19** 471-488.

Beran, J. (1992). Statistical methods for data with long range dependence. *Statist. Science*, **7**, 404-427.

Beran, J. (1994). *Statistics for long-memory processes.* Monographs on Statistics and Applied Probability, **61**. Chapman and Hall, New York.

Berman, S. (1992). *Sojourns and extremes of stochastic processes*, Wadsworth and Brooks/ Cole, Pacific Grove, CA.

Bingham, N., Goldie, C. and Teugels, J.(1987). *Regular Variation.* Cambridge University Press.

Brillinger, D. R. (1975). *Time Series, Data Analysis and Theory*, Holt, Rinehart and Winston, INC.

Brockwell, P. and Davis, R. (1991). *Time Series: Theory and Methods*, Springer-Verlag.

Cheung, Y.-W. (1993). Long memory in foreign exchange rates. *Journal of Business and Economic Statistics* **11** 93-101.

Csörgő, S. and Mielniczuk, J. (1995a). Density estimation under long-range dependence. *Ann. Statist.* **23** 990-999.

Csörgö, S. and Mielniczuk, J. (1995b). Nonparametric regressin under long-range dependent normal errors, *Ann, Statist.* **23** 1000-1014.

Csörgö, S. and Mielniczuk, J. (1995c). Distant long-range dependent sums and regression estimation. *Stoch. Proc. Appl.* **59** 143-155.

Csörgő, S. and Mielniczuk, J. (2000). The smoothing dichotomy in random-design regression with long-memory errors based on moving averages. *Statistica Sinica* **10** 771-787.

Dahlhaus, R. (1995). Efficient location and regression estimation for long range dependent regression models. *Ann. Statist.* **23** 1029–1047.

Dalla,V., Giraitis, L. and Hidalgo, J. (2005). Consistent estimation of the memory parameter for nonlinear time series, *J. Time Series Analysis.* **27** 211-251.

Davydov,Y. (1970). The invariance principal for stationary processes, *Theory Probab. Appl.* **15** 487-498.

Dehling, H., Mikosch, T. and Sørensen, M. (2002), Eds. *Empirical process techniques for dependent data.* Birkhäuser, Boston.

Dobrushin, R.L. and Major, P. (1979). Non-central limit theorems for non-linear functionals of Gaussian fields. *Z. Wahr. verw. Geb.*, **50**, 27-52.

Doukhan, P. (1994). *Mixing: Properties and Examples.* **Lecture notes in Statist. 85**. Springer, New York.

Doukhan, P., Oppenheim, G. and Taqqu, M.S., Eds. (2003). *Theory and applications of long-range dependence.* Birkhäuser, Boston.

Fan, J. and Yao, Q. (1998). Efficient estimation of conditional variance functions in stochastic regression,*Biometrika* **85** No. 3, 645-660.

Fox, R. and Taqqu, M.(1987). Multiple stochastic integrals with dependent integrators. *Journal of Multivariate Analysis* **21** 105-127.

Granger, C.W.J. and Joyeux, R. (1980). An introduction to long memory time series models and fractional differencing, *Journal of Time Series Analysis*, **1**, 15-39.

Gorodetskii, V.V. (1977). On the convergence to semi-stable Gaussian processes. *Th. Probab. Appl.*, **22**, 498-508.

Giraitis, L. (1985). Central limit theorem for functionals of a linear process, *Lithuanian Mathematical Journal* **25** 25-35.

Giraitis, L., Koul, H. and Surgailis, D. (1996). Asymptotic normality of regression estimators with long memory errors. *Statistic & Probability Letters* **29** 317-335.

Hall, P. and Hart, J. (1990). Nonparametric regression with long-range dependence, *Stoch. Proc. Appl.* **36** 339-351.

Härdle, W. (1990) *Applied nonparametric regression.* Cambridge [England] ; New York : Cambridge University Press.

Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (1997). *Wavelets, Approximation and Statistical Applications*, Seminar Berlin-Paris.

Henry, M. and Robinson, P. (1995). Bandwidth choice in Gaussian semiparametric estimation of long range dependence. *Athens Conference on Applied Probability and Time Series Analysis, Vol. II, (1995)*, 220–232, **Lecture Notes in Statist., 115** Springer, New York.

Hidalgo, J. and Robinson, P. (2002). Adapting to unknown disturbance autocorrelation in regression with long memory. *Econometrica* **70** 1545-1581.

Hipel, K.W. and McLeod, A.I. (1978). Preservation of the rescaled adjusted range, 3: Fractional Gaussian noise algorithms, *Water Resources Research* **14** 509-516.

Ho, H. and Hsu, N. (2005). Polynomial trend regression with long memory errors. *Journal of Time Series Analysis* **26** 323-354.

Hosking, J. (1981). Fractional differencing, *Biometrika* **68**, 165-176.

Hosking, J. (1996). Asymptotic distrubutions of the sample mean, autocovariatnces, and autocorrelations of long-memory time series, *Journal of Ecomometrics*, **73** 261-284.

Hurst, H.E. (1951). Long-term storage capacity of reserviors, *Transactions of the American Society of Civil Engineers* **116**, 770-799.

Koul, H.L.(1992). M-estimators in linear models with long range dependent errors. *Statistic & Probability Letters* **14** 153-164.

Koul, H.L. and Mukherjee, K. (1993). Asymptotics of R-, MD- and LAD- estimators in linear regression models with long range dependent errors. *Probab. Theory and Relat. Fields* **95** 535-553.

Koul, H.L. (1996). Asymptotics of M-estimators in non-linear regression with long-range dependent errors. *Athens Conference on Applied Probability and Time*

Series Analysis, Vol. II, (1995), 272–290, **Lecture Notes in Statist.**, **115** Springer, New York.

Koul, H., Baillie, R. and Surgailis, D. (2004). Regression model fitting with a long memory covariate process. *Econometric Theory* **20** 485-512.

Künsch, H. (1987). Statistical aspects of self-similar processes. In *Proc. First World Congress of the Bernoulli Society* VNU Science Press. 67-74.

Li, F. (2006) Testing for the equality of two nonparametric regression curves with long memory errors. *Communications in Statistics: Simulation & Computation*, **35**, (forthcoming).

Luenberger, L. (1979). *Introduction to dynamical systems: theory and applications.* Wiley.

Maxwell, M. and Woodroofe, M. (2000). Central limit theorems for additive functionals of Markov chains,*Ann. Probab.* **28** 713-724.

Maynard, A. and P.C.B. Phillips. (2001). Rethinking an old empirical puzzle: Econometric evidence on the forward discount anomaly. *Journal of Applied Econometrics* **16** 671-708.

Mandelbrot, B.B. and and Van Ness, J.W. (1968). Fractional Brownian motions, fractional Brownian noises and applications, SIAM Review 10, 422-437.

Mandelbrot, B.B. and Wallis, J. (1968). Noah, Joseph and operational hydrology, *Water Resources Research* **4**, 909-918.

Mandelbrot, B.B. and Wallis, J. (1969). Computer experiments with fractional Gaussian noises, Parts 1,2,3, *Water Resources Research* **5**, 228-267.

Noakes, D.J., Hipel, K.W., McLeod, A.I., Jimenez, C. and Yakowitz, S. (1988). Forecasting annual geophysical time series, *International Journal of Forecasting* **4**, 103-115.

Robinson, P. M. (1994). Semiparametric analysis of long memory time series. *Ann. Statist.* **22** 515-539.

Robinson, P. M. (1995a). Log-periodogram regression of time series with long range dependence. *Ann. Statist.*, **23** 1048-1072.

Robinson, P. M. (1997). Large-sample inference for nonparametric regression with dependent errors, *Ann. Statist.* **25** No. 5, 2054-2083.

Robinson, P. M. (2003). *Time series with long memory.* Edited by Peter M. Robinson. Advanced Texts in Econometrics. Oxford University Press, Oxford, U.K.

Robinson, P. and Hidalgo, J. (1997). Time series regression with long-range dependence. *Ann. Statist.* **25** 77-104.

Sansone, G. *Orthogonal functions.* Interscience Publsihers, N.Y. 1959.

Sinai, Ya.G. (1976). Self-similar probability distributions *Theory Probab. Appl.*, **21**, 64-80.

Soulier, P. (2001). Moment bounds and central limit theorem for functions of Gaussian vectors. *Statistics & Probability Letters* **54** 193-203.

Surgailis, D. (1982). Zones of attraction of self-similar multiple integrals, *Luthuanian Math. J.* , **22** 327-340.

Szegö, G. (1959). *Orthogonal pplynomials.* American mathematical society Colloquium publ. **XXIII** Rew.ed New-York.

Taqqu, M. (1975). Weak convergence to fractional Brownian motion and the Rosenblatt process. *Z. Wahrsch. Verw. Gebiete* **31** 287-302.

Taqqu, M. (1986). A bibliographical guide to self-similar processes and long-range dependence *Dependence in probability and statistics. E. Eberlein and M.S. Taqqu (eds.), Birkhüser, Boston* pp. 137-165.

Taqqu,M. and Teverovsky,V. (1997). Robustness of Whittle-type estimators for time series with long-range dependence, *Commun. Statist-Stochastic Models* **13** 723-757.

Uspensky, J.V. (1937).*Introduction to mathematical probability*, New York, London, McGraw-Hill book company, inc.

Willinger, W., Paxson, V. and Taqqu, M. (1998). Self-similarity and Heavy Tails: Structural Modeling of Network Traffic. In *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*, Adler, R., Feldman, R., and Taqqu, M.S., editors, Birkhauser.

Wu, W. (2002). Central limit theorems for functionals of linear processes and their applications, *Statistica Sinica* **12** 635-649.

Yajima, Y. (1988). On estimation of a regression model with long-memory stationary errors. *Ann. Statist.* **16** 791-807.

Yajima, Y. (1991). Asymptotic properties of the LSE in a regression model with long-memory stationary errors. *Ann. Statist.* **19** 158–177.

Zygmund, A. (1968). *Trigonometric Series*, Cambridge At the University Press.