The Estimation and Testing of Persistence in Nonlinear and Cyclical Time Series

Violetta Dalla

London School of Economics

July 2006

A dissertation submitted in partial fulfilment of the requirements of the degree of Doctor of Philosophy at the University of London

Declaration

I hereby declare:

- No part of this doctoral dissertation has been presented to any University for any degree.
- Parts of Chapters 2, 3 and 5 were undertaken as joint work with Professor Javier Hidalgo and Dr Liudas Giraitis.

Violetta Dalla

Abstract

Throughout this thesis, we are concerned with filling some of the gaps in the literature concerning parametric and semiparametric Whittle estimation of long-run and/or cyclical persistence in economic time series. In Chapter 2, we consider local Whittle estimation, and without relying on the assumption of a linear model, we establish sufficient conditions for consistency and provide expansions and rate of convergence for the estimator. In Chapter 3, we apply the results of Chapter 2 to examine the local Whittle estimator for the signal plus noise model and some special cases of it: structural model, nonlinear transformations of a Gaussian process, and long memory stochastic volatility model. Under these specifications, we establish the asymptotic properties of the estimator, and raise several issues concerning its rate of convergence and finite sample bias. In Chapter 4, we employ Monte-Carlo simulations to investigate the finite sample properties of the local Whittle estimator under the linear and nonlinear specifications of Chapters 2 and 3. Furthermore, we apply local Whittle estimation to expected and realized inflation rates, nominal and real interest rates, and transformations of foreign exchange rate returns, in order to assess their long-run persistence and address several issues that have appeared in the empirical literature. Finally, Chapter 5 presents two testing procedures, based on the parametric Whittle method, for the null hypothesis of no persistent component in the data. We derive the asymptotic properties of our test statistics, and moreover introduce and validate a bootstrap scheme for calculating their critical values. A Monte-Carlo study of the finite sample performance of our testing procedures, and an empirical application on the growth rate of industrial production and unemployment rate are also included.

Acknowledgements

I would like to thank my supervisor, Professor Javier Hidalgo for his constant presence, participation and advice during these years. I will be permanently indepted to him for his generous and invaluable help. I would also like to thank Liudas Giraitis for his advice and assistance. Thanks also to Afonso Gonçalves da Silva for his support, and to all my friends and LSE colleagues. Financial assistance through the Greek State Scholarships Foundation (IKY), the ESRC grants R000238212 and R000239936, and the Dennis Sargan Memorial Fund is also gratefully acknowledged. Finally, I would like to thank my parents, Agisilaos and Leoni, my sister, Maria, my grandmother, Marika, and Julian for their encouragement, support and belief.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 The notion of persistence

The persistence property of a time series, defined as the degree of dependence between observations in time, is of undoubted interest for several reasons. First, the degree of persistence gives the practitioner an indication of the existence and strength of mean reversion, and as a by-product, of the sensitivity to shocks of the time series under consideration. Second, a correct understanding of the degree of persistence is a crucial step towards building an appropriate model for the dynamics governing the data. Last, but not least, prior knowledge of the level of persistence is essential for performing correct statistical inference, as different degrees of persistence may give rise to different distributional properties of the same test statistic. Two broad types of persistence are the main focus of this thesis, which we refer to as long-run persistence and cyclical persistence. The former relates to the dependence of observations that are far apart in time, while the latter is concerned with the dependence of observations in the same phase of a cycle.

Numerous empirical studies have found evidence of long-run persistence in macroeconomic and financial time series. It was first pointed out by Granger (1966) that various economic times series, such as industrial production and commodity prices indexes, exhibit strong long-run persistence. Such behaviour has been consequently reported by various authors using different approaches, sample periods and transformations of the data. Among others, we cite Greene and Fielitz (1977) for stock returns, Nelson and Plosser (1982) for different measures of output, wages, industrial production, employment, prices, money stock, stock prices and interest rates, Diebold and Rudebusch (1989) for output, Diebold and Rudebusch (1991) for various measures of income, Sowell (1992) for output, Ding, Granger, and Engle (1993) for the S&P500 series, Backus and Zin (1993) for inflation rate, interest rate and money growth, Cheung (1993) for various exchange rates, Ding and Granger (1996) for various stock returns and exchange rates, Baillie, Chung, and Tieslau (1996) for inflation rate, Andersen and Bollerslev (1997) for exchange rates, Gil-Alaña and Robinson (1997) for output, industrial production, employment, different measures of prices, wages, money stock, velocity, bond yield and stock prices, Lobato and Robinson (1998) for various exchange rates, Lobato and Velasco (2000) for the stock market trading volume, Sun and Phillips (2004) for nominal and real interest rates, inflation and expected inflation rates. In the aforementioned studies, there is an overall agreement that the long-run persistence of the various series examined is strong.

On the other hand, the empirical literature on cyclical persistence is rather limited and it is usually concerned with the seasonality of the data. Strong seasonal behaviour has been reported by Arteche and Robinson (2000) for inflation, and by Arteche (2004) for stock index. Nonseasonal cyclical pattern is evident in various macroeconomic time series and is attributed to business cycle behaviour, see King and Watson (1996) for output growth, employment growth, real balance growth, money supply growth, inflation rate, nominal and real interest rate. The persistence of the business cycle component however was not quantified by King and Watson (1996), although the theoretical business cycle literature emphasizes that the business cycle behaviour is strongly persistent, as deviations from the average level of economic activity are maintained for considerable lengths of time, see for example Diebold and Rudebusch (1999).

1.2 Quantifying persistence

Suppose that we are interested in analyzing the persistence properties of a covariance stationary process $\{x_t\}_{t\in\mathbb{Z}}$ with mean μ_x and variance σ_x^2. The main tool for describing dependence in the time domain is the autocovariance function $\{\gamma_x(\tau)\}_{\tau\in\mathbb{Z}}$ given by

$$\gamma_x(\tau) = E\left(\left(x_t - \mu_x\right)\left(x_{t+\tau} - \mu_x\right)\right). \tag{1.2.1}$$

In this thesis, we focus on the frequency domain approach. To that end, we assume further that $\{x_t\}_{t\in\mathbb{Z}}$ has an absolutely continuous spectral distribution function, so that the spectral density function $f_x(.)$ of $\{x_t\}_{t\in\mathbb{Z}}$ exists, and it is such that

$$\gamma_x(\tau) = \int_{-\pi}^{\pi} e^{i\tau\lambda} f_x(\lambda) d\lambda, \tag{1.2.2}$$

where $f_x(.)$ is a non-negative, even and periodic function of period 2π when extended beyond the range $(-\pi, \pi]$.

The spectral density function is the main tool in the frequency domain for analyzing dependence. It is essentially the Fourier transform of the autocovariance function and therefore, the spectral density function captures the same information about the structure of $\{x_t\}_{t\in\mathbb{Z}}$ as the autocovariance function. Since the spectral density function records the contribution of the components belonging to a given frequency band to the total variation of the process, the decomposition into long-, medium- and short-run comes more naturally, see for example Chapter 7 in Anderson (1971). Notice that the long-run is associated with low frequency components, while a cycle of period T_x corresponds to the frequency $\omega_x = \frac{2\pi}{T_x}$.

If $\{x_t\}_{t\in\mathbb{Z}}$ were a white noise sequence, then it would not exhibit either long-run or cyclical persistence. Notice that for white noise processes, we have $\gamma_x(\tau) = 0$ for all $\tau \neq 0$, and $f_x(\lambda) = c$ for all $\lambda \in [0, \pi]$ and some $0 < c < \infty$. If $\{x_t\}_{t\in\mathbb{Z}}$ followed a covariance stationary Autoregressive Moving Average model of orders p, q ($ARMA(p, q)$), then it is well know that its dependence would be rather weak, resulting to an autocovariance function that is absolutely summable and thus to a spectral density function such that $0 < f_x(\lambda) < \infty$ for all $\lambda \in [0, \pi]$. White noise sequences and covariance stationary $ARMA(p, q)$ models fall in the class of weakly dependent processes.

Here, the concept of weak dependence corresponds to time series that have an autocovariance function, which is absolutely summable. On the other hand, we refer to a time series process being strongly dependent, when its autocovariance function is not absolutely summable. It should be mentioned that the notion of weak/strong dependence is not always associated with summability properties of second moments. For example, Doukhan (1994) and Nze, Bühlmann, and Doukhan (2002) quantify dependence as a measure based on the covariance between functions of the past and the future. An earlier and similar concept was introduced by McLeish (1975), known as mixingale or general near epoch dependence, which measures how fast the conditional moments converge to the unconditional ones. Stronger concepts are those of strong-mixing, see Rosenblatt (1956), or β-mixing, see Volkonskii and Rozanov (1959), which are based on the variation norm between the joint probability function and the product of their marginal.

1.2.1 Long-run persistence

A common "local" parameterization of the spectral density function for the purpose of quantifying long-run persistence is

$$f_x(\lambda) \sim c_{0,x} |\lambda|^{-\alpha_x}, \quad \text{as } \lambda \to 0, \tag{1.2.3}$$

with $-1 < \alpha_x < 1$ and $0 < c_{0,x} < \infty$. Here, the notation $\sim$ means that the ratio of left and right hand side tends to 1. The parameter α_x is referred to as the memory parameter and it quantifies the degree of long-run persistence of the process $\{x_t\}_{t\in\mathbb{Z}}$. Notice that for $0 < \alpha_x < 1$, the spectral density function is unbounded at zero, so that the variation associated with the zero frequency component, i.e. the long-run, is substantial. Actually, the higher the value of the memory parameter, the more of the variation is explained by the long-run component and hence, the stronger the long-run persistence is. This case is usually referred to as $\{x_t\}_{t\in\mathbb{Z}}$ having long memory. For $\alpha_x = 0$, the spectral density function is bounded and bounded away from zero at the zero frequency. Then, the variation of $\{x_t\}_{t\in\mathbb{Z}}$ explained by the long-run component is not significant, and $\{x_t\}_{t\in\mathbb{Z}}$ is said to have short memory. Hence, white noise sequences and covariance stationary $ARMA(p,q)$ models exhibit short memory. For $-1 < \alpha_x < 0$, the spectral density function is equal to zero at the zero frequency, and it is said that $\{x_t\}_{t\in\mathbb{Z}}$ has negative memory. In empirical applications, such a situation is not commonly found in the levels of the data, but arises when the data has been overdifferenced.

The earliest model satisfying (1.2.3) is the fractional noise introduced by Mandelbrot and van Ness (1968), whose autocovariance function satisfies

$$\gamma_x(\tau) = \frac{1}{2}\left(|\tau+1|^{\alpha_x+1} - 2|\tau|^{\alpha_x+1} + |\tau-1|^{\alpha_x+1}\right). \tag{1.2.4}$$

The spectral density function of the fractional noise is complicated, see Sinai (1976), but indeed satisfies (1.2.3).

The most widely used parametric model satisfying (1.2.3) is the Autoregressive Fractionally Integrated Moving Average model of orders p, d, q ($ARFIMA(p,d,q)$) introduced by Adenstedt (1974) and explored by Granger and Joyeux (1980). The latter specification is an extension of the Autoregressive Integrated Moving Average of orders p, d, q ($ARIMA(p,d,q)$) model of Box and Jenkins (1976) with the parameter d allowed to take noninteger values. Under such a specification, $\{x_t\}_{t\in\mathbb{Z}}$ is given by

$$a(L)(1-L)^d x_t = b(L)\varepsilon_t, \tag{1.2.5}$$

where $-\frac{1}{2} < d < \frac{1}{2}$ is referred to as the differencing parameter, L is the lag operator, $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a white noise sequence, and $a(L)$ and $b(L)$ are the autoregressive and moving average polynomials

$$a(L) = 1 - a_1 L - ... - a_p L^p \quad \text{and} \quad b(L) = 1 + b_1 L + ... + b_q L_q, \tag{1.2.6}$$

respectively, all of whose zeros lie outside the unit circle and $a(L)$ and $b(L)$ have no common zeros. Then, the spectral density function is given by

$$f_x(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi}\left|1-e^{-i\lambda}\right|^{-2d}\frac{\left|b\left(e^{-i\lambda}\right)\right|^2}{\left|a\left(e^{-i\lambda}\right)\right|^2}, \quad -\pi < \lambda \leq \pi. \tag{1.2.7}$$

It can be easily shown that (1.2.7) satisfies (1.2.3) with $\alpha_x = 2d$, noticing that $\left|1-e^{-i\lambda}\right|^{-2d} = \left|2\sin\frac{\lambda}{2}\right|^{-2d}$ and $\sin\lambda \sim \lambda$, as $\lambda \to 0$.

Finally, we should mention the extension of the exponential model of Bloomfield (1973) considered by Robinson (1994a), whose spectral density function is given by

$$f_x(\lambda) = \frac{1}{2\pi}\left|1-e^{-i\lambda}\right|^{-2d}\exp\left(\sum_{k=1}^{p-1} c_k \cos\left((k-1)\lambda\right)\right), \quad -\pi < \lambda \leq \pi. \tag{1.2.8}$$

Then, one can easily check, as in (1.2.7), that (1.2.8) satisfies (1.2.3) with $\alpha_x = 2d$.

1.2.2 Cyclical persistence

Cyclical persistence at a known frequency $\omega_x \neq \{0,\pi\}$ is just an extension of the long-run one described in Subsection 1.2.1, as now one needs to be concerned with the frequency ω_x instead of the zero one. A "local" parameterization of the spectral density function for quantifying cyclical persistence is

$$f_x(\lambda) \sim c_{0,\omega,x}\left|\lambda - \omega_x\right|^{-\alpha_{\omega,x}}, \quad \text{as } \lambda \to \omega_x, \tag{1.2.9}$$

with $-1 < \alpha_{\omega,x} < 1$ and $0 < c_{0,\omega,x} < \infty$. We refer to the parameter $\alpha_{\omega,x}$ as the cyclical memory parameter, which quantifies the strength of the cyclical component. As in the case of long-run persistence, when $0 < \alpha_{\omega,x} < 1$, the spectral density function is unbounded at the frequency ω_x. Then, a substantial amount of the variation of $\{x_t\}_{t\in\mathbb{Z}}$ is associated with the cyclical component, and the higher the value of $\alpha_{\omega,x}$ the stronger the cyclical component is. Hence, for $0 < \alpha_{\omega,x} < 1$

we say that $\{x_t\}_{t\in\mathbb{Z}}$ has long cyclical memory. For $\alpha_{\omega,x} = 0$, the spectral density function is bounded and bounded away from zero at the frequency ω_x, so that the cyclical component of $\{x_t\}_{t\in\mathbb{Z}}$ is weak. We refer to this case as $\{x_t\}_{t\in\mathbb{Z}}$ having short cyclical memory, and notice that white noise sequences and covariance stationary $ARMA(p,q)$ models fall in this category. When $-1 < \alpha_{\omega,x} < 0$, the spectral density function is equal to zero at the frequency ω_x, and it is said that $\{x_t\}_{t\in\mathbb{Z}}$ has negative cyclical memory. The latter situation is rather uncommon in practical situations, but might arise if the data has been seasonally overdifferenced, that is, a procedure has been applied to the data to extract the seasonal component but the initial strength of the seasonal component had been overestimated.

A parametric model satisfying (1.2.9) is the Gegenbauer Autoregressive Moving Average model of orders p, d_ω, q ($GARMA(p, d_\omega, q)$) of Gray, Zhang, and Woodward (1989). Under such specification, $\{x_t\}_{t\in\mathbb{Z}}$ is given by

$$a(L)(1 - 2\cos(\omega_x) L + L^2)^{d_\omega} x_t = b(L)\varepsilon_t, \tag{1.2.10}$$

where we refer to $-\frac{1}{2} < d_\omega < \frac{1}{2}$ as the cyclical differencing parameter, and $a(L)$, $b(L)$ and $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ are as defined above in (1.2.5). Then, the spectral density function is given by

$$f_x(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi} \left|1 - 2\cos(\omega_x)e^{-i\lambda} + e^{-2i\lambda}\right|^{-2d_\omega} \frac{\left|b\left(e^{-i\lambda}\right)\right|^2}{\left|a\left(e^{-i\lambda}\right)\right|^2}, \quad -\pi < \lambda \leq \pi. \tag{1.2.11}$$

It can be easily shown that (1.2.11) satisfies (1.2.9) with $\alpha_{\omega,x} = 2d_\omega$, noticing that $\left|1 - 2\cos(\omega_x)e^{-i\lambda} + e^{-2i\lambda}\right|^{-2d_\omega} = \left|4\sin\frac{\lambda-\omega_x}{2}\sin\frac{\lambda+\omega_x}{2}\right|^{-2d}$ and $\sin\lambda \sim \lambda$, as $\lambda \to 0$.

1.2.3 Comments

Notice that models (1.2.3) and (1.2.5) can be considered as special cases of (1.2.9) and (1.2.10) respectively. Hence, the latter specifications can be combined in the case that the nature of the possibly persistent component is not known. One can regard the following "local" specification for the spectral density function

$$f_x(\lambda) \sim c_{0,\omega,x} \left|\lambda - \omega_x\right|^{-\alpha_{\omega,x}}, \quad \text{as } \lambda \to \omega_x, \tag{1.2.12}$$

with $-1 < \alpha_{\omega,x} < 1$, $0 < c_{0,\omega,x} < \infty$ and $\omega_x \in [0,\pi]$. On the other hand, when we are concerned with parametric modelling, we can consider (1.2.10), that is,

$$a(L)(1 - 2\cos(\omega_x) L + L^2)^{d_\omega} x_t = b(L)\varepsilon_t, \tag{1.2.13}$$

with $-\frac{1}{2} < d_\omega < \frac{1}{2}$ when $\omega_x \in (0, \pi)$ and $-\frac{1}{4} < d_\omega < \frac{1}{4}$ when $\omega_x \in \{0, \pi\}$, while $a(L)$, $b(L)$ and $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ are as defined above in (1.2.5). In both (1.2.12) and (1.2.13), the frequency ω_x is unknown, and therefore, ω_x can be treated as another parameter that needs to be estimated.

Before we proceed with overviewing the methods for estimating persistence, there are three points that we need to add. Firstly, we concentrate on covariance stationary processes, and therefore, we require the parameters α_x and $\alpha_{\omega,x}$ to be less than 1. In practical situations, it is likely that one would encounter data sets with nonstationary characteristics. Here, we assume that appropriate transformations can be applied to the data, so that the resulting series are covariance stationary with α_x and/or $\alpha_{\omega,x}$ less than 1.

Secondly, the specifications presented in Subsections 1.2.1 and 1.2.2 are not the only parametric models satisfying (1.2.3) and (1.2.9), respectively. For example, (1.2.5) satisfies (1.2.9) with $\alpha_{\omega,x} = 0$, while (1.2.3) holds for (1.2.10) with $\alpha_x = 0$. Furthermore, the $ARFIMA(p, d, d_\omega, q)$ model, see Robinson (1994c) and Giraitis and Leipus (1995), is given by

$$a(L)(1 - 2\cos(\omega_x) L + L^2)^{d_\omega}(1 - L)^d x_t = b(L)\varepsilon_t, \tag{1.2.14}$$

where d, d_ω, $a(L)$, $b(L)$ and $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ are as defined above in (1.2.5) and (1.2.10), and satisfies (1.2.3) and (1.2.9) with $\alpha_x = 2d$ and $\alpha_{\omega,x} = 2d_\omega$, respectively. Actually, we show in Chapter 3 below, that various nonlinear models satisfy (1.2.3) and (1.2.9) such as linear combinations of $ARFIMA(p, d, q)$ and/or $GARMA(p, d_\omega, q)$ models, as well as nonlinear transformations of $ARFIMA(p, d, q)$ or $GARMA(p, d_\omega, q)$ models.

Lastly, when examining cyclical components at a frequency $\omega_x \neq \{0, \pi\}$, the specifications (1.2.9), (1.2.10), (1.2.12), (1.2.13), and (1.2.14) entail the assumption that cycles with frequencies just above and below ω_x have the same contribution to the total variation of $\{x_t\}_{t\in\mathbb{Z}}$. One could relax such a restriction as it was done in Arteche and Robinson (2000), however we are not going to consider this case here.

1.3 Estimation of persistence

As described above, the parameters α_x and $\alpha_{\omega,x}$ quantify the long-run and cyclical persistence of $\{x_t\}_{t\in\mathbb{Z}}$, respectively. Suppose now that a stretch of data $\{x_1, ..., x_n\}$ is available, where n denotes the sample size, and we are interested in performing inference on α_x and/or $\alpha_{\omega,x}$. In this section, we overview different methods for their estimation that have been proposed and analyzed. The existing methodologies fall in two categories, parametric and semiparametric.

Parametric methods require specifying, up to a finite set of unknown parameters, the spectral density function over the whole range of frequencies. That is, $f_x(\lambda) = f_x(\lambda; \theta_x)$ for all $\lambda \in [0, \pi]$, and examples include the parametric models of Section 1.2. On the other hand, in semiparametric methods the spectral density function is only locally parameterized around a frequency and it is left otherwise unrestricted, as it was done with the specifications (1.2.3) and (1.2.9). Although parametric estimation is more efficient than semiparametric one under correct model specification, it is likely to suffer from inconsistency if the model has been misspecified.

1.3.1 Parametric methods

In the parametric framework, the most popular approach to estimate the memory parameter α_x, along with any other parameters of the model, is based on the Gaussian log-likelihood. If $\{x_t\}_{t\in\mathbb{Z}}$ were a sequence of Gaussian random variables with zero mean, then it would be natural to consider estimates maximizing the log-likelihood

$$-\frac{1}{n}\log|\Sigma_x(\theta_x)| - \frac{1}{n}x'\Sigma_x^{-1}(\theta_x)\,x, \tag{1.3.1}$$

where $x = (x_1, ..., x_n)'$ and $\Sigma_x(\theta_x)$ is the covariance matrix of x, which has (j,k)-th element equal to $\gamma_x(j-k)$. The maximization of (1.3.1) is taken over a compact set of values for θ_x that guarantee the stationarity of $\{x_t\}_{t\in\mathbb{Z}}$.

It can be shown that

$$\frac{1}{n}\log|\Sigma_x(\theta_x)| \to \frac{1}{2\pi}\int_{-\pi}^{\pi}\log f_x(\lambda;\theta_x)d\lambda, \tag{1.3.2}$$

as $n \to \infty$, see Chapter 3 of Hannan (1970). Moreover, the (j,k)-th element of $\Sigma_x(\theta_x)$ satisfies by definition

$$\gamma_x(j-k) = \int_{-\pi}^{\pi} e^{i(j-k)\lambda} f_x(\lambda;\theta_x)d\lambda, \tag{1.3.3}$$

so that the (j,k)-th element of $\Sigma_x^{-1}(\theta_x)$ can be approximated by

$$s_x(j-k) = \frac{1}{(2\pi)^2}\int_{-\pi}^{\pi} e^{i(j-k)\lambda} f_x^{-1}(\lambda;\theta_x)d\lambda. \tag{1.3.4}$$

Then, the matrix $S_x(\theta_x)$, whose (j,k)-th element is $s_x(j-k)$, approximates $\Sigma_x^{-1}(\theta_x)$. Next, introduce the discrete Fourier transform and periodogram of the data

$$w_x(\lambda) = n^{-\frac{1}{2}} \sum_{t=1}^{n} x_t e^{-it\lambda} \quad \text{and} \quad I_x(\lambda) = \frac{1}{2\pi} |w_x(\lambda)|^2. \tag{1.3.5}$$

Since we can write

$$\frac{1}{n} x' S_x(\theta_x) x = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_x(\lambda)}{f_x(\lambda;\theta_x)} d\lambda, \tag{1.3.6}$$

the objective function (1.3.1) can be approximated by

$$-\frac{1}{2\pi} \int_{-\pi}^{\pi} \log f_x(\lambda;\theta_x) d\lambda - \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_x(\lambda)}{f_x(\lambda;\theta_x)} d\lambda, \tag{1.3.7}$$

and maximizing the criterion (1.3.1) is equivalent to minimizing the objective function

$$\int_{-\pi}^{\pi} \log f_x(\lambda;\theta_x) d\lambda + \int_{-\pi}^{\pi} \frac{I_x(\lambda)}{f_x(\lambda;\theta_x)} d\lambda. \tag{1.3.8}$$

If furthermore it is assumed that $\int_{-\pi}^{\pi} \log f_x(\lambda;\theta_x) d\lambda > -\infty$, then it is well known that $\{x_t\}_{t\in\mathbb{Z}}$ can be written as

$$x_t = \sum_{j=0}^{\infty} \varphi_{j,x} \varepsilon_{t-j,x}, \quad \sum_{j=0}^{\infty} \varphi_{j,x}^2 < \infty, \tag{1.3.9}$$

where $\{\varepsilon_{t,x}\}_{t\in\mathbb{Z}}$ is a sequence of uncorrelated random variables with zero mean and variance $\sigma_{\varepsilon_x}^2$, see Chapter 3 of Hannan (1970). Then, we can parameterize the spectral density function of $\{x_t\}_{t\in\mathbb{Z}}$ as

$$f_x(\lambda;\theta_x) = \frac{\sigma_{\varepsilon_x}^2}{2\pi} k_x(\lambda;\psi_x), \tag{1.3.10}$$

where $\theta_x = \left(\psi_x', \sigma_{\varepsilon_x}^2\right)'$ and $k_x(\lambda;\psi_x) = \left| \sum_{j=0}^{\infty} \varphi_{j,x} e^{-ij\lambda} \right|^2$. If moreover $\int_{-\pi}^{\pi} \log k_x(\lambda;\psi_x) d\lambda = 0$, we have that

$$\int_{-\pi}^{\pi} \log f_x(\lambda;\theta_x) d\lambda = 2\pi \log \sigma_{\varepsilon_x}^2 - 2\pi \log(2\pi), \tag{1.3.11}$$

and then the objective function (1.3.8) is up to a constant proportional to

$$\log \sigma_{\varepsilon_x}^2 + \frac{1}{\sigma_{\varepsilon_x}^2} \int_{-\pi}^{\pi} \frac{I_x(\lambda)}{k_x(\lambda;\psi_x)} d\lambda. \tag{1.3.12}$$

As we are interested in inference on the memory parameter α_x which is precluded in ψ_x, one has to minimize over ψ_x the objective function

$$\int_{-\pi}^{\pi} \frac{I_x(\lambda)}{k_x(\lambda;\psi_x)} d\lambda. \tag{1.3.13}$$

Next, for integer j, denote by $\lambda_j = \frac{2\pi j}{n}$ the j-th Fourier frequency. Due to the symmetry around zero and periodicity of the spectral density function, and replacing the integral with a discrete sum evaluated at the Fourier frequencies, the objective function (1.3.13) can be further approximated by

$$\sum_{j=1}^{n-1} \frac{I_x(\lambda_j)}{k_x(\lambda_j;\psi_x)}. \tag{1.3.14}$$

The objective function (1.3.14) has the advantages that is computationally easier to derive by the means of the fast Fourier transform, and does not require the assumption of a known mean for $\{x_t\}_{t\in\mathbb{Z}}$. The latter is the case since the periodogram $I_x(.)$ evaluated at the Fourier frequencies λ_j, $j = 1,..,n-1$, is invariant to location shift in $\{x_t\}_{t\in\mathbb{Z}}$.

The approximation of the Gaussian log-likelihood, resulting to the objective function (1.3.8) and its subsequent forms (1.3.13) and (1.3.14) are due to Whittle (1951) and are therefore referred to as the Whittle likelihoods, while the resulting estimators as the parametric Whittle (PW) estimators. The first major contribution on the asymptotic properties of the PW estimators came from Hannan (1973). His main condition for consistency is the ergodicity of $\{x_t\}_{t\in\mathbb{Z}}$, which is satisfied for the parametric models presented in Section 1.2. Hannan (1973) also showed that the PW estimators are $n^{\frac{1}{2}}$-consistent and asymptotically normal, however under conditions that rule out strong persistence. The latter properties for the PW estimator based on the objective function (1.3.13) were first established under long memory by Fox and Taqqu (1986) for a Gaussian sequence $\{x_t\}_{t\in\mathbb{Z}}$ following a rather general parametric form. Again under the assumption of Gaussianity and long memory of $\{x_t\}_{t\in\mathbb{Z}}$, Dahlhaus (1989) studied both (1.3.1) and (1.3.8) and showed that the Cramér-Rao efficiency bound is still achieved. For the same PW estimator as in Fox and Taqqu (1986), Giraitis and Surgailis (1990) relaxed the Gaussianity of $\{x_t\}_{t\in\mathbb{Z}}$ to a linear process of the form (1.3.9) with the innovation process $\{\varepsilon_{t,x}\}_{t\in\mathbb{Z}}$ being a sequence of identically and independently distributed (i.i.d.) random variables with finite fourth moments. Hosoya (1997) considered multivariate models and allowed for martingale difference innovation sequence $\{\varepsilon_{t,x}\}_{t\in\mathbb{Z}}$. The case when $\{x_t\}_{t\in\mathbb{Z}}$ has negative memory was recently been examined by Velasco and Robinson (2000) for the PW estimator based on (1.3.14) under conditions similar to those in the case of long memory. Lastly, we should mention the work by Giraitis and Taqqu (1999), where the asymptotic properties of the PW estimator examined by Fox and Taqqu (1986) were established for polynomial transformations of a Gaussian long memory sequences. Giraitis and Taqqu (1999) showed that the PW estimator remains consistent, but the $n^{\frac{1}{2}}$-consistency and asymptotic normality were not found to hold for all the transformations considered there.

The above mentioned references mainly concentrate on the case of a long-run component, but can be easily extended to the case of a cyclical component of known frequency ω_x. The case of a persistent component at an unknown frequency was examined by Giraitis, Hidalgo, and Robinson (2001). The authors treated ω_x as another unknown parameter precluded in ψ_x and examined the PW estimator based on the objective function (1.3.14). Under long memory and conditions similar to those in the case of a know frequency, they established the asymptotic properties of the PW estimator, which were found to be unaffected by the lack of knowledge of ω_x.

As far as performing statistical inference on the memory parameter α_x or the cyclical memory parameter $\alpha_{\omega,x}$, one can construct test statistics based on the PW estimators described above and derive, under appropriate assumptions, the asymptotic distribution of the statistic using the theoretical results of the aforementioned references. Hence, given knowledge of the frequency of the possibly persistent component, one can test for different values of α_x or $\alpha_{\omega,x}$ in $(-1, 1)$. When the frequency is unknown, one needs to assume that $\{x_t\}_{t\in\mathbb{Z}}$ has long memory or cyclical long memory, and then inference on values of α_x or $\alpha_{\omega,x}$ falling in $(0, 1)$ can be performed. However, in the latter case, it is not possible to test whether the data does not exhibit a persistent component against the alternative that it does.

Before we proceed to discuss semiparametric methods, we should add that the PW estimators dominate the literature of parametric methods. Another method for the estimation and testing of the memory parameter α_x, in the context of the $ARFIMA(0, d, 0)$ model, is based on the rescaled range statistic R/S and a modified version of it, see Hurst (1951), Mandelbrot (1975) and Lo (1991). The R/S statistic is the range of partial sums of deviations of a time series from its mean, rescaled by its standard deviation. The R/S statistic is easy to compute, and a transformation of it provides a consistent estimate for α_x. However, the resulting estimator of α_x has an asymptotic distribution that is difficult to compute its quantiles from, and furthermore, has a slow rate of convergence.

1.3.2 Semiparametric methods

Semiparametric estimation and inference on the long memory parameter α_x is based on the specification (1.2.3). They rely on the fact that the spectral density function around 0 can be approximated by $c_{0,x} |\lambda|^{-\alpha_x}$, and therefore use information only from a neighbourhood of the zero frequency, as opposed to parametric methods that take into account the whole band of frequencies.

This idea was first initiated in the work of Geweke and Porter-Hudak (1983), see Robinson (1995a) for a precise treatment, who proposed regressing the log-periodogram ordinates $\log (I_x(\lambda_j))$ on a constant and $\log \lambda_j$, for $j = 1, ..., m$, and estimating α_x by the minus of the estimated slope coefficient. Notice that Robinson (1995a) considered trimming the first $j = 1, ..., l$ log-periodogram ordinates. The integer m is referred to as the bandwidth parameter and is taken to satisfy

$$m \to \infty \quad \text{and} \quad m = o(n), \quad \text{as } n \to \infty, \tag{1.3.15}$$

so that information from a degenerating neighbourhood of the zero frequency are taken and the approximation of $f_x(\lambda_j)$ by $c_{0,x}\lambda_j^{-\alpha_x}$ is valid. The estimator of α_x of Geweke and Porter-Hudak (1983) is commonly referred to as the log-periodogram estimator.

The latter approximation was also employed by Künsch (1987), see also Robinson (1995b), who proposed estimating α_x and $c_{0,x}$ by minimizing, over the stationary range of admissible values, the objective function

$$\frac{1}{m}\sum_{j=1}^{m}\left(\log\left(c_{0,x}\lambda_j^{-\alpha_x}\right) + \frac{I_x(\lambda)}{c_{0,x}\lambda_j^{-\alpha_x}}\right). \tag{1.3.16}$$

Observe that (1.3.16) is a local discretized version of the Whittle likelihood (1.3.8), since instead of evaluating the sum at the Fourier frequencies λ_j, $j = 1, ..., n-1$, only the first m of them are employed so that $f_x(\lambda_j)$ in (1.3.8) can be replaced by $c_{0,x}\lambda_j^{-\alpha_x}$. Concentrating on the memory parameter α_x, it can be easily shown that the estimator of α_x resulting from (1.3.16) is that based on minimizing the objective function

$$\log\left(\frac{1}{m}\sum_{j=1}^{m}\frac{I_x(\lambda_j)}{\lambda_j^{-\alpha_x}}\right) + \frac{1}{m}\sum_{j=1}^{m}\log\left(\lambda_j^{-\alpha_x}\right). \tag{1.3.17}$$

The estimator of Künsch (1987) is usually referred to as the Gaussian semiparametric or local Whittle (LW) estimator. Notice again that the assumption of a known mean for $\{x_t\}_{t\in\mathbb{Z}}$ is not required, since the periodogram $I_x(.)$ evaluated at the Fourier frequencies λ_j, $j = 1, .., n-1$, is invariant to location shift in $\{x_t\}_{t\in\mathbb{Z}}$.

The estimators of Geweke and Porter-Hudak (1983) and Künsch (1987) constitute the most popular semiparametric procedures in the literature. Other methods include the FEXP estimator of Robinson (1994a), the averaged periodogram estimator of Robinson (1994b), the estimators of Hidalgo and Yajima (2003) and Hidalgo (2005), and the exact local Whittle estimator of Shimotsu and Phillips (2005). All these estimation methods, with the exception of Hidalgo (2005), were proposed in relation to the memory parameter α_x. They can be easily extended to allow for the estimation of the cyclical memory parameter $\alpha_{\omega,x}$ if the frequency ω_x is known. In Hidalgo (2005), the frequency ω_x is taken to be unknown and a two stage estimation method is proposed. The first stage of this method involves estimation of the unknown frequency ω_x, while in the second, the parameter α_x or $\alpha_{\omega,x}$ is estimated using a method similar to that of Hidalgo and Yajima (2003). An analogous two stage procedure was also proposed in Hidalgo and Soulier (2004), where the unknown frequency is estimated by the method of Yajima (1996) and the parameter α_x or $\alpha_{\omega,x}$ by the estimator of Geweke and Porter-Hudak (1983). The results of Hidalgo and Soulier (2004) and Hidalgo (2005) show that the asymptotic properties of the estimator of α_x or $\alpha_{\omega,x}$ are unaffected by the first stage estimation of ω_x.

The consistency and asymptotic distribution of the aforementioned estimators are well established under appropriate conditions; in Robinson (1995a) for the log-periodogram estimator of Geweke and Porter-Hudak (1983), in Robinson (1995b) for the LW estimator of Künsch (1987), in Moulines and Soulier (1999) and Hurvich and Brodsky (2001) for the FEXP estimator of Robinson (1994a), in Robinson (1994b) and Lobato and Robinson (1996) for the averaged periodogram estimator of Robinson (1994b), and in Hidalgo and Yajima (2003), Hidalgo (2005) and Shimotsu and Phillips (2005) for the corresponding estimators. Hence, the estimation methods of Geweke and Porter-Hudak (1983), Künsch (1987), Robinson (1994b), Hidalgo and Yajima (2003) and, Shimotsu and Phillips (2005) and their asymptotic properties can serve as the basis for constructing statistical procedures for testing for specific values of α_x or $\alpha_{\omega,x}$ in $(-1, 1)$, assuming that frequency of the possibly persistent component is known. If latter is not the case, then the method of Hidalgo and Soulier (2004) or Hidalgo (2005) can be used to construct a statistic technique in order to test for specific values of α_x or $\alpha_{\omega,x}$ in $(0, 1)$. Finally, a test statistic for whether the data does not exhibit a persistent component against the alternative that it does, was provided and examined by Hidalgo (2006).

With few exceptions, the asymptotic properties of the above mentioned semiparametric estimators and test procedures are established under the assumption of Gaussianity or linearity of the process $\{x_t\}_{t\in\mathbb{Z}}$. Notice that by the term linearity, we refer to processes satisfying (1.3.9) with innovation sequence $\{\varepsilon_{t,x}\}_{t\in\mathbb{Z}}$ being a martingale difference satisfying mild conditions, see in more detail Assumption A.5 in Chapter 2. The asymptotic properties of the estimator of Geweke and Porter-Hudak (1983) and the LW estimator of Künsch (1987) were examined for a particular nonlinear model, the sum of a long memory Gaussian or linear process with that of a white noise or i.i.d. or short memory linear sequence, see Deo and Hurvich (2001) and Sun and Phillips (2003) for the estimator of Geweke and Porter-Hudak (1983) and, Arteche (2004) and Hurvich, Moulines, and Soulier (2005) for the LW estimator of Künsch (1987). However, besides these cases, nothing is known about the asymptotic properties of the various semiparametric estimators for nonlinear models.

1.4 Description of the thesis

Throughout this thesis, we are concerned with filling some of the gaps in the literature concerning the estimation and inference of long-run and/or cyclical persistence discussed at the end of Subsections 1.3.1 and 1.3.2. We concentrate on parametric Whittle and local Whittle methods due to their popularity, efficiency and theoretical tractability.

In Chapter 2, we consider the LW estimator of the memory parameter α_x. We establish general conditions that are sufficient for consistency and provide expansions and rate of convergence for the estimator, without relying on the assumption of linearity of the data generating process. As an illustration, we apply our results to the case of a linear process and reaffirm the results obtained by Robinson (1995b).

The practicability of the results in Chapter 2 is demonstrated in Chapter 3. In this chapter, we apply our general results of Chapter 2 in order to assert the asymptotic properties of the LW estimator for nonlinear models. We examine the signal plus noise model and some special cases of it: structural model, nonlinear transformations of a Gaussian process, and long memory stochastic volatility model. Under these specifications we discover that the asymptotic properties of the LW estimator, consistency and asymptotic normality, are unaffected by the presence of the nonlinearity. However, we also find that the rate of convergence and finite sample bias of the LW estimator are worse off as compared to the case of a linear process.

Chapter 4 examines, by the means of Monte-Carlo simulations, the finite sample properties of the LW estimator for the linear and nonlinear specifications considered in Chapters 2 and 3. We find that the consistency property of the estimator is not affected by the presence of nonlinearity. However, we discover that the finite sample properties are worse off as compared to the case of a linear process. Furthermore, we apply LW estimation to real data, expected and realized inflation rates, nominal and real interest rates, and transformations of foreign exchange rate returns, in order to assess their long-run persistence and address several issues that have appeared in the empirical literature.

Finally, Chapter 5 presents two parametric testing procedures for the null hypothesis of no persistent component in the data against the alternative that the data exhibits a persistent component. Our methodologies are based on the Wald and Lagrange multiplier principles and involve PW estimation method. We derive the asymptotic distribution of our test statistics for a wide class of linear processes having a parametrically specified spectral density function, and moreover we establish their consistency and power against local alternatives. As our test statistics are found to have an asymptotic distribution that is nonstandard and model dependent, we introduce a bootstrap scheme for the purposes of calculating valid critical values, and furthermore establish its validity. The finite sample performance of our testing procedures is investigated by the means of Monte-Carlo simulations. Finally, we apply our testing methods to data for the growth rate of industrial production and unemployment rate, and find evidence that these series exhibit persistent components for most of the time periods considered.

As we are examining the PW and LW estimators based on the objective functions (1.3.14) and (1.3.17), respectively, we can assume without loss of generality that $\mu_x = 0$. In what follows C denotes a generic positive finite constant, $\xrightarrow{p}$ denotes convergence in probability, and $\xrightarrow{d}$ convergence in distribution. Moreover, d/dy and $\partial/\partial y$ denote derivative and partial derivative, respectively, for a generic column vector or scalar y, and by $\overline{z}$ we denote the conjugate of a generic z.

Chapter 2

General conditions for local Whittle estimation

2.1 Introduction

As discussed in Chapter 1, the memory parameter α_x characterizes and summarizes the long-run dependence structure of the process $\{x_t\}_{t\in\mathbb{Z}}$, and its consistent estimation is of undoubted interest. The related literature focuses on parametric and semiparametric methods, and various estimators have been proposed and analyzed. In the majority of the cases, assumptions of Gaussianity or linearity of the process are imposed. However, it is not uncommon in empirical applications in Macroeconomics and Finance, that the practitioner deals with time series data, which, possibly after some transformation, do not appear to be generated from a linear process. Therefore, the problem of drawing appropriate inference on α_x arises and, consequently, theoretical justification of the estimator is needed for its use in nonlinear setups.

In the present and next chapters, we address this problem in the context of semiparametric estimation, and in particular, we consider the LW estimator of Künsch (1987), see also Robinson (1995b). Recall that semiparametric methods have the advantage of requiring less a priori known information on the true structure of $\{x_t\}_{t\in\mathbb{Z}}$. Naturally, the latter feature of semiparametric estimation is very appealing in the framework of nonlinearity, as it allows for greater flexibility in the modelling of $\{x_t\}_{t\in\mathbb{Z}}$. Notice also that results for the parametric Whittle were derived by Giraitis and Taqqu (1999) in the case of polynomial transformations of a Gaussian long memory process.

The statistical properties of the LW estimator were initially investigated by Robinson (1995b), where for his proof of consistency, the assumption of linearity of the process was imposed. Some departures from the linear framework have been recently discussed in Arteche (2004) and Hurvich, Moulines, and Soulier (2005). They address the consistent estimation of α_x in the context of the stochastic volatility model for asset returns introduced by Taylor (1994). In both Arteche (2004) and Hurvich, Moulines, and Soulier (2005), the logarithmic squared returns are transformed into a signal plus noise model. Under such specification, the process of interest $\{x_t\}_{t\in\mathbb{Z}}$ is the sum of the signal process $\{y_t\}_{t\in\mathbb{Z}}$ and the noise $\{z_t\}_{t\in\mathbb{Z}}$. The memory parameter α_y of $\{y_t\}_{t\in\mathbb{Z}}$ is taken to exceed that of $\{z_t\}_{t\in\mathbb{Z}}$, so that $\alpha_x = \alpha_y$ and $\{y_t\}_{t\in\mathbb{Z}}$ "signals" the long-run behaviour of $\{x_t\}_{t\in\mathbb{Z}}$. The linearity of the signal process is an assumption made commonly in Arteche (2004) and Hurvich, Moulines, and Soulier (2005) for showing the consistency of the LW estimator. In addition, the structure of the noise is restricted to be a white noise sequence in Hurvich, Moulines, and Soulier (2005) and a linear short memory process in Arteche (2004). However, the latter author restricts the signal and the noise processes to be mutually independent, while the former authors allow for dependence, of a specific form, between the two processes.

Naturally, once the linear framework is abandoned, the consistency of the LW estimator has to be examined on a specific basis, as in Arteche (2004) and Hurvich, Moulines, and Soulier (2005). It would be therefore of interest to establish general sufficient conditions that guarantee the consistency of the LW estimator without relying on a specific linear or nonlinear specification for $\{x_t\}_{t\in\mathbb{Z}}$. Once these sufficient conditions are established, they can be employed to examine the consistency of the estimator in each particular case, without resorting to proving the consistency from first principles. The first objective of this chapter is to provide such general conditions, and show that they are indeed sufficient for the consistency of the LW estimator.

Besides the consistency property, the rate of convergence and asymptotic distribution of an estimator are also of interest. Robinson (1995b), Arteche (2004), and Hurvich, Moulines, and Soulier (2005) established these properties for the LW estimator. Again their results are particular to the specification of $\{x_t\}_{t\in\mathbb{Z}}$, and so our second objective in this chapter is to obtain expansions and rate of convergence for the estimator, without imposing a specific structure on $\{x_t\}_{t\in\mathbb{Z}}$. We can then apply our results to examine the bias, rate of convergence, and asymptotic distribution of the estimator in individual cases.

The remainder of this chapter is as follows. In Section 2.2 we introduce the LW estimator. In Section 2.3, we present and discuss the assumptions. Section 2.4 deals with the theoretical results. As an illustration, in Section 2.5 we apply our findings to the linear process considered by Robinson (1995b), while Section 2.6 contains some final comments. The proofs of Sections 2.4 and 2.5 are found in Appendix 2.A of this chapter, that use a series of technical lemmas placed in Appendix 2.B.

2.2 The local Whittle estimator

For the estimation of the memory parameter α_x, we use the LW estimator, see Künsch (1987) and Robinson (1995b). Recall that, given a set of data $\{x_1, ..., x_n\}$, the estimator is defined as

$$\widehat{\alpha}_x = \underset{\alpha\in[-1,1]}{\arg\min} U_n(\alpha), \tag{2.2.1}$$

where

$$\begin{aligned} U_n(\alpha) &= \log\left(\frac{1}{m}\sum_{j=1}^{m}\frac{I_x(\lambda_j)}{\lambda_j^{-\alpha}}\right) + \frac{1}{m}\sum_{j=1}^{m}\log\left(\lambda_j^{-\alpha}\right) \\ &= \log\left(\frac{1}{m}\sum_{j=1}^{m} j^{\alpha} I_x(\lambda_j)\right) - \frac{\alpha}{m}\sum_{j=1}^{m}\log j. \end{aligned} \tag{2.2.2}$$

Notice that contrary to Robinson (1995b), the minimization of $U_n(\alpha)$ is taken over the interval $[-1,1]$, instead of a closed subinterval of $(-1,1)$ which can be chosen arbitrarily close to $[-1,1]$. From hereafter, we assume as in Robinson (1995b) that the bandwidth parameter $m = m(n)$ is such that

$$m \to \infty \quad \text{and} \quad m = o(n), \quad \text{as } n \to \infty. \tag{2.2.3}$$

The theoretical results of Robinson (1995b) are based on the whitening principle of the normalized periodogram at the Fourier frequencies

$$\eta_{j,x} = \frac{I_x(\lambda_j)}{f_x(\lambda_j)}, \quad j = 1, ..., m. \tag{2.2.4}$$

Roughly speaking, it means that $\{\eta_{j,x}\}_{j=1}^{m}$ behaves as if it were a sequence of uncorrelated random variables with unit mean. This property holds if $\{x_t\}_{t\in\mathbb{Z}}$ were a sequence of i.i.d. random variables, and asymptotically if $\{x_t\}_{t\in\mathbb{Z}}$ were a weakly dependent process. However, for strongly dependent processes and fixed ordinates, Künsch (1986) showed that the normalized periodograms no longer have unit mean and are uncorrelated. Nevertheless, Theorem 2 of Robinson (1995a) implies that the latter bias of the normalized periodograms can be bounded, and that the bound decreases with the distance from the origin, so that for $j \to \infty$ the asymptotic unbiasedness holds. Furthermore, it follows from Theorem 2 of Robinson (1995a) that the correlation between distinct normalized periodograms can also be bounded, and this bound vanishes when $j, k \to \infty$. Therefore, theoretical results for strongly dependent follow as if $\{\eta_{j,x}\}_{j=1}^{m}$ were a sequence of uncorrelated random variables with unit mean.

Notice that under Assumption A.3 below, Lemma 2.6, a version of Theorem 2 of Robinson (1995a), implies that, uniformly in $j \to \infty$, such that $j \leq m$,

$$E(\eta_{j,x}) = 1 + o(1), \quad \text{as } n \to \infty. \tag{2.2.5}$$

If, in addition, it were to be that for $j \neq k$

$$cov(\eta_{j,x}, \eta_{k,x}) \to 0, \quad \text{as } j, k \to \infty, \tag{2.2.6}$$

then from Lemma 2.7 below, it follows under Assumption A.3 that

$$\frac{1}{m} \sum_{j=1}^{m} \eta_{j,x} \xrightarrow{p} 1, \quad \text{as } n \to \infty. \tag{2.2.7}$$

Next, for $j = 1, ..., m$, we denote

$$\eta_{j,x}^{*} = \frac{I_x(\lambda_j)}{c_{0,x} \lambda_j^{-\alpha_x}}. \tag{2.2.8}$$

Observe that under Assumption A.3 and condition (2.2.3) on the bandwidth parameter m, we have that

$$\eta_{j,x}^{*} = \frac{h_x(\lambda_j)}{c_{0,x}} \eta_{j,x} = (1 + o(1)) \eta_{j,x}, \tag{2.2.9}$$

as $n \to \infty$. In addition, under Assumption A.3, we have by Lemma 2.7 that convergence (2.2.7) is equivalent to

$$\frac{1}{m} \sum_{j=1}^{m} \eta_{j,x}^{*} \xrightarrow{p} 1, \quad \text{as } n \to \infty. \tag{2.2.10}$$

The proof of (2.2.10) is one of the key steps in the proof of consistency of the estimator in Robinson (1995b). In particular, assuming that $\{x_t\}_{t \in \mathbb{Z}}$ is a linear process satisfying Assumptions A.3 and A.6 below, Robinson (1995b) showed convergence (2.2.10). Hence, (2.2.10) is needed for the proof of consistency of the estimator and therefore, serves as one of the two sufficient conditions for the consistency of the LW estimator. Notice also that (2.2.5), which holds by Assumption A.3, implies that

$$E(\eta^*_{j,x}) \leq C \tag{2.2.11}$$

uniformly in $j = 1, ..., m$. The latter condition was also established by Robinson (1995b) in his proof of consistency of the LW estimator. The last displayed inequality functions as the second sufficient condition for the consistency of the estimator.

2.3 Assumptions

We introduce the following assumptions:

A.1 Uniformly in $j = 1, ..., m$,

$$E(\eta^*_{j,x}) \leq C. \tag{2.3.1}$$

A.2 We have

$$\frac{1}{m}\sum_{j=1}^{m} \eta^*_{j,x} \xrightarrow{p} 1, \quad \text{as } n \to \infty. \tag{2.3.2}$$

A.3 The spectral density function $f_x(.)$ satisfies

$$f_x(\lambda) = |\lambda|^{-\alpha_x} h_x(\lambda), \quad -\pi < \lambda \leq \pi, \tag{2.3.3}$$

where $-1 < \alpha_x < 1$ and $h_x(\lambda)$, $-\pi < \lambda \leq \pi$ is an even, non-negative function such that

$$h_x(\lambda) \to c_{0,x}, \quad \text{as } \lambda \to 0+, \tag{2.3.4}$$

with $0 < c_{0,x} < \infty$.

A.4 The spectral density function $f_x(.)$ satisfies

$$f_x(\lambda) = \lambda^{-\alpha_x}\left(c_{0,x} + c_{1,x}\lambda^{\beta_x} + o(\lambda^{\beta_x})\right), \quad \text{as } \lambda \to 0+, \tag{2.3.5}$$

with $-1 < \alpha_x < 1$, $0 < c_{0,x} < \infty$, $0 < |c_{1,x}| < \infty$ and $\beta_x \in (0, 2]$.

We now discuss these assumptions. As described in Section 2.2, Assumptions A.1 and A.2 are our main general conditions. We show in Theorem 2.1 below, that they are sufficient for the consistency of the LW estimator. They are based on the periodogram of the data through $\eta^*_{j,x}$, and do not require stationarity or impose a specific structure on $\{x_t\}_{t\in\mathbb{Z}}$ and its spectral density function. It is worth mentioning that for the consistency of the LW estimator, we require essentially the sequence $\{\eta_{j,x}^*\}_{j=1}^m$ to behave as if it were ergodic. This should not come as a surprise, bearing in mind that in the proof of consistency of the parametric Whittle estimator by Hannan (1973), ergodicity of the series $\{x_t\}_{t\in\mathbb{Z}}$ is the main condition.

Assumption A.3 is Assumption A.1 of Robinson (1995b). This assumption is common in the literature on semiparametric estimation of the memory parameter, see Robinson (1994b, 1995a,b), Hidalgo and Yajima (2003), Hidalgo (2005) and Shimotsu and Phillips (2005). As discussed at the end of the previous section, Assumption A.3 implies A.1. Besides in the proof of Theorem 2.1 and equation (2.4.6) of Theorem 2.3, Assumption A.3 is taken to hold true. Assumption A.1 however is easier to check, as we demonstrate in the next chapter for the case of the signal plus noise model.

Assumption A.4 strengthens Assumption A.3 by imposing a known rate of convergence of $\frac{f_x(\lambda)}{c_{0,x}\lambda^{-\alpha_x}}$ to 1, as $\lambda \to 0+$. The parameter β_x characterizes the smoothness of the function $h_x(.)$ in A.3. For example, if $\{x_t\}_{t\in\mathbb{Z}}$ follows an $ARFIMA(p, \frac{\alpha_x}{2}, q)$ process, then Assumption A.4 holds with $\beta_x = 2$. Assumption A.4 is not needed to prove consistency of the estimator, but is required in order to derive expansions, rate of convergence, and the asymptotic distribution of the LW estimator. This assumption is standard in the related literature; it was made by Robinson (1995a,b), Lobato and Robinson (1996), Hidalgo and Yajima (2003), Hidalgo (2005) and Shimotsu and Phillips (2005) to derive the asymptotic distribution of the corresponding semiparametric estimators in the case of Gaussian or linear processes.

2.4 Theoretical results on local Whittle estimation

In this section we present our theoretical results on the consistency, expansions, and rate of convergence for the LW estimator. The proofs of this section are found in Appendix 2.A of this chapter.

2.4.1 Consistency of the local Whittle estimator

We begin by showing the consistency of the LW estimator under the general Assumptions A.1 and A.2. Notice that we could follow the key steps for the proof of consistency by Robinson (1995b), but here we present a different approach. The main idea of our proof is to show that asymptotically, with probability tending to 1, for every $0 < \varepsilon < \frac{1}{2}$, the derivative of the objective function $\frac{\partial}{\partial\alpha}U_n(\alpha)$ is strictly positive when $\alpha \in [\alpha_x + \varepsilon, 1]$ and strictly negative when $\alpha \in [-1, \alpha_x - \varepsilon]$, so that $U_n(\alpha)$ attains a minimum in $(\alpha_x - \varepsilon, \alpha_x + \varepsilon)$.

Theorem 2.1

Suppose that there exists $\alpha_x \in (-1, 1)$ and $0 < c_{0,x} < \infty$ such that Assumptions A.1 and A.2 are satisfied. Then, as $n \to \infty$,

$$\widehat{\alpha}_x \xrightarrow{p} \alpha_x. \tag{2.4.1}$$

Next, assume that $\{x_t\}_{t\in\mathbb{Z}}$ is a covariance stationary process with spectral density function $f_x(.)$. The next theorem shows that if the spectral density function of $\{x_t\}_{t\in\mathbb{Z}}$ satisfies Assumption A.3, then the LW estimator $\widehat{\alpha}_x$ consistently estimates α_x, as long as Assumption A.2 is satisfied. The proof of the theorem is based on Theorem 2.1, and requires showing that Assumption A.3 implies Assumption A.1.

Theorem 2.2

Assume that Assumptions A.2 and A.3 hold. Then, as $n \to \infty$,

$$\widehat{\alpha}_x \xrightarrow{p} \alpha_x. \tag{2.4.2}$$

Assumption A.2 is essential in the proof of Theorem 2.1, but might be difficult to establish in certain cases. So, now we present a simple sufficient condition for Assumption A.2. Denote

$$\Delta_{m,x} = \max_{1\le k\le m} E\left|\sum_{j=1}^{k}\left(\eta^*_{j,x} - E(\eta^*_{j,x})\right)\right|. \tag{2.4.3}$$

Notice that Assumption A.1 implies that $\Delta_{m,x} \le 2\sum_{j=1}^{m} E(\eta^*_{j,x}) \le Cm$. The next proposition shows that $\Delta_{m,x} = o(m)$ together with Assumption A.3 imply Assumption A.2, and therefore, the consistency of the estimator $\widehat{\alpha}_x$.

Proposition 2.1

Suppose that Assumption A.3 holds and $\Delta_{m,x} = o(m)$. Then, $\{x_t\}_{t\in\mathbb{Z}}$ satisfies Assumption A.2 and, as $n \to \infty$,

$$\widehat{\alpha}_x \xrightarrow{p} \alpha_x. \tag{2.4.4}$$

Theorems 2.1, 2.2 and Proposition 2.1 provide general conditions that guarantee the consistency of the LW estimator $\widehat{\alpha}_x$. These conditions can be directly employed to prove the consistency of $\widehat{\alpha}_x$ for specific classes of data generating processes, as we demonstrate below in Section 2.5, and in Chapter 3.

2.4.2 Expansions and convergence rate for the local Whittle estimator

We start with deriving expansions for $\widehat{\alpha}_x$. Denote

$$Q_{m,x} = \frac{1}{m}\sum_{j=1}^{m}\left(\log\left(\frac{j}{m}\right)+1\right)\eta_{j,x}^{*}. \tag{2.4.5}$$

Under the conditions of Theorem 1, we first write $\widehat{\alpha}_x - \alpha_x$ in terms of $Q_{m,x}$ using the mean value theorem. Then, by imposing Assumption A.4, we examine $E\left(Q_{m,x}\right)$, and provide a further expansion for $\widehat{\alpha}_x - \alpha_x$.

Theorem 2.3

Under the assumptions of Theorem 2.1,

$$\widehat{\alpha}_x - \alpha_x = -Q_{m,x}(1+o_p(1)) + O_p\left(\frac{\log m}{m}\right). \tag{2.4.6}$$

If, in addition, the spectral density function $f_x(.)$ satisfies Assumption A.4, then

$$\begin{aligned}\widehat{\alpha}_x - \alpha_x &= -\left(\frac{m}{n}\right)^{\beta_x}\frac{c_{1,x}}{c_{0,x}}B_{\beta_x} - \left(Q_{m,x} - E\left(Q_{m,x}\right)\right)(1+o_p(1)) \\ &\quad + o_p\left(m^{-\frac{1}{2}} + \left(\frac{m}{n}\right)^{\beta_x}\right),\end{aligned} \tag{2.4.7}$$

where

$$B_{\beta_x} = (2\pi)^{\beta_x}\frac{\beta_x}{(\beta_x+1)^2}. \tag{2.4.8}$$

Under the general Assumptions A.1 and A.2, (2.4.6) provides a simple expansion for $\widehat{\alpha}_x - \alpha_x$. It could be used to obtain the asymptotic distribution of $\widehat{\alpha}_x$, if one derives the asymptotic distribution of $Q_{m,x}$. If, in addition, Assumption A.4 is satisfied, then we have a further expansion (2.4.7) for $\widehat{\alpha}_x - \alpha_x$. Observe that the second term in (2.4.7) has mean zero. Hence the first term, i.e. $-\left(\frac{m}{n}\right)^{\beta_x}\frac{c_{1,x}}{c_{0,x}}B_{\beta_x}$, is the dominant term in the finite sample bias of the estimator $\widehat{\alpha}_x$. Expansion (2.4.7) also shows that the rate of convergence of $\widehat{\alpha}_x$ depends on the bias term and the stochastic order of $Q_{m,x} - E\left(Q_{m,x}\right)$.

Remark 2.1

Under the assumptions of Theorem 2.3, the sign of the finite sample bias depends overall on the sign of $c_{1,x}$ given in Assumption A.4. The magnitude of the finite sample bias is determined by $m, n, c_{1,x}, c_{0,x}$ and β_x. In absolute value, the bias decreases when the sample size n increases, increases when the bandwidth parameter m increases, and is also proportional to the ratio $\frac{c_{1,x}}{c_{0,x}}$. With an increase in the smoothness parameter β_x, the term $\left(\frac{m}{n}\right)^{\beta_x}$ decreases, while the term B_{β_x} increases.

Recall from Proposition 2.1 that Assumption A.3 and condition $\Delta_{m,x} = o(m)$ are sufficient for the consistency of the estimator $\widehat{\alpha}_x$, where $\Delta_{m,x}$ is defined in (2.4.3). If, in addition, it is assumed that the spectral density function $f_x(.)$ satisfies Assumption A.4, and $\Delta_{m,x} = o\left(\frac{m}{\log^2 m}\right)$ then, as the next proposition shows, the stochastic order of $Q_{m,x} - E\left(Q_{m,x}\right)$ in (2.4.7) is a function of $\Delta_{m,x}$. Furthermore, we obtain an expansion for $Q_{m,x} - E\left(Q_{m,x}\right)$, which can be employed to obtain, under further assumptions, the asymptotic distribution of $\widehat{\alpha}_x$. Define

$$V_{m,x} = \frac{1}{m}\sum_{j=1}^{m}\left(\log\left(\frac{j}{m}\right)+1\right)\left(\eta_{j,x} - E\left(\eta_{j,x}\right)\right). \tag{2.4.9}$$

Proposition 2.2
If $\Delta_{m,x} = o\left(\frac{m}{\log^2 m}\right)$ and Assumption A.4 holds, then

$$\widehat{\alpha}_x - \alpha_x = O_p\left(\Delta_{m,x}\frac{\log m}{m} + m^{-\frac{1}{2}} + \left(\frac{m}{n}\right)^{\beta_x}\right), \tag{2.4.10}$$

and $Q_{m,x} - E(Q_{m,x})$ in (2.4.7) can be written as

$$Q_{m,x} - E(Q_{m,x}) = V_{m,x} + o_p\left(\left(\frac{m}{n}\right)^{\beta_x}\right). \tag{2.4.11}$$

Expression (2.4.10) is used in Corollary 2.1 below in order to derive, under additional regularity conditions, the rate of convergence of the LW estimator. Before we introduce Corollary 2.1, we formulate conditions in terms of $\{x_t\}_{t\in\mathbb{Z}}$ that control the order of magnitude of $\Delta_{m,x}$ in expression (2.4.10). We now assume that $\{x_t\}_{t\in\mathbb{Z}}$ is a fourth-order stationary sequence, and we denote by $c_x(t_1,t_2,t_3,t_4) = cum(x_{t_1},x_{t_2},x_{t_3},x_{t_4})$ the fourth-order cumulant of the variables $x_{t_1},x_{t_2},x_{t_3},x_{t_4}$ given by

$$\begin{aligned} c_x(t_1,t_2,t_3,t_4) &= E(x_{t_1}x_{t_2}x_{t_3}x_{t_4}) - E(x_{t_1}x_{t_2})E(x_{t_3}x_{t_4}) \\ &\quad - E(x_{t_1}x_{t_3})E(x_{t_2}x_{t_4}) - E(x_{t_1}x_{t_4})E(x_{t_2}x_{t_3}), \end{aligned} \tag{2.4.12}$$

recalling that, without loss of generally, we can assume $E(x_t) = 0$. Denote further

$$D_x = \sum_{t_1,t_2,t_3=-\infty}^{\infty} \left|c_x(t_1,t_2,t_3,0)\right|, \tag{2.4.13}$$

$$D^*_{n,x} = \sum_{t_1,t_2,t_3=-n}^{n} \left|c_x(t_1,t_2,t_3,0)\right|, \tag{2.4.14}$$

and

$$D_{n,x}^{**} = \max_{|t_1|,|t_2| \le n} \sum_{u=-n}^{n} |c_x(t_1, t_2 + u, u, 0)| . \qquad (2.4.15)$$

Note that for a wide class of fourth-order stationary short memory models, e.g. stationary and invertible $ARMA(p,q)$, we have

$$D_x < \infty. \qquad (2.4.16)$$

Observe also that $E(x_t^4) \le C$, implies

$$D_{n,x}^{**} \le Cn. \qquad (2.4.17)$$

We use the quantities $D_{n,x}^{*}$ and $D_{n,x}^{**}$ in Lemma 2.2 to estimate $\Delta_{m,x}$.

Now, we combine expression (2.4.10) of Proposition 2.2 and Lemma 2.2 to derive the rate of convergence of the estimator $\widehat{\alpha}_x$.

Corollary 2.1

Suppose that $\{x_t\}_{t\in\mathbb{Z}}$ is a fourth-order stationary sequence with spectral density function satisfying Assumption A.4. Then, it holds that

$$\widehat{\alpha}_x - \alpha_x = O_p\left(m^{-\frac{1}{2}} \log^2 m + \left(\frac{m}{n}\right)^{\beta_x} + r_m\right), \qquad (2.4.18)$$

where

i) if $\alpha_x = 0$ and $D_x < \infty$,

$$r_m = 0; \qquad (2.4.19)$$

ii) if $\alpha_x \in (-1, 0)$, $D_x < \infty$ and $n^{\frac{1}{2}} m^{-1} \log^2 m = o(1)$,

$$r_m = n^{\frac{1}{2}} m^{-1} \log^2 m = o(1); \qquad (2.4.20)$$

iii) if $\alpha_x \in (0,1)$ and $D_{n,x}^{*} = o\left(\frac{n}{\log^4 n}\right)$,

$$r_m = \left(\frac{D_{n,x}^{*}}{n}\right)^{\frac{1}{2}} \left(\frac{m}{n}\right)^{\alpha_x} \log^2 m = o(1); \qquad (2.4.21)$$

iv) if $\alpha_x \in (0,1)$, $D_{n,x}^{**} = O\left(\frac{n^\gamma}{\log^6 n}\right)$ for some $0 \le \gamma < 1$, and $m^{-1} n^{\frac{\gamma+1}{2}} = O(1)$,

$$r_m = \left(\frac{D_{n,x}^{**}}{n}\right)^{\frac{1}{2}} \left(\frac{m}{n}\right)^{-1+\alpha_x} \log^3 n = o(1). \qquad (2.4.22)$$

Remark 2.2

It is evident from Corollary 2.1 that, in the case of a fourth-order stationary sequence with spectral density function $f_x(.)$ satisfying Assumption A.4, the estimator converges faster the higher the smoothness parameter β_x is, and it has a better rate of convergence under short memory than under long or negative memory.

2.5 An example: Linear process

It is of interest to show that our general results in Subsections 2.4.1 and 2.4.2 hold true for the type of linear processes considered by Robinson (1995b). From hereafter, we say that $\{x_t\}_{t\in\mathbb{Z}}$ is a linear process, if it satisfies Assumption A.3 of Robinson (1995b), that is

A.5 We have

$$x_t = \sum_{j=0}^{\infty} \varphi_{j,x}\varepsilon_{t-j,x}, \quad \sum_{j=0}^{\infty} \varphi_{j,x}^2 < \infty, \tag{2.5.1}$$

where

$$E\left(\varepsilon_{t,x}|F_{t-1}\right) = 0, \quad E\left(\varepsilon_{t,x}^2|F_{t-1}\right) = 1, \quad \text{a.s., } t \in \mathbb{Z}, \tag{2.5.2}$$

in which F_t is the $\sigma-$field of events generated by $\varepsilon_{s,x}$, $s \le t$, and there exists a random variable ε such that $E\left(\varepsilon^2\right) < \infty$ and for all $\eta > 0$ and some $K > 0$, $\Pr\left(|\varepsilon_{t,x}| > \eta\right) \le K \Pr\left(|\varepsilon| > \eta\right).$

Next, we denote $\phi_x(\lambda) = \sum_{j=0}^{\infty} \varphi_{j,x}e^{ij\lambda}$ the transfer function of $\{x_t\}_{t\in\mathbb{Z}}$, and introduce the following conditions:

A.6 In a neighbourhood $(0, \delta)$ of the origin, $f_x(.)$ is differentiable and

$$\frac{d}{d\lambda}\log f_x(\lambda) = O\left(\lambda^{-1}\right), \quad \text{as } \lambda \to 0+. \tag{2.5.3}$$

A.7 In a neighbourhood $(0, \delta)$ of the origin, $\phi_x(.)$ is differentiable and

$$\frac{d\phi_x(\lambda)}{d\lambda} = O\left(\frac{|\phi_x(\lambda)|}{\lambda}\right), \quad \text{as } \lambda \to 0+. \tag{2.5.4}$$

A.8 The sequence $\{\varepsilon_{t,x}\}_{t\in\mathbb{Z}}$ in Assumption A.5 further satisfies

$$E\left(\varepsilon_{t,x}^3|F_{t-1}\right) = \mu_{3,x}, \quad \text{a.s.,} \quad E\left(\varepsilon_{t,x}^4\right) = \mu_{4,x}, \quad t \in \mathbb{Z}, \tag{2.5.5}$$

for finite constants $\mu_{3,x}$ and $\mu_{4,x}$.

Assumption A.6 is Assumption A.2 of Robinson (1995b) and it is employed there for the proof of the consistency of $\widehat{\alpha}_x$. It requires the function $h_x(.)$ to be differentiable in a neighbourhood $(0, \delta)$ of the origin, and to satisfy $\frac{d}{d\lambda}\log h_x(\lambda) = O\left(\lambda^{-1}\right)$, as $\lambda \to 0+$. This assumption is not strong and, for example, it is satisfied when $\{x_t\}_{t\in\mathbb{Z}}$ follows a stationary $ARFIMA(p, \frac{\alpha_x}{2}, d)$ model.

Assumptions A.7 and A.8 are Assumptions A.2' and A.3', respectively, in Robinson (1995b), and are used there to derive the asymptotic distribution of $\widehat{\alpha}_x$. Assumption A.7 strengthens Assumption A.6, since $f_x(\lambda) = \frac{|\phi_x(\lambda)|^2}{2\pi}$. Assumption A.8 further restricts the type of linear processes considered. It implies that the process $\{x_t\}_{t\in\mathbb{Z}}$ is fourth-order stationary, and holds if for example $\{\varepsilon_{t,x}\}_{t\in\mathbb{Z}}$ is a sequence of i.i.d. variables with finite fourth moment.

Robinson (1995b) showed the consistency of the LW estimator for linear processes whose spectral density function satisfy Assumption A.3 and A.6. If furthermore, Assumptions A.4, A.7 and A.8 are satisfied, Robinson (1995b) also established the asymptotic distribution,

$$m^{\frac{1}{2}}(\widehat{\alpha}_x - \alpha_x) \xrightarrow{d} N(0,1), \quad \text{as } n \to \infty, \tag{2.5.6}$$

where the bandwidth parameter m has to be chosen so that $\frac{m^{2\beta_x+1}}{n^{2\beta_x}} \log^2 m = o(1)$.

In the next proposition, we show that the assumptions of Theorem 1 in Robinson (1995b) imply the assumptions of our Theorem 2.1. Furthermore, under Assumptions A.4 and A.8, we derive a bound for $\Delta_{m,x}$ in order to apply Theorem 2.3 and Proposition 2.2 to investigate the rate of convergence of the estimator and provide an expansion for $\widehat{\alpha}_x - \alpha_x$. If furthermore, Assumption A.7 is satisfied, we derive the asymptotic distribution of $V_{m,x}$ given by (2.4.9), which is used to examine further the asymptotic distribution of $\widehat{\alpha}_x$.

Proposition 2.3

Suppose that $\{x_t\}_{t\in\mathbb{Z}}$ is a linear process.

i) If Assumptions A.3 and A.6 hold, then $\{x_t\}_{t\in\mathbb{Z}}$ satisfies Assumptions A.1 and A.2, so that

$$\widehat{\alpha}_x \xrightarrow{p} \alpha_x, \tag{2.5.7}$$

as $n \to \infty$.

ii) Under Assumptions A.4, A.6 and A.8, we have that

$$\Delta_{m,x} = O\left(m^{\frac{1}{2}} \log^{\frac{1}{2}} m + m\left(\frac{m}{n}\right)^{\beta_x}\right). \tag{2.5.8}$$

iii) Under Assumptions A.4, A.7 and A.8,

$$m^{\frac{1}{2}} V_{m,x} \xrightarrow{d} N(0,1), \quad \text{as } n \to \infty. \tag{2.5.9}$$

Proposition 2.3 part i) implies that our general consistency conditions, Assumptions A.1 and A.2, are indeed sufficient in the framework of the consistency theorem of Robinson (1995b).

Moreover, under Assumptions A.4, A.6 and A.8, Proposition 2.3 part ii) implies that

$$\Delta_{m,x} = o\left(\frac{m}{\log^2 m}\right), \quad \text{when } m = o\left(\frac{n}{\log^{\frac{2}{\beta_x}} n}\right). \tag{2.5.10}$$

Notice that the latter requirement on the bandwidth parameter m is not restrictive, as it is satisfied for example, when $m = [n^{\gamma}]$ for any $0 < \gamma < 1$. Hence, relation (2.4.11) of Proposition 2.2 holds, which together with (2.4.7) of Theorem 2.3 yields the expansion

$$\widehat{\alpha}_x - \alpha_x = -\left(\frac{m}{n}\right)^{\beta_x} \frac{c_{1,x}}{c_{0,x}} B_{\beta_x} - V_{m,x}(1 + o_p(1)) + o_p\left(m^{-\frac{1}{2}} + \left(\frac{m}{n}\right)^{\beta_x}\right), \tag{2.5.11}$$

that is valid when $m = o\left(\frac{n}{\log^{\frac{2}{\beta_x}} n}\right)$. Since $E\left(V_{m,x}\right) = 0$, the term $-\left(\frac{m}{n}\right)^{\beta_x} \frac{c_{1,x}}{c_{0,x}} B_{\beta_x}$ determines overall the finite sample bias of the estimator, see also Remark 2.1. In addition, the latter displayed expansion is useful for examining the asymptotic distribution of the estimator $\widehat{\alpha}_x$. We combine (2.5.11) and convergence (2.5.9) of Proposition 2.3 part iii) to conclude the following remark.

Remark 2.3

Under Assumptions A.4, A.7 and A.8:

If $m = o\left(n^{\frac{2\beta_x}{2\beta_x+1}}\right)$, *then*

$$m^{\frac{1}{2}}(\widehat{\alpha}_x - \alpha_x) \xrightarrow{d} N(0,1), \quad \text{as } n \to \infty. \tag{2.5.12}$$

On the other hand, if $\frac{n^{\frac{2\beta_x}{2\beta_x+1}}}{m} = o(1)$ *and* $m = o\left(\frac{n}{\log^{\frac{2}{\beta_x}} n}\right)$, *then*

$$\left(\frac{n}{m}\right)^{\beta_x} (\widehat{\alpha}_x - \alpha_x) \xrightarrow{p} -\frac{c_{1,x}}{c_{0,x}} B_{\beta_x}, \quad \text{as } n \to \infty. \tag{2.5.13}$$

In the special case that $m = \left[n^{\frac{2\beta_x}{2\beta_x+1}}\right]$,

$$m^{\frac{1}{2}}(\widehat{\alpha}_x - \alpha_x) \xrightarrow{d} N\left(-\frac{c_{1,x}}{c_{0,x}} B_{\beta_x}, 1\right), \quad \text{as } n \to \infty. \tag{2.5.14}$$

2.6 Final comments

In this chapter we have described and examined results for the LW estimator of the memory parameter α_x. Without relying on the assumption of linearity of the data generating process, we have established conditions that are sufficient for consistency and we have provided expansions and rate of convergence for the estimator. As an illustration, we have applied our results to the case of a linear process and have reaffirmed the results obtained by Robinson (1995b). Although, we have concentrated on the case of examining long-run persistence, the results can be easily extended to the case of a persistent component of known frequency $\omega_x \neq 0$.

There are two significant aspects that we have not considered in this chapter. Firstly, we have focused on stationary processes, and in particular, we have concentrated on the case $-1 < \alpha_x < 1$. It would be of theoretical and empirical interest to establish similar results to those presented here for higher values of the memory parameter. This problem is likely to be dealt using techniques analogous to those employed here, in conjunction with those of Velasco (1999), who established the asymptotic properties of the LW estimator for linear processes with $\alpha_x > -1$. Using the notion of the pseudo spectral density function as in Velasco (1999), we expect the results on consistency to follow when $-1 < \alpha_x < 2$, while those on expansions and rates of convergence for $-1 < \alpha_x < 1.5$. Examination of the LW estimator under higher values of α_x is likely to require in addition tapering of the observations, as in Velasco (1999).

Secondly, we have assumed knowledge of the possibly persistence component of the data, in the sense that its frequency ω_x is known a priori. The extension to the case of unknown frequency is certainly non-trivial, as the estimation of the frequency ω_x has been studied only for Gaussian or linear processes, see Yajima (1996), Hidalgo and Soulier (2004) and Hidalgo (2005), while the existing results in the semiparametric literature have not concerned LW estimation, see Hidalgo and Soulier (2004) and Hidalgo (2005).

2.A Appendix

This section contains the proofs which use a series of lemmas found in Appendix 2.B below.

Proof of Theorem 2.1. We have

$$\frac{\partial}{\partial \alpha} U_n(\alpha) = \frac{\sum_{j=1}^{m} j^{\alpha} \log j I_x(\lambda_j)}{\sum_{j=1}^{m} j^{\alpha} I_x(\lambda_j)} - \frac{1}{m}\sum_{j=1}^{m} \log j = \frac{\sum_{j=1}^{m} j^{\alpha} \nu_j I_x(\lambda_j)}{\sum_{j=1}^{m} j^{\alpha} I_x(\lambda_j)} := \frac{T_n(\alpha)}{V_n(\alpha)}, \quad (2.A.1)$$

where

$$\nu_j = \log j - \frac{1}{m}\sum_{k=1}^{m} \log k, \tag{2.A.2}$$

$$T_n(\alpha) = \frac{1}{m}\sum_{j=1}^{m} j^{\alpha}\nu_j I_x(\lambda_j) = \frac{1}{m}\sum_{j=1}^{m}\left(\frac{j}{m}\right)^{\alpha-\alpha_x}\nu_j\eta_{j,x}^{*}, \tag{2.A.3}$$

and

$$V_n(\alpha) = \frac{1}{m}\sum_{j=1}^{m} j^{\alpha} I_x(\lambda_j) = \frac{1}{m}\sum_{j=1}^{m}\left(\frac{j}{m}\right)^{\alpha-\alpha_x}\eta_{j,x}^{*}. \tag{2.A.4}$$

Let $0 < \varepsilon < \frac{1}{2}$. By Lemma 2.1 below, for any $\delta > 0$,

$$\Pr\left(\sup_{\alpha\in[-1,1]:\alpha-\alpha_x>-1+\varepsilon}\left|\frac{T_n(\alpha)}{V_n(\alpha)} - \frac{\alpha-\alpha_x}{1+\alpha-\alpha_x}\right| \geq \delta\right) \to 0, \tag{2.A.5}$$

and there exists constant $0 < C(\varepsilon) < \infty$ such that

$$\Pr\left(\sup_{\alpha\in[-1,1]:\alpha-\alpha_x\leq-1+\varepsilon}\frac{T_n(\alpha)}{V_n(\alpha)} \leq -C(\varepsilon)\right) \to 1 \tag{2.A.6}$$

as $n \to \infty$.

Let's consider first the case $\alpha \leq \alpha_x - \varepsilon$. (2.A.5) implies that uniformly in $\alpha \in [-1,1] : -1+\varepsilon < \alpha - \alpha_x \leq -\varepsilon$

$$\left|\frac{T_n(\alpha)}{V_n(\alpha)} - \frac{\alpha-\alpha_x}{1+\alpha-\alpha_x}\right| \leq \varepsilon \tag{2.A.7}$$

with probability tending to 1, as $n \to \infty$, where we have chosen $\delta = \varepsilon$. Notice that the function $\frac{x}{1+x}$ is strictly increasing when $x > -1$. Therefore, for $\alpha \in [-1,1]$: $-1+\varepsilon < \alpha - \alpha_x \leq -\varepsilon$, we obtain that

$$\frac{T_n(\alpha)}{V_n(\alpha)} \leq \varepsilon + \frac{\alpha-\alpha_x}{1+\alpha-\alpha_x} < \varepsilon + \frac{-\varepsilon}{1-\varepsilon} = -\frac{\varepsilon^2}{1-\varepsilon} < 0 \tag{2.A.8}$$

with probability approaching 1, as $n \to \infty$. On the other hand, (2.A.6) implies that uniformly in $\alpha \in [-1,1] : \alpha - \alpha_x \leq -1+\varepsilon$

$$\frac{T_n(\alpha)}{V_n(\alpha)} \leq -C(\varepsilon) \tag{2.A.9}$$

with probability tending to 1, as $n \to \infty$. The last two displayed bounds imply that uniformly in $\alpha \in [-1,1] : \alpha - \alpha_x \leq -\varepsilon$, for some constant $0 < C'(\varepsilon) < \infty$,

$$\frac{\partial}{\partial\alpha}U_n(\alpha) = \frac{T_n(\alpha)}{V_n(\alpha)} \leq -C'(\varepsilon) \tag{2.A.10}$$

with probability approaching 1, as $n \to \infty$, that is

$$\Pr\left(\sup_{\alpha\in[-1,1]:\alpha-\alpha_x\leq-\varepsilon} \frac{\partial}{\partial\alpha}U_n(\alpha) \leq -C'(\varepsilon)\right) \to 1, \tag{2.A.11}$$

as $n \to \infty$.

Next, we consider the case $\alpha - \alpha_x \geq \varepsilon$. Using (2.A.5), it follows that uniformly in $\alpha \in [-1,1] : \alpha - \alpha_x \geq \varepsilon$

$$\left|\frac{T_n(\alpha)}{V_n(\alpha)} - \frac{\alpha-\alpha_x}{1+\alpha-\alpha_x}\right| \leq \frac{\varepsilon}{2} \tag{2.A.12}$$

with probability tending to 1, as $n \to \infty$, where we have chosen $\delta = \frac{\varepsilon}{2}$. Hence, bearing in mind that the function $\frac{x}{1+x}$ is strictly increasing in $x > -1$,

$$\frac{T_n(\alpha)}{V_n(\alpha)} \geq \frac{\alpha-\alpha_x}{1+\alpha-\alpha_x} - \frac{\varepsilon}{2} \geq \frac{\varepsilon}{1+\varepsilon} - \frac{\varepsilon}{2} > \frac{\varepsilon(1-\varepsilon)}{2(1+\varepsilon)} > 0. \tag{2.A.13}$$

The latter displayed inequality implies that uniformly in $\alpha \in [-1,1] : \alpha - \alpha_x \geq \varepsilon$, for some constant $0 < C''(\varepsilon) < \infty$,

$$\frac{\partial}{\partial\alpha}U_n(\alpha) \geq C''(\varepsilon) \tag{2.A.14}$$

with probability tending to 1, as $n \to \infty$, that is

$$\Pr\left(\inf_{\alpha\in[-1,1]:\alpha-\alpha_x\geq\varepsilon} \frac{\partial}{\partial\alpha}U_n(\alpha) \geq C''(\varepsilon)\right) \to 1, \tag{2.A.15}$$

as $n \to \infty$.

Recall that $U_n(\alpha)$ is a continuously differentiable function in α. Hence, from (2.A.11) and (2.A.15), it follows that $U_n(\alpha)$ is strictly decreasing in the interval $[-1, \alpha_x-\varepsilon]$ and strictly increasing in the interval $[\alpha_x+\varepsilon, 1]$, with probability tending to 1, as $n \to \infty$. Therefore, the minimum of $U_n(\alpha)$ falls in the interval $(\alpha_x-\varepsilon, \alpha_x+\varepsilon)$ with probability tending to 1, as $n \to \infty$, so that

$$\Pr(|\widehat{\alpha}_x - \alpha_x| \geq \varepsilon) = \Pr\left(\inf_{\alpha\in[-1,1]:|\alpha-\alpha_x|\geq\varepsilon} U_n(\alpha) < \inf_{\alpha\in[-1,1]:|\alpha-\alpha_x|<\varepsilon} U_n(\alpha)\right) \to 0, \tag{2.A.16}$$

as $n \to \infty$, and for every $0 < \varepsilon < \frac{1}{2}$, which implies (2.4.1) and concludes the proof of this theorem. ■

Proof of Theorem 2.2. In view of Theorem 2.1, it suffices to show that Assumption A.3 implies A.1. Under Assumption A.3, Lemma 2.6 implies that, as $n \to \infty$,

$$E(\eta_{j,x}) = 1 + O\left(\frac{\log j}{j}\right) + o(1) \tag{2.A.17}$$

uniformly in $j = 1, ..., m$. Hence, by Assumption A.3, as $n \to \infty$,

$$\begin{aligned}
E(\eta^*_{j,x}) &= \frac{f_x(\lambda_j)}{c_{0,x}\lambda_j^{-\alpha_x}} E(\eta_{j,x}) \\
&= \frac{c_{0,x}\lambda_j^{-\alpha_x} + o(\lambda_j^{-\alpha_x})}{c_{0,x}\lambda_j^{-\alpha_x}} \left(1 + O\left(\frac{\log j}{j}\right) + o(1)\right) \\
&= (1 + o(1)) \left(1 + O\left(\frac{\log j}{j}\right) + o(1)\right) \qquad (2.A.18) \\
&= 1 + O\left(\frac{\log j}{j}\right) + o(1) \leq C \qquad (2.A.19)
\end{aligned}$$

uniformly in $j = 1, ..., m$, so that Assumption A.1 holds true. ■

Proof of Proposition 2.1. By Theorem 2.2, it suffices to show that Assumption A.2 is satisfied. We know from the proof of Theorem 2.2, that, as $n \to \infty$,

$$E(\eta^*_{j,x}) = 1 + O\left(\frac{\log j}{j}\right) + o(1) \qquad (2.A.20)$$

uniformly in $j = 1, ..., m$. Thus, as $n \to \infty$,

$$\begin{aligned}
\frac{1}{m}\sum_{j=1}^{m} E(\eta^*_{j,x}) &= \frac{1}{m}\sum_{j=1}^{[\log m]} E(\eta^*_{j,x}) + \frac{1}{m}\sum_{j=[\log m]+1}^{m} E(\eta^*_{j,x}) \\
&= \frac{1}{m}\sum_{j=1}^{[\log m]} O(1) + \frac{1}{m}\sum_{j=[\log m]+1}^{m} \left(1 + O\left(\frac{\log j}{j}\right) + o(1)\right) \\
&= O\left(\frac{[\log m]}{m}\right) + O\left(\frac{\log m}{m}\right) \sum_{j=[\log m]+1}^{m} \frac{1}{j} \\
&\quad + \frac{1}{m}(m - [\log m])(1 + o(1)) \\
&= O\left(\frac{[\log m]}{m}\right) + O\left(\frac{\log^2 m}{m}\right) + 1 + o(1) \\
&= 1 + o(1). \qquad (2.A.21)
\end{aligned}$$

On the other hand, by the definition of $\Delta_{m,x}$, we have that

$$E\left|\frac{1}{m}\sum_{j=1}^{m}\left(\eta^*_{j,x} - E(\eta^*_{j,x})\right)\right| \leq Cm^{-1}\Delta_{m,x} = o(1), \qquad (2.A.22)$$

since $\Delta_{m,x} = o(m)$. The last two displayed relations imply that, as $n \to \infty$,

$$\frac{1}{m}\sum_{j=1}^{m} \eta^*_{j,x} \xrightarrow{p} 1 \qquad (2.A.23)$$

and so, Assumption A.2 is satisfied. ■

Proof of Theorem 2.3. Let $0 < \varepsilon < \min\{1-\alpha_x, 1+\alpha_x\}$. From Theorem 2.1 we have that $\widehat{\alpha}_x \xrightarrow{p} \alpha_x$, as $n \to \infty$. Therefore, as $n \to \infty$,

$$\mathcal{I}(|\widehat{\alpha}_x - \alpha_x| < \varepsilon) = 1 + o_p(1) \quad \text{and} \quad \mathcal{I}(|\widehat{\alpha}_x - \alpha_x| \geq \varepsilon) = o_p(1), \tag{2.A.24}$$

where $\mathcal{I}(.)$ denotes the indicator function. We show that

$$(\widehat{\alpha}_x - \alpha_x)\mathcal{I}(|\widehat{\alpha}_x - \alpha_x| < \varepsilon) = -Q_{m,x}(1 + o_p(1)) + O_p\left(\frac{\log m}{m}\right). \tag{2.A.25}$$

Then, the last two displayed equalities imply that

$$\begin{aligned}\widehat{\alpha}_x - \alpha_x &= (\widehat{\alpha}_x - \alpha_x)\mathcal{I}(|\widehat{\alpha}_x - \alpha_x| < \varepsilon) + (\widehat{\alpha}_x - \alpha_x)\mathcal{I}(|\widehat{\alpha}_x - \alpha_x| \geq \varepsilon) \\ &= -Q_{m,x}(1 + o_p(1)) + O_p\left(\frac{\log m}{m}\right) + (\widehat{\alpha}_x - \alpha_x)\, o_p(1) \end{aligned} \tag{2.A.26}$$

and hence,

$$(\widehat{\alpha}_x - \alpha_x)\,(1 - o_p(1)) = -Q_{m,x}(1 + o_p(1)) + O_p\left(\frac{\log m}{m}\right), \tag{2.A.27}$$

which implies that

$$\begin{aligned}\widehat{\alpha}_x - \alpha_x &= \frac{1}{1 - o_p(1)}\left(-Q_{m,x}(1 + o_p(1)) + O_p\left(\frac{\log m}{m}\right)\right) \\ &= -Q_{m,x}(1 + o_p(1)) + O_p\left(\frac{\log m}{m}\right), \end{aligned} \tag{2.A.28}$$

since $(1 - o_p(1))^{-1} = 1 + o_p(1)$, and shows (2.4.6).

To show (2.A.25), we first notice that $|\widehat{\alpha}_x - \alpha_x| < \varepsilon$ implies

$$V_n(\widehat{\alpha}_x) = \frac{1}{m}\sum_{j=1}^{m}\left(\frac{j}{m}\right)^{\widehat{\alpha}_x - \alpha_x}\eta^*_{j,x} \geq \frac{1}{m}\sum_{j=1}^{m}\left(\frac{j}{m}\right)^{\varepsilon}\eta^*_{j,x}. \tag{2.A.29}$$

From Lemma 2.3, because the random variables $\{\eta^*_{j,x}\}_{j=1}^{m}$ satisfy conditions (2.B.50) and (2.B.51), while the function $x^{\varepsilon}, 0 \leq x \leq 1$ satisfies conditions (2.B.52) and (2.B.53), we have that the right hand side of (2.A.29) is $\int_0^1 x^{\varepsilon}dx + o_p(1)$. Thus,

$$V_n(\widehat{\alpha}_x) \geq \frac{1}{1+\varepsilon} + o_p(1), \tag{2.A.30}$$

so that $V_n(\widehat{\alpha}_x) > 0$ with probability tending to 1, as $n \to \infty$. Notice that $\widehat{\alpha}_x \in (-1, 1)$, since $|\widehat{\alpha}_x - \alpha_x| < \varepsilon$ and $0 < \varepsilon < \min\{1-\alpha_x, 1+\alpha_x\}$. Thus, $\frac{\partial}{\partial\alpha}U_n(\widehat{\alpha}_x) = \frac{T_n(\widehat{\alpha}_x)}{V_n(\widehat{\alpha}_x)} = 0$ implies $T_n(\widehat{\alpha}_x) = 0$.

Recall that $T_n(\alpha) = \frac{1}{m}\sum_{j=1}^{m}\left(\frac{j}{m}\right)^{\alpha-\alpha_x}\nu_j\eta_{j,x}^*$, where ν_j are given in (2.A.2). By the mean value theorem,

$$T_n(\widehat{\alpha}_x) - T_n(\alpha_x) = \frac{\partial}{\partial\alpha}T_n(\alpha^*)\left(\widehat{\alpha}_x - \alpha_x\right), \tag{2.A.31}$$

where α^* is an intermediate point between $\widehat{\alpha}_x$ and α_x. From (2.A.31), since $T_n(\widehat{\alpha}_x) = 0$, we have that (2.A.25) holds, if we show that, as $n \to \infty$,

$$T_n(\alpha_x) = Q_{m,x} + O_p\left(\frac{\log m}{m}\right) \tag{2.A.32}$$

and

$$\frac{\partial}{\partial\alpha}T_n(\alpha^*) = 1 + o_p(1). \tag{2.A.33}$$

Applying in the definition of $T_n(\alpha)$ equality

$$\nu_j = \log\left(\frac{j}{m}\right) + 1 + O\left(\frac{\log m}{m}\right), \tag{2.A.34}$$

see (2.B.13) below, we have that

$$\begin{aligned} T_n(\alpha_x) &= \frac{1}{m}\sum_{j=1}^{m}\left(\log\left(\frac{j}{m}\right)+1\right)\eta_{j,x}^* + \frac{1}{m}O\left(\frac{\log m}{m}\right)\sum_{j=1}^{m}\eta_{j,x}^* \\ :&= Q_{m,x} + R_m. \end{aligned} \tag{2.A.35}$$

By Assumption A.1, we have $E(\eta_{j,x}^*) \leq C$ uniformly in $j = 1, ..., m$, which implies that

$$E\left|R_m\right| = \frac{1}{m}O\left(\frac{\log m}{m}\right)\sum_{j=1}^{m}E(\eta_{j,x}^*) = O\left(\frac{\log m}{m}\right), \tag{2.A.36}$$

and proves that $R_m = O_p\left(\frac{\log m}{m}\right)$. Hence, equality (2.A.32) follows.

We now show (2.A.33). From the definition of $T_n(\alpha)$, we have that

$$\begin{aligned} \frac{\partial}{\partial\alpha}T_n(\alpha^*) &= \frac{1}{m}\sum_{j=1}^{m}\log\left(\frac{j}{m}\right)\left(\frac{j}{m}\right)^{\alpha^*-\alpha_x}\nu_j\eta_{j,x}^* \\ &= \frac{1}{m}\sum_{j=1}^{m}\left(p\left(\frac{j}{m};\alpha\right) + q_m\left(\frac{j}{m};\alpha\right)\right)\eta_{j,x}^*, \end{aligned} \tag{2.A.37}$$

where

$$p\left(\frac{j}{m};\alpha\right) = \log\left(\frac{j}{m}\right)\left(\log\left(\frac{j}{m}\right)+1\right) \tag{2.A.38}$$

and

$$q_m\left(\frac{j}{m};\alpha\right) = \log\left(\frac{j}{m}\right)\left(\frac{j}{m}\right)^{\alpha^*-\alpha_x}\nu_j - p\left(\frac{j}{m};\alpha\right). \tag{2.A.39}$$

To examine (2.A.37), we use Lemma 2.4 below. We first notice that the random variables $\{\eta^*_{j,x}\}_{j=1}^m$ satisfy conditions (2.B.50) and (2.B.51). Secondly, the function $p(x;\alpha), 0 \le x \le 1, \alpha \in [-1,1]$ satisfies assumptions (2.B.52) and (2.B.53), since for any fixed $0 < b < 1$, uniformly in $b \le x \le 1$, $\alpha \in [-1,1]$

$$\left|\frac{\partial}{\partial x} p(x;\alpha)\right| \le C, \tag{2.A.40}$$

and for $x \to 0$, uniformly in $\alpha \in [-1,1]$

$$|p(x;\alpha)| \le C\,|\log x|^2 \le C x^{-\gamma}, \tag{2.A.41}$$

for any $0 < \gamma < 1$. Finally, we check that the functions $\{q_m(x;\alpha)\}_{m=1}^{\infty}, 0 \le x \le 1, \alpha \in [-1,1]$ satisfy the assumptions (2.B.67) and (2.B.68). Using relation (2.A.34) we obtain that

$$\begin{aligned}
\left|q_m\left(\frac{j}{m};\alpha\right)\right| &= \left|\log\left(\frac{j}{m}\right)\left(\frac{j}{m}\right)^{\alpha^*-\alpha_x}\left(\log\left(\frac{j}{m}\right)+1+O\left(\frac{\log m}{m}\right)\right)\right.\\
&\quad \left. -p\left(\frac{j}{m};\alpha\right)\right|\\
&\le \left|p\left(\frac{j}{m};\alpha\right)\right|\left|\left(\frac{j}{m}\right)^{\alpha^*-\alpha_x}-1\right|\\
&\quad +\left|\log\left(\frac{j}{m}\right)\left(\frac{j}{m}\right)^{\alpha^*-\alpha_x} O\left(\frac{\log m}{m}\right)\right|.
\end{aligned} \tag{2.A.42}$$

Thus, for any fixed $0 < b < 1$ and uniformly in $b \le x \le 1$, $\alpha \in [-1,1]$ we have

$$|q_m(x;\alpha)| \le C\left|x^{\alpha^*-\alpha_x}-1\right| + |\log x|\, x^{\alpha^*-\alpha_x}\left|O\left(\frac{\log m}{m}\right)\right| = o(1), \tag{2.A.43}$$

because $\alpha^* \xrightarrow{p} \alpha_x$. On the other hand, we obtain from (2.A.42) that, for $x \to 0$ and uniformly in $\alpha \in [-1,1]$

$$\begin{aligned}
|q_m(x;\alpha)| &\le C\,|\log x|^2\left|x^{-|\alpha^*-\alpha_x|}-1\right| + |\log x|\, x^{-|\alpha^*-\alpha_x|}\left|O\left(\frac{\log m}{m}\right)\right|\\
&\le Cx^{-\varepsilon}\,|\log x|^2 \le Cx^{-\gamma'},
\end{aligned} \tag{2.A.44}$$

for any $0 < \gamma' < \varepsilon$, since $|\alpha^* - \alpha_x| < \varepsilon$. So, we can apply Lemma 2.4, to conclude that

$$\frac{1}{m}\sum_{j=1}^{m}\left(p\left(\frac{j}{m};\alpha\right)+q_m\left(\frac{j}{m};\alpha\right)\right)\eta^*_{j,x} \xrightarrow{p} \int_0^1 \log x(\log x+1)dx = 1, \tag{2.A.45}$$

and from (2.A.37) we deduce (2.A.33).

Finally, it remains to show (2.4.7). From (2.4.6) it follows that

$$\begin{aligned}\widehat{\alpha}_x - \alpha_x &= -\left(Q_{m,x} - E(Q_{m,x})\right)(1+o_p(1)) - E(Q_{m,x})(1+o_p(1)) \\ &\quad + O_p\left(\frac{\log m}{m}\right). \qquad (2.A.46)\end{aligned}$$

Next, notice that, under Assumption A.4, Lemma 2.6 implies that uniformly in $j = 1, ..., m$, $E\left(\eta_{j,x}\right) = 1 + O\left(\frac{\log j}{j}\right) + o(1)$, as $n \to \infty$. Thus,

$$\begin{aligned}E(Q_{m,x}) &= \frac{1}{m}\sum_{j=1}^{m}\left(\log\left(\frac{j}{m}\right)+1\right)\frac{h_x(\lambda_j)}{c_{0,x}}E\left(\eta_{j,x}\right) \\
&= \frac{1}{m}\sum_{j=1}^{m}\left(\log\left(\frac{j}{m}\right)+1\right)\left(1+\frac{c_{1,x}}{c_{0,x}}\lambda_j^{\beta_x}+o\left(\lambda_j^{\beta_x}\right)\right) \\
&\quad \times\left(1+O\left(\frac{\log j}{j}\right)+o(1)\right) \\
&= \frac{1}{m}\sum_{j=1}^{m}\left(\log\left(\frac{j}{m}\right)+1\right)\left(1+\frac{c_{1,x}}{c_{0,x}}\lambda_j^{\beta_x}+o\left(\lambda_j^{\beta_x}\right)+o(1)\right) \\
&\quad +\frac{1}{m}\sum_{j=1}^{m}\left(\log\left(\frac{j}{m}\right)+1\right)O\left(\frac{\log j}{j}\right) \\
&= \frac{1}{m}\sum_{j=1}^{m}\left(\log\left(\frac{j}{m}\right)+1\right)\left(1+o\left(\lambda_j^{\beta_x}\right)+o(1)\right) \\
&\quad +\frac{c_{1,x}}{c_{0,x}}(2\pi)^{\beta_x}\left(\frac{m}{n}\right)^{\beta_x}\frac{1}{m}\sum_{j=1}^{m}\left(\log\left(\frac{j}{m}\right)+1\right)\left(\frac{j}{m}\right)^{\beta_x} \\
&\quad +\frac{1}{m}\sum_{j=1}^{m}(\log m+1)\,O\left(\frac{\log j}{j}\right) \\
&= \left(1+o\left(\left(\frac{m}{n}\right)^{\beta_x}\right)+o(1)\right)\frac{1}{m}\sum_{j=1}^{m}\left(\log\left(\frac{j}{m}\right)+1\right) \\
&\quad +\frac{c_{1,x}}{c_{0,x}}(2\pi)^{\beta_x}\left(\frac{m}{n}\right)^{\beta_x}\frac{1}{m}\sum_{j=1}^{m}\left(\log\left(\frac{j}{m}\right)+1\right)\left(\frac{j}{m}\right)^{\beta_x} \\
&\quad +O(\log^2 m)\frac{1}{m}\sum_{j=1}^{m}\frac{1}{j} \\
&= \left(1+o\left(\left(\frac{m}{n}\right)^{\beta_x}\right)+o(1)\right)\left(\int_0^1(\log x+1)dx+O\left(\frac{1}{m}\right)\right)\end{aligned}$$

$$+\frac{c_{1,x}}{c_{0,x}}(2\pi)^{\beta_x}\left(\frac{m}{n}\right)^{\beta_x}\left(\int_0^1(\log x+1)x^{\beta_x}dx+o(1)\right)$$

$$+O(\log^2 m)\frac{1}{m}\int_1^m\frac{1}{x}dx, \qquad (2.A.47)$$

where (2.A.47) follows from Lemma 2.5. Since $\int_0^1(\log x+1)dx=0$ and $\int_0^1(\log x+1)x^{\beta_x}dx=\frac{\beta_x}{(\beta_x+1)^2}$, (2.A.47) yields that

$$E(Q_{m,x})=\left(\frac{m}{n}\right)^{\beta_x}\frac{c_{1,x}}{c_{0,x}}B_{\beta_x}+o\left(m^{-\frac{1}{2}}+\left(\frac{m}{n}\right)^{\beta_x}\right). \qquad (2.A.48)$$

(2.4.7) now follows by (2.A.46) and (2.A.48). ■

Proof of Proposition 2.2. Using (2.4.7), (2.4.10) holds if

$$E\left|Q_{m,x}-E(Q_{m,x})\right|=O\left(\Delta_{m,x}\frac{\log m}{m}\right). \qquad (2.A.49)$$

To show the last displayed bound, we apply summation by parts, that is,

$$\sum_{j=1}^m a_jb_j=\sum_{j=1}^{m-1}(b_j-b_{j+1})\sum_{k=1}^j a_k+b_m\sum_{j=1}^m a_j. \qquad (2.A.50)$$

For $k=1,...,m$, set $S_k=\sum_{j=1}^k\left(\eta^*_{j,x}-E(\eta^*_{j,x})\right)$. Then using (2.A.50) with $a_j=\eta^*_{j,x}-E(\eta^*_{j,x})$ and $b_j=\log\left(\frac{j}{m}\right)+1$, yields that

$$Q_{m,x}-E(Q_{m,x})=\frac{1}{m}\sum_{j=1}^{m-1}\left(\log\left(\frac{j}{m}\right)-\log\left(\frac{j+1}{m}\right)\right)S_j+\frac{1}{m}S_m. \qquad (2.A.51)$$

By the definition of $\Delta_{m,x}$, we have that $E\left|S_k\right|\le\Delta_{m,x}$, for $k=1,...,m$. Therefore,

$$\begin{aligned}E\left|Q_{m,x}-E(Q_{m,x})\right| &\le \frac{1}{m}\sum_{j=1}^m\left|\log\left(1+\frac{1}{j}\right)\right|E\left|S_j\right|+\frac{1}{m}E\left|S_m\right|\\ &\le \Delta_{m,x}\frac{1}{m}\sum_{j=1}^m\frac{1}{j}+\frac{1}{m}\Delta_{m,x}\\ &= O\left(\Delta_{m,x}\frac{\log m}{m}\right), \qquad (2.A.52)\end{aligned}$$

since $\log(1+x)\le x$, for $x\ge 0$.

Next, we prove (2.4.11). It suffices to show that

$$E\left|Q_{m,x}-E(Q_{m,x})-V_{m,x}\right|=o\left(\left(\frac{m}{n}\right)^{\beta_x}\right). \tag{2.A.53}$$

By the definitions of $Q_{m,x}$ and $V_{m,x}$, together with relation $\eta_{j,x}=\frac{c_{0,x}\lambda_j^{-\alpha_x}}{f_x(\lambda_j)}\eta_{j,x}^*$, it follows under Assumption A.3 that

$$\begin{aligned}
Q_{m,x}-E(Q_{m,x})-V_{m,x} &= \frac{1}{m}\sum_{j=1}^{m}\left(\log\left(\frac{j}{m}\right)+1\right)\Big(\left(\eta_{j,x}^*-E(\eta_{j,x}^*)\right)\\
&\quad -\left(\eta_{j,x}-E(\eta_{j,x})\right)\Big)\\
&= \frac{1}{m}\sum_{j=1}^{m}\left(\log\left(\frac{j}{m}\right)+1\right)\left(1-\frac{c_{0,x}}{h_x(\lambda_j)}\right)\\
&\quad \times\left(\eta_{j,x}^*-E(\eta_{j,x}^*)\right).
\end{aligned} \tag{2.A.54}$$

Because by Assumption A.4 we have that, as $n\to\infty$,

$$\begin{aligned}
\frac{1}{c_{0,x}}-\frac{1}{h_x(\lambda_j)} &= \frac{c_{1,x}\lambda_j^{\beta_x}+o\left(\lambda_j^{\beta_x}\right)}{c_{0,x}\left(c_{0,x}+c_{1,x}\lambda_j^{\beta_x}+o\left(\lambda_j^{\beta_x}\right)\right)}\\
&= \frac{c_{1,x}\lambda_j^{\beta_x}+o\left(\lambda_j^{\beta_x}\right)}{c_{0,x}^2+O\left(\lambda_j^{\beta_x}\right)}\\
&= \frac{1}{c_{0,x}^2}\left(c_{1,x}\lambda_j^{\beta_x}+o\left(\lambda_j^{\beta_x}\right)\right)\left(1+o\left(\lambda_j^{\beta_x}\right)\right)\\
&= \frac{c_{1,x}}{c_{0,x}^2}\lambda_j^{\beta_x}+o\left(\lambda_j^{\beta_x}\right),
\end{aligned} \tag{2.A.55}$$

since $(1+O(x))^{-1}=1+o(x)$ for $x\to 0$, we have that the right hand side of (2.A.54) is

$$\frac{1}{m}\sum_{j=1}^{m}\left(\log\left(\frac{j}{m}\right)+1\right)c_{0,x}\left(\frac{c_{1,x}}{c_{0,x}^2}\lambda_j^{\beta_x}+o\left(\lambda_j^{\beta_x}\right)\right)\left(\eta_{j,x}^*-E(\eta_{j,x}^*)\right), \tag{2.A.56}$$

and so

$$\begin{aligned}
E\left|Q_{m,x}-E(Q_{m,x})-V_{m,x}\right| &\le E\left|\frac{1}{m}\sum_{j=1}^{m}\left(\log\left(\frac{j}{m}\right)+1\right)\frac{c_{1,x}}{c_{0,x}}\lambda_j^{\beta_x}\left(\eta_{j,x}^*-E(\eta_{j,x}^*)\right)\right|\\
&\quad +\frac{1}{m}\sum_{j=1}^{m}\left|\log\left(\frac{j}{m}\right)+1\right|o\left(\lambda_j^{\beta_x}\right)E(\eta_{j,x}^*)\\
&:= R_1+R_2.
\end{aligned} \tag{2.A.57}$$

We first examine R_1. For $j = 1, ..., m$, define $p_j = \left(\log\left(\frac{j}{m}\right) + 1\right)\left(\frac{j}{n}\right)^{\beta_x}$. Then,

$$\begin{aligned}
|p_j - p_{j+1}| &\leq \left|\left(\log\left(\frac{j}{m}\right)+1\right)\left(\frac{j}{n}\right)^{\beta_x} - \left(\log\left(\frac{j+1}{m}\right)+1\right)\left(\frac{j}{n}\right)^{\beta_x}\right| + \\
&\quad \left|\left(\log\left(\frac{j+1}{m}\right)+1\right)\left(\frac{j}{n}\right)^{\beta_x} - \left(\log\left(\frac{j+1}{m}\right)+1\right)\left(\frac{j+1}{n}\right)^{\beta_x}\right| \\
&= |\log j - \log(j+1)|\left(\frac{j}{n}\right)^{\beta_x} \\
&\quad + |\log(j+1) - \log m + 1|\left|\left(\frac{j}{n}\right)^{\beta_x} - \left(\frac{j+1}{n}\right)^{\beta_x}\right| \\
&= \left|\log\left(\frac{j+1}{j}\right)\right|\left(\frac{j}{n}\right)^{\beta_x} \\
&\quad + |\log(j+1) - \log m + 1|\left(\frac{j}{n}\right)^{\beta_x}\left|1 - \left(1+\frac{1}{j}\right)^{\beta_x}\right|. \qquad (2.A.58)
\end{aligned}$$

By the mean value theorem, we have for every $j = 1, ..., m$ that

$$1 - \left(1+\frac{1}{j}\right)^{\beta_x} = \beta_x\left(1+\frac{1}{j}\right)^{\delta\beta_x}\log\left(1+\frac{1}{j}\right), \qquad (2.A.59)$$

for some $\delta \in (0, 1)$. Since $\log(1 + x) \leq x$ for every $x > -1$,

$$|p_j - p_{j+1}| \leq C\log m\left(\frac{m}{n}\right)^{\beta_x}\frac{1}{j}. \qquad (2.A.60)$$

For $k = 1, ..., m$, we have by the definition of $\Delta_{m,x}$ that $E|S_k| \leq \Delta_{m,x}$. Applying equality (2.A.50) with $a_j = \eta^*_{j,x} - E(\eta^*_{j,x})$ and $b_j = p_j$ gives

$$\begin{aligned}
R_1 &= CE\left|\frac{1}{m}\sum_{j=1}^{m-1}(p_j - p_{j+1})S_j + \frac{1}{m}p_m S_m\right| \\
&\leq C\left(\frac{m}{n}\right)^{\beta_x}\left(\frac{\log m}{m}\sum_{j=1}^{m-1}\frac{1}{j}E|S_j| + \frac{1}{m}E|S_m|\right) \\
&\leq C\left(\frac{m}{n}\right)^{\beta_x}\Delta_{m,x}\left(\frac{\log m}{m}\sum_{j=1}^{m-1}\frac{1}{j} + \frac{1}{m}\right) \\
&= C\left(\frac{m}{n}\right)^{\beta_x}\Delta_{m,x}\left(\frac{\log m}{m}\left(\int_1^m\frac{1}{x}dx + O(1)\right) + \frac{1}{m}\right) \\
&= O\left(\left(\frac{m}{n}\right)^{\beta_x}\Delta_{m,x}\frac{\log^2 m}{m}\right) = o\left(\left(\frac{m}{n}\right)^{\beta_x}\right), \qquad (2.A.61)
\end{aligned}$$

since by assumption, $\Delta_{m,x} = o\left(\frac{m}{\log^2 m}\right)$.

Now, we consider R_2. Noticing that under Assumption A.4, Lemma 2.6 implies that $E(\eta^*_{j,x}) \leq C$ uniformly in $j = 1, ..., m$, and using Lemma 2.5, we obtain that

$$\begin{aligned} R_2 &= o\left(\left(\frac{m}{n}\right)^{\beta_x}\right)\frac{1}{m}\sum_{j=1}^{m}\left|\log\left(\frac{j}{m}\right)+1\right| \\ &= o\left(\left(\frac{m}{n}\right)^{\beta_x}\right)\left(\int_0^1 |\log x + 1|\, dx + O\left(\frac{1}{m}\right)\right) \\ &= o\left(\left(\frac{m}{n}\right)^{\beta_x}\right)\left(O(1) + O\left(\frac{1}{m}\right)\right) = o\left(\left(\frac{m}{n}\right)^{\beta_x}\right), \end{aligned} \tag{2.A.62}$$

since $\int_0^1 |\log x + 1|\, dx < \infty$. The last two displayed inequalities together with (2.A.57) imply (2.A.53), and complete the proof of (2.4.11). ■

Proof of Corollary 2.1. Noticing that Assumption A.4 is stronger than Assumption A.3, Lemma 2.2 implies that

$$m^{-1}\Delta_{m,x} = O\left(\log(m)\, m^{-\frac{1}{2}} + \left(\frac{D^*_{n,x}}{n}\right)^{\frac{1}{2}}\left(\frac{m}{n}\right)^{\alpha_x}\right), \tag{2.A.63}$$

and in the case $0 \leq \alpha_x < 1$, it holds that

$$m^{-1}\Delta_{m,x} = O\left(\log(m)\, m^{-\frac{1}{2}} + \left(\frac{D^{**}_{n,x}}{n}\right)^{\frac{1}{2}}\left(\frac{m}{n}\right)^{-1+\alpha_x}\log n\right). \tag{2.A.64}$$

Moreover, if $\Delta_{m,x} = o\left(\frac{m}{\log^2 m}\right)$, Proposition 2.2 implies that

$$\widehat{\alpha}_x - \alpha_x = O_p\left(\Delta_{m,x}\frac{\log m}{m} + m^{-\frac{1}{2}} + \left(\frac{m}{n}\right)^{\beta_x}\right). \tag{2.A.65}$$

i) Observe that $D^*_{n,x} < D_x < \infty$. So, from expression (2.A.63) with $\alpha_x = 0$, it follows that

$$\Delta_{m,x} = O\left(\log(m)\, m^{\frac{1}{2}} + mn^{-\frac{1}{2}}\right) = O\left(\log(m)\, m^{\frac{1}{2}}\right) = o\left(\frac{m}{\log^2 m}\right). \tag{2.A.66}$$

We can therefore apply bound (2.A.65) and deduce that

$$\widehat{\alpha}_x - \alpha_x = O_p\left(m^{-\frac{1}{2}}\log m + \left(\frac{m}{n}\right)^{\beta_x}\right), \tag{2.A.67}$$

which proves (2.4.18) with

$$r_m = 0. \tag{2.A.68}$$

ii) We have that $D^*_{n,x} < D_x < \infty$. Hence, (2.A.63) gives

$$\Delta_{m,x} = O\left(\log(m) m^{\frac{1}{2}} + mn^{-\frac{1}{2}} \left(\frac{m}{n}\right)^{\alpha_x}\right) = o\left(\frac{m}{\log^2 m}\right), \tag{2.A.69}$$

under the assumption $\alpha_x \in (-1, 0)$ and $n^{\frac{1}{2}} m^{-1} \log^2 m = o(1)$. Then, (2.A.65) yields

$$\begin{aligned}
\widehat{\alpha}_x - \alpha_x &= O_p\left(m^{-\frac{1}{2}} \log^2 m + n^{-\frac{1}{2}} \left(\frac{m}{n}\right)^{\alpha_x} \log^2 m + \left(\frac{m}{n}\right)^{\beta_x}\right) \\
&= O_p\left(m^{-\frac{1}{2}} \log^2 m + n^{-\frac{1}{2}} \frac{n}{m} \log^2 m + \left(\frac{m}{n}\right)^{\beta_x}\right) \\
&= O_p\left(m^{-\frac{1}{2}} \log^2 m + \left(\frac{m}{n}\right)^{\beta_x} + n^{\frac{1}{2}} m^{-1} \log^2 m\right),
\end{aligned} \tag{2.A.70}$$

which proves (2.4.18) with

$$r_m = n^{\frac{1}{2}} m^{-1} \log^2 m. \tag{2.A.71}$$

iii) From (2.A.63) we have that

$$\Delta_{m,x} = O\left(\log(m) m^{\frac{1}{2}} + m \left(\frac{D^*_{n,x}}{n}\right)^{\frac{1}{2}} \left(\frac{m}{n}\right)^{\alpha_x}\right) = o\left(\frac{m}{\log^2 m}\right). \tag{2.A.72}$$

under the assumptions $D^*_{n,x} = o\left(\frac{n}{\log^4 n}\right)$ and $\alpha_x \in (0, 1)$. We can therefore apply (2.A.65) to obtain that

$$\widehat{\alpha}_x - \alpha_x = O_p\left(m^{-\frac{1}{2}} \log^2 m + \left(\frac{D^*_{n,x}}{n}\right)^{\frac{1}{2}} \left(\frac{m}{n}\right)^{\alpha_x} \log^2 m + \left(\frac{m}{n}\right)^{\beta_x}\right), \tag{2.A.73}$$

which implies that (2.4.18) holds with

$$r_m = \left(\frac{D^*_{n,x}}{n}\right)^{\frac{1}{2}} \left(\frac{m}{n}\right)^{\alpha_x} \log^2 m. \tag{2.A.74}$$

iv) From (2.A.64) it follows that

$$\Delta_{m,x} = O\left(\log(m) m^{\frac{1}{2}} + m \left(\frac{D^{**}_{n,x}}{n}\right)^{\frac{1}{2}} \left(\frac{m}{n}\right)^{-1+\alpha_x} \log n\right) = o\left(\frac{m}{\log^2 m}\right), \tag{2.A.75}$$

since $\alpha_x \in (0, 1)$, $D^{**}_{n,x} = O(\frac{n^\gamma}{\log^6 n})$ for some $0 \le \gamma < 1$, and $m^{-1} n^{\frac{\gamma+1}{2}} = O(1)$. Using (2.A.65), gives

$$\begin{aligned}
\widehat{\alpha}_x - \alpha_x &= O_p\left(m^{-\frac{1}{2}} \log^2 m + \left(\frac{D^{**}_{n,x}}{n}\right)^{\frac{1}{2}} \left(\frac{m}{n}\right)^{-1+\alpha_x} \log n \log^2 m + \left(\frac{m}{n}\right)^{\beta_x}\right) \\
&= O_p\left(m^{-\frac{1}{2}} \log m + \left(\frac{m}{n}\right)^{\beta_x} + \left(\frac{D^{**}_{n,x}}{n}\right)^{\frac{1}{2}} \left(\frac{m}{n}\right)^{-1+\alpha_x} \log^3 n\right),
\end{aligned} \tag{2.A.76}$$

which implies (2.4.18) with

$$r_m = \left(\frac{D^{**}_{n,x}}{n}\right)^{\frac{1}{2}} \left(\frac{m}{n}\right)^{-1+\alpha_x} \log^3 n \tag{2.A.77}$$

and completes the proof of the corollary. ■

Proof of Proposition 2.3. i) Let $I_{\varepsilon_x}(\lambda) = \frac{1}{2\pi n}\left|\sum_{t=1}^{n} \varepsilon_{t,x} e^{it\lambda}\right|^2$ and, for $k = 1, ..., m$, write

$$\sum_{j=1}^{k} \eta_{j,x} = \sum_{j=1}^{k} 2\pi I_{\varepsilon_x}(\lambda_j) + \sum_{j=1}^{k} \left(\eta_{j,x} - 2\pi I_{\varepsilon_x}(\lambda_j)\right) := S_{n,1}(k) + S_{n,2}(k). \tag{2.A.78}$$

We have already shown in the proof of Theorem 2.2 that Assumption A.3 implies A.1.

On the other hand, under Assumption A.3, we have from Lemma 2.7 that Assumption A.2 follows if

$$\frac{1}{m}\sum_{j=1}^{m} \eta_{j,x} \xrightarrow{p} 1, \tag{2.A.79}$$

as $n \to \infty$. To establish (2.A.79), it suffices to show that

$$\frac{1}{m} S_{n,1}(m) \xrightarrow{p} 1 \tag{2.A.80}$$

and

$$\frac{1}{m} S_{n,2}(m) \xrightarrow{p} 0, \tag{2.A.81}$$

as $n \to \infty$. Convergence (2.A.80) follows from Robinson (1995b) pp. 1637-1638. Moreover, from relation (3.17) of Robinson (1995b), it follows that, as $n \to \infty$,

$$E\left|\eta_{j,x} - 2\pi I_{\varepsilon_x}(\lambda_j)\right| \le C\left(\frac{\log j}{j}\right)^{\frac{1}{2}}. \tag{2.A.82}$$

Thus,

$$\begin{aligned}
E\left|\frac{1}{m} S_{n,2}(m)\right| &\le \frac{1}{m}\sum_{j=1}^{m} E\left|\eta_{j,x} - 2\pi I_{\varepsilon_x}(\lambda_j)\right| \\
&\le C\frac{1}{m}\sum_{j=1}^{m}\left(\frac{\log j}{j}\right)^{\frac{1}{2}} \\
&\le Cm^{-\frac{1}{2}} \log^{\frac{1}{2}} m \left(\frac{1}{m}\sum_{j=1}^{m}\left(\frac{j}{m}\right)^{-\frac{1}{2}}\right) \\
&= Cm^{-\frac{1}{2}} \log^{\frac{1}{2}} m \left(\int_0^1 x^{-\frac{1}{2}} dx + o(1)\right) \qquad (2.A.83) \\
&\le Cm^{-\frac{1}{2}} \log^{\frac{1}{2}} m = o(1), \qquad (2.A.84)
\end{aligned}$$

where (2.A.83) follows from Lemma 2.5 below. The inequality (2.A.84) implies (2.A.81).

ii) Using (2.A.78), we have

$$\sum_{j=1}^{k}\eta_{j,x}^{*}=\sum_{j=1}^{k}\eta_{j,x}+\sum_{j=1}^{k}(\eta_{j,x}^{*}-\eta_{j,x})=S_{n,1}\left(k\right)+S_{n,2}\left(k\right)+R_{n}\left(k\right), \tag{2.A.85}$$

where $R_n(k)=\sum_{j=1}^{k}(\eta_{j,x}^{*}-\eta_{j,x})$. Thus,

$$\begin{aligned}\Delta_{m,x} &= \max_{1\le k\le m} E\left|\sum_{j=1}^{k}\left(\eta_{j,x}^{*}-E(\eta_{j,x}^{*})\right)\right| \\ &\le \max_{1\le k\le m}\left(E\left|S_{n,1}(k)-ES_{n,1}(k)\right|+E\left|S_{n,2}(k)-ES_{n,2}(k)\right|\right. \\ &\quad \left.+E\left|R_n(k)-ER_n(k)\right|\right). \end{aligned} \tag{2.A.86}$$

From the proof of relation (4.9) in Robinson (1995b), it is evident that

$$E\left(S_{n,1}(k)-ES_{n,1}(k)\right)^{2}=O(k) \tag{2.A.87}$$

and from Cauchy-Schwarz's inequality, it follows that

$$E\left|S_{n,1}(k)-ES_{n,1}(k)\right|\le\left(E\left(S_{n,1}(k)-ES_{n,1}(k)\right)^{2}\right)^{\frac{1}{2}}=O\left(k^{\frac{1}{2}}\right).$$

Using similar arguments as in (2.A.84), we have that

$$E\left|S_{n,2}(k)-ES_{n,2}(k)\right|\le 2E\left|S_{n,2}(k)\right|\le Ck^{\frac{1}{2}}\log^{\frac{1}{2}}k. \tag{2.A.88}$$

It remains to examine $E\left|R_n(k)-ER_n(k)\right|$. Observe that, under Assumption A.4, Lemma 2.6 implies that, as $n\to\infty$, $E(\eta_{j,x})\le C$ uniformly in $j=1,...,m$. Moreover, under Assumption A.4, as $n\to\infty$,

$$E\left|\eta_{j,x}^{*}-\eta_{j,x}\right|=\left|\frac{c_{1,x}}{c_{0,x}}\lambda_j^{\beta_x}+o(\lambda_j^{\beta_x})\right|E\left(\eta_{j,x}\right)=O\left(\left(\frac{m}{n}\right)^{\beta_x}\right), \tag{2.A.89}$$

which implies that

$$\begin{aligned}E\left|R_n(k)-ER_n(k)\right| &\le 2E\left|R_n(k)\right| \\ &\le C\sum_{j=1}^{k}E\left|\eta_{j,x}^{*}-\eta_{j,x}\right| \\ &= O\left(m\left(\frac{m}{n}\right)^{\beta_x}\right). \end{aligned} \tag{2.A.90}$$

From (2.A.86), (2.A.87), (2.A.88) and (2.A.90), we conclude that

$$\Delta_{m,x} = O\left(m^{\frac{1}{2}} \log^{\frac{1}{2}} m + m\left(\frac{m}{n}\right)^{\beta_x}\right), \qquad (2.A.91)$$

which shows (2.5.8).

iii) We first write

$$\begin{aligned} m^{\frac{1}{2}} V_{m,x} &= m^{-\frac{1}{2}} \sum_{j=1}^{m} \left(\log\left(\frac{j}{m}\right) + 1\right) \left(\eta_{j,x} - 2\pi I_{\varepsilon_x}(\lambda_j)\right) \\ &\quad + m^{-\frac{1}{2}} \sum_{j=1}^{m} \left(\log\left(\frac{j}{m}\right) + 1\right) \left(2\pi I_{\varepsilon_x}(\lambda_j) - 1\right) \\ &\quad + m^{-\frac{1}{2}} \sum_{j=1}^{m} \left(\log\left(\frac{j}{m}\right) + 1\right) \left(1 - E\left(\eta_{j,x}\right)\right) \\ : &= V_{m,1} + V_{m,2} + V_{m,3}. \end{aligned} \qquad (2.A.92)$$

To show (2.5.9), it suffices to show that

$$V_{m,1} \xrightarrow{p} 0, \qquad (2.A.93)$$

$$V_{m,2} \xrightarrow{d} N(0,1), \qquad (2.A.94)$$

and

$$V_{m,3} \xrightarrow{p} 0, \qquad (2.A.95)$$

as $n \to \infty$.

We start by showing (2.A.93). Applying summation by parts (2.A.50) with $a_j = \eta_{j,x} - 2\pi I_{\varepsilon_x}(\lambda_j)$ and $b_j = \log\left(\frac{j}{m}\right) + 1$ implies that

$$\begin{aligned} V_{m,1} &= m^{-\frac{1}{2}} \sum_{j=1}^{m-1} \left(\log\left(\frac{j}{m}\right) - \log\left(\frac{j+1}{m}\right)\right) \sum_{k=1}^{j} \left(\eta_{j,x} - 2\pi I_{\varepsilon_x}(\lambda_j)\right) \\ &\quad + m^{-\frac{1}{2}} \sum_{j=1}^{m} \left(\eta_{j,x} - 2\pi I_{\varepsilon_x}(\lambda_j)\right). \end{aligned} \qquad (2.A.96)$$

From the proof of equation (4.8) in Robinson (1995b), it follows that, as $n \to \infty$,

$$\sum_{k=1}^{j} \left(\eta_{j,x} - 2\pi I_{\varepsilon_x}(\lambda_j)\right) = O_p\left(j^{\frac{1}{3}} \log^{\frac{2}{3}} j + j^{\frac{1}{2}} n^{-\frac{1}{4}}\right). \qquad (2.A.97)$$

Thus, as $n \to \infty$,

$$
\begin{aligned}
V_{m,1} &= m^{-\frac{1}{2}} \sum_{j=1}^{m-1} \log\left(\frac{j}{j+1}\right) O_p\left(j^{\frac{1}{3}} \log^{\frac{2}{3}} j + j^{\frac{1}{2}} n^{-\frac{1}{4}}\right) \\
&\quad + m^{-\frac{1}{2}} O_p\left(m^{\frac{1}{3}} \log^{\frac{2}{3}} m + m^{\frac{1}{2}} n^{-\frac{1}{4}}\right) \\
&= m^{-\frac{1}{2}} O_p\left(\sum_{j=1}^{m-1} \log\left(1+\frac{1}{j}\right)\left(j^{\frac{1}{3}} \log^{\frac{2}{3}} j + j^{\frac{1}{2}} n^{-\frac{1}{4}}\right)\right) \\
&\quad + O_p\left(m^{-\frac{1}{6}} \log^{\frac{2}{3}} m + n^{-\frac{1}{4}}\right) \\
&= m^{-\frac{1}{2}} O_p\left(\sum_{j=1}^{m-1} \frac{1}{j}\left(j^{\frac{1}{3}} \log^{\frac{2}{3}} j + j^{\frac{1}{2}} n^{-\frac{1}{4}}\right)\right) + o_p(1) \\
&= m^{-\frac{1}{2}} O_p\left(\log^{\frac{2}{3}} m \sum_{j=1}^{m-1} j^{-\frac{2}{3}} + n^{-\frac{1}{4}} \sum_{j=1}^{m-1} j^{-\frac{1}{2}}\right) + o_p(1) \\
&= m^{-\frac{1}{2}} O_p\left(\log^{\frac{2}{3}} m \int_0^m x^{-\frac{2}{3}} dx + n^{-\frac{1}{4}} \int_0^m x^{-\frac{1}{2}} dx\right) + o_p(1) \\
&= m^{-\frac{1}{2}} O_p\left(\left(\log^{\frac{2}{3}} m\right) m^{\frac{1}{3}} + n^{-\frac{1}{4}} m^{\frac{1}{2}}\right) + o_p(1) \\
&= O_p\left(\left(\log^{\frac{2}{3}} m\right) m^{-\frac{1}{6}} + n^{-\frac{1}{4}}\right) + o_p(1) = o_p(1), \qquad (2.A.98)
\end{aligned}
$$

as required.

Next, we show (2.A.94). Using equation (2.B.13), we can write

$$
\begin{aligned}
V_{m,2} &= m^{-\frac{1}{2}} \sum_{j=1}^{m} \left(\log\left(\frac{j}{m}\right) - \frac{1}{m}\sum_{k=1}^{m} \log\left(\frac{k}{m}\right)\right)\left(2\pi I_{\varepsilon_x}(\lambda_j) - 1\right) \\
&\quad + m^{-\frac{1}{2}} \sum_{j=1}^{m} \left(\frac{1}{m}\sum_{k=1}^{m} \log\left(\frac{k}{m}\right) + 1\right)\left(2\pi I_{\varepsilon_x}(\lambda_j) - 1\right) \\
&= m^{-\frac{1}{2}} \sum_{j=1}^{m} \nu_j \left(2\pi I_{\varepsilon_x}(\lambda_j) - 1\right) \\
&\quad + m^{-\frac{1}{2}} \left(\frac{1}{m}\sum_{k=1}^{m} \log\left(\frac{k}{m}\right) + 1\right) \sum_{j=1}^{m} \left(2\pi I_{\varepsilon_x}(\lambda_j) - 1\right) \\
&= m^{-\frac{1}{2}} \sum_{j=1}^{m} \nu_j \left(2\pi I_{\varepsilon_x}(\lambda_j) - 1\right) \\
&\quad + m^{-\frac{1}{2}} \left(\int_0^1 \log x dx + o(1) + 1\right) O_p\left(m^{\frac{1}{2}}\right)
\end{aligned}
$$

$$= \quad m^{-\frac{1}{2}} \sum_{j=1}^{m} \nu_j \left(2\pi I_{\varepsilon_x}(\lambda_j) - 1\right) + o_p(1), \tag{2.A.99}$$

where we have used equation (4.9) of Robinson (1995b), that is $\sum_{j=1}^{m} \left(2\pi I_{\varepsilon_x}(\lambda_j) - 1\right) = O_p\left(m^{\frac{1}{2}}\right)$, and that $\frac{1}{m} \sum_{k=1}^{m} \log\left(\frac{k}{m}\right) = \int_0^1 \log x dx + o(1)$ by Lemma 2.5. It follows by Robinson (1995b) pp. 1644-1647 that, as $n \to \infty$,

$$m^{-\frac{1}{2}} \sum_{j=1}^{m} \nu_j \left(2\pi I_{\varepsilon_x}(\lambda_j) - 1\right) \xrightarrow{d} N(0,1), \tag{2.A.100}$$

which together with (2.A.99) imply (2.A.94).

Finally, we show (2.A.95). We have from Theorem 2 of Robinson (1995a) that $E\left(\eta_{j,x}\right) = 1 + O\left(\frac{\log j}{j}\right)$ uniformly in $j = 1, ..., m$, as $n \to \infty$. Thus,

$$\begin{aligned} V_{m,3} &= m^{-\frac{1}{2}} \sum_{j=1}^{m} \left(\log\left(\frac{j}{m}\right) + 1\right) O\left(\frac{\log j}{j}\right) \\ &= O\left(m^{-\frac{1}{2}} \log^2 m \sum_{j=1}^{m} \frac{1}{j}\right) \\ &= O\left(m^{-\frac{1}{2}} \log^3 m\right) = o(1), \end{aligned} \tag{2.A.101}$$

as $n \to \infty$, which proves (2.A.95) and completes the proof of (2.5.9). ■

2.B Appendix

This section contains a series of technical lemmas used in the proofs in Appendix 2.A above.

Lemma 2.1
Suppose that Assumptions A.1 and A.2 hold. Then, as $n \to \infty$, for every $0 < \varepsilon < 1$ and $\delta > 0$,

$$\Pr\left(\sup_{\alpha \in [-1,1]: \alpha - \alpha_x > -1+\varepsilon} \left| \frac{T_n(\alpha)}{V_n(\alpha)} - \frac{\alpha - \alpha_x}{1 + \alpha - \alpha_x} \right| \geq \delta\right) \to 0 \tag{2.B.1}$$

and, for every $0 < \varepsilon < 1$, there exists constant $0 < C(\varepsilon) < \infty$ such that

$$\Pr\left(\sup_{\alpha \in [-1,1]: \alpha - \alpha_x \leq -1+\varepsilon} \frac{T_n(\alpha)}{V_n(\alpha)} \leq -C(\varepsilon)\right) \to 1. \tag{2.B.2}$$

Proof. Define

$$t_m\left(\frac{j}{m};\alpha\right) = \left(\frac{j}{m}\right)^{\alpha-\alpha_x}\nu_j, \tag{2.B.3}$$

$$t\left(\frac{j}{m};\alpha\right) = \left(\frac{j}{m}\right)^{\alpha-\alpha_x}\left(\log\left(\frac{j}{m}\right)+1\right) \tag{2.B.4}$$

and

$$v\left(\frac{j}{m};\alpha\right) = \left(\frac{j}{m}\right)^{\alpha-\alpha_x}, \tag{2.B.5}$$

where ν_j is given by (2.A.2). Then, from the definitions (2.A.3) and (2.A.4) of $T_n(\alpha)$ and $V_n(\alpha)$, we can write

$$T_n(\alpha) = \frac{1}{m}\sum_{j=1}^{m} t_m\left(\frac{j}{m};\alpha\right)\eta_{j,x}^{*} \tag{2.B.6}$$

and

$$V_n(\alpha) = \frac{1}{m}\sum_{j=1}^{m} v\left(\frac{j}{m};\alpha\right)\eta_{j,x}^{*}. \tag{2.B.7}$$

Define further

$$T(\alpha) = \frac{\alpha-\alpha_x}{(1+\alpha-\alpha_x)^2} \quad\text{and}\quad V(\alpha) = \frac{1}{1+\alpha-\alpha_x}. \tag{2.B.8}$$

We start with the proof of (2.B.1). It suffices to show that, as $n \to \infty$,

$$T_n(\alpha) \xrightarrow{p} T(\alpha) \tag{2.B.9}$$

and

$$V_n(\alpha) \xrightarrow{p} V(\alpha), \tag{2.B.10}$$

uniformly in $\alpha \in [-1,1] : \alpha - \alpha_x > -1+\varepsilon$. Then, (2.B.9) and (2.B.10) imply that uniformly in $\alpha \in [-1,1] : \alpha - \alpha_x > -1+\varepsilon$

$$\left|\frac{T_n(\alpha)}{V_n(\alpha)} - \frac{T(\alpha)}{V(\alpha)}\right| \xrightarrow{p} 0, \tag{2.B.11}$$

as $n \to \infty$, which shows (2.B.1).

We first derive (2.B.9). Note that from Lemma 2 of Robinson (1995b), we have that

$$\left|\frac{1}{m}\sum_{k=1}^{m}\log\left(\frac{k}{m}\right)+1\right| = O\left(\frac{\log m}{m}\right), \tag{2.B.12}$$

as $m \to \infty$. Hence,

$$\nu_j = \log\left(\frac{j}{m}\right) - \frac{1}{m}\sum_{k=1}^{m}\log\left(\frac{k}{m}\right) = \log\left(\frac{j}{m}\right)+1+O\left(\frac{\log m}{m}\right), \tag{2.B.13}$$

as $n \to \infty$. Thus,

$$\begin{aligned} t_m\left(\frac{j}{m};\alpha\right) &= \left(\frac{j}{m}\right)^{\alpha-\alpha_x}\left(\log\left(\frac{j}{m}\right)+1+O\left(\frac{\log m}{m}\right)\right) \\ &= t\left(\frac{j}{m};\alpha\right)+\left(\frac{j}{m}\right)^{\alpha-\alpha_x}O\left(\frac{\log m}{m}\right). \end{aligned} \tag{2.B.14}$$

Since $\alpha - \alpha_x > -1+\varepsilon$, (2.B.14) implies that

$$\begin{aligned} t_m\left(\frac{j}{m};\alpha\right) &= t\left(\frac{j}{m};\alpha\right)+\left(\frac{j}{m}\right)^{-1+\varepsilon}O\left(\frac{\log m}{m}\right) \\ &= t\left(\frac{j}{m};\alpha\right)+m^{1-\varepsilon}O\left(\frac{\log m}{m}\right) \\ &= t\left(\frac{j}{m};\alpha\right)+o(1). \end{aligned} \tag{2.B.15}$$

We use equation (2.B.15) to apply Lemma 2.3. We first notice that the variables $\{\eta^*_{j,x}\}_{j=1}^m$ satisfy Assumptions A.1 and A.2 and, therefore satisfy (2.B.50) and (2.B.51). The function $t_m(x;\alpha)$ satisfies (2.B.52) and (2.B.53), since (2.B.15) implies that, for any fixed $0<b<1$, uniformly in $x\in[b,1]$ and $\alpha\in[-1,1]$

$$\left|\frac{\partial}{\partial x}t_m(x;\alpha)\right| \le |\alpha-\alpha_x|\, x^{\alpha-\alpha_x-1}\,|\log x+1| + x^{\alpha-\alpha_x-1}+o(1)\le C, \tag{2.B.16}$$

whereas for $x\to 0$,

$$|t_m(x;\alpha)| \le x^{\alpha-\alpha_x}\,|\log x+1|+o(1)\le x^{-1+\varepsilon}\,|\log x+1|+o(1)\le Cx^{-(1-\frac{\varepsilon}{2})}, \tag{2.B.17}$$

taking into account that $\alpha-\alpha_x>-1+\varepsilon$. So, by Lemma 2.3 we conclude that, as $n\to\infty$,

$$\sup_{\alpha\in[-1,1]}\left|\frac{1}{m}\sum_{j=1}^{m} t_m\left(\frac{j}{m};\alpha\right)\eta^*_{j,x} - \int_0^1 t(x;\alpha)dx\right| \xrightarrow{p} 0, \tag{2.B.18}$$

which implies (2.B.9).

Next, we show (2.B.10). As before, $\{\eta^*_{j,x}\}_{j=1}^m$ satisfy conditions (2.B.50) and (2.B.51) of Lemma 2.3. In addition, for any fixed $0<b<1$, uniformly in $x\in[b,1]$ and $\alpha\in[-1,1]$, we have that

$$\left|\frac{\partial}{\partial x}v(x;\alpha)\right| = |\alpha-\alpha_x|\, x^{\alpha-\alpha_x-1}\le C, \tag{2.B.19}$$

whereas for $x\to 0$,

$$|v(x;\alpha)| = x^{\alpha-\alpha_x}\le x^{-(1-\varepsilon)}, \tag{2.B.20}$$

recalling that $\alpha - \alpha_x > -1 + \varepsilon$. Therefore, from Lemma 2.3, we deduce that, as $n \to \infty$,

$$\sup_{\alpha\in[-1,1]} \left| \frac{1}{m} \sum_{j=1}^{m} v\left(\frac{j}{m};\alpha\right) \eta^*_{j,x} - \int_0^1 v(x;\alpha)dx \right| \xrightarrow{p} 0, \tag{2.B.21}$$

which implies (2.B.10), and completes the proof of (2.B.1).

Finally, we prove (2.B.2). Let $0 < \delta < \frac{1}{e^2}$. Write

$$\begin{aligned} T_n(\alpha) &= \frac{1}{m}\sum_{j=1}^{[\delta m]} t_m\left(\frac{j}{m};\alpha\right)\eta^*_{j,x} + \frac{1}{m}\sum_{j=[\delta m]+1}^{m} t_m\left(\frac{j}{m};\alpha\right)\eta^*_{j,x} \\ &: = T_{n,1}(\alpha) + T_{n,2}(\alpha) \end{aligned} \tag{2.B.22}$$

and

$$\begin{aligned} V_n(\alpha) &= \frac{1}{m}\sum_{j=1}^{[\delta m]} v\left(\frac{j}{m};\alpha\right)\eta^*_{j,x} + \frac{1}{m}\sum_{j=[\delta m]+1}^{m} v\left(\frac{j}{m};\alpha\right)\eta^*_{j,x} \\ &: = V_{n,1}(\alpha) + V_{n,2}(\alpha). \end{aligned} \tag{2.B.23}$$

From (2.B.13) and the definitions of $t_m\left(\frac{j}{m};\alpha\right)$ and $v\left(\frac{j}{m};\alpha\right)$, it follows that, uniformly in $\alpha \in [-1,1] : \alpha - \alpha_x \leq -1 + \varepsilon$ and $j = [\delta m] + 1, ..., m$

$$|\nu_j| \leq C, \quad \left|t_m\left(\frac{j}{m};\alpha\right)\right| \leq C(\varepsilon), \quad \left|v\left(\frac{j}{m};\alpha\right)\right| \leq C(\varepsilon), \tag{2.B.24}$$

for some constant $0 < C(\varepsilon) < \infty$. Using the same argument as in (2.B.65), Assumptions A.1 and A.2 imply that, as $n \to \infty$,

$$\frac{1}{m}\sum_{j=[\delta m]+1}^{m} \eta^*_{j,x} \xrightarrow{p} \int_\delta^1 1dx = 1 - \delta. \tag{2.B.25}$$

Hence, we have that

$$\sup_{\alpha-\alpha_x\leq -1+\varepsilon} |T_{n,2}(\alpha)| \leq C(\varepsilon) \tag{2.B.26}$$

and

$$\sup_{\alpha-\alpha_x\leq -1+\varepsilon} |V_{n,2}(\alpha)| \leq C(\varepsilon), \tag{2.B.27}$$

with probability tending to 1, as $n \to \infty$.

Next, using (2.B.13), we obtain that, for $j = 1, ..., [\delta m]$

$$\nu_j < -2 + o(1) < -1, \tag{2.B.28}$$

since $0 < \delta < \frac{1}{e^2}$. Thus,

$$T_{n,1}(\alpha) \leq -V_{n,1}(\alpha). \tag{2.B.29}$$

Then, if

$$V_{n,1}(\alpha) - C(\varepsilon) > 0 \tag{2.B.30}$$

with probability tending to 1, as $n \to \infty$, (2.B.26), (2.B.27) and (2.B.29) imply that, uniformly in $\alpha \in [-1,1] : \alpha - \alpha_x \leq -1 + \varepsilon$

$$\frac{T_n(\alpha)}{V_n(\alpha)} = \frac{T_{n,1}(\alpha) + T_{n,2}(\alpha)}{V_{n,1}(\alpha) + V_{n,2}(\alpha)} \leq \frac{-V_{n,1}(\alpha) + C(\varepsilon)}{V_{n,1}(\alpha) - C(\varepsilon)} = -1 \tag{2.B.31}$$

with probability tending to 1, as $n \to \infty$, which proves (2.B.2).

It remains to show (2.B.30). Define the function $v'(x;\alpha) = x^{-1+\varepsilon}$ if $0 \leq x \leq \delta$ and $v'(x;\alpha) = 0$ if $\delta < x \leq 1$. Since $\alpha - \alpha_x \leq -1 + \varepsilon$, we have $v\left(\frac{j}{m};\alpha\right) \geq v'\left(\frac{j}{m};\alpha\right)$ for $j = 1, ..., [\delta m]$. Therefore,

$$V_{n,1}(\alpha) = \frac{1}{m}\sum_{j=1}^{[\delta m]} v\left(\frac{j}{m};\alpha\right)\eta_{j,x}^{*} \geq \frac{1}{m}\sum_{j=1}^{m} v'\left(\frac{j}{m};\alpha\right)\eta_{j,x}^{*}. \tag{2.B.32}$$

By assumption, $\{\eta_{j,x}^{*}\}_{j=1}^{m}$ satisfy (2.B.50) and (2.B.51), while the function $v'(x;\alpha), 0 \leq x \leq 1, \alpha \in [-1,1]$ satisfies piecewise conditions (2.B.52) and (2.B.53). Thus, by Lemma 2.3 and Remark 2.4, we obtain that, as $n \to \infty$,

$$V_{n,1}(\alpha) \geq \frac{1}{m}\sum_{j=1}^{m} v'\left(\frac{j}{m};\alpha\right)\eta_{j,x}^{*} \xrightarrow{p} \int_0^1 v'(x;\alpha)\,dx = \frac{\delta^{\varepsilon}}{\varepsilon} > C(\varepsilon), \tag{2.B.33}$$

for $0 < \varepsilon < 1$ small enough. ∎

Lemma 2.2

Suppose that $\{x_t\}_{t\in\mathbb{Z}}$ is a fourth-order stationary sequence with spectral density function $f_x(.)$ satisfying Assumption A.3. Then

$$m^{-1}\Delta_{m,x} = O\left(\log(m)\, m^{-\frac{1}{2}} + \left(\frac{D_{n,x}^{*}}{n}\right)^{\frac{1}{2}}\left(\frac{m}{n}\right)^{\alpha_x}\right). \tag{2.B.34}$$

Moreover, in the case $0 \leq \alpha_x < 1$, it holds that

$$m^{-1}\Delta_{m,x} = O\left(\log(m)\, m^{-\frac{1}{2}} + \left(\frac{D_{n,x}^{**}}{n}\right)^{\frac{1}{2}}\left(\frac{m}{n}\right)^{-1+\alpha_x}\log n\right). \tag{2.B.35}$$

Proof. Denote $v_x(\lambda) = \frac{w_x(\lambda)}{\sqrt{2\pi c_{0,x}|\lambda|^{-\alpha_x}}}$. Then, we can write $\eta^*_{j,x} = |v_x(\lambda_j)|^2$, and for $j = 1, ..., n-1$, we have $E(v_x(\lambda_j)) = E(\overline{v_x}(\lambda_j)) = 0$. From Theorem 2.3.2 of Brillinger (1975), we obtain that

$$\begin{aligned} cov(v_x(\lambda_j)\overline{v_x}(\lambda_j), v_x(\lambda_p)\overline{v_x}(\lambda_p)) &= cum(v_x(\lambda_j), \overline{v_x}(\lambda_j), v_x(\lambda_p), \overline{v_x}(\lambda_p)) \\ &\quad + E(v_x(\lambda_j)v_x(\lambda_p))E(\overline{v_x}(\lambda_j)\overline{v_x}(\lambda_p)) \\ &\quad + E(v_x(\lambda_j)\overline{v_x}(\lambda_p))E(\overline{v_x}(\lambda_j)v_x(\lambda_p)). \quad (2.B.36) \end{aligned}$$

Thus,

$$\begin{aligned} E\left(\sum_{j=1}^{k}(\eta^*_{j,x} - E(\eta^*_{j,x}))\right)^2 &= \sum_{j,p=1}^{k} cov(\eta^*_{j,x}, \eta^*_{p,x}) \\ &= \sum_{j,p=1}^{k} cov\left(v_x(\lambda_j)\overline{v_x}(\lambda_j), v_x(\lambda_p)\overline{v_x}(\lambda_p)\right) \\ &\leq \sum_{j,p=1}^{k} \left(|E(v_x(\lambda_j)v_x(\lambda_p))|^2 + |E(v_x(\lambda_j)\overline{v_x}(\lambda_p))|^2\right) \\ &\quad + \left|\sum_{j,p=1}^{k} cum(v_x(\lambda_j), \overline{v_x}(\lambda_j), v_x(\lambda_p), \overline{v_x}(\lambda_p))\right| \\ &: = R_{n,1}(k) + R_{n,2}(k). \quad (2.B.37) \end{aligned}$$

Then, the Cauchy-Schwarz's inequality implies that

$$\begin{aligned} \Delta_{m,x} &\leq \max_{1\leq k\leq m}\left(E\left(\sum_{j=1}^{k}(\eta^*_{j,x} - E(\eta^*_{j,x}))\right)^2\right)^{\frac{1}{2}} \\ &= \max_{1\leq k\leq m}(R_{n,1}(k) + R_{n,2}(k))^{\frac{1}{2}} \\ &\leq \max_{1\leq k\leq m}\left(R^{\frac{1}{2}}_{n,1}(k) + R^{\frac{1}{2}}_{n,2}(k)\right), \quad (2.B.38) \end{aligned}$$

where the last displayed inequality follows since $(|a| + |b|)^{\frac{1}{2}} \leq |a|^{\frac{1}{2}} + |b|^{\frac{1}{2}}$.

We first consider $R_{n,1}(k)$. By Lemma 2.6, it follows that, as $n \to \infty$,

$$\begin{aligned} R_{n,1}(k) &\leq C\left(\sum_{1\leq p<j\leq k}\left(\frac{\log^2 j}{(j-p)^2} + \frac{\log^2 j}{p^{|\alpha_x|}j^{2-|\alpha_x|}}\right) + \sum_{j=1}^{k}\left(1 + O\left(\frac{\log j}{j}\right) + o(1)\right)\right) \\ &\leq C\left(\log^2 m \sum_{1\leq p<j\leq m}(j-p)^{-2} \sum_{1\leq p<j\leq m} p^{-|\alpha_x|}j^{-2+|\alpha_x|} + m\right) \\ &\leq C\left(\log^2 m \sum_{p=1}^{m}\sum_{j=1}^{\infty} j^{-2} + \log^2 m \sum_{1\leq p<j\leq m} p^{-|\alpha_x|}p^{-1+|\alpha_x|}j^{-1} + m\right) \end{aligned}$$

$$
\begin{aligned}
&= C\left(\log^2(m)\, m + \log^2 m \sum_{1\le p<j\le m} p^{-1} j^{-1} + m\right) \\
&\le C\left(\log^2(m)\, m + \log^2 m \left(\sum_{j=1}^{m} j^{-1}\right)^2 + m\right) \\
&\le C\left(\log^2(m)\, m + \log^4 m + m\right) \le C \log^2(m)\, m. \qquad (2.B.39)
\end{aligned}
$$

Next, we examine $R_{n,2}(k)$. We have

$$
\begin{aligned}
R_{n,2}(k) &\le \frac{1}{c_{0,x}^2}\frac{1}{(2\pi n)^2}\sum_{j,p=1}^{k} \lambda_j^{\alpha_x}\lambda_p^{\alpha_x} \\
&\quad \times \sum_{t_1,t_2,t_3,t_4=1}^{n} \left|e^{i(t_1-t_2)\lambda_j} e^{i(t_3-t_4)\lambda_p} cum(x_{t_1}, x_{t_2}, x_{t_3}, x_{t_4})\right| \\
&\le Cn^{-2}\left(\sum_{j=1}^{k}\lambda_j^{\alpha_x}\right)^2 \sum_{t_1,t_2,t_3,t_4=1}^{n} |c_x(t_1,t_2,t_3,t_4)| \\
&\le Cn^{-2}\left(\left(\frac{m}{n}\right)^{\alpha_x} m \left(\frac{1}{m}\sum_{j=1}^{m}\left(\frac{j}{m}\right)^{\alpha_x}\right)\right)^2 \sum_{t_1,t_2,t_3,t_4=1}^{n} |c_x(t_1,t_2,t_3,t_4)| \\
&= Cn^{-1}\left(\frac{m}{n}\right)^{2\alpha_x} m^2 D_{n,x}^*, \qquad (2.B.40)
\end{aligned}
$$

where (2.B.40) follows, since, for $-1 < \alpha_x < 1$, we have that $\frac{1}{m}\sum_{j=1}^{m}\left(\frac{j}{m}\right)^{\alpha_x} = \int_0^1 x^{\alpha_x} dx + o(1) = O(1)$.

From (2.B.38)-(2.B.40) we obtain that

$$
\Delta_{m,x} \le C\left(\log(m)\, m^{\frac{1}{2}} + m\left(\frac{D_{n,x}^*}{n}\right)^{\frac{1}{2}}\left(\frac{m}{n}\right)^{\alpha_x}\right), \qquad (2.B.41)
$$

which completes the proof of (2.B.34).

Now to prove (2.B.35), we provide a sharper bound for $R_{n,2}(k)$. Define $B_k(t) = \frac{1}{2\pi n}\sum_{j=1}^{k} c_{0,x}^{-1}\lambda_j^{\alpha_x} e^{it\lambda_j}$ for $k = 1, ..., m$. Thus,

$$
\begin{aligned}
R_{n,2}(k) &= \left|\sum_{t_1,t_2,t_3,t_4=1}^{n} B_k(t_1-t_2) B_k(t_3-t_4) c_x(t_1,t_2,t_3,t_4)\right| \\
&= \left|\sum_{t_1,t_2,t_3,t_4=1}^{n} B_k(t_1-t_2) B_k(t_3-t_4)\right. \\
&\qquad \left.\times\, c_x(t_1-t_2, 0, t_3-t_2, t_4-t_2)\right| \qquad (2.B.42)
\end{aligned}
$$

$$\leq \left| n \sum_{u_1,u_2,u_3=-n}^{n} B_k(u_1)B_k(u_2)c_x(u_1,0,u_2+u_3,u_3) \right| \tag{2.B.43}$$

$$\leq n \sum_{u_1,u_2=-n}^{n} |B_k(u_1)B_k(u_2)| \sum_{u_3=-n}^{n} |c_x(u_1,0,u_2+u_3,u_3)|$$

$$\leq nD^{**}_{n,x} \left(\sum_{u=-n}^{n} |B_k(u)| \right)^2, \tag{2.B.44}$$

where equality (2.B.42) follows from the properties of the cumulant and, inequality (2.B.43) by the change of variables, $u_1 = t_1 - t_2$, $u_2 = t_3 - t_4$ and $u_3 = t_4 - t_2$. We show below that, when $0 \leq \alpha_x < 1$,

$$|B_k(t)| \leq C\left(\frac{m}{n}\right)^{\alpha_x} |t|_+^{-1}, \tag{2.B.45}$$

where $|t|_+ = \max\{|t|, 1\}$. Then,

$$R_{n,2}(k) \leq CnD^{**}_{n,x}\left(\frac{m}{n}\right)^{2\alpha_x} \left(\sum_{t=-n}^{n} |t|_+^{-1}\right)^2$$

$$\leq CnD^{**}_{n,x}\left(\frac{m}{n}\right)^{2\alpha_x} \log^2 n, \tag{2.B.46}$$

which together with (2.B.38) and (2.B.39) imply (2.B.35).

It remains to show (2.B.45). Set $s_p = \sum_{j=1}^{p} e^{it\lambda_j}$ for $p = 1, ..., k$. Then,

$$|s_p| = \left| e^{i\lambda_t} \frac{1 - e^{ip\lambda_t}}{1 - e^{i\lambda_t}} \right| \leq \frac{2}{\left| e^{i\frac{\lambda_t}{2}} \left(e^{-i\frac{\lambda_t}{2}} - e^{i\frac{\lambda_t}{2}} \right) \right|}$$

$$= \left| \sin\left(\frac{\lambda_t}{2}\right) \right|^{-1} \leq C\frac{n}{|t|_+}, \tag{2.B.47}$$

where the last inequality follows since $\frac{\sin x}{x} \to 1$, as $x \to 0$. Using (2.A.50) with $a_j = e^{it\lambda_j}$ and $b_j = j^{\alpha_x}$, we have that

$$B_k(t) = Cn^{-1}n^{-\alpha_x} \left(\sum_{j=1}^{k-1} \left(j^{\alpha_x} - (j+1)^{\alpha_x}\right) s_j + k^{\alpha_x} s_k \right). \tag{2.B.48}$$

Hence, it follows from (2.B.47) that

$$|B_k(t)| \leq Cn^{-\alpha_x} |t|_+^{-1} \left(\sum_{j=1}^{k-1} |j^{\alpha_x} - (j+1)^{\alpha_x}| + k^{\alpha_x} \right)$$

$$= Cn^{-\alpha_x} |t|_+^{-1} \left(\sum_{j=1}^{k-1} \left((j+1)^{\alpha_x} - j^{\alpha_x}\right) + k^{\alpha_x} \right)$$

$$= Cn^{-\alpha_x} |t|_+^{-1} \left(2k^{\alpha_x} - 1\right) \leq C\left(\frac{m}{n}\right)^{\alpha_x} |t|_+^{-1}, \tag{2.B.49}$$

as required. ∎

Lemma 2.3

Assume that the random variables $\{y_j\}_{j=1}^m = \{y_{j,m}\}_{j=1}^m$ *are such that*

$$E\left|y_j\right| \leq C \tag{2.B.50}$$

uniformly in $j = 1, ..., m$ *and, for any* $0 < \tau \leq 1$,

$$\frac{1}{[\tau m]}\sum_{j=1}^{[\tau m]} y_j \xrightarrow{p} 1, \tag{2.B.51}$$

as $m \to \infty$. *Suppose further that there exists function* $w(x;\alpha)$, $0 \leq x \leq 1$, $\alpha \in [\alpha_1, \alpha_2] \subset R$ *such that*

$$\sup_{b\leq x\leq 1}\sup_{\alpha\in[\alpha_1,\alpha_2]}\left|\frac{\partial}{\partial x}w(x;\alpha)\right| \leq C, \tag{2.B.52}$$

for any $0 < b < 1$ *and, there exists* $0 < \gamma < 1$ *such that*

$$\sup_{\alpha\in[\alpha_1,\alpha_2]}|w(x;\alpha)| \leq Cx^{-\gamma}, \tag{2.B.53}$$

as $x \to 0$. *Then, as* $m \to \infty$,

$$\sup_{\alpha\in[\alpha_1,\alpha_2]}\left|\frac{1}{m}\sum_{j=1}^{m} w\left(\frac{j}{m};\alpha\right)y_j - \int_0^1 w(x;\alpha)dx\right| \xrightarrow{p} 0. \tag{2.B.54}$$

Proof. First notice that from (2.B.51) we have that, for any $0 < \tau_1 < \tau_2 \leq 1$,

$$\frac{1}{[\tau_2 m]-[\tau_1 m]}\sum_{j=[\tau_1 m]+1}^{[\tau_2 m]} y_j = \frac{1}{[\tau_2 m]-[\tau_1 m]}\left(\sum_{j=1}^{[\tau_2 m]} y_j - \sum_{j=1}^{[\tau_1 m]} y_j\right) \xrightarrow{p} 1, \tag{2.B.55}$$

as $m \to \infty$. Next, we fix $0 < b < 1$ and split the sum in (2.B.54) in two sums,

$$\frac{1}{m}\sum_{j=1}^{m} w\left(\frac{j}{m};\alpha\right)y_j = \frac{1}{m}\sum_{j=1}^{[bm]} w\left(\frac{j}{m};\alpha\right)y_j + \frac{1}{m}\sum_{j=[bm]+1}^{m} w\left(\frac{j}{m};\alpha\right)y_j. \tag{2.B.56}$$

We begin by examining the first sum on the right hand side of equation (2.B.56). In view of assumptions (2.B.50) and (2.B.53), we deduce that, for any $0 < b < 1$,

$$E\left(\sup_{\alpha\in[\alpha_1,\alpha_2]}\left|\frac{1}{m}\sum_{j=1}^{[bm]} w\left(\frac{j}{m};\alpha\right)y_j\right|\right) \leq \frac{1}{m}\sum_{j=1}^{[bm]} C\left(\frac{j}{m}\right)^{-\gamma} E\left|y_j\right|$$

$$
\begin{aligned}
&\le C\frac{1}{m}\sum_{j=1}^{[bm]}\left(\frac{j}{m}\right)^{-\gamma} \\
&\le C\left(\int_0^b x^{-\gamma}dx + o(1)\right) \\
&= \frac{C}{1-\gamma}b^{1-\gamma}, \qquad (2.B.57)
\end{aligned}
$$

because $0 < \gamma < 1$. Thus, as $m \to \infty$,

$$
\sup_{\alpha\in[\alpha_1,\alpha_2]}\left|\frac{1}{m}\sum_{j=1}^{[bm]} w\left(\frac{j}{m};\alpha\right)y_j\right| \xrightarrow{p} 0, \qquad (2.B.58)
$$

as $b \to 0$.

Next, we consider the second sum on the right hand side of equation (2.B.56). Let $\varepsilon > 0$ and define the partition of $[b,1]$, $b = \tau_1 < ... < \tau_P = 1$, where $\tau_{p+1} - \tau_p = \varepsilon$ for $p = 1, ..., P-1$ for some fixed P. Then, we can write

$$
\begin{aligned}
\frac{1}{m}\sum_{j=[bm]+1}^{m} w\left(\frac{j}{m};\alpha\right)y_j &= \frac{1}{m}\sum_{p=1}^{P-1}\sum_{\tau_p<\frac{j}{m}\le\tau_{p+1}} w\left(\frac{j}{m};\alpha\right)y_j \\
&= \frac{1}{m}\sum_{p=1}^{P-1}\sum_{\tau_p<\frac{j}{m}\le\tau_{p+1}} w(\tau_p;\alpha)y_j \\
&\quad +\frac{1}{m}\sum_{p=1}^{P-1}\sum_{\tau_p<\frac{j}{m}\le\tau_{p+1}} \left(w\left(\frac{j}{m};\alpha\right) - w(\tau_p;\alpha)\right)y_j \\
&: = S_1 + S_2. \qquad (2.B.59)
\end{aligned}
$$

We first examine S_1. Since for every $p = 1, ..., P$, we have $\frac{[\tau_p m]}{m} \to \tau_p$, as $m \to \infty$, and using (2.B.55) we obtain that

$$
\begin{aligned}
S_1 &= (1+o_p(1))\sum_{p=1}^{P-1} w(\tau_p;\alpha)(\tau_{p+1}-\tau_p) \\
&= (1+o_p(1))\left(\sum_{p=1}^{P-1}\int_{\tau_p}^{\tau_{p+1}}(w(\tau_p;\alpha)-w(x;\alpha))\,dx + \sum_{p=1}^{P-1}\int_{\tau_p}^{\tau_{p+1}} w(x;\alpha)dx\right) \\
&= (1+o_p(1))\left(O\left(P_m\right) + \int_b^1 w(x;\alpha)dx\right), \qquad (2.B.60)
\end{aligned}
$$

where (2.B.60) follows from the mean value theorem with

$$P_m = \sum_{p=1}^{P-1} \int_{\tau_p}^{\tau_{p+1}} \sup_{b \le x \le 1} \sup_{\alpha \in [\alpha_1, \alpha_2]} \left| \frac{\partial}{\partial x} w(x;\alpha) \right| (x - \tau_p)\, dx = \varepsilon O(1), \tag{2.B.61}$$

by (2.B.52) and since $\tau_{p+1} - \tau_p = \varepsilon$. Hence, we can write (2.B.60) as

$$S_1 = (1 + o_p(1)) \left(\varepsilon O(1) + \int_b^1 w(x;\alpha) dx \right). \tag{2.B.62}$$

In (2.B.62) the term $\delta(\varepsilon) = \varepsilon O(1) \to 0$, as $\varepsilon \to 0$, and thus, we can make $\delta(\varepsilon)$ arbitrarily small by choosing small enough ε. Hence, using assumption (2.B.52), which implies that $\left| \int_b^1 w(x;\alpha) dx \right| < \infty$, we deduce

$$S_1 = \int_b^1 w(x;\alpha) dx + o_p(1) + \delta(\varepsilon), \tag{2.B.63}$$

where $o_p(1) \to 0$, as $m \to \infty$ and $\delta(\varepsilon) \to 0$, as $\varepsilon \to 0$.

Next, we consider S_2. Applying the mean value theorem and then using (2.B.50), we obtain

$$\begin{aligned}
E|S_2| &\le \frac{1}{m} \sum_{p=1}^{P-1} \sum_{\tau_p < \frac{j}{m} \le \tau_{p+1}} \left| w\left(\frac{j}{m};\alpha\right) - w(\tau_p;\alpha) \right| E|y_j| \\
&\le \frac{C}{m} \sum_{p=1}^{P-1} \sum_{\tau_p < \frac{j}{m} \le \tau_{p+1}} \left| \frac{j}{m} - \tau_p \right| \sup_{\tau_p \le \xi \le \frac{j}{m}} \left| \frac{\partial}{\partial \xi} w(\xi;\alpha) \right| \\
&\le \frac{C}{m} \sum_{p=1}^{P-1} \sum_{\tau_p < \frac{j}{m} \le \tau_{p+1}} |\tau_{p+1} - \tau_p| \sup_{b \le \xi \le 1} \sup_{\alpha \in [\alpha_1, \alpha_2]} \left| \frac{\partial}{\partial \xi} w(\xi;\alpha) \right| \\
&\le \frac{C\varepsilon}{m} \sum_{p=1}^{P-1} \sum_{\tau_p < \frac{j}{m} \le \tau_{p+1}} 1 \le C\varepsilon \frac{1}{m} \sum_{j=1}^{m} 1 = C\varepsilon,
\end{aligned} \tag{2.B.64}$$

for any fixed $\varepsilon > 0$, by assumption (2.B.52).

Since ε can be chosen arbitrarily small, (2.B.59), (2.B.63) and (2.B.64) imply that, uniformly in $\alpha \in [\alpha_1, \alpha_2]$,

$$\frac{1}{m} \sum_{j=[bm]+1}^{m} w\left(\frac{j}{m};\alpha\right) y_j \xrightarrow{p} \int_b^1 w(x;\alpha) dx, \tag{2.B.65}$$

as $m \to \infty$. Notice also that, as $b \to 0$,

$$\int_0^1 w(x;\alpha)dx = \int_b^1 w(x;\alpha)dx + \int_0^b w(x;\alpha)dx = \int_b^1 w(x;\alpha)dx + O\left(\int_0^b x^{-\gamma}dx\right)$$
$$= \int_b^1 w(x;\alpha)dx + O(b^{1-\gamma}) = \int_b^1 w(x;\alpha)dx + o(1), \qquad (2.B.66)$$

since $0 < \gamma < 1$.

The latter displayed equation together with (2.B.65) and (2.B.58) complete the proof of the lemma. ∎

Remark 2.4
Lemma 2.3 holds when the function $w(x;\alpha)$ is piecewise differentiable in $(0,1)$ and satisfies piecewise the assumption (2.B.52) and (2.B.53).

Lemma 2.4
Assume that the random variables $\{y_j\}_{j=1}^m$ and the function $w(x;\alpha)$ satisfy the assumptions of Lemma 2.3, and the random variables $\{r_m(x;\alpha)\}_{m=1}^{\infty}, 0 \le x \le 1$, $\alpha \in [\alpha_1, \alpha_2] \subset R$ are such that, as $m \to \infty$,

$$\sup_{b \le x \le 1} \sup_{\alpha \in [\alpha_1,\alpha_2]} |r_m(x;\alpha)| = o_p(1), \qquad (2.B.67)$$

for any $0 < b < 1$ and, there exists $0 < \gamma' < 1$ such that, as $m \to \infty$,

$$\sup_{\alpha \in [\alpha_1,\alpha_2]} |r_m(x;\alpha)| = O_p(x^{-\gamma'}), \qquad (2.B.68)$$

as $x \to 0$. Then, as $m \to \infty$,

$$\sup_{\alpha \in [\alpha_1,\alpha_2]} \left| \frac{1}{m} \sum_{j=1}^m \left(w\left(\frac{j}{m};\alpha\right) + r_m\left(\frac{j}{m};\alpha\right)\right) y_j - \int_0^1 w(x;\alpha)dx \right| \xrightarrow{p} 0. \qquad (2.B.69)$$

Proof. In view of Lemma 2.3, for the proof of (2.B.69) it suffices to check that uniformly in $\alpha \in [\alpha_1, \alpha_2]$

$$\frac{1}{m} \sum_{j=1}^m r_m\left(\frac{j}{m};\alpha\right) y_j \xrightarrow{p} 0, \qquad (2.B.70)$$

as $m \to \infty$. We split the sum in (2.B.70) in two sums,

$$\frac{1}{m} \sum_{j=1}^m r_m\left(\frac{j}{m};\alpha\right) y_j = \frac{1}{m} \sum_{j=1}^{[bm]} r_m\left(\frac{j}{m};\alpha\right) y_j + \frac{1}{m} \sum_{j=[bm]+1}^{m} r_m\left(\frac{j}{m};\alpha\right) y_j, \quad (2.B.71)$$

for some fixed $0 < b < 1$.

We start by examining the first sum on the right hand side of (2.B.71). By (2.B.68) and (2.B.50), we have that

$$\begin{aligned}
E\left(\sup_{\alpha\in[\alpha_1,\alpha_2]}\left|\frac{1}{m}\sum_{j=1}^{[bm]} r_m\left(\frac{j}{m};\alpha\right)y_j\right|\right) &\leq C\frac{1}{m}\sum_{j=1}^{[bm]}\left(\frac{j}{m}\right)^{-\gamma'} E\left|y_j\right| \\
&\leq C\left(m^{-1+\gamma'}\sum_{j=1}^{[bm]} j^{-\gamma'}\right) \\
&\leq C\left(m^{-1+\gamma'}\int_0^{bm} x^{-\gamma'}\,dx\right) \\
&= C\left(m^{-1+\gamma'}\frac{(bm)^{1-\gamma'}}{1-\gamma'}\right) \\
&\leq Cb^{1-\gamma'}. \qquad (2.B.72)
\end{aligned}$$

Since $0 < \gamma' < 1$, we deduce that, as $b \to 0$,

$$\sup_{\alpha\in[\alpha_1,\alpha_2]}\left|\frac{1}{m}\sum_{j=1}^{[bm]} r_m\left(\frac{j}{m};\alpha\right)y_j\right| \xrightarrow{p} 0. \qquad (2.B.73)$$

On the other hand, using (2.B.67) and then (2.B.55), we deduce that, for any fixed $0 < b < 1$, the second sum on the right hand side of (2.B.71) is bounded in absolute value by

$$\left|\frac{1}{m}\sum_{j=[bm]+1}^{m} r_m\left(\frac{j}{m};\alpha\right)y_j\right| \leq o_p(1)\left|\frac{1}{m}\sum_{j=[bm]+1}^{m} y_j\right| = o_p(1), \qquad (2.B.74)$$

uniformly in $\alpha \in [\alpha_1, \alpha_2]$.

Hence, from (2.B.73) and (2.B.74), it follows that (2.B.70) holds uniformly in $\alpha \in [\alpha_1, \alpha_2]$. ■

Lemma 2.5

Let $g(x)$, $0 \leq x \leq 1$ be an integrable function.

i) If there exists $0 \leq \gamma < 1$ such that $\left|\frac{dg(x)}{dx}\right| \leq Cx^{-\gamma}$ for $0 < x < 1$, then

$$\frac{1}{m}\sum_{j=1}^{m} g\left(\frac{j}{m}\right) = \int_0^1 g(x)dx + O\left(m^{-1}\right), \qquad (2.B.75)$$

as $m \to \infty$.

ii) If there exists $1 < \gamma < 2$ *such that* $|g(x)| \leq Cx^{-\gamma+1}$ *for* $0 \leq x \leq 1$, *and* $\left|\frac{dg(x)}{dx}\right| \leq Cx^{-\gamma}$ *for* $0 < x < 1$, *then*

$$\frac{1}{m}\sum_{j=1}^{m} g\left(\frac{j}{m}\right) = \int_0^1 g(x)dx + O(m^{\gamma-2}), \tag{2.B.76}$$

as $m \to \infty$.

Proof. i) The sum on the left hand side of (2.B.75) can be written as

$$\begin{aligned}\frac{1}{m}\sum_{j=1}^{m} g\left(\frac{j}{m}\right) &= \frac{1}{m}\int_0^m g\left(\frac{[x]+1}{m}\right)dx = \int_0^1 g\left(\frac{[mx]+1}{m}\right)dx \\ &= \int_0^1 \left(g\left(\frac{[mx]+1}{m}\right) - g(x)\right)dx + \int_0^1 g(x)dx. \end{aligned} \tag{2.B.77}$$

From the mean value theorem we have that, for some $\xi \in (x, \frac{[mx]+1}{m})$,

$$\begin{aligned}\left|g\left(\frac{[mx]+1}{m}\right) - g(x)\right| &= \left|\frac{dg(\xi)}{d\xi}\right|\left|\frac{[mx]+1}{m} - x\right| \\ &\leq \sup_{\xi\in(x,\frac{[mx]+1}{m})}\left|\frac{dg(\xi)}{d\xi}\right|\left|\frac{[mx]-mx}{m} + \frac{1}{m}\right|, \end{aligned} \tag{2.B.78}$$

which, under the assumptions of the lemma, is bounded by

$$C \sup_{\xi\in(x,\frac{[mx]+1}{m})} \xi^{-\gamma}\left|\frac{[mx]-mx}{m} + \frac{1}{m}\right| \leq Cx^{-\gamma}\frac{1}{m}, \tag{2.B.79}$$

uniformly in $0 < x < 1$, since $0 \leq \gamma < 1$. Hence,

$$\begin{aligned}\left|\int_0^1 \left(g\left(\frac{[mx]+1}{m}\right) - g(x)\right)dx\right| &\leq \int_0^1 \left|g\left(\frac{[mx]+1}{m}\right) - g(x)\right|dx \\ &\leq C\frac{1}{m}\int_0^1 x^{-\gamma}dx = C\frac{1}{m}\frac{1}{1-\gamma} \\ &= O\left(m^{-1}\right). \end{aligned} \tag{2.B.80}$$

The latter displayed bound together with (2.B.77) proves (2.B.75).

ii) In view of (2.B.77), it suffices to show that

$$\int_0^1 \left(g\left(\frac{[mx]+1}{m}\right) - g(x)\right)dx = O(m^{\gamma-2}). \tag{2.B.81}$$

We take the integral in (2.B.81) and split it up in two parts,

$$\int_0^1 \left(g\left(\frac{[mx]+1}{m}\right) - g(x)\right) dx = \int_0^{\frac{1}{m}} \left(g\left(\frac{[mx]+1}{m}\right) - g(x)\right) dx + \int_{\frac{1}{m}}^1 \left(g\left(\frac{[mx]+1}{m}\right) - g(x)\right) dx. \quad (2.B.82)$$

We now examine the first integral on the right hand side of (2.B.82). Since $0 \leq x < \frac{1}{m}$, we have that $[mx] = 0$. Then,

$$\begin{aligned}
\int_0^{\frac{1}{m}} \left| g\left(\frac{[mx]+1}{m}\right) - g(x)\right| dx &= \int_0^{\frac{1}{m}} \left| g\left(\frac{1}{m}\right) - g(x)\right| dx \\
&\leq C \int_0^{\frac{1}{m}} \left(m^{\gamma-1} + x^{-\gamma+1}\right) dx \\
&= O(m^{\gamma-2}). \qquad (2.B.83)
\end{aligned}$$

Consider next the second integral on the right hand side of (2.B.82). Using the same argument as in (2.B.79), we have that

$$\left| g\left(\frac{[mx]+1}{m}\right) - g(x)\right| \leq Cx^{-\gamma}\frac{1}{m} \qquad (2.B.84)$$

uniformly in $0 < x < 1$, since $1 < \gamma < 2$. Thus,

$$\begin{aligned}
\int_{\frac{1}{m}}^1 \left| g\left(\frac{[mx]+1}{m}\right) - g(x)\right| dx &\leq C\frac{1}{m} \int_{\frac{1}{m}}^1 x^{-\gamma} dx \\
&= C\left(\frac{m^{\gamma-2}}{\gamma-1} - \frac{m^{-1}}{\gamma-1}\right) \\
&= O\left(m^{\gamma-2}\right), \qquad (2.B.85)
\end{aligned}$$

since $1 < \gamma < 2$.

From (2.B.82), (2.B.83) and (2.B.85), (2.B.81) follows. ■

The following lemma is an extension of Theorem 2 of Robinson (1995a) that can be found in Abadir, Distaso, and Giraitis (2005). Recall that we denote $v_x(\lambda) = \frac{w_x(\lambda)}{\sqrt{2\pi c_{0,x}|\lambda|^{-\alpha_x}}}$.

Lemma 2.6

Let Assumptions A.3 be satisfied. Then uniformly in $1 \leq k < j \leq m = o(n)$, as $n \to \infty$,

i) $$E\left(\frac{I_x(\lambda_j)}{f_x(\lambda_j)}\right) = 1 + O\left(\frac{\log j}{j}\right) + o(1), \tag{2.B.86}$$

where $o(1) \to 0$ uniformly in $1 \leq j \leq m$, as $n \to \infty$,

ii) $$E\left(v_x(\lambda_j)v_x(\lambda_j)\right) = O\left(\frac{\log j}{j}\right), \tag{2.B.87}$$

iii) $$E\left(v_x(\lambda_j)\overline{v_x(\lambda_k)}\right) = O\left(\frac{\log j}{j-k}\right) + O\left(\frac{\log j}{k^{\frac{|\alpha_x|}{2}} j^{1-\frac{|\alpha_x|}{2}}}\right), \tag{2.B.88}$$

iv) $$E\left(v_x(\lambda_j)v_x(\lambda_k)\right) = O\left(\frac{\log j}{j-k}\right) + O\left(\frac{\log j}{k^{\frac{|\alpha_x|}{2}} j^{1-\frac{|\alpha_x|}{2}}}\right). \tag{2.B.89}$$

Lemma 2.7

Suppose that Assumption A.3 is satisfied.

i) If $cov(\eta_{j,x}, \eta_{k,x}) \to 0$ when $j \neq k$ and $j, k \to \infty$, then

$$\frac{1}{m}\sum_{j=1}^{m} \eta_{j,x} \xrightarrow{p} 1, \tag{2.B.90}$$

as $n \to \infty$.

ii) Convergence (2.B.90) is equivalent to Assumption A.2.

Proof. First notice that under Assumption A.3, it follows by Lemma 2.6 that, as $n \to \infty$,

$$E(\eta_{j,x}) = 1 + O\left(\frac{\log j}{j}\right) + o(1) \tag{2.B.91}$$

uniformly in $j = 1, ..., m$.

i) Write

$$\frac{1}{m}\sum_{j=1}^{m} \eta_{j,x} = \frac{1}{m}\sum_{j=1}^{m} \left(\eta_{j,x} - E(\eta_{j,x})\right) + \frac{1}{m}\sum_{j=1}^{m} E(\eta_{j,x}) := S_{m,1} + S_{m,2}. \tag{2.B.92}$$

To prove (2.B.90), it suffices to show that, as $n \to \infty$,

$$S_{m,1} \xrightarrow{p} 0 \tag{2.B.93}$$

and

$$S_{m,2} \xrightarrow{p} 1. \tag{2.B.94}$$

We first show convergence (2.B.93). Let $\varepsilon > 0$. Since $cov(\eta_{j,x}, \eta_{k,x}) \to 0$ when $j \neq k$ and $j, k \to \infty$, we can find $K_\varepsilon \in \mathbb{N}$ large enough, such that $\left|cov(\eta_{j,x}, \eta_{k,x})\right| < \frac{\varepsilon}{2}$, for any $j \neq k$ and $j, k > K_\varepsilon$. We have that

$$S_{m,1} = \frac{1}{m}\sum_{j=1}^{K_\varepsilon}(\eta_{j,x} - E(\eta_{j,x})) + \frac{1}{m}\sum_{j=K_\varepsilon+1}^{m}(\eta_{j,x} - E(\eta_{j,x})) := R_{m,1} + R_{m,2}. \quad (2.B.95)$$

Then, from (2.B.91) it follows that, for a fixed $K_\varepsilon \in \mathbb{N}$,

$$E\left|R_{m,1}\right| \leq \frac{1}{m}2\sum_{j=1}^{K_\varepsilon} E(\eta_{j,x}) \leq \frac{1}{m}CK_\varepsilon = o(1), \quad (2.B.96)$$

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

$$\begin{aligned} E\left(R_{m,2}^2\right) &= \frac{1}{m^2}\sum_{j,k=K_\varepsilon+1}^{m} cov(\eta_{j,x}, \eta_{k,x}) \\ &\leq \frac{1}{m^2}\sum_{j,k=K_\varepsilon+1}^{m} \left|cov(\eta_{j,x}, \eta_{k,x})\right| \\ &\leq \frac{1}{m^2}\left(\sum_{j,k=K_\varepsilon+1: j\neq k}^{m} \frac{\varepsilon}{2} + \sum_{j=K_\varepsilon+1}^{m} C\right) \\ &\leq \frac{\varepsilon}{2} + \frac{1}{m}C \leq \varepsilon, \end{aligned} \quad (2.B.97)$$

as $n \to \infty$. From (2.B.96) and (2.B.97) we have that, for every $\varepsilon > 0$, $E\left|R_{m,1}\right| \to 0$ and $E\left|R_{m,2}\right| \leq \sqrt{\varepsilon}$, as $n \to \infty$. This implies that $E\left|S_{m,1}\right| \to 0$, as $n \to \infty$, and hence (2.B.93) holds.

Next we show (2.B.94). Using the bound (2.B.91), we have that

$$\begin{aligned} S_{m,2} &= \frac{1}{m}\sum_{j=1}^{m}\left(1 + O\left(\frac{\log j}{j}\right) + o(1)\right) \\ &= 1 + O\left(\frac{1}{m}\sum_{j=1}^{m}\frac{\log j}{j}\right) + o(1) \\ &= 1 + o(1), \end{aligned} \quad (2.B.98)$$

as $n \to \infty$, since

$$\begin{aligned} \frac{1}{m}\sum_{j=1}^{m}\frac{\log j}{j} &\leq \frac{1}{m}\ln m\sum_{j=1}^{m}\frac{1}{j} \leq \frac{\ln m}{m}\left(1 + \int_1^{m+1}\frac{1}{x}dx\right) \\ &= \frac{\ln m\,(1 + \ln(m+1))}{m} = o(1), \end{aligned} \quad (2.B.99)$$

as $n \to \infty$.

ii) By the definitions of $\eta_{j,x}$ and $\eta^*_{j,x}$, we have that

$$E\left|\frac{1}{m}\sum_{j=1}^{m}\left(\eta_{j,x}-\eta^*_{j,x}\right)\right| = \frac{1}{m}E\left|\sum_{j=1}^{m}\left(1-\frac{f_x(\lambda_j)}{c_{0,x}\lambda_j^{-\alpha_x}}\right)\eta_{j,x}\right|$$
$$\leq \frac{1}{m}\sum_{j=1}^{m}\left|1-\frac{f_x(\lambda_j)}{c_{0,x}\lambda_j^{-\alpha_x}}\right| E\left|\eta_{j,x}\right|$$
$$\leq C\frac{1}{m}\sum_{j=1}^{m}\left|1-\frac{f_x(\lambda_j)}{c_{0,x}\lambda_j^{-\alpha_x}}\right| = o(1), \qquad (2.B.100)$$

as $n \to \infty$, given that under Assumption A.3, $E\left|\eta_{j,x}\right| \leq C$ and $1-\frac{f_x(\lambda_j)}{c_{0,x}\lambda_j^{-\alpha_x}} = \frac{\lambda_j^{-\alpha_x}}{c_{0,x}\lambda_j^{-\alpha_x}}\left(c_{0,x}-h_x(\lambda_j)\right) = o(1)$ uniformly in $j = 1, ..., m$, as $n \to \infty$. From (2.B.100) it follows that

$$\frac{1}{m}\sum_{j=1}^{m}\left(\eta_{j,x}-\eta^*_{j,x}\right) \xrightarrow{p} 0, \qquad (2.B.101)$$

as $n \to \infty$. Hence, convergence (2.B.90) is equivalent to Assumption A.2. ■

Chapter 3

Local Whittle estimation for nonlinear time series

3.1 Introduction

In the previous chapter, we described sufficient conditions for the consistency of the LW estimator $\widehat{\alpha}_x$ of the memory parameter α_x, along with rates of convergence and expansions for $\widehat{\alpha}_x - \alpha_x$. The most attractive feature of our conditions is that they do not require the process $\{x_t\}_{t\in\mathbb{Z}}$ to be linear, and are satisfied in the case that it is. We now turn to examine the asymptotic behaviour of the LW estimator $\widehat{\alpha}_x$ of specific nonlinear processes. Under suitable assumptions, we establish consistency, convergence rate and the asymptotic distribution of $\widehat{\alpha}_x$, using our general results in Section 2.4 without needing to derive these asymptotic properties from first principles.

The starting point for our analysis is the signal plus noise model. Under such specification, the process $\{x_t\}_{t\in\mathbb{Z}}$ is written as

$$x_t = y_t + z_t, \tag{3.1.1}$$

where $\{y_t\}_{t\in\mathbb{Z}}$ is the signal process and $\{z_t\}_{t\in\mathbb{Z}}$ is the noise. The memory parameters α_y and α_z of $\{y_t\}_{t\in\mathbb{Z}}$ and $\{z_t\}_{t\in\mathbb{Z}}$, respectively, are such that $\alpha_y > \alpha_z$, so that $\alpha_x = \alpha_y$ and $\{y_t\}_{t\in\mathbb{Z}}$ signals the long-run behaviour of $\{x_t\}_{t\in\mathbb{Z}}$. Notice that since we are interested in estimation and inference on the memory parameter α_x, our pursuit would be trivial if the signal process were to be observed.

Although the model (3.1.1) looks rather simplistic at its first glance, it captures a wide range of data generating mechanisms. For example, macroeconomic and financial data often exhibit long memory and, involve measurement errors either occurring in the collection of the data or arising because the process of interest is not directly observed. Then, $\{y_t\}_{t\in\mathbb{Z}}$ represents the process that is subject to sampling errors or is unobserved, whereas $\{z_t\}_{t\in\mathbb{Z}}$ plays the role of measurement errors, which is usually taken to be a white noise sequence.

Furthermore, observations of macroeconomic and financial time series often exhibit long- and short-run dynamics, and in some cases cyclical/seasonal behaviour. The most common representation for modelling such behaviour is a structural model, under which $\{x_t\}_{t\in\mathbb{Z}}$ is written as

$$x_t = \mu_{t,x} + c_{t,x} + \eta_{t,x}, \tag{3.1.2}$$

where $\{\mu_{t,x}\}_{t\in\mathbb{Z}}$ denotes the long-run component of $\{x_t\}_{t\in\mathbb{Z}}$, $\{c_{t,x}\}_{t\in\mathbb{Z}}$ the cyclical/seasonal one and, $\{\eta_{t,x}\}_{t\in\mathbb{Z}}$ the short-run one. Given that $\alpha_\mu > \alpha_c, \alpha_\eta$, it is evident that (3.1.2) is special case of the signal plus noise model (3.1.1) with $y_t = \mu_{t,x}$ and $z_t = c_{t,x} + \eta_{t,x}$. Notice also that allowing for the presence of separate stochastic elements, the long-, medium- and short-run dynamics of $\{x_t\}_{t\in\mathbb{Z}}$ are not restricted to be driven by the same innovation process, as is the case for example with $ARFIMA(p, \frac{\alpha_x}{2}, q)$ and $ARFIMA(p, \frac{\alpha_x}{2}, \frac{\alpha_{\omega,x}}{2}, q)$ models.

In addition, the signal plus noise model (3.1.1) can incorporate nonlinear specifications, other than the model (3.1.2). If for example $\{x_t\}_{t\in\mathbb{Z}}$ is a nonlinear instantaneous transformation $G(.)$ of a Gaussian process $\{\xi_t\}_{t\in\mathbb{Z}}$, then as long as moments up to second order exist, $\{x_t\}_{t\in\mathbb{Z}}$ admits a Hermite expansion

$$x_t = G(\xi_t) = \sum_{k=k_0}^{\infty} \frac{c_k}{k!} H_k(\xi_t), \tag{3.1.3}$$

where $H_k(.)$ is the k-th Hermite polynomial, c_k is the k-th Hermite coefficient and k_0 is the Hermite rank, whose definitions are found in Section 3.4 below. Observe that if $\{x_t\}_{t\in\mathbb{Z}}$ exhibits long memory, then (3.1.3) can be written as (3.1.1) with $y_t = \frac{c_{k_0}}{k_0!} H_{k_0}(\xi_t)$ and $z_t = \sum_{k=k_0+1}^{\infty} \frac{c_k}{k!} H_k(\xi_t)$. From the properties of the Hermite polynomials given in (3.4.6) and (3.4.7), it follows that $\{y_t\}_{t\in\mathbb{Z}}$ and $\{z_t\}_{t\in\mathbb{Z}}$ are uncorrelated from each other, whereas the proof of Theorem 3.3 implies that the memory parameters of $\{y_t\}_{t\in\mathbb{Z}}$ and $\{z_t\}_{t\in\mathbb{Z}}$ satisfy $\alpha_y > \alpha_z$.

The evidence of nonlinearity and long-run persistence are particularly strong in the empirical finance literature. Returns series have often been found to display short memory in the levels but exhibit long memory when nonlinear transformations are taken, the squared and logarithmic squared being the most popular ones. A model capturing such features is the long memory stochastic volatility (LMSV) proposed independently by Breidt, Crato, and de Lima (1998) and Harvey (1998), which is a particular case of the class of stochastic volatility models introduced by Taylor (1994). Under the LMSV specification, the return sequence $\{r_t\}_{t\in\mathbb{Z}}$ is written as

$$r_t = \varepsilon_t \sigma_t, \tag{3.1.4}$$

where $\{\sigma_t\}_{t\in\mathbb{Z}}$ is the volatility process and $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is usually taken to be a sequence of i.i.d. variables with mean zero, finite variance and independent of $\{\sigma_t\}_{t\in\mathbb{Z}}$. The latter assumption guarantees that the returns is a sequence of uncorrelated zero mean random variables. The volatility process is further restricted to be of the form

$$\sigma_t = \exp\left(\frac{\xi_t}{2}\right), \tag{3.1.5}$$

where $\{\xi_t\}_{t\in\mathbb{Z}}$ is a long memory process independent of $\{\varepsilon_t\}_{t\in\mathbb{Z}}$. Then, the logarithmic squared returns are given by

$$\log r_t^2 = \xi_t + 2\log|\varepsilon_t|, \tag{3.1.6}$$

which is a special case of the signal plus noise model (3.1.1) with $x_t = \log r_t^2$, $y_t = \xi_t$ and $z_t = 2\log|\varepsilon_t|$, noticing that $\{z_t\}_{t\in\mathbb{Z}}$ is a sequence of i.i.d. random variables.

The asymptotic properties of the LW estimator were examined by Arteche (2004) and Hurvich, Moulines, and Soulier (2005) in the context of the LMSV model (3.1.4)-(3.1.5). As in (3.1.6), the authors transform the logarithmic squared returns into a signal plus noise model, and take the signal to follow a long memory linear process. In Arteche (2004), the noise process is taken to be a short memory linear process independent of the signal. On the other hand, Hurvich, Moulines, and Soulier (2005) restrict the noise to be a sequence of uncorrelated random variables, but allow for a specific form of contemporaneous correlation between the noise process and the innovation sequence in the linear representation of the signal. Under further regularity conditions, the aforementioned authors established the consistency and asymptotic distribution of the LW estimator.

The models considered by Arteche (2004) and Hurvich, Moulines, and Soulier (2005) constitute the only nonlinear representations under which the asymptotic properties of the LW estimator have been derived. Due to the restrictions imposed on the noise process and its dependence with the signal, the results of Arteche (2004) and Hurvich, Moulines, and Soulier (2005) cannot be employed for the models (3.1.2) and (3.1.3), and in the case of the LMSV model, rely heavily on the functional form (3.1.5) of the volatility process and the logarithmic squared transformation (3.1.6) employed.

In this chapter, we examine LW estimation for the signal plus noise model (3.1.1) using our general results of Section 2.4. We establish the consistency and asymptotic normality of the estimator using assumptions far less stringent than those in Arteche (2004) and Hurvich, Moulines, and Soulier (2005). In particular, we allow for the signal and noise processes to possibly have long, short or negative memory, the dependence between the signal and the noise series is left unrestricted, while no assumptions are imposed on the structure of the noise process other than mild conditions on its spectral density function. The flexibility of our conditions in the signal plus noise model allows us to further examine the LW estimator in the models (3.1.2) and (3.1.3), along with the LMSV model (3.1.4), without resorting to specification (3.1.5) and a logarithmic squared transformation of the return series.

Section 3.2 is devoted on studying the LW estimator for the signal plus noise model (3.1.1). The sections thereafter examine the LW estimator for the various subcases of the signal plus noise model described above. In particular, Section 3.3 considers the structural model (3.1.2), while Section 3.4 examines nonlinear transformations of a Gaussian process (3.1.3) and, the LMSV model (3.1.4) is further investigated in Section 3.5. Some concluding remarks are placed in Section 3.6, while the proofs of Sections 3.2-3.5 are found in Appendix 3.A of this chapter, that use a series of technical lemmas placed in Appendix 3.B. Throughout this chapter, we make use of the assumptions given in the previous chapter, see Section 2.3.

3.2 Signal plus noise process

Suppose that our observed process $\{x_t\}_{t\in\mathbb{Z}}$ is given by

$$x_t = y_t + z_t, \tag{3.2.1}$$

where $\{y_t\}_{t\in\mathbb{Z}}$ and $\{z_t\}_{t\in\mathbb{Z}}$ are covariance stationary processes having well defined spectral density functions and memory parameters that satisfy $\alpha_y > \alpha_z$.

The first part of the next theorem shows that if $\{y_t\}_{t\in\mathbb{Z}}$ and $\{z_t\}_{t\in\mathbb{Z}}$ satisfy Assumption A.3, then LW estimation applied to the process $\{x_t\}_{t\in\mathbb{Z}}$ consistently estimates $\alpha_x = \alpha_y$, as long as the signal process $\{y_t\}_{t\in\mathbb{Z}}$ satisfies Assumption A.2. The proof of consistency is based on demonstrating that $\{x_t\}_{t\in\mathbb{Z}}$ satisfies Assumptions A.1 and A.2 of Theorem 2.1. Furthermore, using (2.4.6) of Theorem 2.3, we provide an expansion for $\widehat{\alpha}_x - \alpha_y$ in terms of $\widehat{\alpha}_y - \alpha_y$, see (3.2.5) below. The latter expansion allows us to further investigate the rate of convergence and asymptotic distribution of $\widehat{\alpha}_x$, without needing to impose any additional assumptions on the noise process $\{z_t\}_{t\in\mathbb{Z}}$. In particular, in part ii) of the theorem, we further assume that $\{y_t\}_{t\in\mathbb{Z}}$ satisfies Assumptions A.4 and $\Delta_{m,y} \leq Cm^{\gamma}$, for some $0 < \gamma < 1$, and establish convergence rate for $\widehat{\alpha}_x$ by combining expansion (3.2.5) and Proposition 2.2 for $\{y_t\}_{t\in\mathbb{Z}}$. If moreover, $\{y_t\}_{t\in\mathbb{Z}}$ is a linear process satisfying Assumptions A.6 and A.8 and the bandwidth parameter m is such that $m = o\left(\frac{n}{\log^{\frac{2}{\beta_y}} n}\right)$, then part iii) of the theorem provides convergence rate for $\widehat{\alpha}_x$. Under additional conditions, Assumption A.7 on $\{y_t\}_{t\in\mathbb{Z}}$ and $m = o\left(n^{\frac{2r}{2r+1}}\right)$ with $r = \min\left\{\beta_y, \frac{\alpha_y - \alpha_z}{2}\right\}$, the asymptotic distribution of $\widehat{\alpha}_x$ is also derived. The results of part iii) of the theorem use expansion (3.2.5) in conjunction with Theorem 2.3 and Propositions 2.2 and 2.3 applied on $\{y_t\}_{t\in\mathbb{Z}}$.

Theorem 3.1

Suppose that $\{x_t\}_{t\in\mathbb{Z}}$ is as defined in (3.2.1), and that the spectral densities $f_y(.)$ and $f_z(.)$ are such that

$$f_y(\lambda) = c_{0,y}\lambda^{-\alpha_y} + o(\lambda^{-\alpha_y}), \quad f_z(\lambda) \leq C\lambda^{-\alpha_z}, \quad \text{as } \lambda \to 0+, \tag{3.2.2}$$

with

$$-1 < \alpha_z < \alpha_y < 1. \tag{3.2.3}$$

i) If $\{y_t\}_{t\in\mathbb{Z}}$ satisfies Assumption A.2, then, as $n \to \infty$,

$$\widehat{\alpha}_x \xrightarrow{p} \alpha_y. \tag{3.2.4}$$

Moreover,

$$\widehat{\alpha}_x - \alpha_y = (\widehat{\alpha}_y - \alpha_y)(1 + o_p(1)) + O_p\left(\left(\frac{m}{n}\right)^{\frac{\alpha_y-\alpha_z}{2}} + \frac{\log m}{m}\right). \tag{3.2.5}$$

ii) If $\{y_t\}_{t\in\mathbb{Z}}$ satisfies Assumptions A.4, and $\Delta_{m,y} \leq Cm^{\gamma}$ for some $0 < \gamma < 1$, then

$$\widehat{\alpha}_x - \alpha_y = O_p\left(m^{\gamma-1}\log m + m^{-\frac{1}{2}} + \left(\frac{m}{n}\right)^{\beta_y} + \left(\frac{m}{n}\right)^{\frac{\alpha_y-\alpha_z}{2}}\right). \tag{3.2.6}$$

iii) If $\{y_t\}_{t\in\mathbb{Z}}$ is a linear process satisfying Assumptions A.4, A.6 and A.8, and $m = o\left(\frac{n}{\log^{\frac{2}{\beta_y}} n}\right)$, then

$$\widehat{\alpha}_x - \alpha_y = O_p\left(m^{-\frac{1}{2}} + \left(\frac{m}{n}\right)^{r}\right), \tag{3.2.7}$$

where $r = \min\left\{\beta_y, \frac{\alpha_y - \alpha_z}{2}\right\}$. *If furthermore, Assumption A.7 holds, then*

$$m^{\frac{1}{2}}(\widehat{\alpha}_x - \alpha_y) \xrightarrow{d} N(0,1), \quad \text{as } n \to \infty, \tag{3.2.8}$$

with $m = o\left(n^{\frac{2r}{2r+1}}\right)$.

Theorem 3.1 shows that the LW estimator $\widehat{\alpha}_x$ is a consistent estimator of α_y. The assumptions imposed on the signal process essentially require that $\{y_t\}_{t\in\mathbb{Z}}$ is such, that if we applied LW estimation to $\{y_t\}_{t\in\mathbb{Z}}$, then the memory parameter α_y would be consistently estimated. On the other hand, the assumptions on the noise process are very mild, as they only require that the spectral density function $f_z(.)$ to be such that $\{z_t\}_{t\in\mathbb{Z}}$ is indeed the noise process, see (3.2.2). By imposing further assumptions only on the signal process, part ii) and iii) establish convergence rates, while in part iii) we also derive the asymptotic distribution of $\widehat{\alpha}_x$ by assuming that the signal process $\{y_t\}_{t\in\mathbb{Z}}$ satisfies the conditions of Robinson (1995b) in his theorem concerning the asymptotic distribution. Notice that the asymptotic properties of the LW estimator $\widehat{\alpha}_x$ are unaffected by the presence of the additive noise process. However, there are two important points that arise from the results of Theorem 3.1.

Remark 3.1

Relations (3.2.6) and (3.2.7) of Theorem 3.1 indicate that the presence of the noise $\{z_t\}_{t\in\mathbb{Z}}$ *worsens the rate of convergence of the LW estimator. For example, if the assumptions of Theorem 3.1 part iii) hold and we further assume that* $\beta_y = 2$, *as is done frequently in empirical applications, then the rate of convergence of* $\widehat{\alpha}_x$ *is* $O_p\left(m^{-\frac{1}{2}} + \left(\frac{m}{n}\right)^{\frac{\alpha_y - \alpha_z}{2}}\right)$, *whereas without the noise process, the rate of convergence of* $\widehat{\alpha}_x$ *would be of order* $O_p\left(m^{-\frac{1}{2}} + \left(\frac{m}{n}\right)^2\right)$. *Relations (3.2.6) and (3.2.7) also suggest that the bigger the difference between the memory parameter of the signal and the noise, the better is the rate of convergence of the estimator.*

Remark 3.2

In part iii) of Theorem 3.1, the condition on the bandwidth parameter m *for establishing the asymptotic distribution of* $\widehat{\alpha}_x$ *has been affected by the presence of the noise process. Here* $m = o\left(n^{\frac{2r}{2r+1}}\right)$ with $r = \min\left\{\beta_y, \frac{\alpha_y - \alpha_z}{2}\right\}$, *while in the case without the noise, the condition would be* $m = o\left(n^{\frac{2\beta_y}{2\beta_y+1}}\right)$. *If we assume that* $\beta_y =$ 2, *then the condition on the bandwidth parameter becomes* $m = o\left(n^{\frac{\alpha_y - \alpha_z}{\alpha_y - \alpha_z + 1}}\right)$. *Hence, in the implementation of (3.2.8), the problem of bandwidth choice arises, as* $\alpha_y - \alpha_z$ *is an unknown quantity. The latter is the reason why the rule for choosing* *the optimal bandwidth suggested by Henry and Robinson (1996) cannot be implemented, and notice also that if it could, it would suggest a bandwidth parameter smaller than in the case without the noise process.*

Three are the main advantages of our conditions in Theorem 3.1. Firstly, the dependence between the signal and the noise process is left unrestricted. Secondly, we do not require the noise to be a linear process exhibiting short memory. Lastly, the signal is not assumed to be long memory linear process, although linearity of the signal is needed for establishing the asymptotic distribution of $\widehat{\alpha}_x$.

Our results and remarks are in line with that of Arteche (2004) and Hurvich, Moulines, and Soulier (2005), but require less restrictive conditions than those imposed by these authors. In the proof of consistency in Arteche (2004), the signal $\{y_t\}_{t\in\mathbb{Z}}$ is taken to be a linear process satisfying Assumptions A.4 and A.6 with $\alpha_y > 0$, the noise $\{z_t\}_{t\in\mathbb{Z}}$ is required to be a linear process satisfying Assumption A.3 and A.6 with $\alpha_z = 0$ and, the signal and noise processes are assumed to be independent from each other. Hence, our consistency conditions in part i) of Theorem 3.1 are more general than those of Arteche (2004), noticing that under his assumption on the signal process, we have from Proposition 2.3 that Assumption A.2 is satisfied for $\{y_t\}_{t\in\mathbb{Z}}$. Moreover, for deriving the asymptotic distribution of $\widehat{\alpha}_x$, Arteche (2004) further imposed Assumptions A.7 and A.8 on $\{y_t\}_{t\in\mathbb{Z}}$, so that our result in part iii) of Theorem 3.1 relies on less stringent conditions. On the other hand, for the proof of consistency of $\widehat{\alpha}_x$, Hurvich, Moulines, and Soulier (2005) assumed that $\{y_t\}_{t\in\mathbb{Z}}$ is a linear process satisfying Assumptions A.3 with $\alpha_y > 0$ and their Assumption H.3, while $\{z_t\}_{t\in\mathbb{Z}}$ is taken to be a white noise sequence contemporaneously correlated with the innovation process $\{\varepsilon_{t,y}\}_{t\in\mathbb{Z}}$ in the linear representation of $\{y_t\}_{t\in\mathbb{Z}}$. The authors also established the asymptotic distribution of the LW estimator, by further restricting the structure on the noise process $\{z_t\}_{t\in\mathbb{Z}}$ and its dependence with $\{\varepsilon_{t,y}\}_{t\in\mathbb{Z}}$, and assuming that the signal satisfies Assumption A.8. Notice that Assumption H.3 of Hurvich, Moulines, and Soulier (2005) is a milder version of our Assumptions A.4, A.6 and A.7. Besides this point, our results in Theorem 3.1 part i) and iii) are more general than those of Hurvich, Moulines, and Soulier (2005).

We have already noted that the presence of the noise $\{z_t\}_{t\in\mathbb{Z}}$ does not alter the asymptotic properties of the LW estimator $\widehat{\alpha}_x$, but does worsen its convergence rate. In order to investigate further the finite sample behaviour of $\widehat{\alpha}_x$, we now restrict the dependence between the signal and noise processes as in Arteche (2004) and Hurvich, Moulines, and Soulier (2005).

Let's first assume that the series $\{y_t\}_{t\in\mathbb{Z}}$ and $\{z_t\}_{t\in\mathbb{Z}}$ are uncorrelated from each other. Then, the spectral density function of $\{x_t\}_{t\in\mathbb{Z}}$ satisfies

$$f_x(\lambda) = f_y(\lambda) + f_z(\lambda), \tag{3.2.9}$$

for all $\lambda \in (-\pi, \pi]$. If $\{y_t\}_{t\in\mathbb{Z}}$ and $\{z_t\}_{t\in\mathbb{Z}}$ satisfy Assumption A.4 and $\alpha_y - \alpha_z < \beta_y$, we have that, as $\lambda \to 0+$,

$$\begin{aligned} f_x(\lambda) &= c_{0,y}\lambda^{-\alpha_y} + c_{1,y}\lambda^{\beta_y-\alpha_y} + o(\lambda^{\beta_y-\alpha_y}) \\ &\quad + c_{0,z}\lambda^{-\alpha_z} + c_{1,z}\lambda^{\beta_z-\alpha_z} + o\left(\lambda^{\beta_z-\alpha_z}\right) \\ &= \lambda^{-\alpha_y}\left(c_{0,y} + c_{0,z}\lambda^{\alpha_y-\alpha_z} + o\left(\lambda^{\alpha_y-\alpha_z}\right)\right), \end{aligned} \tag{3.2.10}$$

so that $\{x_t\}_{t\in\mathbb{Z}}$ satisfies Assumption A.4 with $\alpha_x = \alpha_y$, $c_{0,x} = c_{0,y}$, $c_{1,x} = c_{0,z}$ and $\beta_x = \alpha_y - \alpha_z$ there. If the assumptions of Theorem 3.1 part i) hold, then we conclude from (2.4.7) of Theorem 2.3 and Remark 2.1 that the finite sample bias of $\widehat{\alpha}_x$ is mainly determined by $-\left(\frac{m}{n}\right)^{\alpha_y-\alpha_z}\frac{c_{0,z}}{c_{0,y}}B_{\alpha_y-\alpha_z}$.

Let's assume next that $\{y_t\}_{t\in\mathbb{Z}}$ and $\{z_t\}_{t\in\mathbb{Z}}$ are correlated processes in the sense that the innovation terms $\{\varepsilon_{t,y}\}_{t\in\mathbb{Z}}$ and $\{\varepsilon_{t,z}\}_{t\in\mathbb{Z}}$ in the representation (2.5.1) are such that $E\left(\varepsilon_{t,y}\varepsilon_{s,z}\right) = \rho$ if $t = s$ and, $E\left(\varepsilon_{t,y}\varepsilon_{s,z}\right) = 0$ if $t \neq s$. Then, the spectral density function of $\{x_t\}_{t\in\mathbb{Z}}$ satisfies

$$f_x(\lambda) = f_y(\lambda) + f_z(\lambda) + 2\rho\,\mathrm{Re}\left(\phi_y(\lambda)\overline{\phi_z(\lambda)}\right), \tag{3.2.11}$$

for all $\lambda \in (-\pi, \pi]$. If $\{y_t\}_{t\in\mathbb{Z}}$ and $\{z_t\}_{t\in\mathbb{Z}}$ satisfy Assumption A.4 and $\frac{\alpha_y-\alpha_z}{2} < \beta_y$, we have that, as $\lambda \to 0+$,

$$\begin{aligned} f_x(\lambda) &= c_{0,y}\lambda^{-\alpha_y} + c_{1,y}\lambda^{\beta_y-\alpha_y} + o(\lambda^{\beta_y-\alpha_y}) \\ &\quad + c_{0,z}\lambda^{-\alpha_z} + c_{1,z}\lambda^{\beta_z-\alpha_z} + o\left(\lambda^{\beta_z-\alpha_z}\right) \\ &\quad + 2\rho\sqrt{c_{0,y}c_{0,z}}\lambda^{-\frac{\alpha_y+\alpha_z}{2}}\left(1 + o\left(1\right)\right) \\ &= \lambda^{-\alpha_y}\left(c_{0,y} + 2\rho\sqrt{c_{0,y}c_{0,z}}\lambda^{\frac{\alpha_y-\alpha_z}{2}} + o\left(\lambda^{\frac{\alpha_y-\alpha_z}{2}}\right)\right), \end{aligned} \tag{3.2.12}$$

so that $\{x_t\}_{t\in\mathbb{Z}}$ satisfies Assumption A.4 with $\alpha_x = \alpha_y$, $c_{0,x} = c_{0,y}$, $c_{1,x} = 2\rho\sqrt{c_{0,y}c_{0,z}}$ and $\beta_x = \frac{\alpha_y-\alpha_z}{2}$ there. If the assumptions of Theorem 3.1 part i) hold, then we have from (2.4.7) of Theorem 2.3 and Remark 2.1 that the finite sample bias of $\widehat{\alpha}_x$ is overall determined by $-\left(\frac{m}{n}\right)^{\frac{\alpha_y-\alpha_z}{2}} 2\rho\sqrt{\frac{c_{0,z}}{c_{0,y}}}B_{\frac{\alpha_y-\alpha_z}{2}}$.

Remark 3.3

In the signal plus noise model, under the assumptions discussed above, the finite sample bias of the estimator $\widehat{\alpha}_x$ tends to be negative if the signal and noise processes *are uncorrelated or positively dependent, and positive if the two processes are negatively dependent. The finite sample bias increases with an increase in the signal-to-noise ratio $\frac{c_{0,z}}{c_{0,y}}$, and convergences faster to 0 the bigger the difference $\alpha_y - \alpha_z$ is. Furthermore, the finite sample bias tends to be smaller and rate of convergence is faster, when the signal and noise processes are uncorrelated.*

Theorem 3.1, together with results in Section 3.4, enable us to study the asymptotic behaviour of the LW estimator under the nonlinear models discussed in the Introduction of this chapter.

3.3 Structural model

Suppose now that our observed process $\{x_t\}_{t\in\mathbb{Z}}$ is decomposed as

$$x_t = \mu_{t,x} + c_{t,x} + \eta_{t,x}, \tag{3.3.1}$$

where $\{\mu_{t,x}\}_{t\in\mathbb{Z}}$ is the trend of $\{x_t\}_{t\in\mathbb{Z}}$, $\{c_{t,x}\}_{t\in\mathbb{Z}}$ is its cyclical component of known frequency $\omega_x \neq 0$ and, $\{\eta_{t,x}\}_{t\in\mathbb{Z}}$ corresponds to the short-run component of $\{x_t\}_{t\in\mathbb{Z}}$. The processes $\{\mu_{t,x}\}_{t\in\mathbb{Z}}$, $\{c_{t,x}\}_{t\in\mathbb{Z}}$ and $\{\eta_{t,x}\}_{t\in\mathbb{Z}}$ are taken to be covariance stationary having well defined spectral density functions. Given the decomposition (3.3.1) of $\{x_t\}_{t\in\mathbb{Z}}$ and the interpretation of its three components, it is natural to assume that the cyclical and short-run component do not exhibit long-run persistence, so that $\alpha_{\mu_x} > \alpha_{c_x}, \alpha_{\eta_x} = 0$. Then, model (3.3.1) is a special case of the signal plus noise model (3.1.1) with $y_t = \mu_{t,x}$ and $z_t = c_{t,x} + \eta_{t,x}$. Notice that we are allowing for the cyclical component to be persistent, i.e. $\alpha_{\omega,c_x} > 0$.

The next theorem shows that, under assumptions similar to those in Theorem 3.1, the LW estimator $\widehat{\alpha}_x$ is a consistent estimator of the memory parameter α_{μ_x} and is asymptotically normally distributed. The proof of the next theorem is a simple extension of Theorem 3.1 and is therefore omitted.

Theorem 3.2
Suppose that $\{x_t\}_{t\in\mathbb{Z}}$ is as defined in (3.3.1), and that the spectral densities $f_{\mu_x}(.)$, $f_{c_x}(.)$ and $f_{\eta_x}(.)$ satisfy Assumption A.3 with $\alpha_{\mu_x} > \alpha_{c_x}, \alpha_{\eta_x} = 0$.
i) If $\{\mu_{t,x}\}_{t\in\mathbb{Z}}$ satisfies Assumption A.2, then, as $n \to \infty$,

$$\widehat{\alpha}_x \xrightarrow{p} \alpha_{\mu_x}. \tag{3.3.2}$$

Moreover,

$$\widehat{\alpha}_x - \alpha_{\mu_x} = (\widehat{\alpha}_{\mu_x} - \alpha_{\mu_x})(1 + o_p(1)) + O_p\left(\left(\frac{m}{n}\right)^{\frac{\alpha_{\mu_x}}{2}} + \frac{\log m}{m}\right). \tag{3.3.3}$$

ii) If $\{\mu_{t,x}\}_{t\in\mathbb{Z}}$ satisfies Assumption A.4 and $\Delta_{m,\mu_{t,x}} \leq Cm^{\gamma}$ for some $0 < \gamma < 1$, then

$$\widehat{\alpha}_x - \alpha_{\mu_x} = O_p\left(m^{\gamma-1}\log m + m^{-\frac{1}{2}} + \left(\frac{m}{n}\right)^{\beta_{\mu_x}} + \left(\frac{m}{n}\right)^{\frac{\alpha_{\mu_x}}{2}}\right). \tag{3.3.4}$$

iii) If $\{\mu_{t,x}\}_{t\in\mathbb{Z}}$ is a linear process satisfying Assumptions A.4, A.6 and A.8, and $m = o\left(\frac{n}{\log^{\frac{2}{\beta_{\mu_x}}} n}\right)$, then

$$\widehat{\alpha}_x - \alpha_{\mu_x} = O_p\left(m^{-\frac{1}{2}} + \left(\frac{m}{n}\right)^{r}\right), \tag{3.3.5}$$

where $r = \min\left\{\beta_{\mu_x}, \frac{\alpha_{\mu_x}}{2}\right\}$. If furthermore, Assumption A.7 holds, then

$$m^{\frac{1}{2}}(\widehat{\alpha}_x - \alpha_{\mu_x}) \xrightarrow{d} N(0,1), \quad \text{as } n \to \infty, \tag{3.3.6}$$

with $m = o\left(n^{\frac{2r}{2r+1}}\right)$.

Similar observations to those following Theorem 3.1 can be made here. However, there is one important point that we need to address in the case that the frequency ω_x of the cyclical component is rather small. The LW estimator consistently estimates the memory parameter α_{μ_x}, even if the cyclical component $\{c_{t,x}\}_{t\in\mathbb{Z}}$ is stronger than the trend $\{\mu_{t,x}\}_{t\in\mathbb{Z}}$. Essentially, for very big samples we have that $\lambda_m \ll \omega_x$, so that no information is included from the frequency ω_x and $\{c_{t,x}\}_{t\in\mathbb{Z}}$ acts as an additive noise with $\alpha_{c_x} = 0$. However, for small samples and when the frequency ω_x is small, it can happen that $\omega_x \ll \lambda_m$. Suppose now that the spectral density function of $\{c_{t,x}\}_{t\in\mathbb{Z}}$ is such that

$$f_{c_x}(\lambda) \sim c_{0,\omega,c_x}\left|\lambda - \omega_x\right|^{-\alpha_{\omega,x}}, \quad \text{as } \lambda \to \omega_x. \tag{3.3.7}$$

Hence, if $\omega_x \ll \lambda_m$, the cyclical component $\{c_{t,x}\}_{t\in\mathbb{Z}}$ behaves as if it satisfied Assumption A.3 with $\alpha_{c_x} = \alpha_{\omega,x}$. This observation leads to the following remark.

Remark 3.4

Suppose that the sample size n and the frequency ω_x are such that $\omega_x \ll \lambda_m$. If $\alpha_{\mu_x} > \alpha_{\omega,c_x}$, then the cyclical component $\{c_{t,x}\}_{t\in\mathbb{Z}}$ still acts as an additive noise and the trend component $\{\mu_{t,x}\}_{t\in\mathbb{Z}}$ as the signal. However, when $\alpha_{\mu_x} < \alpha_{\omega,c_x}$, these roles are reversed, so that $\{c_{t,x}\}_{t\in\mathbb{Z}}$ behaves as the signal and $\{\mu_{t,x}\}_{t\in\mathbb{Z}}$ as the noise. In the latter case, the estimator $\widehat{\alpha}_x$ is likely to be biased towards α_{ω,c_x} in finite samples.

3.4 Nonlinear functions of a Gaussian process

In this section, we consider estimation and inference on the memory parameter α_x, when the process $\{x_t\}_{t\in\mathbb{Z}}$ is an instantaneous nonlinear transformation of a Gaussian process $\{\xi_t\}_{t\in\mathbb{Z}}$. More formally,

$$x_t = G(\xi_t), \tag{3.4.1}$$

where $\{\xi_t\}_{t\in\mathbb{Z}}$ is a stationary zero mean and unit variance Gaussian sequence, and $G : \mathbb{R} \longrightarrow \mathbb{R}$ is a measurable function such that

$$E(G(\xi_t)) = 0 \quad \text{and} \quad E(G^2(\xi_t)) < \infty. \tag{3.4.2}$$

Notice that the conditions on the moments of $\{\xi_t\}_{t\in\mathbb{Z}}$ are not restrictive, since the sequence $\{\xi_t\}_{t\in\mathbb{Z}}$ can always be standardized by a transformation $G(.)$. Also, although we assume that $\{\xi_t\}_{t\in\mathbb{Z}}$ is a Gaussian sequence, the resulting process $\{x_t\}_{t\in\mathbb{Z}}$ is strictly stationary having finite second moments and unrestricted marginal probability density function.

We start by determining the relationship between the autocovariance functions of $\{x_t\}_{t\in\mathbb{Z}}$ and $\{\xi_t\}_{t\in\mathbb{Z}}$. For that purpose, we use the Hermite expansion technique on $\{x_t\}_{t\in\mathbb{Z}}$, developed by Taqqu (1979) and Dobrushin and Major (1979).

Hermite expansion

Under the assumptions $E(G(\xi_t)) = 0$ and $E(G^2(\xi_t)) < \infty$, the sequence $\{x_t\}_{t\in\mathbb{Z}}$ can be written as the sum of Hermite polynomials

$$x_t = \sum_{k=1}^{\infty} \frac{c_k}{k!} H_k(\xi_t), \tag{3.4.3}$$

where $H_k(x)$ is the k-th Hermite polynomial defined as

$$H_k(x) = (-1)^k e^{\frac{x^2}{2}} \frac{d^k(e^{-\frac{x^2}{2}})}{dx^k}, \quad x \in \mathbb{R}, \tag{3.4.4}$$

and c_k is the k-th Hermite coefficient given by

$$c_k = E(G(\xi_t)H_k(\xi_t)), \quad k \geq 0. \tag{3.4.5}$$

Note that the summation in (3.4.3) starts from $k = 1$, since we have assumed that $E(G(\xi_t)) = 0$, which implies that $c_0 = E(G(\xi_t)) = 0$, given that $H_0(x) = 1$. The Hermite polynomials have the following properties:

$$E(H_k(\xi_t)) = 0, \quad E(H_k(\xi_t)H_k(\xi_s)) = k!\gamma_\xi^k(t-s), \quad k \geq 1, \tag{3.4.6}$$

and

$$E(H_k(\xi_t)H_l(\xi_s)) = 0, \quad k,l \geq 1, \ k \neq l. \tag{3.4.7}$$

The minimal integer $k_0 \geq 1$ such that $c_{k_0} \neq 0$ is referred to as the Hermite rank of $G(.)$. Then, (3.4.3) can be written as

$$x_t = \sum_{k=k_0}^{\infty} \frac{c_k}{k!} H_k(\xi_t). \tag{3.4.8}$$

The assumptions on $\{x_t\}_{t\in\mathbb{Z}}$ and $\{\xi_t\}_{t\in\mathbb{Z}}$, together with the two properties (3.4.6) and (3.4.7), imply that

$$\gamma_x(\tau) = \sum_{k=k_0}^{\infty} \frac{c_k^2}{k!}\gamma_\xi^k(\tau) \quad \text{and} \quad \gamma_x(0) = E(x_t^2) = \sum_{k=k_0}^{\infty} \frac{c_k^2}{k!} < \infty. \tag{3.4.9}$$

In the next proposition we further relate the autocovariance function of $\{x_t\}_{t\in\mathbb{Z}}$ with that of $\{\xi_t\}_{t\in\mathbb{Z}}$.

Proposition 3.1

The autocovariance function of $\{x_t\}_{t\in\mathbb{Z}}$ satisfies

$$\gamma_x(\tau) = \gamma_\xi^{k_0}(\tau)\left(\frac{c_{k_0}^2}{k_0!} + R_\tau\right), \tag{3.4.10}$$

where

$$R_\tau \to 0, \quad \text{as } \tau \to \infty. \tag{3.4.11}$$

We now use Proposition 3.1 to investigate the summability properties of the autocovariance function of $\{x_t\}_{t\in\mathbb{Z}}$ and $\{\xi_t\}_{t\in\mathbb{Z}}$, which allows us to describe the connection between the dependence structure of $\{x_t\}_{t\in\mathbb{Z}}$ and $\{\xi_t\}_{t\in\mathbb{Z}}$. The following proposition shows why the Hermite rank k_0 of the function $G(.)$ plays an important role in describing the dependence structure of $\{x_t\}_{t\in\mathbb{Z}}$.

Proposition 3.2

It holds that

$$\frac{1}{\gamma_x(0)}\sum_{\tau\in\mathbb{Z}}|\gamma_x(\tau)| \leq \sum_{\tau\in\mathbb{Z}}\left|\gamma_\xi(\tau)\right|^{k_0} \leq C\sum_{\tau\in\mathbb{Z}}|\gamma_x(\tau)|. \tag{3.4.12}$$

From the results of Proposition 3.2, we have the following remark.

Remark 3.5

Proposition 3.2 shows that if $\sum_{\tau\in\mathbb{Z}}\left|\gamma_\xi(\tau)\right| < \infty$, then $\sum_{\tau\in\mathbb{Z}}|\gamma_x(\tau)| < \infty$. On the other hand, when $\sum_{\tau\in\mathbb{Z}}\left|\gamma_\xi(\tau)\right| = \infty$, Proposition 3.2 implies that if the Hermite rank k_0 is *such that* $\sum_{\tau\in\mathbb{Z}} |\gamma_\xi(\tau)|^{k_0} < \infty$, *then* $\sum_{\tau\in\mathbb{Z}} |\gamma_x(\tau)| < \infty$, *while if* $\sum_{\tau\in\mathbb{Z}} |\gamma_\xi(\tau)|^{k_0} = \infty$, *then* $\sum_{\tau\in\mathbb{Z}} |\gamma_x(\tau)| = \infty$. *Hence, transformations of a Gaussian process do not increase the degree of persistence.*

We introduce the following assumptions:

S.1 The autocovariance function $\gamma_x(.)$ of $\{x_t\}_{t\in\mathbb{Z}}$ is absolutely summable,

$$\sum_{\tau\in\mathbb{Z}} |\gamma_x(\tau)| < \infty. \tag{3.4.13}$$

L.1 The spectral density function $f_x(.)$ satisfies Assumption A.3 with $0 < \alpha_x < 1$.

L.2 The spectral density function $f_\xi(.)$ satisfies Assumption A.4 with $0 < \alpha_\xi < 1$ and $\beta_\xi = 2$. Furthermore, the autocovariance function $\gamma_\xi(.)$ of $\{\xi_t\}_{t\in\mathbb{Z}}$ has the property

$$\gamma_\xi(\tau) \sim c_\xi \tau^{-1+\alpha_\xi}, \quad \text{as } \tau \to \infty, \tag{3.4.14}$$

with $0 < |c_\xi| < \infty$.

In the current and next section, we refer to $\{x_t\}_{t\in\mathbb{Z}}$ having short or long memory, when the Assumptions S.1 or L.1-L.2 are satisfied, respectively. In the short memory case, Assumption S.1 is very general as it does not place directly any restrictions on the underlying process $\{\xi_t\}_{t\in\mathbb{Z}}$. It requires the spectral density function of $\{x_t\}_{t\in\mathbb{Z}}$ to be bounded and bounded away from zero in $(-\pi, \pi]$, which implies that Assumption A.3 is satisfied with $\alpha_x = 0$, but excludes the possibility that $\{x_t\}_{t\in\mathbb{Z}}$ has a persistent component of frequency $\omega_x \neq 0$. In the case of long memory, the assumption placed on the spectral density function of $\{x_t\}_{t\in\mathbb{Z}}$ is the one that we have considered throughout the previous and the current chapters. We also restrict the spectral density function of $\{\xi_t\}_{t\in\mathbb{Z}}$ to satisfy Assumption A.4 with smoothness parameter $\beta_\xi = 2$ and the autocovariance function $\gamma_\xi(.)$ of $\{\xi_t\}_{t\in\mathbb{Z}}$ to have the long memory type decay. Assumption L.2 is needed to establish the relationship between the memory parameters α_x and α_ξ. Notice that Assumption L.2 is satisfied, if for example, $\{\xi_t\}_{t\in\mathbb{Z}}$ is a Gaussian $ARFIMA(p, \frac{\alpha_\xi}{2}, q)$ process.

Next, denote by $f_\xi^{(*k)}(.)$ the k-th order convolution of the spectral density function $f_\xi(.)$,

$$f_\xi^{(*k)}(\lambda) = \int_{-\pi}^{\pi} \ldots \int_{-\pi}^{\pi} f_\xi(\lambda - l_1 - \ldots - l_{k-1}) f_\xi(l_1) \ldots f_\xi(l_{k-1}) dl_1 \ldots dl_{k-1}, \tag{3.4.15}$$

where we assume that $f_\xi(.)$ is periodically extended in $\mathbb{R}$. The next proposition establishes the relationship between the spectral density functions $f_x(.)$ and $f_\xi(.)$.

Proposition 3.3

Under Assumption S.1 or L.2, the spectral density function $f_x(.)$ can be written as

$$f_x(\lambda) = \sum_{k=k_0}^{\infty} \frac{c_k^2}{k!} f_\xi^{(*k)}(\lambda). \tag{3.4.16}$$

Lemma 3.1 in Appendix 3.B examines, under Assumption L.2, the behaviour of $f_\xi^{(*k)}(.)$, for $k \geq 2$. Next, we combine Lemma 3.1 with Proposition 3.3 to analyze the spectral density function of $\{x_t\}_{t\in\mathbb{Z}}$ around the zero frequency.

Proposition 3.4

Suppose that $\{\xi_t\}_{t\in\mathbb{Z}}$ satisfies Assumption L.2.

i) If $k_0(1-\alpha_\xi) < 1$, then the spectral density function of $\{x_t\}_{t\in\mathbb{Z}}$ satisfies

$$f_x(\lambda) = \frac{c_{k_0}^2}{k_0!} f_\xi^{(*k_0)}(\lambda)(1+o(1)) = c_{k_0}^2 s_{k_0} \lambda^{-\alpha_x} + o\left(\lambda^{-\alpha_x}\right), \tag{3.4.17}$$

as $\lambda \to 0+$, with

$$\alpha_x = 1 - k_0(1-\alpha_\xi), \tag{3.4.18}$$

and some $0 < s_{k_0} < \infty$.

ii) If $k_0(1-\alpha_\xi) = 1$, then for any arbitrarily small $\delta > 0$, the spectral density function of $\{x_t\}_{t\in\mathbb{Z}}$ satisfies

$$f_x(\lambda) \leq C\lambda^{-\delta}, \tag{3.4.19}$$

as $\lambda \to 0+$.

iii) If $k_0(1-\alpha_\xi) > 1$, then the spectral density function of $\{x_t\}_{t\in\mathbb{Z}}$ is bounded,

$$f_x(\lambda) \leq C, \tag{3.4.20}$$

for all $\lambda \in (-\pi, \pi]$.

Finally, we combine the results of the propositions above with the results in Section 3.4 and Theorem 3.1 to investigate the asymptotic properties of $\widehat{\alpha}_x$. In the case of short memory, only Assumption S.1 is required for the proof of consistency, while in the long memory case, assumptions in the spirit of Theorem 3.1 are imposed. We also derive the rate of convergence and asymptotic distribution of $\widehat{\alpha}_x$ in the special case that $\{x_t\}_{t\in\mathbb{Z}}$ exhibits long memory and the Hermite rank of $G(.)$ is 1, again by placing assumptions in line with Theorem 3.1.

Theorem 3.3
Suppose that $\{x_t\}_{t\in\mathbb{Z}}$ *is as defined in (3.4.1).*
i) If $\{x_t\}_{t\in\mathbb{Z}}$ *satisfies Assumption S.1, then, as* $n \to \infty$,

$$\widehat{\alpha}_x \xrightarrow{p} \alpha_x = 0. \tag{3.4.21}$$

ii) If $\{x_t\}_{t\in\mathbb{Z}}$ *satisfies Assumptions L.1 and L.2, and* m *is such that* $n^\gamma \leq m = o(n)$ *for some* $1 - \frac{1}{k_0} < \gamma < 1$, *then, as* $n \to \infty$,

$$\widehat{\alpha}_x \xrightarrow{p} \alpha_x. \tag{3.4.22}$$

iii) In the case of ii), if $k_0 = 1$, $\{\xi_t\}_{t\in\mathbb{Z}}$ *satisfies Assumption A.6, and* m *is such that* $n^\gamma \leq m = o(n)$ *for some* $\gamma > 0$, *then*

$$\widehat{\alpha}_x - \alpha_x = O_p\left(m^{-\frac{1}{2}} + \left(\frac{m}{n}\right)^r\right), \tag{3.4.23}$$

where $r = \min\left\{\frac{\alpha_x}{2}, \frac{1-\alpha_x}{2}\right\}$. *Moreover, if* $\{\xi_t\}_{t\in\mathbb{Z}}$ *satisfies Assumption A.7, then*

$$m^{\frac{1}{2}}(\widehat{\alpha}_x - \alpha_x) \xrightarrow{d} N(0,1), \quad \text{as } n \to \infty, \tag{3.4.24}$$

with $m = o\left(n^{\frac{2r}{2r+1}}\right)$.

Theorem 3.3 establishes, under appropriate conditions, the consistency of the LW estimator $\widehat{\alpha}_x$. In part iii) of this theorem, we also derive the asymptotic distribution of $\widehat{\alpha}_x$ in the case of long memory and when the Hermite rank of $G(.)$ is 1. Hence, the asymptotic properties of the LW estimator are unaffected by the presence of the nonlinearity, at least in the cases that we examine here. However, as in the case of the signal plus noise model, there are three points that are worth raising.

Remark 3.6
Under the assumptions of part ii) of Theorem 3.3, we expect the finite sample bias of the estimator $\widehat{\alpha}_x$ *to be negative, see Remark 3.3.*

Remark 3.7
Under the assumption of part iii) of Theorem 3.3, the existence of nonlinearity worsens the rate of convergence of $\widehat{\alpha}_x$ *as compared to the linear case. Moreover, the rate of convergence is faster when* α_x *is closer to* $\frac{1}{2}$ *and slower when* α_x *is close to the boundary points 0 and 1.*

Remark 3.8
Part iii) of Theorem 3.3 implies that the bandwidth parameter m *needed to implement (3.4.24) depends on the unknown memory parameter* α_x, *so that the issue*

of bandwidth choice arises. The optimal bandwidth choice procedure suggested by Henry and Robinson (1996) cannot be implemented, and even if it could, it would suggest a bandwidth parameter smaller than in the linear case.

3.5 Long memory stochastic volatility model

In this section, we consider the LMSV model of Breidt, Crato, and de Lima (1998) and Harvey (1998) under which the return series $\{r_t\}_{t\in\mathbb{Z}}$ is such that

$$r_t = \varepsilon_t \sigma_t, \tag{3.5.1}$$

where $\{\sigma_t\}_{t\in\mathbb{Z}}$ is the volatility process and $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a sequence of i.i.d. variables with mean zero and finite variance. The volatility process is furthermore restricted to be of the form

$$\sigma_t = \exp\left(\xi_t\right), \tag{3.5.2}$$

where $\{\xi_t\}_{t\in\mathbb{Z}}$ is a stationary long memory Gaussian process independent of $\{\varepsilon_t\}_{t\in\mathbb{Z}}$.

As was already mentioned above, the logarithmic squared returns

$$\log r_t^2 = 2\xi_t + 2\log|\varepsilon_t| \tag{3.5.3}$$

is a special case of the signal plus noise model (3.1.1) with $x_t = \log r_t^2$, $y_t = 2\xi_t$ and $z_t = 2\log|\varepsilon_t|$, and under this specification the properties of the LW estimator were analyzed by Arteche (2004) and Hurvich, Moulines, and Soulier (2005). However, there are evidence in the empirical literature that powers of the absolute return series, and not just their logarithmic squared transformation, exhibit long memory. Also, it is evident that the signal plus noise decomposition of the logarithmic squared returns (3.5.3) relies heavily on the form of the volatility process (3.5.2).

Here, we analyze the properties of the LW estimator for the model (3.5.1), allowing for a more general specification of the volatility process and considering the p-th power of the absolute return series. In particular, we assume that

$$\sigma_t = G\left(\xi_t\right), \tag{3.5.4}$$

where $\{\xi_t\}_{t\in\mathbb{Z}}$ is a stationary long memory Gaussian process with zero mean, unit variance and independent of $\{\varepsilon_t\}_{t\in\mathbb{Z}}$, while $G(.)$ is a strictly positive function. We study

$$|r_t|^p = |\varepsilon_t|^p G^p\left(\xi_t\right), \tag{3.5.5}$$

where $p > 0$ is such that $\{G^p(\xi_t)\}_{t\in\mathbb{Z}}$ is long memory process with finite second moment. Observe that we can write

$$|r_t|^p = E(|\varepsilon_t|^p)\, G^p(\xi_t) + (|\varepsilon_t|^p - E(|\varepsilon_t|^p))\, G^p(\xi_t). \tag{3.5.6}$$

Then, (3.5.6) is a signal plus noise model $x_t = y_t + z_t$ with $x_t = |r_t|^p$, $y_t = E(|\varepsilon_t|^p)\, G^p(\xi_t)$ and $z_t = (|\varepsilon_t|^p - E(|\varepsilon_t|^p))\, G^p(\xi_t)$. Observe that the signal process $\{y_t\}_{t\in\mathbb{Z}}$ is a nonlinear transformation of a Gaussian sequence, and denote by $k_{0,p}$ the Hermite rank of $G^p(.)$. Furthermore, under the assumption that $\{\xi_t\}_{t\in\mathbb{Z}}$ and $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ are independent, we have that the noise process $\{z_t\}_{t\in\mathbb{Z}}$ is a sequence of uncorrelated random variables, while the signal process $\{y_t\}_{t\in\mathbb{Z}}$ is uncorrelated from the noise $\{z_t\}_{t\in\mathbb{Z}}$. We can therefore combine the results of Theorem 3.1 and 3.3 to establish the asymptotic properties of the LW estimator for the model (3.5.6).

Theorem 3.4

Suppose that $\{|r_t|^p\}_{t\in\mathbb{Z}}$ is as defined in (3.5.5).

i) If $\{G^p(\xi_t)\}_{t\in\mathbb{Z}}$ satisfies Assumptions L.1 and L.2, and m is such that $n^\gamma \leq m = o(n)$ for some $1 - \frac{1}{k_{0,p}} < \gamma < 1$, then, as $n \to \infty$,

$$\widehat{\alpha}_{|r|^p} \xrightarrow{p} \alpha_{|r|^p} = 1 - k_{0,p}(1 - \alpha_\xi). \tag{3.5.7}$$

ii) If in case of i), $k_{0,p} = 1$, $\{\xi_t\}_{t\in\mathbb{Z}}$ satisfies Assumption A.6, and m is such that $n^\gamma \leq m = o(n)$ for some $\gamma > 0$, then

$$\widehat{\alpha}_{|r|^p} - \alpha_{|r|^p} = O_p\left(m^{-\frac{1}{2}} + \left(\frac{m}{n}\right)^r\right), \tag{3.5.8}$$

where $r = \min\left\{\frac{\alpha_x}{2}, \frac{1-\alpha_x}{2}\right\}$. Moreover, if $\{\xi_t\}_{t\in\mathbb{Z}}$ satisfies Assumption A.7, then

$$m^{\frac{1}{2}}(\widehat{\alpha}_{|r|^p} - \alpha_{|r|^p}) \xrightarrow{d} N(0,1), \quad \text{as } n \to \infty, \tag{3.5.9}$$

with $m = o\left(n^{\frac{2r}{2r+1}}\right)$.

Theorem 3.4 part i) establishes, under appropriate conditions, the consistency of the LW estimator $\widehat{\alpha}_{|r|^p}$. In part ii) of this theorem, we derive under further conditions the rate of convergence and asymptotic distribution of $\widehat{\alpha}_{|r|^p}$ when the Hermite rank of $G^p(.)$ is 1. Remarks similar to those following Theorem 3.3 can also be made here. There is an additional comment that needs to be pointed out.

Remark 3.9

Theorem 3.4 part i) suggests that different p-th power transformations may lead to different long memory parameters. Furthermore, in the case that $k_{0,p_1} = k_{0,p_2}$ for all

$p_1, p_2 > 0$, *we expect the signal-to-noise ratio to depend on* p, *so that the finite sample bias of the LW estimator may vary across different* p*-th power transformations, see Remark 3.3.*

Notice that in the case of (3.5.2), we have $k_{0,p} = 1$ for all $p > 0$. Hence, $\alpha_{|r|^p} = \alpha_\xi$ for all $p > 0$, in line with the results of Harvey (1998), who established that for all $p > 0$, the autocovariance functions $\gamma_{|r|^p}(\tau)$ and $\gamma_\xi(\tau)$ of $\{|r_t|^p\}_{t\in\mathbb{Z}}$ and $\{\xi_t\}_{t\in\mathbb{Z}}$, respectively, have the same rate of decay to 0, as $\tau \to \infty$.

Different p-th power transformations of asset returns have been found to be display different levels of persistence, see Ding, Granger, and Engle (1993) and Ding and Granger (1996). The results presented here raise the question as to whether this fact is actually driven by either of the two following sources. Firstly, the volatility process is not given by (3.5.2) but rather by (3.5.4) with a function $G(.)$ such that different p-th transformations give rise to different Hermite ranks $k_{0,p}$ and therefore to different memory parameters. Secondly, the volatility process is of the form (3.5.2) and the different p-th power transformations produce finite sample bias of different magnitude.

3.6 Final comments

In this chapter, we have applied our general results of Chapter 2 in order to assert the asymptotic properties of the LW estimator for several nonlinear models. We have examined the signal plus noise model and some special cases of it: structural model, nonlinear transformations of a Gaussian process, and LMSV model. Under these specifications we have discovered that the asymptotic properties, consistency and asymptotic normality, are unaffected by the presence of the nonlinearity. We have also found that the rate of convergence and finite sample bias of the estimator are worse off when compared to the case of a linear process. Hence, in order to achieve the same level of accuracy as in the linear case, a large sample size is required. In addition, the issue of bandwidth choice has arisen. It should be also added, that in the case of nonlinear transformations of a Gaussian process, we have established the asymptotic distribution of the LW estimator only when the Hermite rank $k_0 = 1$. It remains an open question whether the asymptotic distribution of the LW estimator is Gaussian in the case of $k_0 \geq 2$. However, the work of Giraitis and Taqqu (1999) on the PW estimator leads us to consider the possibility of a non-Gaussian asymptotic distribution for the LW estimator in the latter case.

The extension to the case of a persistent cyclical component of known frequency $\omega_x \neq 0$ is straightforward, as was already mentioned in the final comments of Chapter 2. In addition, one can expect to retrieve the asymptotic properties of the LW estimator for most of the models presented here when $\alpha_x \geq 1$, as long as the general conditions of Chapter 2 are extended to this case. We should add that Hurvich, Moulines, and Soulier (2005), for their type of signal plus noise model discussed in Section 3.2 above, showed the consistency of the LW estimator when $\alpha_x < 2$ and established its asymptotic distribution when $\alpha_x < 1.5$.

The choice of the bandwidth parameter m is an important issue that we hope to address in the future. The rule of Henry and Robinson (1996) for choosing the optimal bandwidth relies on setting $\beta_x = 2$ and using an iterative procedure that involves estimation of a linear regression model. However, we have seen above that the presumption $\beta_x = 2$ cannot be made in general for nonlinear models, so that estimation of the parameter β_x is also required. The regression model used in the iterative procedure of Henry and Robinson (1996) now becomes nonlinear, and preliminary simulations show that the estimation of β_x is rather imprecise in the latter regression model. The work of Giraitis, Robinson, and Samarov (2000) is likely to provide an estimator of β_x, however further investigation on its finite sample performance is needed. If the latter estimation is precise, then the procedure of Henry and Robinson (1996) can be performed, replacing β_x by its estimate.

Another significant issue that we plan to examine is bias reduction. The results presented in this chapter suggest that the finite sample bias of the LW estimator can be severe for nonlinear models. Hurvich, Moulines, and Soulier (2005), in the context of the signal plus noise model examined there, proposed a modification of the LW estimator by including an additional term to account for the presence of the additive noise. Monte-Carlo experiments performed by the authors indeed showed a reduction in the finite sample bias at the expense of an increase in its dispersion. The methodology of Hurvich, Moulines, and Soulier (2005) is dependent on their signal plus noise model, but is likely to be extended to the nonlinear models examined here. However, its performance in small samples is rather unsatisfactory, as Monte-Carlo experiments performed by Gonçalves da Silva and Robinson (2006) suggest that the modified estimator displays bimodality. And since it is in small samples that a bias reduction method is essential, the usefulness of the method of Hurvich, Moulines, and Soulier (2005) is doubtful. Another possibility, would be to employ the method of Andrews and Sun (2004). The latter authors combined LW estimation with local polynomial approximations, and established the asymptotic properties of their modified estimator. However, linearity of the process and a certain degree of smoothness were required, so that the technique of Andrews and Sun (2004) is not likely to be extendable to the type of nonlinear models considered here. A possible solution to the problem of bias reduction would be first estimating the parameters $c_{0,x}$, $c_{1,x}$ and β_x in order to retrieve the bandwidth parameter m and then calculating the term $-\left(\frac{m}{n}\right)^{\beta_x}\frac{c_{1,x}}{c_{0,x}}B_{\beta_x}$, which has been found in Chapter 2 to be the main determinant of the finite sample bias of the estimator. Then, once the LW estimator is corrected for this term, the finite sample bias is likely to be reduced without affecting the variance and distribution of the LW estimator.

3.A Appendix

This section contains the proofs which use a series of lemmas found in Appendix 3.B below.

Proof of Theorem 3.1. i) We start by showing convergence (3.2.4). By Theorem 2.1, it suffices to show that Assumptions A.1 and A.2 are satisfied with $\alpha_x = \alpha_y$ and $c_{0,x} = c_{0,y}$ there. We have that

$$\begin{aligned} I_x(\lambda_j) &= \frac{1}{2\pi}\left|w_x(\lambda_j)\right|^2 = \frac{1}{2\pi}\left|w_y(\lambda_j)+w_z(\lambda_j)\right|^2 \\ &= \frac{1}{2\pi}\left|w_y(\lambda_j)\right|^2 + \frac{1}{2\pi}\left(\left|w_z(\lambda_j)\right|^2 + w_y(\lambda_j)\overline{w_z(\lambda_j)} + \overline{w_y(\lambda_j)}w_z(\lambda_j)\right) \\ &: = I_y(\lambda_j) + v_j, \end{aligned} \tag{3.A.1}$$

with $v_j = I_z(\lambda_j) + \frac{1}{2\pi}\left(w_y(\lambda_j)\overline{w_z(\lambda_j)} + \overline{w_y(\lambda_j)}w_z(\lambda_j)\right)$ for which,

$$|v_j| \leq I_z(\lambda_j) + \frac{1}{\pi}\left|w_y(\lambda_j)\right|\left|w_z(\lambda_j)\right|. \tag{3.A.2}$$

We first establish Assumption A.1. Taking expectations over (3.A.2), and using the Cauchy-Schwarz inequality, we have

$$\begin{aligned} E\left|v_j\right| &\leq C\left(E\left(I_z(\lambda_j)\right) + \left(E\left(I_y(\lambda_j)\right)E\left(I_z(\lambda_j)\right)\right)^{\frac{1}{2}}\right) \\ &\leq C\left(f_z(\lambda_j) + f_y^{\frac{1}{2}}(\lambda_j)f_z^{\frac{1}{2}}(\lambda_j)\right), \end{aligned} \tag{3.A.3}$$

where the last displayed inequality follows from Lemma 2.6, since under (3.2.2), $\{y_t\}_{t\in\mathbb{Z}}$ and $\{z_t\}_{t\in\mathbb{Z}}$ satisfy Assumption A.3. Now, using (3.2.2), we have that

$$\begin{aligned} E\left|v_j\right| &\leq C\left(\lambda_j^{-\alpha_z} + \lambda_j^{-\frac{\alpha_y}{2}}\lambda_j^{-\frac{\alpha_z}{2}}\right) \\ &\leq C\lambda_j^{-\frac{(\alpha_y+\alpha_z)}{2}}. \end{aligned} \tag{3.A.4}$$

Expressions (3.A.1) and (3.A.4) imply that

$$\begin{aligned} E(\eta_{j,x}^*) &= c_{0,y}^{-1}\lambda_j^{\alpha_y} E\left(I_x(\lambda_j)\right) \\ &\le c_{0,y}^{-1}\lambda_j^{\alpha_y} E\left(I_y(\lambda_j)\right) + c_{0,y}^{-1}\lambda_j^{\alpha_y} E\left|v_j\right| \\ &\le C + C\lambda_j^{\frac{\alpha_y-\alpha_z}{2}} \le C, \end{aligned} \tag{3.A.5}$$

since we have from Lemma 2.6 for $\{y_t\}_{t\in\mathbb{Z}}$ that $c_{0,y}^{-1}\lambda_j^{\alpha_y} E\left(I_y(\lambda_j)\right) \le C$, and $\alpha_z < \alpha_y$. Hence, $\{x_t\}_{t\in\mathbb{Z}}$ satisfies Assumption A.1 with parameters $\alpha_x = \alpha_y$ and $c_{0,x} = c_{0,y}$ there.

It remains to show that $\{x_t\}_{t\in\mathbb{Z}}$ satisfies Assumption A.2 with $c_{0,x} = c_{0,y}$ and $\alpha_x = \alpha_y$. Using (3.A.1), write

$$\frac{1}{m}\sum_{j=1}^{m}\eta_{j,x}^* = \frac{1}{m}\sum_{j=1}^{m}\eta_{j,y}^* + \frac{1}{m}\sum_{j=1}^{m} c_{0,y}^{-1}\lambda_j^{\alpha_y} v_j. \tag{3.A.6}$$

Since $\{y_t\}_{t\in\mathbb{Z}}$ satisfies Assumption A.2, we have that the first sum on the right hand side of (3.A.6) satisfies

$$\frac{1}{m}\sum_{j=1}^{m}\eta_{j,y}^* \xrightarrow{p} 1, \tag{3.A.7}$$

as $n \to \infty$. On the other hand, the bound given in (3.A.4) implies that

$$\begin{aligned} E\left|\frac{1}{m}\sum_{j=1}^{m} c_{0,y}^{-1}\lambda_j^{\alpha_y} v_j\right| &\le C\frac{1}{m}\sum_{j=1}^{m}\lambda_j^{\alpha_y} E\left|v_j\right| \\ &\le C\frac{1}{m}\sum_{j=1}^{m}\lambda_j^{\frac{\alpha_y-\alpha_z}{2}} \\ &\le C\left(\frac{m}{n}\right)^{\frac{\alpha_y-\alpha_z}{2}} = o(1), \end{aligned} \tag{3.A.8}$$

noticing that $\alpha_z < \alpha_y$. Thus, we obtain that the second sum on the right hand side of (3.A.6) satisfies

$$\frac{1}{m}\sum_{j=1}^{m} c_{0,y}^{-1}\lambda_j^{\alpha_y} v_j \xrightarrow{p} 0, \tag{3.A.9}$$

as $n \to \infty$. Using (3.A.7) and (3.A.9) in (3.A.6), we have that

$$\frac{1}{m}\sum_{j=1}^{m}\eta_{j,x}^* \xrightarrow{p} 1, \tag{3.A.10}$$

as $n \to \infty$. Hence, $\{x_t\}_{t\in\mathbb{Z}}$ satisfies Assumption A.2 with $\alpha_x = \alpha_y$ and $c_{0,x} = c_{0,y}$ there.

Next, we show (3.2.5). Since the assumptions of Theorem 2.1 are satisfied with $\alpha_x = \alpha_y$ and $c_{0,x} = c_{0,y}$, we have from expression (2.4.6) of Theorem 2.3 that

$$\widehat{\alpha}_x - \alpha_y = -Q_{m,x}(1 + o_p(1)) + O_p\left(\frac{\log m}{m}\right). \qquad (3.A.11)$$

From (3.A.1) and (3.A.4), it follows that

$$\begin{aligned} Q_{m,x} &= \frac{1}{m}\sum_{j=1}^{m}\left(\log\left(\frac{j}{m}\right)+1\right)\eta_{j,x}^{*} \\ &= \frac{1}{m}\sum_{j=1}^{m}\left(\log\left(\frac{j}{m}\right)+1\right)\eta_{j,y}^{*} + \frac{1}{m}\sum_{j=1}^{m}\left(\log\left(\frac{j}{m}\right)+1\right)c_{0,y}^{-1}\lambda_j^{\alpha_y}v_j \\ &= Q_{m,y} + O_p(1)\frac{1}{m}\sum_{j=1}^{m}\left(\log\left(\frac{j}{m}\right)+1\right)\lambda_j^{\frac{\alpha_y-\alpha_z}{2}} \\ &= Q_{m,y} + O_p\left(\left(\frac{m}{n}\right)^{\frac{\alpha_y-\alpha_z}{2}}\right)\frac{1}{m}\sum_{j=1}^{m}\left(\log\left(\frac{j}{m}\right)+1\right)\left(\frac{j}{m}\right)^{\frac{\alpha_y-\alpha_z}{2}} .(3.A.12) \end{aligned}$$

From Lemma 2.5, it follows that $\frac{1}{m}\sum\limits_{j=1}^{m}\left(\log\left(\frac{j}{m}\right)+1\right)\left(\frac{j}{m}\right)^{\frac{\alpha_y-\alpha_z}{2}} = \int\limits_0^1(\log x+1)x^{\frac{\alpha_y-\alpha_z}{2}}dx + O\left(\frac{1}{m}\right)$. Since $\alpha_z < \alpha_y$, we have that $\left|\int\limits_0^1(\log x+1)x^{\frac{\alpha_y-\alpha_z}{2}}dx\right| < \infty$, and then

$$Q_{m,x} = Q_{m,y} + O_p\left(\left(\frac{m}{n}\right)^{\frac{\alpha_y-\alpha_z}{2}}\right). \qquad (3.A.13)$$

Since $\{y_t\}_{t\in\mathbb{Z}}$ satisfies Assumptions A.2 and A.3, we have that expression (2.4.6) of Theorem 2.3 applies for $\{y_t\}_{t\in\mathbb{Z}}$, that is

$$\widehat{\alpha}_y - \alpha_y = -Q_{m,y}(1 + o_p(1)) + O_p\left(\frac{\log m}{m}\right), \qquad (3.A.14)$$

which together with (3.A.13) imply that

$$\begin{aligned} Q_{m,x} &= \frac{1}{1+o_p(1)}\left(-(\widehat{\alpha}_y-\alpha_y)+O_p\left(\frac{\log m}{m}\right)\right)+O_p\left(\left(\frac{m}{n}\right)^{\frac{\alpha_y-\alpha_z}{2}}\right) \\ &= -(\widehat{\alpha}_y-\alpha_y)(1+o_p(1))+O_p\left(\left(\frac{m}{n}\right)^{\frac{\alpha_y-\alpha_z}{2}}+\frac{\log m}{m}\right). \qquad (3.A.15) \end{aligned}$$

Thus, from (3.A.11), we have that

$$\begin{aligned} \widehat{\alpha}_x - \alpha_y &= \left((\widehat{\alpha}_y-\alpha_y)(1+o_p(1))+O_p\left(\left(\frac{m}{n}\right)^{\frac{\alpha_y-\alpha_z}{2}}+\frac{\log m}{m}\right)\right)(1+o_p(1)) \\ &= (\widehat{\alpha}_y-\alpha_y)(1+o_p(1))+O_p\left(\left(\frac{m}{n}\right)^{\frac{\alpha_y-\alpha_z}{2}}+\frac{\log m}{m}\right), \qquad (3.A.16) \end{aligned}$$

as required.

ii) Since $\{y_t\}_{t\in\mathbb{Z}}$ satisfies Assumption A.4 and $\Delta_{m,y} \leq Cm^{\gamma}$, for some $0<\gamma<1$, we have that $\{y_t\}_{t\in\mathbb{Z}}$ satisfies the assumptions of Proposition 2.1. Hence, from part i) of this theorem, it follows that $\{x_t\}_{t\in\mathbb{Z}}$ satisfies Assumptions A.1 and A.2, and that expression (3.2.5) is valid. Notice also that by the assumptions of the theorem, $\{y_t\}_{t\in\mathbb{Z}}$ satisfies Assumption A.4 and $\Delta_{m,y} = o\left(\frac{m}{\log^2 m}\right)$. Therefore, the conditions of Proposition 2.2 hold for $\{y_t\}_{t\in\mathbb{Z}}$ and, (2.4.10) implies that

$$\begin{aligned}\widehat{\alpha}_y - \alpha_y &= O_p\left(\Delta_{m,y}\frac{\log m}{m} + m^{-\frac{1}{2}} + \left(\frac{m}{n}\right)^{\beta_y}\right) \\ &= O_p\left(m^{\gamma-1}\log m + m^{-\frac{1}{2}} + \left(\frac{m}{n}\right)^{\beta_y}\right), \qquad (3.A.17)\end{aligned}$$

for some $0<\gamma<1$. The latter displayed bound, together with (3.2.5), imply that

$$\begin{aligned}\widehat{\alpha}_x - \alpha_y &= O_p\left(m^{\gamma-1}\log m + m^{-\frac{1}{2}} + \left(\frac{m}{n}\right)^{\beta_y}\right)(1+o_p(1)) \\ &\quad + O_p\left(\left(\frac{m}{n}\right)^{\frac{\alpha_y-\alpha_z}{2}} + \frac{\log m}{m}\right) \\ &= O_p\left(m^{\gamma-1}\log m + m^{-\frac{1}{2}} + \left(\frac{m}{n}\right)^{\beta_y} + \left(\frac{m}{n}\right)^{\frac{\alpha_y-\alpha_z}{2}}\right). \qquad (3.A.18)\end{aligned}$$

iii) Since $\{y_t\}_{t\in\mathbb{Z}}$ is a linear process satisfying Assumptions A.4, A.6 and A.8, it follows from (2.4.7) of Theorem 2.3 that

$$\widehat{\alpha}_y - \alpha_y = O_p\left(\left(\frac{m}{n}\right)^{\beta_y}\right) - \left(Q_{m,y} - E\left(Q_{m,y}\right)\right)(1+o_p(1)) + o_p\left(m^{-\frac{1}{2}}\right), \qquad (3.A.19)$$

whereas (2.5.8) of Proposition 2.3 implies that $\Delta_{m,y} = O\left(m^{\frac{1}{2}}\log^{\frac{1}{2}} m + m\left(\frac{m}{n}\right)^{\beta_y}\right) = o\left(\frac{m}{\log^2 m}\right)$, since $m = o\left(\frac{n}{\log^{\frac{2}{\beta_y}} n}\right)$. So, from (2.4.11) of Proposition 2.2 we deduce that

$$Q_{m,y} - E\left(Q_{m,y}\right) = V_{m,y} + o_p\left(\left(\frac{m}{n}\right)^{\beta_y}\right), \qquad (3.A.20)$$

which together with (3.A.19), implies that

$$\widehat{\alpha}_y - \alpha_y = -V_{m,y}(1+o_p(1)) + O_p\left(\left(\frac{m}{n}\right)^{\beta_y}\right) + o_p\left(m^{-\frac{1}{2}}\right). \qquad (3.A.21)$$

Then, from (3.2.5) we have that

$$
\begin{aligned}
\widehat{\alpha}_x - \alpha_y &= -V_{m,y}(1+o_p(1)) + \left(O_p\left(\left(\frac{m}{n}\right)^{\beta_y}\right) + o_p\left(m^{-\frac{1}{2}}\right)\right)(1+o_p(1)) \\
&\quad + O_p\left(\left(\frac{m}{n}\right)^{\frac{\alpha_y-\alpha_x}{2}} + \frac{\log m}{m}\right) \\
&= -V_{m,y}(1+o_p(1)) + O_p\left(\left(\frac{m}{n}\right)^{r}\right) + o_p\left(m^{-\frac{1}{2}}\right). \qquad (3.A.22)
\end{aligned}
$$

Since $\{y_t\}_{t\in\mathbb{Z}}$ is a linear process satisfying Assumptions A.4, A.7 and A.8, we have from Proposition 2.3 part iii) that

$$
m^{\frac{1}{2}} V_{m,y} \xrightarrow{d} N(0,1), \quad \text{as } n \to \infty. \qquad (3.A.23)
$$

Hence, $V_{m,y} = O_p\left(m^{-\frac{1}{2}}\right)$ and (3.A.22) implies that

$$
\widehat{\alpha}_y - \alpha_y = O_p\left(m^{-\frac{1}{2}} + \left(\frac{m}{n}\right)^{r}\right), \qquad (3.A.24)
$$

while for $m = o\left(n^{\frac{2r}{2r+1}}\right)$, it follows from (3.A.22) and (3.A.23) that

$$
m^{\frac{1}{2}}\left(\widehat{\alpha}_x - \alpha_y\right) \xrightarrow{d} N(0,1), \quad \text{as } n \to \infty, \qquad (3.A.25)
$$

which completes the proof of this theorem. ■

Proof of Proposition 3.1. From (3.4.9), we have that

$$
\begin{aligned}
\gamma_x(\tau) &= \frac{c_{k_0}^2}{k_0!}\gamma_\xi^{k_0}(\tau) + \sum_{k=k_0+1}^{\infty} \frac{c_k^2}{k!}\gamma_\xi^{k}(\tau) \\
&= \gamma_\xi^{k_0}(\tau)\left(\frac{c_{k_0}^2}{k_0!} + \sum_{k=k_0+1}^{\infty} \frac{c_k^2}{k!}\gamma_\xi^{k-k_0}(\tau)\right) \\
&= \gamma_\xi^{k_0}(\tau)\left(\frac{c_{k_0}^2}{k_0!} + \sum_{k=1}^{\infty} \frac{c_{k+k_0}^2}{(k+k_0)!}\gamma_\xi^{k}(\tau)\right) \\
&= \gamma_\xi^{k_0}(\tau)\left(\frac{c_{k_0}^2}{k_0!} + R_\tau\right), \qquad (3.A.26)
\end{aligned}
$$

which shows (3.4.10) with

$$
R_\tau = \sum_{k=1}^{\infty} \frac{c_{k+k_0}^2}{(k+k_0)!}\gamma_\xi^{k}(\tau). \qquad (3.A.27)
$$

Recall that the process $\{\xi_t\}_{t\in\mathbb{Z}}$ has unit variance, so that $\left|\gamma_\xi(\tau)\right| \leq 1$ for every $\tau \in \mathbb{Z}$. Hence,

$$
|R_\tau| = \left|\sum_{k=1}^{\infty} \frac{c_{k+k_0}^2}{(k+k_0)!}\gamma_\xi^{k}(\tau)\right| \leq \left|\gamma_\xi(\tau)\right| \left|\sum_{k=1}^{\infty} \frac{c_{k+k_0}^2}{(k+k_0)!}\right| \leq C\left|\gamma_\xi(\tau)\right|, \qquad (3.A.28)
$$

because by (3.4.9), $\sum\limits_{k=1}^{\infty} \frac{c_{k+k_0}^2}{(k+k_0)!} \leq \sum\limits_{k=k_0}^{\infty} \frac{c_k^2}{k!} = \gamma_x(0) < \infty$. Since $\{\xi_t\}_{t\in\mathbb{Z}}$ is a sequence of stationary random variables, $\gamma_\xi(\tau) \to 0$ as $\tau \to \infty$, which together with (3.A.28), implies (3.4.11). ■

Proof of Proposition 3.2. By (3.4.10) and (3.A.28), we have that for every $\tau \in \mathbb{Z}$,

$$\gamma_x(\tau) = \gamma_\xi^{k_0}(\tau) \left(\frac{c_{k_0}^2}{k_0!} + R_\tau \right), \tag{3.A.29}$$

where

$$|R_\tau| = \left| \sum_{k=1}^{\infty} \frac{c_{k+k_0}^2}{(k+k_0)!} \gamma_\xi^k(\tau) \right| \leq \sum_{k=1}^{\infty} \frac{c_{k+k_0}^2}{(k+k_0)!}, \tag{3.A.30}$$

since $|\gamma_\xi(\tau)| \leq 1$ for every $\tau \in \mathbb{Z}$. Thus,

$$\left| \frac{c_{k_0}^2}{k_0!} + R_\tau \right| \leq \sum_{k=0}^{\infty} \frac{c_{k+k_0}^2}{(k+k_0)!} = \gamma_x(0). \tag{3.A.31}$$

Then, from (3.A.29), we have that for every $\tau \in \mathbb{Z}$, $|\gamma_x(\tau)| \leq |\gamma_\xi(\tau)|^{k_0} \gamma_x(0)$, which implies that

$$\frac{1}{\gamma_x(0)} \sum_{\tau\in\mathbb{Z}} |\gamma_x(\tau)| \leq \sum_{\tau\in\mathbb{Z}} |\gamma_\xi(\tau)|^{k_0}. \tag{3.A.32}$$

Moreover, since by (3.4.11), $R_\tau \to 0$, as $\tau \to \infty$, there exists $\tau_0 \in \mathbb{N}$ such that, for every $|\tau| \geq \tau_0$, $|R_\tau| \leq \frac{c_{k_0}^2}{2k_0!}$. Thus, for every $|\tau| \geq \tau_0$,

$$\left| \frac{c_{k_0}^2}{k_0!} + R_\tau \right| \geq \left| \frac{c_{k_0}^2}{k_0!} - |R_\tau| \right| = \frac{c_{k_0}^2}{2k_0!}, \tag{3.A.33}$$

which together with (3.A.29) implies that $|\gamma_\xi(\tau)|^{k_0} \leq \frac{2k_0!}{c_{k_0}^2} |\gamma_x(\tau)|$ for every $|\tau| \geq \tau_0$. Hence,

$$\sum_{|\tau|\geq\tau_0} |\gamma_\xi(\tau)|^{k_0} \leq \frac{2k_0!}{c_{k_0}^2} \sum_{|\tau|\geq\tau_0} |\gamma_x(\tau)|, \tag{3.A.34}$$

while for $|\tau| < \tau_0$, we can write

$$\sum_{|\tau|<\tau_0} |\gamma_\xi(\tau)|^{k_0} \leq C \sum_{|\tau|<\tau_0} |\gamma_x(\tau)|, \tag{3.A.35}$$

and conclude that

$$\sum_{\tau\in\mathbb{Z}} |\gamma_\xi(\tau)|^{k_0} \leq C \sum_{\tau\in\mathbb{Z}} |\gamma_x(\tau)|, \tag{3.A.36}$$

which completes the proof of (3.4.12). ■

Proof of Proposition 3.3. We have that $\gamma_\xi(\tau) = \int\limits_{-\pi}^{\pi} e^{i\tau\lambda} f_\xi(\lambda)d\lambda$. Thus,

$$\gamma_\xi^k(\tau) = \int\limits_{-\pi}^{\pi} \dots \int\limits_{-\pi}^{\pi} e^{i\tau(l_1+\dots+l_k)} f_\xi(l_1)\dots f_\xi(l_k) dl_1 \dots dl_k. \tag{3.A.37}$$

Setting $\lambda = l_1 + \dots + l_k$, we get that

$$\begin{aligned}
\gamma_\xi^k(\tau) &= \int\limits_{-k\pi}^{k\pi} \int\limits_{-\pi}^{\pi} \dots \int\limits_{-\pi}^{\pi} e^{i\tau\lambda} \mathcal{I}(-\pi < \lambda - l_1 - \dots - l_{k-1} \le \pi) \\
&\quad \times f_\xi(l_1)\dots f_\xi(l_{k-1}) f_\xi(\lambda - l_1 - \dots - l_{k-1}) dl_1 \dots dl_{k-1} d\lambda \\
&= \int\limits_{-\pi}^{\pi} \dots \int\limits_{-\pi}^{\pi} e^{i\tau\lambda} f_\xi(\lambda - l_1 - \dots - l_{k-1}) \\
&\quad \times f_\xi(l_1)\dots f_\xi(l_{k-1}) dl_1 \dots dl_{k-1} d\lambda & (3.A.38) \\
&= \int\limits_{-\pi}^{\pi} e^{i\tau\lambda} f_\xi^{(*k)}(\lambda) d\lambda, & (3.A.39)
\end{aligned}$$

where in (3.A.38), we set $f_\xi(\lambda) = f_\xi(\lambda + 2\pi)$ for $\lambda \in \mathbb{R}$. Since $\gamma_x(\tau) = \sum\limits_{k=k_0}^{\infty} \frac{c_k^2}{k!} \gamma_\xi^k(\tau)$ by (3.4.9), using (3.A.39) we can write

$$\begin{aligned}
\gamma_x(\tau) &= \sum_{k=k_0}^{\infty} \frac{c_k^2}{k!} \int\limits_{-\pi}^{\pi} e^{i\tau\lambda} f_\xi^{(*k)}(\lambda) d\lambda \\
&= \sum_{k=k_0}^{\infty} \int\limits_{-\pi}^{\pi} \frac{c_k^2}{k!} e^{i\tau\lambda} f_\xi^{(*k)}(\lambda) d\lambda \\
&= \sum_{k=k_0}^{M-1} \int\limits_{-\pi}^{\pi} \frac{c_k^2}{k!} e^{i\tau\lambda} f_\xi^{(*k)}(\lambda) d\lambda + \sum_{k=M}^{\infty} \int\limits_{-\pi}^{\pi} \frac{c_k^2}{k!} e^{i\tau\lambda} f_\xi^{(*k)}(\lambda) d\lambda, & (3.A.40)
\end{aligned}$$

for every $M \in \mathbb{N}$ with $M \ge k_0$.

We first deal with the second sum on the right hand side of (3.A.40). Notice that (3.A.39) implies that

$$f_\xi^{(*k)}(\lambda) = \frac{1}{2\pi} \sum_{\tau \in \mathbb{Z}} e^{-i\tau\lambda} \gamma_\xi^k(\tau). \tag{3.A.41}$$

Under Assumption S.1, we have from Proposition 3.2 that $\sum\limits_{\tau \in \mathbb{Z}} |\gamma_\xi(\tau)|^{k_0} < \infty$ and hence,

$$f_\xi^{(*k)}(\lambda) \le C \sum_{\tau \in \mathbb{Z}} |\gamma_\xi(\tau)|^k \le C \sum_{\tau \in \mathbb{Z}} |\gamma_\xi(\tau)|^{k_0} \le C, \tag{3.A.42}$$

since $|\gamma_\xi(\tau)| \leq 1$. On the other hand, under Assumption L.2, we have that $\gamma_\xi(\tau) \sim c_\xi \tau^{-1+\alpha_\xi}$, as $\tau \to \infty$. Hence, $|\gamma_\xi(\tau)| \leq C|\tau|^{-1+\alpha_\xi}$ for every $\tau \neq 0$. Moreover, we can choose $M \in \mathbb{N}$ in (3.A.40) big enough, such that $M(1-\alpha_\xi) > 1$. Then, for every $k \geq M$,

$$\begin{aligned} f_\xi^{(*k)}(\lambda) &\leq C \sum_{\tau \in \mathbb{Z}} |\gamma_\xi(\tau)|^k \\ &= C\left(\sum_{\tau \in \mathbb{Z}:\tau \neq 0} |\gamma_\xi(\tau)|^k + |\gamma_\xi(0)|^k \right) \\ &\leq C\left(\sum_{\tau \in \mathbb{Z}:\tau \neq 0} |\gamma_\xi(\tau)|^M + 1 \right) \\ &\leq C\left(C \sum_{\tau \in \mathbb{Z}:\tau \neq 0} |\tau|^{-M(1-\alpha_\xi)} + 1 \right) \leq C, \end{aligned} \tag{3.A.43}$$

since $|\gamma_\xi(\tau)| \leq 1$ for every $\tau \in \mathbb{Z}$ and, $\sum_{\tau \in \mathbb{Z}:|\tau| \neq 0} |\tau|^{-M(1-\alpha_\xi)} < \infty$ for $M(1-\alpha_\xi) > 1$. Hence, under Assumption S.1 or L.2, we have that, for every $\tau \in \mathbb{Z}$,

$$\left| \frac{c_k^2}{k!} e^{i\tau\lambda} f_\xi^{(*k)}(\lambda) \right| \leq C \tag{3.A.44}$$

uniformly in λ and $k \geq M$, and

$$\left| \sum_{k=M}^{\infty} \frac{c_k^2}{k!} e^{i\tau\lambda} f_\xi^{(*k)}(\lambda) \right| \leq C \sum_{k=k_0}^{\infty} \frac{c_k^2}{k!} = C\gamma_x(0) = C, \tag{3.A.45}$$

for every λ. The last two displayed bounds imply that, in the second part of the right hand side (3.A.40), we can interchange the integral with the infinite sum, using the Weierstrass M-test, see Apostol (1967) p. 427. Hence,

$$\sum_{k=M}^{\infty} \int_{-\pi}^{\pi} \frac{c_k^2}{k!} e^{i\tau\lambda} f_\xi^{(*k)}(\lambda) d\lambda = \int_{-\pi}^{\pi} \sum_{k=M}^{\infty} \frac{c_k^2}{k!} e^{i\tau\lambda} f_\xi^{(*k)}(\lambda) d\lambda. \tag{3.A.46}$$

On the other hand, the first sum on the right hand side of (3.A.40) is finite and so is the integral. Thus,

$$\sum_{k=k_0}^{M-1} \int_{-\pi}^{\pi} \frac{c_k^2}{k!} e^{i\tau\lambda} f_\xi^{(*k)}(\lambda) d\lambda = \int_{-\pi}^{\pi} \sum_{k=k_0}^{M-1} \frac{c_k^2}{k!} e^{i\tau\lambda} f_\xi^{(*k)}(\lambda) d\lambda. \tag{3.A.47}$$

Using the last two displayed expressions in (3.A.40), we have that

$$\gamma_x(\tau) = \int_{-\pi}^{\pi} e^{i\tau\lambda} \left(\sum_{k=k_0}^{\infty} \frac{c_k^2}{k!} f_\xi^{(*k)}(\lambda) \right) d\lambda. \tag{3.A.48}$$

Recall that by definition $\gamma_x(\tau) = \int_{-\pi}^{\pi} e^{i\tau\lambda} f_x(\lambda) d\lambda$ and $\gamma_x(0) = \int_{-\pi}^{\pi} f_x(\lambda) d\lambda$, so that

$$f_x(\lambda) = \sum_{k=k_0}^{\infty} \frac{c_k^2}{k!} f_\xi^{(*k)}(\lambda), \tag{3.A.49}$$

as required. ■

Proof of Proposition 3.4. By Proposition 3.3, the spectral density function of $\{x_t\}_{t\in\mathbb{Z}}$ can be written as

$$f_x(\lambda) = \sum_{k=k_0}^{\infty} \frac{c_k^2}{k!} f_\xi^{(*k)}(\lambda). \tag{3.A.50}$$

i) We choose the smallest $M \in \mathbb{N}$, $M > k_0$, such that $M(1-\alpha_\xi) > 1$ and, write the sum in (3.4.8) as

$$x_t = \sum_{k=k_0}^{M-1} \frac{c_k}{k!} H_k(\xi_t) + \sum_{k=M}^{\infty} \frac{c_k}{k!} H_k(\xi_t) := y_t + z_t. \tag{3.A.51}$$

Notice that both $\{y_t\}_{t\in\mathbb{Z}}$ and $\{z_t\}_{t\in\mathbb{Z}}$ admit a Hermite expansion. In the case of $\{y_t\}_{t\in\mathbb{Z}}$, the Hermite rank is k_0 and the Hermite coefficients satisfy $c_k = 0$ for every $k \geq M$, while $\{z_t\}_{t\in\mathbb{Z}}$ has Hermite rank greater or equal to M. Also, it is evident from (3.4.7) that $\{y_t\}_{t\in\mathbb{Z}}$ and $\{z_t\}_{t\in\mathbb{Z}}$ are uncorrelated. This implies that the spectral density function of $\{x_t\}_{t\in\mathbb{Z}}$ can be written as

$$f_x(\lambda) = f_y(\lambda) + f_z(\lambda). \tag{3.A.52}$$

Consider first the spectral density function of $\{z_t\}_{t\in\mathbb{Z}}$. Notice that for $k \geq M$, we have that $k(1-\alpha_\xi) > 1$. By Proposition 3.3, we have that

$$f_z(\lambda) = \sum_{k=M}^{\infty} \frac{c_k^2}{k!} f_\xi^{(*k)}(\lambda) \leq \sum_{k=M}^{\infty} \frac{c_k^2}{k!} C \leq C, \tag{3.A.53}$$

since by Lemma 3.1 part iii), $f_\xi^{(*k)}(\lambda) \leq C$ for all $\lambda \in (-\pi, \pi]$ and, $\sum_{k=M}^{\infty} \frac{c_k^2}{k!} \leq \sum_{k=k_0}^{\infty} \frac{c_k^2}{k!} = \gamma_x(0) < \infty$ by (3.4.9).

Consider next the spectral density function of $\{y_t\}_{t\in\mathbb{Z}}$. Suppose first that $k_0 \geq 2$. Recall that $k_0(1-\alpha_\xi) < 1$ and, notice that for $k \leq M-1$, we have that $k(1-\alpha_\xi) \leq 1$. Hence, from Lemma 3.1 part i), it follows that, as $\lambda \to 0+$,

$$f_\xi^{(*k_0)}(\lambda) = c_{0,\xi}^{k_0} C_{k_0} \lambda^{-1+k_0(1-\alpha_\xi)} + o\left(\lambda^{-1+k_0(1-\alpha_\xi)}\right), \tag{3.A.54}$$

for some constant $0 < C_{k_0} < \infty$ and

$$f_\xi^{(*k)}(\lambda) = O(\lambda^{-1+k(1-\alpha_\xi)}) = o(\lambda^{-1+k_0(1-\alpha_\xi)}), \tag{3.A.55}$$

for $k \leq M-1$ such that $k(1-\alpha_\xi) < 1$. If it happens that $(M-1)(1-\alpha_\xi) = 1$, then from Lemma 3.1 part ii) we have that, as $\lambda \to 0+$,

$$f_\xi^{(*(M-1))}(\lambda) = O(\lambda^{-\delta}) = o\left(\lambda^{-1+k_0(1-\alpha_\xi)}\right), \tag{3.A.56}$$

for $0 < \delta < 1 - k_0(1-\alpha_\xi)$. Therefore, as $\lambda \to 0+$,

$$\begin{aligned} f_y(\lambda) &= \frac{c_{k_0}^2}{k_0!} f_\xi^{(*k_0)}(\lambda) + \sum_{k=k_0+1}^{M-1} \frac{c_k^2}{k!} f_\xi^{(*k)}(\lambda) \\ &= \frac{c_{k_0}^2}{k_0!} c_{0,\xi}^{k_0} C_{k_0} \lambda^{-1+k_0(1-\alpha_\xi)}(1+o(1)) + \sum_{k=k_0+1}^{M-1} \frac{c_k^2}{k!} o(\lambda^{-1+k_0(1-\alpha_\xi)}) \\ &= c_{k_0}^2 s_{k_0} \lambda^{-1+k_0(1-\alpha_\xi)} + o(\lambda^{-1+k_0(1-\alpha_\xi)}), \end{aligned} \tag{3.A.57}$$

with $s_{k_0} = \frac{c_{0,\xi}^{k_0} C_{k_0}}{k_0!}$. In the special case that $k_0 = 1$, Assumption L.2 implies that, as $\lambda \to 0+$,

$$f_\xi^{(*1)}(\lambda) = f_\xi(\lambda) = c_{0,\xi}\lambda^{-\alpha_\xi} + o\left(\lambda^{-\alpha_\xi}\right). \tag{3.A.58}$$

Using the same arguments as in (3.A.54)-(3.A.56), we have that, as $\lambda \to 0+$,

$$f_\xi^{(*k)}(\lambda) = O\left(\lambda^{-1+2(1-\alpha_\xi)}\right) = o\left(\lambda^{-\alpha_\xi}\right), \tag{3.A.59}$$

for $2 \leq k \leq M-1$. The last two displayed expressions imply that, as $\lambda \to 0+$,

$$\begin{aligned} f_y(\lambda) &= c_1^2 f_\xi^{(*1)}(\lambda) + \sum_{k=2}^{M-1} \frac{c_k^2}{k!} f_\xi^{(*k)}(\lambda) \\ &= c_1^2 c_{0,\xi} \lambda^{-\alpha_\xi}(1+o(1)) + \sum_{k=k_0+1}^{M-1} \frac{c_k^2}{k!} o(\lambda^{-\alpha_\xi}) \\ &= c_1^2 s_1 \lambda^{-\alpha_\xi} + o\left(\lambda^{-\alpha_\xi}\right), \end{aligned} \tag{3.A.60}$$

with $s_1 = c_{0,\xi}$.

Now, set $\alpha_x = 1 - k_0(1-\alpha_\xi)$. Using the bounds (3.A.53), (3.A.54)-(3.A.56) and (3.A.58)-(3.A.59), we have that

$$\begin{aligned} f_x(\lambda) &= \frac{c_{k_0}^2}{k_0!} f_\xi^{(*k_0)}(\lambda) + \sum_{k=k_0+1}^{M-1} \frac{c_k^2}{k!} f_\xi^{(*k)}(\lambda) + f_z(\lambda) \\ &= \frac{c_{k_0}^2}{k_0!} f_\xi^{(*k_0)}(\lambda) + o\left(f_\xi^{(*k_0)}(\lambda)\right), \end{aligned} \tag{3.A.61}$$

as $\lambda \to 0+$, which implies that

$$f_x(\lambda) = \frac{c_{k_0}^2}{k_0!} f_\xi^{(*k_0)}(\lambda)(1+o(1)), \tag{3.A.62}$$

as $\lambda \to 0+$, and hence, the first part of (3.4.17) is shown. Using the last displayed equality together (3.A.54) and (3.A.58), implies that, as $\lambda \to 0+$

$$\begin{aligned} f_x(\lambda) &= c_{k_0}^2 s_{k_0} \lambda^{-1+k_0(1-\alpha_\xi)}(1+o(1)) + o(\lambda^{-1+k_0(1-\alpha_\xi)}) \\ &= c_{k_0}^2 s_{k_0} \lambda^{-\alpha_x} + o(\lambda^{-\alpha_x}), \end{aligned} \tag{3.A.63}$$

with $s_{k_0} = \frac{c_{0,\xi}^{k_0} C_{k_0}}{k_0!}$, when $k_0 \geq 2$ and $s_{k_0} = c_{0,\xi}$, when $k_0 = 1$, which proves the second part of (3.4.17).

ii) Set $M = k_0 + 1$ and write $x_t = y_t + z_t$ as in (3.A.51). Using similar arguments as in part i), we obtain that

$$f_x(\lambda) = f_y(\lambda) + f_z(\lambda) = \frac{c_{k_0}^2}{k_0!} f_\xi^{(*k_0)}(\lambda) + \sum_{k=k_0+1}^{\infty} \frac{c_k^2}{k!} f_\xi^{(*k)}(\lambda). \tag{3.A.64}$$

Since $k_0(1-\alpha_\xi) = 1$, we have by Lemma 3.1 parts ii) and iii) that, as $\lambda \to 0+$, $f_\xi^{(*k_0)}(\lambda) = O(\lambda^{-\delta})$ for every $\delta > 0$ and, $f_\xi^{(*k)}(\lambda) \leq C$ uniformly in $k \geq M$. Hence, the spectral density function function of $\{x_t\}_{t\in\mathbb{Z}}$ satisfies

$$f_x(\lambda) \leq O(\lambda^{-\delta}) + \sum_{k=k_0+1}^{\infty} \frac{c_k^2}{k!} C = O(\lambda^{-\delta}), \tag{3.A.65}$$

as $\lambda \to 0+$ and, for every $\delta > 0$, noticing that $\sum_{k=k_0+1}^{\infty} \frac{c_k^2}{k!} < \sum_{k=k_0}^{\infty} \frac{c_k^2}{k!} = \gamma_x(0) < \infty$ by (3.4.9).

iii) In this case we assume that $k_0(1-\alpha_\xi) > 1$ and hence, $k(1-\alpha_\xi) > 1$ for every $k \geq k_0$. So, from Lemma 3.1 part iii), it follows that, for all $\lambda \in (-\pi, \pi]$, $f_\xi^{(*k)}(\lambda) \leq C$ uniformly in $k \geq k_0$, which implies that

$$f_x(\lambda) = \sum_{k=k_0}^{\infty} \frac{c_k^2}{k!} f_\xi^{(*k)}(\lambda) \leq \sum_{k=k_0}^{\infty} \frac{c_k^2}{k!} C \leq C, \tag{3.A.66}$$

for all $\lambda \in (-\pi, \pi]$, using again that $\sum_{k=k_0}^{\infty} \frac{c_k^2}{k!} < \infty$ by (3.4.9). ■

Proof of Theorem 3.3. i) Assumption S.1 implies that the spectral density function of $\{x_t\}_{t\in\mathbb{Z}}$ satisfies Assumption A.3 with $\alpha_x = 0$ and $c_{0,x} = f_x(0)$. We show that

$$\Delta_{m,x} = \max_{1\leq k\leq m} E\left|\sum_{j=1}^{k} \frac{1}{c_{0,x}}(I_x(\lambda_j) - E(I_x(\lambda_j)))\right| = o(m). \tag{3.A.67}$$

Then, in view of Proposition 2.1, Assumption A.3 and (3.A.67) imply convergence (3.4.21).

Since $\{x_t\}_{t\in\mathbb{Z}}$ satisfies Assumption S.1, we have from Proposition 3.2 that

$$\sum_{\tau\in\mathbb{Z}} \left|\gamma_\xi(\tau)\right|^{k_0} < \infty. \tag{3.A.68}$$

Let $M > k_0$ be such that $c_{M+1} \neq 0$. Split up the sum in (3.4.8) in two sums,

$$x_t = \sum_{k=k_0}^{M} \frac{c_k}{k!} H_k(\xi_t) + \sum_{k=M+1}^{\infty} \frac{c_k}{k!} H_k(\xi_t) := y_t + z_t. \tag{3.A.69}$$

Using analogous arguments to those in (3.A.1), we write $I_x(\lambda_j) = I_y(\lambda_j) + v_j$, where $v_j = I_z(\lambda_j) + \frac{1}{2\pi}\left(w_y(\lambda_j)\overline{w_z(\lambda_j)} + \overline{w_y(\lambda_j)}w_z(\lambda_j)\right)$. Then,

$$\begin{aligned} E\left|\sum_{j=1}^{k}(I_x(\lambda_j) - E(I_x(\lambda_j)))\right| &= E\left|\sum_{j=1}^{k}((I_y(\lambda_j) - E(I_y(\lambda_j))) + (v_j - E(v_j)))\right| \\ &\leq E\left|S_k\right| + E\left|R_k\right|, \end{aligned} \tag{3.A.70}$$

where

$$S_k = \sum_{j=1}^{k}(I_y(\lambda_j) - E(I_y(\lambda_j))) \quad \text{and} \quad R_k = \sum_{j=1}^{k}(v_j - E(v_j)). \tag{3.A.71}$$

We first consider $E\left|R_k\right|$. From (3.A.3), we obtain that

$$E\left|v_j\right| \leq C\left(f_z(\lambda_j) + f_y^{\frac{1}{2}}(\lambda_j) f_z^{\frac{1}{2}}(\lambda_j)\right). \tag{3.A.72}$$

Recall that $z_t = \sum_{k=M+1}^{\infty} \frac{c_k}{k!} H_k(\xi_t)$ and therefore, expression (3.4.9) is satisfied for $\{z_t\}_{t\in\mathbb{Z}}$ with Hermite rank $M+1$. Then, the autocovariance function of $\{z_t\}_{t\in\mathbb{Z}}$ can be written as

$$\gamma_z(\tau) = \sum_{k=M+1}^{\infty} \frac{c_k^2}{k!} \gamma_\xi^k(\tau), \tag{3.A.73}$$

and therefore,

$$\left|\gamma_z(\tau)\right| \leq \sum_{k=M+1}^{\infty} \frac{c_k^2}{k!} \left|\gamma_\xi(\tau)\right|^k \leq \left|\gamma_\xi(\tau)\right|^{k_0} \sum_{k=M+1}^{\infty} \frac{c_k^2}{k!} = \left|\gamma_\xi(\tau)\right|^{k_0} \epsilon_M, \tag{3.A.74}$$

since $\left|\gamma_\xi(\tau)\right| \leq 1$, where $\epsilon_M = \sum_{k=M+1}^{\infty} \frac{c_k^2}{k!}$. From (3.A.68) we have that $\sum_{\tau\in\mathbb{Z}} \left|\gamma_\xi(\tau)\right|^{k_0} \leq C$, while $\epsilon_M \to 0$, as $M \to \infty$, because by (3.4.9) $\sum_{k=M+1}^{\infty} \frac{c_k^2}{k!} \leq \sum_{k=k_0}^{\infty} \frac{c_k^2}{k!} = \gamma_x(0) < \infty$.

Hence, the spectral density function of $\{z_t\}_{t\in\mathbb{Z}}$ can be bounded by

$$f_z(\lambda) \le \frac{1}{2\pi}\sum_{\tau\in\mathbb{Z}}|\gamma_z(\tau)| \le \frac{1}{2\pi}\sum_{\tau\in\mathbb{Z}}\left|\gamma_\xi(\tau)\right|^{k_0}\epsilon_M \le C\epsilon_M. \tag{3.A.75}$$

On the other hand, since $y_t = \sum_{k=k_0}^{M} \frac{c_k}{k!}H_k(\xi_t)$ and $\left|\gamma_\xi(\tau)\right| \le 1$, applying (3.4.9) and the bound (3.A.68) implies that

$$\begin{aligned} f_y(\lambda) &\le \frac{1}{2\pi}\sum_{\tau\in\mathbb{Z}}\left|\gamma_y(\tau)\right| \\ &\le C\sum_{\tau\in\mathbb{Z}}\sum_{k=k_0}^{M}\frac{c_k^2}{k!}\left|\gamma_\xi(\tau)\right|^{k} \\ &\le C\sum_{\tau\in\mathbb{Z}}\sum_{k=k_0}^{M}\frac{c_k^2}{k!}\left|\gamma_\xi(\tau)\right|^{k_0} \\ &\le C\sum_{\tau\in\mathbb{Z}}\left|\gamma_\xi(\tau)\right|^{k_0} < C, \end{aligned} \tag{3.A.76}$$

using again that $\sum_{k=k_0}^{M}\frac{c_k^2}{k!} \le \sum_{k=k_0}^{\infty}\frac{c_k^2}{k!} < \infty$. The latter displayed bound together with (3.A.75) and (3.A.72) imply that $E\left|v_j\right| \le C\epsilon_M^{\frac{1}{2}}$. Thus,

$$E\left|R_k\right| = E\left|\sum_{j=1}^{k}(v_j - E(v_j))\right| \le 2\sum_{j=1}^{k}E\left|v_j\right| \le 2\sum_{j=1}^{k}C\epsilon_M^{\frac{1}{2}} \le Cm\epsilon_M^{\frac{1}{2}}. \tag{3.A.77}$$

Hence,

$$E\left|R_k\right| = o(m), \tag{3.A.78}$$

as $M \to \infty$.

Next, we examine $E\left|S_k\right|$. Since the spectral density function of $\{y_t\}_{t\in\mathbb{Z}}$ is bounded, we have that $\{y_t\}_{t\in\mathbb{Z}}$ satisfies Assumption A.3 with $\alpha_y = 0$. Recall that $\{\xi_t\}_{t\in\mathbb{Z}}$ is a stationary Gaussian sequence with finite second moment, so that $\{y_t\}_{t\in\mathbb{Z}}$ is fourth-order stationary. Next, using the bound of Lemma 2.2 for $\{y_t\}_{t\in\mathbb{Z}}$ with $\alpha_y = 0$, we have that

$$m^{-1}\Delta_{m,y} = O\left(\log(m)\, m^{-\frac{1}{2}} + n^{-\frac{1}{2}}D_{n,y}^{*\frac{1}{2}}\right), \tag{3.A.79}$$

where

$$\Delta_{m,y} = \max_{1\le k\le m} E\left|\sum_{j=1}^{k}\frac{1}{c_{0,y}}(I_y(\lambda_j) - E(I_y(\lambda_j)))\right| \tag{3.A.80}$$

and

$$D_{n,y}^{*} = \sum_{t_1,t_2,t_3=-n}^{n}\left|c_y(t_1,t_2,t_3,0)\right|. \tag{3.A.81}$$

Then,

$$\begin{aligned} D^*_{n,y} &= \sum_{t_1,t_2,t_3=-n}^{n} |cum(y_{t_1}, y_{t_2}, y_{t_3}, y_0)| \\ &= \sum_{t_1,t_2,t_3=-n}^{n} |cum(y_{t_1+t_4+n}, y_{t_2+t_4+n}, y_{t_3+t_4+n}, y_{t_4+n})| \\ &= \frac{1}{n}\sum_{t_4=1}^{n}\sum_{t_1,t_2,t_3=-n}^{n} |cum(y_{t_1+t_4+n}, y_{t_2+t_4+n}, y_{t_3+t_4+n}, y_{t_4+n})| \\ &\leq \frac{1}{n}\sum_{t_1,t_2,t_3,t_4=1}^{3n} |cum(y_{t_1}, y_{t_2}, y_{t_3}, y_{t_4})|. \end{aligned} \tag{3.A.82}$$

Using the bound (3.B.30) of Lemma 3.2 part i), it follows that, as $n \to \infty$,

$$\begin{aligned} D^*_{n,y} &\leq \frac{1}{n}\sum_{t_1,t_2,t_3,t_4=1}^{3n} |cum(y_{t_1}, y_{t_2}, y_{t_3}, y_{t_4})| \\ &= \frac{1}{n}\sum_{t_1,t_2,t_3,t_4=1}^{3n} \left| cum\left(\sum_{k_1=k_0}^{M}\frac{c_{k_1}}{k_1!}H_{k_1}(\xi_{t_1}), \sum_{k_2=k_0}^{M}\frac{c_{k_2}}{k_2!}H_{k_2}(\xi_{t_2}),\right.\right. \\ &\quad \left.\left.\sum_{k_3=k_0}^{M}\frac{c_{k_3}}{k_3!}H_{k_3}(\xi_{t_3}), \sum_{k_4=k_0}^{M}\frac{c_{k_4}}{k_4!}H_{k_4}(\xi_{t_4})\right)\right| \\ &\leq \frac{1}{n}\sum_{t_1,t_2,t_3,t_4=1}^{3n}\sum_{k_1,k_2,k_4,k_4=k_0}^{M} \left|\frac{c_{k_1}}{k_1!}\frac{c_{k_2}}{k_2!}\frac{c_{k_3}}{k_3!}\frac{c_{k_4}}{k_4!}cum(H_{k_1}(\xi_{t_1}),\right. \\ &\quad \left. H_{k_2}(\xi_{t_2}), H_{k_3}(\xi_{t_3}), H_{k_4}(\xi_{t_4}))\right| \\ &= \frac{1}{n}\sum_{k_1,k_2,k_4,k_4=k_0}^{M}\left|\frac{c_{k_1}}{k_1!}\frac{c_{k_2}}{k_2!}\frac{c_{k_3}}{k_3!}\frac{c_{k_4}}{k_4!}\right|\sum_{t_1,t_2,t_3,t_4=1}^{3n}|cum(H_{k_1}(\xi_{t_1}), \\ &\quad H_{k_2}(\xi_{t_2}), H_{k_3}(\xi_{t_3}), H_{k_4}(\xi_{t_4}))| \\ &= \frac{1}{n}\sum_{k_1,k_2,k_4,k_4=k_0}^{M}\frac{c_{k_1}}{k_1!}\frac{c_{k_2}}{k_2!}\frac{c_{k_3}}{k_3!}\frac{c_{k_4}}{k_4!}o(n^2) = o(n). \end{aligned} \tag{3.A.83}$$

Hence, $D^*_{n,y} = o(n)$ and, from (3.A.79) we deduce that

$$\Delta_{m,y} = O(m^{\frac{1}{2}}\log(m) + mn^{-\frac{1}{2}}D^{*\frac{1}{2}}_{n,y}) = O(m^{\frac{1}{2}}\log(m) + mn^{-\frac{1}{2}}o(n^{\frac{1}{2}})) = o(m). \tag{3.A.84}$$

Therefore,

$$E|S_k| = E\left|\sum_{j=1}^{k}(I_y(\lambda_j) - E(I_y(\lambda_j)))\right| \leq c_{0,y}\Delta_{m,y} = o(m). \tag{3.A.85}$$

The estimates (3.A.78) and (3.A.85) together with (3.A.70) imply that

$$\begin{aligned}\Delta_{m,x} &= \max_{1\leq k\leq m} E\left|\sum_{j=1}^{k}\frac{1}{c_{0,x}}(I_x(\lambda_j)-E(I_x(\lambda_j)))\right| \\ &\leq \max_{1\leq k\leq m} E\left|S_k\right| + \max_{1\leq k\leq m} E\left|R_k\right| = o(m), \end{aligned} \tag{3.A.86}$$

as required.

ii) We first write $x_t = y_t + z_t$ as in (3.A.69) with $M = k_0$. We show that the spectral densities of $\{y_t\}_{t\in\mathbb{Z}}$ and $\{z_t\}_{t\in\mathbb{Z}}$ are such that

$$f_y(\lambda) = c_{0,y}\lambda^{-\alpha_y} + o(\lambda^{-\alpha_y}), \quad \text{as } \lambda \to 0+, \tag{3.A.87}$$

with $\alpha_y = 1 - k_0(1-\alpha_\xi)$ and $c_{0,y} = c_{k_0}^2 s_{k_0}$, for some $0 < s_{k_0} < \infty$, while

$$f_z(\lambda) \leq C\lambda^{-\alpha_z'}, \quad \text{as } \lambda \to 0+, \tag{3.A.88}$$

where

$$0 \leq \alpha_z' < \alpha_y. \tag{3.A.89}$$

In addition, we show that $\{y_t\}_{t\in\mathbb{Z}}$ satisfies Assumption A.2. Then convergence (3.4.22) follows by Theorem 3.1 part i).

We begin with the proof of (3.A.87). Notice that $y_t = \frac{c_{k_0}}{k_0!}H_{k_0}(\xi_t)$, and since $\{\xi_t\}_{t\in\mathbb{Z}}$ satisfies Assumption L.2, we have from Proposition 3.4 part i) that, the spectral density function of $\{y_t\}_{t\in\mathbb{Z}}$ satisfies

$$f_y(\lambda) = c_{k_0}^2 s_{k_0}\lambda^{-\alpha_y} + o\left(\lambda^{-\alpha_y}\right), \quad \text{as } \lambda \to 0+, \tag{3.A.90}$$

with $\alpha_y = 1 - k_0(1-\alpha_\xi)$ and for some $0 < s_{k_0} < \infty$, which shows (3.A.87) with $c_{0,y} = c_{k_0}^2 s_{k_0}$.

Next, we show (3.A.88). Recall that $z_t = \sum_{k=k_0+1}^{\infty} \frac{c_k}{k!}H_k(\xi_t)$ and denote $k_{0,z}$ the Hermite rank of $\{z_t\}_{t\in\mathbb{Z}}$. Since $\{\xi_t\}_{t\in\mathbb{Z}}$ satisfies Assumption L.2, we can apply Proposition 3.4 to $\{z_t\}_{t\in\mathbb{Z}}$. We consider three cases for $k_{0,z}(1-\alpha_\xi)$.

If $k_{0,z}(1-\alpha_\xi) < 1$, then we have by Proposition 3.4 part i) that, as $\lambda \to 0+$,

$$f_z(\lambda) = c_{k_{0,z}}^2 s_{k_{0,z}}\lambda^{-\alpha_z} + o(\lambda^{-\alpha_z}), \tag{3.A.91}$$

with $\alpha_z = 1 - k_{0,z}(1-\alpha_\xi)$ and some $0 < s_{k_{0,z}} < \infty$. Hence, (3.A.88) is satisfied with

$$0 < \alpha_z' = \alpha_z. \tag{3.A.92}$$

Since $k_{0,z} > k_0$, we have that

$$\alpha_z' < \alpha_y, \tag{3.A.93}$$

which together with (3.A.92) shows (3.A.89).

If $k_{0,z}(1-\alpha_\xi) = 1$, then it follows by Proposition 3.4 part ii) that, as $\lambda \to 0+$,

$$f_z(\lambda) \leq C\lambda^{-\delta}, \tag{3.A.94}$$

for any arbitrarily small $\delta > 0$. Therefore, (3.A.88) is satisfied with

$$\alpha_z' = \delta. \tag{3.A.95}$$

We can choose δ arbitrarily small, so that

$$0 < \alpha_z' < \alpha_y, \tag{3.A.96}$$

which shows (3.A.89).

If $k_{0,z}(1-\alpha_\xi) > 1$, then Proposition 3.4 part iii) implies that, for all $\lambda \in (-\pi, \pi]$,

$$f_z(\lambda) \leq C. \tag{3.A.97}$$

Hence, (3.A.88) is satisfied with

$$\alpha_z' = 0. \tag{3.A.98}$$

Since $0 < \alpha_y < 1$, we have that

$$0 = \alpha_z' < \alpha_y, \tag{3.A.99}$$

which implies (3.A.89).

It remains to show that $\{y_t\}_{t\in\mathbb{Z}}$ satisfies Assumption A.2. In the case that $k_0 = 1$, we have that $y_t = c_1\xi_t$. Then, since $\{\xi_t\}_{t\in\mathbb{Z}}$ is a Gaussian sequence with finite moments, we have that $\{y_t\}_{t\in\mathbb{Z}}$ is a linear process. In addition, as discussed above, Assumption A.3 is satisfied for $\{y_t\}_{t\in\mathbb{Z}}$. Hence, Proposition 2.3 part i) implies that $\{y_t\}_{t\in\mathbb{Z}}$ satisfies Assumption A.2. Next, consider the case $k_0 \geq 2$. Using the same argument as in part i), we have that $\{y_t\}_{t\in\mathbb{Z}}$ is fourth-order stationary sequence. In addition, it is evident from (3.A.87) that Assumption A.3 holds for $\{y_t\}_{t\in\mathbb{Z}}$ with $0 < \alpha_y < 1$. Hence, we can apply the bound (2.B.35) of Lemma 2.2 for $\{y_t\}_{t\in\mathbb{Z}}$,

$$m^{-1}\Delta_{m,y} = O\left(\log(m)\, m^{-\frac{1}{2}} + \left(\frac{D_{n,y}^{**}}{n}\right)^{\frac{1}{2}} \left(\frac{m}{n}\right)^{-1+\alpha_y} \log n\right), \tag{3.A.100}$$

where

$$D_{n,y}^{**} = \max_{|t_1|,|t_2|\leq n} \sum_{u=-n}^{n} |c_y(0, t_1, t_2+u, u)|. \tag{3.A.101}$$

Since Assumption A.3 holds for $\{y_t\}_{t\in\mathbb{Z}}$, we have by Proposition 2.1 that, $\{y_t\}_{t\in\mathbb{Z}}$ satisfies Assumption A.2 if

$$m^{-1}\Delta_{m,y} = o(1). \tag{3.A.102}$$

To show (3.A.102), we apply the bound (3.B.31) of Lemma 3.2 part ii), which implies that

$$\begin{aligned} |cum(y_{t_1}, y_{t_2}, y_{t_3}, y_{t_4})| &= C\left|cum(H_{k_0}(\xi_{t_1}), H_{k_0}(\xi_{t_2}), H_{k_0}(\xi_{t_3}), H_{k_0}(\xi_{t_4}))\right| \\ &\leq C\left(\gamma_\xi^2(t_1 - t_3) + \gamma_\xi^2(t_1 - t_4) + \gamma_\xi^2(t_2 - t_3)\right. \\ &\quad \left.+\gamma_\xi^2(t_2 - t_4)\right). \end{aligned} \tag{3.A.103}$$

By Assumption L.2, we have that $\gamma_\xi(\tau) \sim c_\xi \tau^{-1+\alpha_\xi}$, as $\tau \to \infty$. Therefore, $\gamma_\xi^2(\tau) \leq C\tau^{-2(1-\alpha_\xi)}$ for every $\tau \neq 0$. Notice also that $0 < 2(1-\alpha_\xi) < 1$, because $k_0 \geq 2$ and $0 < \alpha_\xi, \alpha_y < 1$ with $\alpha_y = 1 - k_0(1-\alpha_\xi)$. Hence,

$$\begin{aligned} D_{n,y}^{**} &= \max_{|t_1|,|t_2|\leq n} \sum_{u=-n}^{n} |cum(y_0, y_{t_1}, y_{t_2+u}, y_u)| \\ &\leq C \max_{|t_1|,|t_2|\leq n} \sum_{u=-n}^{n} \left(\gamma_\xi^2(-t_2-u) + \gamma_\xi^2(-u) + \gamma_\xi^2(t_1-t_2-u) + \gamma_\xi^2(t_1-u)\right) \\ &\leq C \sum_{\tau=-3n}^{3n} \gamma_\xi^2(\tau) \leq C\sum_{\tau=1}^{3n} \tau^{-2(1-\alpha_\xi)} \\ &\leq C \int_0^{3n} x^{-2(1-\alpha_\xi)}dx \leq Cn^{2\alpha_\xi - 1}, \end{aligned} \tag{3.A.104}$$

since $0 < 2(1-\alpha_\xi) < 1$. Using the latter displayed bound in (3.A.100), implies that

$$\begin{aligned} m^{-1}\Delta_{m,y} &= O\left(\log(m)\, m^{-\frac{1}{2}} + \left(\frac{n^{2\alpha_\xi-1}}{n}\right)^{\frac{1}{2}} \left(\frac{m}{n}\right)^{-1+\alpha_y} \log n\right) \\ &= O\left(\log(m)\, m^{-\frac{1}{2}} + \left(\left(\frac{n}{m}\right)^{k_0} n^{-1}\right)^{1-\alpha_\xi} \log n\right), \end{aligned} \tag{3.A.105}$$

since $\alpha_y = 1 - k_0(1-\alpha_\xi)$. Under the assumptions of the theorem, we have that $n^\gamma \leq m = o(n)$, for some $1 - \frac{1}{k_0} < \gamma < 1$. Hence, there exists some $\varepsilon > 0$ such that $n^{1-\frac{1}{k_0}+\varepsilon} \leq m$. Therefore, $(\frac{n}{m})^{k_0} n^{-1} \leq n^{-\varepsilon k_0} = O(n^{-\varepsilon})$ and so,

$$m^{-1}\Delta_{m,y} = O\left(\log(m)\, m^{-\frac{1}{2}} + n^{-\varepsilon(1-\alpha_\xi)} \log n\right) = o(1), \tag{3.A.106}$$

as required.

iii) We first write $x_t = y_t + z_t$ as in (3.A.69) with $M = k_0 = 1$. As in part ii) of this theorem, (3.A.87)-(3.A.89) hold. Hence, condition (3.2.2) of Theorem 3.1 is satisfied. We have that $y_t = c_1\xi_t$, where $\{\xi_t\}_{t\in\mathbb{Z}}$ is a stationary Gaussian sequence, so that $\{y_t\}_{t\in\mathbb{Z}}$ can be written as a linear process satisfying Assumption A.8. Under Assumption L.2 on the spectral density function of $\{\xi_t\}_{t\in\mathbb{Z}}$, $\{y_t\}_{t\in\mathbb{Z}}$ has spectral density function satisfying Assumption A.4 with $\beta_y = 2$. Moreover, by the assumptions of the theorem, we have that $\{y_t\}_{t\in\mathbb{Z}}$ satisfies Assumption A.6, and $n^\gamma \leq m = o(n)$ for some $\gamma > 0$, so that $m = o\left(\frac{n}{\log^{\frac{2}{\beta_y}} n}\right)$. Hence, we can apply Theorem 3.1 part iii), which yields that

$$\widehat{\alpha}_x - \alpha_y = O_p\left(m^{-\frac{1}{2}} + \left(\frac{m}{n}\right)^{\frac{\alpha_y - \alpha_z'}{2}}\right). \tag{3.A.107}$$

We show that

$$\alpha_y = \alpha_x, \tag{3.A.108}$$

and that the spectral density function of $\{z_t\}_{t\in\mathbb{Z}}$ is such that, as $\lambda \to 0+$,

$$f_z(\lambda) \leq C\lambda^{-\alpha_z'}, \tag{3.A.109}$$

with

$$0 \leq \alpha_z' < \alpha_y, \tag{3.A.110}$$

where

$$\alpha_z' = 0, \quad \text{if } \alpha_\xi < \frac{1}{2}, \tag{3.A.111}$$

$$0 \leq \alpha_z' \leq 2\alpha_\xi - 1, \quad \text{if } \alpha_\xi > \frac{1}{2}, \tag{3.A.112}$$

and

$$\alpha_z' = \delta, \quad \text{if } \alpha_\xi = \frac{1}{2}, \tag{3.A.113}$$

for arbitrarily small $\delta > 0$. Then, from (3.A.107), it follows that

$$\begin{aligned}\widehat{\alpha}_x - \alpha_x &= O_p\left(m^{-\frac{1}{2}} + \left(\frac{m}{n}\right)^{\frac{\alpha_x}{2}} + \left(\frac{m}{n}\right)^{\frac{1-\alpha_x}{2}}\right) \\ &= O_p\left(m^{-\frac{1}{2}} + \left(\frac{m}{n}\right)^{r}\right),\end{aligned} \tag{3.A.114}$$

with $r = \min\left\{\frac{\alpha_x}{2}, \frac{1-\alpha_x}{2}\right\}$, which shows (3.4.23).

As in part ii), see equality (3.A.92), we have that $\alpha_y = 1 - k_0(1-\alpha_\xi)$. Then (3.A.108) follows since $k_0 = 1$. Using the same arguments as in the proof of (3.A.88) and (3.A.89), we obtain expressions (3.A.109) and (3.A.110). It remains to show (3.A.111)-(3.A.113).

Consider first the case $\alpha_\xi < \frac{1}{2}$. Notice that $\{z_t\}_{t\in\mathbb{Z}}$ admits a Hermite expansion with rank $k_{0,z} \geq 2$. Then, (3.4.10) and Assumption L.2 on the autocovariance function of $\{\xi_t\}_{t\in\mathbb{Z}}$ imply that, for every $\tau \neq 0$,

$$|\gamma_z(\tau)| \leq C\left|\gamma_\xi(\tau)\right|^{k_{0,z}} \leq C\gamma_\xi^2(\tau) \leq C\tau^{-2(1-\alpha_\xi)}, \tag{3.A.115}$$

since $\left|\gamma_\xi(\tau)\right| \leq 1$. Thus,

$$\sum_{\tau\in\mathbb{Z}} |\gamma_z(\tau)| = 2\sum_{\tau=1}^{\infty} |\gamma_z(\tau)| + 1 \leq C\left(\sum_{\tau=1}^{\infty} \tau^{-2(1-\alpha_\xi)} + 1\right) < \infty, \tag{3.A.116}$$

because $\alpha_\xi < \frac{1}{2}$ implies that $2(1-\alpha_\xi) > 1$. Hence, the spectral density function of of $\{z_t\}_{t\in\mathbb{Z}}$ is bounded and satisfies (3.A.109) with $\alpha'_z = 0$, which implies (3.A.111).

Next consider the case $\alpha_\xi > \frac{1}{2}$. Notice again that $\{z_t\}_{t\in\mathbb{Z}}$ admits a Hermite expansion with rank $k_{0,z} \geq 2$. Now, as in part ii) of this theorem, we have by Proposition 3.4 part i) that, if $k_{0,z}(1-\alpha_\xi) < 1$, then, as $\lambda \to 0+$,

$$f_z(\lambda) = c_{0,z}\lambda^{-\alpha_z} + o(\lambda^{-\alpha_z}), \tag{3.A.117}$$

with $\alpha_z = 1 - k_{0,z}(1-\alpha_\xi)$ and $c_{0,z} = c^2_{k_{0,z}} s_{k_{0,z}}$, for some $0 < s_{k_{0,z}} < \infty$. Hence (3.A.109) is satisfied with $\alpha'_z = \alpha_z = 1 - k_{0,z}(1-\alpha_\xi) \leq 1 - 2(1-\alpha_\xi) = 2\alpha_\xi - 1$, since $k_{0,z} \geq 2$. If $k_{0,z}(1-\alpha_\xi) = 1$, then from Proposition 3.4 part ii), we obtain that, as $\lambda \to 0+$,

$$f_z(\lambda) \leq C\lambda^{-\delta}, \tag{3.A.118}$$

for arbitrarily small $\delta > 0$. Therefore, (3.A.109) is satisfied with $\alpha'_z = \delta \leq 2\alpha_\xi - 1$ for small enough $\delta > 0$, since $\alpha_\xi > \frac{1}{2}$. If $k_{0,z}(1-\alpha_\xi) > 1$, then by Proposition 3.4 part iii), it holds that, for all $\lambda \in (-\pi, \pi]$,

$$f_z(\lambda) \leq C. \tag{3.A.119}$$

Thus, (3.A.109) is satisfied with $\alpha'_z = 0 < 2\alpha_\xi - 1$, since $\alpha_\xi > \frac{1}{2}$. In all three cases, we obtain that $\alpha'_z < 2\alpha_\xi - 1$ when $\alpha_\xi > \frac{1}{2}$, which shows (3.A.112).

Finally, consider the case that $\alpha_\xi = \frac{1}{2}$. Since $k_{0,z} \geq 2$, we have that $k_{0,z}(1-\alpha_\xi) \geq 1$. If $k_{0,z}(1-\alpha_\xi) = 1$, then Proposition 3.4 part ii) implies that, as $\lambda \to 0+$,

$$f_z(\lambda) \leq C\lambda^{-\delta}, \tag{3.A.120}$$

for arbitrarily small $\delta > 0$. Therefore, (3.A.109) is satisfied with $\alpha'_z = \delta < \alpha_\xi$ for small enough $\delta > 0$, which shows (3.A.113). If $k_{0,z}(1-\alpha_\xi) > 1$, then by Proposition 3.4 part iii), it holds that, for all $\lambda \in (-\pi, \pi]$,

$$f_z(\lambda) \leq C. \tag{3.A.121}$$

Thus, (3.A.109) is satisfied with $\alpha'_z = \delta < \alpha_\xi$, which implies (3.A.113).

Finally, notice that if in addition Assumption A.7 is satisfied for $\{\xi_t\}_{t\in\mathbb{Z}}$, then using the same arguments as in (3.A.114), it follows from Theorem 3.1 part iii) that

$$m^{\frac{1}{2}}(\widehat{\alpha}_x - \alpha_x) \xrightarrow{d} N(0,1), \quad \text{as } n \to \infty, \tag{3.A.122}$$

with $m = o\left(n^{\frac{2r}{2r+1}}\right)$, which shows (3.4.24) and completes the proof of the theorem. ■

3.B Appendix

This section contains a series of technical lemmas used in the proofs in Appendix 3.A above.

Lemma 3.1

Let $k \geq 2$. Suppose that $\{\xi_t\}_{t\in\mathbb{Z}}$ satisfies Assumption L.2.

i) If $k(1-\alpha_\xi) < 1$, then, as $\lambda \to 0+$,

$$f_\xi^{(*k)}(\lambda) = c_{0,\xi}^k C_k \lambda^{-1+k(1-\alpha_\xi)} + o\left(\lambda^{-1+k(1-\alpha_\xi)}\right), \tag{3.B.1}$$

for some $0 < C_k < \infty$.

ii) If $k(1-\alpha_\xi) = 1$, then, as $\lambda \to 0+$,

$$f_\xi^{(*k)}(\lambda) \leq C\lambda^{-\delta}, \tag{3.B.2}$$

for any $\delta > 0$.

iii) If $k(1-\alpha_\xi) > 1$, then, for every $\lambda \in (-\pi, \pi]$,

$$f_\xi^{(*k)}(\lambda) \leq C. \tag{3.B.3}$$

Proof. Let $0 < \varepsilon < 1$. By definition,

$$\begin{aligned} f_\xi^{(*k)}(\lambda) &= \int_{-\pi}^{\pi} \ldots \int_{-\pi}^{\pi} f_\xi(\lambda - l_1 - \ldots - l_{k-1}) f_\xi(l_1) \ldots f_\xi(l_{k-1}) dl_1 \ldots dl_{k-1} \\ &= \int_{-\varepsilon}^{\varepsilon} \ldots \int_{-\varepsilon}^{\varepsilon} f_\xi(\lambda - l_1 - \ldots - l_{k-1}) f_\xi(l_1) \ldots f_\xi(l_{k-1}) dl_1 \ldots dl_{k-1} \\ &\quad + \int \ldots \int_{\exists \text{ at least one } l_p : |l_p| \geq \varepsilon} f_\xi(\lambda - l_1 - \ldots - l_{k-1}) f_\xi(l_1) \ldots f_\xi(l_{k-1}) dl_1 \ldots dl_{k-1} \\ &:= I_1(\lambda) + I_2(\lambda). \end{aligned} \tag{3.B.4}$$

i) We show that for a given $0 < \varepsilon < 1$, as $|\lambda| \to 0$,

$$I_1(\lambda) = c_{0,\xi}^k |\lambda|^{-1+k(1-\alpha_\xi)} C_\varepsilon(\lambda) + O(\varepsilon^2 |\lambda|^{-1+k(1-\alpha_\xi)} C_\varepsilon(\lambda)), \tag{3.B.5}$$

where $C_\varepsilon(\lambda)$, given in (3.B.15) below, is such that, for any fixed $0 < \varepsilon < 1$, as $|\lambda| \to 0$,

$$C_\varepsilon(\lambda) \to C_k, \tag{3.B.6}$$

with some constant $0 < C_k < \infty$ defined in (3.B.17) below. Then, (3.B.5) and (3.B.6) imply that, for any given $0 < \varepsilon < 1$, as $|\lambda| \to 0$,

$$I_1(\lambda) = c_{0,\xi}^k C_k |\lambda|^{-1+k(1-\alpha_\xi)} + O\left(\varepsilon^2 |\lambda|^{-1+k(1-\alpha_\xi)}\right). \tag{3.B.7}$$

Since ε can be made arbitrarily small, we obtain that, as $|\lambda| \to 0$,

$$I_1(\lambda) = c_{0,\xi}^k C_k |\lambda|^{-1+k(1-\alpha_\xi)} + o\left(|\lambda|^{-1+k(1-\alpha_\xi)}\right). \tag{3.B.8}$$

Furthermore, we show that, for all $\lambda \in (-\pi, \pi]$,

$$I_2(\lambda) = O(1), \tag{3.B.9}$$

which together with (3.B.4) and (3.B.8) imply that, as $\lambda \to 0+$,

$$f_\xi^{(*k)}(\lambda) = c_{0,\xi}^k C_k \lambda^{-1+k(1-\alpha_\xi)} + o\left(\lambda^{-1+k(1-\alpha_\xi)}\right), \tag{3.B.10}$$

and hence part i) of the lemma holds.

We start by showing (3.B.5). By Assumption L.2 we have that, as $|\lambda| \to 0$,

$$f_\xi(\lambda) = |\lambda|^{-\alpha_\xi} \left(c_{0,\xi} + c_{1,\xi} |\lambda|^2 + o\left(|\lambda|^2\right)\right) = c_{0,\xi} |\lambda|^{-\alpha_\xi} \left(1 + O\left(|\lambda|^2\right)\right). \tag{3.B.11}$$

Hence, for $|\lambda| \le \varepsilon$,

$$f_\xi(\lambda) = c_{0,\xi} |\lambda|^{-\alpha_\xi} \left(1 + O\left(\varepsilon^2\right)\right). \tag{3.B.12}$$

Thus,

$$\begin{aligned}
I_1(\lambda) &= \int_{-\varepsilon}^{\varepsilon} \dots \int_{-\varepsilon}^{\varepsilon} f_\xi(\lambda - l_1 - \dots - l_{k-1}) f_\xi(l_1) \dots f_\xi(l_{k-1}) dl_1 \dots dl_{k-1} \\
&= c_{0,\xi}^k \int_{-\varepsilon}^{\varepsilon} \dots \int_{-\varepsilon}^{\varepsilon} |\lambda - l_1 - \dots - l_{k-1}|^{-\alpha_\xi} \left(1 + O\left(\varepsilon^2\right)\right) \\
&\quad \times |l_1|^{-\alpha_\xi} \left(1 + O\left(\varepsilon^2\right)\right) \dots |l_{k-1}|^{-\alpha_\xi} \left(1 + O\left(\varepsilon^2\right)\right) dl_1 \dots dl_{k-1} \\
&= c_{0,\xi}^k \left(1 + O\left(\varepsilon^2\right)\right) \int_{-\varepsilon}^{\varepsilon} \dots \int_{-\varepsilon}^{\varepsilon} |\lambda - l_1 - \dots - l_{k-1}|^{-\alpha_\xi} \\
&\quad \times |l_1|^{-\alpha_\xi} \dots |l_{k-1}|^{-\alpha_\xi} dl_1 \dots dl_{k-1}.
\end{aligned} \tag{3.B.13}$$

By the change of variables $l'_p = \frac{l_p}{\lambda}$, $p = 1, ..., k-1$, we obtain that

$$\begin{aligned} I_1(\lambda) &= c_{0,\xi}^k \left|\lambda\right|^{-1+k(1-\alpha_\xi)} \left(1 + O\left(\varepsilon^2\right)\right) \int_{-\frac{\varepsilon}{\lambda}}^{\frac{\varepsilon}{\lambda}} ... \int_{-\frac{\varepsilon}{\lambda}}^{\frac{\varepsilon}{\lambda}} \left|1 - l'_1 - ... - l'_{k-1}\right|^{-\alpha_\xi} \\ &\quad \times \left|l'_1\right|^{-\alpha_\xi} ... \left|l'_{k-1}\right|^{-\alpha_\xi} dl'_1 ... dl'_{k-1} \\ &= c_{0,\xi}^k \left|\lambda\right|^{-1+k(1-\alpha_\xi)} C_\varepsilon(\lambda) + O\left(\varepsilon^2 \left|\lambda\right|^{-1+k(1-\alpha_\xi)} C_\varepsilon(\lambda)\right), \end{aligned} \tag{3.B.14}$$

which proves (3.B.5) with

$$C_\varepsilon(\lambda) = \int_{-\frac{\varepsilon}{\lambda}}^{\frac{\varepsilon}{\lambda}} ... \int_{-\frac{\varepsilon}{\lambda}}^{\frac{\varepsilon}{\lambda}} \left|1 - l_1 - ... - l_{k-1}\right|^{-\alpha_\xi} \left|l_1\right|^{-\alpha_\xi} ... \left|l_{k-1}\right|^{-\alpha_\xi} dl_1 ... dl_{k-1}. \tag{3.B.15}$$

Next, we show (3.B.6). Since the integrated function in $C_\varepsilon(\lambda)$ is positive, we have that, for any fixed $0 < \varepsilon < 1$,

$$C_\varepsilon(\lambda) \nearrow C_k, \quad \text{as } |\lambda| \searrow 0, \tag{3.B.16}$$

where

$$C_k = \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} \left|1 - l_1 - ... - l_{k-1}\right|^{-\alpha_\xi} \left|l_1\right|^{-\alpha_\xi} ... \left|l_{k-1}\right|^{-\alpha_\xi} dl_1 ... dl_{k-1}. \tag{3.B.17}$$

It remains to establish that

$$C_k < \infty. \tag{3.B.18}$$

We use the inequality

$$\int_{-\infty}^{\infty} \left|y - x\right|^{-\beta_1} \left|x\right|^{-\beta_2} dx \le C \left|y\right|^{1-(\beta_1+\beta_2)}, \tag{3.B.19}$$

for any $0 < \beta_1, \beta_2 < 1$ with $\beta_1 + \beta_2 > 1$, which holds true because, after a change of variables $x' = \frac{x}{y}$, the integral in (3.B.19) becomes

$$\begin{aligned} &\left|y\right|^{1-(\beta_1+\beta_2)} \int_{-\infty}^{\infty} \left|1 - x'\right|^{-\beta_1} \left|x'\right|^{-\beta_2} dx' \\ &= \left|y\right|^{1-(\beta_1+\beta_2)} \left(\int_{x \in \mathbb{R}: |x| \le \frac{1}{2}} \left|1 - x\right|^{-\beta_1} \left|x\right|^{-\beta_2} dx + \int_{x \in \mathbb{R}: \frac{1}{2} < |x| \le 2} \left|1 - x\right|^{-\beta_1} \left|x\right|^{-\beta_2} dx \right. \end{aligned}$$

$$
\begin{aligned}
&+ \int\limits_{x\in\mathbb{R}:|x|>2} |1-x|^{-\beta_1}|x|^{-\beta_2}\,dx \Bigg) \\
&\leq |y|^{1-(\beta_1+\beta_2)} \Bigg(2^{\beta_1} \int\limits_{x\in\mathbb{R}:|x|\leq\frac{1}{2}} |x|^{-\beta_2}\,dx + 2^{\beta_2} \int\limits_{x\in\mathbb{R}:\frac{1}{2}<|x|\leq 2} |1-x|^{-\beta_1}\,dx \\
&+ 2^{\beta_1} \int\limits_{x\in\mathbb{R}:|x|>2} |x|^{-\beta_1}|x|^{-\beta_2}\,dx \Bigg) \\
&\leq C|y|^{1-(\beta_1+\beta_2)}, \qquad (3.B.20)
\end{aligned}
$$

since $0<\beta_1,\beta_2<1$ and $\beta_1+\beta_2>1$. To prove (3.B.18) we use repeatedly the inequality (3.B.19) with $\beta_1=(p-1)\alpha_\xi-(p-2)$ and $\beta_2=\alpha_\xi$ for $p=2,...,k$. Recall that $0<\alpha_\xi<1$, $0<k(1-\alpha_\xi)<1$ and $k\geq k_0+1\geq 2$. The latter inequalities imply that, for $p=2,...,k$, we have $1-\frac{1}{p}<\alpha_\xi<1$ and hence $\frac{1}{p}<(p-1)\alpha_\xi-(p-2)<1$ to show that $0<\beta_1<1$. It is evident that $0<\beta_2<1$ and also, $\beta_1+\beta_2=p\alpha_\xi-(p-2)>1$ for $p=2,...,k$, since $\alpha_\xi>1-\frac{1}{p}$. Hence,

$$
\begin{aligned}
C_k &\leq C\int_{-\infty}^{\infty}...\int_{-\infty}^{\infty}|1-l_2-...-l_{k-1}|^{-(2\alpha_\xi-1)}|l_2|^{-\alpha_\xi}...|l_{k-1}|^{-\alpha_\xi}\,dl_2...dl_{k-1} \\
&\leq ...\leq C^{k-2}\int_{-\infty}^{\infty}|1-l_{k-1}|^{-((k-1)\alpha_\xi-(k-2))}|l_{k-1}|^{-\alpha_\xi}\,dl_{k-1} \\
&\leq C^{k-1}<\infty, \qquad (3.B.21)
\end{aligned}
$$

as required.

Finally, we show (3.B.9). For $p=1,..,k-1$, we define the set $A_p=\{l_1,...,l_{k-1}\in(-\pi,\pi]:|l_p|\geq\varepsilon\}$. Then, for all $\lambda\in(-\pi,\pi]$,

$$
\begin{aligned}
I_2(\lambda) &= \int\limits_{\exists\text{ at least one } l_p:|l_p|\geq\varepsilon}\!\!...\int f_\xi(\lambda-l_1-...-l_{k-1})f_\xi(l_1)...f_\xi(l_{k-1})dl_1...dl_{k-1} \\
&\leq \sum_{p=1}^{k-1}\int...\int_{A_p} f_\xi(\lambda-l_1-...-l_{k-1})f_\xi(l_1)...f_\xi(l_{k-1})dl_1...dl_{k-1}. \qquad (3.B.22)
\end{aligned}
$$

Without loss of generality, we consider the integral

$$
\int...\int_{A_1} f_\xi(\lambda-l_1-...-l_{k-1})f_\xi(l_1)...f_\xi(l_{k-1})dl_1...dl_{k-1}
$$

$$= \int_{-\pi}^{\pi} ... \int_{-\pi}^{\pi} \int_{l_1 \in (-\pi,\pi]:|l_1| \geq \varepsilon} f_\xi(\lambda - l_1 - ... - l_{k-1}) \times f_\xi(l_1)...f_\xi(l_{k-1}) dl_1...dl_{k-1}. \quad (3.B.23)$$

Since $|l_1| \geq \varepsilon$, we have under Assumption L.2 that $f_\xi(l_1) \leq C$. Hence, the last displayed integral is bounded by

$$\begin{aligned}
& C \int_{-\pi}^{\pi} ... \int_{-\pi}^{\pi} \int_{l_1 \in (-\pi,\pi]:|l_1| \geq \varepsilon} f_\xi(\lambda - l_1 - ... - l_{k-1}) f_\xi(l_2)...f_\xi(l_{k-1}) dl_1...dl_{k-1} \\
= \; & C \int_{-\pi}^{\pi} ... \int_{-\pi}^{\pi} f_\xi(l_2)...f_\xi(l_{k-1}) \\
& \times \left(\int_{l_1 \in (-\pi,\pi]:|l_1| \geq \varepsilon} f_\xi(\lambda - l_1 - ... - l_{k-1}) dl_1 \right) dl_2...dl_{k-1} \\
\leq \; & C \int_{-\pi}^{\pi} ... \int_{-\pi}^{\pi} f_\xi(l_2)...f_\xi(l_{k-1}) \left(\int_{-k\pi}^{k\pi} f_\xi(\lambda) d\lambda \right) dl_2...dl_{k-1} \qquad (3.B.24) \\
\leq \; & C \left(\int_{-\pi}^{\pi} f_\xi(\lambda) d\lambda \right)^{k-1} = C, \qquad (3.B.25)
\end{aligned}$$

since $\int_{-\pi}^{\pi} f_\xi(\lambda) d\lambda = Var(\xi_t) = 1$. Recall that in (3.B.24) we set $f_\xi(\lambda) = f_\xi(\lambda + 2\pi)$ for $\lambda \in \mathbb{R}$. From (3.B.22) and (3.B.25), (3.B.9) follows.

ii) As in part i), we have that, for all $\lambda \in (-\pi, \pi]$, $I_2(\lambda) = O(1)$. We show that $I_1(\lambda) = O\left(|\lambda|^{-\delta}\right)$ for every $\delta > 0$, as $|\lambda| \to 0$. Then (3.B.2) follows from (3.B.4). Now, as in part i) we have that, for any $0 < \varepsilon < 1$, and $|\lambda| \leq \varepsilon$,

$$\begin{aligned}
I_1(\lambda) \;=\; & c_{0,\xi}^k (1 + O(\varepsilon^2)) \int_{-\varepsilon}^{\varepsilon} ... \int_{-\varepsilon}^{\varepsilon} |\lambda - l_1 - ... - l_{k-1}|^{-\alpha_\xi} \\
& \times |l_1|^{-\alpha_\xi} ... |l_{k-1}|^{-\alpha_\xi} dl_1...dl_{k-1}. \qquad (3.B.26)
\end{aligned}$$

Let α'_ξ be such that $0 < \alpha_\xi < \alpha'_\xi < 1$ and set $\gamma = \alpha'_\xi - \alpha_\xi$. Then, $0 < k(1 - \alpha'_\xi) < k(1 - \alpha_\xi) = 1$ and, as in (3.B.14),

$$\begin{aligned}
I_1(\lambda) \;\leq\; & c_{0,\xi}^k (1 + O(\varepsilon^2)) \int_{-\varepsilon}^{\varepsilon} ... \int_{-\varepsilon}^{\varepsilon} |\lambda - l_1 - ... - l_{k-1}|^{-\alpha'_\xi} \\
& \times |l_1|^{-\alpha'_\xi} ... |l_{k-1}|^{-\alpha'_\xi} dl_1...dl_{k-1}
\end{aligned}$$

$$= O\left(|\lambda|^{-1+k(1-\alpha'_\xi)}\right) = O\left(|\lambda|^{-1+k(1-\alpha_\xi)-k\gamma}\right)$$
$$= O\left(|\lambda|^{-k\gamma}\right) = O\left(|\lambda|^{-\delta}\right), \qquad (3.B.27)$$

as $|\lambda| \to 0$, where $\delta = k\gamma$. Since γ can be made arbitrarily small, we have that, for every $\delta > 0$, $I_1(\lambda) = O\left(|\lambda|^{-\delta}\right)$, as $|\lambda| \to 0$.

iii) By Assumption L.2, we have that $\left|\gamma_\xi(\tau)\right| \leq C\,|\tau|^{-1+\alpha_\xi}$, for every $|\tau| \neq 0$. Hence,

$$\sum_{\tau\in\mathbb{Z}} \left|\gamma_\xi(\tau)\right|^k \leq C \sum_{\tau\in\mathbb{Z}:\tau\neq 0} |\tau|^{-k(1+\alpha_\xi)} + 1 < C, \qquad (3.B.28)$$

since $k(1-\alpha_\xi) > 1$. The latter displayed inequality and (3.A.41) imply that, for every $\lambda \in (-\pi, \pi]$,

$$f_\xi^{(*k)}(\lambda) \leq C \sum_{\tau\in\mathbb{Z}} \left|\gamma_\xi(\tau)\right|^k \leq C, \qquad (3.B.29)$$

which completes the proof of this lemma. ■

Lemma 3.2

Let $k_0 \geq 2$ and $\{\xi_t\}_{t\in\mathbb{Z}}$, $t \in \mathbb{Z}$ be a stationary zero mean and unit variance Gaussian sequence such that $\sum_{\tau\in\mathbb{Z}} \left|\gamma_\xi(\tau)\right|^{k_0} < \infty$. Then the Hermite polynomials $H_k(.), k \geq k_0$ have the following properties:

i) For any $k_1, ..., k_4 \geq k_0$,

$$\sum_{t_1,t_2,t_3,t_4=1}^{n} \left|cum\left(H_{k_1}(\xi_{t_1}), H_{k_2}(\xi_{t_2}), H_{k_3}(\xi_{t_3}), H_{k_4}(\xi_{t_4})\right)\right| = o\left(n^2\right). \qquad (3.B.30)$$

ii) For any fixed $k \geq 2$, uniformly in $t_1, ..., t_4 \in \mathbb{Z}$,

$$\left|cum(H_k(\xi_{t_1}), H_k(\xi_{t_2}), H_k(\xi_{t_3}), H_k(\xi_{t_4}))\right| \leq C(\gamma_\xi^2(t_1-t_3) + \gamma_\xi^2(t_1-t_4) + \gamma_\xi^2(t_2-t_3) + \gamma_\xi^2(t_2-t_4)) \qquad (3.B.31)$$

Proof. i) Let $j \geq 3$. From the proof of bound (2.9) in Giraitis and Surgailis (1985), p. 201 it is evident that for every $k_1, ..., k_j \geq k_0$,

$$\sum_{t_1,...,t_j=1}^{n} \left|cum\left(H_{k_1}(t_1), ..., H_{k_j}(t_j)\right)\right| = o\left(n^{\frac{j}{2}}\right). \qquad (3.B.32)$$

In our case we have $j = 4$, which proves (3.B.30).

ii) It is well known, see for example Giraitis and Surgailis (1985) pp. 194-195, that

$$cum(H_k(\xi_{t_1}), H_k(\xi_{t_2}), H_k(\xi_{t_3}), H_k(\xi_{t_4})) = \sum_V d_{V_1}...d_{V_r}, \tag{3.B.33}$$

where $V = \{V_1, .., V_r\}$ is a partition of the table W :

$$W = \begin{pmatrix} (1,1) & ... & (1,k) \\ (2,1) & ... & (2,k) \\ (3,1) & ... & (3,k) \\ (4,1) & ... & (4,k) \end{pmatrix} \tag{3.B.34}$$

into subsets (edges) $V_1 = \{(i_{1,1}, j_{1,1}), (i_{2,1}, j_{2,1})\}, .., V_r = \{(i_{1,r}, j_{1,r}), (i_{2,r}, j_{2,r})\}$ where $r = 2k$.

For $l = 1, .., r$, we define $d_{V_l} = \gamma_\xi (t_i - t_{i'})$ when $V_r = \{(i,j), (i',j')\}$. The sum $\sum_V$ in (3.B.33) is taken over all connected partitions $V = \{V_1, .., V_r\}$ without flat edges. Recall that a diagram $V = \{V_1, .., V_r\}$ is called connected when the rows of the table W cannot be divided into two groups, each of which is partitioned by the diagram separately. A diagram $V = \{V_1, .., V_r\}$ is said to have no flat edges if $i_{1,l} \neq i_{2,l}$ for all $l = 1, .., r$.

It suffices to show that for any partition $V = \{V_1, .., V_r\}$,

$$|d_{V_1}...d_{V_r}| \leq C(\gamma_\xi^2 (t_1 - t_3) + \gamma_\xi^2 (t_1 - t_4) + \gamma_\xi^2 (t_2 - t_3) + \gamma_\xi^2 (t_2 - t_4)). \tag{3.B.35}$$

Since the sum in (3.B.33) has a finite number of terms, then (3.B.31) follows from (3.B.33) and (3.B.35). Suppose that there are two edges V_k, V_s of the partition V connecting rows 1, 2 with rows 3, 4 of the table W. Since $|d_{V_l}| \leq 1$ for all $l = 1, .., r$, then

$$|d_{V_1}...d_{V_r}| \leq |d_{V_k} d_{V_s}| \leq \frac{1}{2} \left(|d_{V_k}|^2 + |d_{V_s}|^2\right). \tag{3.B.36}$$

There are four possible cases for the values d_{V_k} : $d_{V_k} = \gamma_\xi (t_1 - t_3)$, $d_{V_k} = \gamma_\xi (t_1 - t_4)$, $d_{V_k} = \gamma_\xi (t_2 - t_3)$, $d_{V_k} = \gamma_\xi (t_2 - t_4)$. Hence, (3.B.35) follows.

It remains to show that each partition $V = \{V_1, .., V_r\}$ has edges V_k, V_s connecting rows 1, 2 with rows 3, 4 of the table W. Since the diagram is connected, there exists at least one edge V_k connecting rows 1, 2 with rows 3, 4. Without loss of generality, we assume that $V_k = \{(1,1), (3,1)\}$ and suppose that none of the other edges of V have this property, i.e. there are no other connections between rows 1, 2 and rows 3, 4. Since there are no other connections between rows 1, 2 and rows 3, 4, we have $k - 1$ elements in row 1 that have to be connected with the k elements of row 2. This is impossible, since any edge has to contain 2 points.

$$\begin{array}{cccc} & (1,2) & \dots & (1,k) \\ (2,1) & (2,2) & \dots & (2,k) \end{array} \tag{3.B.37}$$

Hence, each partition $V = \{V_1, .., V_r\}$ has at least two edges V_k, V_s connecting rows 1, 2 with rows 3, 4 of the table W. ■

Chapter 4

Local Whittle estimation: Monte-Carlo simulations and empirical applications

4.1 Introduction

In the previous two chapters we analyzed the LW estimator of the memory parameter under various setups, linear and nonlinear ones. We found that the consistency property of the estimator is not affected by the presence of nonlinearity. However, our results also suggested that its finite sample properties are likely to be worse off when compared to the case of a linear process. We now turn to establish our findings by means of Monte-Carlo simulations. Furthermore, we apply LW estimation to real data to access their long-run persistence and address several issues that have appeared in the empirical literature.

We start by performing Monte-Carlo simulations to examine the finite sample bias, standard deviation and root mean squared error (RMSE) of the LW estimator under the linear specification and nonlinear setups discussed in Chapter 3. The main purpose of the experiments is to assess whether the LW estimator is indeed consistent for the nonlinear models we have considered, and under these specifications to contrast its finite sample behaviour against that of a linear model. In addition, we want to address the issue of bandwidth choice, which according to our findings is affected by the presence of nonlinearity. The Monte-Carlo simulations also aim to confirm the various remarks reported in Chapters 2 and 3. In particular, we seek to establish whether the LW estimator performs better under short memory than under long or negative memory. Furthermore, we aim to find how the finite sample bias behaves in the signal plus noise model under various combinations of the memory parameters of the signal and the noise and, different forms of dependence between the two processes, as well as various magnitudes of the signal-to-noise ratio. Moreover, in the structural model, we attempt to understand how strong cyclical persistence might affect the finite sample bias of the LW estimator when the frequency of the cyclical component is small. We also seek to establish if in the case of long memory nonlinear transformations of a Gaussian process, the finite sample bias is negative and the LW estimator performs better for moderate values of long memory. Finally, for the LMSV model having an exponential representation for the volatility process, we address the finite sample performance of the LW estimator for different transformations of the return series. Overall, the Monte-Carlo experiments performed here confirm our findings and remarks presented in Chapters 2 and 3.

As a next step, we apply the LW estimator to real data sets. We start with examining the degree of long-run persistence of different measurements of inflation and expected inflation rates. Although, the rational expectations hypothesis implies that all these series should have the same degree of long-run persistence, we find substantial differences in the point estimates of their memory parameters. The latter observation however can be easily explained by our findings on the signal plus noise model. Since the various measures of inflation and expected inflation rates are associated with short-run components of different variation, the differences in the point estimates of their memory parameters are attributed to the effect of the signal-to-noise ratio. We also apply LW estimation to the nominal and real interest rates in order to assess their degree of long-run persistence. Over a particular time span, the estimates of their memory parameters are unrealistically high, as they suggest that the two series have an explosive behaviour. Using a monetary policy function along with the results on the structural model, we provide an explanation as to why these unexpectedly high estimates arise. We conjecture that the LW estimates are actually driven by the strong cyclical component of these series and overestimate the true degree of long-run persistence. Finally, we examine several foreign exchange rates and apply the LW estimator to different transformations of their returns. The observed estimates vary across the different transformation employed, a result that has already been reported in the literature for other exchange rates and asset returns, and that is in line with our findings on the LMSV model.

The remainder of this chapter is as follows. In Section 4.2, we discuss and comment on our Monte-Carlo experiments on linear process, signal plus noise model, structural model, nonlinear transformations of a Gaussian process and LMSV model. Section 4.3 deals with applying the LW estimator on inflation and expected inflation rate, nominal and real interest rate, and various exchanges rates. Some final comments are found in Section 4.4, while Appendix 4.A contains the results of the Monte-Carlo experiments of Section 4.2, and Appendix 4.B the data and the corresponding LW estimates examined in Section 4.3.

4.2 Monte-Carlo simulations

We now present our Monte-Carlo simulations. In the majority of our experiments, we employ the Davies and Harte (1987) algorithm to generate Gaussian $ARFIMA(0, d, 0)$ processes. In the signal plus noise model, when examining the effect of the dependence between the signal and noise processes, as well as in the structural model, we generate Gaussian $ARFIMA(0, d, 0)$ and $GARMA(0, d_\omega, 0)$ processes by truncating at $5,000$ terms the $MA(\infty)$ representation of these models. We carry out $5,000$ replications of sample sizes $n = 128, 512, 2048$ and take the bandwidth parameter to be of the form $m = [n^\gamma]$ with $\gamma = 0.5, 0.525, 0.55, ..., 0.8$. We calculate the Monte-Carlo bias, standard deviation and RMSE. In all the experiments below, the Monte-Carlo standard deviation decreases with increases in n and m. It is almost the same across the various models for a given n and m, and hence any differences in the RMSE across the various models are driven by differences in the bias. The Monte-Carlo standard deviation is not reported here, while the Monte-Carlo bias and RMSE are found in Appendix 4.A.

4.2.1 Linear process

We start by simulating standard Gaussian $ARFIMA(0, \frac{\alpha_x}{2}, 0)$ processes with $\alpha_x = -0.8, -0.4, 0, 0.4, 0.8$. The Monte-Carlo bias and RMSE are reported in Figures 4.1-4.10. The results are in line with those of Robinson (1995b). The Monte-Carlo bias and RMSE decrease in absolute value with an increase in n, suggesting that the LW estimator is consistent. For a given n and m, the magnitude of the bias and RMSE are on similar levels for the different values of the memory parameter α_x. Overall, the bias tends to increase in absolute value with m, while the RMSE decreases as m increases. In the case of short or long memory, the bias is negative, but in case of negative memory the bias tends to be positive. From the different bandwidth parameters employed here, $m = [n^{0.8}]$ gives rise to the minimum RMSE for all samples sizes and memory parameters considered.

4.2.2 Signal plus noise model

Now, we examine the signal plus noise model (3.2.1). First, we simulate the signal as a standard Gaussian $ARFIMA(0, \frac{\alpha_y}{2}, 0)$ process with $\alpha_y = 0.8$ and the noise as a standard Gaussian $ARFIMA(0, \frac{\alpha_z}{2}, 0)$ process with $\alpha_z = -0.8, -0.4, 0, 0.4$. Moreover, the signal and noise processes are taken to be independent from each other. The Monte-Carlo bias and RMSE are found in Figures 4.11-4.18.

Next, we turn to examine the effect of the signal-to-noise ratio on the finite sample properties of the LW estimator in the signal plus noise model. We simulate the signal as a standard Gaussian $ARFIMA(0, \frac{\alpha_y}{2}, 0)$ process with $\alpha_y = 0.8$, and generate the noise process as an i.i.d. Gaussian sequence with variances $\sigma_z^2 = 0.5, 1, 2$ and independent of the signal process. Notice that the case $\sigma_z^2 = 1$ is already considered above, and hence the results are found in Figures 4.15 and 4.16. The outcomes when $\sigma_z^2 = 0.5, 2$ are presented in Figures 4.19-4.22.

Finally, we consider different types of dependence between the signal and the noise processes and examine the finite sample properties of the LW estimator in the signal plus noise model. Here, we simulate the signal as a standard Gaussian $ARFIMA(0, \frac{\alpha_y}{2}, 0)$ process with $\alpha_y = 0.8$ by truncating the $MA(\infty)$ representation of this model. The innovation term in the latter representation, which is a sequence of i.i.d. standard Gaussian variables, is then multiplied by $\rho = -0.5, 0.5$ and standardized to generate the noise process. Hence, when $\rho = -0.5$ the signal and noise processes are negatively dependent, while when $\rho = 0.5$ they are positively dependent. The Monte-Carlo bias and RMSE of these experiments are presented in Figures 4.23-4.26. Notice that we have also generated the noise as a sequence of i.i.d. standard Gaussian variables independent from the innovation sequence in the $MA(\infty)$ representation of the signal. As expected, the results are very similar to those presented in Figures 4.15 and 4.16 and are therefore not reported.

In all these cases, see Figures 4.19-4.26, the Monte-Carlo bias and RMSE decrease with an increase in n, which supports our finding that the LW estimator is consistent. Overall, the bias increases in absolute value with m, in line with Remark 2.1. Comparing the finite sample bias of the experiments on the signal plus noise model with that from the linear process with $\alpha_x = 0.8$, Figures 4.9 and 4.10, it is evident that the magnitude of the Monte-Carlo bias, and therefore that of the

RMSE, is far bigger in the case of the signal plus noise model. This observation confirms Remark 3.1, noticing that the simulated linear and signal processes have $\beta_x, \beta_y = 2$. On the other hand, the Monte-Carlo RMSE tends to first decrease and then increase as m increases. Contrary to the linear model, there is no unique m across the different samples sizes and memory parameters that gives rise to the minimum RMSE. The bandwidth parameters corresponding to minimum RMSE vary from $[n^{0.525}]$ to $[n^{0.775}]$, and hence are smaller than in the case of a linear process. The latter comments on the bandwidth parameter are in line with Remark 3.2, noticing that the signal process in these experiments has $\beta_y = 2$.

Concentrating on Figures 4.11-4.18, it is clear that Monte-Carlo bias is smaller the bigger the difference $\alpha_y - \alpha_z$ is, except in the case $\alpha_z = 0.4$ and $n = 128, 512$. Furthermore, the finite sample bias converges faster to 0 the larger the difference $\alpha_y - \alpha_z$ is. This observations is in line with Remark 3.1, recalling that the signal process simulated here has $\beta_y = 2$. Also, it is worth noticing that for $n = 2048$ and $\alpha_z = -0.8, -0.4, 0$, the bandwidth parameter minimizing the RMSE increases when $\alpha_y - \alpha_z$ increases. The latter comment confirm our findings in Remark 3.2.

It is evident from Figures 4.15, 4.19 and 4.21 that for all n and m, the higher the magnitude of the signal-to-noise ratio the smaller is the Monte-Carlo bias and therefore the RMSE. The latter observation is in line with Remark 3.3. It is also worth mentioning that the higher the value of the signal-to-noise ratio, the bigger is the bandwidth parameter which minimizes the RMSE, see Figures 4.16, 4.20 and 4.22.

The different types of dependences between the signal and noise processes give rise to different finite sample properties for the LW estimator. In particular, when the two processes are negatively dependent the Monte-Carlo bias is overall positive, see Figure 4.23, while when the two processes are uncorrelated or positively dependent the bias is negative, see Figures 4.15 and 4.25. We should add that in the case that the signal and noise processes are negatively dependent, the bias does not seem to decrease with an increase in n, see Figure 4.23. However, we performed for this particular model Monte-Carlo simulations with higher sample sizes, and the finite sample bias does decrease with increase in n. It is likely that the result of Figure 4.23 is due to small sample properties. It is also worth noticing, that the finite sample bias is smaller when the signal and noise processes are independent from each other, compare Figure 4.15 with Figures 4.23 and 4.25. These observations supports our findings in Remark 3.3.

4.2.3 Structural model

The structural model (3.3.1) is the next specification to be considered. We simulate the trend process $\{\mu_{t,x}\}_{t\in\mathbb{Z}}$ as a standard Gaussian $ARFIMA(0, \frac{\alpha_{\mu_x}}{2}, 0)$ with $\alpha_{\mu_x} = 0.4$ and the cyclical process as a standard Gaussian $GARMA(0, \frac{\alpha_{\omega,c_x}}{2}, 0)$ with $\omega_x = 0.15$ and $\alpha_{\omega,c_x} = 0.1, 0.3$ by truncating the $MA(\infty)$ representation of these models. The two processes $\{\mu_{t,x}\}_{t\in\mathbb{Z}}$ and $\{c_{t,x}\}_{t\in\mathbb{Z}}$ are taken to be independent from each other, while for simplicity we do not include the noise process $\{\eta_{t,x}\}_{t\in\mathbb{Z}}$. Notice that we chose ω_x to be small in order to examine Remark 3.4, and we set the particular value according to an empirical observation in Subsection 4.3.2. The results are given in Figures 4.27-4.30.

In theory, one should expect to see results similar to those for a signal plus noise model with $\alpha_y - \alpha_z = 0.4$, since the cyclical component has $\alpha_{c_x} = 0$. However, comparing Figures 4.27 and 4.29 with Figure 4.17, it is clear that the latter is not the case. This rather peculiar behaviour of the finite sample bias of the LW estimator is down to the relatively small frequency ω_x of the cyclical component $\{c_{t,x}\}_{t\in\mathbb{Z}}$. We have that ω_x corresponds approximately to the Fourier frequency λ_3 when $n = 128$, to the Fourier frequency λ_{12} when $n = 512$, and to the Fourier frequency λ_{48} when $n = 2048$. On the other hand, the minimum value for the bandwidth parameters considered here is 11 when $n = 128$, 22 when $n = 512$, and 45 when $n = 2048$. Hence, for these sample sizes, one cannot consider having $\alpha_{c_x} = 0$, since information from the cyclical component is included in the LW estimation. Instead, the cyclical component should be regarded as if it were a trend component with $\alpha_{c_x} = 2\alpha_{\omega,c_x}$, taking into account Remark 3.4 and that $\{c_{t,x}\}_{t\in\mathbb{Z}}$ follows a $GARMA(0, \frac{\alpha_{\omega,c_x}}{2}, 0)$ model.

When $\alpha_{\omega,c_x} = 0.1$, the cyclical component behaves as if it were a noise process with $\alpha_{c_x} = 0.2$. The finite sample bias in this case, see Figure 4.27, decreases slowly in absolute terms with n. However, this is not the case for small and moderate values of m. We believe that this is a small sample problem and it is driven by the small difference between $\alpha_{\mu_x} = 0.4$ and $\alpha_{c_x} = 0.2$. Overall, the finite sample bias decreases in absolute terms with m, resulting to a RMSE that decreases with m. Notice that $m = [n^{0.8}]$ gives rise to the minimum RMSE, see Figure 4.28.

On the other hand, when $\alpha_{\omega,c_x} = 0.3$, the cyclical component behaves as if it were a trend component with $\alpha_{c_x} = 0.6$. Hence, the roles of $\{\mu_{t,x}\}_{t\in\mathbb{Z}}$ and $\{c_{t,x}\}_{t\in\mathbb{Z}}$ are reversed, and $\{c_{t,x}\}_{t\in\mathbb{Z}}$ becomes the signal process, while $\{\mu_{t,x}\}_{t\in\mathbb{Z}}$ is now the noise. Notice that the bigger the bandwidth parameter, the more information from the cyclical component is included in the LW estimation. Hence, for large values of

m, the finite sample bias is positive, as the LW estimator is estimating $\alpha_{c_x} = 0.6$ instead of $\alpha_{\mu_x} = 0.4$. The latter is particularly apparent for the smaller sample size, see Figure 4.29. The Monte-Carlo bias is negative for the low values of the bandwidth parameter, increases with increases in m, and is positive for moderate and large values of m. The finite sample bias in this case decreases in absolute value with n for large values of m. However, this is not the case for small and moderate values of m, although we believe this is due to the small frequency of the strong cyclical component. Overall, the RMSE decreases with increases in m, and the bandwidths $[n^{0.725}]$ and $[n^{0.8}]$ result to the minimum RMSE, see Figure 4.30.

We should add that the rather peculiar behaviour of the LW estimator in this experiment is in line with Remark 3.4. This behaviour is driven by the choice of the frequency ω_x and the relatively small sample sizes, and it is not going to persist for bigger samples. Essentially, if the sample size or the frequency ω_x is big enough so that $\lambda_m \ll \omega_x$, the finite sample behaviour of the LW estimator in the structural model is comparable to that in a signal plus noise model with $\alpha_z = 0$.

4.2.4 Nonlinear functions of a Gaussian process

We now examine the finite sample behaviour of the LW estimator in the case of nonlinear transformations of a Gaussian process (3.4.1). We consider two transformations, the exponential and the squared ones. The process $\{\xi_t\}_{t\in\mathbb{Z}}$ is generated as a standard Gaussian $ARFIMA(0, \frac{\alpha_\xi}{2}, 0)$. In the case of the exponential transformation, we chose $\alpha_\xi = 0, 0.4, 0.8$, and according to Proposition 3.4 we have $\alpha_x = 0, 0.4, 0.8$, respectively, since the Hermite rank is equal to 1. In the case of the squared transformation, we set $\alpha_\xi = 0, 0.3, 0.7, 0.9$ resulting to $\alpha_x = 0, 0, 0.4, 0.8$, respectively, by Proposition 3.4 and since the Hermite rank is equal to 2. The results of the Monte-Carlo experiments are found in Figures 4.31-4.44.

It is evident that, in all but one case, the finite sample bias decreases with an increase in n, while in all cases the RMSE decreases with an increase in n. Although the bias in the case of the squared transformation with $\alpha_\xi = 0.3$ does not seem to decrease with an increase in n, we repeated this experiment with even higher sample sizes and the bias does decreases with an increase in n, suggesting that the outcome of Figure 4.39 is probably due to small samples. Hence, the results confirm that the LW estimator is consistent for these models. Overall, the Monte-Carlo bias and RMSE tend to decrease with increase in m. As far as the choice of m is concerned, the bandwidth parameter that minimizes the RMSE in the case of short memory is $[n^{0.8}]$, the same as for the linear process, while in the long memory case it varies from $[n^{0.75}]$ to $[n^{0.8}]$. The latter finding only partially confirms our Remark 3.8.

It is worth noticing that for both transformations and when $\alpha_\xi = 0$, the finite sample bias and RMSE, see Figures 4.31, 4.32, 4.37 and 4.38, are almost the same as to those of a linear process with $\alpha_x = 0$, see Figures 4.5 and 4.6. In the case of the squared transformation and $\alpha_\xi = 0.3$, the finite sample bias has clearly a different behaviour, but its magnitude and that of the RMSE, see Figures 4.39 and 4.40, are very similar to those of a linear process with $\alpha_x = 0$, see Figures 4.5 and 4.6. However, when the nonlinear transformations exhibit long memory with parameters $\alpha_x = 0.4, 0.8$, the finite sample behaviour of the LW estimator, see Figures 4.33-4.36 and 4.41-4.44, is far worse compared to the linear case with $\alpha_x = 0.4, 0.8$, see Figures 4.7-4.10. Hence, the LW estimator performs better under short memory than under long memory, and the nonlinearity worsens its finite sample performance in the case of long memory. These comments are in line with the Remarks 2.2 and 3.7. Also, notice that in the case of long memory, the finite sample bias is negative and is smaller in the case of moderate values of the memory parameter, see Figures 4.33, 4.35, 4.41 and 4.43, as predicted by Remarks 3.6 and 3.7. Comparing the finite sample bias of the two nonlinear transformations in the case of long memory, Figures 4.33, 4.35, 4.41 and 4.43, it is evident that the bias is smaller in the case of the squared transformation. This should not come as a surprise, bearing in mind that the squared transformation entails only one Hermite polynomial in its Hermite expansion, while the exponential one involves all Hermite polynomials.

4.2.5 Long memory stochastic volatility model

Lastly, we examine the LMSV model (3.5.1) with volatility process following the commonly used exponential form (3.5.2). We generate $\{\xi_t\}_{t\in\mathbb{Z}}$ as a standard Gaussian $ARFIMA(0, \frac{\alpha_\xi}{2}, 0)$ process with $\alpha_\xi = 0, 0.4, 0.8$, while the process $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is drawn as a sequence of i.i.d. standard Gaussian variables. Furthermore, the process $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is taken to be independent of $\{\xi_t\}_{t\in\mathbb{Z}}$. We consider the absolute, squared and logarithmic squared transformations of the return series resulting to memory parameters $\alpha_x = 0, 0.4, 0.8$, according to the findings of Section 3.5. The Monte-Carlo bias and RMSE are presented in Figures 4.45-4.62.

It is evident that in all the cases, the finite sample bias and RMSE decrease with increase in n, confirming that for these models, the LW estimator is consistent both under short and long memory, although the first was not established in Section 3.5.

In the short memory case, the finite sample bias and RMSE decrease in absolute terms with m, see Figures 4.45-4.46, 4.51-4.52 and 4.57-4.58, while in the long memory case they tend to increase in absolute terms, see Figures 4.47-4.50, 4.53-4.56 and 4.59-4.62. Notice also that in all the cases, the finite sample bias is negative, see Figures 4.47, 4.49, 4.53, 4.55, 4.59 and 4.61. This result is expected in the long memory case, since the LMSV model is essentially a signal plus noise model with the signal and noise processes uncorrelated from each other, see Remark 3.3 and the Monte-Carlo experiments in Subsection 4.2.2. As far as the choice of m is concerned, the bandwidth parameter that minimizes the RMSE in the case of short memory is $[n^{0.8}]$, see Figures 4.46, 4.52 and 4.58, the same as for a linear process, while in the long memory case it varies from $[n^{0.5}]$ to $[n^{0.8}]$, see Figures 4.48, 4.50, 4.54, 4.56, 4.60 and 4.62. The latter finding only partially confirms our Remark 3.8.

It also interesting to notice that the Monte-Carlo bias and RMSE are smaller in the short memory case, see Figures 4.45-4.46, 4.51-4.52 and 4.57-4.58, than in the long memory case, see Figures 4.47-4.50, 4.53-4.56 and 4.59-4.62, as predicted by Remark 2.2. In the long memory case and for the absolute and squared returns, the finite sample behaviour of the LW estimator is better for moderate values of the memory parameter, see Figures 4.47-4.50 and 4.53-4.56, in line with Remark 3.7. Furthermore, in the long memory case, the finite sample bias and RMSE of the absolute returns is smaller than that of the squared returns, see Figures 4.47-4.50 and 4.53-4.56. The latter observations asserts our findings in Remark 3.9 and the succeeding discussion.

Finally, it worth noticing that the logarithmic squared transformation is the most preferable in this experiment, as in the long memory case, it gives rise to smaller finite sample bias and RMSE compared to the absolute and squared transformations, see Figures 4.47-4.50, 4.53-4.56 and 4.59-4.62. Recall that under the LMSV representation employed here, all three transformations of the return series can be written as a signal plus noise model with the signal and noise processes being uncorrelated from each other and the noise process being a sequence of uncorrelated variables. However, there is a crucial difference between the signal plus noise decomposition of the logarithmic squared returns and those of the absolute and squared returns: in the case of the logarithmic squared transformation, the signal process is linear, while in the case of the absolute and squared transformation, the signal process is nonlinear. Actually, in the latter case, the signal process is $\exp(\xi_t)$, which according to Remark 3.6 and the results of the Monte-Carlo experiments of Subsection 4.2.4 induces further negative finite sample bias to the LW estimator.

4.3 Empirical applications

Now, we apply LW estimation to various data sets to assess their degree of long-run persistence. We examine inflation and expected inflation rates, nominal and real interests rates, and various exchange rates and their transformed returns.

4.3.1 Inflation and expected inflation rates

One of the first attempts to quantify the long-run persistence of prices and inflation was taken by Nelson and Plosser (1982). The authors considered U.S. data on consumer prices for the period 1860-1970 and GNP deflator for the period 1889-1970, and examined the long-run persistence of these series by performing the Dickey-Fuller test with a linear trend for the null hypothesis of a unit root against the alternative of a stationary $AR(p)$. Using log-transformation of the data, they failed to reject the null hypothesis of a unit root, that is $\alpha_x = 2$. Gil-Alaña and Robinson (1997) extended the data of Nelson and Plosser (1982) up to 1988, and reexamined their degree of long-run persistence by the means of a Lagrange multiplier type of test introduced by Robinson (1994c). The latter test procedure allows for testing various degrees of long-run persistence, as opposed to the Dickey Fuller method which essentially tests for the null hypothesis $\alpha_x = 2$ against the alternative $\alpha_x = 0$, as well as more general specifications for the data generating mechanism. The results of Gil-Alaña and Robinson (1997) suggested values of α_x ranging from 1 to 4.5 depending on the specification used. In a later study, Backus and Zin (1993) considered U.S. data for inflation measured as the growth rate in the implicit price index for consumption of nondurables and services covering the period 1959-1989. The authors estimated $ARFIMA(p, \frac{\alpha_x}{2}, q)$ models by means of maximum likelihood techniques, and found the estimates of α_x ranging from -0.3 to 0.98 depending on the choice of the orders p, q. The $ARFIMA(p, \frac{\alpha_x}{2}, q)$ model with conditional heteroscedasticity in the error term was also considered by Baillie, Chung, and Tieslau (1996) for U.S. data for inflation measured as the growth rate in the consumer price index (CPI) during the period 1948-1990. The authors propose an estimation procedure for this model based on the maximum likelihood method. Their results suggested that the memory parameter α_x varies in the region $(0.6, 1)$. Phillips (1998) examined the data of Baillie, Chung, and Tieslau (1996) for the period 1934-1997, although constructed three-month inflation instead of one-month. He employed LW estimation with bandwidth parameter $m = [n^{0.75}]$ over various time periods. Depending on the time span, his estimates of α_x varied from 1.06 to 1.396. The data of Phillips (1998) were extended by Sun and Phillips (2004) up to 1999, and the degree of long-run persistence was reexamined using the log-periodogram estimate of Geweke and Porter-Hudak (1983) and the exact Whittle estimator of Shimotsu and Phillips (2005) with bandwidth parameters $m = [n^{0.55}], [n^{0.55}]+1, ..., [n^{0.8}]$. The estimates of α_x varied from 0.6 to 1, depending on the choice of estimator and bandwidth parameter. Sun and Phillips (2004) also considered proxy data on expected inflation rate, and found estimates of α_x much higher than those for the realized inflation rate.

Under the assumption of rational expectations, the inflation rate π_t differs from expected inflation rate π_t^e by an unexpected shock e_t, that is

$$\pi_t = \pi_t^e + e_t, \tag{4.3.1}$$

where e_t is usually taken to follow a martingale difference process. The aforementioned studies clearly suggest that the inflation and expected inflation rate exhibit strong long-run persistence, so that $\alpha_\pi, \alpha_{\pi^e} > 0$. On the other hand, the rational expectations hypothesis indicates that the shock e_t is uncorrelated from π_t^e and is such that $\alpha_e = 0$. Hence, (4.3.1) can be considered as a signal plus noise model, where the signal process is the expected inflation rate π_t^e and the noise process is the unexpected shock e_t. As expected inflation rate π_t^e is unobserved, one can use data on the inflation rate π_t to infer the degree of persistence of the expected inflation rate π_t^e. This line of reasoning was adopted also by Sun and Phillips (2004).

However, as Nelson and Schwert (1977) and Sun and Phillips (2004) among others pointed out, when the variance of the forecasting error e_t is large compared to the variance of the expected inflation rate π_t^e, the degree of persistence of the inflation rate π_t is masked by the short-run variability of its component e_t, so that in finite samples estimates of the memory parameter based on π_t are likely to be smaller than those based on π_t^e. For that reason, Sun and Phillips (2004) adjusted their estimation methods to take into account the structure of (4.3.1). The modified estimators were applied again to the data, and the new estimates of the memory parameters of the realized inflation rate π_t were in line with those from the proxy of the expected inflation rate π_t^e.

The LW estimator is no exception to this effect, as we now discuss. Using quarterly data from the Survey of Professional Forecasters[1] we employ the mean of one-period ahead forecasts on inflation rate as a measure of expected inflation rate π_t^e, as in Sun and Phillips (2004). We compare this series with two series for realized inflation rate π_t calculated as the percent change in the price level between two subsequent quarters based on CPI and GDP deflator data[2]. Figure 4.63 plots the data, all in annual rates, for the period 1981Q4 to 2005Q4, when the data from the Survey of Professional Forecasters are available. It is clear from Figure 4.63 that the three series have the same long-run tendencies, so that one would then expect the estimates of the memory parameter of the three series to be similar. However, it is also evident from Figure 4.63 that the measure of expected inflation rate π_t^e is the least volatile, while the realized inflation rate based on CPI data is the most. Notice that the inflation rate calculated from the CPI data is the growth in prices of a particular basket of goods consumed, while the inflation rate calculated from the GDP deflator data takes into account all goods consumed in the economy. Hence, the CPI inflation rate is more sensitive to changes in the economy or external shocks, and therefore is subject to higher short-run variability. Notice for example, that the spikes in the CPI inflation rate series are mainly due to sharp changes in the price of oil, an effect that is not transmitted so dramatically in the inflation rate series based on GDP deflator data. On the other hand, the proxy of expected inflation is the mean value of forecasts provided by several practitioners, resulting to a rather smooth series having moderate short-run variation.

Thus, taking into account Remark 3.3 and the Monte-Carlo experiment of Subsection 4.2.2, we expect the LW estimates of the memory parameter of the proxy of the expected inflation rate to be the higher, while those of the CPI inflation rate to be the lower. Figure 4.64 plots the LW estimators of the memory parameters of the three series across the different bandwidth parameters, where we employ all bandwidths ranging from $[n^{0.5}]$ to $[n^{0.8}]$. Indeed, as Figure 4.64 shows, the estimates of the memory parameters of the three series are rather different, with those of expected inflation rate being the highest and those of the CPI inflation rate being the lowest.

It is clear from this discussion, that assessing the true degree of long-run persistence of an inflation rate series is a rather difficult exercise, given the small sample size available. One might think that the data from the Survey of Professional Fore-

[1] The Survey of Professional Forecasters is found at http://www.phil.frb.org/econ/spf/.

[2] CPI data are taken from U.S. Department of Labor: Bureau of Labor Statistics, Series I.D. CPIAUCSL, http://stats.bls.gov. GDP deflator data are taken from U.S. Department of Commerce: Bureau of Economic Analysis, Series I.D. GDPDEF, http://www.bea.gov.

casters should be used for such purposes, recalling that we employed one-period ahead forecasts on inflation rate as in Sun and Phillips (2004). However, we have also applied LW estimation to current-period and two-period ahead forecasts of inflation rate from the Survey of Professional Forecasters, and the LW estimates are lower than those from the one-period ahead forecasts since the two series have slighter higher short-run variability. Actually, the new estimates are closer to the LW estimates from the inflation based on GDP deflator data.

4.3.2 Nominal and real interest rates

The long-run properties of the short-term nominal and real interest rates have received a great deal of attention. Initially, Fama (1975) employed techniques based on the sample autocorrelation function, and inferred that the real interest rate remains approximately constant over the period 1953-1971. In a later study, covering the longer periods 1953-1979 and 1931-1952, Mishkin (1981) used the same methodology as in Fama (1975), and rejected the constancy in the real interest rate. Subsequent investigations by Rose (1988) provided evidence in support of unit roots in the series of the nominal and real interest rates over various time periods spanning from 1892 to 1986. However, as the critical values of the Dickey-Fuller test employed by Rose (1988) had often been found misleading in finite samples, see for example Schwert (1987), Mishkin (1992) performed a similar analysis using critical values from Monte-Carlo simulations. For data spanning from 1953 to 1990 and for different subperiods, Mishkin (1992) inferred that the nominal interest rate is a unit root process, while the real interest rate is short memory process in all but the post-1982 period where he found evidence supporting the existence of a unit root. More recently, Phillips (1998) concluded that the nominal interest rate has a higher memory parameter than the inflation and real interest rates, for different subperiods over the 1934-1997 period. The outcome, which was clearly at odds with the Fisher equation, which states that the real interest rate is the difference of the nominal interest rate and inflation, was reexamined by Sun and Phillips (2004). With the availability of data on inflation forecasts from the Survey of Professional Forecasters, they found that the memory of the real interest rate had been underestimated, and fell in the nonstationary region. To eliminate this negative bias, they proposed a modified version of the estimators employed, see the discussion in Subsection 4.3.1, and furthermore suggested a testing procedure for the null hypothesis of equal memory parameters. Using data for the period 1934-1999, they inferred that the three variables in the Fisher equation have the same memory parameter, which was found to fall in the nonstationary region.

In a parallel line of research, Garcia and Perron (1996) reanalyzed the data of Rose (1988) over the period 1961-1986, using regime shift techniques, and found that the real interest is constant subject to regime shifts dating 1973 and 1979. Using regression techniques and tests for parameter stability, Huizinga and Mishkin (1986) reported breaks in the process of the real interest rate, where the dates of the breaks were found to be 1979 and 1982. On the other hand, the results of Clarida, Gali, and Gertler (2000) suggest that the relationship of the nominal interest rate and inflation shifted in 1979, while that of the nominal interest rate and output gap in 1987. The latter authors employed the Taylor (1993) rule for their analysis, which has been found to do a fairly accurate job in describing the generating mechanism of the short-term nominal interest rate.

Here, we reanalyze the long-run persistence of the nominal and real interest rates using the LW estimator and taking into account the results of Clarida, Gali, and Gertler (2000). The ex ante real interest rate r_t^e at time t on a given security maturing at time $t+1$ is defined from the Fisher (1930) equation

$$i_t = \pi_t^e + r_t^e, \tag{4.3.2}$$

where i_t is the nominal interest rate on the given security issued at period t, maturing in period $t+1$, and π_t^e is the expected inflation rate between periods t and $t+1$. The empirical analysis of the ex ante real rate of interest r_t^e is complicated by the factor that r_t^e is not directly measurable as the expected inflation rate π_t^e is not observed. One way to overcome this problem is to obtain proxies for the expected inflation rate from some survey, e.g. from the Survey of Professional Forecasters as in Sun and Phillips (2004).

The most common approach is to use data on the realized inflation rate π_t and calculate the ex post real interest rate r_t from

$$i_t = \pi_t + r_t. \tag{4.3.3}$$

As discussed in Subsection 4.3.1, under rational expectations, the forecasting error $e_t = \pi_t - \pi_t^e$ is a martingale difference. In such a case, the ex post and ex ante real interest rates differ by martingale difference process, so that r_t^e and r_t have the same memory parameter. That is, the ex post and ex ante real interest rates have the same long-run properties and one can make inference on the long-term behaviour of the unobserved r_t^e using data on the observed r_t. However, as in the case of inflation and expected inflation rates, when the variance of the forecasting error e_t is large compared to the variance of the ex ante real interest rate r_t^e, the degree of persistence of the ex post real rate of interest r_t is masked by the short-run variability of its component e_t. Therefore, when calculating the ex post real interest rate r_t we use the inflation rate based on the GDP deflator data. For reasons to become clear below, we study the long-run behaviour of the real interest rate mainly in the post-October 1987 period. For this sample period, given the availability of data from the Survey of Professional Forecasters, we present estimation results for the ex ante real interest rate r_t^e using the expected inflation rate π_t^e as calculated in Subsection 4.3.1 above.

The data are quarterly time series spanning the period 1954Q3-2005Q4. We use the average Federal Funds rate in the first month of each quarter, expressed in annual rates, as the nominal interest rate i_t[3]. Figure 4.65 below plots the data for the nominal interest rate i_t and inflation rate π_t along with the resulting ex post real interest rate r_t. The vertical lines in Figure 4.65 stand for the dates 1979Q3 and 1987Q3 when Paul Volcker and Alan Greenspan were appointed Chairman of the Board of Governors of the Federal Reserve System. The sample is split into three subperiods: the pre-Volcker period (1954Q3-1979Q2), Volcker period (1979Q3-1987Q2) and post-Volcker period (1987Q3-2005Q4). Notice that the results of Clarida, Gali, and Gertler (2000) suggest that the relationship of the nominal interest rate i_t and inflation rate π_t shifted in 1979Q3, while that of the nominal interest rate i_t and output gap g_t in 1987Q3. Figures 4.67, 4.69 and 4.71 plot the nominal interest rate i_t, inflation rate π_t and ex post real interest rate r_t for these three periods, while Figure 4.73 plots the nominal interest rate i_t, expected inflation rate π_t^e and ex ante real interest rate r_t^e for the post-Volcker period.

Next, we apply LW estimation for the full sample and for the three subsamples. The resulting estimates are plotted in Figures 4.66, 4.68, 4.70, 4.72 and 4.74 against the different bandwidth parameters, where we employ all bandwidths ranging from $[n^{0.5}]$ to $[n^{0.8}]$. For the full sample and the first two subsamples, see Figures 4.66, 4.68 and 4.70, the estimates of the memory parameter of i_t and π_t are of similar magnitude and are both higher than that of r_t. On the contrary, for the last subperiod, see Figure 4.72, the estimates of the memory parameter of π_t is lower than that of both r_t and i_t. Also, notice that in the first two subperiods the estimates of the memory parameter of r_t fall in the stationary region, while for the last subperiod in the nonstationary region. Similar observations can be made for the ex ante real

[3]Federal Funds rate data are taken from Board of Governors of the Federal Reserve System, Series I.D. FEDFUNDS, http://www.federalreserve.gov.

interest rate r_t^e in the last subperiod, see Figure 4.74. Overall, the estimates of the memory parameter of i_t tend to exceed that of both π_t and π_t^e.

It is rather surprising that for the post-Volcker period the estimates of the memory parameter of i_t, r_t and r_t^e exceed 2, see Figures 4.72 and 4.74. A value of the memory parameter greater than 2 would impose no upper bound on the path of these processes. However, it is clear from Figures 4.71 and 4.73 that the series i_t, r_t and r_t^e do not have an explosive behaviour, so that our estimates of their memory parameters are far from realistic. In addition, it is reasonable to assume that any shocks to these processes eventually dissipate, so that one would anticipate estimates below 2. Notice also from Figures 4.71 and 4.73 that the nominal interest rate i_t seems to exhibit similar long-run tendencies with the inflation rate π_t and the expected inflation rate π_t^e . Then, we would expect estimates of the memory parameter for the nominal interest rate i_t to be on the same level as those for the inflation rate π_t and the expected inflation rate π_t^e, or lower if we take into account that i_t looks more volatile than π_t^e. However, the estimates of the memory parameter are overall higher for i_t than those for π_t and π_t^e, raising doubts as to whether i_t can be cointegrated with π_t and π_t^e.

Here, we seek to provide an explanation for the instability of the estimates of the memory parameters of the real interest rate over the three different sample periods. We start by considering a monetary policy reaction function in the lines of Taylor (1993) and Clarida, Gali, and Gertler (2000). According to the principles of the Taylor rule, the level of the nominal interest rate is a linear function of the gaps between expected inflation and output and their perspective target levels. Therefore, we can write

$$i_t = i^* + \beta(\pi_t^e - \pi^*) + \gamma g_t + \eta_t, \tag{4.3.4}$$

where i^* and π^* are the target levels of the nominal interest rate and inflation, respectively, g_t is the output gap and η_t is the error term capturing the short-run dynamics of the nominal interest rate i_t. Hence, the ex ante real interest rate r_t^e is determined by

$$r_t^e = a + (\beta - 1)\pi_t^e + \gamma g_t + \eta_t, \tag{4.3.5}$$

where $a = i^* - \beta\pi^*$, while the ex post real interest rate r_t is given by

$$r_t = a + (\beta - 1)\pi_t + \gamma g_t + \varsigma_t, \tag{4.3.6}$$

where $\varsigma_t = \eta_t - \beta e_t$. The results of Clarida, Gali, and Gertler (2000) suggest that for the pre-Volcker period $\beta = 1$ and $\gamma = 0$, for the Volcker period $\beta > 1$ and $\gamma = 0$, while for the post-Volcker period $\beta > 1$ and $\gamma > 0$.

Hence, for the pre-Volcker period, (4.3.4) is essentially a signal plus noise type of model, and the nominal interest rate i_t and expected inflation rate π_t^e should have a common long-run component, which is not transmitted however in the ex post real interest rate r_t. The plot of the data, Figure 4.67, clearly support this statement. On the other hand, the LW estimates for the nominal interest rate i_t and inflation rate π_t, see Figure 4.68, are not identical, but the LW estimates for the ex post real interest rate r_t are lower than those for the nominal interest rate i_t and inflation rate π_t. The discrepancy between the estimated memory parameters of the nominal interest rate i_t and inflation rate π_t is likely to be attributed to the higher short-run variability of the inflation rate π_t as opposed to that of the nominal interest rate i_t.

For the Volcker period, the results of Clarida, Gali, and Gertler (2000) imply that both the nominal i_t and ex post real r_t interest rates should inherit the long-run component of the inflation rate π_t. The plot of the data, Figure 4.69, support this argument, but the LW estimates of the three series do not. However, given the very small sample size for this period, one can hardly make any definite statements about the magnitudes of the estimated memory parameters. We can say at least, that the order of the estimated memory parameters is in line with our Monte-Carlo experiments on the signal plus noise. Notice that the nominal interest rate i_t reacts more than one-to-one to expected inflation rate π_t^e, the inflation rate π_t reacts one-to-one to expected inflation rate π_t^e, while the ex post real interest rate r_t reacts less than one-to-one to expected inflation rate π_t^e. On the other hand, Figure 4.69 suggests that the short-run variability of the ex post real interest rate r_t is the highest, followed by that of the nominal interest rate i_t and inflation rate π_t. The latter observations imply that the signal-to-noise ratio is the lowest for the ex post real interest rate r_t, and according to our findings and Monte-Carlo experiments on the signal plus noise model, we should see the LW estimates for the ex post real interest rate r_t to be the lowest ones.

Finally, for the post-Volcker period, the results of Clarida, Gali, and Gertler (2000) imply that both the nominal i_t and ex post r_t (ex ante r_t^e) real interest rates should inherit the long-run component of the (expected π_t^e) inflation rate π_t, while their cyclical component is driven by output gap g_t. The plot of the nominal interest rate i_t, expected inflation rate π_t^e, ex ante real interest rate r_t^e and output gap g_t over the period 1987Q3-2005Q4, see Figure 4.75, supports this statement. Here, we have calculated output gap g_t as the percent deviation between actual GDP and its corresponding target[4]. Hence, (4.3.4) and (4.3.5) can be considered as types of

[4] GDP data are taken from U.S. Department of Commerce: Bureau of Economic Analysis,

structural models (3.3.1) with $\mu_i = \mu_{r^e} = \mu_{\pi^e}$ and $c_i = c_{r^e} = c_g$. One would then expect the memory parameters of the nominal interest rate i_t, expected inflation rate π_t^e and ex ante real interest rate r_t^e to be similar. However, as we saw already above, the LW estimates of the memory parameters of the nominal i_t and ex ante real r_t^e interest rates are unrealistically high and rather different from those from expected inflation rate π_t^e.

Remark 3.4 and the Monte-Carlo experiments on the structural model (3.3.1) can easily explain this discrepancy. First, notice that the shaded areas in Figure 4.75 correspond to recessions according to the NBER business cycle chronology. The duration of the only business cycle recorded in this period is 128 months, calculated either from peak to peak or from trough to trough. This duration corresponds to a period of approximately 42 quarters, giving a frequency $\omega = 0.15$, which corresponds approximately to just the second Fourier frequency for this particular sample. Secondly, notice in Figure 4.75 that most of the variability of the nominal i_t and ex ante real r_t^e interest rates is explained by the cyclical component rather than the long-run one, although this not the case for the expected inflation rate π_t^e. The latter observations are further supported by the plot of the sample autocorrelation function and periodogram of the nominal interest rate i_t, expected inflation rate π_t^e, ex ante real interest rate r_t^e and output gap g_t, see Figures 4.77 and 4.78. The sample autocorrelation function of the nominal interest rate i_t, ex ante real interest rate r_t^e and output gap g_t have a very similar cyclical pattern of approximate period 42, i.e. the period of the business cycle, which is not observed in the sample autocorrelation function of the expected inflation rate π_t^e. It is also worth noticing that the periodogram of nominal interest rate i_t, ex ante real interest rate r_t^e and output gap g_t peak at the same frequency 0.19, which is very close to the frequency $\omega = 0$ of the business cycle, contrary to the periodogram of the expected inflation rate π_t^e which peaks at the first frequency.

To summarize, in the post-Volcker period, the nominal i_t and ex ante real r_t^e interest rates exhibit a cyclical component of approximate frequency 0.15, and this cyclical component is stronger than the trend one. Taking into account that the frequency of the business cycle corresponds approximately to just the second Fourier frequency used in the LW estimation of the memory parameter, Remark 3.4 and the Monte-Carlo experiments on the structural model (3.3.1) suggest that the LW estimates are likely to be subject to substantial positive bias driven by the strong

Series I.D. GDPC96, http://www.bea.doc.gov. Potential GDP data are taken from U.S. Congress: Congressional Budget Office, Series I.D. GDPPOT, http://www.cbo.gov.

cyclical component. Furthermore, as output gap g_t seems to be the source of the cyclical component, one would expect LW estimation on this variable to retrieve estimates close to those of the nominal i_t and ex ante real r_t^e interest rates. Figure 4.76 plots the LW estimates of the nominal interest rate i_t, expected inflation rate π_t^e, ex ante real interest rate r_t^e and output gap g_t for this period. Indeed, the LW estimates of output gap g_t are very similar to those of the nominal i_t and ex ante real r_t^e interest rates.

4.3.3 Exchange rates

A common finding in the empirical literature of asset returns is that the returns r_t themselves are martingale differences or at least short memory processes, but absolute returns $|r_t|$, power transformations $|r_t|^p$ for $p > 0$, and logarithmic squared transformation $\log r_t^2$ exhibit long memory. Recall, that as was discussed in Chapter 3, a model capable of generating such behaviour is the LMSV model of Breidt, Crato, and de Lima (1998) and Harvey (1998). For various stock returns, Ding, Granger, and Engle (1993) and Ding and Granger (1996) established that, among the various power transformations, this long memory property is strongest when $p = 1$. Ding and Granger (1996) also considered the foreign exchange rate returns for the Deutschmark with the US dollar and found instead that for power transformation $p = \frac{1}{4}$ this property is strongest. Here, we consider the foreign exchange rates for the UK pound with the US dollar (UK£/US$) and for the Japanese yen with the US dollar (JP¥/US$).

The data are monthly and span the period 1971M1-2006M5[5]. We construct the returns r_t as the difference of the logarithmic exchange rate between period t and $t-1$, and examine the absolute returns $|r_t|$, the squared returns r_t^2, the quartered returns $|r_t|^{\frac{1}{4}}$, and the logarithmic squared returns $\log r_t^2$. The data are plotted in Figures 4.79-4.83 for the UK£/US$ exchange rate, and in Figures 4.85-4.89 for the JP¥/US$ exchange rate. We employ all bandwidths ranging from $[n^{0.5}]$ to $[n^{0.8}]$ and apply LW estimation on the absolute returns $|r_t|$, the squared returns r_t^2, the quartered returns $|r_t|^{\frac{1}{4}}$, and the logarithmic squared returns $\log r_t^2$ for both exchange rate series. The results are presented in Figure 4.84 for the UK£/US$ exchange rate, and in Figure 4.90 for the JP¥/US$ exchange rate. The observation of Ding and Granger (1996) that the exchange rate of the Deutschmark with the US dollar exhibit strongest long-run persistence when $p = \frac{1}{4}$ across different p-

[5] UK£/US$ and JP¥/US$ exchange rate data are taken from Board of Governors of the Federal Reserve System, Series I.D. EXUSUK and EXJPUS, respectively, http://www.federalreserve.gov.

th power transformations holds clearly for JP¥/US$ exchange rate, and partially for the UK£/US$ exchange rate. The estimated memory parameters are highest when $p = \frac{1}{4}$ compared to $p = 1, 2$ for all bandwidths parameters in the case of the JP¥/US$ exchange rate, see Figure 4.90, and for moderate and big values of the bandwidth parameter in the case of the UK£/US$ exchange rate, see Figure 4.84. Notice also, that for moderate and big values of bandwidths parameter, the estimated memory parameters of the logarithmic squared returns tend to exceed those of the p-th power transformations, particularly in the case of the JP¥/US$ exchange rate.

The results here cannot directly support the use of the LMSV model (3.5.1) of Breidt, Crato, and de Lima (1998) and Harvey (1998) with the exponential specification of the volatility process. But as the LW estimates of the various transformations of the returns of these series are in line with our theoretical findings of Section 3.5 and the corresponding Monte-Carlo simulations of Subsection 4.2.5, one cannot exclude this specification for the UK£/US$ and JP¥/US$ exchange rates. Furthermore, one cannot say with certainty that the memory parameters of the different transformations of the returns series are distinct. It could be the case that the differences in the LW estimates across the various transformation are driven by differences in the magnitude of the finite sample bias of the LW estimator.

4.4 Final comments

In this chapter, we have analyzed by the means of Monte-Carlo simulations, the finite sample properties of the LW estimator for the linear and nonlinear specifications discussed in Chapters 2 and 3. The results of the Monte-Carlo experiments have confirmed the theoretical findings and remarks of Chapters 2 and 3. The LW estimator is consistent for the nonlinear models considered, but the nonlinearity worsens its finite sample performance as compared to the linear model. Hence, in the presence of nonlinearity, a larger sample is required to achieve the same level of accuracy as in the linear case. The outcome of the Monte-Carlo simulations has also suggested that for nonlinear models, the bandwidth parameter that minimizes the RMSE is, compared to that for the linear case, expected to be of moderate value, possibly in the range $[n^{0.5}] - [n^{0.75}]$. However, we should notice that smaller values of the bandwidth parameter are likely to be required if short-run dynamics are further included in the simulated processes.

We have applied the LW estimation to inflation and expected inflation rates, nominal and real interest rates, and transformations of the returns of two foreign exchange rate series. Overall, we have found that estimating the true degree of long-run persistence can be a difficult task given the small sample sizes and presence of nonlinearity. However, there are some points that can be inferred with certainty from this empirical exercise. Firstly, one should not employ data on the CPI series in order to assess the degree of long-run persistence of the inflation series. Instead, one should use data on the GDP deflator series, or employ some proxy for the expected inflation rate. In general, caution has to be exercised when the data are subject to substantial short-run variability. The latter variability masks the true degree of long-run persistence, leading to underestimates of the memory parameter. Secondly, in data where the cyclical component is stronger than the trend and has a period that is big compared to the sample size, one might want to refrain from estimating the degree of long-run persistence. The long-run component is masked by the cyclical one, and the LW estimates are driven by the strong cyclical behaviour. The latter has been found for the U.S. nominal and real interest rate over the post-1987 sample period, but we believe it is likely to be manifested in other data sets. Finally, one cannot say with certainty if the property that a certain p-th power transformation of return series exhibits the highest degree of long-run dependence, is a stylized fact of return series or just a small sample behaviour of the LW estimator.

The Monte-Carlo experiments and empirical applications presented here indicate the need of an estimation procedure for the memory parameter which is subject to smaller finite sample bias. Certainly, it would be preferable if such a procedure did not increase the dispersion, at least not substantially, and did not need to incorporate the different nonlinear structures. The issue of bandwidth choice remains an open problem, and is the main reason for not presenting confidence intervals in our empirical applications. Clearly, a bias reduction method and a general rule for bandwidth choice are important matters that we plan to examine in future research.

A further issue that we hope to address in the future, is the parametric and/or semiparametric modelling of persistence of business cycle behaviour. Currently, it is common amongst practitioner to employ the Markov switching model of Hamilton (1990) for modelling business cycle behaviour, see Diebold and Rudebusch (1999) and the references therein. However, the latter model does not allow for smooth transition between the different states of the economy, and furthermore it is not clear how persistence is quantified in this setup. On the other hand, models (1.2.9) and (1.2.10) can be used to quantify business cycle persistence. However, these specifications entail the assumption that the business cycle is periodic and symmetric, which is clearly violated as business cycles have variable duration and recoveries tend to last longer than recessions. Perhaps, a combination of these models might give an answer to this problem.

4.A Appendix

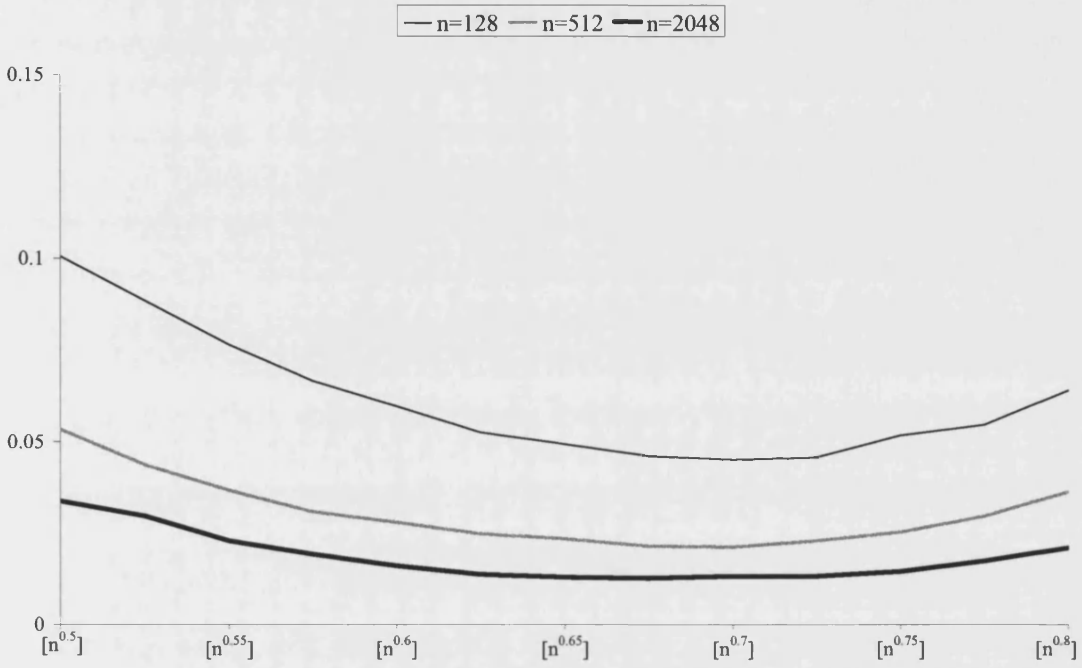


Figure 4.1: Bias of LW estimator; linear process with $\alpha_x = -0.8$.

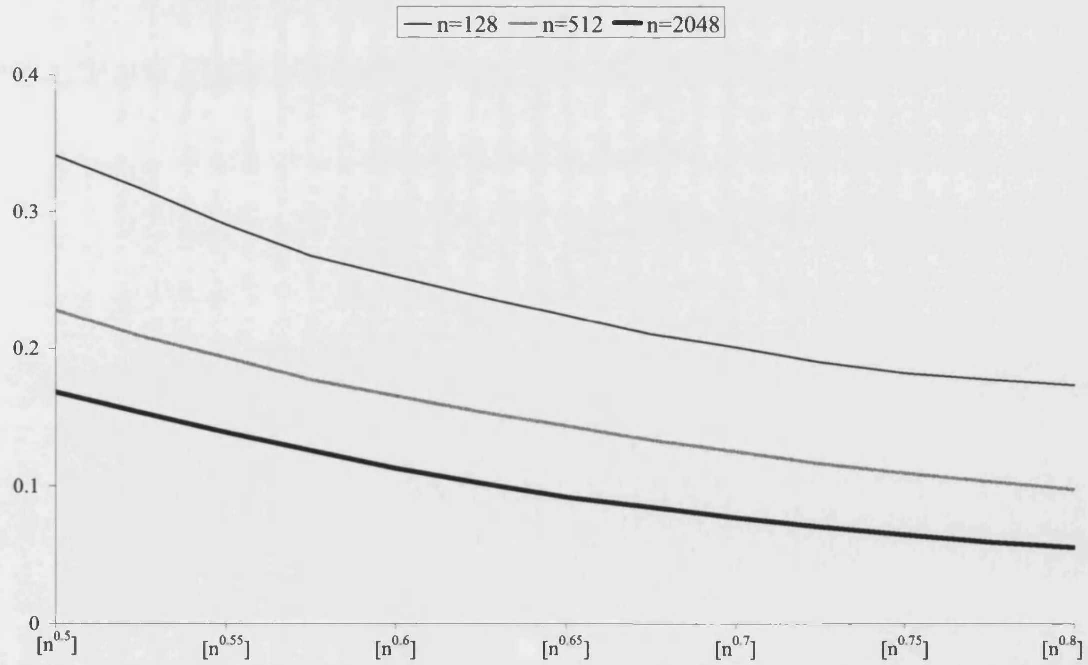


Figure 4.2: RMSE of LW estimator; linear process with $\alpha_x = -0.8$.

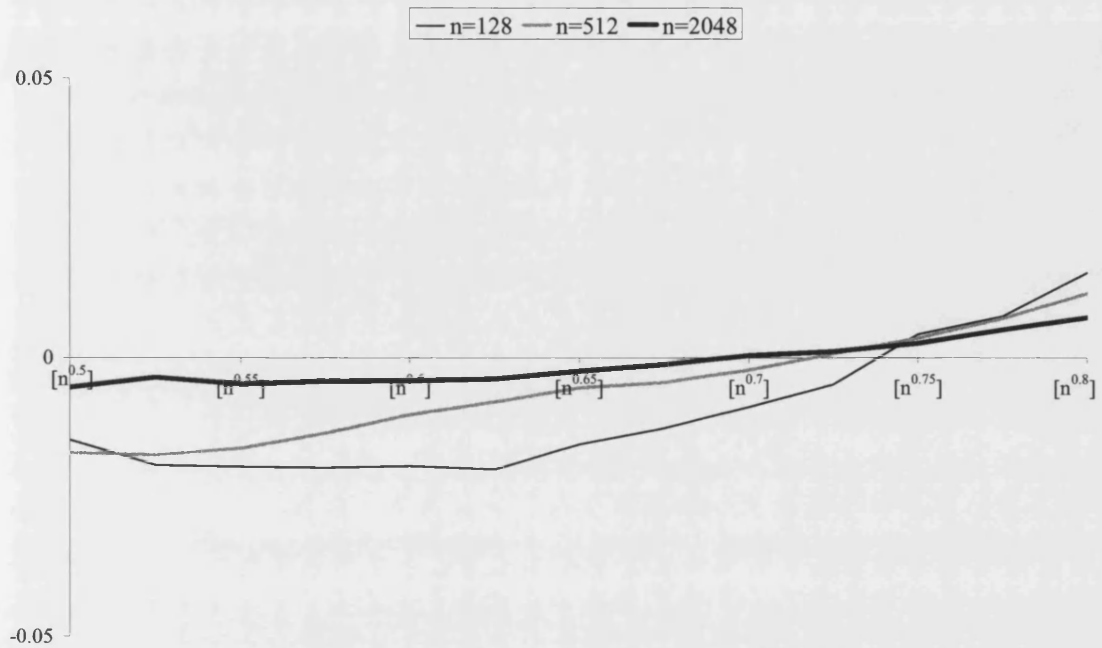


Figure 4.3: Bias of LW estimator; linear process with $\alpha_x = -0.4$.

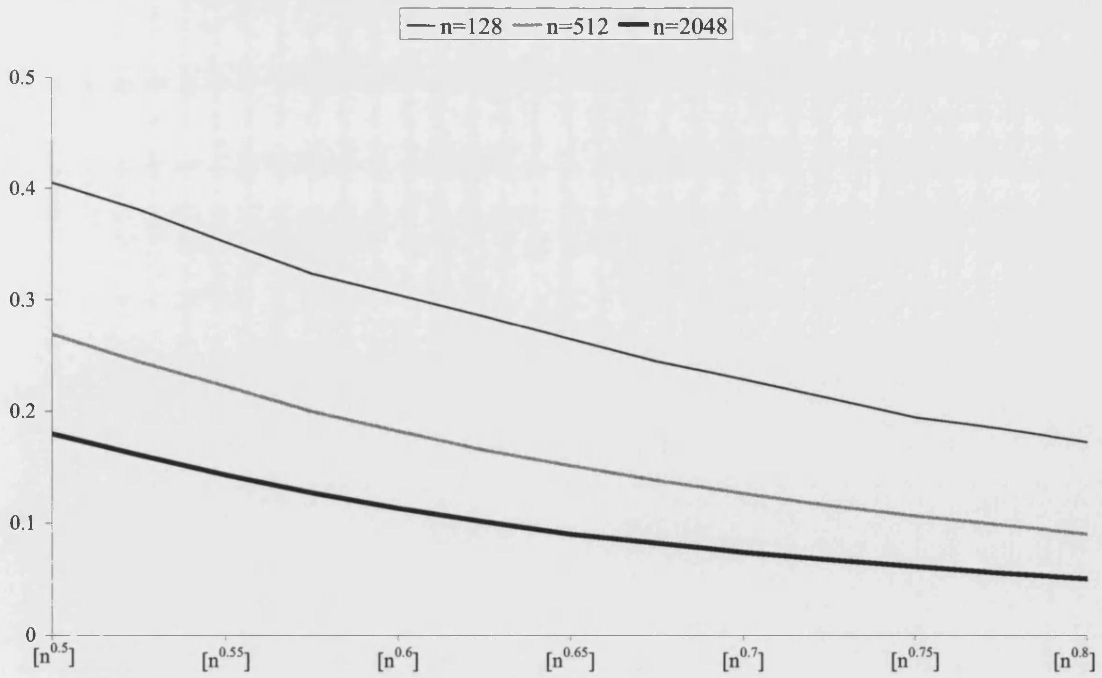


Figure 4.4: RMSE of LW estimator; linear process with $\alpha_x = -0.4$.

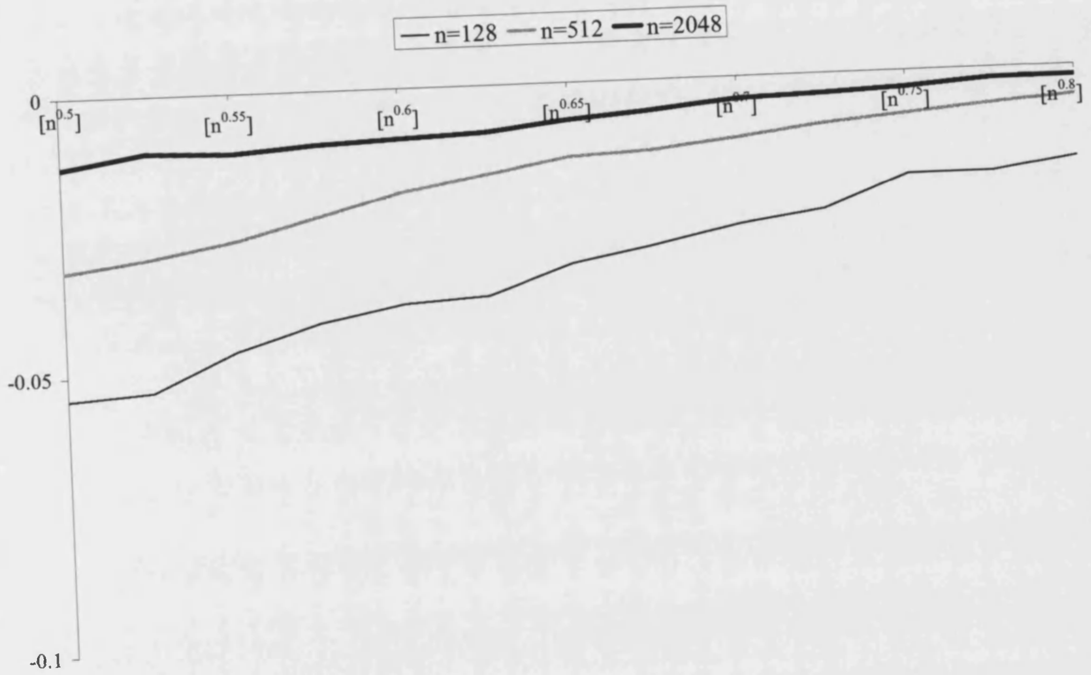


Figure 4.5: Bias of LW estimator; linear process with $\alpha_x = 0$.

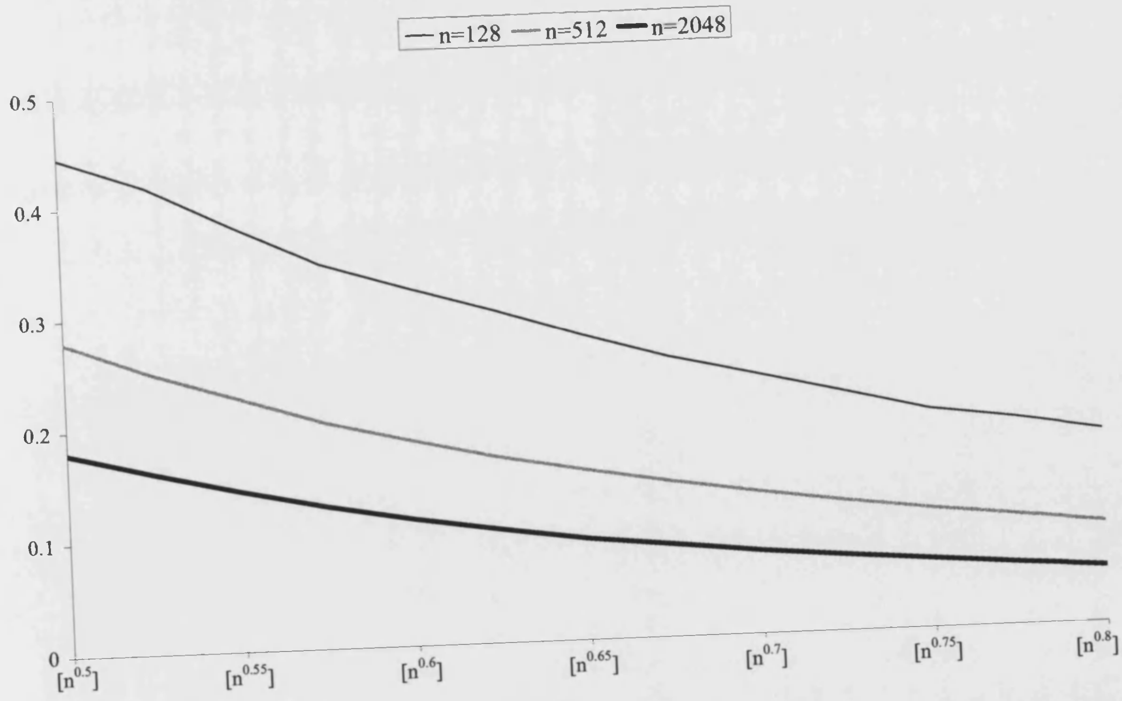


Figure 4.6: RMSE of LW estimator; linear process with $\alpha_x = 0$.

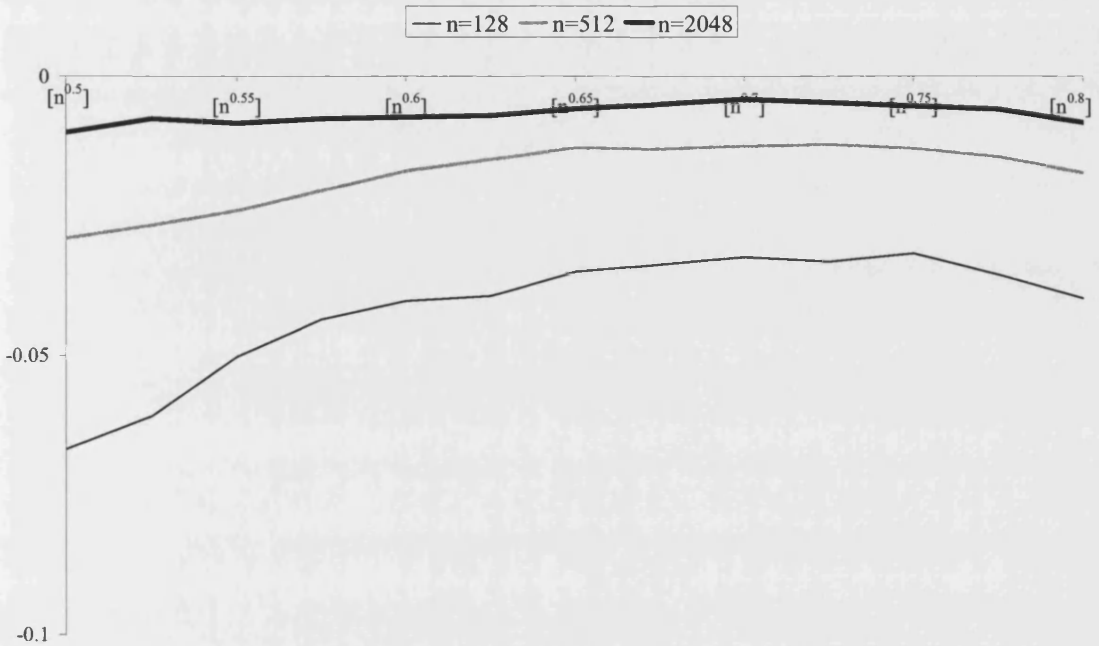


Figure 4.7: Bias of LW estimator; linear process with $\alpha_x = 0.4$.

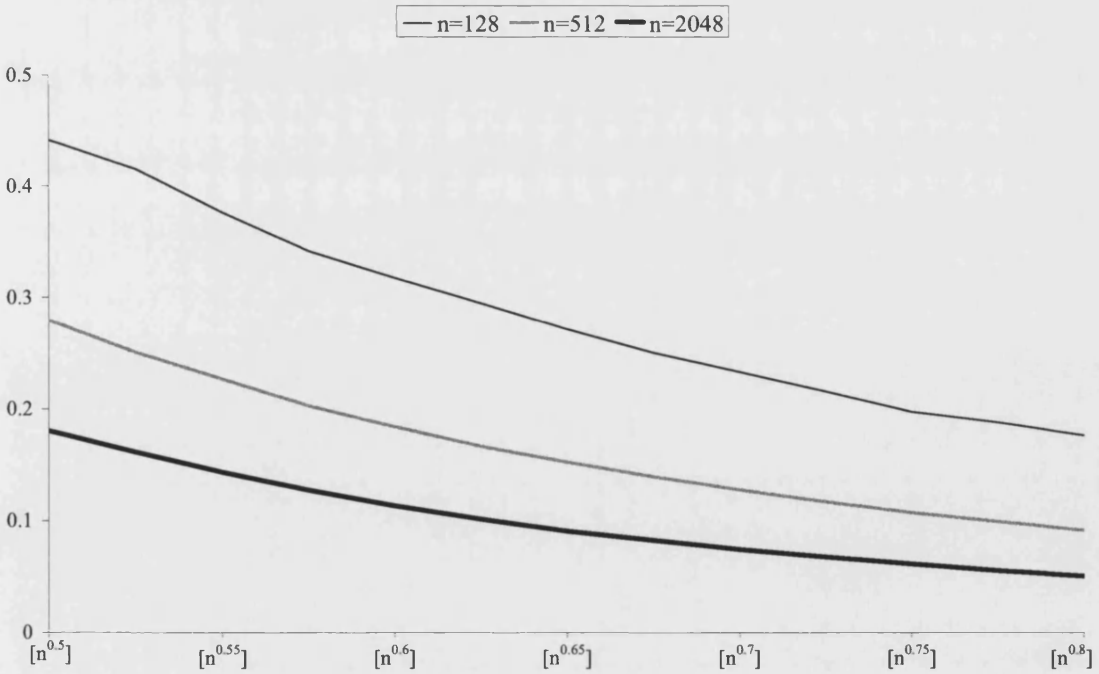


Figure 4.8: RMSE of LW estimator; linear process with $\alpha_x = 0.4$.

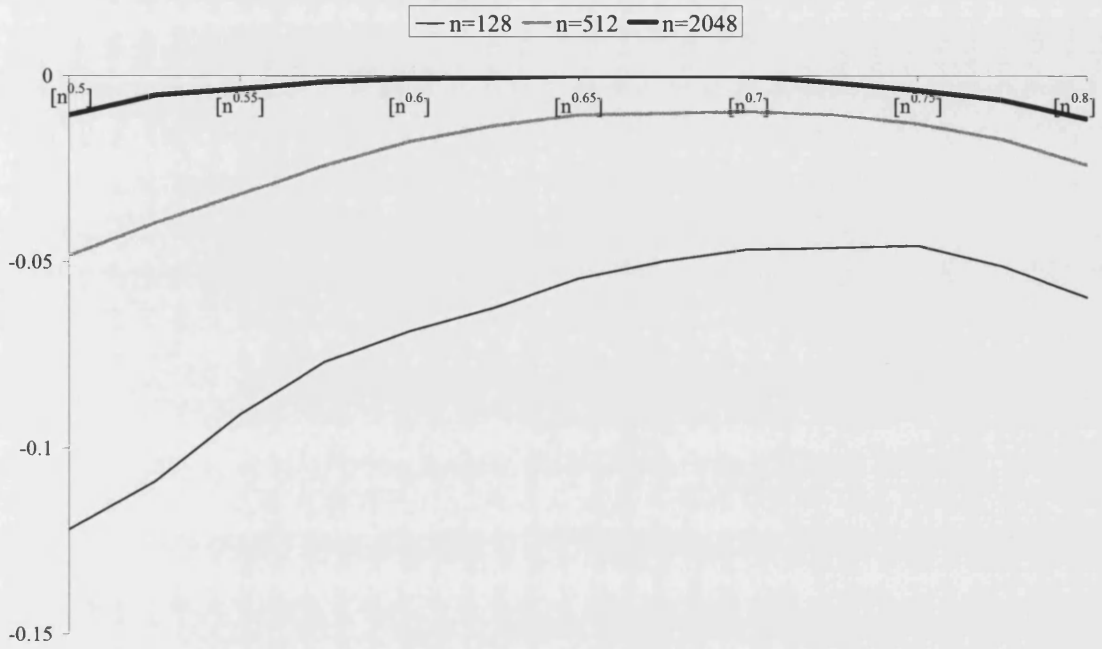


Figure 4.9: Bias of LW estimator; linear process with $\alpha_x = 0.8$.

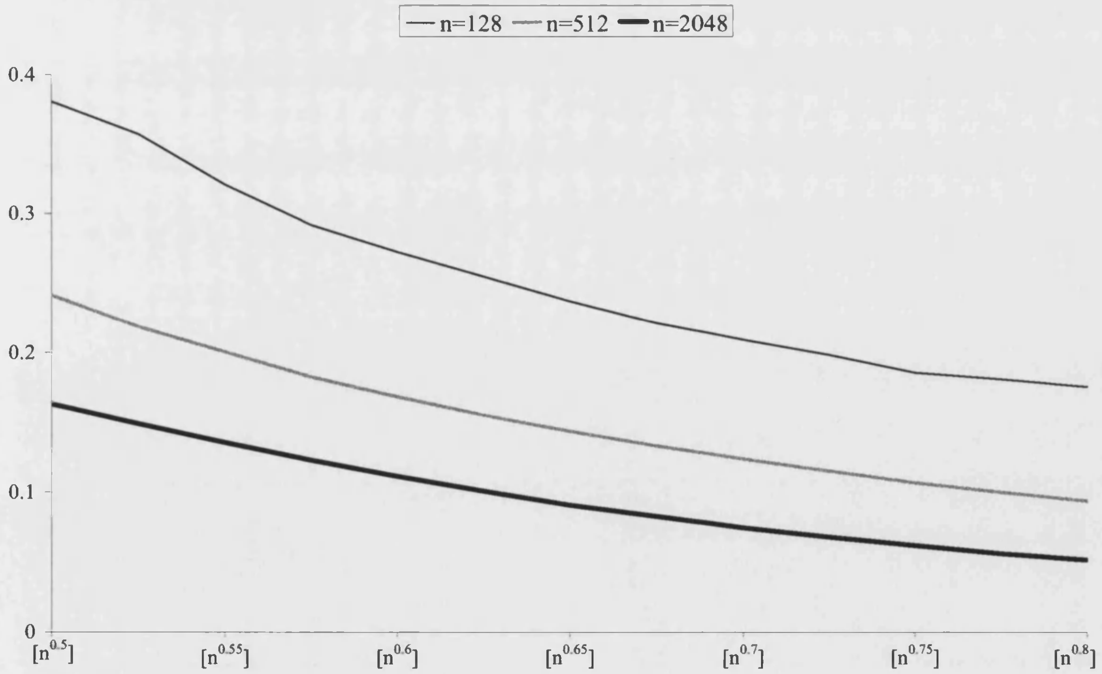


Figure 4.10: RMSE of LW estimator; linear process with $\alpha_x = 0.8$.

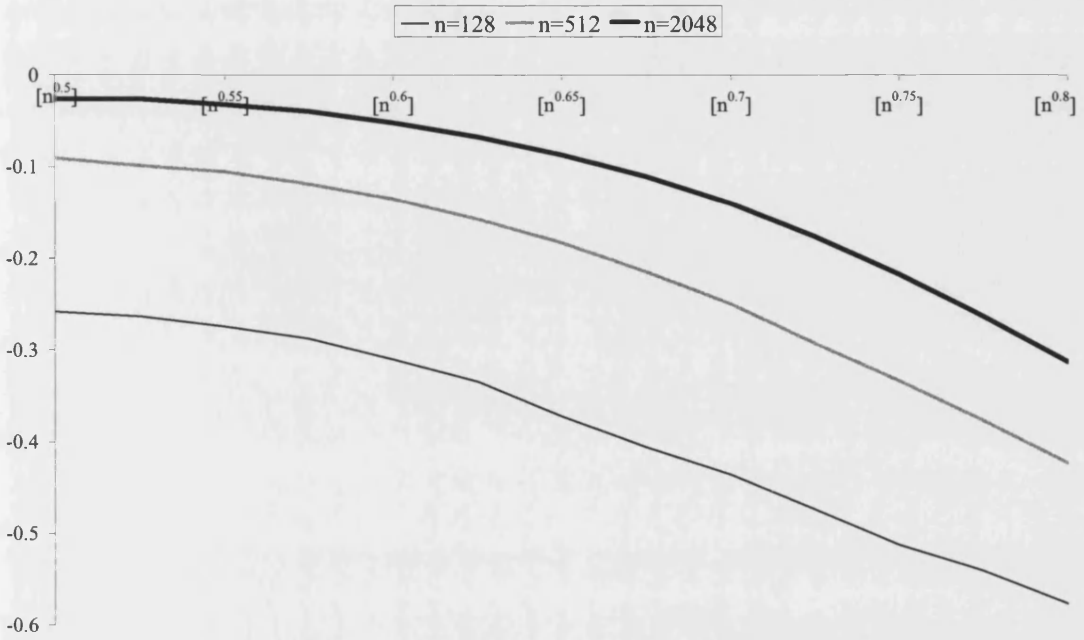


Figure 4.11: Bias of LW estimator; signal plus noise model with $\alpha_y = 0.8$ and $\alpha_z = -0.8$.

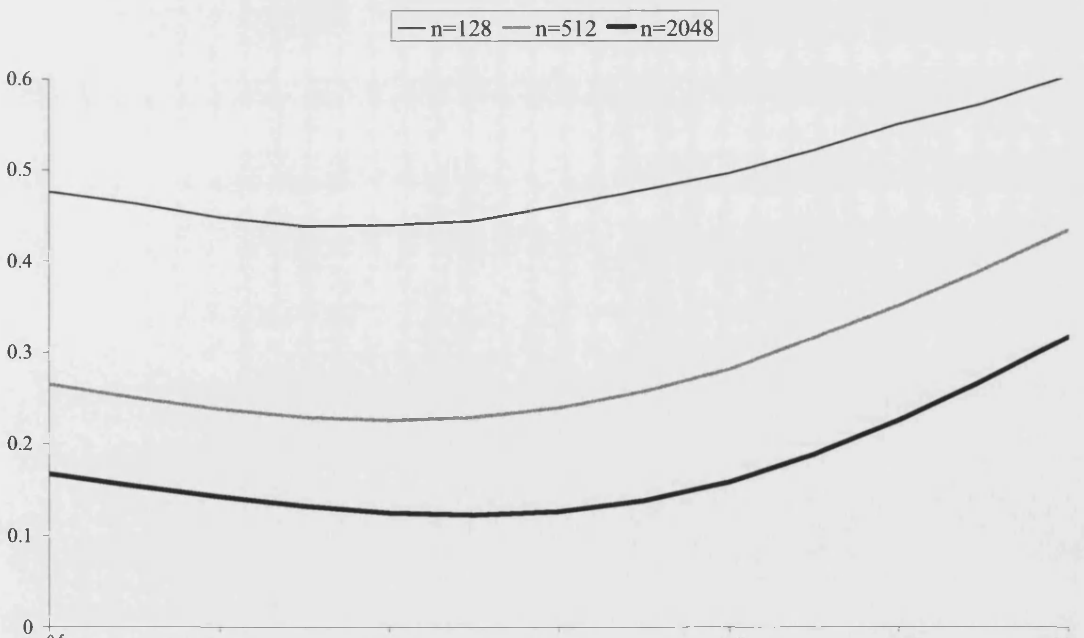


Figure 4.12: RMSE of LW estimator; signal plus noise model with $\alpha_y = 0.8$ and $\alpha_z = -0.8$.

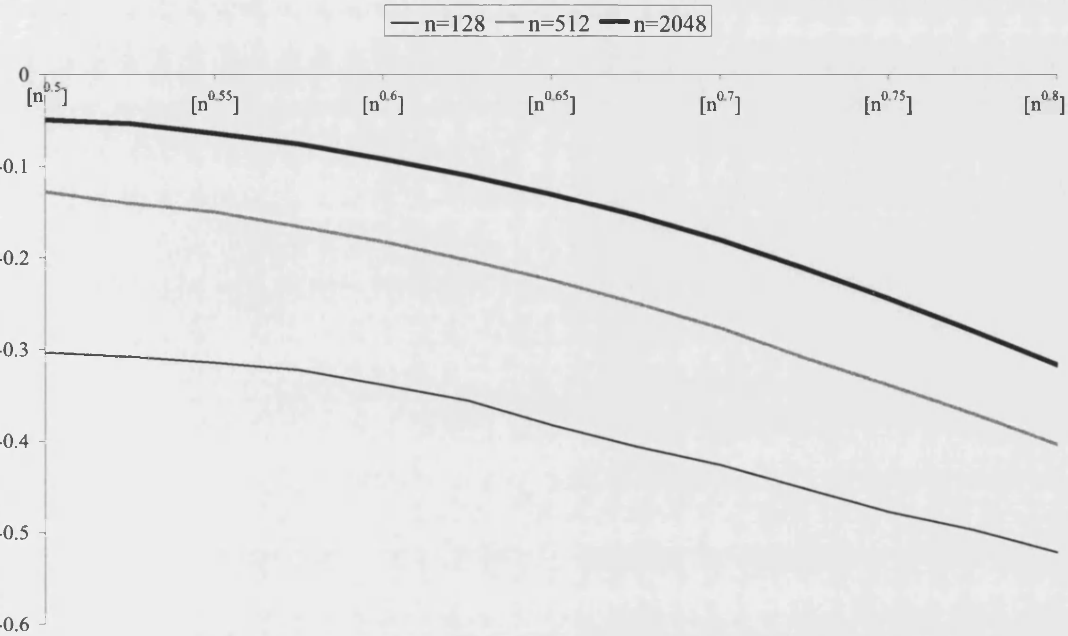


Figure 4.13: Bias of LW estimator; signal plus noise model with $\alpha_y = 0.8$ and $\alpha_z = -0.4$.

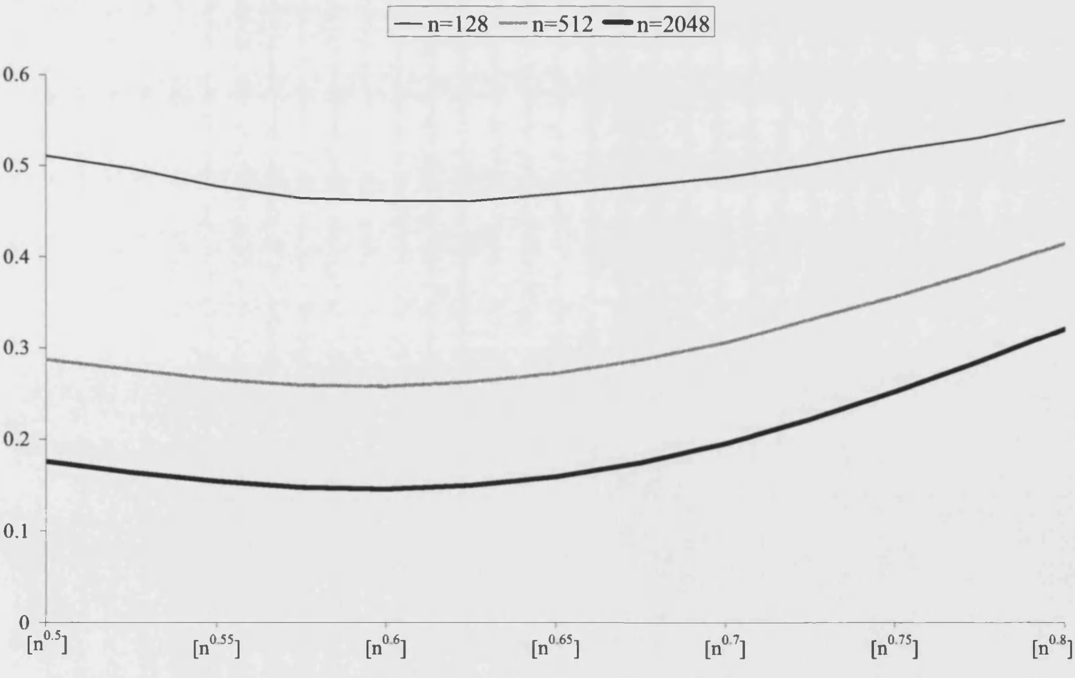


Figure 4.14: RMSE of LW estimator; signal plus noise model with $\alpha_y = 0.8$ and $\alpha_z = -0.4$.

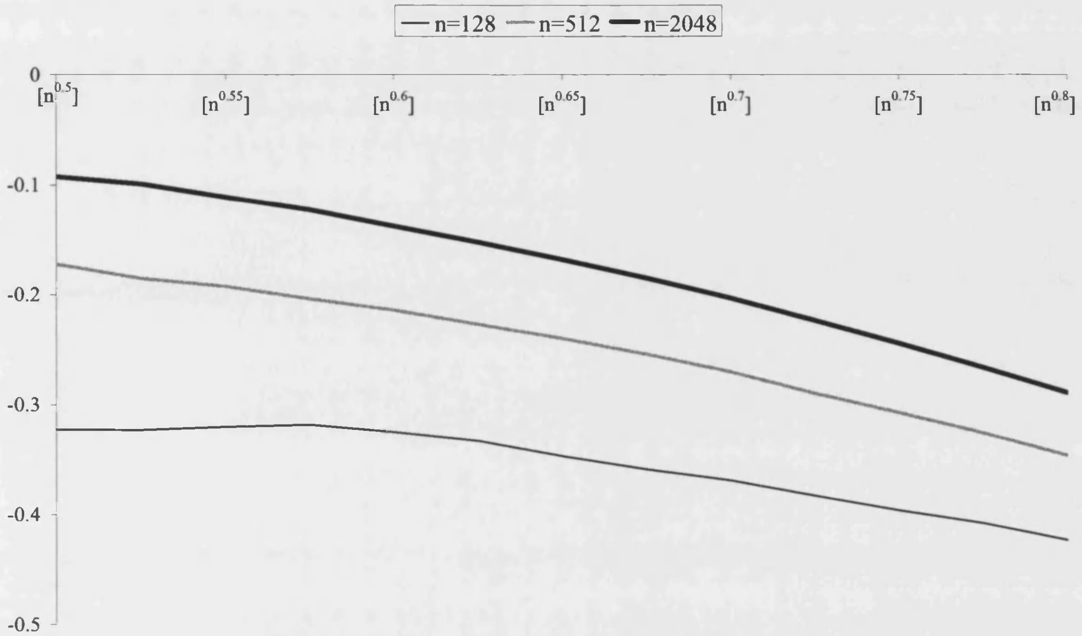


Figure 4.15: Bias of LW estimator; signal plus noise model with $\alpha_y = 0.8$ and $\alpha_z = 0$.

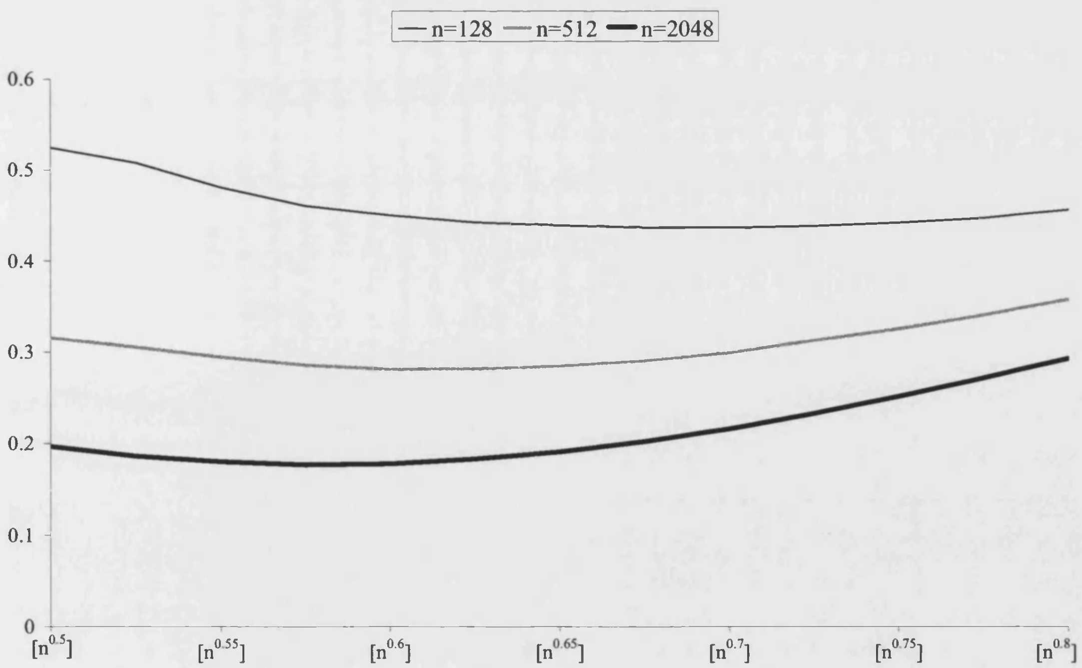


Figure 4.16: RMSE of LW estimator; signal plus noise model with $\alpha_y = 0.8$ and $\alpha_z = 0$.

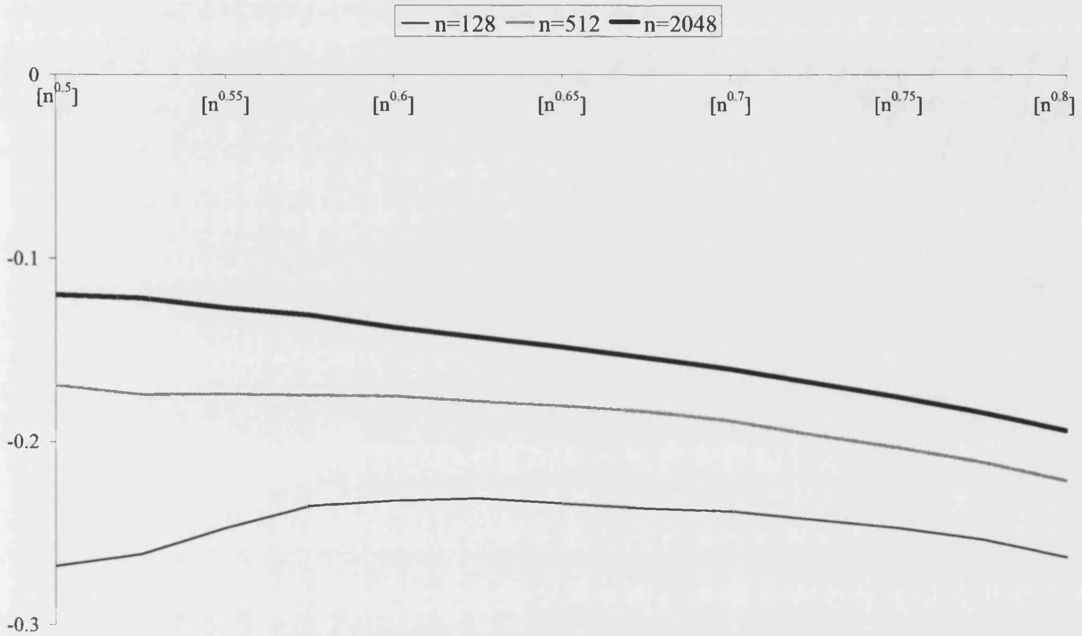


Figure 4.17: Bias of LW estimator; signal plus noise model with $\alpha_y = 0.8$ and $\alpha_z = 0.4$.

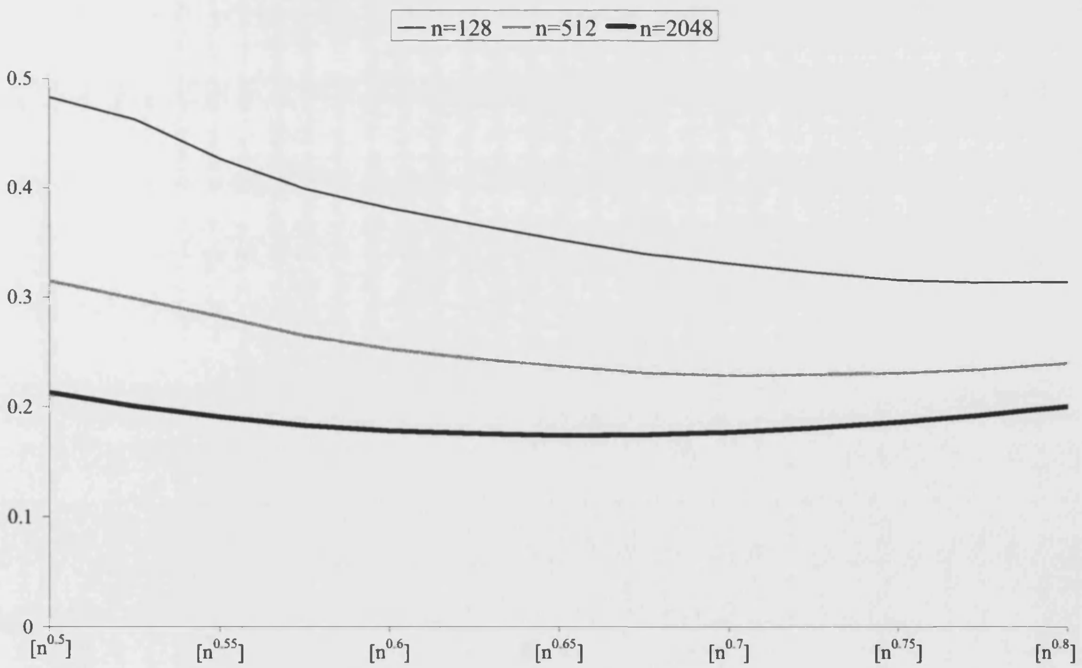


Figure 4.18: RMSE of LW estimator; signal plus noise model with $\alpha_y = 0.8$ and $\alpha_z = 0.4$.

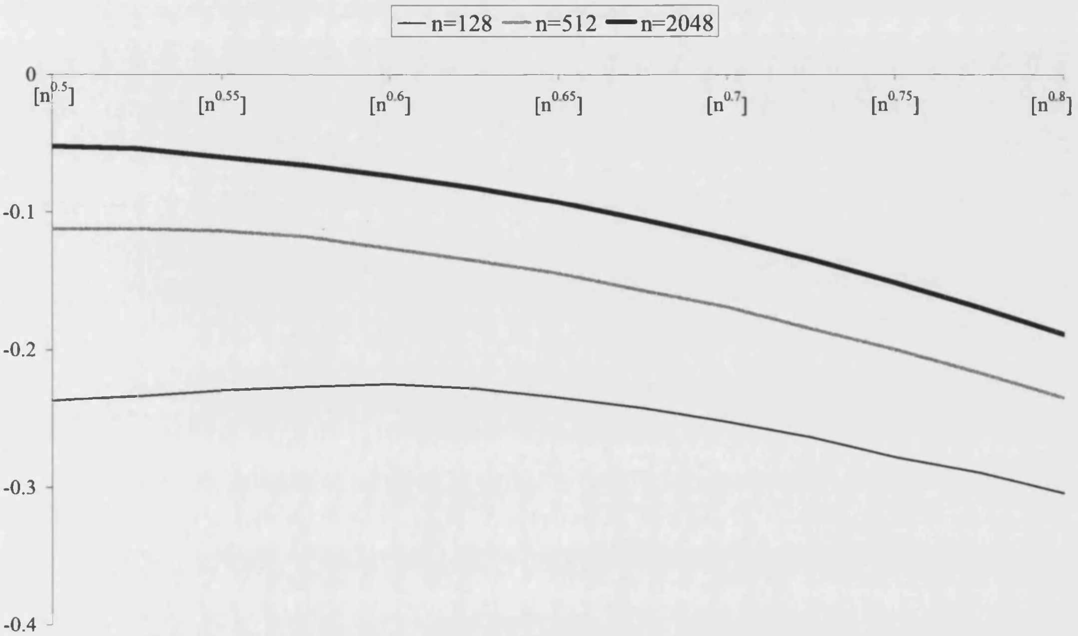


Figure 4.19: Bias of LW estimator; signal plus noise model with $\alpha_y = 0.8$, $\alpha_z = 0$ and signal-to-ratio = 2.

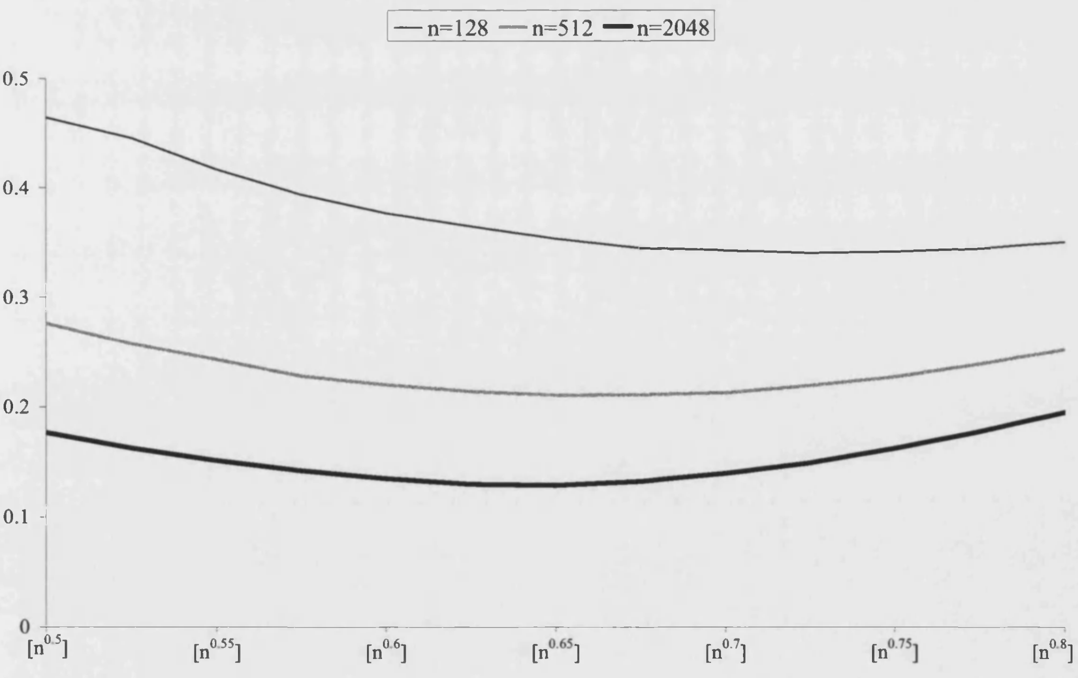


Figure 4.20: RMSE of LW estimator; signal plus noise model with $\alpha_y = 0.8$, $\alpha_z = 0$ and signal-to-ratio = 2.

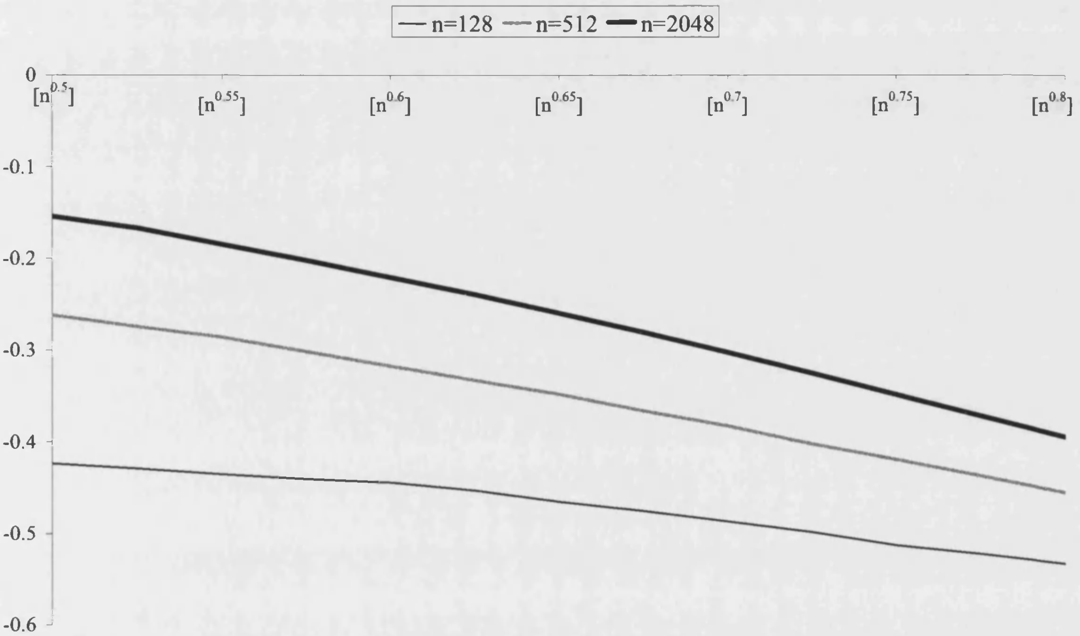


Figure 4.21: Bias of LW estimator; signal plus noise model with $\alpha_y = 0.8$, $\alpha_z = 0$ and signal-to-ratio = 0.5.

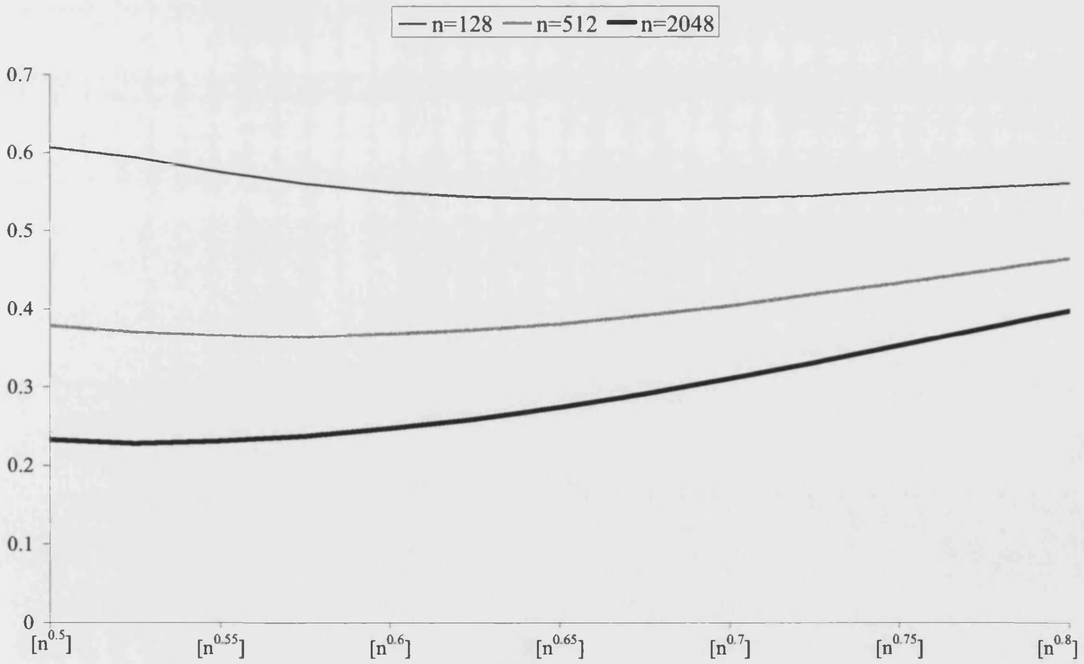


Figure 4.22: RMSE of LW estimator; signal plus noise model with $\alpha_y = 0.8$, $\alpha_z = 0$ and signal-to-ratio = 0.5.

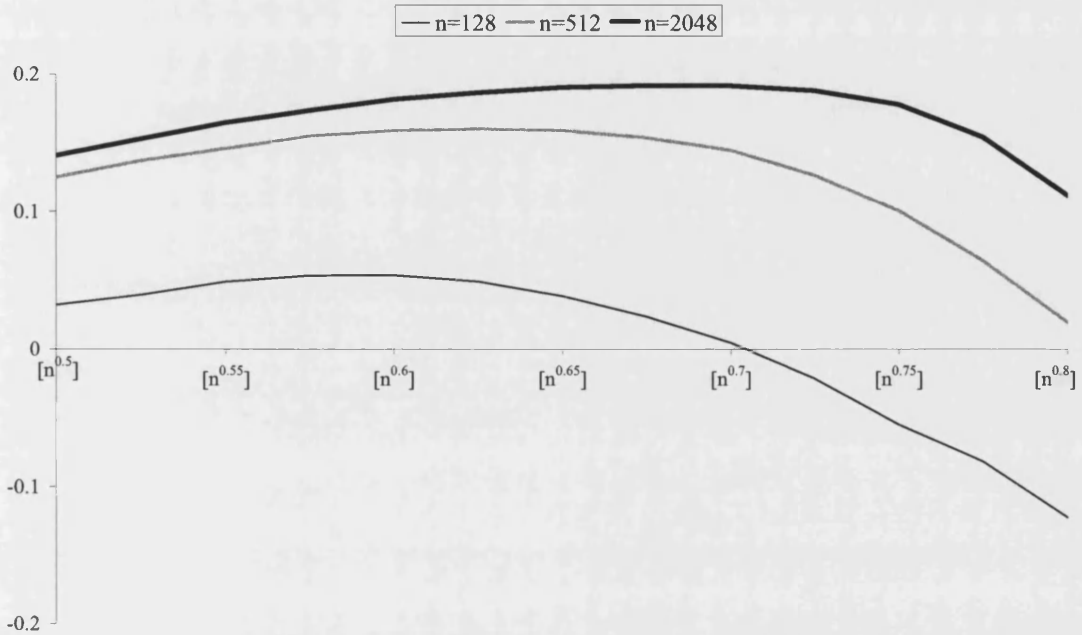


Figure 4.23: Bias of LW estimator; signal plus noise model with $\alpha_y = 0.8$, $\alpha_z = 0$ and $\rho = -0.5$.

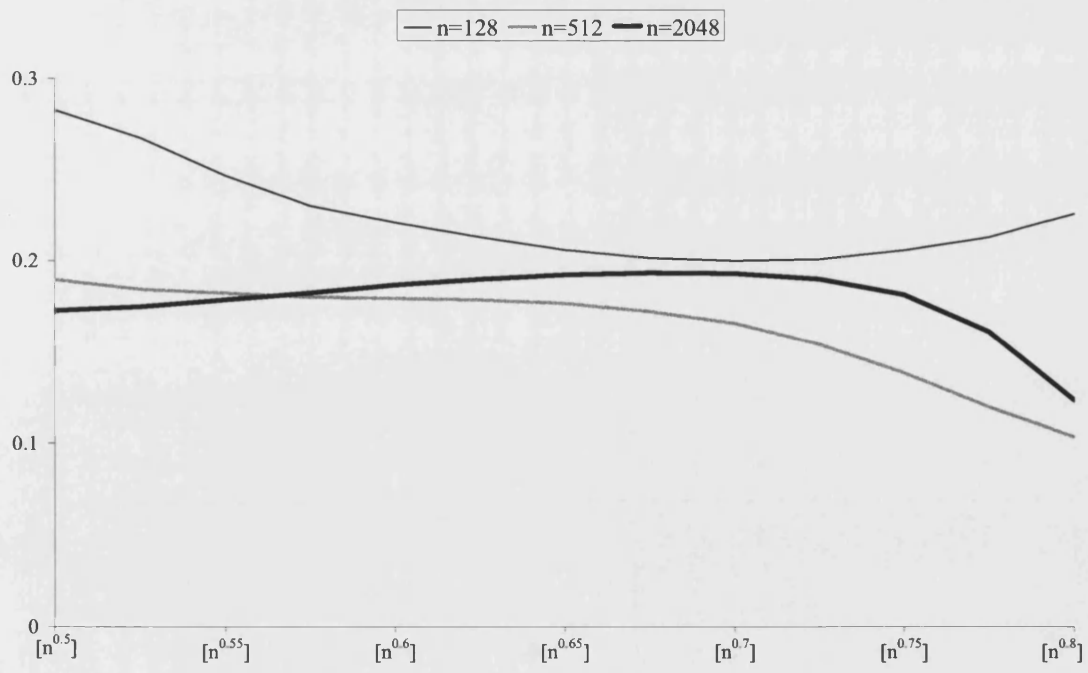


Figure 4.24: RMSE of LW estimator; signal plus noise model with $\alpha_y = 0.8$, $\alpha_z = 0$ and $\rho = -0.5$.

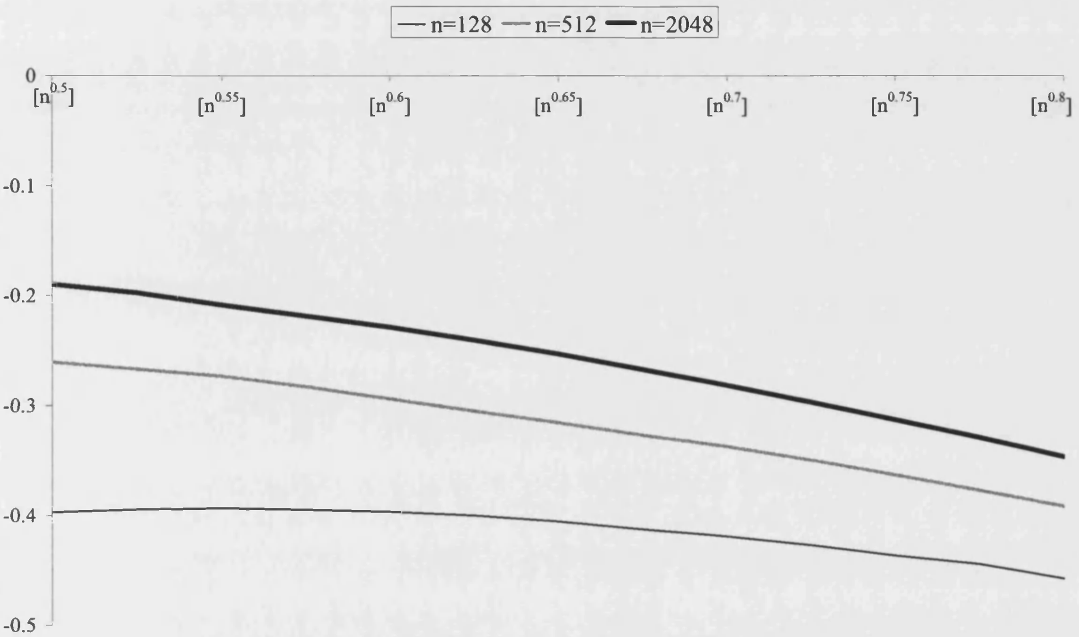


Figure 4.25: Bias of LW estimator; signal plus noise model with $\alpha_y = 0.8$, $\alpha_z = 0$ and $\rho = 0.5$.

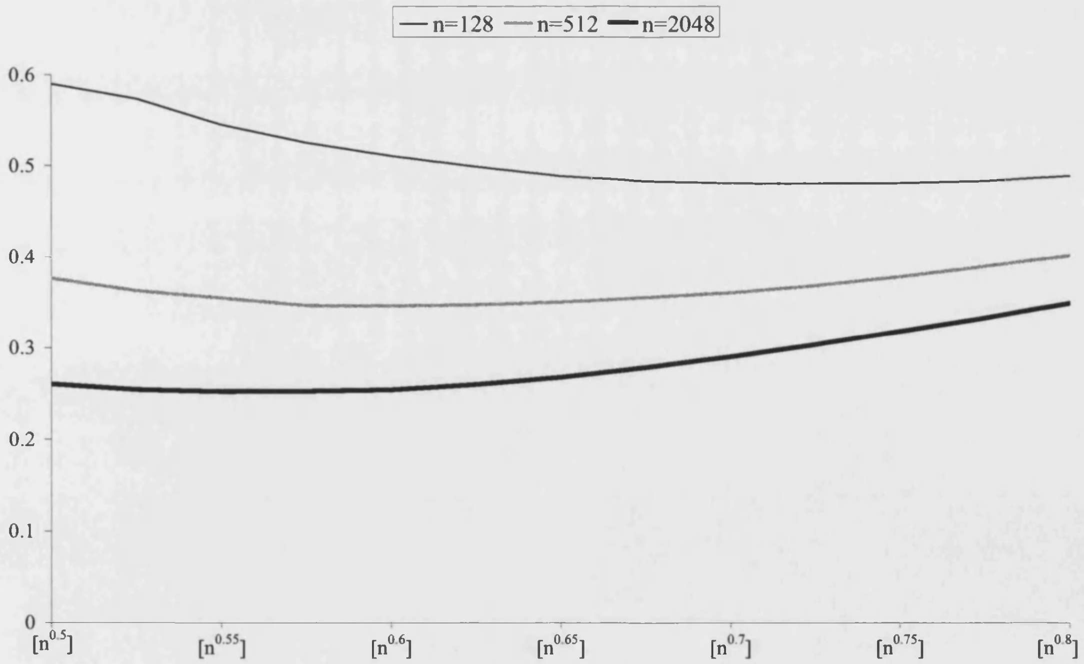


Figure 4.26: RMSE of LW estimator; signal plus noise model with $\alpha_y = 0.8$, $\alpha_z = 0$ and $\rho = 0.5$.

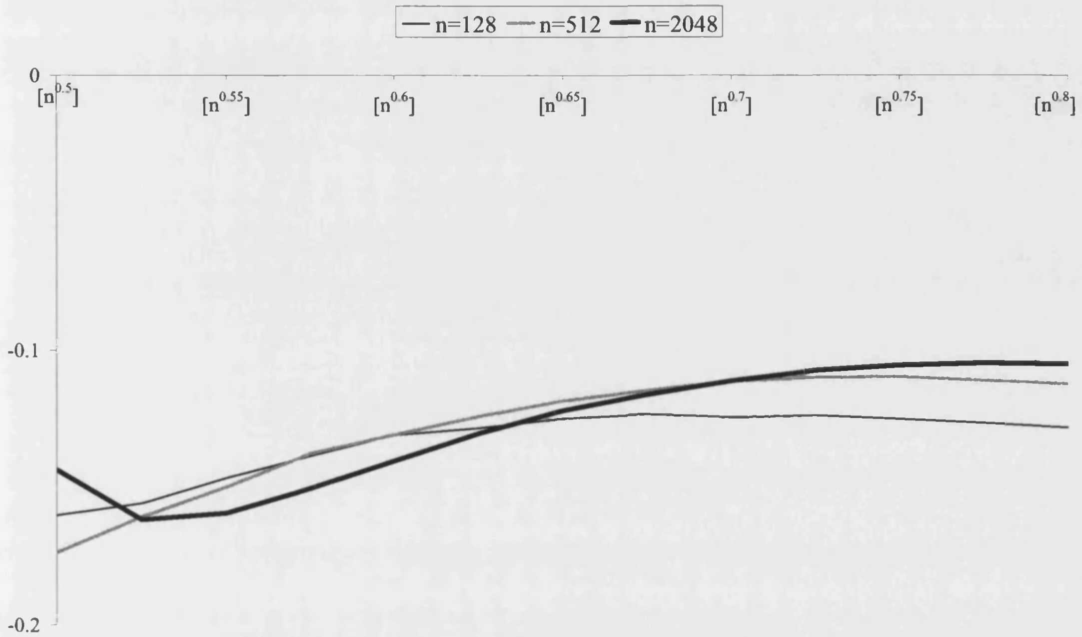


Figure 4.27: Bias of LW estimator; structural model with $\alpha_{\mu_x} = 0.4$, $\alpha_{\omega,c_x} = 0.1$ and $\omega = 0.15$.

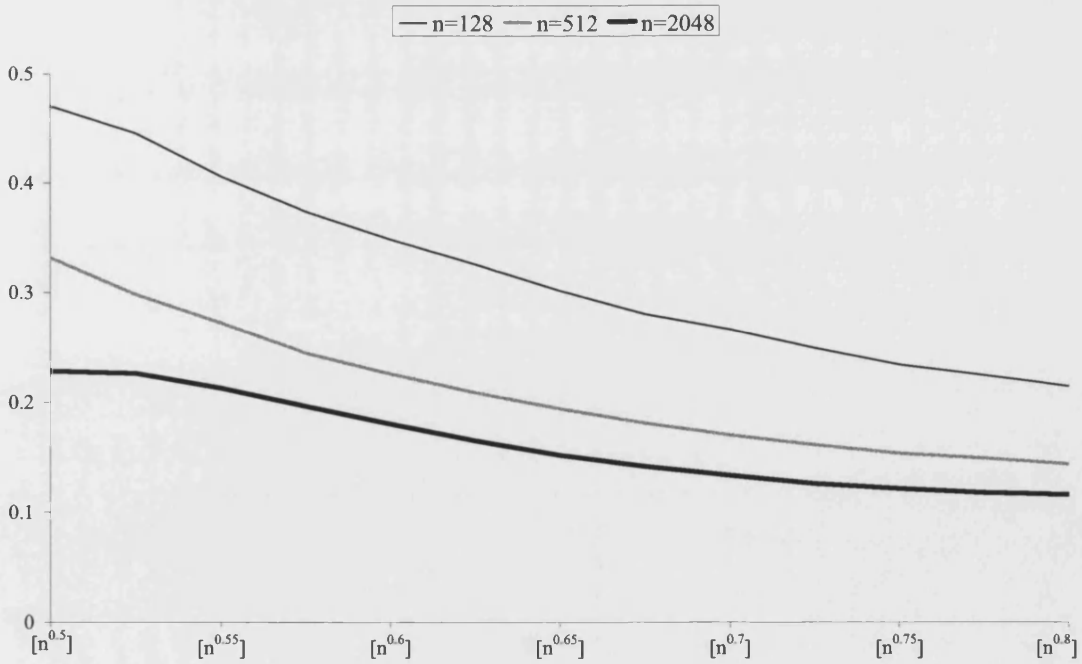


Figure 4.28: RMSE of LW estimator; structural model with $\alpha_{\mu_x} = 0.4$, $\alpha_{\omega,c_x} = 0.1$ and $\omega = 0.15$.

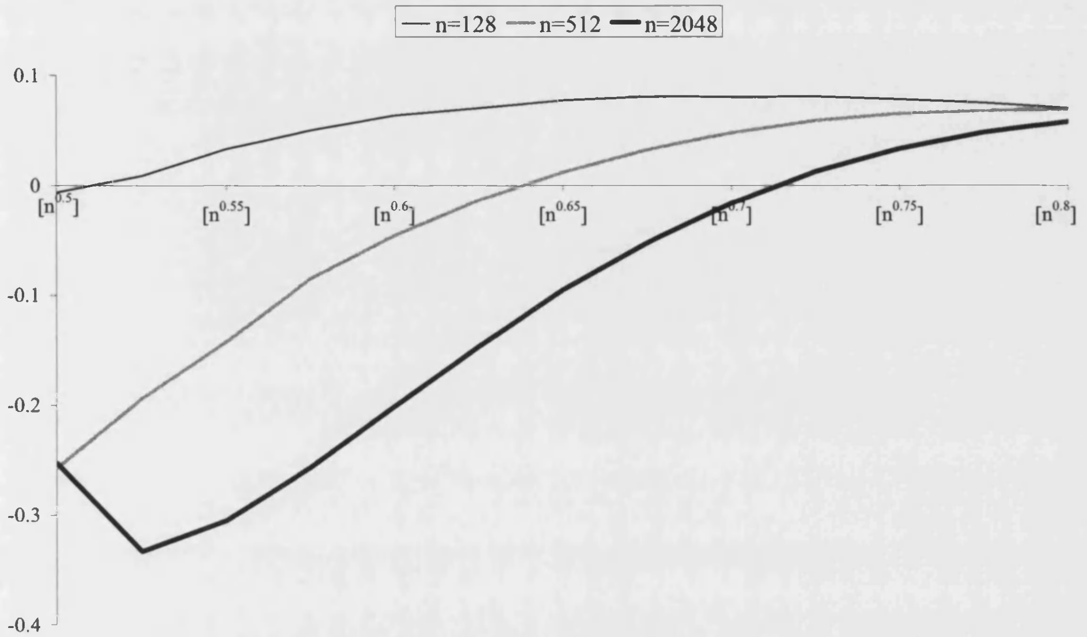


Figure 4.29: Bias of LW estimator; structural model with $\alpha_{\mu_x} = 0.4$, $\alpha_{\omega,c_x} = 0.3$ and $\omega = 0.15$.

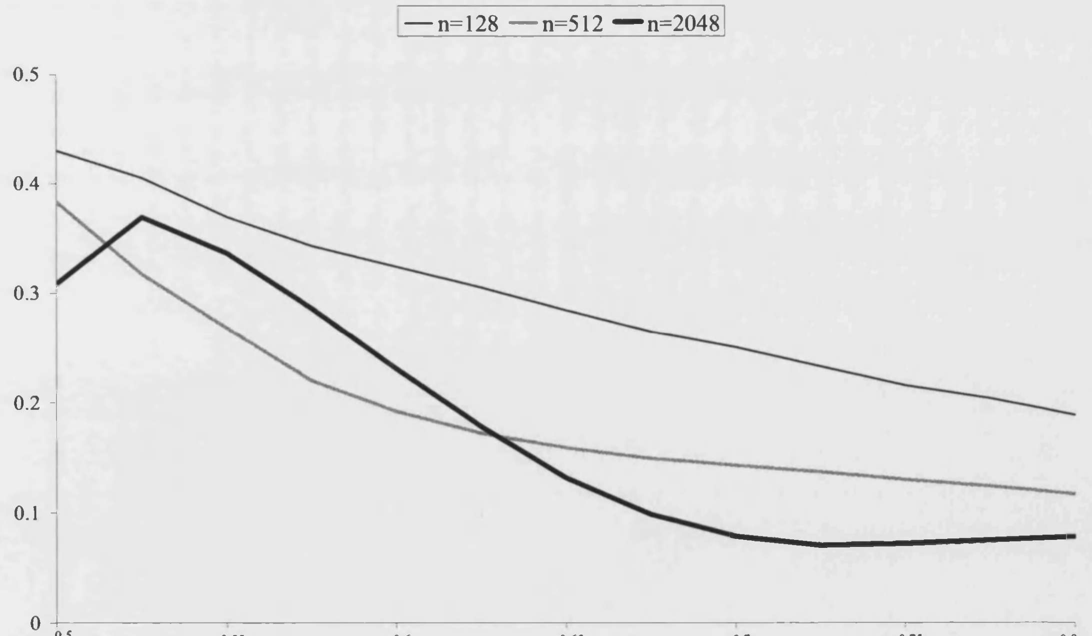


Figure 4.30: RMSE of LW estimator; structural model with $\alpha_{\mu_x} = 0.4$, $\alpha_{\omega,c_x} = 0.3$ and $\omega = 0.15$.

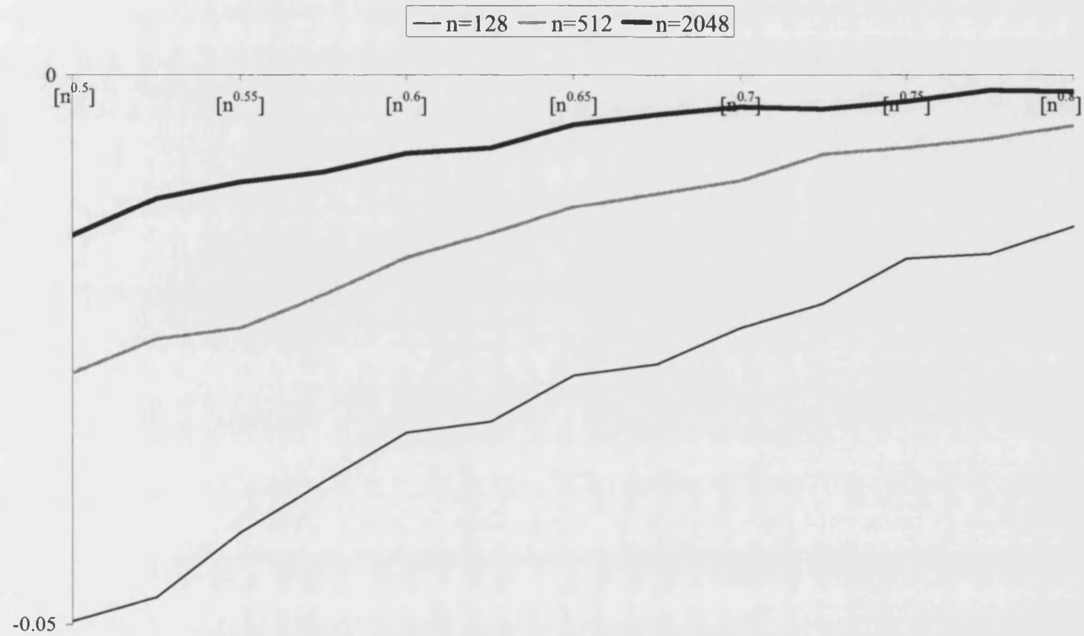


Figure 4.31: Bias of LW estimator; exponential of Gaussian process with $a_x = 0$ and $a_\xi = 0$.

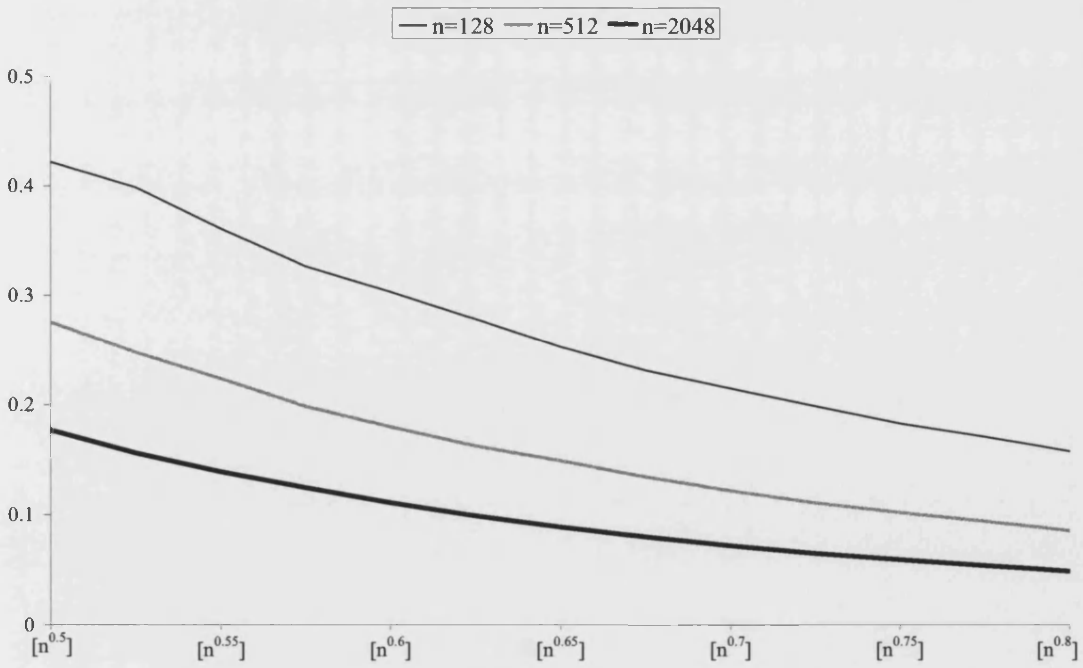


Figure 4.32: RMSE of LW estimator; exponential of Gaussian process with $\alpha_x = 0$ and $\alpha_\xi = 0$.

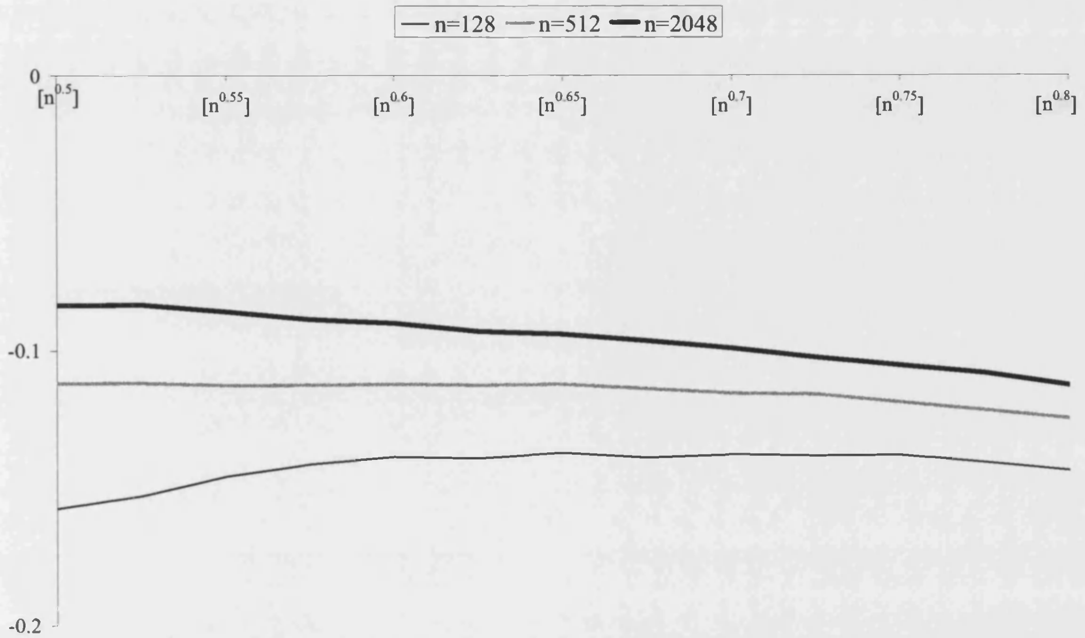


Figure 4.33: Bias of LW estimator; exponential of Gaussian process with $a_x = 0.4$ and $a_\xi = 0.4$.

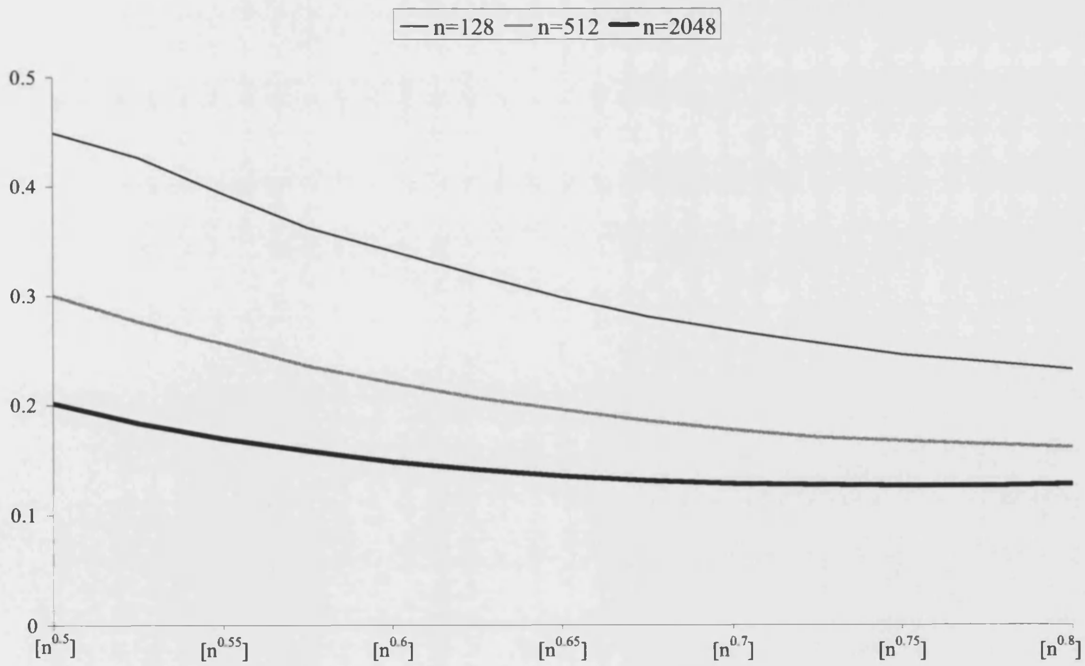


Figure 4.34: RMSE of LW estimator; exponential of Gaussian process with $a_x = 0.4$ and $a_\xi = 0.4$.

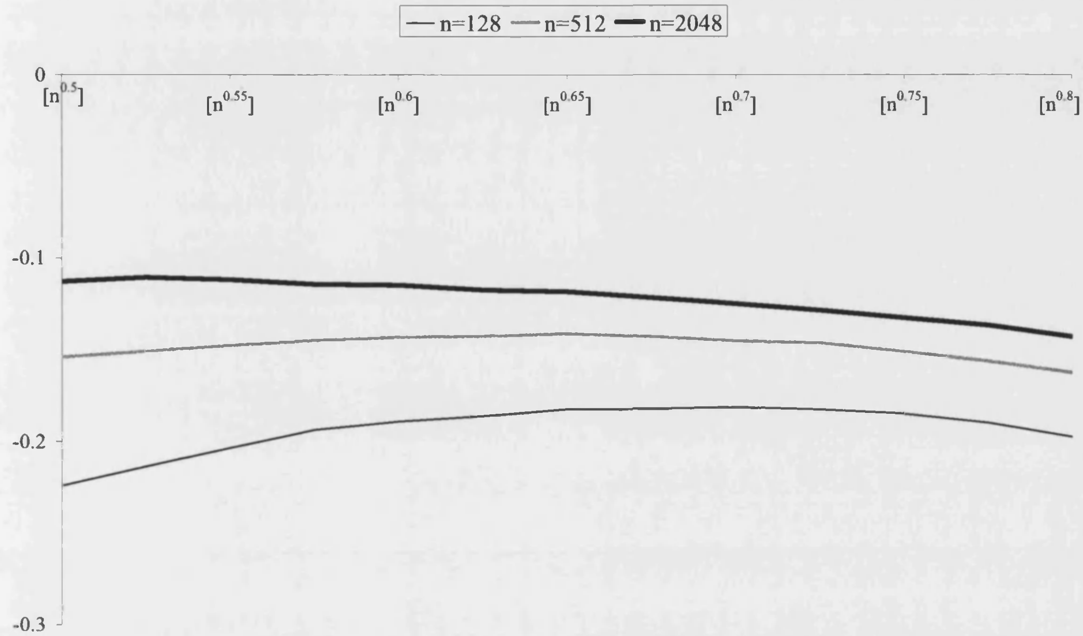


Figure 4.35: Bias of LW estimator; exponential of Gaussian process with $a_x = 0.8$ and $a_\xi = 0.8$.

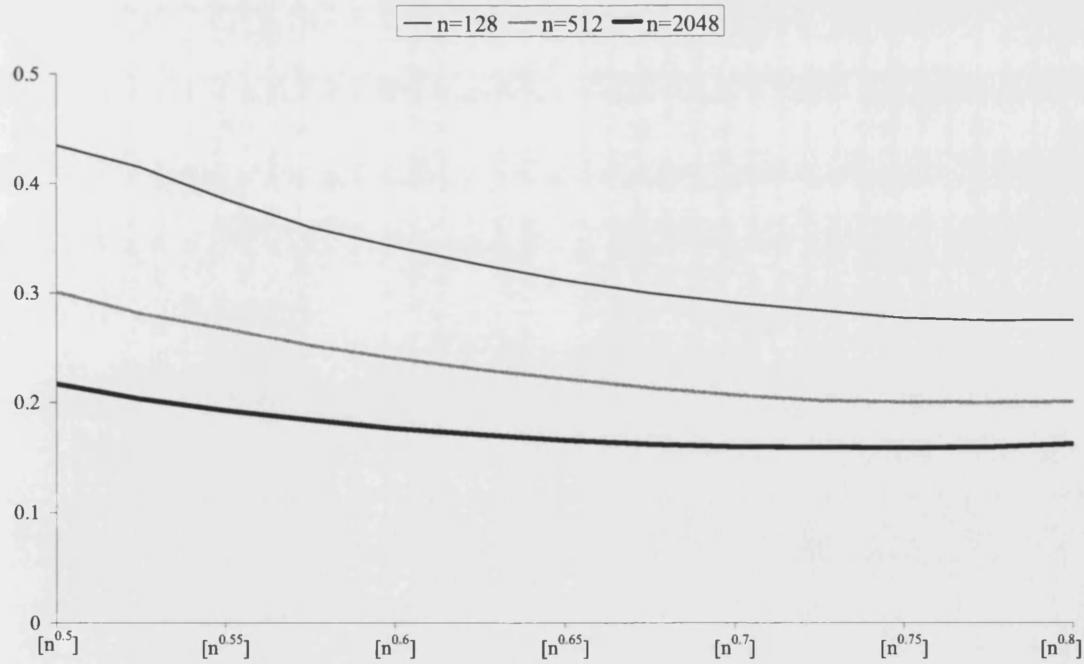


Figure 4.36: RMSE of LW estimator; exponential of Gaussian process with $a_x = 0.8$ and $a_\xi = 0.8$.

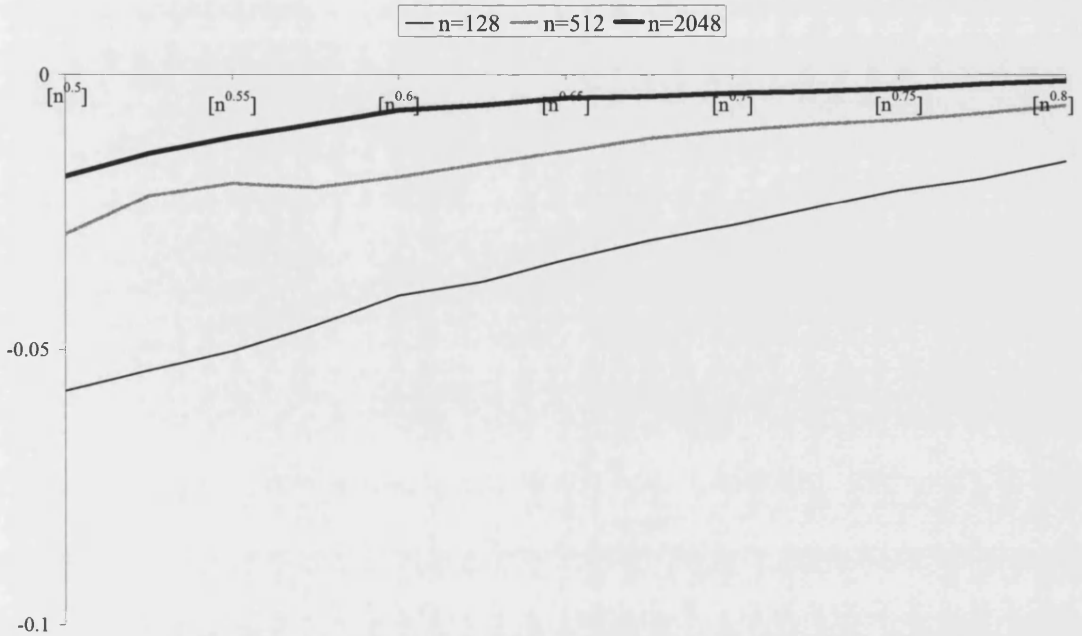


Figure 4.37: Bias of LW estimator; square of Gaussian process with $a_x = 0$ and $a_\xi = 0$.

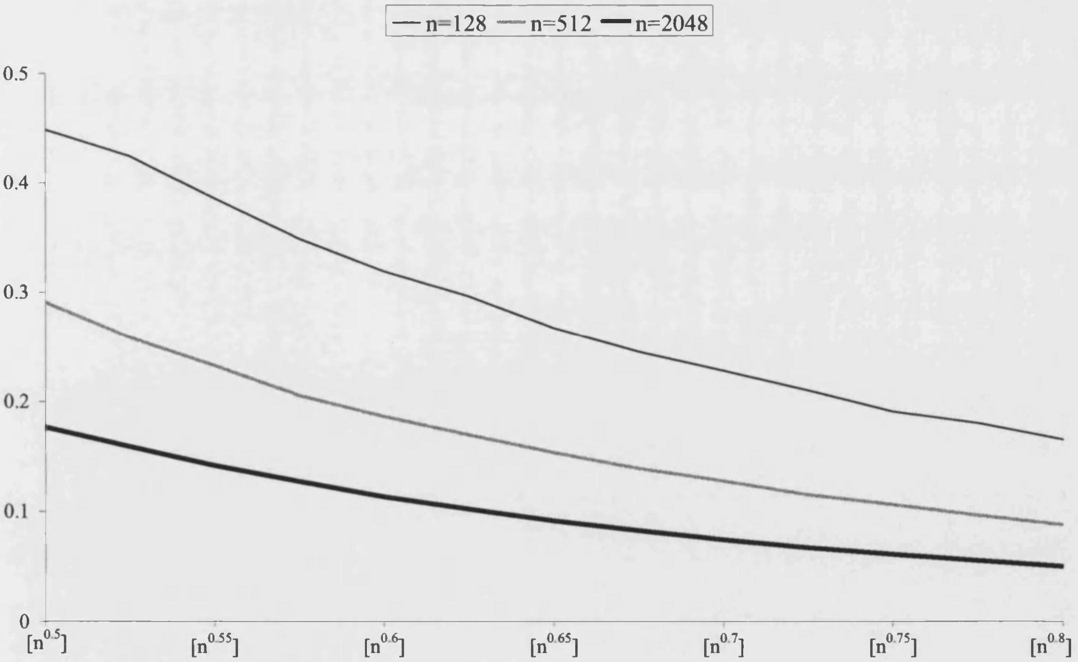


Figure 4.38: RMSE of LW estimator; square of Gaussian process with $a_x = 0$ and $a_\xi = 0$.

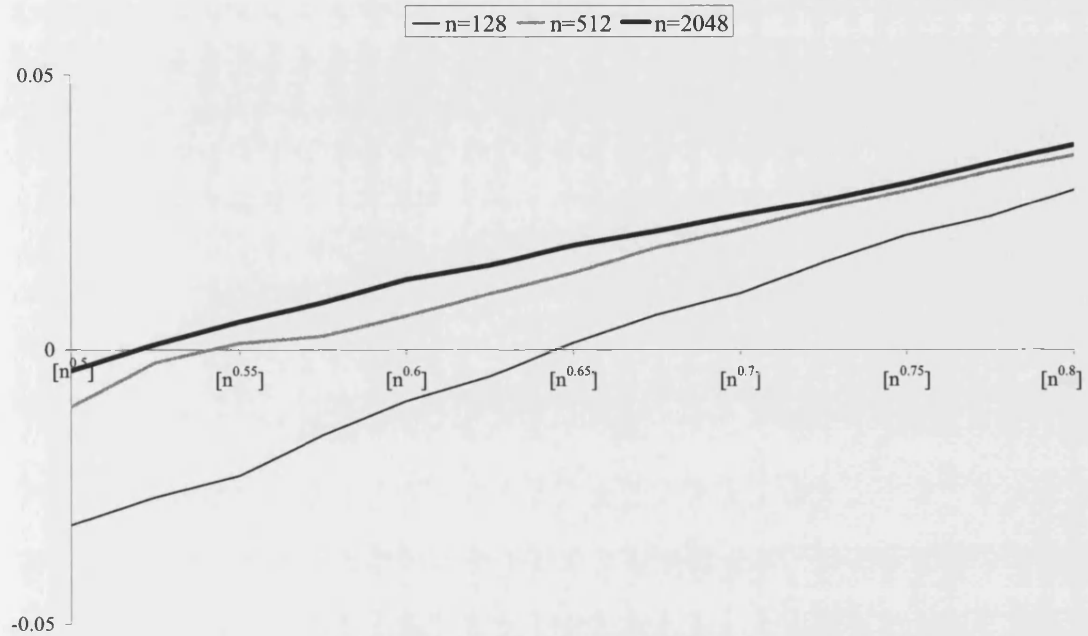


Figure 4.39: Bias of LW estimator; square of Gaussian process with $a_x = 0$ and $a_\xi = 0.3$.

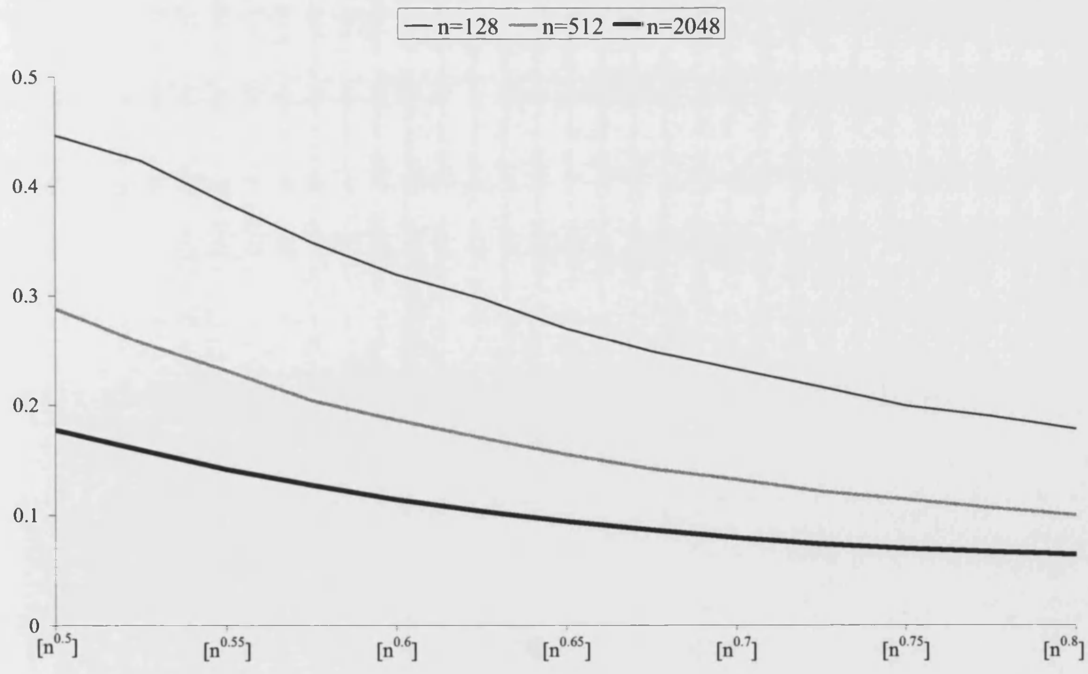


Figure 4.40: RMSE of LW estimator; square of Gaussian process with $a_x = 0$ and $a_\xi = 0.3$.

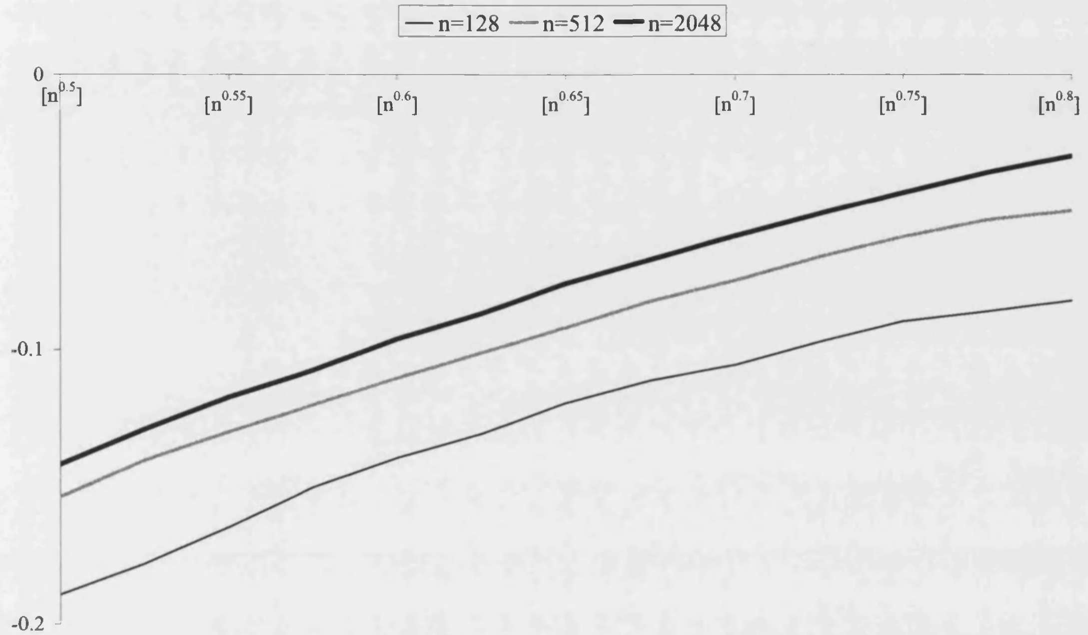


Figure 4.41: Bias of LW estimator; square of Gaussian process with $a_x = 0.4$ and $a_\xi = 0.7$.

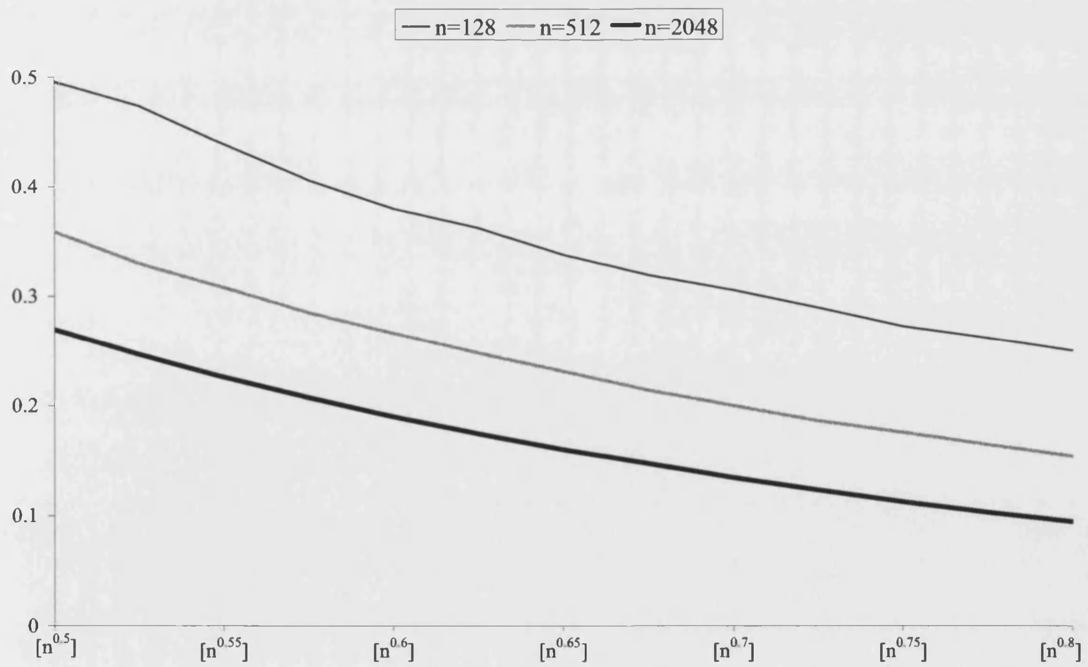


Figure 4.42: RMSE of LW estimator; square of Gaussian process with $a_x = 0.4$ and $a_\xi = 0.7$.

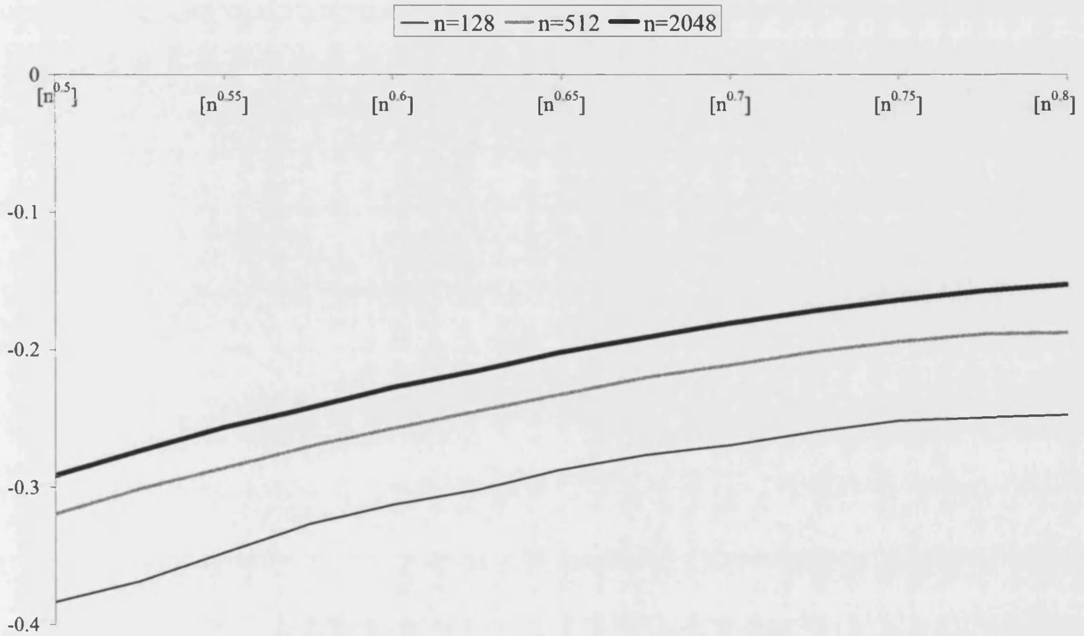


Figure 4.43: Bias of LW estimator; square of Gaussian process with $a_x = 0.8$ and $a_\xi = 0.9$.

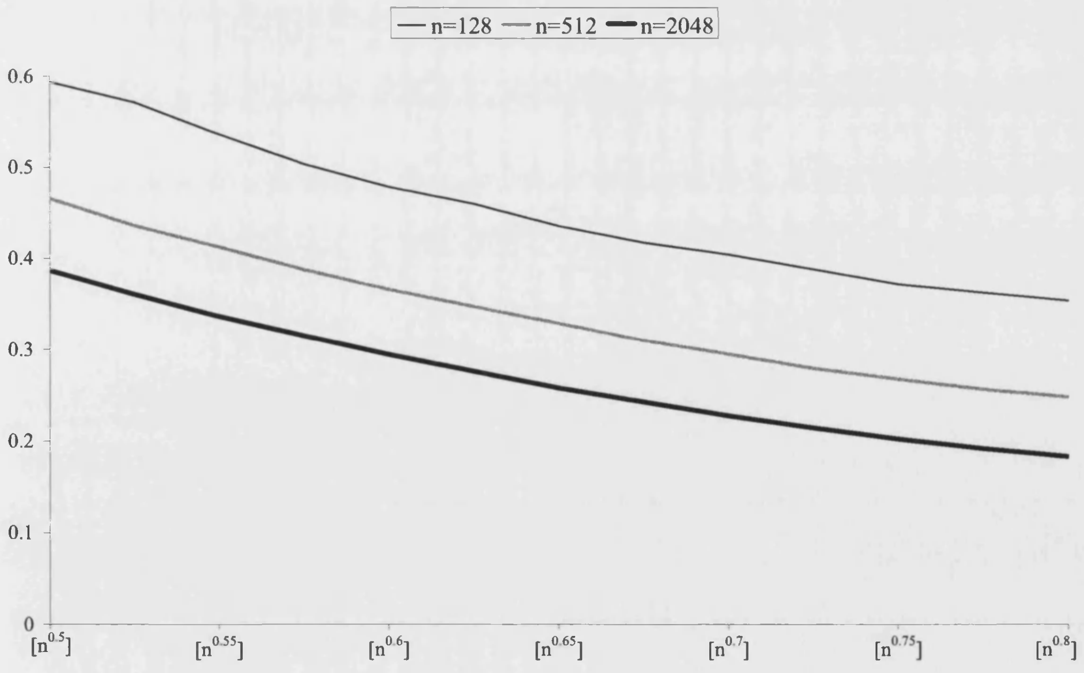


Figure 4.44: RMSE of LW estimator; square of Gaussian process with $a_x = 0.8$ and $a_\xi = 0.9$.

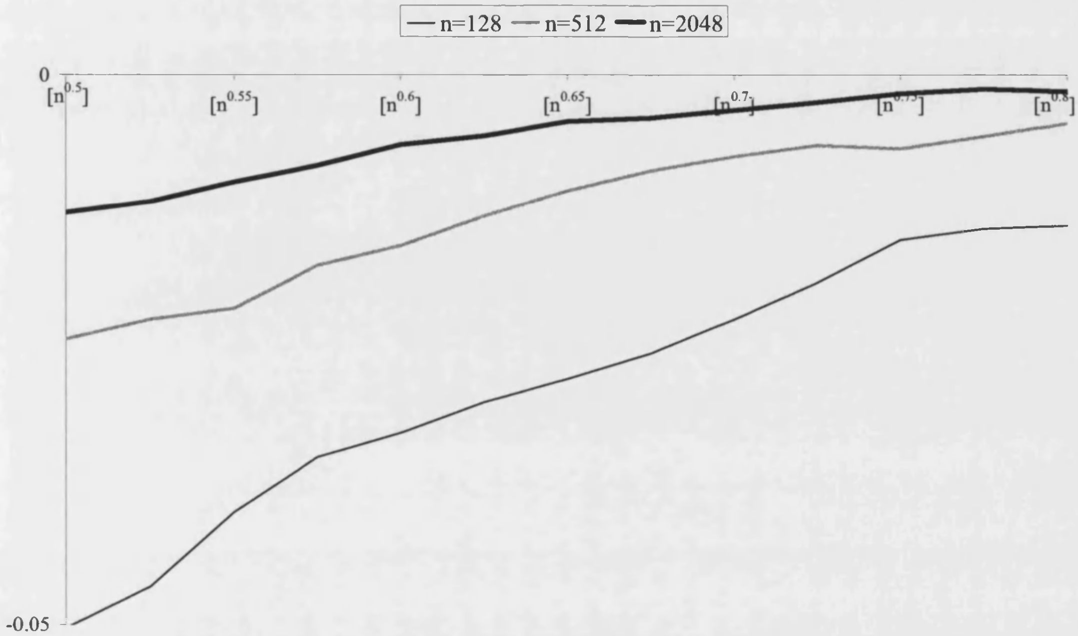


Figure 4.45: Bias of LW estimator; LMSV model, absolute returns with $\alpha_{|r|} = 0$ and $\alpha_{\xi} = 0$.

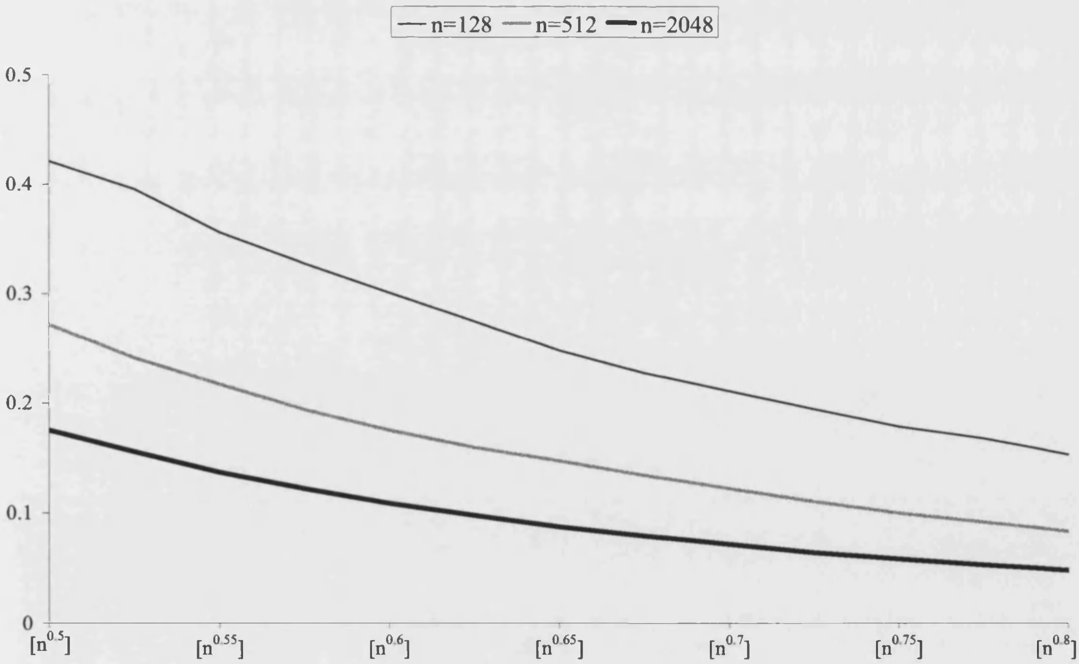


Figure 4.46: RMSE of LW estimator; LMSV model, absolute returns with $\alpha_{|r|} = 0$ and $\alpha_{\xi} = 0$.

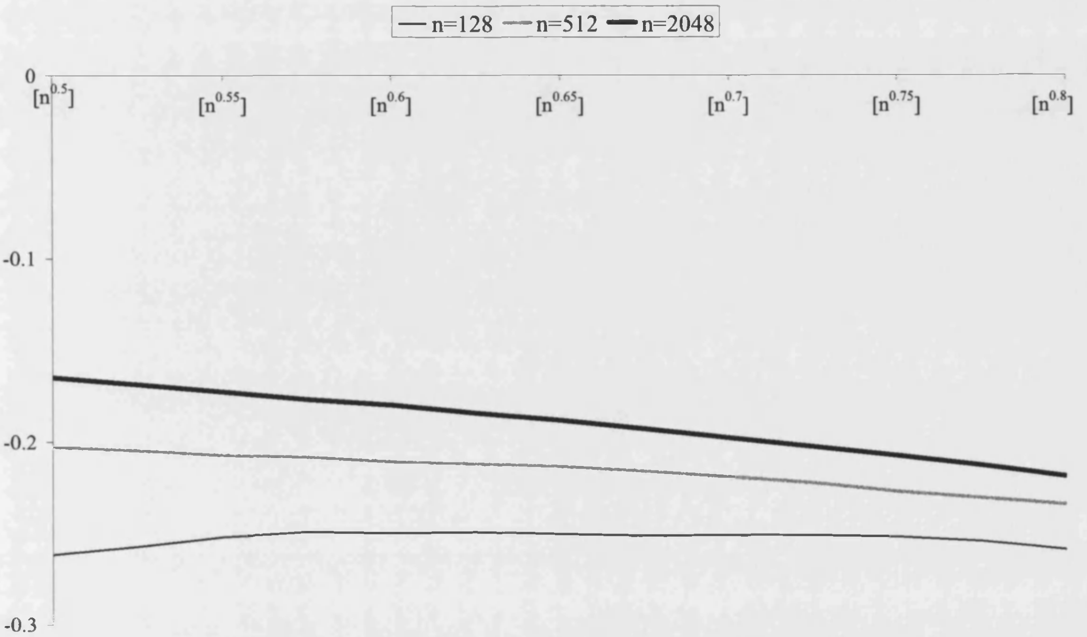


Figure 4.47: Bias of LW estimator; LMSV model, absolute returns with $\alpha_{|r|} = 0.4$ and $\alpha_{\xi} = 0.4$.

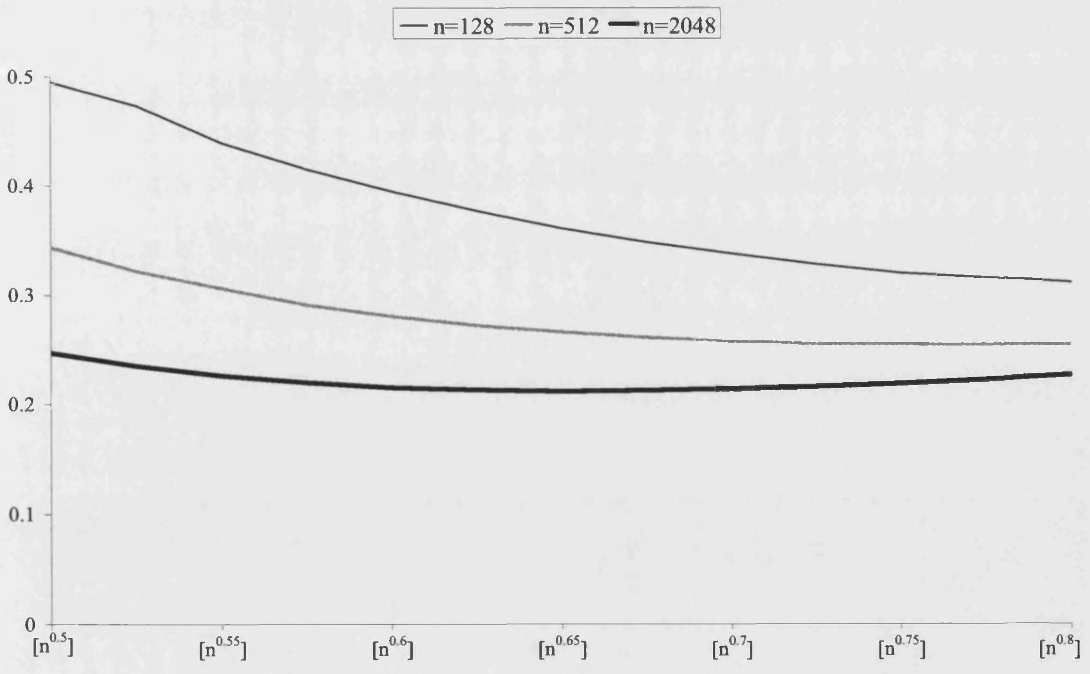


Figure 4.48: RMSE of LW estimator; LMSV model, absolute returns with $\alpha_{|r|} = 0.4$ and $\alpha_{\xi} = 0.4$.

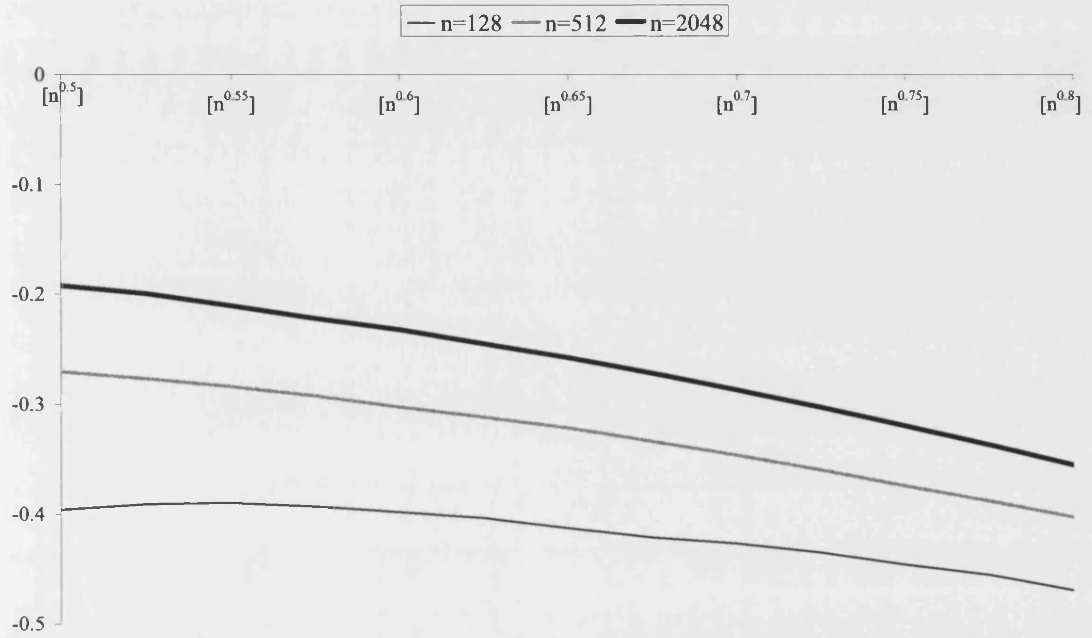


Figure 4.49: Bias of LW estimator; LMSV model, absolute returns with $\alpha_{|r|} = 0.8$ and $\alpha_\xi = 0.8$.

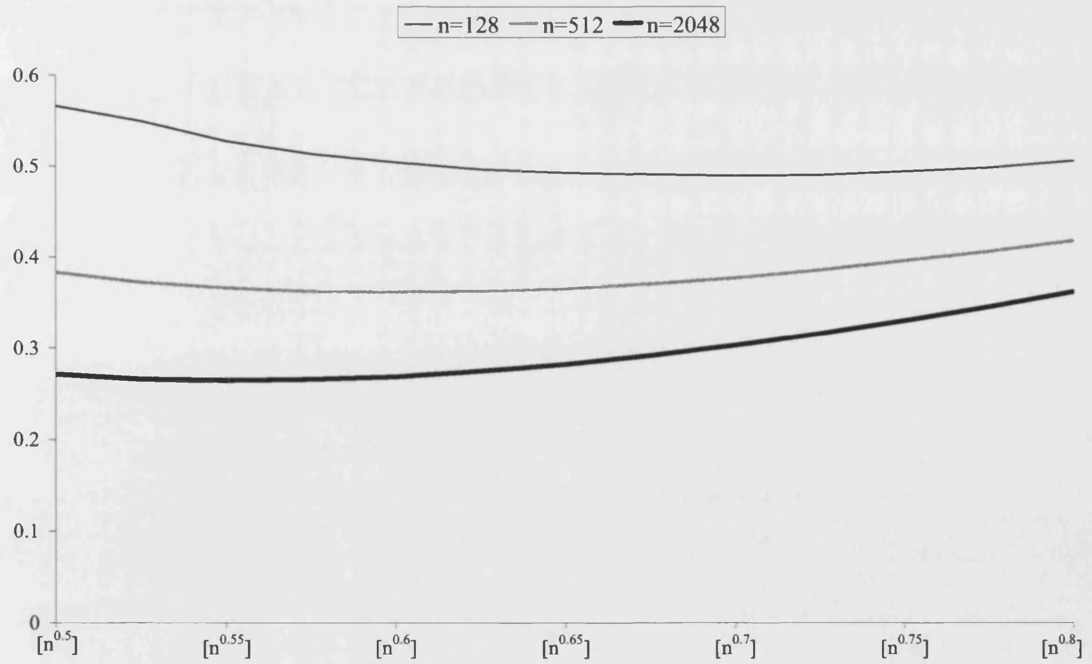


Figure 4.50: RMSE of LW estimator; LMSV model, absolute returns with $\alpha_{|r|} = 0.8$ and $\alpha_\xi = 0.8$.

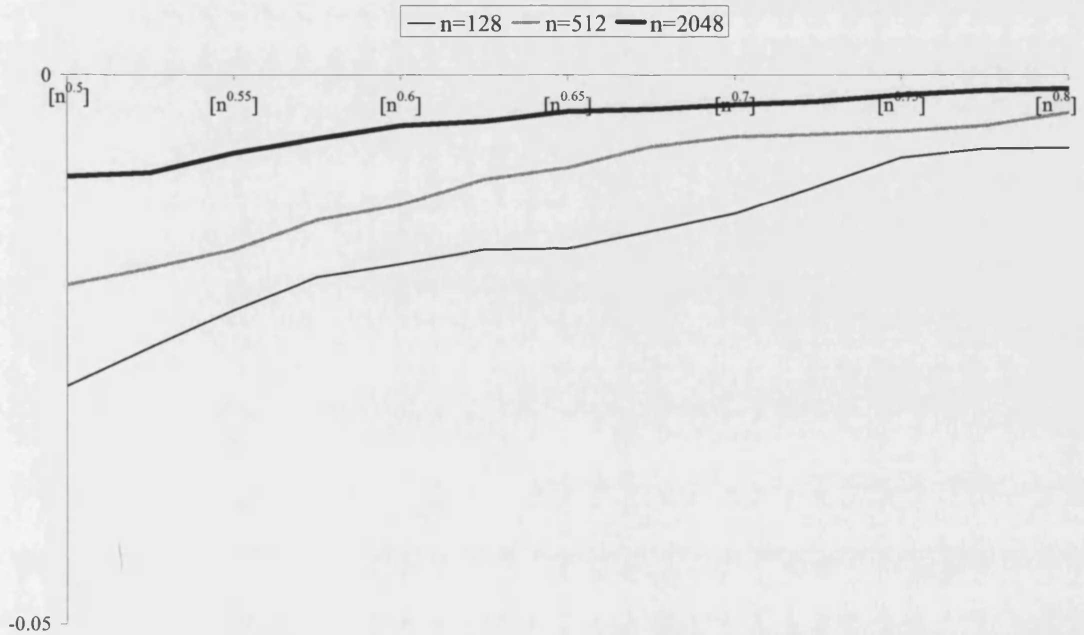


Figure 4.51: Bias of LW estimator; LMSV model, squared returns with $\alpha_{r^2} = 0$ and $\alpha_\xi = 0$.

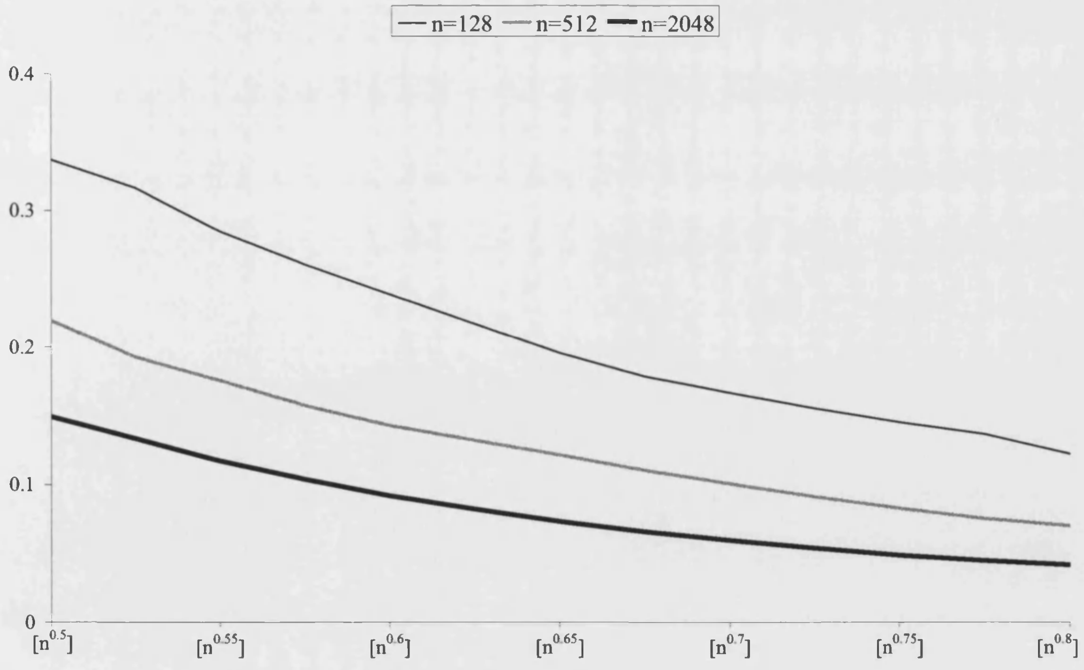


Figure 4.52: RMSE of LW estimator; LMSV model, squared returns with $\alpha_{r^2} = 0$ and $\alpha_\xi = 0$.

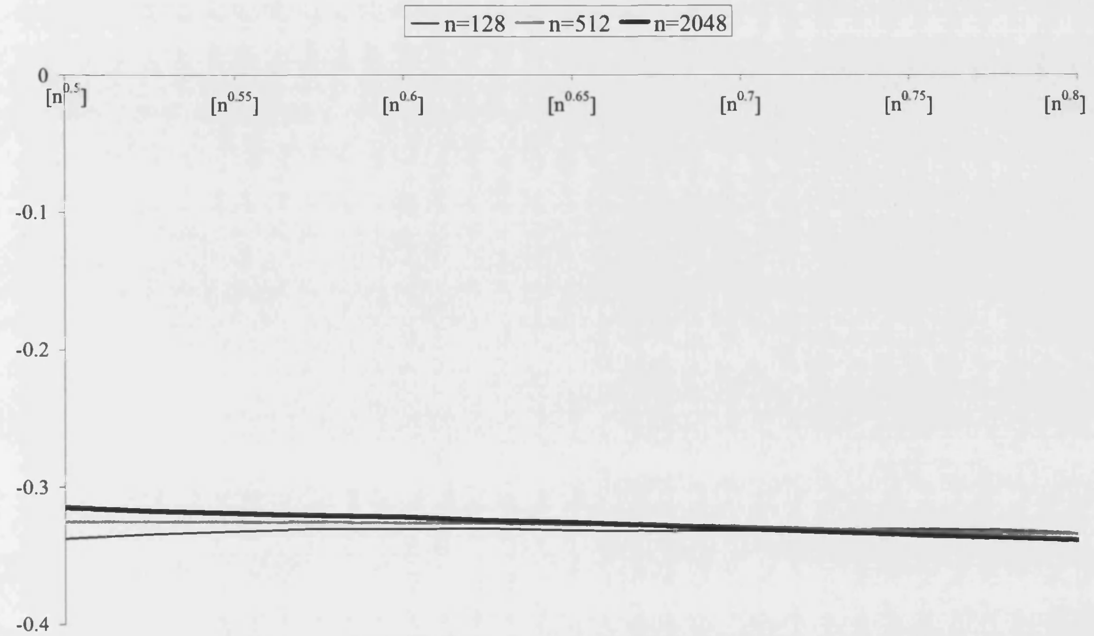


Figure 4.53: Bias of LW estimator; LMSV model, squared returns with $\alpha_{r^2} = 0.4$ and $\alpha_\xi = 0.4$.

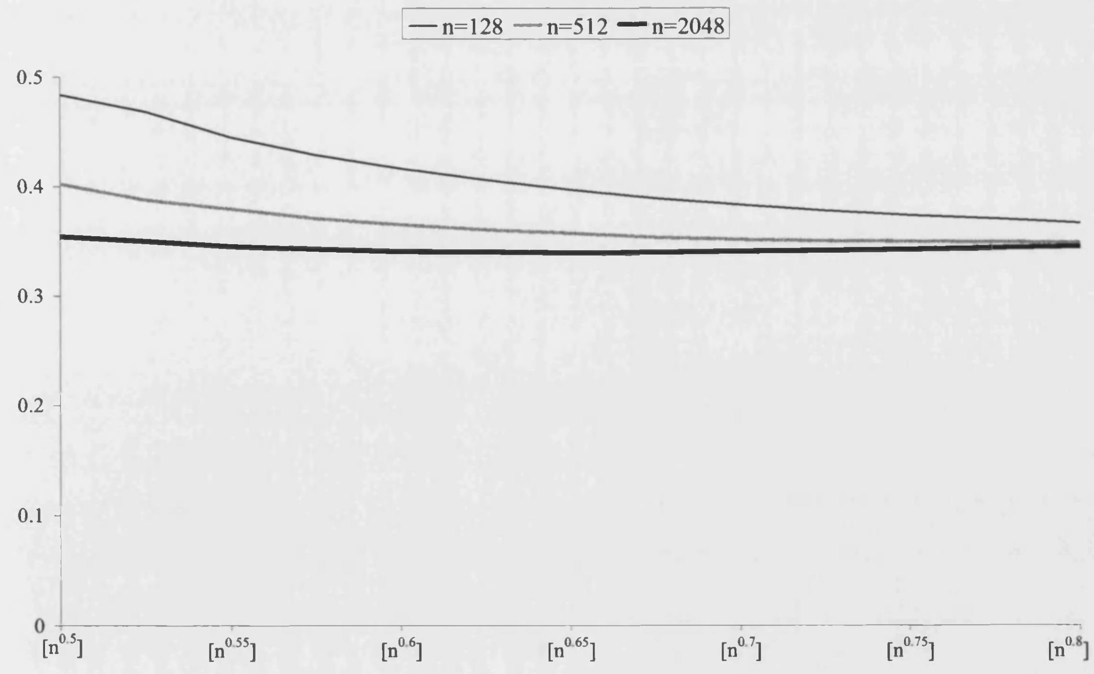


Figure 4.54: RMSE of LW estimator; LMSV model, squared returns with $\alpha_{r^2} = 0.4$ and $\alpha_\xi = 0.4$.

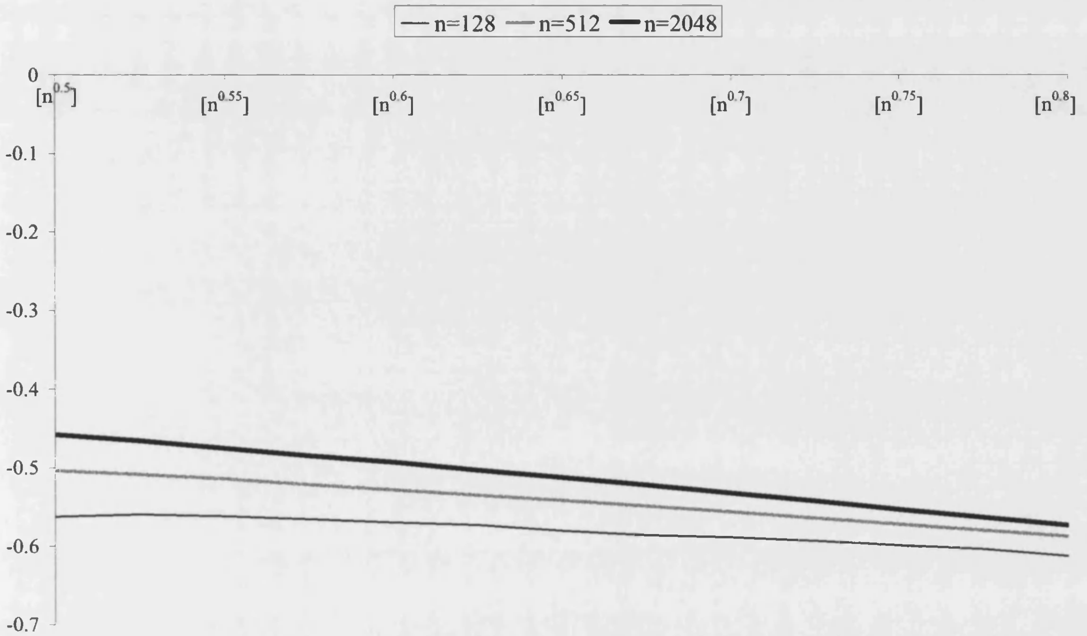


Figure 4.55: Bias of LW estimator; LMSV model, squared returns with $\alpha_{r^2} = 0.8$ and $\alpha_\xi = 0.8$.

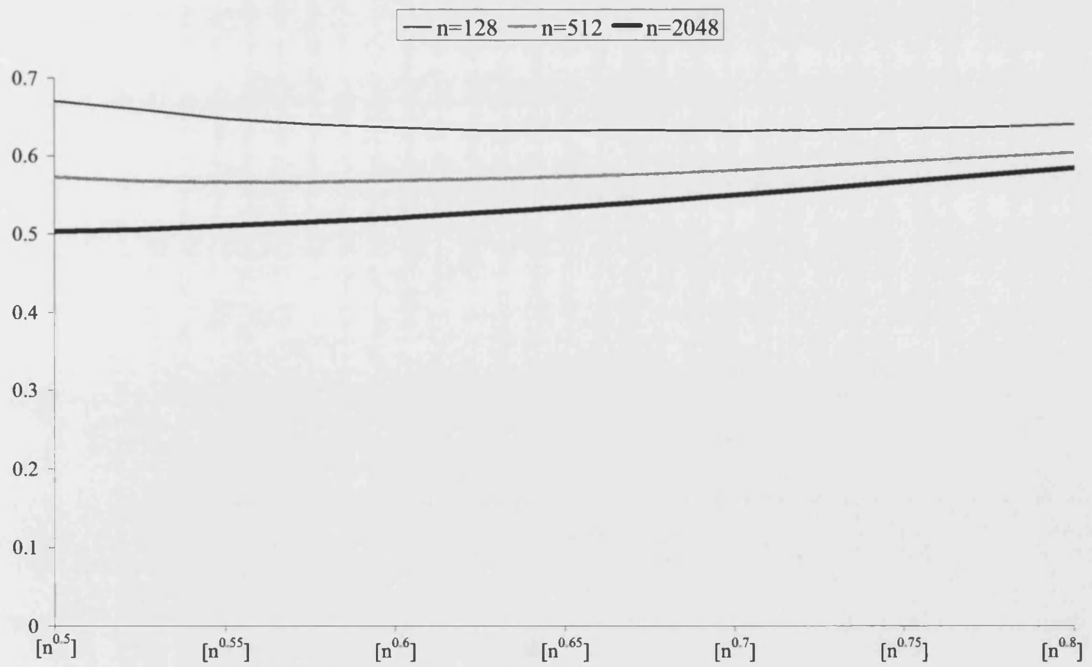


Figure 4.56: RMSE of LW estimator; LMSV model, squared returns with $\alpha_{r^2} = 0.8$ and $\alpha_\xi = 0.8$.

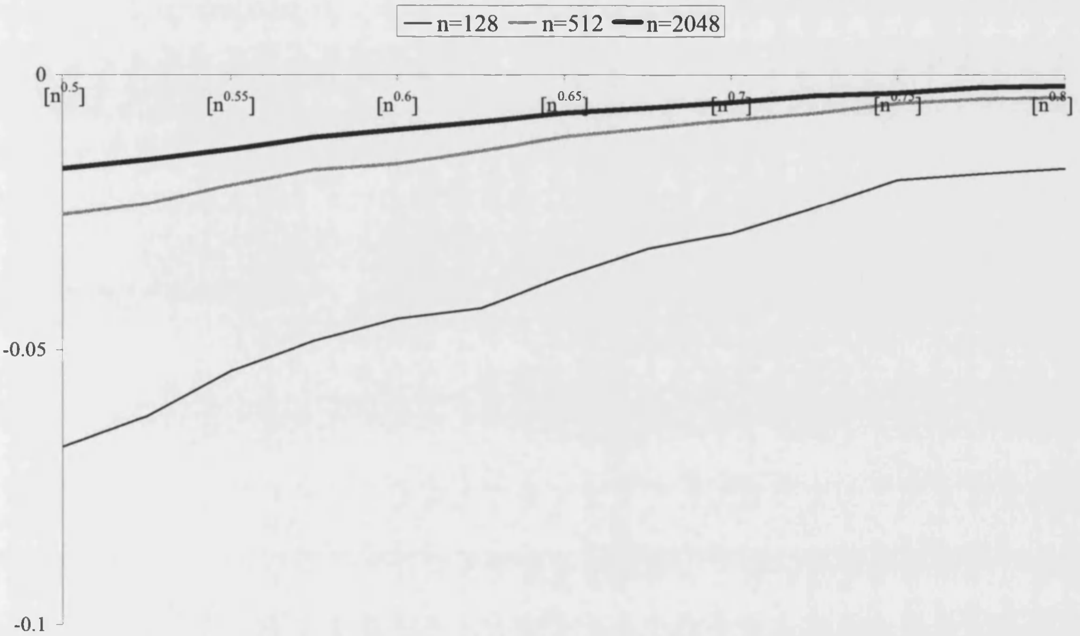


Figure 4.57: Bias of LW estimator; LMSV model, log-squared returns with $\alpha_{\log r^2} = 0$ and $\alpha_\xi = 0$.

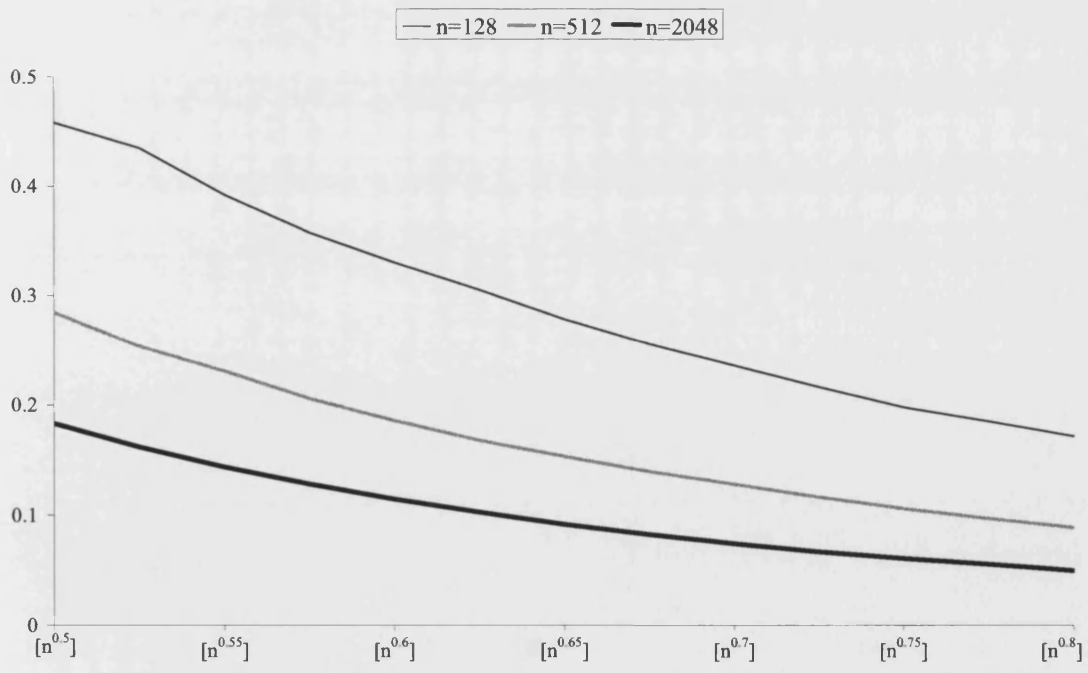


Figure 4.58: RMSE of LW estimator; LMSV model, log-squared returns with $\alpha_{\log r^2} = 0$ and $\alpha_\xi = 0$.

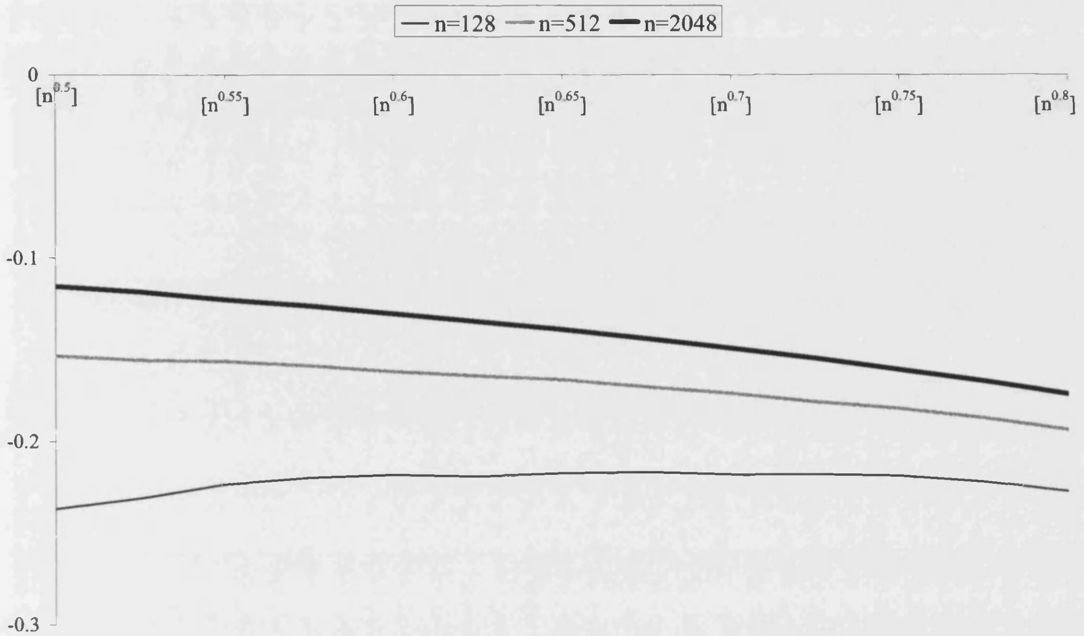


Figure 4.59: Bias of LW estimator; LMSV model, log-squared returns with $\alpha_{\log r^2} = 0.4$ and $\alpha_{\xi} = 0.4$.

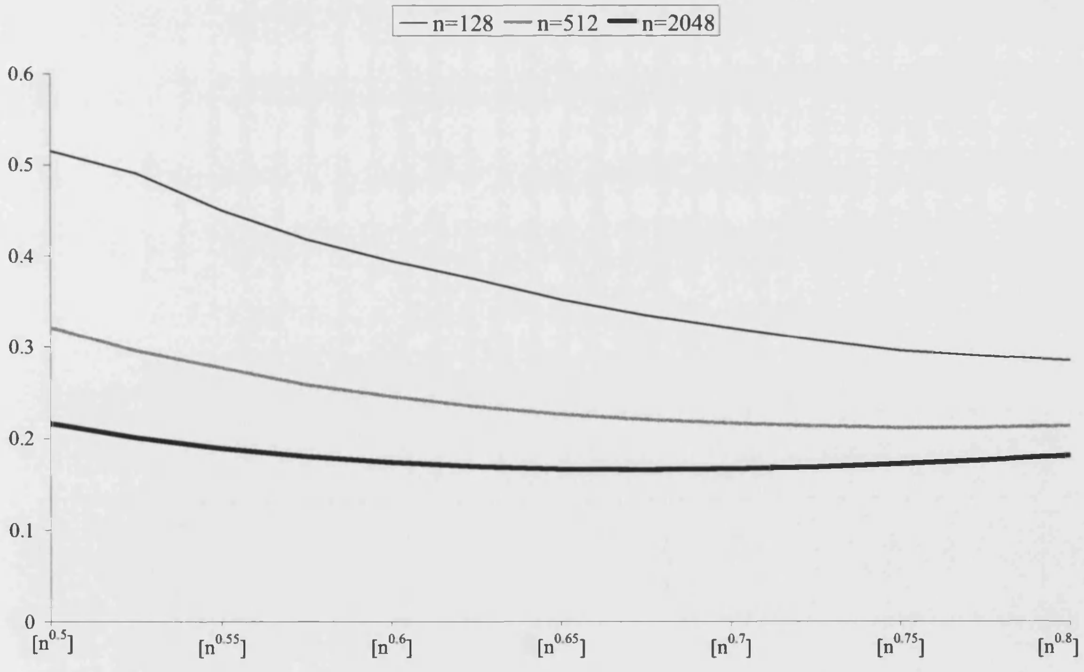


Figure 4.60: RMSE of LW estimator; LMSV model, log-squared returns with $\alpha_{\log r^2} = 0.4$ and $\alpha_{\xi} = 0.4$.

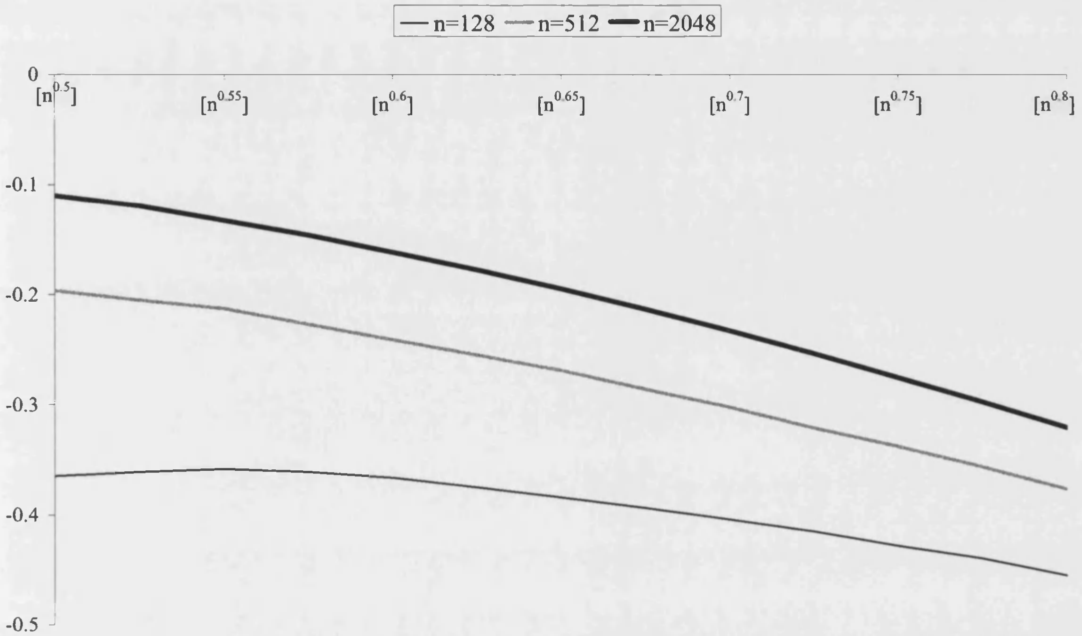


Figure 4.61: Bias of LW estimator; LMSV model, log-squared returns with $\alpha_{\log r^2} = 0.8$ and $\alpha_\xi = 0.8$.

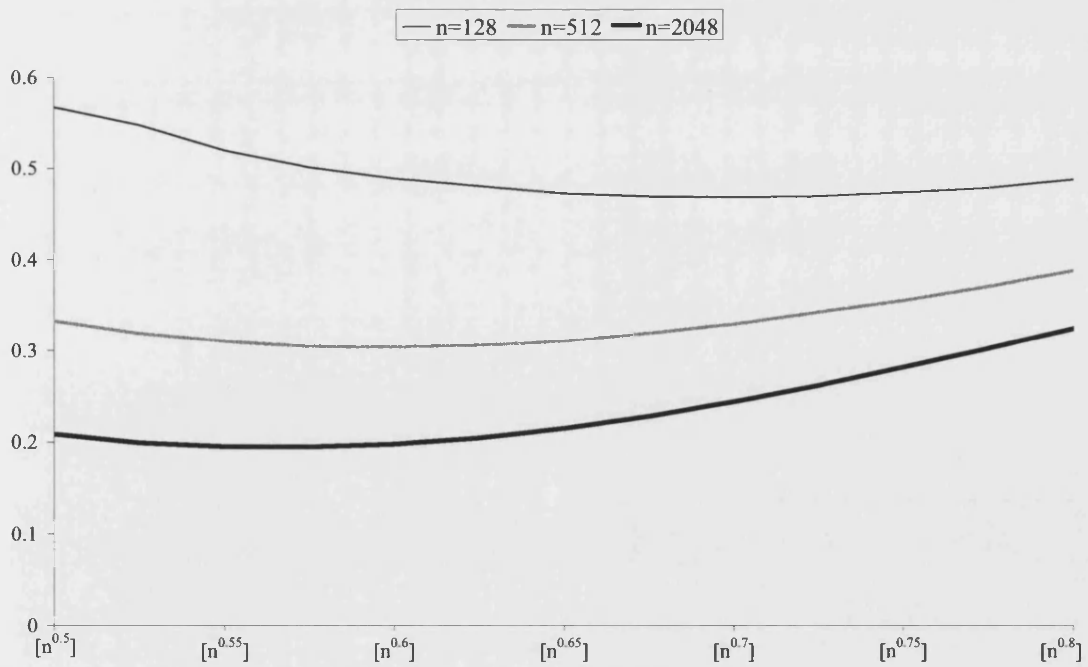


Figure 4.62: RMSE of LW estimator; LMSV model, log-squared returns with $\alpha_{\log r^2} = 0.8$ and $\alpha_\xi = 0.8$.

4.B Appendix

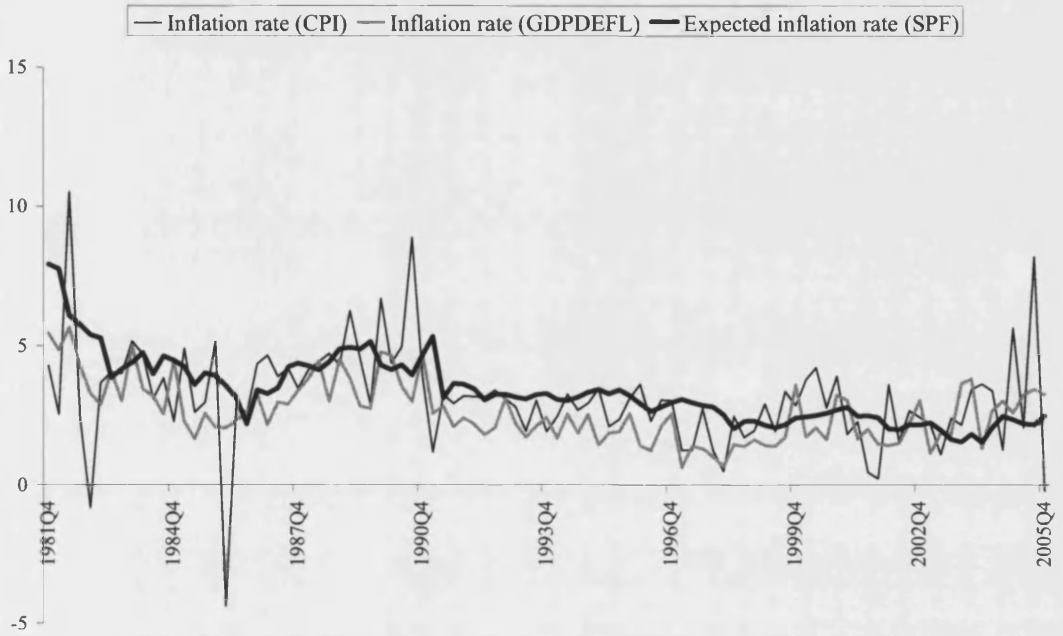


Figure 4.63: Data on inflation rate (CPI), inflation rate (GDPDEFL) and expected inflation rate (SPF) for the period 1981Q4-2005Q4.

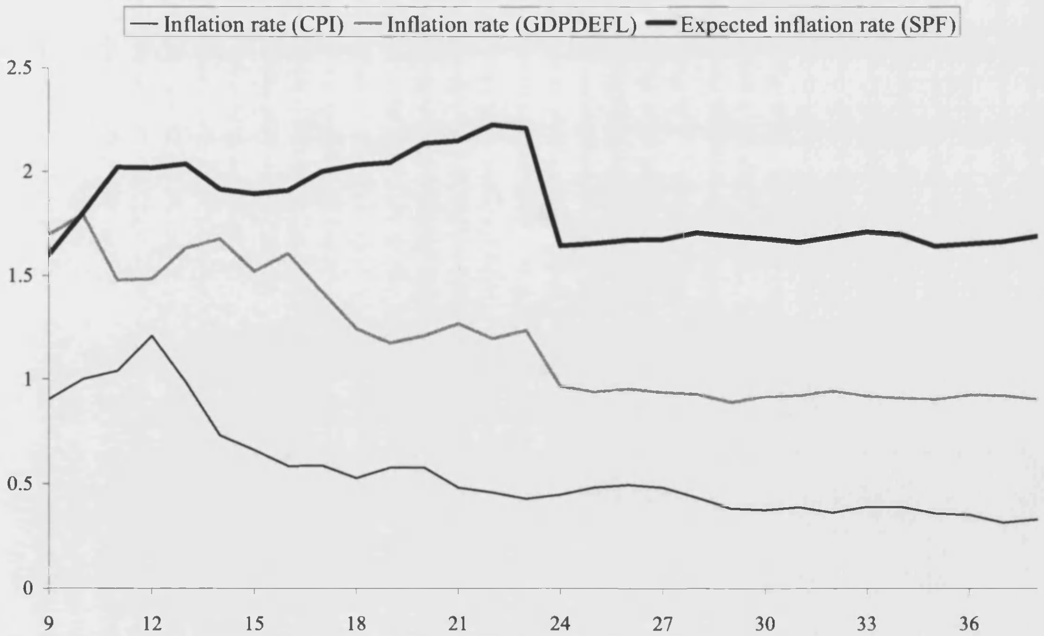


Figure 4.64: LW estimates for the data in Figure 4.63.

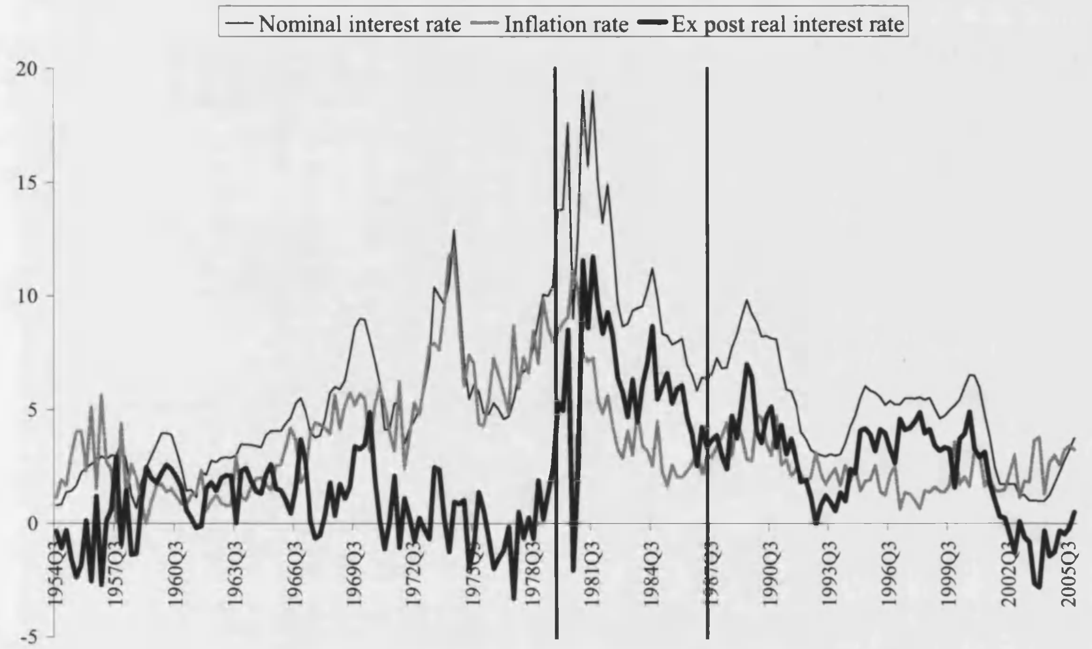


Figure 4.65: Data on nominal interest rate, inflation rate and ex post real interest for the period 1954Q3-2005Q4.

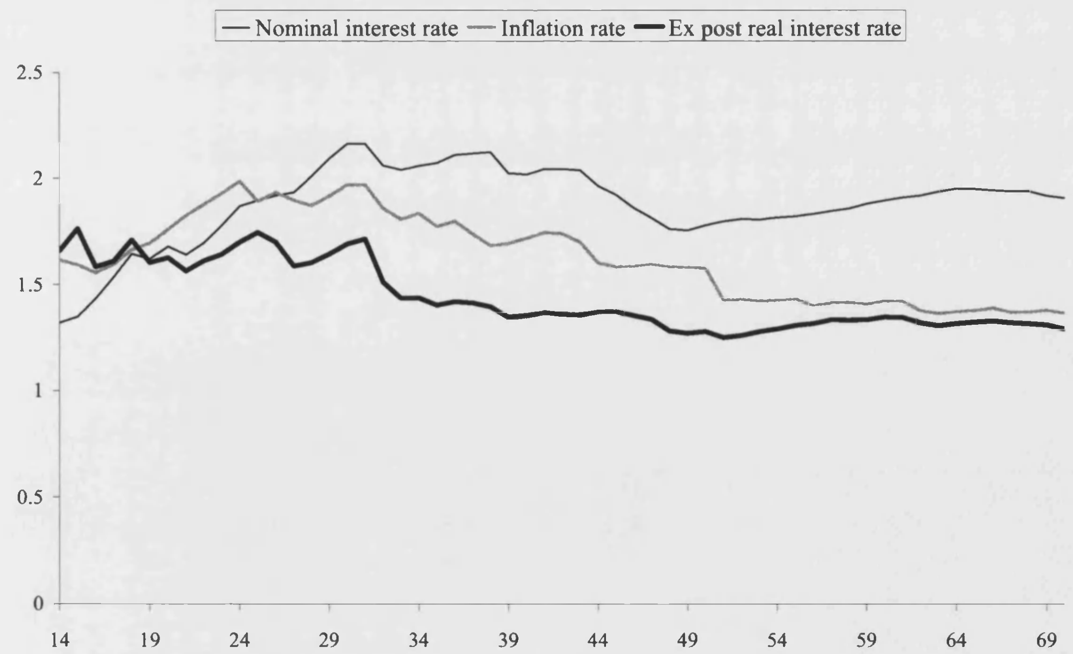


Figure 4.66: LW estimates for the data in Figure 4.65.

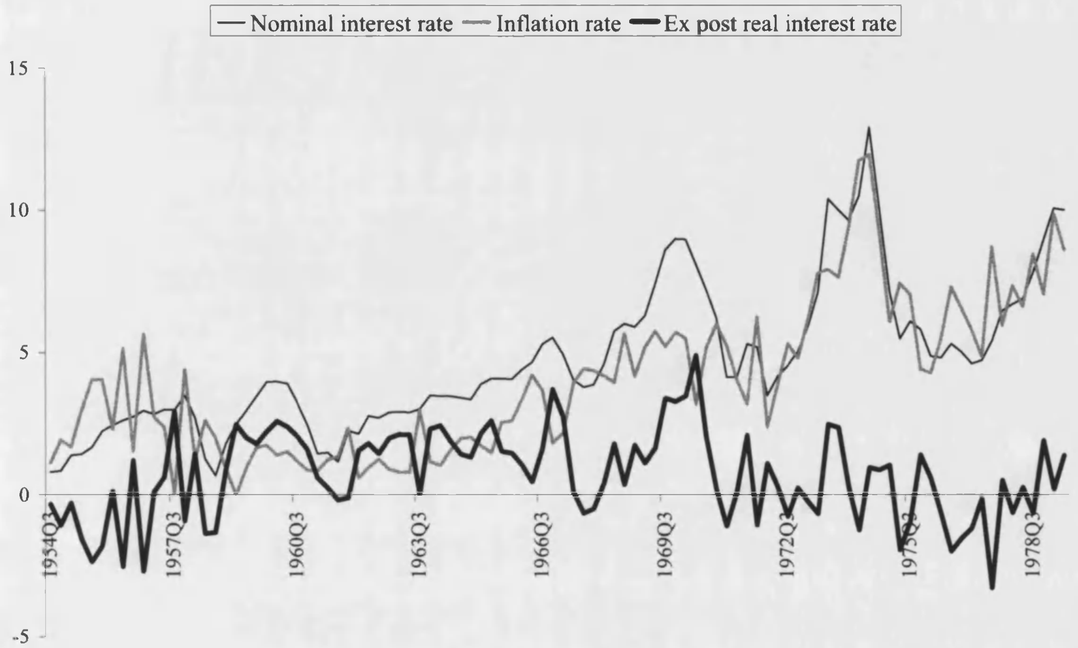


Figure 4.67: Data on nominal interest rate, inflation rate and ex post real interest for the period 1954Q3-1979Q2.

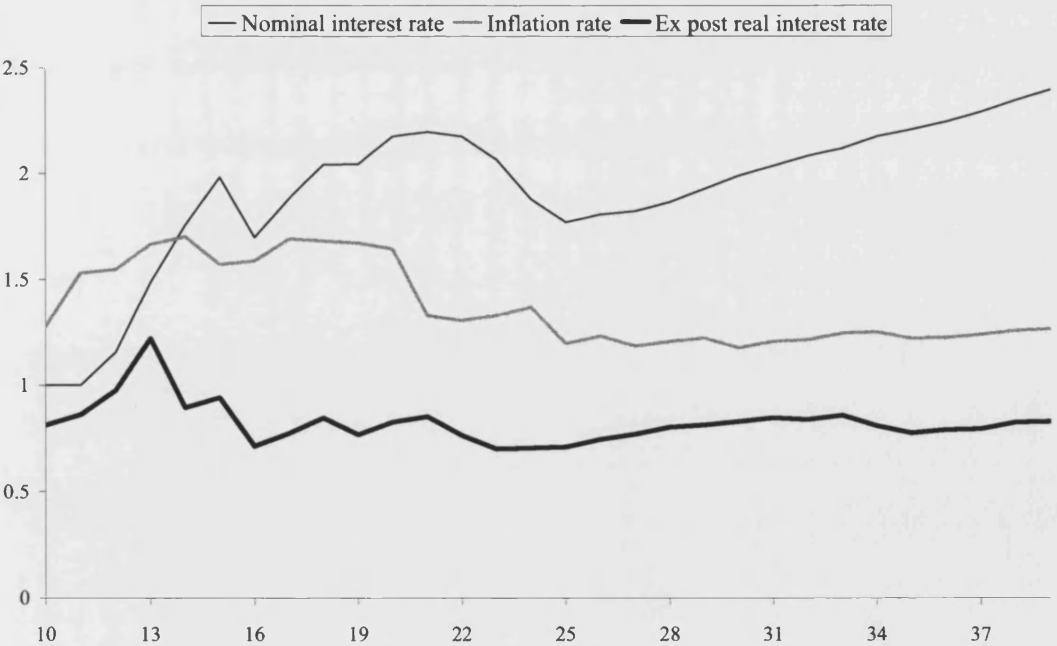


Figure 4.68: LW estimates for the data in Figure 4.67.

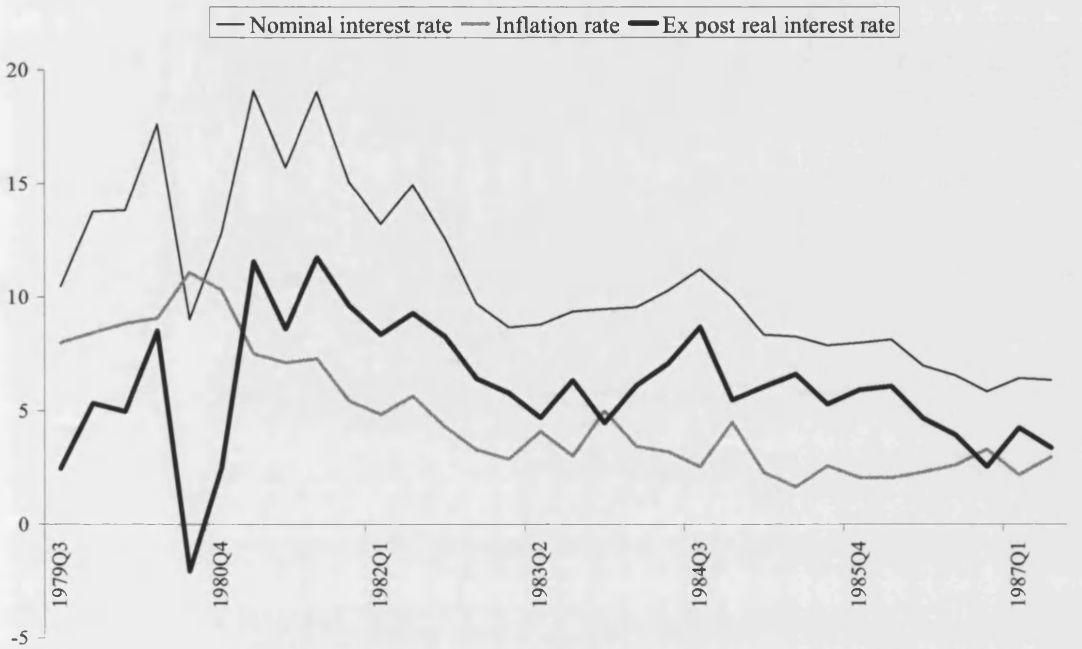


Figure 4.69: Data on nominal interest rate, inflation rate and ex post real interest for the period 1979Q3-1987Q2.

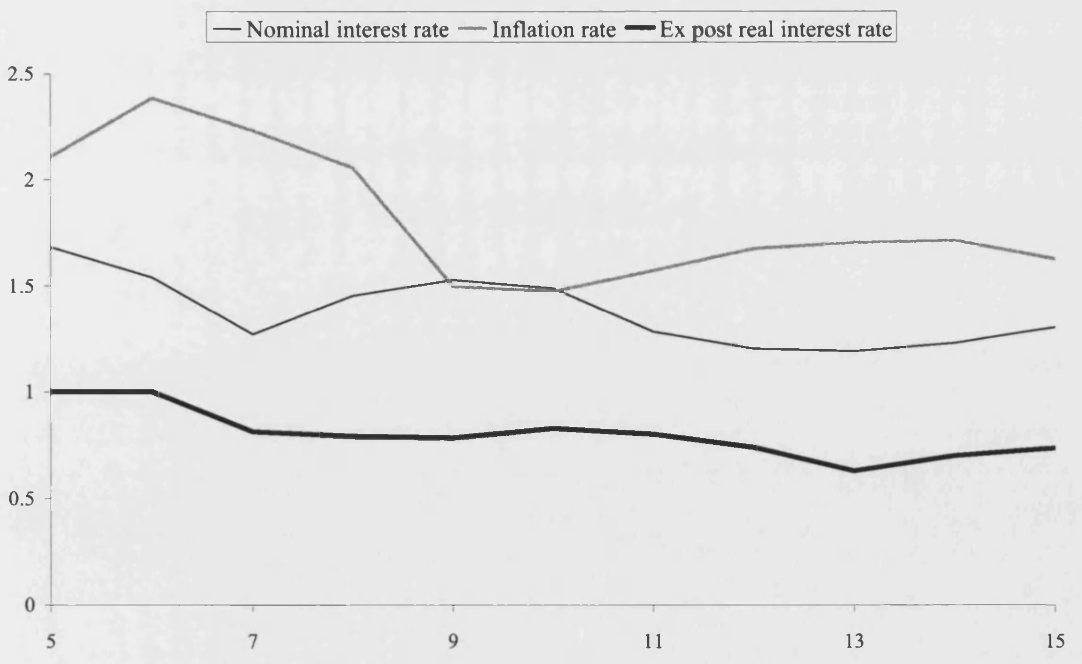


Figure 4.70: LW estimates for the data in Figure 4.69.

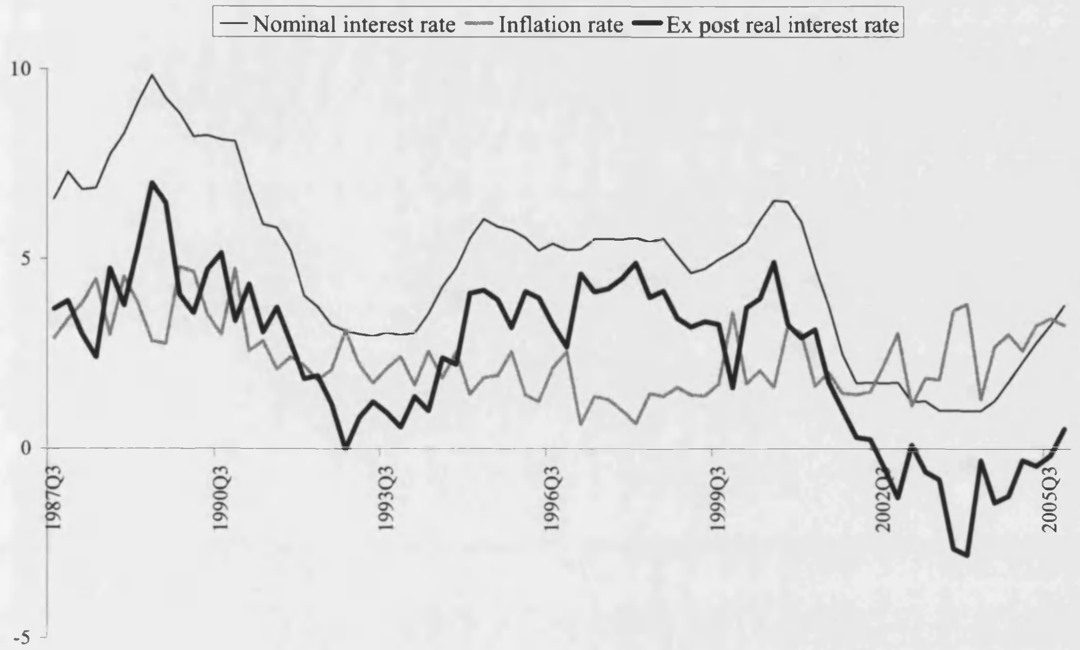


Figure 4.71: Data on nominal interest rate, inflation rate and ex post real interest for the period 1987Q3-2005Q4.

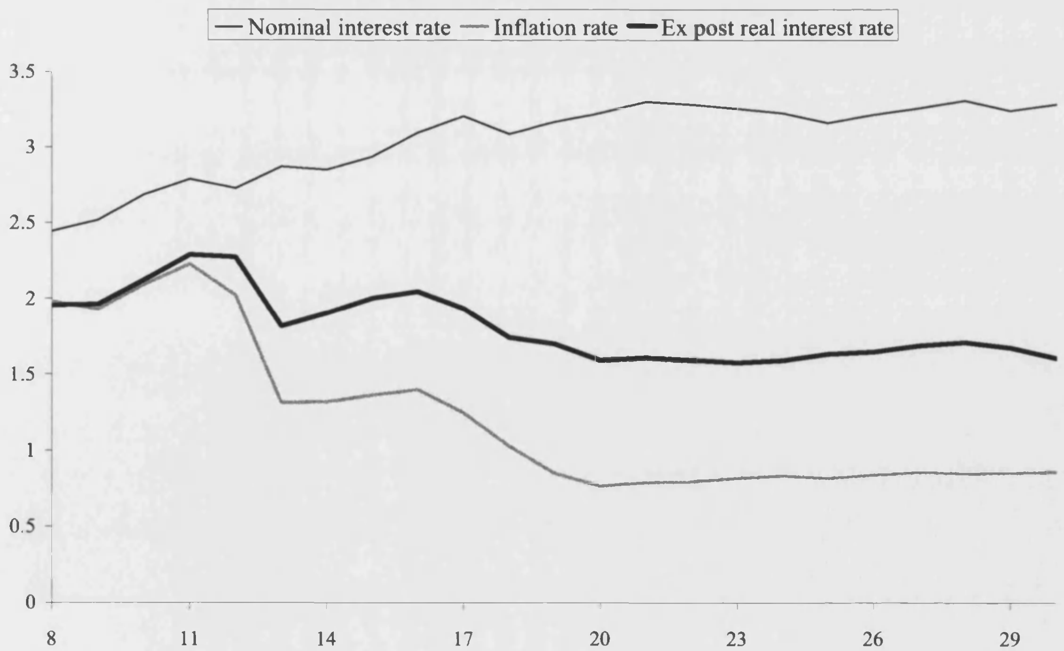


Figure 4.72: LW estimates for the data in Figure 4.71.

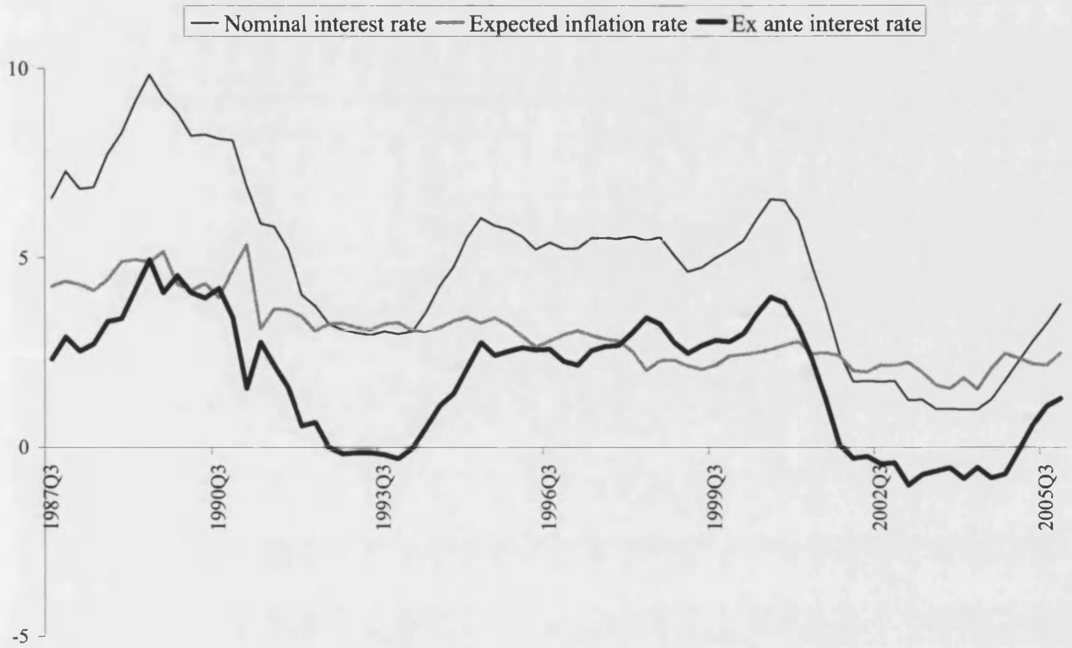


Figure 4.73: Data on nominal interest rate, expected inflation rate and ex ante real interest for the period 1987Q3-2005Q4.

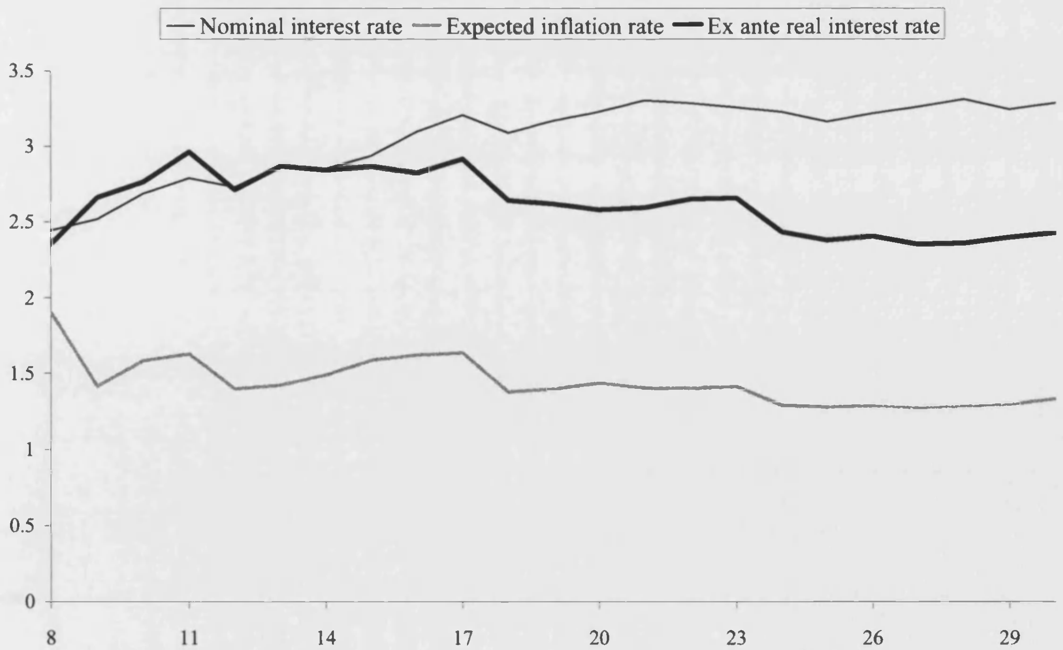


Figure 4.74: LW estimates for the data in Figure 4.73.

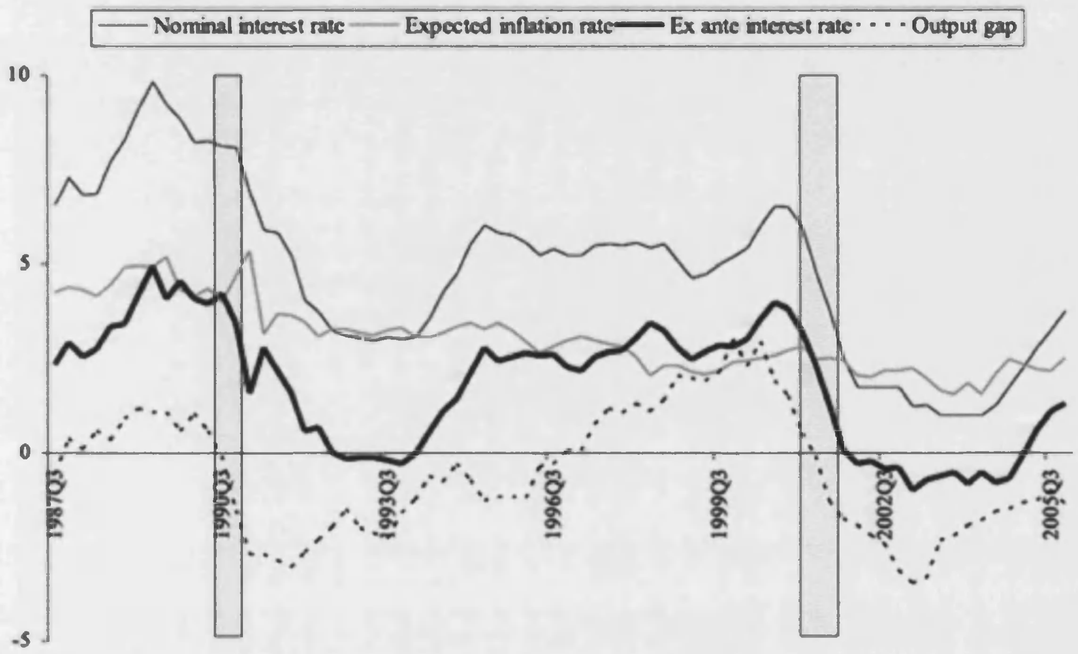


Figure 4.75: Data on nominal interest rate, expected inflation rate, ex ante real interest and output gap for the period 1987Q3-2005Q4.

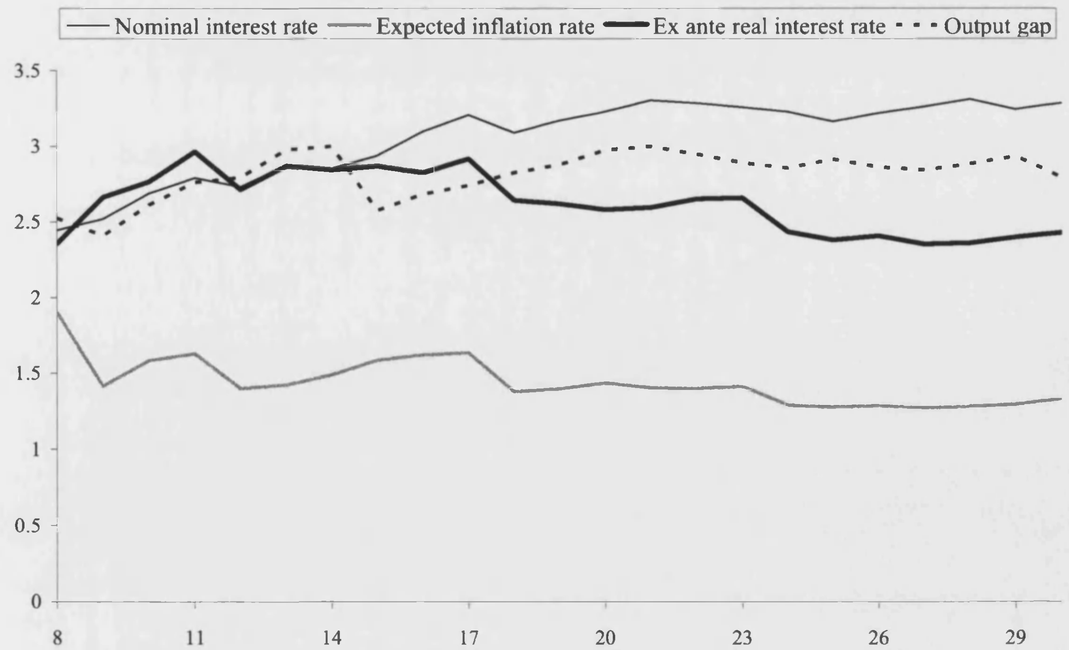


Figure 4.76: LW estimates for the data in Figure 4.75.

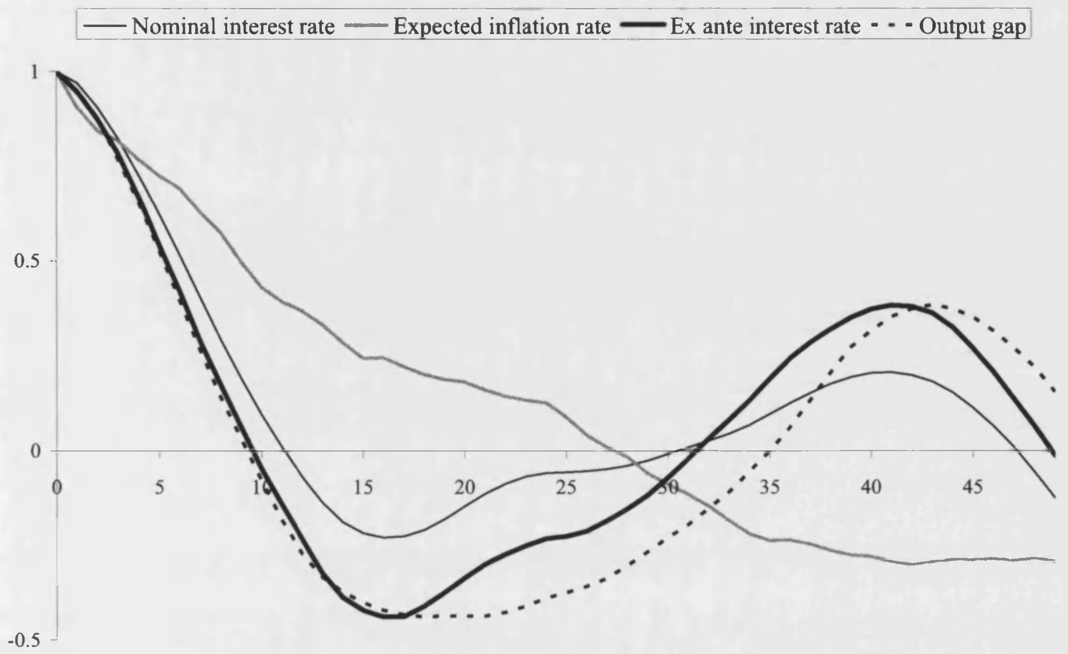


Figure 4.77: Sample autocorrelation function of nominal interest rate, expected inflation rate, ex ante real interest and output gap for the period 1987Q3-2005Q4.

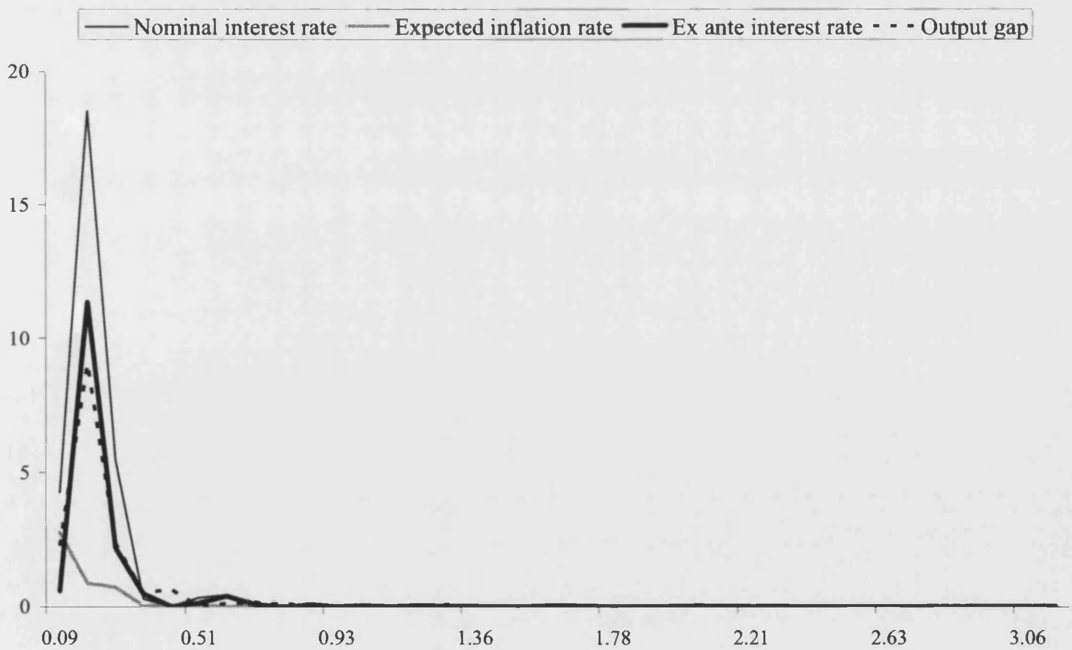


Figure 4.78: Periodogram of nominal interest rate, expected inflation rate, ex ante real interest and output gap for the period 1987Q3-2005Q4.

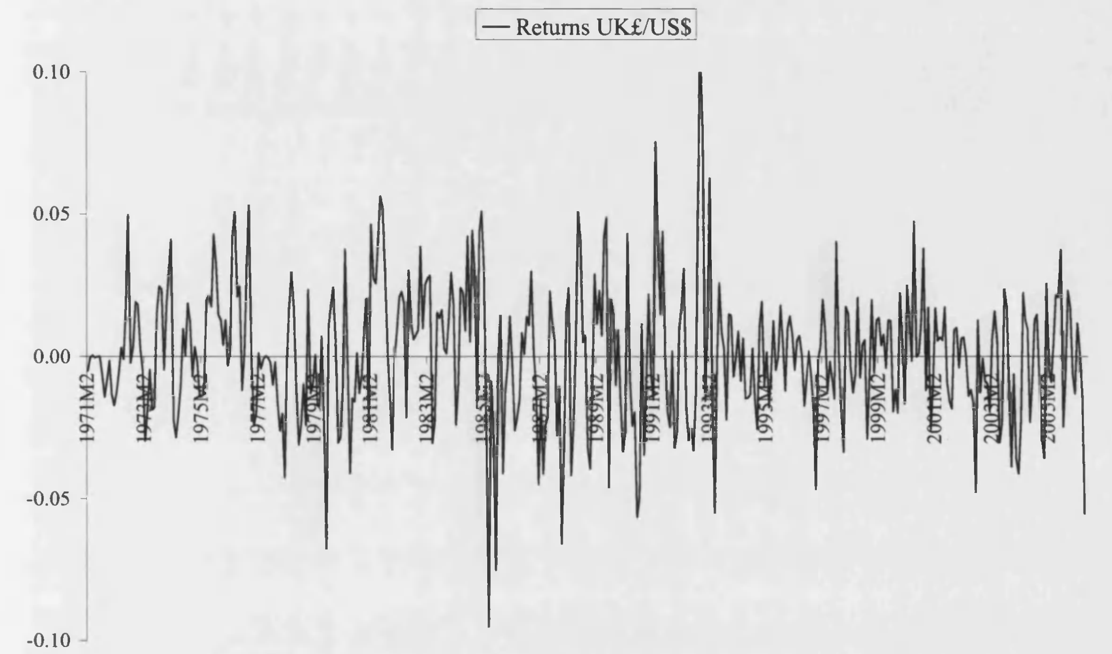
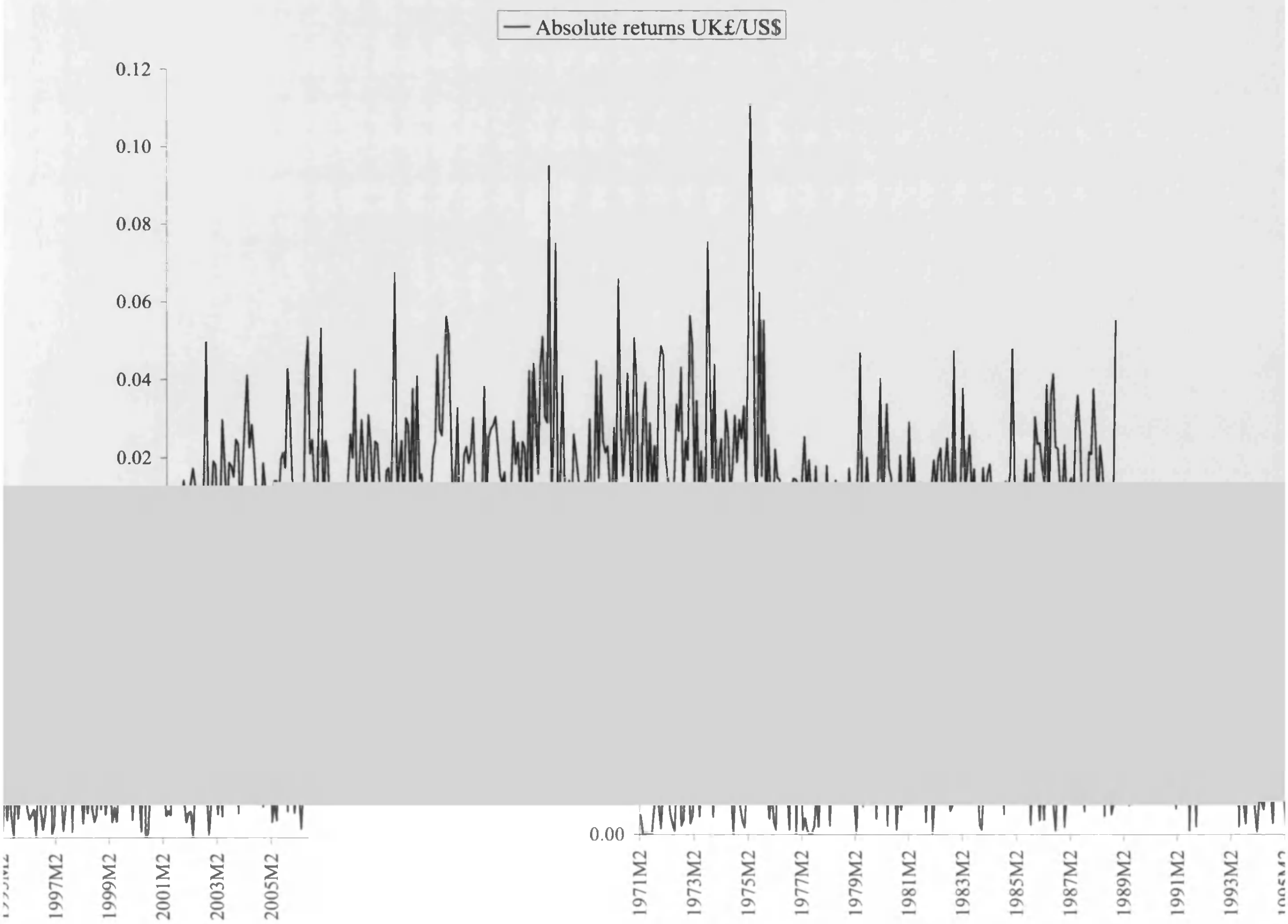


Figure 4.79: Data on UK£/US$ foreign exchange rate returns r_t for the period 1971M2-2006M5.



psolute returns $|r_t|$ for the

Figure 4.80: Data on UK£/US$ foreign exchange rate al

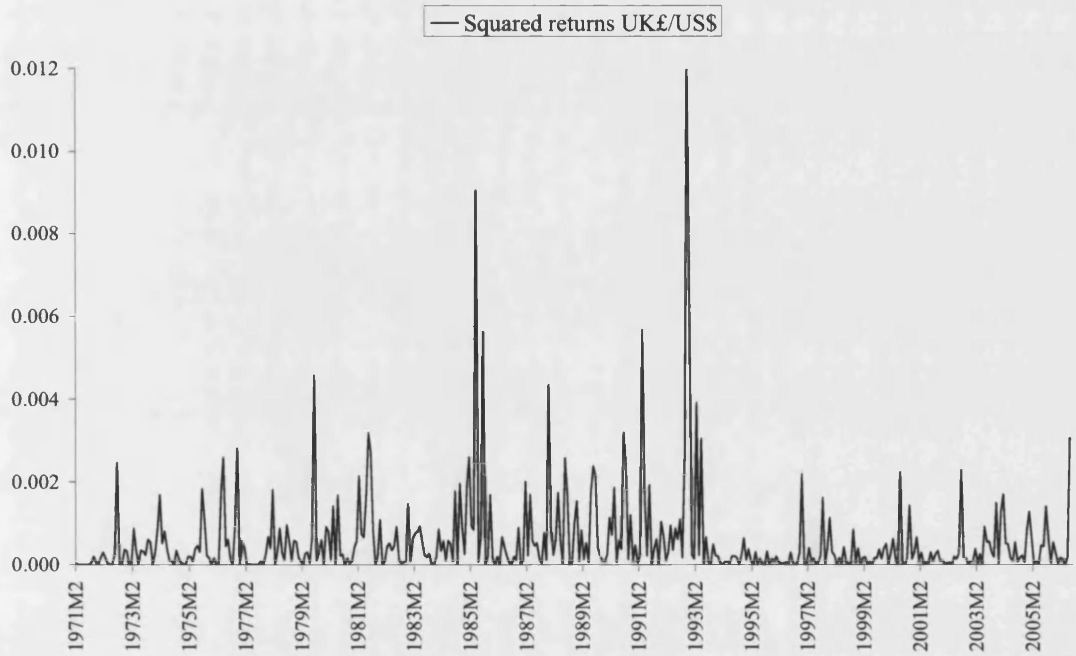
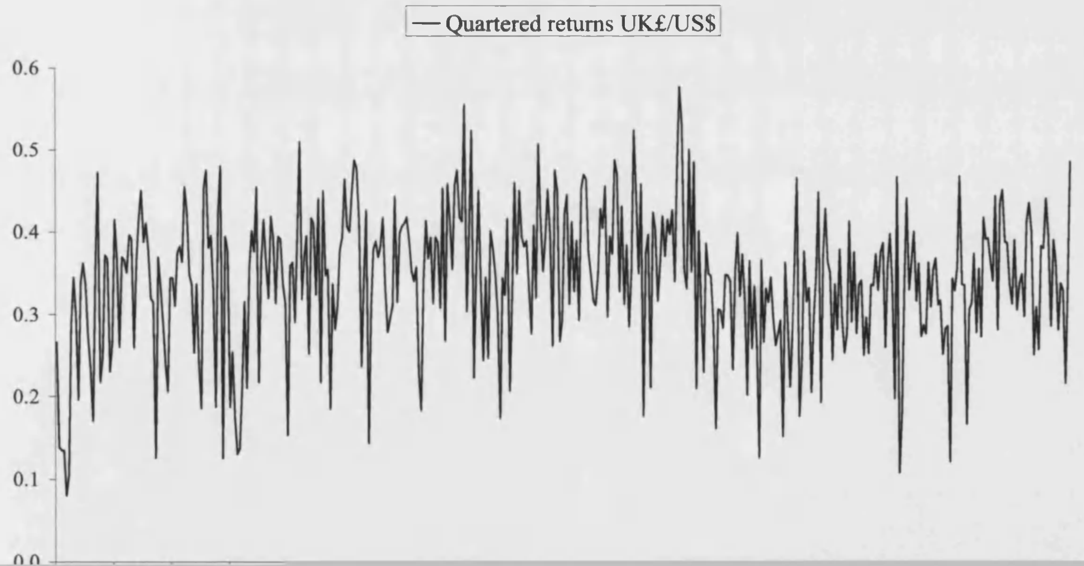


Figure 4.81: Data on UK£/US$ foreign exchange rate squared returns r_t^2 for the period 1971M2-2006M5.



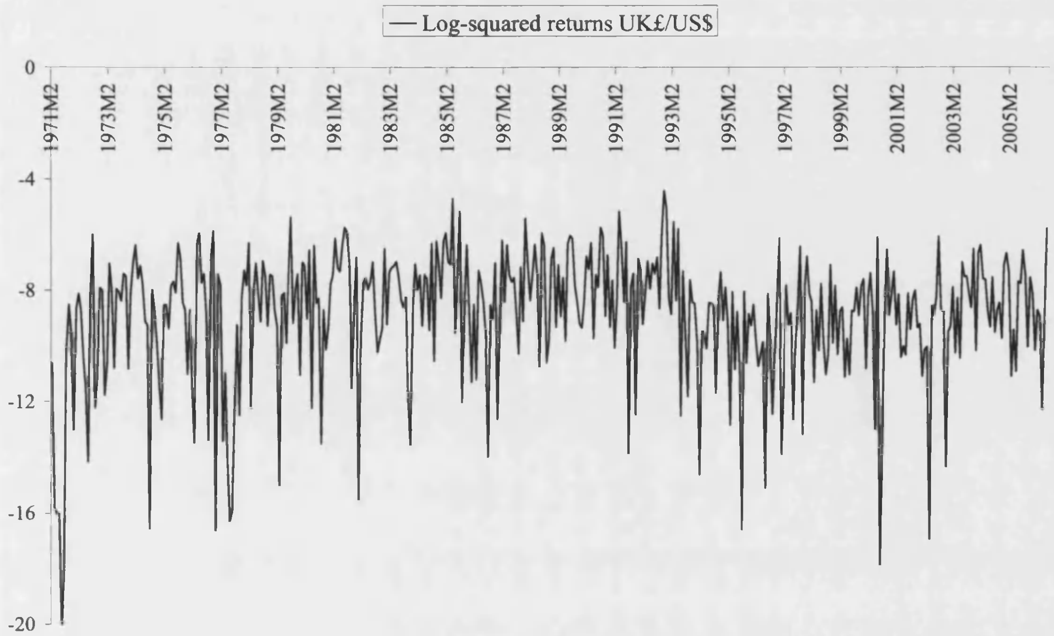


Figure 4.83: Data on UK£/US$ foreign exchange rate log-squared returns $\log r_t^2$ for the period 1971M2-2006M5.

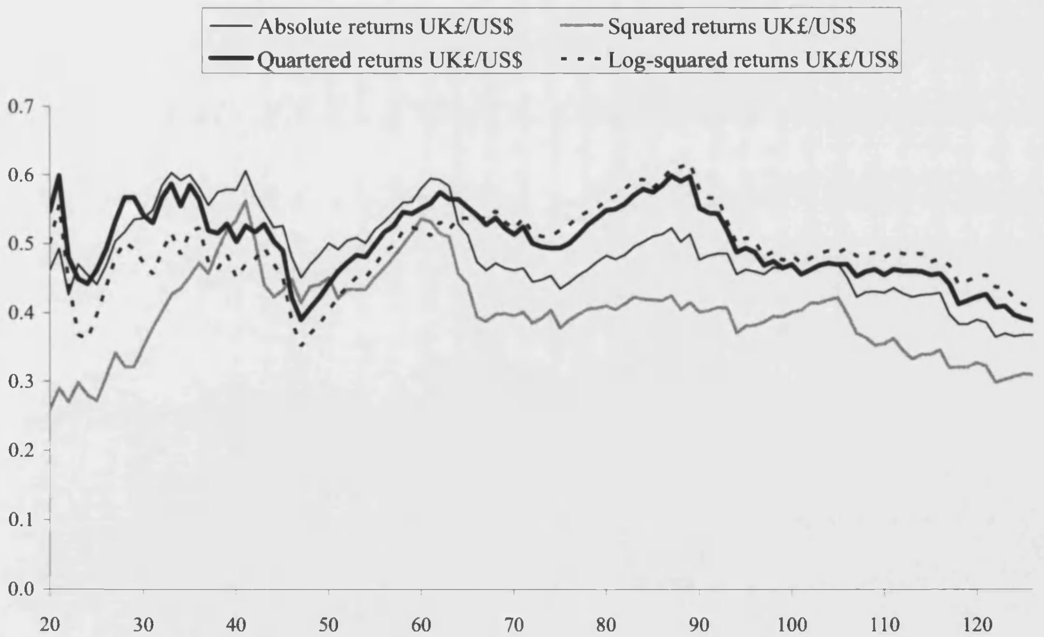


Figure 4.84: LW for UK£/US$ foreign exchange rate absolute returns $|r_t|$, squared returns r_t^2, quartered returns $|r_t|^{\frac{1}{4}}$ and log-squared returns $\log r_t^2$ for the period 1971M2-2006M5.

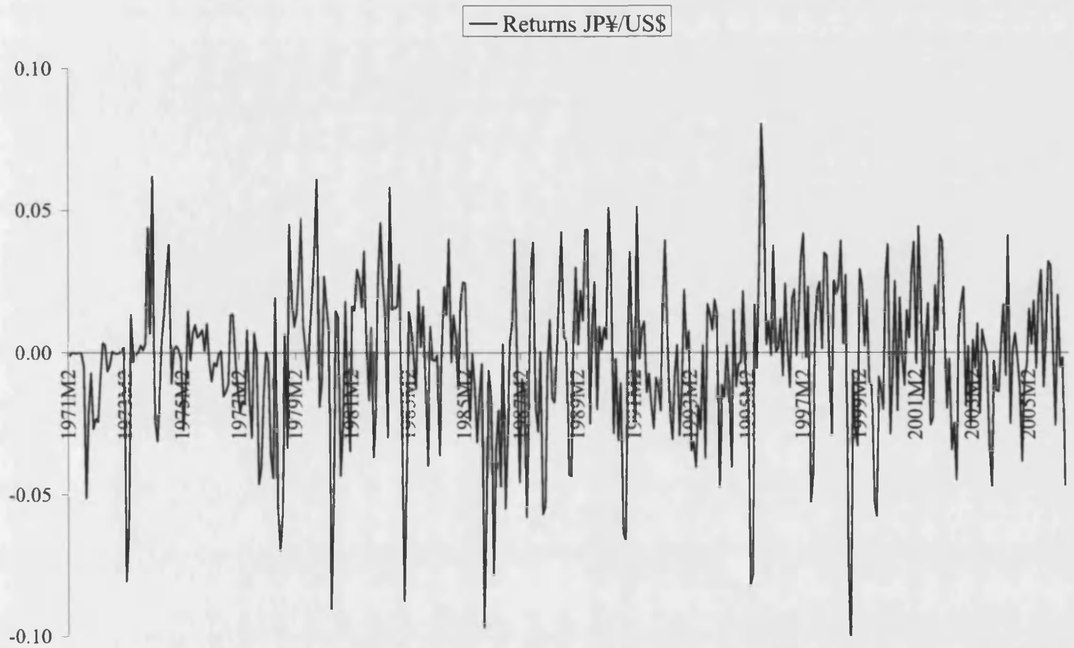


Figure 4.85: Data on JP¥/US$ foreign exchange rate returns r_t for the period 1971M2-2006M5.

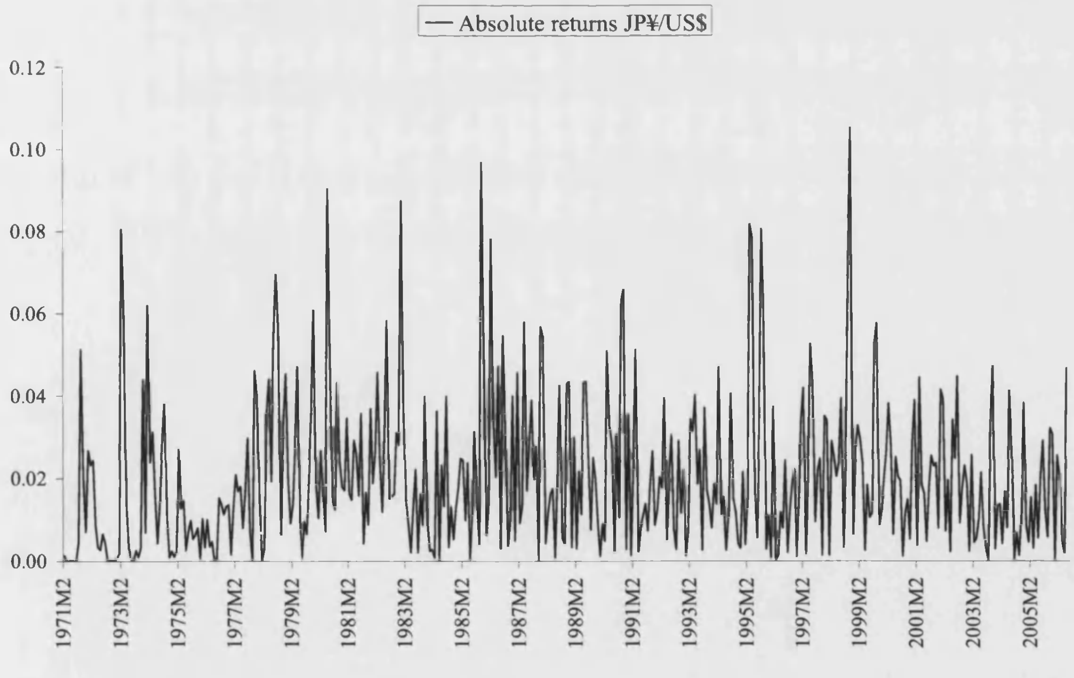


Figure 4.86: Data on JP¥/US$ foreign exchange rate absolute returns $|r_t|$ for the period 1971M2-2006M5.

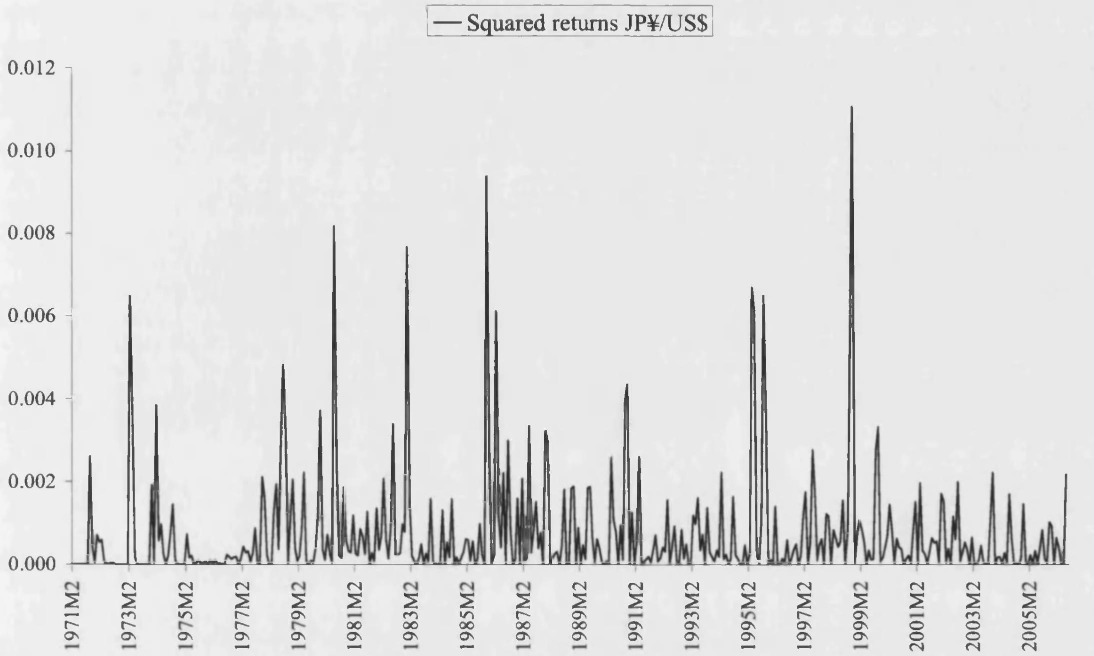
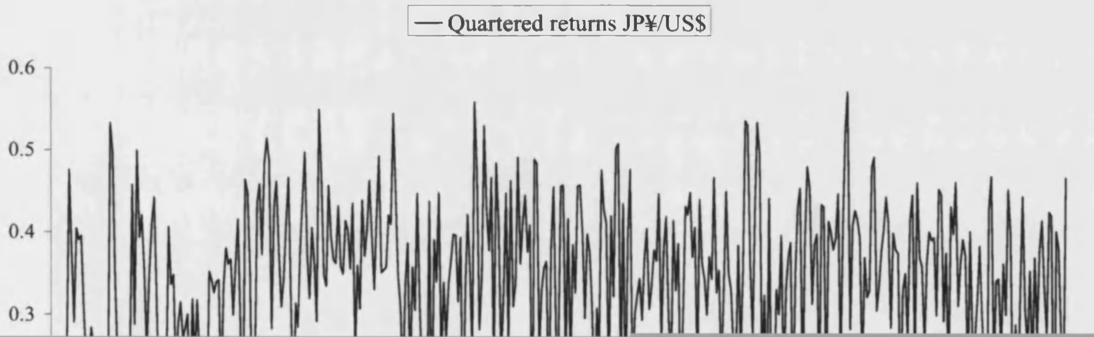


Figure 4.87: Data on JP¥/US$ foreign exchange rate squared returns r_t^2 for the period 1971M2-2006M5.



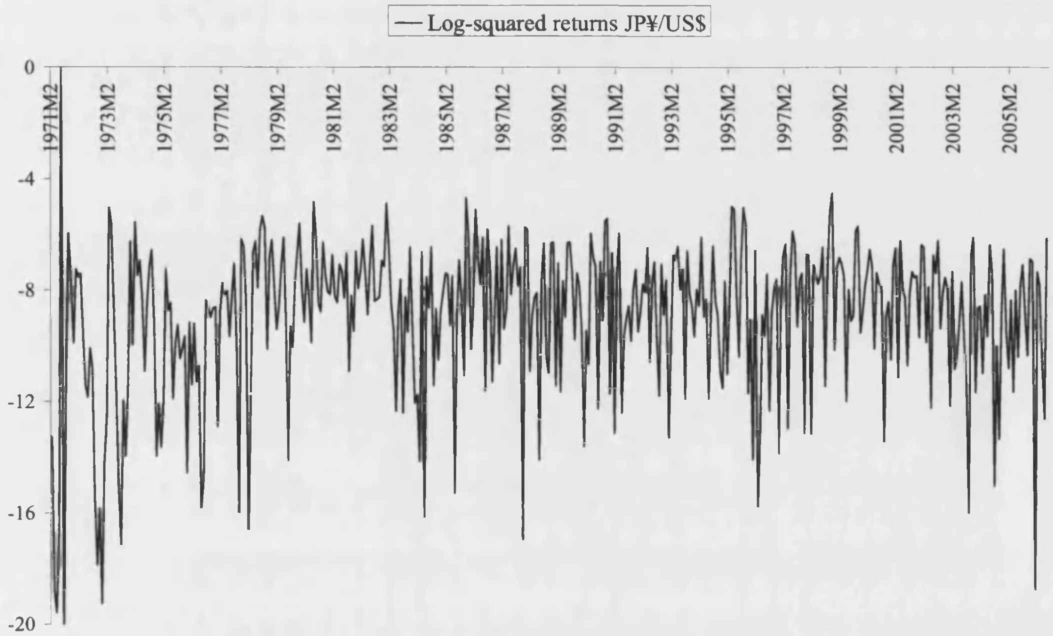


Figure 4.89: Data on JP¥/US$ foreign exchange rate log-squared returns $\log r_t^2$ for the period 1971M2-2006M5.

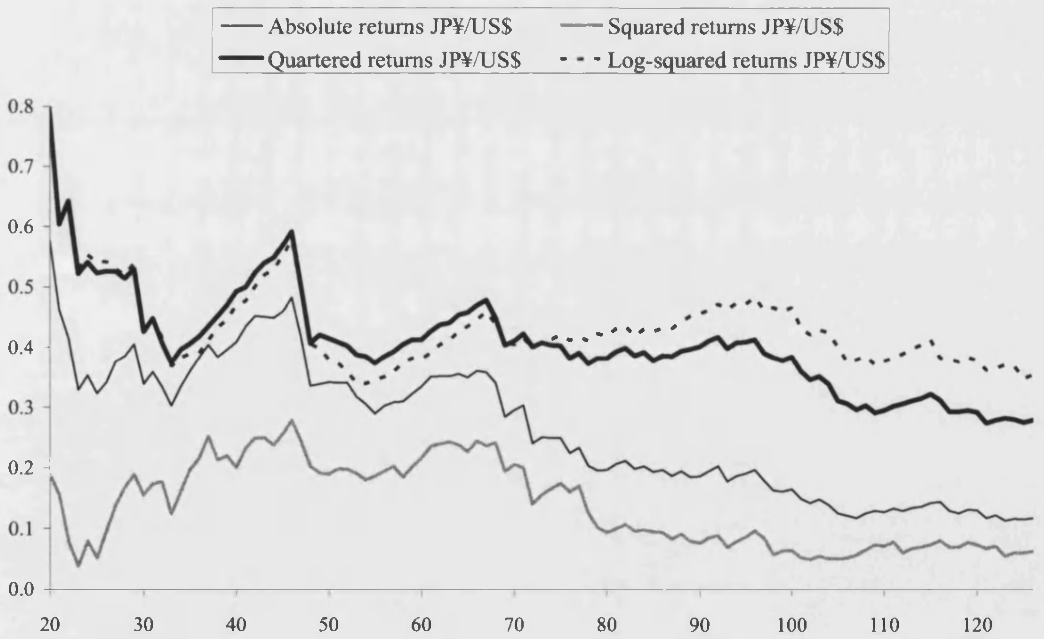


Figure 4.90: LW for JP¥/US$ foreign exchange rate absolute returns $|r_t|$, squared returns r_t^2, quartered returns $|r_t|^{\frac{1}{4}}$ and log-squared returns $\log r_t^2$ for the period 1971M2-2006M5.

Chapter 5

Parametric bootstrap tests for weak persistence

5.1 Introduction

As was already discussed in the Introduction of this thesis, parametric estimation and inference on the memory and cyclical memory parameters α_x and $\alpha_{\omega,x}$ are dominated by the parametric Whittle method. The asymptotic properties of the PW estimator are well established for Gaussian or linear processes having a rather general parametrically specified spectral density function. In the case that the frequency of the possibly persistence component is known, see Hannan (1973), Fox and Taqqu (1986), Dahlhaus (1989), Giraitis and Surgailis (1990), Hosoya (1997) and, Velasco and Robinson (2000), the asymptotic distribution of the PW estimator can then be used to test for values of α_x or $\alpha_{\omega,x}$ in $(-1, 1)$. When the process exhibits long or cyclical long memory and the frequency of the persistent component is unknown, see Giraitis, Hidalgo, and Robinson (2001), the asymptotic distribution of the PW estimator can be employed to test for values of α_x or $\alpha_{\omega,x}$ in $(0, 1)$.

However, it is not possible to test whether the data does not exhibit a persistent component against the alternative that it does, when the practitioner is not certain about the frequency of the possible persistent component. The latter situation is likely to arise in practice, when the practitioner deals with macroeconomic and financial data sets that are subject to business cycles fluctuations. Notice that business cycle chronology is not available for all countries, and even when it is, the dating of the peaks or troughs of the business cycle are usually reported few months after they have taken place, and are calculated based on specific measures of economic activity, so that the business cycle component of a certain series does not necessarily have a frequency matching that from the business cycle chronology.

The main objective of this chapter is to provide testing procedures for the hypothesis that the data does not exhibit a persistent component against the alternative that it does, noticing also that these hypotheses can be reinterpreted as the data being weakly dependent against being strongly dependent. We describe two different tests; the first one is a Wald (W) type of test, whereas the second one is based on the Lagrange multiplier (LM) principle. The tests are based on whether the supremum of a sequence of random variables is significantly greater than zero. In particular, the W and the LM types of tests are based on the supremum of a sequence of PW estimators and the score function, respectively, when it is believed that the frequency of the possibly persistent component falls in $[0, \pi]$. Notice that the W type test presented here relies on the same principle as that of Hidalgo (2006) in the semiparametric context. As a next step, we establish the asymptotic distribution of these test statistics, as well as examine their consistency and power properties.

However, the asymptotic distribution of our test statistics is nonstandard and therefore difficult to implement in practice for the purposes of calculating critical values for our hypothesis testing. Taking also into account that the tests are based on the supremum of a sequence of random variables, one might expect that the rate of convergence of the finite sample distribution to the asymptotic one to be slow, see Hall (1979) who established that the rate is logarithmic for a related statistic. Hence, the second objective of this chapter is to propose and examine a bootstrap approach to our hypothesis testing. Bootstrap schemes, introduced by Efron (1979), have become a routine method for approximating the distribution of a statistical quantity, partly due to the increasing computational power. At a theoretical level, bootstrap algorithms have attracted considerable effort to their development for two reasons. Firstly, they are capable of approximating the finite sample distribution of statistics more effectively than those based on their asymptotic counterparts. Secondly, they allow computation of valid asymptotic quantiles of the limiting distribution in nonstandard situations, and in particular, when the limiting distribution is unknown or if known, the practitioner is unable to compute its quantiles.

The remainder of this chapter is as follows. In Section 5.2, we formulate our hypotheses of interest and introduce the W and LM type of tests. In Section 5.3, we present and discuss the assumptions, while Section 5.4 deals with the theoretical properties of the test statistics. In Section 5.5 we present a bootstrap algorithm and establish its validity. Section 5.6 contains Monte-Carlo simulations, while Section 5.7 applies our testing procedure on data for industrial production and unemployment rate. Section 5.8 contains some final comments. The proofs of Sections 5.4 and 5.5 are placed in Appendix 5.A of this chapter, that use a series of technical lemmas placed in Appendix 5.B, while the results from the Monte-Carlo simulations and empirical applications are found in Appendix 5.C.

Before we proceed, we note some changes in notation for expositional simplicity. From hereafter, the subscript x referring to the sequence $\{x_t\}_{t\in\mathbb{Z}}$ is suppressed, but subscript referring to any other sequences remain. Also, we abbreviate $\xi\left(\lambda_j;.\right)$ by $\xi_j\left(.\right)$ for a generic function $\xi(.;.)$. Moreover, since the frequency of the possible persistent component is not assumed to be known, we denote by α_0 the true memory parameter of either the long-run or cyclical component, and by ω_0 the frequency of the possible persistent component. Furthermore, $\alpha_0, \omega_0, \psi_0, \theta_0, \sigma^2_{0,\varepsilon}$ denote the true value of the parameters, while $\alpha, \omega, \psi, \theta, \sigma^2_\varepsilon$ denote any admissible value. Finally, we denote by $b_{(j)}$ the j-th element of a generic vector b, by $B_{(j,k)}$ the (j,k)-th element of a generic matrix B, and by ∇_b the vector of partial derivatives $\partial/\partial b$ for a generic column vector or scalar b.

5.2 Test statistics

Suppose that the spectral density function $f(.)$ of $\{x_t\}_{t\in\mathbb{Z}}$ is known up to a finite set of parameters $\left(\psi_0', \sigma^2_{0,\varepsilon}\right)'$, and is of the form

$$f(\lambda) = \frac{\sigma^2_{0,\varepsilon}}{2\pi}k(\lambda;\psi_0), \tag{5.2.1}$$

where $\sigma^2_{0,\varepsilon} > 0$, $\psi_0 \in \Psi \subseteq \mathbb{R}^{k+1}$, and $k(\lambda;\psi_0)$ is a known even function. Recall that (5.2.1) captures a wide range of parametrically specified models, including $ARMA(p,q)$, $ARFIMA(p,d,q)$ and $GARMA(p,d_\omega,q)$ models. Actually, any covariance stationary process with no deterministic component and well defined spectral density function satisfies (5.2.1).

Under the null hypothesis of no persistent component in $\{x_t\}_{t\in\mathbb{Z}}$, we have that $f(.)$ is bounded and bounded away from zero for all $\lambda \in [0,\pi]$. This means that there exists constant $0 < K < \infty$ such that

$$K^{-1} < f(\lambda) < K, \quad \text{for all } \lambda \in [0,\pi]. \tag{5.2.2}$$

Hence, our hypothesis of interest can be written as

$$H_0 : \left\{K^{-1} < k(\lambda;\psi_0) < K, \quad \text{for all } \lambda \in [0,\pi]\right\}. \tag{5.2.3}$$

On the other hand, under the alternative hypothesis of a persistent component in $\{x_t\}_{t\in\mathbb{Z}}$, the spectral density function $f(.)$ is unbounded at some frequency $\omega_0 \in [0,\pi]$. Then, we can consider that there exist $\omega_0 \in [0,\pi]$ and $0 < \alpha_0 < 1$ such that

$$f(\lambda) \sim c_0 \left|\lambda - \omega_0\right|^{-\alpha_0}, \quad \text{as } \lambda \to \omega_0, \tag{5.2.4}$$

for some $0 < c_0 < 1$. Hence, our alternative hypothesis can be formulated as

$$H_1 : \left\{\exists\ \omega_0 \in [0,\pi] \text{ and } 0 < \alpha_0 < 1 : k(\lambda;\psi_0) = g\left(\lambda;\alpha_0,\omega_0\right) h\left(\lambda;\theta_0\right),\right. \tag{5.2.5}$$

where $\psi_0 = \left(\alpha_0, \theta_0'\right)'$, while the functions $h\left(.;\theta_0\right)$ and $g\left(.;\alpha_0,\omega_0\right)$ are such that

$$K^{-1} < h\left(\lambda;\theta_0\right) < K, \quad \text{for all } \lambda \in [0,\pi], \tag{5.2.6}$$

and

$$g\left(\lambda;\alpha_0,\omega_0\right) \sim c_0 \left|\lambda - \omega_0\right|^{-\alpha_0}, \quad \text{as } \lambda \to \omega_0, \tag{5.2.7}$$

for some $0 < c_0 < 1\big\}$.

5.2.1 Wald test $\mathcal{T}_W$

Assuming that $f(.)$ follows model (5.3.1)-(5.3.3) in Condition C.1 below, suppose that for a given ω_0 we have an estimator of $\left(\alpha_0, \theta_0', \sigma_{0,\varepsilon}^2\right)'$, and denote by λ_s the closest Fourier frequency to ω_0. For example, given a stretch of data $\{x_1,...,x_n\}$, we can use the PW estimator discussed in the Introduction. Recall that the PW estimator is defined as

$$\left(\widehat{\alpha}_s, \widehat{\theta'}\right)' = \underset{(\alpha,\theta')'\in\Pi\times\Theta}{\arg\min}\ Q\left(\alpha,\theta,s\right) \quad \text{and} \quad \widehat{\sigma}_\varepsilon^2 = Q\left(\widehat{\alpha}_s, \widehat{\theta}, s\right), \tag{5.2.8}$$

where

$$Q\left(\alpha,\theta,s\right) = \frac{2\pi}{\tilde{n}} \sum_{j=1}^{\tilde{n}} \frac{I_j}{g_j\left(\alpha,\lambda_s\right) h_j\left(\theta\right)}, \tag{5.2.9}$$

where the sets Π and Θ are given in Condition C.1 below, and $\tilde{n} = \left[\frac{n}{2}\right]$ with $[.]$ denoting the integer part.

Next, suppose that $\widehat{\alpha}_s$ is computed for all $s = 0,1,...,\tilde{n}$. Because under suitable regularity conditions, given in Section 5.3 below, the PW estimator is consistent, we expect that under H_0, $\widehat{\alpha}_s \approx 0$ for all $s = 0, 1, ..., \tilde{n}$, whereas under H_1, there exists an s such that $\widehat{\alpha}_s > 0$. So, a test statistic for the null hypothesis (5.2.3) against the alternative hypothesis (5.2.5) can be based on whether $\widehat{\alpha}_s$ is significantly greater than zero for some $s = 0, 1, ..., \tilde{n}$. More precisely, the test statistic is given by

$$\mathcal{T}_W = \sup_{s=0,1,...,\tilde{n}} \widehat{\alpha}_s, \tag{5.2.10}$$

rejecting H_0 if $\mathcal{T}_W$ is greater than some critical value.

5.2.2 Lagrange multiplier test $\mathcal{T}_{LM}$

The statistic described in (5.2.10) involves the estimation of $\alpha_0(\lambda_s)$, the memory parameter associated with the component of frequency λ_s, along with any other parameter of the model, for all $s = 0, 1, ..., \tilde{n}$, which can be highly computing intensive as nonlinear optimization algorithms are employed. Moreover, as the asymptotic distribution of $\mathcal{T}_W$ is not standard, see Theorem 5.1 below, to obtain critical values, bootstrap methods are required. Because of that, the implementation of the W type of test can be a prohibitive task in computing time. Thus, we introduce a LM type of test which does not require the estimation of $\alpha_0(\lambda_s)$ under H_0. This is computationally simpler and bootstrap algorithms are easier to implement.

To that end, consider the PW estimator of $(\theta_0', \sigma_{0,\varepsilon}^2)'$ under H_0, that is

$$\widetilde{\theta} = \arg\min_{\theta\in\Theta} \frac{2\pi}{\tilde{n}} \sum_{j=1}^{\tilde{n}} \frac{I_j}{h_j(\theta)} \quad \text{and} \quad \widetilde{\sigma}_\varepsilon^2 = \frac{2\pi}{\tilde{n}} \sum_{j=1}^{\tilde{n}} \frac{I_j}{h_j\left(\widetilde{\theta}\right)}, \tag{5.2.11}$$

given the model (5.3.1)-(5.3.3) for $f(.)$ in Condition C.1 below. Next, for all $s = 0, 1, ..., \tilde{n}$, consider the first derivative of Whittle objective function (5.2.9) with respect to $(\alpha, \theta')'$, and denote

$$q(\alpha, \theta, s) = \frac{1}{\widetilde{\sigma}_\varepsilon^2 \tilde{n}} \sum_{j=1}^{\tilde{n}} q_j(\alpha, \theta, s) \frac{I_j}{h_j(\theta)} \tag{5.2.12}$$

and

$$V(\alpha, \theta, s) = \frac{1}{\tilde{n}} \sum_{j=1}^{\tilde{n}} q_j(\alpha, \theta, s) q_j'(\alpha, \theta, s), \tag{5.2.13}$$

where

$$q_j(\alpha, \theta, s) = -\frac{\partial \log\left(g_j(\alpha, \lambda_s) h_j(\theta)\right)}{\partial (\alpha, \theta')'}. \tag{5.2.14}$$

Notice that under standard suitable conditions, $V(\alpha,\theta,s)/(2\pi)^2$ is an estimate of the asymptotic covariance matrix of $\tilde{n}^{\frac{1}{2}}q(\alpha,\theta,s)$.

Next, suppose that we fix a frequency λ_s for some $s=0,1,...,\tilde{n}$. Then, the LM test for the hypothesis $\widetilde{H}_0 : \alpha_0(\lambda_s)=0$, becomes whether $\widetilde{q}_s$ is not significantly different than zero, where

$$\widetilde{q}_s = \left(V^{-1}_{(1,1)}\left(0,\widetilde{\theta},s\right)\right)^{\frac{1}{2}} \tilde{n}^{\frac{1}{2}} q_{(1)}\left(0,\widetilde{\theta},s\right), \tag{5.2.15}$$

recalling that $q_{(1)}(.)$ denotes the first component of the vector $q(.)$ and, $V^{-1}_{(1,1)}(.)$ is the element $(1,1)$ of the matrix $V^{-1}(.)$. Because under suitable regularity conditions, given in Section 5.3 below, the PW estimator is consistent, then under H_0, we expect that $\widetilde{q}_s \approx 0$ for all $s=0,1,...,\tilde{n}$.

Notice that our hypothesis testing is one sided. So, in the same way that for the W type of test we reject the null hypothesis if $\widehat{\alpha}_s > 0$ for some $s=0,1,...,\tilde{n}$, we need to find the direction of departure from the null hypothesis, that is the sign that $\widetilde{q}_s$ takes under the alternative hypothesis. It is clear that the sign of $\widetilde{q}_s$ is that of

$$\frac{1}{\tilde{n}}\sum_{j=1}^{\tilde{n}} q_{j_{(1)}}\left(0,\widetilde{\theta},s\right)\frac{I_j}{h_j\left(\widetilde{\theta}\right)}, \tag{5.2.16}$$

where $q_{j_{(1)}}(.)$ is the first element of the vector $q_j(.)$. Notice that because Condition C.5 below implies that $\{x_t\}_{t\in\mathbb{Z}}$ is ergodic, see Stout (1974) Theorem 3.5.8, we have that $\widetilde{\sigma}^2_\varepsilon \xrightarrow{p} \sigma^2_{0,\varepsilon} > 0$. Because we are under the alternative hypothesis, there exists an $s=0,1,...,\tilde{n}$ for which $\alpha_0(\lambda_s) > 0$. Suppose for simplicity that $s=0$ and $h(\lambda;\theta_0)=1$. Then, we have that the last displayed expression becomes

$$\begin{aligned} -\frac{1}{\tilde{n}}\sum_{j=1}^{\tilde{n}}\frac{\partial \log g_j(0,0)}{\partial\alpha}I_j &= -\frac{\sigma^2_{0,\varepsilon}}{2\pi\tilde{n}}\sum_{j=1}^{\tilde{n}}\frac{\partial \log g_j(0,0)}{\partial\alpha}g_j(\alpha_0(0),0) \\ &\quad -\frac{1}{\tilde{n}}\sum_{j=1}^{\tilde{n}}\frac{\partial \log g_j(0,0)}{\partial\alpha}g_j(\alpha_0(0),0) \\ &\quad \times\left(\frac{I_j}{g_j(\alpha_0(0),0)}-\frac{\sigma^2_{0,\varepsilon}}{2\pi}\right). \end{aligned} \tag{5.2.17}$$

Now, under Conditions C.1, C.5 and C.6, Lemma 3 of Delgado, Hidalgo, and Velasco (2005) implies that the second term on the right of the last displayed equation is

$$-\frac{1}{\tilde{n}}\sum_{j=1}^{\tilde{n}}\frac{\partial \log g_j(0,0)}{\partial\alpha}g_j(\alpha_0(0),0)\left(I_{j,\varepsilon}-\frac{\sigma^2_{0,\varepsilon}}{2\pi}\right)+o_p\left(n^{-\frac{1}{2}}\right). \tag{5.2.18}$$

Moreover, by standard results on $I_{j,\varepsilon}$, see for example Giraitis, Hidalgo, and Robinson (2001), the last expression is $O_p\left(n^{-\frac{1}{2}}\right)$. On the other hand, under Condition C.1 below, it follows that $g\left(\lambda;\alpha_0\left(0\right),0\right) = \left|2\sin\left(\frac{\lambda}{2}\right)\right|^{-2\alpha_0(0)}$. Then, from the mean value theorem, we have that the first term on the right of (5.2.17) is

$$\begin{aligned}\frac{\sigma^2_{0,\varepsilon}}{\pi\tilde{n}}\sum_{j=1}^{\tilde{n}}\left|2\sin\left(\frac{\lambda_j}{2}\right)\right|^{-2\alpha_0(0)}\log\left|2\sin\left(\frac{\lambda_j}{2}\right)\right| &= \frac{\sigma^2_{0,\varepsilon}}{\pi}\left\{\frac{1}{\tilde{n}}\sum_{j=1}^{\tilde{n}}\log\left|2\sin\left(\frac{\lambda_j}{2}\right)\right|\right.\\ &\quad -2\alpha_0\left(0\right)\frac{1}{\tilde{n}}\sum_{j=1}^{\tilde{n}}\left|2\sin\left(\frac{\lambda_j}{2}\right)\right|^{-2\overline{\alpha_0(0)}}\\ &\quad \left.\times\log^2\left|2\sin\left(\frac{\lambda_j}{2}\right)\right|\right\}, \qquad (5.2.19)\end{aligned}$$

where $\overline{\alpha_0\left(0\right)}$ is an intermediate point between 0 and $\alpha_0\left(0\right)$. Next, by an obvious extension of Lemma 2 of Robinson (1995b), see also Lemma 1 of Delgado, Hidalgo, and Velasco (2005), and since $\int_0^\pi \log\left|2\sin\left(\frac{\lambda}{2}\right)\right| d\lambda = 0$, the first term on the right of the last displayed equality is $O\left(\frac{\log\tilde{n}}{\tilde{n}}\right)$, whereas the second term is strictly negative since $\alpha_0\left(0\right) > 0$ and $\frac{1}{\tilde{n}}\sum_{j=1}^{\tilde{n}}\left|2\sin\left(\frac{\lambda_j}{2}\right)\right|^{-2\overline{\alpha_0(0)}}\log^2\left|2\sin\left(\frac{\lambda_j}{2}\right)\right|$ is bounded away from zero. Hence, we conclude that $q_{(1)}\left(0,0\right) \xrightarrow{p} c < 0$, as $n \to \infty$.

Therefore, the test for the null hypothesis (5.2.3) against the alternative hypothesis (5.2.5) is based on the test statistic

$$\mathcal{T}_{LM} = \sup_{s=0,1,\ldots,\tilde{n}} -\widetilde{q}_s, \qquad (5.2.20)$$

rejecting H_0 if $\mathcal{T}_{LM}$ is greater than some critical value.

5.3 Conditions

We introduce the following regularity conditions.

C.1 The spectral density function $f(.)$ satisfies

$$f(\lambda) = \frac{\sigma^2_{0,\varepsilon}}{2\pi}k(\lambda;\psi_0,\omega_0), \quad -\pi < \lambda \leq \pi, \qquad (5.3.1)$$

where

$$k(\lambda;\psi_0,\omega_0) = g\left(\lambda;\alpha_0,\omega_0\right)h\left(\lambda;\theta_0\right), \qquad (5.3.2)$$

with

$$g(\lambda;\alpha_0,\omega_0) = \left|4\sin\frac{\lambda-\omega_0}{2}\sin\frac{\lambda+\omega_0}{2}\right|^{-\alpha_0}, \tag{5.3.3}$$

where $\theta_0 \in \Theta$, a compact set in $\mathbb{R}^k$, $\alpha_0 \in \Pi$, where $\Pi = [0,1)$ if $\omega_0 \in (0,\pi)$ and $\Pi = [0,1/2)$ if $\omega_0 = \{0,\pi\}$. Furthermore, the function $h(\lambda;\theta)$, $-\pi < \lambda \leq \pi$, is even in λ, bounded and bounded away from zero, and the derivatives $\nabla_\theta h(\lambda;\theta)$, $\nabla_\lambda h(\lambda;\theta)$, $\nabla_\theta\nabla_\lambda h(\lambda;\theta)$ and $\nabla_\theta\nabla'_\theta h(\lambda;\theta)$ are continuous.

C.2 For all $\theta \in \Theta$, the function $h(.;.)$ satisfies

$$\int_{-\pi}^{\pi} \log(h(\lambda;\theta))\,d\lambda = 0. \tag{5.3.4}$$

C.3 It holds that

$$\inf_{(\alpha,\theta')'\in\Pi\times\Theta} \int_{-\pi}^{\pi} \frac{k(\lambda;\alpha_0,\theta_0,\omega_0)}{k(\lambda;\alpha,\theta,\omega_0)} d\lambda = 1, \tag{5.3.5}$$

and the set

$$\{\lambda : k(\lambda;\alpha_0,\theta_0,\omega_0) \neq k(\lambda;\alpha,\theta,\omega_0)\} \quad \text{for } (\alpha_0,\theta_0') \neq (\alpha,\theta'), \tag{5.3.6}$$

has positive Lebesgue measure. Also, the matrix

$$\Omega = \frac{1}{4\pi}\int_{-\pi}^{\pi} \nabla_\psi \log k(\lambda;\psi_0,\omega_0)\nabla_{\psi'} k(\lambda;\psi_0,\omega_0)d\lambda \tag{5.3.7}$$

is positive definite for all $\omega_0 \in [0,\pi]$.

C.4 θ_0 is an interior point of the compact set $\Theta \in \mathbb{R}^k$.

C.5 The process $\{x_t\}_{t\in\mathbb{Z}}$ is defined as

$$x_t = \sum_{j=0}^{\infty} \varphi_j \varepsilon_{t-j}, \quad \sum_{j=0}^{\infty} \varphi_j^2 < \infty, \quad \varphi_0 = 1, \tag{5.3.8}$$

where $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a sequence of ergodic processes satisfying

$$\begin{aligned} E(\varepsilon_t|F_{t-1}) &= 0, \quad E(\varepsilon_t^2|F_{t-1}) = \sigma^2_{0,\varepsilon}, \quad \text{a.s., } t\in\mathbb{Z}, \\ E\left(|\varepsilon_t|^\ell |F_{t-1}\right) &= \mu_\ell < \infty, \quad \ell = 3,...,8, \quad \text{a.s., } t\in\mathbb{Z}, \end{aligned} \tag{5.3.9}$$

where F_t is the $\sigma-$algebra of events generated by $\{\varepsilon_s, s \leq t\}$, and

$$cum(\varepsilon_{t_1},\varepsilon_{t_2},\varepsilon_{t_3},\varepsilon_{t_4}) = \begin{cases} \kappa, & \text{if } t_1 = t_2 = t_3 = t_4 \\ 0, & \text{otherwise} \end{cases}. \tag{5.3.10}$$

C.6 The transfer function $\phi(\lambda) = \sum_{j=0}^{\infty} \varphi_j e^{ij\lambda}$ is differentiable in $\lambda \in [0,\pi] \backslash \{\omega_0\}$ and

$$\frac{d\phi(\lambda)}{d\lambda} = O\left(\frac{|\phi(\lambda)|}{|\lambda - \omega_0|}\right). \tag{5.3.11}$$

We now comment on our conditions. Condition C.1 covers a wide range of short memory models having spectral density function $h(.;.)$, including covariance stationary and invertible $ARMA(p,q)$ and Bloomfield (1973) ones, although it allows for models with autocovariance function decaying to zero much slower than the previous two models. Condition C.2 is not very strong and implies that $\int_{-\pi}^{\pi} \log\left(k(\lambda;\psi,\omega_0)\right) d\lambda = 0$, since $\int_{-\pi}^{\pi} \log\left|2\sin\left(\frac{\lambda\pm\omega_0}{2}\right)\right| d\lambda = 0$. It follows that $\sigma_{0,\varepsilon}^2$ is the one-step mean square linear prediction error, see Hannan (1970) pp. 121-123. Notice also that this condition is employed in the construction of the PW estimator, see Subsection 1.3.1. The first part of Condition C.3 is an identifiability assumption, while the second part is required in order for the variance in the asymptotic distribution of the PW estimator of ψ_0 to be well defined. Condition C.4 is trivial and very mild. Condition C.5 requires the process $\{x_t\}_{t\in\mathbb{Z}}$ to be linear and is an extension of Assumptions A.5 and A.8 in Section 2.5. The normalization $\varphi_0 = 1$ is consistent with (5.3.4), while the finite eighth moments assumed for $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ are needed to show the tightness condition of some process indexed by $\tau \in [0,1]$, required to show Theorem 5.1 below. Notice that the assumptions on $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ in Conditions C.5 are satisfied for a sequence of i.i.d. random variables with finite eighth moments. Condition C.6 is extension of Assumption A.7 in Section 2.5 to the whole band of frequencies $[0,\pi]$ excluding ω_0.

We should notice that Conditions C.1-C.6 are essentially those of Giraitis, Hidalgo, and Robinson (2001) employed there to establish the consistency and asymptotic normality of the PW estimator in (5.2.8). Overall, these conditions are standard in the literature on PW estimation, see Hannan (1973), Fox and Taqqu (1986), Dahlhaus (1989), Giraitis and Surgailis (1990), Hosoya (1997) and, Velasco and Robinson (2000), and are satisfied for processes whose short memory component, having spectral density function $h(.;.)$, is a covariance stationary and invertible $ARMA(p,q)$ or Bloomfield (1973) model. There are only two differences between the assumptions of Giraitis, Hidalgo, and Robinson (2001) and ours. Firstly, we require $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ to have eighth finite moments, while in Giraitis, Hidalgo, and Robinson (2001) finite moments slightly above four are needed. Secondly, we restrict the cumulant of $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ to be of the form (5.3.10), an assumption that is not imposed by Giraitis, Hidalgo, and Robinson (2001). Notice also that as in Giraitis, Hidalgo, and Robinson (2001), we could have changed Condition C.1 to

C.1′ The spectral density function $f(.)$ satisfies

$$f(\lambda) = \frac{\sigma^2_{0,\varepsilon}}{2\pi} k(\lambda; \psi_0, \omega_0), \quad -\pi < \lambda \leq \pi, \tag{5.3.12}$$

where

$$k(\lambda; \psi_0, \omega_0) = \begin{cases} |\lambda - \omega_0|^{-\alpha_0} h_1(\lambda; \psi_0, \omega_0), & \text{if } 0 \leq \lambda \leq \pi \\ |\lambda + \omega_0|^{-\alpha_0} h_1(\lambda; \psi_0, \omega_0), & \text{if } -\pi \leq \lambda \leq 0 \end{cases} \tag{5.3.13}$$

such that for $-\pi < \lambda \leq \pi$, $\theta_0 \in \Theta$, $\alpha_0 \in [0,1)$, $h_1(\lambda; \psi, \omega_0)$ is even in λ, bounded and bounded away from zero, $\nabla_\psi h_1(\lambda; \psi, \omega_0)$ and $\nabla_\psi \nabla_{\psi'} h_1(\lambda; \psi, \omega_0)$ are continuous and bounded, $h_1(\lambda; \psi, \omega_0)$ and $\nabla_\psi h_1(\lambda; \psi, \omega_0)$ satisfy uniform Lipschitz conditions in ω_0 of order greater than $\frac{1}{2}$, while, for $0 < |\lambda| < \pi$, $\theta \in \Theta$, $\omega_0 \in [0,\pi]$, $\nabla_\lambda h_1(\lambda; \psi, \omega_0)$ and $\nabla_\lambda \nabla_\psi h_1(\lambda; \psi, \omega_0)$ are bounded.

For a motivation and comparison with Condition C.1 we refer to Giraitis, Hidalgo, and Robinson (2001).

5.4 Statistical properties of $\mathcal{T}_W$ and $\mathcal{T}_{LM}$

In this section, we establish the asymptotic distributions of the test statistics $\mathcal{T}_W$ and $\mathcal{T}_{LM}$ and discuss their consistency and power properties under the conditions of Section 5.3. First, denote

$$\begin{aligned} A(\tau) &= \begin{pmatrix} A_{(1,1)}(\tau) & A_{(1,2)}(\tau) \\ A_{(2,1)}(\tau) & A_{(2,2)} \end{pmatrix} \\ &= \begin{pmatrix} \int_0^1 \frac{\partial \log|g(\pi x;0,\pi\tau)|}{\partial \alpha} \frac{\partial \log|g(\pi x;0,\pi\tau)|}{\partial \alpha} dx & \int_0^1 \frac{\partial \log|g(\pi x;0,\pi\tau)|}{\partial \alpha} \frac{\partial \log h(\pi x;\theta_0)}{\partial \theta'} dx \\ \int_0^1 \frac{\partial \log h(\pi x;\theta_0)}{\partial \theta} \frac{\partial \log|g(\pi x;0,\pi\tau)|}{\partial \alpha} dx & \int_0^1 \frac{\partial \log h(\pi x;\theta_0)}{\partial \theta} \frac{\partial \log h(\pi x;\theta_0)}{\partial \theta'} dx \end{pmatrix}, \end{aligned} \tag{5.4.1}$$

where $\pi^{-1}\lambda_s \to \tau \in [0,1]$, as $n \to \infty$, and let

$$A^{-1}(\tau) = \begin{pmatrix} A^{-1}_{(1,1)}(\tau) & A^{-1}_{(1,2)}(\tau) \\ A^{-1}_{(2,1)}(\tau) & A^{-1}_{(2,2)}(\tau) \end{pmatrix}. \tag{5.4.2}$$

Also, write

$$\begin{aligned} \mathcal{C}(\tau_1,\tau_2) &= A^{-1}_{(1,1)}(\tau_1) A^{-1}_{(1,1)}(\tau_2) \mathcal{K}(\tau_1,\tau_2) + A^{-1}_{(1,1)}(\tau_2) A^{-1}_{(1,2)}(\tau_1) A_{(2,1)}(\tau_2) \\ &\quad + A^{-1}_{(1,1)}(\tau_1) A_{(1,2)}(\tau_1) A^{-1}_{(2,1)}(\tau_2) + A^{-1}_{(1,2)}(\tau_1) A_{(2,2)} A^{-1}_{(2,1)}(\tau_2), \end{aligned} \tag{5.4.3}$$

where

$$\mathcal{K}(\tau_1,\tau_2) = \int_0^1 \log|g(\pi x;0,\pi\tau_1)| \log|g(\pi x;0,\pi\tau_2)| \, dx. \tag{5.4.4}$$

The next theorem establishes, under Conditions C.1-C.6, the asymptotic distribution of our test statistics $\mathcal{T}_W$ and $\mathcal{T}_{LM}$ under H_0 given in (5.2.3).

Theorem 5.1

Assume that Conditions C.1-C.6 are satisfied. Then, under H_0 given in (5.2.3), we have that

i)
$$\tilde{n}^{\frac{1}{2}}\mathcal{T}_W \xrightarrow{d} \max_{\tau\in[0,1]} \mathcal{G}(\tau), \quad \text{as } n \to \infty, \tag{5.4.5}$$

ii)
$$\tilde{n}^{\frac{1}{2}}\mathcal{T}_{LM} \xrightarrow{d} \max_{\tau\in[0,1]} \mathcal{G}(\tau), \quad \text{as } n \to \infty, \tag{5.4.6}$$

where $\mathcal{G}(\tau)$ is a process such that for fixed τ, $\mathcal{G}(\tau)$ is distributed as $X(\tau)\mathcal{I}(X(\tau)\geq 0)$, where $X(\tau)$ is a Gaussian process with covariance structure given by $\mathcal{C}(\tau_1,\tau_2)$.

One basic requirement for any test is its consistency. Also it is useful to learn about its power function against local alternatives to gain some insight about the characteristics of the test. To this end, consider

$$H_a : \left\{\exists\ \omega_0 \in [0,\pi] : f(\lambda) = \frac{\sigma_{0,\varepsilon}^2}{2\pi} g(\lambda;\alpha(n),\omega_0)\, h(\lambda;\theta_0)\right., \tag{5.4.7}$$

where $\alpha(n) = \alpha_0\tilde{n}^{-\frac{1}{2}}$ with $\left.\alpha_0 > 0\right\}$. Under the same conditions and using similar techniques as in Theorem 5.1, we establish in the following corollary the asymptotic distribution of our test statistics $\mathcal{T}_W$ and $\mathcal{T}_{LM}$ under H_a given above.

Corollary 5.1

Assume that Conditions C.1-C.6 are satisfied. Then, under H_a given in (5.4.7), we have that

i)
$$\tilde{n}^{\frac{1}{2}}\mathcal{T}_W \xrightarrow{d} \max_{\tau\in[0,1]} \left(\mathcal{G}(\tau) + \mathcal{C}(\tau,\tau)^{-\frac{1}{2}}\alpha_0(\tau)\right), \quad \text{as } n \to \infty, \tag{5.4.8}$$

ii)
$$\tilde{n}^{\frac{1}{2}}\mathcal{T}_{LM} \xrightarrow{d} \max_{\tau\in[0,1]} \left(\mathcal{G}(\tau) + \mathcal{C}(\tau,\tau)^{-\frac{1}{2}}\alpha_0(\tau)\right), \quad \text{as } n \to \infty, \tag{5.4.9}$$

where $\alpha_0(\tau) = \alpha_0\mathcal{I}(\tau = \pi^{-1}\omega_0)$.

From the results of Corollary 5.1, it is straightforward to observe that the tests are consistent. This is the case because for fixed alternatives, that is, $\alpha(n) = \alpha_0 > 0$, $\mathcal{T}_W \xrightarrow{p} \mathcal{C}(\tau,\tau)^{-\frac{1}{2}}\alpha_0 > 0$. So, we have that for any $z > 0$,

$$\Pr\left(\tilde{n}^{\frac{1}{2}}\mathcal{T}_W > z\right) \to 1, \tag{5.4.10}$$

as $n \to \infty$. Similarly, we have that for any $z > 0$,

$$\Pr\left(\tilde{n}^{\frac{1}{2}}\mathcal{T}_{LM} > z\right) \to 1, \tag{5.4.11}$$

as $n \to \infty$.

The results of Theorem 5.1 give the (asymptotic) justification of the tests. However, the rate of convergence of the finite sample distribution to the asymptotic one might be slow, see Hall (1979) who established that the rate is logarithmic for the maximum of independent standard Gaussian variables. So, critical values relying on the asymptotic distribution might be a poor approximation to those of the finite sample distribution. One solution would be to employ Edgeworth expansions. However, Hall (1990) showed that they do not perform well, compared to bootstrap schemes, at the tails of the distribution, which is the most important region when performing hypothesis testing. On the other hand, when exploring the properties of the bootstrap for the maximum of the kernel density estimator, Hall (1991) found that the bootstrap performs better than Edgeworth expansions in terms of their accuracy to the finite sample distribution of the suprema. In addition, because the asymptotic distribution of $\tilde{n}^{\frac{1}{2}}\mathcal{T}_W$ and $\tilde{n}^{\frac{1}{2}}\mathcal{T}_{LM}$ is nonstandard and model dependent, it seems necessary to rely on Monte-Carlo algorithms to compute asymptotically valid critical values. For all these reasons, in the next section we propose a bootstrap scheme and establish its validity.

5.5 Bootstrap algorithm for $\mathcal{T}_W$ and $\mathcal{T}_{LM}$

Bootstrap methods have become a routine method for approximating the distribution of a statistical quantity, partly due to the increasing computation power. At a theoretical level, bootstrap algorithms have attracted considerable effort to their development, as they are capable of approximating the finite sample distribution of statistics more effectively than those based on their asymptotic counterparts, and also because they allow for the computation of valid asymptotic quantiles of the limiting distribution in nonstandard situations. In particular, when the limiting distribution is unknown or if known, the practitioner is unable to compute its quantiles.

The first contribution on bootstrap methods in the i.i.d. setup goes back to Efron (1979). The basic idea of the bootstrap is, given a stretch of data $\mathcal{Z}_n = \{z_1, ..., z_n\}$ say, to treat the data as if it were the true population, and to carry out Monte-Carlo experiments in which pseudo-data is drawn randomly with replacement from $\mathcal{Z}_n$. An obvious drawback of this method is the rather restrictive assumption that the data are taken from an i.i.d. population. To circumvent this problem in the case that homoscedasticity is violated, Wu (1986) proposed the wild bootstrap.

In the time series context, the data cannot certainly be assumed to be independent and the need of bootstrap methods that preserve the time dependence structure arises. For this reason, Künsch (1989) proposed to resample not from the data themselves, but from blocks of data so that inside the block the dependence structure of the series is maintained. A similar method was put forward by Politis and Romano (1994), where blocks of data were again constructed, but contrary to Künsch (1989) they did not resample from these blocks of data, and instead considered the blocks as subseries. However, both these methods suffer from the disadvantage that for their implementation the choice of block length is needed. It is also evident that when examining long-run or cyclical persistence, the length of the block has to be chosen big enough to capture the relevant information. However, this would leave the practitioner with few blocks available for resampling or to act as subseries, which in turn results to a poor performance of the bootstrap method.

In later studies, various bootstrap methods have been proposed that do not rely on the construction of blocks of data, but require a priori knowledge of that data generating mechanism of the time series. Examples include the residual based bootstrap of Franke and Kreiss (1992) and the sieve bootstrap of Bühlmann (1997). In Franke and Kreiss (1992) the data are assumed to follow a covariance stationary $ARMA(p,q)$ model, and the bootstrap algorithm is performed by resampling randomly with replacement the residuals of the $ARMA(p,q)$ model. In Bühlmann (1997) the data generating process is assumed to be linear, and is approximated by an $AR(p)$ model in which $p = p(n)$ increases with increases in the sample size. The residuals of the latter model are constructed and the bootstrap scheme is performed in line with that of Franke and Kreiss (1992). The latter bootstrap methods are based on the time domain, but frequency domain bootstrap algorithms have also been proposed. Examples include the studies of Franke and Härdle (1992), Dahlhaus and Janas (1996) and Hidalgo and Kreiss (2004) for covariance stationary linear processes with bootstrap algorithms utilizing the fact the the normalized periodogram ordinates behave as if they were a series of uncorrelated processes, and therefore the resampling is performed on the normalized periodogram ordinates.

In our context, the resampling method must be such that the conditional distribution, given $\mathcal{X}_n = \{x_1, ..., x_n\}$, of the bootstrap statistic, say $\tilde{n}^{\frac{1}{2}}\mathcal{T}_W^*$ the bootstrap analogue of $\tilde{n}^{\frac{1}{2}}\mathcal{T}_W$, consistently estimates the distribution of $\max_{\tau\in[0,1]} \mathcal{G}(\tau)$ under the null hypothesis H_0 and the local alternatives H_a. That is, $\tilde{n}^{\frac{1}{2}}\mathcal{T}_W^* \xrightarrow{d^*} \max_{\tau\in[0,1]} \mathcal{G}(\tau)$ in probability under $H_0 \cup H_a$, where $\xrightarrow{d^*}$ denotes convergence in bootstrap distribution, that is,

$$\Pr\left(\tilde{n}^{\frac{1}{2}}\mathcal{T}_W^* \leq z \middle| \mathcal{X}_n\right) \xrightarrow{p} G(z), \tag{5.5.1}$$

at each continuity point z of $G(z) = \Pr\left(\max_{\tau\in[0,1]} \mathcal{G}(\tau) \le z\right)$, as defined in Giné and Zinn (1990). A second requirement is that under H_1, $\tilde{n}^{\frac{1}{2}}\mathcal{T}_W^*$ must also converge in bootstrap distribution, although possibly to a different one than under H_0. Likewise for the bootstrap analogue of $\mathcal{T}_{LM}$, denoted by $\mathcal{T}_{LM}^*$. To achieve the first requirement, due to the (pseudo) Gaussian behaviour of the limit distribution, one key condition is that the resampling algorithm should preserve the correlation structure of the original data x. On the other hand, the second requirement would be guaranteed if we were capable to bootstrap under the null hypothesis H_0.

To achieve both aims, we propose the following bootstrap algorithm. Suppose that we are under the null hypothesis H_0, so that $f(\lambda) = \frac{\sigma_{0,\varepsilon}^2}{2\pi} h(\lambda;\theta_0)$. Recall that the discrete Fourier transform $w(.)$ of the data $\{x_1, ..., x_n\}$, given in (1.3.5), satisfies

$$x_t = n^{-\frac{1}{2}} \sum_{j=1}^{n} e^{it\lambda_j} w_j. \tag{5.5.2}$$

Notice also, that under Condition C.5, we can apply the approximation

$$w_j \approx \Phi\left(e^{-i\lambda_j}\right) w_{j,\varepsilon}, \tag{5.5.3}$$

where $\Phi(z) = \sum_{j=0}^{\infty} \varphi_j z^j$ and $w_{j,\varepsilon} = n^{-\frac{1}{2}} \sum_{t=1}^{n} \varepsilon_t e^{-it\lambda_j}$, and the approximation error is $o_p(1)$ uniformly in λ_j, see the proof of Theorem 10.3.1. in Brockwell and Davis (1991). Here, $\approx$ should be read as "approximately". Hence, combining (5.5.2) with the latter approximation, we have that

$$x_t \approx \widetilde{x}_t =: n^{-\frac{1}{2}} \sum_{j=1}^{n} e^{it\lambda_j} \Phi\left(e^{-i\lambda_j}\right) w_{j,\varepsilon}. \tag{5.5.4}$$

We now show that $\{\widetilde{x}_t\}_{t\in\mathbb{Z}}$ preserves (asymptotically) the covariance structure of $\{x_t\}_{t\in\mathbb{Z}}$. First, notice that under the assumptions on the innovation sequence $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ in Condition C.5, it follows that $E\left(w_{j,\varepsilon}\overline{w_{k,\varepsilon}}\right) = \sigma_{0,\varepsilon}^2 \mathcal{I}(j=k)$. Hence, the autocovariance function of $\{\widetilde{x}_t\}_{t\in\mathbb{Z}}$ satisfies

$$\begin{aligned}
\gamma_{\widetilde{x}}(\tau) &= E\left(\widetilde{x}_t \widetilde{x}_{t+\tau}\right) \\
&= \frac{\sigma_{0,\varepsilon}^2}{n} \sum_{j=1}^{n} e^{i\tau\lambda_j} \left|\Phi\left(e^{-i\lambda_j}\right)\right|^2 && (5.5.5) \\
&\approx \frac{\sigma_{0,\varepsilon}^2}{2\pi} \int_{-\pi}^{\pi} h(\lambda;\theta_0)\, e^{i\tau\lambda} d\lambda = \gamma(\tau), && (5.5.6)
\end{aligned}$$

because $\left|\Phi\left(e^{-i\lambda}\right)\right|^2 = h\left(\lambda;\theta_0\right)$, and under the assumptions on $h\left(.;.\right)$ in Condition C.1, we have that $\frac{1}{n}\sum_{j=1}^{n} h\left(\lambda_j;\theta_0\right)e^{i\tau\lambda_j} \rightarrow \frac{1}{2\pi}\int_0^{2\pi} h\left(\lambda;\theta_0\right)e^{i\tau\lambda}d\lambda$, as $n \rightarrow \infty$, see Problem 1.7.14 of Brillinger (1975).

Thus, if in the right side of (5.5.4), $\Phi\left(e^{-i\lambda_j}\right)$ was replaced by a consistent estimator, the problem of obtaining a bootstrap sample $\{x_1^*,...,x_n^*\}$, becomes one of performing a valid bootstrap algorithm for $\{w_{1,\varepsilon},...,w_{n,\varepsilon}\}$.

The previous arguments suggest the following bootstrap algorithm.

Step 1: Let $\widetilde{x}^* = \{\widetilde{x}_1^*,...,\widetilde{x}_n^*\}$ be a random sample with replacement from the empirical distribution of the standardized data $\{\breve{x}_1,...,\breve{x}_n\}$

$$\breve{x}_t = \frac{x_t-\overline{x}}{\overline{\sigma}}, \quad \text{with } \overline{x} = \frac{1}{n}\sum_{t=1}^{n} x_t,\ \overline{\sigma}^2 = \frac{1}{n}\sum_{t=1}^{n}\left(x_t-\overline{x}\right)^2, \tag{5.5.7}$$

and obtain the discrete Fourier transform of $\widetilde{x}^*$ as

$$w_{j,\widetilde{x}^*} = n^{-\frac{1}{2}}\sum_{t=1}^{n}\widetilde{x}_t^* e^{-it\lambda_j}, \tag{5.5.8}$$

for $j = 1,...,n$.

Step 2: For $t = 1,...,n$, compute

$$x_t^* = \widetilde{\sigma}_\varepsilon n^{-\frac{1}{2}}\sum_{j=1}^{n} e^{it\lambda_j}\widetilde{\Phi}\left(e^{-i\lambda_j}\right)w_{j,\widetilde{x}^*}, \tag{5.5.9}$$

where

$$\widetilde{\Phi}\left(e^{-i\lambda_j}\right) = 1+\widetilde{\varphi}_1 e^{-i\lambda_j}+...+\widetilde{\varphi}_n e^{-in\lambda_j}, \tag{5.5.10}$$

with

$$\widetilde{\varphi}_l = \frac{1}{n}\sum_{j=-\tilde{n}+1}^{\tilde{n}-1} h^{\frac{1}{2}}\left(\lambda_j;\widetilde{\theta}\right)e^{i\ell\lambda_j}, \tag{5.5.11}$$

for $l = 1,...,n$, while $\widetilde{\sigma}_\varepsilon^2$ and $\widetilde{\theta}$ are given in (5.2.11).

Step 3: For $j = 1,...,\tilde{n}$, compute the periodogram of the bootstrap sample $\{x_1^*,...,x_n^*\}$ evaluated at the Fourier frequencies

$$I_j^* = \frac{1}{2\pi n}\left|\sum_{t=1}^{n} x_t^* e^{it\lambda_j}\right|^2, \tag{5.5.12}$$

and consider the Whittle objective function

$$Q^*\left(\alpha,\theta,s\right) = \frac{2\pi}{\tilde{n}}\sum_{j=1}^{\tilde{n}}\frac{I_j^*}{g_j\left(\alpha,\lambda_s\right)h_j\left(\theta\right)}. \tag{5.5.13}$$

To obtain the bootstrap analogue of $\mathcal{T}_W$, consider for all $s = 0, 1, ..., \tilde{n}$

$$\left(\widehat{\alpha}_s^*, \widehat{\theta}^{*\prime}\right)' = \underset{(\alpha,\theta')' \in \Pi \times \Theta}{\arg\min}\; Q^*(\alpha, \theta, s) \quad \text{and} \quad \widehat{\sigma}_\varepsilon^{*2} = Q^*\left(\widehat{\alpha}_s^*, \widehat{\theta}^*, s\right), \tag{5.5.14}$$

whereas to obtain the corresponding bootstrap for $\mathcal{T}_{LM}$, let

$$\widetilde{\theta}^* = \arg\min_{\theta\in\Theta} \frac{2\pi}{\tilde{n}} \sum_{j=1}^{\tilde{n}} \frac{I_j^*}{h_j(\theta)} \quad \text{and} \quad \widetilde{\sigma}_\varepsilon^{*2} = \frac{2\pi}{\tilde{n}} \sum_{j=1}^{\tilde{n}} \frac{I_j^*}{h_j\left(\widetilde{\theta}^*\right)}. \tag{5.5.15}$$

Then, for all $s = 0, 1, ..., \tilde{n}$, denote

$$q^*(\alpha, \theta, s) = \frac{1}{\widetilde{\sigma}_\varepsilon^{*2}\tilde{n}} \sum_{j=1}^{\tilde{n}} q_j(\alpha, \theta, s) \frac{I_j^*}{h_j(\theta)}, \tag{5.5.16}$$

where $q_j(.)$ is given in (5.2.14). From here, for all $s = 0, 1, ..., \tilde{n}$, compute

$$\widetilde{q}_s^* = \left(V_{(1,1)}^{-1}\left(0, \widetilde{\theta}^*, s\right)\right)^{\frac{1}{2}} \tilde{n}^{\frac{1}{2}} q_{(1)}^*\left(0, \widetilde{\theta}^*, s\right), \tag{5.5.17}$$

recalling that $q^{(1)*}(.)$ denotes the first element of the vector $q^*(.)$ and the matrix $V(.)$ is defined in (5.2.13).

Step 4: Compute the bootstrap test statistics

$$\mathcal{T}_W^* = \sup_{s=0,1,...,\tilde{n}} \widehat{\alpha}_s^*, \tag{5.5.18}$$

$$\mathcal{T}_{LM}^* = \sup_{s=0,1,...,\tilde{n}} -\widetilde{q}_s^*. \tag{5.5.19}$$

The bootstrap scheme, described in Steps 1 to 4 above, is similar to the residual-based bootstrap of Franke and Kreiss (1992) and the sieve bootstrap of Bühlmann (1997), but contrary to them, it is performed in the frequency domain and does not require the choice of initial values. If the coefficients φ_l, $l = 1, ..., n$, were easily obtained from the vector of parameters θ_0, say $\varphi_\ell(\theta_0)$, then $\widetilde{\Phi}\left(e^{-i\lambda_j}\right)$ in Step 2 can be computed as

$$\widetilde{\Phi}\left(e^{-i\lambda_j}\right) = 1 + \varphi_1\left(\widetilde{\theta}\right) e^{-i\lambda_j} + ... + \varphi_n\left(\widetilde{\theta}\right) e^{-in\lambda_j}. \tag{5.5.20}$$

An example of the latter is when $\{x_t\}_{t\in\mathbb{Z}}$ follows an $AR(1)$ model.

One possible criticism, when compared to residual-based and sieve bootstraps, is that we do our resampling from the standardized original data $\{x_1, ..., x_n\}$ and not from estimates of the innovations $\{\varepsilon_1, ..., \varepsilon_n\}$. Because of that, we envisage that we can modify our algorithm to allow bootstrapping from estimates of $\{\varepsilon_1, ..., \varepsilon_n\}$. This method to bootstrap the data $\{x_1, ..., x_n\}$ may be preferable over that given above, as it may capture higher order moment properties of $\{x_t\}_{t\in\mathbb{Z}}$.

Remark 5.1

The bootstrap algorithm given by Steps 1 to 4 can be modified as follows. For $t = 1, ..., n$, compute the residuals

$$\widehat{\varepsilon}_t = n^{-\frac{1}{2}} \sum_{j=1}^{n} e^{it\lambda_j} \widetilde{\Phi}\left(e^{i\lambda_j}\right) \left|\widetilde{\Phi}\left(e^{-i\lambda_j}\right)\right|^{-2} w_j, \tag{5.5.21}$$

and obtain the standardized residuals

$$\widetilde{\varepsilon}_t = \frac{\widehat{\varepsilon}_t - \overline{\widehat{\varepsilon}}}{\sigma_{\widehat{\varepsilon}}}, \quad \text{with } \overline{\widehat{\varepsilon}} = \frac{1}{n}\sum_{t=1}^{n} \widehat{\varepsilon}_t,\ \sigma_{\widehat{\varepsilon}}^2 = \frac{1}{n}\sum_{t=1}^{n}\left(\widehat{\varepsilon}_t - \overline{\widehat{\varepsilon}}\right)^2. \tag{5.5.22}$$

Then in Step 1, instead of resampling with replacement from $\{\check{x}_1, ..., \check{x}_n\}$ to obtain $\{\widetilde{x}_1^*, ..., \widetilde{x}_n^*\}$, we resample with replacement from with $\{\widetilde{\varepsilon}_1, ..., \widetilde{\varepsilon}_n\}$ obtaining a bootstrap sample $\{\widetilde{\varepsilon}_1^*, \widetilde{\varepsilon}_2^*, ..., \widetilde{\varepsilon}_n^*\}$. Then, for $j = 1, ..., n$, we compute

$$w_{j,\widetilde{\varepsilon}^*} = n^{-\frac{1}{2}} \sum_{t=1}^{n} \widetilde{\varepsilon}_t^* e^{-it\lambda_j}. \tag{5.5.23}$$

Hereafter, proceed as in Steps 2 to 4 replacing $\{w_{1,\widetilde{x}^*}, ..., w_{n,\widetilde{x}^*}\}$ with $\{w_{1,\widetilde{\varepsilon}^*}, ..., w_{n,\widetilde{\varepsilon}^*}\}$.

Next, we establish ,under Conditions C.1-C.6, the asymptotic distribution of our bootstrap test statistics $\mathcal{T}_W^*$ and $\mathcal{T}_{LM}^*$ under $H_0 \cup H_1$ *given in* (5.2.3) and (5.2.5).

Theorem 5.2

Assume that Conditions C.1-C.6 are satisfied. Then, under $H_0 \cup H_1$ given in (5.2.3) and (5.2.5), we have that, as $n \to \infty$,

i) $$\tilde{n}^{\frac{1}{2}} \mathcal{T}_W^* \xrightarrow{d^*} \max_{\tau\in[0,1]} \mathcal{G}(\tau), \quad \text{in probability,} \tag{5.5.24}$$

ii) $$\tilde{n}^{\frac{1}{2}} \mathcal{T}_{LM}^* \xrightarrow{d^*} \max_{\tau\in[0,1]} \mathcal{G}(\tau), \quad \text{in probability.} \tag{5.5.25}$$

Theorem 5.2 indicates that the bootstrap test statistics given in $\mathcal{T}_W^*$ and $\mathcal{T}_{LM}^*$ are consistent. That is, let $c_{n,1-\beta}^{f,W}$ $\left(c_{n,1-\beta}^{f,LM}\right)$ and $c_{1-\beta}^{a}$ be such that

$$\Pr\left(\tilde{n}^{\frac{1}{2}} \mathcal{T}_W > c_{n,1-\beta}^{f,W}\right) = \beta \quad \left(\Pr\left(\tilde{n}^{\frac{1}{2}} \mathcal{T}_{LM} > c_{n,1-\beta}^{f,LM}\right) = \beta\right) \tag{5.5.26}$$

and, as $n \to \infty$,

$$\Pr\left(\tilde{n}^{\frac{1}{2}} \mathcal{T}_W > c_{1-\beta}^{a}\right) \to \beta \quad \left(\Pr\left(\tilde{n}^{\frac{1}{2}} \mathcal{T}_{LM} > c_{1-\beta}^{a}\right) \to \beta\right), \tag{5.5.27}$$

respectively. Thus, Theorems 5.1 and 5.2 imply that $c_{n,1-\beta}^{f,W} \to c_{1-\beta}^{a}$ $\left(c_{n,1-\beta}^{f,LM} \to c_{1-\beta}^{a}\right)$ and $c_{1-\beta}^{*,W} \xrightarrow{p} c_{1-\beta}^{a}$ $\left(c_{1-\beta}^{*,LM} \xrightarrow{p} c_{1-\beta}^{a}\right)$, respectively, where $c_{1-\beta}^{*,W}$ $\left(c_{1-\beta}^{*,LM}\right)$ is defined by

$$\Pr\left(\tilde{n}^{\frac{1}{2}} \mathcal{T}_W^* > c_{1-\beta}^{*,W}\right) = \beta \quad \left(\Pr\left(\tilde{n}^{\frac{1}{2}} \mathcal{T}_{LM}^* > c_{1-\beta}^{*,LM}\right) = \beta\right). \tag{5.5.28}$$

Typically, the finite sample distribution of $\tilde{n}^{\frac{1}{2}}\mathcal{T}_W^*$ $\left(\tilde{n}^{\frac{1}{2}}\mathcal{T}_{LM}^*\right)$ is not available, so that the critical value $c_{1-\beta}^{*,W}$ $\left(c_{1-\beta}^{*,LM}\right)$ is obtained by standard Monte-Carlo simulations, which can be approximated as accurately as desired. To that end, consider the b-th bootstrap sample $\widetilde{x}^{*(b)} = \left\{\widetilde{x}_1^{*(b)}, ..., \widetilde{x}_n^{*(b)}\right\}$, $b = 1, ..., B$, and compute $\mathcal{T}_W^{*(b)}$ $\left(\mathcal{T}_{LM}^{*(b)}\right)$ as in (5.5.18) (in (5.5.19)) for each $b = 1, ..., B$. Then, $c_{1-\beta}^{*,W}$ $\left(c_{1-\beta}^{*,LM}\right)$ is approximated by the value $c_{1-\beta}^{*B,W}$ $\left(c_{1-\beta}^{*B,LM}\right)$ satisfying

$$\frac{1}{B}\sum_{b=1}^{B}\mathcal{I}\left(\tilde{n}^{\frac{1}{2}}\mathcal{T}_W^{*(b)} > c_{1-\beta}^{*B,W}\right) = \beta \quad \left(\frac{1}{B}\sum_{b=1}^{B}\mathcal{I}\left(\tilde{n}^{\frac{1}{2}}\mathcal{T}_{LM}^{*(b)} > c_{1-\beta}^{*B,LM}\right) = \beta\right). \quad (5.5.29)$$

5.6 Monte-Carlo simulations

In order to investigate the finite sample performance of our testing procedures, we perform Monte-Carlo simulations. Throughout our Monte Carlo experiments, we employ $2,000$ replications with samples sizes $n = 64, 128$. For the calculation of the bootstrap critical values we generate $1,000$ bootstrap samples, that is, we choose $B = 1,000$ in (5.5.29), while we consider the significance levels $\beta = 10\%, 5\%, 1\%$. We employ the Davies and Harte (1987) algorithm to generate any Gaussian $ARFIMA(0, d, 0)$ and $GARMA(0, d_\omega, 0)$ components of the models simulated below.

To assess the empirical size of the W type of test $\mathcal{T}_W^*$, we simulate Gaussian i.i.d. processes. The empirical power of $\mathcal{T}_W^*$ is examined under the Gaussian $ARFIMA\,(0, d, 0)$ model with $d = 0.1, 0.2, 0.3, 0.4$. Under these sets of models, we also investigate the finite sample performance of the LM type of test $\mathcal{T}_{LM}^*$. Furthermore, we explore the empirical size of $\mathcal{T}_{LM}^*$ under the Gaussian $AR\,(1)$ and $MA\,(1)$ models with autoregressive parameter $a_1 = 0.5$ and moving average parameter $b_1 = 0.5$, respectively, and its empirical power under the Gaussian $GARMA(0, d_\omega, 0)$, $ARFIMA(1, d, 0)$, $ARFIMA(0, d, 1)$, $GARMA(1, d_\omega, 0)$ and $GARMA(0, d_\omega, 1)$ models with parameters $a_1 = 0.5$, $b_1 = 0.5$, $d, d_\omega = 0.1, 0.2, 0.3, 0.4$ and $\omega = \frac{\pi}{2}$.

When $\{x_t\}_{t\in\mathbb{Z}}$ is simulated as either an i.i.d. sequence or an $ARFIMA(0, d, 0)$ model or a $GARMA(0, d_\omega, 0)$ model, we have under the null hypothesis H_0 that $\{x_t\}_{t\in\mathbb{Z}}$ is a sequence of i.i.d. random variables. Hence, the bootstrap procedure is performed by resampling randomly with replacement from the data. On the other hand, when $\{x_t\}_{t\in\mathbb{Z}}$ is simulated as an $AR(1)$ $(MA\,(1))$ process, the residuals from the $AR(1)$ $(MA\,(1))$ model can be easily obtained. This is also the case when $\{x_t\}_{t\in\mathbb{Z}}$ is taken to follow an $ARFIMA(1, d, 0)$ $(ARFIMA(0, d, 1))$ model or a $GARMA(1,d_\omega,0)$ $(GARMA(0,d_\omega,1))$ model, since $\{x_t\}_{t\in\mathbb{Z}}$ becomes an $AR(1)$ $(MA(1))$ process under the null hypothesis H_0. Hence, when the simulated model involves an $AR(1)$ $(MA(1))$ short memory component, the following bootstrap procedure can be performed.

Step 1′: Let $e_t = x_t - \widetilde{a}_1 x_{t-1}$ $\left(e_t = x_t - \widetilde{b}_1 e_{t-1}\right)$, for $t = 2,...,n$, and $e_1 = (1-\widetilde{a}_1^2)^{\frac{1}{2}} x_1$ $(e_1 = x_1)$, where $\widetilde{a}_1$ $\left(\widetilde{b}_1\right)$ is the PW estimator of a_1 (b_1) given by (5.2.11). Notice that in this model $\theta_0 = a_1$ $(\theta_0 = b_1)$. For $t = 2,...,n$, compute the standardized residuals

$$\breve{e}_t = \frac{e_t - \overline{e}}{\overline{\sigma}_e}, \quad \text{with } \overline{e} = \frac{1}{n}\sum_{t=1}^{n} e_t,\ \overline{\sigma}_e^2 = \frac{1}{n}\sum_{t=1}^{n}(e_t - \overline{e})^2. \tag{5.6.1}$$

Let $\{\widetilde{e}_1^*,...,\widetilde{e}_n^*\}$ be a random sample with replacement from the empirical distribution of $\{\breve{e}_1,...,\breve{e}_n\}$, and obtain the bootstrap sample $\{x_1^*,...,x_n^*\}$ as

$$\begin{aligned} x_t^* &= \widetilde{a}_1 x_{t-1}^* + \widetilde{e}_t^* \ \left(x_t^* = \widetilde{e}_t^* + \widetilde{b}_1 \widetilde{e}_{t-1}^*\right), \quad t = 2,...,n \\ x_1^* &= \left(1-\widetilde{a}_1^2\right)^{-\frac{1}{2}} \widetilde{e}_1^* \ (x_1^* = \widetilde{e}_1^*). \end{aligned} \tag{5.6.2}$$

Step 2′: Exactly as Step 3 in Section 5.5, but with $\{x_1^*,...,x_n^*\}$ as generated in Step 1′. Then, for all $s = 0,1,...,\tilde{n}$, compute

$$\widetilde{q}_s^* = \left(V_{(1,1)}^{-1}\left(0,\widetilde{\theta}^*,s\right)\right)^{\frac{1}{2}} \tilde{n}^{\frac{1}{2}} q_{(1)}^*\left(0,\widetilde{\theta}^*,s\right), \tag{5.6.3}$$

noticing that in this model $\widetilde{\theta}^* = \widetilde{a}_1^*$ $\left(\widetilde{\theta}^* = \widetilde{b}_1^*\right)$.

Step 3′: Compute the bootstrap statistic

$$\mathcal{T}_{LM}^* = \sup_{s=0,1,...,\tilde{n}} -\widetilde{q}_s^*. \tag{5.6.4}$$

Observe that the bootstrap scheme described in Steps 1′ to 3′ generates the bootstrap sample $\{x_1^*,...,x_n^*\}$ according to the residual based bootstrap of Franke and Kreiss (1992). We refer to our bootstrap algorithm in Section 5.5 as Method 1, which we compare to the one given by Steps 1′ to 3′, referred to as Method 2. The results of our experiments are given in Tables 5.1-5.32 found in Appendix 5.C.

Tables 5.1-5.4 present the empirical size of the $\mathcal{T}_W^*$ and $\mathcal{T}_{LM}^*$ tests, whereas Tables 5.5-5.32 illustrate the empirical power of the $\mathcal{T}_W^*$ and $\mathcal{T}_{LM}^*$ tests, for the various models described above. Overall, the empirical size and power of the $\mathcal{T}_W^*$ and $\mathcal{T}_{LM}^*$ tests are very satisfactory, and the performance of both tests improves with an increase in the sample size. Notice also that the bootstrap method described in Section 5.5 appears to be a good competitor to the residual based bootstrap of Franke and Kreiss (1992), as the empirical size and power of the $\mathcal{T}_{LM}^*$ test are on the same level for the two methods, see Tables 5.3 and 5.4, and Tables 5.17-5.32. As Tables 5.1 and 5.2 illustrate, in terms of the empirical size, the $\mathcal{T}_{LM}^*$ test tends to perform better than the $\mathcal{T}_W^*$ type test, in line with the consensus that LM type of tests have better size performance as compared to W type of tests. On the contrary, our simulations results go against the principle that W type of tests have better power against LM type of tests, see Tables 5.5-5.8 and Tables 5.9-5.12. This observation is probably driven by the fact that our W type of test is based on the PW estimator $\widehat{\alpha}_s$, $s = 0, 1, ..., \widetilde{n}$, which tends to be downwards biased in finite samples, see Giraitis, Hidalgo, and Robinson (2001).

Regarding the size of the $\mathcal{T}_{LM}^*$ test, we observe that its performance does not seem to be affected by the various short-run models, see Tables 5.2-5.4. When we consider the power performance of the $\mathcal{T}_{LM}^*$ test relative to the short memory component and the frequency of the persistent component, we observe the following. Firstly, the power is low when the short memory component follows an $AR(1)$ model and $\omega = 0$, see Tables 5.17-5.20. This is consistent with the empirical observation that an $AR(1)$ model with a high value of a_1 and an $ARFIMA(0, d, 0)$ model with $d > 0$ cannot be easily discriminated. This is also the conclusion obtained from the Monte-Carlo experiment considered by Lobato and Robinson (1998). However, notice that when the short memory component follows a $MA(1)$ model and $\omega = 0$ the power is quite high, see Tables 5.25-5.28, and is slightly lower that in the case with no short-run component, see Tables 5.9-5.12. Secondly, as the frequency ω moves away from 0, we see that the influence of the short-run $AR(1)$ or $MA(1)$ component seems to be negligible, see Tables 5.21-5.24 and 5.29-5.32, as compared to the case of no short-run component, see Tables 5.13-5.16. This may be somehow expected, as the location of the maximum of the spectral density function for an $AR(1)$ model with positive a_1 and our $GARMA(0, d_\omega, 0)$ models with $\omega = \frac{\pi}{2}$ are very different. Observe also that the empirical power is higher when $\omega = \frac{\pi}{2}$ than when $\omega = 0$, compare Tables 5.9-5.12, 5.17-5.20 and 5.25-5.28 against Tables 5.13-5.16, 5.21-5.24 and 5.29-5.32, respectively. Finally, it is worth noticing that for both $\mathcal{T}_W^*$ and $\mathcal{T}_{LM}^*$ tests, the empirical power increases with increases in d or d_ω, see Tables 5.5-5.32. The latter is expected as the "distance" between the null and alternative hypotheses becomes greater as d or d_ω becomes bigger.

We should also mention that we have performed Monte-Carlo simulations with the bootstrap method suggested in Remark 5.1. However, we did not find any significant difference with the results presented here, even in the case that the distribution of $\{x_t\}_{t\in\mathbb{Z}}$ was chosen to follow a χ_1^2 or a χ_8^2.

5.7 Empirical applications

We now apply our testing procedures to various data sets to assess whether they exhibit persistent components. We examine the rate of growth of industrial production and unemployment rate. An important complication in the implementation of our testing procedures is the choice of the short-run component of the data. It is common in empirical applications to consider that the latter component follows an $ARMA(p,q)$ model. The orders p and q of an $ARMA(p,q)$ model are usually chosen on the basis on some information criterion, like the AIC of Akaike (1973), the BIC of Schwarz (1978) and the HIC of Hannan and Quinn (1979), the latter applicable when $q = 0$. Beran, Bhansali, and Ocker (1998) considered these three information criteria for the $ARFIMA(p,d,0)$ model and showed that the AIC does not provide a consistent estimate of p, but the BIC and HIC do. Hence, if one is certain that the possible persistent component of the data is the long-run one, the BIC and HIC methods can be employed when $q = 0$. The results of Beran, Bhansali, and Ocker (1998) are likely to be extendable for our model (5.3.1)-(5.3.3) with short-run component following an $ARMA(p,q)$ model, but the implementation of the information criterions is likely to be computationally intensive. For this reason, we choose instead various models for the short-run component of the data, and in particular, we consider i.i.d., $AR(1)$, and $MA(1)$ models. The results on the empirical applications are presented in Appendix 5.C.

5.7.1 Industrial production

The series of industrial production for the U.S. has been found by various authors to exhibit strong long-run component, see for example Nelson and Plosser (1982) and Gil-Alaña and Robinson (1997). This strong long-run behavior is clear from the data of the series, found in Figure 5.1, where we employ monthly data for the U.S. spanning the period 1960M1-2006M5[1]. As the data clearly exhibit nonstationary behavior, we consider examining the growth rate of industrial production, measured in percentage points and calculated as the first differences of the logarithmic of

[1]Industrial production data are taken from Board of Governors of the Federal Reserve System, Series I.D. INDPRO, http://www.federalreserve.gov.

industrial production, see Figure 5.2. It is clear from Figure 5.2 that the growth rate of industrial production does not exhibit any breaks in its mean level, but seems to be less volatile in the post 1984 period. This is also the conclusion derived by van Dijk, Osborn, and Sensier (2002), who employed tests for changes in the mean level and volatility of various economic time series, and in the case of the U.S. industrial production growth rate report no change in the mean level of the series, and one change in its volatility occurring around 1984M3. The vertical line in Figures 5.1 and 5.2 correspond to this date.

Thus, we consider applying our testing procedures to the full sample of the data and two subsamples, the first one covering the period 1960M1-1984M3, and the second one spanning the period 1984M4-2006M4. The results are given in Tables 5.33-5.35. For the full sample period, see Table 5.33, the null hypothesis of no persistent component is rejected at all significance levels and under all short-run models considered. For the sample covering the period 1960M1-1984M3, see Table 5.34, the null hypothesis is rejected at all significance levels, when the short-run component is i.i.d. or a $MA(1)$ model. In the case of an $AR(1)$ short-run component, the $\mathcal{T}_W^*$ statistic rejects the null hypothesis at the 10% and 5% significance levels, but fails to reject at the 1% significance level, while the $\mathcal{T}_{LM}^*$ statistic fails to reject the null hypothesis at all significance levels. Finally, for the sample period 1960M1-1984M3, see Table 5.35, the null hypothesis is rejected in all but one case; the $\mathcal{T}_W^*$ statistic fails to reject the null hypothesis at the 1% significance level and when the short-run component follows a $MA(1)$ model.

Given the overwhelming evidence against the null hypothesis, and our simulation results in Section 5.6 on the low power of our tests in the case of a short-run $AR(1)$ component, we conclude that the data exhibit a persistent component for the various sample periods considered. Hence, the growth rate of industrial production exhibits a persistent component, so that $ARMA(p,q)$ models are not likely to provide a good fit of these data.

5.7.2 Unemployment rate

There are conflicting evidence as to whether the series of unemployment rate for the U.S. exhibits a strong long-run component, see for example Nelson and Plosser (1982), Gil-Alaña and Robinson (1997) and Papell, Murray, and Ghiblawi (2000). Various authors have also reported breaks in the unemployment rate both in its mean and variance. In particular, Papell, Murray, and Ghiblawi (2000) found that the series of unemployment rate is subject to two breaks in its mean, with the dates of the breaks being 1974 and 1986, while van Dijk, Osborn, and Sensier (2002) report a break in the variance of the growth of the unemployment rate at 1986M2. Here, we employ monthly data for the U.S. spanning the period 1960M1-2006M5[2]. Figure 5.3 plots that data, while Figure 5.4 plots the growth of the unemployment rate, calculated as the first differences, while the vertical lines in the figures correspond to the dates of the breaks.

We apply our testing procedures to the full sample of the data and three subsamples, the first one covering the period 1960M1-1973M12, the second one spanning the period 1974M1-1986M2, and the third one extending the period 1986M3-200M5. The results are presented in Tables 5.36-5.39. For the full sample period, see Table 5.36, the null hypothesis of no persistent component in the data is rejected in all but one case; the $\mathcal{T}_W^*$ statistic fails to reject the null hypothesis at the 1% significance level and when the short-run component follows an $AR(1)$ model. For the sample covering the period 1960M1-1973M12, see Table 5.37, the null hypothesis is rejected at all significance levels, when the short-run component is i.i.d. or a $MA(1)$ model. In the case of an $AR(1)$ short-run component, the $\mathcal{T}_W^*$ statistic fails to reject the null hypothesis at all significance levels, while the $\mathcal{T}_{LM}^*$ statistic rejects the null hypothesis at the 10% and 5% significance levels, but fails to reject at the 1% significance level. For sample period 1974M1-1986M2, see Table 5.38, the null hypothesis is rejected at all significance levels, when the short-run component is an i.i.d. or $MA(1)$ model. The $\mathcal{T}_W^*$ statistic fails to reject the null hypothesis at all significance levels, when an $AR(1)$ model is assumed for the short-run component, but the $\mathcal{T}_{LM}^*$ statistic rejects the null hypothesis in this case at all significance levels. Finally, for the sample spanning the period 1986M3-2006M5, see Table 5.39, we reject the null hypothesis at all significance levels when an i.i.d. or $MA(1)$ model is chosen for the short-run. However, when the short-run component follows an $AR(1)$ model, both tests fail to reject the null hypothesis at all significance levels.

For the full sample and the first two subsamples, there are strong evidence against the null hypothesis, taking into account the results of the Monte-Carlo simulations in Section 5.6 on the power of our tests in the case of an $AR(1)$ short-run component. Hence, we conclude that for these periods, the unemployment rate exhibits a persistent component. The results for the last subsample are mixed, and therefore a solid conclusion cannot be made. The data for the last subperiod

[2]Unemployment rate data are taken from U.S. Department of Labor: Bureau of Labor Statistics, Series I.D. UNRATE, http://stats.bls.gov.

certainly require further investigation.

5.8 Final comments

In this chapter, we have presented two testing procedures for the null hypothesis of no persistent component in the data against the alternative that the data exhibits a persistent component. Our methodologies have been based on the W and LM principles and involved the PW estimation method. We have derived the asymptotic distribution of our test statistics for a wide class of linear processes having parametrically specified spectral density function, and moreover we have established their consistency and power against local alternatives. As our test statistics were found to have an asymptotic distribution that is nonstandard and model dependent, we have introduced a bootstrap scheme for the purposes of calculating valid critical values, and furthermore we have established its validity. The finite sample performance of our testing procedures has been investigated by the means of Monte-Carlo simulations, and has been found overall to be very satisfactory. Finally, we have applied our testing methods to data for the growth rate of industrial production and unemployment rate, and we have found evidence that these series exhibit persistent components for most of the time periods considered. In the practical implementation of our testing procedures, the issue of choosing the appropriate model for the short-run component of the data arised. Clearly, this is an important problem that we hope to address in the future.

Although we have only considered the situation when $\{x_t\}_{t\in\mathbb{Z}}$ is observed, it appears that the same results should hold true when $\{x_t\}_{t\in\mathbb{Z}}$ are the errors of a regression model. That is, consider the linear regression model $y_t = \beta' z_t + x_t$, $t = 1, ..., n$, where $\{x_t\}_{t\in\mathbb{Z}}$ follows a $ARFIMA\,(p, d, q)$. When $d = 0$, it is well known that, under suitable regularity conditions, the GLS estimator $\widehat{\beta}_{GLS}$ of β, and the PW estimator of the parameters of the $ARMA\,(p, q)$ process of $\{x_t\}_{t\in\mathbb{Z}}$ are (asymptotically) independent. Moreover, Robinson and Hidalgo (1997) showed that the same holds when $\{x_t\}_{t\in\mathbb{Z}}$ exhibits strong dependence. So, this observation leads us to think that the results obtained in the paper are likely to hold true when $\widehat{x}_t = y_t - \widehat{\beta}'_{GLS} z_t$ is used instead of the unobserved $\{x_t\}_{t\in\mathbb{Z}}$.

We finish mentioning two more issues. Firstly, although we have only considered stationary processes, our tests statistics should also detect nonstationary alternatives. Velasco and Robinson (2000) established, under a certain class of nonstationary linear models, that the PW estimator remains $n^{\frac{1}{2}}$-consistent and asymptotically normally distributed, if tapering of the data is employed. Hence, following Velasco and Robinson (2000), we expect that the same type of results presented here, hold under nonstationary alternatives. Secondly, we have concentrated on linear models, but it would be of theoretical and empirical interest to extend our results for nonlinear models. However, this task is of great difficulty, as for nonlinear models, the PW estimator is not always $n^{\frac{1}{2}}$-consistent and asymptotically normally distributed, see Giraitis and Taqqu (1999). Moreover, bootstrap algorithms for strongly dependent time series data tend to concentrate on the linear case, and are likely to require knowledge of the nonlinear model.

5.A Appendix

This Section contains the proofs which use a series of lemmas found in Appendix 5.B below.

Proof of Theorem 5.1. i) First notice that for a given λ_s, the closest Fourier frequency to ω_0, it follows under Conditions C.1-C.6 that, as $n \to \infty$,

$$\left(\widehat{\alpha}_s, \left(\widehat{\theta} - \theta_0\right)'\right)' \xrightarrow{p} 0 \quad \text{and} \quad n^{\frac{1}{2}} \left(\widehat{\alpha}_s, \left(\widehat{\theta} - \theta_0\right)'\right)' = O_p(1), \tag{5.A.1}$$

see for example Velasco and Robinson (2000), among others. Next, denote $\ell_{j,n}(\alpha, \theta, s) = \frac{I_j}{g_j(\alpha,\lambda_s)h_j(\theta)}$. Then, proceeding as Giraitis, Hidalgo, and Robinson (2001), we have that for all $s = 0, 1, ..., \tilde{n}$,

$$\begin{aligned}\sum_{j=1}^{\tilde{n}} \ell_{j,n}(\alpha, \theta, s) &= \sum_{j=1}^{\tilde{n}} \ell_{j,n}(0, \theta_0, s) + \sum_{j=1}^{\tilde{n}} (D\ell_{j,n}(0, \theta_0, s))' \left(\alpha_s, (\theta - \theta_0)'\right)' \\ &\quad + \frac{1}{2} \left(\alpha_s, (\theta - \theta_0)'\right) \sum_{j=1}^{\tilde{n}} D^2\ell_{j,n}(0, \theta_0, s) \left(\alpha_s, (\theta - \theta_0)'\right)' \\ &\quad + R_n(\alpha, \theta, s),\end{aligned} \tag{5.A.2}$$

where D and D^2 denote the first and second generalized derivatives as defined in Andrews (1999), respectively, and

$$\sup_{(\alpha,\theta')'\in\Pi\times\Theta:\ n^{\frac{1}{2}}|\alpha|\leq\gamma;\ n^{\frac{1}{2}}\|\theta-\theta_0\|<\gamma} |R_n(\alpha, \theta, s)| = o_p(1), \tag{5.A.3}$$

for any $0 < \gamma < \infty$. Let $\Rightarrow$ denote weak convergence of a sequence of stochastic processes indexed by $\tau \in [0, 1]$. We show below that

$$\tilde{n}^{-\frac{1}{2}} \sum_{j=1}^{\tilde{n}} (D\ell_{j,n}(0, \theta_0, [\tilde{n}.]))' \Rightarrow X(.) \tag{5.A.4}$$

and

$$\frac{1}{\tilde{n}}\sum_{j=1}^{\tilde{n}} D^2\ell_{j,n}\left(0,\theta_0,[\tilde{n}.]\right) \Rightarrow A\left(.\right). \tag{5.A.5}$$

Then, by (5.A.1)-(5.A.5), and proceeding as in Theorem 2 of Andrews (2001), the finite-dimensional limit distributions are those from minimization of

$$\left(c - X\left(\tau\right)\right)' \mathcal{C}^{-1}\left(\tau,\tau\right)\left(c - X\left(\tau\right)\right) \tag{5.A.6}$$

with the constraint that $c \geq 0$. Then, the result follows by the continuous mapping theorem, since sup is a continuous functional in $\mathbb{C}\left[0,1\right]$.

Now, we show (5.A.4) and (5.A.5). Notice that under Conditions C.1, C.5 and C.6, an obvious extension of Lemma 3 of Delgado, Hidalgo, and Velasco (2005) implies that

$$\sup_{s=0,1,\ldots,\tilde{n}}\left|\tilde{n}^{-\frac{1}{2}}\sum_{j=1}^{\tilde{n}}\begin{pmatrix} D\log g_j\left(0,\lambda_s\right) \\ D\log h_j\left(\theta_0\right)\end{pmatrix}\left(\frac{I_j}{h_j\left(\theta_0\right)} - I_{j,\varepsilon}\right)\right| = o_p\left(1\right) \tag{5.A.7}$$

and

$$\sup_{s=0,1,\ldots,\tilde{n}}\left|\tilde{n}^{-\frac{1}{2}}\sum_{j=1}^{\tilde{n}}\begin{pmatrix} D\log g_j\left(0,\lambda_s\right) \\ D\log h_j\left(\theta_0\right)\end{pmatrix}\begin{pmatrix} D\log g_j\left(0,\lambda_s\right) \\ D\log h_j\left(\theta_0\right)\end{pmatrix}' \times\left(\frac{I_j}{h_j\left(\theta_0\right)} - I_{j,\varepsilon}\right)\right| = o_p\left(1\right) \tag{5.A.8}$$

Recall that $\pi^{-1}\lambda_s \to \tau \in \left[0,1\right]$, as $n \to \infty$. Hence, to establish (5.A.4) and (5.A.5), it suffices to show that

$$X_n\left(.\right) \Rightarrow X\left(.\right) \tag{5.A.9}$$

and

$$A_n\left(.\right) \Rightarrow A\left(.\right), \tag{5.A.10}$$

where

$$X_n\left(\tau\right) = \tilde{n}^{-\frac{1}{2}}\sum_{j=1}^{\tilde{n}}\begin{pmatrix} D\log g_j\left(0,\lambda_{[\tilde{n}\tau]}\right) \\ D\log h_j\left(\theta_0\right)\end{pmatrix}\left(I_{j,\varepsilon} - \frac{\sigma_{0,\varepsilon}^2}{2\pi}\right) \tag{5.A.11}$$

and

$$A_n\left(\tau\right) = \frac{1}{\tilde{n}}\sum_{j=1}^{\tilde{n}}\begin{pmatrix} D\log g_j\left(0,\lambda_{[\tilde{n}\tau]}\right) \\ D\log h_j\left(\theta_0\right)\end{pmatrix}\begin{pmatrix} D\log g_j\left(0,\lambda_{[\tilde{n}\tau]}\right) \\ D\log h_j\left(\theta_0\right)\end{pmatrix}' I_{j,\varepsilon}. \tag{5.A.12}$$

We begin with the proof of (5.A.9). It suffices to show that, for any finite collection $\tau_{\ell_1},\ldots,\tau_{\ell_p}$,

$$\left(X_n\left(\tau_{\ell_1}\right),\ldots,X_n\left(\tau_{\ell_p}\right)\right)' \xrightarrow{d} \left(X\left(\tau_{\ell_1}\right),\ldots,X\left(\tau_{\ell_p}\right)\right)', \tag{5.A.13}$$

where $(X(\tau_{\ell_1}),...,X(\tau_{\ell_p}))' \simeq N(0,A)$, with $\simeq$ read as distributed, and the $(\tau_{\ell_1},\tau_{\ell_2})$-th element of A is given by

$$\int_0^1 \begin{pmatrix} D\log g(\pi x, 0, \pi\tau_{\ell_1}) \\ D\log h(\pi x;\theta_0) \end{pmatrix} \begin{pmatrix} D\log g(\pi x, 0, \pi\tau_{\ell_2}) \\ D\log h(\pi x;\theta_0) \end{pmatrix}' dx, \qquad (5.A.14)$$

and moreover, that the process $X_n(\tau)$ is tight in the space $\mathbb{D}[0,1]$ with the Skorohod's metric, and for every $\varepsilon > 0$, it holds that

$$\Pr(|X(1) - X(\tau)| > \varepsilon) \to 0, \quad \text{as } \tau \to 1. \qquad (5.A.15)$$

Then, Theorem 15.6 of Billingsley (1968) implies (5.A.9).

Notice that (5.A.13) holds true using standard results on PW estimation, see Giraitis, Hidalgo, and Robinson (2001) among others. To prove the tightness of the process $X_n(\tau)$, and since the second component of $X_n(\tau)$ does not depend on τ, it suffices to show the tightness of

$$X_{n(1)}(\tau) = \tilde{n}^{-\frac{1}{2}} \sum_{j=1}^{\tilde{n}} D\log g_j\left(0, \lambda_{[\tilde{n}\tau]}\right)\left(I_{j,\varepsilon} - \frac{\sigma^2_{0,\varepsilon}}{2\pi}\right). \qquad (5.A.16)$$

Because the limit process has continuous paths, see below the proof of (5.A.15) and comments that follow, Theorems 15.4 and 15.6 of Billingsley (1968) imply that it is sufficient to check the Kolmogorov-Chentsov's moment condition

$$\begin{aligned} &E\left(\left|X_{n(1)}(\tau) - X_{n(1)}(\tau_1)\right|^2 \left|X_{n(1)}(\tau_2) - X_{n(1)}(\tau)\right|^2\right) \\ &\le K\left(F(\tau_2) - F(\tau_1)\right)^{1+\delta}, \end{aligned} \qquad (5.A.17)$$

for all $0 \le \tau_1 < \tau < \tau_2 \le 1$, some $\delta > 0$, some generic constant $0 < K < \infty$ and some nondecreasing and continuous function $F(.)$.

First, we observe that we can focus on the case $\tilde{n}^{-1} \le \tau_2 - \tau_1$. If $\tau_2 - \tau_1 < \tilde{n}^{-1}$, then either τ_1 and τ lie in the same subinterval $\left[\frac{p-1}{\tilde{n}}, \frac{p}{\tilde{n}}\right)$, with $p = 1,...,\tilde{n}$, or else τ and τ_2 do; in either of these cases the left hand side of (5.A.17) vanishes. Then, notice that the Cauchy-Schwarz inequality implies that the left hand side of (5.A.17) is bounded by

$$\left(E\left|X_{n(1)}(\tau) - X_{n(1)}(\tau_1)\right|^4\right)^{1/2}\left(E\left|X_{n(1)}(\tau_2) - X_{n(1)}(\tau)\right|^4\right)^{1/2}, \qquad (5.A.18)$$

and since $F(.)$ is a nondecreasing function, we have that

$$\left(F(\tau) - F(\tau_1)\right)\left(F(\tau_2) - F(\tau)\right) \le \left(F(\tau_2) - F(\tau_1)\right)^2. \qquad (5.A.19)$$

Hence to show the tightness condition (5.A.17), it suffices to show that

$$E\left|X_{n(1)}(\tau_2)-X_{n(1)}(\tau_1)\right|^4 \le K\left(F(\tau_2)-F(\tau_1)\right)^{1+\delta}, \tag{5.A.20}$$

for all $0 \le \tau_1 < \tau_2 \le 1$.

Choose $F(\tau)=\tau$, and denote $\zeta_j = I_{j,\varepsilon} - \frac{\sigma^2_{0,\varepsilon}}{2\pi}$. Then,

$$\begin{aligned}
E\left|X_{n(1)}(\tau_2)-X_{n(1)}(\tau_1)\right|^4 &= \frac{1}{\tilde{n}^2}E\left|\sum_{j=1}^{\tilde{n}}\log\left(\frac{g_j\left(-1,\lambda_{[\tilde{n}\tau_1]}\right)}{g_j\left(-1,\lambda_{[\tilde{n}\tau_2]}\right)}\right)\zeta_j\right|^4 \\
&= \frac{1}{\tilde{n}^2}\sum_{j_1,\ldots,j_4=1}^{\tilde{n}}\prod_{\ell=1}^{4}\log\left(\frac{g_{j_l}\left(-1,\lambda_{[\tilde{n}\tau_1]}\right)}{g_{j_l}\left(-1,\lambda_{[\tilde{n}\tau_2]}\right)}\right)E\left(\prod_{\ell=1}^{4}\zeta_{j_l}\right) \\
&= \frac{3}{\tilde{n}^2}\sum_{j_1,\ldots,j_4=1}^{\tilde{n}}\prod_{\ell=1}^{4}\log\left(\frac{g_{j_l}\left(-1,\lambda_{[\tilde{n}\tau_1]}\right)}{g_{j_l}\left(-1,\lambda_{[\tilde{n}\tau_2]}\right)}\right) \\
&\quad \times E\left(\zeta_{j_1}\zeta_{j_2}\right)E\left(\zeta_{j_3}\zeta_{j_4}\right) \\
&\quad +\frac{1}{\tilde{n}^2}\sum_{j_1,\ldots,j_4=1}^{\tilde{n}}\prod_{\ell=1}^{4}\log\left(\frac{g_{j_l}\left(-1,\lambda_{[\tilde{n}\tau_1]}\right)}{g_{j_l}\left(-1,\lambda_{[\tilde{n}\tau_2]}\right)}\right) \\
&\quad \times cum\left(\zeta_{j_1},\zeta_{j_2},\zeta_{j_3},\zeta_{j_4}\right).
\end{aligned} \tag{5.A.21}$$

Theorems 2.3.1 and 4.3.2, and in particular the equation (4.3.15), of Brillinger (1975) imply that the cumulants in (5.A.21) are finite. Since the function $\log^q|\sin x|$, $q=1,\ldots,4$, is integrable, we have that

$$\begin{aligned}
E\left|X_{n(1)}(\tau_2)-X_{n(1)}(\tau_1)\right|^4 &\le K\tilde{n}^{-2}\log^4\tilde{n} \\
&\le K\left(\tau_2-\tau_1\right)^{1+\delta},
\end{aligned} \tag{5.A.22}$$

for any $0<\delta<1$, since $\tilde{n}^{-1} < \tau_2-\tau_1$. On the other hand, Proposition 10.3.2 of Brockwell and Davis (1991) implies that the first term on the right of (5.A.21) is

$$3\sigma^4_{0,\varepsilon}\left(\frac{1}{\tilde{n}}\sum_{j=1}^{\tilde{n}}\log^2\left(\frac{g_{j_l}\left(-1,\lambda_{[\tilde{n}\tau_1]}\right)}{g_{j_l}\left(-1,\lambda_{[\tilde{n}\tau_2]}\right)}\right)\right)^2. \tag{5.A.23}$$

Applying twice the inequality $(a+b)^2 \le 2a^2+2b^2$, we have that the last displayed expression is bounded by

$$\begin{aligned}
&24\sigma^4_{0,\varepsilon}\left(\frac{1}{\tilde{n}}\sum_{j=1}^{\tilde{n}}\log^2\left|\frac{\sin\left(\frac{\lambda_j-\lambda_{[\tilde{n}\tau_1]}}{2}\right)}{\sin\left(\frac{\lambda_j-\lambda_{[\tilde{n}\tau_2]}}{2}\right)}\right|\right)^2 \\
&+24\sigma^4_{0,\varepsilon}\left(\frac{1}{\tilde{n}}\sum_{j=1}^{\tilde{n}}\log^2\left|\frac{\sin\left(\frac{\lambda_j+\lambda_{[\tilde{n}\tau_1]}}{2}\right)}{\sin\left(\frac{\lambda_j+\lambda_{[\tilde{n}\tau_2]}}{2}\right)}\right|\right)^2.
\end{aligned} \tag{5.A.24}$$

We examine the first term, being the second identically handled. This term, without the square and except constants, is

$$\frac{1}{\tilde{n}}\sum_{j=1}^{[\tilde{n}\tau_1]}\log^2\left|\frac{\sin\left(\frac{\lambda_{[\tilde{n}\tau_1]}-\lambda_j}{2}\right)}{\sin\left(\frac{\lambda_{[\tilde{n}\tau_2]}-\lambda_j}{2}\right)}\right|+\frac{1}{\tilde{n}}\sum_{j=[\tilde{n}\tau_2]+1}^{\tilde{n}}\log^2\left|\frac{\sin\left(\frac{\lambda_j-\lambda_{[\tilde{n}\tau_1]}}{2}\right)}{\sin\left(\frac{\lambda_j-\lambda_{[\tilde{n}\tau_2]}}{2}\right)}\right|$$
$$+\frac{1}{\tilde{n}}\sum_{j=[\tilde{n}\tau_1]+1}^{[\tilde{n}\tau_2]}\log^2\left|\frac{\sin\left(\frac{\lambda_j-\lambda_{[\tilde{n}\tau_1]}}{2}\right)}{\sin\left(\frac{\lambda_{[\tilde{n}\tau_2]}-\lambda_j}{2}\right)}\right|. \tag{5.A.25}$$

An extension of Lemma 2 of Robinson (1995b), see also Lemma 1 of Delgado, Hidalgo, and Velasco (2005), implies that the difference between (5.A.25) and

$$\int_0^{\tau_1}\log^2\left|\frac{\sin\left(\frac{\pi\tau_1-\pi u}{2}\right)}{\sin\left(\frac{\pi\tau_2-\pi u}{2}\right)}\right|du+\int_{\tau_2}^{1}\log^2\left|\frac{\sin\left(\frac{\pi u-\pi\tau_1}{2}\right)}{\sin\left(\frac{\pi u-\pi\tau_2}{2}\right)}\right|du$$
$$+\int_{\tau_1}^{\tau_2}\log^2\left|\frac{\sin\left(\frac{\pi u-\pi\tau_1}{2}\right)}{\sin\left(\frac{\pi u-\pi\tau_2}{2}\right)}\right|du \tag{5.A.26}$$

is bounded in absolute value by

$$K\tilde{n}^{-1}\log^2 n\leq K\left(\tau_2-\tau_1\right)^{\frac{1+\delta}{2}}, \tag{5.A.27}$$

for any $0<\delta<1$. Hence, the proof of (5.A.20) is completed, if we show that (5.A.26) is bounded by $K\left(\tau_2-\tau_1\right)^{\frac{1+\delta}{2}}$, for any $0<\delta<1$.

Observe that since $u\in(0,\tau_1)$, we have that $\log^2\left|\sin\left(\frac{\pi\tau_1-\pi u}{2}\right)\right|>\log^2\left|\sin\left(\frac{\pi\tau_2-\pi u}{2}\right)\right|$. Hence,

$$\left(\log\left|\sin\left(\frac{\pi\tau_1-\pi u}{2}\right)\right|-\log\left|\sin\left(\frac{\pi\tau_2-\pi u}{2}\right)\right|\right)^2$$
$$\leq\log^2\left|\sin\left(\frac{\pi\tau_1-\pi u}{2}\right)\right|-\log^2\left|\sin\left(\frac{\pi\tau_2-\pi u}{2}\right)\right|. \tag{5.A.28}$$

Then, the first integral in (5.A.26) is bounded by

$$\int_0^{\tau_1}\log^2\left|\sin\left(\frac{\pi\tau_1-\pi u}{2}\right)\right|du-\int_0^{\tau_1}\log^2\left|\sin\left(\frac{\pi\tau_2-\pi u}{2}\right)\right|du$$

$$\begin{aligned}&= H\left(\tau_1\right)-H\left(\tau_2\right)+H\left(\tau_2-\tau_1\right)\\&\leq H\left(\tau_2-\tau_1\right),\end{aligned} \tag{5.A.29}$$

where $H(\tau) = \int_0^\tau \log^2 \left|\sin\left(\frac{\pi u}{2}\right)\right| du$, because $H(\tau_1) < H(\tau_2)$ for $0 \le \tau_1 < \tau_2 \le 1$. Next, we examine the second integral in (5.A.26). Notice that since $u \in (\tau_2, 1)$, we have that $\log^2 \left|\sin\left(\frac{\pi\tau_1 - \pi u}{2}\right)\right| < \log^2 \left|\sin\left(\frac{\pi\tau_2 - \pi u}{2}\right)\right|$. Hence,

$$\left(\log\left|\sin\left(\frac{\pi u - \pi\tau_1}{2}\right)\right| - \log\left(\frac{\pi u - \pi\tau_2}{2}\right)\right)^2$$
$$\le \log^2\left|\frac{\pi u - \pi\tau_2}{2}\right| - \log^2\left|\sin\left(\frac{\pi u - \pi\tau_1}{2}\right)\right|. \qquad (5.A.30)$$

Then, we have that the second integral in (5.A.26) is bounded by

$$\int_{\tau_2}^{1} \log^2\left|\sin\left(\frac{\pi\tau_2 - \pi u}{2}\right)\right| du - \int_{\tau_2}^{1} \log^2\left|\sin\left(\frac{\pi\tau_1 - \pi u}{2}\right)\right| du$$

$$\begin{aligned} &= H(1-\tau_2) - H(1-\tau_1) + H(\tau_2 - \tau_1) \\ &\le H(\tau_2 - \tau_1), \end{aligned} \qquad (5.A.31)$$

because $H(1-\tau_2) < H(1-\tau_1)$ for $0 \le \tau_1 < \tau_2 \le 1$. Finally, by a change of variables, the third integral of (5.A.26) is

$$2\int_0^{\tau_2 - \tau_1} \log^2\left|\sin\left(\frac{\pi u}{2}\right)\right| du = 2H(\tau_2 - \tau_1). \qquad (5.A.32)$$

Since $H(\tau) \le K\tau^{\frac{1+\delta}{2}}$, for $0 \le \tau \le 1$, and for any $0 < \delta < 1$, we conclude that the all the integrals in (5.A.26) are bounded by

$$K(\tau_2 - \tau_1)^{\frac{1+\delta}{2}}, \qquad (5.A.33)$$

for any $0 < \delta < 1$, as required.

To complete the proof of (5.A.9), it remains to establish (5.A.15). It suffices to show that, for every positive $\zeta > 0$,

$$\sup_{0 \le \tau \le 1-\zeta} \zeta^{-1} \Pr\left(|X(\tau+\zeta) - X(\tau)| > \zeta\right) \to 0, \quad \text{as } \zeta \to 0. \qquad (5.A.34)$$

The latter condition implies that the process $X(\tau)$ belongs to the space $\mathbb{C}[0,1]$ by Problem 15.3 in Billingsley (1968). Theorem 5.3 in Billingsley (1968) implies that

$$E|X(\tau+\zeta) - X(\tau)|^4 \le \liminf_{n\to\infty} E\left|X_{n(1)}(\tau+\zeta) - X_{n(1)}(\tau)\right|^4, \qquad (5.A.35)$$

recalling that only the first term of $X_n(.)$ depends on τ. Then, Markov's inequality implies that (5.A.34) holds true, if the right side of the last displayed inequality satisfies

$$\sup_{0\leq\tau\leq 1-\zeta} \zeta^{-1} \liminf_{n\to\infty} E\left|X_{n(1)}(\tau+\zeta) - X_{n(1)}(\tau)\right|^4 \to 0, \quad \text{as } \zeta \to 0. \tag{5.A.36}$$

But, this is the case because (5.A.20) implies that

$$\liminf_{n\to\infty} \zeta^{-1} E\left|X_{n(1)}(\tau+\zeta) - X_{n(1)}(\tau)\right|^4 \leq K\zeta^{\delta}, \tag{5.A.37}$$

with $0 < \delta < 1$.

To complete the proof, we need to establish (5.A.10). The $(1,1)$-th element of $A_n(\tau)$ is

$$\begin{aligned}\frac{1}{\tilde{n}}\sum_{j=1}^{\tilde{n}} \left(D\log g_j\left(0,\lambda_{[\tilde{n}\tau]}\right)\right)^2 I_{j,\varepsilon} &= \frac{\sigma_{0,\varepsilon}^2}{2\pi\tilde{n}}\sum_{j=1}^{\tilde{n}} \left(D\log g_j\left(0,\lambda_{[\tilde{n}\tau]}\right)\right)^2 \\ &\quad -\frac{1}{\tilde{n}}\sum_{j=1}^{\tilde{n}} \left(D\log g_j\left(0,\lambda_{[\tilde{n}\tau]}\right)\right)^2 \\ &\quad \times\left(I_{\varepsilon,j} - \frac{\sigma_0^2}{2\pi}\right).\end{aligned} \tag{5.A.38}$$

The second term on the right hand side of (5.A.38) satisfies

$$\sup_{s=0,1,\ldots,\tilde{n}} \left|\frac{1}{\tilde{n}}\sum_{j=1}^{\tilde{n}} \left(D\log g_j(0,\lambda_s)\right)^2 \left(I_{j,\varepsilon} - \frac{\sigma_{0,\varepsilon}^2}{2\pi}\right)\right| = o_p(1), \tag{5.A.39}$$

since we have already shown that $\tilde{n}^{-\frac{1}{2}}\sum_{j=1}^{\tilde{n}} D\log g_j\left(0,\lambda_{[\tilde{n}\tau]}\right)\left(I_{j,\varepsilon} - \frac{\sigma_{0,\varepsilon}^2}{2\pi}\right) = O_p(1)$. On the other hand, the first term on the right hand side of (5.A.38) satisfies

$$\sup_{s=0,1,\ldots,\tilde{n}} \left|\frac{1}{\tilde{n}}\sum_{j=1}^{\tilde{n}} \left(D\log g_j(0,\lambda_s)\right)^2 - \frac{1}{\pi}\int_0^{\pi} \left(D\log g(\lambda;0,\lambda_s)\right)^2 d\lambda\right| = o(1), \tag{5.A.40}$$

by an extension of Lemma 2 of Robinson (1995b), see also Lemma 1 of Delgado, Hidalgo, and Velasco (2005). The rest of the elements of $A_n(\tau)$ can be similarly handled, and (5.A.10) follows.

ii) Here we show the asymptotic properties of the LM type of test $\mathcal{T}_{LM}$. The proof for $\mathcal{T}_{LM}$ is similar to that for $\mathcal{T}_W$, and we only sketch any differences. We examine the term

$$\tilde{n}^{-\frac{1}{2}}\sum_{j=1}^{\tilde{n}} \frac{\partial\log g_j(0,.)}{\partial\alpha}\frac{I_j}{h_j\left(\widetilde{\theta}\right)}. \tag{5.A.41}$$

Now, the mean value theorem implies that

$$\begin{aligned}\tilde{n}^{-\frac{1}{2}}\sum_{j=1}^{\tilde{n}}\frac{\partial\log g_j(0,.)}{\partial\alpha}\frac{I_j}{h_j\left(\widetilde{\theta}\right)} &= \tilde{n}^{-\frac{1}{2}}\sum_{j=1}^{\tilde{n}}\frac{\partial\log g_j(0,.)}{\partial\alpha}\frac{I_j}{h_j(\theta_0)} \\ &\quad -\tilde{n}^{-\frac{1}{2}}\sum_{j=1}^{\tilde{n}}\frac{\partial\log g_j(0,.)}{\partial\alpha} \\ &\quad \times\frac{\partial\log h_j\left(\overline{\theta}\right)}{\partial\theta'}\frac{I_j}{h_j\left(\overline{\theta}\right)}\left(\widetilde{\theta}-\theta_0\right), \qquad (5.A.42)\end{aligned}$$

where $\overline{\theta}$ is an intermediate point between $\widetilde{\theta}$ and θ_0.

On the other hand, by definition of $\widetilde{\theta}$, we have that

$$\begin{aligned}0 &= \tilde{n}^{-\frac{1}{2}}\sum_{j=1}^{\tilde{n}}\frac{\partial\log h_j\left(\widetilde{\theta}\right)}{\partial\theta'}\frac{I_j}{h_j\left(\widetilde{\theta}\right)} \\ &= \tilde{n}^{-\frac{1}{2}}\sum_{j=1}^{\tilde{n}}\frac{\partial\log h_j(\theta_0)}{\partial\theta'}\frac{I_j}{h_j(\theta_0)} \\ &\quad +\tilde{n}^{-\frac{1}{2}}\sum_{j=1}^{\tilde{n}}\frac{\partial^2\log h_j(\theta_0)}{\partial\theta'\partial\theta}\frac{I_j}{h_j\left(\overline{\theta}\right)}\left(\widetilde{\theta}-\theta_0\right) \\ &\quad -\tilde{n}^{-\frac{1}{2}}\sum_{j=1}^{\tilde{n}}\frac{\partial\log h_j\left(\overline{\theta}\right)}{\partial\theta'}\frac{\partial\log h_j\left(\overline{\theta}\right)}{\partial\theta}\frac{I_j}{h_j\left(\overline{\theta}\right)}\left(\widetilde{\theta}-\theta_0\right). \qquad (5.A.43)\end{aligned}$$

Now, we show that

$$\sup_{\theta\in\Theta}\left|\frac{1}{\tilde{n}}\sum_{j=1}^{\tilde{n}}\frac{\partial\log h_j(\theta)}{\partial\theta'}\frac{\partial\log h_j(\theta)}{\partial\theta}\frac{I_j}{h_j(\theta)} - \frac{\sigma^2_{0,\varepsilon}}{2\pi^2}\int_0^\pi\frac{\partial\log h_j(\theta)}{\partial\theta'}\frac{\partial\log h_j(\theta)}{\partial\theta}\frac{h(\lambda;\theta_0)}{h(\lambda;\theta)}d\lambda\right| = o_p(1). \qquad (5.A.44)$$

By the triangle inequality, the left hand side of (5.A.44) is bounded by

$$\frac{\sigma^2_{0,\varepsilon}}{2\pi}\sup_{\theta\in\Theta}\left|\frac{1}{\tilde{n}}\sum_{j=1}^{\tilde{n}}\frac{\partial\log h_j(\theta)}{\partial\theta'}\frac{\partial\log h_j(\theta)}{\partial\theta}\frac{h_j(\theta_0)}{h_j(\theta)} - \frac{1}{\pi}\int_0^\pi\frac{\partial\log h_j(\theta)}{\partial\theta'}\frac{\partial\log h_j(\theta)}{\partial\theta}\frac{h(\lambda;\theta_0)}{h(\lambda;\theta)}d\lambda\right| \qquad (5.A.45)$$

$$+\sup_{\theta\in\Theta}\left|\frac{1}{\tilde{n}}\sum_{j=1}^{\tilde{n}}\frac{\partial\log h_j(\theta)}{\partial\theta'}\frac{\partial\log h_j(\theta)}{\partial\theta}\frac{h_j(\theta_0)}{h_j(\theta)}\left(\frac{I_j}{h_j(\theta_0)}-\frac{\sigma^2_{0,\varepsilon}}{2\pi}\right)\right| \qquad (5.A.46)$$

Under Condition C.1, we have by Problem 1.7.14 of Brillinger (1975) that the first term of (5.A.46) converges to zero. On the other hand, the second term of (5.A.46) converges to zero in probability, because by standard arguments, the finite dimensional distributions of

$$\tilde{n}^{-\frac{1}{2}} \sum_{j=1}^{\tilde{n}} \frac{\partial \log h_j(\theta)}{\partial \theta'} \frac{\partial \log h_j(\theta)}{\partial \theta} \frac{h_j(\theta_0)}{h_j(\theta)} \left(\frac{I_j}{h_j(\theta_0)} - \frac{\sigma^2_{0,\varepsilon}}{2\pi} \right) \tag{5.A.47}$$

converges to a Gaussian random variable, whereas the Kolmogorov-Chentsov's tightness condition trivially holds true. Hence, the second term of (5.A.46) is $o_p(1)$. Similarly, it can be shown that

$$\sup_{\theta \in \Theta} \left| \frac{1}{\tilde{n}} \sum_{j=1}^{\tilde{n}} \frac{\partial^2 \log h_j(\theta)}{\partial \theta' \partial \theta} \frac{I_j}{h_j(\theta)} \right.$$

$$\left. - \frac{\sigma^2_{0,\varepsilon}}{2\pi^2} \int_0^{\pi} \frac{\partial^2 \log h_j(\theta)}{\partial \theta' \partial \theta} \frac{h(\lambda;\theta_0)}{h(\lambda;\theta)} d\lambda \right| = o_p(1). \tag{5.A.48}$$

Next, because $\overline{\theta} \xrightarrow{p} \theta_0$, it implies that in equation (5.A.43), after solving for $\left(\widetilde{\theta} - \theta_0\right)$, (5.A.41) is equivalent to

$$\left(\tilde{n}^{-\frac{1}{2}} \sum_{j=1}^{\tilde{n}} \frac{\partial \log g_j(0,.)}{\partial \alpha} \frac{I_j}{h_j(\theta_0)}, \tilde{n}^{-\frac{1}{2}} \sum_{j=1}^{\tilde{n}} \frac{\partial \log h_j(\theta_0)}{\partial \theta'} \frac{I_j}{h_j(\theta_0)} \right)$$

$$\times \left(\frac{1}{\frac{1}{\tilde{n}} \sum_{j=1}^{\tilde{n}} \frac{\partial \log g_j(0,.)}{\partial \alpha} \frac{\partial \log h_j(\theta_0)}{\partial \theta'} \left(\frac{1}{\tilde{n}} \sum_{j=1}^{\tilde{n}} \frac{\partial \log h_j(\theta_0)}{\partial \theta'} \frac{\partial \log h_j(\theta_0)}{\partial \theta} + \frac{\partial^2 \log h_j(\theta_0)}{\partial \theta' \partial \theta} \right)^{-1}} \right). \tag{5.A.49}$$

So, examining the weak convergence of (5.A.41) is equivalent to show the weak convergence of

$$\tilde{n}^{-\frac{1}{2}} \sum_{j=1}^{\tilde{n}} \frac{\partial \log g_j(0,.)}{\partial \alpha} \frac{I_j}{h_j(\theta_0)}, \tag{5.A.50}$$

whose proof proceeds exactly as that given in part i) and so it is omitted. ■

Proof of Corollary 5.1. Notice that $\widehat{\alpha}_s = \widehat{\alpha}_s - \alpha(n) + \alpha(n)$ and $\tilde{n}^{\frac{1}{2}} \alpha(n) = \alpha_0$. The proof follows as in the proof of Theorem 5.1 but with $\widehat{\alpha}_s$ replaced by $\widehat{\alpha}_s - \alpha(n)$. ■

The bootstrap construction induces a conditional probability P^*, given the sample $\{x_1, ..., x_n\}$. Any quantities with respect to P^* are denoted with an asterisk *.

Proof of Theorem 5.2. We only examine part i), since in view of Theorem 5.1, part ii) follows almost immediately from i).

First, by Lemma 5.1, we have that

$$\left(\widehat{\alpha}_s^*, \left(\widehat{\theta}^* - \widehat{\theta}\right)'\right)' \xrightarrow{p^*} 0. \tag{5.A.51}$$

On the other hand, proceeding as inHidalgo and Kreiss (2004), we have that

$$\tilde{n}^{\frac{1}{2}}\left(\widehat{\alpha}_s^*, \left(\widehat{\theta}^* - \widehat{\theta}\right)'\right)' = O_{p^*}(1), \tag{5.A.52}$$

and writing $\ell_{j,n}^*(\alpha, \theta, s) = \frac{I_j^*}{g_j(\alpha,\lambda_s)h_j(\theta)}$, we can further obtain

$$\begin{aligned}\sum_{j=1}^{\tilde{n}} \ell_{j,n}^*(\alpha, \theta, s) &= \sum_{j=1}^{\tilde{n}} \ell_{j,n}^*\left(0, \widehat{\theta}, s\right) + \sum_{j=1}^{\tilde{n}}\left(D\ell_{j,n}^*\left(0, \widehat{\theta}, s\right)\right)'\left(\alpha_s, \left(\theta - \widehat{\theta}\right)'\right)' \\ &+\frac{1}{2}\left(\alpha_s, \left(\theta - \widehat{\theta}\right)'\right)\sum_{j=1}^{\tilde{n}} D^2\ell_{j,n}^*\left(0, \widehat{\theta}, s\right)\left(\alpha_s, \left(\theta - \widehat{\theta}\right)'\right)' \\ &+R_n^*(\alpha, \theta, s),\end{aligned} \tag{5.A.53}$$

where for all $s = 0, 1, ..., \tilde{n}$,

$$\sup_{(\alpha,\theta')'\in\Pi\times\Theta:\ n^{\frac{1}{2}}|\alpha|\leq\gamma;\ n^{\frac{1}{2}}\left\|\theta-\widehat{\theta}\right\|<\gamma} |R_n^*(\alpha, \theta, s)| = o_p(1), \tag{5.A.54}$$

for any $0 < \gamma < \infty$. Proceeding as with the proof of Theorem 5.1 and following its arguments, it suffices to show that

$$X_n^*(\tau) = \tilde{n}^{-\frac{1}{2}}\sum_{j=1}^{\tilde{n}} D\log g_j\left(0, \lambda_{[\tilde{n}\tau]}\right)\left(\frac{I_j^*}{h_j\left(\widehat{\theta}\right)} - \frac{\widehat{\sigma}_\varepsilon^2}{2\pi}\right) \tag{5.A.55}$$

and

$$A_n^*(\tau) = \frac{1}{\tilde{n}}\sum_{j=1}^{\tilde{n}}\begin{pmatrix} D\log g_j\left(0, \lambda_{[\tilde{n}\tau]}\right) \\ D\log h_j\left(\widehat{\theta}\right)\end{pmatrix}\begin{pmatrix} D\log g_j\left(0, \lambda_{[\tilde{n}\tau]}\right) \\ D\log h_j\left(\widehat{\theta}\right)\end{pmatrix}'\frac{I_j^*}{h_j\left(\widehat{\theta}\right)} \tag{5.A.56}$$

converge in bootstrap sense to the same processes as

$$X_n(\tau) = \tilde{n}^{-\frac{1}{2}}\sum_{j=1}^{\tilde{n}} D\log g_j\left(0, \lambda_{[\tilde{n}\tau]}\right)\left(\frac{I_j}{g_j\left(0, \lambda_{[\tilde{n}\tau]}\right)h_j(\theta_0)} - \frac{\sigma_{0,\varepsilon}^2}{2\pi}\right) \tag{5.A.57}$$

and

$$A_n(\tau) = \frac{1}{\tilde{n}} \sum_{j=1}^{\tilde{n}} \begin{pmatrix} D \log g_j\left(0, \lambda_{[\tilde{n}\tau]}\right) \\ D \log h_j(\theta_0) \end{pmatrix} \begin{pmatrix} D \log g_j\left(0, \lambda_{[\tilde{n}\tau]}\right) \\ D \log h_j(\theta_0) \end{pmatrix}' \frac{I_j}{h_j(\theta_0)} \tag{5.A.58}$$

respectively.

We begin with the proof that $X_n^*(\tau)$ converges to the same process as $X_n(\tau)$. The proof is split into three lemmas. Lemma 5.2 shows that $X_{n,\ell}^*(\tau)$ has a covariance structure, conditional on x, that converges in probability to $\mathcal{K}(\tau_1, \tau_2)$ given in (5.4.4). Lemma 5.3 shows that the finite dimensional limiting distribution of $X_n^*(.)$ is Gaussian centered at zero. Finally, Lemma 5.4 proves the tightness of $X_n^*(.)$. Therefore, combining Lemmas 5.2 to 5.4, we have that $X_n^*(\tau) \Longrightarrow X(\tau)$ in $\mathbb{D}[0,1]$ in probability, as defined by Giné and Zinn (1990).

Finally, it is straightforward to show that $A_n^*(\tau)$ converges in bootstrap to $A(\tau)$, following the ideas in Lemma 5.2 and Theorem 5.1. ■

5.B Appendix

This Section contains a series of technical lemmas used in the proofs in Appendix 5.A above.

Lemma 5.1

Suppose that $\widehat{\theta}$ convergences almost surely in Θ. Then,

$$\widehat{\theta}^* - \widehat{\theta} = o_{p^*}(1). \tag{5.B.1}$$

Proof. Because by construction, conditional on the sample $\{x_1, ..., x_n\}$, $\{x_t^*\}_{t\in\mathbb{Z}}$ is a sequence of zero mean and unit variance i.i.d. random variables, then it is also ergodic in a quadratic mean sense. Then, proceeding as in the proof of Lemma 1 of Hannan (1973), we have that uniformly in $\theta \in \Theta$

$$\frac{1}{n} \sum_{j=-\tilde{n}+1}^{\tilde{n}-1} \frac{I_j^*}{f_j(\theta)} - \int_{-\pi}^{\pi} \frac{f\left(\lambda; \widehat{\theta}\right)}{f(\lambda; \theta)} d\lambda \xrightarrow{p^*} 0. \tag{5.B.2}$$

Now,following the proof of Theorem 1 of Hannan (1973), we have that

$$\widehat{\theta}^* - \widehat{\theta} = o_{p^*}(1), \tag{5.B.3}$$

because

$$\widehat{\theta} = \arg\min_{\theta\in\Theta} \int_{-\pi}^{\pi} \frac{f\left(\lambda; \widehat{\theta}\right)}{f(\lambda; \theta)} d\lambda, \tag{5.B.4}$$

which concludes the proof. ■

Lemma 5.2

Suppose that Conditions C.1-C.6 are satisfied. Then,

$$E^{*}\left(X_{n}^{*}\left(\tau_{1}\right) X_{n}^{*}\left(\tau_{2}\right)\right) \xrightarrow{p} \mathcal{K}\left(\tau_{1}, \tau_{2}\right), \tag{5.B.5}$$

for $0 \leq \tau_1 \leq \tau_2 \leq 1$.

Proof. Denote $\psi_j(s) = \frac{\partial}{\partial \alpha} \log g_j(0; \lambda_s)$. First, notice that

$$\sum_{t=1}^{n} x_t^{*} e^{-it\lambda_j} = \widehat{\sigma}_{\varepsilon}^{2} \widetilde{\Phi}\left(e^{-i\lambda_j}\right) \sum_{t=1}^{n} \widetilde{x}_t^{*} e^{-it\lambda_j}, \tag{5.B.6}$$

because by definition of x_t^* and that $\sum_{p=1}^{n} e^{ip\lambda_j} = n\mathcal{I}(j = 0, 2n, ...)$, the left hand side of (5.B.6) is

$$\frac{\widehat{\sigma}_{\varepsilon}^{2}}{n} \sum_{t=1}^{n} \left(\sum_{\ell=1}^{n} e^{it\lambda_\ell} \widetilde{\Phi}\left(e^{-i\lambda_j}\right) \sum_{p=1}^{n} \widetilde{x}_p^{*} e^{-ip\lambda_j} \right) e^{-it\lambda_j} = \widehat{\sigma}_{\varepsilon}^{2} \widetilde{\Phi}\left(e^{-i\lambda_j}\right) \sum_{t=1}^{n} \widetilde{x}_t^{*} e^{-it\lambda_j}. \tag{5.B.7}$$

Next, $E^{*}\left(X_{n}^{*}\left(\tau_{1}\right) X_{n}^{*}\left(\tau_{2}\right)\right)$ becomes

$$\frac{1}{\tilde{n}} \sum_{j_1, j_2=1}^{\tilde{n}} \psi_{j_1}(\pi\tau_1) \psi_{j_2}(\pi\tau_2) E^{*}\left(\left(\frac{I_{j1}^{*} - E^{*}\left(I_{j1}^{*}\right)}{h_{j_1}\left(\widehat{\theta}\right)} \right) \left(\frac{I_{j2}^{*} - E^{*}\left(I_{j2}^{*}\right)}{h_{j_2}\left(\widehat{\theta}\right)} \right) \right)$$

$$+\frac{1}{\tilde{n}} \sum_{j_1, j_2=1}^{\tilde{n}} \psi_{j_1}(\pi\tau_1) \psi_{j_2}(\pi\tau_2) \left(\left(\frac{E^{*}\left(I_{j1}^{*}\right)}{h_{j_1}\left(\widehat{\theta}\right)} - \frac{\widehat{\sigma}_{\varepsilon}^{2}}{2\pi} \right) \left(\frac{E^{*}\left(I_{j2}^{*}\right)}{h_{j_2}\left(\widehat{\theta}\right)} - \frac{\widehat{\sigma}_{\varepsilon}^{2}}{2\pi} \right) \right). \tag{5.B.8}$$

Because $\{x_t^*\}_{t\in\mathbb{Z}}$ is a sequence of zero mean and unit variance i.i.d. random variables, the second term of (5.B.8) is

$$\frac{1}{\tilde{n}} \sum_{j_1, j_2=1}^{\tilde{n}} \psi_{j_1}(\pi\tau_1) \psi_{j_2}(\pi\tau_2) \left(\frac{\widehat{\sigma}_{\varepsilon}^{2} \left| \widetilde{\Phi}\left(e^{-i\lambda_{j_1}}\right) \right|^{2}}{2\pi h_{j_1}\left(\widehat{\theta}\right)} - \frac{\widehat{\sigma}_{\varepsilon}^{2}}{2\pi} \right)$$

$$\times \left(\frac{\widehat{\sigma}_{\varepsilon}^{2} \left| \widetilde{\Phi}\left(e^{-i\lambda_{j_2}}\right) \right|^{2}}{2\pi h_{j_2}\left(\widehat{\theta}\right)} - \frac{\widehat{\sigma}_{\varepsilon}^{2}}{2\pi} \right) \tag{5.B.9}$$

which is equal to zero, since (5.B.6) implies that

$$\begin{aligned}\widetilde{\Phi}\left(e^{-i\lambda_j}\right) &= \frac{1}{n}\sum_{q=1}^{n}\sum_{\ell=-\tilde{n}+1}^{\tilde{n}-1} e^{iq\lambda_j} h^{\frac{1}{2}}\left(\lambda_\ell;\widehat{\theta}\right) e^{-iq\lambda_j} \\ &= \sum_{\ell=-\tilde{n}+1}^{\tilde{n}-1} h^{\frac{1}{2}}\left(\lambda_\ell;\widehat{\theta}\right)\frac{1}{n}\sum_{q=1}^{n} e^{iq\lambda_{\ell-j}} \\ &= h^{\frac{1}{2}}\left(e^{-i\lambda_j};\widehat{\theta}\right), \end{aligned} \qquad (5.B.10)$$

since $\sum_{q=1}^{n} e^{iq\lambda_{\ell-j}} = n\mathcal{I}(j=l)$.

Hence, to conclude the proof, we need to show that the first term of (5.B.8) converges in probability to $\mathcal{K}(\tau_1,\tau_2)$. But because $\{x_t^*\}_{t\in\mathbb{Z}}$ is a sequence of zero mean and unit variance i.i.d. random variables, by Proposition 10.3.2 of Brockwell and Davis (1991), this term is

$$\begin{aligned}&\frac{1}{n}\sum_{j=1}^{\tilde{n}}\psi_j(\pi\tau_1)\psi_j(\pi\tau_2) + \frac{\widehat{\kappa}_4}{\tilde{n}^2}\sum_{j_1,j_2=1}^{\tilde{n}}\psi_{j_1}(\pi\tau_1)\psi_{j_2}(\pi\tau_2) \\ &= \mathcal{K}(\tau_1,\tau_2)\left(1+O\left(\tilde{n}^{-1}\log\tilde{n}\right)\right) + \frac{\widehat{\kappa}_4}{\tilde{n}^2}\sum_{j_1,j_2=1}^{\tilde{n}}\psi_{j_1}(\pi\tau_1)\psi_{j_2}(\pi\tau_2), \end{aligned} \qquad (5.B.11)$$

by a straightforward modification of Lemma 2 of Robinson (1995b), see also Lemma 1 of Delgado, Hidalgo, and Velasco (2005). However, because by a well known argument, see Theorem 3.5.8 of Stout (1974), Condition C.1 implies that $\{x_t\}_{t\in\mathbb{Z}}$ is an ergodic sequence, we have that

$$\widehat{\kappa}_4 = \frac{1}{n}\sum_{t=1}^{n}x_t^4 - 3\left(\frac{1}{n}\sum_{t=1}^{n}x_t^2\right)^2 \xrightarrow{p} \kappa_4. \qquad (5.B.12)$$

From here, the conclusion follows because

$$\frac{1}{\tilde{n}}\sum_{j=1}^{\tilde{n}}\psi_j(\pi\tau) - \frac{1}{\pi}\int_0^\pi \log\left|4\sin\left(\frac{\lambda\pm\pi\tau}{2}\right)\right| d\lambda = O\left(\frac{\log\tilde{n}}{\tilde{n}}\right), \qquad (5.B.13)$$

by a straightforward modification of Lemma 2 of Robinson (1995b), see also Lemma 1 of Delgado, Hidalgo, and Velasco (2005), and that $\int_0^\pi \log\left|4\sin\left(\frac{\lambda\pm\pi\tau}{2}\right)\right| d\lambda = 0$ for all $0\leq\tau\leq 1$. ■

Lemma 5.3

Under Conditions C.1-C.6, the finite-dimensional distributions of $X_n^*(.)$ converge in bootstrap law to those of a centered Gaussian process.

Proof. Fix $\tau_1, ..., \tau_q \in [0,1]$ and constants $c_1, ..., c_q$. By the Cramér-Wold device, it suffices to examine the limit distribution of

$$\sum_{p=1}^{q} c_p \left(\tilde{n}^{-\frac{1}{2}} \sum_{j=1}^{\tilde{n}} \psi_j (\tau_p) \left(\frac{2\pi I_j^*}{h_j \left(\widehat{\theta}\right)} - \widehat{\sigma}_\varepsilon^2 \right) \right). \tag{5.B.14}$$

By Lemma 5.2, the (bootstrap) second moment of (5.B.14) converges in probability to

$$\sum_{p_1, p_2 = 1}^{q} c_{p_1} c_{p_2} \mathcal{K} \left(\tau_{p_1}, \tau_{p_2} \right). \tag{5.B.15}$$

So, to complete the proof, it remains to verify that (5.B.14) satisfies the Lindeberg's condition, that is for every $\delta > 0$,

$$\begin{aligned} &\frac{1}{\tilde{n}} \sum_{j=1}^{\tilde{n}} E^* \left| \psi_j (\tau_p) \left(\frac{2\pi I_j^*}{h_j \left(\widehat{\theta}\right)} - \widehat{\sigma}_\varepsilon^2 \right) \right|^2 \\ &\times \mathcal{I} \left(\frac{1}{\tilde{n}} \left| \psi_j (\tau_p) \left(\frac{2\pi I_j^*}{h_j \left(\widehat{\theta}\right)} - \widehat{\sigma}_\varepsilon^2 \right) \right|^2 > \delta \right) \xrightarrow{p} 0, \end{aligned} \tag{5.B.16}$$

or the sufficient condition

$$\frac{1}{\tilde{n}} E^* \left| \left(\frac{2\pi I_j^*}{h_j \left(\widehat{\theta}\right)} - \widehat{\sigma}_\varepsilon^2 \right) \right|^4 \frac{1}{\tilde{n}} \sum_{j=1}^{\tilde{n}} \left| \psi_j (\tau_p) \right|^4 \xrightarrow{p} 0. \tag{5.B.17}$$

But this is the case, since proceeding as in Theorem 5.1 and Lemma 5.4, the left hand side of the last displayed expression is

$$\tilde{n}^{-1} \widehat{\sigma}_\varepsilon^2 E^* \left| \left(\frac{1}{2\pi n} \sum_{t=1}^{n} \widetilde{x}_t^* e^{-it\lambda_j} - 1 \right) \right|^4 = O_p \left(\tilde{n}^{-1} \right), \tag{5.B.18}$$

by Theorems 2.3.1 and 4.3.2 of Brillinger (1975) since $\{x_t^*\}_{t \in \mathbb{Z}}$ is a sequence of zero mean and unit variance i.i.d. random variables and $\{x_t\}_{t \in \mathbb{Z}}$ is ergodic in that $\frac{1}{n} \sum_{t=1}^{n} |x_t|^r - E |x_t|^r = o_p (1)$ for $r = 1, ..., 8$, and $\widehat{\sigma}_\varepsilon^2 = O_p (1)$. ■

Lemma 5.4

Under Conditions C.1-C.6, conditional on $\{x_1, ..., x_n\}$, $X_n^* (.)$ is tight.

Proof. Denote $\zeta_j^* = \frac{1}{n} \sum_{t=1}^{n} \widetilde{x}_t^* e^{-it\lambda_j} - 1$. Proceeding as with the proof of Theorem 5.1, we only need to check the Kolmogorov-Chentsov's condition. That is, we need to show that

$$E^*\left|\tilde{n}^{-\frac{1}{2}}\sum_{j=1}^{\tilde{n}}\left(\psi_j(\tau_1)-\psi_j(\tau_2)\right)\left(\frac{I_j^*}{h_j\left(\widehat{\theta}\right)}-\frac{\widehat{\sigma}_\varepsilon^2}{2\pi}\right)\right|^4$$

$$\leq G_n(\tau_1,\tau_2)\left|\tau_1-\tau_2\right|^{1+\delta}, \tag{5.B.19}$$

for any $0 \leq \tau_1 < \tau_2 \leq 1$, for some $\delta > 0$ and some $G_n(\tau_1, \tau_2)$ which bounded in probability. Now, by definition of I_j^*, the left hand side of (5.B.19) is

$$\left(\frac{\widehat{\sigma}_\varepsilon^2}{2\pi}\right)^4 E^*\left|\tilde{n}^{-\frac{1}{2}}\sum_{j=1}^{\tilde{n}}\left(\psi_j(\tau_1)-\psi_j(\tau_2)\right)\zeta_j^*\right|^4$$

$$=\left(\frac{\widehat{\sigma}_\varepsilon^2}{2\pi}\right)^4\frac{1}{\tilde{n}^2}\sum_{j_1,\dots,j_4=1}^{\tilde{n}}\prod_{\ell=1}^{4}\left(\log\frac{\left|\sin\left(\frac{\lambda_{j_\ell}-\lambda_{[\pi\tau_1]}}{2}\right)\right|}{\left|\sin\left(\frac{\lambda_{j_\ell}-\lambda_{[\pi\tau_2]}}{2}\right)\right|}\right)E^*\left(\prod_{\ell=1}^{4}\zeta_{j_\ell}^*\right)$$

$$+\left(\frac{\widehat{\sigma}_\varepsilon^2}{2\pi}\right)^4\frac{1}{\tilde{n}^2}\sum_{j_1,\dots,j_4=1}^{\tilde{n}}\prod_{\ell=1}^{4}\left(\log\frac{\left|\sin\left(\frac{\lambda_{j_\ell}+\lambda_{[\pi\tau_1]}}{2}\right)\right|}{\left|\sin\left(\frac{\lambda_{j_\ell}+\lambda_{[\pi\tau_2]}}{2}\right)\right|}\right)E^*\left(\prod_{\ell=1}^{4}\zeta_{j_\ell}^*\right). \tag{5.B.20}$$

We examine only the first term on the right hand side, the second being identically handled. That term is

$$\left(\frac{\widehat{\sigma}_\varepsilon^2}{2\pi}\right)^4\frac{3}{\tilde{n}^2}\sum_{j_1,\dots,j_4=1}^{\tilde{n}}\prod_{\ell=1}^{4}\left(\log\frac{\left|\sin\left(\frac{\lambda_{j_\ell}-\lambda_{[\pi\tau_1]}}{2}\right)\right|}{\left|\sin\left(\frac{\lambda_{j_\ell}-\lambda_{[\pi\tau_2]}}{2}\right)\right|}\right)E^*\left\{\zeta_{j_1}^*\zeta_{j_2}^*\right\}E^*\left\{\zeta_{j_3}^*\zeta_{j_4}^*\right\}$$

$$+\left(\frac{\widehat{\sigma}_\varepsilon^2}{2\pi}\right)^4\frac{1}{\tilde{n}^2}\sum_{j_1,\dots,j_4=1}^{\tilde{n}}\prod_{\ell=1}^{4}\left(\log\frac{\left|\sin\left(\frac{\lambda_{j_\ell}-\lambda_{[\pi\tau_1]}}{2}\right)\right|}{\left|\sin\left(\frac{\lambda_{j_\ell}-\lambda_{[\pi\tau_2]}}{2}\right)\right|}\right)cum^*\left(\zeta_{j_1}^*,\zeta_{j_2}^*,\zeta_{j_3}^*,\zeta_{j_4}^*\right). \tag{5.B.21}$$

By Theorems 2.3.1 and 4.3.2 of Brillinger (1975), and in particular equation (4.3.15), and the integrability of $\log^q|\sin x|$, $q = 1, ..., 4$, we obtain that the second term on the right hand of (5.B.21) is bounded in absolute value by

$$K\tilde{n}^{-2}\log^2\tilde{n}\left(\frac{\widehat{\sigma}_\varepsilon^2}{2\pi}\right)^4 \leq K\left(\frac{\widehat{\sigma}_\varepsilon^2}{2\pi}\right)^4(\tau_2-\tau_1)^{1+\delta}\leq G_n(\tau_1,\tau_2)(\tau_2-\tau_1)^{1+\delta}, \tag{5.B.22}$$

for any $0 < \delta < 1$, noticing that $G_n(\tau_1, \tau_2)$ is bounded in probability since $\widehat{\sigma}_\varepsilon^2 = O_p(1)$. So, to complete the proof we need to examine the first term on the right hand side of (5.B.21), which is

$$3\left(\frac{\widehat{\sigma}_\varepsilon^2}{2\pi}\right)^4\left(\frac{1}{\tilde{n}}\sum_{j=1}^{\tilde{n}}\left(\log\frac{\left|\sin\left(\frac{\lambda_j-\lambda_{[\pi\tau_1]}}{2}\right)\right|}{\left|\sin\left(\frac{\lambda_j-\lambda_{[\pi\tau_2]}}{2}\right)\right|}\right)^2\right)^2\leq G_n(\tau_1,\tau_2)\left|\tau_1-\tau_2\right|^{1+\delta}, \tag{5.B.23}$$

proceeding as in Theorem 5.1. ■

5.C Appendix

	$\beta = 10\%$	$\beta = 5\%$	$\beta = 1\%$
$n = 64$	10.65	10.10	10.10
$n = 128$	11.00	5.15	4.15

Table 5.1: Size of $\mathcal{T}_W^*$ test; i.i.d. model.

	$\beta = 10\%$	$\beta = 5\%$	$\beta = 1\%$
$n = 64$	10.60	5.00	1.20
$n = 128$	9.65	5.10	1.40

Table 5.2: Size of $\mathcal{T}_{LM}^*$ test; i.i.d. model.

Method 1				Method 2			
	$\beta = 10\%$	$\beta = 5\%$	$\beta = 1\%$		$\beta = 10\%$	$\beta = 5\%$	$\beta = 1\%$
$n = 64$	10.05	5.40	1.35	$n = 64$	10.05	5.55	1.25
$n = 128$	9.65	4.30	1.45	$n = 128$	10.20	4.60	1.50

Table 5.3: Size of $\mathcal{T}_{LM}^*$ test; $AR(1)$ model.

Method 1				Method 2			
	$\beta = 10\%$	$\beta = 5\%$	$\beta = 1\%$		$\beta = 10\%$	$\beta = 5\%$	$\beta = 1\%$
$n = 64$	9.50	5.55	1.10	$n = 64$	9.85	5.35	1.05
$n = 128$	10.00	4.80	1.60	$n = 128$	9.80	4.70	1.60

Table 5.4: Size of $\mathcal{T}_{LM}^*$ test; $MA(1)$ model.

	$\beta = 10\%$	$\beta = 5\%$	$\beta = 1\%$
$n = 64$	11.80	11.40	11.40
$n = 128$	15.10	7.95	6.45

Table 5.5: Power of $\mathcal{T}_W^*$ test; $ARFIMA(0, 0.1, 0)$ model.

	$\beta = 10\%$	$\beta = 5\%$	$\beta = 1\%$
$n = 64$	26.95	15.20	3.55
$n = 128$	49.80	34.40	11.50

Table 5.6: Power of $\mathcal{T}_W^*$ test; $ARFIMA(0, 0.2, 0)$ model.

	$\beta = 10\%$	$\beta = 5\%$	$\beta = 1\%$
$n = 64$	55.20	39.30	14.20
$n = 128$	86.60	75.30	47.65

Table 5.7: Power of $\mathcal{T}_W^*$ test; $ARFIMA(0, 0.3, 0)$ model.

	$\beta = 10\%$	$\beta = 5\%$	$\beta = 1\%$
$n = 64$	81.50	68.45	39.30
$n = 128$	98.75	96.60	86.25

Table 5.8: Power of $\mathcal{T}_W^*$ test; $ARFIMA(0, 0.4, 0)$ model.

	$\beta = 10\%$	$\beta = 5\%$	$\beta = 1\%$
$n = 64$	14.60	8.05	2.00
$n = 128$	22.90	15.05	5.45

Table 5.9: Power of $\mathcal{T}^*_{LM}$ test; $ARFIMA(0, 0.1, 0)$ model.

	$\beta = 10\%$	$\beta = 5\%$	$\beta = 1\%$
$n = 64$	38.10	26.75	12.90
$n = 128$	68.45	57.55	38.00

Table 5.10: Power of $\mathcal{T}^*_{LM}$ test; $ARFIMA(0, 0.2, 0)$ model.

	$\beta = 10\%$	$\beta = 5\%$	$\beta = 1\%$
$n = 64$	66.70	57.75	38.50
$n = 128$	94.45	91.70	81.10

Table 5.11: Power of $\mathcal{T}^*_{LM}$ test; $ARFIMA(0, 0.3, 0)$ model.

	$\beta = 10\%$	$\beta = 5\%$	$\beta = 1\%$
$n = 64$	86.25	81.15	66.60
$n = 128$	99.35	98.95	97.50

Table 5.12: Power of $\mathcal{T}^*_{LM}$ test; $ARFIMA(0, 0.4, 0)$ model.

	$\beta = 10\%$	$\beta = 5\%$	$\beta = 1\%$
$n = 64$	18.10	9.60	2.20
$n = 128$	28.55	20.00	7.30

Table 5.13: Power of $\mathcal{T}_{LM}^{*}$ test; $GARMA(0, 0.1, 0)$ model.

	$\beta = 10\%$	$\beta = 5\%$	$\beta = 1\%$
$n = 64$	47.45	36.85	17.90
$n = 128$	76.55	67.85	49.35

Table 5.14: Power of $\mathcal{T}_{LM}^{*}$ test; $GARMA(0, 0.2, 0)$ model.

	$\beta = 10\%$	$\beta = 5\%$	$\beta = 1\%$
$n = 64$	80.65	73.75	56.90
$n = 128$	97.55	96.05	90.90

Table 5.15: Power of $\mathcal{T}_{LM}^{*}$ test; $GARMA(0, 0.3, 0)$ model.

	$\beta = 10\%$	$\beta = 5\%$	$\beta = 1\%$
$n = 64$	96.85	94.40	88.35
$n = 128$	99.90	99.90	99.55

Table 5.16: Power of $\mathcal{T}_{LM}^{*}$ test; $GARMA(0, 0.4, 0)$ model.

Method 1				Method 2			
	$\beta = 10\%$	$\beta = 5\%$	$\beta = 1\%$		$\beta = 10\%$	$\beta = 5\%$	$\beta = 1\%$
$n = 64$	11.30	6.20	1.10	$n = 64$	11.60	6.05	1.20
$n = 128$	11.75	6.15	1.05	$n = 128$	11.65	6.30	0.95

Table 5.17: Power of $\mathcal{T}_{LM}^{*}$ test; $ARFIMA(1, 0.1, 0)$ model.

Method 1				Method 2			
	$\beta = 10\%$	$\beta = 5\%$	$\beta = 1\%$		$\beta = 10\%$	$\beta = 5\%$	$\beta = 1\%$
$n = 64$	11.85	6.35	1.20	$n = 64$	12.20	6.70	1.25
$n = 128$	11.95	6.75	1.15	$n = 128$	12.05	6.95	1.05

Table 5.18: Power of $\mathcal{T}_{LM}^{*}$ test; $ARFIMA(1, 0.2, 0)$ model.

Method 1				Method 2			
	$\beta = 10\%$	$\beta = 5\%$	$\beta = 1\%$		$\beta = 10\%$	$\beta = 5\%$	$\beta = 1\%$
$n = 64$	13.65	7.45	1.45	$n = 64$	13.60	7.65	1.35
$n = 128$	12.85	7.30	1.45	$n = 128$	13.35	7.50	1.50

Table 5.19: Power of $\mathcal{T}_{LM}^{*}$ test; $ARFIMA(1, 0.3, 0)$ model.

Method 1				Method 2			
	$\beta = 10\%$	$\beta = 5\%$	$\beta = 1\%$		$\beta = 10\%$	$\beta = 5\%$	$\beta = 1\%$
$n = 64$	15.55	9.00	2.05	$n = 64$	15.90	9.75	1.95
$n = 128$	17.60	10.20	2.60	$n = 128$	17.85	10.40	2.85

Table 5.20: Power of $\mathcal{T}_{LM}^{*}$ test; $ARFIMA(1, 0.4, 0)$ model.

Method 1				Method 2			
	$\beta = 10\%$	$\beta = 5\%$	$\beta = 1\%$		$\beta = 10\%$	$\beta = 5\%$	$\beta = 1\%$
$n = 64$	19.70	10.85	2.80	$n = 64$	19.65	10.70	2.70
$n = 128$	27.05	18.10	5.95	$n = 128$	26.50	17.75	6.15

Table 5.21: Power of $\mathcal{T}_{LM}^{*}$ test; $GARMA(1, 0.1, 0)$ model.

Method 1				Method 2			
	$\beta = 10\%$	$\beta = 5\%$	$\beta = 1\%$		$\beta = 10\%$	$\beta = 5\%$	$\beta = 1\%$
$n = 64$	46.85	34.50	16.55	$n = 64$	47.00	34.80	16.00
$n = 128$	72.55	62.65	43.85	$n = 128$	72.55	62.70	43.40

Table 5.22: Power of $\mathcal{T}_{LM}^{*}$ test; $GARMA(1, 0.2, 0)$ model.

Method 1				Method 2			
	$\beta = 10\%$	$\beta = 5\%$	$\beta = 1\%$		$\beta = 10\%$	$\beta = 5\%$	$\beta = 1\%$
$n = 64$	79.15	71.60	51.75	$n = 64$	79.00	71.15	52.20
$n = 128$	96.55	94.95	88.05	$n = 128$	96.55	94.90	87.70

Table 5.23: Power of $\mathcal{T}_{LM}^{*}$ test; $GARMA(1, 0.3, 0)$ model.

Method 1				Method 2			
	$\beta = 10\%$	$\beta = 5\%$	$\beta = 1\%$		$\beta = 10\%$	$\beta = 5\%$	$\beta = 1\%$
$n = 64$	95.90	93.65	84.35	$n = 64$	96.05	93.40	84.10
$n = 128$	99.90	99.85	99.15	$n = 128$	99.90	99.90	99.20

Table 5.24: Power of $\mathcal{T}_{LM}^{*}$ test; $GARMA(1, 0.4, 0)$ model.

Method 1				Method 2			
	$\beta = 10\%$	$\beta = 5\%$	$\beta = 1\%$		$\beta = 10\%$	$\beta = 5\%$	$\beta = 1\%$
$n = 64$	15.30	9.05	2.45	$n = 64$	15.10	8.65	2.85
$n = 128$	19.25	11.65	4.95	$n = 128$	19.35	11.45	4.85

Table 5.25: Power of $\mathcal{T}_{LM}^{*}$ test; $ARFIMA(0, 0.1, 1)$ model.

Method 1				Method 2			
	$\beta = 10\%$	$\beta = 5\%$	$\beta = 1\%$		$\beta = 10\%$	$\beta = 5\%$	$\beta = 1\%$
$n = 64$	33.30	23.25	11.25	$n = 64$	33.75	23.50	11.60
$n = 128$	56.45	45.60	27.05	$n = 128$	56.45	45.60	27.20

Table 5.26: Power of $\mathcal{T}_{LM}^{*}$ test; $ARFIMA(0, 0.2, 1)$ model.

Method 1				Method 2			
	$\beta = 10\%$	$\beta = 5\%$	$\beta = 1\%$		$\beta = 10\%$	$\beta = 5\%$	$\beta = 1\%$
$n = 64$	57.60	47.95	31.60	$n = 64$	57.85	48.75	31.25
$n = 128$	87.05	81.55	68.65	$n = 128$	87.50	81.80	69.25

Table 5.27: Power of $\mathcal{T}_{LM}^{*}$ test; $ARFIMA(0, 0.3, 1)$ model.

Method 1				Method 2			
	$\beta = 10\%$	$\beta = 5\%$	$\beta = 1\%$		$\beta = 10\%$	$\beta = 5\%$	$\beta = 1\%$
$n = 64$	79.10	72.15	57.25	$n = 64$	79.20	72.70	57.50
$n = 128$	98.60	97.55	92.00	$n = 128$	98.55	97.50	92.15

Table 5.28: Power of $\mathcal{T}_{LM}^{*}$ test; $ARFIMA(0, 0.4, 1)$ model.

Method 1				Method 2			
	$\beta = 10\%$	$\beta = 5\%$	$\beta = 1\%$		$\beta = 10\%$	$\beta = 5\%$	$\beta = 1\%$
$n = 64$	21.00	13.20	3.55	$n = 64$	21.05	13.35	3.85
$n = 128$	30.65	21.25	8.40	$n = 128$	30.20	20.95	8.40

Table 5.29: Power of $\mathcal{T}_{LM}^*$ test; $GARMA(0, 0.1, 1)$ model.

Method 1				Method 2			
	$\beta = 10\%$	$\beta = 5\%$	$\beta = 1\%$		$\beta = 10\%$	$\beta = 5\%$	$\beta = 1\%$
$n = 64$	51.75	39.80	21.10	$n = 64$	52.45	40.05	21.05
$n = 128$	75.35	67.30	48.00	$n = 128$	75.50	67.55	48.25

Table 5.30: Power of $\mathcal{T}_{LM}^*$ test; $GARMA(0, 0.2, 1)$ model.

Method 1				Method 2			
	$\beta = 10\%$	$\beta = 5\%$	$\beta = 1\%$		$\beta = 10\%$	$\beta = 5\%$	$\beta = 1\%$
$n = 64$	80.65	74.35	57.30	$n = 64$	80.55	74.20	58.15
$n = 128$	97.00	95.00	88.85	$n = 128$	96.95	95.05	89.10

Table 5.31: Power of $\mathcal{T}_{LM}^*$ test; $GARMA(0, 0.3, 1)$ model.

Method 1				Method 2			
	$\beta = 10\%$	$\beta = 5\%$	$\beta = 1\%$		$\beta = 10\%$	$\beta = 5\%$	$\beta = 1\%$
$n = 64$	96.35	93.40	87.90	$n = 64$	96.35	93.75	88.70
$n = 128$	99.90	99.85	99.15	$n = 128$	99.90	99.85	99.30

Table 5.32: Power of $\mathcal{T}_{LM}^*$ test; $GARMA(0, 0.4, 1)$ model.

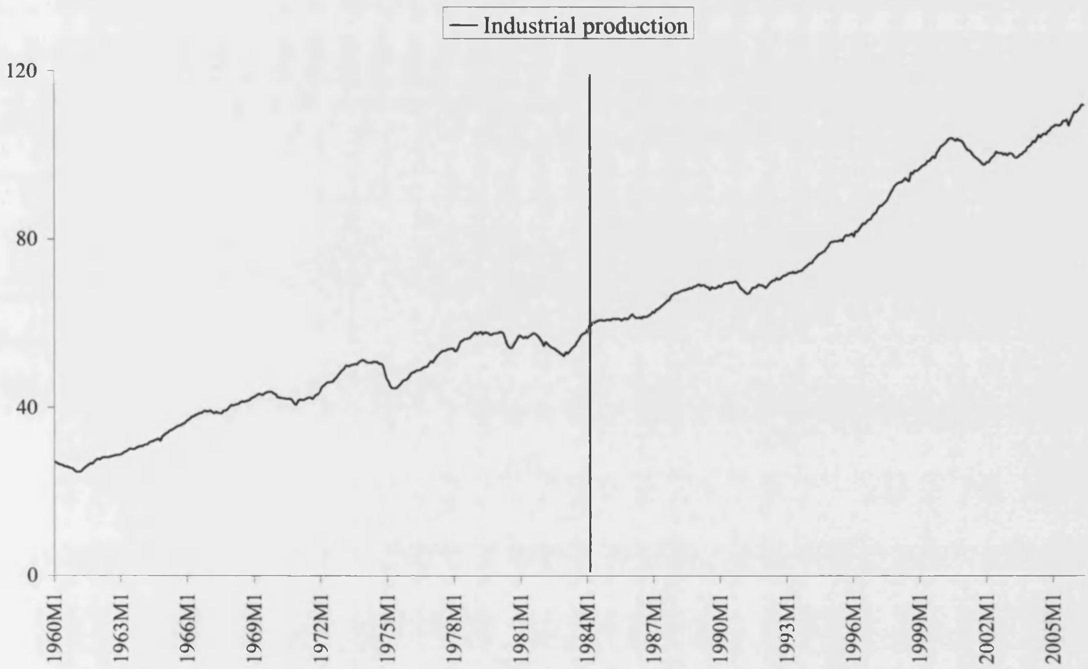


Figure 5.1: Data on industrial production for the period 1960M1-2006M5.

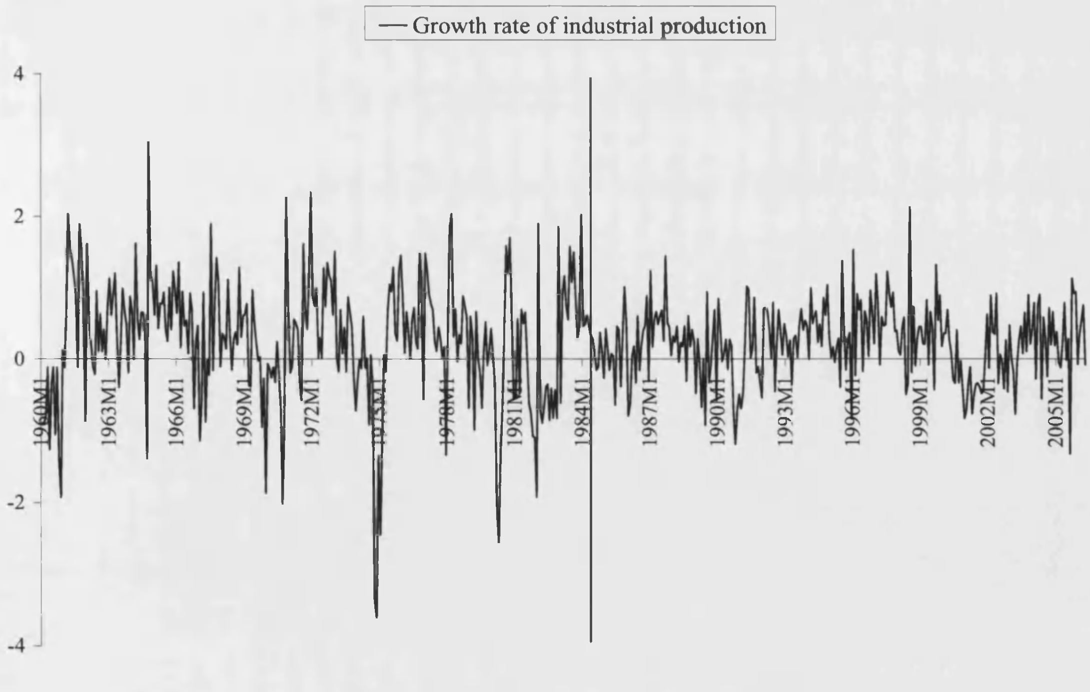


Figure 5.2: Data on growth rate of industrial production for the period 1960M1-2006M4.

	$\mathcal{T}_W^*$	$c_{90\%}^{*B,W}$	$c_{95\%}^{*B,W}$	$c_{99\%}^{*B,W}$
i.i.d.	0.33	0.12	0.13	0.15
$AR(1)$	0.41	0.20	0.23	0.29
$MA(1)$	0.41	0.20	0.25	0.39

	$\mathcal{T}_{LM}^*$	$c_{90\%}^{*B,W}$	$c_{95\%}^{*B,W}$	$c_{99\%}^{*B,W}$
i.i.d.	1.56	0.32	0.37	0.45
$AR(1)$	0.77	0.43	0.48	0.56
$MA(1)$	1.30	0.43	0.48	0.58

Table 5.33: $\mathcal{T}_W^*$ and $\mathcal{T}_{LM}^*$ statistics and their bootstrap critical vaules for growth rate of industrial production for the period 1960M1-2006M4.

	$\mathcal{T}_W^*$	$c_{90\%}^{*B,W}$	$c_{95\%}^{*B,W}$	$c_{99\%}^{*B,W}$
i.i.d.	0.42	0.15	0.17	0.22
$AR(1)$	0.36	0.26	0.30	0.42
$MA(1)$	0.37	0.25	0.29	0.37

	$\mathcal{T}_{LM}^*$	$c_{90\%}^{*B,W}$	$c_{95\%}^{*B,W}$	$c_{99\%}^{*B,W}$
i.i.d.	1.53	0.32	0.37	0.47
$AR(1)$	0.36	0.42	0.47	0.59
$MA(1)$	0.91	0.41	0.46	0.55

Table 5.34: $\mathcal{T}_W^*$ and $\mathcal{T}_{LM}^*$ statistics and their bootstrap critical vaules for growth rate of industrial production for the period 1960M1-1984M3.

	$\mathcal{T}_W^*$	$c_{90\%}^{*B,W}$	$c_{95\%}^{*B,W}$	$c_{99\%}^{*B,W}$
i.i.d.	0.15	0.16	0.19	0.24
$AR(1)$	0.45	0.24	0.28	0.35
$MA(1)$	0.42	0.28	0.33	0.47

	$\mathcal{T}_{LM}^*$	$c_{90\%}^{*B,W}$	$c_{95\%}^{*B,W}$	$c_{99\%}^{*B,W}$
i.i.d.	0.43	0.19	0.23	0.28
$AR(1)$	0.83	0.40	0.45	0.58
$MA(1)$	0.88	0.39	0.45	0.58

Table 5.35: $\mathcal{T}_W^*$ and $\mathcal{T}_{LM}^*$ statistics and their bootstrap critical vaules for growth rate of industrial production for the period 1984M4-2006M4.

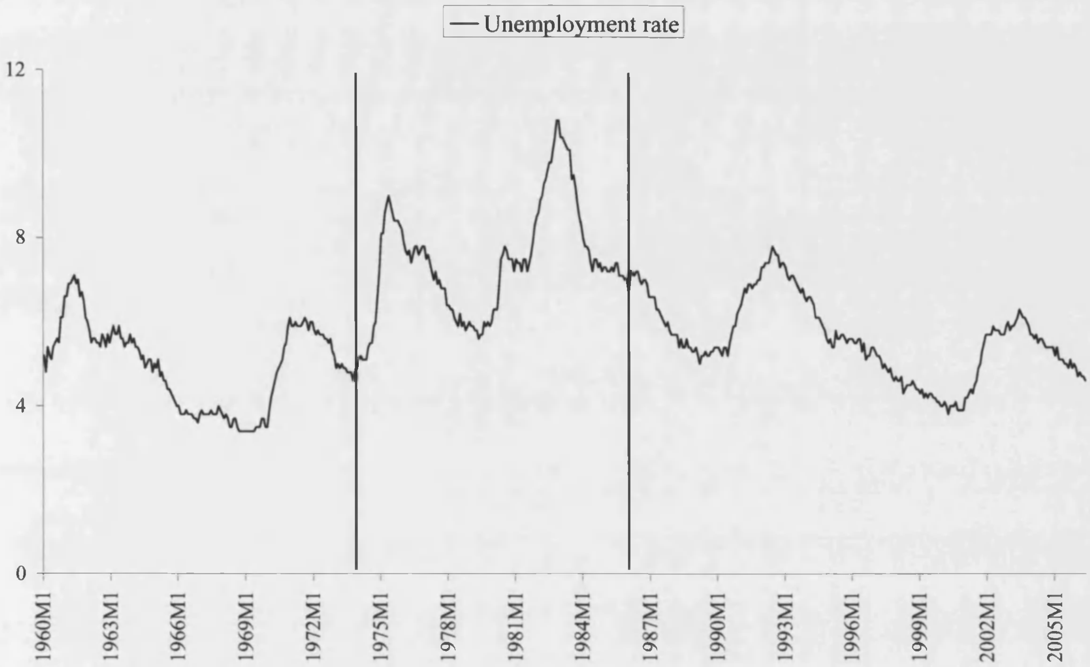


Figure 5.3: Data on unemployment rate for the period 1960M1-2006M5.

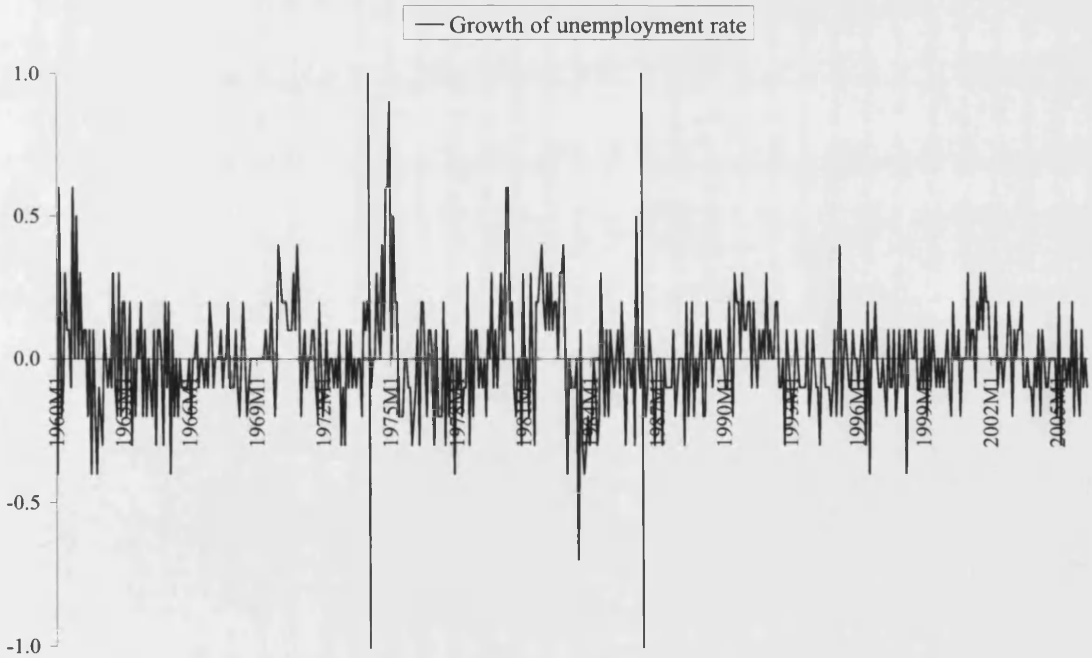


Figure 5.4: Data on growth of unemployment rate for the period 1960M1-2006M4.

	$\mathcal{T}_W^*$	$c_{90\%}^{*B,W}$	$c_{95\%}^{*B,W}$	$c_{99\%}^{*B,W}$
i.i.d.	0.99	0.12	0.14	0.17
$AR(1)$	0.14	0.11	0.13	0.17
$MA(1)$	0.99	0.13	0.15	0.19

	$\mathcal{T}_{LM}^*$	$c_{90\%}^{*B,W}$	$c_{95\%}^{*B,W}$	$c_{99\%}^{*B,W}$
i.i.d.	15.56	0.41	0.46	0.56
$AR(1)$	1.05	0.50	0.56	0.68
$MA(1)$	9.71	0.42	0.49	0.96

Table 5.36: $\mathcal{T}_W^*$ and $\mathcal{T}_{LM}^*$ statistics and their bootstrap critical vaules for unemployment rate for the period 1960M1-2006M5.

	$\mathcal{T}_W^*$	$c_{90\%}^{*B,W}$	$c_{95\%}^{*B,W}$	$c_{99\%}^{*B,W}$
i.i.d.	0.99	0.21	0.24	0.29
$AR(1)$	0.15	0.18	0.22	0.32
$MA(1)$	0.99	0.23	0.26	0.35

	$\mathcal{T}_{LM}^*$	$c_{90\%}^{*B,W}$	$c_{95\%}^{*B,W}$	$c_{99\%}^{*B,W}$
i.i.d.	5.06	0.40	0.46	0.55
$AR(1)$	0.49	0.36	0.41	0.54
$MA(1)$	4.30	0.41	0.46	0.60

Table 5.37: $\mathcal{T}_W^*$ and $\mathcal{T}_{LM}^*$ statistics and their bootstrap critical vaules for unemployment rate for the period 1960M1-1973M12.

	$\mathcal{T}_W^*$	$c_{90\%}^{*B,W}$	$c_{95\%}^{*B,W}$	$c_{99\%}^{*B,W}$
i.i.d.	0.99	0.21	0.24	0.34
$AR(1)$	0.16	0.21	0.26	0.42
$MA(1)$	0.99	0.23	0.28	0.37

	$\mathcal{T}_{LM}^*$	$c_{90\%}^{*B,W}$	$c_{95\%}^{*B,W}$	$c_{99\%}^{*B,W}$
i.i.d.	5.29	0.41	0.46	0.62
$AR(1)$	0.68	0.35	0.39	0.53
$MA(1)$	3.44	0.37	0.42	0.55

Table 5.38: $\mathcal{T}_W^*$ and $\mathcal{T}_{LM}^*$ statistics and their bootstrap critical vaules for unemployment rate for the period 1974M1-1986M2.

	$\mathcal{T}_W^*$	$c_{90\%}^{*B,W}$	$c_{95\%}^{*B,W}$	$c_{99\%}^{*B,W}$
i.i.d.	0.99	0.18	0.21	0.27
$AR(1)$	0.07	0.17	0.19	0.23
$MA(1)$	0.99	0.20	0.23	0.34

	$\mathcal{T}_{LM}^*$	$c_{90\%}^{*B,W}$	$c_{95\%}^{*B,W}$	$c_{99\%}^{*B,W}$
i.i.d.	5.73	0.32	0.36	0.47
$AR(1)$	0.20	0.38	0.42	0.52
$MA(1)$	5.20	0.39	0.46	0.66

Table 5.39: $\mathcal{T}_W^*$ and $\mathcal{T}_{LM}^*$ statistics and their bootstrap critical vaules for unemployment rate for the period 1986M3-2006M5.

Bibliography

ABADIR, K. M., W. DISTASO, AND L. GIRAITIS (2005): "Nonstationarity-extended local Whittle estimation," Preprint.

ADENSTEDT, R. K. (1974): "On large-sample estimation for the mean of a stationary random sequence," *Annals of Mathematical Statistics*, 2(6), 1095–1107.

AKAIKE, H. (1973): "Information theory and an extension of the maximum likelihood principle," in *2nd International Symposium on Information Theory*, ed. by B. N. Petrov, and F. Csaki, pp. 267–281, Akademiai Kiado, Budapest.

ANDERSEN, T. G., AND T. BOLLERSLEV (1997): "Heterogeneous information arrivals and return volatility dynamics: Uncovering the long-run in high frequency returns," *Journal of Finance*, 52(3), 975–1005.

ANDERSON, T. W. (1971): *The Statistical Analysis of Time Series.* Wiley, New York.

ANDREWS, D. W. K. (1999): "Estimation when a parameter is on a boundary," *Econometrica*, 67(6), 1341–1383.

——— (2001): "Testing when a parameter is on the boundary of the maintained hypothesis," *Econometrica*, 69(3), 683–734.

ANDREWS, D. W. K., AND Y. SUN (2004): "Adaptive local polynomial Whittle estimation of long-range dependence," *Econometrica*, 72(2), 569–614.

APOSTOL, T. M. (1967): *Calculus*, vol. 1: One-variable Calculus, with an introduction to Linear Algebra. J. Wiley, 2nd edn.

ARTECHE, J. (2004): "Gaussian semiparametric estimation in long memory in stochastic volatility and signal plus noise models," *Journal of Econometrics*, 119(1), 131–154.

Arteche, J., and P. M. Robinson (2000): "Semiparametric inference in seasonal and cyclical long memory processes," *Journal of Time Series Analysis*, 21(1), 1–25.

Backus, D. K., and S. E. Zin (1993): "Long-memory inflation uncertainty: Evidence from the term structure of interest rates," *Journal of Money, Credit and Banking*, 25(3), 681–700.

Baillie, R. T., C. Chung, and M. A. Tieslau (1996): "Analysing inflation by the fractionally integrated ARFIMA-GARCH model," *Journal of Applied Econometrics*, 11(1), 23–40.

Beran, J., R. J. Bhansali, and D. Ocker (1998): "On unified model selection for stationary and nonstationary short- and long-memory autoregressive processes," *Biometrika*, 85(4), 921–934.

Billingsley, P. (1968): *Convergence of Probability Measures*. Wiley, New York, 1st edn.

Bloomfield, P. (1973): "An exponential model for the spectrum of a scalar time series," *Biometrika*, 60(2), 217–226.

Box, G. E. P., and G. M. Jenkins (1976): *Time Series Analysis: Forecasting and Control*. Holden-Day, San Fransisco, revised edn.

Breidt, F. J., N. Crato, and P. de Lima (1998): "The detection and estimation of long memory in stochastic volatility," *Journal of Econometrics*, 83(2), 325–348.

Brillinger, D. R. (1975): *Time Series, Data Analysis and Theory*. Holden-Day, San Fransisco.

Brockwell, P. J., and R. A. Davis (1991): *Time Series: Theory and Methods*, Spinger Series in Statistics. Springer-Verlag, New York, 2nd edn.

Bühlmann, P. (1997): "Sieve bootstrap for time series," *Bernoulli Journal of Mathematical Statistics and Probability*, 3(2), 123–148.

Cheung, Y. (1993): "Long memory in foreign-exchange rates," *Journal of Business and Economic Statistics*, 11(1), 93–101.

Clarida, R., J. Gali, and M. Gertler (2000): "Monetary policy rules and macroeconomic stability: Evidence and some theory," *Quarterly Journal of Economics*, 115(1), 147–180.

DAHLHAUS, R. (1989): "Efficient parameter estimation for self-similar processes," *Annals of Statistics*, 17(4), 1749–1766.

DAHLHAUS, R., AND D. JANAS (1996): "A frequency domain bootstrap for ration statistics in time series analysis," *Annals of Statistics*, 24(5), 1934–1963.

DAVIES, R. B., AND D. S. HARTE (1987): "Tests for Hurst effect," *Biometrika*, 74(1), 95–101.

DELGADO, M. A., J. HIDALGO, AND C. VELASCO (2005): "Distribution free goodness-of-fit tests for linear processes," *Annals of Statistics*, 33(6), 2568–2609.

DEO, R., AND C. HURVICH (2001): "On the log periodogram regression estimator of the memory parameter in long memory stochastic volatility models," *Econometric Theory*, 17(4), 686–710.

DIEBOLD, F. X., AND G. D. RUDEBUSCH (1989): "Long memory and persistence in aggregate output," *Journal of Monetary Economics*, 24(2), 189–209.

——— (1991): "Is consumption too smooth? Long memory and the Deaton paradox," *Review of Economics and Statistics*, 73(1), 1–9.

——— (1999): *Business Cycles: Durations, Dynamics, and Forecasting.* Princeton University Press, New Jersey.

DING, Z., AND C. W. J. GRANGER (1996): "Modelling volatility persistence of speculative returns: A new approach," *Journal of Econometrics*, 73(1), 185–215.

DING, Z., C. W. J. GRANGER, AND R. F. ENGLE (1993): "A long memory property of stock market returns and a new model," *Journal of Empirical Finance*, 1(1), 83–106.

DOBRUSHIN, R. L., AND P. MAJOR (1979): "Non-central limit theorems for non-linear functionals of Gaussian fields," *Zeitschrift für Wahrsscheinlichkeit Theorie und verwandte Gebiete*, 50(1), 27–52.

DOUKHAN, P. (1994): *Mixing: Properties and Examples*, vol. 85 of *Lecture Notes in Statistics.* Springer-Verlag, Berlin.

EFRON, B. (1979): "Bootstrap methods: Another look at the Jackknife," *Annals of Statistics*, 7(1), 1–26.

FAMA, E. F. (1975): "Short-term interest rates as predictors of inflation," *The American Economic Review*, 65(3), 269–282.

FISHER, I. (1930): *The Theory of Interest.* MacMillan, New York.

FOX, R., AND M. S. TAQQU (1986): "Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series," *Annals of Statistics*, 14(2), 517–532.

FRANKE, J., AND W. HÄRDLE (1992): "On bootstrapping kernel spectral estimates," *Annals of Statistics*, 20(1), 121–145.

FRANKE, J., AND J. P. KREISS (1992): "Bootstrapping stationary autoregressive moving average models," *Journal of Time Series Analysis*, 13, 297–319.

GARCIA, R., AND P. PERRON (1996): "An analysis of the real interest rate under regime shifts," *Review of Economics and Statistics*, 78(1), 111–125.

GEWEKE, J., AND S. PORTER-HUDAK (1983): "The estimation and application of long memory time series models," *Journal of Time Series Analysis*, 4(4), 221–238.

GIL-ALAÑA, L. A., AND P. M. ROBINSON (1997): "Testing of unit root and other nonstationary hypotheses in macroeconomic time series," *Journal of Econometrics*, 80(2), 241–268.

GINÉ, E., AND J. ZINN (1990): "Bootstrapping general empirical measures," *Annals of Probability*, 18(2), 851–869.

GIRAITIS, L., J. HIDALGO, AND P. M. ROBINSON (2001): "Gaussian estimation of parametric spectral density with unknown pole," *Annals of Statistics*, 29(4), 987–1023.

GIRAITIS, L., AND R. LEIPUS (1995): "A generalized fractionally differencing approach in long-memory modelling," *Lithuanian Mathematical Journal*, 36(1), 53–65.

GIRAITIS, L., P. M. ROBINSON, AND A. SAMAROV (2000): "Adaptive semiparametric estimation of the memory parameter," *Journal of Multivariate Analysis*, 72(2), 183–207.

GIRAITIS, L., AND D. SURGAILIS (1985): "CLT and other limit theorems for functionals of Gaussian process," *Zeitschrift für Wahrsscheinlichkeit Theorie und verwandte Gebiete*, 70, 191–212.

——— (1990): "A central limit theorem for quadratic forms in strongly dependent linear variables and application to asymptotical normality of Whittle's estimate," *Probability Theory and Related Fields*, 86, 87–104.

GIRAITIS, L., AND M. S. TAQQU (1999): "Whittle estimation for finite-variance non-Gaussian time series with long memory," *Annals of Statistics*, 27(1), 178–203.

GONÇALVES DA SILVA, A., AND P. M. ROBINSON (2006): "Finite sample performance in cointegration analysis of nonlinear time series with long memory," forthcoming in *Econometric Reviews*.

GRANGER, C. W. J. (1966): "The typical spectral shape of an economic variable," *Econometrica*, 34(1), 150–161.

GRANGER, C. W. J., AND R. JOYEUX (1980): "An introduction to long-memory time series and fractional differencing," *Journal of Time Series Analysis*, 1(1), 15–29.

GRAY, H. L., N. F. ZHANG, AND W. A. WOODWARD (1989): "On generalized fractional processes," *Journal of Time Series Analysis*, 10, 233–257.

GREENE, M. T., AND B. D. FIELITZ (1977): "Long-term dependence in common stock returns," *Journal of Financial Economics*, 4(3), 339–349.

HALL, P. (1979): "On the rate of convergence of normal extremes," *Journal of Applied Probability*, 16(2), 433–439.

——— (1990): "On the relative performance of bootstrap and Edgeworth approximations of a distribution function," *Journal of Multivariate Analysis*, 35(1), 108–129.

——— (1991): "On convergence rates of suprema," *Probability Theory and Related Fields*, 89(4), 447–455.

HAMILTON, J. D. (1990): "Analysis of time series subject to changes in regime," *Journal of Econometrics*, 45(1), 39–70.

HANNAN, E. J. (1970): *Multiple Time Series*. Wiley, New York.

——— (1973): "The asymptotic theory of linear time series models," *Journal of Applied Probability*, 10, 130–145.

HANNAN, E. J., AND B. G. QUINN (1979): "The determination of the order of an autoregression," *Journal of the Royal Statistical Society. Series B (Methodological)*, 41(2), 190–195.

HARVEY, A. C. (1998): "Long memory in stochastic volatility," in *Forecasting volatility in financial markets*, ed. by J. Knight, and S. Satchell, pp. 307–320, Oxford. Butterworth-Heinemann.

HENRY, M., AND P. M. ROBINSON (1996): "Bandwidth choice in Gaussian semiparametric estimation of long range dependence," in *Athens Conference on Applied Probablity and Time Series Analysis. Vol. II: Time Series Analysis*, ed. by P. M. Robinson, and M. Rosenblatt, vol. 115 of *Lecture Notes in Statistics*, pp. 220–232. Springer-Verlag.

HIDALGO, J. (2005): "Semiparametric estimation for stationary processes whose spectra have an unknown pole," *Annals of Statistics*, 33(4), 1843–1889.

——— (2006): "A nonparametric test for weak dependence and its bootstrap analogue," forthcoming in *Journal of Time Series Analysis.*

HIDALGO, J., AND J. P. KREISS (2004): "Bootstrap specification tests for covariance stationary linear processes," forthcoming in *Journal of Econometrics.*

HIDALGO, J., AND P. SOULIER (2004): "Estimation of the location and exponent of the spectral singularity of a long memory process," *Journal of Time Series Analysis*, 25(1), 55–81.

HIDALGO, J., AND Y. YAJIMA (2003): "Semiparametric estimation of the long-range parameter," *Annals of the Institute of Statistical Mathematics*, 55(4), 705–736.

HOSOYA, Y. (1997): "Limit theory with long-range dependence and statistical inference of related models," *Annals of Statistics*, 25, 105–137.

HUIZINGA, J., AND F. S. MISHKIN (1986): "Monetary policy regime shifts and the unusual behavior of real interest rates," *Carnegie-Rochester Conference Series on Public Policy*, 24, 231–274.

HURST, H. E. (1951): "Long-term storage capacity of reservoirs," *Transactions of the American Society of Civil Engineers*, 116, 770–808.

HURVICH, C. M., AND J. BRODSKY (2001): "Broadband semiparametric estimation of the memory parameter of a long-memory time series using fractional exponential model," *Journal of Time Series Analysis*, 22(2), 221–249.

HURVICH, C. M., E. MOULINES, AND P. SOULIER (2005): "Estimating long memory in volatility," *Econometrica*, 73(4), 1283–1328.

KING, R. G., AND M. W. WATSON (1996): "Money, prices, interest rates and the business cycle," *Review of Economics and Statistics*, 78(1), 35–53.

KÜNSCH, H. (1986): "Discrimination between monotonic trends and long-range dependence," *Journal of Applied Probability*, 23(4), 1025–1030.

——— (1987): "Statistical aspects of self-similar process," in *Proceedings of the 1st World Congress of the Bernoulli Society*, ed. by Y. A. Prohorov, and V. V. Sazonov, vol. 1, pp. 67–74, Utrecht. Science Press.

KÜNSCH, H. R. (1989): "The jackknife and the bootstrap for general stationary observations," *Annals of Statistics*, 17(3), 1217–1241.

LO, A. W. (1991): "Long-term memory in stock market prices," *Econometrica*, 59(5), 1279–1313.

LOBATO, I., AND P. M. ROBINSON (1996): "Averaged periodogram estimation of long memory," *Journal of Econometrics*, 73, 303–324.

LOBATO, I. N., AND P. M. ROBINSON (1998): "A nonparametric test for I(0)," *Review of Economic Studies*, 65, 475–495.

LOBATO, I. N., AND C. VELASCO (2000): "Long memory in stock-market trading volume," *Journal of Business and Economic Statistics*, 18(4), 410–427.

MANDELBROT, B. B. (1975): "Limit theorems on the self-normalized range for weakly and strongly dependent processes," *Probability Theory and Related Fields*, 31(4), 271–285.

MANDELBROT, B. B., AND J. W. VAN NESS (1968): "Fractional Brownian motions, fractional noises and applications," *SIAM Review*, 10(4), 422–437.

MCLEISH, D. L. (1975): "Invariance principles for dependent variables," *Probability Theory and Related Fields*, 32(3), 165–178.

MISHKIN, F. S. (1981): "The real interest rate: an empirical investigation," *Carnegie-Rochester Conference Series on Public Policy*, 15, 151–200.

——— (1992): "Is the Fisher effect for real?," *Journal of Monetary Economics*, 30, 195–215.

MOULINES, E., AND P. SOULIER (1999): "Broad band log-periodogram regression of time series with long range dependence," *Annals of Statistics*, 27(4), 1415–1439.

NELSON, C. R., AND C. I. PLOSSER (1982): "Trends and random walks in macroeconomic time series," *Journal of Monetary Economics*, 10(2), 139–162.

NELSON, C. R., AND G. W. SCHWERT (1977): "Short-term interest rates as predictors of inflation: On testing the hypothesis that the real rate of interest is constant," *American Economic Review*, 67(3), 478–486.

NZE, P. A., P. BÜHLMANN, AND P. DOUKHAN (2002): "Weak dependence beyond mixing and asymptotics for nonparametric regression," *Annals of Statistics*, 30(2), 397–430.

PAPELL, D. H., C. J. MURRAY, AND H. GHIBLAWI (2000): "The structure of unemployment," *Review of Economics and Statistics*, 82(2), 309–315.

PHILLIPS, P. C. B. (1998): "Econometric analysis of Fisher's equation," Cowles Foundation Discussion Paper No. 1180, Yale University.

POLITIS, D. N., AND J. P. ROMANO (1994): "Large sample confidence regions based on subsamples under minimal assumptions," *Annals of Statistics*, 22(4), 2031–2050.

ROBINSON, P. M. (1994a): "Time series with strong dependence," in *Advances in Econometrics*, ed. by C. A. Sims, vol. 1, pp. 47–95, Cambridge. Cambridge University Press.

——— (1994b): "Semiparametric analysis of long-memory time series," *Annals of Statistics*, 22(1), 515–539.

——— (1994c): "Efficient tests of nonstationary hypothesis," *Journal of the American Statistical Association*, 89(428), 1420–1437.

——— (1995a): "Log-periodogram regression of time series with long range dependence," *Annals of Statistics*, 23, 1048–1072.

——— (1995b): "Gaussian semiparametric estimation of long range dependence," *Annals of Statistics*, 24, 1630–1661.

ROBINSON, P. M., AND J. HIDALGO (1997): "Time series regression with long-range dependence," *Annals of Statistics*, 25(1), 77–104.

ROSE, A. K. (1988): "Is the real interest rate stable?," *Journal of Finance*, 43(5), 1095–1112.

ROSENBLATT, M. (1956): "A central limit theorem and a strong mixing condition," *Proceedings of the National Academy of Science of the USA*, 42, 43–47.

SCHWARZ, G. (1978): "Estimating the dimension of a model," *Annals of Statistics*, 6(2), 461–464.

SCHWERT, G. W. (1987): "Effects of model specification on tests for unit roots in macroeconomic data," *Journal of Monetary Economics*, 20(1), 73–103.

SHIMOTSU, K., AND P. C. B. PHILLIPS (2005): "Exact local Whittle estimation of fractional integration," *Annals of Statistics*, 33(4), 1890–1933.

SINAI, Y. G. (1976): "Self-similar probability distributions," *Theory of Probability and its Applications*, 21(1), 64–80.

SOWELL, F. (1992): "Modelling long-run behavior with the fractional ARFIMA model," *Journal of Monetary Economics*, 29(2), 277–302.

STOUT, W. F. (1974): *Almost Sure Convergence*. Academic Press, New York.

SUN, Y., AND P. C. B. PHILLIPS (2003): "Nonlinear log-periodogram regression for pertubed fractional processes," *Journal of Econometrics*, 115(2), 355–389.

——— (2004): "Understanding the Fisher equation," *Journal of Applied Econometrics*, 19(7), 869–886.

TAQQU, M. S. (1979): "Convergence of integrated processes of arbitrary Hermite rank," *Zeitschrift für Wahrsscheinlichkeit Theorie und verwandte Gebiete*, 50, 53–83.

TAYLOR, J. B. (1993): "Discretion versus policy rules in practice," *Carnegie-Rochester Conference Series on Public Policy*, 39, 195–214.

TAYLOR, S. J. (1994): "Modelling stochastic volatility: a review and a comparative study," *Mathematical Finance*, 4, 183–204.

VAN DIJK, D., D. R. OSBORN, AND M. SENSIER (2002): "Changes in variability of the business cycle in the G7 countries," Econometric Institute Report EI 2002-28.

VELASCO, C. (1999): "Gaussian semiparametric estimation for non-stationary time series," *Journal of Time Series Analysis*, 20, 87–127.

VELASCO, C., AND P. M. ROBINSON (2000): "Whittle pseudo-maximum likelihood estimation of nonstationary time series," *Journal of the American Statistical Association*, 95, 1229–1243.

VOLKONSKII, V., AND Y. ROZANOV (1959): "Some limit theorems for random functions. I," *Theory of Probability and its Applications*, 4(2), 178–197.

WHITTLE, P. (1951): *Hypothesis Testing in Time Series Analysis.* Almqvist and Wiksells, Uppsala.

WU, C. F. J. (1986): "Jackknife, bootstrap and other resampling methods in regression analysis," *Annals of Statistics*, 14(4), 1261–1295.

YAJIMA, Y. (1996): "Estimation of the frequency of unbounded spectral densities," *Proceedings of the Business and Economic Statistics Section*, American Statistical Association, 4–7.