The robustness of long memory estimators in the presence of short-range dependence and fat-tails

Wen Hoong Koay & Wei Ying Soh

November 22, 2012

Abstract

When attempting to discern whether a particular time series exhibits long memory or not, the ability of the long memory estimator to isolate the long memory component from other characteristics (e.g., short-run dynamics, leptokurtosis) of the time series becomes of first order importance. We employ a Monte Carlo study to test the robustness and analyze the finite sample behavior of the estimators of long-range dependence for eight estimation methods. We find that no single estimator consistently outperforms the others for our extensive set of simulated time series with Gaussian, generalized hyperbolic, and stable innovations.

1 Introduction

Long memory processes have been observed in many time series in a broad range of disciplines, including physical sciences, behavioral sciences, and network management. Though the effect of innovations are persistent, the shock eventually dies out. Long memory, fractionally integrated, or long-range dependent processes are typically characterized as having hyperbolically decaying autocorrelation functions. Indeed, this has led to the development of a range of methods and estimators to characterize the long-range dependence observed in these time series. It is not unreasonable to suggest that the innovations for time series that appear to have long-range dependence may also have a short memory component or be non-Gaussian. Thus, when attempting to discern whether a particular time series exhibits long memory or not, the ability of the long memory estimator to isolate the long memory component from other characteristics (e.g., short-run dynamics, leptokurtosis) of the time series becomes of first order importance.

Within the context of long-range dependence estimation, Taqqu, Teverovsky, and Willinger (1995) argue that real-world data are rarely well-behaved, in that the time series is unlikely to be Gaussian and solely possess a long memory component. Real-world data are likely to possess complex dynamics that potentially renders inference of a long memory process in the time series unreliable. That is, estimation of the parameters that characterize long memory may inappropriately lead to spurious identification of long-range dependence though absent from the actual time series, and vice-versa. We employ a Monte Carlo study to test the robustness and analyze the finite sample behavior of the estimators of long-range dependence for eight estimation methods. The estimation methods we consider and test in our paper are as follows: (i) rescaled range analysis (R/S); (ii) detrended fluctuation analysis (DFA); (iii) Higuchi's fractal dimension method; (iv) modified periodogram method; (v) Whittle method; (vi) Geweke-Porter-Hudak method; (vii) approximate maximum likelihood estimation; and (viii) smoothed periodogram method. We vary sample path lengths and include both short-range dependence structures and non-Gaussian innovations in simulating the set of time series to which we apply each of the eight estimation methods.

A considerable body of research suggests the existence of long-range dependence in macroeconomic and financial time series. In the macroeconomics literature, Diebold and Rudebusch (1989) document that quarterly post World War II U.S. real GNP data exhibits long memory. Baillie, Chung, and Tieslau (1996b) find evidence of long memory in U.S. Consumer Price Index time series, while Backus and Zin (1993) provide evidence of long memory in the term structure of interest rates. That financial time series can exhibit long-range dependence is of particular interest. The fundamental principle of the efficient markets hypothesis is that prices reflect all publicly available information, thus precluding persistent speculative excess returns. Early empirical tests of stock market efficiency by Fama (1965) and Fisher (1966) find evidence that the first-order autocorrelation of daily, weekly, and monthly stock returns are significantly positive. Lo and MacKinlay (1988) and Conrad and Kaul (1988) document further evidence of significantly positive first-order autocorrelations in stock returns. In particular, the returns of portfolios consisting of small stocks show stronger positive autocorrelation. Despite this seemingly persistent statistical dependence at short-horizons, the efficient markets literature argues that the deviations from zero of the autocorrelation functions are economically insignificant. However, Shiller (1984) and Summers (1986) argue that even under the assumption that stock returns over short-horizon are uncorrelated, the market can still be inefficient if the return generating process exhibits long-range dependence, thereby implying predictability in stock returns over longer time horizons. Empirical investigations (see, e.g., Fama and French (1988), Jegadeesh (1990), and Porterba and Summers (1988)) find mixed evidence on the predictability of long-run stock returns. Empirical anomalous stock return behavior at longer horizons is alternately attributed to speculative bubbles or time-varying conditional expected returns. Lo (1991) emphasizes the potential implications on optimal asset allocation and derivative pricing if asset returns exhibit long-range dependence. In the presence of long memory, optimal asset allocation becomes sensitive to both the investment horizon and the initial conditions, while problems arise in pricing long-term derivative contracts such as options using models based on the assumption that asset prices can be characterized by a geometric Brownian motion. The implications of the numerous empirical studies on the predictability of asset returns illustrates the importance of the robustness of long memory estimators in identifying potential long-range dependence in the presence of other dynamics (e.g., short memory, leptokurtosis), particularly in financial time series.

While there still exists much debate as to whether asset returns are uncorrelated through time, it is generally accepted that asset returns are not independent through time. Taylor (1986) provides evidence that the absolute returns of various financial assets tend to have slowly decaying autocorrelation functions. The returns of financial assets are typically characterized as having time-varying volatility; returns display periods of low volatility and periods of high volatility, and this volatility is highly persistent. This regularity, or stylized fact, of the second moment of the return generating process has attracted significant attention within the econometric literature. Ding, Granger, and Engle (1993) document that squared daily logarithmic returns of the S&P500 index display a slowly decaying autocorrelation function. Similarly, Dacorogna, Muller, Nagler, Olsen, and Pictet (1993) and Andersen and Bollerslev (1997) find that squared logarithmic intra-daily returns for stocks and exchange rates, respectively, have slowly decaying autocorrelation functions. The slow rate of decay in the autocorrelation functions of squared returns at both the daily and intra-daily level are consistent with long memory processes. By extending the standard autoregressive conditional heteroskedasticity (ARCH) and generalized ARCH (GARCH) models of Engle (1982) and Bollerslev (1986), respectively, several models of the conditional variance process have been proposed to model the long memory in asset return volatility. Baillie, Bollerslev, and Mikkelsen's (1996) fractionally integrated GARCH and Davidson's (2004) hyperbolic GARCH imply a hyperbolic rate of decay in the autocorrelation function of the conditional variance process. Andersen, Bollerslev, Diebold, and Labys (2003), and Chiriac and Voev (2011) model long memory in realized volatility time series using fractionally integrated autoregressive moving average (ARFIMA) models. Indeed, these studies within the volatility literature estimate and model long-range dependence using maximum likelihood estimation of the GARCH and ARMA class of models. Whether the parametric estimation of long memory is appropriate is jointly considered in our study.

Long memory properties have also been observed in several properties of the microstructure of financial markets. Lobato and Velasco (2000) document that stock market trading volume exhibits long-range dependence. Lillo and Farmer (2004) find that individual buy- and sell-orders on the London Stock Exchange are positively autocorrelated, and have a hyperbolically decaying autocorrelation function. Furthermore, Engle and Russell (1998) analyze the time duration between financial transactions and provide evidence suggesting clustering and long memory in trade durations.

The remainder of the paper is organized as follows. In Section 2, we provide a formal definition of long memory and discuss the estimation methods. We provide detail of the properties of the simulated time series and the simulation approach of our Monte Carlo study in Section 3. The results are presented in Section 4 and the paper concluded in Section 5.

2 Estimation methods for long memory

We first provide a formal definition of long memory processes to aid discussion of the eight estimation methods analyzed in this paper. Long memory may be defined in terms of the asymptotic behavior of the autocorrelation function. McLeod and Hipel (1978) show that if, for a discrete time series X_t, with autocorrelation ρ_k, where k indexes the lag order,

$$\lim_{n\to\infty} \sum_{k=-n}^{n} |\rho_k| \tag{1}$$

is non-finite, the process has long memory. More generally, a defining characteristic of long memory processes has been taken to be an autocovariance function

γ_k that has a slower than exponential rate of decay. A process X_t that satisfies the following difference equation

$$(1-L)^d X_t = \varepsilon_t, \qquad \varepsilon_t \sim iid(0, \sigma_\varepsilon^2) \tag{2}$$

where L is the lag operator and d is the fractional differencing parameter, is said to be fractionally differenced of order d when d assumes non-integer values. Hosking (1981) shows that for values of $-0.5 < d < 0.5$ the process given by equation (2) is stationary and invertible. The autocorrelation coefficients of X_t are of the same sign as the fractional differencing parameter d. The ARFIMA(p, d, q) time series models of Granger and Joyeux (1980) and Hosking (1981) satisfy the definition of long memory given by equation (1) for $0 < d < 0.5$. For these values of d, the series has a positive autocorrelation function with a hyperbolic rate of decay. A value of $d = 0$ implies no long-range dependence. For $-0.5 < d < 0$, the process has negative autocorrelations and is antipersistent. The autocovariance function implied by equation (2) is given by

$$\gamma_k \sim ck^{2d-1} \quad as\ k \to \infty\ , \tag{3}$$

where $-0.5 < d < 0.5$ and c is some constant. Thus, it can be clearly seen in (3) that the autocovariance function decays at a hyperbolic rate governed by the value of d. Values of d closer to 0.5 imply greater long-range dependence in the time series. When $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$ equation (2) represents, in standard notation, an ARFIMA$(0, d, 0)$ process with no short-range dependence and Gaussian innovations. We simulate the time series in our Monte Carlo study using an ARFIMA(p, d, q) process. A detailed discussion of the properties of the simulated ARFIMA(p, d, q) process (i.e., the chosen values for d, ARMA coefficients, and the type of innovation of the process) and the simulation methodology are reserved for Section 3.

Long memory may also be characterized in terms of the Hurst coefficient H, where $0 < H < 1$. Brownian motion is a continuous-time stochastic process with independent increments that are Gaussian. Mandelbrot and Van Ness (1968) develop fractional Brownian motion that incorporates correlated Gaussian increments and nests Brownian motion. For values of $0 < H < 0.5$, the increments are negatively correlated; for values of $0.5 < H < 1$, the increments are positively correlated; for $H = 0.5$, the process reduces to Brownian motion with uncorrelated increments. Mandelbrot and Van Ness (1968) show that the autocovariance function of fractional Brownian motion is given by

$$\gamma_k \approx |k|^{2H-2} \tag{4}$$

for high order lags, k. As such, for large values of k, the autocovariance function decays at a slow hyperbolic rate. A simple relation between H and d exists. For finite variance processes, $H = d + 1/2$, and for infinite variance processes (e.g., stable distribution with $\alpha < 2$), $H = d + 1/\alpha$.

In describing long memory processes in terms of the autocovariance function, the properties of the process X_t are analyzed within the time-domain (i.e., how the process changes with respect to time). The properties of a covariance-stationary process may, equivalently, be analyzed in the frequency-domain (see, e.g., Andersen (1971), Bloomfield (1976), and Fuller (1976)). Frequency-domain, or spectral analysis, aims to decompose the observed series (i.e., X_t) into simpler periodic components, such as $\sin(\omega)$ and $\cos(\omega)$, with particular frequencies ω_j in order to determine the contribution of these components to the behavior of X_t. The spectral density, $s_X(\omega_j)$ characterizes the portion of the variance of X_t that can be attributed to the periodic components, or cycles, with frequency ω_j. A long memory process has a spectral density that diverges to infinity as the frequency tends to zero,

$$\lim_{\omega \to 0} s_X(\omega) = \infty \quad \cdot \tag{5}$$

By noting that $1/\omega \propto T$, where T is the period of the cycle, (5) implies that, for long memory processes, low-frequency, or equivalently long period, components explain much of the variation in X_t.

The periodogram is an estimate of the spectral density. Long-range dependent processes should have periodograms that satisfy

$$s_X(\omega) \propto |\omega|^{1-2H} \tag{6}$$

for ω close to zero. Ordinary least squares regression of the log-log plot of the periodogram against functions of the frequency give an estimate of the long memory of the series.

To summarize the preceding discussion, long-range dependence corresponds to values of the fractional differencing parameter $0 < d < 0.5$, or equivalently, values of the Hurst coefficient $0.5 < H < 1$. A long memory process may also be analyzed in the frequency-domain. The estimation methods discussed below either estimate d or H, thus the distinction between these parameters becomes relevant when comparing the long memory parameter estimates. Rescaled range analysis, detrended fluctuation analysis, Higuchi's fractal dimension, and the modified periodogram method seek to estimate H. The Whittle estimator, Geweke-Porter-Hudak method, approximate maximum likelihood estimation, and the smoothed periodogram method estimate d.

2.1 Rescaled range (R/S) analysis

The classical statistical measure of long memory developed by Hurst (1951) is the range over standard deviation statistic, and is also known as the rescaled range or R/S statistic. The R/S analysis estimation method involves calculation of the R/S statistic to derive a nonparametric estimate of H. The R/S statistic is calculated by dividing a time series into k non-overlapping blocks of equal length n. The sample average $\overline{X}_n = 1/n \sum_{j=1}^{t} X_j$ is then calculated for each of the k blocks. The range of the partial sum of deviations from the mean of each of the k blocks, is given by

$$R_n = \max_{1<k\leqslant n} \sum_{j=1}^{k} (X_j - \overline{X}_n) - \min_{1<k\leqslant n} \sum_{j=1}^{k} (X_j - \overline{X}_n) \quad . \tag{7}$$

The range R_n is then scaled by the sample standard deviation of the block, s_n. The R/S statistic is hence defined as

$$\frac{R_n}{s_n} = \frac{1}{s_n}\left[\max_{1<k\leqslant n} \sum_{j=1}^{k} (X_j - \overline{X}_n) - \min_{1<k\leqslant n} \sum_{j=1}^{k} (X_j - \overline{X}_n)\right] \quad . \tag{8}$$

Hurst (1951) shows that

$$log\left[E\left(\frac{R_n}{s_n}\right)\right] \approx c + Hlog(n) \ , \tag{9}$$

where c is some constant. That is, R/S analysis involves identifying the long-range dependence of a process by plotting the logarithm of the R/S statistics against logarithms of the block length n and estimating the slope coefficient H by ordinary least squares regression.

Mandelbrot and Wallis (1969) show using Monte Carlo simulation that the estimate of H obtained through R/S analysis is able to identify long-range dependence, and that the estimator of H is robust to time series with large skewness and kurtosis. However, Davies and Harte (1987) and Lo (1991) argue that R/S analysis is prone to spuriously identifying long memory if the time series has short-range dependence. Davies and Harte (1987) show that even for a stationary AR(1) (i.e., a Hurst coefficient of 0.5) with an autoregressive parameter of 0.3, R/S analysis rejects the null of no long memory 47 percent of the time. Lo (1991) provides additional Monte Carlo results that suggest R/S estimator of H is not robust in the presence of short-range dependence and heteroskedasticity.

2.2 Detrended fluctuation analysis (DFA)

The DFA method of Peng, Havlin, Stanley, and Goldberger (1995) involves first calculating the partial sum of the time series then dividing the entire time series into k blocks of length n. Within each of the k blocks the "local trend" is derived by fitting an ordinary least squares regression of the form $\beta_0+\beta_1 t$ to the partial sum series of the block. The least squares regression can be extended to a polynomial of higher order to account for higher order local trends. We fit a least squares regression of polynomial order 1 (i.e., $\beta_0+\beta_1 t$) in our analysis as is done in the original paper by Peng et al. (1995). The fluctuation of the series $F(n)$ is then calculated as the root-mean-square deviation

$$F(n) = \left[\frac{1}{n}\sum_{t=1}^{n}(Y_t - \beta_0 - \beta_1 t)\right]^{\frac{1}{2}} . \tag{10}$$

Equation (10) is the sample variance of the residuals within the block. Estimation of the local trend and calculation of the fluctuation is successively repeated for varying block lengths n.

Peng et al. (1995) argue that $F(n)$ typically increases as the length of the block increases; Taqqu et al. (1995) show that $F(n) \propto n^{\alpha}$, where α is the scaling exponent. Ordinary least squares regression of the slope of the log-log plot of $F(n)$ against n provides an estimate of α. For values of $0 < \alpha < 1$, α has the same interpretation as H. For $0 < \alpha < 0.5$ the process is antipersistent; values of $0.5 < \alpha < 1$ imply the process has long memory; whereas an estimate of $\alpha = 0.5$ indicates a random walk. Peng et al. (1995) suggest that though estimates of the slope for time series with short memory but no long memory may initially deviate from 0.5, the slope approaches 0.5 as n increases. Taqqu et al. (1995) document that while the DFA method estimate for long-range dependent Gaussian time series is efficient, the estimator suffers from bias when the process is also short-range dependent.

2.3 Higuchi's fractal dimension method

Higuchi (1988) suggested a measure of long memory based on the fractal dimension D of a series. The basic principle underlying the Higuchi method is that a fractal curve can be regarded as a reduced scale version of the original curve. Thus, based on this scaling property the following proportion holds

$$L(n) \propto n^{-D} \ , \tag{11}$$

where $D = 2 - H$. Similar to R/S analysis and DFA, an estimate of H is given by the ordinary least squares estimator of the slope of a log-log plot of $L(n)$ against n.

2.4 Modified periodogram method

The classical R/S analysis, DFA, and the Higuchi method provide estimates of H based on ordinary least squares regression of log-transformed values within the time-domain. The basic linear regression principle of these three methods can also be applied within the frequency-domain to obtain an estimate of the long memory parameter. The standard periodogram method involves plotting the log periodogram for different frequencies ω_j against the logarithm of ω_j and estimating the slope, $1 - 2H$, through ordinary least squares regression. Given that the proportion in (6) holds only for low frequencies, the log periodograms at high frequencies may skew inference on the long-range dependence of the series. In practice, the slope is estimated for low frequencies to account for the influence of log periodogram points with high frequencies.

The modified periodogram method attempts to correct for points corresponding to large ω_j in the log-log periodogram-frequency plot by dividing the frequency axis into logarithmically spaced blocks and averaging the log periodogram values within the blocks. Low-frequency points are typically not averaged. Ordinary least squares regression is then fitted to the low-frequency and averaged log periodograms to estimate the slope, $1 - 2H$.

2.5 Whittle method

Fox and Taqqu (1986) propose an approximate maximum likelihood estimation approach within the frequency domain. To avoid potential confusion, we refer to Fox and Taqqu's (1986) approximate MLE method within the frequency-domain as the Whittle method and Chung and Baillie's (1993) conditional sum-of-squares method (discussed in Section 2.7) within the time-domain as approximate MLE. The Whittle method provides a parametric estimate of d by minimizing the function

$$L(\omega;\theta) = \int_{-\pi}^{\pi} \frac{I(\omega)}{f(\omega;\theta)} d\omega \tag{12}$$

where $I(\omega)$ is the periodogram, $f(\omega;\theta)$ is the assumed spectral density when ω is close to zero and θ is a vector of unknown parameters that includes d and potentially ARMA coefficients if the series is assumed to be an ARFIMA(p, d, q) process with $p > 0$ and/or $q > 0$. Fox and Taqqu (1986) provide a proof of the Whittle estimator's asymptotically normal distribution. Cheung and Diebold (1994) compare the Whittle method to Sowell's (1992) exact maximum likelihood method and find that when specifying an ARFIMA$(0, d, 0)$ model where the mean parameter μ needs to be estimated, there is little difference between the performance of the Whittle and exact maximum likelihood estimators of d. This is a particularly relevant result given that, in practice, the exact specification of the true underlying process is unknown.

2.6 Geweke-Porter-Hudak (GPH) method

Geweke and Porter-Hudak (1983) propose a semiparametric method of estimating the fractional differencing parameter d within the frequency domain based on a least squares regression of the log periodogram for frequencies ω_j close to zero on a deterministic trigonometric function of ω_j. The ordinary least squares estimator of the slope coefficient is the estimate of $-d$. Geweke and Porter-Hudak (1983) only consider the first m log periodograms (i.e., the lowest frequencies ω_j) in the estimation of the slope coefficient given that the proportion in (6) only holds for low frequencies. Hurvich, Deo, and Brodsky (1998) argue that the choice of the truncation point m implies a bias-variance tradeoff in the estimate of d. Geweke and Porter-Hudak (1983) establish the consistency and asymptotic normality of the estimator for $-0.5 < d < 0$, though Hurvich et al.

(1998) show that the GPH estimator is only consistent and is asymptotically normal for $-0.5 < d < 0.5$ under certain conditions.

2.7 Approximate maximum likelihood estimation (MLE)

Sowell (1992) derives the unconditional exact likelihood function for stationary fractionally integrated time series under the assumption of normality. This allows the simultaneous estimation of the fractional differencing, AR, and MA parameters of an ARFIMA(p, d, q) process by exact maximum likelihood estimation. In Monte Carlo experiments, Sowell (1992) shows that the root-mean-square error for the exact MLE of d is generally lower than the estimates of d from the GPH method and Whittle method but notes the significant computing time required for large samples. The approximate MLE method refers to Chung and Baillie's (1993) conditional sum-of-squares estimator. The authors estimate ARFIMA(p, d, q) processes with p and q equal to 1 or 2 by minimizing

$$\begin{aligned} S(\lambda) &= \frac{1}{2} log\sigma^2 + \frac{1}{2\sigma^2} \sum_{t=1}^{T} \varepsilon_t^2 \\ &= \frac{1}{2} log\sigma^2 + \frac{1}{2\sigma^2} \sum_{t=1}^{T} \left[\phi(L)\theta(L)^{-1}(1-L)^d(y_t - \mu)\right]^2 \ , \end{aligned} \tag{13}$$

which is shown to be asymptotically equivalent to exact MLE. However, compared to exact MLE, approximate MLE is computationally efficient given the simplicity of $S(\lambda)$ relative to the exact log-likelihood function derived by Sowell (1992). Chung and Baillie (1993) further argue that the approximate MLE method can be extended to include non-Gaussian conditional densities with ARCH effects, making the method attractive for modeling real-world processes. However, the method requires full specification of the model. Chung and Baillie (1993) note in their Monte Carlo study that the estimates of d and the ARMA parameters are biased when μ is unspecified and needs to simultaneously be estimated. Their simulation study also shows that the method is subject to significant small-sample bias. As noted in Section 2.5, Fox and Taqqu's (1986) Whittle method is approximate frequency-domain MLE. It is of interest to investigate whether the Whittle method estimator shares the same properties as the approximate time-domain MLE. For this reason, we choose to analyze the performance of Chung and Baillie's (1993) approximate time-domain MLE rather than Sowell's (1992) exact MLE.

2.8 Smoothed periodogram method

Reisen (1994) proposes an estimate of the fractional differencing parameter d in the ARIMA(p, d, q) model based on the smoothed periodogram using regression methods similar to the Geweke-Porter-Hudak technique discussed in Section 2.6. Instead of truncating the log periodogram abruptly at a particular point m, as is done by Geweke and Porter-Hudak (1983), Reisen (1994) gradually truncates the estimated spectral density by applying a smoothing function before estimating the slope coefficient $-d$. Using Monte Carlo simulation Hassler (1993) and Reisen (1994) demonstrate that the estimate of d based on the smoothed periodogram has smaller bias and variance relative to the estimate of d based on the GPH method. Heyde and Dai (1996) document that the smoothed periodogram estimator is highly robust to short-range dependence structures, though this robustness deteriorates somewhat slightly for processes with greater long memory.

3 Methodology

Following Granger and Joyeux (1980), an ARFIMA$(0, d, 0)$ process X_t can be defined as

$$X_t = (1 - L)^d \varepsilon_t \ , \tag{14}$$

where $t \geq 1$ and $\varepsilon_t \sim iid(0, \sigma_\varepsilon^2)$. The fractional differencing parameter d determines the long-range dependence. Denoting k as the lag order and $d \neq 0$, the differencing operator $(1 - L)^d$ is defined by the Binomial theorem series

$$\begin{aligned} (1 - L)^d &= \sum_{k=0}^{\infty} (-1)^k \binom{d}{k} L^k \\ &= \sum_{k=0}^{\infty} \pi_k L^k \ , \end{aligned} \tag{15}$$

where

$$\pi_k = \frac{\Gamma(-d + k)}{\Gamma(-d)\Gamma(K + 1)} \ , \tag{16}$$

and $\Gamma(\cdot)$ is the gamma function. Rewriting equation (14) as

$$\begin{aligned} X_t &= \left(\sum_{k=0}^{\infty} \pi_k L^k \right) \varepsilon_t \\ &= \varepsilon_t + \pi_1 \varepsilon_{t-1} + \pi_2 \varepsilon_{t-2} + \cdots \ , \end{aligned} \tag{17}$$

the ARFIMA process is re-expressed as an infinite moving-average. The process is stationary for $d < 0.5$, else the variance of equation (17) will be explosive. Hosking (1981) subsequently shows that the process is invertible, and has an infinite autoregressive representation if $d > -0.5$. The ARFIMA(p, d, q) process is thus covariance-stationary and invertible for $-0.5 < d < 0.5$. For our study, we set $d = 0$ or $d = 0.4$ to simulate our time series.

An ARFIMA$(0, d, 0)$ process of sample path length n can be generated by first simulating n innovations ε_t then recursively applying equation (17) to the series of simulated ε_t, given a truncation point $\tau \leq n$ which corresponds to the maximum lag of the series. Another widely used long memory process is fractional Gaussian noise. The long-range dependence is described by the Hurst coefficient. The corresponding fractional differencing parameter is $d = H - 1/2$. However, simulation of fractional Gaussian noise lacks the flexibility of including heavy-tailed and skewed distribution of the innovation. As such, we opt to simulate the time series using the ARFIMA specification. We refer readers interested in the fractional Gaussian noise simulation method to Beran (1994).

The distribution of the innovations ε_t determine the type of ARFIMA series generated. Three types of innovations are considered in this study, notably those from normal, generalized hyperbolic and stable distributions. The normal and generalized hyperbolic distributions generate innovations with finite variance, whereas the stable distribution generates innovations with infinite variance. Both generalized hyperbolic distribution and stable distribution exhibit heavy tails.

3.1 Generalized hyperbolic distribution

The generalized hyperbolic (GH) distribution was introduced by Barndorff-Nielsen (1977) to examine the distribution of the logarithm of particle sizes. If a random variable Y has a GH distribution, where

$$Y \sim GH(\lambda, \alpha, \beta, \delta, \mu) \ , \tag{18}$$

then the probability density function of Y is defined as

$$\begin{aligned} p_Y(y; \lambda, \alpha, \beta, \delta, \mu) &= a(\lambda, \alpha, \beta, \delta, \mu) \left(\delta^2 + (y-\mu)^2\right)^{(\lambda - \frac{1}{2})/2} \\ &\times K_{\lambda - \frac{1}{2}}(a\sqrt{\delta^2 + (y-\mu)^2}) exp(\beta(y-\mu)) \ , \end{aligned} \tag{19}$$

where

$$a(\lambda, \alpha, \beta, \delta, \mu) = \frac{(\alpha^2 - \beta^2)^{\lambda/2}}{\sqrt{2\pi}\alpha^{\lambda-1/2}\delta^{\lambda}K_{\lambda}(\delta\sqrt{\alpha^2 - \beta^2})} \tag{20}$$

and K is a modified Bessel function. The parameters δ and μ describe the scale and location of the distribution, while α and β determine the shape and skewness. The parameter λ characterizes the sub-classes of the GH distribution and determines the kurtosis. If $\lambda = 1$, the GH distribution reduces to the hyperbolic distribution. If $\lambda = -1/2$, the GH distribution becomes a normal inverse Gaussian distribution. Eberlein and Keller (1995) and Barndorff-Nielsen and Prause (2001) adopt the hyperbolic and normal inverse Gaussian distributions to model returns of financial assets, and conclude that they are a better fit than the normal distribution as the GH distribution displays both fat tails and skewness. Barndorff-Nielsen and Shephard (2012) document that the GH distribution is infinitely divisible, which is a necessary and sufficient condition to construct a Lévy process.[1] Mencia and Sentana (2004) discuss the flexibility of using multivariate GH distributions in allowing for tail-dependence parameterization within a financial econometrics context. In sum, the flexibility of the GH distribution in capturing the properties of financial markets motivates the use of GH innovations in our study. We set $\lambda = -1.4$, $\alpha = 0.32$, and $\beta = 0$ which correspond to the estimated parameters in Eberlein and Prause (1998) when the GH distribution is fitted to the returns of the S&P500 index.

3.2 Stable distribution

Similar to GH distribution, the stable distribution allows for skewness and heavy tails. With the exception of three sub-classes of stable distribution (Gaussian, Cauchy, and Lévy) the stable class of distributions lacks analytic expressions for the density and distribution functions. Levy (1924) characterizes this class of distribution and finds that its tail probabilities approximate those of the Pareto distribution. Rachev and Mittnik (2000) define a random variable X as having a stable distribution if it has the characteristic function (i.e., the inverse Fourier

[1] A Lévy process is the continuous-time version of the random walk. It has been widely adopted in option pricing theory, term structure and credit risk modeling due to its flexibility in modelling jumps. See, e.g., Eberlein and Raible (1999), Eberlein and Keller (1995), and Barndorff-Nielsen et al. (2001) for a detailed discussion.

transform of the probability density function)

$$E[\exp(i\theta X)] = \begin{cases} \exp\left\{-\delta^\alpha |\theta|^\alpha \left(1 - i\beta \text{sign}(\theta) \tan(\frac{\pi\alpha}{2}) + i\mu\theta\right)\right\} & \text{if } \alpha \neq 1 \\ \exp\left\{-\delta^\alpha |\theta|^\alpha \left(1 + i\beta \text{sign}(\theta) \log(\frac{\pi\alpha}{2}) + i\mu\theta\right)\right\} & \text{if } \alpha = 0 \end{cases}, \tag{21}$$

with

$$\text{sign}(\theta) = \begin{cases} 1 & \text{if } \theta > 0 \\ 0 & \text{if } \theta = 0 \\ -1 & \text{if } \theta < 0 \end{cases} \cdot \tag{22}$$

The stability parameter $0 < \alpha \leq 2$ determines the height of the tails of distribution, whereas $-1 \leq \beta \leq 1$ determines the skewness. The parameters δ and μ describe the scale and the location. A detailed discussion of the methods available to estimate the parameters of the stable distribution are provided by Nolan (1997).

Rachev and Mittnik (2000) argue that financial asset returns can be interpreted as the aggregated outcome of a mass of information arriving continuously in time. The use of the stable distribution in modeling asset returns is then supported by the generalized Central Limit Theorem, which states that the only possible non-trivial limit of normalized sums of independent identically distributed terms is stable. The success of Mandelbrot (1963) in modeling cotton prices with the Lévy stable distribution also motivates the inclusion of stable innovations in our simulation study. Similar to the GH distribution, the stable distribution is fairly flexible given the number of the parameters that characterize it. Additionally, the invariant property of α with respect to sampling intervals allows consistency in modeling financial time series with different frequencies. Despite these attractive properties, the second and higher moments are not defined for the Stable distribution with $\alpha < 2$. The extension to a multivariate stable distribution is also non-trivial, restricting empirical work to univariate cases. Accepting these drawbacks, we proceed with the stable innovations with $\alpha = 1.8$ and $\beta = 0$ which corresponds to Rachev and Mittnik's (2000) parameter estimates when the stable distribution is fitted to S&P500 index returns.

3.3 Short-range dependence

The fractionally integrated process X_t defined in equation (14) can only capture the hyperbolic rate of decay of the autocorrelation function. This, however, is restrictive in modeling short-run dynamics, motivating the need for the versatile ARFIMA(p, d, q) model which captures both the short memory and long memory of the time series. An ARFIMA(p, d, q) process Y_t is, in standard notation, defined as

$$\phi(L)Y_t = \theta(L)(1-L)^d \varepsilon_t \ , \tag{23}$$

where

$$\begin{aligned}
&\varepsilon_t \sim iid(0, \sigma_\varepsilon^2) \\
&\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p \\
&\theta(L) = 1 + \theta_1 L + \cdots + \theta_q L^q \ ,
\end{aligned} \tag{24}$$

and all p and q roots of $\phi(L)$ and $\theta(L)$, respectively, lie outside the unit circle. This gives an ARMA representation where the error term is fractionally integrated of order d, i.e.,

$$Y_t - \phi_1 Y_{t-1} - \cdots - \phi_p Y_{t-p} = (1-L)^d \varepsilon_t + \theta_1 (1-L)^d \varepsilon_{t-1} + \cdots + \theta_q (1-L)^d \varepsilon_{t-q} \ . \tag{25}$$

Hence, to simulate an ARFIMA(p, d, q) process, one can first simulate the series of fractionally integrated error terms using equation (17), then recursively apply equation (25) to the series $(1-L)^d \varepsilon_t$ to obtain the series Y_t. We simulate an ARFIMA$(1, d, 1)$ process with $\phi = \theta = 0.5$ to analyze the behavior and robustness of each estimator under the presence of short-term dynamics.

To summarize, our Monte Carlo study involves simulating the following processes and applying each of the eight estimation methods to the simulated time series.

(i) Gaussian white noise
(ii) Gaussian ARFIMA$(0, 0.4, 0)$
(iii) GH noise ($\lambda = -1.4$, $\alpha = 0.32$, $\beta = 0$)
(iv) GH ARFIMA$(0, 0.4, 0)$ ($\lambda = -1.4$, $\alpha = 0.32$, $\beta = 0$)
(v) Stable noise ($\alpha = 1.8$, $\beta = 0$)
(vi) Stable ARFIMA$(0, 0.4, 0)$ ($\alpha = 1.8$, $\beta = 0$)
(vii) Gaussian ARFIMA$(1, 0, 1)$ ($\phi = \theta = 0.5$)
(viii) Gaussian ARFIMA$(1, 0.4, 1)$ ($\phi = \theta = 0.5$)

(ix) GH ARFIMA$(1, 0, 1)$ ($\lambda = -1.4$, $\alpha = 0.32$, $\beta = 0$, $\phi = \theta = 0.5$)
(x) GH ARFIMA$(1, 0.4, 1)$ ($\lambda = -1.4$, $\alpha = 0.32$, $\beta = 0$, $\phi = \theta = 0.5$)
(xi) Stable ARFIMA$(1, 0, 1)$ ($\alpha = 1.8$, $\beta = 0$, $\phi = \theta = 0.5$)
(xii) Stable ARFIMA$(1, 0.4, 1)$ ($\alpha = 1.8$, $\beta = 0$, $\phi = \theta = 0.5$)

The simulated series (i) to (vi) have no short-range dependence. Series (vii) to (xii) have short-range dependence.

4 Simulation results

For each type of the series, we run 1000 simulations with sample length $n = 250,\ 500,\ 1000,\ 2000$. The estimates d or H for each simulation are then computed using the eight different techniques described in Section 2. To assess the performance of each estimator, the average bias, variance and mean-squared-errors (MSE) of estimates are computed. Let d_0 equal the true fractional differencing parameter and d_k be the estimate of d for simulation k. Then

$$\hat{d} = \frac{1}{1000}\sum_{k=1}^{1000} d_k \tag{26}$$

$$bias = \hat{d} - d_0 \tag{27}$$

$$\hat{\sigma}^2 = \frac{1}{1000-1}\sum_{k=1}^{1000}(d_k - \hat{d})^2 \tag{28}$$

$$MSE = \hat{\sigma}^2 + bias\ . \tag{29}$$

When H is estimated, we compute the statistics on the basis of $H_0 = d_0 + 1/\alpha$, where α is the stability parameter of the stable distribution. For the Gaussian and GH distributions, $\alpha = 2$.

The computations are carried out in R, where existing packages in generating GH and stable innovations are freely available. To reduce computation time, parallel computing in R is adopted. This allows each replication to be generated by different processors concurrently. As a result, each sample is stored in different vectors, allowing d for each sample to be estimated simultaneously using the parallelization technique.

The results are presented in Figures 1-4 for methods estimating H and in Figures 5-8 for methods estimating d. In each figure, the boxplots and MSE stacked bars for each series is provided. The boxplots shows the deviations from d_0, where the band in the box represents the median; the bottom and top of the box correspond to the first and the third quartiles; the upper (lower) whisker extends from the box to the highest (lowest) value that is within the 1.5 inter-quartile range (IQR) of the third (first) quartile and; outliers that fall beyond the whiskers are plotted as points. The stacked bars show the MSE of each estimator, where the yellow (top) portion is the average bias and the grey (bottom) portion is the variance.

4.1 R/S analysis

Figure 1 shows the performance of R/S analysis. In general, the estimates are highly biased. The estimates of d are biased upward for uncorrelated series, and downward for fractionally integrated series. However, the narrow boxplots show that the estimates do not vary much. The MSE plots confirm this as the bias proportion of the estimates dominates the bar graphs. Additionally, the performance does not improve with increase in sample sizes, suggesting that the estimates are not consistent. When d is estimated for heavy-tailed series, while not suffering a loss of efficiency, the estimates remain biased and inconsistent. The results are worst when short-range dynamics are included. Furthermore the estimation method spuriously identifies the existence of long-range dependence when there is only short-run dependence. The high MSE confirms this when ARFIMA$(1, 0, 1)$ is used. Collectively, the simulation results suggests that R/S analysis is extremely inaccurate in capturing the long-range dependence and is not suitable for analysis of financial time series.

4.2 Detrended fluctuation analysis

The results for estimates obtained using detrended fluctuation analysis are provided in Figure 2. In general the estimates are precise but inaccurate. The variance proportion of MSE tend to be dominated by the bias proportion when n is large for most of the series. The method is upwardly biased for both Gaussian and GH ARFIMA$(0, d, 0)$ series. Increasing the sample path length does not improve the accuracy of the estimates suggesting that DFA is not consistent for the two finite variance series. DFA, however, appears to perform well for the stable ARFIMA$(0, d, 0)$ series since the narrow boxplots suggest that the deviations from d_0 cluster around zero. The estimates for the stable series are prone to more outliers in comparison to the Gaussian and GH series, showing a difference between the performance of DFA estimates for finite and infinite variance series. Consequently, while the MSE of the estimates is lower for the stable innovations series, it is advised that a large sample size is required when using DFA to account for large outliers. When estimating d for series with short-run dependence, the performance deteriorates. The estimates for all series have an upward bias, and beyond the d stationary range of 0.5 when both short and long memories exist. We may then use it as an indicator of short-run dynamics contamination if DFA returns an estimate of d that is beyond 0.5.

4.3 Higuchi's fractal dimension method

The Higuchi method is one of the four methods which give the long memory parameter that is fairly accurate under most conditions. Figure 3 summarizes the results. The boxplots show that deviations from the true H_0 are concentrated near the zero line for all twelve series. However, as the Higuchi algorithm restricts the values of H to the interval (0,1), the deviations from H_0 is negatively skewed for $H_0 = 0.9$. Additionally, the bias proportion of MSE remains relatively low compared to the variance proportion. The precision of all estimates improves with increasing sample size. Both the decreasing length of boxplots and decreasing MSE confirm this. Interestingly, when the long memory of the GH ARFIMA$(0, 0, 0)$ and GH ARFIMA$(1, 0, 1)$ series are estimated, the estimates becomes more upwardly biased while for stable ARFIMA$(0, 0.4, 0)$ and stable ARFIMA$(1, 0.4, 1)$, the estimates becomes more downwardly biased. This suggests that the estimator behaves inconsistently under different underlying distributions. Overall, unlike the previous estimators R/S and DFA, the estimates from the performance of the Higuchi method remains largely unaffected by the inclusion of short-run dynamics.

4.4 Modified periodogram method

Figure 4 shows the performance of the estimators obtained by the modified periodogram method. When there is no short-range dependence, the estimates are fairly accurate with low variability. They are however biased downward for stable series. It is then likely that the method estimates $H = d + 1/2$ instead of $H = d + 1/\alpha$. Consistently, increasing sample sizes lead to smaller MSE that is driven by a decreasing variance proportion of the MSE. However, similar to DFA, when short-run dynamics are introduced, the estimates are inaccurate and exceed the stationary boundary values for series with both short and long memory. The method is unable to differentiate the true long-range dependence parameter when the series is auto-correlated in the short-run. As a result, the MSE are large with a large bias proportion. Despite this poor aspect of performance, the variability remains low, suggesting that the method has high precision irrespective of heavy-tailed distributions and short-range dependence.

4.5 Whittle method

The Whittle estimator with the assumed model of ARFIMA$(0, d, 0)$ estimates the parameter d of the series without short-range dependence accurately with high precision. Figure 5 shows that the deviations cluster around zero, with only few outliers. The MSE for GH and Stable ARFIMA$(0, d, 0)$ are similar to the estimates for the Gaussian series, suggesting that the presence of non-Gaussian innovations does not affect the efficiency of estimators adversely.

When ARFIMA$(1, d, 1)$ series are used, the Whittle estimator with the assumed model of ARFIMA$(0, d, 0)$ returns values that are beyond the boundary value of d. We respecify the estimator by assuming the ARFIMA$(1, d, 1)$ model. With small sample sizes, the estimator is biased downward for the six series with short-range dependence. Additionally, the MSE are large with high variance and bias proportion. When sample size increases, the bias proportion diminishes, suggesting that the estimator is consistent. While the variability decreases, outliers to the left remain. A larger number of samples are then required to account for such variability. The estimator is again robust to non-Gaussian series, where the MSE for the series with GH and stable innovations are similar.

We note that the variability for the two specifications of Whittle estimators differs. Even with $n = 2000$, the MSE for the ARFIMA$(0, d, 0)$ specification remain smaller than the MSE of the ARFIMA$(1, d, 1)$ specification. An intuitive explanation is that the latter specification requires estimations of more coefficients (ϕ, d, θ), whereas the former specification only needs to estimate the fractional differencing parameter. A problem arises if we assume the specification of ARFIMA(1, d, 1) to estimate the series without any short-range dependence. In unreported results, the estimator provides non-zero estimates for ϕ and θ which are almost equal with non-significant t-statistics, inducing the common factor problem. While the d estimate remains unbiased, the precision drops. In practice however, it is recommended to over-specify the order of the model rather than to under-specify it to avoid large biases.

4.6 Geweke-Porter-Hudak method

Figure 6 provides the performance of the GPH estimator. The estimates are fairly unbiased with small sample as the deviations from the true fractional differencing paramater d_0 cluster around zero and the MSE are dominated by the variance proportion. The variability however is large, with occurrences of large outliers that are beyond the boundary value. Both the accuracy and precision improve larger samples are used. The narrower boxplots and smaller MSE confirm this. The biases and variability remain approximately the same for Gaussian and non-Gaussian series, suggesting that the introduction of heavy-tailed distributions does not affect the efficiency of estimates. Additionally, the performance is generally unaffected with the inclusion of short-range dependence. Overall, the GPH estimator is unbiased, but inefficient with small sample, requiring a large sample size to account for the potential variability.

4.7 Approximate maximum likelihood estimation

As with the Whittle estimator, we first assume the ARFIMA$(0, d, 0)$ specification for the process. Figure 7 summarizes the results. The method is highly accurate in estimating the parameter d when the series are without any short-term autocorrelations. Even with short sample length, the estimates generated are unbiased with low variance. The boxplots representing the middle 50% of the estimates are narrow with only a few outliers. The precision, however, is higher in estimating series with $d_0 = 0$. Interestingly, the estimates remain both efficient and consistent for the GH and stable time series, despite violating the normality assumption.

When the parameters are estimated for the ARFIMA$(1, d, 1)$ series, approximate MLE returns boundary values of d of 0.5 for each series, suggesting the need to respecify the assumed model. The results presented in Figure 7 are based on ARFIMA$(1, d, 1)$ specification when the estimating the parameters for the series with short-range dependence. The plots suggest that the method is biased downward when $d_0 = 0.4$ for small sample sizes. When the series is not fractionally integrated, approximate MLE obtains estimates of the d accurately with high precision. However, the outliers are larger if compared to the series without short-range dependence. When sample size increases, the bias proportion of the MSE decreases, generating consistent estimates. In unreported results, the variability of estimates increase if we over-specify the MLE as an ARFIMA$(2, d, 2)$ but the estimates remain unbiased. Consistently, the introduction of heavy-tailed distribution does not affect the efficiency of estimates. This suggests that the MLE is robust under non-Gaussian series.

4.8 Smoothed periodogram method

The results for the smoothed periodogram method are shown in Figure 8. Generally, the estimates are fairly unbiased, even when the series length is short. The performance is largely similar to that of the GPH estimator, with large variations and few outliers that are beyond the boundary values when $n = 250$. The efficiency of estimates does not suffer for the GH and stable distributions, suggesting that the smoothed periodogram estimator is robust under heavy-tailed distributions. Additionally, the method is able to differentiate long-range dependence from short-run dynamics accurately. The MSE remains approximately similar when short-range dependence is introduced. As the series length increases, the variability decreases, improving the precision of estimates. Overall, the smoothed periodogram method performs relatively well for situations where there are heavy-tails and additional short-range dependence.

4.9 Discussion

Table 1 summarizes the root-mean-squared-errors (RMSE) of all the estimators for each six series with no short-memory for series length $n = 2000$. Table 2 summarizes the root-mean-squared-errors (RMSE) of all the estimators for each six series with short-memory for series length $n = 2000$. The figures in bold represent the three lowest RMSE for each series. The approximate MLE generates the lowest RMSE compared to other estimators. Additionally, as suggested by Figure 7, the estimator is unbiased, with negligible bias proportion compared to the variance proportion. The estimates do not suffer a loss of efficiency when non-Gaussian series is adopted, including series with infinite variance. The next best method is the Whittle estimator which generates unbiased estimates that are robust to heavy-tailed series, but with higher variability compared to approximate MLE. These methods, however, require specification of the short-run dynamics to remain consistent. From a statistical standpoint, to generate reliable estimates it is recommended to first over-specify the order of short-run dynamics p and q before working towards a more parsimonious model by reducing the lag order of the AR and MA coefficients. The sensitivity of estimates based on the different specifications is beyond the scope of our paper, though we feel this area deserves attention in future research.

While R/S analysis generates estimates with low MSE (due to low variability) for the series with both short-range and long-range dependence, they are biased and remain inconsistent as sample sizes grow. The estimates are at best used as a crude measurement of long-range dependence. Instead, estimates from the GPH and smoothed periodogram methods are accurate, despite the low precision. As such, it is recommended to have sufficiently large sample sizes when these methods are used, to account for the potential large variability.

4.10 Density plots

Having discussed the results of all estimators, we provide a brief discussion on the shape of the distribution of the estimates. We focus only on the density of the four unbiased estimators, notably the Whittle estimator, GPH, approximate MLE and smoothed periodogram methods. The densitiy plots complement the boxplots in revealing multimodality of our estimates.

Figures 9 and 10 compare the density plots of these estimates for ARFIMA(0, 0.4, 0) and ARFIMA$(1, 0.4, 1)$ series. Consistent with the boxplots, all the distributions are centered around the true d. The GPH and smoothed periodogram estimates are more dispersed compared to approximate MLE and Whittle estimates of d. Additionally, the estimates behave similarly for non-Gaussian series. Notably, none of the distributions of estimates however have more than one peak. The general results are consistent with our previous discussions in supporting the estimates from the Whittle method and approximate MLE as the more accurate and precise measures of long-range dependence due to the lower variability. Despite this, large occurrences that are beyond the boundary value remain, especially when short-range dependence is included. It is then best, in practice, to use several different methods to estimate long-range dependence since this can provide a better perspective on the structure of the time series.

4.11 Strong persistence in short-range dependence

As a robustness check, we test the performance of four unbiased estimators when series that are highly persistence in the short-run are used. *A posteriori*, we specify the model for Whittle estimator and approximate MLE as an ARFIMA$(1, d, 1)$. Only series with GH innovations are considered. The following three series with $n = 2000$ are simulated
(i) GH ARFIMA$(1, 0, 1)$ ($\lambda = -1.4$, $\alpha = 0.32$, $\beta = 0$, $\phi = \theta = 0.9$)

(ii) GH ARFIMA$(1, 0.4, 1)$ ($\lambda = -1.4$, $\alpha = 0.32$, $\beta = 0$, $\phi = \theta = 0.9$)

(iii) GH ARFIMA$(1, 0, 1)$ ($\lambda = -1.4$, $\alpha = 0.32$, $\beta = 0$, $\phi = \theta = -0.9$) .

The first series corresponds to a time series with highly persistent short-range dependence but without any long-range dependence. This allows us to test whether the estimators spuriously identify of long memory when the dynamics are only driven by large AR and MA coefficients. The second series is similar but with inclusion of long-range dependence. The third simulated series correspond to a series with long-range dependence but with inclusion of highly antipersistent short memory. The negative autocorrelations may then dampen the effects of long memory, affecting the estimates.

Figure 11 summarizes the results. When the first series GH ARFIMA$(1, 0, 1)$ is used, the estimates from approximate MLE are largely unaffected by the large AR and MA coefficients. The Whittle estimator remains unbiased but produces large outliers that are beyond the boundary value.[2]Despite having a narrow boxplots, the numbers of large outliers greatly increase the MSE of the estimates. In contrast, both GPH and smoothed periodogram methods are biased upwards, suggesting that the strong short-term persistence affects the accuracy of the estimates.

When long-range dependence by setting $d_0 = 0.4$ in the second simulated series, the second column in Figure 11 shows that the variability of Whittle estimator decreases, but the outlier problem remains. Similar to the first series, both GPH and smoothed periodogram overestimate d_0 when ϕ and θ assume large values. The MLE however is now biased downward, and is no longer accurate. When the series is strongly negatively autocorrelated in the short-run, the Whittle estimator performs the best whereas the remaining estimators underestimate the true parameter. This suggests that strong negative autocorrelations offset the long-range dependence, thus decreasing the efficiency of estimates. Overall, strong short-run dynamics indeed affect both accuracy and precision of the estimates of long-range dependence. This further strengthens the need to adopt different methods to estimate d.

5 Conclusion

We study the robustness of several long-range dependence estimators when finite and infinite variance series are considered, as well as series with short-run de-

[2]The outliers are not included in the boxplots as the values are too large, affecting the range of vertical axis.

pendence. Two main conclusions are drawn from our simulations. First, many of the estimators are relatively robust with respect to deviations from Gaussian series. The performance of each estimator is not very different when series with underlying fat-tailed distribution is adopted instead of Gaussian. When Stable series is used, estimators of the Hurst parameter are more likely to generate $H = d + 1/2$ rather than $H = d + 1/\alpha$. Compared to estimators of d, they are fairly inaccurate, even when the simple Gaussian white noise is the process being estimated. The second conclusion is that most estimators are affected by existence of short-run dynamics with the exception of the GPH and smoothed periodogram methods. Both Whittle estimator and approximate MLE require specification for the short-run dynamics to generate values within the stationary and invertible range of d. We suggest in practice to first over-specify these methods and then lower the order of p and q when the AR and MA coefficients are insignificant or share a common root to work down to a more parsimonious model. When the series is strongly autocorrelated in the short-run, the Whittle estimator and MLE with correct specification suffer a loss of efficiency. Similar behavior of the estimates is exhibited by both the GPH and smoothed periodogram methods which are generally unbiased for series with mild short-range dependence. Overall, it is best to adopt several methods when estimating the long-range dependence of a series, instead of solely relying on one estimation method. The performance of estimators and improved identification steps not used in this study is a subject for further research.

References

Andersen, T. G. and T. Bollerslev (1997): "Intraday periodicity and volatility persistence in financial markets," *Journal of Empirical Finance*, 4, 115–158.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2003): "Modeling and forecasting realized volatility," *Econometrica*, 71, 579–625.

Andersen, T. W. (1971): *The Statistical Analysis of Time Series*, Wiley, New York.

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, 681–700.

Baillie, R. T., T. Bollerslev, and H. O. Mikkelsen (1996a): "Fractionally integrated generalized autoregressive conditional heterskedasticity," *Journal of Econometrics*, 74, 3–30.

Baillie, R. T., C.-F. Chung, and M. A. Tieslau (1996b): "Analysing inflation by the fractionally integrated ARFIMA-GARCH model," *Journal of Applied Econometrics*, 11, 23–40.

Barndorff-Nielsen, O. (1977): "Exponentially decreasing distributions for the logarithm of particle size," *Proceedings of the Royal Society*, 353, 401–419.

Barndorff-Nielsen, O., B. Christiansen, and E. Nicolato (2001): "General stochastic volatility and explicit option prices in Heath-Jarrow-Morton term structure analysis," Aarhus University.

Barndorff-Nielsen, O. and K. Prause (2001): "Apparent scaling," *Finance and Stochastics*, 5, 103–113.

Barndorff-Nielsen, O. and N. Shephard (2012): *Basics of Levy Processes*, Unpublished.

Beran, J. (1994): *Statistics for Long-memory Processes*, Chapman & Hall, New York.

Bloomfield, P. (1976): *Fourier Analysis of Time Series: An Introduction*, Wiley, New York.

Bollerslev, T. (1986): "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

Cheung, Y.-W. and F. X. Diebold (1994): "On maximum likelihood estimation of the differencing parameter of fractionally-integrated noise with unknown mean," *Journal of Econometrics*, 62, 301–316.

Chiriac, R. and V. Voev (2011): "Modelling and forecasting multivariate realized volatility," *Journal of Applied Econometrics*, 26, 922–947.

Chung, C.-F. and R. T. Baillie (1993): "Small sample bias in conditional sum-of-squares estimators of fractionally integrated ARMA models," *Empirical Economics*, 18, 791–806.

Conrad, J. and G. Kaul (1988): "Long-term market overreaction or biases in computed returns?" *Journal of Finance*, 48, 39–63.

Dacorogna, M. M., U. A. Muller, R. J. Nagler, R. B. Olsen, and O. V. Pictet (1993): "A geographical model for the daily and weekly seasonal volatility in the foreign exchange market," *Journal of International Money and Finance*, 12, 413–438.

Davidson, J. (2004): "Moment and memory properties of linear conditional heteroscedasticity models, and a new model," *Journal of Business & Economic Statistics*, 22, 16–29.

Davies, R. B. and D. S. Harte (1987): "Tests for Hurst effect," *Biometrika*, 74, 95–101.

Diebold, F. X. and G. D. Rudebusch (1989): "Long memory and persistence in aggregate output," *Journal of Monetary Economics*, 24, 189–209.

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, 83–106.

Eberlein, E. and U. Keller (1995): "Hyperbolic distributions in finance," *Bernoulli*, 1, 281–299.

Eberlein, E. and K. Prause (1998): "The generalized hyperbolic model: Financial derivatives and risk measures," Freiburg University.

Eberlein, E. and S. Raible (1999): "Term structure models driven by general Levy processes," *Mathematical Finance*, 9, 31–53.

Engle, R. F. (1982): "Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation," *Econometrica*, 50, 987–1008.

Engle, R. F. and J. R. Russell (1998): "Autoregressive conditional duration: A new model for irregularly spaced transaction data," *Econometrica*, 66, 1127–1162.

Fama, E. F. (1965): "The behavior of stock-market prices," *Journal of Business*, 38, 34–105.

Fama, E. F. and K. F. French (1988): "Permanent and temporary components of stock prices," *Journal of Political Economy*, 96, 246–273.

Fisher, L. (1966): "Some new stock-market indexes," *Journal of Business*, 39, 191–225.

Fox, R. and M. S. Taqqu (1986): "Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series," *The Annals of Statistics*, 14, 517–532.

Fuller, W. A. (1976): *Introduction to Statistical Time Series*, Wiley, New York.

Geweke, J. and S. Porter-Hudak (1983): "The estimation and application of long memory time series models," *Journal of Time Series Analysis*, 4, 221–238.

Granger, C. W. J. and R. Joyeux (1980): "An introduction to long-memory time series models and fractional differencing," *Journal of Time Series Analysis*, 1, 15–29.

Hassler, U. (1993): "Regression of spectral estimators with fractionally integrated time series," *Journal of Time Series Analysis*, 14, 369–380.

Heyde, C. C. and W. Dai (1996): "On the robustness to small trends of estimation based on the smoothed periodogram," *Journal of Time Series Analysis*, 17, 141–150.

Higuchi, T. (1988): "Approach to an irregular time series on the basis of the fractal theory," *Physica D*, 31, 277–283.

Hosking, J. R. M. (1981): "Fractional differencing," *Biometrika*, 68, 165–176.

Hurst, H. (1951): "Long term storage capacity of reservoirs," *Transactions of the American Society of Civil Engineers*, 116, 770–779.

Hurvich, C. M., R. Deo, and J. Brodsky (1998): "The mean squared error of Geweke and Porter-Hudak's estimator of the memory parameter of a long-memory time series," *Journal of Time Series Analysis*, 19, 19–46.

Jegadeesh, N. (1990): "Evidence of predictable behavior in security returns," *Journal of Finance*, 45, 881–898.

Levy, P. (1924): "Theorie des erreurs. La loi de Gauss et les Lois exceptionelles," *Bulletin de la Societe Mathematique de France*, 52, 49–85.

Lillo, F. and J. D. Farmer (2004): "The long memory of the efficient market," *Studies in Nonlinear Dynamics & Econometrics*, 8, 1–33.

Lo, A. W. (1991): "Long-term memory in stock market prices," *Econometrica*, 59, 1279–1313.

Lo, A. W. and A. C. MacKinlay (1988): "Stock market prices do not follow random walks: Evidence from a simple specification test," *Review of Financial Studies*, 1, 41–66.

Lobato, I. N. and C. Velasco (2000): "Long memory in stock-market trading volume," *Journal of Business & Economic Statistics*, 18, 410–427.

Mandelbrot, B. B. (1963): "The variation of certain speculative prices," *Journal of Business*, 36, 394–419.

Mandelbrot, B. B. and J. W. Van Ness (1968): "Fractional brownian motion, fractional noises and applications," *SIAM Review*, 10, 422–437.

Mandelbrot, B. B. and J. Wallis (1969): "Computer experiments with fractional gaussian noises," *Water Resources Research*, 5, 228–267.

McLeod, A. I. and K. W. Hipel (1978): "Preservation of the rescaled adjusted range," *Water Resources Research*, 14, 491–508.

Mencia, F. and E. Sentana (2004): "Estimating and testing of dynamic models with generalised hyperbolic innovations," CEMFI, Madrid.

Nolan, J. (1997): "Numerical calculations of stable densities and distribution functions," *Communications in Statistics & Stochastic Models*, 13, 759–774.

Peng, C.-K., S. Havlin, H. E. Stanley, and A. L. Goldberger (1995): "Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series," *Chaos*, 5, 82–87.

Porterba, J. M. and L. H. Summers (1988): "Mean reversion in stock prices: Evidence and implications," *Journal of Financial Economics*, 22, 27–59.

Rachev, S. and S. Mittnik (2000): *Stable Paretian Models in Finance*, Wiley, New York.

Reisen, V. A. (1994): "Estimation of the fractional difference parameter in the ARIMA(p,d,q) model using the smoothed periodogram," *Journal of Time Series Analysis*, 15, 335–350.

Shiller, R. J. (1984): "Stock prices and social dynamics," *Brookings Papers on Economic Activity*, 2, 421–436.

Sowell, F. (1992): "Maximum likelihood estimation of stationary univariate fractionally integrated time series," *Journal of Econometrics*, 53, 165–188.

Summers, L. H. (1986): "Does the stock market rationally reflect fundamental values?" *Journal of Finance*, 41, 591–601.

Taqqu, M. S., V. Teverovsky, and W. Willinger (1995): "Estimators for long-range dependence: An empirical study," *Fractals*, 3, 785–798.

Taylor, S. J. (1986): *Modeling financial time series*, Wiley, Chichester.

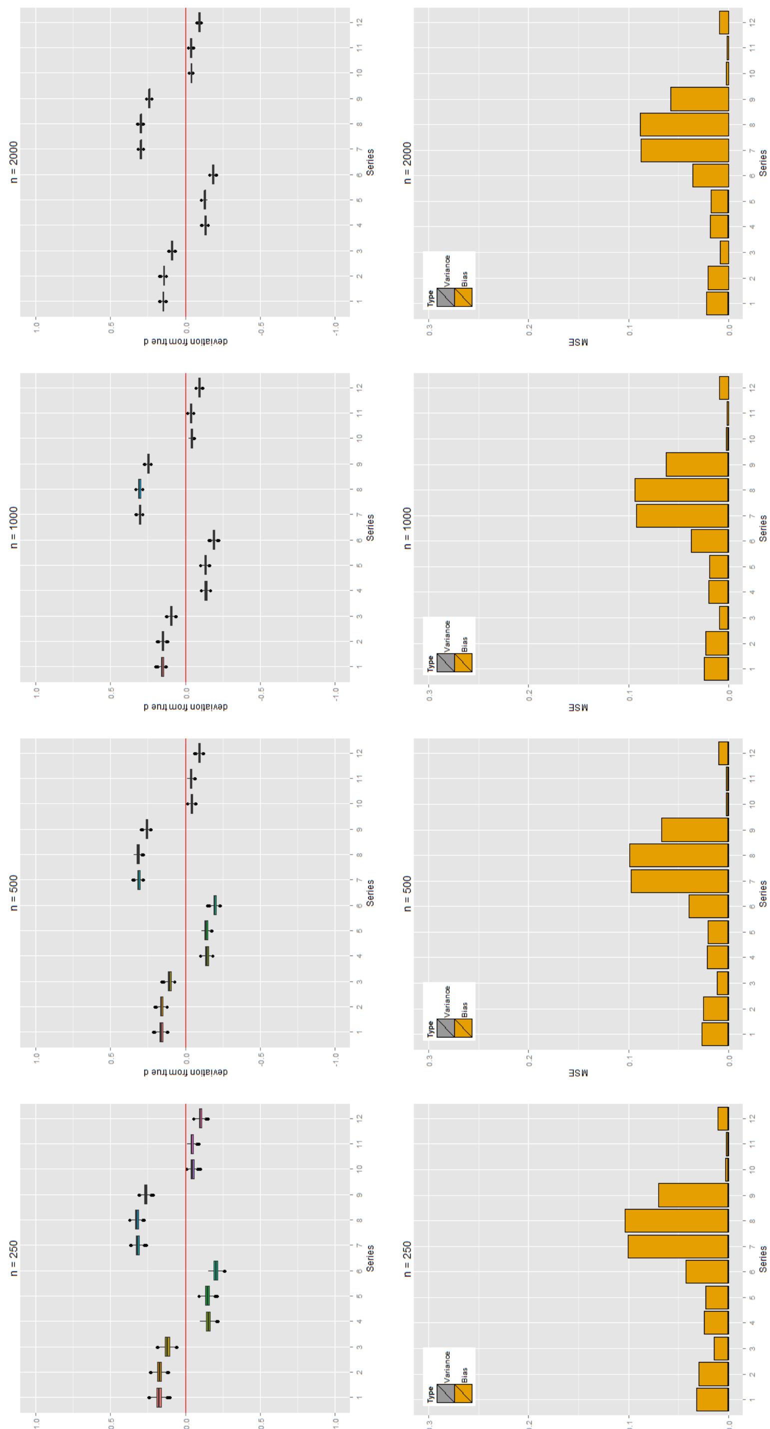

Figure 1: The boxplots and MSE stacked bars for each of the twelve simulated series for R/S Analysis. The numbers on the horizontal axis correspond to the series (i) through (xii), respectively.

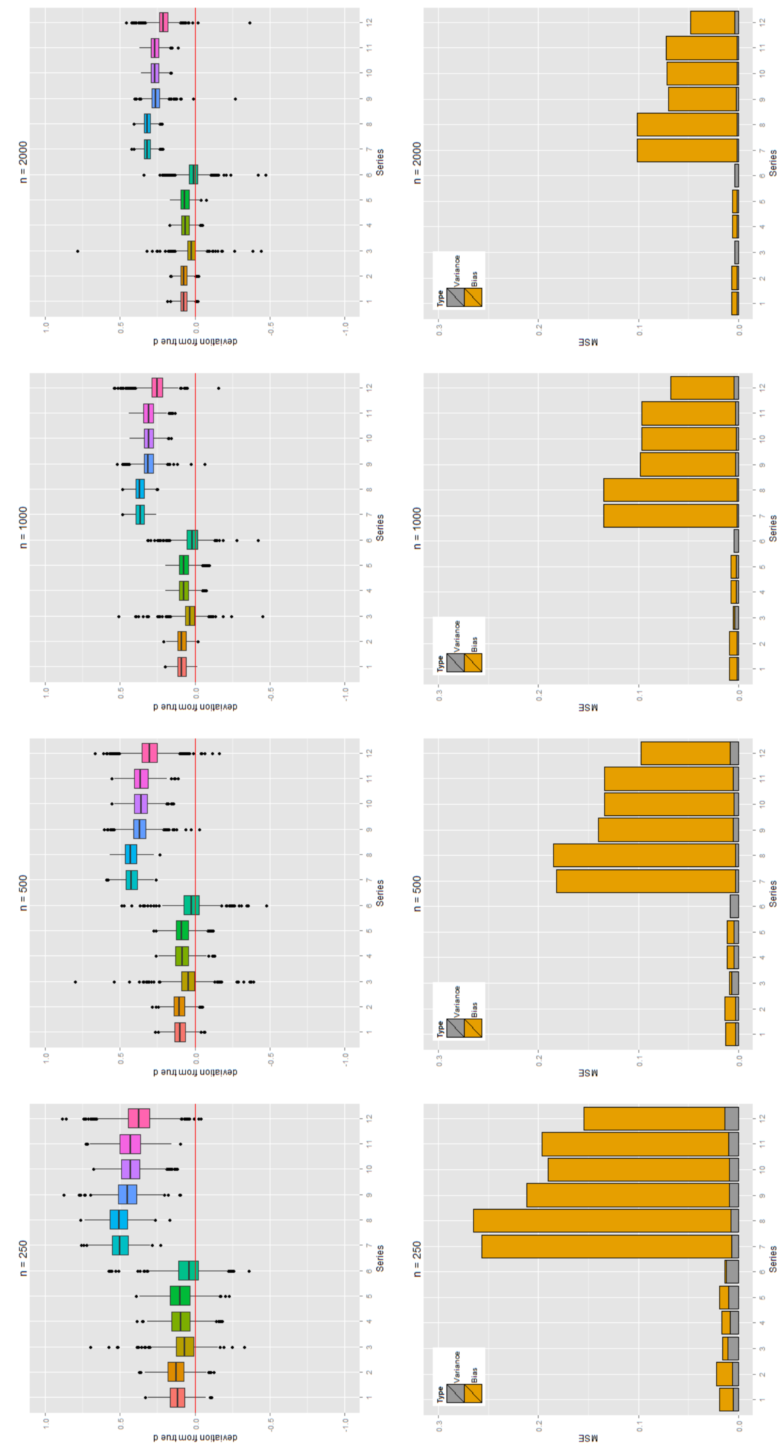

Figure 2: The boxplots and MSE stacked bars for each of the twelve simulated series for DFA. The numbers on the horizontal axis correspond to the series (i) through (xii), respectively.

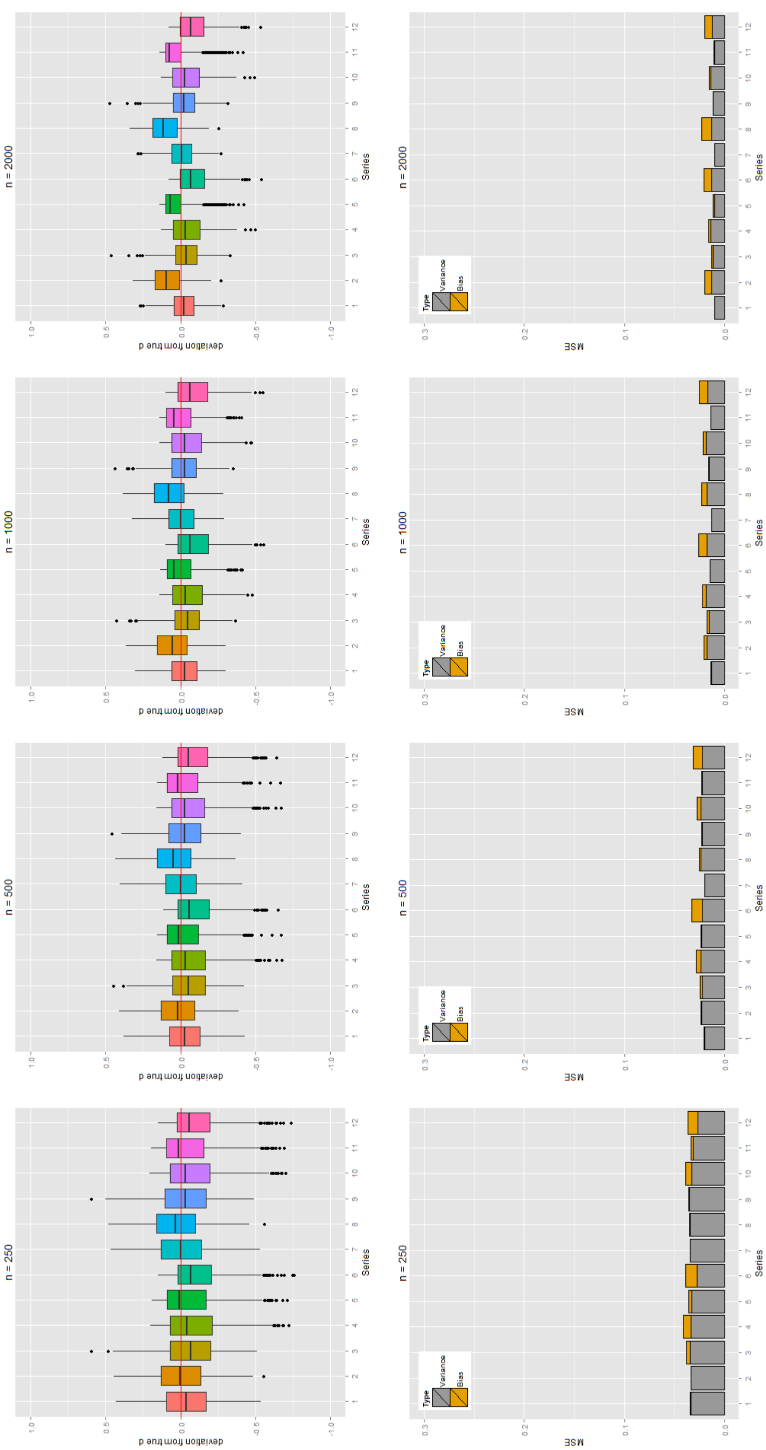

Figure 3: The boxplots and MSE stacked bars for each of the twelve simulated series for the Higuchi method. The numbers on the horizontal axis correspond to the series (i) through (xii), respectively.

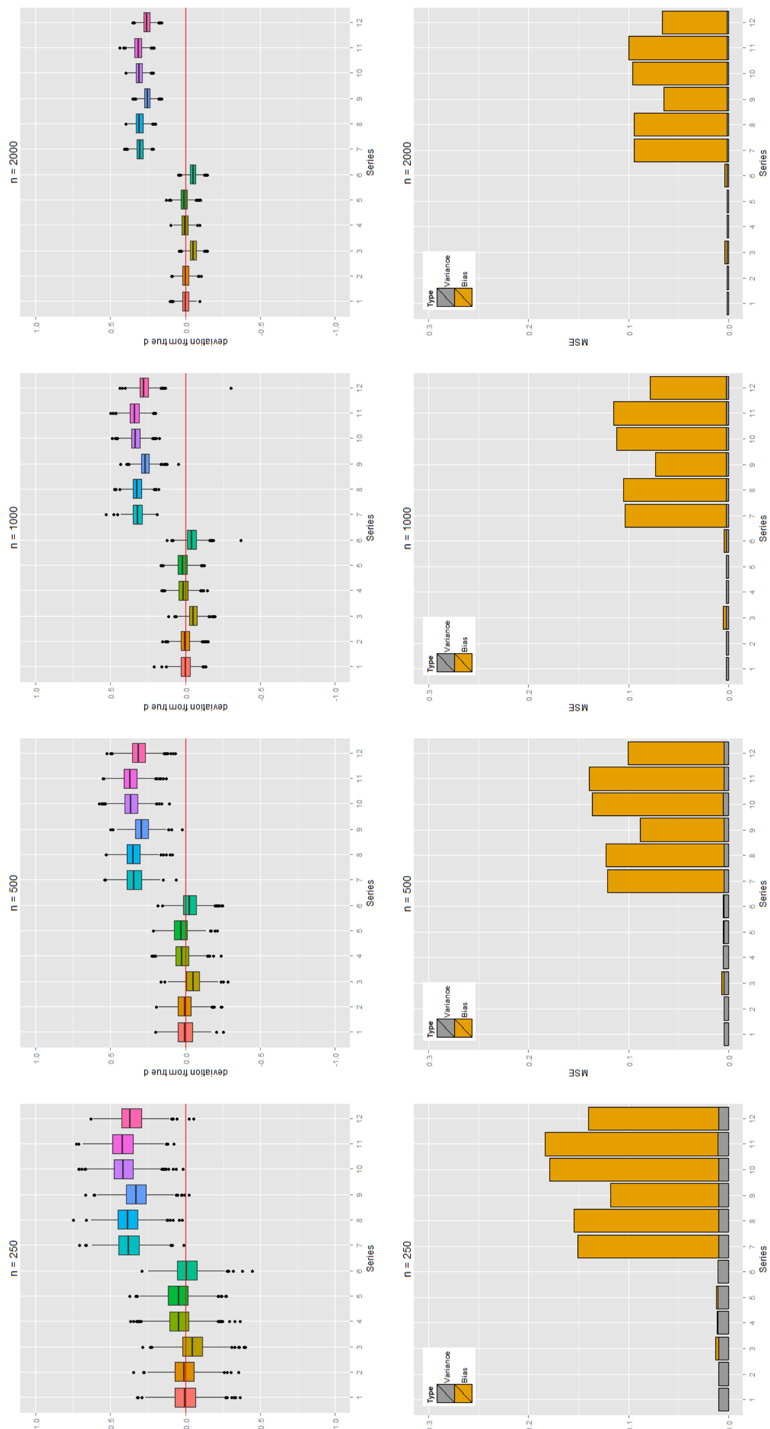

Figure 4: The boxplots and MSE stacked bars for each of the twelve simulated series for the modified periodogram method. The numbers on the horizontal axis correspond to the series (i) through (xii), respectively.

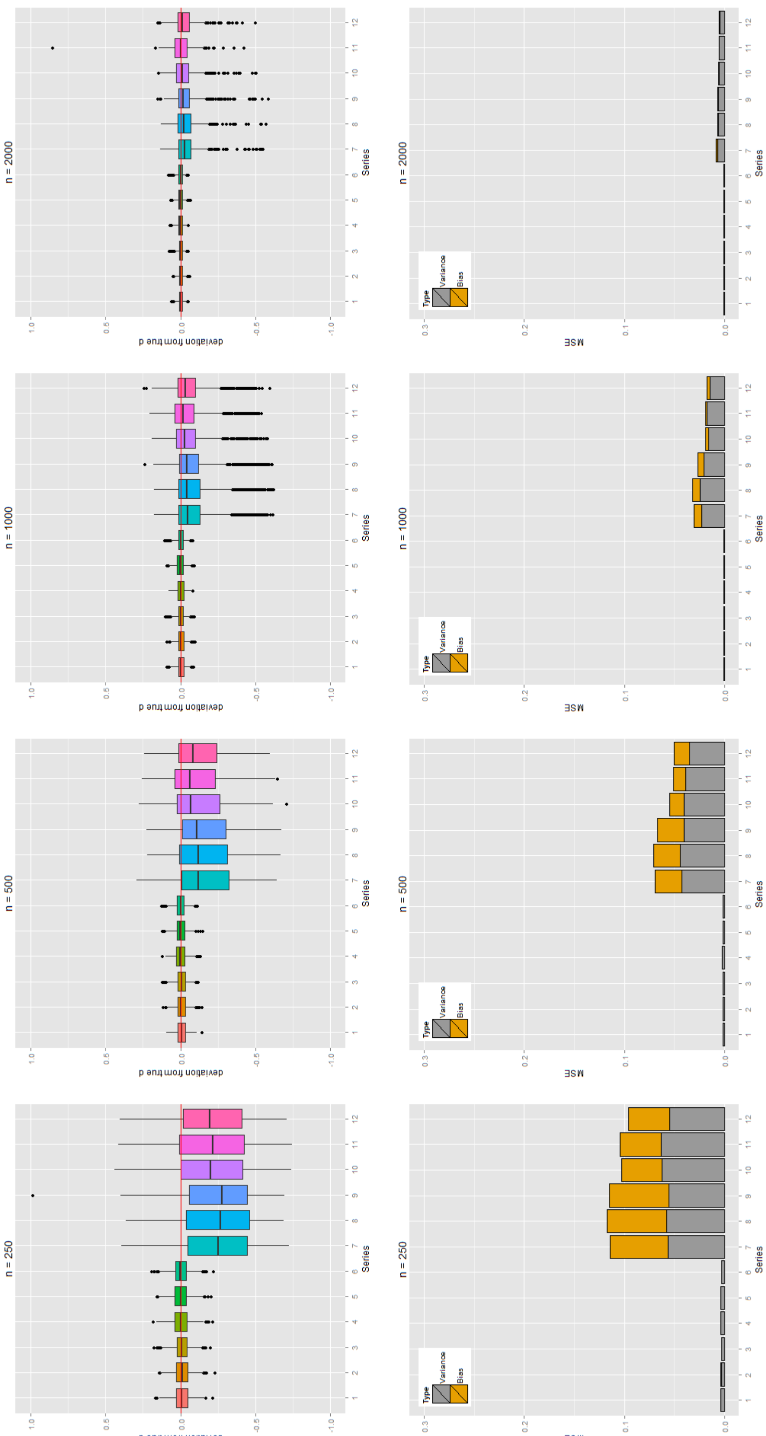

Figure 5: The boxplots and MSE stacked bars for each of the twelve simulated series for the Whittle method. The numbers on the horizontal axis correspond to the series (i) through (xii), respectively.

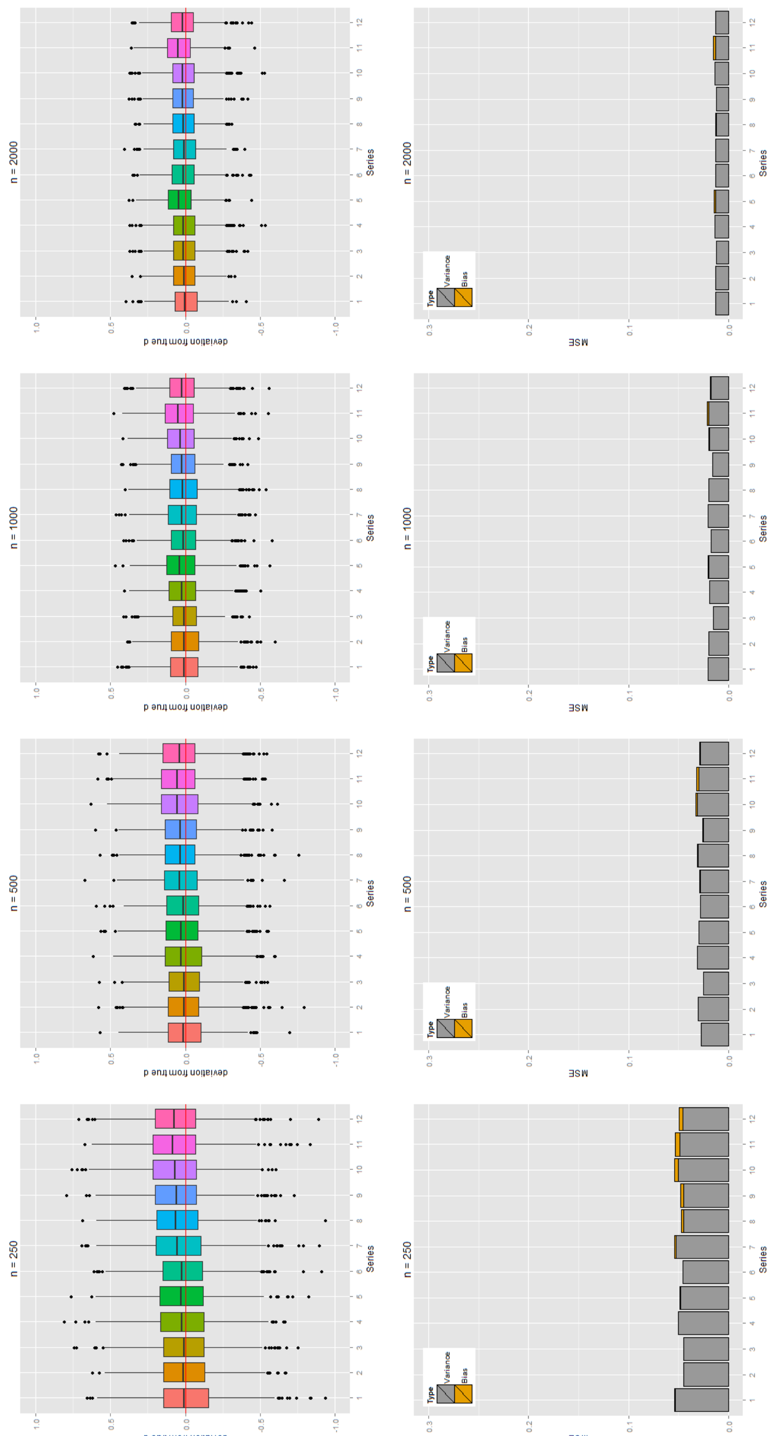

Figure 6: The boxplots and MSE stacked bars for each of the twelve simulated series for the Geweke-Porter-Hudak method. The numbers on the horizontal axis correspond to the series (i) through (xii), respectively.

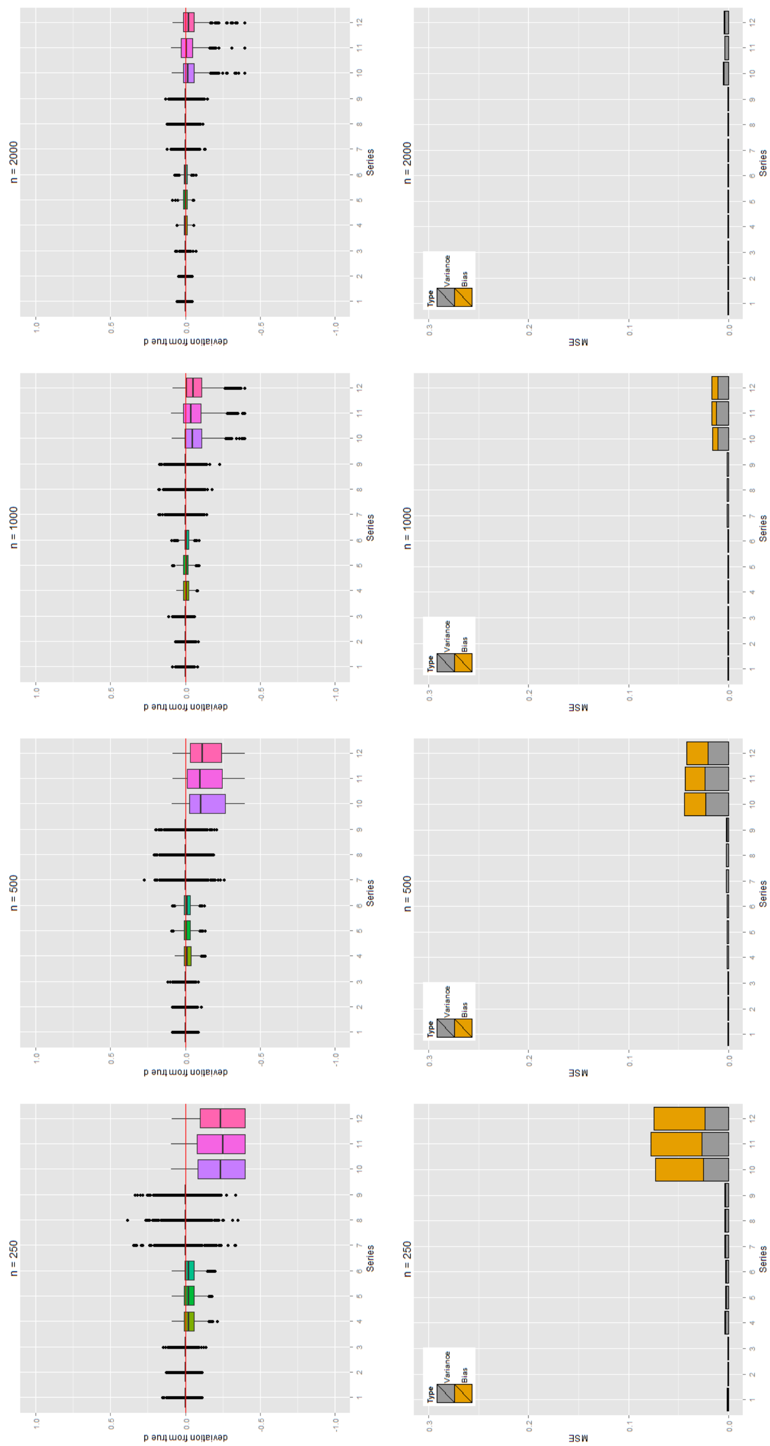

Figure 7: The boxplots and MSE stacked bars for each of the twelve simulated series for approximate MLE. The numbers on the horizontal axis correspond to the series (i) through (xii), respectively.

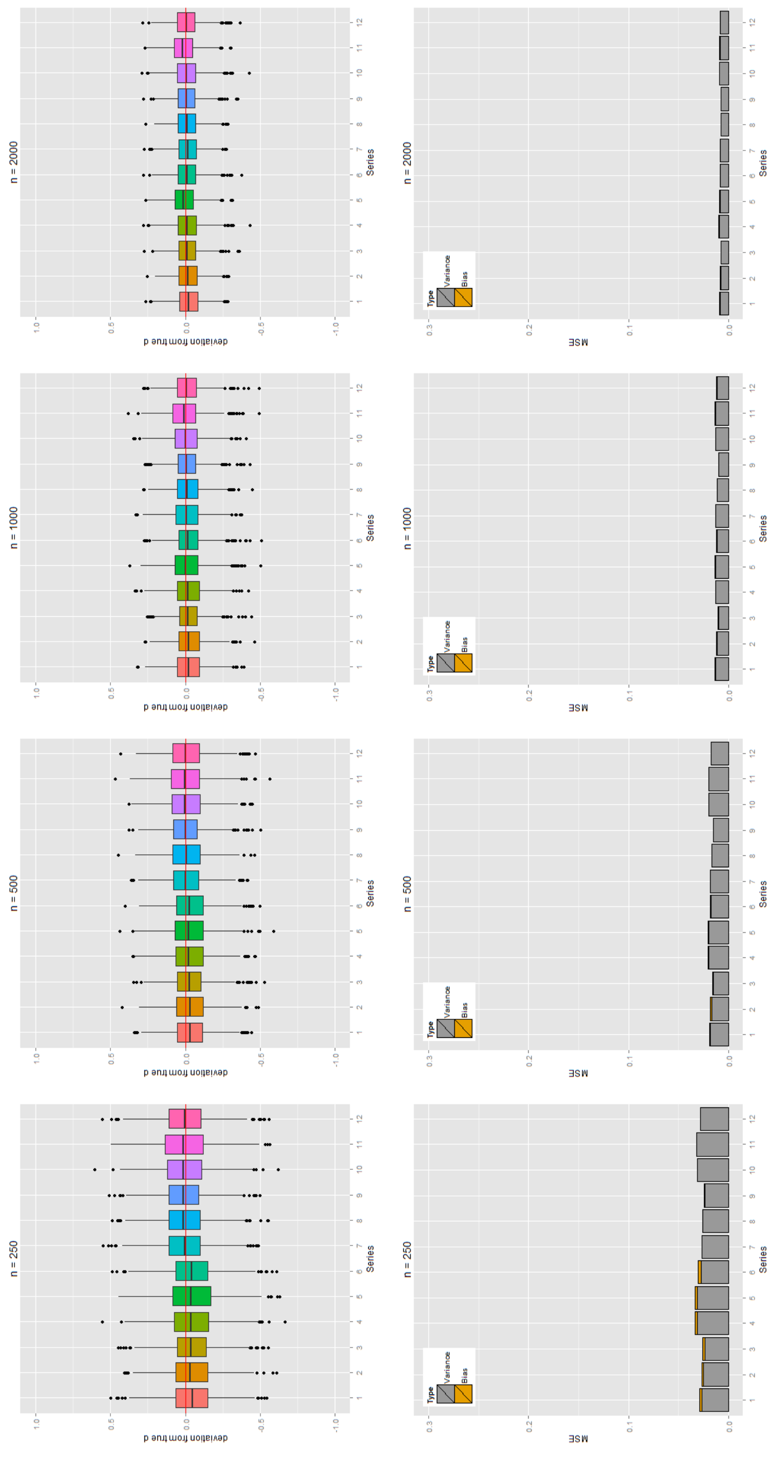

Figure 8: The boxplots and MSE stacked bars for each of the twelve simulated series for the smoothed periodogram method. The numbers on the horizontal axis correspond to the series (i) through (xii), respectively.

Table 1: Root-mean-squared errors for the eight estimators for the series with no short-range dependence. Only the RMSE for n=2000 are shown.

	Series (n=2000)					
	1	**2**	**3**	**4**	**5**	**6**
			R/S Analysis			
Variance	4.77E-05	4.26E-05	4.14E-05	4.96E-05	4.47E-05	4.15E-05
Bias	0.02165	0.020192	0.007724	0.018111	0.017405	0.03513
RMSE	0.147301	0.142247	0.088121	0.134763	0.132096	0.187542
			DFA			
Variance	0.000876	0.000884	0.003289	0.00135	0.001432	0.003125
Bias	0.005661	0.005593	0.000462	0.004121	0.004194	6.40E-05
RMSE	0.080848	0.080482	**0.0612**	0.073969	0.075011	**0.0565**
			Higuchi method			
Variance	0.009366	0.012413	0.010911	0.01326	0.009934	0.012344
Bias	0.000524	0.007277	0.001412	0.002466	0.000846	0.007826
RMSE	0.099448	0.140321	0.111008	0.125403	0.10383	0.142019
			Modified Periodogram method			
Variance	0.000913	0.000955	0.000834	0.000952	0.001056	0.000858
Bias	3.01E-06	1.42E-06	0.002986	3.80E-06	7.38E-05	0.00269
RMSE	**0.0303**	**0.0309**	0.061807	**0.0309**	**0.0336**	0.059567
			Whittle estimator			
Variance	0.000303	0.000345	0.000285	0.00035	0.000373	0.000296
Bias	6.88E-06	9.71E-06	7.09E-06	1.68E-09	7.13E-07	5.47E-07
RMSE	**0.0176**	**0.0188**	**0.0171**	**0.0187**	**0.0193**	**0.0172**
			GPH method			
Variance	0.012543	0.012313	0.011939	0.01357	0.012745	0.01262
Bias	3.42E-06	9.05E-06	6.07E-05	4.70E-05	0.001375	0.00013
RMSE	0.11201	0.111006	0.109541	0.116692	0.118829	0.112917
			Approximate MLE			
Variance	7.77E-05	7.78E-05	7.77E-05	0.000302	0.000347	0.000268
Bias	2.11E-07	9.61E-08	2.40E-07	9.56E-06	1.20E-08	1.26E-05
RMSE	**0.0088**	**0.0088**	**0.0088**	**0.0176**	**0.0186**	**0.0167**
			Smoothed Periodogram method			
Variance	0.008064	0.007329	0.007349	0.009028	0.008431	0.008102
Bias	0.000452	0.000416	0.000255	0.000276	6.85E-05	0.000186
RMSE	0.092284	0.088005	0.087197	0.096459	0.092191	0.091037

Table 2: Root-mean-squared errors for the eight estimators for the series with short-range dependence. Only the RMSE for n=2000 are shown.

	Series (n=2000)					
	7	8	9	10	11	12
			R/S Analysis			
Variance	3.07E-05	3.00E-05	2.79E-05	2.17E-05	2.31E-05	2.17E-05
Bias	0.087312	0.087907	0.057713	0.001723	0.00147	0.008861
RMSE	0.295538	0.296542	0.240294	**0.0418**	**0.0386**	**0.0942**
			DFA			
Variance	0.000913	0.000919	0.00182	0.001397	0.001485	0.00315
Bias	0.100532	0.100366	0.068337	0.070379	0.070419	0.044346
RMSE	0.318503	0.318254	0.264871	0.26791	0.26815	0.217936
			Higuchi method			
Variance	0.009411	0.012397	0.010856	0.01306	0.009733	0.012146
Bias	3.01E-05	0.01049	0.000445	0.002182	0.000945	0.007408
RMSE	0.097163	0.151285	**0.1063**	0.123461	0.103335	0.139837
			Modified Periodogram method			
Variance	0.000912	0.000953	0.000833	0.000949	0.00105	0.000864
Bias	0.093102	0.093491	0.063718	0.095191	0.098968	0.064879
RMSE	0.306617	0.307317	0.25407	0.310065	0.316256	0.256404
			Whittle estimator			
Variance	0.006788	0.005642	0.005909	0.005087	0.004928	0.004457
Bias	0.001206	0.000989	0.007333	0.000443	1.81E-05	0.005955
RMSE	**0.0894**	**0.0814**	0.115075	**0.0744**	**0.0703**	0.102041
			GPH method			
Variance	0.012703	0.012253	0.011894	0.013599	0.012756	0.012629
Bias	1.69E-05	7.88E-05	0.000187	0.000167	0.001846	0.000285
RMSE	0.112783	0.111047	0.109911	0.117329	0.120839	0.113642
			Approximate MLE			
Variance	0.000544	0.000517	0.000513	0.003941	0.003306	0.003723
Bias	3.19E-07	3.14E-12	4.48E-07	0.00083	0.000232	0.000863
RMSE	**0.0233**	**0.0227**	**0.0227**	**0.0691**	**0.0595**	**0.0677**
			Smoothed Periodogram method			
Variance	0.008076	0.007342	0.007347	0.009031	0.00843	0.008114
Bias	0.000235	0.000214	0.000102	0.000117	0.000201	6.18E-05
RMSE	**0.0912**	**0.0869**	**0.0863**	0.095647	0.092902	**0.0904**

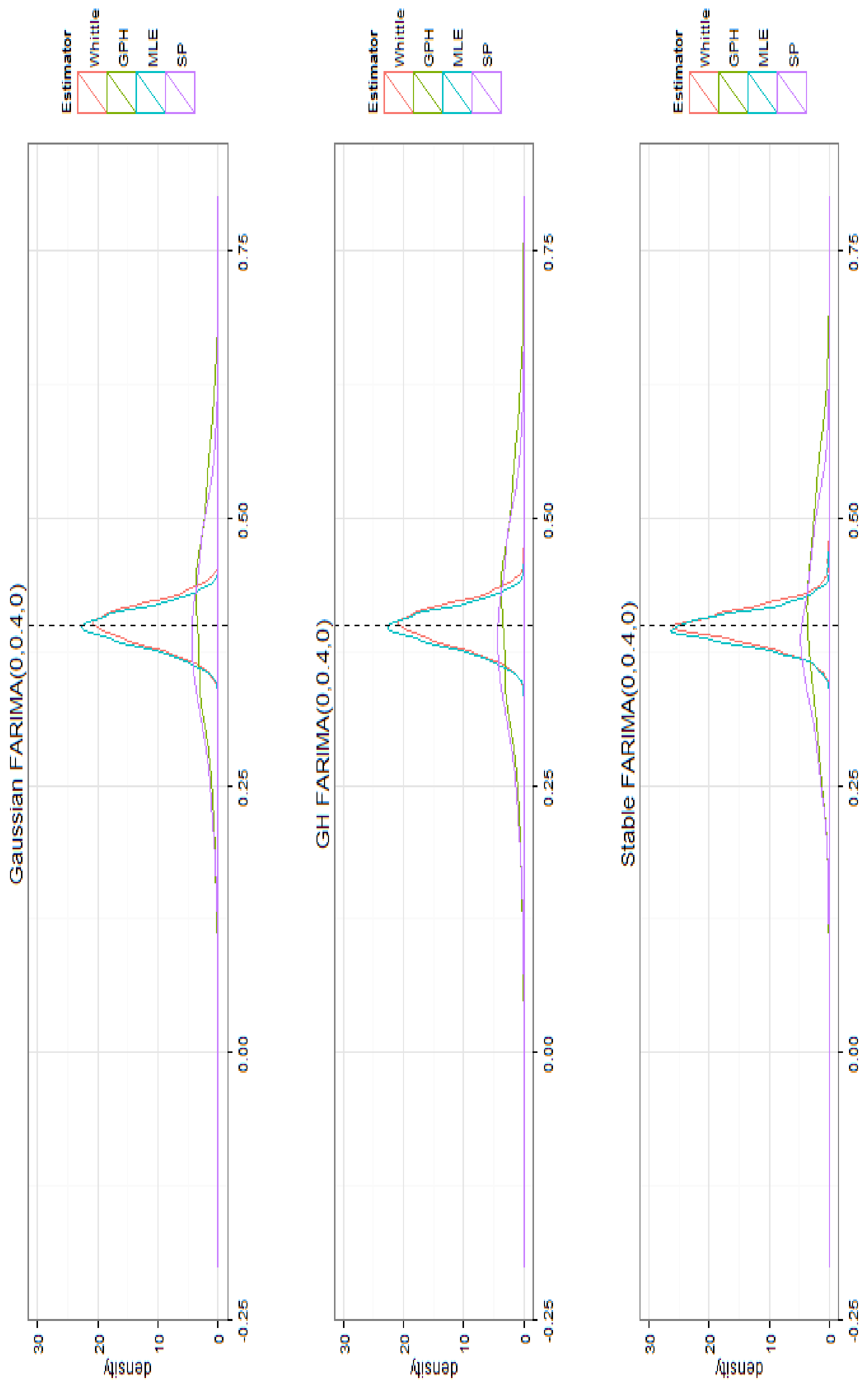

Figure 9: The density plots for the Whittle, GPH, approximate ML, and smoothed periodogram estimators for $d = 0.4$ and no short-range dependence in the simulated series. The dotted line is $d = 0.4$

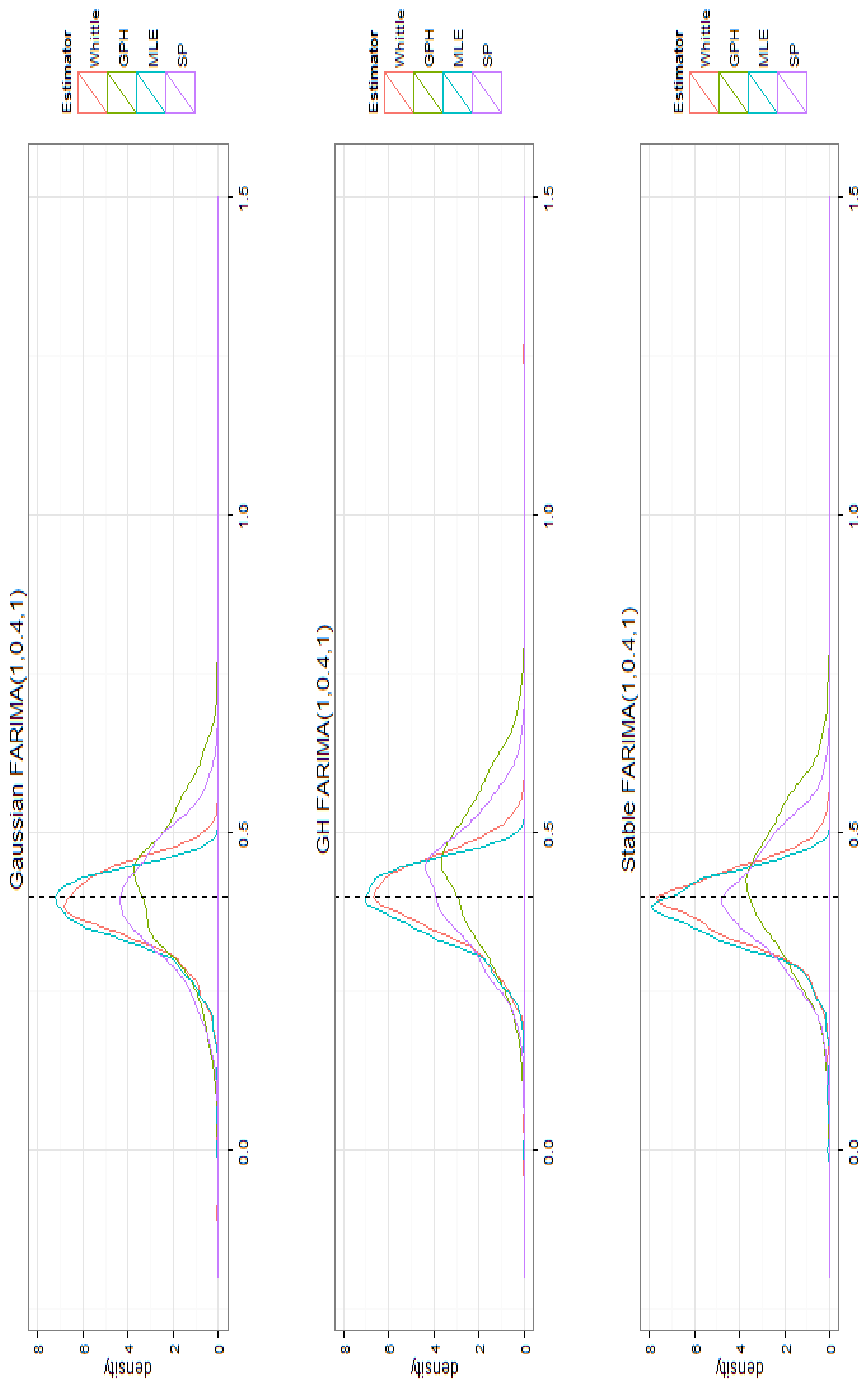

Figure 10: The density plots for the Whittle, GPH, approximate ML, and smoothed periodogram estimators for $d = 0.4$ and short-range dependence in the simulated series. The dotted line is $d = 0.4$

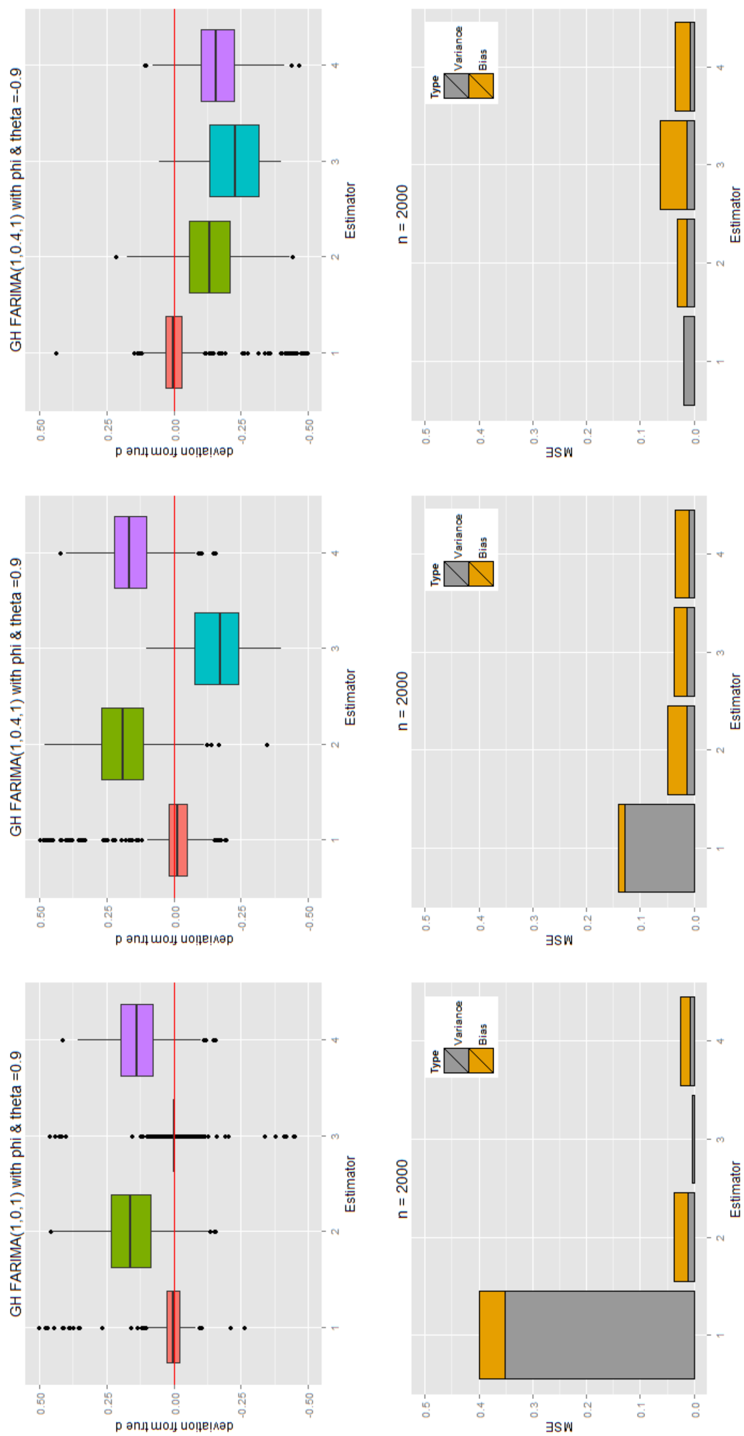

Figure 11: The boxplots and MSE stacked bars for the Whittle method, GPH method, approximate MLE, and smoothed periodogram method, where the numbers on the horizontal axis correspond to the estimators, respectively.