12-2020

Asymptotic properties of least squares estimator in local to unity processes with fractional Gaussian noises

Xiaohu WANG

Weilin XIAO

Jun YU
Singapore Management University, yujun@smu.edu.sg

SMU ECONOMICS & STATISTICS

Asymptotic Properties of Least Squares Estimator in Local to Unity Processes with Fractional Gaussian Noises

Xiaohu Wang, Weilin Xiao, Jun Yu

December 2020

Paper No. 27-2020

Asymptotic Properties of Least Squares Estimator in Local to Unity Processes with Fractional Gaussian Noises[1]

Xiaohu Wang
Fudan University

Weilin Xiao
Zhejiang University

Jun Yu
Singapore Management University

December 23, 2020

[1]Xiaohu Wang, School of Economics, Fudan University, Shanghai, China, Email: wang_xh@fudan.edu.cn. Weilin Xiao, School of Management, Zhejiang University, Hangzhou, 310058, China. Email: wlxiao@zju.edu.cn. Jun Yu, School of Economics and Lee Kong Chian School of Business, Singapore Management University, 90 Stamford Rd, Singapore. Email for Jun Yu: yujun@smu.edu.sg. URL: http://www.mysmu.edu/faculty/yujun/.

Abstract

This paper derives asymptotic properties of the least squares estimator of the autoregressive parameter in local to unity processes with errors being fractional Gaussian noises with the Hurst parameter H. It is shown that the estimator is consistent when $H \in (0,1)$. Moreover, the rate of convergence is n when $H \in [0.5,1)$. The rate of convergence is n^{2H} when $H \in (0,0.5)$. Furthermore, the limit distribution of the centered least squares estimator depends on H. When $H = 0.5$, the limit distribution is the same as that obtained in Phillips (1987a) for the local to unity model with errors for which the standard functional central theorem is applicable. When $H > 0.5$ or when $H < 0.5$, the limit distributions are new to the literature. Simulation studies are performed to check the reliability of the asymptotic approximation for different values of sample size.

JEL classification: C22

Keywords: Least squares, Local to unity, Fractional Brownian motion, Fractional Ornstein-Uhlenbeck process.

1 Introduction

In this paper, we consider the following model:

$$X_t = \rho_n X_{t-1} + \varepsilon_t, \rho_n = \exp(-c/n), t = 1, ..., n, \tag{1}$$

where $\varepsilon_t = \sigma u_t$, u_t is a fractional Gaussian noise (FGN) with mean zero, variance one, and covariance function being

$$\gamma_u(k) := \mathbb{E}(u_t u_s) = \frac{1}{2}\left[(k+1)^{2H} + (k-1)^{2H} - 2k^{2H}\right] \text{ with } k = |t-s|, \tag{2}$$

and $H \in (0, 1)$ is called the Hurst parameter. When $H = 0.5$, $\gamma_u(k) = 0$ for any $k \neq 0$. Given that u_t is normally distributed, $\{u_t\}$ form a sequence of independent and identically distributed (i.i.d.) variables with the standard normal distribution $N(0, 1)$. Whereas, when $H \neq 0.5$, $\gamma_u(k) \neq 0$ for any k and

$$\gamma_u(k) \sim H(2H-1)k^{2H-2}, \text{ for large } k. \tag{3}$$

That is, $\gamma_u(k)$ decays at a hyperbolic rate as k goes to infinity. When $H > 0.5$, $\gamma_u(k) > 0$ and $\sum_{k=-\infty}^{\infty} \gamma_u(k) = \infty$, giving rise to the terminology of 'long-range-dependent' errors. When $H < 0.5$, it has $\gamma_u(k) < 0$ for $k \neq 0$ and $\sum_{k=-\infty}^{\infty} \gamma_u(k) = 0$, giving rise to the terminology of 'anti-persistent' errors. An FGN is obtained as the increments of the fractional Brownian motion (fBm) $B^H(t)$ that is a zero-mean Gaussian process with the covariance function

$$Cov\left(B^H(t), B^H(s)\right) = \frac{1}{2}\left(|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right) \quad \forall t, s \geq 0. \tag{4}$$

That is, $u_t = B^H(t) - B^H(t-1)$.

Model (1) is related to the local to unity model of Phillips (1987a) and Chan and Wei (1988) by replacing the noises where the classical central limit theorem is applicable with fractional Gaussian noises. Model (1) is also related to the fractional unit root model of Sowell (1990) by replacing the AR coefficient of unity with the AR coefficient of local to unity. Although we replace the $I(d)$ noises of Sowell (1990) with the FGN, the results in this paper also apply to $I(d)$ errors as it will become clear later. Model (1) is also related to the model of Park (2003) where $\rho_n = 1 - m/n$ if we assume m is fixed in his model.

Let $\widehat{\rho}_n$ denote the least squares (LS) estimator of ρ_n that takes the form of

$$\widehat{\rho}_n = \sum_{t=1}^{n} X_{t-1}X_t / \sum_{t=1}^{n} X_{t-1}^2.$$

Hence, the centered least squares estimator is

$$\widehat{\rho}_n - \rho_n = \sum_{t=1}^{n} X_{t-1}\varepsilon_t / \sum_{t=1}^{n} X_{t-1}^2. \tag{5}$$

The goal of this paper is to derive the asymptotic properties of $\widehat{\rho}_n$ and $\widehat{\rho}_n - \rho_n$ under $n \to \infty$. As it is well expected for local to unity model, the initial value of X_t significantly affects the finite sample distribution of $\widehat{\rho}_n - \rho_n$. To capture the impact of the initial value on asymptotics, we set the initial value of X_t to be $X_0 = O_p\left(n^H\right)$ and

$$n^{-H}\frac{X_0}{\sigma} \xrightarrow{p} J_c(0),$$

where $J_c(0)$ is a constant (such as zero) or $O_p(1)$.

The rest of the paper is organized as follows. Section 2 reviews the results in the literature. The asymptotic properties of the normalized $\widehat{\rho}_n - \rho_n$ are developed in Section 3. Section 4 obtains the finite sample properties of the normalized $\widehat{\rho}_n - \rho_n$. Section 5 concludes. The Appendix collects proofs of the main results.

Throughout the paper, we use $\xrightarrow{p}$, $\xrightarrow{d}$, $\Rightarrow$, $\sim$ to denote convergence in probability, convergence in distribution, convergence in functional space, and equivalent in distribution. We use $[nr]$ to denote the integral part of nr.

2 A Literature Review

Phillips (1987a) considers the following local to unit root model

$$X_t = \rho_n X_{t-1} + v_t, \rho_n = \exp(-c/n), X_0 = O_p(1), \tag{6}$$

where $\{v_t\}$ is a strong mixing sequence with mixing coefficients α_m satisfying $\sum_{m=1}^{\infty}\alpha_m^{1-2/\beta} < \infty$ and $\sup_t |v_t|^{\beta+\delta} < \infty$ for some $\beta > 2$ and $\delta > 0$. There are two important features in Model (6). First, since $\rho_n = 1 - c/n + O(n^{-2})$, the autoregressive coefficient depends on n and converges to unity as $n \to \infty$. Second, the functional central limit theorem is applied to $\{v_t\}$. An interesting special case of Model (6) is when $\{v_t\}$ are i.i.d. with $E|v_t|^{\beta} < \infty$ for some $\beta > 2$. In this case, according to Phillips (1987a), as $n \to \infty$,

$$n\left(\widehat{\rho}_n - \rho_n\right) \xrightarrow{d} \frac{\int_0^1 J_c(r)dW(r)}{\int_0^1 J_c(r)^2 dr} = \frac{\left\{J_c(1)^2 + 2c\int_0^1 J_c(r)^2 dr - 1\right\}/2}{\int_0^1 J_c(r)^2 dr}. \tag{7}$$

where $J_c(r)$ denotes an Ornstein-Uhlenbeck (OU) process defined by the stochastic differential equation

$$dJ_c(r) = -cJ_c(r)dr + dW(r), J_c(0) = 0, \tag{8}$$

with $W(r)$ being a standard Brownian motion.

Sowell (1990) considers the following unit root model with $\rho = 1$:

$$X_t = \rho X_{t-1} + \sigma v_t,\ v_t = (1-L)^{-d}\epsilon_t,\ \epsilon_t \overset{i.i.d.}{\sim} (0,1),\ X_0 = O_p(1), \tag{9}$$

where L is the lag operator with $(1-L)^{-d}$ defined as

$$(1-L)^{-d} = \sum_{j=0}^{\infty} \frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)} L^j \text{ for } d \in (-0.5, 0.5).$$

In this model, the error term v_t is assumed to follow a fractional integrated process of order d, or an $I(d)$ process. With $\widehat{\rho}$ being the LS estimator of ρ, Sowell (1990) and Marinucci and Robinson (1999) show that, as $n \to \infty$,

$$n(\widehat{\rho} - 1) \xrightarrow{d} \frac{\int_0^1 W(r)dW(r)}{\int_0^1 W(r)^2 dr}, \text{ if } d = 0, \tag{10}$$

$$n(\widehat{\rho} - 1) \xrightarrow{d} \frac{\frac{1}{2}B^H(1)^2}{\int_0^1 B^H(r)^2 dr}, \text{ if } d > 0, \tag{11}$$

$$n^{2H}(\widehat{\rho} - 1) \xrightarrow{d} -\frac{H\frac{\Gamma(0.5+H)}{\Gamma(1.5-H)}}{\int_0^1 B^H(r)^2 dr}, \text{ if } d < 0, \tag{12}$$

where $H = d + 0.5$.[1]

Setting $c = 0$ in (7) or setting $d = 0$ in (10) can lead to the well-known result for the unit root model obtained in Phillips (1987b) as

$$n(\widehat{\rho} - 1) \xrightarrow{d} \frac{\int_0^1 W(r)dW(r)}{\int_0^1 W(r)^2 dr} = \frac{\frac{1}{2}\left(W(1)^2 - 1\right)}{\int_0^1 W(r)^2 dr}.$$

[1]Equations (10)-(12) are different from those reported in Theorem 3 in Sowell (1990). This is because, as remarked in Section 3 of Marinucci and Robinson (1999), the partial sum of an $I(d)$ process, adjusted an appropriate normalizing term, should converge to the Type I fBm denoted by $B^H(t)$ in the present paper, not to the Type II fBm adopted in Sowell (1990).

3 Asymptotic Properties

To develop the asymptotic properties of the centered LS estimator $\widehat{\rho}_n - \rho_n$ defined in (5), we first introduce the limit behavior of the partial sum process $\sum_{t=1}^{[nr]} u_t$ for any $r \in [0,1]$. As $u_t = B^H(t) - B^H(t-1)$, we have

$$\begin{aligned} n^{-H} \sum_{t=1}^{[nr]} u_t &= n^{-H} \sum_{t=1}^{[nr]} \left\{ B^H(t) - B^H(t-1) \right\} = n^{-H} B^H([nr]) \\ &\sim B^H\left(\frac{[nr]}{n}\right) \Rightarrow B^H(r), \quad \text{as } n \to \infty, \end{aligned} \tag{13}$$

where equivalent in distribution comes from the self-similarity property of the fBm $B^H(t)$.

The convergence result in (13) is the source of the asymptotic theory developed in the present paper. Sowell (1990) gives a similar weak convergence result for the partial sum process $\sum_{t=1}^{[nr]} u_t$ when $u_t \sim I(d)$; see also Marinucci and Robinson (1999). Therefore, all the results in our paper applies to the case where $u_t \sim I(d)$. It is important to note that Sowell uses the result of Davydov (1970) to establish the weak convergence while we do not need to resort to Davydov (1970) as our errors are normally distributed.

The result in (13) compares with Donsker's functional central limit theorem, which states that,

$$n^{-0.5} \sum_{t=1}^{[nr]} \epsilon_t \Rightarrow W(r) = B^{0.5}(r), \quad \text{as } n \to \infty, \tag{14}$$

where ϵ_t is a sequence of i.i.d. random variables with mean zero and variance one.

Define a fractional OU (fOU) process through the following stochastic differential equation

$$dJ_c^H(t) = -cJ_c^H(t)dt + dB^H(t), J_c^H(0) = O_p(1). \tag{15}$$

Cheridito et al. (2003) proved that, for $t > 0$, the differential equation (15) has a unique solution and takes the form of

$$J_c^H(t) = e^{-ct} J_c^H(0) + \int_0^t e^{-c(t-s)} dB^H(s),$$

where the integral is a path-wise Riemann-Stieltjes integral. It is worthwhile to mention that, when $H = 0.5$, $J_c^H(t)$ becomes the traditional OU process studied in Phillips (1987a). If in addition, $c = 0$, the process $J_c^H(t)$ is a standard Brownian motion.

Lemma 1 *Let $\{X_t\}$ be the time series generated by (1) and (2). Then, as $n \to \infty$,*

1. $n^{-H} X_{[nr]} \Rightarrow \sigma J_c^H(r)$, *for any* $r \in [0,1]$;
2. $n^{-1-H} \sum_{t=1}^n X_t \Rightarrow \sigma \int_0^1 J_c^H(r) dr$;
3. $n^{-1-2H} \sum_{t=1}^n X_t^2 \Rightarrow \sigma^2 \int_0^1 J_c^H(r)^2 dr$;
4. $n^{-2H} \sum_{t=1}^n X_{t-1}\varepsilon_t \Rightarrow \begin{cases} \sigma^2 \left(J_c^H(1)^2 - J_c^H(0)^2 + 2c\int_0^1 J_c^H(r)^2 dr - 1 \right)/2, & \text{if } H = 0.5 \\ \sigma^2 \left(J_c^H(1)^2 - J_c^H(0)^2 + 2c\int_0^1 J_c^H(r)^2 dr \right)/2, & \text{if } H > 0.5 \end{cases}$;
5. $n^{-1} \sum_{t=1}^n X_{t-1}\varepsilon_t \xrightarrow{p} -\sigma^2/2$, *if* $H < 0.5$.

Remark 1 *This lemma is related to Lemma 1 in Phillips (1987a) with several difference. First, compared with Lemma 1.a-1.c of Phillips (1987a), $J_c(r)$ is replaced with $J_c^H(r)$ in Lemma 1, and the rate of convergence for $X_{[nr]}$, $\sum_{t=1}^n X_t$, $\sum_{t=1}^n X_t^2$ is n^{-H}, n^{1-H}, n^{-1-2H} respectively. Second, the rate and the limit of $\sum_{t=1}^n X_{t-1}\varepsilon_t$ depend on H. When $H \geq 0.5$, the rate of $\sum_{t=1}^n X_{t-1}\varepsilon_t$ is n^{-2H}. The limit has one additional term (i.e., $-\sigma^2/2$) when $H = 0.5$ than when $H > 0.5$. This difference reflects in the limit of $n^{-2H}\sum_{t=1}^n \varepsilon_t^2$. When $H = 0.5$, the limit of $n^{-2H}\sum_{t=1}^n \varepsilon_t^2$ is σ^2. When $H > 0.5$, the limit of $n^{-2H}\sum_{t=1}^n \varepsilon_t^2$ is zero. Third, the initial value $J_c^H(0)$, which is the limit of $n^{-H}X_0/\sigma$, plays an explicit role in the limit of $n^{-2H}\sum_{t=1}^n X_{t-1}\varepsilon_t$ when $H \geq 0.5$.*

Remark 2 *When $H = 0.5$, $J_c^H(r) = J_c(r)$ and Part 5 of Lemma 1 becomes irrelevant. In this case, the results in Parts 1-3 of Lemma 1 are exactly the same as those in Lemma 1.a-1.c in Phillips (1987a). If we further let $J_c^H(0) = 0$, the result in Part 4 of 1 becomes*

$$n^{-1}\sum_{t=1}^n X_{t-1}\varepsilon_t \Rightarrow \sigma^2 \left(J_c(1)^2 + 2c\int_0^1 J_c(r)^2 dr - 1 \right)/2 = \sigma^2 \int_0^1 J_c(r) dW(r),$$

which is the same as that in Lemma 1.d of Phillips (1987a).

Remark 3 *The convergence result in Part 1 of Lemma 1 is the key to the development of the results in the rest of the Lemma. With slight adjustments, the result in Part 1 can be extended to the case where u_t becomes an $I(d)$ process. When $u_t \sim I(d)$, Davydov (1970) has established the weak convergence result as $n^{-H}\left(\frac{\Gamma(1-2d)}{(1+2d)\Gamma(1+d)\Gamma(1-d)}\right)^{-1/2}\sum_{t=1}^{[nr]} u_t \Rightarrow B^H(r)$ when $n \to \infty$. Consequently, with the use of the continuous mapping theorem, it can be proved easily that $n^{-H}\left(\frac{\Gamma(1-2d)}{(1+2d)\Gamma(1+d)\Gamma(1-d)}\right)^{-1/2} X_{[nr]} \Rightarrow \sigma J_c^H(r)$.*

Theorem 2 *Let $\{X_t\}$ be the time series generated by (1) and (2). Then, as $n \to \infty$, if $H = 0.5$,*

$$n\left(\widehat{\rho}_n - \rho_n\right) \Rightarrow \frac{\left(J_c^H(1)^2 - J_c^H(0)^2 + 2c\int_0^1 J_c^H(r)^2 dr - 1\right)/2}{\int_0^1 J_c^H(r)^2 dr}; \tag{16}$$

if $H > 0.5$

$$n\left(\widehat{\rho}_n - \rho_n\right) \Rightarrow \frac{\left(J_c^H(1)^2 - J_c^H(0)^2 + 2c\int_0^1 J_c^H(r)^2 dr\right)/2}{\int_0^1 J_c^H(r)^2 dr}; \tag{17}$$

if $H < 0.5$,

$$n^{2H}\left(\widehat{\rho}_n - \rho_n\right) \Rightarrow \frac{-1/2}{\int_0^1 J_c^H(r)^2 dr}. \tag{18}$$

Remark 4 *When we compare Theorem 2 to Theorem 1 in Phillips (1987a), we have a few observations. First, when $H = 0.5$, $\widehat{\rho}_n - \rho_n$ has the same convergence rate and the same limiting distribution as those in Phillips (1987a). Second, when $H > 0.5$, the convergence rate of $\widehat{\rho}_n - \rho_n$ is n, which is the same as that when $H = 0.5$. However, the limit has one less term in the numerator comparing to the case of $H = 0.5$. When $H < 0.5$, the rate of convergence in $\widehat{\rho}_n - \rho_n$ is n^{2H}, which is slower than that when $H \geq 0.5$. The numerator in the limit has two less terms than that when $H = 0.5$.*

Remark 5 *If $c = 0$, then $\rho_n = \exp(-c/n) = 1$. In this case, the model in (1) becomes a unit root process with FGNs. With the further assumption that $J_c^H(0) = 0$, the results in Theorem 2 becomes*

$$n\left(\widehat{\rho}_n - \rho_n\right) \Rightarrow \frac{\frac{1}{2}B^H(1)^2}{\int_0^1 B^H(r)^2 dr} \qquad \text{when} \quad H > 0.5, \tag{19}$$

$$n^{2H}\left(\widehat{\rho}_n - \rho_n\right) \Rightarrow \frac{-1/2}{\int_0^1 B^H(r)^2 dr} \qquad \text{when} \quad H < 0.5, \tag{20}$$

The result in (19) is the same as that developed in Sowell (1990) and Marinucci and Robinson (1999) for the unit root process with $I(d)$ errors when $d = H - 1/2 > 0$. However, when $H < 0.5$ our limiting result in (20) is slightly different with that obtained in Sowell (1990) and Marinucci and Robinson (1999) when $d = H - 1/2 < 0$; see (12) in the present paper. The difference arises because the $I(d)$ process used in Sowell (1990) has different variance and long-run variance from those of the FGN. The variance *and the long-run variance of an $I(d)$ process is* $\frac{\Gamma(1-2d)}{\Gamma(1-d)^2}$ *and* $O(n^{2H})\frac{\Gamma(1-2d)}{(1+2d)\Gamma(1+d)\Gamma(1-d)}$, *respectively. The ratio of* $\frac{\Gamma(1-2d)}{\Gamma(1-d)^2}$ *and* $\frac{\Gamma(1-2d)}{(1+2d)\Gamma(1+d)\Gamma(1-d)}$, *divided by 2, gives*

$$\frac{(1+2d)\Gamma(1+d)}{2\Gamma(1-d)} = \frac{H\Gamma(0.5+H)}{\Gamma(1.5-H)},$$

which is the numerator of the limit in (11) that has been derived by Marinucci and Robinson (1999).

Remark 6 *There is a discontinuity in the limit theory when H passes 0.5. When H increases to 0.5, the rate of convergence moves from n^{2H} to n. The limit involves two additional terms in the numerator, $\left[J_c(1)^2 - J_c(0)^2\right]/2$ and $c\int_0^1 J_c^H(r)^2 dr$. When H further increases from 0.5, the rate of convergence stays at n. The limit involves one less term in the numerator as the term $-1/2$ is gone when $H > 0.5$.*

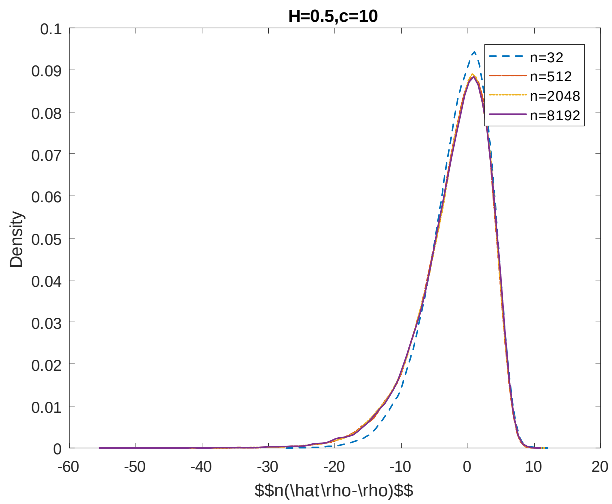
4 Monte Carlo Studies

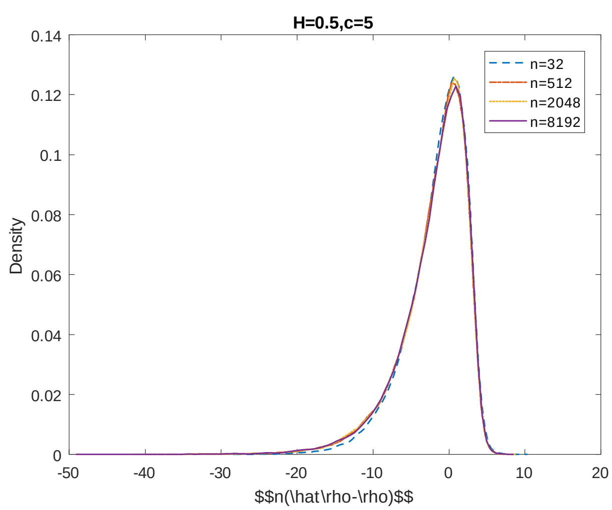
To check how well the limit distribution perform in finite sample, we carry out several Monte Carlo studies. In all studies, we simulate data from Model (1) and (2). For each time series simulated, we estimate ρ_n and calculate $n\left(\widehat{\rho}_n - \rho_n\right)$ when $H \geq 0.5$ and $n^{2H}\left(\widehat{\rho}_n - \rho_n\right)$ when $H < 0.5$. Four different sample sizes are considered, namely, $n = 32, 512, 2048, 8192$. Three values are considered for H, namely $H = 0.5, 0.9, 0.1$.[2] Two values are considered for c, namely, $c = 10, 5$. The $200,000$ replications are used to obtain density of $n\left(\widehat{\rho}_n - \rho_n\right)$ or $n^{2H}\left(\widehat{\rho}_n - \rho_n\right)$.

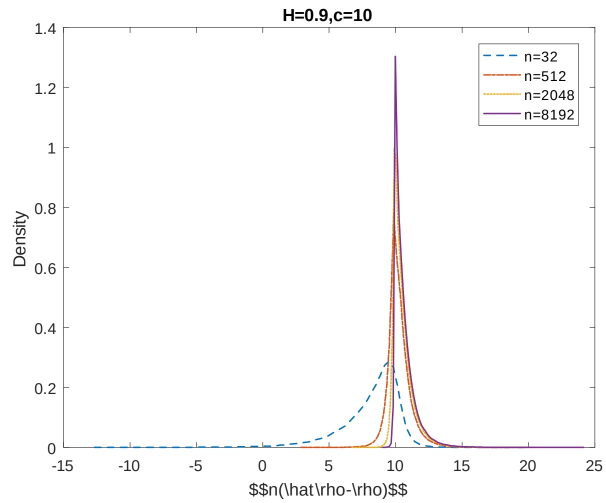
Figures 1-2 display the density of $n\left(\widehat{\rho}_n - \rho_n\right)$ when $H = 0.5$ and $c = 10, 5$. When $c = 10$, the densities are almost identical when $n \geq 512$. The density when $n = 32$ is close to that when n is larger, suggesting the limit distribution provides accurate approximations to the finite sample distribution when the sample size is as small as 32. In all cases, the density is left-skewed.

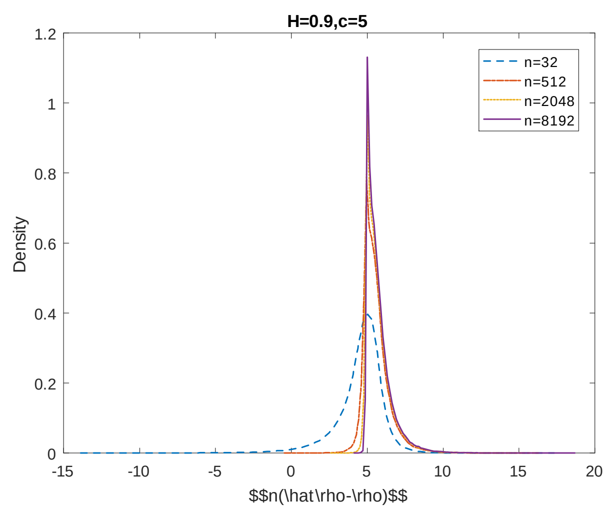
Figures 3-4 display the density of $n\left(\widehat{\rho}_n - \rho_n\right)$ when $H = 0.9$ and $c = 10, 5$. For both values of c, the density when $n = 32$ is very different from that when $n = 8192$. The density for $n = 2048$ is very close to that for $n = 8192$. For small values of n, the density is left-skewed. Interestingly, the density becomes right-skewed when n is

[2] The choice of $H = 0.1$ is empirically relevant for modeling logarithmic realized volatility, as found in Gatheral et al. (2018) and Wang et al. (2019).


H=0.5,c=10
Density
$$n(\hat\rho-\rho)$$
n=32
n=512
n=2048
n=8192
0.1
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
-60
-50
-40
-30
-20
-10
0
10
20

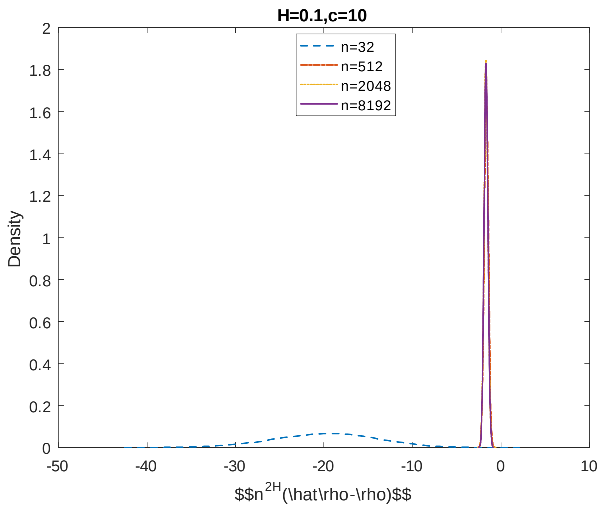

H=0.5,c=5
Density
$$n(\hat\rho-\rho)$$
n=32
n=512
n=2048
n=8192
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
-50
-40
-30
-20
-10
0
10
20

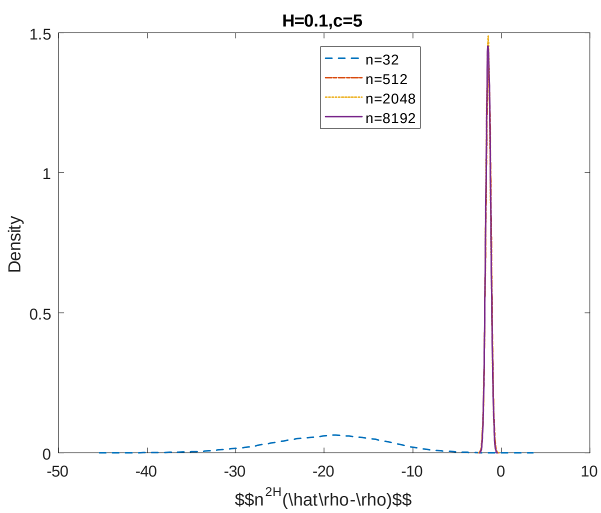




larger. Although the same rate applies to $H = 0.5$ and to $H > 0.5$, the convergence of the density is much slower when $H > 0.5$ than that when $H = 0.5$. This study indicates that the asymptotic distribution approximates the finite sample distribution less accurately when $H > 0.5$ than when $H = 0.5$ if n is small.

Figures 5-6 display the density of $n^{2H}\left(\widehat{\rho}_n - \rho_n\right)$ when $H = 0.1$ and $c = 10, 5$. For both values of c, the density when $n = 32$ is hugely different from those for other values of n, suggesting one would make a terrible mistake by using the limit distribution to approximate the finite sample distribution when $n = 32$. However, the densities for $n = 512, 2048, 8192$ are nearly identical.


H=0.1,c=10
n=32
n=512
n=2048
n=8192
Density
$$n^{2H}(\hat\rho-\rho)$$


H=0.1,c=5
n=32
n=512
n=2048
n=8192
Density
$$n^{2H}(\hat\rho-\rho)$$

5 Conclusions

In this paper we study the properties of the least squares estimator of the autoregressive parameter in local to unity processes when errors are assumed to be fractional Gaussian noises with the Hurst parameter H. It is shown that the estimator is consistent when $H \in (0,1)$. Moreover, the rate of convergence is n when $H \in [0.5,1)$ whereas the rate of convergence is n^{2H} when $H \in (0,0.5)$. This result suggests that the estimator has a slower rate of consistency when $H \in (0,0.5)$ than when $H \in [0.5,1)$.

Furthermore, the limit distribution of the centered least squares estimator depends on H. When $H = 0.5$, the limit distribution is the same as that obtained in Phillips (1987a) for the local to unity model with errors for which the standard functional central theorem is applicable. When $H > 0.5$ or when $H < 0.5$, the limit distributions are new to the literature. The limit distribution for $H > 0.5$ has one less term than that for $H = 0.5$. The limit distribution for $H < 0.5$ has two less terms than that for $H = 0.5$. Simulation studies are performed to check the reliability of the asymptotic approximation. When $H > 0.5$, a large sample size is needed for the limit distribution to provide an accurate approximation to the finite sample distribution. When $H = 0.5$, a small sample size is enough for the limit distribution to provide an accurate approximation to the finite sample distribution. When $H < 0.5$, a moderate sample size is needed for the limit distribution to provide an accurate approximation to the finite sample distribution.

Appendix

Proof of Lemma 1. To prove Lemma 1.1, we first note that

$$\begin{aligned}
X_t &= \rho_n X_{t-1} + \varepsilon_t = \rho_n^t X_0 + \sum_{j=0}^{t-1} \rho_n^j \varepsilon_{t-j} = \rho_n^t X_0 + \sum_{s=1}^{t} \rho_n^{t-s} \varepsilon_s \\
&= \rho_n^t X_0 + \sigma \sum_{s=1}^{t} \rho_n^{t-s} \left[B^H(s) - B^H(s-1) \right] \\
&= \rho_n^t X_0 + n^H \sigma \sum_{s=1}^{t} \rho_n^{t-s} \left[B^H\left(\frac{s}{n}\right) - B^H\left(\frac{s-1}{n}\right) \right] \\
&= \rho_n^t X_0 + n^H \sigma \sum_{s=1}^{t} \rho_n^{t-s} \int_{(s-1)/n}^{s/n} dB^H(r),
\end{aligned}$$

where the fifth equation is from the similarity property of the fractional Brownian motion We then have

$$
\begin{aligned}
n^{-H}X_t &= e^{-ct/n}\frac{X_0}{n^H}+\sigma\sum_{s=1}^{t}\int_{(s-1)/n}^{s/n} e^{-c(t-s)/n}dB^H(r) \\
&= e^{-ct/n}\frac{X_0}{n^H}+\sigma\sum_{s=1}^{t}\int_{(s-1)/n}^{s/n} e^{-c(t/n-r)}e^{-c(r-s/n)}dB^H(r) \\
&= e^{-ct/n}\frac{X_0}{n^H}+\sigma\sum_{s=1}^{t}\int_{(s-1)/n}^{s/n} e^{-c(t/n-r)}\left[1+O\left(1/n\right)\right]dB^H(r) \\
&= e^{-ct/n}\frac{X_0}{n^H}+\sigma\int_{0}^{t/n} e^{-c(t/n-r)}dB^H(r)+O_p\left(1/n\right) \\
&= \sigma e^{-ct/n}\left[J_c^H\left(0\right)+o_p\left(1\right)\right]+\sigma\int_{0}^{t/n} e^{-c(t/n-r)}dB^H(r)+O_p\left(1/n\right) \\
&= \sigma J_c^H\left(t/n\right)+O_p\left(1\right),
\end{aligned}
$$

where the third equation is from the Taylor expansion of $e^{-c(r-s/n)}$ and the last equation comes from the definition of the fOU process $J_c^H\left(t/n\right)$ given in (15). Therefore, for $r\in[0,1]$, we have

$$
n^{-H}X_{[nr]}=\sigma J_c^H\left(\frac{[nr]}{n}\right)+O_p\left(1\right)\Rightarrow\sigma J_c^H\left(r\right), \quad \text{as } \; n\to\infty.
$$

This proves Lemma 1.1.

Then, the convergence results in Lemma 1.2-1.3 can be obtained straightforwardly by using the continuous mapping theorem (Billingsley, 1968, p. 30).

To prove the results in Lemma 1.4-1.5, we first have

$$
\begin{aligned}
X_t^2 &= \left(\rho_n X_{t-1}+\varepsilon_t\right)^2=\rho_n^2X_{t-1}^2+2\rho_nX_{t-1}\varepsilon_t+\varepsilon_t^2 \\
&= X_{t-1}^2+\left(\rho_n^2-1\right)X_{t-1}^2+2\rho_nX_{t-1}\varepsilon_t+\varepsilon_t^2,
\end{aligned}
$$

and

$$
\sum_{t=1}^{n}X_{t-1}\varepsilon_t=\frac{1}{2\rho_n}\left\{X_n^2-X_0^2-\left(\rho_n^2-1\right)\sum_{t=1}^{n}X_{t-1}^2-\sum_{t=1}^{n}\varepsilon_t^2\right\}.
$$

It is crucially important to note that $\sum_{t=1}^{n}\varepsilon_t^2=O_p\left(1/n\right)$ for all values of $H\in(0,1)$ and

$$
\begin{aligned}
\frac{1}{n}\sum_{t=1}^{n}\varepsilon_t^2 &= n^{-1}\sigma^2\sum_{t=1}^{n}\left[B^H(t)-B^H(t-1)\right]^2 \\
&= n^{-1+2H}\sigma^2\sum_{t=1}^{n}\left[B^H\left(\frac{t}{n}\right)-B^H\left(\frac{t-1}{n}\right)\right]^2\xrightarrow{p}\sigma^2,
\end{aligned}
$$

where the convergence result is from Proposition 4.2 in Vittasaari (2015). As a result,

$$n^{-2H}\sum_{t=1}^{n}\varepsilon_t^2 \xrightarrow{p} \begin{cases} 0 & \text{when } H>0.5 \\ \sigma^2 & \text{when } H=0.5 \\ +\infty & \text{when } H=0.5 \end{cases}.$$

This is the reason why $\sum_{t=1}^{n}X_{t-1}\varepsilon_t$ having distinct asymptotic behaviors when H takes various values.

When $H=0.5$, the four items in the decomposition of $\sum_{t=1}^{n}X_{t-1}\varepsilon_t$ have a same order and, as $n\to\infty$,

$$\begin{aligned} n^{-2H}\sum_{t=1}^{n}X_{t-1}\varepsilon_t &= \frac{1}{2\rho_n}\left\{\frac{X_n^2-X_0^2}{n^{2H}} - n\left(\rho_n^2-1\right)\frac{1}{n^{1+2H}}\sum_{t=1}^{n}X_{t-1}^2 - n^{-2H}\sum_{t=1}^{n}\varepsilon_t^2\right\} \\ &\Rightarrow \frac{\sigma^2}{2}\left\{J_c^H(1)^2 - J_c^H(0)^2 + 2c\int_0^1 J_c^H(r)^2dr - 1\right\}, \end{aligned}$$

where the convergence result comes from Lemma 1.1 and Lemma 1.3, together with the limit of $n^{-2H}\sum_{t=1}^{n}\varepsilon_t^2$ obtained above.

Whereas, when $H>0.5$, $\sum_{t=1}^{n}\varepsilon_t^2 = O_p(n)$ is asymptotically dominated by the other terms in the decomposition of $\sum_{t=1}^{n}X_{t-1}\varepsilon_t$. Hence, it disappears in the limit of $n^{-2H}\sum_{t=1}^{n}X_{t-1}\varepsilon_t$ that takes the form of

$$\begin{aligned} n^{-2H}\sum_{t=1}^{n}X_{t-1}\varepsilon_t &= \frac{1}{2\rho_n}\left\{\frac{X_n^2-X_0^2}{n^{2H}} - n\left(\rho_n^2-1\right)\frac{1}{n^{1+2H}}\sum_{t=1}^{n}X_{t-1}^2 - n^{-2H}\sum_{t=1}^{n}\varepsilon_t^2\right\} \\ &\Rightarrow \frac{\sigma^2}{2}\left\{J_c^H(1)^2 - J_c^H(0)^2 + 2c\int_0^1 J_c^H(r)^2dr\right\}. \end{aligned}$$

In contrast, when $H<0.5$, $\sum_{t=1}^{n}\varepsilon_t^2 = O_p(1/n)$ asymptotically dominates the other terms in the decomposition of $\sum_{t=1}^{n}X_{t-1}\varepsilon_t$. Hence,

$$n^{-1}\sum_{t=1}^{n}X_{t-1}\varepsilon_t = \frac{1}{2\rho_n}\left\{o_p(1) - n^{-1}\sum_{t=1}^{n}\varepsilon_t^2\right\} \xrightarrow{p} -\frac{\sigma^2}{2}.$$

The proof of Lemma 1 is complete.

Proof of Theorem 2. The theorem is the direct consequence of Lemma 1.3-1.5. In particular, (16) and (17) follow from Lemma 1.3-1.4 and (18) follow from Lemma 1.3 and Lemma 1.5.

References

Billingsley, P., 1968. *Convergence of Probability Measures.* Wiley, New York.

Chan, N. H, , Wei, C. Z., 1988. Limiting Distributions of Least Squares Estimates of Unstable Autoregressive Processes. *Annals of Statistics* 16, 367-401.

Cheridito, P., Kawaguchi, H., Maejima, M., 2003. Fractional Ornstein–Uhlenbeck processes. *Electronical Journal Probability* 8(3), 1-14.

Davydov, Y. A., 1970. The Invariance Principle for Stationary Processes, *Theory of Probability and Its Applications* 15, 487-489.

Gatheral, J., T. Jaisson, and Rosenbaum, M., 2018. Volatility is rough. *Quantitative Finance* 18 (6), 933-949.

Marinucci, D., and Robinson, P. M., 1999. Alternative forms of fractional Brownian motion. *Journal of Statistical Planning and Inference*, 80, 111-122.

Park, J. Y., 2003. Weak unit roots. Working Paper, Department of Economics, Rice University.

Phillips, P. C. B., 1987a. Toward a unified asymptotic theory for autoregression. *Biometrika* 74, 533-547.

Phillips, P. C. B., 1987b. Time series regression with a unit root. *Econometrica* 55, 277-301.

Sowell, F., 2000. The Fractional Unit Root Distribution, *Econometrica* 58, 495-505.

Wang, X., Xiao, W., and Yu, J., 2019. Estimation and inference of fractional continuous-time model with discrete-sampled data. Working paper, SMU.

Viitasaari, L., 2019. Necessary and sufficient conditions for limit theorems for quadratic variations of Gaussian sequences. *Probability Survey* 16, 62–98.