Methodol Comput Appl Probab
DOI 10.1007/s11009-012-9306-7

Limit Theory for Moderate Deviations from a Unit Root Under Innovations with a Possibly Infinite Variance

Sai-Hua Huang · Tian-Xiao Pang · Chengguo Weng

Received: 5 October 2011 / Revised: 11 August 2012 / Accepted: 30 September 2012

Abstract An asymptotic theory was given by Phillips and Magdalinos (J Econom 136(1):115–130, 2007) for autoregressive time series $Y_t = \rho Y_{t-1} + u_t, t = 1, \ldots, n$, with $\rho = \rho_n = 1 + c/k_n$, under $(2+\delta)$-order moment condition for the innovations u_t, where $\delta > 0$ when $c < 0$ and $\delta = 0$ when $c > 0$, $\{u_t\}$ is a sequence of independent and identically distributed random variables, and $(k_n)_{n \in \mathbb{N}}$ is a deterministic sequence increasing to infinity at a rate slower than n. In the present paper, we established similar results when the truncated second moment of the innovations $l(x) = \mathsf{E}[u_1^2 I\{|u_1| \leq x\}]$ is a slowly varying function at ∞, which may tend to infinity as $x \to \infty$. More interestingly, we proposed a new pivotal for the coefficient ρ in case $c < 0$, and formally proved that it has an asymptotically standard normal distribution and is nuisance parameter free. Our numerical simulation results show that the distribution of this pivotal approximates the standard normal distribution well under normal innovations.

Keywords Unit root process · Moderate deviation · Convergence rate · Limiting distribution

AMS 2000 Subject Classifications 62F12 · 60F05

S.-H. Huang
Department of Ocean Science and Engineering, Zhejiang University,
Zijingang Campus, Hangzhou 310058, People's Republic of China
e-mail: ivyhuang21@163.com

T.-X. Pang (✉)
Department of Mathematics, Zhejiang University,
Yuquan Campus, Hangzhou 310027, People's Republic of China
e-mail: txpang@zju.edu.cn

C. Weng
Department of Statistics and Actuarial Science,
University of Waterloo, Waterloo, N2L 3G1, Canada
e-mail: chengguo.weng@uwaterloo.ca

Published online: 10 October 2012

1 Introduction

Consider the following autoregressive model with order one AR(1):

$$Y_t = \rho Y_{t-1} + u_t, \quad t = 1, \ldots, n, \tag{1.1}$$

where Y_t is the observation at time t, $\{u_t\}$ is a sequence of independent and identically distributed (i.i.d.) random variables and ρ is an unknown parameter. Usually, ρ is estimated by its least square estimate (LSE)

$$\hat{\rho} = \frac{\sum_{t=1}^{n} Y_{t-1} Y_t}{\sum_{t=1}^{n} Y_{t-1}^2}. \tag{1.2}$$

If the u_t's are normally distributed, $\hat{\rho}$ is also the maximum likelihood estimate of ρ. There has been considerable interest in the asymptotic properties of $\hat{\rho}$ both in the statistics and in the econometrics literature.

When $\mathsf{E}[u_t] = 0$, $0 < \mathsf{E}\left[u_t^2\right] = \sigma^2 < \infty$ and Y_0 is a constant, it is well known that the AR(1) process 1.1 with $|\rho| < 1$ is asymptotically stationary and the LSE $\hat{\rho}$ satisfies the following limiting property:

$$\sqrt{n}(\hat{\rho} - \rho) \xrightarrow{d} N(0, 1 - \rho^2), \quad \text{as } n \to \infty; \tag{1.3}$$

see Anderson (1959). Here and in what follows, the notation $\xrightarrow{d}$ denotes convergence in distribution, and $\xrightarrow{p}$ denotes convergence in probability, respectively. When $\rho = 1$, model 1.1 becomes a unit root model, which is non-stationary, and the convergence rate of $\hat{\rho}$ is quite different from the stationary case. In this case,

$$n(\hat{\rho} - \rho) \xrightarrow{d} \frac{W^2(1) - 1}{2\int_0^1 W^2(r)dr}, \quad \text{as } n \to \infty, \tag{1.4}$$

where $\{W(t)\}$ is a standard Wiener process; see White (1958) and Rao (1978).

Over the last two decades, regression asymptotics with root near unity have been playing an important role in time series analysis. Motivated by the classical Poisson approximation, Chan and Wei (1987) considered AR(1) process with $\rho = 1 + c/n$ and a fixed constant c so that

$$Y_t = (1 + c/n) Y_{t-1} + u_t, \tag{1.5}$$

which is called nearly non-stationary AR(1) process in their paper. By assuming $\{u_t\}$ to be a martingale difference sequence with respect to an increasing sequence of σ-fields $\{\mathcal{F}_t\}$ such that

$$\frac{1}{n}\sum_{t=1}^{n} \mathsf{E}\left[u_t^2 | \mathcal{F}_{t-1}\right] \xrightarrow{p} 1 \tag{1.6}$$

as $n \to \infty$ and $Y_0 = 0$, Chan and Wei (1987) established the following results:

$$\tau_n := \sqrt{\sum_{t=1}^{n} Y_{t-1}^2}(\hat{\rho} - \rho) \xrightarrow{d} \frac{\int_0^1 (1+bt)^{-1} W(t) dW(t)}{\sqrt{\int_0^1 (1+bt)^{-2} W^2(t) dt}} \text{ as } n \to \infty, \tag{1.7}$$

and

$$\tau_n^{'} := n(\hat{\rho} - \rho) \xrightarrow{d} \frac{\int_0^1 (1+bt)^{-1} W(t) dW(t)}{\int_0^1 (1+bt)^{-2} W^2(t) dt} \text{ as } n \to \infty, \tag{1.8}$$

where $b = e^{2c} - 1$. The convergence result 1.8 is not explicitly expressed in their paper, but it can be easily recovered from their proof for Eq. 1.7.

Lately, motivated by bridging the $\sqrt{n}$ and n convergence rates for the stationary case and non-stationary case, Giraitis and Phillips (2006) and Phillips and Magdalinos (2007) proposed a model similar to Eq. 1.5 by assuming $\rho = \rho_n = 1 + c/k_n$ with a constant c, $Y_0 = o_p(\sqrt{k_n})$, and $(k_n)_{n \in \mathbb{N}}$ is a sequence increasing to ∞ such that $k_n = o(n)$. For $c < 0$, Phillips and Magdalinos (2007) established

$$\sqrt{nk_n}(\hat{\rho} - \rho) \xrightarrow{d} N(0, -2c), \text{ as } n \to \infty, \tag{1.9}$$

under i.i.d. innovations u_t with $\mathsf{E}[u_t] = 0$ and $\mathsf{E}\left[|u_t|^{2+\delta}\right] < \infty$ for some $\delta > 0$. Note that, by taking $k_n = n^\alpha$ with $0 \le \alpha \le 1$, the above result indeed bridges the $\sqrt{n}$ and n convergence rates. It is also worth mentioning that the limiting distributions are substantial different between these two cases $k_n = n^\alpha$ with $0 \le \alpha < 1$ and $k_n = n$. For $c < 0$, similar results have been established by Giraitis and Phillips (2006) for the innovations to be a martingale difference sequence with a finite second moment. In the case with $c > 0$, Phillips and Magdalinos (2007) established

$$\frac{k_n \rho^n}{2c}(\hat{\rho} - \rho) \xrightarrow{d} C, \text{ as } n \to \infty, \tag{1.10}$$

where the random variable C is a standard Cauchy random variable.

Motivated by and Phillips and Magdalinos (2007), we are interested in the following two questions. First, we shall investigate whether the results 1.9 and 1.10 are still valid when the i.i.d. innovation sequence $\{u_t\}$ has a second moment which is possibly infinite. By adopting the truncation approach, we successfully achieve the results; for more details, see Theorem 2.1 in the sequel. Second, we are interested in the statistical inference on the parameter ρ in the model. It would be difficult to conduct any hypothesis test or construct confidence intervals for the interesting parameter ρ by directly using result 1.9, since there is one unknown nuisance parameter c in the limiting distribution. To overcome such an obstacle, we proposed a new pivot for ρ and, for the case with $c < 0$, we formally established its limiting distribution, which is a standard normal distribution and hence nuisance parameter free; the findings are summarized in Theorem 2.2 in the sequel. To demonstrate the finite sample properties of the pivotal, a simulation study is presented in the second part of Section 2. The results show that the distribution of this pivotal approximates the standard normal distribution well under normal innovations.

The rest of the paper is organized as follows. Section 2 are our main results and the numerical example. All the lemmas and proof are relegated to an Appendix.

2 Main Results and a Numerical Example

2.1 Main Results

To introduce our main results, we impose the following conditions:

C1. In $\rho := \rho_n = 1 + c/k_n$, c is a constant, $(k_n)_{n\in\mathbb{N}}$ is a real sequence increasing to ∞ with $k_n = o(n)$;
C2. The innovations $\{u_t\}$ are i.i.d. random variables with $\mathsf{E}[u_1] = 0$, and their truncated second moment $l(x) = \mathsf{E}[u_1^2 I\{|u_1| \le x\}]$ is a slowly varying function of x at ∞, where $I\{\cdot\}$ stands for the indicator function;
C3. $Y_0 = o_p(\sqrt{k_n})$.

Remark 2.1 Although the above condition **C2** allows the second moment of the innovations to be infinite, it implies that $\mathsf{E}\left[|u_1|^\delta\right] < \infty$ for any $0 \le \delta < 2$. Hence, the innovation random variables which are in the domain of attraction of a stable law with index $\alpha < 2$ are ruled out by this condition. This means that our results do not apply to a model with very heavy-tailed innovations.

Theorem 2.1 *Assume that conditions* **C1–C3** *are satisfied by process 1.1.*

(a) *If additionally $c < 0$, the asymptotic normality of the LSE $\hat\rho$ given in Eq. 1.9 is satisfied.*
(b) *If additionally $c > 0$, the limiting property of the LSE $\hat\rho$ given in Eq. 1.10 is satisfied.*

Remark 2.2 The result in part (a) of Theorem 2.1 bridges the $\sqrt{n}$ and n convergence rates for $\hat\rho$ respectively in an asymptotically stationary AR(1) model and the corresponding unit root process when the innovations $\{u_t\}$ are i.i.d. with zero mean and a truncated second moment $l(x)$ slowly varying function at ∞. The proofs in these two special cases can be found in Davis and Resnick (1985) and Wang (2006) respectively.

As mentioned in the introduction section, we may be interested in the statistical inference on the parameter ρ in process 1.1. To this end, we propose an innovative pivot for ρ as shown below:

$$\beta_n := \sqrt{\frac{n\sum_{t=1}^n Y_{t-1}^2}{\sum_{t=1}^n (Y_t - \hat\rho Y_{t-1})^2}}(\hat\rho - \rho). \tag{2.1}$$

We formally established its limiting distribution as stated in Theorem 2.2 below.

Theorem 2.2 *Assume that conditions* **C1–C3** *are satisfied by process 1.1, and $c < 0$. Then the pivot β_n given in Eq. 2.1 is asymptotically normal:*

$$\beta_n \xrightarrow{d} N(0, 1), \quad as\ n \to \infty. \tag{2.2}$$

Remark 2.3 Note that the pivotal quantity β_n only contains the interesting parameter ρ, and the limiting distribution is nuisance parameter free. Thus, β_n is a potentially useful statistics for conducting hypothesis test and constructing confidence intervals for the parameter ρ. A simulation study is conducted in the subsequent numerical example to illustrate the finite sample properties of β_n.

2.2 A Numerical Example

To demonstrate the finite sample properties of the pivot β_n defined in Eq. 2.1, we experiment on the AR(1) process 1.1 with the following numerical setup:

$$\begin{cases} c = \{-0.1, \ -1.0\}, \\ Y_0 = 0, \\ k_n \in \left\{\log n, \ n^{0.5}, \ n^{0.75}\right\}, \\ u_t \in \{N(0,1), \ t(2)\}, \end{cases} \tag{2.3}$$

where $t(2)$ denotes a student-t random variable of freedom 2 and hence its variance is infinite. It is easy to verify that each combination in the above numerical setup satisfies all the conditions in Theorem 2.2.

In our numerical experiment, we consider a set of different sample size n, and for each n, we independently generate 100,000 sequences of $\{Y_0, Y_1, \cdots, Y_n\}$ according to the AR(1) process 1.1 with parameter values as specified in Eq. 2.3, and then compute the values of statistics $\hat{\rho}$ and β_n for each of those generated sequences so that we have 100,000 values of $\hat{\rho}$ and β_n. To investigate how well the standard normal distribution approximates that of β_n, we use those 100,000 values of β_n to estimate its moments (mean, standard deviation, skewness, and kurtosis), and the probability $\Pr\left(\beta_n \le \Phi^{-1}(\tau)\right)$ for a set of different values of τ, where $\Phi^{-1}(\tau)$ denotes the τ-quantile of the standard normal distribution. If the approximation works well, $\Pr\left(\beta_n \le \Phi^{-1}(\tau)\right)$ is expected to be very close to τ. To illustrate how useful the pivot β_n can be in estimating the parameter ρ, we construct the asymptotic confidence interval for ρ by using the limiting distribution of β_n, and then compute the empirical coverage probability (ECP) of such an asymptotic confidence interval. Specifically, we first construct a 95 % confidence interval $[l_n, u_n]$ of ρ for each of those 100,000 generated sequences of $\{Y_0, Y_1, \cdots, Y_n\}$, where

$$l_n = \hat{\rho} - \Psi^{-1}(0.975) \cdot \widehat{\sigma}_n, \ \ u_n = \hat{\rho} + \Psi^{-1}(0.975) \cdot \widehat{\sigma}_n, \ \text{ and } \widehat{\sigma}_n = \sqrt{\frac{\sum_{t=1}^{n}(Y_t - \hat{\rho} Y_{t-1})^2}{n\sum_{t=1}^{n} Y_{t-1}^2}}.$$

Then, we compute the ECP as the proportion of those intervals including the true value of ρ, and ideally it is expected to be close to 95 %.

All the experiment results are reported in Tables 1 and 2, on which we have the following observations and comments.

(i) The normal approximation has fairly similar overall validity (or lack thereof) for both the normal and t innovations. First, the mean values of β_n in both tables are all negative and diminish to 0 as the sample size n increases. Such an observation is consistent to the well known fact that the LSE $\hat{\rho}$ is asymptotically unbiased and yet always negatively biased particularly when ρ is close to 1. Second, except the kurtosis column with $k_n = \log n$ in Table 2, all the moments (mean, standard deviation, skewness and kurtosis) of β_n in both

Table 1 Simulated moments and inverse quantiles of β_n for $N(0, 1)$ innovations, where Skew., Kurt. and ECP are respectively short for Skewness, Kurtosis and Empirical Coverage Probability

k_n	c	n	ρ	Mean	St.D.	Skew.	Kurt.	Simulated probability $\Pr(\beta_n \le \Phi^{-1}(\tau))$											ECP
								0.01	0.025	0.05	0.1	0.25	0.5	0.75	0.9	0.95	0.975	0.99	(%)
$k_n = \log n$	$c = -0.1$	100	0.9783	−0.3336	0.9647	0.0786	3.1752	0.0188	0.0436	0.0832	0.1589	0.3624	0.6436	0.8555	0.9502	0.9761	0.9888	0.9959	94.48
		200	0.9811	−0.2848	0.9679	0.0172	3.0818	0.0173	0.0416	0.0791	0.1495	0.3427	0.6177	0.8414	0.9464	0.9759	0.9888	0.9959	94.79
		400	0.9833	−0.2371	0.9751	0.0275	3.0365	0.0153	0.0373	0.0738	0.1420	0.3267	0.5977	0.8255	0.9399	0.9724	0.9877	0.9952	95.03
		800	0.9850	−0.1861	0.9856	0.0004	2.9940	0.0151	0.0359	0.0685	0.1339	0.3108	0.5752	0.8088	0.9316	0.9681	0.9847	0.9945	94.85
		3200	0.9876	−0.1120	0.9938	−0.0045	3.0122	0.0130	0.0313	0.0610	0.1194	0.2858	0.5437	0.7869	0.9198	0.9615	0.9813	0.9928	95.00
		6400	0.9886	−0.0846	0.9966	0.0031	3.0161	0.0121	0.0296	0.0583	0.1148	0.2768	0.5349	0.7765	0.9146	0.9581	0.9797	0.9925	95.02
	$c = -1$	100	0.7829	−0.1256	0.9961	−0.0111	3.0756	0.0139	0.0331	0.0630	0.1224	0.2903	0.5490	0.7905	0.9220	0.9634	0.9823	0.9929	94.92
		200	0.8113	−0.0995	1.0002	−0.0067	3.0693	0.0134	0.0317	0.0616	0.1176	0.2812	0.5395	0.7812	0.9173	0.9595	0.9803	0.9922	94.84
		400	0.8331	−0.0792	1.0003	−0.0113	3.0155	0.0126	0.0304	0.0583	0.1148	0.2763	0.5305	0.7741	0.9131	0.9580	0.9795	0.9922	94.99
		800	0.8504	−0.0542	0.9979	−0.0118	3.0036	0.0115	0.0287	0.0558	0.1102	0.2669	0.5208	0.7673	0.9100	0.9562	0.9790	0.9918	94.95
		3200	0.8761	−0.0333	1.0004	−0.0066	2.9862	0.0108	0.0266	0.0540	0.1059	0.2613	0.5133	0.7608	0.9053	0.9531	0.9771	0.9910	95.01
		6400	0.8859	−0.0242	1.0005	0.0001	3.0165	0.0109	0.0269	0.0523	0.1040	0.2577	0.5101	0.7579	0.9043	0.9521	0.9760	0.9906	95.08
$k_n = n^{0.5}$	$c = -0.1$	100	0.9900	−0.3789	0.9686	0.1563	3.2227	0.0199	0.0464	0.0887	0.1694	0.3856	0.6680	0.8636	0.9503	0.9756	0.9884	0.9956	94.19
		200	0.9929	−0.3593	0.9616	0.1090	3.1558	0.0183	0.0448	0.0861	0.1640	0.3750	0.6557	0.8614	0.9514	0.9773	0.9893	0.9959	94.45
		400	0.9950	−0.3404	0.9587	0.0962	3.1308	0.0174	0.0435	0.0839	0.1609	0.3652	0.6481	0.8571	0.9508	0.9775	0.9896	0.9959	94.61
		800	0.9965	−0.3111	0.9628	0.0673	3.0712	0.0167	0.0413	0.0814	0.1551	0.3560	0.6313	0.8489	0.9492	0.9764	0.9892	0.9959	94.79
		3200	0.9982	−0.2566	0.9684	0.0243	3.0261	0.0156	0.0392	0.0754	0.1443	0.3330	0.6057	0.8327	0.9434	0.9746	0.9882	0.9957	94.90
		6400	0.9988	−0.2198	0.9789	0.0139	3.0116	0.0157	0.0376	0.0723	0.1389	0.3206	0.5900	0.8192	0.9374	0.9706	0.9864	0.9951	94.89
	$c = -1$	100	0.9000	−0.1941	0.9862	−0.0009	3.0716	0.0156	0.0366	0.0696	0.1343	0.3116	0.5788	0.8126	0.9323	0.9684	0.9858	0.9945	94.92
		200	0.9293	−0.1643	0.9920	−0.0162	3.0446	0.0151	0.0357	0.0680	0.1294	0.3032	0.5648	0.8008	0.9290	0.9665	0.9835	0.9938	94.78
		400	0.9500	−0.1497	0.9914	0.0134	3.0170	0.0137	0.0332	0.0652	0.1273	0.2973	0.5616	0.7968	0.9256	0.9653	0.9832	0.9935	95.00
		800	0.9646	−0.1252	0.9953	−0.0075	2.9891	0.0136	0.0328	0.0634	0.1228	0.2894	0.5500	0.7885	0.9215	0.9618	0.9814	0.9933	94.86
		3200	0.9823	−0.0950	0.9944	0.0082	3.0007	0.0123	0.0304	0.0590	0.1162	0.2800	0.5389	0.7812	0.9169	0.9594	0.9804	0.9925	95.00
		6400	0.9875	−0.0778	0.9954	0.0049	2.9868	0.0117	0.0291	0.0579	0.1139	0.2737	0.5310	0.7756	0.9136	0.9580	0.9795	0.9919	95.04

Table 1 (continued)

k_n	c	n	ρ	Mean	St.D.	Skew.	Kurt.	Simulated probability $\Pr(\beta_n \leq \Phi^{-1}(\tau))$											ECP
								0.01	0.025	0.05	0.1	0.25	0.5	0.75	0.9	0.95	0.975	0.99	(%)
$k_n = n^{0.75}$	$c = -0.1$	100	0.9968	−0.4084	0.9806	0.2185	3.1924	0.0205	0.0487	0.0928	0.1792	0.4092	0.6804	0.8627	0.9489	0.9749	0.9876	0.9953	93.89
		200	0.9981	−0.4034	0.9736	0.2016	3.1332	0.0193	0.0472	0.0920	0.1769	0.4058	0.6781	0.8615	0.9502	0.9765	0.9887	0.9959	94.15
		400	0.9989	−0.4037	0.9708	0.2117	3.1401	0.0188	0.0474	0.0918	0.1764	0.4054	0.6793	0.8627	0.9502	0.9767	0.9888	0.9957	94.14
		800	0.9993	−0.3959	0.9711	0.2178	3.1256	0.0179	0.0456	0.0900	0.1758	0.4022	0.6762	0.8608	0.9495	0.9757	0.9886	0.9956	94.30
		3200	0.9998	−0.3948	0.9641	0.1806	3.1539	0.0191	0.0464	0.0899	0.1733	0.3967	0.6754	0.8638	0.9523	0.9776	0.9891	0.9959	94.28
		6400	0.9999	−0.3825	0.9672	0.1709	3.1328	0.0185	0.0453	0.0889	0.1726	0.3889	0.6702	0.8614	0.9509	0.9766	0.9887	0.9956	94.34
	$c = -1$	100	0.9684	−0.3029	0.9676	0.0445	3.1308	0.0177	0.0421	0.0801	0.1533	0.3497	0.6279	0.8469	0.9477	0.9757	0.9887	0.9960	94.65
		200	0.9812	−0.2852	0.9679	0.0175	3.0821	0.0173	0.0416	0.0791	0.1496	0.3429	0.6179	0.8414	0.9465	0.9759	0.9888	0.9959	94.72
		400	0.9888	−0.2735	0.9675	0.0392	3.0579	0.0161	0.0398	0.0771	0.1480	0.3385	0.6148	0.8372	0.9447	0.9752	0.9890	0.9956	94.92
		800	0.9934	−0.2559	0.9719	0.0245	3.0206	0.0157	0.0386	0.0759	0.1450	0.3343	0.6054	0.8316	0.9423	0.9737	0.9880	0.9955	94.94
		3200	0.9976	−0.2309	0.9734	0.0173	3.0157	0.0150	0.0380	0.0727	0.1402	0.3238	0.5945	0.8240	0.9399	0.9728	0.9872	0.9952	94.92
		6400	0.9986	−0.2101	0.9806	0.0121	3.0122	0.0156	0.0372	0.0715	0.1369	0.3168	0.5864	0.8170	0.9358	0.9697	0.9862	0.9948	94.89

Table 2 Simulated moments and inverse quantiles of β_n for $t(2)$ innovations, where Skew., Kurt. and ECP are respectively short for Skewness, Kurtosis and Empirical Coverage Probability

k_n	c	n	ρ	Mean	St.D.	Skew.	Kurt.	Simulated probability $\Pr(\beta_n \leq \Phi^{-1}(\tau))$											ECP
								0.01	0.025	0.05	0.1	0.25	0.5	0.75	0.9	0.95	0.975	0.99	
$k_n = \log n$	$c = -0.1$	100	0.9783	−0.2851	0.9075	−0.0561	3.2922	0.0149	0.0347	0.0674	0.1341	0.3243	0.6241	0.8624	0.9582	0.9827	0.9925	0.9975	95.78
		200	0.9811	−0.2468	0.9100	−0.0969	3.2245	0.0143	0.0334	0.0653	0.1267	0.3079	0.6039	0.8497	0.9564	0.9822	0.9926	0.9976	95.91
		400	0.9833	−0.2089	0.9190	−0.0913	3.1989	0.0135	0.0319	0.0611	0.1203	0.2986	0.5844	0.8371	0.9500	0.9790	0.9914	0.9972	95.95
		800	0.9850	−0.1657	0.9272	−0.0883	3.1690	0.0123	0.0296	0.0578	0.1144	0.2833	0.5672	0.8226	0.9429	0.9759	0.9895	0.9967	95.98
		3200	0.9876	−0.0990	0.9434	−0.0853	3.3083	0.0116	0.0269	0.0519	0.1028	0.2621	0.5392	0.7990	0.9300	0.9681	0.9858	0.9948	95.94
		6400	0.9886	−0.0762	0.9475	−0.0799	3.4585	0.0110	0.0261	0.0497	0.0990	0.2559	0.5303	0.7915	0.9265	0.9656	0.9839	0.9941	95.87
	$c = -1$	100	0.7829	−0.1032	0.9140	−0.1395	3.5317	0.0120	0.0263	0.0498	0.0963	0.2493	0.5393	0.8140	0.9393	0.9724	0.9875	0.9955	96.12
		200	0.8113	−0.0824	0.9250	−0.1763	3.7095	0.0120	0.0259	0.0486	0.0940	0.2429	0.5285	0.8055	0.9347	0.9702	0.9859	0.9948	96.19
		400	0.8331	−0.0623	0.9308	−0.1573	4.1029	0.0111	0.0243	0.0464	0.0910	0.2404	0.5224	0.7982	0.9293	0.9673	0.9847	0.9938	96.00
		800	0.8504	−0.0505	0.9348	−0.1578	4.0927	0.0114	0.0240	0.0452	0.0896	0.2345	0.5189	0.7942	0.9284	0.9656	0.9834	0.9934	95.92
		3200	0.8761	−0.0270	0.9425	−0.1274	4.6394	0.0108	0.0230	0.0433	0.0852	0.2260	0.5093	0.7886	0.9249	0.9639	0.9818	0.9919	95.81
		6400	0.8859	−0.0269	0.9489	−0.0976	4.9272	0.0104	0.0231	0.0431	0.0858	0.2282	0.5105	0.7879	0.9224	0.9622	0.9806	0.9914	95.76
$k_n = n^{0.5}$	$c = -0.1$	100	0.9900	−0.3279	0.9186	0.0290	3.2760	0.0156	0.0376	0.0729	0.1441	0.3491	0.6473	0.8670	0.9573	0.9818	0.9917	0.9973	95.42
		200	0.9929	−0.3167	0.9164	0.0007	3.2132	0.0149	0.0367	0.0730	0.1424	0.3441	0.6398	0.8643	0.9584	0.9823	0.9923	0.9975	95.56
		400	0.9950	−0.3057	0.9169	−0.0032	3.1871	0.0146	0.0368	0.0718	0.1418	0.3377	0.6328	0.8627	0.9577	0.9819	0.9919	0.9971	95.52
		800	0.9965	−0.2795	0.9173	−0.0388	3.1272	0.0144	0.0346	0.0696	0.1377	0.3273	0.6178	0.8555	0.9565	0.9820	0.9923	0.9975	95.77
		3200	0.9982	−0.2331	0.9279	−0.0577	3.1016	0.0137	0.0340	0.0660	0.1293	0.3112	0.5959	0.8391	0.9505	0.9790	0.9911	0.9972	95.72
		6400	0.9988	−0.2067	0.9370	−0.0504	3.0939	0.0139	0.0331	0.0642	0.1255	0.3035	0.5843	0.8293	0.9457	0.9761	0.9898	0.9967	95.68
	$c = -1$	100	0.9000	−0.1602	0.9126	−0.1356	3.3504	0.0127	0.0292	0.0559	0.1075	0.2733	0.5632	0.8273	0.9463	0.9776	0.9906	0.9968	96.14
		200	0.9293	−0.1380	0.9167	−0.1410	3.3803	0.0120	0.0281	0.0537	0.1049	0.2658	0.5541	0.8200	0.9433	0.9758	0.9897	0.9966	96.16
		400	0.9500	−0.1267	0.9267	−0.1123	3.4258	0.0119	0.0273	0.0534	0.1052	0.2674	0.5498	0.8134	0.9386	0.9730	0.9884	0.9957	96.11
		800	0.9646	−0.1109	0.9327	−0.0955	3.2879	0.0118	0.0271	0.0535	0.1033	0.2620	0.5438	0.8076	0.9345	0.9709	0.9868	0.9956	95.97
		3200	0.9823	−0.0827	0.9409	−0.0848	3.4780	0.0109	0.0259	0.0501	0.0993	0.2545	0.5334	0.7966	0.9290	0.9670	0.9845	0.9943	95.87
		6400	0.9875	−0.0753	0.9462	−0.0712	3.3087	0.0107	0.0260	0.0505	0.0996	0.2535	0.5290	0.7919	0.9263	0.9654	0.9844	0.9943	95.84

Table 2 (continued)

k_n	c	n	ρ	Mean	St.D.	Skew.	Kurt.	Simulated probability $\Pr(\beta_n \le \Phi^{-1}(\tau))$											ECP
								0.01	0.025	0.05	0.1	0.25	0.5	0.75	0.9	0.95	0.975	0.99	
$k_n = n^{0.75}$	$c = -0.1$	100	0.9968	−0.3564	0.9357	0.0966	3.2005	0.0161	0.0393	0.0778	0.1545	0.3743	0.6572	0.8654	0.9560	0.9806	0.9911	0.9970	95.12
		200	0.9981	−0.3589	0.9357	0.0975	3.1315	0.0156	0.0395	0.0792	0.1563	0.3759	0.6584	0.8643	0.9566	0.9812	0.9916	0.9972	95.11
		400	0.9989	−0.3655	0.9369	0.1228	3.1358	0.0158	0.0402	0.0800	0.1586	0.3798	0.6621	0.8656	0.9563	0.9803	0.9911	0.9967	95.08
		800	0.9993	−0.3600	0.9360	0.1115	3.1008	0.0154	0.0394	0.0796	0.1581	0.3748	0.6598	0.8645	0.9557	0.9804	0.9914	0.9970	95.18
		3200	0.9998	−0.3609	0.9369	0.1134	3.1227	0.0158	0.0403	0.0805	0.1585	0.3740	0.6614	0.8645	0.9558	0.9802	0.9911	0.9969	95.14
		6400	0.9999	−0.3610	0.9383	0.1041	3.1373	0.0162	0.0417	0.0809	0.1593	0.3725	0.6605	0.8661	0.9549	0.9804	0.9911	0.9968	95.07
	$c = -1$	100	0.9684	−0.2566	0.9060	−0.0934	3.2820	0.0145	0.0331	0.0643	0.1277	0.3123	0.6084	0.8549	0.9575	0.9826	0.9928	0.9977	95.97
		200	0.9812	−0.2471	0.9100	−0.0967	3.2244	0.0143	0.0334	0.0653	0.1268	0.3080	0.6039	0.8498	0.9564	0.9821	0.9926	0.9976	95.91
		400	0.9888	−0.2428	0.9165	−0.0744	3.1779	0.0140	0.0332	0.0649	0.1280	0.3108	0.6010	0.8467	0.9541	0.9810	0.9919	0.9973	95.86
		800	0.9934	−0.2288	0.9204	−0.0764	3.1207	0.0131	0.0326	0.0640	0.1277	0.3063	0.5934	0.8413	0.9516	0.9797	0.9916	0.9974	95.89
		3200	0.9976	−0.2095	0.9301	−0.0669	3.1050	0.0133	0.0328	0.0634	0.1245	0.3022	0.5846	0.8326	0.9474	0.9779	0.9906	0.9970	95.78
		6400	0.9986	−0.1978	0.9380	−0.0528	3.0954	0.0137	0.0325	0.0631	0.1236	0.3006	0.5812	0.8265	0.9445	0.9755	0.9895	0.9967	95.70

tables show a reasonably small deviation from their corresponding nominal values (0, 1, 0 and 3 respectively) in the standard normal distribution. Third, the probability $\Pr\left(\beta_n \leq \Phi^{-1}(\tau)\right)$ is always larger than τ throughout both tables, with a deviation smaller in some cases and larger in the others under normal innovations than under $t(2)$ innovations.

(ii) If we were use the asymptotic result from Theorem 2.2 to conduct a hypothesis test on the parameter ρ, we will be more concerned on those probabilities around the lower and upper 1 %—5 % regions in the tables than the others. To see this, consider a null hypothesis $H_0 : \rho = \rho_0$ for some constant ρ_0 and an alternative hypothesis $H_1 : \rho < \rho_0$. We will reject the null hypothesis H_0, if $\beta_n \leq \Psi^{-1}(\alpha)$ is observed for a given significant level α, which is usually either 5 % or 1 %. Obviously, the probability of type one error in such a hypothesis test is $\Pr\left(\beta_n \leq \Psi^{-1}(\alpha)\right)$, which, as shown in Tables 1 and 2, is unanimously estimated to be larger than α, the expected significant level we set at the inception of the hypothesis test. If we change the alternative hypothesis H_1 to be $\rho > \rho_0$, we will then reject the null hypothesis upon a sample with $\beta_n \geq \Psi^{-1}(1-\alpha)$. In this case, the probability of type one error $\Pr\left(\beta_n \geq \Psi^{-1}(1-\alpha)\right)$ is unanimously estimated to be smaller than the expected significant level α, according to the upper 1 %—5 % regions in Tables 1 and 2.

(iii) Finally, we note that the empirical coverage probabilities displayed in the last column of the tables are all close to its nominal value of 95 %, particularly in the case with normal innovations. It implies that the limiting distribution we established in Theorem 2.2 works very well in terms of constructing confidence intervals for the parameter ρ, even under a moderate sample size of 100. It is worth noting that, although the probability $\Pr\left(\beta_n \leq \Psi^{-1}(\alpha)\right)$ deviates from α quite substantially at both $\alpha = 0.025$ and 0.975 in some panels of the tables (e.g., $k_n = n^{0.75}$), the resulting confidence interval still leads to an ECP very close to its nominal value of 95 %.

Acknowledgements The authors would like to thank the two anonymous referees for their constructive and illuminating comments which have significantly improved the present paper. Pang acknowledges funding support from the National Natural Science Foundation of China (No. 11001236), Research Fund for the Doctoral Program of Higher Education of China (No. 20090101120011) and Zhejiang Provincial Natural Science Foundation of China (No. R6090034 and No. Y6100091). Weng thanks the University of Waterloo Start-up Grant, the funding support from the Natural Sciences and Engineering Research Council of Canada, and the MOE Project of Key Research Institute of Humanities and Social Sciences at Universities (No11JJD790053).

Appendix: Lemmas and Proofs

The proofs of our main results are based on the idea of using truncated random variables. Let

$$l(x) = \mathsf{E}\left[u_1^2 I\{|u_1| \leq x\}\right], \quad b = \inf\{x \geq 1 : l(x) > 0\}, \tag{A.1}$$

and

$$\eta_j = \inf\left\{s : s \geq b+1, \frac{l(s)}{s^2} \leq \frac{1}{j}\right\}, \text{ for } j = 1, 2, 3, \cdots, \tag{A.2}$$

then it is easy to see that $nl(\eta_n) \leq \eta_n^2$ for all $n \geq 1$. In addition, we denote

$$V_n^2 = \sum_{i=1}^{n} u_i^2 \tag{A.3}$$

and

$$\begin{cases} u_t^{(1)} = u_t I\{|u_t| \leq \eta_n\} - \mathsf{E}\left[u_t I\{|u_t| \leq \eta_n\}\right], \\ u_t^{(2)} = u_t I\{|u_t| > \eta_n\} - \mathsf{E}\left[u_t I\{|u_t| > \eta_n\}\right]. \end{cases} \tag{A.4}$$

Throughout the rest of the Appendix, we denote A a positive constant whose value can be different in different places, and ρ_n always denotes $1 + c/k_n$. Moreover, we will use $[t]$ to denote the integer part of a real number t.

The following four lemmas will be useful in our proofs of the main results and the first one is due to Csörgő et al. (2003).

Lemma A.1 *Let X be a random variable. Then the following statements are equivalent:*

(a) $E\left[X^2 I\{|X| \leq x\}\right]$ *is a slowly varying function of x at ∞,*
(b) $x^2 P(|X| > x) = o(l(x))$,
(c) $xE\left[|X| I\{|X| > x\}\right] = o(l(x))$,
(d) $E\left[|X|^n I\{|X| \leq x\}\right] = o(x^{n-2} l(x))$ *for $n > 2$,*

where $l(x)$ is as defined in Eq. A.1 with u_1 replaced by X.

Lemma A.2 *Assume that conditions* **C1**, **C2** *and* **C3** *are satisfied by process 1.1 with $c < 0$, we have for each $s \in [0, 1]$, as $n \to \infty$,*

(a) $\dfrac{1}{\sqrt{nk_n} l(\eta_n)} \sum_{t=2}^{[ns]} \left(\sum_{i=1}^{t-1} \rho_n^{t-1-i} u_i^{(1)} \right) u_t^{(2)} \xrightarrow{p} 0,$

(b) $\dfrac{1}{\sqrt{nk_n} l(\eta_n)} \sum_{t=2}^{[ns]} \left(\sum_{i=1}^{t-1} \rho_n^{t-1-i} u_i^{(2)} \right) u_t^{(1)} \xrightarrow{p} 0,$

(c) $\dfrac{1}{\sqrt{nk_n} l(\eta_n)} \sum_{t=2}^{[ns]} \left(\sum_{i=1}^{t-1} \rho_n^{t-1-i} u_i^{(2)} \right) u_t^{(2)} \xrightarrow{p} 0.$

Lemma A.3 *Assume that conditions* **C1**, **C2** *and* **C3** *are satisfied by process 1.1 with $c < 0$, we have for each $s \in [0, 1]$, as $n \to \infty$,*

(a) $Y_{[ns]}^2 / V_n^2 \xrightarrow{p} 0,$

(b) $\sum_{t=1}^{[ns]} Y_{t-1} u_t / V_n^2 \xrightarrow{p} 0.$

Lemma A.4 *Assume that conditions* **C1**, **C2** *and* **C3** *are satisfied by process 1.1 with $c < 0$, we have as $n \to \infty$,*

(a) $\sum_{t=1}^{n} Y_{t-1}^2 / (k_n V_n^2) \xrightarrow{p} 1/(-2c),$

(b) $\sqrt{n} \sum_{t=1}^{n} Y_{t-1} u_t / (\sqrt{k_n} V_n^2) \xrightarrow{d} N(0, 1/(-2c)).$

Proof of Lemma A.2 Before we proceed, we first recall that $V_n^2/(nl(\eta_n)) \xrightarrow{p} 1$ as $n \to \infty$; see (3.4) in Giné et al. (1997). Now, let us consider the part (a). Obviously, it follows from Lemma A.1 that

$$\begin{aligned}
&\frac{1}{\sqrt{nk_n}l(\eta_n)} \sum_{t=2}^{[ns]} \left(\sum_{i=1}^{t-1} \rho_n^{t-1-i} u_i^{(1)} \right) u_t^{(2)} \\
&= \frac{1}{\sqrt{nk_n}l(\eta_n)} \sum_{t=2}^{[ns]} \left(\sum_{i=1}^{t-1} \rho_n^{t-1-i} u_i^{(1)} \right) u_t I\{|u_t| > \eta_n\} \\
&\quad + o\left(\frac{l(\eta_n)}{\eta_n} \right) \cdot \frac{1}{\sqrt{nk_n}l(\eta_n)} \sum_{t=2}^{[ns]} \sum_{i=1}^{t-1} \rho_n^{t-1-i} u_i^{(1)} \\
&:= I + II.
\end{aligned}$$

A direct application of Cauchy–Schwarz inequality yields

$$|I| \le \frac{V_{[ns]}}{\sqrt{nk_n}l(\eta_n)} \sqrt{ \sum_{t=2}^{[ns]} \left(\sum_{i=1}^{t-1} \rho_n^{t-1-i} u_i^{(1)} \right)^2 I\{|u_t| > \eta_n\} },$$

where $V_{[ns]}$ is as defined in Eq. A.3. Moreover, using Jensen inequality and recalling $nl(\eta_n) \le \eta_n^2$, we have

$$\begin{aligned}
&\frac{1}{\sqrt{k_n l(\eta_n)}} \cdot \mathsf{E}\left[\sqrt{ \sum_{t=2}^{[ns]} \left(\sum_{i=1}^{t-1} \rho_n^{t-1-i} u_i^{(1)} \right)^2 I\{|u_t| > \eta_n\} } \right] \\
&\le \frac{1}{\sqrt{k_n l(\eta_n)}} \sqrt{ \sum_{t=2}^{[ns]} \sum_{i=1}^{t-1} \rho_n^{2(t-1-i)} l(\eta_n)[1 + o(1)] \mathsf{P}(|u_t| > \eta_n) } \\
&\le \frac{A}{\sqrt{k_n l(\eta_n)}} \sqrt{ [ns] k_n l(\eta_n) \cdot o\left(\frac{1}{n} \right) } = o(1),
\end{aligned}$$

since $1 - \rho_n^2 = -2c/k_n[1 + O(k_n^{-1})]$, which together with $V_n^2/(nl(\eta_n)) \xrightarrow{p} 1$ as $n \to \infty$ yields $I \xrightarrow{p} 0$. Hence, we only need to show $II \xrightarrow{p} 0$ for part (a). To this end, we calculate the variance of II as follows,

$$\begin{aligned}
\mathsf{Var}(II) &= o(\frac{l^2(\eta_n)}{\eta_n^2}) \cdot \frac{1}{nk_n l^2(\eta_n)} \sum_{i=1}^{[ns]} \left(\sum_{t=i+1}^{[ns]} \rho_n^{t-1-i} \right)^2 l(\eta_n)[1 + o(1)] \\
&= o(\frac{l(\eta_n)}{n}) \cdot \frac{[ns]k_n^2}{nk_n l(\eta_n)} = o(1),
\end{aligned}$$

by which the proof of part (a) is complete.

The proof of part (b) can be achieved by a similar argument, and hence omitted. Finally, the proof of part (c) follows from Lemma A.1 as follows:

$$\frac{1}{\sqrt{nk_n}l(\eta_n)}\mathsf{E}\left[\left|\sum_{t=2}^{[ns]}\left(\sum_{i=1}^{t-1}\rho_n^{t-1-i}u_i^{(2)}\right)u_t^{(2)}\right|\right]$$
$$\leq \frac{A}{\sqrt{nk_n}l(\eta_n)}\sum_{t=2}^{[ns]}\sum_{i=1}^{t-1}\rho_n^{t-1-i}\cdot o\left(\frac{l^2(\eta_n)}{\eta_n^2}\right)$$
$$\leq \frac{A[ns]k_n}{\sqrt{nk_n}l(\eta_n)}\cdot o\left(\frac{l(\eta_n)}{n}\right)=o\left(\sqrt{\frac{k_n}{n}}\right)=o(1).$$

□

Proof of Lemma A.3 To prove part (a), obviously it is sufficient to show

$$Y_{[ns]}^2/(nl(\eta_n))=o_p(1). \tag{A.5}$$

To this end, for each $s\in[0,1]$, we write

$$\frac{Y_{[ns]}^2}{nl(\eta_n)}=\frac{1}{nl(\eta_n)}\left(\rho_n^{[ns]}Y_0+\sum_{j=1}^{[ns]}\rho_n^{[ns]-j}u_j\right)^2$$
$$\leq \frac{2}{nl(\eta_n)}\rho_n^{2[ns]}Y_0^2+\frac{2}{nl(\eta_n)}\left(\sum_{j=1}^{[ns]}\rho_n^{[ns]-j}u_j\right)^2. \tag{A.6}$$

The first term on the right hand side of Eq. A.6 is obviously equal to $o_p(k_n/n)=o_p(1)$, since $Y_0=o_p(\sqrt{k_n})$. Thus, by the second term on the right hand side of Eq. A.6, we only need to show

$$\frac{1}{\sqrt{nl(\eta_n)}}\sum_{j=1}^{[ns]}\rho_n^{[ns]-j}u_j=o_p(1). \tag{A.7}$$

To this end, we first write

$$\frac{1}{\sqrt{nl(\eta_n)}}\sum_{j=1}^{[ns]}\rho_n^{[ns]-j}u_j=\frac{1}{\sqrt{nl(\eta_n)}}\sum_{j=1}^{[ns]}\rho_n^{[ns]-j}u_j^{(1)}+\frac{1}{\sqrt{nl(\eta_n)}}\sum_{j=1}^{[ns]}\rho_n^{[ns]-j}u_j^{(2)}$$
$$:=III+IV. \tag{A.8}$$

In view of Lemma A.1, we have

$$\mathsf{E}\left[|IV|\right]\leq\frac{2}{\sqrt{nl(\eta_n)}}\cdot o\left(\frac{l(\eta_n)}{\eta_n}\right)\cdot\sum_{j=1}^{[ns]}\rho_n^{[ns]-j}$$
$$\leq\frac{Ak_n}{\sqrt{nl(\eta_n)}}\cdot o\left(\frac{l(\eta_n)}{\eta_n}\right)=o\left(\frac{k_n}{n}\right)=o(1) \tag{A.9}$$

by recalling that $nl(\eta_n) \le \eta_n^2$ for $n \ge 1$. Moreover,

$$\mathsf{Var}(III) = \frac{1}{nl(\eta_n)} \sum_{j=1}^{[ns]} \rho_n^{2([ns]-j)} \cdot l(\eta_n)[1+o(1)] = O\left(\frac{k_n}{n}\right) = o(1). \quad \text{(A.10)}$$

Combining Eqs. A.8–A.10 yields Eq. A.7. Hence, part (a) is proved.

For part (b), it suffices to show

$$\frac{\sum_{t=1}^{[ns]} Y_{t-1}u_t}{nl(\eta_n)} = o_p(1). \quad \text{(A.11)}$$

To this end, we write

$$\begin{aligned}
&\frac{\sum_{t=1}^{[ns]} Y_{t-1}u_t}{nl(\eta_n)} \\
&= \frac{1}{nl(\eta_n)} \sum_{t=2}^{[ns]} \left(\rho_n^{t-1} Y_0 + \sum_{j=1}^{t-1} \rho_n^{t-1-j} u_j\right) u_t + o_p(1) \\
&= \frac{Y_0}{\sqrt{k_n}} \sqrt{\frac{k_n}{n}} \frac{1}{\sqrt{n}l(\eta_n)} \sum_{t=2}^{[ns]} \rho_n^{t-1} u_t + \frac{1}{nl(\eta_n)} \sum_{t=2}^{[ns]} \left(\sum_{j=1}^{t-1} \rho_n^{t-1-j} u_j\right) u_t + o_p(1),
\end{aligned} \quad \text{(A.12)}$$

and further express the first term on the right hand side of Eq. A.12 as follows

$$\begin{aligned}
\frac{1}{\sqrt{n}l(\eta_n)} \sum_{t=2}^{[ns]} \rho_n^{t-1} u_t &= \frac{1}{\sqrt{n}l(\eta_n)} \sum_{t=2}^{[ns]} \rho_n^{t-1} u_t^{(1)} + \frac{1}{\sqrt{n}l(\eta_n)} \sum_{t=2}^{[ns]} \rho_n^{t-1} u_t^{(2)} \\
&:= V + VI.
\end{aligned} \quad \text{(A.13)}$$

It follows from Lemma A.1 again that

$$\mathsf{E}[|VI|] \le \frac{C}{\sqrt{n}l(\eta_n)} \cdot o\left(\frac{l(\eta_n)}{\eta_n}\right) \cdot \sum_{t=2}^{[ns]} \rho_n^{t-1} = o\left(\frac{k_n}{n\sqrt{l(\eta_n)}}\right) = o(1). \quad \text{(A.14)}$$

In addition,

$$\mathsf{Var}(V) = \frac{1}{nl^2(\eta_n)} \sum_{t=2}^{[ns]} \rho_n^{2(t-1)} l(\eta_n)[1+o(1)] = O\left(\frac{k_n}{nl(\eta_n)}\right) = o(1). \quad \text{(A.15)}$$

Combining Eqs. A.14 and A.15 with the assumptions $Y_0/\sqrt{k_n} = o_p(1)$ and $k_n = o(n)$ implies that the first term on the right hand side of Eq. A.12 converges to zero in probability as $n \to \infty$. The second term on the right hand side of Eq. A.12 can be rewritten as follows:

$$\begin{aligned}
&\frac{1}{nl(\eta_n)} \sum_{t=2}^{[ns]} \left(\sum_{j=1}^{t-1} \rho_n^{t-1-j} u_j \right) u_t \\
&= \frac{1}{nl(\eta_n)} \sum_{t=2}^{[ns]} \left(\sum_{j=1}^{t-1} \rho_n^{t-1-j} u_j^{(1)} \right) u_t^{(1)} + \frac{1}{nl(\eta_n)} \sum_{t=2}^{[ns]} \left(\sum_{j=1}^{t-1} \rho_n^{t-1-j} u_j^{(1)} \right) u_t^{(2)} \\
&\quad + \frac{1}{nl(\eta_n)} \sum_{t=2}^{[ns]} \left(\sum_{j=1}^{t-1} \rho_n^{t-1-j} u_j^{(2)} \right) u_t^{(1)} + \frac{1}{nl(\eta_n)} \sum_{t=2}^{[ns]} \left(\sum_{j=1}^{t-1} \rho_n^{t-1-j} u_j^{(2)} \right) u_t^{(2)}.
\end{aligned} \tag{A.16}$$

It follows from Lemma A.2 that all the terms except the first one on the above expression converge to zero in probability as $n \to \infty$, by noting that $k_n = o(n)$. Regarding its first term, we have

$$\begin{aligned}
&\mathsf{Var} \left(\frac{1}{nl(\eta_n)} \sum_{t=1}^{[ns]} \left(\sum_{j=1}^{t-1} \rho_n^{t-1-j} u_j^{(1)} \right) u_t^{(1)} \right) \\
&\qquad \leq \frac{A}{n^2} \sum_{t=1}^{[ns]} \sum_{j=1}^{t-1} \rho_n^{2(t-1-j)} = O\left(\frac{[ns]k_n}{n^2} \right) = o(1).
\end{aligned} \tag{A.17}$$

This implies that the second term on the right hand side of Eq. A.12 also converges to zero in probability as $n \to \infty$. Thus, the proof of part (b) is complete. □

Proof of Lemma A.4 For part (a), by squaring Eq. 1.1 and summing over $t \in \{1, \ldots, n\}$, it follows from the given conditions and Lemma A.3 that

$$\begin{aligned}
\frac{1-\rho_n^2}{V_n^2} \sum_{t=1}^{n} Y_{t-1}^2 &= \frac{Y_0^2}{V_n^2} - \frac{Y_n^2}{V_n^2} + \frac{1}{V_n^2} \sum_{t=1}^{n} u_t^2 + \frac{2\rho_n}{V_n^2} \sum_{t=1}^{n} Y_{t-1} u_t \\
&= 1 + o_p(1) + \frac{Y_0^2}{nl(\eta_n)} \frac{nl(\eta_n)}{V_n^2} \\
&= 1 + o_p(1),
\end{aligned} \tag{A.18}$$

which together with the fact $1 - \rho_n^2 = -2c/k_n[1 + O(k_n^{-1})]$ leads to the desired result.

For part (b), it suffices to show

$$\frac{1}{\sqrt{nk_n} l(\eta_n)} \sum_{t=1}^{n} Y_{t-1} u_t \xrightarrow{d} N(0, 1/(-2c)). \tag{A.19}$$

To this end, we write

$$\frac{1}{\sqrt{nk_n}l(\eta_n)}\sum_{t=1}^{n}Y_{t-1}u_t$$

$$=\frac{1}{\sqrt{nk_n}l(\eta_n)}Y_0u_1+\frac{1}{\sqrt{nk_n}l(\eta_n)}\sum_{t=2}^{n}\left(\sum_{i=1}^{t-1}\rho_n^{t-1-i}u_i+\rho_n^{t-1}Y_0\right)u_t$$

$$=o_p(1)+\frac{1}{\sqrt{nk_n}l(\eta_n)}Y_0\sum_{t=2}^{n}\rho_n^{t-1}u_t+\frac{1}{\sqrt{nk_n}l(\eta_n)}\sum_{t=2}^{n}\left(\sum_{i=1}^{t-1}\rho_n^{t-1-i}u_i\right)u_t. \tag{A.20}$$

Applying Eqs. A.13–A.15 with $s=1$, one has

$$\frac{1}{\sqrt{n}l(\eta_n)}\sum_{t=2}^{n}\rho_n^{t-1}u_t\xrightarrow{p}0, \tag{A.21}$$

which together with the assumption $Y_0=o_p(\sqrt{k_n})$ implies that the second term on the right hand side of Eq. A.20 converges to zero in probability as $n\to\infty$. As for the third term on the right hand side of Eq. A.20, we can do the decomposition as we did in Eq. A.16 as follows:

$$\frac{1}{\sqrt{nk_n}l(\eta_n)}\sum_{t=2}^{n}\left(\sum_{i=1}^{t-1}\rho_n^{t-1-i}u_i\right)u_t$$

$$=\frac{1}{\sqrt{nk_n}l(\eta_n)}\sum_{t=2}^{n}\left(\sum_{i=1}^{t-1}\rho_n^{t-1-i}u_i^{(1)}\right)u_t^{(1)}+\frac{1}{\sqrt{nk_n}l(\eta_n)}\sum_{t=2}^{n}\left(\sum_{i=1}^{t-1}\rho_n^{t-1-i}u_i^{(1)}\right)u_t^{(2)}$$

$$+\frac{1}{\sqrt{nk_n}l(\eta_n)}\sum_{t=2}^{n}\left(\sum_{i=1}^{t-1}\rho_n^{t-1-i}u_i^{(2)}\right)u_t^{(1)}+\frac{1}{\sqrt{nk_n}l(\eta_n)}\sum_{t=2}^{n}\left(\sum_{i=1}^{t-1}\rho_n^{t-1-i}u_i^{(2)}\right)u_t^{(2)}. \tag{A.22}$$

From Lemma A.2, all the terms except the first one in the above equation converge to 0 in probability. Hence, we only need to consider the first term on the right hand side of Eq. A.22. Define

$$\xi_{nt}=\frac{1}{\sqrt{nk_n}}\left(\sum_{i=1}^{t-1}\rho_n^{t-1-i}\frac{u_i^{(1)}}{\sqrt{l(\eta_n)}}\right)\frac{u_t^{(1)}}{\sqrt{l(\eta_n)}},$$

which is clearly a martingale difference array with respect to the filtration $\mathcal{F}_t \equiv \sigma(u_1, \ldots, u_t)$ for $t \geq 2$. Since the mean of ξ_{nt} is zero and the variance of $\sum_{t=2}^{n} \xi_{nt}$ is given by

$$\begin{aligned}
&\mathsf{Var}\left(\frac{1}{\sqrt{nk_n}} \sum_{t=2}^{n} \left(\sum_{i=1}^{t-1} \rho_n^{t-1-i} \frac{u_i^{(1)}}{\sqrt{l(\eta_n)}}\right) \frac{u_t^{(1)}}{\sqrt{l(\eta_n)}}\right) \\
&= \frac{1}{nk_n} \sum_{t=2}^{n} \sum_{i=1}^{t-1} \rho_n^{2(t-1-i)} [1 + o(1)] \\
&= \frac{1}{-2c} + o(1), \qquad \text{(A.23)}
\end{aligned}$$

it is sufficient to show the Lindeberg condition

$$\sum_{t=2}^{n} \mathsf{E}_{\mathcal{F}_{t-1}} \left[\xi_{nt}^2 I\{|\xi_{nt}| > \eta\}\right] = o_p(1) \;\; \text{for any } \eta > 0$$

holds for proving

$$\frac{1}{\sqrt{nk_n} l(\eta_n)} \sum_{t=2}^{n} \left(\sum_{i=1}^{t-1} \rho_n^{t-1-i} u_i^{(1)}\right) u_t^{(1)} \xrightarrow{d} N(0, 1/(-2c)), \qquad \text{(A.24)}$$

which, combined with the decomposition in Eq. A.22, will lead to the desired result of part (b). To this end, we write

$$\begin{aligned}
&\sum_{t=2}^{n} \mathsf{E}_{\mathcal{F}_{t-1}} \left[\xi_{nt}^2 I\{|\xi_{nt}| > \eta\}\right] \\
&= \frac{1}{nk_n l^2(\eta_n)} \sum_{t=2}^{n} \left(\sum_{i=1}^{t-1} \rho_n^{t-1-i} u_i^{(1)}\right)^2 \mathsf{E}_{\mathcal{F}_{t-1}} \left[\left(u_t^{(1)}\right)^2 I\{|\xi_{nt}| > \eta\}\right] \\
&\leq \max_{2 \leq t \leq n} \mathsf{E}_{\mathcal{F}_{t-1}} \left[\left(\frac{u_t^{(1)}}{\sqrt{l(\eta_n)}}\right)^2 I\{|\xi_{nt}| > \eta\}\right] \cdot \frac{1}{nk_n l(\eta_n)} \sum_{t=2}^{n} \left(\sum_{i=1}^{t-1} \rho_n^{t-1-i} u_i^{(1)}\right)^2 .
\end{aligned} \qquad \text{(A.25)}$$

Using the same argument we used in the proof of part (a), it is not difficult to show

$$\frac{1}{nk_n l(\eta_n)} \sum_{t=2}^{n} \left(\sum_{i=1}^{t-1} \rho_n^{t-1-i} u_i^{(1)}\right)^2 \xrightarrow{p} \frac{1}{-2c} \qquad \text{(A.26)}$$

by recalling again that $V_n^2/(nl(\eta_n)) \to 1$ in probability. Thus, we only need to show

$$\max_{2 \leq t \leq n} \mathsf{E}_{\mathcal{F}_{t-1}} \left[\left(\frac{u_t^{(1)}}{\sqrt{l(\eta_n)}}\right)^2 I\{|\xi_{nt}| > \eta\}\right] = o_p(1).$$

Observe that

$$I\{|\xi_{nt}| > \eta\} \leq I\left\{\frac{|u_t^{(1)}|}{\sqrt{l(\eta_n)}} \geq M\eta\right\} + I\left\{\frac{1}{\sqrt{nk_n}}|\sum_{i=1}^{t-1}\rho_n^{t-1-i}\frac{u_i^{(1)}}{\sqrt{l(\eta_n)}}| > \frac{1}{M}\right\}$$

for any fixed $M > 0$; thus, we have

$$\begin{aligned}\max_{2\leq t\leq n} \mathsf{E}_{\mathcal{F}_{t-1}}&\left[\left(\frac{u_t^{(1)}}{\sqrt{l(\eta_n)}}\right)^2 I\{|\xi_{nt}| > \eta\}\right]\\ &\leq \mathsf{E}\left[\left(\frac{u_t^{(1)}}{\sqrt{l(\eta_n)}}\right)^2 I\left\{\frac{|u_t^{(1)}|}{\sqrt{l(\eta_n)}} \geq M\eta\right\}\right]\\ &\quad + A\cdot\max_{2\leq t\leq n} I\left\{\frac{1}{\sqrt{nk_n}}\left|\sum_{i=1}^{t-1}\rho_n^{t-1-i}\frac{u_i^{(1)}}{\sqrt{l(\eta_n)}}\right| > \frac{1}{M}\right\} \qquad (A.27)\end{aligned}$$

since $\mathsf{E}(u_t^{(1)}/\sqrt{l(\eta_n)})^2 = O(1)$. Obviously, the first term on the right hand side of Eq. A.27 is $o(1)$ as $M \to \infty$. Thus, we only need to show

$$\max_{2\leq t\leq n}\frac{1}{nk_n}\left(\sum_{i=1}^{t-1}\rho_n^{t-1-i}\frac{u_i^{(1)}}{\sqrt{l(\eta_n)}}\right)^2 = o_p(1). \qquad (A.28)$$

Since the second moment of $u_i^{(1)}/\sqrt{l(\eta_n)}$ is finite, we can follow the proof of (15) in Phillips and Magdalinos (2007) and use Eq. A.26 to show Eq. A.28 easily with details omitted. □

Proof of Theorem 2.1

(a) Since

$$\sqrt{nk_n}(\hat{\rho}-\rho) = \sqrt{nk_n}\frac{\sum_{t=1}^n Y_{t-1}u_t}{\sum_{t=1}^n Y_{t-1}^2} = \frac{\sqrt{n}\sum_{t=1}^n Y_{t-1}u_t/(\sqrt{k_n}V_n^2)}{\sum_{t=1}^n Y_{t-1}^2/(k_nV_n^2)},$$

the desirable result follows immediately from Lemma A.4.

(b) Following the idea in the proof of Theorem 4.3 in Phillips and Magdalinos (2007), and using the truncation approach we applied in the proof of part (a), the proof of this part can be obtained in a similar way. The key step is to modify the definitions of X_n and Y_n in Phillips and Magdalinos (2007) as follows,

$$X_n = \frac{1}{\sqrt{k_n}V_n}\sum_{t=1}^n \rho_n^{-(n-t)-1}u_t \quad\text{and}\quad Y_n = \frac{1}{\sqrt{k_n}V_n}\sum_{t=1}^n \rho_n^{-t}u_t,$$

and to show that $(X_n, Y_n) \xrightarrow{d} (X, Y)$, as $n \to \infty$, where X and Y are two independent $N(0, 1/(2c))$ random variable. We omit the details. □

Proof of Theorem 2.2 Note that

$$\beta_n = \sqrt{\frac{n\sum_{t=1}^n Y_{t-1}^2}{\sum_{t=1}^n (Y_t - \hat{\rho}Y_{t-1})^2}}(\hat{\rho} - \rho)$$
$$= \frac{V_n}{\sqrt{\sum_{t=1}^n (Y_t - \hat{\rho}Y_{t-1})^2}} \frac{\sqrt{n}\sum_{t=1}^n Y_{t-1}u_t/(\sqrt{k_n}V_n^2)}{\sqrt{\sum_{t=1}^n Y_{t-1}^2/(k_nV_n^2)}},$$

and

$$\frac{\sqrt{n}\sum_{t=1}^n Y_{t-1}u_t/(\sqrt{k_n}V_n^2)}{\sqrt{\sum_{t=1}^n Y_{t-1}^2/(k_nV_n^2)}} \xrightarrow{d} N(0, 1)$$

by Lemma A.4. It suffices to show

$$\frac{\sum_{t=1}^n (Y_t - \hat{\rho}Y_{t-1})^2}{V_n^2} \xrightarrow{p} 1. \tag{A.29}$$

To this end, we write

$$\sum_{t=1}^n (Y_t - \hat{\rho}Y_{t-1})^2 = \sum_{t=1}^n [(Y_t - \rho Y_{t-1}) + (\rho - \hat{\rho})Y_{t-1}]^2$$
$$= V_n^2 + 2(\rho - \hat{\rho})\sum_{t=1}^n Y_{t-1}u_t + (\rho - \hat{\rho})^2\sum_{t=1}^n Y_{t-1}^2.$$

Hence, in view of the Cauchy–Schwarz inequality, we only need to prove

$$\frac{(\rho - \hat{\rho})^2\sum_{t=1}^n Y_{t-1}^2}{V_n^2} \xrightarrow{p} 0 \tag{A.30}$$

for showing Eq. A.29. Observe that

$$\frac{(\rho - \hat{\rho})^2\sum_{t=1}^n Y_{t-1}^2}{V_n^2} = \frac{1}{n}\cdot\frac{\sum_{t=1}^n Y_{t-1}^2}{k_nV_n^2}\cdot[\sqrt{nk_n}(\hat{\rho} - \rho)]^2,$$

Theorem 2.1 and Lemma A.4 guarantee that Eq. A.30 is true, and the proof is complete. □

References

Anderson TW (1959) On asymptotic distributions of estimates of parameters of stochastic difference equations. Ann Math Stat 30:676–687

Chan NH, Wei CZ (1987) Asymptotic inference for nearly non-stationary AR(1) processes. Ann Stat 15(3):1050–1063

Csörgő M, Szyszkowicz B, Wang QY (2003) Donsker's theorem for self-normalized partial sums processes. Ann Probab 31(3):1228–1240

Davis R, Resnick S (1985) More limit theory for the sample correlation function of moving averages. Stoch Process Appl 20(2):257–279

Giné E, Götze F, Mason DM (1997) When is the student *t*-statistic asymptotically standard normal? Ann Probab 25(3):1514–1531

Giraitis L, Phillips PCB (2006) Uniform limit theory for stationary autoregression. J Time Series Anal 27:51–60. CORRIGENDUM: Volume 27, Issue 6, pages i–ii, November 2006

Phillips PCB, Magdalinos T (2007) Limit theory for moderate deviations from a unit root. J Econom 136(1):115–130

Rao MM (1978) Asymptotic distribution of an estimator of the boundary parameter of an unstable process. Ann Stat 6:185–190

Wang GW (2006) A note on unit root tests with heavy-tailed GARCH errors. Stat Probab Lett 76(10):1075–1079

White JS (1958) The limiting distribution of the serial correlation coefficient in the explosive case. Ann Math Stat 29:1188–1197