

## GEOMETRIC ERGODICITY AND CONDITIONAL SELF-WEIGHTED M-ESTIMATOR OF A GRCAR(P) MODEL WITH HEAVY-TAILED ERRORS

XIAOYAN LI<sup>a</sup> JIAZHU PAN<sup>b\*</sup> AND ANCHAO SONG<sup>c</sup>

<sup>a</sup>College of Mathematics and Statistics, Chongqing University, Chongqing, 401331, China

<sup>b</sup>Department of Mathematics and Statistics, University of Strathclyde, Glasgow, G1 1XH, UK

<sup>c</sup>School of Public Health and Management, Chongqing Medical University, Chongqing, 400016, China

We establish the geometric ergodicity for general stochastic functional autoregressive (linear and nonlinear) models with heavy-tailed errors. The stationarity conditions for a generalized random coefficient autoregressive model (GRCAR( $p$ )) are presented as a corollary. And then, a conditional self-weighted M-estimator for parameters in the GRCAR( $p$ ) is proposed. The asymptotic normality of this estimator is discussed by allowing infinite variance innovations. Simulation experiments are carried out to assess the finite-sample performance of the proposed methodology and theory, and a real heavy-tailed data example is given as illustration.

Received 5 February 2022; Revised 2 January 2023; Accepted 13 January 2023

Keywords: asymptotic normality; generalized random coefficient autoregressive model; geometric ergodicity; self-weighted M-estimator; stochastic functional autoregression.

JEL. C22; C32.

MOS subject classification: 62M10; 60F05.

### 1. INTRODUCTION

Suppose that  $\{y_t, t \geq 1 - p\}$  are observations from the generalized random coefficient autoregressive (GRCAR) model with order  $p$  defined by

$$y_t = \phi_{t0} + \phi_{t1}y_{t-1} + \phi_{t2}y_{t-2} + \cdots + \phi_{tp}y_{t-p} + \varepsilon_t, \quad t = 1, 2, \dots \quad (1)$$

where  $\{(\phi'_t, \varepsilon_t)' = (\phi_{t0}, \phi_{t1}, \dots, \phi_{tp}, \varepsilon_t)', t \geq 1\}$  is a sequence of independent and identically distributed (i.i.d.) random vectors with  $E\phi_t = \phi = (\phi_0, \phi_1, \dots, \phi_p)'$ . Here, it is assumed that  $(\phi'_t, \varepsilon_t)'$  is independent of  $\mathcal{F}_{t-1} = \sigma(y_{t-1}, y_{t-2}, \dots, y_{1-p})$ . We are interested in stationarity of the above model and estimation of unknown parameter vector  $\phi = (\phi_0, \phi_1, \dots, \phi_p)'$  with its true value  $\phi_0$ . Note that, in model (1), the random coefficients are permitted to be correlated with the error process.

Model (1) with  $p = 1$  (GRCAR(1)) was first introduced by Hwang and Basawa (1998), and it includes Markovian bilinear model and random coefficient exponential autoregressive model as special cases. When  $\text{Var}(\phi_t) = \mathbf{0}$ , model (1) becomes the ordinary autoregressive (AR( $p$ )) model. There have been a lot of perfect theoretical achievements about it. For example, Ling (2005) proposed a self-weighted least absolute deviation estimator and showed its asymptotic normality. The method has been used in many references, such as Pan *et al.* (2007), Pan and Chen (2013), Pan *et al.* (2015). Wang and Hu (2017) proposed a self-weighted M-estimator for the AR( $p$ ) model and established the asymptotic normality of this estimator. When  $\text{Var}(\phi_t) \neq \mathbf{0}$  with  $p = 1$ , and  $\phi_t$  is independent

\* Correspondence to: Jiazhu Pan, Department of Mathematics and Statistics, University of Strathclyde, 26 Richmond Street, Glasgow G1 1XH, UK.  
Email: jiazhu.pan@strath.ac.uk

of  $\varepsilon_t$ , model (1) becomes the first-order random coefficient autoregressive (RCAR(1)) model (see Nicholls and Quinn (1982)), which has been frequently used to describe the random perturbations of dynamical systems in economics and biology (see Tong (1990), Yu *et al.* (2011), Zhang *et al.* (2015), Araveeporn (2017)). As a generalization of RCAR model and AR models, GRCAR model has become one type of important models in nonlinear time series, since it allows dependence between random errors and random coefficients. Estimation of parameters and asymptotic properties of GRCAR models have been studied in the literature. For instance, Hwang and Basawa (1997) established the local asymptotic normality of a class of GRCAR models. Zhao and Wang (2012) constructed confidence regions for the parameters by using empirical likelihood method. Zhao *et al.* (2013) considered the problem of testing the constancy of coefficients in a GRCAR model by empirical likelihood method. Zhao *et al.* (2018) studied the variable selection problem in GRCAR models. Zhao *et al.* (2019) proposed a weighted least squares estimate and empirical likelihood (EL) based weights through using some auxiliary information for GRCAR models. Moreover, time series models with heavy-tailed errors, even when  $E(\varepsilon_t^2)$  is infinite, are often found and studied in economic and financial modeling. Wu (2013) studied M-estimation for general ARMA processes with infinite variance. Yang and Ling (2017) investigated the self-weighted least absolute deviation estimation for heavy-tailed threshold autoregressive models. Fu *et al.* (2021) studied the asymptotic properties for the conditional self-weighted M-estimator of GRCAR(1) model with possibly heavy-tailed errors. However, general easy-to-check conditions for stationarity and limiting distributions of robust parameter estimators for statistical inference of GRCAR( $p$ ) with heavy-tailed errors are still open problems.

This article aims to reach two targets. First, we establish the geometric ergodicity of general stochastic functional autoregressive (linear and nonlinear) models with possibly heavy-tailed error terms under a mild moment condition. Moreover, the stationarity conditions for GRCAR( $p$ ) are implied as a corollary of our general result. Second, motivated by Yang and Ling (2017), Ling (2005), Wang and Hu (2017) and Fu *et al.* (2021), we prove the asymptotic property of a self-weighted M-estimator (SM-estimator) for GRCAR( $p$ ) with possible infinite variance, and show that the limiting distribution of SM-estimator is asymptotically normal. Simulation results and a real data example are given to support our methodology.

The contents of this article are organized as follows. Section 2 presents the main results. Section 3 reports the simulation results. Section 4 shows a real data example. All proofs of our main results are given in Section 5.

## 2. MAIN RESULTS

### 2.1. Geometric Ergodicity

We first establish geometric ergodicity of general stochastic functional autoregressive models (including linear and nonlinear) under a mild moment condition. Then the geometric ergodicity which can imply the stationarity conditions of model (1) is given as a corollary of the main theorem.

Consider a general stochastic functional autoregressive model defined as follows

$$\begin{cases} y_t = \varphi_t(y_{t-1}, \dots, y_{t-p}) + \varepsilon_t, & t \geq 1 \\ (y_0, y_{-1}, \dots, y_{-p+1})' \in \mathcal{R}^p, \end{cases} \tag{2}$$

where  $\{\varphi_t\}$  is a sequence of i.i.d. stochastic functions such that  $\{(\varphi_t(\cdot), \varepsilon_t)'\}$  being of i.i.d. random vectors, and  $(\varphi_t(\cdot), \varepsilon_t)'$  is independent of  $\mathcal{F}_{t-1} = \sigma(y_{t-1}, y_{t-2}, \dots, y_{1-p})$ . It can be seen that both linear and nonlinear autoregressive models are included. This model can be rewritten in vector form as follows:

$$\begin{cases} X_t = \Phi_t(X_{t-1}) + \varepsilon_t U, & t \geq 1, \\ X_0 \in \mathcal{R}^p, \end{cases} \tag{3}$$

where  $\Phi_t(X_{t-1}) = (\varphi_t(y_{t-1}, \dots, y_{t-p}), y_{t-1}, \dots, y_{t-p+1})'$ ,  $X_t = (y_t, \dots, y_{t-p+1})'$ ,  $U = (1, 0, \dots, 0)'$  Under the above conditions, model (3) is a homogeneous Markov chain. It is easily seen that stationarity for  $\{y_t, t \geq 1-p\}$  is

equivalent to that for  $\{X_t, t \geq 0\}$ . Furthermore, the geometric ergodicity for model (2) is equivalent to that for the model (3).

**Theorem 2.1.** Suppose model (3) satisfies

- (i) There exists some norm  $\|\cdot\|_v$  on the  $p$ -dimensional vector space, constants  $0 < \rho < 1, 0 < \delta < 1$  and  $c \geq 0$ , such that

$$E\|\Phi_t(\mathbf{x})\|_v^\delta \leq \rho\|\mathbf{x}\|_v^\delta + c, \forall \mathbf{x} \in \mathcal{R}^p; \quad (4)$$

- (ii) The density function of  $\varepsilon_t$  is continuous and positive everywhere, and  $E|\varepsilon_t|^\delta < \infty$  for  $\delta$  in (i).

Then, model (3) is geometrically ergodic, which implies  $\{y_t\}$  in model (2) is stationary and geometrically ergodic.

Under the conditions of Theorem 2.1, more concretely, we have the following corollary.

**Corollary 2.2.** Suppose model (2) satisfies

- (i) There exists a constant vector  $\varphi = (\varphi_1, \dots, \varphi_p)'$  and a  $0 < \delta < 1$  satisfying

$$1 - \varphi_1 z - \dots - \varphi_p z^p \neq 0, \quad |z| \leq 1 \quad (5)$$

such that

$$\lim_{\|\mathbf{x}\| \rightarrow \infty} \frac{E|\varphi_t(\mathbf{x}) - \varphi' \mathbf{x}|^\delta}{\|\mathbf{x}\|} = 0 \quad (6)$$

and for any  $K > 0$ ,

$$\sup_{\|\mathbf{x}\| \leq K} E|\varphi_t(\mathbf{x}) - \varphi' \mathbf{x}|^\delta < \infty. \quad (7)$$

- (ii) The density function of  $\varepsilon_t$  is continuous and positive everywhere, and  $E|\varepsilon_t|^\delta < \infty$  for  $\delta$  in (i).

Then,  $\{y_t\}$  in model (2) is stationary and geometrically ergodic.

In Corollary 2.2, when  $\varphi_t(\mathbf{x}) = \phi_{t0} + (\phi_{t1}, \dots, \phi_{tp})' \mathbf{x}, \mathbf{x} \in \mathcal{R}^p$ , we can get the stationarity conditions for GRCAR( $p$ ) as another corollary of our general result.

**Corollary 2.3.** Suppose model (1) satisfies

(C.1)

- (i)  $1 - \phi_1 z - \dots - \phi_p z^p \neq 0$ , for  $|z| \leq 1$ ;  
 (ii)  $E|\phi_{it} - \phi_i|^\delta < \infty, i = 1, 2, \dots, p, E|\phi_{t0}|^\delta < \infty$ , for a constant  $0 < \delta < 1$ ;  
 (iii) The density function of  $\varepsilon_t$  is continuous and positive everywhere, and  $E|\varepsilon_t|^\delta < \infty$  for  $\delta$  in (ii).  
 Then,  $\{y_t\}$  in model (1) is stationary and geometrically ergodic.

**Remark 1.** Theorem 2.1 establishes the geometric ergodicity of general stochastic functional autoregressive (linear and nonlinear) models with possibly heavy-tailed error terms under a mild moment condition. The stationarity conditions for GRCAR( $p$ ) in Corollary 2.3 is a consequence of Theorem 2.1. In Corollary 2.3, the moment condition is very weak. We only require a finite moment of order  $\delta (0 < \delta < 1)$  about the error  $\varepsilon_t$ , which includes the Cauchy distribution. The condition on random coefficients makes the model not too far away from the linear AR. This is a reasonable requirement for any (non-parametric or parametric) AR-type model.

**Remark 2.** We note that the GRCAR is a quite broad kind of models for time series data. A special case of this type of models can be used to describe conditionally heteroscedastic structure. For example, consider the model

defined as:  $y_t = (\phi_0 + \beta_0 \varepsilon_t) + (\phi_1 + \beta_1 \varepsilon_t)y_{t-1} + \varepsilon_t$ ,  $\varepsilon_t \sim N(0, \sigma^2)$ . The conditional mean and conditional variance of this model are

$$E(y_t | \mathcal{F}_{t-1}) = \phi_0 + \phi_1 y_{t-1};$$

$$Var(y_t | \mathcal{F}_{t-1}) = \sigma^2(1 + \beta_0 + \beta_1 y_{t-1})^2.$$

It is shown that this very special case of the GRCAR has similar function to AR(1)–ARCH(1), but we don't need to restrict the parameters  $\beta_0$  and  $\beta_1$  to be non-negative. Furthermore, if the assumption on the distribution of error is changed to  $\varepsilon_t \sim t_2$ , the model becomes a model with infinite variance.

### 2.2. Conditional Self-weighted M-estimation

Denote  $\mathbf{X}_{t-1} = (1, y_{t-1}, \dots, y_{t-p})'$ . Then the model (1) becomes  $y_t = \boldsymbol{\phi}'_t \mathbf{X}_{t-1} + \varepsilon_t$ , where  $\boldsymbol{\phi}_t = (\phi_{t0}, \phi_{t1}, \dots, \phi_{tp})'$ . Define the objective function

$$L_n(\boldsymbol{\phi}) = \sum_{t=1}^n \omega_t \rho(y_t - \boldsymbol{\phi}'_t \mathbf{X}_{t-1}), \tag{8}$$

where  $\boldsymbol{\phi} = E(\boldsymbol{\phi}_t) = (\phi_0, \phi_1, \dots, \phi_p)'$  and  $\omega_t$  is a positive function which is measurable to  $\mathcal{F}_{t-1} = \sigma(y_{t-1}, y_{t-2}, \dots)$  and  $\rho(\cdot)$  is assumed to be a suitable nonnegative convex function. The conditional self-weighted M-estimator  $\hat{\boldsymbol{\phi}}_{SM}$  of  $\boldsymbol{\phi}$  is defined by

$$\hat{\boldsymbol{\phi}}_{SM} = \arg \min_{\boldsymbol{\phi} \in \Theta} L_n(\boldsymbol{\phi})$$

where  $\Theta \subset \mathcal{R}^{p+1}$  is the parameter space containing the true value  $\boldsymbol{\phi}_0$ .

### 2.3. Asymptotic Normality of SM-estimation

To derive the asymptotic property of  $\hat{\boldsymbol{\phi}}_{SM}$ , we need the following assumptions:

- (C.2) Let  $\rho(\cdot)$  be a convex function on  $\mathbb{R}$  with left derivative  $\psi_-$  and right derivative  $\psi_+$ . Choose a function  $\psi$  such that  $\psi_- \leq \psi \leq \psi_+$ .
- (C.3) Suppose that  $G(t) := E\psi(\varepsilon_1 + t)$  exists,  $G(t)$  has a derivative  $\lambda > 0$  at  $t = 0$  and  $G(0) = 0$ .
- (C.4)  $E\psi^2(\varepsilon_1) = \tau < \infty$  and  $E(\psi(\varepsilon_1 + t) - \psi(\varepsilon_1))^2 \rightarrow 0$ , as  $t \rightarrow 0$ .
- (C.5)  $\omega_t = g(y_{t-1}, \dots, y_{t-p})$  is a measurable and positive function on  $\mathbb{R}^p$  such that  $E(\omega_t + \omega_t^2)(\|\mathbf{X}_t\| + \|\mathbf{X}_t\|^2) < \infty$ , where  $\|\mathbf{v}\|$  denotes the Euclidean norm of a vector  $\mathbf{v}$ .

**Theorem 2.4.** Under the (C.1)–(C.5), we have

$$\sqrt{n}(\hat{\boldsymbol{\phi}}_{SM} - \boldsymbol{\phi}_0) \xrightarrow{L} N\left(0, \frac{\tau}{\lambda^2} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Omega} \boldsymbol{\Sigma}^{-1}\right) \tag{9}$$

where  $\boldsymbol{\Sigma} = E(\omega_t \mathbf{X}_t \mathbf{X}'_t)$ ,  $\boldsymbol{\Omega} = E(\omega_t^2 \mathbf{X}_t \mathbf{X}'_t)$  and  $\xrightarrow{L}$  denotes convergence in distribution.

**Remark 3.** Assumption (C.1) does not rule out the possibility that  $\varepsilon_t$  has an infinite variance, and even  $E|\varepsilon_t|$  is infinite. Theorem 2.4 establishes the asymptotic property of SM-estimators for parameters in GRAR( $p$ ) models with possible heavy-tailed errors.

**Remark 4.** It is worth mentioning that Assumptions (C.2)–(C.4) are traditional assumptions for an M-estimation in a linear model, which can be found in many references, for examples, Bai *et al.* (1992), Wu (2007) and Wang

and Zhu (2018). Examples of  $\rho(x)$  satisfying assumptions include  $\rho(x) = x^2$ ,  $\rho(x) = |x|$  and  $\rho(x) = \frac{1}{2}x^2I(|x| \leq m) + (m|x| - \frac{1}{2}m^2)I(|x| > m)$ , which then correspond to the conditional self-weighted least-squares estimator, conditional self-weighted least absolute deviation estimator and conditional self-weighted Huber estimator respectively. Assumption (C.5) is standard on the weight  $\omega_t$  for the self-weighted method in IVAR models which allows  $Ey_t^2$  to be infinite by properly choosing weight function  $\omega_t$ . Firstly, the purpose of the weight  $\omega_t$  is to downweight the leverage points in  $X_t$  such that the covariance matrices  $\mathbf{\Omega}$  and  $\mathbf{\Sigma}$  in Theorem 2.4 above are finite. Secondly, the  $\omega_t$  allow us to approximate  $L_n(\hat{\phi}_n)$  by a quadratic form. In addition, Theorem 2.4 generalizes the results of Ling (2005), Wang and Hu (2017) and Fu *et al.* (2021).

**Remark 5.** For the case  $\rho(x) = x^2$  and  $E\varepsilon_1^2 = \sigma^2 < \infty$ , taking  $\psi(x) = 2x$ ,  $\lambda = 2$ . Applying Theorem 2.4, we have

$$\sqrt{n}(\hat{\phi}_{SM} - \phi_0) \xrightarrow{L} N(0, \sigma^2 \mathbf{\Sigma}^{-1} \mathbf{\Omega} \mathbf{\Sigma}^{-1}) \quad (10)$$

which shows the asymptotic property of the conditional self-weighted least-squares estimator of parameters in a GRCAR( $p$ ) model with finite variance.

For the case  $\rho(x) = |x|$  and  $E\varepsilon_1^2 = \infty$ , taking  $\psi(x) = \text{sign}(x)$ . Suppose that the errors  $\varepsilon_t$  have zero median with a density  $f(x)$  satisfying  $\sup_{x \in \mathbb{R}} |f'(x)| < \infty$ . Then  $\lambda = 2f(0)$  and  $\tau = 1$ . Using Theorem 2.4, we have

$$\sqrt{n}(\hat{\phi}_{SM} - \phi_0) \xrightarrow{L} N\left(0, \frac{1}{4f(0)^2} \mathbf{\Sigma}^{-1} \mathbf{\Omega} \mathbf{\Sigma}^{-1}\right) \quad (11)$$

which shows the asymptotic property of the conditional self-weighted least absolute deviation estimator for a GRCAR( $p$ ) model with infinite variance.

For the case  $\rho(x) = \frac{1}{2}x^2I(|x| \leq m) + (m|x| - \frac{1}{2}m^2)I(|x| > m)$ , taking  $\psi(x) = -mI(x < -m) + xI(|x| \leq m) + mI(x > m)$ , and

$$\lambda = \int_{-m}^m dF(x), \tau = m^2 - \int_{-m}^m (m^2 - x^2)dF(x). \quad (12)$$

Using Theorem 2.4, we have

$$\sqrt{n}(\hat{\phi}_{SM} - \phi_0) \xrightarrow{L} N\left(0, \frac{\tau}{\lambda^2} \mathbf{\Sigma}^{-1} \mathbf{\Omega} \mathbf{\Sigma}^{-1}\right) \quad (13)$$

which includes the asymptotic property of the conditional self-weighted Huber estimator for a GRCAR( $p$ ) model with finite variance or infinite variance.

### 3. SIMULATION STUDIES

We conduct some simulation studies in finite samples through Monte Carlo experiments. What we are interested in are the accuracy and sampling distribution of the proposed estimator. The results show that our method performs well.

Data are generated from the following GRCAR ( $p$ ) models:

Model A:  $y_t = (\phi_0 + 0.1\varepsilon_t) + (\phi_1 + 0.1\varepsilon_t)y_{t-1} + \varepsilon_t$ . Consider  $\varepsilon_t \sim N(0, 1)$  and  $\varepsilon_t \sim t_2$ .

Model B:  $y_t = (\phi_0 + 0.1u_t) + (\phi_1 + 0.1u_t)y_{t-1} + \varepsilon_t$ , where  $u_t \sim N(0, 1)$ . Consider  $\varepsilon_t \sim N(0, 1)$ ,  $\varepsilon_t \sim t_2$  and  $\varepsilon_t \sim \text{Cauchy}(0, 1)$ .

The true values of parameters in model A are  $(\phi_0, \phi_1) = (0, 0.5)$ , and  $(\phi_0, \phi_1) = (0, -0.5)$  in model B. In model A, the random coefficients are correlated with the error process. The random coefficients are independent

Table I. Bias, SDs and ADs of the SM estimators for model A

			$\varepsilon_t \sim N(0, 1)$				$\varepsilon_t \sim t_2$			
			$\hat{\phi}_0^{SM_1}$	$\hat{\phi}_0^{SM_2}$	$\hat{\phi}_1^{SM_1}$	$\hat{\phi}_1^{SM_2}$	$\hat{\phi}_0^{SM_1}$	$\hat{\phi}_0^{SM_2}$	$\hat{\phi}_1^{SM_1}$	$\hat{\phi}_1^{SM_2}$
$w_{t1}$	$n = 200$	Bias	0.000	0.000	-0.006	-0.005	0.001	0.001	-0.004	-0.005
		SD	0.103	0.084	0.093	0.077	0.123	0.118	0.066	0.065
		AD	0.116	0.075	0.103	0.067	0.137	0.077	0.076	0.043
	$n = 400$	Bias	0.001	0.001	-0.003	-0.002	0.001	0.000	-0.002	-0.003
		SD	0.072	0.059	0.065	0.053	0.086	0.082	0.044	0.043
		AD	0.079	0.053	0.070	0.047	0.093	0.054	0.051	0.030
$w_{t2}$	$n = 200$	Bias	0.001	0.001	-0.002	0.000	0.000	0.002	-0.008	-0.007
		SD	0.111	0.090	0.090	0.111	0.132	0.125	0.105	0.099
		AD	0.125	0.081	0.151	0.098	0.114	0.082	0.076	0.064
	$n = 400$	Bias	0.002	0.002	0.001	0.001	0.002	0.002	-0.003	-0.002
		SD	0.078	0.064	0.064	0.076	0.092	0.086	0.044	0.065
		AD	0.086	0.057	0.101	0.068	0.076	0.057	0.051	0.044
$w_{t3}$	$n = 200$	Bias	-0.001	-0.001	-0.003	-0.004	0.000	0.001	-0.004	-0.005
		SD	0.110	0.089	0.095	0.078	0.141	0.134	0.078	0.075
		AD	0.126	0.081	0.107	0.069	0.162	0.090	0.090	0.050
	$n = 400$	Bias	0.000	0.001	-0.002	-0.002	-0.001	0.000	-0.002	-0.002
		SD	0.078	0.064	0.067	0.055	0.099	0.093	0.053	0.052
		AD	0.086	0.057	0.073	0.049	0.110	0.063	0.061	0.035
$w_{t4}$	$n = 200$	Bias	-0.002	-0.001	-0.004	-0.004	-0.001	0.002	-0.004	-0.005
		SD	0.118	0.096	0.094	0.078	0.152	0.144	0.072	0.070
		AD	0.135	0.087	0.106	0.069	0.174	0.096	0.082	0.046
	$n = 400$	Bias	0.000	0.001	-0.002	-0.002	-0.001	0.000	-0.002	-0.002
		SD	0.083	0.068	0.066	0.054	0.106	0.100	0.048	0.048
		AD	0.093	0.061	0.073	0.048	0.119	0.068	0.056	0.032

of the error process in model B. Here three distributions are given for reference:  $N(0, 1)$  has finite expectation and finite variance;  $t_2$  has finite expectation but infinite variance; Cauchy(0,1) only have a finite moment of order  $\delta(0 < \delta < 1)$ . We set the sample sizes  $n = 200$  and  $n = 400$ . The number of replications is 2000.

Tables I and II list the biases, standard deviations (SD) and asymptotic standard deviations (ADs) of the conditional self-weighted least absolute deviation estimator ( $SM_1$ ) and conditional self-weighted Huber estimator ( $SM_2$ ) with the following choice of weight functions respectively:

$$w_{1t} = \begin{cases} 1, & |y_{t-1}| \leq K, \\ \frac{K^3}{|y_{t-1}|^3}, & |y_{t-1}| > K, \end{cases}$$

$$w_{2t} = I\{|y_{t-1}| \leq K\},$$

$$w_{3t} = \frac{1}{1 + y_{t-1}^2},$$

$$w_{4t} = \frac{1}{(1 + y_{t-1})^2}$$

where  $K$  is the 0.9 quantile of data  $\{|y_1|, \dots, |y_n|\}$ . The weight  $w_{1t}$  is similar to Ling (2005). The weights  $w_{2t}$ ,  $w_{3t}$  and  $w_{4t}$  were considered by Yang and Ling (2017). The tuning parameter of Huber estimator is taken as  $m = 1.5$ .

Table II. Bias, SDs and ADs of the  $SM$  estimators for model B

			$\varepsilon_t \sim N(0, 1)$				$\varepsilon_t \sim t_2$				$\varepsilon_t \sim Cauchy(0, 1)$			
			$\hat{\phi}_0^{SM_1}$	$\hat{\phi}_0^{SM_2}$	$\hat{\phi}_1^{SM_1}$	$\hat{\phi}_1^{SM_2}$	$\hat{\phi}_0^{SM_1}$	$\hat{\phi}_0^{SM_2}$	$\hat{\phi}_1^{SM_1}$	$\hat{\phi}_1^{SM_2}$	$\hat{\phi}_0^{SM_1}$	$\hat{\phi}_0^{SM_2}$	$\hat{\phi}_1^{SM_1}$	$\hat{\phi}_1^{SM_2}$
$w_{11}$	$n = 200$	Bias	0.001	-0.001	0.006	0.006	0.001	0.000	0.005	0.004	0.000	0.000	0.003	0.003
		SD	0.095	0.075	0.092	0.076	0.110	0.106	0.068	0.064	0.127	0.133	0.045	0.044
		AD	0.109	0.075	0.106	0.073	0.131	0.076	0.079	0.046	0.156	0.077	0.045	0.022
	$n = 400$	Bias	0.000	0.001	0.003	0.003	0.000	0.001	0.002	0.002	-0.002	-0.002	0.002	0.002
		SD	0.066	0.053	0.065	0.053	0.078	0.075	0.048	0.045	0.090	0.095	0.031	0.031
		AD	0.074	0.053	0.072	0.051	0.089	0.054	0.054	0.032	0.106	0.055	0.031	0.015
$w_{12}$	$n = 200$	Bias	0.002	0.000	0.006	0.006	-0.002	0.000	0.005	0.004	-0.001	-0.001	0.005	0.005
		SD	0.102	0.081	0.136	0.110	0.117	0.112	0.097	0.093	0.133	0.138	0.064	0.064
		AD	0.118	0.081	0.155	0.107	0.139	0.081	0.117	0.069	0.161	0.081	0.070	0.035
	$n = 400$	Bias	-0.001	0.001	0.005	0.004	0.000	0.001	0.003	0.003	-0.002	-0.002	0.003	0.003
		SD	0.071	0.058	0.095	0.076	0.083	0.079	0.070	0.067	0.093	0.100	0.044	0.045
		AD	0.080	0.057	0.105	0.074	0.094	0.057	0.080	0.048	0.110	0.057	0.048	0.024
$w_{13}$	$n = 200$	Bias	0.000	-0.001	0.006	0.005	-0.002	-0.001	0.004	0.002	0.000	0.000	0.004	0.005
		SD	0.101	0.079	0.094	0.077	0.125	0.121	0.079	0.073	0.164	0.175	0.076	0.073
		AD	0.116	0.079	0.107	0.074	0.150	0.087	0.089	0.052	0.200	0.102	0.072	0.036
	$n = 400$	Bias	-0.001	0.001	0.002	0.002	-0.003	-0.001	0.001	0.001	0.000	0.000	0.002	0.002
		SD	0.070	0.057	0.065	0.053	0.090	0.086	0.055	0.051	0.115	0.124	0.051	0.050
		AD	0.079	0.056	0.073	0.052	0.102	0.062	0.061	0.037	0.136	0.072	0.049	0.026
$w_{14}$	$n = 200$	Bias	0.000	-0.001	0.006	0.005	-0.003	-0.001	0.003	0.002	0.001	0.001	0.004	0.005
		SD	0.108	0.085	0.094	0.077	0.136	0.131	0.074	0.068	0.176	0.188	0.066	0.062
		AD	0.125	0.085	0.107	0.074	0.161	0.094	0.082	0.048	0.215	0.109	0.060	0.030
	$n = 400$	Bias	-0.001	0.001	0.002	0.002	-0.004	-0.002	0.001	0.001	0.001	0.001	0.002	0.002
		SD	0.075	0.061	0.065	0.053	0.097	0.092	0.051	0.048	0.125	0.134	0.044	0.043
		AD	0.085	0.060	0.073	0.052	0.110	0.067	0.056	0.034	0.146	0.077	0.041	0.021

We define

$$\begin{aligned} \hat{\Sigma} &= \frac{1}{n} \sum_{t=1}^n \omega_t X_t X_t', \hat{\Omega} = \frac{1}{n} \sum_{t=1}^n \omega_t^2 X_t X_t', \\ \hat{\tau} &= \frac{1}{n} \sum_{t=1}^n \psi^2(\hat{\varepsilon}_t), \hat{G}(r) = \frac{1}{n} \sum_{t=1}^n \psi(\hat{\varepsilon}_t + r), \end{aligned} \tag{14}$$

$\hat{\lambda}$  is the derivative of  $\hat{G}(r)$  at  $r = 0$  and  $\{\hat{\varepsilon}_t\}$  is the sequence of residuals in  $GRCAR(p)$ . The ADs are calculated by (10)-(14). We estimate  $f(0)$  by

$$\hat{f}_n(0) = \frac{1}{\hat{\sigma}_\omega b_n n} \sum_{t=1}^n \omega_t K\left(\frac{y_t - \hat{\phi}_n' X_{t-1}}{b_n}\right)$$

where  $\hat{\sigma}_\omega = \frac{1}{n} \sum_{t=1}^n \omega_t$ ,  $K(x) = e^{-x}/(1 + e^{-x})^2$  and  $b_n = 1.06 \times n^{-1/5}$ . For the choice of the optimal bandwidth and its motivation, we refer to Silverman (1986,p.40) and Pan *et al.* (2007). Tables I and II show that all the biases are very small and all the SDs and ADs are very close no matter  $E\varepsilon_t^2$  is finite or infinite and no matter the random coefficients are correlated with the error process or independent of the error process. All the biases, SDs and ADs become smaller, when  $n$  increases from 200 to 400. And the  $SM_2$  estimators perform better than the  $SM_1$  estimators. All of the estimators based on  $w_{1t}$  are more efficient than the others.

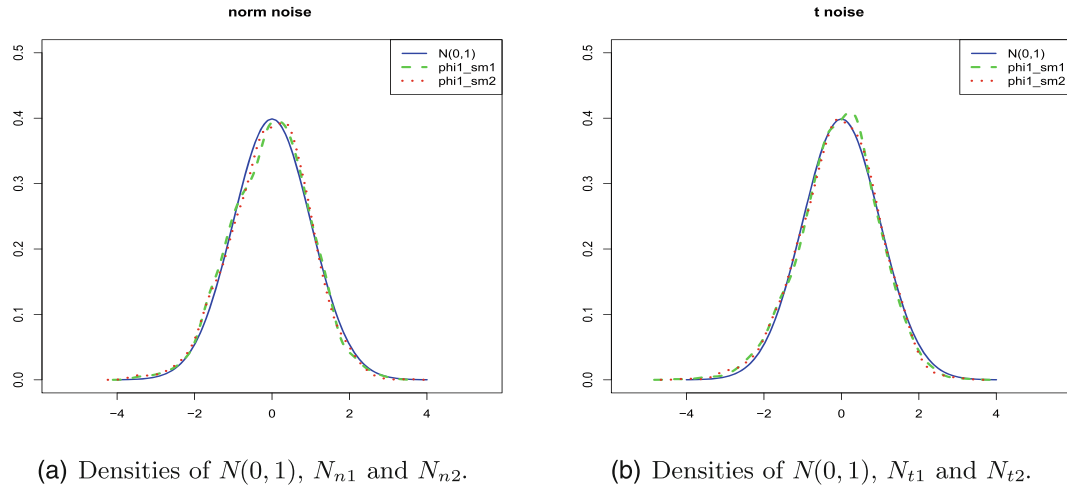


Figure 1. The sampling distribution for model A

To get an overall view on the sampling distributions of the  $SM_1$  estimators and the  $SM_2$  estimators, we simulate 2000 replications for the case  $\phi_1 = 0.5$  and  $n = 400$  when the error distributions are  $N(0, 1)$  and  $t_2$  for model A, and for the case  $\phi_1 = -0.5$  and  $n = 400$  when the error distributions are  $N(0, 1)$ ,  $t_2$  and Cauchy for model B. Denote  $N_{n1} = \sqrt{n}(\hat{\phi}_1^{SM_1} - 0.5)/\hat{\sigma}_{SM_1}$ ,  $N_{n2} = \sqrt{n}(\hat{\phi}_1^{SM_2} - 0.5)/\hat{\sigma}_{SM_2}$ ;  $N_{t1} = \sqrt{n}(\hat{\phi}_1^{SM_1} - 0.5)/\hat{\sigma}_{SM_1}$ ,  $N_{t2} = \sqrt{n}(\hat{\phi}_1^{SM_2} - 0.5)/\hat{\sigma}_{SM_2}$ ;  $N_{c1} = \sqrt{n}(\hat{\phi}_1^{SM_1} - 0.5)/\hat{\sigma}_{SM_1}$ ,  $N_{c2} = \sqrt{n}(\hat{\phi}_1^{SM_2} - 0.5)/\hat{\sigma}_{SM_2}$ , when the error distribution is  $N(0, 1)$ ,  $t_2$  and Cauchy respectively, where  $\hat{\sigma}_{SM_1}$  and  $\hat{\sigma}_{SM_2}$  are the SDs of  $\hat{\phi}_1^{SM_1}$  and  $\hat{\phi}_1^{SM_2}$  respectively. Figure 1 shows the density curves of model A. The density curves of model B are presented in Figure 2. We can see that the density of  $N(0, 1)$  is approximated reasonably well by those of  $N_{n1}, N_{n2}, N_{t1}, N_{t2}$  and  $N_{c1}, N_{c2}$  in both model A and model B.

In conclusion, the numeric results show that the conditional self-weighted M-estimators perform well in finite sample no matter with finite variance or infinite variance.

#### 4. REAL DATA ANALYSIS

The proposed methodology is applied to modeling a real dataset. We consider the Hang Seng Index (HSI) for the stock market which is the most influential index in the Hong Kong stock market and one of the most important indices in the Asian financial markets as well. This index has been extensively investigated in the literature. Our dataset consists of the daily Hang Seng closing index from 7 May 2020 to 31 December 2021, which was downloaded at <https://cn.investing.com/>. There are 412 available observations in total, which are denoted by  $x_1, x_2, \dots, x_{412}$ . The first 392 data are selected as the training sample to build model, and next 20 data are used as test sample to evaluate the model. We take the following steps to analyze this dataset by the GRCAR model and method proposed in this article.

**Step 1. Data transformation:** The sample time plot for the data  $\{x_t\}$  is shown in Figure 3(a). It can be seen that the time series is not stationary because of various levels. To get a stationary time series, let  $y_t = 100 \times (\log(x_t/x_{t-1}))$ . The sample path plot for the data  $\{y_t\}$  is shown in Figure 3(b). Figure 3(b) indicates that  $\{y_t\}$  is close to stationarity.

**Step 2. Model identification:** The plot of sample autocorrelation function (ACF) and sample partial autocorrelation function (PACF) of  $\{y_t\}$  are presented in Figure 3(c) and (d) respectively. Figure 3 can provide some important information for tentative identification of the orders of a stable AR model. Based on the sample ACF and PACF plots, it is reasonable to consider fitting an AR(3) autocorrelation structure to  $\{y_t\}$ . Since stock data are affected by various factors, the coefficients of the autoregressive model may also change randomly over



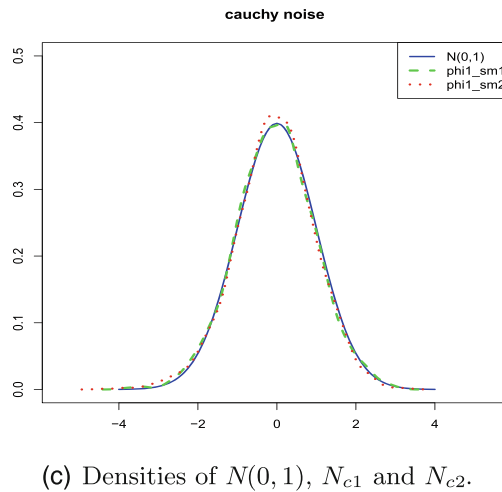
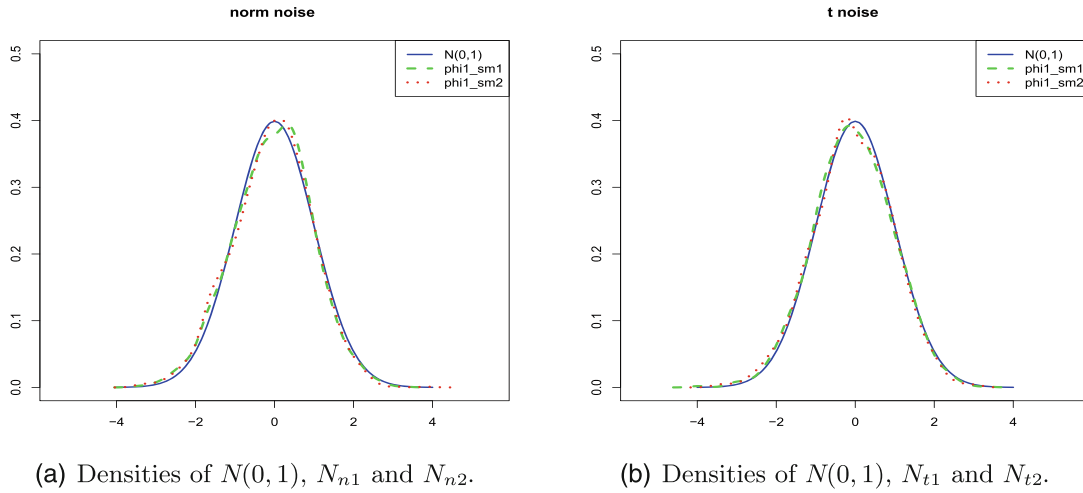


Figure 2. The sampling distribution for model B

time, even the coefficients may be correlated with the error, so we can try to fit a GRCAR(3) model instead of AR(3). However, we realize that the data may be heavy-tailed and how to determine the order of autoregression for a time series with infinite variance is a problem which needs further study. Here we just use the sample PACF to give a rough indication of the order in which a GRCAR model might be fitted.

**Step 3. Heavy-tail test:** To test whether  $\{y_t\}$  has a heavy-tailed distribution, we use Hill's estimator (see Drees *et al.* (2000) and Resnick (2000)) to estimate the tail index of  $y_t$ . Let  $y_{(1)} > y_{(2)} > \dots > y_{(n)}$  be the order statistics of  $y_t, t = 1, \dots, n$ . The estimators of the right-tail and left-tail indices are defined as

$$H_{1k} = \left\{ \frac{1}{k} \sum_{i=1}^k \log \left( \frac{y_{(i)}}{y_{(k+1)}} \right) \right\}^{-1},$$

$$H_{2k} = \left\{ \frac{1}{k} \sum_{i=1}^k \log \left( \frac{y_{(n-i+1)}}{y_{(n-k)}} \right) \right\}^{-1},$$

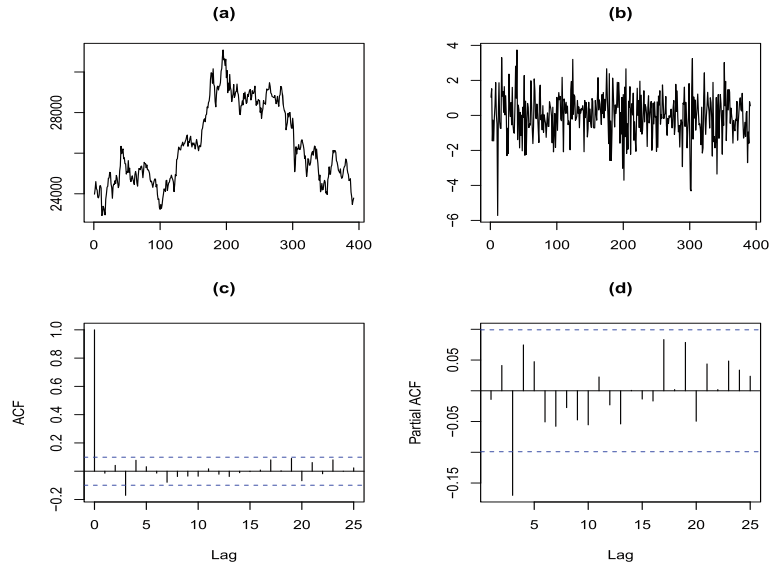


Figure 3. (a) Sample path of data  $x_t$ , (b) sample path of data  $y_t$ , (c) sample ACF of  $y_t$ , (d) sample PACF of  $y_t$

respectively. Figure 4 displays the Hill estimates of the right-tail and left-tail indices when  $1 \leq k \leq 200$ . From Figure 4, we can see that both the right and left tail indices are most likely less than 2. Hence,  $y_t$  should have a heavy tail. Therefore, it may be more appropriate to suppose these data are generated from a process with infinite variance rather than to assume this data have finite variance.

Based on the above discussion, we can fit a GRCAR(3) model to the data:

$$y_t = (\phi_0 + \varepsilon_t) + (\phi_1 + \varepsilon_t)y_{t-1} + (\phi_2 + \varepsilon_t)y_{t-2} + (\phi_3 + \varepsilon_t)y_{t-3} + \varepsilon_t, \varepsilon_t \sim t_2. \tag{15}$$

**Step 4. Parameter estimation:** The unknown parameters are estimated by different methods on the training data. We calculated the mean absolute errors (MAEs) of predicted values for transformed data based on 1000 repetitions. The results are shown in Table III. From them, we can see that the self-weighted estimators perform better, especially  $SM_2$ . So we choose the  $SM_2$  for further analysis. The estimates are

$$(\phi_0, \phi_1, \phi_2, \phi_3) = (0.0493, 0.0059, 0.0214, -0.1608),$$

whose corresponding asymptotic standard deviations are 0.0028, 0.0019, 0.0018, 0.0018 respectively.

**Step 5. Model diagnostics:** The absolute value of the eigenvalues for the corresponding matrix in model (15) are 0.5549, 0.5383 and 0.5383, which are all less than one. Therefore, this model satisfies the stationary conditions of Corollary 2.3. Figure 5 presents the residuals of the fitted model (15), the normal  $Q-Q$  plot of the residuals, the sample autocorrelation function (ACF) of the residuals and the sample ACF of the squared residuals. From that, we can see the model (15) fits the data reasonably well.

**Step 6. Prediction:** We use the above model (15) to predict  $y_{393}, y_{394}, \dots, y_{412}$  in the test set. As the coefficient of model (15) is random, the predicted values of Hang Seng Index from 3 December 2021 to 31 December 2021,  $x_{393}, x_{394}, \dots, x_{412}$ , were calculated by taking the average of 1000 repetitions. We compare our GRCAR(3) model with the AR(3) model and AR(3)-ARCH(2) model. The predictive performance of different models is presented in Figure 6. We can see that the predicted value of model (15) captures the change trend of the real value and most of the predicted values are very close to the true values. Figure 6 and Table IV also show that the GRCAR(3) model performs better than AR(3) model and AR(3)-ARCH(2) model for this dataset.

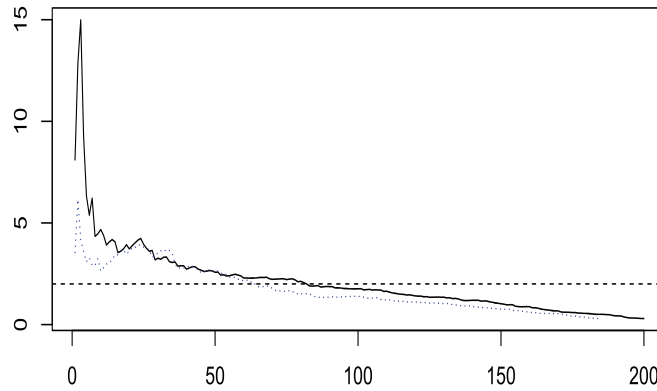


Figure 4. Hill estimates of the right tail index  $H_{1k}$ (black) and the left tail index  $H_{2k}$ (blue)

Table III. Mean absolute errors (MAE) of predicted values on the training set for transformed data

Methods	$SM_1$	$SM_2$	LAD	LS
MAE	0.9691	0.9529	0.9800	0.9996

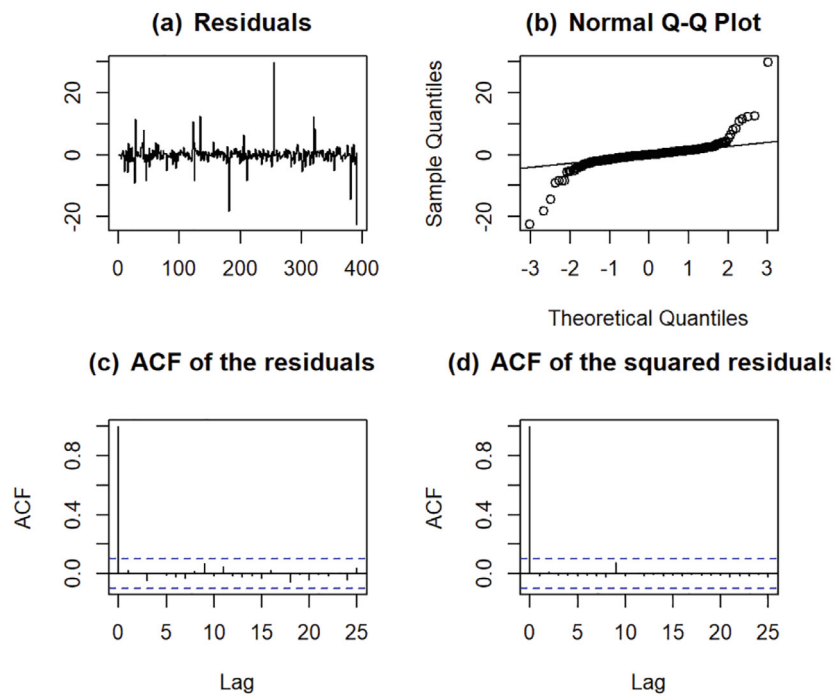


Figure 5. (a) The residuals from the fitted GRCAR(3) model, (b) the normal  $Q-Q$  plot of the residuals, (c) the sample autocorrelation function (ACF) of the residuals and (d) the sample ACF of the squared residuals

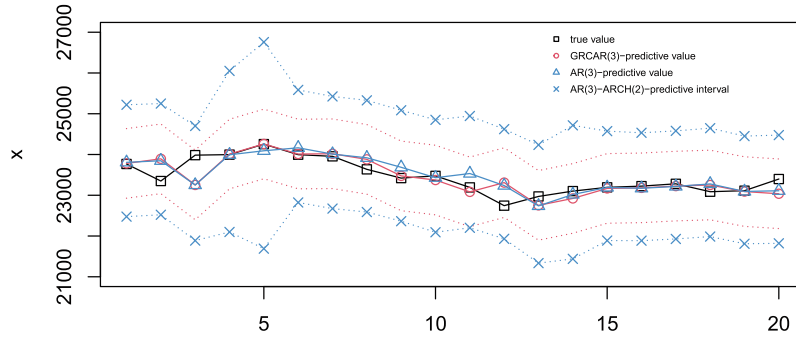


Figure 6. The predictive performance of different models

Table IV. Some indicators of predictive accuracy on test set about the original data

	Mean of residual	SD of residual	MAE
AR(3)	-32.60	282.93	200.98
GRCAR(3)	8.07	279.34	176.75

In summary, our model and method perform well in analysis and forecasting of time series data with heavy-tailed distributions.

### 5. PROOFS OF THEORETICAL RESULTS

This section presents the proofs of our theoretical results.

*Proof of Theorem 2.1.* It is easy to see that  $\{X_t\}$  defined by (3) is a homogenous Markov chain. By the conditions of Theorem 2.1, this Markov chain is  $u_p$ -irreducible and aperiodic, and all bounded sets with positive  $u_p$ -measure in  $R^p$  are small sets. Take the test function  $g(x) = \|x\|_v^\delta$ . It holds that

$$\begin{aligned}
 E \{ g(X_t) | X_{t-1} = x \} &= E \{ g(\Phi_t(x) + \varepsilon_t U) \} \\
 &= E \| \Phi_t(x) + \varepsilon_t U \|_v^\delta \\
 &\leq E \| \Phi_t(x) \|_v^\delta + E \| \varepsilon_t U \|_v^\delta \\
 &\leq \rho \| x \|_v^\delta + c + (E | \varepsilon_t |^\delta) \| U \|_v^\delta \\
 &= \rho \| x \|_v^\delta + c + c' \\
 &= \alpha \| x \|_v^\delta - [(\alpha - \rho) \| x \|_v^\delta - c - c'],
 \end{aligned}$$

where  $0 < \rho < \alpha < 1$ ,  $c' = (E | \varepsilon_t |^\delta) \| U \|_v^\delta$ . Let  $C = \{ x : \| x \|_v^\delta \leq k \}$ ,  $k > \max \left\{ 1, \frac{c+c'}{\alpha-\rho} \right\}$ , then  $C$  is a small set, and

$$\begin{aligned}
 E \{ g(X_t) | X_{t-1} = x \} &\leq \alpha \| x \|_v^\delta - c_1, \forall x \notin C, \\
 E \{ g(X_t) | X_{t-1} = x \} &\leq c_2, \forall x \in C,
 \end{aligned}$$

where  $c_1 = (\alpha - \rho)k - c - c'$ ,  $c_2 = \rho k + c + c'$ . By the Lyapunov shift criteria (see Meyn and Tweedie (1994)), the model (3) is geometrically ergodic, which implies  $\{y_t\}$  in model (2) is stationary and geometrically ergodic. ■

*Proof of Corollary 2.2.* We only need to verify the condition (i) of Theorem 2.1. Denote

$$\mathbf{A} = \begin{pmatrix} \varphi_1 & \varphi_2 & \cdots & \varphi_{p-1} & \varphi_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}.$$

The condition (5) of Corollary 2.2 is equivalent to  $\lambda^p - \varphi_1 \lambda^{p-1} - \cdots - \varphi_p \neq 0, |\lambda| \geq 1$ . It implies the roots  $\lambda_1, \dots, \lambda_p$  of  $|\lambda \mathbf{I} - \mathbf{A}| = 0$  satisfy  $|\lambda_i| < 1, i = 1, 2, \dots, p$ . Put  $\rho = \max\{|\lambda_i|, i = 1, 2, \dots, p\}$ , then  $0 < \rho < 1$ . Thus there exists a positively definite matrix  $\mathbf{V}_{p \times p}$  and  $0 < \rho < 1$  such that  $\mathbf{A}'\mathbf{V}\mathbf{A} \leq \rho^2 \mathbf{V}$  (see Ciarlet (1982)). Furthermore,

$$\forall \mathbf{x} \in R^p, \mathbf{x}'\mathbf{A}'\mathbf{V}\mathbf{A}\mathbf{x} \leq \rho^2 \mathbf{x}'\mathbf{V}\mathbf{x}.$$

Define a norm  $\|\cdot\|_v$  as follows

$$\|\mathbf{x}\|_v = (\mathbf{x}'\mathbf{V}\mathbf{x})^{1/2}, \forall \mathbf{x} \in R^p.$$

Then

$$\|\mathbf{A}\mathbf{x}\|_v \leq \rho \|\mathbf{x}\|_v, \forall \mathbf{x} \in R^p.$$

Let  $H_t(\mathbf{x}) = \Phi_t(\mathbf{x}) - \mathbf{A}\mathbf{x} = (\varphi_t(\mathbf{x}) - \varphi'_t \mathbf{x}, 0, \dots, 0)'$ , where  $\varphi'_t = (\varphi_1, \dots, \varphi_p)$ . By the norm equivalence, there is a positive constant  $M$  such that

$$\frac{E\|H_t(\mathbf{x})\|_v^\delta}{\|\mathbf{x}\|_v^\delta} \leq \frac{ME\|H_t(\mathbf{x})\|^\delta}{\|\mathbf{x}\|^\delta} = \frac{ME|\varphi_t(\mathbf{x}) - \varphi'_t \mathbf{x}|^\delta}{\|\mathbf{x}\|^\delta}. \quad (16)$$

By (6), when  $\|\mathbf{x}\| \rightarrow \infty$ , the right-hand side of the above inequality tends to 0. Hence, there exists a constant  $K_0$  such that

$$E\|H_t(\mathbf{x})\|_v^\delta \leq \frac{1}{2} (1 - \rho^\delta) \|\mathbf{x}\|_v^\delta$$

for  $\|\mathbf{x}\|_v^\delta > K_0$ . Therefore,

$$\begin{aligned} E\|\Phi_t(\mathbf{x})\|_v^\delta &= E\|\mathbf{A}\mathbf{x} + H_t(\mathbf{x})\|_v^\delta \\ &\leq \|\mathbf{A}\mathbf{x}\|_v^\delta + E\|H_t(\mathbf{x})\|_v^\delta \\ &\leq \rho^\delta \|\mathbf{x}\|_v^\delta + E\|H_t(\mathbf{x})\|_v^\delta \\ &\leq \begin{cases} \frac{1}{2}(1 + \rho^\delta)\|\mathbf{x}\|_v^\delta, & \|\mathbf{x}\|_v^\delta \geq K_0, \\ \rho^\delta \|\mathbf{x}\|_v^\delta + M_1, & \|\mathbf{x}\|_v^\delta < K_0. \end{cases} \end{aligned}$$

By (7) and (16), when  $\|\mathbf{x}\|_v^\delta < K_0$ , there exists a constant  $M_1 \geq 0$  such that  $E\|H_t(\mathbf{x})\|_v^\delta \leq M_1$ . This implies that the condition (i) of Theorem 2.1 holds. By Theorem 2.1, we get the result of this corollary. ■

*Proof of Corollary 2.3.* We only need to verify the condition (i) of corollary 2.2.

Define

$$\varphi_t(\mathbf{x}) = \phi_{t0} + (\phi_{t1}, \dots, \phi_{tp}) \mathbf{x}, \mathbf{x} \in R^p.$$

Then

$$\begin{aligned} |\varphi_t(\mathbf{x}) - \varphi'_t \mathbf{x}|^\delta &= \left| \phi_{t0} + (\phi_{t1} - \phi_1, \dots, \phi_{tp} - \phi_p) \mathbf{x} \right|^\delta \\ &\leq |\phi_{t0}|^\delta + \left| (\phi_{t1} - \phi_1, \dots, \phi_{tp} - \phi_p) \mathbf{x} \right|^\delta. \end{aligned}$$

Therefore

$$\begin{aligned} \lim_{\|\mathbf{x}\| \rightarrow \infty} \frac{E|\varphi_t(\mathbf{x}) - \varphi'_t \mathbf{x}|^\delta}{\|\mathbf{x}\|} &\leq \lim_{\|\mathbf{x}\| \rightarrow \infty} \frac{E|\phi_{t0}|^\delta}{\|\mathbf{x}\|} + \lim_{\|\mathbf{x}\| \rightarrow \infty} \frac{E\left| (\phi_{t1} - \phi_1, \dots, \phi_{tp} - \phi_p) \mathbf{x} \right|^\delta}{\|\mathbf{x}\|} \\ &\leq \lim_{\|\mathbf{x}\| \rightarrow \infty} \frac{E|\phi_{t0}|^\delta}{\|\mathbf{x}\|} + \lim_{\|\mathbf{x}\| \rightarrow \infty} \frac{E\left\| (\phi_{t1} - \phi_1, \dots, \phi_{tp} - \phi_p) \right\|^\delta \|\mathbf{x}\|^\delta}{\|\mathbf{x}\|} \\ &\leq \lim_{\|\mathbf{x}\| \rightarrow \infty} \frac{E|\phi_{t0}|^\delta}{\|\mathbf{x}\|} + \lim_{\|\mathbf{x}\| \rightarrow \infty} \frac{E\left( \sum_{i=1}^p |\phi_{ti} - \phi_i|^\delta \right) \|\mathbf{x}\|^\delta}{\|\mathbf{x}\|} \\ &\leq \lim_{\|\mathbf{x}\| \rightarrow \infty} \frac{E|\phi_{t0}|^\delta}{\|\mathbf{x}\|} + \lim_{\|\mathbf{x}\| \rightarrow \infty} \frac{\sum_{i=1}^p E|\phi_{ti} - \phi_i|^\delta}{\|\mathbf{x}\|^{1-\delta}} \\ &= 0, \end{aligned}$$

where the second inequality uses the Schwarz inequality, the third inequality uses the Triangle inequality, and the last equation holds by the condition (ii) of Corollary 2.3. Also,

$$\sup_{\|\mathbf{x}\| \leq K} E|\varphi_t(\mathbf{x}) - \varphi'_t \mathbf{x}|^\delta \leq \sup_{\|\mathbf{x}\| \leq K} \left[ E|\phi_{t0}|^\delta + \sum_{i=1}^p E|\phi_{ti} - \phi_i|^\delta \right] < \infty.$$

This implies that the condition (i) of Corollary 2.2 holds. By Corollary 2.2,  $\{y_t\}$  in model (1) is stationary and geometrically ergodic. ■

In the following, we give two lemmas, which will be used frequently in the proof of Theorem 2.4. The first lemma is directly taken from Davis *et al.* (1992).

**Lemma 5.1.** Let  $V_n(\cdot)$  and  $V(\cdot)$  be stochastic process on  $R^{p+1}$  and suppose that  $V_n(\cdot) \xrightarrow{L} V(\cdot)$  on  $C(R^{p+1})$ . Let  $\xi_n$  minimize  $V_n(\cdot)$  and  $\xi_n$  minimize  $V(\cdot)$ . If  $V_n(\cdot)$  is convex for each  $n$  and  $\xi$  is unique with probability one, then  $\xi_n \xrightarrow{L} \xi$  on  $R^{p+1}$ .

*Proof.* See Davis *et al.* (1992). ■

**Lemma 5.2.** Under the conditions (C.1)–(C.5), we have, as  $n \rightarrow \infty$ ,

- (a)  $\frac{1}{n} \sum_{t=1}^n \omega_t X_t X_t' \xrightarrow{P} \Sigma, \frac{1}{n} \sum_{t=1}^n \omega_t^2 X_t X_t' \xrightarrow{P} \Omega;$
- (b) for any fixed  $(p + 1) \times 1$  vector  $C$  such that  $C' \Omega C > 0, \max_{1 \leq t \leq n} \frac{|\omega_t C' X_t|}{\sqrt{n}} \xrightarrow{P} 0;$
- (c)  $\frac{1}{n} \sum_{t=1}^n \omega_t X_t \psi(\varepsilon_t) \xrightarrow{L} N(0, \tau \Omega).$

*Proof.* By applying the conditions (C.1) with  $y_t$  being stationary and ergodic, it is easy to get (a) and (b). So we omit the proofs of (a) and (b), and only give the proof of (c). Put  $\zeta_{nt} = \frac{1}{\sqrt{n}}\omega_t C' X_t \psi(\varepsilon_t)$ . Then

$$\sum_{t=1}^n \zeta_{nt} = C' \frac{1}{\sqrt{n}} \sum_{t=1}^n \omega_t X_t \psi(\varepsilon_t),$$

and  $\zeta_{nt}$ ,  $1 \leq t \leq n$  is a sequence of martingale differences with respect to  $\mathcal{F}_{t-1}$ . By (a), it follows that

$$\begin{aligned} \sum_{t=1}^n E\left(\frac{1}{n}\omega_t^2 C' X_t X_t' C \psi^2(\varepsilon_t) | \mathcal{F}_{t-1}\right) &= \frac{1}{n} \sum_{t=1}^n (\omega_t^2 C' X_t X_t' C) E\psi^2(\varepsilon_t) \\ &= \tau \frac{1}{n} \sum_{t=1}^n (\omega_t^2 C' X_t X_t' C) \xrightarrow{p} \tau v, \end{aligned} \quad (17)$$

where  $v = C' \Omega C$ . Put  $\xi_t = \omega_t C' X_t$ , then for any  $\eta > 0$ , we have

$$\begin{aligned} \sum_{t=1}^n E(\zeta_{nt}^2 I(|\zeta_{nt}| > \eta) | \mathcal{F}_{t-1}) &= \frac{1}{n} \sum_{t=1}^n \xi_t^2 E(\psi^2(\varepsilon_t) I(|\xi_t \psi(\varepsilon_t)| > \eta \sqrt{n}) | \mathcal{F}_{t-1}) \\ &\leq \max_{1 \leq t \leq n} E\left(\psi^2(\varepsilon_t) I(|\xi_t \psi(\varepsilon_t)| > \eta \sqrt{n}) | \mathcal{F}_{t-1}\right) \frac{1}{n} \sum_{t=1}^n \xi_t^2. \end{aligned} \quad (18)$$

Notice that

$$I(|\xi_t \psi(\varepsilon_t)| > \eta \sqrt{n}) \leq I(|\psi(\varepsilon_t)| > \eta M) + I\left(\frac{|\xi_t|}{\sqrt{n}} > \frac{1}{M}\right)$$

for any fixed  $M > 0$ . It implies that, for  $1 \leq t \leq n$ ,

$$\begin{aligned} &E\left(\psi^2(\varepsilon_t) I\left(\left|\frac{\xi_t \psi(\varepsilon_t)}{\sqrt{n}}\right| > \eta\right) \middle| \mathcal{F}_{t-1}\right) \\ &\leq E\left(\psi^2(\varepsilon_t) I(|\psi(\varepsilon_t)| > \eta M)\right) + E\left(\psi^2(\varepsilon_t) I\left(\frac{|\xi_t|}{\sqrt{n}} > \frac{1}{M}\right) \middle| \mathcal{F}_{t-1}\right) \\ &\leq E\left(\psi^2(\varepsilon_t) I(|\psi(\varepsilon_t)| > \eta M)\right) + \tau \cdot \max_{1 \leq t \leq n} I\left(\frac{|\xi_t|}{\sqrt{n}} > \frac{1}{M}\right) \\ &\leq E\left(\psi^2(\varepsilon_t) I(|\psi(\varepsilon_t)| > \eta M)\right) + \tau \cdot I\left(\max_{1 \leq t \leq n} \frac{|\xi_t|}{\sqrt{n}} > \frac{1}{M}\right). \end{aligned} \quad (19)$$

This leads to

$$\begin{aligned} &\max_{1 \leq t \leq n} E\left(\psi^2(\varepsilon_t) I\left(\left|\frac{\xi_t \psi(\varepsilon_t)}{\sqrt{n}}\right| > \eta\right) \middle| \mathcal{F}_{t-1}\right) \\ &\leq E\left(\psi^2(\varepsilon_t) I(|\psi(\varepsilon_t)| > \eta M)\right) + \tau \cdot I\left(\max_{1 \leq t \leq n} \frac{|\xi_t|}{\sqrt{n}} > \frac{1}{M}\right). \end{aligned} \quad (20)$$

Thus

$$E \left\{ \max_{1 \leq t \leq n} E \left( \psi^2(\varepsilon_t) I \left( \left| \frac{\xi_t \psi(\varepsilon_t)}{\sqrt{n}} \right| > \eta \right) \middle| \mathcal{F}_{t-1} \right) \right\} \leq E \left( \psi^2(\varepsilon_1) I (|\psi(\varepsilon_1)| > \eta M) \right) + \tau \cdot P \left( \max_{1 \leq t \leq n} \frac{|\xi_t|}{\sqrt{n}} > \frac{1}{M} \right).$$

Notice  $E\psi^2(\varepsilon_1) < \infty$ . Then for any  $\varepsilon > 0$ , there exists  $M = M(\varepsilon)$  such that

$$E \left( \psi^2(\varepsilon_1) I (|\psi(\varepsilon_1)| > \eta M) \right) < \varepsilon,$$

which implies that

$$E \left\{ \max_{1 \leq t \leq n} E \left( \psi^2(\varepsilon_t) I \left( \left| \frac{\xi_t \psi(\varepsilon_t)}{\sqrt{n}} \right| > \eta \right) \middle| \mathcal{F}_{t-1} \right) \right\} \leq \tau \cdot P \left( \max_{1 \leq t \leq n} \frac{|\xi_t|}{\sqrt{n}} > \frac{1}{M} \right) + \varepsilon. \tag{21}$$

From (b) and (21), we have

$$\limsup_{n \rightarrow \infty} E \left\{ \max_{1 \leq t \leq n} E \left( \psi^2(\varepsilon_t) I \left( \left| \frac{\xi_t \psi(\varepsilon_t)}{\sqrt{n}} \right| > \eta \right) \middle| \mathcal{F}_{t-1} \right) \right\} \leq \varepsilon.$$

Then

$$\lim_{n \rightarrow \infty} E \left\{ \max_{1 \leq t \leq n} E \left( \psi^2(\varepsilon_t) I \left( \left| \frac{\xi_t \psi(\varepsilon_t)}{\sqrt{n}} \right| > \eta \right) \middle| \mathcal{F}_{t-1} \right) \right\} = 0.$$

Thus

$$\max_{1 \leq t \leq n} E \left( \psi^2(\varepsilon_t) I \left( \left| \frac{\xi_t \psi(\varepsilon_t)}{\sqrt{n}} \right| > \eta \right) \middle| \mathcal{F}_{t-1} \right) = o_p(1), \tag{22}$$

which, combining with (22) and (a), derives that, for any  $\eta > 0$ ,

$$\sum_{t=1}^n E \left( \zeta_{nt}^2 I(|\zeta_{nt}| > \eta) \middle| \mathcal{F}_{t-1} \right) = o_p(1). \tag{23}$$

Therefore, by applying the martingale central limit theorem, (18) and (23), we have

$$\frac{1}{n} \sum_{t=1}^n \omega_t X_t \psi(\varepsilon_t) \xrightarrow{L} N(0, \tau \Omega).$$

The proof of the Lemma 5.2 is completed. ■

*Proof of Theorem 2.4.* Denote  $\hat{\beta}_n = \sqrt{n}(\hat{\phi}_{SM} - \phi_0)$  and

$$V_n(\mu) = \sum_{t=1}^n \omega_t \left( \rho \left( \varepsilon_t - \frac{1}{\sqrt{n}} \mu' X_t \right) - \rho(\varepsilon_t) \right), \tag{24}$$



where  $\hat{\beta}_n$  is the value of  $\mu$  that minimizes the convex objective function  $V_n(\mu)$  and  $\mu \in \mathcal{R}^{p+1}$ . Put

$$A_n = \frac{1}{\sqrt{n}} \sum_{t=1}^n \omega_t X_t \psi(\varepsilon_t),$$

$$B_t(\mu) = \omega_t \int_0^{-\frac{\mu' X_t}{\sqrt{n}}} (\psi(\varepsilon_t + s) - \psi(\varepsilon_t)) ds.$$

Then

$$V_n(\mu) = -\mu' A_n + \sum_{t=1}^n B_t(\mu)$$

$$= -\mu' A_n + \sum_{t=1}^n E(B_t(\mu) | \mathcal{F}_{t-1}) + \sum_{t=1}^n (B_t(\mu) - E(B_t(\mu) | \mathcal{F}_{t-1})). \tag{25}$$

From the condition (C.3), we obtain

$$\sum_{t=1}^n E(B_t(\mu) | \mathcal{F}_{t-1}) = \sum_{t=1}^n \omega_t \int_0^{-\frac{\mu' X_t}{\sqrt{n}}} E(\psi(\varepsilon_t + s)) ds$$

$$= \sum_{t=1}^n \omega_t \int_0^{-\frac{\mu' X_t}{\sqrt{n}}} \lambda s(1 + o(1)) ds$$

$$= \frac{\lambda \mu'}{2n} \left( \sum_{t=1}^n \omega_t X_t X_t' \right) \mu (1 + o(1))$$

$$= \frac{\lambda \mu'}{2} \left( \frac{1}{n} \sum_{t=1}^n \omega_t X_t X_t' \right) \mu + o_p(1).$$

Note that  $\{B_t(\mu) - E(B_t(\mu) | \mathcal{F}_{t-1}), 1 \leq t \leq n\}$  is a sequence of martingale differences. Then we get

$$\sum_{t=1}^n E(B_t^2(\mu) | \mathcal{F}_{t-1}) = \sum_{t=1}^n E \left( \omega_t^2 \left( \int_0^{-\frac{\mu' X_t}{\sqrt{n}}} (\psi(\varepsilon_t + s) - \psi(\varepsilon_t)) ds \right)^2 \middle| \mathcal{F}_{t-1} \right)$$

$$\leq \sum_{t=1}^n E \left( \omega_t^2 \int_0^{-\frac{\mu' X_t}{\sqrt{n}}} ds \int_0^{-\frac{\mu' X_t}{\sqrt{n}}} (\psi(\varepsilon_t + s) - \psi(\varepsilon_t))^2 ds \middle| \mathcal{F}_{t-1} \right)$$

$$\leq \sum_{t=1}^n \omega_t^2 \frac{|\mu' X_t|}{\sqrt{n}} \left| \int_0^{-\frac{\mu' X_t}{\sqrt{n}}} E(\psi(\varepsilon_t + s) - \psi(\varepsilon_t))^2 ds \right|$$

$$= \mu' \left( \frac{1}{n} \sum_{t=1}^n \omega_t^2 X_t X_t' \right) \mu \cdot o(1). \tag{26}$$

From Lemma 5.1 and (26), we can obtain

$$\sum_{t=1}^n EB_t^2(\mu) \leq \mu' \left( \frac{1}{n} \sum_{t=1}^n E\omega_t^2 X_t X_t' \right) \mu \cdot o(1) = \mu' \Omega \mu \cdot o(1) \rightarrow 0.$$

Therefore, we have

$$\sum_{t=1}^n EB_t^2(\boldsymbol{\mu}) \rightarrow 0.$$

Thus

$$\begin{aligned} E\left(\sum_{t=1}^n (B_t(\boldsymbol{\mu}) - E(B_t(\boldsymbol{\mu}) | \mathcal{F}_{t-1}))\right)^2 &= \sum_{t=1}^n E(B_t(\boldsymbol{\mu}) - E(B_t(\boldsymbol{\mu}) | \mathcal{F}_{t-1}))^2 \\ &\leq 2 \sum_{t=1}^n EB_t^2(\boldsymbol{\mu}) \rightarrow 0. \end{aligned}$$

Hence (25) can be rewritten to be

$$V_n(\boldsymbol{\mu}) = \frac{\lambda}{2} \boldsymbol{\mu}' \left( \frac{1}{n} \sum_{t=1}^n \omega_t X_t X_t' \right) \boldsymbol{\mu} - \boldsymbol{\mu}' A_n + o_p(1). \tag{27}$$

Because of Lemma 5.2, we can get

$$V_n(\boldsymbol{\mu}) \xrightarrow{L} V(\boldsymbol{\mu}) = \frac{\lambda}{2} \boldsymbol{\mu}' \boldsymbol{\Sigma} \boldsymbol{\mu} - \boldsymbol{\mu}' A,$$

where  $A \sim N(0, \tau \boldsymbol{\Omega})$ . Note that  $V(\boldsymbol{\mu})$  has a unique minimum at  $\boldsymbol{\mu} = \frac{1}{\lambda} \boldsymbol{\Sigma}^{-1} A$  almost surely and  $V_n(\boldsymbol{\mu})$  has convex sample paths due to the condition (C.2). Applying Lemma 5.1, we have

$$\hat{\boldsymbol{\beta}}_n = \sqrt{n}(\hat{\boldsymbol{\phi}}_{SM} - \boldsymbol{\phi}_0) \xrightarrow{L} \frac{1}{\lambda} \boldsymbol{\Sigma}^{-1} A \sim N\left(0, \frac{\tau}{\lambda^2} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Omega} \boldsymbol{\Sigma}^{-1}\right).$$

So the proof of Theorem 2.4 is completed. ■

### 6. CONCLUDING REMARKS

This article establishes the geometric ergodicity of general stochastic functional autoregressive models (including linear and nonlinear) under a broad condition. Furthermore, the stationary conditions and a self-weighted M-estimator for GRCAR(p) models with possibly heavy-tailed errors are proposed. The proposed estimator is shown to be asymptotically normal. The simulation study and a real data example showed that our theory and methodology perform well in practice. A general approach to stationarity and estimation for GRCAR models with heavy-tailed errors is presented. The methodology and results could be extended further to other time series models such as heavy-tailed GRCARMA models.

### ACKNOWLEDGEMENTS

The authors thank the Editor, the Co-Editor and the Referee(s) for their insightful comments and suggestions that make us improve our article significantly. The second author’s work was partially supported by the National Natural Science Foundation of China (Grant No. 12171161).

### DATA AVAILABILITY STATEMENT

Our dataset consists of the daily Hang Seng closing index from 7 May 2020 to 31 December 2021. The data were downloaded at <https://cn.investing.com> and are available publicly.

## REFERENCES

- Araveeporn A. 2017. Comparing random coefficient autoregressive model with and without autocorrelated errors by Bayesian analysis. *Statistical Journal of the IAOS* **33**:537–545.
- Bai ZD, Rao CR, Wu Y. 1992. M-estimation of multivariate linear regression parameters under a convex discrepancy function. *Statistica Sinica* **2**:237–254.
- Ciarlet PG. 1982. *Introduction à l'analyse numérique matricielle et à l'optimisation* MASSON, Paris.
- Davis RA, Knight K, Liu J. 1992. M-estimation for autoregressions with infinite variance. *Stochastic Processes and their Applications* **40**:145–180.
- Drees H, de Haan L, Resnick S. 2000. How to make a Hill plot. *Annals of Statistics* **28**:254–274.
- Fu KA, Li T, Ni C, He W, Wu R. 2021. Asymptotics for the conditional self-weighted m-estimator of GRCA(1) models with possibly heavy-tailed errors. *Statistical Papers* **62**:1407–1419.
- Hwang SY, Basawa IV. 1997. The local asymptotic normality of a class of generalized random coefficient autoregressive processes. *Statistics and Probability Letters* **34**:165–170.
- Hwang SY, Basawa IV. 1998. Parameter estimation for generalization random coefficient autoregressive processes. *Journal of Statistical Planning and Inference* **68**:323–337.
- Ling S. 2005. Self-weighted LAD estimation for infinite variance autoregressive models. *Journal of the Royal Statistical Society Series B* **67**:381–393.
- Meyn SP, Tweedie RL. 1994. State-dependent criteria for convergence of Markov chains. *Annals of Applied Probability* **4**:149–168.
- Nicholls DF, Quinn BG. 1982. *Random Coefficient Autoregressive Models: An Introduction. Lectures Notes in Statistics*, Vol. 11. Springer, New York.
- Pan B, Chen M. 2013. Self-weighted quasi-maximum exponential likelihood estimator for ARFIMA-GARCH model. *Journal of Statistical Planning and Inference* **143**:716–729.
- Pan J, Wang H, Yao Q. 2007. Weighted least absolute deviations estimation for ARMA models with infinite variance. *Econometric Theory* **23**:852–879.
- Pan BG, Chen M, Wang Y, Xia W. 2015. Weight least absolute deviations estimation for ARFIMA time series with finite or infinite variance. *Journal of the Korean Statistical Society* **144**:1–11.
- Resnick SI. 2000. Heavy tail modeling and teletraffic data. *Annals of Statistics* **25**:1805–1849.
- Silverman BW. 1986. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.
- Tong H. 1990. *Nonlinear Time Series*. Oxford University Press, Oxford.
- Wang X, Hu S. 2017. Asymptotics of self-weighted M-estimators for autoregressive models. *Metrika* **80**:83–92.
- Wang K, Zhu Y. 2018. M-estimation in high-dimensional linear model. *Journal of Inequalities and Applications*, Article number: 225. <https://doi.org/10.1186/s13660-018-1819-3>.
- Wu WB. 2007. M-estimation of linear models with dependent errors. *Annals of Statistics* **35**:495–521.
- Wu R. 2013. M-estimation for general ARMA processes with infinite variance. *Scandinavian Journal of Statistics: Theory and Applications* **40**:571–591.
- Yang Y, Ling S. 2017. Self-weighted LAD-based inference for heavy-tailed threshold autoregressive models. *Journal of Econometrics* **197**:368–381.
- Yu C, Deng M, Yau SST. 2011. DNA sequence comparison by a novel probabilistic method. *Information Sciences* **181**:1484–1492.
- Zhang L, Yu C, Sun JQ. 2015. Generalized Weierstrass–Mandelbrot function model for actual stocks markets indexes with nonlinear characteristics. *Fractals* **23**:1550006.
- Zhao Z, Wang D. 2012. Statistical inference for generalized random coefficient autoregressive model. *Mathematical and Computer Modelling* **56**:152–166.
- Zhao Z, Wang D, Peng C. 2013. Coefficient constancy test in generalized random coefficient autoregressive model. *Applied Mathematics and Computation* **219**:10283–10292.
- Zhao Z, Liu Y, Peng C. 2018. Variable selection in generalized random coefficient autoregressive models. *Journal of Inequalities and Applications*, Article number: 82. <https://doi.org/10.1186/s13660-018-1680-4>.
- Zhao ZW, Yang L, Peng CX, Fu ZH. 2019. Empirical likelihood-based inference in generalized random coefficient autoregressive model with conditional moment restrictions. *Journal of Computational and Applied Mathematics* **348**:146–160.