Statistica Sinica Preprint No: SS-2022-0049	
Title	Quasi-maximum Likelihood Inference for Linear Double Autoregressive Models
Manuscript ID	SS-2022-0049
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202022.0049
Complete List of Authors	Hua Liu, Songhua Tan and Qianqian Zhu
Corresponding Authors	Qianqian Zhu
E-mails	zhu.qianqian@mail.shufe.edu.cn

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0049

Quasi-maximum Likelihood Inference for Linear Double Autoregressive Models

Hua Liu, Songhua Tan and Qianqian Zhu

School of Statistics and Management

Shanghai University of Finance and Economics

Abstract: This study investigates the quasi-maximum likelihood inference, including estimation, model selection, and diagnostic checking, for linear double autoregressive (DAR) models, where all asymptotic properties are established under only a fractional moment of the observed process. We propose an exponential quasi-maximum likelihood estimator (E-QMLE) for the linear DAR model, and establish its consistency and asymptotic normality. Based on the E-QMLE, we propose a Bayesian information criterion for model selection, and construct a mixed portmanteau test to check the adequacy of the fitted models. Inference tools based on the Gaussian quasi-maximum likelihood estimator (G-QMLE) are also discussed, for comparison. Moreover, we compare the proposed E-QMLE with the G-QMLE and the existing doubly weighted quantile regression estimator in terms of their asymptotic efficiency and numerical performance. Simulation studies illustrate the finite-sample performance of the proposed inference tools, and a real example using a Bitcoin return series shows their usefulness.

Key words and phrases: Double autoregressive models; Model selection; Portmanteau test; Quasi-maximum likelihood estimation.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0049

1. Introduction

Many conditional heteroscedastic models have been proposed to capture the time-varying volatility of financial and economic time series, with the autoregressive conditional heteroscedastic (ARCH) and generalized autoregressive conditional heteroscedastic (GARCH) models proving popular (Engle, 1982; Bollerslev, 1986). However, the conditional mean and conditional heteroscedasticity usually appear simultaneously in time series, and ignoring the conditional mean can lead to an inaccurate inference for the volatility (Li, Ling, and McAleer, 2002). Therefore, it is of vital importance to jointly model the conditional mean and volatility. The linear double autoregressive (DAR) model was proposed by Zhu, Zheng, and Li (2018) for this purpose, having the form

$$y_t = \sum_{i=1}^{p} \alpha_i y_{t-i} + \eta_t \left(\omega + \sum_{i=1}^{p} \beta_i |y_{t-i}| \right), \tag{1.1}$$

where $\omega > 0$, $\beta_i \geq 0$ for $1 \leq i \leq p$, and $\{\eta_t\}$ are independent and identically distributed ($i.i.d.$) random variables. Model (1.1) assumes a linear structure for the conditional standard deviation, which makes it less sensitive to extreme values, thus leading to a more robust inference than those form models with a linear structure for the conditional variance; see also Taylor (2008) and Xiao and Koenker (2009). Model (1.1) has a novel property that it enjoys a larger parameter space than those of the AR and AR-ARCH models (Zhu, Zheng, and Li, 2018). For example, with $p = 1$, it can be stationary, even if $|\alpha_1| > 1$, which is impossible for causal AR and AR-ARCH models.

A doubly weighted quantile regression estimator (DWQRE) was introduced by Zhu, Zheng, and Li (2018) for model (1.1), with $\omega = 1$ for identification. The DWQRE linearly combines the self-weighted quantile regression estimators at multiple quantile levels using weighting matrices, and requires only a fractional moment on the observed process $\{y_t\}$ to

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0049

establish its asymptotic properties. This novel property leads to robust inferences using model (1.1), and thus can be used for heavy-tailed data. Zhu, Zheng, and Li (2018) showed that as the total number of quantile levels goes to infinity, the optimal DWQRE can approach the efficiency of the maximum likelihood estimator (MLE) under certain conditions, defined in (2.5) in Section 2.3. However, combining the self-weighted quantile regression estimators at infinite quantile levels is infeasible in practice, and, more importantly, condition (2.5) implies that the DWQRE is, in general, less efficient than the MLE if the conditional mean structure exists. Furthermore, the two-step estimation procedure of the DWQRE is more complex than the one-step MLE or quasi-maximum likelihood estimation (QMLE) methods. To the best of our knowledge, no studies have examined the QMLE for model (1.1). This study addresses this gap in the literature.

Model (1.1) was originally motivated by the DAR model proposed by Ling (2004, 2007a), which is defined as

$$y_t = \sum_{i=1}^{p} \alpha_i y_{t-i} + \varepsilon_t \sqrt{\omega + \sum_{i=1}^{p} \beta_i y_{t-i}^2}, \tag{1.2}$$

where $\omega > 0$, $\beta_i \geq 0$ for $1 \leq i \leq p$, and $\{\varepsilon_t\}$ are $i.i.d.$ random variables. Model (1.2) is a special case of the ARMA-ARCH models in Weiss (1986), and has been extended by several studies, including the threshold DAR (Li, Ling, and Zakoïan, 2015; Li, Ling, and Zhang, 2016), mixture DAR (Li et al., 2017), linear DAR (Zhu, Zheng, and Li, 2018), augmented DAR (Jiang, Li, and Zhu, 2020), and asymmetric linear DAR (Tan and Zhu, 2022) models. In contrast to the linear DAR model (1.1), model (1.2) assumes a linear structure for the conditional variance, which may make it sensitive to extreme values. However, similarly to model (1.1), it enjoys a larger parameter space than those of the causal AR and AR-ARCH models. For the estimation of model (1.2), Ling (2007a) proposed a Gaussian quasi-

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0049

maximum likelihood estimator (G-QMLE), and established its asymptotic normality under a fractional moment of y_t and $E(\varepsilon_t^4) < \infty$. To reduce the moment condition of ε_t for a more robust estimation, Zhu and Ling (2013) proposed an exponential quasi-maximum likelihood estimator (E-QMLE) for model (1.2), where its asymptotic normality requires only that $E(\varepsilon_t^2) < \infty$. Note that we need only a finite fractional moment of the process $\{y_t\}$ in order to establish the asymptotic normality of the G-QMLE or E-QMLE, yielding a robust estimation for model (1.2).

The G-QMLE and E-QMLE essentially represent the least square estimation and the least absolute deviation estimation, respectively, and are popular for investigating the G-QMLE or E-QMLE for various time series models. For example, Aue and Horváth (2011) proposed the G-QMLE for random coefficient AR models, Francq and Zakoian (2004) and Francq and Zakoian (2019) studied the G-QMLE for GARCH and ARMA-GARCH models, Ling (2007b) proposed a self-weighted G-QMLE for ARMA-GARCH models, and Zhu and Ling (2011) investigated the E-QMLE for ARMA-GARCH models. Note that the G-QMLE and E-QMLE for the GARCH or ARMA-GARCH models usually require a second or fourth moment condition on the observed process in order to establish the asymptotic normality, whereas we need only a finite fractional moment for the DAR model (1.2). Therefore, we examine whether the robust property of QMLEs for model (1.2) can be preserved for the linear DAR model (1.1). Hopefully, the QMLEs for model (1.1) are robust and more convenient in terms of computation than the DWQRE. Model selection and diagnosis are another two key elements of the classical Box–Jenkins procedure, and hence need to be investigated based on QMLEs for model (1.1). This study contributes to the literature in three ways.

First, we propose an E-QMLE for model (1.1) in Section 2.1, and establish its asymptotic

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0049

normality under $E(|y_t|^\kappa) < \infty$ and $E(\eta_t^2) < \infty$. In particular, to derive the asymptotic normality of the E-QMLE, we adopt the bracketing method of Pollard (1985) to overcome the difficulty of the nondifferentiable and nonconvex objective function; see also Zhu and Ling (2011, 2013). For comparison with the E-QMLE, we introduce the G-QMLE for model (1.1) in Section 2.2, and obtain its asymptotic normality under $E(|y_t|^\kappa) < \infty$, for some $\kappa > 0$ and $E(\eta_t^4) < \infty$. We also compare the asymptotic efficiency of the E-QMLE with that of the DWQRE in Section 2.3. Simulation studies indicate that no estimator dominates in terms of asymptotic efficiency, but the E-QMLE is much more efficient than the G-QMLE, but slightly less efficient than the DWQRE when the data are more heavy tailed. Although all three estimators can be used to fit heavy-tailed data with $E(|y_t|^\kappa) < \infty$, in practice, we suggest choosing the most suitable option according to the fat tailedness of the fitted residuals and the computational complexity. Because the proposed E-QMLE offers a good trade-off between robustness and computational complexity, it is preferred for most financial and economic time series with heavy tails. The real-data application in Section 5 further illustrates this point.

Second, based on the E-QMLE, we propose a Bayesian information criterion (BIC) for model selection in Section 2.4, and show that it enjoys selection consistency without any moment condition on the process $\{y_t\}$. As the first stage of the Box–Jenkins procedure, order selection is crucial for fitting time series in practice, and the BIC is widely used to select orders for time series models (Poskitt and Tremayne, 1983; Cryer and Chan, 2008). Schwarz (1978) introduced the BIC for likelihood functions in the exponential family, Machado (1990) extended the BIC to a wide class of likelihood functions, and Machado (1993) studied the BIC based on objective functions that define M-estimators. Note that the E-QMLE belongs

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0049

to the class of M-estimators, motivating us to propose a BIC that is robust, in line with the estimation procedure. We further show that the proposed BIC is asymptotically consistent in terms of estimating the true order. Simulation studies indicate that, even if the sample size is not very large, the proposed BIC exhibits satisfactory performance.

Our third contribution is to construct a robust portmanteau test, in Section 3, for checking the adequacy of the fitted models, with no moment conditions imposed on the process $\{y_t\}$. It is well known that diagnostic checking is important for time series modeling, and portmanteau tests are commonly used for this purpose; see Box et al. (2008). For pure conditional mean models, the portmanteau test is usually constructed using the sample autocorrelation functions (ACFs) of the residuals (Ljung and Box, 1978), whereas we use the ACFs of the squared or absolute residuals for pure volatility models (Li and Mak, 1994; Li, 2004; Li and Li, 2005). Wong and Ling (2005) proposed a mixed portmanteau statistic for testing the adequacy of fitted time series models using residuals and squared residuals. To ensure the robustness of the test based on the E-QMLE, in the sense that its asymptotic properties require only $E(\eta_t^2) < \infty$, we adopt the ACFs of the absolute rather than the squared residuals to check the adequacy of the volatility part. As a result, we propose a mixed portmanteau test based on the ACFs of the residuals and the absolute residuals to simultaneously detect misspecifications of the conditional mean and volatility in the fitted model. The asymptotic properties of the test are established in Section 3.

Section 4 evaluates the finite-sample performance of the proposed inference tools by means of simulation, and a real-data example is given in Section 5. Section 6 concludes the paper. All technical details and additional simulation results are relegated to the Supplementary Material. Throughout the paper, $\to_p$ ($\to_{\mathcal{L}}$) denotes convergence in probability

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0049

(distribution), and $o_p(1)$ denotes a sequence of random variables converging to zero in probability. The data from Section 5 and the programs used to analyze them are available from https://github.com/Tansonghua-sufe/Linear-double-autoregression.

2. Model estimation

In this section, we propose an E-QMLE for model (1.1) in Section 2.1. Then, we compare the E-QMLE with the G-QMLE and the DWQRE of Zhu, Zheng, and Li (2018) in Sections 2.2–2.3, respectively. Finally, order selection is discussed in Section 2.4.

2.1 Exponential quasi-maximum likelihood estimation

Let $\boldsymbol{\theta} = (\boldsymbol{\alpha}', \boldsymbol{\delta}')'$ be the unknown parameter vector of model (1.1), and let $\boldsymbol{\theta}_0 = (\boldsymbol{\alpha}_0', \boldsymbol{\delta}_0')'$ be the true parameter vector, where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_p)'$ and $\boldsymbol{\delta} = (\omega, \beta_1, \beta_2, \ldots, \beta_p)'$. Denote the parameter space by $\Theta = \Theta_\alpha \times \Theta_\delta$, where $\Theta_\alpha \subset \mathbb{R}^p$ and $\Theta_\delta \subset \mathbb{R}_+^{p+1}$, with $\mathbb{R}_+ = (0, +\infty)$. Assume that $\{y_1, \ldots, y_n\}$ are generated by model (1.1), with η_t having a zero median and satisfying $E(|\eta_t|) = 1$. When η_t follows the standard double exponential distribution, the negative conditional log-likelihood function (ignoring a constant) can be written as

$$L_n^E(\boldsymbol{\theta}) = \frac{1}{n-p} \sum_{t=p+1}^{n} \ell_t^E(\boldsymbol{\theta}) \quad \text{and} \quad \ell_t^E(\boldsymbol{\theta}) = \ln h_t(\boldsymbol{\delta}) + \frac{|\varepsilon_t(\boldsymbol{\alpha})|}{h_t(\boldsymbol{\delta})}, \tag{2.1}$$

where

$$\varepsilon_t(\boldsymbol{\alpha}) = y_t - \sum_{i=1}^{p} \alpha_i y_{t-i} \quad \text{and} \quad h_t(\boldsymbol{\delta}) = \omega + \sum_{i=1}^{p} \beta_i |y_{t-i}|. \tag{2.2}$$

Let $\widehat{\boldsymbol{\theta}}_n = \operatorname{argmin}_{\boldsymbol{\theta} \in \Theta} L_n^E(\boldsymbol{\theta})$. Because we do not assume that η_t follows the standard double exponential distribution (i.e., the Laplace distribution), $\widehat{\boldsymbol{\theta}}_n$ is called the E-QMLE of $\boldsymbol{\theta}_0$; see also Zhu and Ling (2011, 2013).

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0049

Assumption 1. *Θ is compact, with $\underline{\omega} \le \omega \le \overline{\omega}$ and $\underline{\beta} \le \beta_i \le \overline{\beta}$, for $i = 1, \ldots, p$, where $\underline{\omega}, \overline{\omega}, \underline{\beta}$, and $\overline{\beta}$ are some positive constants, and $\boldsymbol{\theta}_0$ is an interior point in Θ.*

Assumption 2. *$\{y_t : t = 1, 2, ...\}$ is strictly stationary and ergodic, with $E(|y_t|^\kappa) < \infty$, for some $0 < \kappa \le 1$.*

Assumption 3. *(i) η_t has a zero median with $E(|\eta_t|) = 1$; (ii) the density function of η_t is continuous and positive everywhere on $\mathbb{R}$ satisfying $\sup_{x\in\mathbb{R}} f(x) < \infty$; (iii) $E(\eta_t^2) < \infty$.*

Assumption 1 is standard in the literature on quasi-maximum likelihood estimation; see also Ling (2007a) and Zhu and Ling (2013). For general distributions of η_t, it is difficult to derive a necessary and sufficient condition for the strict stationarity of y_t, owing to the nonlinearity of model (1.1). A sufficient condition for Assumption 2 is given in Theorem 2 of Zhu, Zheng, and Li (2018), that is, $\sum_{i=1}^{p} \max\{E(|\alpha_i - \beta_i\eta_t|^\kappa), E(|\alpha_i + \beta_i\eta_t|^\kappa)\} < 1$, for some $0 < \kappa \le 1$. Assumption 3 is a general setup for establishing the consistency and asymptotic normality of the E-QMLE; see also Zhu and Ling (2013).

Theorem 1. *If Assumptions 1, 2, and 3(i) hold, then $\widehat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}_0$ almost surely as $n \to \infty$.*

Let $\kappa_1 = E(\eta_t)$ and $\kappa_2 = E(\eta_t^2) - 1$. Denote $\boldsymbol{Y}_{1t} = h_t^{-1}(\boldsymbol{\delta}_0)(y_{t-1}, \ldots, y_{t-p})'$ and $\boldsymbol{Y}_{2t} = h_t^{-1}(\boldsymbol{\delta}_0)(1, |y_{t-1}|, \ldots, |y_{t-p}|)'$. Define the $(2p+1) \times (2p+1)$ matrices

$$\Sigma = \operatorname{diag}\left\{ f(0)E(\boldsymbol{Y}_{1t}\boldsymbol{Y}_{1t}'), \frac{1}{2}E(\boldsymbol{Y}_{2t}\boldsymbol{Y}_{2t}') \right\} \text{ and } \Omega = \begin{pmatrix} E(\boldsymbol{Y}_{1t}\boldsymbol{Y}_{1t}') & \kappa_1 E(\boldsymbol{Y}_{1t}\boldsymbol{Y}_{2t}') \\ \kappa_1 E(\boldsymbol{Y}_{2t}\boldsymbol{Y}_{1t}') & \kappa_2 E(\boldsymbol{Y}_{2t}\boldsymbol{Y}_{2t}') \end{pmatrix}.$$

Theorem 2. *If Assumptions 1–3 hold, then*

(i) $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = O_p(1)$;

(ii) $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \to_{\mathcal{L}} N(\mathbf{0}, \Xi)$ *as* $n \to \infty$, *where* $\Xi = \Sigma^{-1}\Omega\Sigma^{-1}/4$.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0049

Theorem 2 shows that the asymptotic normality of the proposed E-QMLE is established under a fractional moment of y_t, with $E(\eta_t^2) < \infty$, for model (1.1). Therefore, the E-QMLE is robust to heavy-tailed data, and can be used for $E(\eta_t^4) = \infty$ and $E(\eta_t^2) < \infty$, where the G-QMLE introduced in Section 2.2 is no longer applicable. In addition, if $\kappa_1 = 0$, then the asymptotic covariance in Theorem 2 reduces to the block diagonal matrix $\Gamma_0 = \text{diag}\{4f^2(0)E(\boldsymbol{Y}_{1t}\boldsymbol{Y}_{1t}'), \kappa_2^{-1}E(\boldsymbol{Y}_{2t}\boldsymbol{Y}_{2t}')\}^{-1}$.

To estimate the asymptotic covariance of $\widehat{\boldsymbol{\theta}}_n$, define the residuals fitted by the E-QMLE as $\widehat{\eta}_t = \varepsilon_t(\widehat{\boldsymbol{\alpha}}_n)/h_t(\widehat{\boldsymbol{\delta}}_n)$, and then $\widehat{\kappa}_1 = (n-p)^{-1}\sum_{t=p+1}^n \widehat{\eta}_t$ and $\widehat{\kappa}_2 = (n-p)^{-1}\sum_{t=p+1}^n \widehat{\eta}_t^2 - 1$. We can estimate the density function $f(x)$ using the kernel density estimator $\widehat{f}(x) = (nb_n)^{-1}\sum_{t=p+1}^n K((x-\widehat{\eta}_t)/b_n)$, where $K(\cdot)$ is the kernel function and b_n is the bandwidth. Finally, we use sample averages to replace the expectations in Σ and Ω, $\widehat{\boldsymbol{\delta}}_n$ to replace $\boldsymbol{\delta}_0$, $\widehat{\kappa}_i$ to replace κ_i, for $i = 1, 2$, and $\widehat{f}(0)$ to replace $f(0)$. Then, we can obtain estimates of Σ and Ω, denoted by $\widehat{\Sigma}_n$ and $\widehat{\Omega}_n$, respectively. Remark 1 provides basic conditions on the kernel function $K(\cdot)$ and the bandwidth b_n such that $\widehat{\Sigma}_n$ and $\widehat{\Omega}_n$ are consistent estimators of Σ and Ω, respectively; see the Supplementary Material for the proof.

Remark 1. *Suppose the conditions in Theorem 2 hold and* $\sup_x |f'(x)| < \infty$. *If there exists a positive number* L *such that* $|K(x) - K(y)| \le L|x-y|$, *for any* x, y, *and* $b_n \to 0$ *and* $nb_n^4 \to \infty$ *as* $n \to \infty$, *then* $\widehat{\Sigma}_n \to_p \Sigma$ *and* $\widehat{\Omega}_n \to_p \Omega$ *as* $n \to \infty$; *see also Corollary 1 in Zhu and Ling (2013).*

Numerous choices for the kernel function $K(\cdot)$ and bandwidth b_n satisfy the conditions in Remark 1. The kernel density estimator is robust to the selection of the kernel functions, but sensitive to that of the bandwidths. Thus, we suggest using the optimal bandwidth related to the selected kernel function in practice. Here, we use the Gaussian kernel function and its

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0049

rule-of-thumb bandwidth $b_n = 0.9n^{-1/5}\min\{s, \widehat{R}/1.34\}$ for our numerical studies in Sections 4–5, where s and $\widehat{R}$ are the sample standard deviation and interquartile of the residuals $\{\widehat{\eta}_t\}$, respectively; see also Zhu, Zheng, and Li (2018).

2.2 Comparison with the Gaussian quasi-maximum likelihood estimation

A Gaussian quasi-maximum likelihood estimation provides a popular QMLE called the G-QMLE. Assuming that $\{y_1, \ldots, y_n\}$ are generated by model (1.1), with $E(\eta_t) = 0$ and $\text{var}(\eta_t) = 1$, the G-QMLE of $\boldsymbol{\theta}_0$ is defined as $\widetilde{\boldsymbol{\theta}}_n = (\widetilde{\boldsymbol{\alpha}}_n', \widetilde{\boldsymbol{\delta}}_n')' = \text{argmin}_{\boldsymbol{\theta}\in\Theta} L_n^G(\boldsymbol{\theta})$, where

$$L_n^G(\boldsymbol{\theta}) = \frac{1}{n-p}\sum_{t=p+1}^{n} \ell_t^G(\boldsymbol{\theta}) \quad \text{and} \quad \ell_t^G(\boldsymbol{\theta}) = \ln h_t(\boldsymbol{\delta}) + \frac{\varepsilon_t^2(\boldsymbol{\alpha})}{2h_t^2(\boldsymbol{\delta})},$$

and $\varepsilon_t(\boldsymbol{\alpha})$ and $h_t(\boldsymbol{\delta})$ are defined as in (2.2); see also Ling (2007a) and Tan and Zhu (2022). Instead of Assumption 3 for the E-QMLE, we require the following assumption to establish the asymptotic properties of the G-QMLE.

Assumption 4. *(i) η_t has a zero mean and unit variance; (ii) $E(\eta_t^4) < \infty$.*

Let $\kappa_3 = E(\eta_t^3)$ and $\kappa_4 = E(\eta_t^4) - 1$. Define the $(2p+1)\times(2p+1)$ matrices

$$\Sigma_1 = \text{diag}\{E(\boldsymbol{Y}_{1t}\boldsymbol{Y}_{1t}'), 2E(\boldsymbol{Y}_{2t}\boldsymbol{Y}_{2t}')\} \text{ and } \Omega_1 = \begin{pmatrix} E(\boldsymbol{Y}_{1t}\boldsymbol{Y}_{1t}') & \kappa_3 E(\boldsymbol{Y}_{1t}\boldsymbol{Y}_{2t}') \\ \kappa_3 E(\boldsymbol{Y}_{2t}\boldsymbol{Y}_{1t}') & \kappa_4 E(\boldsymbol{Y}_{2t}\boldsymbol{Y}_{2t}') \end{pmatrix}.$$

Similarly to Theorem 2 of Tan and Zhu (2022), we can show the consistency and asymptotic normality of the G-QMLE $\widetilde{\boldsymbol{\theta}}_n$ under fractional moments of y_t for model (1.1).

Theorem 3. *Suppose that Assumptions 1, 2, and 4(i) hold. Then,*

(i) $\widetilde{\boldsymbol{\theta}}_n \to_p \boldsymbol{\theta}_0$ as $n \to \infty$;

(ii) furthermore, if Assumption 4(ii) holds and the matrix $D = \begin{pmatrix} 1 & \kappa_3 \\ \kappa_3 & \kappa_4 \end{pmatrix}$ is positive definite, then $\sqrt{n}(\widetilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \to_{\mathcal{L}} N(\boldsymbol{0}, \Xi_1)$ as $n \to \infty$, where $\Xi_1 = \Sigma_1^{-1}\Omega_1\Sigma_1^{-1}$.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0049

We next compare the E-QMLE with the G-QMLE. Because the two estimators need different conditions on the innovation, we assume $\kappa_1 = E(\eta_t) = 0$, and reparametrize model (1.1) under Assumption 3 to ensure that the innovation term has a zero mean and unit variance. Specifically, we consider the following reparametrized model:

$$y_t = \sum_{i=1}^{p} \alpha_i y_{t-i} + \eta_t^* \left(\omega^* + \sum_{i=1}^{p} \beta_i^* |y_{t-i}| \right), \tag{2.3}$$

where $\eta_t^* = \eta_t / \sqrt{E(\eta_t^2)}$ and $\boldsymbol{\delta}^* = (\omega^*, \beta_1^*, \beta_2^*, \ldots, \beta_p^*)' = \boldsymbol{\delta}\sqrt{E(\eta_t^2)}$. Denote $\boldsymbol{\theta}^* = (\boldsymbol{\alpha}', \boldsymbol{\delta}^{*\prime})'$ as the unknown parameter vector of model (2.3) and $\boldsymbol{\theta}_0^* = (\boldsymbol{\alpha}_0', \boldsymbol{\delta}_0^{*\prime})'$ as the true value. Let $\widetilde{\boldsymbol{\theta}}_n^*$ be the G-QMLE of $\boldsymbol{\theta}_0^*$, $\kappa_3^* = E(\eta_t^{*3})$, $\kappa_4^* = E(\eta_t^{*4}) - 1$, and $\boldsymbol{Y}_{it}^* = \boldsymbol{Y}_{it} / \sqrt{E(\eta_t^2)}$, for $i = 1$ and 2. If $E(\eta_t^{*4}) < \infty$ and Assumptions 1–2 hold, then by Theorem 3, we have $\sqrt{n}(\widetilde{\boldsymbol{\theta}}_n^* - \boldsymbol{\theta}_0^*) \to_{\mathcal{L}} N(\mathbf{0}, \Xi_1^*)$ as $n \to \infty$, where $\Xi_1^* = (\Sigma_1^*)^{-1}\Omega_1^*(\Sigma_1^*)^{-1}$, with

$$\Sigma_1^* = \operatorname{diag}\{E(\boldsymbol{Y}_{1t}^*\boldsymbol{Y}_{1t}^{*\prime}), 2E(\boldsymbol{Y}_{2t}^*\boldsymbol{Y}_{2t}^{*\prime})\} \text{ and } \Omega_1^* = \begin{pmatrix} E(\boldsymbol{Y}_{1t}^*\boldsymbol{Y}_{1t}^{*\prime}) & \kappa_3^* E(\boldsymbol{Y}_{1t}^*\boldsymbol{Y}_{2t}^{*\prime}) \\ \kappa_3^* E(\boldsymbol{Y}_{2t}^*\boldsymbol{Y}_{1t}^{*\prime}) & \kappa_4^* E(\boldsymbol{Y}_{2t}^*\boldsymbol{Y}_{2t}^{*\prime}) \end{pmatrix}.$$

Note that $\boldsymbol{\theta}_0 = R\boldsymbol{\theta}_0^*$, where $R = \operatorname{diag}\{I_p, [E(\eta_t^2)]^{-1/2} I_{p+1}\}$, with I_m being the $m \times m$ identity matrix. Then, we have $\sqrt{n}(\widetilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \to_{\mathcal{L}} N(\mathbf{0}, R\Xi_1^* R')$ as $n \to \infty$, where $\widetilde{\boldsymbol{\theta}}_n = R\widetilde{\boldsymbol{\theta}}_n^*$ is the G-QMLE of $\boldsymbol{\theta}_0$ for model (1.1) under Assumption 3. As a result, it is sufficient to compare the asymptotic covariance Ξ with $R\Xi_1^* R'$ for some specific cases:

(i) If η_t follows the standard Laplace distribution, such that $\kappa_1 = 0, \kappa_2 = 1$, and $f(0) = 1/2$, then the E-QMLE reduces to the MLE, with the asymptotic covariance $\Xi = \operatorname{diag}\{[E(\boldsymbol{Y}_{1t}\boldsymbol{Y}_{1t}')]^{-1}, [E(\boldsymbol{Y}_{2t}\boldsymbol{Y}_{2t}')]^{-1}\}$ attaining the Cramér–Rao lower bound, indicating that the E-QMLE is more efficient than the G-QMLE.

(ii) If η_t is standard normal, such that $\kappa_3 = 0, \kappa_4 = 2$, and $\Omega_1 = \Sigma_1$, then the G-QMLE reduces to the MLE with the asymptotic covariance $R\Xi_1^* R' = \Sigma_1^{-1}$, attaining the Cramér–Rao lower bound, and thus the G-QMLE is more efficient than the E-QMLE.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0049

(iii) For other situations of η_t, it is difficult to compare the asymptotic relative efficiency (ARE) of $\widehat{\boldsymbol{\theta}}_n$ to $\widetilde{\boldsymbol{\theta}}_n$. However, given the true parameter vector $\boldsymbol{\theta}_0$ and the density function $f(\cdot)$, we can obtain a theoretical value of $f(0)$, and estimate the matrices in Ξ and $R\Xi_1^*R'$ using sample averages and empirical values of κ_i or κ_i^*, based on a large generated sequence. Then, the ARE of $\widehat{\boldsymbol{\theta}}_n$ to $\widetilde{\boldsymbol{\theta}}_n$ can be calculated by $\text{ARE}(\widehat{\boldsymbol{\theta}}_n, \widetilde{\boldsymbol{\theta}}_n) = (|R\Xi_1^*R'|/|\Xi|)^{1/(2p+1)}$, where $|\cdot|$ is the determinant of a matrix; see Serfling (2009). The simulation results in Section 4 indicate that the E-QMLE is asymptotically more efficient than the G-QMLE when the data are more heavy tailed.

2.3 Comparison with the DWQRE

Consider the DWQRE of Zhu, Zheng, and Li (2018). With the identification condition that $\omega = 1$ in model (1.1), they consider the following reparametrized model:

$$y_t = \sum_{i=1}^{p} \alpha_i y_{t-i} + \varepsilon_t \left(1 + \sum_{i=1}^{p} \beta_i^\star |y_{t-i}|\right), \tag{2.4}$$

where $\varepsilon_t = \omega\eta_t$ and $\beta_i^\star = \beta_i/\omega$. Denote $\boldsymbol{\theta}^\star = (\boldsymbol{\alpha}', \boldsymbol{\beta}^{\star\prime})'$ as the unknown parameter vector of model (2.4) and $\boldsymbol{\theta}_0^\star = (\boldsymbol{\alpha}_0', \boldsymbol{\beta}_0^{\star\prime})'$ as the true value, where $\boldsymbol{\beta}^\star = (\beta_1^\star, \ldots, \beta_p^\star)'$. The DWQRE of $\boldsymbol{\theta}_0^\star$ is defined as $\check{\boldsymbol{\theta}}_n^\star = \sum_{k=1}^K \pi_k \check{\boldsymbol{\theta}}_{\tau_k n}^\star$, which combines the self-weighted quantile regression estimators $\check{\boldsymbol{\theta}}_{\tau_k n}^\star$ at K quantile levels using weighting matrices π_k, the optimal choices of which π_k^{opt} are defined in Theorem 4 of Zhu, Zheng, and Li (2018). The asymptotic properties of the DWQRE are also established under the finite fractional moment of $\{y_t\}$, which makes it possible to handle heavy-tailed data. Specifically, Zhu, Zheng, and Li (2018) show that the optimal DWQRE $\check{\boldsymbol{\theta}}_n^{\star opt} = \sum_{k=1}^K \check{\pi}_k^{opt} \check{\boldsymbol{\theta}}_{\tau_k n}^\star$ satisfies $\sqrt{n}(\check{\boldsymbol{\theta}}_n^{\star opt} - \boldsymbol{\theta}_0^\star) \to_{\mathcal{L}} N(\mathbf{0}, \Xi_2)$ as $n \to \infty$, where Ξ_2 is defined in their Theorem 4, and $\check{\pi}_k^{opt}$ are the estimated optimal weighting matrices. Moreover, they verify that $\check{\boldsymbol{\theta}}_n^{\star opt}$ approaches the efficiency of the MLE if $K \to \infty$ and the

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0049

following condition holds:

$$E(h_t^{-1}\boldsymbol{Y}_{1t}) = 0 \quad \text{and} \quad E(\boldsymbol{Y}_{at}'\boldsymbol{Y}_{1t}) = 0, \tag{2.5}$$

where $\boldsymbol{Y}_{at} = h_t^{-1}(|y_{t-1}|, \ldots, |y_{t-p}|)'$, with $h_t = 1 + \sum_{i=1}^{p} \beta_{i0}^{\star}|y_{t-i}|$. Condition (2.5) implies that the parameters in the conditional mean (i.e., $\boldsymbol{\alpha}$) and conditional scale (i.e., $\boldsymbol{\beta}^{\star}$) can be estimated separately, without loss of efficiency, which is satisfied when η_t is symmetrically distributed about zero and all α_i in model (2.4) are zero. However, for other general cases, the asymptotic covariance Ξ_2 cannot attain the Cramér–Rao lower bound.

Next, we compare the E-QMLE with the DWQRE. Note that if η_t follows the standard Laplace distribution, then the E-QMLE reduces to the MLE, such that the asymptotic covariance Ξ attains the Cramér–Rao lower bound. Therefore, the E-QMLE is asymptotically more efficient than the DWQRE if η_t follows the standard Laplace distribution. In contrast, the DWQRE is asymptotically more efficient than the E-QMLE only if Condition (2.5) holds, with infinite quantile levels used for estimation, and η_t does not follow the standard Laplace distribution. For general comparisons, denote the E-QMLE of model (2.4) as $\widehat{\boldsymbol{\theta}}_n^{\star} = g(\widehat{\boldsymbol{\theta}}_n)$, where $g : \mathbb{R}^p \times \mathbb{R}_+^{p+1} \to \mathbb{R}^p \times \mathbb{R}_+^p$ is a measurable transformation, such that $g(\boldsymbol{\theta}) = \boldsymbol{\theta}^{\star}$. Then, using the delta method, we have $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n^{\star} - \boldsymbol{\theta}_0^{\star}) \to_{\mathcal{L}} N(\boldsymbol{0}, \dot{g}(\boldsymbol{\theta}_0)\Xi\dot{g}'(\boldsymbol{\theta}_0))$ as $n \to \infty$, where $\dot{g}(\boldsymbol{\theta}_0)$ is the first derivative of $g(\boldsymbol{\theta}_0)$, defined as the following $2p \times (2p+1)$ matrix:

$$\dot{g}(\boldsymbol{\theta}_0) = \begin{pmatrix} I_p & 0_{p\times 1} & 0_{p\times p} \\ 0_{p\times p} & -\omega_0^{-2}\boldsymbol{\beta}_0 & \omega_0^{-1}I_p \end{pmatrix},$$

where $0_{m\times n}$ is an $m \times n$ zero matrix. Therefore, we can use the ARE of $\widehat{\boldsymbol{\theta}}_n^{\star}$ to $\check{\boldsymbol{\theta}}_n^{\star opt}$, defined by $\text{ARE}(\widehat{\boldsymbol{\theta}}_n^{\star}, \check{\boldsymbol{\theta}}_n^{\star opt}) = [|\Xi_2|/|\dot{g}(\boldsymbol{\theta}_0)\Xi\dot{g}'(\boldsymbol{\theta}_0)|]^{1/(2p)}$, to compare the E-QMLE with the optimal DWQRE. Here, $\text{ARE}(\widehat{\boldsymbol{\theta}}_n^{\star}, \check{\boldsymbol{\theta}}_n^{\star opt})$ can be computed by simulation, as for $\text{ARE}(\widehat{\boldsymbol{\theta}}_n, \widetilde{\boldsymbol{\theta}}_n)$ in Section 2.2. The simulation results in Section 4 indicate that neither model dominates for general

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0049

situations.

In addition, the two-step estimation procedure makes the DWQRE more complex in terms of computation than QMLEs obtained using a one-step estimation. In particular, given the same bounded maximum number of iterations in the optimization, $O(n)$ operations are required to obtain the E-QMLE, while $O(nK) + O(K^2)$ operations are needed for the DWQRE. Clearly, the computational load of the DWQRE can become much larger than that of the E-QMLE as $K \to \infty$, which makes using infinite quantile levels to construct the DWQRE infeasible in practice. As a result, the E-QMLE is preferred, because it can be more efficient for moderately heavy-tailed data and the computation is simpler than that of the DWQRE.

2.4 Model selection

For model (1.1) fitted using the E-QMLE, we introduce the following BIC for selection the order p:

$$\mathrm{BIC}(p) = 2(n - p_{\max})L_n^E(\widehat{\boldsymbol{\theta}}_n^p) + (2p+1)\ln(n - p_{\max}), \tag{2.6}$$

where $\widehat{\boldsymbol{\theta}}_n^p$ is the E-QMLE with the order set to p, $L_n^E(\widehat{\boldsymbol{\theta}}_n^p) = (n - p_{\max})^{-1}\sum_{t=p_{\max}+1}^{n} \ell_t^E(\widehat{\boldsymbol{\theta}}_n^p)$, and $p_{\max}$ is a predetermined positive integer; see also Machado (1993) and Zhu, Zheng, and Li (2018). Let $\widehat{p}_n = \arg\min_{1\le p\le p_{\max}} \mathrm{BIC}(p)$. The selection consistency of the BIC is given in the following theorem.

Theorem 4. *Let p_0 be the true order and p_{max} be a predetermined positive integer. Under the conditions of Theorem 2, if $p_{max} \ge p_0$, then $P(\widehat{p}_n = p_0) \to 1$ as $n \to \infty$.*

Theorem 4 shows that the BIC in (2.6) is robust in a similar way to the E-QMLE, that is, its selection consistency requires only $E(|y_t|^\kappa) < \infty$, for any $\kappa > 0$. The results of our

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0049

simulation studies in Section 4 indicate that the BIC performs well in finite samples.

Remark 2. *The BIC can also be defined for model* (1.1) *fitted using the G-QMLE. In this case, the BIC is* $BIC^G(p) = 2(n - p_{\max})L_n^G(\widetilde{\boldsymbol{\theta}}_n^p) + (2p + 1)\ln(n - p_{\max})$, *where* $\widetilde{\boldsymbol{\theta}}_n^p$ *is the G-QMLE with the order set to* p, *and* $L_n^G(\widetilde{\boldsymbol{\theta}}_n^p) = (n - p_{\max})^{-1}\sum_{t=p_{\max}+1}^{n}\ell_t^G(\widetilde{\boldsymbol{\theta}}_n^p)$; *see also Tan and Zhu (2022). Let* $\widehat{p}_n^G = \arg\min_{1\le p\le p_{\max}} BIC^G(p)$. *Similarly to Theorem 3 of Tan and Zhu (2022), we can prove the selection consistency of* $BIC^G(p)$ *under the conditions of Theorem 3.*

3. Model checking

To check the adequacy of the linear DAR models at (1.1) fitted using the E-QMLE, we construct a mixed portmanteau test to jointly detect possible misspecifications in the conditional mean and the conditional standard deviation; see also Wong and Ling (2005). We can conduct a diagnostic test of the conditional mean by checking the significance of the sample ACFs of the residuals (Ljung and Box, 1978); a similar test of the conditional standard deviation can be done by checking the significance of the sample ACFs of the absolute residuals for robustness (Li and Li, 2005).

The ACFs of $\{\eta_t\}$ and $\{|\eta_t|\}$ at lag k are defined by $\rho_k = \mathrm{cov}(\eta_t, \eta_{t-k})/\mathrm{var}(\eta_t)$ and $\gamma_k = \mathrm{cov}(|\eta_t|, |\eta_{t-k}|)/\mathrm{var}(|\eta_t|)$, respectively. If the data-generating process is specified correctly by model (1.1), then $\{\eta_t\}$ and $\{|\eta_t|\}$ are independent and identically distributed (*i.i.d.*), such that $\rho_k = 0$ and $\gamma_k = 0$ hold, for any $k \ge 1$. Define the error function as $\eta_t(\boldsymbol{\theta}) = \varepsilon_t(\boldsymbol{\alpha})/h(\boldsymbol{\delta})$. For model (1.1) fitted using the E-QMLE, the corresponding residuals are computed as $\widehat{\eta}_t = \varepsilon_t(\widehat{\boldsymbol{\alpha}}_n)/h_t(\widehat{\boldsymbol{\delta}}_n)$. Accordingly, the residual ACF and absolute residual ACF at lag k are

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0049

defined as

$$\widehat{\rho}_k = \frac{\sum_{t=p+k+1}^{n}(\widehat{\eta}_t - \widehat{\eta}_1)(\widehat{\eta}_{t-k} - \widehat{\eta}_1)}{\sum_{t=p+1}^{n}(\widehat{\eta}_t - \widehat{\eta}_1)^2} \text{ and } \widehat{\gamma}_k = \frac{\sum_{t=p+k+1}^{n}(|\widehat{\eta}_t| - \widehat{\eta}_2)(|\widehat{\eta}_{t-k}| - \widehat{\eta}_2)}{\sum_{t=p+1}^{n}(|\widehat{\eta}_t| - \widehat{\eta}_2)^2},$$

respectively, where $\widehat{\eta}_1 = (n-p)^{-1}\sum_{t=p+1}^{n}\widehat{\eta}_t$ and $\widehat{\eta}_2 = (n-p)^{-1}\sum_{t=p+1}^{n}|\widehat{\eta}_t|$. Note that $\widehat{\rho}_k$ is the sample version of ρ_k, whereas $\widehat{\gamma}_k$ is the sample version of γ_k. If the value of $\widehat{\rho}_k$ (or $\widehat{\gamma}_k$) deviates far from zero, then the conditional mean (or standard deviation) structure in model (1.1) may be misspecified.

For a prespecified positive integer M, denote $\widehat{\boldsymbol{\rho}} = (\widehat{\rho}_1, \ldots, \widehat{\rho}_M)'$ and $\widehat{\boldsymbol{\gamma}} = (\widehat{\gamma}_1, \ldots, \widehat{\gamma}_M)'$. Let $\sigma_1^2 = \text{var}(\eta_t)$ and $\sigma_2^2 = \text{var}(|\eta_t|)$. Define the $M \times (2p+1)$ matrices $U_\rho = (\boldsymbol{U}'_{\rho 1}, \ldots, \boldsymbol{U}'_{\rho M})'$ and $U_\gamma = (\boldsymbol{U}'_{\gamma 1}, \ldots, \boldsymbol{U}'_{\gamma M})'$, where $\boldsymbol{U}_{\rho k} = -\left(E\left[(\eta_{t-k} - \kappa_1)\boldsymbol{Y}'_{1t}\right], \kappa_1 E\left[(\eta_{t-k} - \kappa_1)\boldsymbol{Y}'_{2t}\right]\right)$ and $\boldsymbol{U}_{\gamma k} = -\left(\boldsymbol{0}'_p, E\left[(|\eta_{t-k}| - 1)\boldsymbol{Y}'_{2t}\right]\right)$, for $1 \le k \le M$, with $\boldsymbol{0}_p$ being a p-dimensional zero vector. Denote the $2M \times (2M + 2p + 1)$ matrix

$$V = \begin{pmatrix} I_M & 0_{M\times M} & U_\rho/\sigma_1^2 \\ 0_{M\times M} & I_M & U_\gamma/\sigma_2^2 \end{pmatrix}.$$

Let $\boldsymbol{G}_t = (\boldsymbol{Y}'_{1t}[I(\eta_t < 0) - I(\eta_t > 0)], \boldsymbol{Y}'_{2t}(1 - |\eta_t|))'$ and $G = E(\boldsymbol{v}_t\boldsymbol{v}'_t)$, where

$$\boldsymbol{v}_t = \big[(\eta_t - \kappa_1)(\eta_{t-1} - \kappa_1)/\sigma_1^2, \ldots, (\eta_t - \kappa_1)(\eta_{t-M} - \kappa_1)/\sigma_1^2,$$
$$(|\eta_t| - 1)(|\eta_{t-1}| - 1)/\sigma_2^2, \ldots, (|\eta_t| - 1)(|\eta_{t-M}| - 1)/\sigma_2^2, -\boldsymbol{G}'_t\Sigma_2^{-1}/2\big]'.$$

Theorem 5. *Suppose model* (1.1) *is specified correctly. Under the conditions of Theorem 2, then* $\sqrt{n}(\widehat{\boldsymbol{\rho}}', \widehat{\boldsymbol{\gamma}}')' \to_{\mathcal{L}} N(\boldsymbol{0}, VGV')$ *as* $n \to \infty$.

Theorem 5 can be used to check the significance of $\widehat{\rho}_k$ or $\widehat{\gamma}_k$ individually. Consistent estimators of V and G, denoted by $\widehat{V}$ and $\widehat{G}$, respectively, can be constructed by replacing the expectations with the sample averages and η_t with $\widehat{\eta}_t$. Then, we can estimate the asymptotic covariance in Theorem 5, and construct confidence intervals for $\widehat{\rho}_k$ and $\widehat{\gamma}_k$ accordingly.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0049

To check the first M lags jointly, we construct the following portmanteau test statistic:

$$Q(M) = n \begin{pmatrix} \widehat{\boldsymbol{\rho}} \\ \widehat{\boldsymbol{\gamma}} \end{pmatrix}' \left(\widehat{V}\widehat{G}\widehat{V}'\right)^{-1} \begin{pmatrix} \widehat{\boldsymbol{\rho}} \\ \widehat{\boldsymbol{\gamma}} \end{pmatrix}.$$

Theorem 5 and the continuous mapping theorem imply that $Q(M) \to_{\mathcal{L}} \chi^2_{2M}$ as $n \to \infty$, where χ^2_{2M} is the chi-squared distribution with $2M$ degrees of freedom. Therefore, if $Q(M)$ exceeds the $(1-\tau)$th quantile of the χ^2_{2M} distribution, we can reject the null hypothesis that ρ_k and γ_k $(1 \le k \le M)$ are jointly nonsignificant at level τ.

Remark 3. *The diagnostic tools can also be derived for model* (1.1)*, fitted using the G-QMLE. In this case, the residual ACF and absolute residual ACF at lag k are defined as*

$$\widetilde{\rho}_k = \frac{\sum_{t=p+k+1}^{n}(\widetilde{\eta}_t - \widetilde{\eta}_1)(\widetilde{\eta}_{t-k} - \widetilde{\eta}_1)}{\sum_{t=p+1}^{n}(\widetilde{\eta}_t - \widetilde{\eta}_1)^2} \text{ and } \widetilde{\gamma}_k = \frac{\sum_{t=p+k+1}^{n}(|\widetilde{\eta}_t| - \widetilde{\eta}_2)(|\widetilde{\eta}_{t-k}| - \widetilde{\eta}_2)}{\sum_{t=p+1}^{n}(|\widetilde{\eta}_t| - \widetilde{\eta}_2)^2},$$

respectively, where $\widetilde{\eta}_t = \varepsilon_t(\widetilde{\boldsymbol{\alpha}}_n)/h_t(\widetilde{\boldsymbol{\delta}}_n)$, $\widetilde{\eta}_1 = (n-p)^{-1}\sum_{t=p+1}^{n}\widetilde{\eta}_t$, and $\widetilde{\eta}_2 = (n-p)^{-1}\sum_{t=p+1}^{n}|\widetilde{\eta}_t|$. For a given positive integer M, denote $\widetilde{\boldsymbol{\rho}} = (\widetilde{\rho}_1, \ldots, \widetilde{\rho}_M)'$ and $\widetilde{\boldsymbol{\gamma}} = (\widetilde{\gamma}_1, \ldots, \widetilde{\gamma}_M)'$. Let $\tau_1 = E[\mathrm{sgn}(\eta_t)]$ and $\tau_2 = E(|\eta_t|)$. Define the $M \times (2p+1)$ matrices $U^G_\rho = (\boldsymbol{U}^{G\prime}_{\rho 1,G}, \ldots, \boldsymbol{U}^{G\prime}_{\rho M})'$ and $U^G_\gamma = (\boldsymbol{U}^{G\prime}_{\gamma 1}, \ldots, \boldsymbol{U}^{G\prime}_{\gamma M})'$, where $\boldsymbol{U}^G_{\rho k} = -\left(E(\eta_{t-k}\boldsymbol{Y}'_{1t}), \boldsymbol{0}'_{p+1}\right)$ and $\boldsymbol{U}^G_{\gamma k} = -(\tau_1 E[(|\eta_{t-k}| - \tau_2)\boldsymbol{Y}'_{1t}], \tau_2 E[(|\eta_{t-k}| - \tau_2)\boldsymbol{Y}'_{2t}])$, for $1 \le k \le M$. Let $\boldsymbol{G}_{1t} = (-\boldsymbol{Y}'_{1t}\eta_t, \boldsymbol{Y}'_{2t}(1-\eta_t^2))'$ and $G_1 = E(\boldsymbol{v}_{1t}\boldsymbol{v}'_{1t})$, where $\boldsymbol{v}_{1t} = (\eta_t\eta_{t-1}, \ldots, \eta_t\eta_{t-M}, (|\eta_t| - \tau_2)(|\eta_{t-1}| - \tau_2)/\sigma_2^2, \ldots, (|\eta_t| - \tau_2)(|\eta_{t-M}| - \tau_2)/\sigma_2^2, -\boldsymbol{G}'_{1t}\Sigma_1^{-1})'$. Similarly to Theorem 7 of Tan and Zhu (2022), $\sqrt{n}(\widetilde{\boldsymbol{\rho}}', \widetilde{\boldsymbol{\gamma}}')' \to_{\mathcal{L}} N(\boldsymbol{0}, V_1 G_1 V_1')$ as $n \to \infty$ if model (1.1) *is specified correctly and the conditions of Theorem 3 hold, where V_1 is defined as V, with U_ρ and U_γ replaced with U^G_ρ and U^G_γ, respectively. Then, we can check the significance of $\widetilde{\rho}_k$ or $\widetilde{\gamma}_k$ individually. In addition, we can use the following portmanteau test statistic to check the first M lags jointly:*

$$Q^G(M) = n \begin{pmatrix} \widetilde{\boldsymbol{\rho}} \\ \widetilde{\boldsymbol{\gamma}} \end{pmatrix}' \left(\widehat{V}_1\widehat{G}_1\widehat{V}_1'\right)^{-1} \begin{pmatrix} \widetilde{\boldsymbol{\rho}} \\ \widetilde{\boldsymbol{\gamma}} \end{pmatrix},$$

where $\widehat{V}_1$ *and* $\widehat{G}_1$ *are consistent estimators of* V_1 *and* G_1*, respectively. It can be shown that* $Q^G(M) \to_{\mathcal{L}} \chi^2_{2M}$ *as* $n \to \infty$ *under the null hypothesis that* ρ_k *and* γ_k *are jointly nonsignificant for* $1 \le k \le M$ *at level* τ*.*

Remark 4. *In practice, the choice of* M *may affect the performance of the portmanteau tests* $Q(M)$ *and* $Q^G(M)$*. For other portmanteau tests, the selection of* M *remains an open issue. Some general rules have been provided to choose* M *for Box–Pierce and Ljung–Box tests. For example, Box et al. (2015) recommend taking values of* M *between 10 and 20, and Tsay (2005) suggests using several choices of* M *and a general rule of* $M \approx \ln(n)$*, owing to its satisfactory power performance in simulation studies.*

Motivated by Tsay (2005), we conduct simulation studies to evaluate the size and power of $Q(M)$ *and* $Q^G(M)$ *with respect to the sample size* n *and the lag order* M*. We find that the size of a test is insensitive to* n*, and that the power is linearly increasing with respect to* n*. Moreover, the choice of* $M > 20$ *usually makes* $Q(M)$ *and* $Q^G(M)$ *undersize, and the logarithmic power is linearly decreasing with respect to* M*; see Section S1 of the Supplementary Material for more details. In general,* M *should be large enough to capture possible correlations among residuals and absolute residuals, but not be too large because of the resulting power loss. Thus, we suggest using multiple choices of* M*, such as* $(M_1, M_2, \ldots, M_J) = (\lfloor \ln(n) \rfloor, 2\lfloor \ln(n) \rfloor, \ldots, J\lfloor \ln(n) \rfloor)$*, where* $\lfloor x \rfloor$ *denotes the largest integer not greater than* x*, and* J *is the maximum value of* j*, such that* $M_j = j\lfloor \ln(n) \rfloor \le 20$*.*

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0049

4. Simulation experiments

4.1 E-QMLE and G-QMLE

The first experiment examines the finite-sample performance of the E-QMLE $\widehat{\boldsymbol{\theta}}_n$ and the G-QMLE $\widetilde{\boldsymbol{\theta}}_n$ in Sections 2.1–2.2, respectively, for which the data-generating process is

$$y_t = 0.5y_{t-1} + \eta_t(1 + 0.4|y_{t-1}|),$$

where $\{\eta_t\}$ are *i.i.d.* normal, Laplace or Student's t_3 distributed random variables. Here, $\{\eta_t\}$ are standardized with median zero and $E(|\eta_t|) = 1$ to evaluate the E-QMLE $\widehat{\boldsymbol{\theta}}_n$, and $\{\eta_t\}$ are standardized with mean zero and $\text{var}(\eta_t) = 1$ to evaluate the G-QMLE $\widetilde{\boldsymbol{\theta}}_n$. The sample size is set to $n = 500$ or 1000, with 1000 replications for each sample size.

Table 1 reports the biases, empirical standard deviations (ESDs), and asymptotic standard deviations (ASDs) of $\widehat{\boldsymbol{\theta}}_n$ and $\widetilde{\boldsymbol{\theta}}_n$ for different innovation distributions and sample sizes. We find that as the sample size increases, most of the biases, ESDs, and ASDs of both estimators $\widehat{\boldsymbol{\theta}}_n$ and $\widetilde{\boldsymbol{\theta}}_n$ become smaller, and the ESDs become closer to the corresponding ASDs. For the E-QMLE $\widehat{\boldsymbol{\theta}}_n$, the ESDs and ASDs of the scale-type estimators $\widehat{\omega}_n$ and $\widehat{\beta}_n$ increase as the distribution of η_t becomes more heavy tailed, while those of the location-type estimator $\widehat{\alpha}_n$ are smallest for the Laplace distribution. The mixed performance of the E-QMLE is probably because the heavier tail of $\{\eta_t\}$ makes the E-QMLE less efficient, but it becomes more efficient as it reduces to the MLE, if η_t follows the Laplace distribution. For the G-QMLE $\widetilde{\boldsymbol{\theta}}_n$, the ESDs and ASDs increase as the distribution of η_t becomes more heavy tailed. This is expected, because the G-QMLE reduces to the MLE if η_t follows the normal distribution, and the G-QMLE becomes less efficient as the tail of $\{\eta_t\}$ becomes heavier. Note that if η_t follows the Student's t_3 distribution, then $E(\eta_t^4) = \infty$ and the G-QMLE is not applicable,

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0049

which results in the inferior performance of the G-QMLE in this case. Similar observations can be found for other innovation distributions in the Supplementary Material.

4.2 Asymptotic efficiency comparison

The second experiment compares the ARE of the E-QMLE with those of the G-QMLE and DWQRE. We generate a sequence of sample size $n = 10000$ from the following model:

$$y_t = 0.1y_{t-1} + \varepsilon_t(1 + 0.2|y_{t-1}|),$$

where $\{\varepsilon_t\}$ are $i.i.d.$ random variables with the mixture distribution and the probability density function (pdf)

$$f(x) = (1 - \delta)\phi(x) + \delta m(x),$$

where $\delta \in [0, 1]$ is a constant, $\phi(x)$ is the pdf of $N(0, 1)$, and $m(x)$ is the pdf of $N(0, 6)$, standard Laplace, or t_3 distribution.

Figure 1 plots the ARE$(\widehat{\boldsymbol{\theta}}_n, \widetilde{\boldsymbol{\theta}}_n)$ and ARE$(\widehat{\boldsymbol{\theta}}^{\star}_n, \check{\boldsymbol{\theta}}^{\star opt}_n)$ defined in Sections 2.2 and 2.3 for $\delta = k/20$, with $k = 0, 1, \ldots, 20$, and different settings of $m(x)$, where the optimal DWQRE $\check{\boldsymbol{\theta}}^{\star opt}_n$ is obtained using $K = 9$ quantile levels. We have the following findings: (1) ARE$(\widehat{\boldsymbol{\theta}}_n, \widetilde{\boldsymbol{\theta}}_n)$ can be either larger or smaller than one, which suggests that neither the E-QMLE nor the G-QMLE dominate; furthermore the G-QMLE is more efficient than the E-QMLE when ε_t is closer to normal, whereas the E-QMLE becomes much more efficient as ε_t becomes more heavy tailed; (2) ARE$(\widehat{\boldsymbol{\theta}}^{\star}_n, \check{\boldsymbol{\theta}}^{\star opt}_n)$ can be either greater or less than one, which implies that neither the E-QMLE nor the DWQRE dominate; furthermore, the E-QMLE is more efficient than the DWQRE when ε_t approaches the Laplace distribution, but becomes less efficient than the DWQRE when ε_t becomes more heavy tailed; (3) when $\delta = 1$ and $m(x)$ is the pdf of a standard Laplace distribution, then ARE$(\widehat{\boldsymbol{\theta}}_n, \widetilde{\boldsymbol{\theta}}_n) > 1$ and ARE$(\widehat{\boldsymbol{\theta}}^{\star}_n, \check{\boldsymbol{\theta}}^{\star opt}_n) > 1$, indicating

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0049

that the E-QMLE is the most efficient. This is because the E-QMLE reduces to the MLE when ε_t follows a standard Laplace distribution, and thus its asymptotic covariance attains the Cramér–Rao lower bound; and (4) when $\delta = 0$, then $\text{ARE}(\widehat{\boldsymbol{\theta}}_n, \widetilde{\boldsymbol{\theta}}_n) < 1$, implying that the G-QMLE is the most efficient. This is because the G-QMLE reduces to the MLE when $\varepsilon_t \sim N(0,1)$, such that its asymptotic covariance attains the Cramér–Rao lower bound.

4.3 Model selection

In the third experiment, we evaluate the performance of the proposed model selection methods in Section 2.4, where the data-generating process is

$$y_t = 0.1y_{t-1} + 0.2y_{t-2} + \eta_t(1 + 0.1|y_{t-1}| + 0.2|y_{t-2}|),$$

and the innovations $\{\eta_t\}$ are defined as in the first experiment. We consider three sample sizes, $n = 300, 500$, and 1000, and generate 1000 replications for each sample size. The BIC in (2.6) and BIC^G in Remark 2 are employed to select the order p with $p_{\max} = 5$. As a result, the underfitted, correctly selected, and overfitted models by BIC (or BIC^G) correspond to $\widehat{p}_n$ (or $\widehat{p}_n^G$) being 1, 2, and greater than 2, respectively.

Table 2 provides the percentages of underfitting, correct selection, and overfitting cases by the BIC and BIC^G. Both BICs select the correct model in most of the replications when the sample size is as small as $n = 300$, and their performance improves as the sample size increases. Moreover, in terms of different distributions for the innovation η_t, the BIC in (2.6) performs best when η_t follows the Laplace distribution, whereas the BIC^G performs best when η_t follows the normal distribution, especially for small sample sizes. This is expected, because the E-QMLE reduces to the MLE if η_t follows the Laplace distribution, whereas the G-QMLE reduces to the MLE if η_t is normally distributed. Thus, no model misspecifications

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0049

appear when deriving the BIC and BIC^G, respectively, for these two cases. In addition, owing to the inferior performance of the G-QMLE when η_t follows the Student's t_3 distribution, BIC^G performs worst in this situation. We also consider other innovation distributions for both BICs in Section S1 of the Supplementary Material; the findings support those reported above.

4.4 Portmanteau tests

In the fourth experiment, we study the proposed mixed portmanteau tests $Q(M)$ and $Q^G(M)$. The data are generated from

$$y_t = 0.1y_{t-1} + c_1 y_{t-2} + \eta_t(1 + 0.2|y_{t-1}| + c_2|y_{t-2}|),$$

where all other settings are the same as those in the first experiment. We fit a linear DAR model with $p = 1$ using the exponential or Gaussian quasi-maximum likelihood estimation. Here, the case of $c_1 = c_2 = 0$ corresponds to the size of the test, $c_1 \neq 0$ corresponds to misspecifications in the conditional mean, and $c_2 > 0$ corresponds to misspecifications in the conditional standard deviation. Two departure levels, 0.1 and 0.3, are considered for both c_1 and c_2.

The rejection rates of $Q(6)$ and $Q^G(6)$ at the 5% significance level are summarized in Table 3. We have the following findings: First, all sizes are close to the nominal rate, except for those of $Q^G(6)$ in the t_3 case, and the power increases as the sample size n or the departure level increases. This is expected, because the G-QMLE is not applicable if η_t follows the t_3 distribution with $E(\eta_t^4) = \infty$, which makes the size inaccurate. Second, for the same level of departures, $Q(6)$ and $Q^G(6)$ are more powerful in terms of detecting the misspecification in the conditional mean $(c_1 \neq 0, c_2 = 0)$ than they are in doing so in the conditional standard

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0049

deviation ($c_1 = 0$, $c_2 > 0$). Third, as the innovation distribution becomes more heavy tailed, $Q(6)$ and $Q^G(6)$ perform worse in terms of detecting misspecifications in the conditional standard deviation. This seems consistent with the results in the first experiment, where the estimation performance of the scale parameters ω_0 and β_0 worsens as the innovation distribution becomes more heavy tailed. In addition, we also evaluate $Q(M)$ and $Q^G(M)$ with $M = 12$ and 18 in Section S1 of the Supplementary Material; once again, the findings are similar.

5. An empirical example

In this section, we apply the proposed inference tools to the weekly closing prices of Bitcoin (BTC) from July 18, 2010, to August 16, 2020, with 527 observations in total. We focus on their log returns (after mean adjustment), denoted by $\{y_t\}$. The time plot of $\{y_t\}$ in Figure 2 suggests evidence of volatility clustering, and the kurtosis of $\{y_t\}$ is 9.3, indicating that the tail of $\{y_t\}$ is much heavier than that of a normal distribution. Moreover, we can determine the autocorrelation from the sample partial autocorrelation functions (PACFs) of both $\{y_t\}$ and $\{|y_t|\}$. Therefore, we fit the data set $\{y_t\}$ using a linear DAR model.

We first employ the exponential quasi-maximum likelihood estimation method in Section 2.1 to fit $\{y_t\}$. Based on $p_{\max} = 10$, the proposed BIC in (2.6) selects the order $p = 3$ for the linear DAR model, and the fitted model is

$$\begin{aligned} y_t &= 0.0815_{0.0504} y_{t-1} + 0.1401_{0.0487} y_{t-2} + 0.0693_{0.0471} y_{t-3} + \widehat{\eta}_t \widehat{\sigma}_t, \\ \widehat{\sigma}_t &= 0.0435_{0.0065} + 0.2192_{0.0664} |y_{t-1}| + 0.1895_{0.0645} |y_{t-2}| + 0.1616_{0.0624} |y_{t-3}|, \end{aligned} \tag{5.1}$$

where the subscripts are the standard errors of the estimated coefficients, and $\{\widehat{\sigma}_t\}$ and $\{\widehat{\eta}_t\}$ are the fitted volatilities and residuals, respectively. The QQ plots of the fitted residuals

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0049

$\{\widehat{\eta}_t\}$ against Students' t_2, t_3, and t_4 distributions are shown in Figure 3. Here, the residuals are approximately t_3 distributed and the tail is heavier than t_4, but much lighter than t_2, possibly indicating that $E(\eta_t^2) < \infty$ and $E(\eta_t^4) = \infty$. We further employ the Kernelized Stein discrepancy (KSD) test proposed by Luo et al. (2021) to check whether η_t follows the standard Laplace or normal distribution. To calculate the KSD test statistic defined by (2.7) of Luo et al. (2021), we choose $n_0 = n = 526$ and use the Gaussian kernel $k(x, y) = \exp\{-\|x - y\|^2/(2\sigma^2)\}$, with σ being the median of the residual distance. The p-value is calculated using the parametric bootstrap, and is 0.37 for the standard Laplace distribution test, and less than 0.01 for the normal distribution test, suggesting that η_t follows the standard Laplace distribution and the E-QMLE may reduce to the MLE for the data $\{y_t\}$. Moreover, we perform mixed portmanteau tests $Q(M)$ for $M = 6, 12$, and 18, as in Section 3, and their p-values are $0.56, 0.71$, and 0.19, respectively. In addition, Figure 4 plots the residual ACFs $\widehat{\rho}_k$ and $\widehat{\gamma}_k$ up to lag 18, all of which fall within their corresponding 95% pointwise CIs, except for $\widehat{\gamma}_3$, which is slightly beyond its 95% CI. Clearly, almost all residual ACFs are nonsignificant, both individually and jointly, at the 5% significance level, and hence the fitted model at (5.1) is adequate.

For comparison, the Gaussian quasi-maximum likelihood estimation method in Section 2.2 is also used to fit $\{y_t\}$. Based on $p_{\max} = 10$, the BIC^G in Remark 2 selects the same order $p = 3$ for the linear DAR model, and obtains the following fitted model:

$$\begin{aligned} y_t &= 0.1098_{0.0579} y_{t-1} + 0.1268_{0.0547} y_{t-2} + 0.1733_{0.0586} y_{t-3} + \widetilde{\eta}_t \widetilde{\sigma}_t, \\ \widetilde{\sigma}_t &= 0.0821_{0.0146} + 0.2348_{0.1324} |y_{t-1}| + 0.1674_{0.1260} |y_{t-2}| + 0.2519_{0.1348} |y_{t-3}|, \end{aligned} \tag{5.2}$$

where the subscripts are the standard errors of the estimated coefficients, and $\{\widetilde{\sigma}_t\}$ and $\{\widetilde{\eta}_t\}$ are the fitted volatilities and residuals, respectively. The diagnosis from the residuals $\{\widetilde{\eta}_t\}$ in

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0049

the Supplementary Material indicates that the fitted model at (5.2) is adequate. Moreover, we estimate the linear DAR model of order three using the DWQRE of Zhu, Zheng, and Li (2018) for comparison. To facilitate a comparison between the E-QMLE and DWQRE, the volatility coefficients of the fitted model using the DWQRE are reparametrized to ensure $E(|\eta_t|) = 1$. As a result, the fitted model using the DWQRE method based on quantile levels $\tau_k = k/10$, for $k = 1, \ldots, 9$, is given by

$$\begin{aligned} y_t &= 0.1311_{0.0447} y_{t-1} + 0.0854_{0.0394} y_{t-2} + 0.0627_{0.0357} y_{t-3} + \check{\eta}_t \check{\sigma}_t, \\ \check{\sigma}_t &= 0.0384 + 0.3329_{0.0972} |y_{t-1}| + 0.2020_{0.0762} |y_{t-2}| + 0.1209_{0.0630} |y_{t-3}|, \end{aligned} \tag{5.3}$$

where the subscripts are the standard errors of the estimated coefficients, $\{\check{\sigma}_t\}$ are the fitted volatilities, and $\{\check{\eta}_t\}$ are the standardized residuals, such that $E(|\eta_t|) = 1$. To compare the efficiency of the three estimation methods, we approximate the $\mathrm{ARE}(\widehat{\boldsymbol{\theta}}_n, \widetilde{\boldsymbol{\theta}}_n)$ and $\mathrm{ARE}(\widehat{\boldsymbol{\theta}}^{\star}_n, \check{\boldsymbol{\theta}}^{\star opt}_n)$ defined in Sections 2.2 and 2.3, respectively, using parameter estimates and sample averages. We have $\widehat{\mathrm{ARE}}(\widehat{\boldsymbol{\theta}}_n, \widetilde{\boldsymbol{\theta}}_n) \approx 1.6223$ and $\widehat{\mathrm{ARE}}(\widehat{\boldsymbol{\theta}}^{\star}_n, \check{\boldsymbol{\theta}}^{\star opt}_n) \approx 0.8967$, implying that the E-QMLE is slightly less efficient than the DWQRE and that both are more efficient than the G-QMLE for fitting $\{y_t\}$. In addition, note that the conditional mean structures of the three fitted models (5.1)–(5.3) are significant at the 5% significance level, which suggests that the BTC market was not efficient during the examined period (Urquhart, 2016).

For financial time series, an important application of linear DAR models is to forecast risk measures, such as the value-at-risk (VaR). The VaR is actually a tail quantile of the loss series' conditional distribution, and thus the τth conditional quantile of y_t, denoted by $Q_{y_t}(\tau \mid \mathcal{F}_{t-1})$, is the negative τth VaR. To examine the forecasting performance of the linear DAR model estimated using the E-QMLE and G-QMLE, we conduct one-step-ahead predictions using a rolling forecasting procedure, with a fixed moving window of size 350.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0049

Specifically, we estimate the linear DAR model using the E-QMLE (or G-QMLE) for each moving window, and calculate the one-week-ahead forecast of the τth conditional quantile of y_{t+1} using $\widehat{Q}_{y_{t+1}}(\tau \mid \mathcal{F}_t) = \widehat{\mu}_{t+1} + \widehat{\sigma}_{t+1}\widehat{b}_\tau$ (or $\widetilde{Q}_{y_{t+1}}(\tau \mid \mathcal{F}_t) = \widetilde{\mu}_{t+1} + \widetilde{\sigma}_{t+1}\widetilde{b}_\tau$), where $\widehat{\mu}_{t+1}$ (or $\widetilde{\mu}_{t+1}$) and $\widehat{\sigma}_{t+1}$ (or $\widetilde{\sigma}_{t+1}$) are the predicted conditional mean and standard deviation, respectively, using the E-QMLE (or G-QMLE), and $\widehat{b}_\tau$ (or $\widetilde{b}_\tau$) is the τth sample quantile of the residuals $\{\widehat{\eta}_t\}$ (or $\{\widetilde{\eta}_t\}$). For example, the rolling one-week-ahead forecasts at $\tau = 5\%$ are displayed in Figure 2. Here, the negative VaRs based on the E-QMLE and G-QMLE are close to each other and change according to the volatility of the data, and y_t occasionally falls below its one-week negative VaR forecast.

We next compare the forecasting performance of the proposed E-QMLE and G-QMLE with that of the DWQRE. We conduct a rolling forecasting procedure with a fixed moving window of size 350 for the DWQRE approach, and compute the one-week-ahead forecast of the τth conditional quantile of y_{t+1} using $\check{Q}_{y_{t+1}}(\tau \mid \mathcal{F}_t) = \check{\mu}_{t+1} + \check{\sigma}_{t+1}\check{b}_\tau$, where $\check{\mu}_{t+1}$ and $\check{\sigma}_{t+1}$ are the predicted conditional mean and conditional standard deviation, respectively, using the DWQRE, and $\check{b}_\tau$ is the τth sample quantile of the corresponding residuals. To evaluate the forecasting performance of the three estimation methods, we calculate the empirical coverage rate (ECR), and perform VaR backtests for the forecasts at $\tau = 5\%$, 10%, 90%, and 95%. Specifically, the ECR is calculated as the proportion of observations that fall below the corresponding conditional quantile forecast for the last 176 data points. We use three VaR backtests, namely, the likelihood ratio tests for correct unconditional coverage (UC) in Kupiec (1995), correct conditional coverage (CC) in Christoffersen (1998), and dynamic quantile (DQ) test in Engle and Manganelli (2004). Let $H_t = I(y_t < Q_{y_t}(\tau \mid \mathcal{F}_{t-1}))$ be the hit, where $I(\cdot)$ is an indicator function. The UC test examines the accuracy of the VaR

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0049

forecasts using the null hypothesis that $E(H_t) = \tau$. The null hypothesis of the CC test is that, conditional on $\mathcal{F}_{t-1}$, $\{H_t\}$ are *i.i.d.* Bernoulli random variables with success probability τ. For the DQ test, we regress H_t on regressors including a constant, three lagged hits H_{t-i}, for $i = 1, 2, 3$, and the VaR forecast at the time point t. The null hypothesis of the DQ test is that the intercept is equal to τ and all regression coefficients are zero.

Table 4 reports the ECRs and p-values of the three VaR backtests for one-week-ahead forecasts obtained using the three estimation methods at four quantile levels. The proposed E-QMLE performs well at all four quantile levels, with p-values not less than 0.1 for the backtests. In addition, the proposed G-QMLE performs well, except for $\tau = 95\%$, with the p-value of the DQ test slightly smaller than 0.1, whereas the DWQRE performs less satisfactorily at three quantile levels in terms of the DQ tests. For the ECRs, those of the E-QMLE are closest to the nominal quantile level, except for $\tau = 90\%$. Overall, the linear DAR model fitted using the proposed E-QMLE method outperforms that of the DWQRE in forecasting VaRs. The G-QMLE method performs worse than the E-QMLE, probably because the G-QMLE is not suitable for the data.

In addition, to compare the forecasting performance of the linear DAR model with that of the DAR model (1.2) and the AR-GARCH model fitted using QMLEs, we apply the DAR model of order three and an AR(3)-GARCH(1, 1) model to the data using the E-QMLE and G-QMLE. The results based on the same rolling forecasting procedure are reported in Table 4. Here, both DAR models are comparable in the VaR backtests, and the linear DAR model fitted using the E-QMLE outperforms the DAR model in the ECRs for $\tau = 10\%, 90\%$, and 95%. This demonstrates the forecasting superiority of the linear DAR model for heavy-tailed data, which possibly benefits from its linear structure on the conditional standard deviation

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0049

instead of on the conditional variance for the DAR model. Furthermore, the linear DAR model is competitive with the AR-GARCH model in both the ECRs and the backtests.

In summary, the proposed E-QMLE procedure of a linear DAR model seems to be more reasonable and suitable for the considered BTC data set in terms of fitting and forecasting.

6. Conclusion

We have proposed two QMLEs, namely, the E-QMLE and the G-QMLE, for linear DAR models, which are simpler to compute than the DWQRE of Zhu, Zheng, and Li (2018). Under only a finite fractional moment of the process $\{y_t\}$, we establish the consistency and asymptotic normality for both QMLEs. We compare the E-QMLE with the G-QMLE and DWQRE in terms of asymptotic efficiency, and provide practical suggestions on choosing a suitable estimator. Moreover, we propose two BICs for order selection and two mixed portmanteau tests to check the adequacy of the fitted models based on the two QMLEs, and obtain their asymptotic properties without any moment conditions on the observed process. A real-data example confirms the usefulness and superiority of the proposed robust inference tools in terms of data fitting and forecasting.

The robust inference tools presented here can be extended in two directions. First, there is a practical need to consider a linear DAR model with different orders for the conditional location and scale components. The proposed robust inference tools can be adapted to such an extension using a self-weighting approach. Second, the linear DAR model can be generalized to a vector form to jointly model multivariate time series. In the framework of a vector linear DAR model, it would be interesting to investigate whether the good properties of robust inference can be preserved. We leave these topics for future research.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0049

Supplementary Material

The online Supplementary Material contains additional results for the simulation and the empirical analysis, as well as technical details for Theorems 1–5 and Remarks 1–3.

Acknowledgments

We are grateful to the co-editor, associate editor, and two anonymous referees for their valuable comments and suggestions. Liu and Tan are co-first authors, and contributed to the article equally. Zhu's research was supported by NSFC grant 12001355, Chenguang Program 19CG44, and Shanghai Research Center for Data Science and Decision Technology.

References

Aue, A. and L. Horváth (2011). Quasi-likelihood estimation in stationary and nonstationary autoregressive models with random coefficients. *Statistica Sinica 21*, 973–999.

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

Box, G. E., G. M. Jenkins, G. C. Reinsel, and G. M. Ljung (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Box, G. E. P., G. M. Jenkins, and G. C. Reinsel (2008). *Time Series Analysis, Forecasting and Control* (4th ed.). New York: Wiley.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International economic review 39*, 841–862.

Cryer, J. D. and K.-S. Chan (2008). *Time series analysis: with applications in R*. Springer Science & Business Media.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica 50*, 987–1007.

Engle, R. F. and S. Manganelli (2004). CAViaR: conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics 22*, 367–381.

Francq, C. and J.-M. Zakoian (2004). Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli 10*, 605–637.

Francq, C. and J. M. Zakoian (2019). *Garch models: structure, statistical inference and financial applications, 2nd edition*. Wiley.

Jiang, F., D. Li, and K. Zhu (2020). Non-standard inference for augmented double autoregressive models with null volatility coefficients. *Journal of Econometrics 215*, 165–183.

Kupiec, P. (1995). Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives 3*, 73–84.

Li, D., S. Ling, and J.-M. Zakoïan (2015). Asymptotic inference in multiple-threshold double autoregressive models. *Journal of Econometrics 189*, 415–427.

Li, D., S. Ling, and R. Zhang (2016). On a threshold double autoregressive model. *Journal of Business and Economic Statistics 34*, 68–80.

Li, G. and W. K. Li (2005). Diagnostic checking for time series models with conditional heteroscedasticity estimated by the least absolute deviation approach. *Biometrika 92*, 691–701.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0049

Li, G., Q. Zhu, Z. Liu, and W. K. Li (2017). On mixture double autoregressive time series models. *Journal of Business and Economic Statistics 35*, 306–317.

Li, W. K. (2004). *Diagnostic Checks in Time Series*. Boca Raton: Chapman and Hall.

Li, W. K., S. Ling, and M. McAleer (2002). Recent theoretical results for time series models with GARCH errors. *Journal of Economic Surveys 16*, 245–269.

Li, W. K. and T. Mak (1994). On the squared residual autocorrelations in non-linear time series with conditional heteroskedasticity. *Journal of Time Series Analysis 15*, 627–636.

Ling, S. (2004). Estimation and testing stationarity for double-autoregressive models. *Journal of the Royal Statistical Society, Series B 66*, 63–78.

Ling, S. (2007a). A double AR(p) model: structure and estimation. *Statistica Sinica 17*, 161–175.

Ling, S. (2007b). Self-weighted and local quasi-maximum likelihood estimators for ARMA-GARCH/IGARCH models. *Journal of Econometrics 140*, 849–873.

Ljung, G. M. and G. E. P. Box (1978). On a measure of lack of fit in time series models. *Biometrika 65*, 297–303.

Luo, D., K. Zhu, H. Gong, and D. Li (2021). Testing error distribution by kernelized stein discrepancy in multivariate time series models. *Journal of Business & Economic Statistics 0*, 1–15.

Machado, J. A. (1993). Robust model selection and M-estimation. *Econometric Theory 9*, 478–493.

Machado, J. A. F. (1990). Model selection: Consistency and robustness properties of the schwarz information criterion for generalized M-estimation. *Ph.D. thesis, University of Illinois, Urbana-Champaign.*

Pollard, D. (1985). New ways to prove central limit theorems. *Econometric Theory 1*, 295–313.

Poskitt, D. S. and A. R. Tremayne (1983). On the posterior odds of time series models. *Biometrika 70*, 157–162.

Schwarz, G. (1978). Estimaing the dimentions of a model. *Annals of Statistics 6*, 461–464.

Serfling, R. J. (2009). *Approximation theorems of mathematical statistics*, Volume 162. John Wiley & Sons.

Tan, S. and Q. Zhu (2022). Asymmetric linear double autoregression. *Journal of Time Series Analysis 43*, 371–388.

Taylor, S. J. (2008). *Modelling financial time series*. New York: World Scientific.

Tsay, R. S. (2005). *Analysis of financial time series*. John wiley & sons.

Urquhart, A. (2016). The inefficiency of bitcoin. *Economics Letters 148*, 80–82.

Weiss, A. A. (1986). Asymptotic theory for ARCH models: Estimation and testing. *Econometric Theory 2*, 107–131.

Wong, H. and S. Ling (2005). Mixed portmanteau tests for time-series models. *Journal of Time Series Analysis 26*, 569–579.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0049

Xiao, Z. and R. Koenker (2009). Conditional quantile estimation for generalized autoregressive conditional heteroscedasticity models. *Journal of the American Statistical Association 104*, 1696–1712.

Zhu, K. and S. Ling (2011). Global self-weighted and local quasi-maximum exponential likelihood estimators for ARMA-GARCH/IGARCH models. *The Annals of Statistics 39*, 2131–2163.

Zhu, K. and S. Ling (2013). Quasi-maximum exponential likelihood estimators for a double ar(p) model. *Statistica Sinica 23*, 251–270.

Zhu, Q., Y. Zheng, and G. Li (2018). Linear double autoregression. *Journal of Econometrics 207*, 162–174.

Hua Liu, Shanghai University of Finance and Economics, School of Statistics and Management

E-mail: liuhua@163.sufe.edu.cn

Songhua Tan, Shanghai University of Finance and Economics, School of Statistics and Management

E-mail: tansonghua@163.sufe.edu.cn

Qianqian Zhu, Shanghai University of Finance and Economics, School of Statistics and Management

E-mail: zhu.qianqian@mail.shufe.edu.cn

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0049

Table 1: Biases (×10), ESDs, and ASDs of the E-QMLE $\widehat{\boldsymbol{\theta}}_n$ and G-QMLE $\widetilde{\boldsymbol{\theta}}_n$ when the innovations follow the normal, Laplace, or Student's t_3 distribution.

		Normal			Laplace			t_3		
	n	Bias	ESD	ASD	Bias	ESD	ASD	Bias	ESD	ASD
					E-QMLE					
α	500	-0.002	0.065	0.069	-0.013	0.044	0.051	-0.039	0.052	0.056
	1000	0.006	0.047	0.048	-0.009	0.031	0.036	-0.017	0.037	0.039
ω	500	0.075	0.071	0.072	0.063	0.088	0.088	0.018	0.102	0.100
	1000	0.034	0.050	0.051	0.036	0.061	0.062	0.010	0.073	0.072
β	500	-0.059	0.045	0.045	-0.073	0.056	0.056	-0.027	0.070	0.066
	1000	-0.021	0.032	0.032	-0.036	0.039	0.039	-0.018	0.049	0.047
					G-QMLE					
α	500	0.000	0.050	0.051	-0.001	0.055	0.054	-0.002	0.061	0.058
	1000	0.000	0.036	0.036	0.000	0.039	0.038	0.000	0.042	0.042
ω	500	0.048	0.064	0.063	0.037	0.091	0.089	-0.279	0.201	0.150
	1000	0.026	0.044	0.045	0.035	0.064	0.064	-0.169	0.162	0.125
β	500	-0.060	0.052	0.051	-0.106	0.082	0.081	0.074	0.250	0.157
	1000	-0.025	0.037	0.036	-0.065	0.057	0.058	0.022	0.177	0.129

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0049

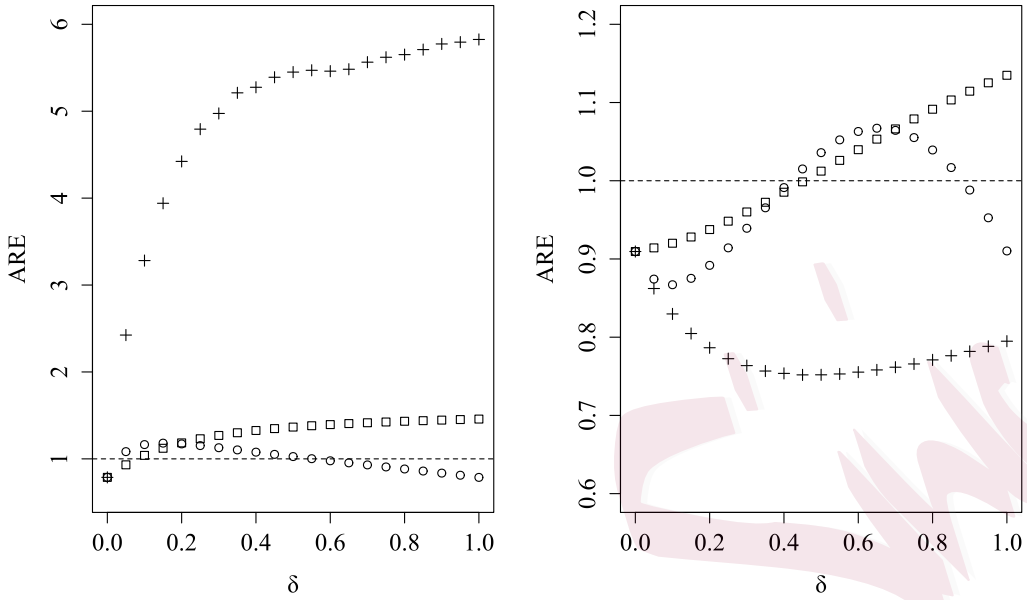

Figure 1: The ARE$(\widehat{\boldsymbol{\theta}}_n, \widetilde{\boldsymbol{\theta}}_n)$ (left panel) and ARE$(\widehat{\boldsymbol{\theta}}_n^{\star}, \check{\boldsymbol{\theta}}_n^{\star opt})$ (right panel) for $\delta = k/20$ ($k = 0, 1, \ldots, 20$), where $m(x)$ is the pdf of the $N(0, 6)$ ($\circ$), standard Laplace ($\square$), or t_3 ($+$) distribution.

Table 2: Percentages of underfitted, correctly selected, and overfitted models by the BIC and BIC^G when the innovations follow the normal, Laplace, or Student's t_3 distribution.

		Normal			Laplace			t_3		
	n	Under	Exact	Over	Under	Exact	Over	Under	Exact	Over
BIC	300	40.7	58.9	0.4	12.8	87.1	0.1	15.3	84.4	0.3
	500	15.1	84.5	0.4	2.1	97.4	0.5	3.0	96.1	0.9
	1000	1.0	99.0	0.0	0.0	100.0	0.0	0.0	99.8	0.2
BIC^G	300	33.5	66.5	0.0	32.2	66.6	1.2	32.7	58.2	9.1
	500	10.1	89.8	0.1	9.8	88.4	1.8	11.8	78.1	10.1
	1000	0.0	100.0	0.0	0.3	98.7	1.0	0.4	83.9	15.7

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0049

Table 3: Rejection rates of the tests $Q(6)$ and $Q^G(6)$ at the 5% significance level, where the innovations follow the normal, Laplace, or Student's t_3 distribution.

			Normal		Laplace		t_3	
	c_1	c_2	500	1000	500	1000	500	1000
Q	0.0	0.0	0.041	0.042	0.049	0.051	0.049	0.048
	0.1	0.0	0.194	0.443	0.216	0.462	0.207	0.469
	0.3	0.0	0.996	1.000	0.995	1.000	0.977	0.998
	0.0	0.1	0.123	0.340	0.099	0.228	0.076	0.173
	0.0	0.3	0.898	1.000	0.633	0.975	0.487	0.874
Q^G	0.0	0.0	0.058	0.047	0.076	0.053	0.155	0.133
	0.1	0.0	0.228	0.449	0.252	0.485	0.340	0.532
	0.3	0.0	0.999	1.000	0.999	1.000	0.991	1.000
	0.0	0.1	0.106	0.194	0.091	0.132	0.151	0.167
	0.0	0.3	0.733	0.993	0.366	0.837	0.338	0.630

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0049

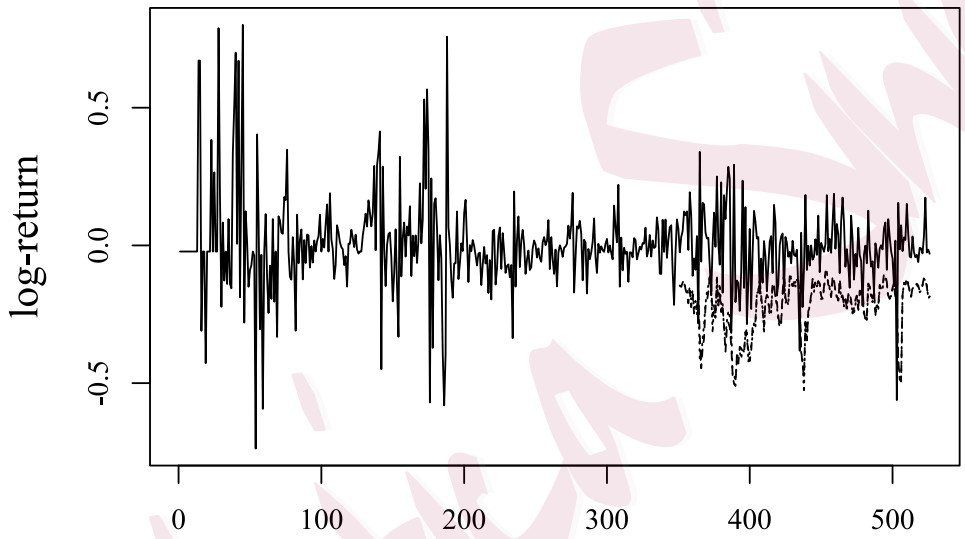

Figure 2: Time plot for centered weekly log returns in percentage (black line) of BTC from July 18, 2010, to August 16, 2020, with one-week negative VaR forecasts at the level of 5% based on the G-QMLE (dotted line) and the E-QMLE (dashed line) from March 26, 2017, to August 16, 2020.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0049

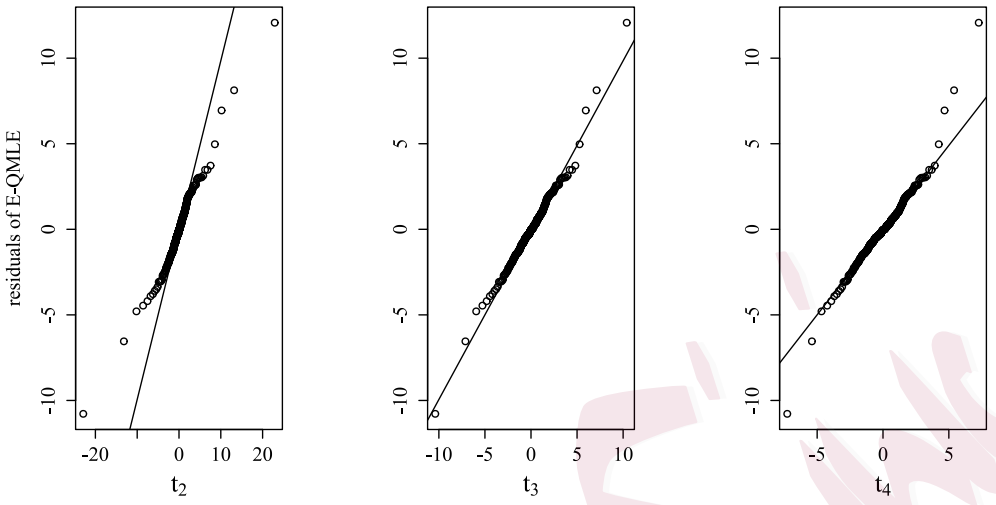

Figure 3: QQ plots of the residuals $\{\widehat{\eta}_t\}$ against the Student's t_2 (left panel), t_3 (middle panel), and t_4 (right panel) distributions.

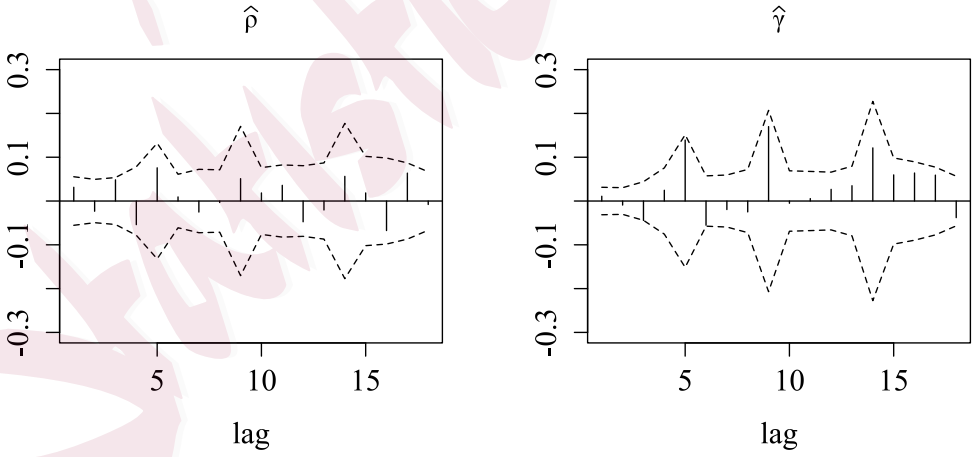

Figure 4: Residual ACF plots for $\widehat{\rho}_l$ (left panel) and $\widehat{\gamma}_l$ (right panel), where the dashed lines are the corresponding 95% pointwise confidence intervals.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0049

Table 4: Empirical coverage rates (%) and p-values of three VaR backtests at the 5%, 10%, 90%, and 95% conditional quantiles. M1, M2, and M3 represent the linear DAR model fitted using the E-QMLE, G-QMLE, and DWQRE, respectively; M4 and M5 represent the DAR model fitted using the E-QMLE and G-QMLE, respectively; and M6 and M7 represent the AR-GARCH model fitted using the E-QMLE and G-QMLE, respectively. The ECRs closest to the nominal level are marked in bold.

	$\tau = 5\%$				$\tau = 10\%$				$\tau = 90\%$				$\tau = 95\%$			
	ECR	UC	CC	DQ	ECR	UC	CC	DQ	ECR	UC	CC	DQ	ECR	UC	CC	DQ
M1	5.68	0.68	0.79	0.13	**10.23**	0.92	0.99	0.48	88.07	0.41	0.34	0.19	**94.89**	0.95	0.61	0.37
M2	6.25	0.46	0.71	0.22	10.80	0.73	0.94	0.31	88.64	0.55	0.48	0.29	93.75	0.46	0.37	0.09
M3	6.25	0.46	0.71	0.07	9.09	0.68	0.82	0.45	**89.77**	0.92	0.99	0.06	94.32	0.68	0.50	0.07
M4	5.68	0.68	0.50	0.71	11.93	0.41	0.67	0.54	86.93	0.20	0.33	0.67	94.32	0.68	0.50	0.73
M5	6.25	0.46	0.37	0.48	9.65	0.88	0.83	0.48	87.50	0.29	0.23	0.56	93.75	0.46	0.37	0.50
M6	6.82	0.29	0.56	0.14	9.66	0.88	0.83	0.50	89.20	0.73	0.09	0.34	95.45	0.79	0.66	0.71
M7	**5.11**	0.95	0.76	0.26	9.66	0.87	0.83	0.54	**89.77**	0.92	0.75	0.87	96.02	0.52	0.61	0.75