Econometric Theory, **38**, 2022, 536–561.
doi:10.1017/S0266466621000244

TEST FOR ZERO MEDIAN OF ERRORS IN AN ARMA–GARCH MODEL

YAOLAN MA
North Minzu University

MO ZHOU
Zhejiang University

LIANG PENG
Georgia State University

RONGMAO ZHANG
Zhejiang University
Minnan Normal University

Because the ARMA–GARCH model can generate data with some important properties such as skewness, heavy tails, and volatility persistence, it has become a benchmark model in analyzing financial and economic data. The commonly employed quasi maximum likelihood estimation (QMLE) requires a finite fourth moment for both errors and the sequence itself to ensure a normal limit. The self-weighted quasi maximum exponential likelihood estimation (SWQMELE) reduces the moment constraints by assuming that the errors and their absolute values have median zero and mean one, respectively. Therefore, it is necessary to test zero median of errors before applying the SWQMELE, as changing zero mean to zero median destroys the ARMA–GARCH structure. This paper develops an efficient empirical likelihood test without estimating the GARCH model but using the GARCH structure to reduce the moment effect. A simulation study confirms the effectiveness of the proposed test. The data analysis shows that some financial returns do not have zero median of errors, which cautions the use of the SWQMELE.

1. INTRODUCTION

Many economic and financial variables often exhibit skewness, heavy tails, time dependence, and volatility persistence. For capturing these stylized facts, a popular time series model is the so-called ARMA(r,s)–GARCH(p,q) model, defined as

We thank three reviewers for helpful comments, which improved the presentation. Peng's research was partly supported by the Simons Foundation and the NSF grant of DMS-2012448. Ma's research was partly supported by the First-Class Disciplines Foundation of Ningxia (No. NXYLXK2017B09), the Natural Science Foundation of Ningxia (No. 2019AAC03130), NSSFC (No. 20BTJ026), and General Scientific Research Project of North Minzu University (No. 2018XYZSX06). Zhang's research was partly supported by NSFC (No. 11771390/11371318), ZPNSFC (No. LZ21A010002), Ten Thousand Talents Plan of Zhejiang Province (No. 2018R52042), and the Fundamental Research Funds for the Central Universities. Corresponding address: Zhejiang University and Minnan Normal University, China; e-mail: rmzhang@zju.edu.cn.

© The Author(s), 2021. Published by Cambridge University Press.

https://doi.org/10.1017/S0266466621000244 Published online by Cambridge University Press

$$\begin{cases} X_t = \mu + \sum_{i=1}^{r} \phi_i X_{t-i} + \sum_{j=1}^{s} \psi_j \varepsilon_{t-j} + \varepsilon_t, \\ \varepsilon_t = \sigma_t \eta_t,\ \sigma_t^2 = w_1 + \sum_{i=1}^{p} a_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} b_j \sigma_{t-j}^2, \end{cases} \tag{1.1}$$

where $\mu \in \mathbb{R}$, $\phi_i \in \mathbb{R}$ $(i = 1, \ldots, r)$, $\psi_j \in \mathbb{R}$ $(j = 1, \ldots, s)$, $w_1 > 0$, $a_i \geq 0$ $(i = 1, 2, \ldots, p)$, $b_j \geq 0$ $(j = 1, 2, \ldots, q)$, and $\{\eta_t\}$ is a sequence of independent and identically distributed (i.i.d.) random variables with mean zero and variance one. Because of $\mathbb{E}(\eta_t) = 0$ and $\mathbb{E}(\eta_t^2) = 1$, equation (1.1) models the conditional mean and conditional standard deviation of $\{X_t\}$ parametrically.

Since the seminal papers of Engle (1982) and Bollerslev (1986), researchers have paid much attention to the probabilistic properties and statistical inferences of GARCH and ARMA-GARCH models. Basrak, Davis, and Mikosch, 2002 study the tail behavior and the sample autocovariance function of a GARCH model. Mikosch and Starica (2000), Berkes, Horváth, and Kokoszka, 2003; Chan, Peng, and Zhang, 2012, and Zhang, Li, and Peng (2019) estimate the tail index of a GARCH(1,1) sequence. Hall and Yao (2003) show that the quasi-maximum likelihood estimation (QMLE) for a GARCH sequence has a stable law limit when $\mathbb{E}(\eta_t^4) = \infty$. Francq and Zakoian (2004) prove that the QMLE for the ARMA–GARCH model has a normal limit when both $\mathbb{E}(\varepsilon_t^4) < \infty$ and $\mathbb{E}(\eta_t^4) < \infty$.

Real data applications show that $\sum_{i=1}^{p} a_i + \sum_{j=1}^{q} b_j$ is quite close to one, indicating that the assumption of $\mathbb{E}(\varepsilon_t^4) < \infty$ is problematic. This motivates Ling (2007) to propose a self-weighted local quasi-maximum likelihood estimation for an ARMA–GARCH model, which has a normal limit when $\mathbb{E}(\eta_t^4) < \infty$ and $\mathbb{E}(|\varepsilon_t|^\tau) < \infty$ for some $\tau > 0$. It is not surprising that an estimation based on $\mathbb{E}(\eta_t) = 0$ and $\mathbb{E}(\eta_t^2) = 1$ cannot have a normal limit when $\mathbb{E}(\eta_t^4) = \infty$. To derive an inference with a normal limit by allowing $\mathbb{E}(\eta_t^4) = \infty$ and $\mathbb{E}(\varepsilon_t^4) = \infty$, one has to impose different conditions on η_t. A common trick is to develop a median-based inference. Peng and Yao (2003) derive the asymptotic normality of the least absolute deviation estimation (LADE) for a GARCH sequence by assuming that the median of η_t^2, rather than variance, is one. Zhu and Ling (2011) derive the asymptotic normality of the self-weighted quasi maximum exponential likelihood estimation (SWQMELE) for an ARMA–GARCH sequence by assuming that the median of η_t is zero rather than $\mathbb{E}(\eta_t) = 0$.

When the median of η_t in (1.1) is $d_0 \neq 0$, a simple transformation to have zero median is

$$X_t = \mu + \sum_{i=1}^{r} \phi_i X_{t-i} + \sum_{j=1}^{s} \psi_j \widetilde{\varepsilon}_{t-j} + \widetilde{\varepsilon}_t + \sum_{j=1}^{s} \psi_j d_0 \sigma_{t-j} + d_0 \sigma_t,\ \widetilde{\varepsilon}_t = \sigma_t(\eta_t - d_0),$$

https://doi.org/10.1017/S0266466621000244 Published online by Cambridge University Press

which destroys the ARMA–GARCH structure due to the term $\sum_{j=1}^{s}\psi_j d_0\sigma_{t-j} + d_0\sigma_t$ and makes the SWQMELE not applicable. We refer to Fan, Qi, and Xiu (2014) for more details on the transformation effect for skewed data. Therefore, it is necessary to test whether η_t in (1.1) has zero median before employing the SWQMELE to fit the model.

Because zero median of η_t is equivalent to zero median of ε_t in the model (1.1), this paper develops a robust test for zero median of ε_t without knowing p and q and without estimating the unknown parameters of $w_1, \{a_i\}_{i=1}^{p}$, and $\{b_j\}_{j=1}^{q}$. To estimate the median of ε_t with a normal limit and without inferring the GARCH model, we cannot use the least squares estimation for parameter $\theta = (\mu, \phi_1, \ldots, \phi_r, \psi_1, \ldots, \psi_s)'$ in the ARMA model, as Zhu and Ling (2015) show that it has a non-normal limit when $\mathbb{E}(\varepsilon_t^4) = \infty$. Instead, we propose to estimate θ by the weighted least squares estimation in Huang et al. (2020) and then estimate the median of ε_t by a weighted median. Because the employed weight function is designed to reduce the moment effect of σ_t, although we do not infer the GARCH model, the proposed estimator for the median of ε_t will have a normal limit even when $\mathbb{E}(\varepsilon_t^2) = \infty$. Nonetheless and unsurprisingly, its asymptotic variance is complicated. Therefore, this paper develops an empirical likelihood test that can be conducted without estimating the asymptotic variance. We refer to Owen (2001) for an overview of the empirical likelihood method, which has been proved to be powerful in hypothesis tests. For applications of the empirical likelihood method to ARMA–GARCH models, we refer to Chan and Ling (2006) for a GARCH model, Li, Liang, and He (2012) for an AR-ARCH model, and Zhang et al. (2019) for the tail index of a GARCH(1,1) sequence.

In summary, this paper develops an empirical likelihood test for zero median of errors in an ARMA–GARCH model to ensure a proper application of the SWQMELE. The developed test is robust to volatility and heavy tails because we do not infer the GARCH model and allow heavy tailed errors. The test is efficient too, as the empirical likelihood method does not estimate the asymptotic variance explicitly and has been proved to be efficient.

We organize this paper as follows. Section 2 presents the methodologies and asymptotic results. Sections 3 and 4 are a simulation study and some data analyses, respectively. Some conclusions are summarized in Section 5. All proofs appear in Section 6.

2. METHODOLOGIES AND ASYMPTOTIC RESULTS

Let $\mathcal{F}_t$ denote the σ-field generated by $\{\varepsilon_s : s \le t\}$. Put $\theta = (\mu, \phi_1, \ldots, \phi_r, \psi_1, \ldots, \psi_s)'$, $\delta = (w_1, a_1, \ldots, a_p, b_1, \ldots, b_q)'$, and $\gamma = (\theta', \delta')'$ with γ_0 denoting the true value.

Write $\eta_t(\gamma_0) = \eta_t$, $\varepsilon_t(\theta_0) = \varepsilon_t$, and $\sigma_t(\gamma_0) = \sigma_t$. Define $\phi(z) = 1 - \sum_{i=1}^{r}\phi_i z^i$, $\psi(z) = 1 + \sum_{j=1}^{s}\psi_j z^j$, $a(z) = \sum_{i=1}^{p} a_i z^i$, and $b(z) = 1 - \sum_{j=1}^{q} b_j z^j$.

Given the observations $\{X_1, \ldots, X_n\}$ and the initial values $\{X_0, X_{-1}, \ldots\}$ taken as zero in our simulation study and data analysis, we write the parametric form of

https://doi.org/10.1017/S0266466621000244 Published online by Cambridge University Press

(1.1) as

$$\begin{cases} \varepsilon_t(\theta) = X_t - \mu - \sum_{i=1}^{r} \phi_i X_{t-i} - \sum_{j=1}^{s} \psi_j \varepsilon_{t-j}(\theta), \\ \sigma_t^2(\gamma) = w_1 + \sum_{i=1}^{p} a_i \varepsilon_{t-i}^2(\theta) + \sum_{j=1}^{q} b_j \sigma_{t-j}^2(\gamma), \ \eta_t(\gamma) = \dfrac{\varepsilon_t(\theta)}{\sigma_t(\gamma)}. \end{cases}$$

For testing whether ε_t has zero median, we cannot estimate θ by the least squares estimation, as Zhang and Ling (2015) show that it does not have a normal limit when $\mathbb{E}(\varepsilon_t^4) = \infty$. Instead, we propose to use the weighted least squares estimation by minimizing $\sum_{t=1}^{n} w_{t-1}^{-2} \varepsilon_t^2(\theta)$, which is equivalent to solve the score equations

$$\sum_{t=1}^{n} w_{t-1}^{-2} \varepsilon_t(\theta) \frac{\partial \varepsilon_t(\theta)}{\partial \theta} = 0,$$

where w_t is $\mathcal{F}_t$-measurable and will be defined later. As $\varepsilon_t(\theta) = \sigma_t(\gamma)\eta_t(\gamma)$, the estimation procedure above employs one w_{t-1} to reduce the moment effect of σ_t in $\varepsilon_t(\theta)$ and another one to reduce that in $\dfrac{\partial \varepsilon_t(\theta)}{\partial \theta}$. Hence, the proposed weighted least squares estimation, denoted by $\hat{\theta}$, has a normal limit due to $\mathbb{E}(\eta_t^2) = 1$ even when $\mathbb{E}(\varepsilon_t^2) = \infty$. Next, using $\hat{\theta}$, we can estimate the median of ε_t by solving

$$\sum_{t=1}^{n} w_{t-1}^{-1} \operatorname{sgn}(\varepsilon_t(\hat{\theta}) - d) = 0,$$

where $\operatorname{sgn}(x)$ is the sign function. Again, the weight w_{t-1} reduces the moment effect due to the plug-in estimator $\hat{\theta}$. So, the resulted median estimator has a normal limit without requiring $\mathbb{E}(\varepsilon_t^2) < \infty$, but its asymptotic variance is complicated. To avoid estimating the asymptotic variance, we propose to employ the empirical likelihood method based on estimating equations in Qin and Lawless (1994) as follows. Note that the empirical likelihood method is less computationally intensive than a bootstrap method. The residual-based bootstrap method is not applicable, as we do not estimate the GARCH model.

Put

$$\boldsymbol{D}_{t,1}(\theta) = w_{t-1}^{-2} \varepsilon_t(\theta) \frac{\partial \varepsilon_t(\theta)}{\partial \theta}, \ D_{t,2}(\theta, d) = w_{t-1}^{-1} \operatorname{sgn}(\varepsilon_t(\theta) - d), \ \text{and}$$
$$\boldsymbol{D}_t(\theta, d) = (\boldsymbol{D}_{t,1}'(\theta), D_{t,2}(\theta, d))',$$

where d is the median of $\varepsilon_t(\theta)$. Then we define the empirical likelihood function for θ and d as

$$L(\theta, d) = \sup \left\{ \prod_{t=1}^{n} (np_t) : p_1 \geq 0, \ldots, p_n \geq 0, \sum_{t=1}^{n} p_t = 1, \sum_{t=1}^{n} p_t \boldsymbol{D}_t(\theta, d) = \boldsymbol{0} \right\}.$$

https://doi.org/10.1017/S0266466621000244 Published online by Cambridge University Press

Using the Lagrange multiplier technique, we obtain $p_t = \dfrac{1}{n\{1+\lambda' \boldsymbol{D}_t(\theta,d)\}}$ and the log empirical likelihood ratio

$$l(\theta,d) = -2\log(L(\theta,d)) = 2\sum_{t=1}^{n}\log\{1+\lambda'\boldsymbol{D}_t(\theta,d)\},$$

where $\lambda = \lambda(\theta,d)$ satisfies

$$\sum_{t=1}^{n}\frac{\boldsymbol{D}_t(\theta,d)}{1+\lambda'\boldsymbol{D}_t(\theta,d)} = 0.$$

As we are interested in d, we consider the profile empirical likelihood ratio $l_p(d) = \min_{\theta\in\Theta} l(\theta,d)$.

Because $\mathbb{E}(\varepsilon_t(\theta_0)|\mathcal{F}_{t-1}) = 0$ and $\mathbb{E}(D_{t,2}(\theta_0,0)|\mathcal{F}_{t-1}) = w_{t-1}^{-1}\mathbb{E}(\mathrm{sgn}(\eta_t)|\mathcal{F}_{t-1}) = 0$ under the model assumption of $E\eta_t = 0$ and the null hypothesis of zero median of ε_t, we expect that the Wilks theorem holds for the above empirical likelihood method. However, if one tests for a nonzero value of d, the Wilks theorem does not hold because of $\mathbb{E}(D_{t,2}(\theta_0,d)|\mathcal{F}_{t-1}) \neq 0$.

Let

$$\boldsymbol{A}_t = \begin{pmatrix} a_1\eta_t^2+b_1 & b_2 & \cdots & b_{q-1} & b_q & a_2 & a_3 & \cdots & a_p \\ 1 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & 0 & 0 & \cdots & 0 \\ \eta_t^2 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix},$$

and γ be the Lyapunov exponent of the random matrices $\{\boldsymbol{A}_t\}$, namely, for any norm $\|\cdot\|$ on the space of $(p+q-1)\times(p+q-1)$ matrices,

$$\gamma = \inf\left\{\frac{1}{n}\mathbb{E}(\ln\|\boldsymbol{A}_1\cdots\boldsymbol{A}_n\|), n\in\mathrm{N}\right\},$$

where $\|\boldsymbol{A}_t\| = \sup_{|\boldsymbol{x}|=1}|\boldsymbol{A}_t\boldsymbol{x}|$. To prove the Wilks theorem of the proposed empirical likelihood test, we need the following regularity conditions.

Assumption 1. θ_0 is an interior point in Θ, and for each $\theta\in\Theta$, $\phi(z)\neq 0$ and $\psi(z)\neq 0$ when $|z|\leq 1$, and $\phi(z)$ and $\psi(z)$ have no common root with $\phi_r \neq 0$ or $\psi_s \neq 0$.

Assumption 2. $w_1 > 0$, $\gamma < 0$, and $\mathbb{E}(\ln(\max(|\eta_1|,1))) < \infty$.

https://doi.org/10.1017/S0266466621000244 Published online by Cambridge University Press

Assumption 3. $\mathbb{E}(w_{t-1}^{-4}\xi_{\rho,t-1}^{4}) < \infty$ for any $\rho \in (0,1)$, where $\xi_{\rho,t} = 1 + \sum_{i=0}^{\infty} \rho^i |X_{t-i}|$, w_t is positive and $\mathcal{F}_t$-measurable, $\inf_{t\geq 1} w_t > 0$, and $\{w_t\}$ is a stationary sequence.

Assumption 4. $\{\eta_t\}$ is a sequence of independent and identically distributed random variables with mean zero and variance one.

Assumption 5. $\{\eta_t\}$ has a continuous density function $g(x)$ satisfying $g(0) > 0$ and $\sup_{x\in\mathbb{R}} g(x) < \infty$.

Assumptions 1 and 2 ensure that there exists a unique strictly stationary causal solution to the first and second equations of (1.1), respectively (see Theorem 3.1 of Basrak et al. (2002) for the GARCH model). Assumption 3 allows the weight to reduce the moment effect of σ_t. Assumption 4 means that we consider the classical ARMA–GARCH model for conditional mean and conditional standard deviation. Assumption 5 is the standard condition for median estimation.

THEOREM 1. *Under Assumptions 1–5 and the null hypothesis of* $H_0 : d_0 = 0$, $l_p(0)$ *converges in distribution to a chi-squared limit with one degree of freedom as* $n \to \infty$.

For testing whether ε_t in (1.1) has zero median, the theorem above suggests rejecting the null hypothesis at level α whenever $l_p(0) \geq \chi^2_{1,1-\alpha}$ with $\chi^2_{1,1-\alpha}$ denoting the $(1-\alpha)$th quantile of a chi-squared distribution with one degree of freedom.

To investigate the local power of the proposed test, we consider the following local alternative hypothesis

$$H_a : d_0 = \frac{M_1}{\sqrt{n}} \text{ for some constant } M_1. \tag{2.2}$$

The following theorem shows that the proposed test has nontrivial power, its power tends to one as $|M_1| \to \infty$, and the weight function impacts the test power but in a complicated way.

THEOREM 2. *Suppose that Assumptions 1–5 hold for model (1.1). Under the alternative hypothesis of (2.2),* $l_p(0)$ *converges in distribution to a non-central chi-squared limit with one degree of freedom and the noncentrality parameter*

$$(\mathbf{0}', 2g(0)M_1\mathbb{E}(w_{t-1}^{-1})\mathbb{E}(\sigma_t^{-1}))\Lambda^{-1}(\mathbf{0}', 2g(0)M_1\mathbb{E}(w_{t-1}^{-1})\mathbb{E}(\sigma_t^{-1}))',$$

where $\Lambda = \mathbb{E}\big(\boldsymbol{D}_1(\theta_0,0)\boldsymbol{D}_1^{'}(\theta_0,0)\big)$.

Finally, we provide a class of weight functions. It is clear from Ling (2007), Zhu and Ling (2011, 2015) that the key idea of choosing the weight function is to

https://doi.org/10.1017/S0266466621000244 Published online by Cambridge University Press

bound $\xi_{\rho,t}$ defined in Assumption 3. There are many different choices, including the one in Ling (2007). Here, we propose to employ the following weight function

$$w_t(h) = \max(C, \sum_{i=0}^{t-1} e^{\log(h)\log^2(i+1)}|X_{t-i}|) \text{ for some } h \in (0,1) \text{ and } t = 1, \ldots, n, \tag{2.3}$$

where C is chosen as the 90% quantile of $\{|X_t|\}_{t=1}^n$, and $w_0(h) = 1$. Like He et al. (2020), we can show that the above weight function with the sample quantile C replaced by the corresponding quantile (denoted by $\widetilde{C}$) satisfies Assumption 3. Further, we can show estimators with C and $\widetilde{C}$ have the same asymptotic distribution. Like kernel density estimation, it is challenging to choose h in terms of coverage probability, which requires derivations of the Edgeworth expansion for the empirical likelihood ratio. Nevertheless, simulations below show that use of $h = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6$ provides accurate size and good power.

3. SIMULATION STUDY

This section examines the finite sample performance of the proposed empirical likelihood test in terms of size and power.

We generate 5000 random samples with sample size $n = 1000$ or 2000 from the ARMA(1,0)–GARCH(1,1) and ARMA(1,1)–GARCH(1,1) models with $\mu = 0.1, \phi_1 = 0.5, \psi_1 = 0.2$, $w_1 = 0.1$, $a_1 = 0.2$ or 0.1, $b_1 = 0.8$, $\eta_t = (V - \mathbb{E}(V))/\sqrt{\text{Var}(V)}$, where

$$V = \frac{I(U < \delta)V_1}{1/(\alpha_1 - 1)} - \frac{I(U \geq \delta)V_2}{1/(\alpha_2 - 1)}$$

with $U \sim \text{Uniform}(0,1)$, $V_1 \sim \text{Pareto}(1, \alpha_1)$, and $V_2 \sim \text{Pareto}(1, \alpha_2)$ being independent. Note that $\mathbb{E}\varepsilon_t^2$ is finite for $a_1 = 0.1$ and infinite for $a_1 = 0.2$. It is easy to check that

$$\mathbb{E}(V) = 2\delta - 1, \ \mathbb{E}(V^2) = \frac{2\delta(\alpha_1 - 1)}{\alpha_1 - 2} + \frac{2(1-\delta)(\alpha_2 - 1)}{\alpha_2 - 2},$$

$$P(\eta_t < 0) = 1 - \delta + \delta P\left(V_1 < \frac{2\delta - 1}{\alpha_1 - 1}\right) = 1 - \delta\left(1 + \frac{2\delta - 1}{\alpha_1 - 1}\right)^{-\alpha_1} \text{ for } 2\delta - 1 \geq 0,$$

and

$$P(\eta_t < 0) = \delta P\left(V_2 > \frac{1 - 2\delta}{\alpha_2 - 1}\right) = \delta\left(1 + \frac{1 - 2\delta}{\alpha_2 - 1}\right)^{-\alpha_2} \text{ for } 2\delta - 1 < 0.$$

Therefore, η_t has the right tail index α_1, the left tail index α_2, zero median if $\delta = 0.5$, and is asymmetric when $\alpha_1 \neq \alpha_2$. Some calculations show that $P(\eta_t < 0) \neq 1/2$ for $\delta \neq 1/2$, $\alpha_1 > 2$, and $\alpha_2 > 2$, i.e., η_t has nonzero median if $\delta \neq 1/2$. In contrast, if η_t has a standardized skewed t distribution with mean zero and variance one,

https://doi.org/10.1017/S0266466621000244 Published online by Cambridge University Press

TABLE 1. Test sizes for ARMA(1,0)–GARCH(1,1) and ARMA(1,1)–GARCH(1,1) models with $\delta = 0.5$

				ARMA(1,0)–GARCH(1,1)		ARMA(1,1)–GARCH(1,1)	
a_1	α_1	n	Level	$h = 0.1$	$h = 0.3$	$h = 0.1$	$h = 0.3$
0.2	2.2						
		1000	10%	0.1282	0.1292	0.1448	0.1458
			5%	0.0738	0.0752	0.0824	0.0868
			1%	0.0164	0.0174	0.0250	0.0264
		2000	10%	0.1194	0.1162	0.1326	0.1322
			5%	0.0608	0.0604	0.0744	0.0742
			1%	0.0148	0.0128	0.0212	0.0202
0.2	4.5						
		1000	10%	0.1160	0.1154	0.1168	0.1188
			5%	0.0586	0.0590	0.0626	0.0664
			1%	0.0122	0.0136	0.0194	0.0170
		2000	10%	0.1012	0.1040	0.1120	0.1156
			5%	0.0530	0.0520	0.0630	0.0612
			1%	0.0126	0.0098	0.0156	0.0144
0.1	2.2						
		1000	10%	0.1276	0.1298	0.1408	0.1424
			5%	0.0730	0.0708	0.0804	0.0856
			1%	0.0164	0.0186	0.0230	0.0262
		2000	10%	0.1194	0.1148	0.1378	0.1378
			5%	0.0606	0.0580	0.0756	0.0654
			1%	0.0128	0.0132	0.0208	0.0200
0.1	4.5						
		1000	10%	0.1144	0.1146	0.1164	0.1206
			5%	0.0590	0.0608	0.0654	0.0668
			1%	0.0128	0.0128	0.0172	0.0170
		2000	10%	0.1002	0.1022	0.1112	0.1134
			5%	0.0530	0.0492	0.0624	0.0594
			1%	0.0116	0.0098	0.0150	0.0156

then it has the same right and left tail index and is symmetric in the case of zero median. Therefore, the proposed distributions for η_t are more flexible in modeling tails and asymmetry than the widely used skewed t distributions.

We take $\alpha_1 = 2.2$ or 4.5, $\alpha_2 = 2.2$, and $\delta = 0.3,$ or 0.4, or 0.5. Hence, the case of $\delta = 0.5$ corresponds to the test size, and other choices of δ stand for the test power. We use the R package "emplik" to calculate the empirical likelihood function and the R function "optim" to compute the profile empirical likelihood function. We

https://doi.org/10.1017/S0266466621000244 Published online by Cambridge University Press

TABLE 2. Test powers for ARMA(1,0)–GARCH(1,1) and ARMA(1,1)–GARCH(1,1) models with $\delta = 0.3$

				ARMA(1,0)–GARCH(1,1)		ARMA(1,1)–GARCH(1,1)	
a_1	α_1	n	Level	$h = 0.1$	$h = 0.3$	$h = 0.1$	$h = 0.3$
0.2	2.2						
		1000	10%	0.8368	0.8344	0.8358	0.8322
			5%	0.7672	0.7584	0.7728	0.7648
			1%	0.5960	0.5918	0.5908	0.5880
		2000	10%	0.9394	0.9374	0.9478	0.9448
			5%	0.9082	0.9038	0.9164	0.9160
			1%	0.8110	0.8130	0.8218	0.8202
0.2	4.5						
		1000	10%	0.9514	0.9532	0.9498	0.9474
			5%	0.9178	0.9152	0.9086	0.9078
			1%	0.7806	0.7792	0.7696	0.7670
		2000	10%	0.9986	0.9990	0.9976	0.9986
			5%	0.9960	0.9970	0.9950	0.9952
			1%	0.9812	0.9822	0.9804	0.9820
0.1	2.2						
		1000	10%	0.8422	0.8382	0.8460	0.8382
			5%	0.7786	0.7654	0.7658	0.7734
			1%	0.6118	0.6000	0.6102	0.5988
		2000	10%	0.9426	0.9362	0.9508	0.9466
			5%	0.9114	0.9042	0.9200	0.9176
			1%	0.8252	0.8172	0.8318	0.8234
0.1	4.5						
		1000	10%	0.9590	0.9536	0.9584	0.9500
			5%	0.9282	0.9146	0.9226	0.9120
			1%	0.8040	0.7920	0.7958	0.7810
		2000	10%	0.9992	0.9990	0.9984	0.9984
			5%	0.9978	0.9962	0.9964	0.9962
			1%	0.9848	0.9854	0.9860	0.9838

use the weight function $w_t(h)$ in (2.3) with $h = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6$. We report the empirical sizes of the profile empirical likelihood test at levels 10%, 5%, and 1% in Table 1 and the empirical powers in Tables 2 and 3. To save space, we only report results for $h = 0.1$ and 0.3 because results for other h's are similar.

Table 1 shows that *(i)* the size accuracy improves as the sample size becomes larger, *(ii)* the size for $\alpha_1 = 4.5$ is more accurate than that for $\alpha_1 = 2.2$, *(iii)* the size is robust to h, *(iv)* the size for the ARMA (1,0) model is more accurate than that

https://doi.org/10.1017/S0266466621000244 Published online by Cambridge University Press

TABLE 3. Test powers for ARMA(1,0)–GARCH(1,1) and ARMA(1,1)–GARCH(1,1) models with $\delta = 0.4$

				ARMA(1,0)–GARCH(1,1)		ARMA(1,1)–GARCH(1,1)	
a_1	α_1	n	Level	$h = 0.1$	$h = 0.3$	$h = 0.1$	$h = 0.3$
0.2	2.2						
		1000	10%	0.3926	0.3926	0.4004	0.4040
			5%	0.2930	0.2958	0.3048	0.3038
			1%	0.1424	0.1406	0.1400	0.1426
		2000	10%	0.5632	0.5576	0.5658	0.5656
			5%	0.4472	0.4462	0.4594	0.4586
			1%	0.2402	0.2432	0.2658	0.2664
0.2	4.5						
		1000	10%	0.4326	0.4380	0.4420	0.4398
			5%	0.3184	0.3264	0.3144	0.3274
			1%	0.1424	0.1492	0.1358	0.1414
		2000	10%	0.6854	0.6834	0.6870	0.6816
			5%	0.5656	0.5722	0.5774	0.5806
			1%	0.3178	0.3318	0.3346	0.3372
0.1	2.2						
		1000	10%	0.4022	0.3996	0.4108	0.4086
			5%	0.3020	0.2996	0.3098	0.3040
			1%	0.1506	0.1462	0.1484	0.1468
		2000	10%	0.5708	0.5610	0.5686	0.5642
			5%	0.4606	0.4488	0.4680	0.4578
			1%	0.2550	0.2446	0.2658	0.2630
0.1	4.5						
		1000	10%	0.4544	0.4474	0.4594	0.4508
			5%	0.3418	0.3324	0.3448	0.3380
			1%	0.1602	0.1570	0.1498	0.1480
		2000	10%	0.7048	0.6954	0.7094	0.6942
			5%	0.5984	0.5826	0.6028	0.5830
			1%	0.3486	0.3444	0.3676	0.3512

for the ARMA (1,1) model, and *(v)* the tail heaviness of η_t rather than ε_t has an impact on the size. Also, we conclude from Tables 2 and 3 that *(i)* the test power improves as the sample size increases, *(ii)* the test for $\delta = 0.3$ has a larger power than that for $\delta = 0.4$ as the median of errors in the former case is farther away from zero than the latter, and *(iii)* the test for $\alpha_1 = 4.5$ is more powerful than that for $\alpha_1 = 2.2$. Again, the tail heaviness of η_t rather than ε_t impacts the power because we use the weight function to reduce the moment effect of σ_t.

https://doi.org/10.1017/S0266466621000244 Published online by Cambridge University Press

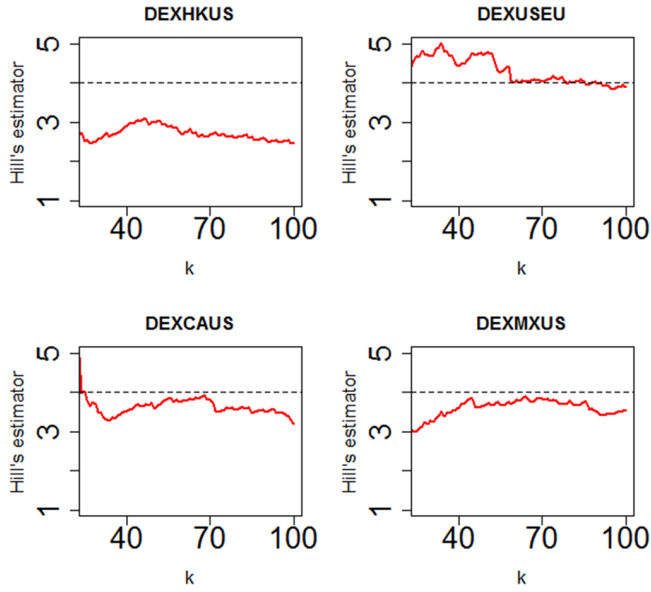


FIGURE 1. The Hill estimates for the HKD/USD, USD/EUD, CANADA/USD, and MEXICAN/USD exchange rates.

4. APPLICATIONS

4.1. Exchange Rates

This section studies the daily HKD/USD, USD/EUD, CANADA/USD, and MEXICAN/USD exchange rates from October 5, 2009 to October 4, 2019. Denote the log-returns ($\times 100$) of each series by $\{X_t\}$. We first estimate the tail index of $\{X_t\}$ by the Hill estimator $\hat{\alpha}(k)$ in Hill (1975) defined as

$$\hat{\alpha}(k) = \left[\frac{1}{k}\sum_{i=1}^{k}\log\left(\frac{X_{(n-i+1)}}{X_{(n-k)}}\right)\right]^{-1}$$

with $\{X_{(t)}\}$ being the ascending order statistics of $\{X_t\}$. We plot the Hill estimates against various k in Figure 1, indicating that the tail index of HKD/USD, CANADA/USD, and MEXICAN/USD exchange rates is less than 4, i.e., $\mathbb{E}(X_t^4) = \infty$. Therefore, when we fit an ARMA–GARCH model by using the QMLE, the inference may not have a normal limit. To explore the possibility of using the SWQMELE to fit an ARMA–GARCH model, we have to test whether the GARCH model has zero median.

To conduct such a test, we first use the function "auto.arima" in the R package "forecast" with Akaike information criterion (AIC) to obtain the appropriate orders for the ARMA model. The results are an ARMA(2,2) model for

https://doi.org/10.1017/S0266466621000244 Published online by Cambridge University Press

TABLE 4. p values of the empirical likelihood test for daily exchange rates

Exchange rate	ARMA model	$h = 0.1$	$h = 0.2$	$h = 0.3$	$h = 0.4$	$h = 0.5$
HKD/USD	ARMA(2,2)	0.1754	0.2971	0.1935	0.2555	0.6919
USD/EUD	ARMA(1,1)	0.6431	0.6178	0.8581	0.8583	0.9567
CANADA/USD	ARMA(5,4)	0.2100	0.1267	0.1044	0.0841	0.0153
MEXICAN/USD	ARMA(3,2)	0.0134	0.0157	0.0158	0.0107	0.0109

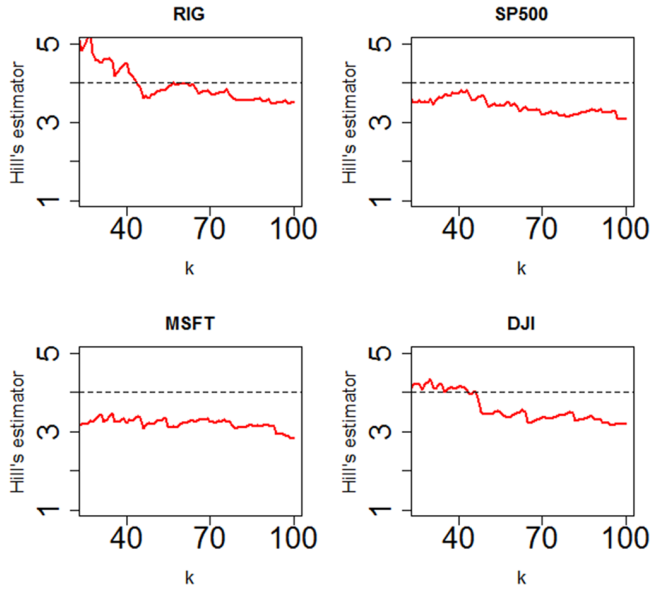


FIGURE 2. The Hill estimates for the Transocean (RIG), S&P 500, Microsoft Stock (MSFT), and Dow Jones Index (DJI).

HKD/USD exchange rates, an ARMA(1,1) model for USD/EUD exchange rates, an ARMA(5,4) model for CANADA/USD exchange rates, and an ARMA(3,2) for MEXICAN/USD exchange rates. Next, we compute the p value of the proposed profile empirical likelihood test for zero median of errors using the weight function $w_t(h)$ in (2.3) with $h = 0.1, 0.2, 0.3, 0.4$, and 0.5. Table 4 shows that the null hypothesis of zero median is rejected for MEXICAN/USD exchange rates, while HKD/USD and USD/EDU exchange rates do not reject the null hypothesis. Therefore, one should be cautious in using the SWQMELE to fit an ARMA–GARCH model to the MEXICAN/USD exchange rates, as the ARMA part no longer models the conditional mean.

https://doi.org/10.1017/S0266466621000244 Published online by Cambridge University Press

TABLE 5. p values of the empirical likelihood tests for stocks and indices

Stock or index	ARMA model	$h = 0.1$	$h = 0.2$	$h = 0.3$	$h = 0.4$	$h = 0.5$
RIG	ARMA(1,0)	0.9743	0.9013	0.9756	0.9811	0.7316
S&P 500	ARMA(3,1)	0.0937	0.1037	0.0996	0.0761	0.0551
MSFT	ARMA(1,1)	0.9624	0.7056	0.9750	0.8013	0.4259
DJI	ARMA(3,1)	0.0044	0.0675	0.2006	0.0216	0.0169

4.2. Stock Market Indices

This subsection considers the daily close prices of the following stocks and indices: the Transocean (RIG), S&P 500, Microsoft Stock (MSFT), and Dow Jones Index (DJI) from October 28, 2009 to July 29, 2019. Like Section 4.1, we first plot the Hill estimates for each series in Figure 2, which indicates that the tail index of S&P 500 and MSFT is less than 4. Hence, one may employ the SWQMELE to fit an ARMA–GARCH model after confirming that the GARCH model has zero median.

As before, we first use the function "auto.arima" with AIC to obtain the appropriate orders for the ARMA model, and then we compute the p value of the proposed empirical likelihood test for zero median of errors with $h = 0.1$, 0.2, 0.3, 0.4, and 0.5. Table 5 shows that we should be cautious about using the SWQMELE to fit an ARMA–GARCH model to S&P500 and DJI, as zero median of errors may not be true.

5. CONCLUSIONS

To fit an ARMA–GARCH model with fewer moments, one has to impose conditions on the median rather than the mean of the errors. Because changing the zero mean to zero median is significant for skewed data and destroys the ARMA–GARCH structure, this paper develops an empirical likelihood test for zero median of errors without estimating the GARCH model and the asymptotic variance and by allowing heavy tailed errors. Hence, the proposed test is robust and efficient. A simulation study confirms the good finite sample performance in terms of size and power. The data analysis shows that some real financial data do not have zero median of errors, which cautions the use of the SWQMELE.

APPENDIX

A. PROOFS

Before proving Theorems 1 and 2, we need some lemmas. Lemma 1 derives bounds for allowing the use of weight functions to reduce the moment effect. Lemma 2 shows that the sample mean and sample variance of $\boldsymbol{D}_t(\theta, 0)$ can be approximated by those of $\boldsymbol{D}_t(\theta_0, 0)$.

https://doi.org/10.1017/S0266466621000244 Published online by Cambridge University Press

Lemma 3 derives the asymptotic normality of $\sum_{t=1}^{n} \boldsymbol{D}_t(\theta_0, 0)$. For profiling the nuisance parameters, Lemma 4 develops an approximation for $\sum_{t=1}^{n} D_{t,2}(\theta, 0)$.

LEMMA 1. *Under conditions of Theorem 1, there exist a constant $\rho \in (0,1)$, a constant $C > 0$, and a neighborhood Θ_0 of θ_0 such that*

$$\sup_{\Theta_0} |\varepsilon_t(\theta)| \le C\xi_{\rho,t-1}, \ \sup_{\Theta_0} \left\| \frac{\partial \varepsilon_t(\theta)}{\partial \theta} \right\| \le C\xi_{\rho,t-1}, \text{ and } \sup_{\Theta_0} \left\| \frac{\partial^2 \varepsilon_t(\theta)}{\partial \theta \partial \theta'} \right\| \le C\xi_{\rho,t-1},$$

where $\xi_{\rho,t}$ is defined in Assumption 3.

Proof. See Ling (2007). ■

LEMMA 2. *Let $\mathcal{B}_0 = \{\theta : \|\theta - \theta_0\| \le \frac{C_0}{\sqrt{n}}\}$ for some constant $C_0 > 0$. Under conditions of Theorem 1, we have*

(i) $\max_{1\le t\le n} \sup_{\mathcal{B}_0} \|\boldsymbol{D}_t(\theta, 0)\| = o_p(\sqrt{n})$.

(ii) $\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{D}_t(\theta, 0) = \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{D}_t(\theta_0, 0) + O_p(n^{-1/2})$ *holds uniformly for* $\theta \in \mathcal{B}_0$.

(iii) $\frac{1}{n}\sum_{t=1}^{n} \{\boldsymbol{D}_t(\theta, 0)\boldsymbol{D}_t'(\theta, 0)\} = E\{\boldsymbol{D}_1(\theta_0, 0)\boldsymbol{D}_1'(\theta_0, 0)\} + o_p(1)$ *holds uniformly for* $\theta \in \mathcal{B}_0$.

Proof. (i) We first prove that

$$\max_{1\le t\le n} \sup_{\mathcal{B}_0} \|\boldsymbol{D}_{t,1}(\theta)\| = o_p(\sqrt{n}). \tag{A.1}$$

By Lemma 1, we have for n large enough,

$$\begin{aligned}\sup_{\mathcal{B}_0} \|\boldsymbol{D}_{t,1}(\theta)\| &= \sup_{\mathcal{B}_0} \left\| w_{t-1}^{-2} \varepsilon_t(\theta) \frac{\partial \varepsilon_t(\theta)}{\partial \theta} \right\| \le \sup_{\mathcal{B}_0} w_{t-1}^{-2} \|\varepsilon_t(\theta)\| \left\| \frac{\partial \varepsilon_t(\theta)}{\partial \theta} \right\| \\ &\le C^2 w_{t-1}^{-2} \xi_{\rho,t-1}^2 =: V(t).\end{aligned}$$

By the Chebyshev inequality, for any $\epsilon > 0$,

$$\begin{aligned}&P\left(\max_{1\le t\le n} V(t) > \sqrt{n}\epsilon \right) \\ \le &\frac{1}{n\epsilon^2} \sum_{t=1}^{n} \mathbb{E}\left\{ V^2(t) I(V(t) > \sqrt{n}\epsilon) \right\} \\ \le &\frac{1}{\epsilon^2} \max_{1\le t\le n} \mathbb{E}\left\{ V^2(t) I(V(t) > \sqrt{n}\epsilon) \right\} \to 0\end{aligned}$$

as $n \to \infty$, where the last inequality follows from the fact that

$$\mathbb{E}\big(V^2(t)\big) = C^4 \mathbb{E}\big(w_{t-1}^{-4} \xi_{\rho,t-1}^4\big) < \infty,$$

by Assumption 3. Hence, (A.1) follows.

https://doi.org/10.1017/S0266466621000244 Published online by Cambridge University Press

Because

$$\sup_{\mathcal{B}_0}\left|D_{t,2}(\theta,0)\right| = \sup_{\mathcal{B}_0}\left|w_{t-1}^{-1}\mathrm{sgn}(\varepsilon_t(\theta))\right| \le w_{t-1}^{-1}\sup_{\mathcal{B}_0}\left|\mathrm{sgn}(\varepsilon_t(\theta)) - \mathrm{sgn}(\varepsilon_t)\right| + w_{t-1}^{-1} \le 3w_{t-1}^{-1},$$

it follows from Assumption 3 that

$$\max_{1\le t\le n}\sup_{\mathcal{B}_0}\left|D_{t,2}(\theta,0)\right| = o_p(\sqrt{n}). \tag{A.2}$$

Hence, by (A.1) and (A.2), we have

$$\max_{1\le t\le n}\sup_{\mathcal{B}_0}\|\boldsymbol{D}_t(\theta,0)\| = o_p(\sqrt{n}).$$

(ii) Note that

$$\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{D}_t(\theta,0) - \frac{1}{n}\sum_{t=1}^{n}\boldsymbol{D}_t(\theta_0,0) = \frac{1}{n}\sum_{t=1}^{n}\left(\boldsymbol{D}_{t,1}^{'}(\theta) - \boldsymbol{D}_{t,1}^{'}(\theta_0), D_{t,2}(\theta,0) - D_{t,2}(\theta_0,0)\right)'.$$

For the proof of (ii), it suffices to show that

$$\sup_{\mathcal{B}_0}\left\|\frac{1}{n}\sum_{t=1}^{n}\left(\boldsymbol{D}_{t,1}(\theta) - \boldsymbol{D}_{t,1}(\theta_0)\right)\right\| = O_p(\frac{1}{\sqrt{n}})$$

and

$$\sup_{\mathcal{B}_0}\left|\frac{1}{n}\sum_{t=1}^{n}\left(D_{t,2}(\theta,0) - D_{t,2}(\theta_0,0)\right)\right| = O_p(\frac{1}{\sqrt{n}}).$$

Using Taylor expansion and Lemma 1, we have

$$\begin{aligned}
&\sup_{\mathcal{B}_0}\left\|\boldsymbol{D}_{t,1}(\theta) - \boldsymbol{D}_{t,1}(\theta_0)\right\| \\
=&\sup_{\mathcal{B}_0}\left\|w_{t-1}^{-2}\varepsilon_t(\theta)\frac{\partial\varepsilon_t(\theta)}{\partial\theta} - w_{t-1}^{-2}\varepsilon_t\frac{\partial\varepsilon_t(\theta_0)}{\partial\theta}\right\| \\
=&\sup_{\mathcal{B}_0}w_{t-1}^{-2}\left\|\left(\varepsilon_t(\theta) - \varepsilon_t\right)\frac{\partial\varepsilon_t(\theta)}{\partial\theta} + \varepsilon_t\left(\frac{\partial\varepsilon_t(\theta)}{\partial\theta} - \frac{\partial\varepsilon_t(\theta_0)}{\partial\theta}\right)\right\| \\
=&\sup_{\mathcal{B}_0}w_{t-1}^{-2}\left\|(\theta-\theta_0)'\frac{\partial\varepsilon_t(\theta^*)}{\partial\theta}\frac{\partial\varepsilon_t(\theta)}{\partial\theta} + \varepsilon_t\frac{\partial^2\varepsilon_t(\theta^{**})}{\partial\theta\partial\theta'}(\theta-\theta_0)\right\| \\
\le& w_{t-1}^{-2}\left\{\left(\sup_{\mathcal{B}_0}\left\|\frac{\partial\varepsilon_t(\theta)}{\partial\theta}\right\|^2\right)\left(\sup_{\mathcal{B}_0}\|\theta-\theta_0\|\right) + \left(\sup_{\mathcal{B}_0}\left\|\frac{\partial^2\varepsilon_t(\theta)}{\partial\theta\partial\theta'}\right\|\right)\left(\sup_{\mathcal{B}_0}|\varepsilon_t(\theta)|\right)\left(\sup_{\mathcal{B}_0}\|\theta-\theta_0\|\right)\right\} \\
\le&\frac{2C_0C^2}{\sqrt{n}}w_{t-1}^{-2}\xi_{\rho,t-1}^2,
\end{aligned}$$

where θ^* and θ^{**} lie between θ_0 and θ. Then, it follows from Assumption 3 that

$$\frac{1}{n}\sum_{t=1}^{n}\sup_{\mathcal{B}_0}\left\|\boldsymbol{D}_{t,1}(\theta) - \boldsymbol{D}_{t,1}(\theta_0)\right\| = O_p\left(\frac{1}{\sqrt{n}}\right).$$

https://doi.org/10.1017/S0266466621000244 Published online by Cambridge University Press

By Lemma 1, we have

$$\left|\frac{\partial \varepsilon_t(\theta^*)}{\partial \theta'}(\theta-\theta_0)\right| \le \sup_{\mathcal{B}_0}\left\|\frac{\partial \varepsilon_t(\theta)}{\partial \theta}\right\|\|\theta-\theta_0\| \le \frac{CC_0}{\sqrt{n}}\xi_{\rho,t-1}.$$

Thus, it follows from Taylor expansion and Assumptions 3 and 5 that

$$\begin{aligned}
&\mathbb{E}\sup_{\mathcal{B}_0} w_{t-1}^{-1}\left|\operatorname{sgn}(\varepsilon_t(\theta))-\operatorname{sgn}(\varepsilon_t)\right| \\
&= \mathbb{E}\sup_{\mathcal{B}_0} w_{t-1}^{-1}\Big[\left|I(\varepsilon_t(\theta)>0)-I(\varepsilon_t>0)\right|+\left|I(\varepsilon_t(\theta)<0)-I(\varepsilon_t<0)\right|\Big] \\
&= \mathbb{E}\sup_{\mathcal{B}_0} w_{t-1}^{-1}\left[\left|I\left(\frac{\partial \varepsilon_t(\theta^*)}{\partial \theta'}(\theta_0-\theta)<\varepsilon_t\le 0\right)-I\left(0<\varepsilon_t\le\frac{\partial \varepsilon_t(\theta^*)}{\partial \theta'}(\theta_0-\theta)\right)\right|\right. \\
&\quad \left.+\left|I\left(0\le\varepsilon_t<\frac{\partial \varepsilon_t(\theta^*)}{\partial \theta'}(\theta_0-\theta)\right)\right|-I\left(\frac{\partial \varepsilon_t(\theta^*)}{\partial \theta'}(\theta_0-\theta)\le\varepsilon_t<0\right)\right] \\
&\le 2\mathbb{E}\sup_{\mathcal{B}_0} w_{t-1}^{-1} I\big(|\varepsilon_t|\le|(\partial\varepsilon_t(\theta^*)/\partial\theta')(\theta_0-\theta)|\big) \\
&= 2\mathbb{E}\big\{w_{t-1}^{-1} I\big(|\eta_t|\le CC_0 n^{-\frac{1}{2}}\xi_{\rho,t-1}/\sigma_t\big)\big\} \\
&= \frac{2g(0)CC_0}{\sqrt{n}}\mathbb{E}\left\{w_{t-1}^{-1}\frac{\xi_{\rho,t-1}}{\sigma_t}\right\}+o\left(\frac{1}{\sqrt{n}}\right)=O\left(\frac{1}{\sqrt{n}}\right), \qquad \textbf{(A.3)}
\end{aligned}$$

where θ^* lies between θ_0 and θ. Hence, by (A.3), we have

$$\frac{1}{n}\sum_{t=1}^{n}\sup_{\mathcal{B}_0}\left|D_{t,2}(\theta,0)-D_{t,2}(\theta_0,0)\right| = \frac{1}{n}\sum_{t=1}^{n}\sup_{\mathcal{B}_0} w_{t-1}^{-1}\left|\operatorname{sgn}(\varepsilon_t(\theta))-\operatorname{sgn}(\varepsilon_t)\right| = O_p\left(\frac{1}{\sqrt{n}}\right)$$

and complete the proof of (ii).

(iii) For proving part (iii), it suffices to show that

$$\sup_{\mathcal{B}_0}\left\|\frac{1}{n}\sum_{t=1}^{n}\big\{\boldsymbol{D}_{t,1}(\theta)D_{t,2}(\theta,0)\big\}-\mathbb{E}\big\{\boldsymbol{D}_{t,1}(\theta_0)D_{t,2}(\theta_0,0)\big\}\right\| = o_p(1) \qquad \textbf{(A.4)}$$

because the other sums of products can be proved in the same way. Note that

$$\begin{aligned}
\boldsymbol{D}_{t,1}(\theta) &= w_{t-1}^{-2}\varepsilon_t(\theta)\frac{\partial \varepsilon_t(\theta)}{\partial \theta} \\
&= w_{t-1}^{-2}(\varepsilon_t(\theta)-\varepsilon_t)\frac{\partial \varepsilon_t(\theta)}{\partial \theta}+w_{t-1}^{-2}\varepsilon_t\left(\frac{\partial \varepsilon_t(\theta)}{\partial \theta}-\frac{\partial \varepsilon_t(\theta_0)}{\partial \theta}\right)+w_{t-1}^{-2}\varepsilon_t\frac{\partial \varepsilon_t(\theta_0)}{\partial \theta} \\
&=: L_1(t)+L_2(t)+L_3(t)
\end{aligned}$$

and

$$\begin{aligned}
D_{t,2}(\theta,0) &= w_{t-1}^{-1}\operatorname{sgn}(\varepsilon_t(\theta)) = w_{t-1}^{-1}\big\{\operatorname{sgn}(\varepsilon_t(\theta))-\operatorname{sgn}(\varepsilon_t)\big\} \\
&\quad + w_{t-1}^{-1}\operatorname{sgn}(\varepsilon_t) =: M_1(t)+M_2(t).
\end{aligned}$$

https://doi.org/10.1017/S0266466621000244 Published online by Cambridge University Press

By Taylor expansion, (A.3), and Lemma 1, we have

$$\sup_{\mathcal{B}_0}\|L_1(t)+L_2(t)\| = \sup_{\mathcal{B}_0} w_{t-1}^{-2}\left\|(\theta-\theta_0)'\frac{\partial\varepsilon_t(\theta^*)}{\partial\theta}\frac{\partial\varepsilon_t(\theta)}{\partial\theta}+\varepsilon_t\frac{\partial^2\varepsilon_t(\widetilde{\theta})}{\partial\theta\partial\theta'}(\theta-\theta_0)\right\|$$
$$\leq \frac{2C_0C^2}{\sqrt{n}} w_{t-1}^{-2}\xi_{\rho,t-1}^2, \tag{A.5}$$

where θ^* and $\widetilde{\theta}$ lie between θ_0 and θ,

$$\|L_3(t)\| \leq \sup_{\mathcal{B}_0} w_{t-1}^{-2}\left\|\frac{\partial\varepsilon_t(\theta)}{\partial\theta}\right\|\|\varepsilon_t(\theta)\| \leq C^2 w_{t-1}^{-2}\xi_{\rho,t-1}^2 \tag{A.6}$$

and

$$\sup_{\mathcal{B}_0}|M_1(t)| = \sup_{\mathcal{B}_0} w_{t-1}^{-1}\left|\operatorname{sgn}(\varepsilon_t(\theta))-\operatorname{sgn}(\varepsilon_t)\right| \leq 2w_{t-1}^{-1} I\left\{|\eta_t| \leq CC_0 n^{-\frac{1}{2}}\frac{\xi_{\rho,t-1}}{\sigma_t}\right\}. \tag{A.7}$$

Thus, by (A.5)–(A.7), and Assumption 3, we can show that

$$\frac{1}{n}\sum_{t=1}^{n}\sup_{\mathcal{B}_0}\left\{|M_1(t)|(\|L_1(t)+L_2(t)\|)\right\} \xrightarrow{p} 0, \tag{A.8}$$

$$\frac{1}{n}\sum_{t=1}^{n}\sup_{\mathcal{B}_0}\left\{|M_1(t)|(\|L_3(t)\|)\right\} \xrightarrow{p} 0, \tag{A.9}$$

and

$$\frac{1}{n}\sum_{t=1}^{n}\sup_{\mathcal{B}_0}\left\{|M_2(t)|(\|L_1(t)+L_2(t)\|)\right\} \xrightarrow{p} 0. \tag{A.10}$$

Therefore, it follows from (A.8)–(A.10) that

$$\sup_{\mathcal{B}_0}\left\|\frac{1}{n}\sum_{t=1}^{n}\{\boldsymbol{D}_{t,1}(\theta)D_{t,2}(\theta,0)\}-\mathbb{E}\{\boldsymbol{D}_{t,1}(\theta_0)D_{t,2}(\theta_0,0)\}\right\|$$
$$=\sup_{\mathcal{B}_0}\left\|\frac{1}{n}\sum_{t=1}^{n}\{M_1(t)(L_1(t)+L_2(t))\}+\frac{1}{n}\sum_{t=1}^{n}\{M_2(t)(L_1(t)+L_2(t))\}\right.$$
$$\left.+\frac{1}{n}\sum_{t=1}^{n}\{M_1(t)L_3(t)\}+\frac{1}{n}\sum_{t=1}^{n}\{M_2(t)L_3(t)\}-\mathbb{E}\{D_{t,1}(\theta_0,0)D_{t,2}(\theta_0,0)\}\right\|$$
$$=\sup_{\mathcal{B}_0}\left\|\frac{1}{n}\sum_{t=1}^{n}\{M_2(t)L_3(t)\}-\mathbb{E}\{D_{t,1}(\theta_0,0)D_{t,2}(\theta_0,0)\}\right\|+o_p(1)$$
$$=o_p(1),$$

i.e., (A.4) holds. ■

LEMMA 3. *Suppose conditions of Theorem 1 hold. Then, as* $n\to\infty$

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{D}_t(\theta_0,0) \xrightarrow{d} N\left(\mathbf{0},\mathbb{E}\{\boldsymbol{D}_1(\theta_0,0)\boldsymbol{D}_1'(\theta_0,0)\}\right).$$

https://doi.org/10.1017/S0266466621000244 Published online by Cambridge University Press

Proof. Note that $\mathcal{F}_t$ is the σ-filed generated by the sequence $\{\varepsilon_t, \varepsilon_{t-1}, \dots\}$. Then,

$$\mathbb{E}\big(\boldsymbol{D}_{t,1}(\theta_0)\big|\mathcal{F}_{t-1}\big) = \mathbb{E}\left(w_{t-1}^{-2}\varepsilon_t\frac{\partial\varepsilon_t(\theta_0)}{\partial\theta}\middle|\mathcal{F}_{t-1}\right) = w_{t-1}^{-2}\frac{\partial\varepsilon_t(\theta_0)}{\partial\theta}\mathbb{E}\big(\varepsilon_t\big|\mathcal{F}_{t-1}\big) = 0$$

and

$$\mathbb{E}\big(D_{t,2}(\theta_0,0)\big|\mathcal{F}_{t-1}\big) = \mathbb{E}\Big(w_{t-1}^{-1}\operatorname{sgn}(\varepsilon_t)\big|\mathcal{F}_{t-1}\Big) = 0.$$

Therefore, $\{\boldsymbol{D}_t(\theta_0,0)\}$ are martingale differences.

By Lemma 2 (i) and (iii), we have

$$\max_{1\le t\le n}\|\frac{1}{\sqrt{n}}\boldsymbol{D}_t(\theta_0,0)\| = o_p(1),$$

$$\frac{1}{n}\sum_{t=1}^{n}\big\{\boldsymbol{D}_t(\theta_0,0)\boldsymbol{D}_t^{'}(\theta_0,0)\big\} = \mathbb{E}\big\{\boldsymbol{D}_1(\theta_0,0)\boldsymbol{D}_1^{'}(\theta_0,0)\big\} + o_p(1),$$

and the dominated convergence theorem implies that

$$\mathbb{E}\left\|\max_{1\le t\le n}\frac{1}{n}\boldsymbol{D}_t(\theta_0,0)\boldsymbol{D}_t^{'}(\theta_0,0)\right\| = o(1).$$

Therefore, the conditions of the central limit theorem of martingale differences are satisfied. The conclusion follows from the central limit theorem of martingale differences (See Hall and Heyde, 1980). ■

LEMMA 4. *Under conditions of Theorem 1, as $n\to\infty$, we have*

$$\sup_{\mathcal{B}_0}\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\big\{w_{t-1}^{-1}\big[\operatorname{sgn}(\varepsilon_t(\theta)) - \operatorname{sgn}(\varepsilon_t)\big]\big\} - 2g(0)\mathbb{E}\left\{\frac{w_{t-1}^{-1}}{\sigma_t}\frac{\partial\varepsilon_t(\theta_0)}{\partial\theta'}\right\}[\sqrt{n}(\theta-\theta_0)]\right\| = o_p(1).$$

Proof. Let

$$d_{1t}(\theta) = w_{t-1}^{-1}\left\{I\left(\frac{1}{\sigma_t}\frac{\partial\varepsilon_t(\theta^*)}{\partial\theta'}(\theta_0-\theta) < \eta_t\right) - I\big(\eta_t > 0\big)\right\}$$

and

$$d_{2t}(\theta) = w_{t-1}^{-1}\left\{I\left(\frac{1}{\sigma_t}\frac{\partial\varepsilon_t(\theta^*)}{\partial\theta'}(\theta_0-\theta) > \eta_t\right) - I\big(\eta_t < 0\big)\right\}.$$

Then,

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}w_{t-1}^{-1}\big[\operatorname{sgn}(\varepsilon_t(\theta)) - \operatorname{sgn}(\varepsilon_t)\big] = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\big[d_{1t}(\theta) - d_{2t}(\theta)\big].$$

It suffices to show that

$$\sup_{\mathcal{B}_0}\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\Big\{[d_{1t}(\theta) - d_{2t}(\theta)] - \mathbb{E}\big[(d_{1t}(\theta) - d_{2t}(\theta))\big|\mathcal{F}_{t-1}\big]\Big\}\right\| = o_p(1) \qquad \textbf{(A.11)}$$

https://doi.org/10.1017/S0266466621000244 Published online by Cambridge University Press

and

$$\sup_{\mathcal{B}_0}\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathbb{E}\big[(d_{1t}(\theta)-d_{2t}(\theta))\big|\mathcal{F}_{t-1}\big]-2g(0)\mathbb{E}\left\{\frac{w_{t-1}^{-1}}{\sigma_t}\frac{\partial\varepsilon_t(\theta_0)}{\partial\theta'}\right\}[\sqrt{n}(\theta-\theta_0)]\right\|=o_p(1). \tag{A.12}$$

For (A.11), we only need to show that

$$\sup_{\mathcal{B}_0}\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\{d_{it}(\theta)-\mathbb{E}(d_{it}(\theta)|\mathcal{F}_{t-1})\}\right\|=:\sup_{\mathcal{B}_0}\|M_{n,i}(\theta)\|=o_p(1) \tag{A.13}$$

for $i=1,2$. Using the same argument in proving Lemma 2.2 in Zhu and Ling (2011), we can show that for any $\eta>0$,

$$\sup_{\|\theta-\theta_0\|<\eta}\frac{\|M_{n,i}(\theta)\|}{1+\sqrt{n}\|\theta-\theta_0\|}=o_p(1) \tag{A.14}$$

for $i=1,2$. When $\theta\in\mathcal{B}_0$, there exists an $\eta_0>0$ such that $\|\theta-\theta_0\|\le C_0/\sqrt{n}<\eta_0$. By (A.14), we have

$$\sup_{\mathcal{B}_0}\|M_{n,i}(\theta)\|\le(C_0+1)\sup_{\mathcal{B}_0}\frac{\|M_{n,i}(\theta)\|}{1+\sqrt{n}\|\theta-\theta_0\|}=o_p(1),$$

which implies (A.13).

Next, we prove (A.12). By Taylor expansion and Assumption 5, we have

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathbb{E}\big[d_{1t}(\theta)-d_{2t}(\theta)\big|\mathcal{F}_{t-1}\big]=2g(0)\left\{\frac{1}{n}\sum_{t=1}^{n}\frac{w_{t-1}^{-1}}{\sigma_t}\frac{\partial\varepsilon_t(\theta^*)}{\partial\theta'}\right\}[\sqrt{n}(\theta-\theta_0)](1+o_p(1)). \tag{A.15}$$

Using (A.15), Theorem 3.1 in Ling and McAleer (2003), and the dominated convergence theorem, we can easily prove (A.12). Hence, the lemma holds. ■

Proof of Theorem 1. Put $\theta=\theta_0+\dfrac{\boldsymbol{u}}{\sqrt{n}}$, where $\boldsymbol{u}$ is a $(r+s+1)$-dimensional vector. Recall $\Lambda=\mathbb{E}\{\boldsymbol{D}_1(\theta_0,0)\boldsymbol{D}_1'(\theta_0,0)\}$ in Theorem 2. Write

$$h(\theta,0,\lambda)=\frac{1}{n}\sum_{t=1}^{n}\frac{\boldsymbol{D}_t(\theta,0)}{1+\lambda'\boldsymbol{D}_t(\theta,0)}$$

and $\delta_t(\theta,0)=\lambda'\boldsymbol{D}_t(\theta,0)$, where λ is a solution of the equations $h(\theta,0,\lambda)=\mathbf{0}$.

First, we show that $\|\lambda\|=O_p(1/\sqrt{n})$. Write $\lambda=\varrho\boldsymbol{v}$ with $\|\boldsymbol{v}\|=1$ and define $Z_n(\theta,0)=\max_{1\le t\le n}\|\boldsymbol{D}_t(\theta,0)\|$ and $\boldsymbol{T}_n(\theta,0)=\dfrac{1}{n}\sum_{t=1}^{n}\boldsymbol{D}_t(\theta,0)\boldsymbol{D}_t'(\theta,0)$. Observe that

$$0=\|h(\theta,0,\varrho\boldsymbol{v})\|\ge\|\boldsymbol{v}'h(\theta,0,\varrho\boldsymbol{v})\|=\left|\frac{1}{n}\sum_{t=1}^{n}\frac{\boldsymbol{v}'\boldsymbol{D}_t(\theta,0)}{1+\varrho\boldsymbol{v}'\boldsymbol{D}_t(\theta,0)}\right|$$

$$=\left|\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{v}'\boldsymbol{D}_t(\theta,0)-\frac{1}{n}\sum_{t=1}^{n}\frac{\varrho\boldsymbol{v}'\boldsymbol{D}_t(\theta,0)\boldsymbol{D}_t'(\theta,0)\boldsymbol{v}}{1+\varrho\boldsymbol{v}'\boldsymbol{D}_t(\theta,0)}\right|$$

https://doi.org/10.1017/S0266466621000244 Published online by Cambridge University Press

$$\geq \frac{1}{n}\left|\sum_{t=1}^{n}\frac{\varrho \boldsymbol{v}'\boldsymbol{D}_t(\theta,0)\boldsymbol{D}_t'(\theta,0)\boldsymbol{v}}{1+\varrho \boldsymbol{v}'\boldsymbol{D}_t(\theta,0)}\right| - \frac{1}{n}\left|\sum_{t=1}^{n}\boldsymbol{v}'\boldsymbol{D}_t(\theta,0)\right|$$

$$\geq \frac{\varrho \boldsymbol{v}'\boldsymbol{T}_n(\theta,0)\boldsymbol{v}}{1+\varrho Z_n(\theta,0)} - \frac{1}{n}\left|\sum_{t=1}^{n}\boldsymbol{v}'\boldsymbol{D}_t(\theta,0)\right|.$$

Thus, it follows from Lemma 2 (ii) and Lemma 3 that

$$\begin{aligned}\frac{\varrho \boldsymbol{v}'\boldsymbol{T}_n(\theta,0)\boldsymbol{v}}{1+\varrho Z_n(\theta,0)} &\leq \frac{1}{n}\left|\sum_{t=1}^{n}\boldsymbol{v}'\boldsymbol{D}_t(\theta,0)\right| \\ &\leq \frac{1}{n}\left|\sum_{t=1}^{n}\boldsymbol{v}'\boldsymbol{D}_t(\theta_0,0)\right| + \frac{1}{n}\left|\sum_{t=1}^{n}\boldsymbol{v}'\big(\boldsymbol{D}_t(\theta,0)-\boldsymbol{D}_t(\theta_0,0)\big)\right| \\ &= O_p(1/\sqrt{n})\end{aligned}$$

uniformly for $\theta \in \mathcal{B}_0$.

Let a be the smallest eigenvalue of Λ. By Lemma 2 (iii), we have $\boldsymbol{v}'\boldsymbol{T}_n(\theta,0)\boldsymbol{v} \geq a + o_p(1)$ uniformly for $\theta \in \mathcal{B}_0$. Hence, it follows that

$$\varrho = \|\lambda\| = O_p(1/\sqrt{n}) \tag{A.16}$$

uniformly for $\theta \in \mathcal{B}_0$.

Secondly, we derive the solution formula for λ as a function of $\boldsymbol{D}_t(\theta,0)$. It follows from Lemma 2 (i) and (A.16) that

$$\max_{1\leq t\leq n}|\delta_t(\theta,0)| \leq \|\lambda\| \max_{1\leq t\leq n}\|\boldsymbol{D}_t(\theta,0)\| = O_p(1/\sqrt{n})o_p(\sqrt{n}) = o_p(1) \tag{A.17}$$

uniformly for $\theta \in \mathcal{B}_0$. By Lemma 2 (ii) and Lemma 3, we have

$$\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{D}_t(\theta,0) = \frac{1}{n}\sum_{t=1}^{n}\boldsymbol{D}_t(\theta_0,0) + O_p(1/\sqrt{n}) = O_p(1/\sqrt{n}) + O_p(1/\sqrt{n}) = O_p(1/\sqrt{n})$$

uniformly for $\theta \in \mathcal{B}_0$. It follows from (A.17) that

$$\begin{aligned}\mathbf{0} = h(\theta,0,\lambda) &= \frac{1}{n}\sum_{t=1}^{n}\boldsymbol{D}_t(\theta,0)\left(1-\lambda'\boldsymbol{D}_t(\theta,0)+\frac{\delta_t^2(\theta,0)}{1+\delta_t(\theta,0)}\right) \\ &= \frac{1}{n}\sum_{t=1}^{n}\boldsymbol{D}_t(\theta,0) - \boldsymbol{T}_n(\theta,0)\lambda + \frac{1}{n}\sum_{t=1}^{n}\frac{\boldsymbol{D}_t(\theta,0)\delta_t^2(\theta,0)}{1+\delta_t(\theta,0)} \\ &= \frac{1}{n}\sum_{t=1}^{n}\boldsymbol{D}_t(\theta,0) - \boldsymbol{T}_n(\theta,0)\lambda + o_p(1)O_p(1/\sqrt{n}) \\ &= \frac{1}{n}\sum_{t=1}^{n}\boldsymbol{D}_t(\theta,0) - \boldsymbol{T}_n(\theta,0)\lambda + o_p(1/\sqrt{n}).\end{aligned}$$

https://doi.org/10.1017/S0266466621000244 Published online by Cambridge University Press

In view of Lemma 2 (iii), $\boldsymbol{T}_n^{-1}(\theta,0) \geq C$ with probability tending to one. Hence,

$$\lambda = \boldsymbol{T}_n^{-1}(\theta,0)\Big(\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{D}_t(\theta,0)\Big) + \boldsymbol{R}_n,$$

where $\|\boldsymbol{R}_n\| = o_p(1/\sqrt{n})$.

Thirdly, we derive an expression for $l(\theta,0)$ as a function of $\boldsymbol{D}_t(\theta,0)$. It follows from Taylor expansion and (A.17) that

$$\begin{aligned}
l(\theta,0) &= -2\ln L(\theta,0) \\
&= 2\sum_{t=1}^{n}\ln\big(1+\delta_t(\theta,0)\big) \\
&= 2\lambda'\left(\sum_{t=1}^{n}\boldsymbol{D}_t(\theta,0)\right) - n\lambda'\boldsymbol{T}_n(\theta,0)\lambda + 2\sum_{t=1}^{n}U_t(\theta,0) \\
&= 2\Big\{\Big[\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{D}_t(\theta,0)\Big]'\boldsymbol{T}_n^{-1}(\theta,0) + \boldsymbol{R}_n'\Big\}\Big(\sum_{t=1}^{n}\boldsymbol{D}_t(\theta,0)\Big) \\
&\quad - n\Big\{\Big[\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{D}_t(\theta,0)\Big]'\boldsymbol{T}_n^{-1}(\theta,0) + \boldsymbol{R}_n'\Big\}\boldsymbol{T}_n(\theta,0)\Big\{\boldsymbol{T}_n^{-1}(\theta,0)\Big[\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{D}_t(\theta,0)\Big] + \boldsymbol{R}_n\Big\} \\
&\quad + 2\sum_{t=1}^{n}U_t(\theta,0) \\
&= 2n\Big(\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{D}_t(\theta,0)\Big)'\boldsymbol{T}_n^{-1}(\theta,0)\Big(\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{D}_t(\theta,0)\Big) + 2\boldsymbol{R}_n'\Big(\sum_{t=1}^{n}\boldsymbol{D}_t(\theta,0)\Big) \\
&\quad - n\Big(\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{D}_t(\theta,0)\Big)'\boldsymbol{T}_n^{-1}(\theta,0)\Big(\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{D}_t(\theta,0)\Big) - n\Big(\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{D}_t(\theta,0)\Big)'\boldsymbol{R}_n \\
&\quad - n\boldsymbol{R}_n'\Big(\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{D}_t(\theta,0)\Big) - n\boldsymbol{R}_n'\boldsymbol{T}_n(\theta,0)\boldsymbol{R}_n + 2\sum_{t=1}^{n}U_t(\theta,0) \\
&= \Big(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{D}_t(\theta,0)\Big)'\boldsymbol{T}_n^{-1}(\theta,0)\Big(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{D}_t(\theta,0)\Big) - n\boldsymbol{R}_n'\boldsymbol{T}_n(\theta,0)\boldsymbol{R}_n + 2\sum_{t=1}^{n}U_t(\theta,0),
\end{aligned}$$

where $U_t(\theta,0)$ is the remainder term. Because $\|\boldsymbol{R}_n\| = o_p(1/\sqrt{n})$, we have

$$n\boldsymbol{R}_n'\boldsymbol{T}_n(\theta,0)\boldsymbol{R}_n = o_p(1).$$

For the remainder term $U_t(\theta,0)$, we will show that

$$2\sum_{t=1}^{n}U_t(\theta,0) = o_p(1).$$

By Lemma 2 (iii), it follows that

https://doi.org/10.1017/S0266466621000244 Published online by Cambridge University Press

$$\begin{aligned}\sum_{t=1}^{n}\|\boldsymbol{D}_t(\theta,0)\|^3 &\le \max_{1\le t\le n}\|\boldsymbol{D}_t(\theta,0)\|\left(\sum_{t=1}^{n}\|\boldsymbol{D}_t(\theta,0)\|^2\right)\\ &= no_p(\sqrt{n})\left(\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{D}_t'(\theta,0)\boldsymbol{D}_t(\theta,0)\right)\\ &= no_p(\sqrt{n})tr\left(\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{D}_t'(\theta,0)\boldsymbol{D}_t(\theta,0)\right)\\ &= no_p(\sqrt{n})tr\left(\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{D}_t(\theta,0)\boldsymbol{D}_t'(\theta,0)\right)\\ &= no_p(\sqrt{n})O_p(1) = o_p(n^{\frac{3}{2}}),\end{aligned}$$

where *tr* denotes the trace of a matrix. Hence, we have

$$|2\sum_{t=1}^{n}U_t(\theta,0)| \le 2B\sum_{t=1}^{n}|\delta_t(\theta,0)|^3 \le 2B\sum_{t=1}^{n}\|\lambda\|^3\|\boldsymbol{D}_t(\theta,0)\|^3 = O_p(n^{-\frac{3}{2}})o_p(n^{\frac{3}{2}}) = o_p(1)$$

by noting that $\|\lambda\|^3 = O_p(n^{-\frac{3}{2}})$.

Hence, it follows from Lemma 2 (iii) that

$$\begin{aligned}l(\theta,0) &= \left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{D}_t(\theta,0)\right)'\boldsymbol{T}_n^{-1}(\theta,0)\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{D}_t(\theta,0)\right)+o_p(1)\\ &= \left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{D}_t(\theta,0)\right)'\Lambda^{-1}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{D}_t(\theta,0)\right)+o_p(1)\end{aligned}$$

uniformly for $\theta\in\mathcal{B}_0$. In particular, we have

$$l(\theta_0,0) = \left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{D}_t(\theta_0,0)\right)'\Lambda^{-1}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{D}_t(\theta_0,0)\right)+o_p(1).$$

Finally, we derive the expansion of the minimizer $\hat{\theta}$ of $l(\theta,0)$. Let $\boldsymbol{S}_n(\theta) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{D}_t(\theta,0)$. Then

$$\begin{aligned}&l(\theta,0)-l(\theta_0,0)\\ &= \boldsymbol{S}_n'(\theta)\Lambda^{-1}\boldsymbol{S}_n(\theta)-\boldsymbol{S}_n'(\theta_0)\Lambda^{-1}\boldsymbol{S}_n(\theta_0)+o_p(1)\\ &= \big(\boldsymbol{S}_n(\theta)-\boldsymbol{S}_n(\theta_0)\big)'\Lambda^{-1}\boldsymbol{S}_n(\theta)+\boldsymbol{S}_n'(\theta_0)\Lambda^{-1}\boldsymbol{S}_n(\theta)\\ &\quad-\boldsymbol{S}_n'(\theta_0)\Lambda^{-1}\boldsymbol{S}_n(\theta_0)+o_p(1)\\ &= \big(\boldsymbol{S}_n(\theta)-\boldsymbol{S}_n(\theta_0)\big)'\Lambda^{-1}\big(\boldsymbol{S}_n(\theta)-\boldsymbol{S}_n(\theta_0)\big)+\big(\boldsymbol{S}_n(\theta)-\boldsymbol{S}_n(\theta_0)\big)'\Lambda^{-1}\boldsymbol{S}_n(\theta_0)\\ &\quad+\boldsymbol{S}_n'(\theta_0)\Lambda^{-1}\big(\boldsymbol{S}_n(\theta)-\boldsymbol{S}_n(\theta_0)\big)+o_p(1)\end{aligned}$$

https://doi.org/10.1017/S0266466621000244 Published online by Cambridge University Press

uniformly for $\theta \in \mathcal{B}_0$. Note that

$$\boldsymbol{S}_n(\theta) - \boldsymbol{S}_n(\theta_0) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(\boldsymbol{D}_{t,1}'(\theta) - \boldsymbol{D}_{t,1}'(\theta_0), D_{t,2}(\theta, 0) - D_{t,2}(\theta_0, 0) \right)'.$$

Let $\Sigma_1 = \mathbb{E}\left\{ \frac{\partial \boldsymbol{D}_{t,1}(\theta_0)}{\partial \theta'} \right\} = \mathbb{E}\left\{ w_{t-1}^{-2} \left(\frac{\partial \varepsilon_t(\theta_0)}{\partial \theta} \frac{\partial \varepsilon_t(\theta_0)}{\partial \theta'} + \varepsilon_t \frac{\partial^2 \varepsilon_t(\theta_0)}{\partial \theta \partial \theta'} \right) \right\}$. By Taylor expansion and Lemma 2 (ii), it follows that

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(\boldsymbol{D}_{t,1}(\theta) - \boldsymbol{D}_{t,1}(\theta_0) \right) &= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\{ \frac{\partial \boldsymbol{D}_{t,1}(\theta_0)}{\partial \theta'} (\theta - \theta_0) + o_p\left(\frac{1}{\sqrt{n}}\right) \right\} \\ &= \left(\frac{1}{n} \sum_{t=1}^{n} \frac{\partial \boldsymbol{D}_{t,1}(\theta_0)}{\partial \theta'} \right) \boldsymbol{u} + o_p(1) \\ &= \Sigma_1 \boldsymbol{u} + o_p(1) \end{aligned}$$

uniformly for $\theta \in \mathcal{B}_0$. Define $\Sigma_2 = 2g(0)\mathbb{E}\left\{ \frac{w_{t-1}^{-1}}{\sigma_t} \frac{\partial \varepsilon_t(\theta_0)}{\partial \theta'} \right\}$. By Lemma 4, we have

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(D_{t,2}(\theta, 0) - D_{t,2}(\theta_0, 0) \right) &= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} w_{t-1}^{-1} \left(\operatorname{sgn}(\varepsilon_t(\theta)) - \operatorname{sgn}(\varepsilon_t) \right) \\ &= \Sigma_2 \boldsymbol{u} + o_p(1) \end{aligned}$$

uniformly for $\theta \in \mathcal{B}_0$.

Put $\Gamma = \left(\Sigma_1, \Sigma_2 \right)$. Then,

$$\boldsymbol{S}_n(\theta) - \boldsymbol{S}_n(\theta_0) = \left(\Sigma_1', \Sigma_2' \right)' \boldsymbol{u} + o_p(1) = \Gamma' \boldsymbol{u} + o_p(1)$$

uniformly in $\theta \in \mathcal{B}_0$, implying that

$$l(\theta, 0) - l(\theta_0, 0) = \boldsymbol{u}' \Gamma \Lambda^{-1} \Gamma' \boldsymbol{u} + \boldsymbol{u}' \Gamma \Lambda^{-1} \boldsymbol{S}_n(\theta_0) + \boldsymbol{S}_n'(\theta_0) \Lambda^{-1} \Gamma' \boldsymbol{u} + o_p(1).$$

Like the proof of Lemma 1 of Qin and Lawless (1994), the minimizer $\hat{\theta} = \theta_0 + \frac{\boldsymbol{u}}{\sqrt{n}}$ must lie in $\mathcal{B}_0$, that is,

$$\hat{\boldsymbol{u}} = -\left(\Gamma \Lambda^{-1} \Gamma' \right)^{-1} \left(\Gamma \Lambda^{-1} \boldsymbol{S}_n(\theta_0) \right) + o_p(1),$$

and

$$\begin{aligned} l(\hat{\theta}, 0) &= \boldsymbol{S}_n'(\theta_0) \Lambda^{-1} \boldsymbol{S}_n(\theta_0) - \left(\Gamma \Lambda^{-1} \boldsymbol{S}_n(\theta_0) \right)' \left(\Gamma \Lambda^{-1} \Gamma' \right)^{-1} \left(\Gamma \Lambda^{-1} \boldsymbol{S}_n(\theta_0) \right) + o_p(1) \\ &= \left(\Lambda^{-\frac{1}{2}} \boldsymbol{S}_n(\theta_0) \right)' \left(\boldsymbol{I} - \Lambda^{-\frac{1}{2}} \Gamma' (\Gamma \Lambda^{-1} \Gamma')^{-1} \Gamma \Lambda^{-\frac{1}{2}} \right) \left(\Lambda^{-\frac{1}{2}} \boldsymbol{S}_n(\theta_0) \right) + o_p(1). \end{aligned}$$

From Lemma 3, $\Lambda^{-\frac{1}{2}} \boldsymbol{S}_n(\theta_0)$ converges in distribution to a multivariate standard normal distribution, and

$$\begin{aligned} tr\left(\boldsymbol{I} - \Lambda^{-\frac{1}{2}} \Gamma' (\Gamma \Lambda^{-1} \Gamma')^{-1} \Gamma \Lambda^{-\frac{1}{2}} \right) &= tr(\boldsymbol{I}) - tr\left(\Lambda^{-\frac{1}{2}} \Gamma' (\Gamma \Lambda^{-1} \Gamma')^{-1} \Gamma \Lambda^{-\frac{1}{2}} \right) \\ &= (r + s + 2) - (r + s + 1) = 1. \end{aligned}$$

Therefore, $l(\hat{\theta}, 0) \xrightarrow{d} \chi_1^2$ as $n \to \infty$. ■

https://doi.org/10.1017/S0266466621000244 Published online by Cambridge University Press

Proof of Theorem 2. Following the proof of Theorem 1, we have

$$l(\hat{\theta},0)=\big(\Lambda^{-\frac{1}{2}}\boldsymbol{S}_n^*(\theta_0)\big)'\big(\boldsymbol{I}-\Lambda^{-\frac{1}{2}}\Gamma'(\Gamma\Lambda^{-1}\Gamma')^{-1}\Gamma\Lambda^{-\frac{1}{2}}\big)\big(\Lambda^{-\frac{1}{2}}\boldsymbol{S}_n^*(\theta_0)\big)+o_p(1), \tag{A.18}$$

where $\boldsymbol{S}_n^*(\theta_0)=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{D}_t(\theta_0,0)$. Put $D_{t,2}^*=w_{t-1}^{-1}\{\mathrm{sgn}(\varepsilon_t)-\mathbb{E}(\mathrm{sgn}(\varepsilon_t))\}$ and $\boldsymbol{D}_t^*=(\boldsymbol{D}_{t,1}'(\theta_0),D_{t,2}^*)'$. Then

$$\begin{aligned}
&\boldsymbol{S}_n^*(\theta_0)\\
&=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\boldsymbol{D}_{t,1}'(\theta_0),D_{t,2}(\theta_0,0))'\\
&=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\big(\boldsymbol{D}_{t,1}'(\theta_0),w_{t-1}^{-1}\{\mathrm{sgn}(\varepsilon_t)-\mathbb{E}(\mathrm{sgn}(\varepsilon_t)|\mathcal{F}_{t-1})\}\big)'\\
&\quad+\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\boldsymbol{0}',w_{t-1}^{-1}\mathbb{E}((\mathrm{sgn}(\varepsilon_t)-\mathrm{sgn}(\varepsilon_t-d_0))|\mathcal{F}_{t-1}))'\\
&\quad+\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\boldsymbol{0}',w_{t-1}^{-1}\mathbb{E}(\mathrm{sgn}(\varepsilon_t-d_0)|\mathcal{F}_{t-1}))'\\
&=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{D}_t^*+\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\boldsymbol{0}',w_{t-1}^{-1}\mathbb{E}\{(\mathrm{sgn}(\eta_t)-\mathrm{sgn}(\eta_t-d_0/\sigma_t))|\mathcal{F}_{t-1}\})'\\
&\quad+\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\boldsymbol{0}',w_{t-1}^{-1}\{\mathbb{E}(\mathrm{sgn}(\eta_t-d_0/\sigma_t)|\mathcal{F}_{t-1})-\mathbb{E}(\mathrm{sgn}(\eta_t-d_0/\sigma_t))\})'\\
&=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{D}_t^*+\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\boldsymbol{0}',2w_{t-1}^{-1}\{F_\eta(d_0/\sigma_t)-F_\eta(0)\})'\\
&\quad-\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\boldsymbol{0}',2w_{t-1}^{-1}\{F_\eta(d_0/\sigma_t)-\mathbb{E}(F_\eta(d_0/\sigma_t))\})',
\end{aligned} \tag{A.19}$$

where $\boldsymbol{0}$ is a $(r+s+1)$-vector and F_η is the distribution of η_t.

Note that $\mathbb{E}\{w_{t-1}^{-1}[\mathrm{sgn}(\varepsilon_t)-\mathbb{E}(\mathrm{sgn}(\varepsilon_t)|\mathcal{F}_{t-1})]|\mathcal{F}_{t-1}\}=0$, it follows that $\{\boldsymbol{D}_t^*\}$ is a sequence of martingale differences. Like Lemma 3, we can show that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{D}_t^*\xrightarrow{d}N\Big(\boldsymbol{0},\mathbb{E}\big\{\boldsymbol{D}_1(\theta_0,0)\boldsymbol{D}_1'(\theta_0,0)\big\}\Big)$$

by noting that $d_0\to 0$. On the other hand, because $d_0/\sigma_t\le M/(n\omega_1)^{1/2}$ holds uniformly for all $1\le t\le n$, it follows from Taylor expansion that

$$\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\big\{w_{t-1}^{-1}(F_\eta(d_0/\sigma_t)-F_\eta(0))\big\}-g(0)M_1\mathbb{E}\{w_{t-1}^{-1}\sigma_t^{-1}\}\right|=o_p(1)$$

https://doi.org/10.1017/S0266466621000244 Published online by Cambridge University Press

and

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{t=1}^{n} w_{t-1}^{-1}\{F_\eta(d_0/\sigma_t) - \mathbb{E}(F_\eta(d_0/\sigma_t))\} \\
&= \frac{1}{\sqrt{n}}\sum_{t=1}^{n} w_{t-1}^{-1} g(0)[(d_0/\sigma_t) - \mathbb{E}(d_0/\sigma_t)](1+o_p(1)) \\
&= \frac{M_1 g(0)}{n}\sum_{t=1}^{n} w_{t-1}^{-1}(\sigma_t^{-1} - \mathbb{E}(\sigma_t^{-1}))(1+o_p(1)) \\
&= M_1 g(0)\mathbb{E}\{w_{t-1}^{-1}(\sigma_t^{-1} - \mathbb{E}(\sigma_t^{-1}))\} + o_p(1),
\end{aligned}$$

where the last equality follows by ergodicity. Thus, $\lambda^{-\frac{1}{2}}\boldsymbol{S}_n^*(\theta_0)$ converges to a multivariate normal distribution with mean $\Lambda^{-1/2}(\mathbf{0}', 2g(0)M_1\mathbb{E}[w_{t-1}^{-1}\mathbb{E}(\sigma_t^{-1})])'$ and covariance $\boldsymbol{I}$. By (A.18), we have $l(\hat{\theta}, 0)$ converges to a noncentral chi-squared limit with one degree of freedom and the noncentrality parameter

$$(\mathbf{0}', 2g(0)M_1\mathbb{E}(w_{t-1}^{-1})\mathbb{E}(\sigma_t^{-1}))\Lambda^{-1}(\mathbf{0}', 2g(0)M_1\mathbb{E}(w_{t-1}^{-1})\mathbb{E}(\sigma_t^{-1}))'.$$

■

REFERENCES

Basrak, B., R. A. Davis & T. Mikosch (2002) Regular variation of GARCH processes. *Stochastic Processes and their Applications* 99, 95–116.

Berkes, I., L. Horváth & P. Kokoszka (2003) Estimation of the maximal moment exponent of a GARCH(1,1) sequence. *Econometric Theory* 19, 565–586.

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

Chan, N. H. & S. Ling (2006) Empirical likelihood for GARCH models. *Econometric Theory* 22, 403–428.

Chan, N. H., L. Peng & R. Zhang (2012) Interval estimation of the tail index of a GARCH(1,1) model. *Test* 21, 546–565.

Engle, R. F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation. *Econometrica* 50, 987–1008.

Fan, J., L. Qi & D. Xiu (2014) Quasi maximum likelihood estimation of GARCH models with heavy-tailed likelihoods. *Journal of Business and Economics Statistics* 32, 178–205.

Francq, C. & J. Zakoïan (2004) Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* 10, 847–860.

Hall, P. & C. C. Heyde (1980) *Martingale Limit Theory and Its Application* (Academic Press).

Hall, P. & Q. Yao (2003) Inference in ARCH and GARCH models with heavy-tailed errors. *Econometrica* 71, 285–317.

He, Y., Y. Hou, L. Peng & H. Shen (2020) Inference for conditional value-at-risk of a predictive regression. *Annals of Statistics* 8, 3442–3464.

Hill, B. (1975) A simple general approach to inference about the tail of a distribution. *Annals of Statistics* 3, 1163–1174.

Huang, H., X. Leng, X. Liu & L. Peng (2020) Unified inference for an AR process with possible infinite variance GARCH errors. *Journal of Financial Econometrics* 18, 425–470.

https://doi.org/10.1017/S0266466621000244 Published online by Cambridge University Press

Li, J., W. Liang & S. He (2012) Empirical likelihood for AR-ARCH models based on LAD estimation. *Acta Mathematicae Applicatae Sinica* 28, 371–382.

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

Ling, S. & M. McAleer (2003) Asymptotic theory for a new vector ARMA-GARCH model. *Econometric Theory* 19, 280–310.

Mikosch, T. & C. Starica (2000) Limit theory for the sample autocorrelations and extremes of a GARCH(1,1) sequence. *Annals of Statistics* 28, 1427–1451.

Owen, A. B. (2001) *Empirical Likelihood*. Chapman Hall.

Peng, L. & Q. Yao (2003) Least absolute deviations estimation for ARCH and GARCH models. *Biometrika* 90, 967–975.

Qin, J. & J. Lawless (1994) Empirical likelihood and general estimating equations. *Annals of Statistics* 22, 300–325.

Zhang, R., C. Li & L. Peng (2019) Inference for the tail index of a GARCH(1,1) model and an AR(1) model with ARCH(1) errors. *Econometric Reviews* 38, 151–169.

Zhang, R. & S. Ling (2015) Asymptotic inference for AR models with heavy-tailed G-GARCH noises. *Econometric Theory* 31, 880–890.

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

Zhu, K. & S. Ling (2015) LADE-based inference for ARMA models with unspecified and heavy-tailed heteroscedastic noises. *Journal of the American Statistical Association* 110, 784–794.

https://doi.org/10.1017/S0266466621000244 Published online by Cambridge University Press