entropy

MDPI

Article

On Entropy Test for Conditionally Heteroscedastic Location-Scale Time Series Models

Sangyeol Lee * and Minjo Kim

Department of Statistics, Seoul National University, Gwanak-gu, Daehak-dong, Seoul 08826, Korea; kminjjo@naver.com

* Correspondence: sylee@stats.snu.ac.kr; Tel.: +82-2-880-8814

Received: 9 June 2017; Accepted: 26 July 2017; Published: 28 July 2017

Abstract: This study considers the goodness of fit test for a class of conditionally heteroscedastic location-scale time series models. For this task, we develop an entropy-type goodness of fit test based on residuals. To examine the asymptotic behavior of the test, we first investigate the asymptotic property of the residual empirical process and then derive the limiting null distribution of the entropy test.

Keywords: conditionally heteroscedastic location-scale time series models; goodness of fit test; entropy test; residual empirical process

1. Introduction

In this study, we consider the goodness of fit (GOF) test on the innovations of location-scale time series models with heteroscedasticity. These models accommodate a broad class of financial time series models (see Noh and Lee [1] and Kim and Lee [2]). Correct information on the innovation distribution is considerably important in analyzing time series. For example, in the parameter estimation, one conventionally uses the Gaussian quasi-maximum likelihood estimator (QMLE), which undermines the accuracy of estimation when their innovation distributions are deviated far from the normal distribution. To overcome this difficulty, a different likelihood function has been considered as an alternative—see Lee and Lee [3] who use a family of normal mixtures, and Lee and Kim [4] who use asymmetric skew t distribution (ASTD) and asymmetric exponential power distribution (AEPD) families. The family of normal and Student's t distributions has been widely used in the literature—see Hansen [5], who uses a skew Student's t distribution in generalized autoregressive conditionally heteroscedastic (GARCH)-type models (Bollerslev [6]), and also the papers cited in Kim and Lee [2,7].

The GOF test has a long history, and has been playing a central role in matching given data sets with the best-fitted distribution families (see D'Agostino and Stephens [8] for a review). Among the GOF tests, the empirical process-based GOF test has long been popular because the classical Kolmogorov–Smirnov and Cramér–von Mises tests can be generated from the empirical process. Recently, Lee, Vonta and Karagrigoriou [9] proposed an entropy-based GOF test and demonstrated that it outperforms the classical tests in various situations. Lee, Lee and Park [10] and Lee and Oh [11] later applied the entropy test to GARCH-type models, and all confirmed its validity empirically. Further, Lee and Kim [4] used the entropy test for iid random variables following ASTD and AEPD families to demonstrate that ASTD accommodates AEPD to a greater degree than the other way around. Although the asymptotic theorems for the entropy test are established for GARCH models (Lee, Lee and Park [10]), those are not yet attempted in general location-scale time series models. Motivated by this, we are led to investigate the asymptotic behavior of the residual empirical process from the location-scale model and then verify the limiting null distribution of the entropy test—see Durbin [12], Lee and Wei [13], and Lee and Taniguchi [14] for relevant references.

The remainder of this paper is organized as follows. Section 2 investigates the asymptotic behavior of the residual empirical process and derives the limiting null distribution of the entropy test. Section 3 proves the theorem in Section 2. Section 4 provides concluding remarks.

2. Entropy Test for Location-Scale Models

2.1. Main Result

Let us consider the conditional location-scale model:

$$Y_t = g_t(\beta_{1,0}) + h_t(\beta_0)\eta_t, \ \ t \in Z, \tag{1}$$

where $g : \mathbb{R}^\infty \times \Theta_1 \to \mathbb{R}$ and $h : \mathbb{R}^\infty \times \Theta^m \to \mathbb{R}^+$ are measurable functions, $\Theta^m = \Theta_1 \times \Theta_2$ with compact subsets $\Theta_1 \subset \mathbb{R}^{d_1}$ and $\Theta_2 \subset \mathbb{R}^{d_2}$, $g_t(\beta_{1,0}) = g(Y_{t-1}, Y_{t-2}, \cdots; \beta_{1,0})$ and $h_t(\beta_0) = h(Y_{t-1}, Y_{t-2}, \cdots; \beta_0)$, where $\beta_0 = (\beta_{1,0}^T, \beta_{2,0}^T)^T$ denotes the true model parameter belonging to Θ^m; $\{\eta_t\}$ is a sequence of iid random variables with zero mean and unit variance. In what follows, we assume that $\{Y_t : t \in Z\}$ is strictly stationary and ergodic and that η_t is independent of past observations Ω_s for $s < t$. In this section, we consider the entropy-based GOF test proposed by Lee, Vonta and Karagrigoriou [9] for the location-scale models in (1). To this end, we set up the hypotheses:

$$\mathcal{H}_0 : F_\eta \in \{F_\vartheta : \vartheta \in \Theta^d\} \quad \text{vs.} \quad \mathcal{H}_1 : not \ \ \mathcal{H}_0, \tag{2}$$

where F_η denotes the innovation distribution of the model and F_ϑ can be any family of distributions.

To carry out the test, inspired by Rosenblatt [15], we check whether the transformed random variables $U_t = F_{\vartheta_0}\left(\frac{Y_t - g_t(\beta_{1,0})}{h_t(\beta_0)}\right)$ follow a uniform distribution on $[0,1]$, say, $U[0,1]$, where ϑ_0 and β_0 are the true parameters. Since the parameters are unknown, by replacing those with their estimates, we check the departure from $U[0,1]$ based on $\hat{U}_t := F_{\hat{\vartheta}_n}(\hat{\eta}_t)$ with $\hat{\eta}_t = \frac{Y_t - \tilde{g}_t(\hat{\beta}_{1,n})}{\tilde{h}_t(\hat{\beta}_n)}$, where $\tilde{g}_t(\beta_1) = g(Y_t, Y_{t-1}, \ldots, Y_1, 0, \ldots; \beta_1)$ and $\tilde{h}_t(\beta) = h(Y_t, Y_{t-1}, \ldots, Y_1, 0, \ldots; \beta)$ with $\beta = (\beta_1^T, \beta_2^T)^T \in \Theta^m$: see Francq and Zakoian [16], who take this approach of using initial values for GARCH models.

The entropy-based GOF test is constructed based on the Boltzmann–Shannon entropy defined by

$$H(f) = -\int_{-\infty}^{\infty} f(x) \log(f(x)) dx \tag{3}$$

for any density function f. It is noteworthy that the $H(f)$ actually measures the distance between a distribution with density f and the uniform distribution. Lee, Vonta and Karagrigoriou [9] construct a GOF test using an approximation form of the integral in (3). For any distribution F, we introduce

$$S^w(F) = -\sum_{i=1}^{m} w_i (F(s_i) - F(s_{i-1})) \log\left(\frac{F(s_i) - F(s_{i-1})}{s_i - s_{i-1}}\right), \tag{4}$$

where the w_i's are weights with $0 \le w_i \le 1$ and $\sum_{i=1}^m w_i = 1$, m is the number of disjoint intervals for partitioning the data range, and $-\infty < a \le s_0 \le \cdots \le s_m \le b < \infty$ are preassigned partition points. Note that the argument in (4) is a good approximation of that in (3) when w_i are all equal to 1—see Section 2.1 of Lee, Vonta and Karagrigoriou [9], and also their Remark 1 concerning the role of weights w.

Further, we define the residual empirical process:

$$\hat{V}_n(r) = \sqrt{n}(\hat{F}_n(r) - r), \quad 0 \le r \le 1 \tag{5}$$

with $\hat{F}_n(r) = \frac{1}{n}\sum_{t=1}^{n} I(F_{\hat{\vartheta}_n}(\hat{\eta}_t) \leq r)$, where $\hat{\vartheta}_n$ is any consistent estimator of ϑ_0 under the null; for example, the maximum likelihood estimator (MLE). We then define the entropy test by $\hat{T}_n := \sqrt{n}\sup_{w\in W}|S^w(\hat{F}_n)|$.

To derive the null limiting distribution of the entropy test, we impose the regularity conditions as follows:

(C1) **(i)** For some random variable V and constant $\kappa \in (0,1)$, $\sup_{\beta_1\in\Theta_1}\left|g_t(\beta_1) - \tilde{g}_t(\beta_1)\right| \leq V\kappa^t$ for all $t \geq 1$;
(ii) For some random variable V and constant $\kappa \in (0,1)$, $\sup_{\beta\in\Theta^m}\left|h_t(\beta) - \tilde{h}_t(\beta)\right| \leq V\kappa^t$ for all $t \geq 1$.

(C2) **(i)** For all $t \geq 1$, $\tilde{g}_t(\beta_1)$ and $\tilde{h}_t(\beta)$ are differentiable in β_1 and β on some neighborhoods N_1 of $\beta_{1,0}$ and N_2 of β_0;
(ii) There exists a random variable V and constant $\kappa \in (0,1)$ such that for all $t \geq 1$, $\sup_{\beta_1\in N_1}\|\partial g_t(\beta_1)/\partial\beta_1 - \partial\tilde{g}_t(\beta_1)/\partial\beta_1\| \leq V\kappa^t$ and $\sup_{\beta\in N_2}\|\partial h_t(\beta)/\partial\beta - \partial\tilde{h}_t(\beta)/\partial\beta\| \leq V\kappa^t$.

(C3) **(i)** For all $t \in Z$, $g_t(\beta_1)$ and $h_t(\beta)$ are twice differentiable in $\beta_1 \in N_{\delta 1}$ and $\beta \in N_2$, where N_1 and N_2 are the ones in **(C2)**(**i**);
(ii) $E[\sup_{\beta_1\in N_1}\|\partial g_1(\beta_1)/\partial\beta_1\|^2] < \infty$ and $E[\sup_{\beta\in N_2}\|\partial h_1(\beta)/\partial\beta\|^2] < \infty$;
(iii) $E[\sup_{\beta_1\in N_1}\|\partial^2 g_1(\beta_1)/\partial\beta_1\partial\beta_1^T\|] < \infty$ and $E[\sup_{\beta\in N_2}\|\partial^2 h_1(\beta)/\partial\beta\partial\beta^T\|] < \infty$.

(C4) **(i)** F_{ϑ_0} is continuous and has a positive density f_{ϑ_0};
(ii) f_{ϑ_0} and $x \to xf_{\vartheta_0}(x)$ are uniformly continuous on $(-\infty,\infty)$;
(iii) For some $L > 0$, $\sup_x|f_{\vartheta_0}(x)| \leq L$, $\sup_x|xf_{\vartheta_0}(x)| \leq L$, and $\sup_x|x^2 f'_{\vartheta_0}(x)| \leq L$.

(C5) F_ϑ is twice continuously differentiable with respect to ϑ and there exists $L > 0$ such that $\sup_x\|\frac{\partial^2 F_\vartheta(x)}{\partial\vartheta\partial\vartheta}\| \leq L$ and $x \to \|\frac{\partial F_\vartheta(x)}{\partial\vartheta}\|$ is uniformly continuous on $(-\infty,\infty)$.

(C6) Under the null, $\sqrt{n}(\hat{\beta}_n - \beta_0) = O_p(1)$ and $\sqrt{n}(\hat{\vartheta}_n - \vartheta_0) = O_p(1)$.

Remark 1. *The above conditions can be found in Kim and Lee [2]. They show that a class of GARCH and TGARCH models with ASTD and AEPD innovations satisfy the regularity conditions and the MLE is asymptotically normal.*

Below is the main result of this section: see the proof in Section 2.2.

Theorem 1. *Under* **(C1)**$\sim$**(C6)**, *we have*

$$\hat{V}_n(r) = V_n(r) + R_n(r), \quad 0 \leq r \leq 1 \tag{6}$$

where $V_n(r) = \sqrt{n}(F_n(r) - r)$ *with* $F_n(r) = \frac{1}{n}\sum_{t=1}^{n} I(F_{\vartheta_0}(\eta_t) \leq r)$ *and*

$$\begin{aligned}
R_n(r) &= \sqrt{n}(\hat{\beta}_{1n} - \beta_{1,0})^T E\Big[\frac{1}{h_1(\beta_0)}\frac{\partial g_1(\beta_{1,0})}{\partial\beta_1}\Big] f_{\vartheta_0}(F_{\vartheta_0}^{-1}(r)) \\
&+ \sqrt{n}(\hat{\beta}_n - \beta_0)^T E\Big[\frac{1}{h_1(\beta_0)}\frac{\partial h_1(\beta_0)}{\partial\beta}\Big] F_{\vartheta_0}^{-1}(r) f_{\vartheta_0}(F_{\vartheta_0}^{-1}(r)) \\
&+ \sqrt{n}(\hat{\vartheta}_n - \vartheta_0)^T \frac{\partial F_{\vartheta_0}(F_{\vartheta_0}^{-1}(r))}{\partial\theta} + o_p(1) \ \ uniformly\ in\ r.
\end{aligned}$$

Moreover, we are led to the following result, the detailed proof of which is omitted for brevity because it is essentially the same as that of Lee, Vonta and Karagrigoriou [9] and Lee, Lee and Park [10].

Theorem 2. *Suppose that the assumptions in Theorem 1 hold. Then, under $\mathcal{H}_0$, if $\max_{1\le i\le m}|s_i - s_{i-1}| \to 0$ as $m \to \infty$, we have that for all large m, as $n \to \infty$,*

$$\hat{T}_n \overset{d}{\approx} \sup_{w\in W'} \Big| \sum_{i=1}^{m} w_i(\mathcal{B}(s_i) - \mathcal{B}(s_{i-1}))\Big|, \tag{7}$$

where W' is any finite subset of the class of all weights and $\mathcal{B}$ is the Brownian bridge on $[0,1]$.

Here, the symbol $A_n := A_{n,m} \overset{d}{\approx} A := A_m$ as $n \to \infty$ indicates that the limiting distribution of A_n is approximately the same as the distribution of A as n tends to ∞. More precisely, we can write $A_n = A_{1,n,m} + A_{2,n,m}$, where $A_{1,n,m} \overset{d}{\to} A$ as $n \to \infty$ and $\lim_{m\to\infty}\lim_{n\to\infty} P(|A_{2,n,m}| > \delta) = 0$ for all $\delta > 0$.

Remark 2. *As seen in the proof of Theorem 2 of Lee, Lee and Park [10], one can easily check that owing to Theorem 1, under the null,*

$$\sup_{w\in W'} \Big| \sqrt{n} S^w(\hat{F}_n) + \sum_{i=1}^{m} w_i\Big(V_n(s_i) - V_n(s_{i-1})\Big) + \sum_{i=1}^{m} w_i\Big(R_n(s_i) - R_n(s_{i-1})\Big)\Big| = o_p(1),$$

wherein the term: $\sum_{i=1}^m w_i\Big(R_n(s_i) - R_n(s_{i-1})\Big)$ becomes negligible as n tends to infinity when m is large. This yields Theorem 2.

2.2. Proof of Theorem 1

We reexpress $\hat{V}_n(r)$ as follows:

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}[I[\hat{\eta}_t \le x] - F_{\hat{\vartheta}_n}(x)] = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}[I[\eta_t \le x] - F_{\vartheta_0}(x)] + A_n(x) + B_n(x) + C_n(x),$$

where

$$\begin{aligned}
A_n(x) &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\Big\{I\Big[\eta_t \le \frac{\tilde{g}_t(\hat{\beta}_{1,n}) - g_t(\beta_{1,0})}{h_t(\beta_0)} + \frac{\tilde{h}_t(\hat{\beta}_n)}{h_t(\beta_0)}x\Big] - F_{\vartheta_0}\Big(\frac{\tilde{g}_t(\hat{\beta}_{1,n}) - g_t(\beta_{1,0})}{h_t(\beta_0)} + \frac{\tilde{h}_t(\hat{\beta}_n)}{h_t(\beta_0)}x\Big) \\
&\quad + F_{\vartheta_0}(x) - I[\eta_t \le x]\Big\}, \\
B_n(x) &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\Big[F_{\vartheta_0}\Big(\frac{\tilde{g}_t(\hat{\beta}_{1,n}) - g_t(\beta_{1,0})}{h_t(\beta_0)} + \frac{\tilde{h}_t(\hat{\beta}_n)}{h_t(\beta_0)}x\Big) - F_{\vartheta_0}(x)\Big], \\
C_n(x) &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}[F_{\vartheta_0}(x) - F_{\hat{\vartheta}_n}(x)].
\end{aligned}$$

Since $\sup_{x\in R}|A_n(x)| = o_p(1)$ owing to Lemma 1 below, we handle the two terms $B_n(x)$ and $C_n(x)$. Let

$$a_{tn}(\beta_1) = \frac{\tilde{g}_t(\hat{\beta}_1) - g_t(\beta_{1,0})}{h_t(\beta_0)} \quad \text{and} \quad b_{tn}(\beta) = \frac{\tilde{h}_t(\beta_1) - h_t(\beta_0)}{h_t(\beta_0)},$$

and let ζ_n be a sequence of positive integer numbers with $\zeta_n = o(\sqrt{n})$ and $\zeta_n \to \infty$ as $n \to \infty$. We express $B_n(x) = B_{1,n}(x) + B_{2,n}(x) + o_p(1)$, where

$$\begin{aligned} B_{1,n}(x) &= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} f_{\vartheta_0}(x)(a_{tn}(\hat{\beta}_{1,n}) + b_{tn}(\hat{\beta}_n)x), \qquad (8) \\ B_{2,n}(x) &= \frac{1}{2\sqrt{n}} \sum_{t=\zeta_n+1}^{n} f'_{\vartheta_0}(x_t^*)(a_{tn}(\hat{\beta}_{1,n}) + b_{tn}(\hat{\beta}_n)x)^2 \end{aligned}$$

for some x_t^* between x and $a_{tn}(\hat{\beta}_{1,n}) + b_{tn}(\hat{\beta}_n)x + x$. By Taylor's theorem, we can express

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} f_{\vartheta_0}(x)a_{tn}(\hat{\beta}_{1,n}) = \sqrt{n}(\hat{\beta}_{1,n} - \beta_{1,0})^T f_{\vartheta_0}(x) \frac{1}{n} \sum_{t=1}^{n} \frac{1}{h_t(\beta_0)} \frac{\partial g_t(\beta_{1,0})}{\partial \beta_1} + R_{1,n}(x) + R_{2,n}(x) \qquad (9)$$

with

$$\begin{aligned} R_{1,n}(x) &= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} f_{\vartheta_0}(x) \Big(\frac{\tilde{g}_t(\hat{\beta}_{1,n}) - g_t(\hat{\beta}_{1,n})}{h_t(\beta_0)} \Big), \\ R_{2,n}(x) &= \sqrt{n}(\hat{\beta}_{1,n} - \beta_{1,0})^T f_{\vartheta_0}(x) \frac{1}{n} \sum_{t=1}^{n} \frac{1}{h_t(\beta_0)} \frac{\partial^2 g_t(\beta_{1,n}^*)}{\partial \beta_1 \partial \beta_1^T} (\hat{\beta}_{1,n} - \beta_{1,0}) \end{aligned}$$

for some $\beta_{1,n}^*$ between $\beta_{1,0}$ and $\hat{\beta}_{1,n}$. Then, owing to **(C1)(i)** and **(C4)(iii)**,

$$\sup_{x \in R} |R_{1,n}(x)| \leq \frac{1}{\sqrt{n}} \sum_{t=1}^{n} V\kappa^t \sup_{x \in R} |f_{\vartheta_0}(x)| = o_p(1),$$

and due to the ergodic theorem, Lemma 4 of Amemiya [17], (**C3**)(**iii**), (**C4**)(**iii**), and (**C6**), we get $\sup_{x \in R} |R_{2,n}(x)| = o_p(1)$, so that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} f_{\vartheta_0}(x)a_{tn}(\hat{\beta}_{1,n}) = \sqrt{n}(\hat{\beta}_{1,n} - \beta_{1,0})^T f_{\vartheta_0}(x) \frac{1}{n} \sum_{t=1}^{n} \frac{1}{h_t(\beta_0)} \frac{\partial g_t(\beta_{1,0})}{\partial \beta_1} + o_p(1). \qquad (10)$$

Similarly, it can be easily seen that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} x f_{\vartheta_0}(x)b_{tn}(\hat{\beta}_n) = \sqrt{n}(\hat{\beta}_n - \beta_0)^T x f_{\vartheta_0}(x) \frac{1}{n} \sum_{t=1}^{n} \frac{1}{h_t(\beta_0)} \frac{\partial h_t(\beta_0)}{\partial \beta} + o_p(1). \qquad (11)$$

Next, we analyze $B_{2,n}(x)$. Owing to the ergodic theorem, Lemma 4 of Amemiya [17], (**C1**)(**i**), (**C3**)(**ii**), (**C4**)(**iii**), and (**C6**), we have

$$\begin{aligned} & \sup_{x \in R} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} f'_{\vartheta_0}(x) a_{tn}^2(\hat{\beta}_{1,n}) \\ \leq\ & \frac{K}{\sqrt{n}} \sum_{t=1}^{n} \sup_{x \in R} |f'_{\vartheta_0}(x)| \Big(\frac{\tilde{g}_t(\hat{\beta}_{1,n}) - g_t(\hat{\beta}_{1,n})}{h_t(\beta_0)} \Big)^2 + \frac{K}{\sqrt{n}} \sum_{t=1}^{n} \sup_{x \in R} |f'_{\vartheta_0}(x)| \Big(\frac{g_t(\hat{\beta}_{1,n}) - g_t(\beta_{1,0})}{h_t(\beta_0)} \Big)^2 \end{aligned}$$

for some $K > 0$, which is no more than

$$\frac{K}{\sqrt{n}} \sum_{t=1}^{n} \sup_{x \in R} |f'_{\vartheta_0}(x)| V^2 \kappa^t + K\sqrt{n}(\hat{\beta}_{1,n} - \beta_{1,0})^T \sup_{x \in R} |f'_{\vartheta_0}(x)| \frac{1}{n} \sum_{t=1}^{n} \frac{1}{h_t^2(\beta_0)} \frac{\partial g_t(\beta_{1,n}^*)}{\partial \beta_1} \frac{\partial g_t(\beta_{1,n}^*)}{\partial \beta_1^T} (\hat{\beta}_{1,n} - \beta_{1,0})$$

$= o_p(1)$, where $\beta_{1,n}^*$ is an intermediate point between $\hat{\beta}_{1,n}$ and $\beta_{1,0}$.

Meanwhile, since $\max_{\zeta_n \le t \le n} |a_{tn}(\hat{\beta}_{1,n})| = o_p(1)$ and $\max_{\zeta_n \le t \le n} |b_{tn}(\hat{\beta}_n)| = o_p(1)$, we can find (large) $M > 0$, such that on the event $\mathcal{E} := (\max_{\zeta_n \le t \le n} |a_{tn}(\hat{\beta}_{1,n})| \le \delta_1, \max_{\zeta_n \le t \le n} |b_{tn}(\hat{\beta}_n)| \le \delta_2)$, $\delta_1 > 0, 1 < \delta_2 < 1$, with probability tending to 1,

$$\sup_{|x| \ge M} \Big(\frac{x}{x_t^*}\Big)^2 := D < \infty. \tag{12}$$

Hence, owing to the ergodic theorem, Lemma 4 of Amemiya [17], (**C1**)(**ii**), (**C3**)(**ii**), (**C4**)(**iii**), and (**C6**), we can have that on $\mathcal{E}$ and for $|x| \ge M$,

$$\begin{aligned}
&\Big|\frac{1}{\sqrt{n}} \sum_{t=\zeta_n+1}^{n} x^2 f'_{\vartheta_0}(x_t^*) b_{tn}^2(\hat{\beta}_n)\Big| \\
\le\ & \frac{K}{\sqrt{n}} \sum_{t=\zeta_n+1}^{n} x^2 |f'_{\vartheta_0}(x_t^*)| \Big(\frac{\tilde{h}_t(\hat{\beta}_n) - h_t(\hat{\beta}_n)}{h_t(\beta_0)}\Big)^2 + \frac{K}{\sqrt{n}} \sum_{t=\zeta_n+1}^{n} x^2 |f'_{\vartheta_0}(x_t^*)| \Big(\frac{h_t(\hat{\beta}_n) - h_t(\beta_0)}{h_t(\beta_0)}\Big)^2 \\
\le\ & \frac{K}{\sqrt{n}} \sum_{t=\zeta_n+1}^{n} x_t^{*2} |f'_{\vartheta_0}(x_t^*)| \Big(\frac{x}{x_t^*}\Big)^2 V^2 \kappa^t \\
+\ & \sqrt{n}(\hat{\beta}_n - \beta_0)^T \frac{K}{n} \sum_{t=\zeta_n+1}^{n} x_t^{*2} |f'_{\vartheta_0}(x_t^*)| \Big(\frac{x}{x_t^*}\Big)^2 \Big(\frac{1}{h_t^2(\beta_0)} \frac{\partial h_t(\beta_n^*)}{\partial \beta} \frac{\partial h_t(\beta_n^*)}{\partial \beta^T}\Big)(\hat{\beta}_n - \beta_0),
\end{aligned}$$

for some $K > 0$ and intermediate vector β_n^* between $\hat{\beta}_n$ and β_0, which is negligible. Because for $|x| < M$, $\frac{1}{\sqrt{n}} \sum_{t=\zeta_n+1}^{n} x^2 f'_{\vartheta_0}(x_t^*) b_{tn}^2(\hat{\beta}_n) = o_p(1)$, it holds that $\sup_{x \in R} |B_{2,n}(x)| = o_p(1)$, which together with (10) and (11) indicates

$$\begin{aligned}
B_n(x) &= \sqrt{n}(\hat{\beta}_{1,n} - \beta_{1,0})^T f_{\vartheta_0}(x) \frac{1}{n} \sum_{t=1}^{n} \frac{1}{h_t(\beta_0)} \frac{\partial g_t(\beta_{1,0})}{\partial \beta_1} \\
&+ \sqrt{n}(\hat{\beta}_n - \beta_0)^T x f_{\vartheta_0}(x) \frac{1}{n} \sum_{t=1}^{n} \frac{1}{h_t(\beta_0)} \frac{\partial h_t(\beta_0)}{\partial \beta} + o_p(1).
\end{aligned} \tag{13}$$

Since $C_n(x) = \sqrt{n}(\hat{\vartheta}_n - \vartheta_0)^T \frac{\partial F_{\vartheta_0}(x)}{\partial \vartheta} + o_p(1)$, owing to (**C5**) and (**C6**), we establish the theorem.

Lemma 1. *Under the assumptions in Theorem 1, we have* $\sup_{x \in R} |A_n(x)| = o_p(1)$.

Proof of Lemma 1. Due to (**C6**), for any $\epsilon > 0$, there exists $L > 0$ such that $P(\hat{\beta}_n \in \mathcal{N}_{L/\sqrt{n}}) \ge 1 - \epsilon$, where $N_{L/\sqrt{n}} = \mathcal{N}^1_{L/\sqrt{n}} \times \mathcal{N}^2_{L/\sqrt{n}}$ is a compact neighborhood of β_0 with $||\beta - \beta_0|| \le L/\sqrt{n}$ for all $\beta \in \mathcal{N}_{L/\sqrt{n}}$. For a positive real number ι, we partition $\mathcal{N}_{L/\sqrt{n}}$ into a finite number, say, $q(\iota)$ of subsets $I_n^1 = I_{1,n}^1 \times I_{2,n}^1, \ldots, I_n^{q(\iota)} = I_{1,n}^{q(\iota)} \times I_{2,n}^{q(\iota)}$ with diameter less than $\frac{\iota}{\sqrt{n}}$. Set

$$d_{1tn} = \frac{\iota}{\sqrt{n}} \sup_{\beta_1 \in N^1_{L/\sqrt{n}}} \Big\|\frac{\partial \tilde{g}_t(\beta_1)}{\partial \beta_1}\Big\| h_t^{-1}(\beta_0) \quad \text{and} \quad d_{2tn} = \frac{\iota}{\sqrt{n}} \sup_{\beta \in N_{L/\sqrt{n}}} \Big\|\frac{\partial \tilde{h}_t(\beta)}{\partial \beta}\Big\| h_t^{-1}(\beta_0).$$

Let $N(n)$ be an integer such that $N(n) = [n^{1/2+d}] + 1$, where $d \in (0, 1/2)$ and $[x]$ is the largest integer that does not exceed x. We divide the interval $[0, \infty)$ into $N(n)$ parts by the points $0 = x_0 < x_1 < \cdots < x_{N(n)} = \infty$ with $F_{\vartheta_0}(x_i) = iN(n)^{-1}$.

Then, for any points $\beta_n^j = (\beta_{1,n}^{j^T}, \beta_{2,n}^{j^T})^T$ in I_n^j, we have

$$
\begin{aligned}
A_n(x) &\le \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \Big\{ I\Big[\eta_t \le a_{tn}(\beta_{1,n}^j) + b_{tn}(\beta_n^j)x_{r+1} + d_{1tn} + d_{2tn}x_{r+1} + x_{r+1}\Big] \\
&- F_{\vartheta_0}\Big(a_{tn}(\hat\beta_{1,n}) + b_{tn}(\hat\beta_n)x + x\Big) + F_{\vartheta_0}(x) - I[\eta_t \le x]\Big\}; \qquad (14)\\
A_n(x) &\ge \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \Big\{ I\Big[\eta_t \le a_{tn}(\beta_{1,n}^j) + b_{tn}(\beta_n^j)x - d_{1tn} - d_{2tn}x_r + x_r\Big] \\
&- F_{\vartheta_0}\Big(a_{tn}(\hat\beta_{1,n}) + b_{tn}(\hat\beta_n)x + x\Big) + F_{\vartheta_0}(x) - I[\eta_t \le x]\Big\}. \qquad (15)
\end{aligned}
$$

Putting $A_n^{'} = \sup_x A_n^{'}(x)$, for $\hat\beta_n \in N_{L/\sqrt{n}}$, we can express

$$
A_n^{'} \le \max_{1\le j\le q(\iota)} \sup_r \sup_{x\in[x_r,x_{r+1})} (A_{1,n}^{'}(x) + A_{2,n}^{'}(x) + A_{3,n}^{'}(x) + A_{4,n}^{'}(x) + A_{5,n}^{'}(x))
$$

with

$$
\begin{aligned}
A_{1,n}^{'}(x) = &\Big| \tfrac{1}{\sqrt{n}} \textstyle\sum_{t=1}^{n} \Big\{ I\Big[\eta_t \le a_{tn}(\beta_{1,n}^j) + b_{tn}(\beta_n^j)x_{r+1} + d_{1tn} + d_{2tn}x_{r+1} + x_{r+1}\Big] \\
&-F_{\vartheta_0}\Big(a_{tn}(\beta_{1,n}^j) + b_{tn}(\beta_n^j)x_{r+1} + d_{1tn} + d_{2tn}x_{r+1} + x_{r+1}\Big) + F_{\vartheta_0}(x_{r+1}) - I[\eta_t \le x_{r+1}]\Big\}\Big|, \\
A_{2,n}^{'}(x) = &\Big| \tfrac{1}{\sqrt{n}} \textstyle\sum_{t=1}^{n} \Big\{ F_{\vartheta_0}\Big(a_{tn}(\beta_{1,n}^j) + b_{tn}(\beta_n^j)x_{r+1} + d_{1tn} + d_{2tn}x_{r+1} + x_{r+1}\Big) \\
&-F_{\vartheta_0}(a_{tn}(\hat\beta_{1,n}) + b_{tn}(\hat\beta_n)x + x)\Big\}\Big|, \\
A_{5,n}^{'}(x) = &\Big| \tfrac{1}{\sqrt{n}} \textstyle\sum_{t=1}^{n} \Big\{ I[\eta_t \le x_r] - F_{\vartheta_0}(x_r) + F_{\vartheta_0}(x) - I[\eta_t \le x]\Big\}\Big|,
\end{aligned}
$$

and $A_{3,n}^{'}(x)$ and $A_{4,n}(x)$ are the same as $A_{1,n}(x)$ and $A_{3,n}(x)$, with x_{r+1} and d_{itn} replaced by x_r and $-d_{itn}$, $i = 1,2$, respectively.

To show $A_n^{'} = o_p(1)$, we verify that $\max_{1\le j\le q(\lambda)} \sup_r \sup_{x_r < x \le x_{r+1}} A_{i,n}^{'}(x) = o_p(1)$, $i = 1,\ldots,5$. Below, we only provide the proof for the cases of $i = 1,2,5$, since the cases of $i = 3,4$ can be handled similarly.

We first deal with $A_{2,n}^{'}(x)$. By the mean value theorem, we can see that $\max_{1\le j\le q(\iota)} \sup_r \sup_{x\in[x_r,x_{r+1})} A_{2,n}^{'}(x)$ is no more than

$$
\begin{aligned}
&\max_{1\le j\le q(\iota)} \sup_r \sup_{x\in[x_r,x_{r+1})} \Big| \tfrac{1}{\sqrt{n}} \textstyle\sum_{t=\zeta_n+1}^{n} \Big\{ F_{\vartheta_0}\Big(a_{tn}(\beta_{1,n}^j) + b_{tn}(\beta_n^j)x_{r+1} + d_{1tn} + d_{2tn}x_{r+1} + x_{r+1}\Big) \\
&-F_{\vartheta_0}\Big(a_{tn}(\beta_{1,n}^j) + b_{tn}(\beta_n^j)x_r + d_{1tn} + d_{2tn}x_r + x_r\Big)\Big| + \Delta_n
\end{aligned} \qquad (16)
$$

with

$$
\begin{aligned}
\Delta_n = &\max_{1\le j\le q(\iota)} \sup_r \sup_{x\in[x_r,x_{r+1})} \Big| \frac{1}{\sqrt{n}} \sum_{t=\zeta_n+1}^{n} \Big\{ F_{\vartheta_0}\Big(a_{tn}(\beta_{1,n}^j) + b_{tn}(\beta_n^j)x_r + d_{1tn} + d_{2tn}x_r + x_r\Big) \\
- &F_{\vartheta_0}(a_{tn}(\hat\beta_{1,n}) + b_{tn}(\hat\beta_n)x_r + x_r)\Big\}\Big|.
\end{aligned}
$$

Note that the term in (16) is $o_p(1)$ due to Lemma 1, and

$$
\Delta_n \le max_{1\le j\le q(\iota)} \sup_r \sup_{x\in[x_r,x_{r+1})} \sum_{t=\zeta_n+1}^{n} f_{\vartheta_0}(x_t^*) \sup_x (1+|x|)(f_{\vartheta_0}(x) f\Big\{\Big|\frac{x}{x_t^*}\Big| d_{2tn} + d_{1tn}\Big\},
$$

where x_t^* is a real number between $a_{tn}(\beta_{1,n}^j) + b_{tn}(\beta_n^j)x_r + d_{1tn} + d_{2tn}x_r + x_r$ and $a_{tn}(\hat{\beta}_{1,n}) + b_{tn}(\hat{\beta}_n)x_r + x_r$. Using an argument similar to that in (12), we can see that $II_n = \iota O_p(1)$, which can be made arbitrarily small by taking sufficiently small ι. Hence, we get $\max_{1\le j\le q(\iota)} \sup_r \sup_{x_r<x\le x_{r+1}} A'_{2,n}(x) = o_p(1)$.

Next, because $|F_{\hat{\vartheta}_n}(x_{r+1}) - F_{\hat{\vartheta}_n}(x_r)| \le n^{-1/2-d}$, we can write

$$\begin{aligned}&\sup_r \sup_{x_r<x\le x_{r+1}} A'_{5,n}(x)\\ &\le \sup_r \frac{1}{\sqrt{n}} \left|\sum_{t=1}^n \{I(\eta_t \le x_{r+1}) - F_{\vartheta_0}(x_{r+1}) - F_{\vartheta_0}(x_r) + I(\eta_t \le x_r)\}\right| + o(1) = o_p(1).\end{aligned} \tag{17}$$

Hence, it remains to show that

$$\max_{1\le j\le q(\iota)} \sup_r \sup_{x_r<x\le x_{r+1}} A'_{1,n}(x) = o_p(1). \tag{18}$$

Put

$$\begin{aligned} e_{tn} &= I\left[\eta_t \le a_{tn}(\beta_{1,n}^j) + b_{tn}(\beta_n^j)x_r + d_{1tn} + d_{2tn}x_r + x_r\right] \\ &- F_{\vartheta_0}\left(a_{tn}(\beta_{1,n}^j) + b_{tn}(\beta_n^j)x_r + d_{1tn} + d_{2tn}x_r + x_r\right) + F_{\vartheta_0}(x_r) - I[\eta_t \le x_r]\Big\}, \quad 1 \le t \le n; \end{aligned}$$

and $S_{kn} = \sum_{t=1}^k e_{tn}$. Note that $\{S_{kn}; k = 1, \ldots, n\}, n \ge 1$, forms an array of martingale differences. Then, we get

$$P\Big(\max_{1\le j\le q(\iota)} \sup_x A'_{1,n}(x) \ge \epsilon\Big) = P\Big(\max_{1\le j\le q(\iota)} \max_r n^{-1/2}|S_{nn}| \ge \epsilon\Big), \tag{19}$$

and further, applying Rosenthal's inequality (Hall and Heyde [18], p. 23),

$$E[S_{nn}^4] \le C\Big(E\Big[\sum_{t=1}^n E(e_{tn}^2|\Omega_t)\Big]^2 + \sum_{t=1}^n E(e_{tn}^4)\Big), \quad C > 0. \tag{20}$$

By the mean value theorem, we can have $E(e_{tn}^2|\Omega_t) \le K f_{\vartheta_0}(x_t^*)||a_{tn}(\beta_{1,n}^j) + b_{tn}(\beta_n^j)x_r + d_{1tn} + d_{2tn}x_r|$ for some $K > 0$ and x_t^* between x_r and $a_{tn}(\beta_{1,n}^j) + b_{tn}(\beta_n^j)x_r + d_{1tn} + d_{2tn}x_r + x_r$, so that $E\Big[\sum_{t=1}^n E(e_{tn}^2|\Omega_t)\Big]^2 = O(n)$, by using an argument such as that in (12). Therefore, since $\sum_{t=1}^n E(e_{tn}^4) \le 16n$, we have $E[S_{nn}^4] = O(n)$ by (20). This, together with (19), validates the lemma. $\square$

Lemma 2. *Under the assumptions in Theorem 1, for every $d \in (0, 1/2)$, we have*

$$\begin{aligned}\Lambda_n := \sup_{\beta\in\mathcal{N}_{L/\sqrt{n}}} \sup_{(x,y)\in B_{d,n}} &\Big|\frac{1}{\sqrt{n}}\sum_{t=1}^n [F_{\vartheta_0}(a_{tn}(\beta_1) + b_{tn}(\beta)x + d_{1tn} + d_{2tn}x + x)\\ &- F_{\vartheta_0}(a_{tn}(\beta_1) + b_{tn}(\beta)y + d_{1tn} + d_{2tn}y + y)]\Big| = o_p(1)\end{aligned}$$

where $B_{d,n} = \{(x, y) \in \mathbb{R}^2 : |F_{\vartheta_0}(x) - F_{\vartheta_0}(y)| \le n^{-1/2-d}\}$.

Proof of Lemma 2. The lemma can be proven by using **(C2)**∼**(C4)** and the second-order Taylor's expansion theorem centered at x and y. We omit the details for brevity. $\square$

3. Discussion

In implementation, following the idea of Lee, Vonta and Karagrigoriou [9] and Lee, Lee and Park [10], we generate independent and identically distributed (i.i.d.) r.v.s w_{ij}, $j = 1, \cdots, J$, from $U[0, 1]$, where J

is a large integer (e.g., 1000), and then use $\tilde{w}_{ij} = \frac{w_{ij}}{w_{1j}+\cdots+w_{mj}}$ and $s_i = i/m$, $i = 1, \cdots, m$ to apply the test:

$$\begin{aligned}\hat{T}_n &= \sqrt{n}\max_{1\le j\le J}\left|\sum_{i=1}^m \tilde{w}_{ij}\left(\hat{F}_n\left(\tfrac{i}{m}\right) - \hat{F}_n\left(\tfrac{i-1}{m}\right)\right)\log m\left(\hat{F}_n\left(\tfrac{i}{m}\right) - \hat{F}_n\left(\tfrac{i-1}{m}\right)\right)\right| \\ &\overset{d}{\approx} \sup_{w\in W}\left|\sum_{i=1}^m w_i\left(\mathcal{B}\left(\tfrac{i}{m}\right) - \mathcal{B}\left(\tfrac{i-1}{m}\right)\right)\right|.\end{aligned} \tag{21}$$

The choice of m could be an important issue because the test performance might be sensitive to m. Here, we use $m = [n^{1/3}]$ because this has produced reasonably good results in our previous studies. The critical values could be obtained through Monte Carlo simulations as follows:

(i) From the data $X_1, \ldots, X_n$, estimate β and ϑ by suitable estimators $\hat{\beta}_n$ and $\hat{\vartheta}_n$; for example, MLE (Kim and Lee [2]).
(ii) Generate $\eta_1^*, \ldots, \eta_n^*$ from $F_{\hat{\vartheta}_n}(\cdot)$ and $Y_1^*, \ldots, Y_n^*$ using the equation: $Y_t^* = \tilde{g}_t(\hat{\beta}_{1,n}) + \tilde{h}_t(\hat{\beta}_n)\eta_t^*$. Then, obtain $\hat{T}_n$, denoted by $\hat{T}_n^*$, with the preassigned m in (21) based on these random variables.
(iii) Repeat the above procedure B times and calculate the $100(1-p)\%$ percentile of the obtained B number of $\hat{T}_n^*$ values.
(iv) Reject H_0 if the value of $\hat{T}_n$ obtained from the original observations is larger than the obtained $100(1-p)\%$ percentile in (iii).

The good performance of the entropy test for GARCH-type models can be seen in our previous works: Lee, Lee and Park [10], Lee and Oh [11], and Lee and Kim [4]. However, more refined empirical studies are required to see the performance of the above procedure in various location-scale models. Meanwhile, verifying the weak consistency of the $\hat{T}_n^*$ can be an important issue. The proof would be similar to that in Lee and Kim [4], which, however, needs much more careful analysis. All these issues are worth further investigation and are left as our future project.

4. Conclusions

In this study, we considered the entropy-based test for location-scale time series models and showed that it converges weakly to a functional of a Brownian bridge. As mentioned earlier, the bootstrap test in this setting deserves special attention owing to its importance in implementation. Furthermore, a modification of the entropy test based on integrated distributions is worth further investigation. We leave these issues to our future project.

Acknowledgments: We would like to thank the referees for their careful reading and valuable comments that improve the quality of the paper. This research is supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and future Planning (No. 2015R1A2A2A010003894).

Author Contributions: Sangyeol Lee and Minjo Kim conceived the topic, designed analysis tools and wrote the paper. Both authors have read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

GOF	goodness of fit test
GARCH	generalized autoregressive conditionally heteroscedastic
QMLE	quasi-maximum likelihood estimator
ASTD	asymmetric skew *t* distribution
AEPD	asymmetric exponential power distribution

References

1. Noh, J.; Lee, S. Quantile regression for location-scale time series models with conditional hetroscedasticity. *Scand. J. Stat.* **2015**, *43*, 700–720.
2. Kim, M.; Lee, S. Inferences for conditionally heteroscedastic location-scale time series models with ASTD and AEPD innovations. Unpublished work, 2016.
3. Lee, T.; Lee, S. Normal mixture quasi maximum likelihood estimator for GARCH models. *Scand. J. Stat.* **2009**, *36*, 157–170.
4. Lee, S.; Kim, M. On entropy based goodness of fit test for asymmetric student-*t* and exponential power distributions. *J. Stat. Comput. Simul.* **2017**, *87*, 187–197.
5. Hansen, B.E. Autoregressiveconditional density estimation. *Int. Econ. Rev. (Philadelphia)* **1994**, *35*, 705–730.
6. Bollerslev, T. Generalized autoregressive conditional heteroskedasticity. *J. Econ.* **1986**, *31*, 307–327.
7. Kim, M.; Lee, S. Comparison study of semiparametric methods to estimate conditional VaR and ES. *Korean J. Appl. Stat.* **2016**, *29*, 171–180. (In Korean)
8. D'Agostino, R.B.; Stephens, M.A. *Goodness-of-Fit Techniques*; Marcel Dekker, Inc.: New York, NY, USA, 1986.
9. Lee, S.; Vonta, I.; Karagrigoriou, A. A maximum entropy type test of fit. *Comput. Stat. Data Anal.* **2011**, *55*, 2635–2643.
10. Lee, J.; Lee, S.; Park, S. Maximum entropy test for GARCH models. *Stat. Method* **2015**, *22*, 8–16.
11. Lee, S.; Oh, H. Entropy test and residual empirical process for autoregressive conditional duration models. *Comput. Stat. Data Anal.* **2015**, *86*, 1–12.
12. Durbin, J. Weak convergence of the sample distribution function when parameters are estimated. *Ann. Stat.* **1973**, *1*, 279–290.
13. Lee, S.; Wei, C.Z. On residual empirical processes of stochastic regression models with applications to time series. *Ann. Stat.* **1999**, *27*, 237–261.
14. Lee, S.; Taniguchi, M. Asymptotic theory for ARCH-SM models: LAN and residual empirical process. *Stat. Sin.* **2005**, *15*, 215–234.
15. Rosenblatt, M. Remarks on a multivaraite transformation. *Ann. Math. Stat.* **1952**, *23*, 470–472.
16. Francq, C.; Zakoïan, J.M. Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* **2004**, *10*, 605–637.
17. Amemiya, T. Regression analysis when the dependent variable is truncated normal. *Econometrica* **1973**, *41*, 997–1016.
18. Hall, P.; Heyde. C.C. *Martingale Limit Theory and Its Application*; Academic Press: New York, NY, USA, 1980.