Biometrika (2017), **xx**, x, *pp*. 1–16
Printed in Great Britain

A robust goodness-of-fit test for generalized autoregressive conditional heteroscedastic models

BY YAO ZHENG, WAI KEUNG LI AND GUODONG LI

Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong

yaozheng@connect.hku.hk hrntlwk@hku.hk gdli@hku.hk

SUMMARY

Estimation of time series models with heavy-tailed innovations has been widely discussed, but corresponding goodness-of-fit tests have attracted less attention, primarily because the autocorrelation function commonly used in constructing goodness-of-fit tests necessarily imposes certain moment conditions on the innovations. As a bounded random variable has finite moments of all orders, we address the problem by first transforming the residuals with a bounded function. More specifically, we consider the sample autocorrelation function of the transformed absolute residuals of a fitted generalized autoregressive conditional heteroscedastic model. With the corresponding residual empirical distribution function naturally employed as the transformation, a robust goodness-of-fit test is then constructed. The asymptotic distributions of the test statistic under the null hypothesis and local alternatives are derived, and Monte Carlo experiments are conducted to examine finite-sample properties. The proposed test is shown to be more powerful than existing tests when the innovations are heavy-tailed.

Some key words: Conditional heteroscedastic model; Goodness-of-fit test; Heavy tail; Residual empirical process; Robustness.

1. INTRODUCTION

The heavy-tail phenomenon has attracted considerable attention in time series analysis, and great efforts have been made in model fitting and parameter estimation; see, e.g., Davis & Resnick (1986) and Ling (2005). The generalized autoregressive conditional heteroscedastic model (Engle, 1982; Bollerslev, 1986) is well-known for its success in capturing time-dependent conditional variances or scales, a feature often observed in financial data; see Zivot (2009) and Guo et al. (2017). Although a stationary generalized autoregressive conditional heteroscedastic process with Gaussian innovations can be heavy-tailed (He & Teräsvirta, 1999; Basrak et al., 2002), numerous empirical studies have shown that the residuals $\{\widehat{\varepsilon}_t\}$ of fitted generalized autoregressive conditional heteroscedastic models of financial returns appear to have high or even nonexistent kurtosis; see, e.g., Mittnik & Paolella (2003), Mikosch & Stărică (2000) and §6 of this paper. Various robust estimators that allow $E(\varepsilon_t^4) = \infty$ yet still achieve $\sqrt{n}$-consistency have been introduced. For example, the least absolute deviations estimator in Peng & Yao (2003) and the Pearsonian quasi-maximum likelihood estimator in Zhu & Li (2015) require only a finite fractional moment of ε_t, i.e., $E(|\varepsilon_t|^{2\gamma}) < \infty$ for some $\gamma > 0$, and the Laplacian quasi-maximum likelihood estimator in Berkes & Horváth (2004) requires $E(\varepsilon_t^2) < \infty$.

In contrast to the many studies on robust parameter estimation, research on the corresponding goodness-of-fit test, despite its importance, is still quite limited, primarily because the autocorrelation function commonly used in constructing the test imposes certain moment conditions on the innovations. As a bounded random variable has finite moments of all orders, we can remove such conditions through a bounded transformation. Although the distribution function of the innovations is a natural transformation, it is unknown in practice, so, an alternative is to employ the empirical distribution function of the residuals. For conditional heteroscedastic models, diagnostic tools constructed from the sample autocorrelation functions of squared residuals (Li & Mak, 1994) and absolute residuals (Li & Li, 2005) are particularly popular. However, the former require $E(\varepsilon_t^4) < \infty$ (Li, 2004), and the latter $E(\varepsilon_t^2) < \infty$ (Li & Li, 2005). Even worse, the convergence rates of these residual sample autocorrelation functions can become extremely slow under generalized autoregressive conditional heteroscedastic alternatives if $E(\varepsilon_t^4) = \infty$ (Davis & Mikosch, 1998; Basrak et al., 2002), possibly undermining the power of the corresponding test.

To address these problems, this paper constructs a robust goodness-of-fit test based on the sample autocorrelation function of transformed absolute residuals, where the transformation is the residual empirical distribution function. This test is shown to be asymptotically equivalent to the test where the transformation is the true distribution function of $|\varepsilon_t|$. We also derive the asymptotic power of the test based on transformed absolute residuals with any known function, which include existing methods such as those based on squared and absolute residual autocorrelations (Li & Mak, 1994; Li & Li, 2005) as special cases. Doing so makes it possible to theoretically compare the commonly-used goodness-of-fit tests in the literature. Our asymptotic analysis is crucially reliant on Lemmas A1 and A2 in the Appendix, which provide useful results for weighted residual empirical processes of generalized autoregressive conditional heteroscedastic models, and hence are of independent interest.

2. GOODNESS-OF-FIT TEST BASED ON TRANSFORMED ABSOLUTE RESIDUALS

2·1. *Goodness-of-fit test based on residual empirical processes*

Our null hypothesis is that the observed time series $\{y_1, \ldots, y_n\}$ is generated by the model

$$H_0: \quad y_t = \varepsilon_t h_t^{1/2}, \quad h_t = \omega_0 + \sum_{i=1}^{p} \alpha_{0i} y_{t-i}^2 + \sum_{j=1}^{q} \beta_{0j} h_{t-j}, \tag{1}$$

where $\{\varepsilon_t\}$ is a sequence of innovations. Denote by $\theta = (\omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)^{\mathrm{T}} \in \Theta$ the parameter vector of model (1), where the parameter space $\Theta \subset \mathbb{R}_+^{p+q+1}$ is a compact set, with $\mathbb{R}_+ = (0, \infty)$, and the true parameter vector, $\theta_0 = (\omega_0, \alpha_{01}, \ldots, \alpha_{0p}, \beta_{01}, \ldots, \beta_{0q})^{\mathrm{T}}$, is an interior point of Θ. We call model (1) the GARCH(p, q) model.

Assumption 1. Model (1) satisfies the following conditions: (i) the innovations $\{\varepsilon_t\}$ are independent and identically distributed with ε_t^2 following a non-degenerate distribution and $E(|\varepsilon_t|^{2\gamma}) < \infty$ for some $\gamma > 0$; (ii) $\{y_t\}$ is a strictly stationary and ergodic process; (iii) $\sum_{j=1}^{q} \beta_j < 1$ for all $\theta \in \Theta$; and (iv) the polynomials $\sum_{j=1}^{p} \alpha_{0j} z^j$ and $1 - \sum_{j=1}^{q} \beta_{0j} z^j$ have no common root.

A necessary and sufficient condition for Assumption 1(ii) is given in Bougerol & Picard (1992), and Assumption 1(iv) is for the identifiability of model (1) (Berkes et al., 2003; Francq & Zakoïan, 2004). We further restrict the innovations $\{\varepsilon_t\}$ of model (1) so that the estimator converges to θ_0 as $n \to \infty$; see Francq & Zakoïan (2010, pp. 231–235). For example, we assume $E(\varepsilon_t) = 0$ and $\mathrm{var}(\varepsilon_t) = 1$ for the Gaussian quasi-maximum likelihood estimator (Hall & Yao, 2003); $\mathrm{median}(|\varepsilon_t|) = 1$ for the least absolute deviations estimator (Peng & Yao, 2003; Chen &

Zhu, 2015); $E(\varepsilon_t) = 0$ and $E(|\varepsilon_t|) = 1$ for the Laplacian quasi-maximum likelihood estimator (Berkes & Horváth, 2004).

Define the functions

$$\varepsilon_t(\theta) = y_t / h_t^{1/2}(\theta), \quad h_t(\theta) = \omega + \sum_{i=1}^{p} \alpha_i y_{t-i}^2 + \sum_{j=1}^{q} \beta_j h_{t-j}(\theta). \tag{2}$$

Then $h_t(\theta_0) = h_t$ and $\varepsilon_t(\theta_0) = \varepsilon_t$. Because the recursive equation in (2) depends on past observations that are infinitely far away, in practice initial values are needed for $\{y_0^2, \ldots, y_{1-p}^2, h_0, \ldots, h_{1-q}\}$. For simplicity, we set them to zero and denote the corresponding functions by $\widetilde{\varepsilon}_t(\theta)$ and $\widetilde{h}_t(\theta)$; fixing these initial values does not affect our asymptotic results.

Let $\widehat{\theta}_n = (\widehat{\omega}, \widehat{\alpha}_1, \ldots, \widehat{\alpha}_p, \widehat{\beta}_1, \ldots, \widehat{\beta}_q)^{\mathrm{T}}$ be an estimator for model (1). The residuals of the fitted model are $\widehat{\varepsilon}_t = \widetilde{\varepsilon}_t(\widehat{\theta}_n) = y_t / \widehat{h}_t^{1/2}$, where $\widehat{h}_t = \widetilde{h}_t(\widehat{\theta}_n)$. In the literature, the sample autocorrelation function of absolute or squared residuals is commonly used to check the adequacy of fitted conditional heteroscedastic models, whereas that of the residuals usually has very low power (Li & Li, 2008). Hence, we focus on the absolute residuals $|\widehat{\varepsilon}_t|$. We first transform them with the residual empirical distribution function,

$$\widehat{G}_n(x) = \frac{1}{n} \sum_{t=1}^{n} I(|\widehat{\varepsilon}_t| \leq x) \quad (0 \leq x < \infty), \tag{3}$$

and obtain $\widehat{G}_n(|\widehat{\varepsilon}_t|)$. Let $G(\cdot)$ be the distribution function of $|\varepsilon_t|$, so $E\{G(|\varepsilon_t|)\} = 0{\cdot}5$. The sample autocorrelation function of $\{\widehat{G}_n(|\widehat{\varepsilon}_t|)\}$ at lag k can be defined as $\widehat{\rho}_k = \widehat{\gamma}_k / \widehat{\gamma}_0$, where the sample autocovariance function is

$$\widehat{\gamma}_k = \frac{1}{n} \sum_{t=k+1}^{n} \left\{ \widehat{G}_n(|\widehat{\varepsilon}_t|) - 0{\cdot}5 \right\} \left\{ \widehat{G}_n(|\widehat{\varepsilon}_{t-k}|) - 0{\cdot}5 \right\} \quad (k \geq 0). \tag{4}$$

Note that $\widehat{\gamma}_k$ would take the same value if the squared residuals $\widehat{\varepsilon}_t^2$ were used in (3) and (4).

Andreou & Werker (2015) considered the f-rank autocorrelation coefficients (Hallin & Puri, 1994) of the residuals and squared residuals of autoregressive models with generalized autoregressive conditional heteroscedastic errors, which are fitted by the Gaussian quasi-maximum likelihood method. The f-rank autocorrelation coefficients in Andreou & Werker (2015) have a symmetric form only when the reference distribution is Gaussian. The proposed $\widehat{\rho}_k$ has a symmetric and simple form, which can be interpreted as the Spearman rank correlation coefficient (Wald & Wolfowitz, 1943; Bartels, 1982; Dufour & Roy, 1985; Hallin et al., 1985). Andreou & Werker (2015) used the local asymptotic normality approach (Le Cam & Yang, 1990; van der Vaart, 1998; Andreou & Werker, 2012) to derive the limiting distributions of residual-based statistics. To apply the method of Andreou & Werker (2015), we would have to assume that the residuals are based on the true values of $\{y_0, y_{-1}, \ldots\}$, which are unobservable in practice. This problem is avoided by our asymptotic approach.

For a predetermined positive integer M, we first derive the asymptotic null distribution of $\widehat{\rho} = (\widehat{\rho}_1, \ldots, \widehat{\rho}_M)^{\mathrm{T}}$. Let $\mathcal{F}_t$ be the σ-field generated by $\{\varepsilon_t, \varepsilon_{t-1}, \ldots\}$, and $g(\cdot)$ be the density function of $|\varepsilon_t|$.

Assumption 2. Under H_0, the estimator $\widehat{\theta}_n$ admits the representation,

$$n^{1/2}(\widehat{\theta}_n - \theta_0) = n^{-1/2} \sum_{t=1}^{n} \xi_t + o_p(1),$$

where $\{\xi_t, \mathcal{F}_t\}$ is a strictly stationary and ergodic martingale difference sequence with $\Gamma = \mathrm{var}(\xi_t) < \infty$.

Assumption 3. The density g satisfies the following conditions: (i) $\lim_{x\to 0} xg(x) = 0$; (ii) $\lim_{x\to\infty} xg(x) = 0$; and (iii) g is continuous on $(0, \infty)$.

Let $\kappa = E\{|\varepsilon_t|g(|\varepsilon_t|)\}$ and let

$$\Sigma = I_M + 144\{0{\cdot}25\kappa^2 D\Gamma D^{\mathrm{T}} + 0{\cdot}5\kappa(DQ^{\mathrm{T}} + QD^{\mathrm{T}})\},$$

where I_M is the $M \times M$ identity matrix, $D = (d_1, \ldots, d_M)^{\mathrm{T}}$ and $Q = (q_1, \ldots, q_M)^{\mathrm{T}}$, with

$$d_k = E\left\{\frac{0{\cdot}5 - G(|\varepsilon_{t-k}|)}{h_t}\frac{\partial h_t(\theta_0)}{\partial\theta}\right\}, \quad q_k = E\left[\{G(|\varepsilon_t|) - 0{\cdot}5\}\{G(|\varepsilon_{t-k}|) - 0{\cdot}5\}\xi_t\right].$$

THEOREM 1. *Suppose that H_0 and Assumptions* 1 *through* 3 *hold. If Σ is positive definite, then $n^{1/2}\widehat{\rho} \to N(0, \Sigma)$ in distribution as $n \to \infty$.*

Because $g(x) = f(x) + f(-x)$ for $0 \le x < \infty$, where $f(\cdot)$ is the density function of ε_t, we can estimate κ by $\widehat{\kappa} = n^{-1}\sum_{t=1}^n |\widehat{\varepsilon}_t|\{\widehat{f}_n(|\widehat{\varepsilon}_t|) + \widehat{f}_n(-|\widehat{\varepsilon}_t|)\}$, where $\widehat{f}_n(\cdot)$ is the kernel density estimator of $f(\cdot)$. Let $\xi_t = \xi_t(\theta_0)$, i.e., the function $\xi_t(\theta)$ evaluated at θ_0. Let $\widetilde{\xi}_t(\theta)$ be obtained by replacing $\{y_0^2, \ldots, y_{1-p}^2, h_0, \ldots, h_{1-q}\}$ with their initial values in $\xi_t(\theta)$, and denote $\widehat{\xi}_t = \widetilde{\xi}_t(\widehat{\theta}_n)$. We can estimate Γ, D and Q respectively by $\widehat{\Gamma} = n^{-1}\sum_{t=1}^n \widehat{\xi}_t\widehat{\xi}_t^{\mathrm{T}}$, $\widehat{D} = (\widehat{d}_1, \ldots, \widehat{d}_M)^{\mathrm{T}}$ and $\widehat{Q} = (\widehat{q}_1, \ldots, \widehat{q}_M)^{\mathrm{T}}$, where $\widehat{d}_k = n^{-1}\sum_{t=k+1}^n \widehat{h}_t^{-1}\{0{\cdot}5 - \widehat{G}_n(|\widehat{\varepsilon}_{t-k}|)\}\partial\widetilde{h}_t(\widehat{\theta}_n)/\partial\theta$ and $\widehat{q}_k = n^{-1}\sum_{t=k+1}^n\{\widehat{G}_n(|\widehat{\varepsilon}_t|) - 0{\cdot}5\}\{\widehat{G}_n(|\widehat{\varepsilon}_{t-k}|) - 0{\cdot}5\}\widehat{\xi}_t$. Under the conditions of Theorem 1, it can be shown that $\widehat{\kappa} = \kappa + o_p(1)$, $\widehat{\Gamma} = \Gamma + o_p(1)$, $\widehat{D} = D + o_p(1)$ and $\widehat{Q} = Q + o_p(1)$. Thus, a consistent estimator $\widehat{\Sigma}$ of Σ can be obtained, leading us to construct the test statistic

$$Q(M) = n\widehat{\rho}^{\mathrm{T}}\widehat{\Sigma}^{-1}\widehat{\rho},$$

which under H_0 is asymptotically distributed as χ_M^2, the chi-squared distribution with M degrees of freedom. One could also employ $n^{1/2}\widehat{\rho}_k/\widehat{\Sigma}_{kk}^{1/2}$ to examine the significance of the residual autocorrelation at lag k individually, where $\widehat{\Sigma}_{kk}$ is the kth diagonal element of $\widehat{\Sigma}$.

2·2. *Goodness-of-fit test based on predetermined transformations*

We can also consider the transformation with any predetermined function $\Psi(\cdot)$. The sample autocorrelation function of $\{\Psi(|\widehat{\varepsilon}_t|)\}$ at lag k can be defined as $\widehat{\rho}_k^{\Psi} = \widehat{\gamma}_k^{\Psi}/\widehat{\gamma}_0^{\Psi}$, where

$$\widehat{\gamma}_k^{\Psi} = \frac{1}{n}\sum_{t=k+1}^n \{\Psi(|\widehat{\varepsilon}_t|) - \widehat{\mu}_\Psi\}\{\Psi(|\widehat{\varepsilon}_{t-k}|) - \widehat{\mu}_\Psi\} \quad (k \ge 0),$$

with $\widehat{\mu}_\Psi = n^{-1}\sum_{t=1}^n \Psi(|\widehat{\varepsilon}_t|)$, is the sample autocovariance function. Let $\widehat{\rho}_\Psi = (\widehat{\rho}_1^{\Psi}, \ldots, \widehat{\rho}_M^{\Psi})^{\mathrm{T}}$.

Denote the first and second derivatives of Ψ by ψ and $\dot{\psi}$. Let $\mu_\Psi = E\{\Psi(|\varepsilon_t|)\}$, $\sigma_\Psi^2 = \mathrm{var}\{\Psi(|\varepsilon_t|)\}$, $\kappa_\Psi = E\{|\varepsilon_t|\psi(|\varepsilon_t|)\}$ and

$$\Sigma_\Psi = I_M + \sigma_\Psi^{-4}\left\{0{\cdot}25\kappa_\Psi^2 D_\Psi\Gamma D_\Psi^{\mathrm{T}} + 0{\cdot}5\kappa_\Psi\left(D_\Psi Q_\Psi^{\mathrm{T}} + Q_\Psi D_\Psi^{\mathrm{T}}\right)\right\},$$

where $D_\Psi = (d_1^\Psi, \ldots, d_M^\Psi)^{\mathrm{T}}$ and $Q_\Psi = (q_1^\Psi, \ldots, q_M^\Psi)^{\mathrm{T}}$, with

$$d_k^\Psi = E\left\{\frac{\mu_\Psi - \Psi(|\varepsilon_{t-k}|)}{h_t}\frac{\partial h_t(\theta_0)}{\partial\theta}\right\}, \quad q_k^\Psi = E\left[\{\Psi(|\varepsilon_t|) - \mu_\Psi\}\{\Psi(|\varepsilon_{t-k}|) - \mu_\Psi\}\xi_t\right].$$

Assumption 4. There exists $m > 0$ such that function $\Psi^*(x) = |\dot{\psi}(x)|x^2 + |\psi(x)|x$ satisfies $\Psi^*(x) \leq Cx^m$ as $x > 1$ and $\Psi^*(x) \leq C$ as $0 \leq x \leq 1$, where $C > 0$ is a constant, and $E(|\varepsilon_t|^m) < \infty$ and $E\{\Psi^2(|\varepsilon_t|)\} < \infty$.

THEOREM 2. *Suppose that H_0 and Assumptions 1, 2 and 4 hold. If Σ_Ψ is positive definite, then $n^{1/2}\widehat{\rho}_\Psi \to N(0, \Sigma_\Psi)$ in distribution as $n \to \infty$.*

Similarly, we can obtain a consistent estimator $\widehat{\Sigma}_\Psi$ of the asymptotic covariance matrix Σ_Ψ using sample averages. Thus, a goodness-of-fit test, $Q_\Psi(M) = n\widehat{\rho}_\Psi^{\mathrm{T}}\widehat{\Sigma}_\Psi^{-1}\widehat{\rho}_\Psi$, can be constructed.

The first interesting example is $\Psi(x) = x^c$ for some $c > 0$, and Assumption 4 is implied by $E(|\varepsilon_t|^{2c}) < \infty$. This example includes existing tests based on absolute and squared residuals, which correspond to cases with $c = 1$ and 2, respectively; see Li & Li (2005) and Li (2004). From the proof of Theorem 1, when Ψ is bounded, Theorem 2 still holds if, instead of Assumption 4, the derivative ψ satisfies the conditions on the density g in Assumption 3. For Theorem 4 in §3, the conditions can be similarly substituted.

Motivated by transformation $\widehat{G}_n$ in the previous subsection, we can also consider $\Psi = G$, although G is unknown in practice. Let G_n denote the empirical distribution function of $\{|\varepsilon_t|\}$, defined as $G_n(x) = n^{-1}\sum_{t=1}^n I(|\varepsilon_t| \leq x)$ for $0 \leq x < \infty$. From the proofs of Theorems 1 and 2, it can be readily verified that $n^{1/2}\widehat{\rho}_k$, $n^{1/2}\widehat{\rho}_k^{G_n}$ and $n^{1/2}\widehat{\rho}_k^{G}$ are asymptotically equivalent:

PROPOSITION 1. *Suppose that H_0 and Assumptions 1 and 3 hold with $n^{1/2}(\widehat{\theta}_n - \theta_0) = O_p(1)$. Then, $n^{1/2}(\widehat{\gamma}_k - \widehat{\gamma}_k^G) = o_p(1)$ and $n^{1/2}(\widehat{\gamma}_k - \widehat{\gamma}_k^{G_n}) = o_p(1)$ for any positive integer k. Moreover, $\widehat{\gamma}_0$, $\widehat{\gamma}_0^G$ and $\widehat{\gamma}_0^{G_n}$ all converge in probability to $1/12$ as $n \to \infty$.*

To apply joint tests $Q(M)$ and $Q_\Psi(M)$, we can consider several specific values of order M or select M as

$$\widetilde{M} = \underset{d_{\min} \leq M \leq d_{\max}}{\operatorname{argmax}} \{Q(M) - M\log n\}, \quad \widetilde{M}_\Psi = \underset{d_{\min} \leq M \leq d_{\max}}{\operatorname{argmax}} \{Q_\Psi(M) - M\log n\}, \tag{5}$$

where integer M is searched over a fixed range $[d_{\min}, d_{\max}]$ for $d_{\min} \geq 1$ and some large enough $d_{\max}$. As shown in §5, the performance of the automatic tests is insensitive to the choice of $d_{\max}$.

COROLLARY 1. *(i) Under the conditions of Theorem 1, $Q(\widetilde{M}) \to \chi^2_{d_{min}}$ in distribution as $n \to \infty$; and (ii) under the conditions of Theorem 2, $Q_\Psi(\widetilde{M}_\Psi) \to \chi^2_{d_{min}}$ in distribution as $n \to \infty$.*

In §3, we demonstrate that under the local alternatives, $n^{1/2}\widehat{\rho}$ is asymptotically normal with a possible shift in the mean, $\varUpsilon = (\varUpsilon_1, \ldots, \varUpsilon_M)^{\mathrm{T}}$; see Theorem 3. As a result, $\lim_{n\to\infty} \mathrm{pr}(\widetilde{M} = d_{\min}) = 1$, which may be undesirable for particular local alternatives with $\varUpsilon_1 = \cdots = \varUpsilon_{d_{\min}} = 0$ and $\varUpsilon_K \neq 0$ for some $d_{\min} < K \leq d_{\max}$, as in such cases $Q(\widetilde{M})$ would have no power. The test $Q_\Psi(\widetilde{M}_\Psi)$ would suffer from the same problem, which can be avoided by using a smaller penalty, e.g., the Akaike-information-criterion-type penalty, $2M$, to ensure that the probability of choosing a value of M larger than $d_{\min}$ is nonzero. However, as shown in §5, doing so may lead to seriously inflated Type I error rates.

In practice, the aforementioned problem can be remedied by choosing a proper $d_{\min}$. Suppose that sample autocorrelation function $\widehat{\rho}_k$ falls clearly outside the 95% confidence interval at certain lags. To guarantee that the joint test, $Q(\widetilde{M})$, takes into account at least one of the lags, we need only choose $d_{\min}$ as the smallest such lag by simply examining the plot of the residual autocorrelations, $\widehat{\rho}_k$; if no such exists, then we may set $d_{\min} = 1$.

3. ASYMPTOTIC POWER UNDER LOCAL ALTERNATIVES

To study the power of the proposed test, we consider the following local alternatives. For each n, the observed time series $\{y_{1,n}, \ldots, y_{n,n}\}$ is generated by

$$H_{1n}: \quad y_{t,n} = \varepsilon_t h_{t,n}^{1/2}, \quad h_{t,n} = \omega_0 + \sum_{i=1}^{p} \alpha_{0i} y_{t-i,n}^2 + \sum_{j=1}^{q} \beta_{0j} h_{t-j,n} + n^{-1/2} s_{t,n}, \tag{6}$$

where the subscript n is used to emphasize the dependence of $y_{t,n}, h_{t,n}$ and $s_{t,n}$ on n. For simplicity, we consider $s_{t,n} = s(y_{t-1,n}^2, \ldots, y_{t-p^*,n}^2, h_{t-1,n}, \ldots, h_{t-q^*,n})$ for some positive integers $p^* > p$ and $q^* > q$, where the function s satisfies the following condition.

Assumption 5. The function s and all elements of its gradient ∇s are nonnegative everywhere.

Assumption 6. There exists a positive integer n_0 such that, for each $n \geq n_0$, $\{y_{t,n}\}$ and $\{h_{t,n}\}$ are strictly stationary and ergodic processes, and $E(s_{t,n_0}^{\delta_0}) < \infty$ for some constant $\delta_0 > 0$ independent of n.

The nonnegativity of s guarantees that $h_{t,n} \geq 0$; see Nelson & Cao (1992) for a discussion of the relaxation of the nonnegativity constraints on the parameters of generalized autoregressive conditional heteroscedastic models. The condition $\nabla s \geq 0$ is used to simplify our technical proofs, and we can similarly derive asymptotic results for other cases of ∇s. The finite fractional moment of s_{t,n_0} in Assumption 6 ensures those of $y_{t,n}$ and $h_{t,n}$, which are needed in our proofs.

Similar to (2), we define the functions

$$\varepsilon_{t,n}(\theta) = y_{t,n}/h_{t,n}^{1/2}(\theta), \quad h_{t,n}(\theta) = \omega + \sum_{i=1}^{p} \alpha_i y_{t-i,n}^2 + \sum_{j=1}^{q} \beta_j h_{t-j,n}(\theta).$$

For simplicity, with the initial values set to be independent of n, we denote the resulting functions by $\widetilde{\varepsilon}_{t,n}(\theta)$ and $\widetilde{h}_{t,n}(\theta)$, respectively. Under H_{1n}, the residuals are calculated as $\widehat{\varepsilon}_t = \widetilde{\varepsilon}_{t,n}(\widehat{\theta}_n) = y_{t,n}/\widehat{h}_{t,n}^{1/2}$, where $\widehat{h}_{t,n} = \widetilde{h}_{t,n}(\widehat{\theta}_n)$.

Whilst $h_t = h_t(\theta_0)$ in (2), we can show that the departure $n^{-1/2} s_{t,n}$ in (6) results in

$$h_{t,n} - h_{t,n}(\theta_0) = n^{-1/2} \sum_{k=0}^{\infty} e_1^{\mathrm{T}} B_0^k e_1 s_{t-k,n} \geq 0,$$

where $e_1 = (1, 0, \ldots, 0)^{\mathrm{T}}$, and

$$B_0 = \begin{pmatrix} \beta_{01} & \cdots & \beta_{0q-1} & \beta_{0q} \\ & I_{q-1} & & 0 \end{pmatrix}$$

is a $q \times q$ matrix. Define the nonnegative, $\mathcal{F}_{t-1}$-measurable random variables

$$r_{t,n} = \frac{n^{1/2}\{h_{t,n} - h_{t,n}(\theta_0)\}}{h_{t,n}(\theta_0)} = \frac{1}{h_{t,n}(\theta_0)} \sum_{k=0}^{\infty} e_1^{\mathrm{T}} B_0^k e_1 s_{t-k,n}.$$

Let $s_t = s(y_{t-1}^2, \ldots, y_{t-p^*}^2, h_{t-1}, \ldots, h_{t-q^*})$ and

$$r_t = \frac{1}{h_t} \sum_{k=0}^{\infty} e_1^{\mathrm{T}} B_0^k e_1 s_{t-k}. \tag{7}$$

Assumption 7. There exist processes $\{r_{t,n}^{(l)} : t = 1 \ldots, n\}$ and $\{r_{t,n}^{(u)} : t = 1 \ldots, n\}$ for each n satisfying the following conditions: (i) all $r_{t,n}^{(l)}$ and $r_{t,n}^{(u)}$ are $\mathcal{F}_{t-1}$-measurable; (ii) the processes $\{r_{t,n_0}^{(l)}\}$ and $\{r_{t,n_0}^{(u)}\}$ are strictly stationary and ergodic with $r_{t,n_0}^{(l)} \leq r_{t,n} \leq r_{t,n_0}^{(u)}$ for all $n \geq n_0$; and (iii) for each fixed t, $r_{t,n}^{(l)}$ increases monotonically with n, whereas $r_{t,n}^{(u)}$ decreases monotonically with n, i.e., $r_{t,n}^{(l)} \leq r_{t,n+1}^{(l)} \leq r_{t,n+1}^{(u)} \leq r_{t,n}^{(u)}$ for all n, and $\lim_{n\to\infty} r_{t,n}^{(l)} = \lim_{n\to\infty} r_{t,n}^{(u)} = r_t$ with probability one.

PROPOSITION 2. *Consider the case of* $s_{t,n} = a_0 + \sum_{i=1}^{p^*} a_i y_{t-i,n}^2 + \sum_{j=1}^{q^*} a_{p^*+j} h_{t-j,n}$, *where* $a_0, a_1, \ldots, a_{p^*+q^*}$ *are nonnegative constants. Under Assumptions* 1 *and* 6, *if* $q > 0$, *then the conditions in Assumption* 7 *hold and* $E\{(r_{t,n_0}^{(u)})^m\} < \infty$ *for any* $m > 0$.

For other forms of $s_{t,n}$, Assumption 7 can also be readily verified, although additional moment restrictions on $y_{t,n}$ may be required.

Assumption $2'$. Under H_{1n}, the estimator $\widehat{\theta}_n$ admits the representation,

$$n^{1/2}(\widehat{\theta}_n - \theta_0) = n^{-1/2} \sum_{t=1}^{n} \xi_{t,n} + \Delta + o_p(1),$$

where $\{\xi_{t,n}, \mathcal{F}_t : t = 1 \ldots, n\}$ is a strictly stationary and ergodic martingale difference sequence for each sufficiently large n, $\lim_{n\to\infty} \text{var}(\xi_{t,n}) = \Gamma$, and $\Delta \in \mathbb{R}^{p+q+1}$ is a constant vector.

It is possible to derive the explicit form of Δ under additional regularity conditions of estimator $\widehat{\theta}_n$ and those of the underlying model in (6). Specifically, assuming that model (6) is locally asymptotically normal (van der Vaart, 1998), by Le Cam's third lemma, the shift $\Delta = \lim_{n\to\infty} \text{cov}\{n^{-1/2} \sum_{t=1}^{n} \xi_t, \Delta^{(n)}(\theta_0)\}$, where $\Delta^{(n)}(\theta_0) = -0{\cdot}5 n^{-1/2} \sum_{t=1}^{n} \{1 + \varepsilon_t f'(\varepsilon_t)/f(\varepsilon_t)\} h_t^{-1} \partial h_t(\theta_0)/\partial\theta$ is the central sequence of the GARCH(p, q) model (Drost & Klaassen, 1997).

Let $V = (v_1, \ldots, v_M)^{\mathrm{T}}$ with $v_k = E\left[\{0{\cdot}5 - G(|\varepsilon_{t-k}|)\} r_t\right]$, and $V_\Psi = (v_1^\Psi, \ldots, v_M^\Psi)^{\mathrm{T}}$ with $v_k^\Psi = E\left[\{\mu_\Psi - \Psi(|\varepsilon_{t-k}|)\} r_t\right]$.

THEOREM 3. *Suppose that* H_{1n} *and Assumptions* 1, $2'$, 3 *and* 5 *through* 7 *hold with* $E\{(r_{t,n_0}^{(u)})^{4+\delta_1}\} < \infty$ *for some* $\delta_1 > 0$. *If* Σ *is positive definite, then* $n^{1/2}\widehat{\rho} \to N(\Upsilon, \Sigma)$ *in distribution as* $n \to \infty$, *where* $\Upsilon = 6\kappa(D\Delta - V)$, *with* κ, D *and* Σ *defined as in Theorem* 1.

THEOREM 4. *Suppose that* H_{1n} *and Assumptions* 1, $2'$ *and* 4 *through* 7 *hold with* $E\{(r_{t,n_0}^{(u)})^{4+\delta_1}\} < \infty$ *for some* $\delta_1 > 0$. *If* Σ_Ψ *is positive definite, then* $n^{1/2}\widehat{\rho}_\Psi \to N(\Upsilon_\Psi, \Sigma_\Psi)$ *in distribution as* $n \to \infty$, *where* $\Upsilon_\Psi = 0{\cdot}5\kappa_\Psi(D_\Psi\Delta - V_\Psi)/\sigma_\Psi^2$, *with* κ_Ψ, D_Ψ, σ_Ψ *and* Σ_Ψ *defined as in Theorem* 2.

We can show that, under H_{1n}, the consistency of estimators $\widehat{\Sigma}$ and $\widehat{\Sigma}_\Psi$ in the previous section still holds, and hence $Q(M)$ and $Q_\Psi(M)$ converge to the noncentral χ_M^2 distribution with noncentrality parameter $c_\Psi = \Upsilon_\Psi^{\mathrm{T}} \Sigma_\Psi^{-1} \Upsilon_\Psi$ as $n \to \infty$, where $\Psi = G$ for $Q(M)$. In other words, the local power is determined by the value of c_Ψ.

4. TWO APPLICATIONS

This section applies the asymptotic results in §2 and §3 to generalized autoregressive conditionally heteroscedastic models fitted by the Laplacian quasi-maximum likelihood method (Berkes & Horváth, 2004) and least absolute deviations method (Peng & Yao, 2003).

We first derive the asymptotic distributions of these two estimators under H_{1n}. Denote

$$J = E\left\{\frac{1}{h_t^2}\frac{\partial h_t(\theta_0)}{\partial\theta}\frac{\partial h_t(\theta_0)}{\partial\theta^{\mathrm{T}}}\right\}, \quad \lambda = E\left\{\frac{r_t}{h_t}\frac{\partial h_t(\theta_0)}{\partial\theta}\right\},$$

where r_t is defined as in (7). For model (1), the Laplacian quasi-maximum likelihood estimator (Berkes & Horváth, 2004) is defined as $\widehat{\theta}_n^{\mathrm{LQML}} = \mathrm{argmin}_{\theta\in\Theta}\, n^{-1}\sum_{t=1}^n\{\log \widetilde{h}_t^{1/2}(\theta) + |y_t|/\widetilde{h}_t^{1/2}(\theta)\}$, where the identifiability conditions are $E(\varepsilon_t) = 0$ and $E(|\varepsilon_t|) = 1$. Under H_0 and Assumption 1, if $E(\varepsilon_t^2) < \infty$, then we can show that

$$n^{1/2}(\widehat{\theta}_n^{\mathrm{LQML}} - \theta_0) = \frac{2J^{-1}}{n^{1/2}}\sum_{t=1}^n \frac{|\varepsilon_t| - 1}{h_t}\frac{\partial h_t(\theta_0)}{\partial\theta} + o_p(1),$$

which converges in distribution to $N[0, 4\{E(\varepsilon_t^2) - 1\}J^{-1}]$ as $n \to \infty$.

THEOREM 5. *Suppose that H_{1n} and Assumptions 1 and 5 through 7 hold. If $E(r_{t,n_0}^{(u)}) < \infty$, then $\widehat{\theta}_n^{LQML} \to \theta_0$ almost surely as $n \to \infty$. Moreover, if $E(\varepsilon_t^2) < \infty$ and $E\{(r_{t,n_0}^{(u)})^{2+\delta_1}\} < \infty$ for some $\delta_1 > 0$, then*

$$n^{1/2}(\widehat{\theta}_n^{LQML} - \theta_0) = \frac{2J^{-1}}{n^{1/2}}\sum_{t=1}^n \frac{|\varepsilon_t| - 1}{h_{t,n}(\theta_0)}\frac{\partial h_{t,n}(\theta_0)}{\partial\theta} + J^{-1}\lambda + o_p(1),$$

which converges in distribution to $N[J^{-1}\lambda, 4\{E(\varepsilon_t^2) - 1\}J^{-1}]$ as $n \to \infty$.

For model (1), the least absolute deviations estimator in Peng & Yao (2003) is defined as $\widehat{\theta}_n^{\mathrm{LAD}} = \mathrm{argmin}_{\theta\in\Theta}\, n^{-1}\sum_{t=1}^n |\log y_t^2 - \log \widetilde{h}_t(\theta)|$, where the identifiability condition is $\mathrm{median}(|\varepsilon_t|) = 1$. Under H_0 and Assumption 1, if $g(1) > 0$, then it can be verified that

$$n^{1/2}(\widehat{\theta}_n^{\mathrm{LAD}} - \theta_0) = \frac{\{g(1)J\}^{-1}}{n^{1/2}}\sum_{t=1}^n \frac{\mathrm{sgn}(|\varepsilon_t| - 1)}{h_t}\frac{\partial h_t(\theta_0)}{\partial\theta} + o_p(1),$$

which converges in distribution to $N[0, \{g(1)\}^{-2}J^{-1}]$ as $n \to \infty$, where $\mathrm{sgn}(x) = I(x > 0) - I(x < 0)$ is the sign function; see Chen & Zhu (2015).

THEOREM 6. *If H_{1n} and Assumptions 1 and 5 through 7 hold, then $\widehat{\theta}_n^{LAD} \to \theta_0$ almost surely as $n \to \infty$. Moreover, if $g(1) > 0$ and $E\{(r_{t,n_0}^{(u)})^{4+\delta_1}\} < \infty$ for some $\delta_1 > 0$, then*

$$n^{1/2}(\widehat{\theta}_n^{LAD} - \theta_0) = \frac{\{g(1)J\}^{-1}}{n^{1/2}}\sum_{t=1}^n \frac{\mathrm{sgn}(|\varepsilon_t| - 1)}{h_{t,n}(\theta_0)}\frac{\partial h_{t,n}(\theta_0)}{\partial\theta} + J^{-1}\lambda + o_p(1),$$

which converges in distribution to $N[J^{-1}\lambda, \{g(1)\}^{-2}J^{-1}]$ as $n \to \infty$.

Given Theorems 5 and 6, the estimators $\widehat{\theta}_n^{\mathrm{LQML}}$ and $\widehat{\theta}_n^{\mathrm{LAD}}$ both satisfy Assumptions 2 and 2′ with $\Delta = J^{-1}\lambda$, and we can then obtain the asymptotic distributions of $n^{1/2}\widehat{\rho}$ and $n^{1/2}\widehat{\rho}_\Psi$ under both H_0 and H_{1n}. Moreover, Theorems 1–4 ensure that the proposed statistic, $n^{1/2}\widehat{\rho}$, has the same asymptotic distributions as $n^{1/2}\widehat{\rho}_\Psi$ with $\Psi = G$ under both H_0 and H_{1n}. Thus, we focus on $n^{1/2}\widehat{\rho}_\Psi$ in the following discussion.

By Theorems 1–4, under both H_0 and H_{1n}, the asymptotic covariance matrix of $n^{1/2}\widehat{\rho}_\Psi$ is

$$\Sigma_\Psi = I_M + \sigma_\Psi^{-4}\big(\kappa_\Psi^2\{E(\varepsilon_t^2)-1\} + 2\kappa_\Psi E\left[\{\mu_\Psi - \Psi(|\varepsilon_t|)\}(|\varepsilon_t|-1)\right]\big)D_\Psi J^{-1} D_\Psi^{\mathrm{T}}$$

for the Laplacian quasi-maximum likelihood estimator $\widehat{\theta}_n^{\mathrm{LQML}}$, and is

$$\Sigma_\Psi = I_M + \sigma_\Psi^{-4}\left\{\frac{\kappa_\Psi^2}{4g^2(1)} + \frac{\kappa_\Psi}{g(1)}E\left[\{\mu_\Psi - \Psi(|\varepsilon_t|)\}\mathrm{sgn}(|\varepsilon_t|-1)\right]\right\}D_\Psi J^{-1} D_\Psi^{\mathrm{T}}$$

for the least absolute deviations estimator $\widehat{\theta}_n^{\mathrm{LAD}}$. When $\Psi = G$, $\sigma_\Psi^2 = 1/12$. Moreover, under H_{1n}, the asymptotic distributions of $n^{1/2}\widehat{\rho}_\Psi$ for both estimators are shifted by

$$\varUpsilon_\Psi = 0{\cdot}5\kappa_\Psi(D_\Psi J^{-1}\lambda - V_\Psi)/\sigma_\Psi^2.$$

We now consider the case when $\varUpsilon_\Psi$ is nonzero. Let $b_1 = \mathrm{argmin}_{b\in\mathbb{R}^{p+q+1}} E\{(r_t - X_t^{\mathrm{T}}b)^2\}$ and $b_2 = \mathrm{argmin}_{b\in\mathbb{R}^{p+q+1}} E[\{\Psi(|\varepsilon_{t-k}|) - X_t^{\mathrm{T}}b\}^2]$, where $X_t = h_t^{-1}\partial h_t(\theta_0)/\partial\theta$. Define the partial covariance (Fan & Yao, 2003),

$$\mathrm{pcov}\{r_t, \Psi(|\varepsilon_{t-k}|) \mid X_t\} = E[(r_t - X_t^{\mathrm{T}}b_1)\{\Psi(|\varepsilon_{t-k}|) - X_t^{\mathrm{T}}b_2\}]. \tag{8}$$

Because $b_1 = J^{-1}\lambda$, the kth element of the term $D_\Psi J^{-1}\lambda - V_\Psi$, i.e., $d_k^{\Psi\mathrm{T}}J^{-1}\lambda - v_k^\Psi$, can be written as $-\mathrm{pcov}\{r_t, \Psi(|\varepsilon_{t-k}|) \mid X_t\}$. Moreover, as $\kappa_\Psi > 0$, the kth element of $\varUpsilon_\Psi$ is zero if and only if the partial covariance in (8) is zero.

Consider the example in Proposition 2, where we have $s_t = s_{1,t} + s_{2,t}$, with $s_{1,t} = a_0 + \sum_{i=1}^p a_i y_{t-i}^2 + \sum_{j=1}^q a_{p^*+j}h_{t-j}$ and $s_{2,t} = \sum_{i=p+1}^{p^*} a_i y_{t-i}^2 + \sum_{j=q+1}^{q^*} a_{p^*+j}h_{t-j}$. Then, $r_t = X_t^{\mathrm{T}}a + h_t^{-1}\sum_{k=0}^\infty e_1^{\mathrm{T}}B_0^k e_1 s_{2,t-k}$, where $a = (a_0, a_1, \ldots, a_p, a_{p^*+1}, \ldots, a_{p^*+q})^{\mathrm{T}}$. As a result, when $s_{2,t} = 0$, i.e., the model is correctly specified, the partial covariance in (8) is zero for all $k > 0$, and the test $Q_\Psi(M)$ has no power. When the model is misspecified, i.e., $s_{2,t} \neq 0$, by a method similar to the proof of identifiability for generalized autoregressive conditional heteroscedastic models (Francq & Zakoïan, 2004), we can show that $r_t - X_t^{\mathrm{T}}b_1 \neq 0$ with probability one, provided that Assumption 1 holds. Thus, (8) becomes nonzero at some k's, resulting in nontrivial power for the test.

Table 1. *Noncentrality parameter c_Ψ $(\times 10^2)$ under different local alternatives of the* GARCH$(1,1)$ *model with* $(\omega_0, \alpha_0, \beta_0) = (1, 0{\cdot}3, 0{\cdot}2)$*, for* $\Psi(x) = G(x), x$ *and* x^2

	$s_{t,n} = G(\lvert y_{t-2,n}\rvert)$			$s_{t,n} = \lvert y_{t-2,n}\rvert$			$s_{t,n} = y_{t-2,n}^2$		
	G	x	x^2	G	x	x^2	G	x	x^2
t_1	3E-05			2E-03			99·52		
$t_{2\cdot5}$	0·05	3E-03		1·17	0·13		31·38	8·45	
t_3	0·07	0·01		1·32	0·27		26·10	12·15	
t_5	0·10	0·03	3E-03	1·42	0·72	0·11	17·07	16·86	3·98
t_7	0·11	0·05	0·01	1·42	0·93	0·26	14·35	16·96	7·91
Normal	0·15	0·10	0·04	1·36	1·25	0·74	9·32	13·62	12·80

Small numbers are written in standard form: e.g., 3E-05 refers to 3×10^{-5}.

In general, the local power of $Q_\Psi(M)$ is determined by the noncentrality parameter $c_\Psi = \varUpsilon_\Psi^{\mathrm{T}}\Sigma_\Psi^{-1}\varUpsilon_\Psi$, which depends on the departure $s_{t,n}$, the underlying model, the estimator $\widehat{\theta}_n$ and the function Ψ. It is difficult to make a direct comparison of the values of c_Ψ across different Ψ's. We next calculate c_Ψ for specific scenarios. Table 1 presents the values of c_Ψ under local alternatives of the GARCH$(1,1)$ model with $(\omega_0, \alpha_0, \beta_0) = (1, 0{\cdot}3, 0{\cdot}2)$ and three types of departure,

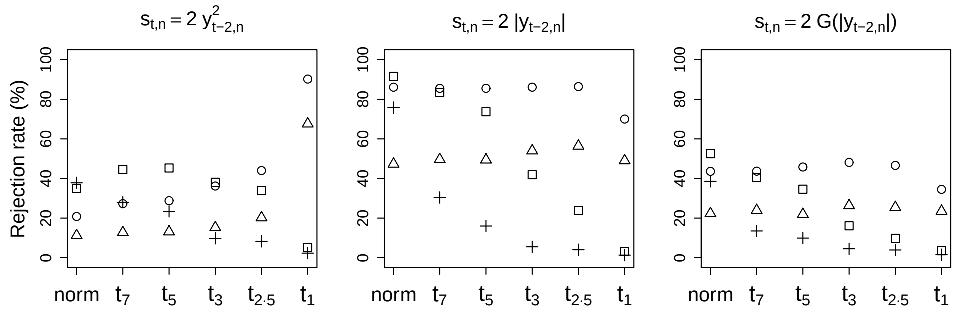

Fig. 1. Power (%) of four goodness-of-fit tests, i.e., $Q(6)$ (circles), $Q_{\text{sgn}}(6)$ (triangles), $Q_{\text{abs}}(6)$ (squares) and $Q_{\text{sqr}}(6)$ (pluses), for six different innovation distributions and three different departure $s_{t,n}$.

i.e., $s_{t,n} = G(|y_{t-2,n}|)$, $|y_{t-2,n}|$ and $y^2_{t-2,n}$, for $\{\varepsilon_t\}$ following the normal distribution with mean zero and Student's t_7, t_5, t_3, $t_{2\cdot5}$ and t_1 distributions, standardized such that median$(|\varepsilon_t|) = 1$. We assume that the model is estimated by the least absolute deviations method, and approximate the quantities in Υ_Ψ and Σ_Ψ by sample averages based on a generated sequence $\{y_1, \ldots, y_n\}$ with $n = 100{,}000$. We set $M = 6$, and compare the three transformations: $\Psi(x) = G(x)$, x and x^2. Some values are left blank in Table 1 owing to violations of the moment conditions of ε_t. It can be seen that $\Psi = G$ dominates all of the transformations when $E(\varepsilon_t^4) = \infty$, and even for moderate-tailed or Gaussian innovations when the departure is $s_{t,n} = G(|y_{t-2,n}|)$ and $|y_{t-2,n}|$. The desirable performance of $\Psi = G$ is also found in other situations; see the Supplementary Material. Moreover, consistent with these results, our first simulation experiment in §5 demonstrates that the proposed test, $Q(M)$, performs favourably compared with existing tests.

5. Simulation experiments

This section presents the results of three simulation experiments carried out to (i) assess the empirical power of the proposed test, $Q(M)$, (ii) evaluate the performance of the automatic method of selecting M, and (iii) verify the asymptotic equivalence in Proposition 1. The least absolute deviations estimator (Peng & Yao, 2003) is employed throughout.

In the first experiment, we compare the power of the proposed test, $Q(M)$, with three existing goodness-of-fit tests: the sign-based test of Chen & Zhu (2015), $Q_{\text{sgn}}(M)$; the test based on absolute residuals in Li & Li (2005), $Q_{\text{abs}}(M)$; and that based on squared residuals in Li (2004), $Q_{\text{sqr}}(M)$. For comparison, M is fixed to six. We generate 1000 replications from

$$y_{t,n} = \varepsilon_t h_{t,n}^{1/2}, \qquad h_{t,n} = 0\cdot01 + 0\cdot03 y^2_{t-1,n} + 0\cdot2 h_{t-1,n} + n^{-1/2} s_{t,n}, \tag{9}$$

where $\{\varepsilon_t\}$ are independent and identically distributed, following the normal distribution with mean zero or Student's t_7, t_5, t_3, $t_{2\cdot5}$ or t_1 distributions, and are standardized such that median$(|\varepsilon_t|) = 1$. We consider departure $s_{t,n} = 2y^2_{t-2,n}$, $2|y_{t-2,n}|$ and $2G(|y_{t-2,n}|)$, and the sample size is $n = 1000$. The density function of ε_t is estimated by the kernel density method with the Gaussian kernel and its rule-of-thumb bandwidth, $h = 0\cdot9 n^{-1/5} \min(\widehat{\sigma}, \widehat{R}/1\cdot34)$, where $\widehat{\sigma}$ and $\widehat{R}$ are the sample standard deviation and interquartile of the residuals $\{\widehat{\varepsilon}_t\}$, respectively; see Silverman (1986). Figure 1 presents the power of the four tests. When the tails of ε_t become heavier, the power of $Q_{\text{abs}}(M)$ and $Q_{\text{sqr}}(M)$ drops dramatically. Although both $Q(M)$ and $Q_{\text{sgn}}(M)$ maintain their power, $Q(M)$ is clearly more powerful, suggesting that the degree of information loss from its transformation of absolute residuals is relatively small. Finally, al-

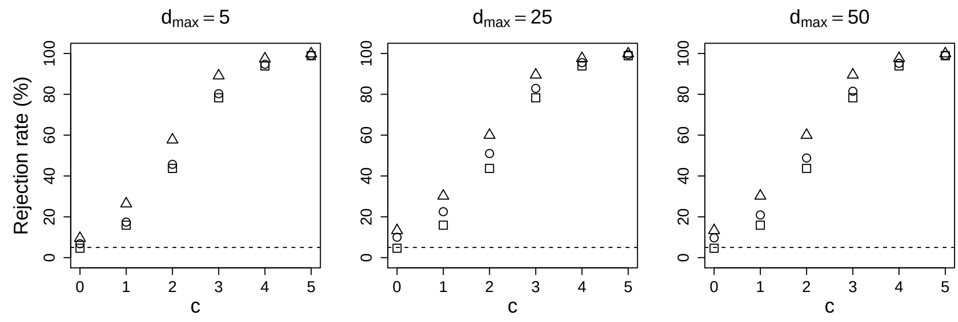

Fig. 2. Rejection rates (%) of the automatic test, $Q(\widetilde{M})$, for $d_{\text{max}} = 5$, 25 and 50, and three selection rules: Bayesian-information-criterion-type (squares), Akaike-information-criterion-type (triangles) and mixed method (circles). The horizontal lines indicate the 5% nominal level.

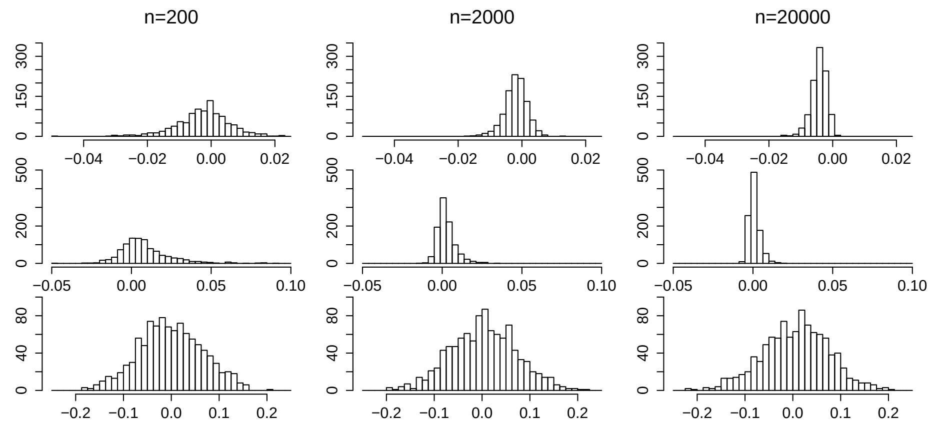

Fig. 3. Histograms of $n^{1/2}(\widehat{\gamma}_1 - \widehat{\gamma}_1^G)$ (top), $n^{1/2}(\widehat{\gamma}_1^{G_n} - \widehat{\gamma}_1^G)$ (middle) and $n^{1/2}\widehat{\gamma}_1^G$ (bottom) under H_0 for sample sizes $n = 200$, 2000 and 20000.

though $Q_{\text{abs}}(M)$ performs well when $s_{t,n} = 2y_{t-2,n}^2$ and ε_t is lighter-tailed, the proposed $Q(M)$ is almost always the most powerful test for the other two types of departure, even when ε_t is moderate-tailed.

The second experiment evaluates the performance of the proposed order selection method. We compare three different methods: (i) the Bayesian-information-criterion-type method in (5), where the penalty term is $M \log n$; (ii) the Akaike-information-criterion-type method, for which the penalty term in (5) is replaced by $2M$; and (iii) the mixed method, for which the penalty term in (5) is replaced by $2M$ if and only if $n^{1/2} \max(|\widehat{\rho}_1|, \ldots, |\widehat{\rho}_{d_{\max}}|) > (\log n)^{1/2}$. We set $d_{\min} = 1$ and $d_{\max} = 5, 25$ or 50. The data are generated from (9) with $s_{t,n} = cy_{t-2,n}^2$, where $c = 0$ corresponds to the size and $c = 1, \ldots, 5$ to the power. The innovations $\{\varepsilon_t\}$ are Student t_3-distributed; the findings under the other innovation distributions from the previous experiment are similar. All other settings are preserved from the previous experiment. Figure 2 shows that the rejection rates are insensitive to the value of $d_{\max}$, but vary for different selection methods. The size of the Bayesian-information-criterion-based automatic test is close to the nominal rate. Although the power of that test is slightly smaller than that of the Akaike-information-criterion-based test, the latter is severely over-sized. The behaviour of the mixed method falls between that of the other two methods. In addition, when comparing the performance of the Bayesian-information-criterion-based automatic test $Q(\widetilde{M})$ for $c = 2$ in Fig. 2 with the left panel of Fig. 1

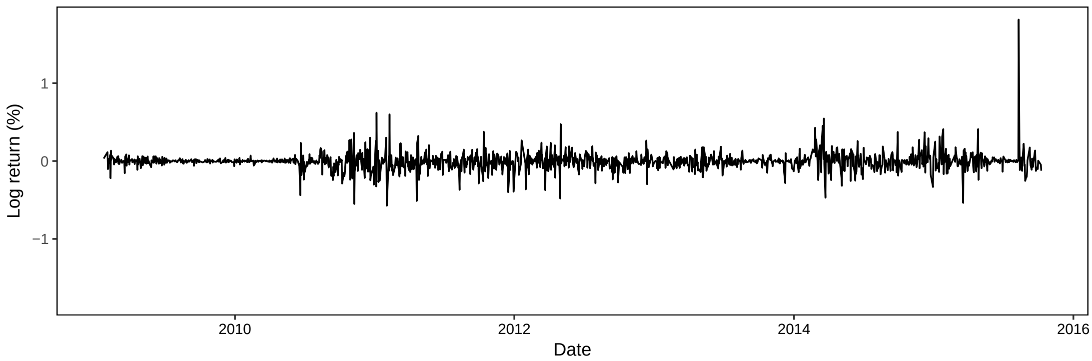

Fig. 4. Daily log returns (%) of the yuan-dollar exchange rates from 23 January 2009 to 9 October 2015.

for $Q(6)$ and Student t_3-distributed innovations, we can see that the automatic test has a power comparable to that with a fixed M. Based on these findings, we recommend using the Bayesian-information-criterion-based method for automatic selection of M.

The third experiment is conducted to verify the asymptotic equivalence of the test $Q_\Psi(M)$ based on transformations $\widehat{G}_n$, G_n and G. We generate 1000 replications from

$$y_t = \varepsilon_t h_t^{1/2}, \qquad h_t = 0{\cdot}01 + 0{\cdot}2y_{t-1}^2 + 0{\cdot}2h_{t-1},$$

where $\{\varepsilon_t\}$ follow the normal distribution with mean zero and $\text{median}(|\varepsilon_t|) = 1$, and the sample sizes are $n = 200$, 2000 and 20000. Figure 3 presents the histograms of $n^{1/2}(\widehat{\gamma}_k - \widehat{\gamma}_k^G)$, $n^{1/2}(\widehat{\gamma}_k^{G_n} - \widehat{\gamma}_k^G)$ and $n^{1/2}\widehat{\gamma}_k^G$ with $k = 1$. It shows that as n increases, the distributions of $n^{1/2}(\widehat{\gamma}_1 - \widehat{\gamma}_1^G)$ and $n^{1/2}(\widehat{\gamma}_1^{G_n} - \widehat{\gamma}_1^G)$ both shrink towards zero, whilst that of $n^{1/2}\widehat{\gamma}_1^G$ maintains the same shape, thereby confirming the asymptotic results in Proposition 1.

Furthermore, three additional simulation studies are provided in the Supplementary Material, wherein we verify the asymptotic distributions of $Q(M)$ under H_0 and H_{1n} and apply the proposed order selection method to all test statistics in the first experiment. In particular, we show that the null distribution of $Q(M)$ is well approximated by the χ^2_M distribution even in small samples and that when n is large, $Q(M)$ converges to a non-central χ^2_M distribution under H_{1n}, although the convergence rate seems slower for heavier-tailed innovation distributions. Finally, the proposed order selection method performs well when applied to other test statistics.

6. AN EMPIRICAL EXAMPLE

This section analyses the daily log returns, in percentage form, of the exchange rate of the Chinese yuan to the United States dollar from 23 January 2009 to 9 October 2015. The sample size is $n = 1520$. Figure 4 shows clear volatility clustering. The sample autocorrelation function lies inside or near the bounds of $\pm 1{\cdot}96/n^{1/2}$ at the first 30 lags, so a pure generalized autoregressive conditionally heteroscedastic model is suggested.

We fit four models using the least absolute deviations method: the $\text{GARCH}(1,1)$ model, and the autoregressive conditional heteroscedastic models of orders $p = 6$, 7 and 8, defined as $y_t = \varepsilon_t h_t^{1/2}$, $h_t = \omega_0 + \sum_{i=1}^p \alpha_{0i} y_{t-i}^2$ and denoted by $\text{ARCH}(p)$. The estimated coefficients with standard errors are listed in Table 2. Before conducting goodness-of-fit tests, we first plot the sample autocorrelation functions of the absolute residuals transformed by $\Psi(x) = \widehat{G}_n(x), \text{sgn}(x-1), x$ and x^2, with their corresponding 95% confidence bands. Figure 5 shows that residual autocorre-

Table 2. *Estimation results* $(\times 10^2)$ *for all fitted models with standard errors*

	ARCH(6)		ARCH(7)		ARCH(8)		GARCH	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
ω	0·01	2E-03	0·01	2E-03	0·01	2E-03	2E-03	6E-04
α_1	19·13	3·11	18·68	3·09	17·25	2·99	11·50	1·70
α_2	9·30	2·24	9·19	2·22	8·65	2·20		
α_3	5·78	1·86	4·94	1·73	5·38	1·81		
α_4	3·59	1·50	2·60	1·37	2·56	1·37		
α_5	0·04	0·68	4E-06	0·76	7E-05	0·79		
α_6	5·02	1·39	5·03	1·44	4·31	1·40		
α_7			1·10	0·72	0·70	0·71		
α_8					1·59	0·83		
β_1							69·34	3·00

SE, standard error; small numbers are written in standard form: e.g., 2E-03 refers to 2×10^{-3}.

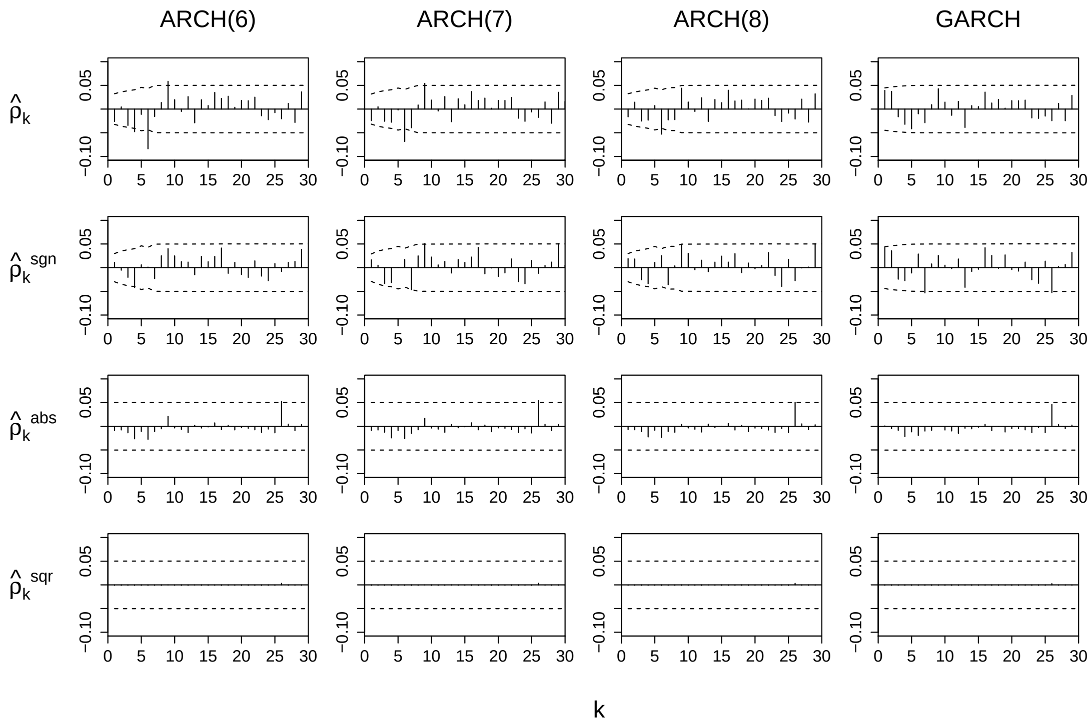

Fig. 5. Sample autocorrelation functions of absolute residuals transformed by $\Psi = \widehat{G}_n, \mathrm{sgn}(x-1), x$ and x^2 (from top to bottom) for four fitted models, with corresponding 95% confidence bands.

lation function $\widehat{\rho}_k$ falls noticeably outside the confidence band at lag $k = 6$ for all fitted ARCH(p) models, yet falls inside the band at all lags for the fitted GARCH$(1,1)$ model. In contrast, $\widehat{\rho}_k^{\mathrm{sgn}}$, $\widehat{\rho}_k^{\mathrm{abs}}$ and $\widehat{\rho}_k^{\mathrm{sqr}}$ all either lie inside the confidence bands or stand out only slightly. The last two sample autocorrelation functions in particular are very small at almost all lags.

We next compare the performance of the proposed test, based on $Q(M)$, with those of the tests based on $Q_{\mathrm{sgn}}(M)$, $Q_{\mathrm{abs}}(M)$ and $Q_{\mathrm{sqr}}(M)$. For each test, we employ the Bayesian-information-criterion-type method in (5) to select M, and use $d_{\min} = 6$ because $\widehat{\rho}_k$ first falls outside its confidence band at $k = 6$ in Fig. 5; $d_{\max}$ is set to 30. Table 3 presents the p-values of these tests with automatically selected orders $\widetilde{M}$, indicated by superscript A. We also report the p-values for the tests with $M = 9$, because $\widehat{\rho}_k$ for both of the fitted ARCH(6) and ARCH(7) models is significant at lag nine. The p-values of Q_{abs} and Q_{sqr} are all close to or even equal to unity. Although Q_{sgn} has smaller p-values, it fails to reject any of the fitted ARCH(p) models at the 5% significance level. By contrast, the inadequacy of the fitted ARCH(6) and ARCH(7) models is successfully detected by our proposed test for both $M = \widetilde{M}$ and 9, which indicates that $\Psi = \widehat{G}_n$ achieves a better performance in detecting possible autocorrelation structures.

Table 3. *p-values of four goodness-of-fit tests with selected order $\widetilde{M}$ or $M = 9$*

	Q^A	Q^A_{sgn}	Q^A_{abs}	Q^A_{sqr}	$Q(9)$	$Q_{\text{sgn}}(9)$	$Q_{\text{abs}}(9)$	$Q_{\text{sqr}}(9)$
ARCH(6)	0·0014	0·4130	0·8210	1·0000	0·0015	0·3349	0·9242	1·0000
ARCH(7)	0·0185	0·2790	0·8666	1·0000	0·0125	0·0787	0·9483	1·0000
ARCH(8)	0·0904	0·1872	0·9139	1·0000	0·0981	0·1187	0·9805	1·0000
GARCH	0·1329	0·1367	0·9474	1·0000	0·1272	0·1034	0·9925	1·0000

Finally, we evaluate the tail-heaviness of ε_t. The Pickands and Hill estimates of the tail index are calculated for the squared residuals of the fitted GARCH$(1, 1)$ model. The implication is $E(\varepsilon_t^2) < \infty$ and $E(\varepsilon_t^4) = \infty$; see the Supplementary Material and Resnick (2007) for details. We also adopt the strict stationarity tests in Francq & Zakoïan (2012) based on least absolute deviations, and confirm the stationarity of the observed log returns at the 1% significance level. Moreover, $\widehat{\alpha}_1\widehat{\sigma}^2 + \widehat{\beta}_1 = 3{\cdot}5$, which is much greater than one, implying that the observed sequence has an infinite second-order moment. This phenomenon, together with the heavy-tailedness of ε_t, may have led to the considerable volatility exhibited in Fig. 4.

7. CONCLUSION AND DISCUSSION

For a time series model, let $\{\varepsilon_t\}$ and $\{\widehat{\varepsilon}_t\}$ denote the innovations and corresponding residuals, respectively. In constructing a goodness-of-fit test, the sample autocorrelation function of $\{\widehat{\varepsilon}_t\}$, $\{|\widehat{\varepsilon}_t|\}$ or $\{\widehat{\varepsilon}_t^2\}$ is usually employed. However, to ensure the existence of the autocorrelation function of $\{\varepsilon_t\}$, $\{|\varepsilon_t|\}$ or $\{\varepsilon_t^2\}$, a finite second- or even fourth-order moment is unavoidable. The essence of our idea in this paper is to transform the residuals before calculating the conventional autocorrelation function. Such transformation is simple to perform, and yet leads to a rich class of tests through various transformations. When the absolute residuals are transformed by their corresponding empirical distribution function, no moment condition for ε_t is required, and the resultant goodness-of-fit test is applicable to arbitrarily heavy-tailed innovations.

There is an extensive body of literature on time series models with innovations of infinite variance, such as the infinite variance autoregressive (Davis & Resnick, 1986; Ling, 2005) and autoregressive moving-average (Zhu & Ling, 2015) models. The corresponding estimators may not even be $\sqrt{n}$-consistent. To the best of our knowledge, there is no goodness-of-fit test that is well-suited to such situations yet, and we thus propose that the method in this paper be adopted to resolve this problem, which we leave for future research.

ACKNOWLEDGEMENT

We thank the editor, associate editor and two anonymous referees for their invaluable comments, which led to the paper being substantially improved, and acknowledge the Hong Kong Research Grants Council for partial support.

SUPPLEMENTARY MATERIAL

Supplementary material available at *Biometrika* online includes additional results on the noncentrality parameter, additional simulation studies, tail index estimation in the empirical example, and all technical proofs.

APPENDIX

Three important lemmas

Lemmas A1 and A2 below can be used to derive asymptotic distributions of weighted residual empirical processes for generalized autoregressive conditionally heteroscedastic models, and hence are of independent interest. Lemma A3 provides a Hájek projection for the Spearman rank autocorrelation coefficient.

LEMMA A1. *Suppose that H_0 and Assumptions 1 and 3 hold with $n^{1/2}(\widehat{\theta}_n - \theta_0) = O_p(1)$. If $\{w_t\}$ is a strictly stationary and ergodic process $\{w_t\}$ with $0 \le w_t \le 1$ and $w_t \in \mathcal{F}_{t-1}$, then*

$$\sup_{0 \le x < \infty} \left| n^{-1/2} \sum_{t=1}^{n} w_t \left\{ I(|\widehat{\varepsilon}_t| \le x) - I(|\varepsilon_t| \le x) \right\} - 0{\cdot}5 x g(x) d_w^{\mathrm{T}} n^{1/2}(\widehat{\theta}_n - \theta_0) \right| = o_p(1),$$

where $d_w = E\{w_t h_t^{-1} \partial h_t(\theta_0)/\partial\theta\}$.

LEMMA A2. *Suppose that H_0 and Assumptions 1 and 3 hold with $n^{1/2}(\widehat{\theta}_n - \theta_0) = O_p(1)$. If $\{w_t\}$ is a strictly stationary and ergodic process $\{w_t\}$ with $0 \le w_t \le 1$, and w_t is independent of $\mathcal{F}_t$, then*

$$\sup_{0 \le x < \infty} \left| n^{-1/2} \sum_{t=1}^{n} w_t \left\{ I(|\widehat{\varepsilon}_t| \le x) - I(|\varepsilon_t| \le x) \right\} - E(w_t) x g(x) d_0^{*\mathrm{T}} n^{1/2}(\widehat{\theta}_n - \theta_0) \right| = o_p(1),$$

where $d_0^* = 0{\cdot}5 E\{h_t^{-1} \partial h_t(\theta_0)/\partial\theta\}$.

LEMMA A3. *Let $X_1, \ldots, X_n$ be a sample of independent observations with distribution function $F(x)$ and empirical distribution function $F_n(x) = n^{-1}\sum_{t=1}^{n} I(X_t \le x)$, $-\infty < x < \infty$. It then holds that, for any positive integer k,*

$$n^{-1/2} \sum_{t=k+1}^{n} \{F_n(X_t)F_n(X_{t-k}) - F(X_t)F(X_{t-k})\} = -n^{-1/2} \sum_{t=k+1}^{n} \{F(X_t) - 0{\cdot}5\} + o_p(1).$$

REFERENCES

ANDREOU, E. & WERKER, B. J. (2012). An alternative asymptotic analysis of residual-based statistics. *The Review of Economics and Statistics* **94**, 88–99.

ANDREOU, E. & WERKER, B. J. (2015). Residual-based rank specification tests for AR-GARCH type models. *Journal of Econometrics* **185**, 305–331.

BARTELS, R. (1982). The rank version of von Neumann's ratio test for randomness. *Journal of the American Statistical Association* **77**, 40–46.

BASRAK, B., DAVIS, R. A. & MIKOSCH, T. (2002). Regular variation of GARCH processes. *Stochastic Processes and their Applications* **99**, 95–115.

BERKES, I. & HORVÁTH, L. (2004). The efficiency of the estimators of the parameters in GARCH processes. *The Annals of Statistics* **32**, 633–655.

BERKES, I., HORVÁTH, L. & KOKOSZKA, P. (2003). GARCH processes: structure and estimation. *Bernoulli* **9**, 201–227.

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

BOUGEROL, P. & PICARD, N. (1992). Stationarity of GARCH processes and of some nonnegative time series. *Journal of Econometrics* **52**, 115–127.

CHEN, M. & ZHU, K. (2015). Sign-based portmanteau test for ARCH-type models with heavy-tailed innovations. *Journal of Econometrics* **189**, 313–320.

DAVIS, R. A. & MIKOSCH, T. (1998). The sample autocorrelations of heavy-tailed processes with applications to ARCH. *The Annals of Statistics* **26**, 2049–2080.

DAVIS, R. A. & RESNICK, S. I. (1986). Limit theory for the sample covariance and correlation function of moving averages. *The Annals of Statistics* **14**, 533–558.

DROST, F. C. & KLAASSEN, C. A. (1997). Efficient estimation in semiparametric GARCH models. *Journal of Econometrics* **81**, 193–221.

DUFOUR, J.-M. & ROY, R. (1985). Some robust exact results on sample autocorrelations and tests of randomness. *Journal of Econometrics* **29**, 257–273.

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

FAN, J. & YAO, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. New York: Springer.

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

FRANCQ, C. & ZAKOÏAN, J.-M. (2010). *GARCH Models: Structure, Statistical Inference and Financial Applications*. Chichester: John Wiley & Sons.

FRANCQ, C. & ZAKOÏAN, J.-M. (2012). Strict stationarity testing and estimation of explosive and stationary generalized autoregressive conditional heteroscedasticity models. *Econometrica* **80**, 821–861.

GUO, S., BOX, J. L. & ZHANG, W. (2017). A dynamic structure for high dimensional covariance matrices and its application in portfolio allocation. *Journal of the American Statistical Association* **112**, 235–253.

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

HALLIN, M., INGENBLEEK, J.-F. & PURI, M. L. (1985). Linear serial rank tests for randomness against ARMA alternatives. *The Annals of Statistics* **13**, 1156–1181.

HALLIN, M. & PURI, M. L. (1994). Aligned rank tests for linear models with autocorrelated error terms. *Journal of Multivariate Analysis* **50**, 175–237.

HE, C. & TERÄSVIRTA, T. (1999). Fourth moment structure of the GARCH(p, q) process. *Econometric Theory* **15**, 824–846.

LE CAM, L. & YANG, G. L. (1990). *Asymptotics in Statistics.* New York: Springer.

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

LI, G. & LI, W. K. (2008). Least absolute deviation estimation for fractionally integrated autoregressive moving average time series models with conditional heteroscedasticity. *Biometrika* **95**, 399–414.

LI, W. K. (2004). *Diagnostic Checks in Time Series*. New York: Chapman & Hall/CRC.

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

LING, S. (2005). Self-weighted least absolute deviation estimation for infinite variance autoregressive models. *Journal of the Royal Statistical Society, Series B* **67**, 381–393.

MIKOSCH, T. & STĂRICĂ, C. (2000). Limit theory for the sample autocorrelations and extremes of a GARCH$(1, 1)$ process. *The Annals of Statistics* **28**, 1427–1451.

MITTNIK, S. & PAOLELLA, M. S. (2003). Prediction of financial downside-risk with heavy-tailed conditional distributions. In *Handbook of Heavy Tailed Distributions in Finance*, S. T. Rachev, ed. Amsterdam: Elsevier, pp. 385–404.

NELSON, D. B. & CAO, C. Q. (1992). Inequality constraints in the univariate GARCH model. *Journal of Business & Economic Statistics* **10**, 229–235.

PENG, L. & YAO, Q. (2003). Least absolute deviation estimation for ARCH and GARCH models. *Biometrika* **90**, 967–975.

RESNICK, S. I. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. New York: Springer-Verlag.

SILVERMAN, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.

VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. New York: Cambridge University Press.

WALD, A. & WOLFOWITZ, J. (1943). An exact test for randomness in the non-parametric case based on serial correlation. *Annals of Mathematical Statistics* **14**, 378–388.

ZHU, K. & LI, W. K. (2015). A new Pearson-type QMLE for conditionally heteroskedastic models. *Journal of Business and Economic Statistics* **33**, 552–565.

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

ZIVOT, E. (2009). Practical issues in the analysis of univariate GARCH Models. In *Handbook of Financial Time Series*, T. Mikosch, J.-P. Kreiß, R. A. Davis & T. G. Andersen, eds. New York: Springer, pp. 113–155.