Hausman tests for the error distribution in conditionally heteroskedastic models

Zhu, Ke

Institute of Applied Mathematics, Chinese Academy of Sciences

30 September 2015

Hausman tests for the error distribution in conditionally heteroskedastic models

BY KE ZHU

Institute of Applied Mathematics, Chinese Academy of Sciences, Haidian District, Zhongguancun, Beijing, China

kzhu@amss.ac.cn

ABSTRACT

This paper proposes some novel Hausman tests to examine the error distribution in conditionally heteroskedastic models. Unlike the existing tests, all Hausman tests are easy-to-implement with the limiting null distribution of χ^2, and moreover, they are consistent and able to detect the local alternative of order $n^{-1/2}$. The scope of the Hausman test covers all Generalized error distributions and Student's t distributions. The performance of each Hausman test is assessed by simulated and real data sets.

Some key words: Conditionally heteroskedastic model; Consistent test; GARCH model; Goodness-of-fit test; Hausman test; Nonlinear time series.

1. INTRODUCTION

Assume that $\{y_t : t = 0, \pm 1, \pm 2, \cdots\}$ is generated by a conditionally heteroskedastic model:

$$y_t = \sigma_t \eta_t \quad \text{and} \quad \sigma_t = \sigma(y_{t-1}, y_{t-2}, \cdots; \theta_0), \tag{1.1}$$

where η_t being independent of $\{y_j; j < t\}$ is a sequence of i.i.d. random variables, the parameter space $\Theta \in \mathcal{R}^m$ is compact, the true value θ_0 is an interior point in Θ, and $\sigma : \mathcal{R}^\infty \times \Theta \to (0, \infty)$. Many existing models, such as (G)ARCH model in Engle (1982) and Bollerslev (1986), asymmetric power GARCH model in Ding, Granger, and Engle (1993), and asymmetric log-GARCH model in Geweke (1986) to name but a few, are embedded into model (1.1); see, e.g., Bollerslev, Chou, and Kroner (1992) and Francq and Zakoïan (2010). In many applications, the knowledge of the distribution of η_t is crucial for determining the optimal prediction of y_t [e.g., Christoffersen and Diebold (1997)], the value at risk of y_t [e.g., Engle (2004)], and the pricing of financial derivatives written on y_t [e.g., Zhu and Ling (2015)]. All of these are widely used to guide our decisions in practice. Thus, it is necessary to testing the goodness-of-fit hypothesis:

$$H_0 : \eta_t \sim F_0 \quad \text{v.s.} \quad H_1 : \eta_t \not\sim F_0, \tag{1.2}$$

where F_0 is a known distribution function.

Goodness-of-fit testing for the distribution of observable or non-observable random variables has attracted a considerable interest in the literature; see, e.g., D'Agostino and Stephens (1986) and the references therein. The often used technique is based on the empirical process, and this leads to the so-called Kolmogorov-Smirnov (KS) test statistic in general. For the observable i.i.d. random variables, the limiting distribution of KS test statistic is asymptotically distribution free (ADF); and for the unobservable i.i.d. errors in AR or MA models, Boldin (1982, 1989) and Koul (2002) have shown that this ADF property still holds based on the residual sequence. However, when the unobservable i.i.d. errors like $\{\eta_t\}$ in model (1.1) are from a special non-linear model, the ADF property of KS test statistic does not hold any more; see, e.g., Koul (1996) for threshold AR models, Horváth, Kokoszka, and Teyssière (2001) for ARCH models, Berkes and Horváth (2001) for GARCH models, and many others. Particularly, based on the bootstrap-assisted test, this unsatisfactory phenomenon has been verified by Monte Carlo studies in Horváth, Kokoszka, and Teyssière (2004) and Klar, Lindner, and Meintanis (2012) for (G)ARCH models. To retain the property of ADF, Horváth and Zitikis (2006) have constructed a nonparametric Cramér-von Mises type goodness-of-fit test for GARCH models; and meanwhile, Koul and Ling (2006) have proposed a weighted KS test statistic for a class of GARCH and ARMA-GARCH models; however, the former method calls for a good choice of bandwidth used in the kernel-type density estimator of the residual, and the latter method, loosely speaking, is not ADF, since its limiting distribution still relies on F_0.

In this paper, we propose some novel Hausman tests to detect H_0 in spirit of Hausman (1978). The idea to construct the Hausman test is as follows: first, we choose a quasi maximum likelihood estimator (QMLE) $\widetilde{\theta}_n$ of model (1.1); second, we rescale (η_t, H_0) to $(\eta_t^\dagger, H_0^\dagger)$ such that the structure of y_t is unchanged and $\eta_t^\dagger$ satisfies the identification condition of $\widetilde{\theta}_n$; third, we calculate the MLE $\widehat{\theta}_n$ of model (1.1) under $H_0^\dagger$, and formulate the Hausman test by measuring the difference between $\widetilde{\theta}_n$ and $\widehat{\theta}_n$. In the aforementioned procedure, the choice of QMLE is flexible, and we use the generalized QMLE (GQMLE) in Francq and Zakoïan (2013) and the least absolute deviation estimator (LADE) in Peng and Yao (2003) to propose the so-called GQMLE-based and LADE-based Hausman tests, respectively. Under suitable conditions, we show that each Hausman test is ADF with a limiting null distribution of χ^2, and that it is consistent and able to detect the local alternative of order $n^{-1/2}$. Our Hausman testing procedure is easy-to-implement, and its scope covers all Generalized error distributions and Student's t distributions. The performance of this testing procedure is assessed by simulated and real data sets.

This paper is organized as follows. Sections 2 and 3 propose and study the GQMLE-based and LADE-based Hausman test statistics, respectively. Simulation results are reported in Section 4. A real example on S&P 500 stock index is given in Section 5. Concluding remarks are offered in Section 6. All of the proofs are given in Appendix. Throughout the paper, some symbols are conventional. A' is the transpose of matrix A, $|A| = (tr(A'A))'$ is the Euclidean norm of matrix A, $\|A\|_s = (E|A|^s)^{1/s}$ is the L_s-norm ($s \leq 1$) of a random matrix A, $o_p(1)(O_p(1))$ denotes a sequence of random numbers converging to zero (bounded) in probability, $\to_d$ denotes convergence in distribution, $I(\cdot)$ is the indicator function, and $\mathrm{sgn}(\cdot) = I(\cdot > 0) - I(\cdot < 0)$ is the sign function.

2. GQMLE-BASED HAUSMAN TESTS

This section proposes some Hausman tests to detect H_0 in spirit of Hausman (1978). To accomplish it, we need two estimators: a quasi maximum likelihood estimator (QMLE) and a MLE, which are introduced in the following subsection.

2·1. *Preliminary*

Let Θ be a compact space and $\sigma_t(\theta) = \sigma(y_{t-1}, y_{t-2}, \cdots; \theta)$. First, we choose the QMLE as the generalized QMLE (GQMLE) in Francq and Zakoïan (2013) given by

$$\widetilde{\theta}_{n,r} = \begin{cases} \arg\min_{\theta\in\Theta} \sum_{t=1}^{n} \left[\log\{\widetilde{\sigma}_t^r(\theta)\} + \frac{|y_t|^r}{\widetilde{\sigma}_t^r(\theta)}\right], & \text{if } r > 0, \\ \arg\min_{\theta\in\Theta} \sum_{t=1}^{n} \left[\log|y_t| - \log\widetilde{\sigma}_t(\theta)\right]^2, & \text{if } r = 0, \end{cases} \tag{2.1}$$

where $\widetilde{\sigma}_t(\theta) := \sigma(y_{t-1}, y_{t-2}, \cdots, y_1, \widetilde{y}_0, \widetilde{y}_{-1}, \cdots; \theta)$ is calculated based on the observations $\{y_s\}_{s=1}^{n}$ and the arbitrary initial values $\{\widetilde{y}_s\}_{s\leq 0}$. Here, the objective function in (2.1) is written on the assumption that η_t has the density function

$$h(x) = \begin{cases} c|x|^{\lambda-1}\exp\left(-\lambda|x|^r/r\right), & \text{if } r > 0, \\ \sqrt{\lambda/\pi}|2x|^{-1}\exp(-\lambda\log|x|^2), & \text{if } r = 0, \end{cases}$$

where λ and c are two positive normalization constants; see Francq and Zakoïan (2013, p.349). Particularly, when $r = 2$, $c = 1/2$, and $\lambda = 1$, $\widetilde{\theta}_{n,r}$ reduces to the Gaussian QMLE; and when $r = 1$, $c = 1/2$, and $\lambda = 1$, $\widetilde{\theta}_{n,r}$ reduces to the Laplacian QMLE.

As shown in Francq and Zakoïan (2013), the identifiability condition for $\widetilde{\theta}_{n,r}$ is as follows:

Assumption 2.1. $E|\eta_t|^r = 1$ when $r > 0$, and $E\log|\eta_t| = 0$ when $r = 0$.

We now assume that model (1.1) holds under Assumption 2.1, and $\theta_{0,r}$ is the corresponding true parameter, where the subscript r in $\theta_{0,r}$ is involved to indicate the chosen GQMLE method. Note that under H_0, η_t does not satisfy Assumption 2.1 in general. Hence, we have to consider an equivalent rescaling version of H_0, under which the corresponding rescaling innovation satisfies Assumption 2.1. In order to accomplish it, we denote η_t as $\eta_t^{(0)}$ if $\eta_t \sim F_0$, and let $\eta_{t,r}^{\dagger} := \eta_t^{(0)}/\kappa_r$ be the rescaling form of $\eta_t^{(0)}$, where κ_r is the rescaling parameter defined by

$$\kappa_r = \begin{cases} \left[E|\eta_t^{(0)}|^r\right]^{1/r}, & \text{if } r > 0, \\ \exp\left(E \log |\eta_t^{(0)}|\right), & \text{if } r = 0. \end{cases}$$

The following Assumption guarantees that model (1.1) can be re-parametrized so that the structure of y_t is unchanged after this rescaling transformation.

Assumption 2.2. There is a function Π such that, for any $\theta \in \Theta$, $K > 0$, and real sequence $\{x_s\}_{s\geq 1}$, $K\sigma(x_1, x_2, \cdots; \theta) = \sigma(x_1, x_2, \cdots; \theta^{\dagger})$ with $\theta^{\dagger} = \Pi(\theta, K)$.

Assumption 2.2 is quite mild, and it holds for the standard GARCH model and most of its extensions; see, e.g., Francq and Zakoïan (2013, p.353) for a specific illustration. By Assumption 2.2, we know that under H_0,

$$y_t = \eta_t^{(0)}\sigma(y_{t-1}, y_{t-2}, \cdots; \theta_0) = \eta_{t,r}^{\dagger}\left[\sigma(y_{t-1}, y_{t-2}, \cdots; \theta_{0,r})\right],$$

where $\theta_{0,r} = \Pi(\theta_0, \kappa_r)$; and so H_0 in (1.2) is equivalent to its rescaling version:

$$H_{0,r}^{\dagger} : \eta_t \sim F_0(\kappa_r x), \tag{2.2}$$

where $F_0(\kappa_r x)$ is the distribution of $\eta_{t,r}^{\dagger}$. We will consider the rescaling null hypothesis $H_{0,r}^{\dagger}$ instead of H_0 subsequently, since the identifiability condition of the GQMLE in Assumption 2.1 holds under $H_{0,r}^{\dagger}$.

Next, we consider the MLE under $H_{0,r}^{\dagger}$ in (2.2). In this case, the density of η_t is $f_{0,r}^{\dagger}(x) = \kappa_r f_0(\kappa_r x)$ with $f_0(x) = F_0'(x)$, and hence the MLE is

$$\widehat{\theta}_{n,r} := \arg\min_{\theta\in\Theta} \frac{1}{n}\sum_{t=1}^{n}\left[\log\widetilde{\sigma}_t(\theta) - \log f_{0,r}^{\dagger}\left(\frac{y_t}{\widetilde{\sigma}_t(\theta)}\right)\right]. \tag{2.3}$$

We are now ready to give three subsections below to study $\widetilde{\theta}_{n,r}$, $\widehat{\theta}_{n,r}$, and the related Hausman test based on $\widetilde{\theta}_{n,r}$ and $\widehat{\theta}_{n,r}$, respectively.

2·2. *Technical conditions for the GQMLE*

Assume that $\theta_{0,r}$ is an interior point in Θ. We give four assumptions for the strong consistency and asymptotic normality of $\widetilde{\theta}_{n,r}$.

Assumption 2.3. y_t is strictly stationary and ergodic.

Assumption 2.4. (i) Almost surely (a.s.), $\sigma_t(\theta) \in (\underline{\omega}, \infty]$ for some $\underline{\omega} > 0$ and any $\theta \in \Theta$; (ii) $\sigma_t(\theta_{0,r})/\sigma_t(\theta) = 1$ a.s. if and only if $\theta = \theta_{0,r}$; (iii) $\sigma_t(\theta)$ has continuous second-order derivatives with respective to θ (a.s.); (iv) if $x'(\partial \sigma_t^2(\theta)/\partial \theta_i)_{i=1,\cdots,m} = 0$ (a.s.) for any $x \in \mathcal{R}^m$, then $x = 0$.

Assumption 2.5. There exist constants $C_0 > 0$ and $\rho \in (0,1)$, and a neighborhood $V(\theta_{0,r})$ of $\theta_{0,r}$ such that

$$\sup_{\theta \in \Theta} |\Delta_t(\theta)| \leq C_0 \rho^t \text{ and } \sup_{\theta \in V(\theta_{0,r})} \left\| \frac{\partial \Delta_t(\theta)}{\partial \theta} \right\| \leq C_0 \rho^t \quad (a.s.),$$

where $\Delta_t(\theta) = \widetilde{\sigma}_t(\theta) - \sigma_t(\theta)$.

Assumption 2.6. (i) $E|\eta_t|^{2r} < \infty$; (ii) $E|y_t|^{2\delta_0} < \infty$ for some $\delta_0 > 0$; (iii) the following variables have finite expectation:

$$\sup_{\theta \in V(\theta_{0,r})} \left\| \frac{1}{\sigma_t(\theta)} \frac{\partial \sigma_t(\theta)}{\partial \theta} \right\|^4, \quad \sup_{\theta \in V(\theta_{0,r})} \left\| \frac{1}{\sigma_t(\theta)} \frac{\partial^2 \sigma_t(\theta)}{\partial \theta \partial \theta'} \right\|^2, \quad \sup_{\theta \in V(\theta_{0,r})} \left| \frac{\sigma_t(\theta_{0,r})}{\sigma_t(\theta)} \right|^{2r}.$$

Assumptions 2.3-2.6 are taken from Francq and Zakoïan (2013). Assumption 2.3 is a basic set-up for time series models. Assumption 2.4 exhibits some conditions for the volatility function $\sigma_t(\theta)$, among which conditions (i) and (iii) hold for most of heteroskedastic models, condition (ii) is to prove the strong consistency of $\widetilde{\theta}_{n,r}$, and condition (iv) is to guarantee the invertibility of the asymptotic variance of $\widetilde{\theta}_{n,r}$. Assumption 2.5 provides the sufficient condition to make the initial values $\{\widetilde{y}_s\}_{s\leq 0}$ ignorable. Assumption 2.6 lists some sufficient technical conditions for the proofs. Particularly, Assumptions 2.4, 2.5 and 2.6(iii) have been verified for standard GARCH model and many extensions; see, e.g., Ling (2007), Hamadeh, and Zakoïan (2011), and Francq, Wintenberger, and Zakoïan (2013). Under Assumptions 2.1 and 2.3-2.6, Theorem 1 and Proposition 2 in Francq and Zakoïan (2013) have shown that $\widetilde{\theta}_{n,r}$ is strongly consistent to $\theta_{0,r}$ and asymptotically normal.

2·3. *Technical conditions for the MLE*

We need the following assumption to guarantee the weak convergence of $\widehat{\theta}_{n,r}$:

Assumption 2.7. There exists a unique interior point $\theta_{*,r} \in \Theta$ such that $\widehat{\theta}_{n,r} - \theta_{*,r} = o_p(1)$.

In general, $\theta_{*,r} \neq \theta_{0,r}$; but we can have $\theta_{*,r} = \theta_{0,r}$ under $H^{\dagger}_{0,r}$, if Assumption 2.8 below holds.

Assumption 2.8. (i) $f^{\dagger}_{0,r}(x)$ is twice differentiable with $|k_{i,r}(x)| \leq C_1(1+|x|^{\delta_1})(i=1,2)$ for all $x \in \mathcal{R}^*$ and some constants $C_1 > 0, \delta_1 \in \mathcal{R}$, where $\mathcal{R}^* = \mathcal{R} \setminus \{0\}$,

$$k_{1,r}(x) = \frac{x}{f^{\dagger}_{0,r}(x)} \frac{\partial f^{\dagger}_{0,r}(x)}{\partial x} \quad \text{and} \quad k_{2,r}(x) = x^2 \frac{\partial}{\partial x}\left[\frac{1}{f^{\dagger}_{0,r}(x)} \frac{\partial f^{\dagger}_{0,r}(x)}{\partial x}\right];$$

(ii) $E|\eta_t|^{2\delta_1} < \infty$.

Assumption 2.8(i) is a mild condition, and it holds when $f_0(x)$ is the density of Generalized error distribution, Student t_ν distribution, or more generally any distribution having the density

$$h(x) = K_1|x|^{\lambda_0}\exp(K_2|x|^{\lambda_1}) \quad \text{for any } \lambda_0, \lambda_1 \in \mathcal{R}, \tag{2.4}$$

where K_1 and K_2 are two normalizing constants. Particularly, Assumption 2.8(ii) holds with $\delta_1 = \lambda_1$ under (2.4).

Based on Assumptions 2.3-2.6, 2.8 and Assumption 2.9 below, Francq and Zakoïan (2013) have showed that $\widehat{\theta}_{n,r}$ is consistent to $\theta_{0,r}$ (i.e., $\theta_{*,r} = \theta_{0,r}$) and asymptotically normal.

Assumption 2.9. $E[g_r(\eta_t,\sigma)] < E[g_r(\eta_t,1)], \quad \forall\sigma > 0 \quad \text{and} \quad \sigma \neq 1, \quad \text{where} \quad g_r(x,\sigma) = \log\{\frac{1}{\sigma} f_{0,r}^{\dagger}(\frac{x}{\sigma})\}$.

Under $H_{0,r}^{\dagger}$, the true density of η_t is exactly $f_{0,r}^{\dagger}(\cdot)$. In this case, Assumption 2.9 holds directly by Jansen's inequality, and hence $\theta_{*,r} = \theta_{0,r}$. In general, Assumption 2.9 entails a moment condition on η_t, which shall be different from the moment condition $E|\eta_t|^r = 1$ in Assumption 2.1; and then this implies that $\theta_{*,r} \neq \theta_{0,r}$. To see it clearly, we give two illustrating examples below, and for more discussions on Assumption 2.9, we refer to Berkes and Horváth (2004) and Francq and Zakoïan (2013).

Example 2.1. Suppose that

$$H_0 : \eta_t \sim \text{Generalized error distribution, i.e.,}$$
$$f_0(x) = \frac{w}{2u\Gamma(1/w)} \exp\left[-\left(\frac{|x|}{u}\right)^w\right] \text{ for } u, w > 0,$$

where $\Gamma(\cdot)$ is the gamma function. In this case, we can easily show that if $\kappa_r \in (0,\infty)$, Assumption 2.9 is equivalent to the moment condition

$$E|\eta_t|^w = \frac{u^w}{w\kappa_r^w}, \tag{2.5}$$

which is the identification condition for $\widehat{\theta}_{n,r}$. For instance, consider two important special cases of H_0:

Case 1: $\eta_t \sim \mathrm{N}(0,1)$ [i.e., $u = \sqrt{2}$ and $w = 2$];

Case 2: $\eta_t \sim \mathrm{Laplace}(0,1)$ [i.e., $u = 1$ and $w = 1$].

In Case 1, condition (2.5) becomes $E|\eta_t|^2 = (E|\eta_t^{(0)}|^r)^{-2/r}$, where $\eta_t^{(0)} \sim \mathrm{N}(0,1)$; moreover, if $r = 2$, condition (2.5) and Assumption 2.1 coincide, and hence $\theta_{*,2} = \theta_{0,2}$. In Case 2, condition (2.5) becomes $E|\eta_t| = (E|\eta_t^{(0)}|^r)^{-1/r}$, where $\eta_t^{(0)} \sim \mathrm{Laplace}(0,1)$; moreover, if $r = 1$, condition (2.5) and Assumption 2.1 coincide, and hence $\theta_{*,1} = \theta_{0,1}$.

Example 2.2. Suppose that

$$H_0 : \eta_t \sim \text{Student's t distribution, i.e.,}$$
$$f_0(x) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\nu\pi}\Gamma(\nu/2)}\left(1+\frac{x^2}{\nu}\right)^{-(1+\nu)/2} \quad \text{for } \nu > 0.$$

In this case, we can easily show that if $\kappa_r \in (0, \infty)$, Assumption 2.9 is equivalent to the moment condition

$$E\left[\frac{1}{\kappa_r^2\eta_t^2+\nu}\right] = \frac{1}{1+\nu}, \tag{2.6}$$

which is the identification condition for $\widehat{\theta}_{n,r}$. Particularly, when $\kappa_r = 1$, condition (2.6) is the identification condition for $\widehat{\theta}_{n,r}$ based on $\eta_t \sim t_v$.

2·4. *Asymptotic theory of the Hausman test*

In this subsection, we propose the GQMLE-based Hausman test by measuring the difference between $\widetilde{\theta}_{n,r}$ and $\widehat{\theta}_{n,r}$. To accomplish it, we need the following theorem:

THEOREM 2.1. *Suppose that (i) Assumptions* 2.1*,* 2.3-2.6*, and* 2.8 *hold; and (ii)* $\tau_r \neq 0$ *and* $E[k_{2,r}(\eta_t)] \neq 1$*. Then, under* $H_{0,r}^{\dagger}$*, we have*

$$\sqrt{n}\left(\widetilde{\theta}_{n,r} - \widehat{\theta}_{n,r}\right) \to_d N(0, \tau_r\mathcal{J}_r^{-1}) \quad \textit{as } n \to \infty,$$

where

$$\tau_r = \begin{cases} E\left[\frac{|\eta_t|^r-1}{r} + \frac{1+k_{1,r}(\eta_t)}{1-E[k_{2,r}(\eta_t)]}\right]^2, & \textit{if } r > 0, \\ E\left[\log|\eta_t| + \frac{1+k_{1,r}(\eta_t)}{1-E[k_{2,r}(\eta_t)]}\right]^2, & \textit{if } r = 0, \end{cases} \quad \textit{and } \mathcal{J}_r = E\left[\frac{1}{\sigma_t^2(\theta_{0,r})}\frac{\partial\sigma_t(\theta_{0,r})}{\partial\theta}\frac{\partial\sigma_t(\theta_{0,r})}{\partial\theta'}\right].$$

Remark 2.1. For the null hypothesis H_0 in Example 2.1, we have

$$k_{1,r}(x) = -\frac{w|\kappa_r x|^w}{u^w} \quad \text{and} \quad k_{2,r}(x) = (w-1)k_{1,r}(x).$$

For the null hypothesis H_0 in Example 2.2, we have

$$k_{1,r}(x) = -\frac{(1+\nu)\kappa_r^2 x^2}{\kappa_r^2 x^2 + \nu} \quad \text{and} \quad k_{2,r}(x) = k_{1,r}(x) + \frac{2(1+\nu)\kappa_r^4 x^4}{(\kappa_r^2 x^2 + \nu)^2}.$$

The value of κ_r involved in $k_{1,r}(x)$ and $k_{2,r}(x)$ depends on $F_0(\cdot)$ in H_0 and r, and it can be easily calculated via a numerical integration for a specific pair of $(F_0(\cdot), r)$; see Table 1 in Section 3 below.

Based on Theorem 2.1, our GQMLE-based Hausman test is proposed as follows:

$$\mathcal{H}_{n,r} = n\left(\widetilde{\theta}_{n,r} - \widehat{\theta}_{n,r}\right)' [\widetilde{\tau}_{n,r}^{-1}\widetilde{\mathcal{J}}_{n,r}] \left(\widetilde{\theta}_{n,r} - \widehat{\theta}_{n,r}\right), \tag{2.7}$$

where $\widetilde{\tau}_{n,r}$ and $\widetilde{\mathcal{J}}_{n,r}$ are the sample counterparts of τ_r and $\mathcal{J}_r$, respectively, given by

$$\widetilde{\tau}_{n,r} = \begin{cases} \frac{1}{n}\sum_{t=1}^{n}\left[\frac{|\widetilde{\eta}_{t,r}|^r - 1}{r} + \frac{1+k_{1,r}(\widetilde{\eta}_{t,r})}{1-\widetilde{k}_{n,r}}\right]^2, & \text{if } r > 0, \\ \frac{1}{n}\sum_{t=1}^{n}\left[\log|\widetilde{\eta}_{t,r}| + \frac{1+k_{1,r}(\widetilde{\eta}_{t,r})}{1-\widetilde{k}_{n,r}}\right]^2, & \text{if } r = 0, \end{cases}$$

$$\text{and } \widetilde{\mathcal{J}}_{n,r} = \frac{1}{n}\sum_{t=1}^{n}\left[\frac{1}{\widetilde{\sigma}_t^2(\widetilde{\theta}_{n,r})}\frac{\partial\widetilde{\sigma}_t(\widetilde{\theta}_{n,r})}{\partial\theta}\frac{\partial\widetilde{\sigma}_t(\widetilde{\theta}_{n,r})}{\partial\theta'}\right]$$

with $\widetilde{\eta}_{t,r} = y_t/\widetilde{\sigma}_t(\widetilde{\theta}_{n,r})$ and $\widetilde{k}_{n,r} = n^{-1}\sum_{t=1}^{n} k_{2,r}(\widetilde{\eta}_{t,r})$. It is not hard to see that $\widetilde{\tau}_{n,r}$ and $\widetilde{\mathcal{J}}_{n,r}$ are consistent estimators of τ_r and $\mathcal{J}_r$, respectively. Note that H_0 and $H_{0,r}^{\dagger}$ are equivalent under Assumption 2.2. Hence, by Theorem 2.1, the following corollary is straightforward:

COROLLARY 2.1. *Suppose that Assumption* 2.2 *and the conditions in Theorem* 2.1 *hold. Then, under H_0, we have*

$$\mathcal{H}_{n,r} \to_d \chi_m^2 \quad \text{as } n \to \infty,$$

*where m is the dimension of θ_0 in model (*1.1*), and χ_s^2 is a chi-square distribution with degree s.*

Remark 2.2. Besides the GQMLE, our Hausman tests could use many other QMLEs of model (1.1); see, e.g., Fan, Qi, and Xiu (2014), Zhu and Li (2015), and references therein. For instance, we will use the least absolute deviation estimator (LADE) as the QMLE to construct the Hausman test in Section 3. The reason that we use the GQMLE or LADE as the QMLE, since we can easily find the re-scaling parameter κ_r in both cases such that the rescaling version of $\eta_t^{(0)}$ satisfies the identification condition of the chosen QMLE method. This is key for our Hausman test, which requires that both QMLE and MLE converge to the same parameter under the rescaling null hypothesis.

To carry out the GQMLE-based Hausman testing procedure, one computes (2.7) and compares it to the upper critical value $c_{m,\alpha}$ for the χ_m^2 distribution at a given significance level α, where $c_{m,\alpha}$ is chosen by $P(\chi_m^2 > c_{m,\alpha}) = \alpha$. If $\mathcal{H}_{n,r} > c_{m,\alpha}$, then we reject H_0; otherwise, we can not reject H_0.

Furthermore, we study the asymptotic power of $\mathcal{H}_{n,r}$ by considering the alternative hypothesis

$$H_{1,r} : \theta_{0,r} - \theta_{*,r} \neq 0,$$

and the local alternative hypothesis

$$H_{1n,r} : \theta_{0,r} - \theta_{*,r} = \frac{\Delta}{\sqrt{n}} \text{ for some constant vector } \Delta \in \mathcal{R}^m.$$

Although there are other ways to construct alternatives in terms of the distribution function $F_0(\cdot)$ directly (see, e.g., Koul and Ling (2006)), the proceeding two alternatives are meaningful, because H_0 and $H_{0,r}^{\dagger}$ are equivalent under Assumption 2.2; and when $H_{0,r}^{\dagger}$ fails, $f_{0,r}^{\dagger}(\cdot)$ is not the true density of η_t, and then $\theta_{*,r}$ tends to deviate from $\theta_{0,r}$ in general.

Below, we make one more technical assumption, which is stronger than Assumption 2.7.

Assumption 2.10. As $n \to \infty$, $\sqrt{n}[(\widetilde{\theta}_{n,r} - \theta_{0,r}) - (\widehat{\theta}_{n,r} - \theta_{*,r})] \to_d \xi_r$ (a distribution).

COROLLARY 2.2. *Suppose that (i) Assumptions* 2.1, 2.3-2.6 *and* 2.10 *hold; and (ii)* $\tau_r \neq 0$ *and* $E[k_{2,r}(\eta_t)] \neq 1$. *Then, under* $H_{1,r}$, *we have* $\lim_{n\to\infty} \mathcal{H}_{n,r} = \infty$; *and under* $H_{1n,r}$, *we have*

$$\mathcal{H}_{n,r} \to_d (\xi_r + \Delta)'(\tau_r^{-1}\mathcal{J}_r)(\xi_r + \Delta) \quad \text{as } n \to \infty,$$

and consequently, $\lim_{|\Delta|\to\infty} \lim_{n\to\infty} \mathcal{H}_{n,r} = \infty$.

The proof of Corollary 2.2 is directly from continuous mapping theorem. From this corollary, we know that $\mathcal{H}_{n,r}$ can consistently detect $H_{1,r}$, and has the nontrivial local power to detect $H_{1n,r}$. Since $\mathcal{H}_{n,r}$ gains the power under alternatives as long as $\theta_{0,r} \neq \theta_{*,r}$, we are of interest to unveil the condition that $\theta_{0,r} = \theta_{*,r}$ for those two illustrating examples in subsection 2.3.

Example 2.1 (*con't*). For each $u, w > 0$, define

$$\mathcal{A}_r = \{\eta_t : \eta_t \text{ satisfies Assumption 2.1 and the moment condition (2.5)}\}.$$

Clearly, if $\eta_t \in \mathcal{A}_r$, we have $\theta_{0,r} = \theta_{*,r}$. Under $H^{\dagger}_{0,r}$, $\eta_t \in \mathcal{A}_r$ for all r, and this guarantees that $\mathcal{H}_{n,r}$ has a desirable size performance. Under alternatives, we should choose a suitable r to avoid $\eta_t \in \mathcal{A}_r$ so that $\mathcal{H}_{n,r}$ is not lack of power. For instance, if we are testing the null hypothesis that $\eta_t \sim \mathrm{N}(0,1)$ (or Laplace$(0,1)$), we should not choose $r = 2$ (or 1). In general, for a well chosen r, the probability that $\eta_t \in \mathcal{A}_r$ under alternatives shall be very low.

Example 2.2 (*con't*). For each $\nu > 0$, define

$$\mathcal{B}_r = \{\eta_t : \eta_t \text{ satisfies Assumption 2.1 and the moment condition (2.6)}\}.$$

Clearly, if $\eta_t \in \mathcal{B}_r$, we have $\theta_{0,r} = \theta_{*,r}$. As the discussion for Example 2.1, we should choose a suitable r to avoid $\eta_t \in \mathcal{B}_r$ under alternatives so that $\mathcal{H}_{n,r}$ is not lack of power.

From Examples 2.1-2.2, we know that for most of the choices of r, we do not face the dilemma that $\theta_{0,r} = \theta_{*,r}$ under the alternative. To further relieve the concern that $\theta_{0,r} = \theta_{*,r}$ for a single chosen r, one can implement $\mathcal{H}_{n,r}$ for different choices of r. Needless to say, the finite performance of $\mathcal{H}_{n,r}$ depends on the choice of r. Simulation studies in Section 4 imply that we should choose a smaller (or larger) r when the tail of η_t is heavier (or lighter).

3. LADE-BASED HAUSMAN TESTS

In this section, we choose the LADE in Peng and Yao (2003) as the QMLE to construct our Hausman test, where the LADE is given by

$$\widetilde{\theta}_{n,l} = \arg\min_{\theta\in\Theta} \sum_{t=1}^{n} \left|\log y_t^2 - \log[\widetilde{\sigma}_t(\theta)]^2\right|, \tag{3.1}$$

and $\widetilde{\sigma}_t(\theta)$ is defined as in (2.1). Here, the subscript l in $\widetilde{\theta}_{n,l}$ is involved to indicate the chosen LADE method. Compared with the GQMLE $\widetilde{\theta}_{n,r}$ in (2.1), the LADE $\widetilde{\theta}_{n,l}$ in (3.1) only needs a finite fractional moment of η_t for its asymptotic normality, and hence it applies for very heavy-tailed η_t; see, e.g., Linton, Pan, and Wang (2010), Francq and Zakoïan (2013), and Chen and Zhu (2015) for more discussions on the LADE.

As shown in Peng and Yao (2003), the identifiability condition for $\widetilde{\theta}_{n,l}$ is as follows:

Assumption 3.1. median$(\eta_t^2) = 1$.

Following the same idea as in Section 2, we assume that model (1.1) holds under Assumption 3.1, and $\theta_{0,l}$ is the corresponding true parameter. Let

$$\kappa_l := \sqrt{\text{median}([\eta_t^{(0)}]^2)}$$

be the rescaling parameter, and $\eta_{t,l}^{\dagger} := \eta_t^{(0)}/\kappa_l$ be the rescaling form of $\eta_t^{(0)}$. As for $H_{0,r}^{\dagger}$ in (2.2), under Assumption 2.2, H_0 in (1.2) is equivalent to its rescaling version:

$$H_{0,l}^{\dagger} : \eta_t \sim F_0(\kappa_l x), \tag{3.2}$$

where $F_0(\kappa_l x)$ is the distribution of $\eta_{t,l}^{\dagger}$, and Assumption 3.1 holds under $H_{0,l}^{\dagger}$.

Next, we consider the MLE under $H_{0,l}^{\dagger}$ in (3.2). In this case, the density of η_t is $f_{0,l}^{\dagger}(x) = \kappa_l f_0(\kappa_l x)$, and hence the MLE is

$$\widehat{\theta}_{n,l} := \arg\min_{\theta\in\Theta} \frac{1}{n} \sum_{t=1}^{n} \left[\log \widetilde{\sigma}_t(\theta) - \log f_{0,l}^{\dagger}\left(\frac{y_t}{\widetilde{\sigma}_t(\theta)}\right)\right]. \tag{3.3}$$

For $i=1,2$, let $k_{i,l}(\cdot)$ be defined in the same way as $k_{i,r}(\cdot)$ in Assumption 2.8, with $f_{0,r}^{\dagger}(\cdot)$ being replaced by $f_{0,l}^{\dagger}(\cdot)$. The following theorem measures the difference between $\widetilde{\theta}_{n,l}$ and $\widehat{\theta}_{n,l}$.

THEOREM 3.1. *Suppose that (i) Assumptions* 2.3-2.5, 2.6*(ii)-(iii) and* 3.1 *hold; (ii)* $E|\eta_t|^{2\delta_2}<\infty$ *for some* $\delta_2>0$*; (iii) Assumption* 2.8 *holds with* $f_{0,r}^{\dagger}(\cdot)$*,* $k_{1,r}(\cdot)$ *and* $k_{2,r}(\cdot)$ *being replaced by* $f_{0,l}^{\dagger}(\cdot)$*,* $k_{1,l}(\cdot)$ *and* $k_{2,l}(\cdot)$*, respectively; (iv) the probability density function* $g(\cdot)$ *of* $\log\eta_t^2$ *satisfying* $g(0)>0$ *and* $\sup_{x\in\mathcal{R}}g(x)<\infty$*, is continuous at zero; (v)* $\tau_l\neq 0$ *and* $E[k_{2,l}(\eta_t)]\neq 1$*. Then, under* $H_{0,l}^{\dagger}$*, we have*

$$\sqrt{n}\left(\widetilde{\theta}_{n,l}-\widehat{\theta}_{n,l}\right)\rightarrow_d N(0,\tau_l\mathcal{J}_l^{-1})\quad \textit{as } n\rightarrow\infty,$$

where

$$\tau_l=E\left[\frac{sgn(\eta_t^2-1)}{4g(0)}+\frac{1+k_{1,l}(\eta_t)}{1-E[k_{2,l}(\eta_t)]}\right]^2 \textit{ and } \mathcal{J}_l=E\left[\frac{1}{\sigma_t^2(\theta_{0,l})}\frac{\partial\sigma_t(\theta_{0,l})}{\partial\theta}\frac{\partial\sigma_t(\theta_{0,l})}{\partial\theta'}\right].$$

Based on Theorem 3.1, our LADE-based Hausman test is proposed as follows:

$$\mathcal{H}_{n,l}=n\left(\widetilde{\theta}_{n,l}-\widehat{\theta}_{n,l}\right)'[\widetilde{\tau}_{n,l}^{-1}\widetilde{\mathcal{J}}_{n,l}]\left(\widetilde{\theta}_{n,l}-\widehat{\theta}_{n,l}\right), \tag{3.4}$$

where $\widetilde{\tau}_{n,l}$ and $\widetilde{\mathcal{J}}_{n,l}$ are the sample counterparts of τ_l and $\mathcal{J}_l$, respectively, given by

$$\begin{aligned}\widetilde{\tau}_{n,l}&=\frac{1}{n}\sum_{t=1}^{n}\left[\frac{sgn(\widetilde{\eta}_{t,l}^2-1)}{4\widetilde{g}_n(0)}+\frac{1+k_{1,l}(\widetilde{\eta}_{t,l})}{1-\widetilde{k}_{n,l}}\right]^2\\ \text{and } \widetilde{\mathcal{J}}_{n,l}&=\frac{1}{n}\sum_{t=1}^{n}\left[\frac{1}{\widetilde{\sigma}_t^2(\widetilde{\theta}_{n,l})}\frac{\partial\widetilde{\sigma}_t(\widetilde{\theta}_{n,l})}{\partial\theta}\frac{\partial\widetilde{\sigma}_t(\widetilde{\theta}_{n,l})}{\partial\theta'}\right]\end{aligned}$$

with $\widetilde{\eta}_{t,l}=y_t/\widetilde{\sigma}_t(\widetilde{\theta}_{n,l})$, $\widetilde{k}_{n,l}=n^{-1}\sum_{t=1}^{n}k_{2,l}(\widetilde{\eta}_{t,l})$, and

$$\widetilde{g}_n(0)=\frac{1}{nb_n}\sum_{t=1}^{n}K\left(\frac{\log\widetilde{\eta}_{t,l}^2}{b_n}\right).$$

Here $K(x)$, with $\int_{-\infty}^{\infty}K(x)dx=1$ and $\int_{-\infty}^{\infty}|x|K(x)<\infty$, is a kernel function and $b_n(>0)$ is the bandwidth with order $O(n^{-1/5})$. It is not hard to see that $\widetilde{\tau}_{n,l}$ and $\widetilde{\mathcal{J}}_{n,l}$ are consistent estimators of τ_l and $\mathcal{J}_l$, respectively. Note that H_0 and $H_{0,l}^{\dagger}$ are equivalent under Assumption 2.2. Hence, by Theorem 3.1, the following corollary is straightforward:

COROLLARY 3.1. *Suppose that Assumption* 2.2 *and the conditions in Theorem* 3.1 *hold. Then, under H_0, we have*

$$\mathcal{H}_{n,l} \to_d \chi^2_m \quad \textit{as } n \to \infty,$$

Remark 3.1. For the null hypothesis H_0 in Examples 2.1 and 2.2, the expressions of $k_{1,l}(x)$ and $k_{2,l}(x)$ are the same as $k_{1,r}(x)$ and $k_{2,r}(x)$ in Remark 2.1, respectively, except that κ_r is replaced by κ_l. Also, the moment conditions in (2.5) and (2.6) remain valid for $\mathcal{H}_{n,l}$, except that κ_r is replaced by κ_l.

Remark 3.2. In order to calculate $\mathcal{H}_{n,r}$ and $\mathcal{H}_{n,l}$, the values of κ_r and κ_l are involved, respectively. For the often used $F_0(\cdot)$ in applications, the values of κ_r and κ_l are reported in Table 1 below, and they are easily calculated via a numerical integration. For other cases of $F_0(\cdot)$, we can obtain the values of κ_r and κ_l in a similar way.

Remark 3.3. Unlike $\mathcal{H}_{n,r}$, $\mathcal{H}_{n,l}$ relies on the choice of bandwidth b_n. Hereafter, we choose b_n as in Fan and Yao (2003, p.201). Simulation studies in Section 4 imply that this choice of b_n has a good finite sample performance.

To carry out the LADE-based Hausman testing procedure, one computes (3.4) and compares it to the upper critical value $c_{m,\alpha}$ at a given significance level α. If $\mathcal{H}_{n,l} > c_{m,\alpha}$, then we reject H_0; otherwise, we can not reject H_0.

In the end, we make the following assumption:

Table 1. *The values of κ_r and κ_l for the often used distribution function $F_0(\cdot)$*

		Distribution function $F_0(\cdot)$ under H_0													
	r	N$(0,1)$	Laplace$(0,1)$	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{15}	t_{20}
κ_r	0.0	0.5301	0.5618	1.0000	0.7074	0.6374	0.6068	0.5897	0.5788	0.5712	0.5657	0.5615	0.5582	0.5484	0.5437
	0.2	0.5922	0.6527	1.2836	0.8253	0.7296	0.6892	0.6671	0.6532	0.6436	0.6366	0.6313	0.6271	0.6150	0.6091
	0.4	0.6492	0.7416	1.6889	0.9494	0.8205	0.7686	0.7407	0.7234	0.7115	0.7029	0.6964	0.6913	0.6765	0.6694
	0.6	0.7020	0.8290		1.0843	0.9120	0.8462	0.8117	0.7905	0.7761	0.7658	0.7580	0.7519	0.7342	0.7258
	0.8	0.7514	0.9150		1.2360	1.0055	0.9231	0.8809	0.8554	0.8382	0.8259	0.8167	0.8095	0.7888	0.7790
	1.0	0.7979	1.0000			1.1027	1.0000	0.9491	0.9186	0.8984	0.8839	0.8731	0.8647	0.8408	0.8295
	1.2	0.8421	1.0842			1.2053	1.0779	1.0166	0.9807	0.9570	0.9402	0.9277	0.9181	0.8906	0.8778
	1.4	0.8841	1.1676			1.3157	1.1573	1.0842	1.0420	1.0145	0.9952	0.9809	0.9698	0.9387	0.9241
	1.6	0.9243	1.2503				1.2392	1.1521	1.1029	1.0712	1.0491	1.0328	1.0202	0.9851	0.9688
	1.8	0.9629	1.3325				1.3245	1.2209	1.1637	1.1274	1.1022	1.0837	1.0696	1.0302	1.0121
	2.0	1.0000	1.4142					1.2910	1.2248	1.1832	1.1547	1.1339	1.1181	1.0742	1.0541
	2.2	1.0359	1.4955					1.3629	1.2863	1.2390	1.2068	1.1835	1.1658	1.1171	1.0950
	2.4	1.0706	1.5764					1.4371	1.3486	1.2949	1.2587	1.2326	1.2129	1.1592	1.1349
	2.6	1.1042	1.6569						1.4119	1.3510	1.3104	1.2814	1.2595	1.2004	1.1739
	2.8	1.1368	1.7372						1.4767	1.4077	1.3622	1.3300	1.3058	1.2410	1.2122
	3.0	1.1686	1.8171							1.4650	1.4142	1.3785	1.3519	1.2810	1.2497
κ_l		0.6745	0.6931	1.0000	0.8165	0.7649	0.7407	0.7267	0.7176	0.7111	0.7064	0.7027	0.6998	0.6912	0.6870

† For the distribution of t_ν, the values of κ_r are absent when $r \geq \nu/2$, according to Assumption 2.6(i).

Assumption 3.2. There exists a unique interior point $\theta_{*,l} \in \Theta$ such that $\sqrt{n}[(\widetilde{\theta}_{n,l} - \theta_{0,l}) - (\widehat{\theta}_{n,l} - \theta_{*,l})] \to_d \xi_l$ (a distribution) as $n \to \infty$.

Note that the proceeding assumption implies that $\widehat{\theta}_{n,l} - \theta_{*,l} = o_p(1)$. We now study the asymptotic power of $\mathcal{H}_{n,l}$ by considering the alternative hypothesis

$$H_{1,l} : \theta_{0,l} - \theta_{*,l} \neq 0,$$

and the local alternative hypothesis

$$H_{1n,l} : \theta_{0,l} - \theta_{*,l} = \frac{\Delta}{\sqrt{n}} \text{ for some constant vector } \Delta \in \mathcal{R}^m.$$

COROLLARY 3.2. *Suppose that Assumption* 3.2 *and conditions (i)-(ii) and (iv)-(v) in Theorem* 3.1 *hold. Then, under $H_{1,l}$, we have $\lim_{n\to\infty} \mathcal{H}_{n,l} = \infty$; and under $H_{1n,l}$, we have*

$$\mathcal{H}_{n,l} \to_d (\xi_l + \Delta)'(\tau_l^{-1}\mathcal{J}_l)(\xi_l + \Delta) \quad \textit{as } n \to \infty,$$

and consequently, $\lim_{|\Delta|\to\infty} \lim_{n\to\infty} \mathcal{H}_{n,l} = \infty$.

The proof of Corollary 3.2 is directly from continuous mapping theorem. From this corollary, we know that $\mathcal{H}_{n,l}$ can consistently detect $H_{1,l}$, and has the nontrivial local power to detect $H_{1n,l}$. Again, as for $\mathcal{H}_{n,r}$, $\mathcal{H}_{n,l}$ is lack of power when $\theta_{0,l} = \theta_{*,l}$ under alternatives, and this situation shall happen with a small chance in applications. Simulation studies in Section 4 imply that $\mathcal{H}_{n,l}$ has a good finite performance especially when η_t is heavy-tailed.

4. Simulation study

In this section, we examine the performance of the test statistics $\mathcal{H}_{n,r}$ and $\mathcal{H}_{n,l}$ in finite samples through Monte Carlo experiments. We generate 5000 replications of sample size $n = 1000$ and 2000 from the following GARCH$(1, 1)$ model:

$$y_t = \eta_t \sigma_t \quad \text{and} \quad \sigma_t^2 = \omega_0 + \alpha_0 y_{t-1}^2 + \beta_0 \sigma_{t-1}^2, \tag{4.1}$$

where $(\omega_0, \alpha_0, \beta_0) = (0.025, 0.25, 0.5)$ as in Koul and Ling (2006), and η_t is i.i.d. and generated, respectively, as follows:

Case 1 : $\eta_t \sim$ standardized $[\mathrm{N}(0, 1) + \lambda t_5]$ such that $E\eta_t^2 = 1$;

Case 2 : $\eta_t \sim$ standardized $[\mathrm{N}(0, 1) + \lambda t_3]$ such that $E\eta_t^2 = 1$;

Case 3 : $\eta_t \sim$ standardized $[\text{Laplace}(0, 1) + \lambda t_5]$ such that $E|\eta_t| = 1$;

Case 4 : $\eta_t \sim$ standardized $[\text{Laplace}(0, 1) + \lambda t_3]$ such that $E|\eta_t| = 1$;

Case 5 : $\eta_t \sim t_{8+5\lambda}$;

Case 6 : $\eta_t \sim t_{8-5\lambda}$.

Here, λ is chosen to be $0.0, 0.2, 0.4, 0.6, 0.8$ or 1.0. For each case, the null hypothesis H_0 corresponds to the scenario that $\lambda = 0$, and its alternatives are the scenarios that $\lambda > 0$. In view of Assumption 2.6(i), we choose $\mathcal{H}_{n,r}$ with $r = 0.0, 0.6$ and 1.2 for all cases, and also $r = 1.8$ and

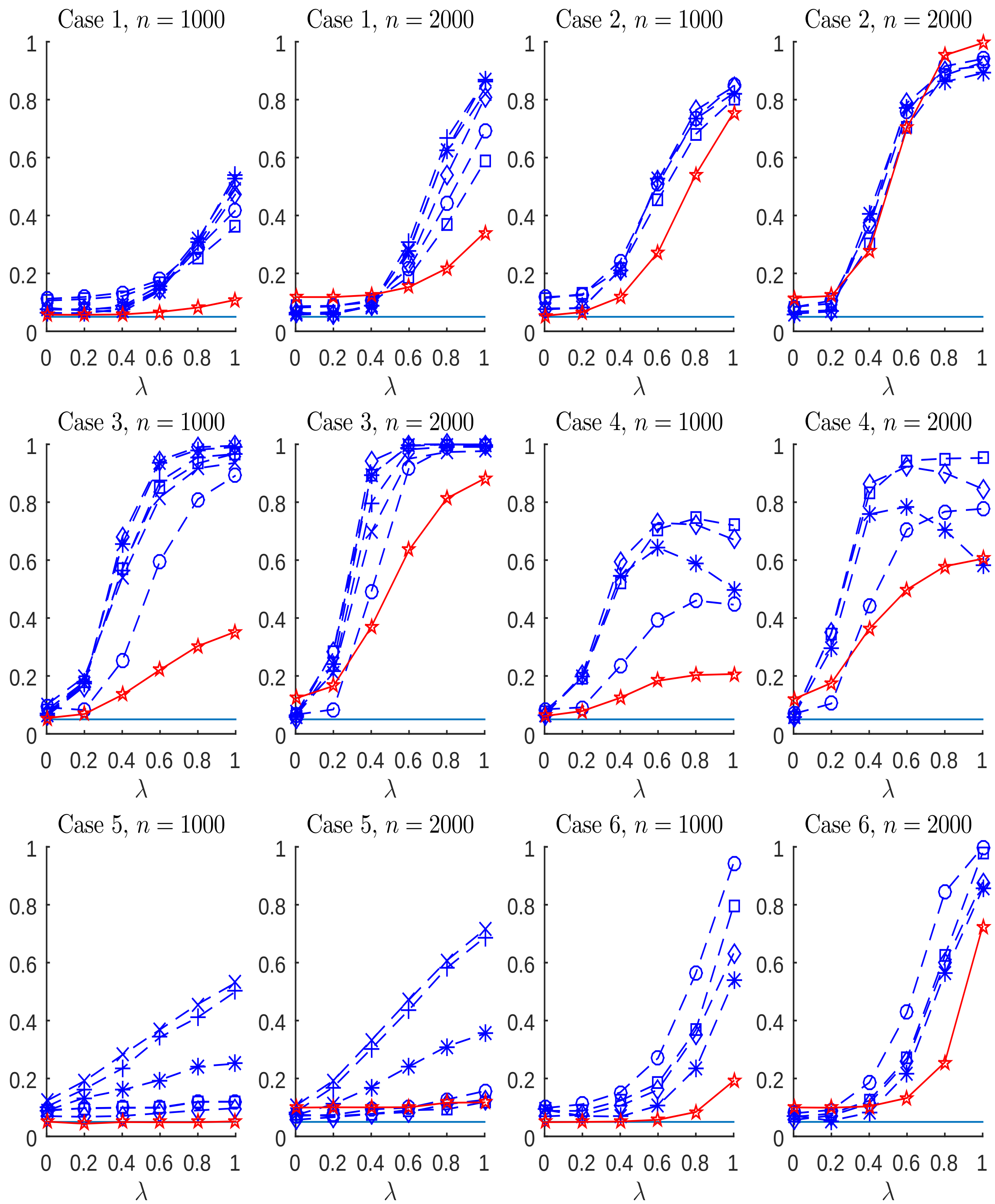

Fig. 1. The power and size plot in each case for $\mathcal{H}_{n,0}$ (dashed square line), $\mathcal{H}_{n,0.6}$ (dashed diamond line), $\mathcal{H}_{n,1.2}$ (dashed star line), $\mathcal{H}_{n,1.8}$ (dashed plus line), $\mathcal{H}_{n,2.4}$ (dashed cross line), $\mathcal{H}_{n,l}$ (dashed circle line), and $\mathcal{K}_n$ (solid pentangle line). Here, the horizontal solid line is the significance level $\alpha = 5\%$.

2.4 for Cases 1, 3 and 5. As a comparison, we also consider the weighted KS test statistic $\mathcal{K}_n$ in Koul and Ling (2006). In all calculations, we set the significance level $\alpha = 5\%$.

Figure 1 reports the result of all test statistics, and our findings from this figure are as follows:

(i) all test statistics have a precise size performance especially when the sample size n is large, and our Hausman tests may be slightly over-sized when n is small.

(ii) the power of all test statistics becomes large when the value of λ (or n) increases.

(iii) $\mathcal{H}_{n,l}$ has a comparable power performance to $\mathcal{H}_{n,r}$ with small r in Cases 1 and 2, and it is the worse one among all Hausman tests in Cases 3 and 4. However, when the tail of η_t becomes much heavier as in Case 6, $\mathcal{H}_{n,l}$ is the most powerful one among all test statistics.

(iv) $\mathcal{H}_{n,r}$ with large (or small) r exhibits a power advantage over others when the tail of η_t becomes lighter (or heavier) as shown in Case 5 (or 6). In Cases 1 and 3, $\mathcal{H}_{n,r}$ with large r has a comparable power performance with the one with small r.

(v) Except Case 2 in which $\mathcal{K}_n$ has a comparable power performance with $\mathcal{H}_{n,r}$ and $\mathcal{H}_{n,l}$, $\mathcal{K}_n$ in general has the worse power performance among all test statistics, especially in Case 5. It is also worth noting that the power advantage of $\mathcal{H}_{n,r}$ or $\mathcal{H}_{n,l}$ over $\mathcal{K}_n$ is remarkably significant when the value of λ is greater than 0.2 or 0.4.

Overall, our simulation studies reveal that $\mathcal{H}_{n,r}$ with large (or small) r has a good performance when the tail of η_t is light (or heavy), and $\mathcal{H}_{n,l}$ has a desirable performance when the tail of η_t is heavy, while $\mathcal{K}_n$ is generally less powerful than $\mathcal{H}_{n,r}$ and $\mathcal{H}_{n,l}$ in all examined alternatives.

5. APPLICATION

This section studies the daily S&P 500 index in U.S. stock market. The data sets we considered are divided into two groups by the 1987's crash. The first group is collected from January 3, 1979 to December 31, 1986, and the second group is collected from January 2, 1987 to December 30, 1994. Since the log-return ($\times 100$) of the data set in the first group exhibits some correlations in its conditional mean, it is filtered by an ARMA$(2,1)$ model with the least square estimation method. Likewise, the log-return $(\times 100)$ of the data set in the second group is filtered by an ARMA$(1,4)$ model. Consequently, we denote the residuals from each fitted ARMA model by $\{y_t\}_{t=1}^n$, where n is the sample size. Table 2 gives the summary statistics for each y_t, from which we find that the p-values of the Li-Mak portmanteau tests are close to zero. Hence, it implies that y_t has the ARCH effect in each group.

Table 2. *The summary of y_t in each group.*

y_t	n	mean	sd	skewness	kurtosis	$Q_{\text{lb}}(6)^{\dagger}$	$Q_{\text{lb}}(18)$	$Q_{\text{lm}}(6)^{\ddagger}$	$Q_{\text{lm}}(18)$
ex-1987	2012	0.0000	0.8765	0.0560	4.7662	0.8934	0.8750	0.0000	0.0000
post-1987	2019	0.0004	1.0657	-5.2596	111.81	0.8102	0.2405	0.0000	0.0000

† The p-value of Ljung-Box portmanteau test $Q_{\text{lb}}(M)$ in Ljung and Box (1978).
‡ The p-value of Li-Mak portmanteau test $Q_{\text{lm}}(M)$ in Li and Mak (1994).

Next, we fit each $\{y_t\}_{t=1}^n$ by a GARCH$(1,1)$ model in (4.1) with the Gaussian QMLE method, and find that the p-values of the Li-Mak portmanteau tests $Q_{\text{lm}}(6)$ and $Q_{\text{lm}}(18)$ are 0.7026 and 0.6293 for the ex-1987 data set, and 0.9876 and 0.9996 for the post-1987 data set. Hence, we can conclude that the GARCH$(1,1)$ model is adequate to fit both data sets. Furthermore, we are of interest to test the distribution of η_t in model (4.1). We consider four different null hypotheses, respectively, as follows:

$$H_0^{(1)} : \eta_t \sim \text{N}(0,1); \quad H_0^{(2)} : \eta_t \sim \text{Laplace}(0,1);$$
$$H_0^{(3)} : \eta_t \sim t_5; \qquad H_0^{(4)} : \eta_t \sim t_8.$$

We apply $\mathcal{H}_{n,r}$ with $r = 0, 0.6, 1.2, 1.8$ or 2.4, $\mathcal{H}_{n,l}$, and $\mathcal{K}_n$ to detect each null hypothesis above. The corresponding results are given in Table 3. From this table, we can find that (i) for the ex-1987 data set, only $H_0^{(4)}$ is accepted by all test statistics, while the other hypotheses are strongly rejected by the Hausman test, especially the GQMLE-based one with large r; (ii) for the post-1987 data set, none of hypotheses is accepted by the Hausman test, especially the LADE-based one and GQMLE-based one with small r. It is worth noting that (i) for the ex-1987 data set, $\mathcal{K}_n$

can reject $H_0^{(1)}$ only at 10% level and can not reject $H_0^{(3)}$ and $H_0^{(4)}$ at that level; (ii) for the post-1987 data set, $\mathcal{K}_n$ only has the marginal ability to reject $H_0^{(4)}$ at 10% level, and has no ability to reject $H_0^{(2)}$ and $H_0^{(3)}$ at that level.

Table 3. *The values of test statistics for null hypotheses $H_0^{(i)}$ ($i = 1, 2, 3, 4$).*

	ex-1987							post-1987						
	Tests†							Tests						
	$\mathcal{H}_{n,0}$	$\mathcal{H}_{n,0.6}$	$\mathcal{H}_{n,1.2}$	$\mathcal{H}_{n,1.8}$	$\mathcal{H}_{n,2.4}$	$\mathcal{H}_{n,l}$	$\mathcal{K}_n^{\ddagger}$	$\mathcal{H}_{n,0}$	$\mathcal{H}_{n,0.6}$	$\mathcal{H}_{n,1.2}$	$\mathcal{H}_{n,1.8}$	$\mathcal{H}_{n,2.4}$	$\mathcal{H}_{n,l}$	$\mathcal{K}_n$
$H_0^{(1)}$	14.215	23.012	25.567	19.189	16.208	22.487	2.6980	212.17	92.716	44.359	20.327	8.0543	197.50	11.125
	[0.0026]	**[0.0000]**	**[0.0000]**	**[0.0002]**	**[0.0010]**	**[0.0001]**		**[0.0000]**	**[0.0000]**	**[0.0000]**	**[0.0001]**	[0.0449]	**[0.0000]**	
$H_0^{(2)}$	54.463	90.940	98.181	91.792	64.000	56.538	2.9095	27.460	19.243	10.522	6.4201	4.1448	3.1184	1.9454
	[0.0000]	**[0.0000]**	**[0.0000]**	**[0.0000]**	**[0.0000]**	**[0.0000]**		**[0.0000]**	**[0.0002]**	[0.0146]	[0.0929]	[0.2462]	[0.3737]	
$H_0^{(3)}$	2.2091	0.0984	20.559	19.746	20.965	1.1025	0.9077	21.321	5.3959	3.2087	4.4935	3.4764	17.393	1.5854
	[0.5302]	[0.9920]	**[0.0001]**	**[0.0002]**	**[0.0001]**	[0.7765]		**[0.0001]**	[0.1450]	[0.3606]	[0.2129]	[0.3238]	**[0.0006]**	
$H_0^{(4)}$	0.1281	0.2755	1.3522	0.3121	0.2116	0.8880	0.9354	46.286	17.338	3.7204	5.7531	3.9882	42.493	2.4896
	[0.9883]	[0.9646]	[0.7168]	[0.9577]	[0.9757]	[0.8283]		**[0.0000]**	**[0.0006]**	[0.2933]	[0.1243]	[0.2627]	**[0.0000]**	

† The p-value of the Hausman test is in the square bracket, and its value less than 1% is in bold face.

‡ The 10%, 5% and 1% upper percentiles of $\mathcal{K}_n$ are 2.382, 2.804 and 3.737 for $\eta_t \sim \mathrm{N}(0, 1)$, 2.344, 2.781 and 3.149 for $\eta_t \sim \mathrm{Laplace}(0, 1)$, 2.428, 2.852 and 3.691 for $\eta_t \sim t_5$, and 2.464, 2.897 and 3.793 for $\eta_t \sim t_8$, respectively.

In view of these facts, we shall fit the ex-1987 data set by a GARCH$(1, 1)$ model with $\eta_t \sim t_8$. Table 4 reports the related results for this fitted model, from which we can see that the sample skewness and kurtosis of residuals are -0.0006 and 4.3151, which are close to 0 and 4.5 (the skewness and kurtosis of t_8 distribution), respectively. To gain more evidence, we apply the three-step estimation method in Fan, Qi, and Xiu (2014) to the ex-1987 data set with the auxiliary innovation being t_8, and find that the estimate of η_f (see, eqn (6) in that paper) is 1.0008. This suggests that the true distribution of η_t has the same tail thickness as t_8, and so it is consistent to our findings. By using the same method, we also find that the true distribution of η_t has the same tail thickness as $t_{4.48}$ for the post-1987 data set. Thus, it motivates us to consider one more null hypothesis for the post-1987 data set:

$$H_0^{(5)} : \eta_t \sim t_{4.48}.$$

However, some additional results (not reported here but available upon requirement) show that $H_0^{(5)}$ is rejected by both $\mathcal{H}_{n,0}$ and $\mathcal{H}_{n,l}$ with p-values less than 1%. Nevertheless, we try to fit the post-1987 data set by a GARCH$(1,1)$ model with $\eta_t \sim t_{4.48}$, and the corresponding results are given in Table 4. Clearly, the sample kurtosis of residuals is slightly larger than 15.5 (the kurtosis of $t_{4.48}$ distribution), while the sample skewness of residuals is much less than 0 (the skewness of $t_{4.48}$ distribution). Thus, the failure of $t_{4.48}$ in fitting η_t for the post-1987 data set may be due to its inability to fit asymmetric data set.

Table 4. *The summary of the fitted GARCH$(1,1)$ model for each y_t.*

		MLE†			Residuals				
y_t	η_t	$\widehat{\omega}_n$	$\widehat{\alpha}_n$	$\widehat{\beta}_n$	IC‡	mean	sd	skewness	kurtosis
ex-1987	t_8	0.0143	0.0281	0.9384	0.0001	-0.0057	1.3136	-0.0006	4.3151
		(0.0048)	(0.0078)	(0.0159)					
post-1987	$t_{4.48}$	0.0076	0.0347	0.9241	0.0000	0.0087	1.8466	-1.3243	16.275
		(0.0022)	(0.0069)	(0.0128)					

† The standard deviation of the MLE is in the open bracket.
‡ According to (2.6), the identification condition of the MLE with $\eta_t \sim t_\nu$ is $E[1/(\eta_t^2+\nu)] = 1/(1+\nu)$. IC stands for the sample value of $\{E[1/(\eta_t^2+\nu)] - 1/(1+\nu)\}$ based on residuals.

In summary, we find that the error distribution is t_8 in fitted GARCH$(1,1)$ model for the ex-1987 data set, and we also expect that the error distribution in fitted GARCH$(1,1)$ model for the post-1987 data set may be a skewed one with the tail thickness as $t_{4.48}$.

6. CONCLUDING REMARKS

In this paper, we propose the novel GQMLE-based Hausman test statistic $\mathcal{H}_{n,r}$ and LADE-based Hausman test statistic $\mathcal{H}_{n,l}$ for checking the error distribution in conditionally heteroskedastic models. Both test statistics are shown to have the limiting null distribution χ^2, and so they are ADF. Moreover, both test statistics are consistent and able to detect the local alternative of order $n^{-1/2}$. Simulation studies reveal that our Hausman test statistics have a power advantage over the weighted KS test statistic $\mathcal{K}_n$ under most of the examined alternatives. By studying the S&P 500 stock index from 1979 to 1994, our Hausman test statistics find that based on the fitted GARCH$(1,1)$ model, the error distribution is t_8 for the ex-1987 data set, and it may be a skewed distribution with the tail thickness as $t_{4.48}$ for the post-1987 data set.

It is worth noting that both simulation study and real application imply that $\mathcal{H}_{n,l}$ has a better performance when the error is more heavy-tailed, while $\mathcal{H}_{n,r}$ with a large (or small) r has a better performance when the error is less (or more) heavy-tailed. It means that practitioners may select the range of r by looking at errors' tail index (e.g., Hill's estimators of the robust LADE-based errors). Needless to say, it is always better to try different choices of r in real applications, and this can give us more information on the distribution of the error term.

As one natural extension, we may consider our Hausman testing procedure for the error distribution in other time series models, such as the heteroskedastic model with a conditional mean, the heteroskedastic model without intercept (e.g., Hafner and Preminger (2015)), the non-stationary heteroskedastic model (e.g, Francq and Zakoïan (2012)), and the multivariate heteroskedastic model. This extension is interesting and left for future research.

APPENDIX: PROOFS

Proof of Theorem 2.1. Under Assumptions 2.1 and 2.3-2.6, Theorem 1 in Francq and Zakoïan (2013) showed that

$$\sqrt{n}\left(\widetilde{\theta}_{n,r}-\theta_{0,r}\right)=\begin{cases}-\Sigma_{0,r}^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{1}{\sigma_t(\theta_{0,r})}\frac{\partial\sigma_t(\theta_{0,r})}{\partial\theta}[1-|\eta_t|^r]+o_p(1), & \text{if } r>0,\\ \mathcal{J}_r^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{1}{\sigma_t(\theta_{0,r})}\frac{\partial\sigma_t(\theta_{0,r})}{\partial\theta}\log|\eta_t|+o_p(1), & \text{if } r=0,\end{cases} \tag{A1}$$

where $\Sigma_{0,r}=r\mathcal{J}_r>0$. Moreover, by Assumptions 2.3-2.6 and 2.8, and the same arguments as for Theorem 1 in Francq and Zakoian (2013), we have under $H_{0,r}^{\dagger}$, $E[k_{1,r}(\eta_t)]=-1$ and

$$\sqrt{n}\left(\widehat{\theta}_{n,r}-\theta_{0,r}\right)=-\Sigma_{1,r}^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{1}{\sigma_t(\theta_{0,r})}\frac{\partial\sigma_t(\theta_{0,r})}{\partial\theta}[1+k_{1,r}(\eta_t)]+o_p(1), \tag{A2}$$

where $\Sigma_{1,r}=\{1-E[k_{2,r}(\eta_t)]\}\mathcal{J}_r>0$.

Furthermore, by (A1)-(A2) and the central limit theorem for martingale difference sequence, we have

$$\sqrt{n}\left(\widetilde{\theta}_{n,r}-\widehat{\theta}_{n,r}\right)\to_d N(0,E\{W_r\mathcal{J}_rW_r\}) \tag{A3}$$

as $n \to \infty$, where $E\{W_r \mathcal{J}_r W_r\} > 0$ and

$$W_r = \begin{cases} [|\eta_t|^r - 1]\Sigma_{0,r}^{-1} + [1 + k_1(\eta_t)]\Sigma_{1,r}^{-1}, & \text{if } r > 0, \\ \log|\eta_t|\mathcal{J}_r^{-1} + [1 + k_1(\eta_t)]\Sigma_{1,r}^{-1}, & \text{if } r = 0. \end{cases}$$

Now, the conclusion holds from (A3) and the direct calculation.

Proof of Theorem 3.1**.** Under Assumptions 2.1 and 2.3-2.6, Lemma A.1 in Chen and Zhu (2015) showed that

$$\sqrt{n}\left(\widetilde{\theta}_{n,l} - \theta_{0,l}\right) = -\Sigma_{0,l}^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{1}{\sigma_t(\theta_{0,l})}\frac{\partial\sigma_t(\theta_{0,l})}{\partial\theta}\text{sgn}(1-\eta_t^2) + o_p(1), \tag{A4}$$

where $\Sigma_{0,l} = 4g(0)\mathcal{J}_l > 0$. Moreover, as for (A2), we have under $H_{0,l}^{\dagger}$, $E[k_{1,l}(\eta_t)] = -1$ and

$$\sqrt{n}\left(\widehat{\theta}_{n,l} - \theta_{0,l}\right) = -\Sigma_{1,l}^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{1}{\sigma_t(\theta_{0,l})}\frac{\partial\sigma_t(\theta_{0,l})}{\partial\theta}[1 + k_{1,l}(\eta_t)] + o_p(1), \tag{A5}$$

where $\Sigma_{1,l} = \{1 - E[k_{2,l}(\eta_t)]\}\mathcal{J}_l > 0$.

Furthermore, by (A4)-(A5) and the central limit theorem for martingale difference sequence, we have

$$\sqrt{n}\left(\widetilde{\theta}_{n,l} - \widehat{\theta}_{n,l}\right) \to_d N(0, E\{W_l\mathcal{J}_l W_l\}) \tag{A6}$$

as $n \to \infty$, where $E\{W_l\mathcal{J}_l W_l\} > 0$ and $W_l = \text{sgn}(\eta_t^2 - 1)\Sigma_{0,l}^{-1} + [1 + k_{1,l}(\eta_t)]\Sigma_{1,l}^{-1}$. Now, the conclusion holds from (A6) and the direct calculation.

REFERENCES

BERKES, I. & HORVÁTH, L. (2001). Strong approximation of the empirical process of GARCH sequences. *Annals of Applied Probability* **11**, 789-809.

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

BOLDIN, M.V. (1982). Estimation of the distribution of noise in an autoregressive scheme. *Theory of Probability and Its Applications* **27**, 866-871.

BOLDIN, M.V. (1989). On hypothesis testing in a moving average scheme by the Kolmogorov-Smirnov and the ω^2 tests. *Theory of Probability and Its Applications* **34**, 699-704.

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

BOLLERSLEV, T., CHOU, R.Y. & KRONER, K.F. (1992). ARCH modelling in finance: A review of the theory and empirical evidence. *Journal of Econometrics* **52**, 5–59.

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

CHRISTOFFERSEN, P.F. & DIEBOLD, F.X. (1997). Optimal prediction under asymmetric loss. *Econometric Theory* **13**, 808-817,

D'AGOSTINO, R.B. & STEPHENS, M.A. (1986). Goodness-of-Fit Techniques. Dekker, New York.

DING, Z., GRANGER, C.W.J. & ENGLE, R.F. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance* **1**, 83-106.

ENGLE, R.F. (1982). Autoregressive conditional heteroskedasticity with estimates of variance of U.K. inflation. *Econometrica* **50**, 987-1008.

ENGLE, R.F. (2004). Risk and volatility: Econometric models and financial practice. *American Economic Review* **94**, 405-420.

FAN, J., QI, L. & XIU, D. (2014). Quasi maximum likelihood estimation of GARCH models with heavy-tailed likelihoods (with discussion). *Journal of Business and Economic Statistics* **32**, 178-205.

FAN, J. & YAO, Q. (2003). *Nonlinear time series: Nonparametric and parametric methods*. Springer, New York.

FRANCQ, C., WINTENBERGER, O. & ZAKOÏAN, J.M. (2013). Garch models without positivity constraints: exponential or log garch? *Journal of Econometrics* **177**, 34-46.

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

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

FRANCQ, C. & ZAKOÏAN, J.M. (2013). Optimal predictions of powers of conditionally heteroscedastic processes. *Journal of the Royal Statistical Society: Series B* **75**, 345-367.

GEWEKE, J. (1986). Modeling the persistence of conditional variances: A comment. *Econometric Review* **5**, 57-61.

HAFNER, C.M. & PREMINGER, A. (2015). An ARCH model without intercept. *Economics Letters* **129**, 13-17.

HAMADEH, T. & ZAKOÏAN, J.M. (2011). Asymptotic properties of LS and QML estimators for a class of nonlinear GARCH processes. *Journal of Statistical Planning and Inference* **141**, 488-507.

HAUSMAN, J.A. (1978). Specification tests in econometrics. *Econometrica* **46**, 1251-1271.

HORVÁTH, L., KOKOSZKA, P. & TEYSSIÈRE, G. (2001). Empirical process of the squared residuals of an ARCH sequence. *Annals of Statistics* **29**, 445-469.

HORVÁTH, L., KOKOSZKA, P. & TEYSSIÈRE, G. (2004). Bootstrap misspecification tests for ARCH based on the empirical process of squared residuals. *Journal of Statistical Computation and Simulation* **74**, 469-485.

HORVÁTH, L. & ZITIKIS, R. (2006). Testing goodness of fit based on densities of GARCH innovations. *Econometric theory* **22**, 457-482.

KLAR, B., LINDNER, F., & MEINTANIS, S.G. (2012). Specification tests for the error distribution in GARCH models. *Computational Statistics and Data Analysis* **56**, 3587-3598.

KOUL, H.L. (1996). Asymptotics of some estimators and sequential residual empiricals in nonlinear time series. *Annals of Statistics* **24** 380-404.

KOUL, H.L. (2002). *Weighted Empirical Processes in Dynamic Nonlinear Models*, second ed., in: Lecture Notes in Statistics, vol. 166, Springer-Verlag, New York.

KOUL, H.L. & LING, S. (2006). Fitting an error distribution in some heteroscedastic time series models. *Annals of Statistics* **34**, 994-1012.

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

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

LINTON, O., PAN, J. & WANG, H. (2010). Estimation for a non-stationary semi-strong GARCH$(1,1)$ model with heavy-tailed errors. *Econometric Theory* **26**, 1-28.

LJUNG, G.M. & BOX, G.E.P. (1978). On a measure of lack of fit in time series models. *Biometrika* **65**, 297-303.

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

ZAKOÏAN, J.M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and Control* **18**, 931-955.

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

ZHU, K. & LING, S. (2015). Model-based pricing for financial derivatives. *Journal of Econometrics* **187**, 447-457.