Sign-based portmanteau test for ARCH-type models with heavy-tailed innovations

Chen, Min and Zhu, Ke

Chinese Academy of Sciences

8 October 2013

SIGN-BASED PORTMANTEAU TEST FOR ARCH-TYPE MODELS WITH HEAVY-TAILED INNOVATIONS

By Min Chen and Ke Zhu

Chinese Academy of Sciences

This paper proposes a sign-based portmanteau test for diagnostic checking of ARCH-type models estimated by the least absolute deviation approach. Under the strict stationarity condition, the asymptotic distribution is obtained. The new test is applicable for very heavy-tailed innovations with only finite fractional moments. Simulations are undertaken to assess the performance of the sign-based test, as well as a comparison with other two portmanteau tests. A real empirical example for exchange rates is given to illustrate the practical usefulness of the test.

1. Introduction. After the seminal work of Engle (1982) and Bollerslev (1986), the following ARCH-type model has been widely used in economics and finance:

$$\varepsilon_t = \eta_t\sqrt{h_t} \quad \text{and} \quad h_t = h(\varepsilon_{t-1}, \varepsilon_{t-2}, \cdots; \theta_0), \tag{1.1}$$

where η_t being independent of $\{\varepsilon_j; j < t\}$ is a sequence of i.i.d. random variables, $\theta_0 \in \mathcal{R}^m$ is a parameter vector belonging to a parameter space Θ and $h : \mathcal{R}^\infty \times \Theta \to (0, \infty)$. The variable h_t is generally referred as the conditional variance of ε_t in the econometrics literature. Many existing models, such as GARCH model (Bollerslev (1986)), asymmetric power GARCH model (Ding et al. (1993)) and asymmetric log-GARCH model (Geweke (1986)), are embedded into model (1.1); see e.g., Bollerslev et al. (1992) and Francq and Zakoïan (2010) for more discussions in this context.

Due to the widespread use of model (1.1), a fundamental problem for practitioners is to check its adequacy. The portmanteau test initially proposed by Box and Pierce (1970) and Ljung and Box (1978) is for testing the i.i.d. assumption of η_t, and has become a popular tool for diagnostic checking of model (1.1). Li and Mak (1994) studied a portmanteau test for the Gaussian QMLE-type fitted GARCH model by using the square-residual autocorrelations; Ling and Li (1997) extended this method to the multivariate ARCH models; Carbon and Francq (2011) further investigate the portmanteau test for the asymmetric power GARCH model; see also Hong and Li

Keywords and phrases: ARCH-type model; heavy-tailed innovation; LAD estimator; model diagnostics; sign-based portmanteau test.

(2003), Escanciano (2007) and Ling and Tong (2011) for other diagnostic checking methods of model (1.1).

Although all of the aforementioned tests have achieved a great success, a necessary set-up for them is that $E\eta_t^4 < \infty$. This is because the asymptotic normality of the Gaussian QMLE in model (1.1) needs the condition that $E\eta_t^4 < \infty$; see, e.g., Hall and Yao (2003), Francq and Zakoïan (2004), and Ling (2007). Recently, more and more empirical studies have documented the very heavy-tailed innovations in financial time series; see Rachev (2003) and the reference therein. However, relatively few references have considered the diagnostic checking of model (1.1) when $E\eta_t^4 = \infty$. Based on the LAD-type estimator in Peng and Yao (2003), Li and Li (2005) proposed two portmanteau tests for GARCH models when $E\varepsilon_t^2 < \infty$ and $E\eta_t^4 < \infty$. However, none of portmanteau tests is valid when $E\eta_t^2 = \infty$ or $E\varepsilon_t^2 = \infty$ up to now.

In this paper, we first derive the limiting distribution of the autocorrelation functions of the sign of $\hat{\eta}_t^2 - 1$, where the residual $\hat{\eta}_t$ is obtained from model (1.1) fitted by the LAD approach in Peng and Yao (2003). Based on this, we further propose a sign-based portmanteau test statistic for model (1.1), and obtain its asymptotic distribution under the strict stationarity condition. The new test is applicable for very heavy-tailed innovations with only finite fractional moments of η_t (i.e., $E|\eta_t|^{2\iota} < \infty$ for some $\iota > 0$). Simulations are undertaken to assess the performance of the sign-based test, as well as a comparison with other two portmanteau tests in Li and Li (2005). A real empirical example for exchange rates is given to illustrate the practical usefulness of the test. To our best knowledge, our sign-based portmanteau test is the first one for testing the adequacy of the fitted ARCH-type model when $E\eta_t^2 = \infty$.

This paper is organized as follows. Section 2 derives our main results and hence the sign-based portmanteau test. Section 3 reports the simulation results. A real example is provided in Section 4. The proofs are presented in the Appendix. Throughout the paper, some symbols are conventional. A' is the transpose of 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 an indicator function.

2. Main results. Let $\theta \in \Theta$ be the unknown parameter of model (1.1). Given the observations $\{\varepsilon_n, ..., \varepsilon_1\}$ and the initial values $Y_0 \equiv \{\varepsilon_0, \varepsilon_{-1}, ...\}$, we can rewrite the parametric model (1.1) as

$$\eta_t(\theta) = \varepsilon_t / \sqrt{h_t(\theta)} \text{ and } h_t(\theta) = h(\varepsilon_{t-1}, \varepsilon_{t-2}, \cdots; \theta).$$

Here, $\eta_t(\theta_0) = \eta_t$ and $h_t(\theta_0) = h_t$. Assume that Θ is compact and the true value θ_0 is an interior point in Θ. Following Peng and Yao (2003), the least absolute deviation

(LAD) estimator of θ_0, denoted by $\hat{\theta}_n$, is defined as

$$\hat{\theta}_n = \arg\min_{\Theta} L_n(\theta),\ \ L_n(\theta) = \frac{1}{n}\sum_{t=1}^{n}\left|\log \varepsilon_t^2 - \log h_t(\theta)\right|.$$

Compared to Gaussian QMLE, the LAD estimator $\hat{\theta}_n$ is generally more robust and requires a weaker moment condition of η_t; see also Fan et al. (2013) for other robust alternative QML estimators in GARCH models. Let $z_t = \log \eta_t^2$. We first introduce the following assumptions:

ASSUMPTION 2.1. *(i) Almost surely (a.s.), $h_t(\theta) \geq \underline{w}$ for some $\underline{w} > 0$ and all $\theta \in \Theta$. Moreover, $h_t(\theta) = h_t(\theta_0)$ a.s. if and only if $\theta = \theta_0$; (ii) if $x'(\partial h_t(\theta)/\partial\theta_i)_{i=1,\cdots,m} = 0$ a.s. for any $x \in \mathcal{R}^m$, then $x = 0$.*

ASSUMPTION 2.2. *median$(z_t) = 0$ and the probability density function $f(x)$ of z_t satisfying $f(0) > 0$ and $\sup_{x\in R} f(x) < \infty$, is continuous at zero.*

ASSUMPTION 2.3. *ε_t is strictly stationary and ergodic.*

ASSUMPTION 2.4. *(i) $E\log|\varepsilon_t| < \infty$; (ii) $E[\sup_\theta|\log h_t(\theta)|] < \infty$; (iii)*

$$E\left[\sup_\theta\left\|\frac{1}{h_t(\theta)}\frac{\partial h_t(\theta)}{\partial\theta}\right\|\right]^2 < \infty \quad \textit{and} \quad E\left[\sup_\theta\left\|\frac{1}{h_t(\theta)}\frac{\partial^2 h_t(\theta)}{\partial\theta\partial\theta'}\right\|\right] < \infty.$$

Assumption 2.1 imposes some basic requirements on the function $h_t(\theta)$, and they are satisfied by most ARCH-type models; see, e.g., Francq and Zakoïan (2004, 2013). Assumption 2.2 is a general set-up for the LAD-type estimator; see, e.g., Peng and Yao (2003), Li and Li (2008) and Zhu and Ling (2011). Assumption 2.3 is weaker than the moment condition $E\varepsilon_t^2 < \infty$ as in Peng and Yao (2003) and Li and Li (2005, 2008), and its necessary and sufficient condition is provided in Bougerol and Picard (1992) for GARCH models; see also Hamadeh and Zakoïan (2011) and Francq et al. (2013) for sufficient conditions in asymmetric power GARCH/log-GARCH models, respectively. Assumption 2.4 gives some technical moment conditions, which have been verified for GARCH models in Ling (2007), asymmetric power GARCH models in Hamadeh and Zakoïan (2011) and asymmetric log-GARCH models in Francq et al. (2013) provided that Assumptions 2.1 and 2.3 and Assumption 2.5 below hold.

ASSUMPTION 2.5. *$E|\eta_t|^{2\iota} < \infty$ for some $\iota > 0$.*

Note that Assumption 2.5 as in Berkes and Horváth (2004) and Linton et al. (2010) allows for the very heavy-tailed η_t. As an independent interest, the strong consistency and asymptotic normality of $\hat{\theta}_n$ are derived in Lemma A.1 based on Assumptions 2.1-2.4.

Next, let $\xi_t = sgn(\eta_t^2 - 1)$, where $sgn(x) = I(x > 0) - I(x < 0)$. Since $median(\eta_t^2) = 1$ by Assumption 2.2, $\{\xi_t\}$ is a sequence of i.i.d. random variables with mean zero and variance one. Thus, we can propose a portmanteau test for model (1.1) by using the residual-autocorrelation functions of $\{\xi_t\}$. Denote the residuals $\hat{\eta}_t \triangleq \eta_t(\hat{\theta}_n)$ and $\hat{\xi}_t \triangleq sgn(\hat{\eta}_t^2 - 1)$. Then, the lag-$l$ residual autocorrelation function can be defined as

$$\hat{\rho}_l^* = \frac{\sum_{t=l+1}^n \left(\hat{\xi}_t - \bar{\xi}_n\right)\left(\hat{\xi}_{t-l} - \bar{\xi}_n\right)}{\sum_{t=1}^n (\hat{\xi}_t - \bar{\xi}_n)^2},$$

where $\bar{\xi}_n = n^{-1}\sum_{t=1}^n \hat{\xi}_t$. Note that $\hat{\theta}_n - \theta_0 = o_p(1)$ by Lemma A.1. Under Assumptions 2.1-2.4, by Theorem 3.1 in Ling and McAleer (2003) and the dominated convergence theorem, we can show that $\bar{\xi}_n = E(\xi_t) + o_p(1) = o_p(1)$ and

$$\frac{1}{n}\sum_{t=1}^n \left(\hat{\xi}_t - \bar{\xi}_n\right)^2 = var(\xi_t) + o_p(1) = 1 + o_p(1), \tag{2.1}$$

and hence theoretically we only need to consider

$$\hat{\rho}_l = \frac{1}{n}\sum_{t=l+1}^n \hat{\xi}_t \hat{\xi}_{t-l}.$$

Denote $\hat{\rho} = (\hat{\rho}_1, \cdots, \hat{\rho}_M)'$. We are now ready to give our main result on the limiting distribution of $\hat{\rho}$ in the following theorem:

THEOREM 2.1. *Suppose that Assumptions 2.1-2.4 hold. Then,*

$$\sqrt{n}\hat{\rho} \to_d N\left(0, I_M - X\Sigma^{-1}X'\right) \quad \text{as } n \to \infty,$$

where $X = (X_1, \cdots, X_M)'$ *and*

$$\Sigma = E\left[\frac{1}{h_t^2(\theta_0)}\frac{\partial h_t(\theta_0)}{\partial \theta}\frac{\partial h_t(\theta_0)}{\partial \theta'}\right] \text{ with } X_l = E\left[\frac{\xi_{t-l}}{h_t(\theta_0)}\frac{\partial h_t(\theta_0)}{\partial \theta}\right] \text{ for } l \geq 1.$$

PROOF. See the Appendix. □

REMARK 2.1. *In practice, the initial values* Y_0 *are unknown, and can be replaced by any constants. Unless stated otherwise, we set the initial values* $Y_0 \equiv 0$*, and denote* *the corresponding $h_t(\theta)$ as $\tilde{h}_t(\theta)$. Following the same argument as in Zhu (2011), we can show that this will not affect our asymptotic result in Theorem 2.1, if*

$$(i)\ \sup_{\theta}\left\|\frac{1}{\tilde{h}_t(\theta)}\frac{\partial \tilde{h}_t(\theta)}{\partial\theta}-\frac{1}{h_t(\theta)}\frac{\partial h_t(\theta)}{\partial\theta}\right\| \le O(\rho^t)R_t$$

$$(ii)\ \sup_{\theta}\left\|\frac{1}{\tilde{h}_t(\theta)}\frac{\partial^2 \tilde{h}_t(\theta)}{\partial\theta\partial\theta'}-\frac{1}{h_t(\theta)}\frac{\partial^2 h_t(\theta)}{\partial\theta\partial\theta'}\right\| \le O(\rho^t)R_t,$$

for some constant $\rho \in (0,1)$ and positive random variable R_t such that $ER_t^2 < \infty$. Particularly, based on Assumptions 2.1, 2.3 and 2.5, conditions (i)-(ii) have been verified for GARCH models in Ling (2007), asymmetric power GARCH models in Hamadeh and Zakoïan (2011), and asymmetric log-GARCH models in Francq et al. (2013).

Given the observations $\{\varepsilon_n, \cdots, \varepsilon_1\}$, we then can estimate the matrixes X and Σ by their sample means X_n and Σ_n, respectively. Under Assumptions 2.1-2.4, by a similar argument as for (2.1), we can show that $\hat{X}_n = X + o_p(1)$ and $\hat{\Sigma}_n = \Sigma + o_p(1)$. Thus, from Theorem 2.1, the following corollary is straightforward.

COROLLARY 2.1. *Suppose that Assumptions 2.1-2.4 hold. Then,*

$$S(M) \triangleq n\hat{\rho}'\left(I_M - \hat{X}_n\hat{\Sigma}_n^{-1}\hat{X}_n'\right)^{-1}\hat{\rho} \to_d \chi^2(M) \quad \textit{as } n \to \infty.$$

We call $S(M)$ in Corollary 2.1 the sign-based portmanteau test statistic. Unlike the portmanteau tests $Q(M)$ and $Q^2(M)$ in Li and Li (2005), the limiting distribution of $S(M)$ only requires a fractional moment of η_t and it is still valid when $E\varepsilon_t^2 = \infty$. Thus, $S(M)$ is applicable for the very heavy-tailed ε_t and η_t. Also, it is worthy noting that no estimation for $f(0)$ is needed in calculation of $S(M)$.

3. Simulation. In this section, we first examine the asymptotic result in Theorem 2.1. We generate 1000 replications of sample size $n = 200$ and 400 from model (3.1) and fit each replication by using the LAD method:

$$\varepsilon_t = \eta_t\sqrt{h_t},\ \ h_t = 0.01 + 0.2\varepsilon_{t-1}^2 + 0.2h_{t-1}, \tag{3.1}$$

where η_t is chosen to be the re-scaled $N(0,1)$, t_3, t_2 and t_1, respectively, such that it satisfies $median(\eta_t^2) = 1$. In this case, it is not hard to check that the conditions in Assumption 2.2 are satisfied. The asymptotic standard deviations of the residual autocorrelations $\hat{\rho}$ are calculated from Theorem 2.1 with $M = 6$. Table 1 lists the sample standard deviations (SD) and the average estimated asymptotic standard deviations (AD) of $\hat{\rho}$ for all lags. From Table 1, we can see that all pairs of AD and SD are close to each other for n as small as 200. As n increases from 200 to 400, all of the SDs and ADs become smaller.

TABLE 1
SDs and ADs (×10) for model (3.1)

			Lags					
η_t	n		1	2	3	4	5	6
$N(0,1)$	200	SD	0.441	0.629	0.697	0.655	0.693	0.701
		AD	0.435	0.615	0.674	0.689	0.696	0.699
	400	SD	0.313	0.450	0.464	0.487	0.494	0.504
		AD	0.299	0.438	0.474	0.486	0.492	0.495
t_3	200	SD	0.451	0.662	0.666	0.661	0.677	0.671
		AD	0.452	0.653	0.680	0.691	0.696	0.699
	400	SD	0.313	0.471	0.474	0.493	0.475	0.510
		AD	0.311	0.464	0.480	0.488	0.493	0.496
t_2	200	SD	0.456	0.684	0.649	0.682	0.696	0.701
		AD	0.457	0.663	0.683	0.691	0.696	0.699
	400	SD	0.323	0.477	0.474	0.495	0.500	0.498
		AD	0.316	0.470	0.483	0.489	0.493	0.495
t_1	200	SD	0.495	0.665	0.693	0.695	0.676	0.671
		AD	0.478	0.667	0.691	0.697	0.699	0.700
	400	SD	0.346	0.458	0.458	0.482	0.495	0.512
		AD	0.336	0.472	0.490	0.494	0.495	0.496

Next, we compare the finite sample performance of our sign-based test $S(M)$ with those of two portmanteau tests $Q(M)$ and $Q^2(M)$ in Li and Li (2005). We choose our null model as

$$\varepsilon_t = \eta_t\sqrt{h_t} \quad \text{and} \quad h_t = 0.01 + \alpha\varepsilon_{t-1}^2 + 0.8h_{t-1}, \tag{3.2}$$

and use the following two models to study the powers for all tests:

$$\varepsilon_t = \eta_t\sqrt{h_t} \quad \text{and} \quad h_t = 0.01 + \alpha\varepsilon_{t-1}^2 + 0.2\varepsilon_{t-2}^2 + 0.8h_{t-1}, \tag{3.3}$$

$$\varepsilon_t = \eta_t\sqrt{h_t} \quad \text{and} \quad h_t = 0.01 + \alpha\varepsilon_{t-1}^2 + 0.2\varepsilon_{t-2}^2, \tag{3.4}$$

where η_t is chosen as in model (3.1). In order to make sure that $E(\eta_t^2)\alpha + 0.8 \approx 1$ for N(0,1) and t_3 distributions, we take $\alpha = 0.08$ and 0.03, respectively. For t_1 and t_2 distributions, we take $\alpha = 0.03$ as for t_3 distribution. Based on these choices of α, we generate 1000 replications of sample size $n = 200, 400$ and 1000 from each model and fit each replication by a GARCH(1,1) model with the LAD method. The significance level $\underline{\alpha} = 0.05$ and $M = 6$. In all calculations (hereafter), $f(0)$ is estimated by using the default syntax "ksdensity" in MatLab. The empirical power and sizes of these tests are reported in Table 2. Their sizes correspond to the results for model (3.2).

TABLE 2
Empirical size and power ($\times 100$) *for* $S(M)$, $Q(M)$ *and* $Q^2(M)$

		model(3.2)			model(3.3)			model(3.4)		
η_t	n	$S(M)$	$Q(M)$	$Q^2(M)$	$S(M)$	$Q(M)$	$Q^2(M)$	$S(M)$	$Q(M)$	$Q^2(M)$
$N(0,1)$	200	4.60	6.00	10.6	10.3	19.38	16.8	19.2	43.5	34.3
	400	4.70	5.60	7.70	15.6	42.0	36.6	33.2	76.0	65.5
	1000	4.70	5.10	6.30	35.2	86.7	80.4	65.3	99.6	99.1
t_3	200	6.20	6.20	6.50	14.3	18.7	10.1	35.2	58.4	23.3
	400	5.70	6.90	7.10	26.3	40.7	12.9	64.1	84.1	29.0
	1000	5.50	5.10	5.40	58.2	82.5	14.2	96.8	98.3	39.7
t_2	200	6.50	6.60	5.80	16.7	12.8	6.20	44.1	38.3	12.0
	400	5.10	8.80	6.90	26.7	19.3	6.50	76.3	55.5	11.1
	1000	5.60	5.90	4.60	61.3	34.9	6.30	99.9	76.6	11.1
t_1	200	4.80	6.00	3.30	20.8	6.20	3.50	76.8	9.20	3.90
	400	6.50	6.20	3.50	33.1	4.60	2.90	97.4	8.50	3.60
	1000	4.20	4.60	1.60	72.0	3.80	1.10	100.0	6.40	2.10

From Table 2, it is clear that the sizes of $S(M)$ are always close to their nominal ones, while the sizes of $Q(M)$ and $Q^2(M)$ are not precise when n is small. For the power of these tests, it is generally as expected. First, except $Q(M)$ and $Q^2(M)$ in the case that $E\eta_t^2 = \infty$, all the powers become large as n increases. Second, $Q(M)$ is the most powerful test among these three tests when $E\eta_t^2 < \infty$. Third, $Q^2(M)$ is more powerful than $S(M)$ when $E\eta_t^4 < \infty$, while its power is less than that of $S(M)$ when $E\eta_t^4 = \infty$. Forth, $S(M)$ becomes more powerful when η_t is more heavy-tailed, but $Q(M)$ and $Q^2(M)$ lose their power substantially when $E\eta_t^2 = \infty$. Overall, $S(M)$ has a very good performance, especially when η_t is very heavy-tailed.

4. A real example. In this section, we study the daily exchange rate of United States Dollars (USD) to Chinese Yuan (CNY) from December 19, 2008 to May 13, 2010, which has in total 351 observations; see Figure 1 (a). Its 100 times log return, denoted by $\{\varepsilon_t\}_{t=1}^{350}$, is plotted in Figure 1 (b). To begin with, we first plot the kernel density of ε_t in Figure 2. Compared with the corresponding normal density, we know that ε_t is more heavy-tailed than the normal distribution. Thus, the Gaussian QMLE is not suitable in this case. Here, we consider the LAD estimation for the following ARCH model with $r = 3, r = 4$ and $r = 5$:

$$\varepsilon_t = \eta_t \sqrt{h_t} \quad \text{and} \quad h_t = \alpha_0 + \sum_{i=1}^{r} \alpha_i \varepsilon_{t-i}^2. \tag{4.1}$$

Our major interest concerns which of the three models can fit the data adequately. Table 3 presents all estimation results for these three fitted models. To check the adequacy of these models, the values of $S(M)$, $Q(M)$ and $Q^2(M)$ with $M = 6$

and $M = 12$ are also reported in the same table. From Table 3, we find that an ARCH(5) model is adequate according to all three statistics. However, S(M) implies both ARCH(3) and ARCH(4) models are not adequate, but this can not be detected

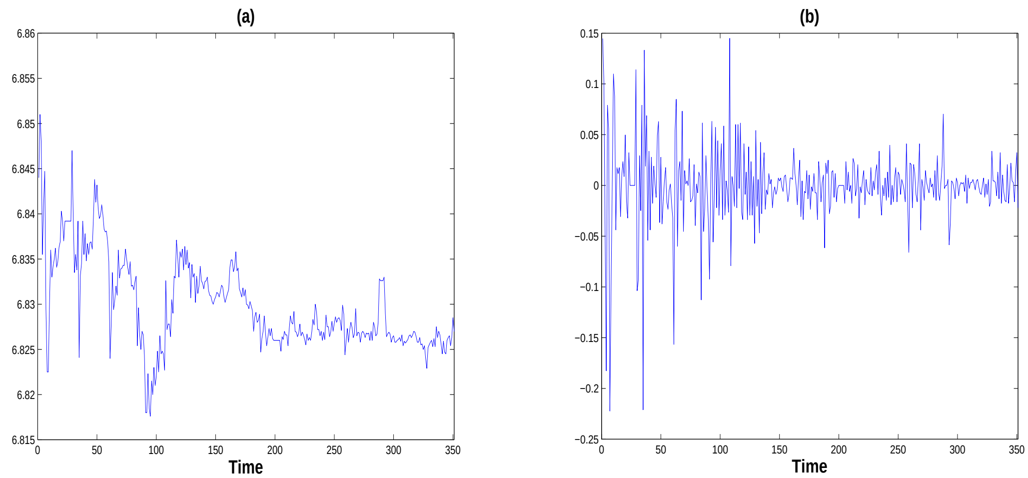

FIG 1. *(a) the daily exchange rate of USD/CNY from December 19, 2008 to May 13, 2010, and (b) its 100 times log return.*

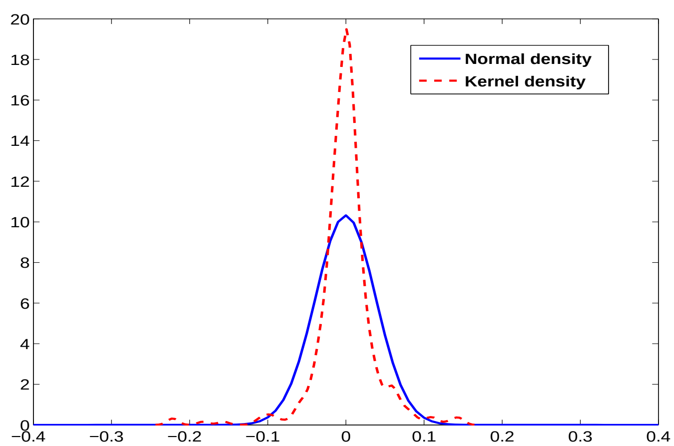

FIG 2. *The kernel density of ε_t and the normal density with the same mean and variance.*

by $Q(M)$ or $Q^2(M)$. To see the reason, Figure 3 plots the Hill's estimator $\hat{H}_\eta(k)$ with the largest k data of $\{\hat{\eta}_t^2\}$ for ARCH(3) model and ARCH(4) model, where

$$\hat{H}_\eta(k) = \frac{k}{\sum_{j=1}^{k}(\log \tilde{\eta}_{350-j} - \log \tilde{\eta}_{350-k})},$$

and $\tilde{\eta}_j$ is the j-th order statistic of $\hat{\eta}_t^2$. From Figure 3, we can see that the tail of η_t^2 in ARCH(3) model or ARCH(4) model is most likely less than 1, i.e., $E\eta_t^2 = \infty$. Thus,

$S(M)$ is more powerful than $Q(M)$ or $Q^2(M)$ under this heavy-tailed situation.

TABLE 3
Results for all fitted model(4.1)

	Models					
	$r=3$		$r=4$		$r=5$	
Parameters	$\hat{\theta}_n$	AD	$\hat{\theta}_n$	AD	$\hat{\theta}_n$	AD
α_0[a]	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
α_1	0.2701	0.0801	0.2517	0.0769	0.1904	0.0678
α_2	0.0394	0.0395	0.0308	0.0364	0.0297	0.0355
α_3	0.0788	0.0400	0.0658	0.0399	0.0636	0.0401
α_4			0.0295	0.0285	0.0001	0.0218
α_5					0.0918	0.0421
$(S(6), S(12))$[b]	(13.16, 17.59)		(19.87, 22.63)		(9.93, 11.32)	
$(Q(6), Q(12))$[b]	(12.52, 17.62)		(10.76, 15.48)		(5.22, 10.45)	
$(Q^2(6), Q^2(12))$[b]	(0.98, 1.50)		(0.71, 1.08)		(0.42, 0.69)	

[a] The estimator $\hat{\alpha}_{0n}$ and its AD are less than 10^{-4} for each model.
[b] The 95% upper percentages for $\chi^2(6)$ and $\chi^2(12)$ are 12.59 and 21.03, respectively.

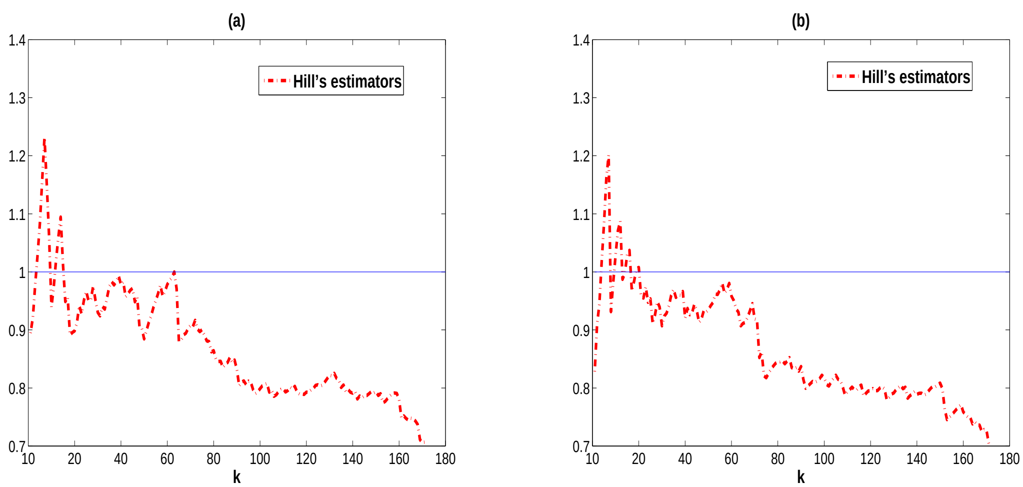

FIG 3. *(a) the Hill's estimators for η_t^2 in ARCH(3) model, and (b) the Hill's estimators for η_t^2 in ARCH(4) model.*

Acknowledgements. The first author's work is supported by the National Natural Science Foundation of China (No.10990012, 11021161 and 11371354), and the second author's work is supported by the National Natural Science Foundation of China (No.11201459).

APPENDIX

LEMMA A.1. *Suppose that Assumptions 2.1-2.4 hold. Then, (i) $\hat{\theta}_n \to \theta_0$ a.s. as $n \to \infty$; (ii) it follows that*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{\Sigma^{-1}}{2f(0)\sqrt{n}} \sum_{t=1}^{n} \frac{\xi_t}{h_t(\theta_0)} \frac{\partial h_t(\theta_0)}{\partial \theta} + o_p(1), \quad \text{(A.1)}$$

and it entails $\sqrt{n}(\hat{\theta}_n - \theta_0) \to_d N(0, [2f(0)]^{-2}\Sigma^{-1})$ as $n \to \infty$, where Σ is defined as in Theorem 2.1.

PROOF. Let $B_\eta(\theta) \in \Theta$ be an open neighborhood of θ with radius $\eta > 0$. We first verify the following three claims to prove (i):

$$(a) \quad E\left[\sup_{\theta\in\Theta} l_t(\theta)\right] < \infty;$$

$$(b) \quad E[l_t(\theta)] \text{ has a unique minimum at } \theta_0;$$

$$(c) \quad E\left[\sup_{\theta\in B_\eta(\theta^*)} |l_t(\theta) - l_t(\theta^*)|\right] \to 0 \text{ as } \eta \to 0,$$

where $l_t(\theta) = |\log \varepsilon_t^2 - \log h_t(\theta)|$. Clearly, claim (a) follows directly from Assumption 2.5(i)-(ii). For claim (b), by using the inequality $\mathbb{E}\,|X - a| \geq E\,|X - median(X)|$ for all random variable X and real number a, we can show that

$$\begin{aligned} E[l_t(\theta)] &= E\left[E\left(|z_t - \log[h_t(\theta)/h_t]|\,\middle|\mathcal{F}_{t-1}\right)\right] \\ &\geq E\left[E\left(|z_t|\middle|\mathcal{F}_{t-1}\right)\right] = E[l_t(\theta_0)], \end{aligned}$$

where the inequality holds since z_t has median 0 by Assumption 2.2, and the equation holds if and only if $\log[h_t(\theta)/h_t] = 0$ a.s., which implies that $\theta = \theta_0$ by Assumption 2.1(i). Moreover, by Taylor's expansion, triangle's inequality and Assumption 2.4(iii), it is straightforward to see that claim (c) holds. Now, based on claims (a)-(c), following the same argument as for Theorem 2.1 in Zhu and Ling (2011), we can show that (i) holds.

Next, we use the same argument as for Theorem 2.2 in Zhu and Ling (2011) to prove (ii). Let $H_n(u) = n\,[L_n(\theta_0 + u) - L_n(\theta_0)] \triangleq \sum_{t=1}^{n} A_t(u)$, where $u \in \Lambda \triangleq \{u : u+\theta_0 \in \Theta\}$. Denote $Z_{1t}(s) = I(z_t < s) - I(z_t > s)$ and $Z_{2t}(s) = I(z_t \leq s) - I(z_t \leq 0)$. Then, by Taylor's expansion and using the identity

$$|x - y| - |x| = -y[I(x > 0) - I(x < 0)] + 2\int_0^y [I(x \leq s) - I(x \leq 0)]ds$$

for $x \neq 0$, it follows that

$$A_t(u) = q_t(u)Z_{1t}(0) + 2\int_0^{q_t(u)} Z_{2t}(s)ds, \quad \text{(A.2)}$$

where $q_t(u) = q_{1t}(u) + q_{2t}(u)$ with

$$q_{1t}(u) = \frac{u'}{h_t}\frac{\partial h_t}{\partial \theta}(\theta_0), \quad q_{2t}(u) = \frac{u'}{2}\left[\frac{1}{h_t}\frac{\partial^2 h_t}{\partial\theta\partial\theta'}(\zeta^*) - \frac{1}{h_t^2}\frac{\partial h_t}{\partial\theta}\frac{\partial h_t}{\partial\theta'}(\zeta^*)\right]u,$$

and ζ^* lies between θ_0 and $\theta_0 + u$.

Furthermore, let $\mathcal{F}_t = \sigma(\eta_i; i \le t)$ and $W_t(u) = 2\int_0^{q_{1t}(u)} Z_{2t}(s)ds$. Since $Z_{1t}(0) = -\xi_t$, by (A.2) we have

$$\sum_{t=1}^{n} A_t(u) = (\sqrt{n}u)'\mathcal{S}_n(\theta_0) + \Pi_{1n}(u) + \Pi_{2n}(u) + \Pi_{3n}(u), \tag{A.3}$$

where

$$\begin{aligned}
\mathcal{S}_n(\theta_0) &= -\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left[\frac{1}{h_t}\frac{\partial h_t}{\partial\theta}(\theta_0)\right]\xi_t,\\
\Pi_{1n}(u) &= \sum_{t=1}^{n}\left\{W_t(u) - E[W_t(u)|\mathcal{F}_{t-1}]\right\},\\
\Pi_{2n}(u) &= \sum_{t=1}^{n} E[W_t(u)|\mathcal{F}_{t-1}],\\
\Pi_{3n}(u) &= -\sum_{t=1}^{n} q_{2t}(u)\xi_t + 2\sum_{t=1}^{n}\int_{q_{1t}(u)}^{q_t(u)} Z_{2t}(s)ds.
\end{aligned}$$

Let $u_n = \hat{\theta}_n - \theta_0$. By (i), Assumptions 2.1-2.4, and the same argument as for Lemmas 2.2-2.3 in Zhu and Ling (2011), we can show that $\Pi_{1n}(u_n) = o_p(\sqrt{n}\|u_n\| + n\|u_n\|^2)$, $\Pi_{2n}(u_n) = (\sqrt{n}u_n)'[f(0)\Sigma](\sqrt{n}u_n)$, and $\Pi_{3n}(u_n) = o_p(n\|u_n\|^2)$, where Σ is positive definite by Assumption 2.1(ii). Thus, by (A.3) and the same argument as Theorem 2.2 in Zhu and Ling (2011), it follows that (ii) holds. □

LEMMA A.2. *Suppose that Assumptions 2.1-2.4 hold. Then,*

$$\sqrt{n}\hat{\rho}_l = \sqrt{n}\rho_l - 2f(0)X_l'\sqrt{n}(\hat{\theta}_n - \theta_0) + o_p(1)$$

for any integer $l \ge 1$, where ρ_l is defined in the same way as $\hat{\rho}_l$ with ξ_t replacing $\hat{\xi}_t$, and X_l is defined as in Theorem 2.1.

PROOF. Rewrite

$$\begin{aligned}
\sqrt{n}\hat{\rho}_l - \sqrt{n}\rho_l &= \frac{1}{\sqrt{n}}\sum_{t=l+1}^{n}\hat{\xi}_{t-l}\left(\hat{\xi}_t - \xi_t\right) + \frac{1}{\sqrt{n}}\sum_{t=l+1}^{n}\xi_t\left(\hat{\xi}_{t-l} - \xi_{t-l}\right)\\
&\triangleq I_{1n} + I_{2n} \text{ say.}
\end{aligned}$$

Then, $I_{1n} = \Delta_{1n} + \Delta_{2n}$, where

$$\Delta_{1n} = \frac{1}{\sqrt{n}} \sum_{t=l+1}^{n} E\left[\hat{\xi}_{t-l}\left(\hat{\xi}_t - \xi_t\right)|\mathcal{F}_{t-1}\right],$$

$$\Delta_{2n} = \frac{1}{\sqrt{n}} \sum_{t=l+1}^{n} \left\{\hat{\xi}_{t-l}\left(\hat{\xi}_t - \xi_t\right) - E\left[\hat{\xi}_{t-l}\left(\hat{\xi}_t - \xi_t\right)|\mathcal{F}_{t-1}\right]\right\}.$$

We first consider Δ_{1n}. Let $u = \theta - \theta_0$, and $G(\cdot)$ and $g(\cdot)$ be the c.d.f. and p.d.f. of η_t, respectively. Note that by Taylor's expansion, we have

$$\eta_t(u+\theta_0) = \eta_t\sqrt{\frac{h_t}{h_t(u+\theta_0)}} = \eta_t\sqrt{\frac{h_t}{h_t + u'\partial h_t(\zeta^*)/\partial\theta}},$$

where ζ^* lies between θ_0 and $u+\theta_0$. Thus, by the double expectation and Taylor's expansion again, it follows that

$$\begin{aligned} & E\left[I(-1<\eta_t(u+\theta_0)<1) - I(-1<\eta_t<1)|\mathcal{F}_{t-1}\right] \\ &= \left[G\left(\sqrt{1+\frac{u'}{h_t}\frac{\partial h_t(\zeta^*)}{\partial\theta}}\right) - G(1)\right] + \left[G(-1) - G\left(-\sqrt{1+\frac{u'}{h_t}\frac{\partial h_t(\zeta^*)}{\partial\theta}}\right)\right] \\ &= \left[\frac{g(\zeta_{1t}^*)}{2\zeta_{1t}^*} - \frac{g(\zeta_{2t}^*)}{2\zeta_{2t}^*}\right]\frac{u'}{h_t}\frac{\partial h_t(\zeta^*)}{\partial\theta}, \end{aligned} \tag{A.4}$$

where ζ_{1t}^* lies between 1 and $\sqrt{1+(u'/h_t)\partial h_t(\zeta^*)/\partial\theta}$, and ζ_{2t}^* lies between -1 and $-\sqrt{1+(u'/h_t)\partial h_t(\zeta^*)/\partial\theta}$. Similarly, we can show that

$$E\left[I(\eta_t(u+\theta_0)>1) - I(\eta_t>1)|\mathcal{F}_{t-1}\right] = -\frac{g(\zeta_{1t}^*)}{2\zeta_{1t}^*}\frac{u'}{h_t}\frac{\partial h_t(\zeta^*)}{\partial\theta}, \tag{A.5}$$

$$E\left[I(\eta_t(u+\theta_0)<-1) - I(\eta_t<-1)|\mathcal{F}_{t-1}\right] = \frac{g(\zeta_{2t}^*)}{2\zeta_{2t}^*}\frac{u'}{h_t}\frac{\partial h_t(\zeta^*)}{\partial\theta}. \tag{A.6}$$

Let $w_t(u) = sgn(\eta_t^2(u+\theta_0)-1)$. Since $w_{t-l}(u) \in \mathcal{F}_{t-1}$, by (A.4)-(A.6), we know that

$$\begin{aligned} & \frac{1}{\sqrt{n}} \sum_{t=l+1}^{n} E\left\{w_{t-l}(u)\left[w_t(u) - w_t(0)\right]|\mathcal{F}_{t-1}\right\} \\ &= -\left\{\frac{1}{n}\sum_{t=l+1}^{n}\left[\frac{g(\zeta_{1t}^*)}{\zeta_{1t}^*} - \frac{g(\zeta_{2t}^*)}{\zeta_{2t}^*}\right]\frac{w_{t-l}(u)}{h_t}\frac{\partial h_t(\zeta^*)}{\partial\theta'}\right\}(\sqrt{n}u). \end{aligned} \tag{A.7}$$

Furthermore, for any $M>0$, by a similar argument as for (2.1), we can show that

$$\begin{aligned} \sup_{\sqrt{n}\|u\|\le M} \Bigg| & \frac{1}{n}\sum_{t=l+1}^{n}\left[\frac{g(\zeta_{1t}^*)}{\zeta_{1t}^*} - \frac{g(\zeta_{2t}^*)}{\zeta_{2t}^*}\right]\frac{w_{t-l}(u)}{h_t}\frac{\partial h_t(\zeta^*)}{\partial\theta'} \\ & - E\left[[g(1)+g(-1)]\frac{w_{t-l}(0)}{h_t}\frac{\partial h_t(\theta_0)}{\partial\theta'}\right]\Bigg| = o_p(1). \end{aligned} \tag{A.8}$$

Recall that $\hat{u}_n = \hat{\theta}_n - \theta_0$. Since $g(1) + g(-1) = 2f(0)$ and $\sqrt{n}\hat{u}_n = O_p(1)$ by Lemma A.1(ii), by (A.7)-(A.8), it follows that

$$\Delta_{1n} = \frac{1}{\sqrt{n}} \sum_{t=l+1}^{n} E\left\{w_{t-l}(\hat{u}_n)\left[w_t(\hat{u}_n) - w_t(0)\right]|\mathcal{F}_{t-1}\right\}$$

$$= -2f(0)X_l'(\sqrt{n}\hat{u}_n) + o_p(1). \tag{A.9}$$

Next, we consider Δ_{2n}. Since $\{\hat{\xi}_{t-l}(\hat{\xi}_t - \xi_t) - E[\hat{\xi}_{t-l}(\hat{\xi}_t - \xi_t)|\mathcal{F}_{t-1}]\}$ is a martingale difference sequence, it is not hard to see that

$$E[\Delta_{2n}^2] \le \frac{1}{n} \sum_{t=l+1}^{n} E\left[\hat{\xi}_{t-l}^2(\hat{\xi}_t - \xi_t)^2\right] \le \frac{1}{n} \sum_{t=l+1}^{n} E\left\{E\left[(\hat{\xi}_t - \xi_t)^2|\mathcal{F}_{t-1}\right]\right\} \to 0$$

as $n \to \infty$, where the last relation holds by the dominated convergence theorem. Thus, it follows that $\Delta_{2n} = o_p(1)$, which implies $I_{1n} = -2f(0)X_l'(\sqrt{n}\hat{u}_n) + o_p(1)$ by (A.9). Moreover, by a similar argument as for I_{1n}, we can show that $I_{2n} = o_p(1)$, and hence the conclusion holds. □

Proof of Theorem 2.1. First, by Lemma A.2, we have

$$\sqrt{n}\hat{\rho} = \sqrt{n}\rho - 2f(0)X\sqrt{n}(\hat{\theta}_n - \theta_0) + o_p(1),$$

where $\rho = (\rho_1, \cdots, \rho_M)'$. Next, by Lemma A.1, it follows that $\sqrt{n}\hat{\rho} = VZ_n + o_p(1)$, where

$$V = [I_M, -X\Sigma^{-1}] \text{ and } Z_n = \sqrt{n}\left[\rho', \frac{1}{n}\sum_{t=1}^{n} \frac{\xi_t}{h_t(\theta_0)} \frac{\partial h_t(\theta_0)}{\partial \theta'}\right]'.$$

Finally, the conclusion holds by the martingale central limit theorem. □

REFERENCES

[1] Berkes, I. and Horváth, L. (2004) The efficiency of the estimators of the parameters in GARCH processes. *Annals of Statistics* **32**, 633-655.

[2] Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**, 307-327.

[3] Bollerslev, T., Chou, R.Y. and Kroner, K.F. (1992). ARCH modeling in finance: A review of the theory and empirical evidence. *Journal of Econometrics* **52**, 5-59.

[4] Bougerol, P. and Picard, N. (1992) Stationarity of GARCH processes and of some non-negative time series. *Journal of Econometrics* **52**, 115-127.

[5] Box, G.E.P. and Pierce, D.A. (1970) Distribution of the residual autocorrelations in autoregressive integrated moving average time series models. *Journal of American Statistical Association* **65**, 1509-1526.

[6] Carbon, M. and Francq, C. (2011) Portmanteau goodness-of-fit test for asymmetric power GARCH models. *Austrian Journal of Statistics* **40**, 55-64.

[7] Ding, Z., Granger, C.W.J. and Engle, R.F. (1993) A long memory property of stock market returns and a new model. *Journal of Empirical Finance* **1**, 83-106.

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

[9] Escanciano, J.C. (2007) Joint and marginal specification tests for conditional mean and varnance models. *Journal of Econometrics* **143**, 74-87.

[10] Fan, J., Qi, L. and Xiu, D. (2013) Quasi maximum likelihood estimation of GARCH models with heavy-Tailed likelihoods. *Journal of Business and Economic Statistics.* forthcoming.

[11] Francq, C., Wintenberger, O. and Zakoïan, J-M. (2013) Garch models without positivity constraints: exponential or log garch? *Journal of Econometrics.* forthcoming.

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

[13] Francq, C. and Zakoïan, J.M. (2010) GARCH Models: Structure, Statistical Inference and Financial Applications. Wiley, Chichester, UK.

[14] Francq, C. and Zakoïan, J.M. (2013) Optimal predictions of powers of conditionally heteroscedastic processes. *Journal of the Royal Statistical Society* **B 75**, 345-367.

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

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

[17] Hamadeh, T. and 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.

[18] Hong, Y. and Lee, T.H. (2003) Diagnostic checking for adequacy of nonlinear time series models. *Econometric Theory* **19**, 1065-1121.

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

[20] Li, G. and Li, W.K. (2008) Least absolute deviation estimation for fractionally integrated autoregressive moving average time series models with conditional heteroscedasticity. *Biometrika* **95**, 399-414.

[21] Li, W.K. and 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.

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

[23] Ling, S. and Li, W.K. (1997) Diagnostic checking of nonlinear multivariate time series with multivariate arch errors. *Journal of Time Series Analysis* **18**, 447-464.

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

[25] Ling, S. and Tong, H. (2011) Score based goodness-of-fit tests for time series. *Statistica Sinica* **21**, 1807-1829.

[26] Linton, O., Pan, J. and Wang, H. (2010) Estimation for a non-stationary semi-strong GARCH(1,1) model with heavy-tailed errors. *Econometric Theory* **26**, 1-28.

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

[28] Peng, L. and Yao, Q.W. (2003) Least absolute deviations estimation for ARCH and GARCH models. *Biometrika* **90**, 967-975.

[29] Rachev, S.T. (2003) Handbook of Heavy Tailed Distributions in Finance. Elsevier/North-Holland.

[30] Zhu, K. (2011) On the LAD estimation and likelihood ratio test for time series models. Thesis Dissertion. Hong Kong University of Science and Technology.

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

Chinese Academy of Sciences
Institute of Applied Mathematics
Haidian District, Zhongguancun
Bei Jing, China
E-mail: mchen@amss.ac.cn
kzhu@amss.ac.cn