Biometrika (2008), **95**, 2, *pp*. 399–414
doi: 10.1093/biomet/asn014

Least absolute deviation estimation for fractionally integrated autoregressive moving average time series models with conditional heteroscedasticity

BY GUODONG LI AND WAI KEUNG LI

Department of Statistics and Actuarial Science,
The University of Hong Kong, Pokfulam Road, Hong Kong
ligd@hku.hk hrntlwk@hku.hk

SUMMARY

We consider a unified least absolute deviation estimator for stationary and nonstationary fractionally integrated autoregressive moving average models with conditional heteroscedasticity. Its asymptotic normality is established when the second moments of errors and innovations are finite. Several other alternative estimators are also discussed and are shown to be less efficient and less robust than the proposed approach. A diagnostic tool, consisting of two portmanteau tests, is designed to check whether or not the estimated models are adequate. The simulation experiments give further support to our model and the results for the absolute returns of the Dow Jones Industrial Average Index daily closing price demonstrate their usefulness in modelling time series exhibiting the features of long memory, conditional heteroscedasticity and heavy tails.

Some key words: ARFIMA; Conditional heteroscedasticity; Heavy tail; GARCH; Least absolute deviation; Long memory.

1. INTRODUCTION

The fractionally integrated autoregressive moving average, ARFIMA or fractional ARIMA, process was proposed by McLeod & Hipel (1978), Granger & Joyeux (1980) and Hosking (1981), and is one of the most popular models for explaining the phenomenon of long memory in diverse fields of statistical application, especially in the field of finance; see Robinson (2003). On the other hand, since Engle (1982), it has been accepted that many financial time series have a time-varying conditional variance. In fact, some financial time series may exhibit the features of both long memory and time-varying conditional variance; these include the consumer price index inflation series in Baillie et al. (1996), the Swiss Euromarket interest rate in Hauser & Hunst (2001) and the absolute log return sequences in Tsay (2002). The generalized autoregressive conditional heteroscedasticity, GARCH, models (Bollerslev, 1986) are usually considered for modelling the phenomenon of time-varying conditional variance and it is natural to consider the ARFIMA–GARCH model defined as follows:

$$\phi(B)(1-B)^d Y_t = \psi(B)\varepsilon_t, \tag{1}$$

$$\varepsilon_t = u_t h_t^{1/2}, \quad h_t = \alpha_0 + \sum_{i=1}^{r} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{s} \beta_j h_{t-j}, \tag{2}$$

where $\alpha_0 > 0$, $\alpha_i \geqslant 0$ $(i = 1, \ldots, r)$ $\beta_j \geqslant 0$ $(j = 1, \ldots, s)$ B is the backward shift operator, $\phi(B) = 1 - \sum_{i=1}^{p} \phi_i B^i$, $\psi(B) = 1 + \sum_{j=1}^{q} \psi_j B^j$ and $(1 - B)^d = \sum_{k=0}^{\infty}[\Gamma(k - d)/\{\Gamma(k + 1)\Gamma(-d)\}]B^k$. Note that $\{\varepsilon_t\}$ is the error sequence and members of the innovation sequence $\{u_t\}$ are independently and identically distributed.

The parameter d describes the extent of long memory generated by the ARFIMA–GARCH model and is called the memory parameter. The process generated by models (1) and (2) is short memory, long memory, stationary or nonstationary, respectively, when $d \in (-0{\cdot}5, 0)$, $(0, \infty)$, $(-0{\cdot}5, 0{\cdot}5)$ or $(0{\cdot}5, \infty)$; see Ling & Li (1997). It is important to estimate d as well as the other parameters. When the errors $\{\varepsilon_t\}$ are independent, many procedures, including time domain and frequency domain methods, have been developed for model (1); see Bhardwaj & Swanson (2006) and references therein. Under the normality of u_t, Ling & Li (1997) established the asymptotic normality of the maximum likelihood estimators. Beran & Feng (2001) considered local polynomial estimation of semiparameteric models with an ARFIMA–GARCH error. However, both papers required the condition $E(\varepsilon_t^4) < \infty$ resulting in a more restricted parameter space for model (2); see Ling (2007). Francq & Zakoian (2004) discussed the asymptotic normality of the Gaussian quasi-maximum likelihood estimators of the autoregressive moving average models with GARCH errors, and finiteness of $E(\varepsilon_t^4)$ was needed although finiteness of only a smaller-order moment was required for pure GARCH models. The condition $E(\varepsilon_t^4) < \infty$ seems unavoidable for such estimation approaches and it is necessary to develop a different method for ARFIMA–GARCH models with a less restricted parameter space.

Furthermore, recent empirical evidence has increasingly shown that some financial time series may be so heavy-tailed that the fourth moment of the innovation u_t is infinite; see Mittnik et al. (1998) and Mikosch & Starica (2000). Most existing estimation methods for fractional ARIMA and ARFIMA–GARCH models are sensitive to outliers. Haldrup & Nielsen (2007) showed, by simulation experiments, that some commonly used estimators of the fractional ARIMA models were not robust to outliers and the estimators of memory parameters may be biased. Granger et al. (1999) applied the fractional ARIMA models to several price indices, and Ling & Li (1997) fitted a ARFIMA–GARCH model to the Hong Kong Heng Seng Index. In order to obtain a good estimator, outliers were removed before estimation in both papers. However, it is well known that these outliers may include useful information; see Embrechts et al. (1997). For pure GARCH processes with $E(u_t^4) = \infty$, Hall & Yao (2003) showed that Gaussian quasi-maximum likelihood estimation may not be asymptotically normal and the convergence rate is slower than the standard rate of $n^{1/2}$. The same may happen for models (1) and (2) with infinite fourth moment for u_t. Peng & Yao (2003) constructed three least absolute deviation estimators for the pure GARCH models and established their asymptotic normality under only finite second moments of ε_t and u_t. This approach may be useful for providing robust estimation for ARFIMA–GARCH models.

2. THE LEAST ABSOLUTE DEVIATION ESTIMATION

Let $l = p + q + r + s + 2$ and denote the parameter vector of models (1) and (2) by $\lambda = (\gamma', \delta')'$, where $\gamma = (d, \phi_1, \ldots, \phi_p, \psi_1, \ldots, \psi_q)'$, $\delta = (\alpha_0, \alpha_1, \ldots, \alpha_r, \beta_1, \ldots, \beta_s)'$ and λ is an l-dimensional vector. Assume that the parameter space Θ is a compact set of R^l, the true parameter vector λ_0 is an interior point of Θ and each λ in Θ satisfies the following two assumptions.

Assumption 1. We assume that $\alpha_i > 0$, $i = 0, 1, \ldots, r$, $\beta_j > 0$, $j = 1, \ldots, s$, $E(\varepsilon_t^2) < \infty$, and the polynomials $\sum_{i=1}^{r} \alpha_i z^i$ and $1 - \sum_{j=1}^{s} \beta_j z^j$ have no common root.

Assumption 2. We assume that $d > -0{\cdot}5$ and $d \notin J - 0{\cdot}5$, where J is the collection of all positive integers. All roots of the polynomials $\phi(z) = 1 - \sum_{i=1}^{p} \phi_i z^i$ and $\psi(z) = 1 + \sum_{j=1}^{q} \psi_j z^j$ lie outside the unit circle and there is no common factor between $\phi(z)$ and $\psi(z)$.

Denote Eu_t^2 by σ^2. Then the condition $\sigma^2 \sum_{i=1}^{r} \alpha_i + \sum_{j=1}^{s} \beta_j < 1$ is necessary and sufficient for $\{\varepsilon_t\}$ in (2) to exist as a unique strictly stationary sequence with a finite second moment Li & Li (2005). The cases with $d \in J - 0{\cdot}5$ are very complicated (Beran, 1995; Ling & Li, 1997). Hence we exclude them from the parameter space and leave them for possible future research.

Under Assumptions 1 and 2, model (1) is invertible and then ε_t can be written as

$$\varepsilon_t = \phi(B)\psi^{-1}(B) \sum_{k=0}^{\infty} \frac{\Gamma(k-d)}{\Gamma(k+1)\Gamma(-d)} Y_{t-k}.$$

When the true parameter vector in the above equation is replaced by $\lambda \in \Theta$, ε_t can be considered as a function on Θ and hence can be denoted by $\varepsilon_t(\gamma)$ or $\varepsilon_t(\lambda)$. Similarly, we can define the function $h_t(\lambda)$ by the iterative equation $\alpha_0 + \sum_{i=1}^{r} \alpha_i \varepsilon_{t-i}^2(\lambda) + \sum_{j=1}^{s} \beta_j h_{t-j}(\lambda)$. These two functions depend on the infinite past values of the sequences $\{Y_t\}$ and $\{\varepsilon_t\}$. They are unobservable in real applications and some initial values are needed. For simplicity, we set the initial values of $\{Y_t\}$ and $\{\varepsilon_t\}$ to zero and replace h_t and ε_t^2 for $t \leqslant 0$ by $(1/n)\sum_{i=1}^{n} \varepsilon_i^2$. This will not affect the results in the following derivation; see Bollerslev (1986) and Ling & Li (1997). Furthermore, to save space and forestall confusion, we denote $\varepsilon_t(\lambda_0)$, $h_t(\lambda_0)$, $\partial\varepsilon_t(\lambda_0)/\partial\gamma$ and $\partial h_t(\lambda_0)/\partial\lambda$ respectively by ε_t, h_t, $\partial\varepsilon_t/\partial\gamma$ and $\partial h_t/\partial\lambda$.

There are two different approaches to defining the least absolute deviation estimators: one is based on the sum of absolute errors and the other on Laplace quasi-maximum likelihood estimation. These two methods are consistent with each other for linear models, but they may be totally different for nonlinear models. For the difference between these two methods, see Peng & Yao (2003) and Berkes & Horvath (2004) for the least absolute deviation estimators of pure GARCH models. We first consider the sum of absolute errors for models (1) and (2). Peng & Yao (2003) designed three least absolute deviation estimators by rewriting the GARCH model in the forms of regression. A natural extension of the best of these three to the ARFIMA–GARCH model is given by

$$\hat{\lambda}_{\mathrm{PY}} = \arg\min_{\lambda\in\Theta} \sum_{t=1}^{n} |\log \varepsilon_t^2(\lambda) - \log h_t(\lambda)|.$$

However, in order to derive its asymptotic normality, the condition $E(1/|u_t|) < \infty$ is required and this excludes many familiar continuous distributions such as the normal distribution and Student's t-distributions. The other two estimators considered require a finite fourth moment condition. Hence, this paper does not pursue this direction and focuses only on the method of Laplace quasi-maximum likelihood estimation. By temporarily assuming a Laplace distribution with density $g(x) = 0{\cdot}5\, e^{-|x|}$ for the innovation u_t, we can define the least absolute deviation estimator as

$$\hat{\lambda}_n = \arg\min_{\lambda\in\Theta} L_n(\lambda),\ L_n(\lambda) = \sum_{t=1}^{n} \left[\frac{|\varepsilon_t(\lambda)|}{\surd\{h_t(\lambda)\}} + \frac{1}{2}\log h_t(\lambda) \right].$$

To investigate the asymptotic distribution of $\hat{\lambda}_n$, we need another assumption.

Assumption 3. The median of u_t is equal to zero, $E|u_t| = 1$, $E(u_t^2) = \sigma^2 < \infty$ and the probability density function $f(x)$ of u_t is continuous at the origin.

When $|d| < 0{\cdot}5$, under Assumptions 1 and 2, the processes $\{Y_t\}$ and $\{\varepsilon_t\}$ generated by models (1) and (2) are strictly stationary and ergodic with a finite second moment; see Ling & Li (1997).

Let $\mu = Eu_t$ and $\sigma^2_{|u|} = \text{var}(|u_t|)$. We define the matrices

$$\Omega_\varepsilon = E\left(\frac{1}{h_t}\frac{\partial\varepsilon_t}{\partial\gamma}\frac{\partial\varepsilon_t}{\partial\gamma'}\right), \quad \Omega_1 = \begin{pmatrix}\Omega_\varepsilon & 0\\ 0 & 0\end{pmatrix}, \quad \Omega_2 = E\left(\frac{1}{4h_t^2}\frac{\partial h_t}{\partial\lambda}\frac{\partial h_t}{\partial\lambda'}\right),$$

$$\Omega_3 = E\left\{\frac{1}{2h_t^{3/2}}\left(\frac{\partial\varepsilon_t}{\partial\lambda}\frac{\partial h_t}{\partial\lambda'} + \frac{\partial h_t}{\partial\lambda}\frac{\partial\varepsilon_t}{\partial\lambda'}\right)\right\}, \quad \Sigma_1 = \Omega_1 + \sigma^2_{|u|}\Omega_2 - \mu\Omega_3,$$

$$\Sigma_2 = f(0)\Omega_1 + 0{\cdot}5\Omega_2,$$

where $f(0)$ is the value of the probability density function $f(x)$ of u_t evaluated at zero, the matrix Ω_ε is $(p+q+1)\times(p+q+1)$ and other matrices are $l \times l$. Following the method in Francq & Zakoian (2004), under Assumptions 1 and 2, we can show that the matrices Σ_1 and Σ_2 are positive definite.

THEOREM 1. *Suppose that $\{Y_t\}$ are generated by models (1) and (2). Under Assumptions 1–3, if $|d| < 0{\cdot}5$, then there exists a sequence of local minimizers $\{\hat\lambda_n\}$ of $L_n(\lambda)$ such that*

$$\sqrt{n}(\hat\lambda_n - \lambda_0) \to N(0, 0{\cdot}25\Sigma_2^{-1}\Sigma_1\Sigma_2^{-1})$$

in distribution as $n \to \infty$.

Following a suggestion from a referee, we have not assumed that $E(u_t) = 0$ in the above theorem and the quantity $\mu = E(u_t)$ is included in the covariance matrix as a parameter.

For a complicated ARFIMA–GARCH model with many parameters, we may encounter computational difficulty in finding the least absolute deviation estimator $\hat\lambda_n$. If the innovation u_t is further assumed to have a symmetric distribution, then $\mu = 0$ and Ω_2 is a block-diagonal matrix since $E\{h_t^{-2}(\partial h_t/\partial\gamma)(\partial h_t/\partial\delta')\}$ is equal to zero. Hence, the matrices Σ_1, Σ_2 and $\Sigma_2^{-1}\Sigma_1\Sigma_2^{-1}$ are all block-diagonal. This implies that we can separately minimize the score function $L_n(\lambda)$ with respect to γ and δ without incurring an asymptotic loss of efficiency.

For the general case, a two-stage estimation approach seems more plausible in which we first apply least absolute deviation estimation to the ARFIMA part to find a minimizer $\tilde\gamma$, and then apply the approach of Peng & Yao (2003) to the residuals $\{\varepsilon_t(\tilde\gamma)\}$ to find the estimator $\tilde\delta$. Alternatively, we may obtain $\tilde\gamma_{\text{LS}} = \arg\min \sum_{t=1}^n \varepsilon_t^2(\gamma)$ by least squares. However, to derive the asymptotic normality of $\tilde\gamma_{\text{LS}}$, we need to show that $n^{-1/2}\sum_{t=1}^n \varepsilon_t(\partial\varepsilon_t/\partial\gamma)$ converges in distribution to a normal distribution with a finite variance $E\{\varepsilon_t^2(\partial\varepsilon_t/\partial\gamma)(\partial\varepsilon_t/\partial\gamma')\}$. If $\{\varepsilon_t\}$ are independent, then the condition $E\varepsilon_t^2 < \infty$ is enough for the finiteness of $E\{\varepsilon_t^2(\partial\varepsilon_t/\partial\gamma)(\partial\varepsilon_t/\partial\gamma')\}$, but the assumption $E\varepsilon_t^4 < \infty$ is unavoidable for the ARFIMA–GARCH case. The simulation results in § 4 suggest that both two-stage estimators are inferior to the least absolute deviation estimator $\hat\lambda_n$. This is not surprising since, by ignoring the full model, two-stage methods should have lower efficiency. In real applications, the two-stage estimator can be used as an initial estimator and then we can use the local quadratic approximation (Fan & Li, 2001) to minimize

$$\sum_{t=1}^n \left[\frac{\varepsilon_t^2(\gamma)}{|\varepsilon_t(\gamma^{(m)})|\sqrt{\{h_t(\gamma,\delta)\}}} + \frac{1}{2}\log h_t(\gamma,\delta)\right]$$

iteratively, where $\gamma^{(m)}$ is the minimizer in the mth iteration and $\gamma^{(0)} = \tilde\gamma$. The above score function is sufficiently smooth and algorithms such as Newton–Raphson can be employed for the optimization.

When $d > 0{\cdot}5$, the process $\{\varepsilon_t\}$ generated by (2) is still stationary. However, the process $\{Y_t\}$ generated by models (1) and (2) is nonstationary.

Let $d = m + d_f$, where $|d_f| < 0{\cdot}5$ and m is a positive integer. If $U_t = (1-B)^m Y_t$, then U_t follows the model $\phi(B)(1-B)^{d_f}U_t = \psi(B)\varepsilon_t$. This means that, after mth-order differencing, the nonstationary process $\{Y_t\}$ will be transformed to a stationary ARFIMA(p, d_f, q)–GARCH(r, s) process $\{U_t\}$. We next consider the asymptotic behaviour of the least absolute deviation estimation for the process $\{U_t\}$.

Let $\gamma^* = (d_f, \phi_1, \ldots, \phi_p, \psi_1, \ldots, \psi_q)'$ and $\lambda^* = (\gamma^{*\prime}, \delta')'$, where the first elements d in both λ and γ are replaced by d_f. Denote by Σ_1^* and Σ_2^* the corresponding matrices associated with $\{U_t\}$ instead of $\{Y_t\}$. By an argument similar to that in Beran (1995) and Ling & Li (1997), it can be shown that $\Sigma_1 = \Sigma_1^*$ and $\Sigma_2 = \Sigma_2^*$. Let $L_n^*(\lambda)$ be the corresponding score function. Then, by Theorem 1, there exists a sequence of local minimizers $\{\hat{\lambda}_n^*\}$ of $L_n^*(\lambda)$ such that $\sqrt{n}(\hat{\lambda}_n^* - \lambda_0) \to N(0, 0{\cdot}25\Sigma_2^{-1}\Sigma_1\Sigma_2^{-1})$ in distribution as $n \to \infty$.

Let $\hat{\lambda}_n = \hat{\lambda}_n^* + (m, 0, \ldots, 0)'$. Then $\hat{\lambda}_n$ is a local minimizer of $L_n(\lambda)$ and hence we have the following results for the nonstationary ARFIMA–GARCH models.

THEOREM 2. *Suppose that $\{Y_t\}$ are generated by models (1) and (2). Under Assumptions* 1–3, *if $d > 0{\cdot}5$, then there exists a sequence of local minimizers $\{\hat{\lambda}_n\}$ of $L_n(\lambda)$ such that*

$$\sqrt{n}(\hat{\lambda}_n - \lambda_0) \to N\left(0, 0{\cdot}25\Sigma_2^{-1}\Sigma_1\Sigma_2^{-1}\right)$$

in distribution as $n \to \infty$, where the matrices Σ_1 and Σ_2 are given as in Theorem 1.

The fractional ARIMA processes may sometimes include an unknown mean μ_Y, leading to the fractional ARIMA form

$$\phi(B)(1-B)^{d_f}\{(1-B)^m Y_t - \mu_Y\},$$

where $|d_f| < 0{\cdot}5$ and $m \geqslant 0$ is an integer. We follow Beran (1995) and Ling & Li (1997) in dealing with this case. Let $U_t = (1-B)^m Y_t$ and $\hat{U} = (n-m)^{-1}\sum_{t=m+1}^n U_t$. Then $\hat{U}$ is a consistent estimator of μ_Y. We can centre the sequence $\{U_t\}$ on $\hat{U}$ and then the methodology introduced before can be used. The simulation results in § 4 show that the estimators obtained by this method are very similar to those obtained when the mean is known.

3. TWO PORTMANTEAU TESTS

This section constructs two portmanteau tests for checking whether or not the fitted ARFIMA–GARCH models in the previous section are adequate. One test is based on the residual autocorrelations and the other is based on the absolute residual autocorrelations.

Let $\hat{\varepsilon}_t$ and $\hat{h}_t$ be the corresponding values when the parameter vector λ in functions $\varepsilon_t(\lambda)$ and $h_t(\lambda)$ is replaced by $\hat{\lambda}_n$, the least absolute deviation estimator from § 2. From the proof of Theorem 1, up to $o_p(1)$,

$$\sqrt{n}(\hat{\lambda}_n - \lambda_0) \simeq \frac{1}{2}n^{-1/2}\Sigma_2^{-1}\sum_{t=1}^n\left\{\frac{1}{2}(|u_t|-1)h_t^{-1}\frac{\partial h_t}{\partial\lambda} - \mathrm{sgn}(u_t)h_t^{-1/2}\frac{\partial\varepsilon_t}{\partial\lambda}\right\}. \quad (3)$$

Note that $\{\hat{\varepsilon}_t/\hat{h}_t^{1/2}\}$ is the residual sequence. Then the lag-k residual autocorrelation is

$$\hat{r}_k = \frac{\sum_{t=k+1}^n\left(\hat{\varepsilon}_t/\hat{h}_t^{1/2} - \hat{\mu}_n\right)\left(\hat{\varepsilon}_{t-k}/\hat{h}_{t-k}^{1/2} - \hat{\mu}_n\right)}{\sum_{t=1}^n\left(\hat{\varepsilon}_t/\hat{h}_t^{1/2} - \hat{\mu}_n\right)^2},$$

where $\hat{\mu}_n = n^{-1}\sum_{t=1}^{n}\hat{\varepsilon}_t/\hat{h}_t^{1/2}$, and the lag-$k$ absolute residual autocorrelation is

$$\hat{\rho}_k = \frac{\sum_{t=k+1}^{n}\left(|\hat{\varepsilon}_t|/\hat{h}_t^{1/2}-1\right)\left(|\hat{\varepsilon}_{t-k}|/\hat{h}_{t-k}^{1/2}-1\right)}{\sum_{t=1}^{n}\left(|\hat{\varepsilon}_t|/\hat{h}_t^{1/2}-1\right)^2}.$$

We next consider the asymptotic distributions of the first M residual autocorrelations and absolute residual autocorrelations.

Let $\hat{R} = (\hat{r}_1, \ldots, \hat{r}_M)'$ and $\kappa = E\{u_t(|u_t|-1)\}$. Let $X = (X_1, \ldots, X_M)$, $Z = (Z_1, \ldots, Z_M)$ and $\Omega_4 = X'\Sigma_2^{-1}Z + Z'\Sigma_2^{-1}X$, where $X_k = E\{(u_{t-k}-\mu)h_t^{-1/2}(\partial\varepsilon_t/\partial\lambda)\}$ and $Z_k = E\{(u_{t-k}-\mu)h_t^{-1}(\partial h_t/\partial\lambda)\}$ with $k = 1, \ldots, M$. Along the lines of Li (1992), by the approximation in (3), Taylor expansion, the central limit theorem and the Mann–Wald device, we can show that

$$\sqrt{n}\hat{R} \to N(0, V_1),$$

in distribution as $n \to \infty$, where

$$V_1 = I - \frac{1}{\sigma^4}X'\Sigma_2^{-1}(\Sigma_2 - 0{\cdot}25\Sigma_1)\Sigma_2^{-1}X + \frac{\kappa}{4\sigma^4}\Omega_4.$$

When the innovation u_t is symmetrically distributed, the quantity κ is equal to zero and the last term in the matrix V_1 disappears.

Let $\hat{\rho} = (\hat{\rho}_1, \ldots, \hat{\rho}_M)'$, $\hat{C} = (\hat{C}_1, \ldots, \hat{C}_M)'$ and $C = (C_1, \ldots, C_M)'$, where

$$\hat{C}_k = \frac{1}{n}\sum_{t=k+1}^{n}\left(\frac{|\hat{\varepsilon}_t|}{\hat{h}_t^{1/2}}-1\right)\left(\frac{|\hat{\varepsilon}_{t-k}|}{\hat{h}_{t-k}^{1/2}}-1\right), \quad k = 1, \ldots, M,$$

and C_k is the corresponding value when $\hat{\lambda}_n$ in $\hat{C}_k$ is replaced by the true parameter vector λ_0. Let $Z^* = (Z_1^*, \ldots, Z_M^*)$ and $\Omega_5 = H'\Sigma_2^{-1}Z^* + Z^{*\prime}\Sigma_2^{-1}H$, where $Z_k^* = E\{0{\cdot}5(|u_{t-k}|-1)h_t^{-1/2}(\partial\varepsilon_t/\partial\lambda)\}$ with $k = 1, \ldots, M$. We can show that $n^{-1}\sum(|\hat{\varepsilon}_t|/\hat{h}_t^{1/2}-1)^2 = \sigma_{|u|}^2 + o_p(1)$, and then it is sufficient to derive the asymptotic distributions of $\hat{C}$. However, the vector $\hat{C}$ is a function of $\hat{\lambda}_n$ and this function is not smooth. The classical Taylor expansion cannot be used here and this is just the main difficulty in deriving the asymptotic distribution of $\hat{\rho}$. Fortunately, by the inequalities in the Appendix, we can show that

$$\hat{C} = C - H'(\hat{\lambda}_n - \lambda_0) + o_p(n^{-1/2}), \tag{4}$$

where $H = (H_1, \ldots, H_M)$ and $H_k = E\{0{\cdot}5(|u_{t-k}|-1)h_t^{-1}(\partial h_t/\partial\lambda)\}$ with $k = 1, \ldots, M$. The above equation plays the same role as Taylor expansion in Li (1992) and Li & Mak (1994). Hence, as in Li & Li (2005), by the approximations in (3) and (4), the central limit theorem and the Mann–Wald device, we can obtain that

$$\sqrt{n}\hat{\rho} \to N(0, V_2)$$

in distribution as $n \to \infty$, where

$$V_2 = I - \frac{1}{\sigma_{|u|}^4}H'\Sigma_2^{-1}\left(\sigma_{|u|}^2\Sigma_2 - 0{\cdot}25\Sigma_1\right)\Sigma_2^{-1}H + \frac{\mu}{\sigma_{|u|}^4}\Omega_5.$$

Note that $\mu = 0$ when the innovation u_t is symmetrically distributed.

Based on the asymptotic normality of $\hat{R}$ and $\hat{\rho}$, we can construct two portmanteau tests,

$$Q_r(M) = n\hat{R}'V_1^{-1}\hat{R}, \quad Q_a(M) = n\hat{\rho}'V_2^{-1}\hat{\rho}.$$

It can be shown that, if the ARFIMA–GARCH model in the previous section is correctly specified, the quantities $Q_r(M)$ and $Q_a(M)$ will be asymptotically distributed as $\chi^2(M)$. In practice, we can obtain the exact values of μ, σ^2, $\sigma^2_{|u|}$ and $f(0)$ in the definitions of matrices V_1 and V_2 if the distribution of u_t is known. Otherwise, we can use $n^{-1}\sum \hat{\varepsilon}_t/\hat{h}_t^{1/2}$ to replace μ, $n^{-1}\sum \hat{\varepsilon}_t^2/\hat{h}_t$ to replace σ^2 and $n^{-1}\sum(|\hat{\varepsilon}_t|/\hat{h}_t^{1/2}-1)^2$ to replace $\sigma^2_{|u|}$. For $f(0)$, the sequence $\{\hat{\varepsilon}_t/\hat{h}_t^{1/2}\}$ is first supposed to be independently and identically distributed and then some nonparametric method, such as kernel estimation, can be applied to fit the density function $\hat{f}(x)$. Finally, we can use $\hat{f}(0)$ to replace $f(0)$. The entries of X, Z, H, Z^*, Ω_ε, Ω_2 and Ω_3 can be replaced by the corresponding sample averages, as in Li & Mak (1994). Tse & Zuo (1997) considered the optimal choice of M for portmanteau tests proposed in Li & Mak (1994).

The tests $Q_r(M)$ and $Q_a(M)$ should be useful in checking whether or not the fitted ARFIMA–GARCH models in § 2 are adequate, and the simulation results in § 4 give further support to this fact. It can be seen that $Q_r(M)$ is not sensitive to the misspecification in conditional variances while $Q_a(M)$ is not sensitive to the misspecification in conditional means. In fact, Wong & Ling (2005) observed the same phenomenon for the residual autocorrelations and squared residual autocorrelations, and proposed a combined portmanteau test based on these two types of autocorrelation. Hence, it is also of interest to construct a combined portmanteau test, based on the asymptotic joint distribution of residual autocorrelations and absolute residual autocorrelations. However, this method is more complicated and we leave it for future research.

4. Simulation experiments

4·1. *Performance of the least absolute deviation estimation*

When the innovation u_t is normally distributed, Ling & Li (1997) considered the maximum likelihood estimator of models (1) and (2),

$$\hat{\lambda}_{\text{MLE}} = \arg\min_{\lambda\in\Theta} \sum_{t=1}^{n} \left\{ \frac{\varepsilon_t^2(\lambda)}{h_t(\lambda)} + \log h_t(\lambda) \right\},$$

and its asymptotic normality was also obtained. This estimator can still be used when the normality of u_t is violated, and is the so-called Gaussian quasi-maximum likelihood estimator. Note that the proof in Ling & Li (1997) only needs the conditions $E(\varepsilon_t^4) < \infty$ and $E(u_t^4) < \infty$, and hence $\hat{\lambda}_{\text{MLE}}$ is still asymptotically normal under these conditions.

The first experiment compares the least absolute deviation estimator $\hat{\lambda}_n$ in § 2 with the Gaussian quasi-maximum likelihood estimator $\hat{\lambda}_{\text{MLE}}$ and the following ARFIMA(0,d,0)–GARCH(1,1) process was involved:

$$(1-B)^d Y_t = \varepsilon_t, \quad \varepsilon_t = u_t h_t^{1/2}, \quad h_t = 0{\cdot}5 + 0{\cdot}2\varepsilon_{t-1}^2 + 0{\cdot}7h_{t-1}, \tag{5}$$

where u_t followed a standard normal distribution or a Student's t-distribution with 3 or 5 degrees of freedom. These three distributions were always employed for the innovation u_t except in the next experiment, and were first standardized to have mean 0 and variance 1. The memory parameter d was chosen to be $-0{\cdot}3$ or $0{\cdot}3$ for the stationary case and $d = 0{\cdot}7$ or $1{\cdot}3$ for the nonstationary case. For each combination of innovation distributions and memory parameters, we considered the sample size $n = 600$ and drew 400 independent replications. The subroutine DBCPOL in the IMSL library was employed to perform an exhaustive search for $\hat{\lambda}_{\text{MLE}}$ and $\hat{\lambda}_n$ at the same time. We set initially the value of the parameter d to zero and the parameters in the conditional variance, α_0, α_1 and β_1, to 0·1. The subroutine DBCPOL was also used for optimization in the following experiments and the real example. Since the values of parameters α_0 and α_1

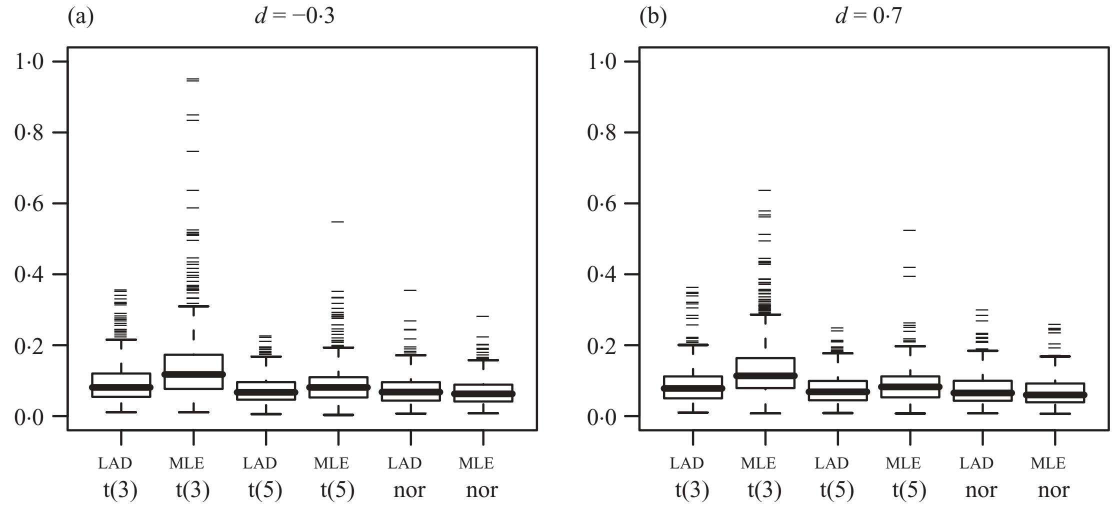

Fig. 1. Boxplots for the average absolute errors of $\hat{\lambda}_n$ and $\hat{\lambda}_{\text{MLE}}$; for (a) a stationary case with $d = -0{\cdot}3$, (b) a nonstationary case with $d = 0{\cdot}7$. Labels $t(3)$, $t(5)$ or 'nor' indicate that the innovation u_t has the t_3, t_5 or $N(0, 1)$ distribution, respectively.

fitted by the least absolute deviation approach are different from 0·5 and 0·2 by a common factor as in Peng & Yao (2003), we define the average absolute error as

$$\tfrac{1}{3}(|\hat{\alpha}_1/\hat{\alpha}_0 - 0{\cdot}4| + |\hat{\beta}_1 - 0{\cdot}2| + |\hat{d} - d|),$$

which can be used to compare the performance of $\hat{\lambda}_n$ with $\hat{\lambda}_{\text{MLE}}$.

Figure 1 displays the boxplots for the average absolute error for one stationary case, $d = -0{\cdot}3$, and one nonstationary case, $d = 0{\cdot}7$. The results for $d = 0{\cdot}3$ and $d = 1{\cdot}3$ were very similar to those in Fig. 1(a) and (b), respectively. There are a few very large values of average absolute errors for $\hat{\lambda}_{\text{MLE}}$ when the errors are distributed as t_3. For convenience of presentation, we have removed these outliers from the figures. The least absolute deviation estimator $\hat{\lambda}_n$ is much superior when $u_t \sim t_3$. This may reflect the fact that the heavier the tails, the slower the convergence rate of the Gaussian quasi-maximum likelihood estimator; see Hall & Yao (2003). Note that the t_5 distribution has a finite fourth moment so that $\hat{\lambda}_{\text{MLE}}$ will enjoy the standard $n^{1/2}$ convergence rate. For this case, $\hat{\lambda}_n$ also performs better. When the error is normally distributed, of course, $\hat{\lambda}_{\text{MLE}}$ is the better choice, but the performance of $\hat{\lambda}_n$ is comparable.

As suggested by a referee, we also conduct an experiment to compare $\hat{\lambda}_n$ with two two-stage estimators. For simplicity, an AR(1)–ARCH(1) model is employed to generate the realizations:

$$Y_t = 0{\cdot}3Y_{t-1} + \varepsilon_t, \quad \varepsilon_t = u_t(0{\cdot}5 + 0{\cdot}2\varepsilon_{t-1}^2)^{1/2}.$$

The innovation sequence $\{u_t\}$ come from a mixed t-distribution, taking the values of $|t_{f_1}|$ and $-|t_{f_2}|$ respectively with probability 0·5. Four different distributions of u_t are considered in this experiment with (f_1, f_2) respectively equal to (3, 3), (5, 5), (3, 5) and (9, 5). Note that the median of u_t is zero, but when $f_1 \neq f_2$, u_t is asymmetrically distributed with $E(u_t) \neq 0$. The following two-stage estimation method, 2LAD, was employed:

$$(\hat{\alpha}_0, \hat{\alpha}_1)' = \arg\min \sum_{t=1}^{n} |\log \varepsilon_t^2(\hat{\phi}) - \log h_t(\hat{\phi}, \alpha_0, \alpha_1)|,$$

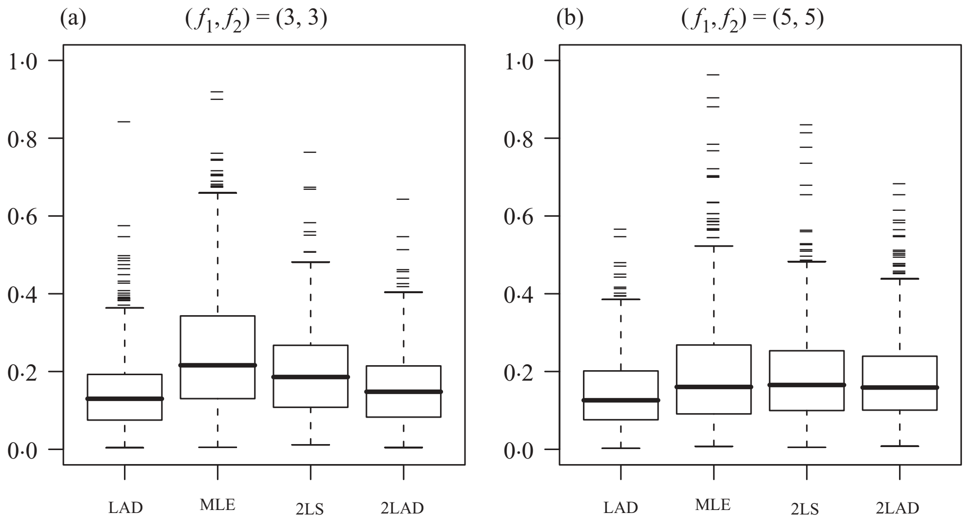

Fig. 2. Boxplots for average absolute errors of the four types of estimator with symmetric innovation u_t for (a) $(f_1, f_2) = (3, 3)$ and (b) $(f_1, f_2) = (5, 5)$, for methods LAD, MLE, 2LS and 2LAD.

where $\hat{\phi}$ is the median of the sequence $\{Y_t/Y_{t-1}, t = 2, \ldots, n\}$. For comparison, another two-stage procedure, 2LS, was also considered:

$$\widetilde{\phi} = \arg\min \sum_{t=1}^{n} \varepsilon_t^2(\phi), \qquad (\widetilde{\alpha}_0, \widetilde{\alpha}_1)' = \arg\min \sum_{t=1}^{n} |\log \varepsilon_t^2(\widetilde{\phi}) - \log h_t(\widetilde{\phi}, \alpha_0, \alpha_1)|,$$

where $\varepsilon_t(\phi) = Y_t - \phi Y_{t-1} - \mu$ since $\mu = Eu_t$ may not be equal to zero. The same adjustment is applied to $\hat{\lambda}_{\text{MLE}}$. The boxplots for the average absolute errors of the above four different methods are presented in Fig. 2 for the cases of symmetric innovation distributions corresponding to $(f_1, f_2) = (3, 3)$ and $(f_1, f_2) = (5, 5)$. For the cases of asymmetric innovation distributions, the results for $(f_1, f_2) = (3, 5)$ and $(f_1, f_2) = (9, 5)$ were very similar to Fig. 2(a) and (b), respectively. The results show that the least absolute deviation estimator $\hat{\lambda}_n$ outperforms the other three. This conclusion is consistent with the argument in § 2.

The final experiment examines the performance of $\hat{\lambda}_n$ in finite-sample cases. The generating process (5) with $d = 0{\cdot}3$ was employed again. Note that the series $\{Y_t\}$ has the long-memory property. The sample size is set to be 300 or 400, and we drew 500 independent replications for each combination of the sample sizes and the innovation distributions. Table 1 presents the estimated biases, the empirical root mean squared errors and the root mean asymptotic variance of the estimators. The biases are all small and the root mean asymptotic variances are very similar to the empirical root mean squared errors. All biases, empirical root mean squared errors and root mean asymptotic variances change little when the series is centred by the sample mean and they decrease as the sample size increases. The empirical root mean squared errors and the root mean asymptotic variances become very similar when the sample size is larger, $n = 400$. We also tried different memory parameters for the generating process (5) and found very similar results.

4·2. *Performance of the portmanteau tests*

In this section, we check the empirical sizes and powers of the two portmanteau tests obtained in § 3. Three different generating processes were involved: the ARFIMA(0,d,0)–GARCH(1,1)

Table 1. *Simulation study. Estimated bias, square root of the mean squared error,* RMSE, *and square root of the mean asymptotic variance,* RMAV, *for the* ARFIMA $(0, 0{\cdot}3, 0)$–GARCH $(1, 1)$ *model, based on* 500 *replications; all figures multiplied by* 10

		Known mean				Unknown mean			
n		$\hat{d}$	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{d}$	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\beta}_1$
					Innovation distribution $t(3)$				
300	BIAS	0·023	0·695	0·072	−0·723	−0·024	0·734	0·063	−0·751
	RMSE	0·457	1·938	0·565	2·068	0·499	1·964	0·568	2·099
	RMAV	0·450	2·257	0·540	2·355	0·462	2·309	0·538	2·386
400	BIAS	−0·009	0·516	0·046	−0·478	−0·045	0·542	0·045	−0·490
	RMSE	0·387	1·565	0·422	1·590	0·423	1·613	0·425	1·614
	RMAV	0·389	1·537	0·455	1·603	0·398	1·550	0·455	1·611
					Innovation distribution $t(5)$				
300	BIAS	−0·045	0·747	0·015	−0·406	−0·115	0·743	0·009	−0·391
	RMSE	0·559	2·063	0·479	1·447	0·589	2·040	0·476	1·414
	RMAV	0·520	2·114	0·486	1·448	0·532	2·147	0·485	1·467
400	BIAS	−0·025	0·626	0·001	−0·279	−0·069	0·629	−0·005	−0·276
	RMSE	0·469	1·765	0·402	1·163	0·479	1·759	0·394	1·154
	RMAV	0·452	1·725	0·416	1·169	0·461	1·733	0·415	1·176
					Innovation distribution $N(0, 1)$				
300	BIAS	−0·035	0·919	0·006	−0·309	−0·129	0·918	0·000	−0·305
	RMSE	0·606	2·317	0·514	1·227	0·622	2·332	0·511	1·233
	RMAV	0·593	2·616	0·472	1·339	0·605	3·091	0·471	1·562
400	BIAS	0·052	0·720	0·004	−0·260	−0·141	0·706	−0·005	−0·246
	RMSE	0·542	1·931	0·423	1·023	0·586	1·922	0·423	1·018
	RMAV	0·513	1·984	0·409	1·055	0·522	1·977	0·407	1·052

process,

$$(1-B)^d Y_t = \varepsilon_t, \varepsilon_t = u_t h_t^{1/2}, \quad h_t = 0{\cdot}3 + 0{\cdot}3\varepsilon_{t-1}^2 + 0{\cdot}3h_{t-1},$$

was used to check the empirical sizes; the ARFIMA(0,d,0)–GARCH(3,1) process,

$$(1-B)^d Y_t = \varepsilon_t, \varepsilon_t = u_t h_t^{1/2}, \quad h_t = 0{\cdot}3 + 0{\cdot}3\varepsilon_{t-1}^2 + 0{\cdot}3\varepsilon_{t-3}^2 + 0{\cdot}3h_{t-1},$$

was used to check the sensitivity for the misspecification of conditional variance, and we call this Type I power; and the ARFIMA(2,d,0)–GARCH(1,1) process,

$$(1-0{\cdot}2B^2)(1-B)^d Y_t = \varepsilon_t, \varepsilon_t = u_t h_t^{1/2}, \quad h_t = 0{\cdot}3 + 0{\cdot}3\varepsilon_{t-1}^2 + 0{\cdot}3h_{t-1},$$

was used to check the sensitivity for the misspecification of conditional mean, and we call this Type II power. The memory parameter d was taken to be $-0{\cdot}3$, $0{\cdot}3$ or $0{\cdot}7$, resulting in series with the short-memory, long-memory or nonstationary property, respectively. Two different sample sizes, $n = 400$ and $n = 600$, were considered and there were 1000 replications for each combination of the memory parameters d, sample sizes n, and the innovation distributions. We estimated all the simulated data with the ARFIMA(0, d, 0)–GARCH(1,1) model by the least absolute deviation approach and calculated the values of $Q_a(M)$ and $Q_r(M)$ with $M = 6$.

Table 2 displays the proportions of rejections based on the upper 5th percentile of the χ_6^2 distribution. All the sizes of $Q_a(6)$ and $Q_r(6)$ in Table 2 are very close to 0·05 especially for the cases with $n = 600$. Type I powers of $Q_a(6)$ and Type II powers of $Q_r(6)$ are all greater than 0·5 when the sample size n is as large as 600. This means that the tests $Q_a(M)$ and $Q_r(M)$ can be used to check respectively whether or not the conditional variance part or the conditional mean part of the fitted model is misspecified. Type I powers of $Q_r(6)$ and Type II powers of $Q_a(6)$

Table 2. *The empirical size and power of $Q_a(6)$ and $Q_r(6)$, based on 1000 replications*

Innovation distribution	n	Size $Q_a(6)$	Size $Q_r(6)$	Type I Power $Q_a(6)$	Type I Power $Q_r(6)$	Type II Power $Q_a(6)$	Type II Power $Q_r(6)$
				Differencing parameter $d = -0{\cdot}3$			
t_3	400	0·043	0·043	0·328	0·078	0·039	0·374
	600	0·046	0·044	0·500	0·088	0·037	0·620
t_5	400	0·055	0·063	0·733	0·099	0·052	0·620
	600	0·053	0·047	0·901	0·104	0·045	0·801
$N(0, 1)$	400	0·058	0·047	0·919	0·105	0·048	0·691
	600	0·045	0·053	0·984	0·099	0·038	0·990
				Differencing parameter $d = 0{\cdot}3$			
t_3	400	0·051	0·038	0·338	0·087	0·039	0·388
	600	0·053	0·043	0·462	0·086	0·034	0·589
t_5	400	0·061	0·043	0·747	0·108	0·047	0·621
	600	0·051	0·052	0·860	0·094	0·039	0·824
$N(0, 1)$	400	0·055	0·056	0·913	0·104	0·063	0·698
	600	0·048	0·053	0·981	0·092	0·049	0·886
				Differencing parameter $d = 0{\cdot}7$			
t_3	400	0·052	0·048	0·319	0·082	0·038	0·346
	600	0·049	0·043	0·475	0·073	0·041	0·588
t_5	400	0·052	0·049	0·742	0·105	0·050	0·599
	600	0·048	0·048	0·900	0·092	0·038	0·815
$N(0, 1)$	400	0·047	0·043	0·923	0·124	0·050	0·701
	600	0·051	0·043	0·987	0·114	0·053	0·875

are no more than 0·15. Hence, only the combination of $Q_a(M)$ and $Q_r(M)$ forms a complete diagnostic tool for checking whether or not the fitted ARFIMA–GARCH model is adequate.

5. An illustrative example

The dataset contains the absolute returns, as a percentage, of the Dow Jones Industrial Average Index daily closing price. There are 2519 observations from January 3, 1995 to December 31, 2004. The mean and standard deviation of the absolute returns are 0·352 and 0·335, respectively. Denote the centralized absolute returns by $\{y_t\}$. The conditional heteroscedasticity in the time series is obvious from Fig. 3(a). Figure 3(b) shows that the sample autocorrelations of the absolute returns are relatively small in magnitude, but decay very slowly. They appear to be significant at the 5% level even after 200 lags, suggesting the presence of long memory (Tsay, 2002). The ARFIMA–GARCH models were considered for the observed series $\{y_t\}$ and the least absolute deviation method was used to find the memory parameter as well as others.

We considered two different models for the conditional mean, ARFIMA$(4, d, 0)$ and ARFIMA$(2, d, 1)$, and two different models for the conditional variance, ARCH(4) and GARCH(1, 1), leading to four models in total: Model 1 combines an ARFIMA$(4, d, 0)$ model with ARCH(4) error; Model 2 combines an ARFIMA$(4, d, 0)$ model with GARCH(1, 1) error; Model 3 combines an ARFIMA$(2, d, 1)$ model with ARCH(4) error; and Model 4 combines an ARFIMA$(2, d, 1)$ model with GARCH(1, 1) error. The methodology in § 2 and § 3 was applied to these four models. We set M to be 15 and the values of μ, σ^2, $\sigma^2_{|u|}$ and $f(0)$ were estimated with the methods mentioned in § 3. The bandwidth was set to be 0·05. The results are summarized in Table 3.

Only Model 4 is not rejected by either of the test statistics $Q_a(15)$ and $Q_r(15)$ at significance level 0·05 and hence is adequate; note that $\chi^2_{15,0{\cdot}95} = 25{\cdot}00$. Consistent with the simulation results

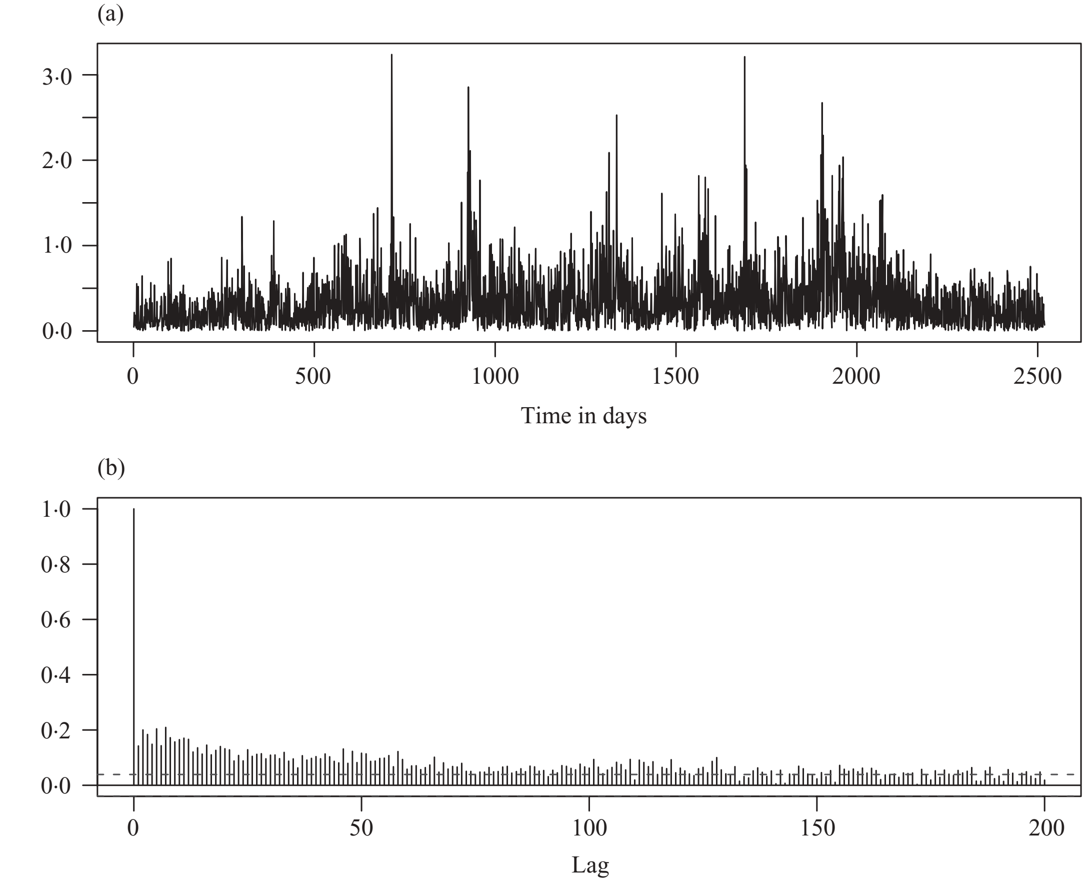

Fig. 3. Daily closing Dow Jones Industrial Average Index, 1995–2004. (a) Time plot of absolute returns, (b) Empirical autocorrelation function of absolute returns.

Table 3. *Modelling results for the absolute returns of Dow Jones Industrial Average Index daily closing price,* 1995–2004

	Model 1		Model 2		Model 3		Model 4	
Parameter	$\hat{\lambda}_n$	SD	$\hat{\lambda}_n$	SD	$\hat{\lambda}_n$	SD	$\hat{\lambda}_n$	SD
d	0·4349	5·672	0·4651	5·541	0·6594	12·40	0·7117	13·50
ϕ_1	−0·4573	6·818	−0·4786	6·570	0·0784	7·118	0·0618	7·815
ϕ_2	−0·2611	6·526	−0·2827	6·393	0·0170	4·334	0·0193	3·912
ϕ_3	−0·1794	5·543	−0·2052	5·591				
ϕ_4	−0·0707	4·280	−0·1004	4·499				
ψ_1					−0·7699	7·925	−0·8004	7·673
α_0	0·0242	1·982	0·0009	0·242	0·0229	1·925	0·0008	0·223
α_1	0·0973	21·50	0·0465	7·264	0·1011	21·60	0·0469	7·201
α_2	0·1045	22·00			0·1086	22·30		
α_3	0·0746	19·50			0·0815	20·10		
α_4	0·0876	20·60			0·0923	20·90		
β_1			0·9084	12·70			0·9100	12·20
$Q_r(15)$	30·43		33·12		17·69		15·96	
$Q_a(15)$	109·40		10·79		97·83		10·88	

SD, estimated asymptotic standard deviation multiplied by 10^3.

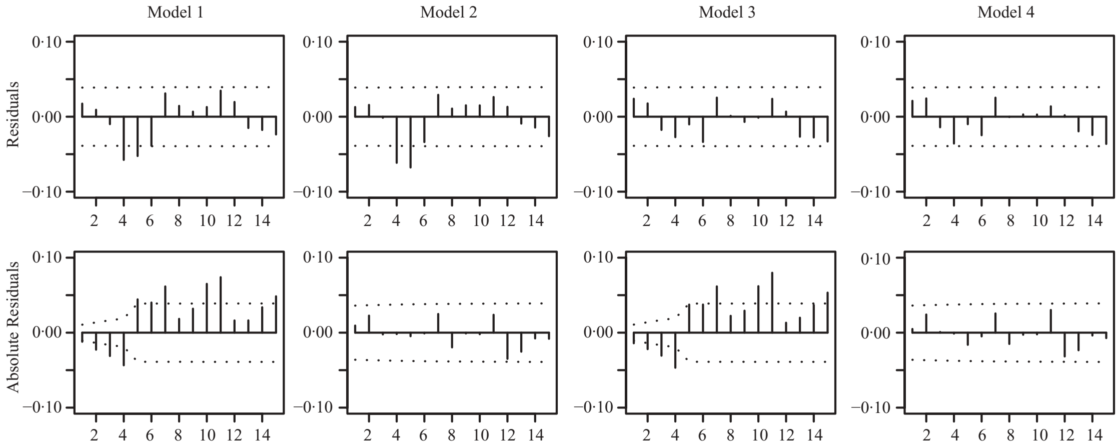

Fig. 4. Dow Jones example. Sample autocorrelation functions of residuals and absolute residuals obtained from fitting Models 1–4. The dotted lines show $\pm 1{\cdot}96 A_i$, the 95% asymptotic confidence intervals, where A_i, $i = 1, \ldots, 15$, is the asymptotic standard error for lag i.

in § 5, at significance level 0·05, the portmanteau test $Q_a(15)$ rejects Models 1 and 3, suggesting that the conditional variances are misspecified, while $Q_r(15)$ rejects Models 1 and 2, suggesting that the conditional means are misspecified. Figure 4 presents the sample autocorrelation functions of residuals and absolute residuals coming from the above four fitted models, along with 95% confidence bands. These plots are consistent with the above findings. We tried several other values of M and similar results were obtained.

The value $\hat{d} = 0{\cdot}7117$ can be considered as a reliable estimate of the true memory parameter since Model 4 is the selected model, indicating that this sequence is probably nonstationary. Note that all the estimated memory parameters are significantly different, with 95% confidence intervals that do not even overlap. Models 1 and 2 even fail to identify that the sequence is nonstationary. Hence it is important to specify the conditional mean and variance correctly when the least absolute deviation method is employed to estimate ARFIMA–GARCH models.

Acknowledgement

W. K. Li thanks the Croucher Foundation for awarding a Senior Research Fellowship and the Hong Kong Research Grant Council for a grant for partial support. The authors thank Dr S. Ling for useful comments and discussion. We thank Professor D.M. Titterington and two referees for comments that led to improvement of the paper.

Appendix

Technical details

Some properties of the derivative functions of $\varepsilon_t(\lambda)$ and $h_t(\lambda)$. Under Assumptions 1 and 2, the functions, $\varepsilon_t(\lambda)$ and $h_t(\lambda)$, defined in § 2 are both meaningful and their first-order derivatives are as follows:

$$\frac{\partial \varepsilon_t(\lambda)}{\partial \phi_j} = -\phi^{-1}(B)\varepsilon_{t-j}(\lambda), \qquad \frac{\partial \varepsilon_t(\lambda)}{\partial \psi_j} = -\psi^{-1}(B)\varepsilon_{t-j}(\lambda),$$

$$\frac{\partial \varepsilon_t(\lambda)}{\partial d} = -\sum_{k=1}^{\infty} \frac{1}{k}\varepsilon_{t-k}(\lambda), \qquad \frac{\partial h_t(\lambda)}{\partial \delta} = \tilde{\varepsilon}_t + \sum_{i=1}^{s} \beta_i \frac{\partial h_{t-i}(\lambda)}{\partial \delta},$$

$$\frac{\partial h_t(\lambda)}{\partial \gamma} = 2\sum_{i=1}^{r} \alpha_i \varepsilon_{t-i}(\lambda)\frac{\partial \varepsilon_{t-i}(\lambda)}{\partial \gamma} + \sum_{i=1}^{s} \beta_i \frac{\partial h_{t-i}(\lambda)}{\partial \gamma},$$

where $\tilde{\varepsilon}_t = (1, \varepsilon_{t-1}^2(\lambda), \ldots, \varepsilon_{t-r}^2(\lambda), h_{t-1}(\lambda), \ldots, h_{t-s}(\lambda))'$; see Ling & Li (1997).

If the condition $|d| < 0{\cdot}5$ is also assumed, the processes $\{Y_t\}$ and $\{\varepsilon_t\}$ generated by models (1) and (2) are strictly stationary and have a finite second moment. Let $\Theta^* = \Theta \cap \{|d| < 0{\cdot}5\}$, which is still compact. Then we can show that

$$E\left\{\sup_{\lambda\in\Theta^*}\left\|\frac{\partial \varepsilon_t(\lambda)}{\partial \gamma}\right\|^2\right\} < \infty, \qquad E\left\{\sup_{\lambda\in\Theta^*}\left\|\frac{1}{\surd\{h_t(\lambda)\}}\frac{\partial \varepsilon_t(\lambda)}{\partial \gamma}\right\|^2\right\} < \infty,$$

$$E\left\{\sup_{\lambda\in\Theta^*}\left\|\frac{\partial^2 \varepsilon_t(\lambda)}{\partial \gamma\partial\gamma'}\right\|^2\right\} < \infty, \qquad E\left\{\sup_{\lambda\in\Theta^*}\left\|\frac{1}{\surd\{h_t(\lambda)\}}\frac{\partial h_t(\lambda)}{\partial \lambda}\right\|^2\right\} < \infty,$$

$$E\left\{\sup_{\lambda\in\Theta^*}\left\|\frac{1}{h_t(\lambda)}\frac{\partial h_t(\lambda)}{\partial \lambda}\right\|^2\right\} < \infty, \qquad E\left\{\sup_{\lambda\in\Theta^*}\left\|\frac{1}{h_t(\lambda)}\frac{\partial^2 h_t(\lambda)}{\partial \lambda\partial\lambda'}\right\|^2\right\} < \infty,$$

where $\|\cdot\|$ is the Euclidean norm. The above inequalities are necessary for the derivations in § 3 and the proof of Theorem 1.

Proof of Theorem 1. For any $v = (v_1', v_2')' \in R^l$, where $v_1 \in R^{p+q+1}$, $v_2 \in R^{r+s+1}$ and $v \neq 0$, let

$$\begin{aligned} S_n(v) &= L_n(\lambda_0 + n^{-1/2}v) - L_n(\lambda_0) \\ &= \sum_{t=1}^{n}\left[\frac{1}{\surd\{h_t(\lambda_0)\}}\left\{|\varepsilon_t(\lambda_0 + n^{-1/2}v)| - |\varepsilon_t(\lambda_0)|\right\}\right] \\ &\quad + \sum_{t=1}^{n}\left[\left\{\frac{1}{\surd\{h_t(\lambda_0+n^{-1/2}v)\}} - \frac{1}{\surd\{h_t(\lambda_0)\}}\right\}\left\{|\varepsilon_t(\lambda_0+n^{-1/2}v)| - |\varepsilon_t(\lambda_0)|\right\}\right] \\ &\quad + \sum_{t=1}^{n}\left[\frac{|\varepsilon_t(\lambda_0)|}{\surd\{h_t(\lambda_0+n^{-1/2}v)\}} + \frac{1}{2}\log h_t(\lambda_0+n^{-1/2}v) \right. \\ &\qquad\left. - \frac{|\varepsilon_t(\lambda_0)|}{\surd\{h_t(\lambda_0)\}} - \frac{1}{2}\log h_t(\lambda_0)\right] \\ &= S_{1n}(v) + S_{2n}(v) + S_{3n}(v), \\ S_{1n}^*(v) &= \sum_{t=1}^{n}\left[\frac{1}{\surd\{h_t(\lambda_0)\}}\left\{\left|\varepsilon_t(\lambda_0) + n^{-1/2}v_1\frac{\partial \varepsilon_t(\lambda_0)}{\partial\gamma}\right| - |\varepsilon_t(\lambda_0)|\right\}\right]. \end{aligned}$$

It holds that, for $x \neq 0$,

$$|x+y| - |x| = y\,\mathrm{sgn}(x) + 2\int_0^{-y}\{I(x \leqslant s) - I(x \leqslant 0)\}ds,$$

where the function sgn(x) is equal to 1 for $x > 0$, 0 for $x = 0$ and -1 for $x < 0$; see Knight (1998). By a method similar to that in Peng & Yao (2003) together with the above equation, we can show that

$$S_{1n}^*(v) = n^{-1/2}v_1'\sum_{t=1}^{n}\mathrm{sgn}(u_t)h_t^{-1/2}\frac{\partial \varepsilon_t}{\partial \gamma} + f(0)v_1'\Omega_\varepsilon v_1 + o_p(1).$$

As in Davis & Dunsmuir (1997), by the inequalities at the beginning of this Appendix, we can show that

$$S_{1n}(v) \to n^{-1/2} v_1' \sum_{t=1}^{n} \mathrm{sgn}(u_t) h_t^{-1/2} \frac{\partial \varepsilon_t}{\partial \gamma} + f(0) v_1' \Omega_\varepsilon v_1$$

in distribution as $n \to \infty$. Hence, $S_{2n}(v) = o_p(1)$.

Note that $S_{3n}(v)$ is a smooth function so that Taylor expansion can be used. After some algebra, the inequalities at the beginning of this Appendix ensure that

$$S_{3n}(v) = -0{\cdot}5\, n^{-1/2} v' \sum_{t=1}^{n} (|u_t| - 1) h_t^{-1} \frac{\partial h_t}{\partial \lambda} + 0{\cdot}5\, v' \Omega_2 v + o_p(1).$$

Let

$$s_n = v' \sum_{t=1}^{n} \left\{ \mathrm{sgn}(u_t) h_t^{-1/2} \frac{\partial \varepsilon_t}{\partial \lambda} - 0{\cdot}5(|u_t| - 1) h_t^{-1} \frac{\partial h_t}{\partial \lambda} \right\}.$$

Note that, under Assumption 3, s_n is a martingale and $(1/n)Es_n^2 = v'\Sigma_1 v > 0$, where the matrix $\Sigma_1 = \Omega_1 + \sigma_{|u|}^2 \Omega_2 - \mu\Omega_3$ is defined in § 2. By the strict stationarity and ergodicity of $\{Y_t\}$ and $\{\varepsilon_t\}$, it holds that

$$\left\{(1/n)Es_n^2\right\}^{-1}\left\{(1/n)E\left(s_n^2 | \mathcal{F}_{n-1}\right)\right\} \to 1,$$

almost surely. Using the central limit theorem of Stout (1974), we can show that $n^{-1/2}s_n \to v'W$ in distribution, where W is a multivariate normal vector with mean 0 and covariance matrix Σ_1. Hence,

$$S_n(v) \to v'W + v'\Sigma_2 v$$

in distribution, where $\Sigma_2 = f(0)\Omega_1 + 0{\cdot}5\Omega_2$ is defined in § 2. Following Lemma 2·2 and Remark 1 of Davis et al. (1992) we complete the proof of Theorem 1. □

References

Baillie, R. T., Chung, C. F. & Tiles, M. A. (1996). Analyzing inflation by the fractionally integrated ARFIMA-GARCH model. *J. Appl. Economet.* **11**, 23–40.

Beran, J. (1995). Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models. *J. R. Statist. Soc.* B **57**, 659–72.

Beran, J. & Feng, Y. H. (2001). Local polynomial estimation with a ARFIMA-GARCH error process. *Bernoulli* **7**, 733–50.

Berkes, I. & Horvath, L. (2004). The efficiency of the estimators of the parameters in GARCH processes. *Ann. Statist.* **32**, 633–55.

Bhardwaj, G. & Swanson, N. R. (2006). An empirical investigation of the usefulness of ARFIMA models for predicting macroeconomic and financial time series. *J. Economet.* **131**, 539–78.

Bollerslev, T. (1986). Generalized autoregression conditional heteroskedasticity. *J. Economet.* **31**, 307–27.

Davis, R. A. & Dunsmuir, W. T. M. (1997). Least absolute deviation estimation for regression with ARMA errors. *J. Theor. Prob.* **10**, 481–97.

Davis, R. A., Knight, K. & Liu, J. (1992). M-estimation for autoregressions with infinite variances. *Stoch. Proces. Applic.* **40**, 145–80.

Embrechts, P., Kluppelberg, C. & Mikosch, T. (1997). *Modelling Extremal Events*. Berlin: Springer.

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

Fan, J. & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Am. Statist. Assoc.* **96**, 1348–60.

Francq, C. & Zakoian, J. M. (2004). Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* **10**, 605–37.

Granger, C. W. J. & Joyeux, R. (1980). An introduction to long-memory time series models and fractional difference. *J. Time Ser. Anal.* **1**, 15–39.

Granger, C. W. J., Spear, S. & Ding, Z. X. (1999). Stylized facts on the temporal and distributional properties of absolute returns: an update. In *Statistics and Finance: An Interface*, Ed. W. S. Chan, W. K. Li and H. Tong, pp. 97–120. London: Imperial College Press.

HALDRUP, N. & NIELSEN, M. O. (2007). Estimation of fractional integration in the presence of data noise. *Comp. Statist. Data Anal.* **51**, 3100–14.

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

HAUSER, M. A. & HURST, R. M. (2001). Forecasting high-frequency financial data with the ARFIMA-ARCH model. *J. Forecasting* **20**, 501–18.

HOSKING, J. R. M. (1981). Fractional differencing. *Biometrika* **68**, 165–76.

KNIGHT, K. (1998). Limiting distributions for l_1 regression estimators under general conditions. *Ann. Statist.* **26**, 755–70.

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

LI, W. K. (1992). On the asymptotic distribution of residual autocorrelations in nonlinear time series modelling. *Biometrika* **79**, 435–7.

LI, W. K. & MAK, T. K. (1994). On the squared residual autocorrelations in nonlinear time series with conditional heteroskedasticity. *J. Time Ser. Anal.* **15**, 627–36.

LING, S. (2007). A double AR(p) model: structure and estimation. *Statist. Sinica* **17**, 161–75.

LING, S. & LI, W. K. (1997). On fractionally integrated autoregressive moving average time series models with conditional heteroscedasticity. *J. Am. Statist. Assoc.* **92**, 1184–93.

MCLEOD, A. I. & HIPEL, K. W. (1978). Preservation of the rescaled adjusted range 1. A reassessment of the hurst phenomenon. *Water Resour. Res.* **14**, 491–508.

MIKOSCH, T. & STARICA, C. (2000). Limit theory for the sample autocorrelations and extremes of a GARCH(1,1) process. *Ann. Statist.* **28**, 1427–51.

MITTNIK, S., RACHEV, S. T. & PAOLELLA, M. S. (1998). Stable Paretian models in finance: some empirical and theoretical aspects. In *A Practical Guide to Heavy Tails*, Ed. R. J. Adler, R. E. Feldman and M. S. Taqqu, pp. 79–110. Boston: Birkhauser.

PENG, L. & YAO, Q. (2003). Least absolute deviations estimation for ARCH and GARCH models. *Biometrika* **90**, 967–75.

ROBINSON, P. M. (2003). Long-memory time series. In *Time Series with Long Memory*, Ed. P. M. Robinson, pp. 4–32. Oxford: Oxford University Press.

STOUT, W. F. (1974). *Almost Sure Convergence*. New York: Academic Press.

TSAY, R. S. (2002). *Analysis of Financial Time Series*. New York: John Wiley & Sons.

TSE, Y. K. & ZUO, X. L. (1997). Testing for conditional heteroscedasticity: some Monte Carlo results. *J. Statist. Comp. Simul.* **58**, 237–53.

WONG, H. & LING, S. (2005). Mixed portmanteau tests for time-series models. *J. Time Ser. Anal.* **26**, 569–79.

[*Received January* 2006. *Revised November* 2007]