Comment

Beth Andrews
Department of Statistics, Northwestern University, Evanston, IL 60208 (bandrews@northwestern.edu)

In their article, Fan, Qi, and Xiu develop non-Gaussian quasi-maximum likelihood estimators for the parameters $\boldsymbol{\theta} = (\sigma, \boldsymbol{\gamma}')' = (\sigma, a_1, \ldots, a_p, b_1, \ldots, b_q)'$ of a GARCH process $\{x_t\}$ where

$$\begin{aligned} x_t &= \sigma v_t \varepsilon_t, \\ v_t^2 &= 1 + \sum_{i=1}^{p} a_i x_{t-i}^2 + \sum_{j=1}^{q} b_j v_{t-j}^2, \end{aligned}$$

and the noise $\{\varepsilon_t\}$ are assumed to be independent and identically distributed with mean zero and variance one. The QMLEs of $\boldsymbol{\gamma}$ are shown to be $\sqrt{T}$-consistent (T represents sample size) and asymptotically Normal under general conditions. When $\mathrm{E}\{\varepsilon_t^4\} < \infty$, the QMLEs of the scale parameter σ are also $\sqrt{T}$-consistent and asymptotically Normal, but, as is the case for Gaussian QMLEs of GARCH model parameters, the estimator of σ has a slower rate of convergence otherwise (Hall and Yao, 2003). As mentioned in Fan et al., a rank-based technique for estimating $\boldsymbol{\theta}$ is presented in Andrews (2012). These rank (R)-estimators are also consistent under general conditions, with the same rates of convergence as the non-Gaussian QMLEs. Hence, the R-estimators have robustness properties similar to the QMLEs. In this note, I make some methodological and efficiency comparisons between the two techniques, and suggest R-estimation be used prior to QMLE for preliminary GARCH estimation. Once a R-estimate has been found, corresponding model residuals can be used to identify one or more suitable noise distributions and QMLE/MLE can then be used. As suggested by Fan et al. in Section 6, one can optimize over a pool of appropriate likelihoods in an effort to improve efficiency. Additionally, MLEs of all elements of $\boldsymbol{\theta}$ are consistent with rate $\sqrt{T}$ under general conditions (Berkes and Horváth, 2004).

Methodological Comparisons The R-estimator $\hat{\boldsymbol{\gamma}}_R$ of $\boldsymbol{\gamma}$ is found by minimizing

$$D_T(\boldsymbol{\gamma}) = \sum_{t=p+1}^{T} \lambda\left(\frac{R_t(\boldsymbol{\gamma})}{T-p+1}\right)\left[\xi_t(\boldsymbol{\gamma}) - \overline{\xi(\boldsymbol{\gamma})}\right], \tag{1}$$

where $\{\xi_t(\boldsymbol{\gamma})\}_{t=p+1}^{T} = \{\log(x_t^2) - \log(v_t^2(\boldsymbol{\gamma}))\}_{t=p+1}^{T}$ are log-transformed residuals, $\{R_t(\boldsymbol{\gamma})\}_{t=p+1}^{T}$ contains the ranks of $\{\xi_t(\boldsymbol{\gamma})\}_{t=p+1}^{T}$, $\overline{\xi(\boldsymbol{\gamma})} = (T-p)^{-1}\sum_{t=p+1}^{T}\xi_t(\boldsymbol{\gamma})$, and λ is a nonconstant and nondecreasing weight function from $(0,1)$ to $\mathbb{R}$. In practice when the noise distribution for the GARCH process is unknown, I recommend using the weight function $\lambda_{t_7}(x) = [7\{F_{t_7}^{-1}((x+1)/2)\}^2 - 5]/[\{F_{t_7}^{-1}((x+1)/2)\}^2 + 5]$, where F_{t_7} represents the distribution function for standardized t_7 noise, or a similar weight function (Andrews, 2012, Remark 7). When compared to other techniques, R-estimation with weight function λ_{t_7} performs well for light, medium, and heavier-tailed noise distributions. Note that, if $\overline{\lambda} = (T-p)^{-1}\sum_{t=p+1}^{T}\lambda((t-p)/(T-p+1))$ and $\{\xi_{(t)}(\boldsymbol{\gamma})\}_{t=p+1}^{T}$ is the series $\{\xi_t(\boldsymbol{\gamma})\}_{t=p+1}^{T}$ ordered from smallest to largest, then equation (1) can also be written as $D_T(\boldsymbol{\gamma}) = \sum_{t=p+1}^{T}[\lambda((t-p)/(T-p+1)) - \overline{\lambda}][\xi_{(t)}(\boldsymbol{\gamma}) - \overline{\xi(\boldsymbol{\gamma})}]$. In addition, D_T is a non-negative, continuous function (Andrews, 2012). Because it tends to be near zero when the elements of $\{\xi_t(\boldsymbol{\gamma})\}$ are similar and gets larger as the values of $\{|\xi_{(t)}(\boldsymbol{\gamma}) - \overline{\xi(\boldsymbol{\gamma})}|\}$ increase, D_T can be thought of as a measure of the dispersion of the residuals $\{\xi_t(\boldsymbol{\gamma})\}$. This rank-based estimation technique is similar to the one introduced in Jaeckel (1972) for estimating linear regression parameters. In Remark 9 of Andrews (2012), the corresponding R-estimator of σ^2 (which, following the notation of Bollerslev [1986], I denote as α_0) is given by $\hat{\sigma}_R^2 = n^{-1}\sum_{t=p+1}^{T} x_t^2/v_t^2(\hat{\gamma}_R)$. Hence,

$$\hat{\sigma}_R = \sqrt{n^{-1}\sum_{t=p+1}^{T}\frac{x_t^2}{v_t^2(\hat{\gamma}_R)}}. \tag{2}$$

It follows that the R-estimate of $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}_R = (\hat{\sigma}_R, \hat{\boldsymbol{\gamma}}_R')'$, can be obtained via a two-step procedure: (a) minimize $D_T(\boldsymbol{\gamma})$ in equation (1) to find $\hat{\boldsymbol{\gamma}}_R$, and (b) obtain $\hat{\sigma}_R$ via equation (2). In contrast, the non-Gaussian QMLEs $\hat{\boldsymbol{\theta}}_T$ proposed by Fan et al. are obtained via a three-step procedure, where optimization (i.e., maximization) is required in all three steps. From this perspective, R-estimation is a simpler method than non-Gaussian QMLE.

Relative Efficiency Let $\tilde{f}$ and $\tilde{F}$ represent the density and distribution functions for $\ln(\varepsilon_t^2)$. In Andrews (2012), I show that, when the distribution for the noise $\{\varepsilon_t\}$ is symmetric about zero and a weight function $\lambda(x) \propto -\tilde{f}'(\tilde{F}^{-1}(x))/\tilde{f}(\tilde{F}^{-1}(x))$ is used, R-estimators of $\boldsymbol{\gamma}$ have the same asymptotic efficiency as MLEs. When ε_t has a standardized t_7 distribution, $\lambda_{t_7}(x) \propto -\tilde{f}'(\tilde{F}^{-1}(x))/\tilde{f}(\tilde{F}^{-1}(x))$. Furthermore, when the weight function λ used for R-estimation and the density f used for QMLE correspond to the same noise distribution, R-estimation tends to be asymptotically as efficient or more efficient than QMLE. This is discussed in Andrews (2012, Remarks 6 and 7) and is also observed in Fan et al. (Section 7.2/Table 3) when λ and f correspond to the standardized t_7 distribution. Note that the asymptotic relative efficiency for R-estimators of $\boldsymbol{\gamma}$ with respect to non-Gaussian QMLEs is given by

$$\mathrm{ARE} = 4\tilde{J}^{-1}\tilde{K}^2 \frac{\mathrm{E}\left(h_1(\varepsilon_t, \eta_f)\right)^2}{\eta_f^2 \left(\mathrm{E}h_2(\varepsilon_t, \eta_f)\right)^2}, \tag{3}$$

where $\tilde{J}$ and $\tilde{K}$ are defined in Andrews (2012) and depend on the choice of weight function λ, and h_1, h_2, and η_f are defined in Fan et al. and depend on f, the density being used for QMLE.

Following Remark 9 in Andrews (2012), when $\mathrm{E}\{\varepsilon_t^4\} < \infty$,

$$\sqrt{T}(\hat{\sigma}_R - \sigma) \xrightarrow{d} \mathrm{N}\left(0, \tilde{J}\tilde{K}^{-2}\frac{\sigma^2}{4}\mathrm{E}\left\{\frac{\partial v_t^2(\boldsymbol{\gamma})/\partial\boldsymbol{\gamma}}{v_t^2(\boldsymbol{\gamma})}\right\}' \left[\mathrm{Var}\left\{\frac{\partial v_t^2(\boldsymbol{\gamma})/\partial\boldsymbol{\gamma}}{v_t^2(\boldsymbol{\gamma})}\right\}\right]^{-1} \mathrm{E}\left\{\frac{\partial v_t^2(\boldsymbol{\gamma})/\partial\boldsymbol{\gamma}}{v_t^2(\boldsymbol{\gamma})}\right\} + \frac{\sigma^2}{4}\mathrm{Var}\{\varepsilon_t^2\}\right)$$

as $T \to \infty$. By Theorem 2 in Fan et al., also when $\mathrm{E}\{\varepsilon_t^4\} < \infty$, the non-Gaussian QMLE of σ has limiting distribution

$$\sqrt{T}(\hat{\sigma}_T - \sigma) \xrightarrow{d} \mathrm{N}\left(0, \frac{\mathrm{E}\left(h_1(\varepsilon_t, \eta_f)\right)^2}{\eta_f^2 \left(\mathrm{E}h_2(\varepsilon_t, \eta_f)\right)^2}\left\{\mathbf{M}^{-1}[1,1] - \sigma^2\right\} + \frac{\sigma^2}{4}\mathrm{Var}\{\varepsilon_t^2\}\right),$$

where $\mathbf{M}^{-1}[1,1]$ represents the element in row one, column one of matrix $\mathbf{M}^{-1}$; $\mathbf{M}$ is defined in the statement of Theorem 2. Via matrix algebra, it can be shown that

$$\sigma^2\mathrm{E}\left\{\frac{\partial v_t^2(\boldsymbol{\gamma})/\partial\boldsymbol{\gamma}}{v_t^2(\boldsymbol{\gamma})}\right\}' \left[\mathrm{Var}\left\{\frac{\partial v_t^2(\boldsymbol{\gamma})/\partial\boldsymbol{\gamma}}{v_t^2(\boldsymbol{\gamma})}\right\}\right]^{-1} \mathrm{E}\left\{\frac{\partial v_t^2(\boldsymbol{\gamma})/\partial\boldsymbol{\gamma}}{v_t^2(\boldsymbol{\gamma})}\right\} = \mathbf{M}^{-1}[1,1] - \sigma^2,$$

so the R-estimator of σ is asymptotically more efficient than the QMLE of σ when the ARE in equation (3) for estimators of $\boldsymbol{\gamma}$ is larger than one. As demonstrated in Table 3 of Fan et al., this is often the case (see also Andrews, 2012, Remark 7).

In Section 7.2, Fan et al. give simulation results for the GARCH(1,1) model with parameters $(\sigma, a_1, b_1) = (0.5, 0.6, 0.3)$ when sample size $T = 250$, 500, and 1000. In these simulations, R-estimation with weight function λ_{t_7} is essentially as efficient as QMLE. For comparison, I considered GARCH(1,1) models with parameters $(\sigma, a_1, b_1) = (0.1, 50.0, 0.4)$ and $(\sigma, a_1, b_1) = (0.1, 10.0, 0.8)$, and the same three values of T. (The model $(\sigma, a_1, b_1) = (0.1, 50.0, 0.4)$ is considered in Andrews [2012], and $(\sigma, a_1, b_1) = (0.1, 10.0, 0.8)$ was selected because, for many observed series, a value of b_1 near one appears appropriate.) In each case, I simulated 1000 GARCH processes with N(0,1) and standardized t_3 noise, and found the corresponding R-estimates and QMLEs, also using the weight function λ_{t_7} for R-estimation and the standardized t_7 density for QMLE. Root mean squared errors for the estimates are listed in Table 1. For these estimation methods,

T	Model Parameters	RMSEs R-estimates (N(0,1), t_3)	RMSEs QMLEs (N(0,1), t_3)
250	$\sigma = 0.1$	0.020, 0.026	0.029, 0.031
	$a_1 = 50.0$	22.004, 27.196	20.135, 27.213
	$b_1 = 0.4$	0.119, 0.184	0.129, 0.192
500	$\sigma = 0.1$	0.014, 0.020	0.021, 0.026
	$a_1 = 50.0$	14.700, 18.399	14.312, 18.544
	$b_1 = 0.4$	0.080, 0.125	0.083, 0.129
1000	$\sigma = 0.1$	0.009, 0.016	0.016, 0.023
	$a_1 = 50.0$	9.465, 12.982	9.496, 13.085
	$b_1 = 0.4$	0.052, 0.084	0.054, 0.086
250	$\sigma = 0.1$	0.090, 0.067	0.113, 0.085
	$a_1 = 10.0$	6.877, 7.716	6.812, 6.932
	$b_1 = 0.8$	0.350, 0.330	0.442, 0.408
500	$\sigma = 0.1$	0.059, 0.046	0.079, 0.056
	$a_1 = 10.0$	5.111, 5.182	5.365, 5.143
	$b_1 = 0.8$	0.222, 0.215	0.299, 0.265
1000	$\sigma = 0.1$	0.033, 0.028	0.047, 0.038
	$a_1 = 10.0$	3.680, 3.564	3.943, 3.745
	$b_1 = 0.8$	0.111, 0.109	0.164, 0.166

Table 1: *Root mean squared errors for R-estimates and QMLEs of GARCH model parameters when the noise distribution is N(0,1) and standardized t_3.*

the value of ARE in equation (3) is 1.041 when the noise $\{\varepsilon_t\}$ are N(0,1), and ARE is 1.052 when the $\{\varepsilon_t\}$ are standardized t_3 (Andrews, 2012). Since the RMSEs in Table 1 for R-estimation are mostly smaller than the corresponding values for QMLE, it appears the asymptotic relative efficiencies for R-estimation with respect to non-Gaussian QMLE can be indicative of finite sample behavior for sample size $250 \leq T \leq 1000$.

Concluding Remarks In Andrews (2012), the limiting distribution for R-estimators is given not only when the true parameter vector is in the interior of its parameter space and the estimators are asymptotically Normal, but also when some GARCH parameters are zero and the limiting distribution is non-Normal. The results are used to develop hypothesis tests for GARCH order selection (Andrews, 2012, Section 3.2). Since R-estimates are straightforward to compute and tend to be relatively efficient, I recommend R-estimation be used not only for preliminary GARCH estimation, but also for order selection when the noise distribution is unknown. If further model accuracy is desired, residuals from R-estimation can be used to identify one or more suitable noise distributions, and then the GARCH model can be estimated via QMLE/MLE.

References

Andrews, B. (2012), "Rank-Based Estimation for GARCH Processes," *Econometric Theory*, 28, 1037–1064.

Berkes, I., and Horváth, L. (2004), "The Efficiency of the Estimators of the Parameters in GARCH Processes," *Annals of Statistics*, 32, 633–655.

Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

Hall, P., and Yao, Q. (2003), "Inference in ARCH and GARCH Models with Heavy-Tailed Errors," *Econometrica*, 71, 285–317.

Jaeckel, L.A. (1972), "Estimating Regression Coefficients by Minimizing the Dispersion of the Residuals," *Annals of Mathematical Statistics*, 43, 1449–1458.