Rejoinder

Jianqing FAN and Lei QI
Princeton University, Princeton, NJ 08544 (jqfan@princeton.edu; lqi@princeton.edu)

Dacheng XIU
University of Chicago, Chicago, IL 60637 (dacheng.xiu@chicagobooth.edu)

1. INTRODUCTION

We would like to thank Beth Andrews, Christian Francq, Jean-Michel Zakoïan, Gabriele Fiorentini, Enrique Sentana, Qiwei Yao, Shiqing Ling, and Ke Zhu for their insightful and stimulating discussions of our article.

The comments span a range of issues, which we summarize as follows. Section 2 discusses the choice of estimators in practice. Section 3 discusses estimation with additional shape parameters. Section 4 comments on finite sample studies. Section 5 includes some final remark on extensions and future work.

2. CHOICE OF ESTIMATORS

In practice, the choice of an estimator depends at least on the trade-off among computational effort, statistical efficiency, finite sample performance, feasibility and flexibility, and information content of the output. Therefore, this is a complicated matter and usually there exists no method that dominates the others from every aspect.

As pointed out by Andrews, Fiorentini, and Sentana, and Francq and Zakoïan, our three-step procedure in general involves three optimization steps, which needs more computational effort than the alternative methods they propose. Nevertheless, with additional effort come certain advantages.

First, η_f is informative. It measures the deviation between the employed likelihood and the true likelihood. Moreover, η_f characterizes the exact bias one would otherwise have when using the commercial softwares to conduct non-Gaussian Quasi Maximum Likelihood Estimator (NGQMLE). Understanding the magnitude of such an important bias motivated our study initially. It is worth mentioning that η_f not only determines the bias of model parameters, it also sheds light on the bias in conditional volatilities ($\widehat{\sigma}_t \xrightarrow{\mathcal{P}} \eta_f \sigma_t$), which is crucial for empirical applications.

Second, our method is very flexible in the choice of likelihood functions. In addition to generalized Gaussian likelihoods, Student's-t densities are allowed, which are smooth everywhere. Asymmetric densities could be employed as well. Such a choice can be motivated from the empirical distribution of the residuals obtained from our first step estimation. As described in Section 7, our method has a potential to further close the gap between NGQMLE and maximum likelihood estimation (MLE) by using information embedded in these residuals.

That being said, we agree that alternative estimators could be more convenient under certain circumstances. For example, given the choice of a specific generalized Gaussian likelihood, the algorithm by Francq and Zakoïan is indeed a short cut. Also, aware of the asymptotic equivalence between our estimator and the one by Fiorentini and Sentana, we find their procedure simple yet general as well. Andrews' approach differs from the likelihood method we consider, offering an equally competitive alternative. The relative asymptotic efficiency between the likelihood method and rank-based approach depends on the choice of likelihood and Andrews' $\lambda(\cdot)$. Ling and Zhu's approach is equivalent to our NGQMLE when generalized Gaussian likelihood is employed with $\eta_f = E|\varepsilon_t|$. Yao suggests choosing Gaussian likelihood when η_f is close to 1. This may be valid for certain situations. In practice, we would prefer to use the likelihood that leads to such an η_f, which usually requires weaker moment conditions than using a Gaussian likelihood.

3. ESTIMATION OF THE SHAPE PARAMETER

Fiorentini and Sentana also suggest estimating additional shape parameters together with model parameters, which might improve the efficiency compared with using a prespecified likelihood. To implement this, we could modify the second and third steps accordingly:

$$(\widehat{\eta}_f, \widehat{\beta}_f) = \operatorname{argmax}_{\eta,\beta} \frac{1}{T} \sum_{t=1}^{T} l_2(\bar{\boldsymbol{x}}_t, \widetilde{\boldsymbol{\theta}}_T, \eta, \beta), \quad (1)$$

$$\widehat{\boldsymbol{\theta}}_T = \operatorname{argmax}_{\boldsymbol{\theta}} \frac{1}{T} \sum_{t=1}^{T} l_3(\bar{\boldsymbol{x}}_t, \widehat{\eta}_f, \boldsymbol{\theta}, \widehat{\beta}_f) = \operatorname{argmax}_{\boldsymbol{\theta}} \frac{1}{T} \sum_{t=1}^{T} \times \left(-\log(\widehat{\eta}_f \sigma v_t) + \log f\left(\frac{x_t}{\widehat{\eta}_f \sigma v_t}, \widehat{\beta}_f \right) \right), \quad (2)$$

where β_f is the pseudo-true parameter defined below:

$$(\eta_f, \beta_f) = \operatorname{argmax}_{\eta,\beta} \{-\log \eta + E(\log f(\varepsilon/\eta, \beta))\}.$$

Suppose we employ the generalized Gaussian likelihood family indexed by a shape parameter β to conduct the NGQMLE:

$$\log f(\varepsilon, \beta) = -|\varepsilon|^\beta/\beta + \text{const},$$

then by concentrating out η in (1) using $\eta_f(\beta) = (E|\varepsilon|^\beta)^{1/\beta}$ for any given β, we have

$$\beta_f = \operatorname{argmax}_\beta \left\{ -\frac{1}{\beta}(\log E|\varepsilon|^\beta + 1) \right\}.$$

Journal of Business & Economic Statistics
April 2014, Vol. 32, No. 2
DOI: 10.1080/07350015.2014.898448

Not to mention the existence of β_f may depend on the density of ε, even when ε is Gaussian, which is nested within the generalized Gaussian family, the optimal β_f is 2.571 instead of 2. As a result, estimation of the pseudo-true parameter may introduce an additional term to the asymptotic variance of γ, which further complicates the inference procedure. Our suggestion agrees with Francq and Zakoïan's, which is to select β that minimizes $(E|\varepsilon|^{2\beta} - (E|\varepsilon|^{\beta})^2)/\beta^2(E|\varepsilon|^{\beta})^2$. The optimal β in case of Gaussian ε is indeed 2, although caution is needed when estimating this β using residuals, in view of the comment by Francq and Zakoïan.

4. FINITE SAMPLE PERFORMANCE

We appreciate the effort by Andrews, Fiorentini, and Sentana, Francq and Zakoïan, Ling and Zhu in comparing the finite sample performance of available estimators. While all simulation studies are well executed, we would like to point out that caution is needed when interpreting simulation results, because simulation studies might have a design bias or an implementation bias. We compared our approach with Andrews' using the matlab code obtained from her website. Our finite sample results in Tables 4 and 5 indicate that neither of the two estimators dominates the other one under all circumstances.

5. EXTENSIONS AND FUTURE WORK

We thank for their insightful comment on applying our approach to nonstationary processes. This is a very exciting topic of research that deserves a more serious investigation. We leave it for future work. Extensions to multivariate generalized autoregressive conditional heteroscedasticity (GARCH) models or GARCH-M models are important, as mentioned by Yao, Fiorentini, and Sentana. The asymptotic theory is far more cumbersome, but a similar theory is expected to be true. As Yao conjectured, model reparameterization enables standard bootstrap inference for heteroscedastic parameters, although asymptotic theory is a feasible alternative. That being said, parameters under the original parameterization depend on the scale parameter σ, which is poorly estimated when noise is heavy-tailed. For this reason, standard bootstrap may not be feasible for parameters using the original specification.