Quasi-Maximum Likelihood Estimation of GARCH Models With Heavy-Tailed Likelihoods

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*)

The non-Gaussian maximum likelihood estimator is frequently used in GARCH models with the intention of capturing heavy-tailed returns. However, unless the parametric likelihood family contains the true likelihood, the estimator is inconsistent due to density misspecification. To correct this bias, we identify an unknown scale parameter η_f that is critical to the identification for consistency and propose a three-step quasi-maximum likelihood procedure with non-Gaussian likelihood functions. This novel approach is consistent and asymptotically normal under weak moment conditions. Moreover, it achieves better efficiency than the Gaussian alternative, particularly when the innovation error has heavy tails. We also summarize and compare the values of the scale parameter and the asymptotic efficiency for estimators based on different choices of likelihood functions with an increasing level of heaviness in the innovation tails. Numerical studies confirm the advantages of the proposed approach.

KEY WORDS: Heavy-tailed error; Quasi-likelihood; Three-step estimator.

1. INTRODUCTION

Volatility has been a crucial variable in modeling financial time series, designing trading strategies, and implementing risk management. It is often observed that volatility tends to cluster together, suggesting that volatility is autocorrelated and changing over time. Engle (1982) proposed autoregressive conditional heteroscedasticity (ARCH) to model volatility dynamics by taking weighted averages of past squared returns. This seminal idea led to a variety of volatility models. Among numerous generalizations and developments, the following GARCH model by Bollerslev (1986) has been commonly used:

$$x_t = v_t \varepsilon_t, \tag{1}$$

$$v_t^2 = c + \sum_{i=1}^{p} \widetilde{a}_i x_{t-i}^2 + \sum_{j=1}^{q} \widetilde{b}_j v_{t-j}^2. \tag{2}$$

In this GARCH(p, q) model, the variance forecast takes the weighted average of not only past square errors but also historical variances. Its simplicity and intuitive appeal make the GARCH model, especially GARCH$(1, 1)$, a workhorse and good starting point in many financial applications.

Earlier literature on inference from ARCH/GARCH models is based on a Maximum Likelihood Estimation (MLE) with the conditional Gaussian assumption on the innovation distribution. However, plenty of empirical evidence has documented heavy-tailed and asymmetric distributions of ε_t, rendering this assumption unjustified; see, for instance, Diebold (1988). Consequently, the MLE using Student's t or generalized Gaussian likelihood functions has been introduced, for example, Engle and Bollerslev (1986), Bollerslev (1987), Hsieh (1989), and Nelson (1991). However, these methods may lead to inconsistent estimates of model parameters in Equation (2) if the distribution of the innovation is misspecified. Alternatively, the Gaussian MLE, regarded as a Gaussian Quasi-Maximum Likelihood Estimator (GQMLE) may be consistent (as seen in Elie and Jeantheau 1995), and asymptotically normal, provided that the innovation has a finite fourth moment, even if the true distribution is far from Gaussian, as shown by Hall and Yao (2003) and Berkes, Horváth, and Kokoszka (2003). The asymptotic theory dates back as early as Weiss (1986) for ARCH models. Lee and Hansen (1994) and Lumsdaine (1996) showed this for GARCH(1, 1) with stronger conditions, and Bollevslev and Wooldbridge (1992) proved the theory for GARCH(p, q) under high-level assumptions.

Nevertheless, this gain in robustness comes with a loss of efficiency. Theoretically, the divergence of Gaussian likelihood from the true innovation density may considerably increase the variance of the estimates, which thereby fails to reach the efficiency of MLE by a wide margin, reflecting the cost of not knowing the true innovation distribution. Engle and Gonzalez-Rivera (1991) suggested a semiparametric procedure that can improve the efficiency of the parameter estimates up to 50% over the GQMLE based on their Monte Carlo simulations, but this is still incapable of capturing the total potential gain in efficiency, as shown by Linton (1993). Drost and Klaassen (1997) put forward an adaptive two-step semiparametric procedure based on a reparameterization of the GARCH(1, 1) model with unknown but symmetric error. Sun and Stengos (2006) further extended this work to asymmetric GARCH models. González-Rivera and Drost (1999) compared the semiparametric procedure's efficiency gain/loss to GQMLE and MLE. However, all of the above results require the innovation error to have a finite fourth moment. Hall and Yao (2003) showed that the GQMLE

© 2014 American Statistical Association
Journal of Business & Economic Statistics
April 2014, Vol. 32, No. 2
DOI: 10.1080/07350015.2013.840239
Color versions of one or more of the figures in the article can be found online at *www.tandfonline.com/r/jbes*.

would converge to a stable distribution asymptotically rather than a normal distribution if such condition fails.

The empirical fact that financial returns generally have heavy tails, often leads to a violation of the conditional normality of the innovation error, which hereby results in a loss of efficiency for the GQMLE. For example, Bollerslev and Wooldbridge (1992) reported that the sample kurtosis of estimated residuals of GQMLE on S&P 500 monthly returns is 4.6, well exceeding the Gaussian kurtosis of 3. As a result, it is intuitively appealing to develop a QMLE based on heavy-tailed likelihoods, so that the loss in efficiency of GQMLE can be reduced.

In contrast to the majority of literature focusing on the GQMLE for inference, there is rather limited attention given to inference using a non-Gaussian QMLE (NGQMLE). This may be partly because the GQMLE is robust against error misspecification, whereas in general the QMLE with a non-Gaussian likelihood family does not yield consistent estimates unless the true innovation density happens to be a member of this likelihood family. Newey and Steigerwald (1997) first considered the identification condition for consistency of the heteroscedastic parameter with an NGQMLE in general conditional heteroscedastic models. They also point out that it is the scale parameter that may not be identified.

A potential remedy for the NGQMLE could be changing model assumptions to maintain a consistent estimation. Berkes and Horváth (2004) showed that an NGQMLE would obtain consistency and asymptotic normality with a different moment condition for the innovation rather than $E(\varepsilon^2) = 1$. However, the condition $E(\varepsilon^2) = 1$ is essential because it enables v_t to bear the natural interpretation of the conditional standard deviation, the notion of volatility. More importantly, a moment condition is part of the model specification, and it should be independent of and determined prior to the choice of likelihood functions. Related works also include Peng and Yao (2003) and Huang, Wang, and Yao (2008) for LAD type of estimators.

We prefer an NGQMLE method which is robust against density misspecification, more efficient than the GQMLE, and yet practical. The main contribution of this article is a novel three-step NGQMLE approach, which meets these desired properties. We introduce a scale adjustment parameter, denoted as η_f, for non-Gaussian likelihood to ensure the identification condition. In the first step, the GQMLE is conducted; then η_f is estimated through the residuals of the GQMLE; we feed the estimated $\widehat{\eta}_f$ into the NGQMLE in the final step. In GQMLE, η_f is always 1; but for NGQMLE, η_f is no longer equal to 1, and how much it deviates from 1 measures how much asymptotic bias would incur by simply using an NGQMLE without such an adjustment.

Also, we adopt the reparameterized GARCH model which separates the volatility scale parameter from the heteroscedastic parameter (see also Newey and Steigerwald 1997 and Drost and Klaassen 1997). Such a parameterization simplifies our derivation for asymptotic normality. The results show that our approach is more efficient than the GQMLE under various innovations. Furthermore, for the heteroscedastic parameter we can always achieve $\sqrt{T}$-consistency and asymptotic normality, whereas the GQMLE has a slower convergence rate when the innovation does not have a fourth moment.

Independently from our work, Francq, Lepage, and Zakoïan (2011) constructed a two-stage non-Gaussian QMLE, which allows the use of generalized Gaussian likelihood. Their estimator, though constructed in a different way compared to ours, turns out to be asymptotically equivalent to our estimator when our likelihood function is selected from the generalized Gaussian class. Our framework, however, can naturally accommodate more general models and likelihood functions. Related work also includes Lee and Lee (2009), where the likelihood is chosen from a parametric mixture Gaussian. Their estimator requires additional optimization over mixture probabilities, and hence is computationally more expensive than ours. Recently, Andrews (2012) constructed a rank-based estimator for the heteroscedastic parameter, which encompasses similar robustness compared to our procedure.

The outline of this article is as follows. Section 2 introduces the model and its assumptions. Section 3 constructs two feasible estimation strategies. Section 4 derives the asymptotic results. Section 5 focuses on the efficiency of the NGQMLE. Section 6 proposes extensions to further improve the asymptotic efficiency of the estimator. Section 7 employs Monte Carlo simulations to verify the theoretical results. Section 8 conducts real data analysis on S&P 500 stocks. Section 9 concludes. The Appendix provides key mathematical proofs.

2 MODEL SETUP

2.1 The GARCH Model

The reparameterized GARCH(p, q) model takes on the parametric form

$$x_t = \sigma v_t \varepsilon_t, \tag{3}$$

$$v_t^2 = 1 + \sum_{i=1}^{p} a_i x_{t-i}^2 + \sum_{j=1}^{q} b_j v_{t-j}^2. \tag{4}$$

The model parameters are summarized in $\boldsymbol{\theta} = \{\sigma, \boldsymbol{\gamma}'\}'$, where σ is the scale parameter and $\boldsymbol{\gamma} = (\boldsymbol{a}', \boldsymbol{b}')'$ is the heteroscedastic parameter. We use subscript 0 to denote the value under the true model throughout the article. The following standard assumptions for GARCH models are made.

Assumption 1. The true parameter $\boldsymbol{\theta_0}$ is in the interior of Θ, which is a compact subset of the $\boldsymbol{R}_+^{1+p+q}$, satisfying $\sigma > 0, a_i \geq 0, b_j \geq 0$. The innovation $\{\varepsilon_t, -\infty < t < \infty\}$ are iid random variables with mean 0, variance 1, and unknown density $g(\cdot)$. In addition, we assume that the GARCH process $\{x_t\}$ is strictly stationary and ergodic.

The elementary conditions for the stationarity and ergodicity of GARCH models have been discussed in Bougerol and Picard (1992). In this case, it immediately implies that

$$\begin{aligned} v_t^2 = {} & \frac{1}{1 - \sum_{j=1}^{q} b_j} + \sum_{i=1}^{p} a_i x_{t-i}^2 \\ & + \sum_{i=1}^{p} a_i \sum_{k=1}^{\infty} \sum_{j_1=1}^{q} \ldots \sum_{j_k=1}^{q} b_{j_1} \ldots b_{j_k} x_{t-i-j_1-\ldots-j_k}^2, \end{aligned} \tag{5}$$

and hence v_t is a function of $\bar{\boldsymbol{x}}_{\boldsymbol{t-1}} = \{x_s, -\infty < s \leq t-1\}$.

2.2 The Likelihood and the Scale Parameter

We consider a parametric family of quasi-likelihood $\{\eta : \frac{1}{\eta} f(\frac{\cdot}{\eta})\}$ indexed by $\eta > 0$, for any given likelihood function f. Here, η is used to adjust the scale of the quasi-likelihood. For a specific likelihood function f, the parameter η_f minimizes the discrepancy between the true innovation density g and the quasi-likelihood family in the sense of Kullback–Leibler Information Distance (KLID); see, for example, White (1982). Or equivalently,

$$\eta_f = \text{argmax}_{\eta>0} E\left\{ -\log\eta + \log f\left(\frac{\varepsilon}{\eta}\right)\right\}, \tag{6}$$

where the expectation is taken under the true density g.

Note that η_f only depends on the KLID of the two densities under consideration but not on the GARCH model. Once η_f is given, the NGQMLE $\widehat{\boldsymbol{\theta}}$ is defined by maximizing the following modified quasi-likelihood[1] with this model parameter η_f:

$$\begin{aligned} L_T(\boldsymbol{\theta}) &= \frac{1}{T}\sum_{t=1}^{T} l(\bar{\boldsymbol{x}}_t, \boldsymbol{\theta}) \\ &= \frac{1}{T}\sum_{t=1}^{T}\left(-\log(\eta_f \sigma v_t) + \log f\left(\frac{x_t}{\eta_f\sigma v_t}\right)\right). \end{aligned} \tag{7}$$

Equivalently, we in fact maximize the quasi-likelihood function selected from the parametric family $\{\eta : \frac{1}{\eta} f(\frac{\cdot}{\eta})\}$. Our approach is a generalization of the GQMLE and the MLE as illustrated in the next proposition.

Proposition 1. If $f \propto \exp(-x^2/2)$ or $f = g$, then $\eta_f = 1$.

Moreover, it can be implied from Newey and Steigerwald (1997) that in general an unscaled non-Gaussian likelihood function applied in this new reparameterization of GARCH(p, q) setting fails to identify the volatility scale parameter σ, resulting in inconsistent estimates. We show in the next section that incorporating η_f into the likelihood function facilitates the identification of the volatility scale parameter.

2.3 Identification

Identification is a critical step for consistency. It requires that the expected quasi-likelihood $\bar{L}(\boldsymbol{\theta}) = E(L_T(\boldsymbol{\theta}))$ has a unique maximum at the true parameter value $\boldsymbol{\theta_0}$. To show that $\boldsymbol{\theta}$ can be identified in the presence of η_f, we make the following assumptions.

Assumption 2. A quasi-likelihood of the GARCH (p, q) model is selected such that the function $Q(\eta) = -\log\eta + E(\log f(\varepsilon/\eta))$ has a unique maximizer $\eta_f > 0$.

This assumption is the key to the identification of σ_0, which literally means there exists a unique likelihood within the proposed family that has the smaller KLID than the rest of the family members.

Remark 1. A number of families of likelihoods, for instance, the Gaussian likelihood with $f \propto e^{-x^2/2}$, standardized t_ν-distribution with $f \propto (1 + x^2/(\nu-2))^{-(\nu+1)/2}$ and $\nu > 2$, and a generalized Gaussian likelihood with $\log f(x) = -|x|^\beta(\Gamma(3/\beta)/\Gamma(1/\beta))^{\beta/2} + \text{const}$, satisfy the requirement of Assumption 2.

Lemma 1. Given Assumption 2, $\bar{L}(\boldsymbol{\theta})$ has a unique maximum at the true value $\boldsymbol{\theta} = \boldsymbol{\theta_0}$.

Therefore, the identification condition is guaranteed, clearing the way for the consistent estimation of all parameters. Moreover, it turns out that model parameter η_f has another interpretation as a bias correction for a simple NGQMLE of the scale parameter in that $\sigma_0\eta_f$ would be reported instead of σ_0. Therefore, a simple implementation without η_f can consistently estimate σ_0 if and only if $\eta_f = 1$. Proposition 1 hence reveals the distinction in identification between the MLE, GQMLE, and QMLE based on alternative distributions.

In general, for an arbitrary likelihood, η_f may not equal 1. It is therefore necessary to incorporate this bias-correction factor η_f into NGQMLE. However, as we have no prior information concerning the true innovation density, η_f is unknown. Next, we propose a feasible three-step procedure that estimates η_f.

3 FEASIBLE ESTIMATION STRATEGIES

3.1 Three-Step Estimation Procedure

To estimate η_f, a sample on the true innovation is desired. According to Proposition 1, without knowing η_f, the residuals from the GQMLE may serve as substitutes for the true innovation sample. In the first step, we conduct Gaussian quasi-likelihood estimation:

$$\begin{aligned} \widetilde{\boldsymbol{\theta}}_{\boldsymbol{T}} &= \text{argmax}_{\boldsymbol{\theta}} \frac{1}{T}\sum_{t=1}^{T} l_1(\bar{\boldsymbol{x}}_t, \boldsymbol{\theta}) \\ &= \text{argmax}_{\boldsymbol{\theta}} \frac{1}{T}\sum_{t=1}^{T}\left(-\log(\sigma v_t) - \frac{x_t^2}{2\sigma^2 v_t^2}\right), \end{aligned} \tag{8}$$

then $\widehat{\eta}_f$ is obtained by maximizing Equation (6) with estimated residuals from the first step:

$$\begin{aligned} \widehat{\eta}_f &= \text{argmax}_{\eta} \frac{1}{T}\sum_{t=1}^{T} l_2(\bar{\boldsymbol{x}}_t, \widetilde{\boldsymbol{\theta}}_{\boldsymbol{T}}, \eta) \\ &= \text{argmax}_{\eta} \frac{1}{T}\sum_{t=1}^{T}\left(-\log(\eta) + \log f\left(\frac{\widetilde{\varepsilon}_t}{\eta}\right)\right), \end{aligned} \tag{9}$$

where $\widetilde{\varepsilon}_t = x_t/(\widetilde{\sigma} v_t(\widetilde{\boldsymbol{\gamma}}))$ is the residuals from GQMLE in the first step. Finally, we maximize non-Gaussian quasi-likelihood

[1] The likelihoods written here and below assume initial values $x_0, \ldots, x_{1-q}, v_0(\theta_0), \ldots, v_{1-p}(\theta_0)$ are given. Empirically, we can take $x_0 = \cdots = x_{1-q} = v_0 = \cdots = v_{1-p} = 1$. This would not affect the asymptotic properties of our estimator, which can be shown using similar arguments in Berkes, Horváth, and Kokoszka (2003) and Hall and Yao (2003).

with plug-in $\widehat{\eta}_f$ and obtain $\widehat{\boldsymbol{\theta}}_{\boldsymbol{T}}$:

$$\begin{aligned}\widehat{\boldsymbol{\theta}}_{\boldsymbol{T}} &= \operatorname{argmax}_{\boldsymbol{\theta}} \frac{1}{T}\sum_{t=1}^{T} l_3(\bar{\boldsymbol{x}}_{\boldsymbol{t}}, \widehat{\eta}_f, \boldsymbol{\theta}) \\ &= \operatorname{argmax}_{\boldsymbol{\theta}} \frac{1}{T}\sum_{t=1}^{T}\left(-\log(\widehat{\eta}_f \sigma v_t) + \log f\left(\frac{x_t}{\widehat{\eta}_f \sigma v_t}\right)\right).\end{aligned} \tag{10}$$

We call $\widehat{\boldsymbol{\theta}}_{\boldsymbol{T}}$ the NGQMLE estimator.

3.2 GMM Implementation

Alternatively, the above three-step procedure can be viewed as a one-step generalized methods of moments (GMM) procedure, by considering the score functions. Denote

$$\widetilde{\boldsymbol{s}}(\bar{\boldsymbol{x}}_{\boldsymbol{t}}, \boldsymbol{\theta}, \eta, \boldsymbol{\phi}) = (s_1(\bar{\boldsymbol{x}}_{\boldsymbol{t}}, \boldsymbol{\theta}), s_2(\bar{\boldsymbol{x}}_{\boldsymbol{t}}, \boldsymbol{\theta}, \eta), s_3(\bar{\boldsymbol{x}}_{\boldsymbol{t}}, \eta, \boldsymbol{\phi}))',$$

where

$$\begin{aligned}&\boldsymbol{s_1}(\bar{\boldsymbol{x}}_{\boldsymbol{t}}, \boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} l_1(\bar{\boldsymbol{x}}_{\boldsymbol{t}}, \boldsymbol{\theta}), \quad s_2(\bar{\boldsymbol{x}}_{\boldsymbol{t}}, \boldsymbol{\theta}, \eta) = \frac{\partial}{\partial \eta} l_2(\bar{\boldsymbol{x}}_{\boldsymbol{t}}, \boldsymbol{\theta}, \eta), \\ &\boldsymbol{s_3}(\bar{\boldsymbol{x}}_{\boldsymbol{t}}, \eta, \boldsymbol{\phi}) = \frac{\partial}{\partial \boldsymbol{\phi}} l_3(\bar{\boldsymbol{x}}_{\boldsymbol{t}}, \eta, \boldsymbol{\phi}),\end{aligned} \tag{11}$$

then the NGQMLE amounts to an exactly identified GMM with the moment condition

$$E(\widetilde{\boldsymbol{s}}(\bar{\boldsymbol{x}}_{\boldsymbol{t}}, \boldsymbol{\theta}, \eta, \boldsymbol{\phi})) = \boldsymbol{0},$$

which can be implemented by minimizing

$$\begin{aligned}(\widetilde{\boldsymbol{\theta}}_{\boldsymbol{T}}, \widehat{\eta}_f, \widehat{\boldsymbol{\phi}}_{\boldsymbol{T}}) = \operatorname{argmin}_{\boldsymbol{\theta},\eta,\boldsymbol{\phi}} &\left(\frac{1}{T}\sum_{t=1}^{T} \widetilde{\boldsymbol{s}}(\bar{\boldsymbol{x}}_{\boldsymbol{t}}, \boldsymbol{\theta}, \eta, \boldsymbol{\phi})\right)' \\ &\times \left(\frac{1}{T}\sum_{t=1}^{T} \widetilde{\boldsymbol{s}}(\bar{\boldsymbol{x}}_{\boldsymbol{t}}, \boldsymbol{\theta}, \eta, \boldsymbol{\phi})\right).\end{aligned} \tag{12}$$

Thus, our proposed estimator is simply $\widehat{\boldsymbol{\theta}}_{\boldsymbol{T}} = \widehat{\boldsymbol{\phi}}_{\boldsymbol{T}}$.

It is worth pointing out that the proposed estimation strategies work not only for the standard GARCH model. In fact, the definition of η_f and the estimation strategies do not depend on the particular function form of $v_t(\boldsymbol{\gamma})$. As long as $v_t(\boldsymbol{\gamma})/v_t(\boldsymbol{\gamma}_0)$ is not a constant for $\boldsymbol{\gamma} \neq \boldsymbol{\gamma}_0$, the model can be identified—see, for example, Newey and Steigerwald (1997)—and our estimation procedure carries over. Hence, our three-step QMLE can handle more general models than those considered by Francq, Lepage, and Zakoïan (2011) and Lee and Lee (2009). The same framework can be extended to multivariate GARCH models, as suggested by Fiorentini and Sentana (2010).

4 ASYMPTOTIC THEORY

We now develop some asymptotic theory to reveal the difference in efficiency between the GQMLE, the NGQMLE, and the MLE. In an effort to demonstrate the idea without delving into mathematical details, for convenience, we make the following regularity condition.

Assumption 3. Let $h(x, \eta) = \log f(x/\eta) - \log \eta$ with $\eta > 0$, $\boldsymbol{k} = (1/\sigma, 1/v_t \cdot \partial v_t/\partial \boldsymbol{\gamma}')'$, and $\boldsymbol{k_0}$ be its value at $\theta = \theta_0$.

1. $h(x, \eta)$ is continuously differentiable up to the second order with respect to η.
2. For any $\eta > 0$, we have

$$E \sup_{\boldsymbol{\theta}\in\mathcal{N}} ||h_1(x_t/\sigma_t(\boldsymbol{\theta}), \eta)\boldsymbol{k}|| < \infty,$$
$$E \sup_{\boldsymbol{\theta}\in\mathcal{N}} ||\nabla_{\boldsymbol{\theta}}(h_1(x_t/\sigma_t(\boldsymbol{\theta}), \eta)\boldsymbol{k})|| < \infty$$

for some neighborhood $\mathcal{N}$ of $\boldsymbol{\theta_0}$, where $h_1(x, \eta)$ is the first-order derivative of $h(x, \eta)$ with respect to η, and $\sigma_t(\boldsymbol{\theta}) = \sigma v_t(\boldsymbol{\theta})$.

3. $0 < E(h_1(\varepsilon, \eta_f))^2 < \infty$, $0 < E|h_2(\varepsilon, \eta_f)| < \infty$, where $h_2(x, \eta)$ is the second-order derivative of $h(x, \eta)$ with respect to η.

The first requirement in Assumption 3 is more general than the usual second-order continuously differentiable condition on f. For example, the generalized Gaussian likelihood family does not have a second-order derivative at 0 when β is smaller than 2. However, members of this likelihood family certainly satisfy Assumption 3.

Identification for the parameters $\boldsymbol{\theta}$ and η jointly is straightforward in view of Lemma 1. The consistency thereby holds.

Theorem 1. Given Assumptions 1, 2, and 3, $(\widetilde{\boldsymbol{\theta}}_{\boldsymbol{T}}, \widehat{\eta}_f, \widehat{\boldsymbol{\theta}}_{\boldsymbol{T}}) \xrightarrow{\mathcal{P}} (\boldsymbol{\theta_0}, \eta_f, \boldsymbol{\theta_0})$, in particular the NGQMLE $\widehat{\boldsymbol{\theta}}_{\boldsymbol{T}}$ is consistent.

To obtain the asymptotic normality, we realize that a finite fourth moment for the innovation is essential in that the first step employs the GQMLE. Although alternative rate efficient estimators may be adopted to avoid moment conditions required in the first step, we prefer the GQMLE for its simplicity and popularity in practice. Theorem 3 in Section 5 discusses the situation without this moment condition.

Theorem 2. Assume that $E(\varepsilon^4) < \infty$ and that Assumptions 1, 2, and 3 are satisfied. Then $(\widetilde{\boldsymbol{\theta}}_{\boldsymbol{T}}, \widehat{\eta}_f, \widehat{\boldsymbol{\theta}}_{\boldsymbol{T}})$ are jointly normal asymptotically. That is,

$$\begin{pmatrix} T^{\frac{1}{2}}(\widetilde{\boldsymbol{\theta}}_{\boldsymbol{T}} - \boldsymbol{\theta_0}) \\ T^{\frac{1}{2}}(\widehat{\eta}_f - \eta_f) \\ T^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}}_{\boldsymbol{T}} - \boldsymbol{\theta_0}) \end{pmatrix} \xrightarrow{\mathcal{L}} N\left(\begin{pmatrix} \boldsymbol{0} \\ 0 \\ \boldsymbol{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma_G} & \boldsymbol{\Pi}' & \boldsymbol{\Xi} \\ \boldsymbol{\Pi} & \Sigma_{\eta_f} & \boldsymbol{\Pi} \\ \boldsymbol{\Xi} & \boldsymbol{\Pi}' & \boldsymbol{\Sigma_2} \end{pmatrix}\right),$$

where $\boldsymbol{M} = E(\boldsymbol{k_0}\boldsymbol{k_0}')$ with $\boldsymbol{k_0}$ defined in Assumption 3,

$$\boldsymbol{\Sigma_G} = \frac{E(\varepsilon^2 - 1)^2}{4}\boldsymbol{M}^{-1}, \tag{13}$$

$$\begin{aligned}\boldsymbol{\Sigma_2} = &\frac{E(h_1(\varepsilon_t, \eta_f))^2}{\eta_f^2 (Eh_2(\varepsilon_t, \eta_f))^2}\boldsymbol{M}^{-1} \\ &+ \sigma_0^2\left(\frac{E(\varepsilon^2 - 1)^2}{4} - \frac{E(h_1(\varepsilon_t, \eta_f))^2}{\eta_f^2 (Eh_2(\varepsilon_t, \eta_f))^2}\right)\boldsymbol{e_1}\boldsymbol{e_1'},\end{aligned} \tag{14}$$

$$\Sigma_{\eta_f} = \eta_f^2 E\left(\frac{\varepsilon^2-1}{2} + \frac{h_1(\varepsilon,\eta_f)}{\eta_f E h_2(\varepsilon,\eta_f)}\right)^2, \tag{15}$$

$$\boldsymbol{\Pi} = \frac{\eta_f\sigma_0}{2} E\left((1-\varepsilon^2)\left(\frac{h_1(\varepsilon,\eta_f)}{\eta_f E h_2(\varepsilon,\eta_f)} + \frac{\varepsilon^2-1}{2}\right)\right)\boldsymbol{e}_1', \tag{16}$$

$$\begin{aligned}\boldsymbol{\Xi} &= \frac{E(h_1(\varepsilon,\eta_f)\cdot(1-\varepsilon^2))}{2\eta_f E(h_2(\varepsilon,\eta_f))}\boldsymbol{M}^{-1} \\ &\quad - \frac{\sigma_0^2}{2} E\left((1-\varepsilon^2)\left(\frac{h_1(\varepsilon,\eta_f)}{\eta_f E h_2(\varepsilon,\eta_f)} + \frac{\varepsilon^2-1}{2}\right)\right)\boldsymbol{e}_1\boldsymbol{e}_1', \end{aligned}\tag{17}$$

and $\boldsymbol{e}_1$ is a unit column vector that has the same length as $\boldsymbol{\theta}$, with the first entry one and all the rest zeros. In addition, in the case where η_f is known, the benchmark asymptotic variance of the infeasible estimator $\widehat{\boldsymbol{\theta}}_T$ is given by

$$\boldsymbol{\Sigma}_1 = \boldsymbol{M}^{-1}\frac{E(h_1(\varepsilon_t,\eta_f))^2}{\eta_f^2(Eh_2(\varepsilon_t,\eta_f))^2}. \tag{18}$$

Remark 2. If we select $\log f = \text{const} - |x|^\beta/\beta$, then we have $\eta_f = (E|x|^\beta)^{1/\beta}$, and

$$\begin{aligned}\boldsymbol{\Sigma}_2 &= \frac{E|x|^{2\beta} - (E|x|^\beta)^2}{\beta^2(E|x|^\beta)^2}\boldsymbol{M}^{-1} \\ &\quad + \sigma_0^2\left(\frac{E(\varepsilon^2-1)^2}{4} - \frac{E|x|^{2\beta} - (E|x|^\beta)^2}{\beta^2(E|x|^\beta)^2}\right)\boldsymbol{e}_1\boldsymbol{e}_1'.\end{aligned}$$

This indicates that our estimator has the same asymptotic efficiency as the estimator given by Francq, Lepage, and Zakoïan (2011), see Theorem 2.1 therein.

5 EFFICIENCY OF THE NGQMLE

Before a thorough efficiency analysis of the NGQMLE $\widehat{\boldsymbol{\theta}}_T$, it is important to discuss the asymptotic property of $\widehat{\eta}_f$. Although $\widehat{\eta}_f$ is obtained using fitted residuals $\widetilde{\varepsilon}_t$ in Equation (9), the asymptotic variance of $\widehat{\eta}_f$ is not necessarily worse than that using the actual innovations ε_t. In fact, using true innovation ε_t, the asymptotic variance of $\widehat{\eta}_f$ is $E(h_1(\varepsilon,\eta_f))^2/(Eh_2(\varepsilon,\eta_f))^2$. Comparing it with Equation (15), we can find that using fitted residuals may obtain better efficiency. One extreme case is to choose the Gaussian likelihood in the second step. Then η_f exactly equals one and the asymptotic variance of $\widehat{\eta}_f$ vanishes.

The parameter η_f also reveals the issue of asymptotic bias incurred by simply using unscaled NGQMLE. From Equation (10), while NGQMLE $\widehat{\boldsymbol{\theta}}_T = (\widehat{\sigma}_T, \widehat{\boldsymbol{\gamma}}_T)$ maximizes the log-likelihood, unscaled NGQMLE would choose the estimator $(\widehat{\eta_f\sigma_T}, \widehat{\boldsymbol{\gamma}}_T)$ to maximize log-likelihood. We can see that the estimation of σ is biased by a factor of $\widehat{\eta}_f$. Such bias will propagate if using the popular original parameterization. Recall

$$\begin{aligned} x_t &= \sigma_t\varepsilon_t, \\ \sigma_t^2 &= \widetilde{c} + \sum_{i=1}^p \widetilde{a}_i x_{t-i}^2 + \sum_{j=1}^q \widetilde{b}_j\sigma_{t-j}^2. \end{aligned}$$

Clearly, we have $\sigma^2 a_i = \widetilde{a}_i$, $b_j = \widetilde{b}_j$ and $\sigma^2 = c$. Therefore, potential model misspecification would result in systematic biases in the estimates of a_i and c if unscaled NGQMLE is applied without η_f. We will highlight this bias in our empirical study.

5.1 Efficiency Gain Over GQMLE

We compare the efficiency of three estimators of $\boldsymbol{\theta}$ including the three-step NGQMLE, one-step (infeasible) NGQMLE with known η_f, and the GQMLE. Their asymptotic variances are $\boldsymbol{\Sigma}_2$, $\boldsymbol{\Sigma}_1$, and $\boldsymbol{\Sigma}_G$, respectively. The difference in asymptotic variances between the first two estimators is

$$\boldsymbol{\Sigma}_2 - \boldsymbol{\Sigma}_1 = \begin{pmatrix} \mu\sigma_0^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{pmatrix}, \tag{19}$$

where

$$\mu = \frac{E(\varepsilon^2-1)^2}{4} - \frac{E(h_1(\varepsilon,\eta_f))^2}{\eta_f^2(Eh_2(\varepsilon,\eta_f))^2}. \tag{20}$$

Effectively, the sign and magnitude of μ summarize the advantage of knowing η_f. μ is usually positive when the true error has heavy tails, while NGQMLE is selected to be a heavy-tailed likelihood, illustrating the loss from not knowing η_f. However, it could also be negative when the true innovation has thin tails, indicating that not knowing η_f is actually better when a heavy tail density is selected. Intuitively, this is because the three-step estimator incorporates a more efficient GQMLE into the estimation procedure. More importantly, the asymptotic variance of $\boldsymbol{\gamma}$ and the covariance between σ and $\boldsymbol{\gamma}$ are not affected by the estimation of η_f. In other words, we achieve the adaptivity property for $\boldsymbol{\gamma}$: with an appropriate NGQMLE, $\boldsymbol{\gamma}$ could be estimated without knowing η_f equally well as if η_f were known.

We next compare the efficiency between GQMLE and NGQMLE. From Equation (18) with f replaced by the Gaussian likelihood, we have

$$\boldsymbol{\Sigma}_G = \frac{E(\varepsilon^2-1)^2}{4}\boldsymbol{M}^{-1}.$$

It follows from Lemma 2 in the Appendix that,

$$\boldsymbol{\Sigma}_G - \boldsymbol{\Sigma}_2 = \mu\begin{pmatrix} \sigma_0^2\bar{\boldsymbol{y}}_0'\boldsymbol{V}\bar{\boldsymbol{y}}_0 & -\sigma_0\bar{\boldsymbol{y}}_0'\boldsymbol{V} \\ -\sigma_0\boldsymbol{V}\bar{\boldsymbol{y}}_0 & \boldsymbol{V}\end{pmatrix}, \tag{21}$$

where $\boldsymbol{y}_0 = v_t(\boldsymbol{\gamma}_0)\frac{\partial v_t(\boldsymbol{\gamma}_0)}{\partial\boldsymbol{\gamma}}$, $\bar{\boldsymbol{y}}_0 = E(\boldsymbol{y}_0)$, $\boldsymbol{V} = \text{var}(\boldsymbol{y}_0)^{-1}$ and hereby the last matrix in Equation (21) is positive definite. Therefore, as long as μ is positive, NGQMLE is more efficient for both σ and $\boldsymbol{\gamma}$.

To summarize the variation of μ, one can draw a line for distributions according to their asymptotic behavior of tails, in other words, according to how heavy their tails are, with thin tails on the left and heavy tails on the right. Then we place the non-Gaussian likelihood, Gaussian likelihood, and true innovation distribution onto this line. The sign and value of μ usually depend on where the true innovation distribution is placed. (a) If it is placed on the right side of non-Gaussian likelihood, then μ is positive and large. (b) If error is on the left side of Gaussian likelihood, then μ is negative and large in absolute value. (c) If error is between non-Gaussian and Gaussian, then the sign of μ can be positive or negative, depending on which distribution in the likelihood is closer to that of the innovation. This seems like a symmetric argument for Gaussian and non-Gaussian likelihood. But in financial applications we know true innovations are heavy-tailed. Even though the non-Gaussian likelihood may not be the innovation distribution, we still can guarantee either (a) happens or (c) happens with innovation closer to a non-Gaussian likelihood. In both cases, we have $\mu > 0$ and NGQMLE is a more efficient procedure than GQMLE.

5.2 Efficiency Gap From the MLE

Denote the asymptotic variance of the MLE as $\boldsymbol{\Sigma_M}$. From Equation (18) with f replaced by the true likelihood g, we have

$$\boldsymbol{\Sigma_M} = \boldsymbol{M}^{-1}\left(\int_{-\infty}^{+\infty} x^2 \frac{\dot{g}^2}{g} dx - 1\right)^{-1} = \boldsymbol{M}^{-1}\left(E\left(h_g^2 - 1\right)\right)^{-1},$$

where $h_g = x\dot{g}(x)/g(x)$. The gap in asymptotic variance between NGQMLE and MLE is given by

$$\begin{aligned}\boldsymbol{\Sigma_2} - \boldsymbol{\Sigma_M} &= \left(\frac{Eh_1(\varepsilon, \eta_f)^2}{\eta_f^2(Eh_2(\varepsilon, \eta_f))^2} - \left(E\left(h_g^2 - 1\right)\right)^{-1}\right)\boldsymbol{M}^{-1} \\ &\quad + \boldsymbol{\sigma_0^2}\left(\frac{E(\varepsilon^2 - 1)^2}{4} - \frac{Eh_1(\varepsilon, \eta_f)^2}{\eta_f^2(Eh_2(\varepsilon, \eta_f))^2}\right)\boldsymbol{e_1 e_1'}. \qquad (22)\end{aligned}$$

An extreme case is that the selected likelihood f happens to be the true innovation density. Being unaware of this, we would still apply a three-step procedure and use the estimated η_f. Therefore, the first term in Equation (22) vanishes, but the second term remains. Consequently, $\widehat{\boldsymbol{\gamma}}$ reaches the efficiency bounds, while the volatility scale $\widehat{\sigma}$ fails, which reflects the penalty of ignorance of the true model. This example also sheds light on the fact that $\widehat{\boldsymbol{\theta}}_{\boldsymbol{T}}$ cannot obtain the efficiency bounds for all parameters unless the true underlying density and the selected likelihood are both Gaussian. This observation agrees with the comparison study in the González-Rivera and Drost (1999) concerning the MLE and their semiparametric estimator.

5.3 The Effect of the First-Step Estimation

We would like to further explore the adaptivity property of the estimator for the heteroscedastic parameter by considering a general first-step estimator. We have shown in Theorem 2 that the efficiency of the estimator for $\boldsymbol{\gamma}$ is not affected by the first-step estimation of η_f using the GQMLE as if η_f were known. The moment conditions given by Assumption 3(ii) depend on the tail of the innovation density g and quasi-likelihood f. It is well known that the asymptotic normality of Gaussian likelihood requires a finite fourth moment. In contrast, Remark 1 implies that any Student's t likelihood with degree of freedom larger than 2 has a bounded second moment, so that no additional moment conditions are needed for the asymptotic normality of NGQMLE. Therefore, we may relax the finite fourth moment requirement on the innovation error by applying another efficient estimator in the first step. Moreover, even if the first step estimator has a lower rate, it may not affect the efficiency of the heteroscedastic parameter $\boldsymbol{\gamma}$, which is always $\sqrt{T}$-consistent and asymptotically normal.

Theorem 3. Given that Assumptions 1, 2, and 3 hold, and suppose that the first step estimator $\widetilde{\boldsymbol{\theta}}$ has an influence function representation:

$$T\lambda_T^{-1}(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta_0}) = \lambda_T^{-1}\sum_{t=1}^{T}\boldsymbol{\Psi_t}(\varepsilon_t) + \boldsymbol{o_P(1)},$$

with the right-hand side converging to a nondegenerate distribution, and $\lambda_T \sim T^{1/\alpha}$ for some $\alpha \in [1, 2]$. Then, the convergence rate for σ is also $T\lambda_T^{-1}$, while the same central limit theorem for γ as in Theorem 2 remains, that is,

$$T^{\frac{1}{2}}(\widehat{\boldsymbol{\gamma}}_{\boldsymbol{T}} - \boldsymbol{\gamma_0}) \xrightarrow{\mathcal{L}} N\left(\boldsymbol{0}, \frac{E(h_1(\varepsilon, \eta_f))^2}{\eta_f^2(Eh_2(\varepsilon, \eta_f))^2}\boldsymbol{V}\right),$$

where $\boldsymbol{V} = (\mathrm{var}(\frac{1}{\nu(\boldsymbol{\gamma_0})}\frac{\partial\nu}{\partial\boldsymbol{\gamma}}|_{\boldsymbol{\gamma}=\boldsymbol{\gamma_0}}))^{-1}$.

Theorem 3 applies to several estimators that have been discussed in the literature. For example, Hall and Yao (2003) discussed the GQMLE with ultra heavy-tailed innovations that violate a finite fourth moment. In their analysis, λ_T is regularly varying at infinity with exponent $\alpha \in [1, 2)$. The resulting GQMLE $\widetilde{\boldsymbol{\theta}}$ suffers lower convergence rates. By contrast, Drost and Klaassen (1997) suggested an M-estimator based on the score function for logistic distribution to avoid moment conditions on the innovations. Both estimators, if applied in the first step, would not affect the efficiency of $\widehat{\boldsymbol{\gamma}}_{\boldsymbol{T}}$.

An infinite fourth moment is not a rare case in empirical financial data modeling. Under such a circumstance, the NGQMLE procedure surely outperforms standard GQMLE, in that NGQMLE can achieve $\sqrt{T}$-rate of convergence for the heteroscedastic parameter $\boldsymbol{\gamma}$, as shown in Theorem 3. Moreover, the asymptotic variance of $\widehat{\boldsymbol{\gamma}}_{\boldsymbol{T}}$ maintains the same form as in Theorem 2. By contrast, the GQMLE suffers a lower rate of convergence for all parameters including $\boldsymbol{\gamma}$ (see Hall and Yao 2003). The key to efficiency gains, which becomes more obvious in the case of $E\varepsilon^4 = \infty$, is to use a likelihood that better mimics the tail nature of innovation error.

6 EXTENSIONS

Here, we discuss two ways to further improve the efficiency of NGQMLE. One is choosing the non-Gaussian likelihood from a pool of candidate likelihoods to adapt to data. The other is an affine combination of NGQMLE and GQMLE according to their covariance matrix in Theorem 2 to minimize the resulting estimator's asymptotic variance.

6.1 Choice of Likelihood

There are two distinctive advantages of choosing a heavy-tailed quasi-likelihood over Gaussian likelihood. First, the $\sqrt{T}$-consistency of NGQMLE of $\boldsymbol{\gamma}$ no longer depends on the finite fourth moment condition, but instead finite $E(h_1(\varepsilon, \eta_f))^2/(\eta_f Eh_2(\varepsilon, \eta_f))^2$. This can be easily met by, for example, choosing generalized Gaussian likelihood with $\beta \leq 1$. Second, even with a finite fourth moment, a heavy-tailed NGQMLE has lower variance than GQMLE if true innovation is heavy-tailed. A prespecified heavy-tailed likelihood can have these two advantages. However, we can adaptively choose this quasi-likelihood to further improve its efficiency. This is done by minimizing the efficiency loss from MLE, which is equivalent to minimizing $E(h_1(\varepsilon,\eta_f))^2/(\eta_f Eh_2(\varepsilon,\eta_f))^2$ over certain families of heavy-tailed likelihoods. We propose an optimal choice of non-Gaussian likelihoods, where candidate likelihoods are from a Student's t family $\{f_\nu^t\}$ with the degree of freedom $\nu > 2$ and generalized Gaussian family $\{f_\beta^{gg}\}$ with $\beta \le 1$. Formally, for true innovation distribution g and candidate likelihood f, define

$$A(f,g) = \frac{E_g h_1(\varepsilon,\eta_f)^2}{\eta_f^2 E_g(h_2(\varepsilon,\eta_f))^2}. \tag{23}$$

Then the optimal likelihood from the t-family and generalized Gaussian family (gg) is chosen:

$$f^* = \text{argmin}_{\nu,\beta}\left\{\left\{A\left(f_\nu^t,\widehat{g}\right)\right\}_{\nu>2},\left\{A\left(f_\beta^{gg},\widehat{g}\right)\right\}_{\beta\le 1}\right\}, \tag{24}$$

where $\widehat{g}$ denotes the empirical distribution of estimated residuals from GQMLE in the first step. Because this procedure of choosing likelihood is adaptive to data, it is expected that the chosen quasi-likelihood results in a more efficient NGQMLE than a prespecified one.

6.2 Aggregating NGQMLE and GQMLE

Another way to further improve the efficiency of NGQMLE is through aggregation. Since both GQMLE and NGQMLE are consistent, an affine combination of the two, with weights chosen according to their joint asymptotic variance, yields a consistent estimator and is more efficient than both. Define the aggregation estimator

$$\widehat{\boldsymbol{\theta}}_{\boldsymbol{T}}^{\boldsymbol{W}} = \boldsymbol{W}\widehat{\boldsymbol{\theta}} + (\boldsymbol{I}-\boldsymbol{W})\widetilde{\boldsymbol{\theta}}, \tag{25}$$

where $\boldsymbol{W}$ is a diagonal matrix with weights $(w_1, w_2, \ldots, w_{1+p+q})$ on the diagonal. From Theorem 2, the optimal weights are chosen from minimizing the asymptotic variance of each component of the aggregation estimator:

$$\begin{aligned} w_i^* &= \text{argmin}_w w^2(\boldsymbol{\Sigma_2})_{i,i} + (1-w)^2(\boldsymbol{\Sigma_G})_{i,i} + 2w(1-w)\boldsymbol{\Xi}_{i,i} \\ &= \frac{(\boldsymbol{\Sigma_G})_{i,i} - \boldsymbol{\Xi}_{i,i}}{(\boldsymbol{\Sigma_2})_{i,i} + (\boldsymbol{\Sigma_G})_{i,i} - 2\boldsymbol{\Xi}_{i,i}}. \end{aligned} \tag{26}$$

It turns out that all optimal aggregation weights w_i^* are the same, which is

$$w^* = \frac{E\left(\frac{1-\varepsilon^2}{2}\left(\frac{1-\varepsilon^2}{2} - \frac{h_1(\varepsilon,\eta_f)}{\eta_f Eh_2(\varepsilon,\eta_f)}\right)\right)}{E\left(\frac{1-\varepsilon^2}{2} - \frac{h_1(\varepsilon,\eta_f)}{\eta_f Eh_2(\varepsilon,\eta_f)}\right)^2}. \tag{27}$$

Proposition 2. The optimal weight of the aggregated estimator satisfies $\boldsymbol{W}^* = w^*\boldsymbol{I}$. Given any consistent estimator $\widehat{\omega}^*$ of ω^*, the asymptotic variance of the aggregated estimator $\widehat{\boldsymbol{\theta}}_{\boldsymbol{T}}^{*} = \widehat{\boldsymbol{W}}^{*}\widehat{\boldsymbol{\theta}} + (\boldsymbol{I}-\widehat{\boldsymbol{W}}^{*})\widetilde{\boldsymbol{\theta}}$ has diagonal terms

$$\boldsymbol{\Sigma}_{i,i}^* = \frac{(\boldsymbol{\Sigma_2})_{i,i}(\boldsymbol{\Sigma_G})_{i,i} - \boldsymbol{\Xi}_{i,i}^2}{(\boldsymbol{\Sigma_2})_{i,i} + (\boldsymbol{\Sigma_G})_{i,i} - 2\boldsymbol{\Xi}_{i,i}}, \quad i = 1,\ldots,1+p+q. \tag{28}$$

Although estimators for σ and γ have different asymptotic properties, the optimal aggregation weights are the same: $w_i^* = w^*$. Also, the weight depends only on non-Gaussian likelihood and innovation distribution, but not on GARCH model specification. The aggregated estimator $\widehat{\boldsymbol{\theta}}_{\boldsymbol{T}}^{*}$ always has a smaller asymptotic variance than both NGQMLE and GQMLE. If data are heavy-tailed, for example, $E\varepsilon^4$ is large or equal to ∞, from Equation (27) it simply assigns weights of approximately 1 for NGQMLE and 0 for GQMLE. In practice, after running NGQMLE with an optimal choice of likelihood, we can estimate the optimal aggregation weight w^* by plugging into Equation (27) the estimated residuals.

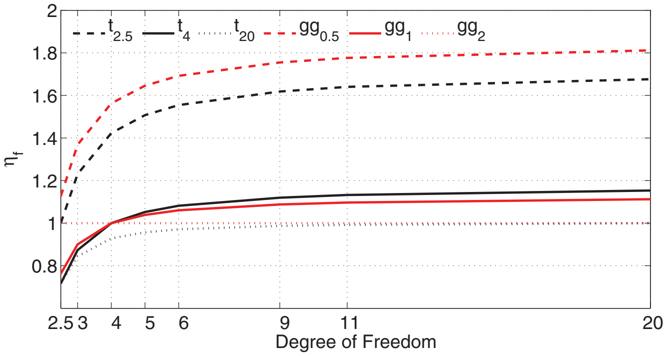

Figure 1. Variations of η_f across Student's t innovations.

7 SIMULATION STUDIES

7.1 Variations of η_f

The scale parameter η_f is generic characteristic of non-Gaussian likelihoods and of the true innovations, and it does not change when using another conditional heteroscedastic model. Here, we numerically evaluate how it varies according to the non-Gaussian likelihoods and innovations.

Figures 1 and 2 show how η_f varies over generalized Gaussian likelihoods and Student's t likelihoods with different parameters, respectively. For each curve, which amounts to fixing a quasi-likelihood, the lighter the tails of innovation errors are, the larger η_f. Furthermore, $\eta_f > 1$ for innovation errors that are lighter than the likelihood, and $\eta_f < 1$ for innovations that are heavier than the likelihood. Therefore, if the non-Gaussian likelihood has heavier tails than true innovation, we should shrink the estimates to obtain consistent estimates. On the other hand, if the quasi-likelihood is lighter than true innovation, we should magnify the estimates.

Tables 1 and 2 provide more values of η_f. For each column (fix an innovation distribution), in most cases the heavier the

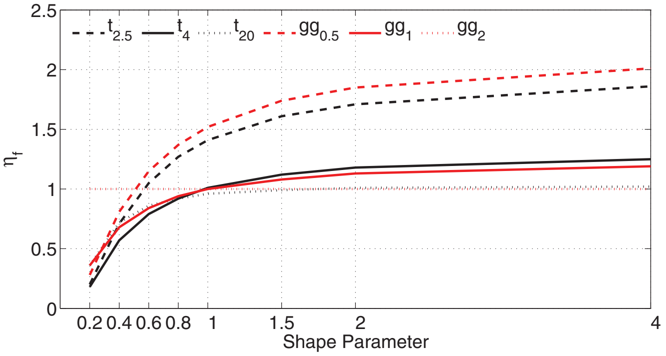

Figure 2. Variations of η_f across generalized Gaussian innovations.

Table 1. η_f for generalized GQMLEs (gg,row) and innovation distributions (column)

	$gg_{0.2}$	$gg_{0.6}$	gg_1	$gg_{1.4}$	$gg_{1.8}$	gg_2	t_3	t_5	t_7	t_{11}
$gg_{0.2}$	1.000	6.237	8.901	10.299	11.125	11.416	8.128	9.963	10.483	10.885
$gg_{0.6}$	0.271	1.000	1.291	1.434	1.515	1.544	1.159	1.384	1.443	1.487
$gg_{1.0}$	0.354	0.844	1.000	1.073	1.114	1.128	0.900	1.040	1.074	1.098
$gg_{1.4}$	0.537	0.873	0.962	1.000	1.022	1.029	0.883	0.977	0.998	1.012
$gg_{1.8}$	0.811	0.952	0.981	0.993	1.000	1.002	0.946	0.985	0.991	0.997

tails of likelihoods are, the larger η_f, but the monotone relationship is not true for some ultra heavy-tail innovations, in which cases η_f shows a "smile." The nonmonotonicity in the likelihood dimension indicates that to determine η_f one needs more information about the likelihood than just the asymptotic behavior of its tails. For example, both the true likelihood and Gaussian likelihood have η_f equal to 1.

7.2 Comparison With GQMLE, MLE, Rank, and Semiparametric Estimator

Here, we compare the efficiency of NGQMLE, GQMLE, MLE, and an adaptive estimator under various innovation error distributions. We fix the quasi-likelihood to be the Student's t distribution with a degree of freedom 7.

We include into the comparison the rank-based estimator by Andrews (2012). The estimator is obtained by minimizing the function D_T below:

$$D_T(\boldsymbol{\theta}) = \sum_{t=P+1}^{T} \lambda\left(\frac{R_t(\boldsymbol{\theta})}{T-p+1}\right)(\xi_t(\boldsymbol{\theta}) - \overline{\xi_t(\boldsymbol{\theta})}),$$

where $\lambda(\cdot)$ is a nonconstant and nondecreasing function from $(0, 1)$ to $\mathbb{R}$, $R_t(\boldsymbol{\theta})$ denotes the rank corresponding to $\xi_t(\boldsymbol{\theta}) = \log(\varepsilon_t^2(\boldsymbol{\theta}))$, and $\overline{\xi_t(\boldsymbol{\theta})} = (n-P)^{-1}\sum_{t=P+1}^{T}\xi_t(\boldsymbol{\theta})$. We follow Andrews (2012) to choose $\lambda(x)$ as $\{7(F_{t_7}^{-1}((x+1)/2))^2 - 5\}/\{(F_{t_7}^{-1}((x+1)/2))^2 + 5\}$, where $F_{t_7}^{-1}$ represents the distribution function for rescaled t_7 noise. The asymptotic relative efficiency of the NGQMLE against the rank-based estimator is given in Table 3, with less-than-one ratios in favor of the NGQMLE. From this table, we can see that the two estimators are almost indistinguishable asymptotically with the rank-based estimator slightly better. This observation does not undermine the appeal of the NGQMLE as it also estimates the scale parameter σ, and a by-product η_f, which gauges how large the bias of non-Gaussian likelihood estimator is without scale adjustment.

We also consider the semiparametric estimator proposed by Drost and Klaassen (1997) in the setting of GARCH(1, 1) with symmetric heavy-tailed errors. Their estimation is done in two steps. The first step runs a GQMLE to obtain estimates of parameter $\widetilde{\boldsymbol{\theta}} = (\widetilde{a}_1, \widetilde{b}_1)$ as well as the noise series $\{\widetilde{\varepsilon}_t\}$, with which they can construct the second-step estimator:

$$\begin{pmatrix}\widehat{a}_1\\ \widehat{b}_1\end{pmatrix} = \begin{pmatrix}\widetilde{a}_1\\ \widetilde{b}_1\end{pmatrix} + \begin{pmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\end{pmatrix} \times \left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{W}_t(\widetilde{\boldsymbol{\theta}})\Psi_t(\widetilde{\varepsilon})\Psi_t(\widetilde{\varepsilon})'\mathbf{W}_t(\widetilde{\boldsymbol{\theta}})'\right)^{-1} \times \frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{W}_t(\widetilde{\boldsymbol{\theta}}) - \frac{1}{T}\sum_{s=1}^{T}\mathbf{W}_s(\widetilde{\boldsymbol{\theta}})\right)\Psi_t(\widetilde{\varepsilon}),$$

where

$$\mathbf{W}_t(\boldsymbol{\theta}) = \begin{pmatrix}\mathbf{H_t}(\boldsymbol{\theta})\left(\mathbf{0}, \frac{\mathbf{1}}{\mathbf{2}}\mathbf{v_t}^{-\mathbf{2}}(\boldsymbol{\theta})\right)\\ \sigma^{-1}\boldsymbol{I}_2\end{pmatrix},$$

$$\mathbf{H}_t(\boldsymbol{\theta}) = \beta\mathbf{H}_{t-1}(\boldsymbol{\theta}) + \begin{pmatrix}x_{t-1}^2\\ v_{t-1}^2(\boldsymbol{\theta})\end{pmatrix}, \quad \mathbf{H}_1(\boldsymbol{\theta}) = \mathbf{0}_{2\times 1},$$

$$\boldsymbol{\Psi}_t(\widetilde{\varepsilon}) = -\begin{pmatrix}l'(\widetilde{\varepsilon}_t)\\ 1 + \widetilde{\varepsilon}_t l'(\widetilde{\varepsilon}_t)\end{pmatrix}, \quad l'(\widetilde{\epsilon}) = \frac{\widehat{g}'(\widetilde{\varepsilon}_t)}{\widehat{g}(\widetilde{\varepsilon}_t)}, \quad \text{and}$$

$$\widehat{g}(\cdot) = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{b_T}K\left(\frac{\cdot - \widetilde{\varepsilon}_t}{b_T}\right).$$

Table 2. η_f for Student's t QMLEs (row) and innovation distributions (column)

	$t_{2.5}$	t_3	t_4	t_5	t_7	t_{11}	$gg_{0.5}$	gg_1	$gg_{1.5}$	gg_2
$t_{2.5}$	1.000	1.231	1.425	1.506	1.584	1.641	0.900	1.414	1.614	1.716
t_3	0.815	1.000	1.151	1.216	1.275	1.318	0.756	1.150	1.301	1.375
t_4	0.715	0.874	1.000	1.054	1.100	1.133	0.697	1.011	1.122	1.174
t_5	0.690	0.836	0.953	1.000	1.043	1.071	0.691	0.966	1.061	1.107
t_7	0.679	0.816	0.922	0.964	1.000	1.024	0.708	0.945	1.018	1.053
t_{11}	0.690	0.823	0.916	0.953	0.980	1.000	0.749	0.941	0.998	1.021
t_{20}	0.720	0.845	0.928	0.958	0.981	0.992	0.811	0.954	0.992	1.007
t_{30}	0.742	0.862	0.939	0.965	0.981	0.992	0.846	0.966	0.993	1.004

Table 3. Asymptotic relative efficiency of Student's 7 QMLEs against rank-based estimators

Student's t innovations	ARE	Generalized Gaussian innovations	ARE
t_{20}	1.006	gg_4	1.201
t_9	0.998	Gauss.	1.027
t_6	1.002	gg_1	1.010
t_4	1.025	$gg_{0.8}$	1.029
t_3	1.065	$gg_{0.4}$	1.140

The choice of the first step estimation requires a finite fourth moment for the innovation. In our comparison, we choose $K(\cdot)$ to be the Gaussian kernel, with a range of bandwidths equal to 0.1, 0.5, and 0.9, to demonstrate the effect of bandwidth selection in finite sample.

Our simulations are conducted using a GARCH(1, 1) model with true parameters $(\sigma, a_1, b_1) = (0.5, 0.6, 0.3)$. For innovation errors we use Student's t and generalized Gaussian distributions of various degrees of freedoms and shape parameters to generate data. For each type of innovation distribution, we run $N = 1000$ simulations each with sample size ranging among 250, 500, and 1000. Tables 4, 5, and 6 report the RMSE comparison of these three estimators.

In the upper panel of Tables 4, 5, and 6, the innovation distributions range from thin-tailed t_{20} (approximately Gaussian) to heavy-tailed t_3. For the first thin-tailed t_{20} case, GQMLE outperforms NGQMLE by a small margin. For all other cases, NGQMLE outperforms GQMLE. In the cases t_9 and t_6, NGQMLE performs nearly as well as MLE, and reduces standard deviations by roughly 20%–30% from GQMLE. In heavier tail cases (t_4 and t_3), since a fourth moment no longer exists, GQMLE is not $\sqrt{T}$-consistent, and its estimation precision quickly deteriorates, sometimes to an intolerable level. Hence, we omit reporting the estimates in the tables. In contrast, NGQMLE using t_4 likelihood does not require a finite fourth moment for $\sqrt{T}$-consistent a_1 and b_1, so standard deviations for a_1 and b_1 are still nearly equal to MLE. Standard deviations of σ are now larger than MLE, but still significantly smaller than GQMLE. The finite sample performance of the rank-based estimator is again indistinguishable compared to that of the NGQMLE, which agrees with the asymptotic relative efficiency in Table 3. The comparison with semiparametric estimator shows that the NGQMLE behaves better, especially in small sample, and the performance of the semiparametric estimator

Table 4. Comparison of RMSE of $\widehat{a}_1$ with Student's t and generalized Gaussian innovations

Innov.	T	GQMLE	NGQMLE	RANK	SEMI(0.1)	SEMI(0.5)	SEMI(0.9)	MLE
t_{20}	250	0.608	0.618	0.637	0.835	0.840	0.789	0.599
	500	0.407	0.391	0.393	0.483	0.445	0.446	0.393
	1000	0.263	0.261	0.262	0.312	0.295	0.291	0.258
t_9	250	0.548	0.496	0.501	0.683	0.860	0.804	0.493
	500	0.379	0.355	0.361	0.467	0.494	0.490	0.352
	1000	0.284	0.250	0.253	0.332	0.307	0.321	0.251
t_6	250	0.750	0.561	0.582	0.744	1.027	0.932	0.562
	500	0.469	0.387	0.397	0.488	0.457	0.448	0.388
	1000	0.325	0.262	0.265	0.326	0.283	0.286	0.263
t_4	250		0.524	0.530				0.520
	500		0.403	0.402				0.399
	1000		0.273	0.268				0.264
t_3	250		0.703	0.685				0.661
	500		0.438	0.423				0.418
	1000		0.297	0.289				0.287
gg_4	250	0.422	0.465	0.424	0.566	0.636	0.657	0.392
	500	0.291	0.329	0.300	0.346	0.319	0.340	0.268
	1000	0.214	0.239	0.217	0.260	0.229	0.234	0.200
Gauss.	250	0.488	0.533	0.529	0.685	0.781	0.767	0.488
	500	0.338	0.343	0.357	0.423	0.402	0.396	0.338
	1000	0.243	0.253	0.254	0.287	0.259	0.257	0.243
gg_1	250	0.661	0.601	0.630	0.682	0.788	0.788	0.597
	500	0.439	0.413	0.417	0.477	0.485	0.472	0.399
	1000	0.324	0.287	0.289	0.344	0.309	0.312	0.285
$gg_{0.8}$	250	0.892	0.720	0.735	0.987	0.992	0.994	0.723
	500	0.587	0.461	0.467	0.564	0.519	0.530	0.450
	1000	0.384	0.316	0.317	0.397	0.338	0.350	0.306
$gg_{0.4}$	250	4.267	1.465	1.365	1.715	1.428	1.509	1.406
	500	3.158	0.801	0.766	1.183	0.790	0.880	0.753
	1000	1.290	0.454	0.431	0.798	0.520	0.573	0.421

NOTE: We report the RMSE for a_1 in this table. The innovation errors follow t-distribution with degrees of freedom equal to 20, 9, 6, 4, and 3, and generalized Gaussian distribution with shape parameters equal to 4, 2, 1, 0.8, and 0.4, respectively. We follow Andrews (2012) to impose the constraints $a_1 > 0$ and $0 < b_1 < 1$ in optimization. For semiparametric estimation, we follow Drost and Klaassen (1997) to impose an additional constraint $a_1\sigma^2 + b_1 < 1$ in the first GQMLE step.

Table 5. Comparison of RMSE of $\widehat{b}_1$ with Student's t and generalized Gaussian innovations

Innov.	T	GQMLE	NGQMLE	RANK	SEMI(0.1)	SEMI(0.5)	SEMI(0.9)	MLE
t_{20}	250	0.285	0.284	0.297	0.698	1.611	1.674	0.281
	500	0.242	0.241	0.250	0.404	0.442	0.542	0.238
	1000	0.198	0.196	0.200	0.261	0.253	0.252	0.196
t_9	250	0.295	0.285	0.295	0.746	1.576	2.083	0.284
	500	0.249	0.243	0.251	0.453	0.499	0.494	0.244
	1000	0.208	0.201	0.204	0.270	0.255	0.248	0.201
t_6	250	0.300	0.284	0.292	0.769	1.208	1.221	0.283
	500	0.262	0.245	0.254	0.519	0.620	0.667	0.245
	1000	0.217	0.199	0.205	0.267	0.301	0.278	0.200
t_4	250		0.283	0.296				0.287
	500		0.233	0.242				0.233
	1000		0.197	0.196				0.194
t_3	250		0.290	0.299				0.286
	500		0.247	0.255				0.245
	1000		0.208	0.210				0.204
gg_4	250	0.253	0.260	0.260	0.662	0.875	1.218	0.252
	500	0.216	0.229	0.221	0.331	0.357	0.374	0.205
	1000	0.180	0.194	0.184	0.230	0.237	0.237	0.171
Gauss.	250	0.269	0.269	0.275	0.940	0.814	0.989	0.269
	500	0.235	0.240	0.251	0.495	0.579	0.425	0.235
	1000	0.185	0.189	0.193	0.248	0.232	0.230	0.185
gg_1	250	0.308	0.291	0.305	0.861	1.407	1.632	0.289
	500	0.255	0.249	0.262	0.438	0.553	0.558	0.249
	1000	0.223	0.213	0.218	0.277	0.276	0.267	0.213
$gg_{0.8}$	250	0.316	0.292	0.309	0.850	1.098	1.311	0.298
	500	0.274	0.251	0.271	0.493	0.638	0.527	0.253
	1000	0.232	0.210	0.218	0.318	0.286	0.282	0.204
$gg_{0.4}$	250	0.336	0.302	0.326	1.060	1.036	0.981	0.306
	500	0.322	0.276	0.301	0.747	0.694	0.640	0.284
	1000	0.295	0.235	0.247	0.416	0.390	0.392	0.237

varies with the choice of the bandwidth. The semiparametric estimator is clearly less attractive compared to the NGQMLE, since the NGQMLE is free of tuning parameters, and its gap in efficiency (compared to MLE) is not large at all.

In the lower panel, the innovations range from thin tailed gg_4 to heavy-tailed $gg_{0.4}$. For innovations with gg_1 and heavier, NGQMLE starts to outperform GQMLE, and in all such cases, the Student t_7 NGQMLE performs very close to MLE. In comparison, GQMLE's performance deteriorates as tails grow heavier, particularly in $gg_{0.8}$ and $gg_{0.4}$, although in these cases the fourth moments are finite. The comparison results with the rank-based estimator, the semiparametric estimator, and the MLE are similar to the cases with Student's t innovations.

8 EMPIRICAL WORK

To demonstrate the empirical relevance of our approach, we estimate values of η_f for S&P 500 components. To do so, we fit GARCH(1, 1) models to daily returns collected from January 2, 2004, to December 30, 2011, so the sample contains 2015 trading days in total. There are 458 stocks selected from the current S&P 500 stocks, as they have been traded since the beginning of the sample. We apply two common likelihood functions that are extensively used in the literature, including Student's t likelihood with degree of freedom 4, and generalized Gaussian likelihood with shape parameter equal to 1. Figure 3 provides an illustration of the empirical distributions of their respective η_f's. It is clear from the plot that most stocks have η_f's larger than 1, indicating the tail in the data is less heavier than the likelihoods we select. On the other hand, most η_f's deviate from 1 by a wide margin, showing that the bias of likelihood estimates without adjustments could be as large as 15%–20%. In summary, the message we want to deliver here is the extra step of η_f adjustment is necessary for non-Gaussian likelihood estimators.

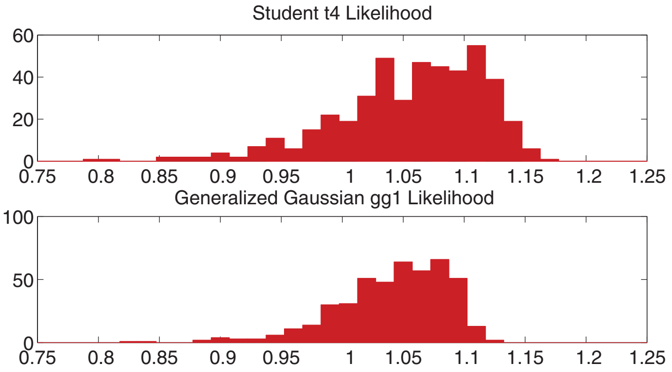

Figure 3. Empirical estimates of η_f across S&P 500 component stocks.

Table 6. Comparison of RMSE of $\widehat{\sigma}$ with Student's t and generalized Gaussian innovations

Innov.	T	GQMLE	NGQMLE	MLE
t_{20}	250	0.130	0.130	0.128
	500	0.111	0.110	0.109
	1000	0.088	0.087	0.087
t_9	250	0.135	0.128	0.128
	500	0.113	0.110	0.110
	1000	0.093	0.088	0.088
t_6	250	0.145	0.133	0.132
	500	0.122	0.112	0.112
	1000	0.095	0.086	0.087
t_4	250	0.150	0.134	0.131
	500	0.129	0.107	0.102
	1000	0.107	0.089	0.084
t_3	250	0.230	0.171	0.132
	500	0.170	0.148	0.109
	1000	0.138	0.103	0.088
gg_4	250	0.114	0.117	0.112
	500	0.094	0.101	0.088
	1000	0.078	0.085	0.073
Gauss.	250	0.122	0.122	0.122
	500	0.104	0.107	0.104
	1000	0.081	0.083	0.081
gg_1	250	0.147	0.135	0.134
	500	0.116	0.113	0.113
	1000	0.100	0.095	0.095
$gg_{0.8}$	250	0.153	0.139	0.140
	500	0.127	0.114	0.115
	1000	0.105	0.093	0.090
$gg_{0.4}$	250	0.189	0.163	0.149
	500	0.172	0.142	0.135
	1000	0.148	0.110	0.106

NOTE: We report the RMSE for σ in this table. The innovation errors follow t-distribution with degrees of freedom equal to 20, 9, 6, 4, and 3, and generalized Gaussian distribution with shape parameters equal to 4, 2, 1, 0.8, and 0.4, respectively.

9 CONCLUSION

This article focuses on GARCH model estimation when the innovation distribution is unknown, and it questions the efficiency of GQMLE and consistency of the popular NGQMLE. It proposes the NGQMLE to tackle both issues. The first step of the proposed procedure runs a GQMLE whose purpose is to identify the unknown scale parameter η_f in the second step. The final step performs an NGQMLE to estimate model parameters. The quasi-likelihood f can be a prespecified heavy-tailed likelihood, properly scaled by η_f or selected from a pool of candidate distributions.

The asymptotic theory of NGQMLE does not depend on any symmetric or unimodal assumptions of innovations. By adopting the different parameterization proposed by Newey and Steigerwald (1997), and incorporating η_f, NGQMLE improves the efficiency of GQMLE. This article shows that the asymptotic behavior of NGQMLE can be broken down to two parts. For the heteroscedastic parameter γ, NGQMLE is always $\sqrt{T}$-consistent and asymptotically normal, whereas $\sqrt{T}$-consistency of GQMLE relies on finite fourth moment assumption. When $E\varepsilon_t^4 < \infty$, NGQMLE outperforms GQMLE in terms of smaller asymptotic variance, provided that innovation distribution is reasonably heavy-tailed, which is common for financial data. For the scale part σ, NGQMLE is not always $\sqrt{T}$-consistent. The article also provides simulations to compare the performance of GQMLE, NGQMLE, semiparametric estimator, and MLE. In most cases, NGQMLE shows an advantage and is close to the MLE.

A. APPENDIX

A.1 Proof of Lemma 1

Proof. Note that

$$\begin{aligned} E_t(l(\bar{\boldsymbol{x}}_t, \boldsymbol{\theta})) &= Q\left(\frac{\eta_f \sigma v_t(\boldsymbol{\gamma})}{\sigma_0 v_t(\boldsymbol{\gamma}_0)}\right) - \log \sigma_0 v_t(\boldsymbol{\gamma}_0) + \log \eta_f \\ &\leq Q(\eta_f) - \log \sigma_0 v_t(\boldsymbol{\gamma}_0) + \log \eta_f \\ &= E_t(l(\bar{\boldsymbol{x}}_t, \boldsymbol{\theta}_0)). \end{aligned}$$

Also, Assumption 1 implies that if $v_t(\boldsymbol{\gamma}) = v_t(\boldsymbol{\gamma}_0)$ a.s., then $\boldsymbol{\gamma} = \boldsymbol{\gamma}_0$, see for example, (Francq and Zakoïan 2004, p. 615). By Assumption 2, the inequality holds with positive probability. Therefore, by iterated expectations, $\bar{L}(\boldsymbol{\theta}) < \bar{L}(\boldsymbol{\theta}_0)$. □

A.2 Proof of Theorem 1

Proof. It is straightforward to verify that $(\boldsymbol{\theta}_0, \eta_f, \boldsymbol{\theta}_0)$ satisfy the following equation:

$$E(\widetilde{\boldsymbol{s}}(\bar{\boldsymbol{x}}_t, \boldsymbol{\theta}, \eta, \boldsymbol{\phi})) = \boldsymbol{0},$$

Assumptions 1 and 2 guarantee the uniqueness, hence identification is established. By Theorem 2.6 in Newey and McFadden (1994) (a generalized version for stationary and ergodic x_t), along with Assumption 3 (ii), we have the desired consistency. □

A.3 Proof of Proposition 1

Proof. Define the likelihood ratio function $G(\eta) = E(\log(\frac{\frac{1}{\eta}f(\frac{\varepsilon}{\eta})}{f(\varepsilon)}))$. Suppose $G(\eta)$ has no local extremal values. And since $\log(x) \leq 2(\sqrt{x} - 1)$,

$$\begin{aligned} E\left(\log\left(\frac{\frac{1}{\eta}f(\frac{\varepsilon}{\eta})}{f(\varepsilon)}\right)\right) &\leq 2E\left(\sqrt{\frac{\frac{1}{\eta}f(\frac{\varepsilon}{\eta})}{f(\varepsilon)}} - 1\right) \\ &= 2\int_{-\infty}^{+\infty}\sqrt{\frac{1}{\eta}f\left(\frac{x}{\eta}\right)f(x)}dx - 2 \\ &\leq -\int_{-\infty}^{+\infty}\left(\sqrt{\frac{1}{\eta}f\left(\frac{x}{\eta}\right)} - \sqrt{f(x)}\right)^2 dx \\ &\leq 0. \end{aligned}$$

The equality holds if and only if $\eta = 1$. Therefore, $\eta = 1$ is the unique maximum of $Q(\eta)$. □

A.4 Proof of Theorem 2

To show the asymptotic normality, we first list some notations and derive a lemma. For convenience, we denote $\boldsymbol{y_0} = \frac{1}{v_t(\boldsymbol{\gamma_0})}\frac{\partial v_t(\boldsymbol{\gamma_0})}{\partial \boldsymbol{\gamma}}$ and $\bar{\boldsymbol{y}}_{\boldsymbol{0}} = E(\boldsymbol{y_0})$, so $\boldsymbol{k_0} = (\frac{1}{\sigma_0}, \boldsymbol{y_0}')'$ and $\bar{\boldsymbol{k}}_{\boldsymbol{0}} = E\boldsymbol{k_0} = (\frac{1}{\sigma_0}, \bar{\boldsymbol{y}}_{\boldsymbol{0}}')'$. Also, let $\boldsymbol{M} = E(\boldsymbol{k_0}\boldsymbol{k_0}')$, $\boldsymbol{N} = \bar{\boldsymbol{k}}_{\boldsymbol{0}}\bar{\boldsymbol{k}}_{\boldsymbol{0}}'$ and $\boldsymbol{V} = \mathrm{var}(\boldsymbol{y_0})^{-1}$. All the expectations above are taken under the true density g. Note that from Section 3.2, we have

$$\boldsymbol{s_1}(\bar{\boldsymbol{x}}_{\boldsymbol{t}}, \boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} l_1(\bar{\boldsymbol{x}}_{\boldsymbol{t}}, \boldsymbol{\theta}) = \left(-1 + \frac{x_t^2}{\sigma^2 v_t^2(\boldsymbol{\theta})}\right)\boldsymbol{k}(\boldsymbol{\theta}), \qquad \text{(A.4)}$$

$$s_2(\bar{\boldsymbol{x}}_{\boldsymbol{t}}, \boldsymbol{\theta}, \eta) = \frac{\partial}{\partial \eta} l_2(\bar{\boldsymbol{x}}_{\boldsymbol{t}}, \boldsymbol{\theta}, \eta) = h_1\left(\frac{x_t}{\sigma v_t(\boldsymbol{\theta})}, \eta\right), \qquad \text{(A.5)}$$

$$\boldsymbol{s_3}(\bar{\boldsymbol{x}}_{\boldsymbol{t}}, \eta, \boldsymbol{\phi}) = \frac{\partial}{\partial \boldsymbol{\phi}} l_3(\bar{\boldsymbol{x}}_{\boldsymbol{t}}, \eta, \boldsymbol{\phi}) = \eta h_1\left(\frac{x_t}{\sigma v_t(\boldsymbol{\phi})}, \eta\right)\boldsymbol{k}(\boldsymbol{\phi}). \qquad \text{(A.6)}$$

First, we need a lemma.

Lemma 2. The following claims hold:

1. The inverse of $\boldsymbol{M}$ in block expression is

$$\boldsymbol{M}^{-1} = \begin{pmatrix} \sigma_0^2(1 + \bar{\boldsymbol{y}}_{\boldsymbol{0}}'\boldsymbol{V}\bar{\boldsymbol{y}}_{\boldsymbol{0}}) & -\sigma_0\bar{\boldsymbol{y}}_{\boldsymbol{0}}'\boldsymbol{V} \\ -\sigma_0\boldsymbol{V}\bar{\boldsymbol{y}}_{\boldsymbol{0}} & \boldsymbol{V} \end{pmatrix}; \qquad \text{(A.7)}$$

2. $\bar{\boldsymbol{k}}_{\boldsymbol{0}}'\boldsymbol{M}^{-1} = \sigma_0\boldsymbol{e_1}'$, $\bar{\boldsymbol{k}}_{\boldsymbol{0}}'\boldsymbol{M}^{-1}\boldsymbol{k_0} = \bar{\boldsymbol{k}}_{\boldsymbol{0}}'\boldsymbol{M}^{-1}\bar{\boldsymbol{k}}_{\boldsymbol{0}} = 1$;
3. $\boldsymbol{M}^{-1}\boldsymbol{N}\boldsymbol{M}^{-1} = \boldsymbol{M}^{-1}\boldsymbol{N}\boldsymbol{M}^{-1}\boldsymbol{N}\boldsymbol{M}^{-1} = \sigma_0^2\boldsymbol{e_1}\boldsymbol{e_1}'$, where $\boldsymbol{e_1}$ is a unit column vector that has the same length as $\boldsymbol{\theta}$, with the first entry one and all the rest zeros.

Proof. The proof uses Moore–Penrose pseudo inverse described in Ben-Israel and Greville (2003). Observe that

$$\boldsymbol{M} = \begin{pmatrix} 0 & \boldsymbol{0} \\ \boldsymbol{0} & \mathrm{var}(\boldsymbol{y_0}) \end{pmatrix} + \bar{\boldsymbol{k}}_{\boldsymbol{0}}\bar{\boldsymbol{k}}_{\boldsymbol{0}}'. \qquad \text{(A.8)}$$

Using the technique of Moore–Penrose pseudo inverse, we have

$$\boldsymbol{M}^{-1} = \begin{pmatrix} 0 & \boldsymbol{0} \\ \boldsymbol{0} & \mathrm{var}(\boldsymbol{y_0}) \end{pmatrix}^{+} + \boldsymbol{H} = \begin{pmatrix} 0 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{V} \end{pmatrix} + \boldsymbol{H}, \qquad \text{(A.9)}$$

where H is formed by the elements below:

$$\beta = 1 + \bar{\boldsymbol{y}}_{\boldsymbol{0}}'\boldsymbol{V}\bar{\boldsymbol{y}}_{\boldsymbol{0}}, \quad \boldsymbol{w} = \left(\sigma_0^{-1}, \boldsymbol{0}\right)', \quad \boldsymbol{m} = \boldsymbol{w},$$
$$\boldsymbol{v} = (0, \bar{\boldsymbol{y}}_{\boldsymbol{0}}'\boldsymbol{V})', \quad \boldsymbol{n} = \boldsymbol{v}.$$

$$\begin{aligned}\boldsymbol{H} &= -\frac{1}{\|\boldsymbol{w}\|^2}\boldsymbol{v}\boldsymbol{w}' - \frac{1}{\|\boldsymbol{m}\|^2}\boldsymbol{m}\boldsymbol{n}' + \frac{\beta}{\|\boldsymbol{w}\|^2\|\boldsymbol{m}\|^2}\boldsymbol{m}\boldsymbol{w}' \\ &= \sigma_0^2\begin{pmatrix} 1 + \bar{\boldsymbol{y}}_{\boldsymbol{0}}'\boldsymbol{V}\bar{\boldsymbol{y}}_{\boldsymbol{0}} & -\sigma_0^{-1}\bar{\boldsymbol{y}}_{\boldsymbol{0}}'\boldsymbol{V} \\ -\sigma_0^{-1}\boldsymbol{V}\bar{\boldsymbol{y}}_{\boldsymbol{0}} & \boldsymbol{0} \end{pmatrix}.\end{aligned}$$

So Equation (A.7) is obtained by plugging $\boldsymbol{H}$ into Equation (A.9). The rest two points of the lemma can be obtained by simple matrix manipulation. □

Next we return to the proof of Theorem 2.

Proof. According to Theorem 3.4 in Newey and McFadden (1994) (a generalized version for stationary and ergodic x_t), $(\widetilde{\boldsymbol{\theta}}_{\boldsymbol{T}}, \widehat{\eta}, \widehat{\boldsymbol{\theta}}_{\boldsymbol{T}})$ are jointly $T^{\frac{1}{2}}$-consistent and asymptotic normal. The asymptotic variance matrix is

$$\boldsymbol{G}^{-1}E\left(\widetilde{\boldsymbol{s}}(\bar{\boldsymbol{x}}_{\boldsymbol{t}}, \boldsymbol{\theta_0}, \eta_f, \boldsymbol{\theta_0})\widetilde{\boldsymbol{s}}(\bar{\boldsymbol{x}}_{\boldsymbol{t}}, \boldsymbol{\theta_0}, \eta_f, \boldsymbol{\theta_0})'\right)\boldsymbol{G}'^{-1}, \qquad \text{(A.10)}$$

where $\boldsymbol{G} = E(\nabla\widetilde{\boldsymbol{s}}(\bar{\boldsymbol{x}}_{\boldsymbol{t}}, \boldsymbol{\theta_0}, \eta_f, \boldsymbol{\theta_0}))$. View this matrix as 3×3 blocks, with asymptotic variances of $(\widetilde{\boldsymbol{\theta}}_{\boldsymbol{T}}, \widehat{\eta}, \widehat{\boldsymbol{\theta}}_{\boldsymbol{T}})$ on the first, second, and third diagonal blocks. We now calculate the second and third diagonal blocks. The expected Jacobian matrix G can be decomposed into

$$\boldsymbol{G} = E\begin{pmatrix} \nabla_{\boldsymbol{\theta}}\boldsymbol{s_1}(\bar{\boldsymbol{x}}_{\boldsymbol{t}}, \boldsymbol{\theta_0}) & \boldsymbol{0} & \boldsymbol{0} \\ \nabla_{\boldsymbol{\theta}}s_2(\bar{\boldsymbol{x}}_{\boldsymbol{t}}, \boldsymbol{\theta_0}, \eta_f) & \nabla_{\eta}s_2(\bar{\boldsymbol{x}}_{\boldsymbol{t}}, \boldsymbol{\theta_0}, \eta_f) & \boldsymbol{0} \\ \boldsymbol{0} & \nabla_{\eta}\boldsymbol{s_3}(\bar{\boldsymbol{x}}_{\boldsymbol{t}}, \eta_f, \boldsymbol{\theta_0}) & \nabla_{\boldsymbol{\phi}}\boldsymbol{s_3}(\bar{\boldsymbol{x}}_{\boldsymbol{t}}, \eta_f, \boldsymbol{\theta_0}) \end{pmatrix}.$$

Denote the corresponding blocks as G_{ij}, $i, j = 1, 2, 3$. Direct calculation yields

$$\begin{aligned}\boldsymbol{G_{11}} &= -2\boldsymbol{M}, \\ \boldsymbol{G_{21}} &= \eta_f Eh_2(\varepsilon, \eta_f)\bar{\boldsymbol{k}}_{\boldsymbol{0}}', \\ G_{22} &= Eh_2(\varepsilon, \eta_f), \\ \boldsymbol{G_{32}} &= \boldsymbol{G_{21}}', \\ \boldsymbol{G_{33}} &= \eta_f^2 Eh_2(\varepsilon, \eta_f)\boldsymbol{M}.\end{aligned}$$

The second diagonal block depends on the second row of $\boldsymbol{G^{-1}}$ and $\widetilde{\boldsymbol{s}}(\bar{\boldsymbol{x}}_{\boldsymbol{t}}, \boldsymbol{\theta_0}, \eta_f, \boldsymbol{\theta_0})$. The second row of $\boldsymbol{G}^{-1}$ is

$$\begin{pmatrix} -G_{22}^{-1}\boldsymbol{G_{21}}\boldsymbol{G_{11}}^{-1} & G_{22}^{-1} & \boldsymbol{0} \end{pmatrix}.$$

So the asymptotic variance of $\widehat{\eta}$ is $G_{22}^{-1}E(\boldsymbol{q_2}\boldsymbol{q_2}')G_{22}'^{-1}$, where

$$\begin{aligned}q_2 &= -\boldsymbol{G_{21}}\boldsymbol{G_{11}}^{-1}\boldsymbol{s_1}(\bar{\boldsymbol{x}}_{\boldsymbol{t}}, \boldsymbol{\theta_0}) + s_2(\bar{\boldsymbol{x}}_{\boldsymbol{t}}, \boldsymbol{\theta_0}, \eta_f) \\ &= \frac{\eta_f}{2}E(h_2(\varepsilon, \eta_f))\bar{\boldsymbol{k}}_{\boldsymbol{0}}'\overline{\boldsymbol{k_0}\boldsymbol{k_0}'}^{-1}(\varepsilon^2 - 1)\boldsymbol{k_0} + h_1(\varepsilon, \eta_f) \\ &= \frac{\eta_f}{2}Eh_2(\varepsilon^2 - 1) + h_1(\varepsilon, \eta_f).\end{aligned}$$

The last step uses the second point of Lemma 2. So Equation (15) is obtained by plugging in the expressions for G_{22} and q_2. Similarly, the third row of $\boldsymbol{G^{-1}}$ is

$$\boldsymbol{G_{33}}^{-1}\begin{pmatrix} \boldsymbol{G_{32}}G_{22}^{-1}\boldsymbol{G_{21}}\boldsymbol{G_{11}}^{-1} & -\boldsymbol{G_{32}}G_{22}^{-1} & \boldsymbol{I} \end{pmatrix}.$$

The asymptotic variance for $\widehat{\boldsymbol{\theta}}$ is $\boldsymbol{G_{33}}^{-1}E(\boldsymbol{q_3}\boldsymbol{q_3}')\boldsymbol{G_{33}}'^{-1}$, where

$$\begin{aligned}\boldsymbol{q_3} &= \boldsymbol{G_{32}}G_{22}^{-1}\left(\boldsymbol{G_{21}}\boldsymbol{G_{11}}^{-1}\boldsymbol{s_1}(\bar{\boldsymbol{x}}_{\boldsymbol{t}}, \boldsymbol{\theta_0}) - \boldsymbol{s_2}(\bar{\boldsymbol{x}}_{\boldsymbol{t}}, \boldsymbol{\theta_0}, \eta_f)\right) \\ &\quad + \boldsymbol{s_3}(\bar{\boldsymbol{x}}_{\boldsymbol{t}}, \eta_f, \boldsymbol{\theta_0}) \\ &= \eta_f h_1(\varepsilon, \eta_f)(\boldsymbol{k_0} \\ &\quad - \bar{\boldsymbol{k}}_{\boldsymbol{0}}) - \frac{\eta_f^2}{2}E(h_2(\varepsilon, \eta_f))\bar{\boldsymbol{k}}_{\boldsymbol{0}}\bar{\boldsymbol{k}}_{\boldsymbol{0}}'(\overline{\boldsymbol{k_0}\boldsymbol{k_0}'})^{-1}\boldsymbol{k_0}(\varepsilon^2 - 1) \\ &= \eta_f h_1(\varepsilon, \eta_f)(\boldsymbol{k_0} - \bar{\boldsymbol{k}}_{\boldsymbol{0}}) - \frac{\eta_f^2}{2}(Eh_2(\varepsilon, \eta_f))(\varepsilon^2 - 1)\bar{\boldsymbol{k}}_{\boldsymbol{0}}.\end{aligned}$$

The last step uses the second point of Lemma 2. Then

$$\begin{aligned}&E\boldsymbol{q_3}\boldsymbol{q_3}' \\ &= \eta_f^2 Eh_1(\varepsilon, \eta_f)^2(\boldsymbol{M} - \boldsymbol{N}) + \frac{\eta_f^4}{4}(Eh_2(\varepsilon, \eta_f))^2E(\varepsilon^2 - 1)\boldsymbol{N} \\ &= \eta_f^2 Eh_1(\varepsilon, \eta_f)^2\boldsymbol{M} + \left(\frac{1}{4}E(\varepsilon^2 - 1)^2 - \eta_f^2 Eh_1(\varepsilon, \eta_f)^2\right)\boldsymbol{N}.\end{aligned}$$

Therefore, Equation (14) is obtained by plugging in the expressions for $\boldsymbol{G_{33}}$, $\boldsymbol{Eq_3q_3}'$, and apply the third point of Lemma 2.

Note that in the case where η_f is known, the asymptotic variance of $\widehat{\boldsymbol{\theta}}$ is

$$\boldsymbol{G_{33}}^{-1}E(\boldsymbol{s_3}(\bar{\boldsymbol{x}}_{\boldsymbol{t}},\eta_f,\boldsymbol{\theta_0})\boldsymbol{s_3}'(\bar{\boldsymbol{x}}_{\boldsymbol{t}},\eta_f,\boldsymbol{\theta_0}))\boldsymbol{G_{33}}'^{-1}$$
$$=\frac{E(h_1(\varepsilon_t,\eta_f))^2}{\eta_f^2(Eh_2(\varepsilon_t,\eta_f))^2}\boldsymbol{M}^{-1}.$$

The asymptotic covariance between $\widehat{\boldsymbol{\theta}}$ and $\widehat{\eta}_f$ is $\boldsymbol{G_{33}}^{-1}E(\boldsymbol{q_3}q_2)G_{22}'^{-1}$. A direct calculation using the second point of Lemma 2 yields

$$\boldsymbol{\Pi}=\frac{\eta_f\sigma_0}{2}E\left((1-\varepsilon^2)\left(\frac{h_1(\varepsilon,\eta_f)}{\eta_f Eh_2(\varepsilon,\eta_f)}+\frac{\varepsilon^2-1}{2}\right)\right)\boldsymbol{e_1'}.$$

The same formula recurs in the asymptotic covariance between $\widetilde{\boldsymbol{\theta}}$ and $\widehat{\eta}_f$, which is $\boldsymbol{G_{11}}^{-1}E(\boldsymbol{q_1}q_2)G_{22}'^{-1}$.

Finally, the asymptotic covariance between $\widetilde{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\theta}}$ is $\boldsymbol{G_{11}}^{-1}E(\boldsymbol{q_1q_3'})\boldsymbol{G_{33}}'^{-1}$, denoted as Ξ. It implies from the third point of Lemma 2 that

$$\boldsymbol{\Xi}=\frac{E(h_1(\varepsilon,\eta_f)(1-\varepsilon^2))}{2\eta_f E(h_2(\varepsilon,\eta_f))}\boldsymbol{M^{-1}}$$
$$-\frac{\sigma_0^2}{2}E\left((1-\varepsilon^2)\left(\frac{h_1(\varepsilon,\eta_f)}{\eta_f Eh_2(\varepsilon,\eta_f)}+\frac{\varepsilon^2-1}{2}\right)\right)\boldsymbol{e_1e_1'},$$

which concludes the proof. □

A.5 Proof of Theorem 3

Proof. Following the similar idea to GMM, we may prove:

$$\begin{pmatrix}\boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0}\\ \lambda_T T^{-\frac{1}{2}}\boldsymbol{G_{21}} & G_{22} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{G_{32}} & \boldsymbol{G_{33}}\end{pmatrix}\begin{pmatrix}T\lambda_T^{-1}(\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta_0})\\ T^{\frac{1}{2}}(\widehat{\eta}_f-\eta_f)\\ T^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta_0})\end{pmatrix}$$
$$=\begin{pmatrix}\frac{1}{\lambda_T}\sum_{t=1}^{T}\boldsymbol{\Psi_t}(\varepsilon_t)+\boldsymbol{o_P(1)}\\ \frac{1}{\sqrt{T}}\sum_{t=1}^{T}h_1(\varepsilon,\eta_f)+o_P(1)\\ -\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\eta_f h(\varepsilon_t,\eta_f)\boldsymbol{k_0}+\boldsymbol{o_P(1)}\end{pmatrix}.$$

Clearly, the corresponding weighting vector for $\sqrt{T}(\widehat{\boldsymbol{\theta}}_{\boldsymbol{T}}-\boldsymbol{\theta_0})$ is

$$\left(\boldsymbol{G_{33}}^{-1}\boldsymbol{G_{32}}G_{22}^{-1}\lambda_T T^{-\frac{1}{2}}\boldsymbol{G_{21}}\quad -\boldsymbol{G_{33}}^{-1}\boldsymbol{G_{32}}G_{22}^{-1}\quad \boldsymbol{G_{33}}^{-1}\right).$$

Note that

$$\boldsymbol{G_{33}}^{-1}\boldsymbol{G_{32}}G_{22}^{-1}\lambda_T T^{-\frac{1}{2}}\boldsymbol{G_{21}}=\lambda_T T^{-\frac{1}{2}}\boldsymbol{M}^{-1}\bar{\boldsymbol{k}}_{\boldsymbol{0}}\bar{\boldsymbol{k}}_{\boldsymbol{0}}'=\lambda_T T^{-\frac{1}{2}}\sigma_0\boldsymbol{e_1}\bar{\boldsymbol{k}}_{\boldsymbol{0}}',$$

and

$$-\boldsymbol{G_{33}}^{-1}\boldsymbol{G_{32}}G_{22}^{-1}=-\sigma_0(\eta_f Eh_2(\varepsilon,\eta_f))^{-1}\boldsymbol{e_1}.$$

Thus, the submatrices corresponding to $\boldsymbol{\gamma}$ parameter are $\boldsymbol{0}$'s. Therefore, the first step has no effect on the central limit theorem of $\widehat{\boldsymbol{\gamma}}_{\boldsymbol{T}}$. The result follows from Lemma 2. In terms of $\widehat{\sigma}_T$, its convergence rate becomes $T\lambda_T^{-1}$. □

A.6 Proof of Proposition 2

Proof. Denote by random variables $\kappa_G=(1-\varepsilon^2)/2$, and $\kappa_2=-h_1(\varepsilon,\eta_f)/E(\eta_f h_2(\varepsilon,\eta_f))$. We show the optimal weights for σ and $\boldsymbol{\gamma}$ are the same. From Lemma 2, Theorem 2, and Equation (27), for σ, the numerator in w_1^* is

$$(\boldsymbol{\Sigma_G})_{1,1}-\boldsymbol{\Xi}_{1,1}$$
$$=\sigma_0^2(1+\bar{\boldsymbol{y}}_0'\boldsymbol{V}\bar{\boldsymbol{y}}_0)E\kappa_G^2-\sigma_0^2E\kappa_G^2+\sigma_0^2\bar{\boldsymbol{y}}_0'\boldsymbol{V}\bar{\boldsymbol{y}}_0E(\kappa_G\kappa_2)$$
$$=\sigma_0^2\bar{\boldsymbol{y}}_0'\boldsymbol{V}\bar{\boldsymbol{y}}_0E(\kappa_G(\kappa_G+\kappa_2)).$$

The denominator in w_1^* is

$$(\boldsymbol{\Sigma_G})_{1,1}+(\boldsymbol{\Sigma_2})_{1,1}-2\boldsymbol{\Xi}_{1,1}$$
$$=\sigma_0^2(1+\bar{\boldsymbol{y}}_0'\boldsymbol{V}\bar{\boldsymbol{y}}_0)\left(E\kappa_G^2+E\kappa_2^2\right)+\sigma_0^2\left(E\kappa_G^2-E\kappa_2^2\right)$$
$$-2\sigma_0^2E\kappa_G^2+2\sigma_0^2\bar{\boldsymbol{y}}_0'\boldsymbol{V}\bar{\boldsymbol{y}}_0E(\kappa_G\kappa_2)$$
$$=\sigma_0^2\bar{\boldsymbol{y}}_0'\boldsymbol{V}\bar{\boldsymbol{y}}_0E\left(\kappa_G^2+\kappa_2^2+2\kappa_G\kappa_2\right).$$

Therefore, we obtain $w_1^*=E(\kappa_G(\kappa_G+\kappa_2))/E(\kappa_G+\kappa_2)^2$. Now we compute the weights corresponding to $\boldsymbol{\gamma}$. For $i=2,\dots,1+p+q$, let $j=i-1$, also from Equation (27),

$$w_i^*=\frac{\boldsymbol{V}_{j,j}E\kappa_G^2+\boldsymbol{V}_{j,j}E(\kappa_G\kappa_2)}{\boldsymbol{V}_{j,j}E\kappa_G^2+\boldsymbol{V}_{j,j}E\kappa_2^2 2\boldsymbol{V}_{j,j}E(\kappa_G\kappa_2)}=\frac{E(\kappa_G(\kappa_G+\kappa_2))}{E(\kappa_G+\kappa_2)^2}.$$

Therefore, all the optimal aggregation weights are the same. The optimal variance follows from direct calculations. Note that the estimation of ω^* does not lead to a larger asymptotic variance because $\widehat{\boldsymbol{\theta}}_{\boldsymbol{T}}^*-(\boldsymbol{W}^*\widehat{\boldsymbol{\theta}}+(\boldsymbol{I}-\boldsymbol{W}^*)\widetilde{\boldsymbol{\theta}})=(\widehat{\boldsymbol{W}}^*-\boldsymbol{W}^*)(\widehat{\boldsymbol{\theta}}-\widetilde{\boldsymbol{\theta}})=o_p(T^{-\frac{1}{2}})$.

□

ACKNOWLEDGMENTS

We are grateful for helpful suggestions from the editor, the associated editor and the referees, extensive discussions with Chris Sims and Enrique Sentana, and thoughtful comments from seminar and conference participants at Princeton University, the 4th CIREQ Time Series Conference, and the 10th World Congress of the Econometric Society. Financial support from NSF Grant DMS-0704337 and DMS-1206464, and the University of Chicago Booth School of Business is gratefully acknowledged.

[Received November 2012. Revised March 2013.]

REFERENCES

Andrews, B. (2012), "Rank-Based Estimation for GARCH Processes," *Econometric Theory*, 28, 1037–1064. [179,185]

Ben-Israel, A., and Greville, T. (2003), *Generalized Inverses Theory and Applications*, New York: Springer. [189]

Berkes, I., and Horváth, L. (2004), "The Efficiency of the Estimators of the Parameters in Garch Processes," *The Annals of Statistics*, 32, 633–655. [179]

Berkes, I., Horváth, L., and Kokoszka, P. (2003), "Garch Processes: Structure and Estimation," *Bernoulli*, 9, 201–227. [178,180]

Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307–327. [178]

——— (1987), "A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return," *The Review of Economics and Statistics*, 69, 542–547. [178]

Bollerslev, T., and Wooldbridge, J. M. (1992), "Quasi-maximum Likelihood Estimation and Inference in Dynamic Models With Time-varying Covariances," *Econometric Reviews*, 11, 143–172. [178,179]

Bougerol, P., and Picard, N. (1992), "Stationarity of Garch Processes and of Some Nonnegative Time Series," *Journal of Econometrics*, 52, 115–127. [179]

Diebold, F. (1988), *Empirical Modeling of Exchange Rate Dynamics*, New York: Springer. [178]

Drost, F. C., and Klaassen, C. A. J. (1997), "Efficient Estimation in Semiparametric Garch Models," *Journal of Econometrics*, 81, 193–221. [178,179,183,185]

Elie, L., and Jeantheau, T. (1995), "Consistency in Heteroskedastic Models," *Comptes Rendus de l 'Académie des Sciences*, 320, 1255–1258. [178]

Engle, R. F. (1982), "Autoregressive Conditional Heteroscedasticity With Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987–1007. [178]

Engle, R. F., and Bollerslev, T. (1986), "Modelling the Persistence of Conditional Variances," *Econometric Reviews*, 5, 1–50. [178]

Engle, R. F., and Gonzalez-Rivera, G. (1991), "Semiparametric Arch Models," *Journal of Business and Economic Statistics*, 9, 345–359. [178]

Fiorentini, G., and Sentana, E. (2010), "On the Efficiency and Consistency of Likelihood Estimation in Multivariate Conditionally Heteroskedastic Dynamic Regression Models," unpublished manuscript, CEMFI. [181]

Francq, C., Lepage, G., and Zakoïan, J.-M. (2011), "Two-stage Non Gaussian QML Estimation of GARCH Models and Testing the Efficiency of the Gaussian QMLE," *Journal of Econometrics*, 165, 246–257. [179,181,182]

Francq, C., and Zakoïan, J.-M. (2004), "Maximum Likelihood Estimation of Pure GARCH and ARMA-GARCH Processes," *Bernoulli*, 10, 605–637. [188]

González-Rivera, G., and Drost, F. C. (1999), "Efficiency Comparisons of Maximum-likelihood-based Estimators in Garch Models," *Journal of Econometrics*, 93, 93–111. [178,183]

Hall, P., and Yao, Q. (2003), "Inference in Arch and Garch Models With Heavy-tailed Errors," *Econometrica*, 71, 285–317. [178,180,183]

Hsieh, D. A. (1989), "Modeling Heteroscedasticity in Daily Foreign-Exchange Rates," *Journal of Business and Economic Statistics*, 7, 307–317. [178]

Huang, D., Wang, H., and Yao, Q. (2008), "Estimating GARCH Models: When to Use What?," *Econometrics Journal*, 11, 27–38. [179]

Lee, S.-W., and Hansen, B. E. (1994), "Asymptotic Theory for the Garch (1, 1) Quasi-maximum Likelihood Estimator," *Econometric Theory*, 10, 29–52. [178]

Lee, T., and Lee, S. (2009), "Normal Mixture Quasi-maximum Likelihood Estimator for GARCH Models," *Scandinavian Journal of Statistics*, 36, 157–170. [179,181]

Linton, O. (1993), "Adaptive Estimation in Arch Models," *Econometric Theory*, 9, 539–569. [178]

Lumsdaine, R. L. (1996), "Consistency and Asymptotic Normality of the Quasi-maximum Likelihood Estimator in Igarch(1, 1) and Covariance Stationary Garch(1, 1) Models," *Econometrica*, 64, 575–596. [178]

Nelson, D. B. (1991), "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, 59, 347–370. [178]

Newey, W. K., and McFadden, D. (1994), "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics* Vol. 4 , chap. 36, eds. R. F. Engle and D. McFadden, North Holland: Elsevier, pp. 2111–2245. [188,189]

Newey, W. K., and Steigerwald, D. G. (1997), "Asymptotic Bias for Quasi-maximum-likelihood Estimators in Conditional Heteroskedasticity Models," *Econometrica*, 65, 587–599. [179,180,181,188]

Peng, L., and Yao, Q. (2003), "Least Absolute Deviations Estimation for ARCH and GARCH Models," *Biometrika*, 90, 967–975. [179]

Sun, Y., and Stengos, T. (2006), "Semiparametric Efficient Adaptive Estimation of Asymmetric GARCH Models," *Journal of Econometrics*, 133, 373–386. [178]

Weiss, A. A. (1986), "Asymptotic Theory for Arch Models: Estimation and Testing," *Econometric Theory*, 2, 107–131. [178]

White, H. (1982), "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1–26. [180]

Comment

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

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

$$x_t = \sigma v_t \varepsilon_t,$$

$$v_t^2 = 1 + \sum_{i=1}^{p} a_i x_{t-i}^2 + \sum_{j=1}^{q} b_j v_{t-j}^2,$$

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 Fan, Qi, and Xiu mention, a rank-based technique for estimating $\boldsymbol{\theta}$ was 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 comment, 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 an 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 Fan, Qi, and Xiu suggest 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).

© 2014 American Statistical Association
Journal of Business & Economic Statistics
April 2014, Vol. 32, No. 2
DOI: 10.1080/07350015.2013.875921