When Simplicity Offers a Benefit, Not a Cost: Closed-Form Estimation of the GARCH(1,1) Model that Enhances the Efficiency of Quasi-Maximum Likelihood

Supplemental Appendix

Todd Prono

A.1.1. Preliminaries

Contained in this Supplemental Appendix are both the statements and proofs of all Lemmas that support the paper's main theorems. Concerning notation, C denotes a generic constant that can assume different values in different places. For matrices $\mathbf{A}$ and $\mathbf{B}$, $\mathbf{A} \geq \mathbf{B}$ means that every element in $\mathbf{A} \geq$ every corresponding element in $\mathbf{B}$. For a vector $\mathbf{y}$, $\delta_{\mathbf{y}}$ denotes the Dirac measure at $\mathbf{y}$. For a random variable $X > 0$ with CDF $F_X(x)$, where $\overline{F}_X(x) = 1 - F_X(x)$, if

$$\lim_{x\to\infty} \frac{\overline{F}_X(tx)}{\overline{F}_X(x)} = t^{-\kappa_0}, \quad \forall\, t > 0, \quad \kappa_0 \geq 0, \tag{1}$$

then X is regularly varying with tail index κ_0. Finally, RV(κ_0) is shorthand for regularly varying with tail index κ_0.

Proposition 1 *For a random variable $X > 0$, assume (1) holds. Then for a $p > 0$, X^p is regularly varying with tail index κ_0/p.*

Proof. Let $Y = X^p$. Then

$$\begin{aligned}
\frac{\overline{F}_Y(ty)}{\overline{F}_Y(y)} &= \frac{P(Y > ty)}{P(Y > y)} \\
&= \frac{P\left(Y^{1/p} > t^{1/p}y^{1/p}\right)}{P\left(Y^{1/p} > y^{1/p}\right)} \\
&= \frac{P\left(X > t^{1/p}x\right)}{P(X > x)} \\
&= \frac{\overline{F}_X\left(t^{1/p}x\right)}{\overline{F}_X(x)} \\
&= \frac{\overline{F}_X(Bx)}{\overline{F}_X(x)},
\end{aligned}$$

in which case, by (1),

$$\lim_{y\to\infty} \frac{\overline{F}_Y(ty)}{\overline{F}_Y(y)} = \lim_{x\to\infty} \frac{\overline{F}_X(Bx)}{\overline{F}_X(x)} = B^{-\kappa_0} = t^{-\kappa_0/p}.$$

■

A.1.2. Regular Variation

Consider the model

$$Y_t = \sigma_t \epsilon_t, \qquad \epsilon_t \sim i.i.d.\ D\left(0,\ 1\right) \tag{2}$$

$$\sigma_t^2 = \omega_0 + \alpha_0 Y_{t-1}^2 + \beta_0 \sigma_{t-1}^2, \tag{3}$$

where D is some unknown distribution. This model is the linear GARCH$(1,1)$ model of Bollerslev (1986). Also note that from (3),

$$\begin{aligned} \sigma_t^2 &= \omega_0 + \sigma_{t-1}^2 \times \left(\alpha_{0,t-1}\epsilon_{t-1}^2 + \beta_0\right) \\ &= \omega_0 + \sigma_{t-1}^2 A_t. \end{aligned} \tag{4}$$

Lemma 2 *For the model in (2) and (3), let Assumptions A1–A2 and A4 hold. Then*

$$E\left(A^{\kappa/2}\right) = 1 \tag{5}$$

has a unique and positive solution κ_0.

Proof. For a $\kappa > 0$, $E\left(A^{\kappa}\right)$ is a continuous and convex (upwards) function of κ. Since Assumption A4 is sufficient for $E\left(A\right) < 1$,

CONDITION C1: $E\left(A^{\kappa}\right) < 1$ for values of κ in some neighborhood of one.

Also, since $P\left(A > 1\right) > 0$, and since there exists a value $\overline{\kappa}$ of κ such that $E\left(A^{\overline{\kappa}/2}\right) = \infty$,

CONDITION C2: $E\left(A^{\kappa}\right) > 1$ for sufficiently large κ.

Conditions C1 and C2 together complete the proof. ■

Lemma 3 *For the model in (2) and (3), under the same Assumptions as Lemma 2,*

$$P\left(\sigma > x\right) \sim Cx^{-\kappa_0}, \tag{6}$$

and

$$P\left(|Y| > x\right) = E\left(|\epsilon|^{\kappa_0}\right) \times P\left(\sigma > x\right), \quad x \to \infty. \tag{7}$$

Proof. Assumption A4 is sufficient to establish the sequence $\left\{\sigma_t^2\right\}$ as strictly stationary (see; e.g., Mikosch, 1999, Corollary 1.4.39). Owing to the method used to establish (5) as having a unique and positive solution, there exists a small $\eta > 0$ such that

$$E\left(A^{\kappa_0/2+\eta}\right) < \infty.$$

Consider then for (4)

CONDITION C3: $\{A_t\}$ is an $i.i.d.$ sequence.

CONDITION C4: σ_{t-1}^2 is independent of A_t for every t.

C3 follows because A_t is only a function of ϵ_{t-1}. The validity of C4 depends on σ_{t-1}^2 being a function of $\epsilon_{t-2},\ \epsilon_{t-3},\ \ldots,\ \epsilon_0$. Given C3 and C4, (4) is a SRE (see also Mikosch and Stărică, 2000). At this point, all of the conditions in Goldie (1991, Theorem 4.1) are satisfied, in which case,

$$P\left(\sigma^2 > x\right) \sim cx^{-\kappa_0/2}. \tag{8}$$

The result in (6) then follows from Proposition 1. Lastly, summarizing a result originally from Breiman (1965), which is also stated as Mikosch (1999, Proposition 1.3.9(b)), consider two non-negative random variables X and Z that are also independent. If X is regularly varying with tail index θ, and $E\left(Z^{\theta+\eta}\right) < \infty$ for an η as defined above, then

$$P\left(XZ > x\right) \sim E\left(Z^{\kappa}\right) P\left(X > x\right). \tag{9}$$

Since $|Y| = \sigma\,|\epsilon|$, (7) immediately follows from (9). ∎

Remark 4 *Lemma 3 collects results available in the literature (see; e.g., Mikosch and Stărică, 2000, Theorem 2.1, and Basrak, Davis, and Mikosch, 2002, Theorem 3.1(B)).*

Next, for $0 \leq h < \infty$, consider

$$\mathbf{Y}_h^{(i)} = \left(\ |Y_0|^i,\quad \sigma_0^i,\quad \ldots,\quad |Y_h|^i,\quad \sigma_h^i\ \right),\quad i = 1, 2,$$

and

$$\widetilde{\mathbf{Y}}_m = \left(\ Y_0,\quad \sigma_0,\quad \ldots,\quad Y_h,\quad \sigma_h\ \right).$$

Lemma 5 *For the model in (2) and (3), under the same Assumptions as Lemma 2, $\mathbf{Y}_h^{(2)}$ is RV$(\kappa_0/2)$, while both $\mathbf{Y}_h^{(1)}$ and $\mathbf{Y}_h$ are RV(κ_0).*

Proof. Given (4) and (8), Mikosch and Stārica̅ (2000, proof of Theorem 2.3(a)) establishes $\mathbf{Y}_h^{(2)}$ as RV$(\kappa_0/2)$, which, in turn, establishes $\mathbf{Y}_h^{(1)}$ as RV(κ_0), given Mikosch (1999, Proposition 1.5.9), the multivariate extension of Proposition 1. Finally, $\mathbf{Y}_h$ is RV(κ_0) by Basrak et al. (2002, proof of Corollary 3.5(B)). ∎

Remark 6 *Lemma 5 pieces together different results available in the literature. The SRE upon which this lemma depends is (4) (see the proof of Lemma 3 that establishes (4) as a valid SRE). In contrast, the SRE upon which Basrak et al. (2002, Corollary 3.5) is based is more closely related to*

$$\begin{aligned}\mathbf{X}_t &= \begin{pmatrix} Y_t^2 \\ \sigma_t^2 \end{pmatrix} = \begin{pmatrix} \alpha\epsilon_t^2 & \beta\epsilon_t^2 \\ \alpha & \beta \end{pmatrix} \mathbf{X}_{t-1} + \begin{pmatrix} \omega\epsilon_t^2 \\ \omega \end{pmatrix} \\ &= \mathbf{A}_t\mathbf{X}_{t-1} + \mathbf{B}_t \end{aligned} \tag{10}$$

which is also a valid SRE, since it, too, satisfies Conditions C3 and C4.

A.1.3. Central Limit Theorem

Lemma 7 *For the model in (2) and (3), under the same Assumptions as Lemma 2, let*

$$\mathbf{Y}_t = \left(\begin{array}{cccccc} Y_t, & \sigma_t, & \ldots, & Y_{t+h}, & \sigma_{t+h} \end{array} \right), \qquad h < \infty,$$

and $\{a_n\}$ *be a sequence of constants satisfying*

$$nP\left(|\mathbf{Y}| > a_n\right) \longrightarrow 1, \qquad n \to \infty,$$

where $|\mathbf{Y}| = \max_{m=0,\ldots,h} |Y_m|$*;* $a_n = n^{1/\kappa_0} L\left(n\right)$*, and* $L\left(\cdot\right)$ *is slowly-varying at* ∞*. Then*

$$N_n := \sum_{t=1}^{n} \delta_{a_n^{-1}\mathbf{Y}_t} \xrightarrow{d} N := \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \delta_{P_i \mathbf{Q}_{i,j}},$$

where

1. $\sum_{i=1}^{\infty} \delta_{P_i}$ *is a Poisson process on* $\mathbb{R}_+$
2. $\sum_{j=1}^{\infty} \delta_{\mathbf{Q}_{i,j}},\ i \in \mathbb{N}$*, is an i.i.d. sequence of point processes on* $\mathbb{R}_+^{h+1} \setminus \{\mathbf{0}\}$ *with common distribution* Q
3. $\sum_{i=1}^{\infty} \delta_{P_i}$ *and* $\sum_{j=1}^{\infty} \delta_{\mathbf{Q}_{i,j}},\ i \in \mathbb{N}$*, are mutually independent*
4. $Q_{i,j} = \left(\left(\begin{array}{cc} Q_{ij,Y}^{(m)}, & Q_{ij,\sigma}^{(m)} \end{array} \right)_{m=0,\ldots,h}\right)$

Proof. The proof proceeds by verifying the conditions of Davis and Mikosch (1998, Theorem 2.8):

CONDITION C8: (joint) regular variation of all finite-dimmensional distributions of $\{\mathbf{Y}_t\}$

CONDITION C9: weak mixing for $\{\mathbf{Y}_t\}$

CONDITION C10: that

$$\lim_{k\to\infty}\lim_{n\to\infty} P\left(\bigvee_{k\leq|t|\leq r_n} |\mathbf{Y}_t| > a_n y \mid |\mathbf{Y}_0| > a_n y \right) = 0, \qquad y > 0, \tag{11}$$

where $\vee_i b_i = \max_i \left(b_i\right)$, and $r_n,\ m_n \to \infty$ are two integer sequences such that $n\phi_{m_n}/r_n \to 0$, $r_n m_n/n \to 0$, and ϕ_n is the mixing rate of $\{\mathbf{Y}_t\}$.

Lemma 5 establishes Condition C8. $\{\mathbf{Y}_t\}$ is strongly mixing by Carrasco and Chen (2002, Corollary 6). Finally, by the definition of the sequence $\{\mathbf{Y}_t\}$ and as in Mikosch and Stărică (2000, proof of Theorem 3.1), it suffices to switch in (11) to the sequence $\left\{\left(\begin{array}{cc} Y_t^2, & \sigma_t^2 \end{array}\right)\right\}$ and to replace $a_n y$ by $a_n^2 y^2$. Consequently, consider the SRE in (10). Recursive substitution establishes

$$\begin{aligned} \mathbf{X}_t &= \prod_{i=1}^{t} \mathbf{A}_i \mathbf{X}_0 + \sum_{i=1}^{t} \prod_{j=i+1}^{t} \mathbf{A}_j \mathbf{B}_i \\ &\equiv \mathbf{I}_{t,1} \mathbf{X}_0 + \mathbf{I}_{t,2} \end{aligned} \tag{12}$$

Condition C10 is then established following Davis, Mikosch, and Basrak (1999, proof of Theorem 3.3). ■

Remark 8 *Lemma 7 is the (nonstandard) CLT upon which (weak) distributional convergence of the GARCH*$(1,1)$ *estimators in Sections 2.1 of the main paper are based and generalizes Mikosch and Stārica̅ (2000, Theorem 3.1) by covering the case of an asymmetric* D*. Given Remark 6, Lemma 7 complements Basrak et al. (2002, Theorem 2.10). Finally, given a continuous mapping argument, implied by Lemma 7 for*

$$\mathbf{Y}_t^{(l)} = \left(\begin{array}{cccccc} Y_t^l, & \sigma_t^l, & \ldots, & Y_{t+h}^l, & \sigma_{t+h}^l \end{array}\right), \qquad l = 2, 3,$$

is that

$$N_n^{(l)} := \sum_{t=1}^{n} \delta_{a_n^{-l} \mathbf{Y}_t^{(l)}} \xrightarrow{d} N^{(l)} := \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \delta_{P_i \mathbf{Q}_{i,j}^{(l)}},$$

where

$$\mathbf{Q}_{i,j}^{(l)} = \left(\left(\begin{array}{cc} \left(Q_{ij,Y}^{(m)}\right)^l, & \left(Q_{ij,\sigma}^{(m)}\right)^l \end{array}\right)_{m=0,\ldots,h}\right).$$

A.1.4. GARCH(1,1) Convergence Results

From the model in (2) and (3) when $\alpha_{1,0} = \alpha_{2,0} = \alpha_0$,

$$X_t = \phi_0 X_{t-1} + V_t, \qquad V_t = W_t - \beta_0 W_{t-1}, \tag{13}$$

where $X_t \equiv Y_t^2 - \gamma_0$, and $\gamma_0 \equiv E\left(Y_t^2\right) = \frac{\omega_0}{1-\phi_0}$.

Lemma 9 *For the model in (2) and (3), under the same Assumptions as Lemma 2,*

$$a_n^{-3} \sum_t \sigma_t^3 - E\left(\sigma^3\right) \xrightarrow{d} V_{0,\sigma},$$

where "$\xrightarrow{d}$*" is weak, and* $V_{0,\sigma}$ *is* $(\kappa_0/3)$ *−stable.*

Proof. For an $\varepsilon > 0$,

$$\begin{aligned} a_n^{-3} \sum_t \sigma_t^3 - E\left(\sigma^3\right) &= a_n^{-3} \sum_t \left(\sigma_t^3 - E\left(\sigma^3\right)\right) \times I_{\{\sigma_t > a_n \varepsilon\}} + a_n^{-3} \sum_t \left(\sigma_t^3 - E\left(\sigma^3\right)\right) \times I_{\{\sigma_t \leq a_n \varepsilon\}} \\ &= Ia + IIa, \end{aligned}$$

where $E\left(\sigma^3\right) < \infty$ by Prono (2018, Lemma 1). Then,

$$\begin{aligned} a_n^{-3} \sum_t E\left(\sigma^3\right) \times I_{\{\sigma_t > a_n \varepsilon\}} &= a_n^{-3} E\left(\sigma^3\right) n \left(n^{-1} \sum_t I_{\{\sigma_t > a_n \varepsilon\}}\right) \\ &\sim a_n^{-3} E\left(\sigma^3\right) n P\left(\sigma_t > a_n \varepsilon\right) \\ &\quad \longrightarrow 0, \end{aligned} \tag{14}$$

where "$\sim$" holds for sufficiently large n, and "$\longrightarrow$" as $n \to \infty$ follows since

$$nP\left(\sigma_t > a_n \varepsilon\right) \longrightarrow \varepsilon^{-\kappa_0}, \qquad n \to \infty, \tag{15}$$

so that

$$Ia = a_n^{-3} \sum_t \sigma_t^3 \times I_{\{\sigma_t > a_n \varepsilon\}} + o_p(1).$$

Next,

$$\begin{aligned} a_n^{-3} \sum_t E\left(\sigma^3\right) \times I_{\{\sigma_t \leq a_n \varepsilon\}} &= n^{\frac{\kappa_0 - 6}{2\kappa_0}} E\left(\sigma^3\right) n^{-1/2} \sum_t I_{\{\sigma_t \leq a_n \varepsilon\}} \\ &\longrightarrow 0 \end{aligned}$$

as $n \to \infty$ by the CLT of Ibragimov and Linnik (1971, Theorem 18.5.3) applied to $n^{-1/2} \sum_t I_{\{\sigma_t \leq a_n \varepsilon\}}$ if $\kappa_0 < 6$, so that

$$IIa = a_n^{-3} \sum_t \sigma_t^3 \times I_{\{\sigma_t \leq a_n \varepsilon\}} + o_p(1).$$

Then, by Markov's Inequality for a $\zeta > 0$,

$$P\left(a_n^{-3} \sum_t \sigma_t^3 \times I_{\{\sigma_t \leq a_n \varepsilon\}} > \zeta\right) \leq n\left(\zeta^{-1} a_n^{-3}\right) E\left(\sigma^3 \times I_{\{\sigma \leq a_n \varepsilon\}}\right). \tag{16}$$

In addition, for $\kappa \equiv \kappa_0/3$, and $r \in (\kappa, 2)$, there exists a constant $C \in (0, \infty)$ such that

$$\begin{aligned} n\left(\zeta^{-1} a_n^{-3}\right) E\left(\sigma^3 \times I_{\{\sigma \leq a_n \varepsilon\}}\right) &\leq nC\left(\zeta^{-1} a_n^{-3}\right)^r E\left(\sigma^{3r} \times I_{\{\sigma \leq a_n \varepsilon\}}\right) \\ &\leq nC\left(\zeta^{-1} a_n^{-3}\right)^r \int_0^{a_n \varepsilon} \sigma^{3r} f(\sigma)\, d\sigma \\ &\leq nC\left(\zeta^{-1} a_n^{-3}\right)^r (-\kappa_0) \int_0^{a_n \varepsilon} \sigma^{3r - \kappa_0 - 1} L(\sigma)\, d\sigma, \end{aligned}$$

where the last inequality follows from Mikosch (1999, Theorem 1.2.9). Since, by Karamata's Theorem,

$$\int_0^{a_n\varepsilon} \sigma^{3r-\kappa_0-1} L(\sigma)\, d\sigma \sim \frac{\sigma^{3r-\kappa_0}}{-(3r-\kappa_0)} L(\sigma)\, \Big|_0^{a_n\varepsilon},$$

then

$$\begin{aligned} n\left(\zeta^{-1}a_n^{-3}\right) E\left(\sigma^3 \times I_{\{\sigma \le a_n\varepsilon\}}\right) &\le C\left(\zeta^{-1}a_n^{-3}\right)^r \left(\frac{\kappa_0}{3r-\kappa_0}\right)(a_n\varepsilon)^{3r}\, nP(\sigma > a_n\varepsilon) \\ &\longrightarrow C\zeta^{-r}\left(\frac{\kappa_0}{3r-\kappa_0}\right)\varepsilon^{3r-\kappa_0}, \\ &\longrightarrow 0, \end{aligned} \tag{17}$$

where the first "$\longrightarrow$" is as $n \to \infty$ and follows from (15), while the second "$\longrightarrow$" is as $\varepsilon \to 0$. As a consequence, $\lim_{n\to\infty} \lim_{\varepsilon\to 0} \sup P\left(a_n^{-3}\sum_t \sigma_t^3 \times I_{\{\sigma_t \le a_n\varepsilon\}} > \zeta\right) = 0$, and

$$a_n^{-3}\sum_t \sigma_t^3 - E\left(\sigma^3\right) = a_n^{-3}\sum_t \sigma_t^3 \times I_{\{\sigma_t > a_n\varepsilon\}} + o_p(1).$$

Finally, let

$$\mathbf{y}_t = \left(\begin{array}{ccccc} y_{t,Y}^{(0)}, & y_{t,\sigma}^{(0)}, & \ldots, & y_{t,Y}^{(h)}, & y_{t,\sigma}^{(h)} \end{array}\right) \in \mathbb{R}^{h+1} \setminus \{\mathbf{0}\}, \tag{18}$$

and define

$$T_{0,\varepsilon,\sigma}\left(\sum_{i=1}^{\infty} n_i \delta_{\mathbf{y}_i}\right) = \sum_{i=1}^{\infty} n_i \left(y_{i,\sigma}^{(0)}\right)^3 \times I_{\left\{y_{i,\sigma}^{(0)} > a_n\varepsilon\right\}}.$$

Since the set $\left\{\mathbf{y} \in \mathbb{R}^{h+1} \setminus \{\mathbf{0}\} : \left|y^{(m)}\right| > \varepsilon\right\}$ for any $m \ge 0$ is bounded away from the origin, and given Vaynman and Beare (2014, Lemma A.2), then

$$\begin{aligned} a_n^{-3}\sum_t \sigma_t^3 - E\left(\sigma^3\right) &= T_{0,\varepsilon,\sigma}(N_n) + o_p(1) \\ &\xrightarrow{d} T_{0,\varepsilon,\sigma}(N) \\ &\xrightarrow{d} V_{0,\sigma}, \end{aligned} \tag{19}$$

where the first "$\xrightarrow{d}$" is as $n \to \infty$ and follows from Lemma 7 and the continuous mapping theorem, while the second "$\xrightarrow{d}$" is as $\varepsilon \to 0$ and follows from Davis and Hsing (1995, Proof of Theorem 3.1, pp 897-898). ■

Lemma 10 *For the model in (2) and (3), under the same Assumptions as Lemma 2,*

$$a_n^{-3}\sum_t Y_t \sigma_{t+m}^2 - E\left(Y_t \sigma_{t+m}^2\right) \xrightarrow{d} \left(V_{m,\mathbf{y}}\right)_{m=1,\ldots,h}, \tag{20}$$

where "$\xrightarrow{d}$" continues to be weak, and $V_{m,\mathbf{y}}$ is $(\kappa_0/3)$ −stable.

Proof. The (weak) convergence result in (20) is established for $m = 1, 2$. Generalizing to cases where $m > 2$ is an extension of the arguments given below. Given (4),

$$\begin{aligned} & a_n^{-3} \sum_t Y_t \sigma_{t+1}^2 - E\left(Y_t \sigma_{t+1}^2\right) \\ = & a_n^{-3} \sum_t \sigma_t^3 \left(\epsilon_t A_{t+1} - \alpha_0 c_3^*\right) \times I_{\{\sigma_t > a_n \varepsilon\}} \\ & + a_n^{-3} \sum_t \sigma_t^3 \left(\epsilon_t A_{t+1} - \alpha_0 c_3^*\right) \times I_{\{\sigma_t \leq a_n \varepsilon\}} \\ & + \alpha_0 c_3^* a_n^{-3} \sum_t \sigma_t^3 - E\left(\sigma^3\right) + o_p(1) \\ = & a_n^{-3} \sum_t \sigma_t^3 \epsilon_t A_{t+1} \times I_{\{\sigma_t > a_n \varepsilon\}} \\ & + a_n^{-3} \sum_t \sigma_t^3 \left(\epsilon_t A_{t+1} - \alpha_0 c_3^*\right) \times I_{\{\sigma_t \leq a_n \varepsilon\}} + o_p(1), \end{aligned}$$

where the first equality relies on

$$a_n^{-1} \sum_t Y_t \xrightarrow{d} V_0, \tag{21}$$

which follows given Lemma 7 and Davis and Hsing (1995, Theorem 3.1) and under which V_0 is κ_0-stable, while the second equality follows from (19). Then for the same $r \in (1, 2)$ in the proof of Lemma 9 and a $\zeta > 0$,

$$\begin{aligned} P\left(a_n^{-3} \left|\sum_t \sigma_t^3 \times I_{\{\sigma_t \leq a_n \varepsilon\}} \times \left(\epsilon_t A_{t+1} - \alpha_0 c_3^*\right)\right| > \zeta\right) & \leq \left(\zeta^{-1} a_n^{-3}\right)^r E\left|\sum_t \sigma_t^3 \times I_{\{\sigma_t \leq a_n \varepsilon\}} \times \left(\epsilon_t A_{t+1} - \alpha_0 c_3^*\right)\right|^r \\ & \leq 2\left(\zeta^{-1} a_n^{-3}\right)^r n E\left(\sigma_t^{3r} \times I_{\{\sigma_t \leq a_n \varepsilon\}}\right) \\ & \quad \times E\left|\alpha_0\left(\epsilon_t^3 - c_3^*\right) + \beta_0 \epsilon_t\right|^r, \end{aligned}$$

where the first inequality follows from Markov's Inequality, and the second inequality follows from von Bahr and Esseen (1965, Theorem 2), since for

$$M_n \equiv \sum_t \sigma_t^3 \times I_{\{\sigma_t \leq a_n \varepsilon\}} \times \left(\epsilon_t A_{t+1} - \alpha_0 c_3^*\right),$$

$E\left(M_{n+1} \mid M_n\right) = M_n \; a.s.$[1] Given (17),

$$\lim_{n \to \infty} \lim_{\varepsilon \to 0} \sup P\left(a_n^{-3} \left|\sum_t \sigma_t^3 \times I_{\{\sigma_t \leq a_n \varepsilon\}} \times \left(\epsilon_t A_{t+1} - \alpha_0 c_3^*\right)\right| > \zeta\right) = 0.$$

Next, given (18), define

$$T_{m,\varepsilon,\mathbf{y}}\left(\sum_{i=1}^{\infty} n_i \delta_{\mathbf{y}_i}\right) = \sum_{i=1}^{\infty} n_i \left(y_{i,Y}^{(0)}\right)\left(y_{i,\sigma}^{(m)}\right)^2 \times I_{\left\{y_{i,\sigma}^{(0)} > a_n \varepsilon\right\}}, \qquad m \geq 1. \tag{22}$$

[1]The applicability of von Bahr and Esseen (1965, Theorem 2) in this general context is first recognized in Vaynman and Beare (2014, proof of Lemma A.1).

Then

$$a_n^{-3}\sum_t Y_t\sigma_{t+1}^2 - E\left(Y_t\sigma_{t+1}^2\right) = T_{1,\varepsilon,\mathbf{y}}\left(N_n\right) + o_p\left(1\right) \tag{23}$$
$$\xrightarrow{d} T_{1,\varepsilon,\mathbf{y}}\left(N\right)$$
$$\xrightarrow{d} V_{1,\mathbf{y}},$$

where the first "$\xrightarrow{d}$" is as $n \to \infty$, the second "$\xrightarrow{d}$" as $\varepsilon \to 0$, and each convergence result follows from the same arguments that support (19). Next, consider

$$\begin{aligned}
& a_n^{-3}\sum_t Y_t\sigma_{t+2}^2 - E\left(Y_t\sigma_{t+2}^2\right) \\
= \; & a_n^{-3}\sum_t Y_t\sigma_{t+1}^2\left(A_{t+2} - E\left(A\right)\right) \times I_{\{\sigma_t > a_n\varepsilon\}} \\
& + a_n^{-3}\sum_t Y_t\sigma_{t+1}^2\left(A_{t+2} - E\left(A\right)\right) \times I_{\{\sigma_t \le a_n\varepsilon\}} \\
& + E\left(A\right)a_n^{-3}\sum_t Y_t\sigma_{t+1}^2 - E\left(Y_t\sigma_{t+1}^2\right) + o_p\left(1\right) \\
= \; & a_n^{-3}\sum_t Y_t\sigma_{t+2}^2 \times I_{\{\sigma_t > a_n\varepsilon\}} \\
& + a_n^{-3}\sum_t Y_t\sigma_{t+1}^2\left(A_{t+2} - E\left(A\right)\right) \times I_{\{\sigma_t \le a_n\varepsilon\}} + o_p\left(1\right) \\
= \; & Ib + IIb + o_p\left(1\right),
\end{aligned}$$

where the first equality, again, relies on (4) and (21), while the second equality follows from (23). For

$$\begin{aligned}
IIb &= \alpha_0\omega_0 a_n^{-3}\sum_t Y_t\left(\epsilon_{t+1}^2 - 1\right) \times I_{\{\sigma_t \le a_n\varepsilon\}} + \alpha_0 a_n^{-3}\sum_t \sigma_t^3\epsilon_t A_{t+1} \times \left(\epsilon_{t+1}^2 - 1\right) \times I_{\{\sigma_t \le a_n\varepsilon\}} \\
&= \alpha_0 a_n^{-3}\sum_t \sigma_t^3\epsilon_t A_{t+1} \times \left(\epsilon_{t+1}^2 - 1\right) \times I_{\{\sigma_t \le a_n\varepsilon\}} + o_p\left(1\right),
\end{aligned}$$

where the second equality relies on the CLT of Ibragimov and Linnik (1971, Theorem 18.5.3). Next, for a $\zeta > 0$,

$$\begin{aligned}
P\left(a_n^{-3}\left|\sum_t \sigma_t^3 \times I_{\{\sigma_t \le a_n\varepsilon\}} \times \left(\alpha_0\epsilon_t^3 + \beta_0\epsilon_t\right) \times \left(\epsilon_{t+1}^2 - 1\right)\right| > \zeta\right) \; \le \; & 2\left(\zeta^{-1}a_n^{-3}\right)^r nE\left(\sigma_t^{3r} \times I_{\{\sigma_t \le a_n\varepsilon\}}\right) \\
& \times E\left|\alpha_0\epsilon_t^3 + \beta_0\epsilon_t\right|^r \times E\left|\epsilon_{t+1}^2 - 1\right|^r,
\end{aligned}$$

by Markov's Inequality and von Bahr and Esseen (1965, Theorem 2), so that

$$\lim_{n\to\infty}\lim_{\varepsilon\to 0}\sup P\left(a_n^{-3}\left|\sum_t \sigma_t^3 \times I_{\{\sigma_t \le a_n\varepsilon\}} \times \left(\alpha_0\epsilon_t^3 + \beta_0\epsilon_t\right) \times \left(\epsilon_{t+1}^2 - 1\right)\right| > \zeta\right) = 0;$$

in which case,

$$\begin{aligned}
a_n^{-3}\sum_t Y_t\sigma_{t+2}^2 - E\left(Y_t\sigma_{t+2}^2\right) &= a_n^{-3}\sum_t Y_t\sigma_{t+2}^2 \times I_{\{\sigma_t > a_n\varepsilon\}} + o_p(1) \\
&= T_{2,\varepsilon,\mathbf{y}}\left(N_n\right) + o_p(1) \\
&\xrightarrow{d} T_{2,\varepsilon,\mathbf{y}}(N) \\
&\xrightarrow{d} V_{2,\mathbf{y}},
\end{aligned}$$

where, as is true elsewhere, the first "$\xrightarrow{d}$" is as $n \to \infty$, and the second "$\xrightarrow{d}$" is as $\varepsilon \to 0$. ■

Lemma 11 *For the model in (2) and (3), under the same Assumptions as Lemma 2,*

$$a_n^{-3}\sum_t Y_t Y_{t+m}^2 - E\left(Y_t Y_{t+m}^2\right) \xrightarrow{d} \alpha_0^{-1}\left(V_{m+1,\mathbf{y}} - \beta_0 V_{m,\mathbf{y}}\right)_{m=1,\ldots,h}, \tag{24}$$

where, as is the case elsewhere, "$\xrightarrow{d}$" is weak, and the limits are $(\kappa_0/3)$ −stable.

Proof. The (weak) convergence result in (24) is established for $m = 1, 2$. Generalizing to $m > 2$ is an extension of the results stated below. From (4),

$$\epsilon_t^2 = \alpha_0^{-1}\left(A_{t+1} - \beta_0\right). \tag{25}$$

in which case,

$$\begin{aligned}
& a_n^{-3}\sum_t Y_t Y_{t+1}^2 - E\left(Y_t Y_{t+1}^2\right) \\
= \; & \alpha_0^{-1}\left(a_n^{-3}\sum_t Y_t\sigma_{t+1}^2 A_{t+2} - E\left(Y_t\sigma_{t+1}^2 A_{t+2}\right) - \beta_0 a_n^{-3}\sum_t Y_t\sigma_{t+1}^2 - E\left(Y_t\sigma_{t+1}^2\right)\right) \\
= \; & \alpha_0^{-1}\left(a_n^{-3}\sum_t Y_t\sigma_{t+2}^2 - E\left(Y_t\sigma_{t+2}^2\right) - \beta_0 a_n^{-3}\sum_t Y_t\sigma_{t+1}^2 - E\left(Y_t\sigma_{t+1}^2\right) - \omega_0 a_n^{-3}\sum_t Y_t\right) \\
= \; & \alpha_0^{-1}\left(a_n^{-3}\sum_t Y_t\sigma_{t+2}^2 - E\left(Y_t\sigma_{t+2}^2\right) - \beta_0 a_n^{-3}\sum_t Y_t\sigma_{t+1}^2 - E\left(Y_t\sigma_{t+1}^2\right) + o_p(1)\right) \\
& \xrightarrow{d} \alpha_0^{-1}\left(V_{2,\mathbf{y}} - \beta_0 V_{1,\mathbf{y}}\right),
\end{aligned}$$

where the second equality relies on (4), the third equality (21), and "$\xrightarrow{d}$" follows from Lemma 10. The same arguments then support

$$\begin{aligned}
& a_n^{-3}\sum_t Y_t Y_{t+2}^2 - E\left(Y_t Y_{t+2}^2\right) \\
= \; & \alpha_0^{-1}\left(a_n^{-3}\sum_t Y_t \sigma_{t+2}^2 A_{t+3} - E\left(Y_t \sigma_{t+2}^2 A_{t+3}\right) - \beta_0 a_n^{-3}\sum_t Y_t \sigma_{t+2}^2 - E\left(Y_t \sigma_{t+2}^2\right)\right) \\
= \; & \alpha_0^{-1}\left(a_n^{-3}\sum_t Y_t \sigma_{t+3}^2 - E\left(Y_t \sigma_{t+3}^2\right) - \beta_0 a_n^{-3}\sum_t Y_t \sigma_{t+2}^2 - E\left(Y_t \sigma_{t+2}^2\right) + o_p(1)\right) \\
& \xrightarrow{d} \alpha_0^{-1}\left(V_{3,\mathbf{y}} - \beta_0 V_{2,\mathbf{y}}\right),
\end{aligned}$$

which completes the proof. ■

Lemma 12 *For the model in (2) and (3), under the same Assumptions as Lemma 2,*

$$a_n^{-3}\sum_t Y_t^3 - E\left(Y_t^3\right) \xrightarrow{d} V_{0,\mathbf{y}},$$

where "$\xrightarrow{d}$" is weak, and $V_{0,Y}$ is $(\kappa_0/3)$ –stable.

Proof.

$$\begin{aligned}
a_n^{-3}\sum_t Y_t^3 - E\left(Y_t^3\right) &= a_n^{-3}\sum_t \sigma_t^3\left(\epsilon_t^3 - c_3^*\right) \times I_{\{\sigma_t \le a_n \varepsilon\}} \\
&\quad + a_n^{-3}\sum_t \sigma_t^3\left(\epsilon_t^3 - c_3^*\right) \times I_{\{\sigma_t > a_n \varepsilon\}} \\
&\quad + a_n^{-3}\sum_t \sigma_t^3 - E\left(\sigma^3\right) \\
&= Ic + IIc + IIIc.
\end{aligned}$$

As relied upon elsewhere, given Markov's Inequality and von Bahr and Esseen (1965, Theorem 2), for a $\zeta > 0$ and a $r \in (\kappa, 2)$ defined in the proof of Lemma 9,

$$\begin{aligned}
P\left(|Ic| > \zeta\right) &\le \left(\zeta^{-1} a_n^{-3}\right)^r E\left|\sum_t \sigma_t^3\left(\epsilon_t^3 - c_3^*\right) \times I_{\{\sigma_t \le a_n \varepsilon\}}\right|^r \\
&\le 2\left(\zeta^{-1} a_n^{-3}\right)^r nE\left(\sigma_t^{3r} \times I_{\{\sigma_t \le a_n \varepsilon\}}\right) \times E\left|\epsilon_t^3 - c_3^*\right|^r
\end{aligned}$$

so that

$$\lim_{n\to\infty}\lim_{\varepsilon\to 0}\sup P\left(|Ic| > \zeta\right) = 0 \tag{26}$$

by the arguments that support (17). Next, given (18),define

$$T_{0,\varepsilon,\mathbf{y}}\left(\sum_{i=1}^{\infty} n_i \delta_{\mathbf{y}_i}\right) = \sum_{i=1}^{\infty} n_i \left(y_{i,Y}^{(0)}\right)^3 \times I_{\left\{y_{i,\sigma}^{(0)} > a_n \varepsilon\right\}}.$$

Then, given Lemma 7,

$$\begin{aligned} a_n^{-3} \sum_t Y_t^3 - E\left(Y_t^3\right) &= a_n^{-3} \sum_t Y_t^3 \times I_{\{\sigma_t > a_n \varepsilon\}} - c_3^* a_n^{-3} \sum_t \sigma_t^3 \times I_{\{\sigma_t > a_n \varepsilon\}} + IIIc + o_p(1) \\ &= T_{0,\varepsilon,\mathbf{y}}(N_n) - c_3^* T_{0,\varepsilon,\sigma}(N_n) + IIIc \\ &\xrightarrow{d} T_{0,\varepsilon,\mathbf{y}}(N) \\ &\xrightarrow{d} V_{0,\mathbf{y}}, \end{aligned}$$

where $T_{0,\varepsilon,\sigma}(N_n)$ is defined in the proof of Lemma 9 and the sequential limiting results (first as $n \to \infty$ and then as $\varepsilon \to 0$) follow from the arguments given in that same proof. ■

Consider

$$\mathbf{Z}_{t-2} = \left(\begin{array}{ccc} Y_{t-2}, & \dots, & Y_{t-h} \end{array} \right)'$$

as a vector of (proper) instruments for X_{t-1} in (13). Then

$$\widehat{\phi}_{IV} = \widehat{\mathbf{F}} \left(n^{-1} \sum_t \widehat{X}_t \mathbf{Z}_{t-2} \right), \qquad \widehat{\mathbf{F}} = \frac{\left(n^{-1} \sum_t \widehat{X}_{t-1} \mathbf{Z}_{t-2} \right)' \widehat{\mathbf{\Lambda}}}{\left(n^{-1} \sum_t \widehat{X}_{t-1} \mathbf{Z}_{t-2} \right)' \widehat{\mathbf{\Lambda}} \left(n^{-1} \sum_t \widehat{X}_{t-1} \mathbf{Z}_{t-2} \right)}. \tag{27}$$

Theorem 13 *Let*

$$\mathbf{F}_0 = B_0^{-1} \mathbf{A}_0',$$

where

$$\mathbf{A}_0 = \mathbf{\Lambda}_0 E\left(X_{t-1} \mathbf{Z}_{t-2}\right), \qquad B_0 = E\left(X_{t-1} \mathbf{Z}_{t-2}\right)' \mathbf{A}_0.$$

In addition, let Assumptions A1–A5 from the main paper hold. Then

$$\widehat{\phi}_{IV} \xrightarrow{a.s.} \phi_0,$$

and

$$n a_n^{-3} \left(\widehat{\phi}_{IV} - \phi_0 \right) \xrightarrow{d} \alpha_0^{-1} \mathbf{F}_0 \mathbf{S}, \tag{28}$$

where $\kappa_0 \in (3,\ 6)$*, "*$\xrightarrow{d}$*" is weak,*

$$\mathbf{S} = \left((V_{m+1,\mathbf{y}} - \beta_0 V_{m,\mathbf{y}})_{m=2,\dots,h} \right),$$

each $(V_{m,\mathbf{y}})_{m=2,\dots,h+1}$ *is defined in Lemma 16, and* S *is jointly* $(\kappa_0/3)$ *−stable. If* $\kappa_0 \in (6,\ \infty)$ *such that* $E\left(Y_t^6\right) < \infty$*, then*

$$\sqrt{n} \left(\widehat{\phi}_{IV} - \phi_0 \right) \xrightarrow{d} N\left(0,\ \frac{\mathbf{A}_0' \mathbf{\Sigma}_{V\mathbf{Z}_{-2}} \mathbf{A}_0}{B_0^2} \right),$$

where

$$\boldsymbol{\Sigma}_{V\mathbf{Z}_{-2}} = E\left(V_t^2 \mathbf{Z}_{t-2}\mathbf{Z}_{t-2}'\right) + 2E\left(V_t V_{t-1}\mathbf{Z}_{t-2}\mathbf{Z}_{t-3}'\right),$$

and V_t is defined in Theorem 1 of the main paper.

Proof. Since

$$\widehat{X}_t = X_t - (\widehat{\gamma} - \gamma_0), \tag{29}$$

$$\widehat{X}_t = c_0 + \phi_0 \widehat{X}_{t-1} - \beta_0 W_{t-1} + W_t, \qquad c_0 = (\widehat{\gamma} - \gamma_0) \times (\phi_0 - 1) \tag{30}$$

given (13). Also, since $\{Y_t\}$ is strongly mixing (see the proof of Theorem 1 in the Appendix of the main paper), then given (29),

$$\begin{aligned} n^{-1}\sum_t \widehat{X}_{t-1}\mathbf{Z}_{t-2} &= n^{-1}\sum_t X_{t-1}\mathbf{Z}_{t-2} - (\widehat{\gamma} - \gamma_0)\, n^{-1}\sum_t \mathbf{Z}_{t-2} \\ &\xrightarrow{a.s.} E\left(X_{t-1}\mathbf{Z}_{t-2}\right) \end{aligned}$$

by the Ergodic Theorem so that $\widehat{\mathbf{F}} \xrightarrow{a.s.} \mathbf{F}_0$. Also, given (30),

$$\begin{aligned} n^{-1}\sum_t \widehat{X}_t\mathbf{Z}_{t-2} &= c_0 n^{-1}\sum_t \mathbf{Z}_{t-2} + \phi_0 n^{-1}\sum_t \widehat{X}_{t-1}\mathbf{Z}_{t-2} - \beta_0 n^{-1}\sum_t W_{t-1}\mathbf{Z}_{t-2} + n^{-1}\sum_t W_t\mathbf{Z}_{t-2} \\ &\xrightarrow{a.s.} \phi_0 E\left(X_{t-1}\mathbf{Z}_{t-2}\right). \end{aligned}$$

Next,

$$\begin{aligned} n a_n^{-3}\left(\widehat{\phi}_{IV} - \phi_0\right) &= \mathbf{F}_0\left(a_n^{-3}\sum_t X_t\mathbf{Z}_{t-2} - E\left(X_t\mathbf{Z}_{t-2}\right)\right) + o_p(1) \\ &= \mathbf{F}_0\left(a_n^{-3}\sum_t Y_t^2\mathbf{Z}_{t-2} - E\left(Y_t^2\mathbf{Z}_{t-2}\right)\right) + o_p(1) \\ &\xrightarrow{d} \alpha_0^{-1}\mathbf{F}_0\mathbf{S} \end{aligned}$$

where the second equality follows from the arguments that support (**XX**) in the proof of Theorem 1 in the Appendix of the main paper, and **S** is jointly $(\kappa_0/3)$ −stable by Lemma 17 and Samorodnitsky and Taqqu

(1994, Theorem 2.1.5(c)). If $\kappa_0 \in (6,\ \infty)$ so that $E\left(Y_t^6\right) < \infty$, then

$$\begin{aligned}
\sqrt{n}\left(\widehat{\phi} - \phi_0\right) &= \sqrt{n}\left(\frac{\phi_0\left(n^{-1}\sum_t X_t \mathbf{Z}_{t-2}\right)' \widehat{\mathbf{A}}}{\widehat{B}} - \phi_0 + \frac{\left(n^{-1}\sum_t V_t \mathbf{Z}_{t-2}\right)' \widehat{\mathbf{A}}}{\widehat{B}} + o_p(1)\right) \\
&= \sqrt{n}\left(\frac{\left(n^{-1}\sum_t V_t \mathbf{Z}_{t-2}\right)' \widehat{\mathbf{A}}}{\widehat{B}} + o_p(1)\right) \\
&\xrightarrow{d} N\left(0,\ \frac{\mathbf{A}_0' \mathbf{\Sigma}_{V\mathbf{Z}_{-2}} \mathbf{A}_0}{B_0^2}\right),
\end{aligned}$$

where the limiting result uses the same CLT from the proof of Theorem 1. ■

Consistency of $\widehat{\phi}_{IV}$ does not depend on consistency of $\widehat{\gamma}$, and $\widehat{\gamma}$ does not impact the limiting distribution of $\widehat{\phi}_{IV}$. Necessary for $B_0 \neq 0$ is $E\left(Y_t^3\right) \neq 0$, which illustrates the lack of identification that results if in A1, D is a symmetric distribution. (28) depends on $j \in (3,\ 6)$ in A1, which is consistent with empirical findings. Lastly, the rate of convergence that applies in (28) is $n^{\frac{\kappa_0 - 3}{3}}$.

References

[1] Basrak, B., R.A Davis & T. Mikosch (2002) Regular variation of GARCH processes. *Stochastic Processes and Their Applications* 99, 95-115.

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

[3] Breiman, L. (1965) On some limit theorems similar to the arc sin law. *Theory Probab. Appl.* 10, 323-331.

[4] Carrasco, M. & X. Chen (2002) Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory* 18, 17-39.

[5] Davis, R.A. & T. Hsing (1995) Point process and partial sum convergence for weakly dependent random variables with infinite variance. *The Annals of Probability* 23, 879-917.

[6] Davis, R.A. & T. Mikosch (1998) The sample autocorrelations of heavy-tailed processes with applications to ARCH. *The Annals of Statistics* 26, 2049-2080.

[7] Davis, R.A., T. Mikosch and B. Basrak (1999) Sample ACF of multivariate stochastic recurrence equations with application to GARCH. Unpublished manuscript.

[8] Goldie, C.M. (1991) Implicit renewal theory and tails of solutions of random equations. *Ann. Appl. Probab.* 1, 126-166.

[9] Ibragimov, I.A. and Y.V. Linnik (1971) *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhoff: Groningen.

[10] Mikosch, T. (1999) Regular Variation, Subexponentiality and their applications in probability theory. Lecture notes for the workshop "Heavy Tails and Queques," EURANDOM, Eindhoven, Netherlands.

[11] Mikosch, T. & C. Stărică (2000) Limit theory for the sample autocorrelations and extremes of a GARCH(1,1) process. *The Annals of Statistics* 28, 1427-1451.

[12] Prono, T. (2018) Closed-form estimators for finite-order arch models as simple and competitive alternatives to qmle: supplemental appendix. Forthcoming in *Studies in Nonlinear Dynamics and Econometrics*.

[13] Resnick, S.I. (2007) *Probabilistic and Statistical Modeling of Heavy Tailed Phenomena*. New York: Springer-Verlag.

[14] Samorodnitsky, G. & M.S. Taqqu (1994) *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance. Stochastic Modeling*. New York: Chapmand and Hall.

[15] Vaynman, I. & B.K. Beare (2014) Stable limit theory for the variance targeting estimator, in Y. Chang, T.B. Fomby & J.Y. Park (eds), *Essays in Honor of Peter C.B. Phillips,* vol. 33 of *Advances in Econometrics:* Emerald Group Publishing Limited, chapter 24, 639-672.

[16] von Bahr, B., & C.G. Esseen (1965) Inequalities for the rth absolute moment of a sum of random variables, $1 \leq r \leq 2$. *Annals of Mathematical Statistics* 36, 299-303.