Robust Estimation and Inference for Heavy Tailed Nonlinear GARCH [1]

Jonathan B. Hill[1,2]

Dept. of Economics, University of North Carolina

Feb. 2012

ABSTRACT

We develop new tail-trimmed QML estimators for nonlinear GARCH models with possibly heavy tailed errors. Tail-trimming allows both identification of the true parameter and asymptotic normality. In heavy tailed cases the rate of convergence is below but arbitrarily close to $\sqrt{n}$, the highest possible amongst M-estimators for GARCH with errors that have an infinite fourth moment, and faster than QML. We present a consistent estimator of the covariance matrix that permits classic inference without knowledge of the rate of convergence. Finally, a simulation study shows our estimators trump existing ones for sharpness and approximate normality, and we apply them to financial returns data.

1. INTRODUCTION It is now widely accepted that log-returns of many macroeconomic and financial time series are heavy tailed, asymmetrically distributed, and exhibit clustering of large values. In broader contexts extremes are encountered in actuarial, meteorological, and telecommunication network data (e.g. Leadbetter et al 1983, Engle and Ng 1993, Glosten et al 1993, Embrechts et al 1997). GARCH-type clustering alone implies higher moments do not exist due to Pareto-like distribution tails (e.g. Basrak et al 2002, Cline 2007), while nonlinear GARCH models have evolved in the literature as essential representations

[1] Dept. of Economics, University of North Carolina-Chapel Hill, www.unc.edu/∼jbhill, jbhill@email.unc.edu.

Key words and phrases: QML; tail trimming; heavy tails; nonlinear GARCH; robust inference.

AMS subject classifications. Primary 62F35; secondary 62F07.

[2] This paper is a substantially reduced version of a circulated working paper by Hill (2011a): "Robust M-Estimation for Heavy Tailed Nonlinear AR-GARCH".

of the above stylized traits, while retaining parsimony (e.g. Nelson 1991, Engle and Ng 1993, Glosten et al 1993, Engle and Rangle 2007).

We develop new methods of robust M-estimation for a nonlinear GARCH(1,1) model:

$$y_t = \sigma_t \epsilon_t \text{ and } \sigma_t^2 = g\left(y_{t-1}, \sigma_{t-1}^2, \theta^0\right) \text{ where } \theta^0 \in \mathbb{R}^q, \tag{1}$$

where g is a known response function and ϵ_t is a random variable representing noise or model error. We assume there exists a unique θ^0 such that $\{y_t, \epsilon_t, \sigma_t^2\}$ is stationary and geometrically β-mixing, the error satisfies $E[\epsilon_t] = 0$ and $E[\epsilon_t^2] = 1$, and θ^0 minimizes our criterion asymptotically. Hence θ^0 is allowed to be a "pseudo-true" parameter and the above model may be mis-specified, for example we do not require ϵ_t or $\epsilon_t^2 - 1$ to be martingale differences (Sawa 1978, White 1982). We are particularly interested in heavy tailed errors $E[\epsilon_t^4] = \infty$, while GARCH feedback itself may also prompt heavy tails in y_t as long as stationarity holds.

In this paper we tail-trim a QML loss function in three ways to obtain asymptotically normal Quasi-Maximum Tail-Trimmed Likelihood [QMTTL] estimators of θ^0. The first method exploits an asymptotic first order expansion by targeting extremes based on score and Jacobian equations, while the second exploits the score and Jacobian constructions and focuses trimming to the source of extremes. The third, however, re-centers the trimmed errors to eliminate bias when the errors are independent and identically distributed, or i.i.d., and uses a Method of Tail-Trimmed Moments criterion for estimation. See Sections 2 and 3. The reader can use the main results to support other loss functions like non-Gaussian ML, LAD and its variants (e.g. Peng and Yao 2003) and Whittle estimation; and extensions of (1) like nonlinear AR-GARCH (Hill 2011a).

In simple cases we show how trimming and distribution tail parameters impact efficiency, while the negligible amount of trimming never affects the asymptotic covariance matrix when $E[\epsilon_t^4] < \infty$. Fixed quantile trimming or truncation may by construction influence efficiency irrespective of higher moments, and may cause bias due to asymmetry in the data generating process (e.g. Ronchetti and Trojani 2001, Ling 2007, Čižek 2008).

The convergence rate of our estimators is $o(\sqrt{n})$ for strong-GARCH models with $E[\epsilon_t^4] = \infty$, but can be assured to be $\sqrt{n}/L(n)$ for slowly varying $L(n) \to \infty$ by following simple trimming rules of thumb. Hence, in general QMTTL converges faster than QML (cf. Hall and Yao 2003) but slower than Peng and Yao's (2003) $\sqrt{n}$-conovergent Log-LAD estimator for linear strong-GARCH. See Section 4.

In Section 5 we show classic inference applies as long as self-normalization is used, a nice convenience since tail thickness and the precise rate of convergence need never be known. We complete the paper with simulation and empirical studies in Sections 6 and 7.

Choosing the trimming portion in practice is not transparent, in particular because the literature almost exclusively focuses on fixed quantile methods for outlier robust estimation (see Agulló et al 2008 for references). In the supplemental material Hill (2012) we therefore adapt a bootstrap mean-squared-error method for selecting a tail-trimming tuning parameter, although our simulation study shows simply picking small parameter values leads to sharp estimates. See also Hill (2011a).

A complete theory of QML for linear and nonlinear strong-GARCH is presented in Lee and Hansen (1994), Straumann and Mikosch (2006) and Meitz and Saikkonen (2009), amongst others, while at least a finite fourth moment $E[\epsilon_t^4] < \infty$ is standard (Francq and Zakoïan 2004). Estimation by QML for semi-strong linear GARCH similarly requires $E[\epsilon_t^4] < \infty$ (Escanciano 2009).

A class of Log-LAD estimators for strong and semi-strong GARCH are $\sqrt{n}$-convergent if $E[\epsilon_t^2] = 1$ and $\ln(\epsilon_t^2)$ is symmetric (Peng and Yao 2003, Linton et al 2010), while Whittle estimation for GARCH requires $E[\epsilon_t^8] < \infty$ (Mikosch and Straumann 2002). Ling (2007) establishes asymptotic normality for the Quasi-Maximum Weighted Likelihood [QMWL] estimator for linear strong GARCH. Since the class of weights depend on lags y_{t-i} and not ϵ_t, the error must have a finite fourth moment. A purely computational treatment of outlier robust M-estimation for GARCH is treated in Boudt and Croux (2010).

The QMTTL criterion for GARCH(1,1), by comparison, is based solely on ϵ_t if we know there are GARCH effects, and otherwise on only ϵ_t, y_{t-1} and y_{t-2} due to the form of first order equations. We only require $E[\epsilon_t^2] < \infty$ if ϵ_t is i.i.d., in which case if $E[\epsilon_t^4] < \infty$ then QMTTL is asymptotically equivalent to QML. If the errors are not i.i.d. then our assumptions are more general than Escanciano (2009) because we allow for nonlinear GARCH with possibly heavy tailed and non-martingale difference errors.

We use the following notation conventions. The indicator function is $I(A) = 1$ if A is true, and otherwise $I(A) = 0$. If A is a square matrix then $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote the minimum and maximum eigenvalues. The L_r-norm of a random $M \times N$ matrix A is $||A||_r = (\sum_{i=1,j=1}^{M,N} E|A_{i,j}|^r)^{1/r}$ and the spectral norm is $||A||$ $= \lambda_{\max}(A'A)^{1/2}$. If z is a scalar we write $(z)_+ := \max\{0, z\}$. K denotes a positive finite constant whose value may change from line to line; $\iota > 0$ is an arbitrarily tiny constant. $\overset{p}{\to}$ and $\overset{d}{\to}$ denote probability and distribution convergence. $x_n \sim a_n$ denotes $x_n/a_n \to 1$; $x_n = o(a_n)$ denotes $x_n/a_n \to 0$, and $x_n = o_p(a_n)$ means $x_n/a_n \overset{p}{\to} 0$. $L(n)$ is a slowly varying function that may change with the context. $\epsilon_t \overset{i.i.d.}{\sim} (0,1)$ states ϵ_t is i.i.d. with zero mean and unit variance. A random variable is *symmetric* if its distribution is symmetric about zero.

2. QUASI-MAXIMUM TAIL-TRIMMED LIKELIHOOD

Assume each $w_t \in \{y_t, \epsilon_t, \sigma_t^2\}$ is stationary and ergodic, and $E|w_t|^\iota < \infty$ for some tiny $\iota > 0$. Recall $\sigma_t^2 = g(y_{t-1}, \sigma_{t-1}^2, \theta^0)$: the response function maps $g : \mathbb{R} \times \mathbb{R}_+ \times \Theta \to \mathbb{R}_+$, where $\Theta \subset \mathbb{R}^q$ is a compact parameter space with $q \geq 1$. The sample is $\{y_t\}_{t=0}^n$ with sample size $n \geq 1$ to simplify notation since we condition on the first observation y_0 and a volatility constant defined below.

Estimation requires an approximation to the volatility process $\{\sigma_t^2\}$ since y_t for $t < 0$ is not observed. Define an iterated volatility process

$$h_t(\theta) = \tilde{\omega} > 0 \text{ for } t = 0, \text{ and } h_t(\theta) = g\left(y_{t-1}, h_{t-1}(\theta), \theta\right) \text{ for } t = 1, 2, ... \tag{2}$$

where $\tilde{\omega}$ is not necessarily an element of θ^0. Assume $g(y, h, \theta)$ is twice differentiable in (y, h, θ), and define

$$h_t^\theta(\theta) := (\partial/\partial\theta) h_t(\theta) \text{ and } h_t^{\theta,\theta}(\theta) := (\partial/\partial\theta) h_t^\theta(\theta).$$

We need stationary solutions of $\{h_t(\theta), h_t^\theta(\theta), h_t^{\theta,\theta}(\theta)\}$ denoted $\{h_t^*(\theta), h_t^{*\theta}(\theta), h_t^{*\theta,\theta}(\theta)\}$, where $h_t^*(\theta)$ solves (2) and $\{h_t^{*\theta}(\theta), h_t^{*\theta,\theta}(\theta)\}$ solve related difference equations, e.g. eq.'s (9) and (10) in Meitz and Saikkonen (2009), hereafter MS (2009). See Appendix A for a proof stationary solutions exist based on Lipschitz type bounds on g and its derivative.

2.1 TAIL-TRIMMED QML

Define error and scaled volatility derivatives

$$\epsilon_t(\theta) := \frac{y_t}{h_t^{1/2}(\theta)}, \ \ \mathfrak{s}_t(\theta) := \frac{1}{h_t(\theta)} h_t^{\theta}(\theta) \in \mathbb{R}^q \ \text{ and } \ \mathfrak{s}_t^{\theta}(\theta) := \frac{\partial}{\partial\theta}\mathfrak{s}_t(\theta) \in \mathbb{R}^{q\times q},$$

and let $l_t(\theta)$ denote QML loss equations

$$l_t(\theta) := \ln h_t(\theta) + \epsilon_t^2(\theta).$$

QML asymptotics for smooth GARCH models is based on the first and second derivatives which we call *score* and *Jacobian* equations, respectively

$$\begin{aligned} m_t(\theta) &:= \frac{\partial}{\partial\theta} l_t(\theta) = -\left(\epsilon_t^2(\theta) - 1\right) \mathfrak{s}_t(\theta) \in \mathbb{R}^q \\ G_t(\theta) &:= \frac{\partial}{\partial\theta} m_t(\theta) = -\left(\epsilon_t^2(\theta) - 1\right) \mathfrak{s}_t^{\theta}(\theta) + \mathfrak{s}_t(\theta)\mathfrak{s}_t(\theta)' \in \mathbb{R}^{q\times q}. \end{aligned} \quad (3)$$

We now drop θ^0 and write $m_t = m_t(\theta^0)$, etc. We assume the analyst has $m_t(\theta)$ and $G_t(\theta)$ in hand since in general $m_t(\theta)$ and $G_t(\theta)$ can be computed by numerical approximation if an analytic expression does not exist.

If we trim an asymptotically vanishing sample portion of $l_t(\theta)$ when a large $m_{i,t}(\theta)$ or $G_{i,j,t}(\theta)$ occurs then a consistent and asymptotically normal estimator of θ^0 is achievable. This is possible because an expanded first order condition shows m_t and G_t can be sufficiently but negligibly trimmed ensuring Gaussian asymptotics and identification of θ^0. We must trim G_t because it need not be integrable for $t \geq 1$: $(\epsilon_t^2 - 1)\mathfrak{s}_t^{\theta}$ may not be integrable if $\mathfrak{s}_t^{\theta}$ is not integrable and ϵ_t is not independent, and $\mathfrak{s}_{i,t}^2$ may not be integrable. A simple example is the case where $\sigma_t = 1$ *a.s.* and $E[\epsilon_t^4] = \infty$ since $\mathfrak{s}_{i,t}$ is a constant or lag of ϵ_t^2 for $t \geq 1$. If $m_t(\theta)$ and $G_t(\theta)$ are simple functions of random variables we may focus trimming there, which is the topic of Section 3.

Denote left, right and two tail observations and their order statistics for any w_t:

$$w_t^{(-)} := w_t I\left(w_t < 0\right) \ \ \textit{and} \ \ w_{(1)}^{(-)} \leq \cdots \leq w_{(n)}^{(-)} \leq 0$$

$$w_t^{(+)} := w_t I\left(w_t \geq 0\right) \ \ \textit{and} \ \ w_{(1)}^{(+)} \geq \cdots \geq w_{(n)}^{(+)} \geq 0$$

$$w_t^{(a)} := |w_t| \ \ \textit{and} \ \ w_{(1)}^{(a)} \geq \cdots \geq w_{(n)}^{(a)} \geq 0$$

The determination of large $m_{i,t}(\theta)$ and $G_{i,j,t}(\theta)$ is made by intermediate order sequences $\{k_{1,i,n}^{(m)}, k_{2,i,n}^{(m)}\}$ and $\{k_{i,j,n}^{(G)}\}$: if $\{k_n\}$ denotes any one of them then assume $1 \leq k_n < n$, $k_n \to \infty$ and $k_n/n \to 0$. See Leadbetter et al (1983) and Hahn et al (1991). Define indicator selection functions

$$\hat{I}_{i,n,t}^{(m)}(\theta) := I\left(m_{i,(k_{1,i,n}^{(m)})}^{(-)}(\theta) \leq m_{i,t}(\theta) \leq m_{i,(k_{2,i,n}^{(m)})}^{(+)}(\theta)\right) \text{ and } \hat{I}_{i,j,n,t}^{(G)}(\theta) := I\left(|G_{i,j,t}(\theta)| \leq G_{i,j,(k_{i,j,n}^{(G)})}^{(a)}(\theta)\right),$$

and a composite trimming indicator

$$\hat{I}_{n,t}(\theta) = \prod_{1 \leq i \leq q} \hat{I}_{i,n,t}^{(m)}(\theta) \times \prod_{1 \leq i \leq j \leq q} \hat{I}_{i,j,n,t}^{(G)}(\theta) := \hat{I}_{n,t}^{(m)}(\theta) \times \hat{I}_{n,t}^{(G)}(\theta),$$

where $\prod_{1 \leq i \leq j \leq q} \hat{I}_{i,j,n,t}^{(G)}(\theta)$ omits redundant indicators. Our first QMTTL estimator therefore solves

$$\hat{\theta}_n = \underset{\theta \in \Theta}{\operatorname{arginf}} \left\{ \frac{1}{n} \sum_{t=1}^{n} l_t(\theta) \hat{I}_{n,t}(\theta) \right\}.$$

Each k_n represents the number of trimmed $l_t(\theta)$ due to large $m_{i,t}(\theta)$ or $G_{i,j,t}(\theta)$. We require $k_n \to \infty$ for asymptotic normality, while negligibility $k_n/n \to 0$ ensures identification of θ^0 asymptotically. Yet choosing a complete policy $\{k_{1,i,n}^{(m)}, k_{2,i,n}^{(m)}, k_{i,j,n}^{(G)}\}$ in practice is an outstanding challenge with few avenues provided by the literature, although in many cases the problem reduces to choosing one, or very few, sequences $\{k_n\}$, depending on the model and symmetry (see Sections 3 and 4). In the supplemental appendix Hill (2012) we adapt a bootstrap method for fixed quantile selection problems to our needs.

Identification of θ^0 is expedited asymptotically if we negligibly asymmetrically trim by $m_{i,t}(\theta)$ since $m_{i,t}(\theta)$ is in general asymmetric, although if $m_{i,t}$ is known to be symmetric then we use

$$\hat{I}_{i,n,t}^{(m)}(\theta) := I\left(|m_{i,t}(\theta)| \leq m_{i,(k_{i,n}^{(m)})}^{(a)}(\theta)\right) \text{ where } m_{i,t}^{(a)}(\theta) := |m_{i,t}(\theta)|, \; k_{i,n}^{(m)} \to \infty \text{ and } k_{i,n}^{(m)} = o(n).$$

In terms of the Jacobian equations, we need only negligibly trim to ensure a probability limit for $1/n \sum G_{i,j,t}(\theta) \hat{I}_{i,j,n,t}^{(G)}(\theta)$.

In order to characterize the limit distribution of $\hat{\theta}_n$ we require non-random thresholds which the order statistics $m_{i,(k_{1,i,n}^{(m)})}^{(-)}(\theta)$, $m_{i,(k_{2,i,n}^{(m)})}^{(+)}(\theta)$ and $G_{i,j,(k_{i,j,n}^{(G)})}^{(a)}(\theta)$ approximate. Let $l_t^*(\theta)$ denote the QML loss evaluated with $h_t^*(\theta)$, and similarly define score and Jacobian functions:

$$m_t^*(\theta) := \frac{\partial}{\partial \theta} l_t^*(\theta) \in \mathbb{R}^q \text{ and } G_t^*(\theta) := \frac{\partial}{\partial \theta} m_t^*(\theta) \in \mathbb{R}^{q \times q}.$$

Although we do not express it, technically these derivatives exist *a.s.* at any point θ.

Now define sequences $\{\mathcal{L}_{i,n}^{(m)}(\theta),\mathcal{U}_{i,n}^{(m)}(\theta)\}$ denoting the lower $k_{1,i,n}^{(m)}/n$ and upper $k_{2,i,n}^{(m)}/n$ quantiles of $m_{i,t}^{*}(\theta)$:

$$P\left(m_{i,t}^{*}(\theta)\leq -\mathcal{L}_{i,n}^{(m)}(\theta)\right)=\frac{k_{1,i,n}^{(m)}}{n}\ \text{ and } P\left(m_{i,t}^{*}(\theta)\geq \mathcal{U}_{i,n}^{(m)}(\theta)\right)=\frac{k_{2,i,n}^{(m)}}{n}.$$

Similarly, $\{\mathcal{C}_{i,j,n}^{(G)}(\theta)\}$ are two-tailed upper $k_{i,j,n}^{(G)}/n$ quantiles of $G_{i,j,t}^{*}(\theta)$: $P(|G_{i,j,t}^{*}(\theta)|\geq \mathcal{C}_{i,j,n}^{(G)}(\theta))=k_{i,j,n}^{(G)}/n$. Under symmetric trimming for $m_{i,t}^{*}(\theta)$ we use $I_{i,n,t}^{(m^{*})}(\theta):=I(|m_{i,t}^{*}(\theta)|\leq \mathcal{C}_{i,n}^{(m)}(\theta))$ where $P(|m_{i,t}^{*}(\theta)|\geq \mathcal{C}_{i,n}^{(m)}(\theta))$ $=k_{i,n}^{(m)}/n$. Thresholds $\{\mathcal{L}_{i,n}^{(m)}(\theta),\mathcal{U}_{i,n}^{(m)}(\theta)\}$ and $\{\mathcal{C}_{i,j,n}^{(G)}(\theta)\}$ exist for any $\{k_n\}$ since we assume $m_{i,t}^{*}(\theta)$ and $G_{i,j,t}^{*}(\theta)$ have smooth distributions, and under a mixing condition the order statistics are consistent for the above thresholds, e.g. $m_{i,(k_{2,i,n}^{(m)})}^{(+)}(\theta)/\mathcal{U}_{i,n}^{(m)}(\theta)\overset{p}{\rightarrow}1$ uniformly on Θ. See below for all assumptions, and see Appendix B for supporting limit theory.

The selection indicators are

$$I_{i,n,t}^{(m^{*})}(\theta):=I\left(-\mathcal{L}_{i,n}^{(m)}(\theta)\leq m_{i,t}^{*}(\theta)\leq \mathcal{U}_{i,n}^{(m)}(\theta)\right)\ \text{ and }\ I_{i,j,n,t}^{(G^{*})}(\theta):=I\left(\left|G_{i,j,t}^{*}(\theta)\right|\leq \mathcal{C}_{i,j,n}^{(G)}(\theta)\right)$$

and a composite indicator is

$$I_{n,t}^{*}(\theta)=\prod_{1\leq i\leq q}I_{i,n,t}^{(m^{*})}(\theta)\times\prod_{1\leq i\leq j\leq q}I_{i,j,n,t}^{(G^{*})}(\theta):=I_{n,t}^{(m^{*})}(\theta)\times I_{n,t}^{(G^{*})}(\theta).$$

Now define deterministically tail-trimmed score and Jacobian equations

$$m_{n,t}^{*}(\theta):=m_{t}^{*}(\theta)I_{n,t}^{*}(\theta)\ \text{ and }\ G_{n,t}^{*}(\theta):=G_{t}^{*}(\theta)I_{n,t}^{*}(\theta),$$

and long run covariance, Jacobian and scale matrices:

$$\mathcal{S}_{n}(\theta):=\frac{1}{n}\sum_{s,t=1}^{n}E\left[m_{n,s}^{*}(\theta)m_{n,t}^{*}(\theta)'\right],\ \mathcal{G}_{n}(\theta):=E\left[G_{n,t}^{*}(\theta)\right]\ \text{ and }\ V_{n}(\theta)=n\mathcal{G}_{n}(\theta)'\mathcal{S}_{n}^{-1}(\theta)\mathcal{G}_{n}(\theta). \tag{4}$$

2.2 ASSUMPTIONS

We now detail all assumptions concerning smoothness, identification, dependence, tail decay and fractile bounds. We say $g(y,h,\theta)$ is Lipschitz in h if $||g(y,h_1,\theta)-g(y,h_2,\theta)||\leq K|h_1-h_2|\ \forall h_1,h_2\in\mathbb{R}_{+}$ and $y,\theta\in\mathbb{R}\times\Theta$. Now let $a,b\in\{y,h,\theta\}$ be indices, drop all arguments, and let g_a and $g_{a,b}$ denote first and second derivatives, e.g. $g_{y,\theta}=(\partial/\partial\theta)(\partial/\partial y)g(y,h,\theta)$. Response differentiability greatly simplifies asymptotic theory under tail trimming, while Lipschitz in h ensures stationary solutions $\{h_t^{*}(\theta),h_t^{*\theta}(\theta),h_t^{*\theta,\theta}(\theta)\}$ exist.

ASSUMPTION 1 (response smoothness).

a. $g(y,h,\theta)$ *is twice continuously differentiable on* $\mathbb{R} \times \mathbb{R}_+ \times \Theta$.

b. $g \leq \rho h + K(1+y^2)$ *for some* $\rho \in (0,1)$ *and* $\inf_{y\in\mathbb{R},h\in\mathbb{R}_+,\theta\in\Theta,} |g| =: \underline{g} > 0$;

c. $||g_a||$ *and* $||g_{a,b}||$ *are bounded by* $K(1+y^2+h)$ *for each* $a,b \in \{y,\theta\}$;

d. g, g_a *and* $g_{a,b}$ *are Lipschitz in* h, *for each* $a,b \in \{y,h,\theta\}$.

Conventional M-estimator asymptotics for semi-strong GARCH follows from stationarity, ergodicity and the existence of higher moments for ϵ_t since then a martingale difference central limit theorem applies (e.g. Straumann and Mikosch 2006, MS 2009). Tail trimming by m_t precludes a martingale difference component even if ϵ_t is i.i.d., hence ergodicity and martingale difference arguments do not suffice, although a mixing CLT does. Further, we require a uniform central limit theorem for trimming indicators $I^*_{n,t}(\theta)$ which forces us to assume $h^*_t(\theta)$ is mixing for any $\theta \in \Theta$. A particularly elegant UCLT by Doukhan et al (1995) requires only β-mixing.

ASSUMPTION 2 (mixing). *Let* $\{y_t, \epsilon_t, h^*_t(\theta)\}$ *be for each* $\theta \in \Theta$ *geometrically* β*-mixing.*

Let $\kappa_w(\theta) \in \{\kappa_{m_i}(\theta), \kappa_{G_{i,j}}(\theta)\}$ be the moment suprema of $w_t(\theta) \in \{m^*_{i,t}(\theta), G^*_{i,j,t}(\theta)\}$: $\kappa_w(\theta) := \arg\sup_{\alpha\geq 0}$ $\{E|w_t(\theta)|^\alpha < \infty\}$, and write $\kappa_w = \kappa_w(\theta^0)$. Notice $\kappa_w(\theta) = \infty$ is possible, for example under exponential tail decay or a bounded support.

ASSUMPTION 3 (distribution).

a. Each y_t, $h^*_t(\theta)$, $m^*_{i,t}(\theta)$, and $G^*_{i,j,t}(\theta)$ *have absolutely continuous finite dimensional distributions, and each* $w_t(\theta) \in \{m^*_{i,t}(\theta), G^*_{i,j,t}(\theta)\}$ *has a uniformly bounded distribution:* $\sup_{\theta\in\Theta}\sup_{a\in\mathbb{R}}\{(\partial/\partial a)P(w_t(\theta) \leq a)\} < \infty$ *and* $\sup_{\theta\in\Theta}\sup_{a\in\mathbb{R}}\{(\partial/\partial\theta)P(w_t(\theta) \leq a)\} < \infty$.

b. Each $w_t(\theta) \in \{m^*_{i,t}(\theta), G^*_{i,j,t}(\theta)\}$ *has a bounded envelope* $E[\sup_{\theta\in\Theta}|w_t(\theta)|^\iota] < \infty$ *for tiny* $\iota > 0$. *Further* $\inf_{\theta\in\Theta}\kappa_w(\theta) > 0$, *and* $\kappa_{m_i} > 1$. *If* $\kappa_{m_i}(\theta) \leq 2$ *or* $\kappa_{G_{i,j}}(\theta) \leq 1$ *then the corresponding* $w_t(\theta) \in \{m^*_{i,t}(\theta), G^*_{i,j,t}(\theta)\}$ *has a power-law tail* $P(|w_t(\theta)| > c)\} = d_w(\theta)c^{-\kappa_w(\theta)}(1+o(1))$. *Define* $\bar{\kappa}_m := 2$ *and* $\bar{\kappa}_G := 1$. *In particular*

$$\sup_{\theta\in\Theta:\kappa_w(\theta)\ \leq\ \bar{\kappa}_w} \left\{\left|c^{\kappa_w(\theta)}P\left(|w_t(\theta)| > c\right) - d_w(\theta)\right|\right\} \to 0 \text{ as } c \to \infty \text{ and } \inf_{\theta\in\Theta:\kappa_w(\theta)\ \leq\ \bar{\kappa}_w}\{d_w(\theta)\} > 0. \tag{5}$$

Remark: Under distribution continuity of the loss $l_t^*(\theta)$ a unique solution $\hat{\theta}_n$ exists with probability one, and thresholds $\{\mathcal{L}_{i,n}^{(m)}(\theta), \mathcal{U}_{i,n}^{(m)}(\theta), \mathcal{C}_{i,j,n}^{(G)}(\theta)\}$ exist for all $\theta \in \Theta$ and any fractiles $\{k_n\}$. Uniform boundedness of the distributions simplifies the verification that $\{I_{i,n,t}^{(m^*)}(\theta), I_{i,j,n,t}^{(G^*)}(\theta)\}$ satisfy a UCLT. Heavy tailed $m_{i,t}^*(\theta)$ and $G_{i,j,t}^*(\theta)$ have power-law tails to ease characterizing trimmed moments in lieu of Karamata's Theorem.

Consistency requires moment smoothness, and we must rule out degeneracy due to trimming.

ASSUMPTION 4 (smoothness and non-degeneracy).

a. $\liminf_{n\rightarrow\infty} \inf_{||\theta-\theta^0||>\delta}\{||E[m_{n,t}^*(\theta)]||/\sup_{\theta\in\Theta}||E[m_{n,t}^*(\theta)]||\} > 0$ *for tiny* $\delta > 0$.

b. $\mathcal{A}_n(\theta) \in \{E|m_{i,n,t}^*(\theta)|^p, \mathcal{S}_n(\theta), \mathcal{G}_n(\theta)\}$ *for any* $p > 0$ *satisfy* $\liminf_{n\rightarrow\infty}\inf_{\theta\in\Theta}\{||\mathcal{A}_n(\theta)||\} > 0$. *Further* $\liminf_{n\rightarrow\infty}\inf_{\theta\in\Theta}\{\lambda_{\min}(\mathcal{S}_n(\theta))\} > 0$ *and* $\sup_{\theta\in\Theta}||E[m_{n,t}^*(\theta)m_{n,t}^{*\prime}(\theta)]\mathcal{S}_n^{-1}(\theta)|| = O(1)$.

Remark 1: Property (a) bounds $E[m_{n,t}^*(\theta)]$ from 0 for θ near θ^0. The scale $\sup_{\theta\in\Theta}||E[m_{n,t}^*(\theta)]||$ ensures a uniform law of large numbers applies to $m_{n,t}^*(\theta)$ when $m_t(\theta)$ is non-integrable, which expedites our proof of consistency.

Remark 2: We prove in Appendix B $||\mathcal{S}_n(E[m_{n,t}^*m_{n,t}^{*\prime}])^{-1}|| = O(\ln(n))$ although if each $E[m_{i,t}^2] < \infty$ then by Assumption 2 $\mathcal{S}_n \sim E[m_{n,t}^*m_{n,t}^{*\prime}]$. We therefore require the converse $||E[m_{n,t}^*m_{n,t}^{*\prime}]\mathcal{S}_n^{-1}|| = O(1)$ under (b) for a central limit theorem when $E[m_{i,t}^2] = \infty$, a standard condition that rules out degeneracy (see Dehling et al 1986 and their references).

In the appendices we show $\hat{\theta}_n$ obtains the expansion $\mathcal{V}_n^{1/2}(\hat{\theta}_n - \theta^0) \sim n^{-1/2}\mathcal{S}_n^{-1/2}\sum_{t=1}^n m_{n,t}^*$, where $\mathcal{V}_n = n\mathcal{G}_n\mathcal{S}_n^{-1}\mathcal{G}_n$, hence $n^{1/2}\mathcal{S}_n^{-1/2}E[m_{n,t}^*] \rightarrow 0$ must hold for asymptotic unbiasedness.

ASSUMPTION 5 (identification). $E[m_t^*(\theta)] = E[(\epsilon_t^2(\theta) - 1)\mathfrak{s}_t^*(\theta)] = 0$ if and only if $\theta = \theta^0$, *a unique interior point of compact* $\Theta \subset \mathbb{R}^q$. *In particular the fractile sequences* $\{k_{j,\epsilon,n}, k_{j,i,n}\}$ *satisfy* $||n^{1/2}\mathcal{S}_n^{-1/2}E[m_{n,t}^*]|| \rightarrow 0$.

Remark: We do not require $E[m_{n,t}^*] = 0$ for finite n since our results are asymptotic, while $E[m_{n,t}^*] \rightarrow E[m_t^*] = 0$ by dominated convergence and the fact that $\{k_{j,\epsilon,n}, k_{j,i,n}\}$ are intermediate order sequences. Notice at least $E[\epsilon_t^2] = 1$ must hold for $E[m_{n,t}^*] \rightarrow 0$. In the strong-GARCH case we only need $E[\epsilon_t^2] = 1$, but if the errors are not i.i.d. then $(\epsilon_t^2 - 1)\mathfrak{s}_t^*$ must be integrable. The latter reduces to $E[\epsilon_t^2] = 1$ in some cases, e.g. a semi-strong ARCH since $||\mathfrak{s}_t^*|| \leq K$ *a.s.*

2.3 MAIN RESULTS

In the Appendices we show $\hat{\theta}_n - \theta^0$ is approximated by $n^{-1/2}\mathcal{S}_n^{-1/2}\sum_{t=1}^n\{m_{n,t}^* - E[m_{n,t}^*]\}$ in order to prove asymptotic normality. This in turn requires Jacobian consistency $1/n\sum_{t=1}^n G_t(\hat{\theta}_n)\hat{I}_{n,t}(\hat{\theta}_n) = \mathcal{G}_n \times (1 + o_p(1))$ hence $\hat{\theta}_n \xrightarrow{p} \theta^0$ from first principles.

THEOREM 2.1 (*QMTTL consistency*). *Under Assumptions 1-5* $\hat{\theta}_n \xrightarrow{p} \theta^0$.

Remark: Consistency requires all assumptions, and in general more assumptions than imposed in Straumann and Mikosch (2006) and MS (2009). This is due to nonlinearities induced by trimming: we require a first order condition under tail-trimming and response differentiability allows for almost sure criterion differentiability at $\hat{\theta}_n$; we require a UCLT to relate $\hat{I}_{n,t}(\theta)$ to $I_{n,t}(\theta)$ in the first order equations, and a ULLN on $m_{n,t}^*(\theta)$ to prove consistency, and both rely on mixing and smoothness properties. It does not appear to be possible to prove consistency under weaker conditions.

THEOREM 2.2 (*QMTTL normality*). *Under Assumptions 1-5* $\mathcal{V}_n^{1/2}(\hat{\theta}_n - \theta^0) \xrightarrow{d} N(0, I_q)$ *where* $\mathcal{V}_n := n\mathcal{G}_n'\mathcal{S}_n^{-1}\mathcal{G}_n$.

Remark 1: Although $\mathcal{V}_n = n\mathcal{G}_n'\mathcal{S}_n^{-1}\mathcal{G}_n$ we cannot conclude the rate is $n^{1/2}$ since if $E[\epsilon_t^4] = \infty$ then $||\mathcal{S}_n|| \to \infty$ and $||\mathcal{G}_n|| \to \infty$ are possible. This implies heterogeneous rates $\mathcal{V}_{i,i,n}^{1/2} \to \infty$ below, at, or above $n^{1/2}$ are possible depending on error-regressor feedback. In many models, however, $\limsup_{n\to\infty} ||\mathcal{G}_n|| < \infty$ if there are GARCH effects (see Francq and Zakoïan 2010), in which case $\mathcal{V}_{i,i,n}^{1/2} = o(n^{1/2})$ if the errors have an infinite fourth moment. See Section 4.

Remark 2: If each $E[(\epsilon_t^2 - 1)^2\mathfrak{s}_{i,t}^2] < \infty$ and $E|G_{i,j,t}| < \infty$ then by geometric β-mixing and dominated convergence

$$\mathcal{G}_n'\mathcal{S}_n^{-1}\mathcal{G}_n \to E[G_t]'\left(E\left[m_t m_t'\right] + 2\sum_{i=1}^{\infty} E\left[m_1 m_{i+1}'\right]\right)^{-1} E[G_t] =: \mathcal{G}'\mathcal{S}^{-1}\mathcal{G} =: \mathcal{V},$$

where $||\mathcal{S}|| < \infty$. The inverse $\mathcal{V}^{-1}$ is the classic asymptotic covariance matrix, hence trimming does not impact efficiency asymptotically under thin tails: $n^{1/2}(\hat{\theta}_n - \theta^0) \xrightarrow{d} N(0, \mathcal{V}^{-1})$. Thus QMTTL is asymptotically equivalent to QML under thin tails (cf. Straumann and Mikosch 2006, Escanciano 2009).

In the case of heavy tails it is impossible to identify unique "rates of convergence" *and* an "asymptotic covariance" from $\mathcal{V}_n$ without first specifying error dependence and trimming fractiles. We treat this topic in the following section where focused trimming substantially simplifies the analysis.

3. FOCUSED TRIMMING

In view of the score and Jacobian equation structures (3) we can always focus trimming on the source of extremes $\epsilon_t(\theta)$, $\mathfrak{s}_{i,t}(\theta)$, and $\mathfrak{s}^{\theta}_{i,j,t}(\theta)$. Depending on the functional form of g this may be laborious, hence the QMTTL method above may be preferred. In many cases, however, focused trimming is exceptionally simple. Further, if ϵ_t is i.i.d. then we can decompose the scale $\mathcal{V}_n$ into components representing an asymptotic covariance matrix and rates of convergence.

3.1 Focused Trimming

Define $\mathfrak{s}^{(a)}_{i,t}(\theta) := |\mathfrak{s}_{i,t}(\theta)|$ and $\mathfrak{s}^{\theta(a)}_{i,j,t}(\theta) := |\mathfrak{s}^{\theta}_{i,j,t}(\theta)|$, and indicators

$$\hat{I}^{(\epsilon)}_{n,t}(\theta) := I\left(\epsilon^{(-)}_{(k^{(\epsilon)}_{1,n})}(\theta) \leq \epsilon_t(\theta) \leq \epsilon^{(+)}_{(k^{(\epsilon)}_{2,n})}(\theta)\right) \text{ and } \hat{I}^{(\mathfrak{s})}_{i,n,t}(\theta) := I\left(|\mathfrak{s}_{i,t}(\theta)| \leq \mathfrak{s}^{(a)}_{i,(k^{(\mathfrak{s})}_{i,n})}(\theta)\right)$$
$$\hat{I}^{(\mathfrak{s}^{\theta})}_{i,j,n,t}(\theta) := I\left(\left|\mathfrak{s}^{\theta}_{i,j,t}(\theta)\right| \leq \mathfrak{s}^{\theta(a)}_{i,j,(k^{(\mathfrak{s}^{\theta})}_{i,j,n})}(\theta)\right),$$

where $\{k^{(\epsilon)}_{1,n}, k^{(\epsilon)}_{2,n}, k^{(\mathfrak{s})}_{i,n}, k^{(\mathfrak{s}^{\theta})}_{i,j,n}\}$ are intermediate order sequences. The Quasi-Maximum Focused Tail-Trimmed Likelihood estimator [QMFTTLE] solves

$$\hat{\theta}_n = \underset{\theta \in \Theta}{\operatorname{argmin}} \left\{ \sum_{t=1}^{n} \left\{\ln h_t(\theta) + \epsilon^2_t(\theta)\right\} \hat{I}^{(\epsilon)}_{n,t}(\theta) \times \hat{I}^{(\mathfrak{s})}_{n,t}(\theta) \times \hat{I}^{(\mathfrak{s}^{\theta})}_{n,t}(\theta) \right\}.$$

Since the trimming mechanism is different relative to Section 2 we use different notation for the trimmed equations with non-random thresholds. Write $\mathfrak{s}^{*}_t(\theta)$ and $\mathfrak{s}^{*\theta}_t(\theta)$ to denote $\mathfrak{s}_t(\theta)$ and $\mathfrak{s}^{\theta}_t(\theta)$ computed with the stationary solutions $\{h^{*}_t(\theta), h^{*\theta}_t(\theta), h^{*\theta,\theta}_t(\theta)\}$, define thresholds by

$$P\left(\epsilon_t(\theta) \leq -\mathcal{L}^{(\epsilon)}_n(\theta)\right) = \frac{k^{(\epsilon)}_{1,n}}{n} \text{ and } P\left(\epsilon_t(\theta) \geq \mathcal{U}^{(\epsilon)}_n(\theta)\right) = \frac{k^{(\epsilon)}_{2,n}}{n}$$
$$P\left(\left|\mathfrak{s}^{*}_{i,t}\right| \leq \mathcal{C}^{(\mathfrak{s})}_{i,n}(\theta)\right) = \frac{k^{(\mathfrak{s})}_{i,n}}{n} \text{ and } P\left(\left|\mathfrak{s}^{*\theta}_{i,j,t}\right| \leq \mathcal{C}^{(\mathfrak{s}^{\theta})}_{i,j,n}(\theta)\right) = \frac{k^{(\mathfrak{s}^{\theta})}_{i,j,n}}{n},$$

and indicators

$$I_{n,t}^{(\epsilon)}(\theta) := I\left(-\mathcal{L}_n^{(\epsilon)}(\theta) \leq \epsilon_t(\theta) \leq \mathcal{U}_n^{(\epsilon)}(\theta)\right)$$
$$I_{i,n,t}^{(\mathfrak{s})}(\theta) := I\left(\left|\mathfrak{s}_{i,t}^*(\theta)\right| \leq \mathcal{C}_{i,n}^{(\mathfrak{s})}(\theta)\right) \quad \text{and} \quad I_{i,j,n,t}^{(\mathfrak{s}^\theta)}(\theta) := I\left(\left|\mathfrak{s}_{i,j,t}^{*\theta}(\theta)\right| \leq \mathcal{C}_{i,j,n}^{(\mathfrak{s})}(\theta)\right)$$
$$I_{n,t}^{(\mathfrak{s})}(\theta) = \prod_{i=1}^{q} I_{i,n,t}^{(\mathfrak{s})}(\theta) \quad \text{and} \quad I_{n,t}^{((\mathfrak{s}^\theta)}(\theta) = \prod_{1\leq i\leq j\leq q} I_{i,j,n,t}^{(\mathfrak{s}^\theta)}(\theta),$$

and defined trimmed versions $\mathfrak{s}_{n,t}^*(\theta) := \mathfrak{s}_t^*(\theta)\, I_{n,t}^{(\mathfrak{s})}(\theta)$ and $\mathfrak{s}_{n,t}^{*\theta}(\theta) := \mathfrak{s}_t^{*\theta}(\theta)\, I_{n,t}^{(\mathfrak{s}^\theta)}(\theta)$. Finally, the tail-trimmed score equations, and long-run covariance and Jacobian matrices are

$$\mathring{m}_{n,t}^*(\theta) = \left(\epsilon_t^2(\theta) - 1\right) I_{n,t}^{(\epsilon)}(\theta) \times \mathfrak{s}_{n,t}^*(\theta) \times I_{n,t}^{(\mathfrak{s}^\theta)}(\theta) \quad \text{and} \quad \mathring{\mathcal{S}}_n(\theta) := \frac{1}{n}\sum_{s,t=1}^{n} E\left[\mathring{m}_{n,s}^*(\theta)\, \mathring{m}_{n,t}^*(\theta)'\right] \tag{6}$$
$$\mathring{\mathcal{G}}_n(\theta) := E\left[\left(\epsilon_t^2(\theta) - 1\right) I_{n,t}^{(\epsilon)}(\theta) \times \mathfrak{s}_{n,t}^{*\theta}(\theta) \times I_{n,t}^{(\mathfrak{s})}(\theta)\right] - E\left[\epsilon_{n,t}^{*2}(\theta)\, \mathfrak{s}_{n,t}^*(\theta)\, \mathfrak{s}_{n,t}^{*\prime}(\theta) \times I_{n,t}^{(\mathfrak{s}^\theta)}(\theta)\right]'.$$

Notice $\mathring{\mathcal{G}}_n \sim -E[\mathfrak{s}_{n,t}^*\mathfrak{s}_{n,t}^{*\prime} I_{n,t}^{(\mathfrak{s}^\theta)}]$ may not hold, unless ϵ_t is i.i.d., or a martingale difference with $E[\epsilon_t^4] < \infty$ and each $E[(\mathfrak{s}_{i,j,t}^{*\theta})^2] < \infty$.

THEOREM 3.1 (*QMFTTL*). *Let Assumptions 1-3 hold, and let Assumptions 4 and 5 apply to* $\mathring{m}_{n,t}^*(\theta)$ *and* $\mathring{\mathcal{G}}_n(\theta)$. *Then* $\mathring{\mathcal{V}}_n^{1/2}(\hat{\theta}_n - \theta^0) \xrightarrow{d} N(0, I_q)$, *where* $\mathring{\mathcal{V}}_n = n\mathring{\mathcal{G}}_n'\mathring{\mathcal{S}}_n^{-1}\mathring{\mathcal{G}}_n$.

Now assume $\epsilon_t \overset{i.i.d.}{\sim} (0,1)$ such that $E[\mathring{m}_{n,t}^*] = E[(\epsilon_t^2 - 1)I_{n,t}^{(\epsilon)}] \times E[\mathfrak{s}_{n,t}^* I_{n,t}^{(\mathfrak{s}^\theta)}]$. Notice each $\mathfrak{s}_{i,n,t}^* I_{n,t}^{(\mathfrak{s}^\theta)} = \mathfrak{s}_{i,t}^* I_{n,t}^{(\mathfrak{s})} I_{n,t}^{(\mathfrak{s}^\theta)}$ is trimmed by the same compound $I_{n,t}^{(\mathfrak{s})} I_{n,t}^{(\mathfrak{s}^\theta)}$, but negligibility $I_{j,n,t}^{(\mathfrak{s})} I_{n,t}^{(\mathfrak{s}^\theta)} \rightarrow 1$ *a.s.* implies by dominated convergence

$$E\left[\mathfrak{s}_{i,t}^{*2} I_{n,t}^{(\mathfrak{s})} I_{n,t}^{(\mathfrak{s}^\theta)}\right] = E\left[\mathfrak{s}_{i,t}^{*2} I_{i,n,t}^{(\mathfrak{s})}\right] \times \left(1 - \frac{E\left[\mathfrak{s}_{i,t}^{*2} I_{i,n,t}^{(\mathfrak{s})}\left(1 - \prod_{j\neq i} I_{j,n,t}^{(\mathfrak{s})} I_{n,t}^{(\mathfrak{s}^\theta)}\right)\right]}{E\left[\mathfrak{s}_{i,t}^{*2} I_{i,n,t}^{(\mathfrak{s})}\right]}\right) \sim E\left[\mathfrak{s}_{i,t}^{*2} I_{i,n,t}^{(\mathfrak{s})}\right]. \tag{7}$$

By the same argument $E[\mathfrak{s}_{i,j,t}^{*\theta} I_{n,t}^{(\mathfrak{s}^\theta)} I_{n,t}^{(\mathfrak{s})}] \sim E[\mathfrak{s}_{i,j,t}^{*\theta} I_{i,j,n,t}^{(\mathfrak{s}^\theta)}]$.

In conjunction with (7) if we strengthen Assumption 5 then $\mathring{\mathcal{S}}_n$ and $\mathring{\mathcal{G}}_n$ reduce to simple forms.

ASSUMPTION 5' (identification). *Define*

$$\mathcal{A}_n := \frac{\left(E\left[\left(\epsilon_t^2 - 1\right)^2 I_{n,t}^{(\epsilon)}\right]\right)^{1/2} \left\|\left\{E\left[\mathfrak{s}_{i,t}^* I_{i,n,t}^{(\mathfrak{s})} \mathfrak{s}_{j,t}^* I_{j,n,t}^{(\mathfrak{s})}\right]\right\}_{i,j=1}^{q}\right\|^{1/2}}{n^{1/2}\left\|\left\{E\left[\mathfrak{s}_{i,t}^* I_{i,n,t}^{(\mathfrak{s})}\right]\right\}_{i=1}^{q}\right\|}, \quad \mathcal{B}_n := \frac{1}{\left\|\left\{E\left[\mathfrak{s}_{i,j,t}^{*\theta} I_{i,j,n,t}^{(\mathfrak{s}^\theta)}\right]\right\}_{i,j=1}^{q}\right\|}$$

and assume the fractile sequences $\{k_{j,\epsilon,n}, k_{j,i,n}\}$ *satisfy*

$$E\left[\left(\epsilon_t^2 - 1\right) I_{n,t}^{(\epsilon)}\right] = o\left(\min\left\{\mathcal{A}_n, \mathcal{B}_n\right\}\right). \tag{8}$$

The first component $\mathcal{A}_n$ of minimand (8) identically covers Assumption 5 in the i.i.d. case. Along with (7) and $E[\epsilon_{n,t}^{*2}] \sim 1$, by dominated convergence we have both $\mathring{\mathcal{S}}_n \sim (E[\epsilon_{n,t}^{*4}]-1) \times \{E[\mathfrak{s}_{i,t}^* I_{i,n,t}^{(\mathfrak{s})} \mathfrak{s}_{j,t}^* I_{j,n,t}^{(\mathfrak{s})}]\}_{i,j=1}^q$ and $||n^{1/2}\mathring{\mathcal{S}}_n^{-1/2} E[\mathring{m}_{n,t}^*]|| \to 0$. The second component $\mathcal{B}_n$ with property (7) implies $\mathring{\mathcal{G}}_n \sim \{E[\mathfrak{s}_{i,t}^* I_{i,n,t}^{(\mathfrak{s})} \mathfrak{s}_{j,t}^* I_{j,n,t}^{(\mathfrak{s})}]\}_{i,j=1}^q$. Combined with Theorem 3.1 this proves the following.

COROLLARY 3.2. *Let* $\epsilon_t \overset{i.i.d.}{\sim} (0,1)$, *let Assumptions 1-3 hold, and let Assumptions 4 and 5′ apply to* $\mathring{m}_{n,t}^*(\theta)$ *and* $\mathring{\mathcal{G}}_n(\theta)$. *Then* $\mathring{\mathcal{V}}_n \sim n(E[\epsilon_{n,t}^{*4}] - 1)^{-1}[E(\mathfrak{s}_{i,t}^* I_{i,n,t}^{(\mathfrak{s})} \mathfrak{s}_{j,t}^* I_{j,n,t}^{(\mathfrak{s})})]_{i,j=1}^q$, *in particular*

$$n^{1/2}\frac{1}{\left(E\left[\epsilon_{n,t}^{*4}\right]-1\right)^{1/2}} \times \left(\left\{E\left[\mathfrak{s}_{i,t}^* I_{i,n,t}^{(\mathfrak{s})} \mathfrak{s}_{j,t}^* I_{j,n,t}^{(\mathfrak{s})}\right]\right\}_{i,j=1}^q\right)^{1/2} \left(\hat{\theta}_n - \theta^0\right) \xrightarrow{d} N(0, I_q). \tag{9}$$

3.2 Examples

We present several models that satisfy the major assumptions. Assume the following error properties throughout for simplicity: $\epsilon_t \overset{i.i.d.}{\sim} (0,1)$, ϵ_t has an absolutely continuous, symmetric distribution, and if $E[\epsilon_t^4] = \infty$ then ϵ_t has tail

$$P\left(|\epsilon_t| \geq \epsilon\right) = d_\epsilon \epsilon^{-\kappa_\epsilon}\left(1 + o(1)\right) \quad \text{where } d_\epsilon > 0 \text{ and } \kappa_\epsilon \in (2, 4]. \tag{10}$$

EXAMPLE 1 (Linear Strong-GARCH): The model is $y_t = \sigma_t \epsilon_t$ with

$$\sigma_t^2 = \omega^0 + \alpha^0 y_{t-1}^2 + \beta^0 \sigma_{t-1}^2, \text{ where } \omega^0 > 0, \alpha^0, \beta^0 \geq 0, \text{ and } E\left[\ln\left(\alpha^0 \epsilon_t^2 + \beta^0\right)\right] < 0. \tag{11}$$

Define $\theta = [\omega, \alpha, \beta]'$, and the iterated process $h_0(\theta) = \omega$ and $h_t(\theta) = \omega + \alpha y_{t-1}^2 + \beta h_{t-1}(\theta)$ for $t \geq 1$. In general $E[\mathfrak{s}_{i,t}^{2+\iota}] < \infty$ and $E[(\mathfrak{s}_{i,j,t}^\theta)^{2+\iota}] < \infty$ for tiny $\iota > 0$ if there are GARCH effects $\alpha^0 + \beta^0 > 0$ (e.g. Francq and Zakoïan 2004), in which case we do not need to trim by these elements: the QMFTTLE solves $\hat{\theta}_n = \text{argmin}_{\theta \in \Theta}\{\sum_{t=1}^n (\ln h_t(\theta) + \epsilon_t^2(\theta))\hat{I}_{n,t}^{(\epsilon)}(\theta)\}$.

If we only impose $\alpha^0 + \beta^0 \geq 0$ then we need to trim by ϵ_t *and* y_{t-1} and y_{t-2} because under $\alpha^0 + \beta^0 = 0$ the scaled derivatives $\mathfrak{s}_{i,t}$ and $\mathfrak{s}_{i,j,t}^\theta$ are constants or proportional to powers of y_{t-1} and y_{t-2}. In this case the indicators are $\hat{I}_{n,t}(\theta) = \hat{I}_{n,t}^{(\epsilon)}(\theta)\hat{I}_{n,t}^{(y)}$ where $\hat{I}_{n,t}^{(y)} = \hat{I}_{1,n,t}^{(y)}\hat{I}_{2,n,t}^{(y)}$ and $\hat{I}_{i,n,t}^{(y)} := I(|y_{t-i}| \leq y_{(k_n^{(y)})}^{(a)})$ with fractile

$k_n^{(y)} \to \infty$ and $k_n^{(y)}/n \to 0$. If we use $\hat{I}_{n,t}(\theta) = \hat{I}_{n,t}^{(\epsilon)}(\theta)\hat{I}_{n,t}^{(y)}$ and there *are* GARCH effects $\alpha^0 + \beta^0 > 0$ then by dominated convergence and error independence the indicator $\hat{I}_{n,t}^{(y)}$ has no impact on the scale asymptotically.

Define $I_{i,n,t}^{(y)} := I(|y_{t-i}| \leq \mathcal{C}_n^{(y)})$ and $I_{n,t}^{(y)} := I_{1,n,t}^{(y)} I_{2,n,t}^{(y)}$, where the thresholds $\mathcal{C}_n^{(y)}$ satisfy $P(|y_t| > \mathcal{C}_n^{(y)}) = k_n^{(y)}/n$. If we trim by ϵ_t, y_{t-1} and y_{t-2} then identification (8) becomes

$$E\left[\left(\epsilon_t^2 - 1\right) I_{n,t}^{(\epsilon)}\right] = o\left(\min\left\{\frac{\left(E\left[\left(\epsilon_t^2 - 1\right)^2 I_{n,t}^{(\epsilon)}\right]\right)^{1/2} \left\|E\left[\mathfrak{s}_t^* \mathfrak{s}_t^{*\prime} I_{n,t}^{(y)}\right]\right\|^{1/2}}{n^{1/2}\left\|E\left[\mathfrak{s}_t^* I_{n,t}^{(y)}\right]\right\|}, \frac{1}{\left\|E\left[\mathfrak{s}_t^{*\theta} I_{n,t}^{(y)}\right]\right\|}\right\}\right). \tag{12}$$

Notice if we assume $\alpha^0 + \beta^0 > 0$ then $E[\mathfrak{s}_t^*]$ and $E[\mathfrak{s}_t^{*\theta}]$ are finite, and along with the Cauchy-Schwartz inequality $|E[(\epsilon_t^2 - 1)I_{n,t}^{(\epsilon)}]| \leq (E[(\epsilon_t^2 - 1)^2 I_{n,t}^{(\epsilon)}])^{1/2}$ we only need $E[(\epsilon_t^2 - 1)I_{n,t}^{(\epsilon)}] = o(1/n^{1/2})$.

THEOREM 3.3 (*Strong-GARCH*). *Consider GARCH model (11) and let (12) hold. Then Assumptions 1-4 hold. Further:*

i. Assume $\alpha^0 + \beta^0 > 0$. *If* $\hat{I}_{n,t}(\theta) = \hat{I}_{n,t}^{(\epsilon)}(\theta)$ *or* $\hat{I}_{n,t}^{(\epsilon)}(\theta)\hat{I}_{n,t}^{(y)}$ *then* $\mathring{\mathcal{V}}_n \sim n(E[\epsilon_{n,t}^{*4}] - 1)^{-1} \times E[\mathfrak{s}_t^* \mathfrak{s}_t^{*\prime}]$.

ii. Assume $\alpha^0 + \beta^0 \geq 0$. *If* $\hat{I}_{n,t}(\theta) = \hat{I}_{n,t}^{(\epsilon)}(\theta)\hat{I}_{n,t}^{(y)}$ *then* $\mathring{\mathcal{V}}_n \sim n(E[\epsilon_t^4 I_{n,t}^{(\epsilon)}] - 1)^{-1} \times E[\mathfrak{s}_t^* \mathfrak{s}_t^{*\prime} I_{n,t}^{(y)}]$.

EXAMPLE 2 (Nonlinear-GARCH): The NGARCH model in Engle and Ng (1993) is $y_t = \sigma_t \epsilon_t$, $\sigma_t^2 = \theta_1^0 + \theta_2^0 \sigma_{t-1}^2 + \theta_3^0(\epsilon_{t-1} - \theta_4^0)^2 \sigma_{t-1}^2$, where $\theta_1^0 > 0$, and the remaining $\theta_i^0 > 0$ and $\theta_3^0 + \theta_2^0 + \theta_3^0(\theta_4^0)^2 < 1$. Then $\{y_t, \sigma_t^2\}$ are stationary and geometrically β-mixing (Carrasco and Chen 2002). Define the iterated process $h_0(\theta) = \theta_1$ and $h_t(\theta) = \theta_1 + \theta_2 h_{t-1}(\theta) + \theta_3(\epsilon_{t-1} - \theta_4)^2 h_{t-1}(\theta)$ for $t \geq 1$. It is easy to verify that $\mathfrak{s}_t$ is uniformly square integrable if there are GARCH effects, and otherwise $\mathfrak{s}_0 = [1, 0, 0, 0]'$ and $\mathfrak{s}_t = [1, \theta_1^0, y_{t-1}^2, 0]'/\theta_1^0$ for $t \geq 1$, hence the simple trimming format in Example 1 carries over.

EXAMPLE 3 (GJR-GARCH): The GJR model in Glosten et al (1993) is $y_t = \sigma_t \epsilon_t$, $\sigma_t^2 = \theta_1^0 + \theta_2^0 y_{t-1}^2 + \theta_3^0 y_{t-1}^2 I(y_{t-1} \leq 0) + \theta_4^0 \sigma_{t-1}^2$, $\theta_1^0 > 0$, and the remaining $\theta_i^0 > 0$. The iterated process is $h_0(\theta) = \theta_1^0$ and $h_t(\theta) = \theta_1 + \theta_2 y_{t-1}^2 + \theta_3 y_{t-1}^2 I(y_{t-1} \leq 0) + \theta_4 h_{t-1}(\theta)$ for $t \geq 1$. If $\theta_2^0 + \theta_3^0 E[y_{t-1}^2 I(y_{t-1} \leq 0)] + \theta_4^0 < 1$ then $\{y_t, \sigma_t^2\}$ are stationary and geometrically β-mixing (Carrasco and Chen 2002), and again the structure of $h_t(\theta)$ ensures the simple trimming format of Example 1 is applicable.

3.3 Convergence Rate for Strong GARCH

Consider the strong GARCH model (11) with $\alpha^0 + \beta^0 > 0$ for brevity. Qualitatively similar conclusions arise from NGARCH and GJR-GARCH models. Theorem 3.3 reveals the rate of convergence is $O(n^{1/2})$. If there are GARCH effects then $E[\mathfrak{s}_{i,t}^{*2}] < \infty$ and the scale is

$$\mathring{\mathcal{V}}_{i,i,n}^{1/2} \sim n^{1/2} \frac{1}{\left(E[\epsilon_{n,t}^{*4}] - 1\right)^{1/2}} \times \left(E\left[\mathfrak{s}_{i,t}^{*2}\right]\right)^{1/2} = O(n^{1/2}).$$

If $E[\epsilon_t^4] = \infty$ then under Paretian tail decay (10) with index $\kappa_\epsilon \in (2,4]$ it follows from Karamata's Theorem

$$\kappa_\epsilon = 4 \ : \ E[\epsilon_{n,t}^{*4}] \sim L(n) \to \infty \text{ is slowly varying}$$

$$\kappa_\epsilon \in (2,4) \ : \ E[\epsilon_{n,t}^{*4}] \sim \left(\frac{\kappa_\epsilon}{4-\kappa_\epsilon}\right)\left(\mathcal{C}_n^{(\epsilon)}\right)^4 \frac{k_n^{(\epsilon)}}{n} = \left(\frac{\kappa_\epsilon}{4-\kappa_\epsilon}\right) d_\epsilon^{4/\kappa_\epsilon} \left(\frac{n}{k_n^{(\epsilon)}}\right)^{4/\kappa_\epsilon - 1}.$$

Hence if $\kappa_\epsilon = 4$ then $\mathring{\mathcal{V}}_{i,i,n}^{1/2} \sim n^{1/2}/L(n)$, and

$$\text{if } \kappa_\epsilon \in (2,4) \text{ then } \mathring{\mathcal{V}}_{i,i,n}^{1/2} \sim \frac{n^{1/2}}{\left(n/k_n^{(\epsilon)}\right)^{2/\kappa_\epsilon - 1/2}} \left(\frac{4-\kappa_\epsilon}{\kappa_\epsilon}\right)^{1/2} d_\epsilon^{-2/\kappa_\epsilon} \left(E\left[\mathfrak{s}_{i,t}^{*2}\right]\right)^{1/2}.$$

There are several key observations. First, as long as $\kappa_\epsilon \in (2,4)$ then elevating intermediate order trimming arbitrarily close to central order trimming will optimize the convergence rate. In general this implies

$$k_n^{(\epsilon)} \sim n/L(n), \text{ for example } k_n^{(\epsilon)} \sim \lambda n/\ln(n) \text{ for } \lambda \in (0,1),$$

ensures $\mathring{\mathcal{V}}_{i,i,n}^{1/2} \sim n^{1/2}/L(n)$ for any $\kappa_\epsilon \in (2,4]$. Hall and Yao (2003) show the QML rate is $n^{1-2/\kappa_\epsilon}/L(n)$ $\leq n^{1/2}/L(n)$ for any $\kappa_\epsilon \in (2,4]$, with strict inequality if $\kappa_\epsilon < 4$. Hence, QMFTTL is asymptotically more efficient when $\kappa_\epsilon < 4$. The Log-LAD estimator in Peng and Yao (2003: Theorem 1), however, is $n^{1/2}$-convergent for model (11).

Second, if $\kappa_\epsilon < 4$ then

$$\frac{n^{1/2}}{\left(n/k_n^{(\epsilon)}\right)^{2/\kappa_\epsilon - 1/2}} \left(\hat{\theta}_n - \theta^0\right) \xrightarrow{d} N\left(0, \left(\frac{\kappa_\epsilon}{4-\kappa_\epsilon}\right) d_\epsilon^{4/\kappa_\epsilon} \left(E\left[\mathfrak{s}_t^* \mathfrak{s}_t^{*\prime}\right]\right)^{-1}\right) = N\left(0, \mathcal{V}\left(\kappa_\epsilon, d_\epsilon\right)\right),$$

say. For a given $E[\mathfrak{s}_t^* \mathfrak{s}_t^{*\prime}]$, a larger tail scale d_ϵ is associated with a larger variance $\mathcal{V}(\kappa_\epsilon, d_\epsilon)$, while faster tail decay leads to a lower variance $(\partial/\partial\kappa_\epsilon)(\kappa_\epsilon/(4-\kappa_\epsilon))d_\epsilon^{4/\kappa_\epsilon} < 0$ *if and only if* $d_\epsilon < e^{\kappa_\epsilon/(4-\kappa_\epsilon)}$.

Third, if $\kappa_\epsilon < 4$ and we use the parametric fractile form $k_n^{(\epsilon)} \sim \lambda n/\ln(n)$ for $\lambda \in (0,1]$ then

$$\frac{n^{1/2}}{(\ln(n))^{2/\kappa_\epsilon - 1/2}} \left(\hat{\theta}_n - \theta^0\right) \xrightarrow{d} N\left(0, \lambda^{-(2/\kappa_\epsilon - 1/2)} \left(\frac{\kappa_\epsilon}{4-\kappa_\epsilon}\right) d_\epsilon^{4/\kappa_\epsilon} \left(E\left[\mathfrak{s}_t^* \mathfrak{s}_t^{*\prime}\right]\right)^{-1}\right) = N\left(0, \mathcal{V}\left(\lambda, \kappa_\epsilon, d_\epsilon\right)\right).$$

We can always diminish the asymptotic variance $\mathcal{V}(\lambda, \kappa_\epsilon, d_\epsilon)$ by removing more extremes per sample in the sense $\lambda \nearrow 1$.

4. METHOD OF MOMENTS WITH RE-CENTERING

Our final estimator exploits focused tail trimmed equations imbedded in a method of moments criterion. This gives us the advantage of re-centering to remove bias if the errors are i.i.d., so let $\epsilon_t \overset{i.i.d.}{\sim} (0,1)$ for the sake of discussion. Define $\Im_t := \sigma(y_\tau : \tau \leq t)$.

Define trimmed and re-centered equations

$$\begin{aligned}\hat{m}^*_{n,t}(\theta) &= \left(\epsilon_t^2(\theta)\hat{I}^{(\epsilon)}_{n,t}(\theta)\hat{I}^{(\mathfrak{s})}_{n,t}(\theta) - \frac{1}{n}\sum_{t=1}^n \epsilon_t^2(\theta)\hat{I}^{(\epsilon)}_{n,t}(\theta)\hat{I}^{(\mathfrak{s})}_{n,t}(\theta)\right) \\ &\quad \times \left(\mathfrak{s}_t(\theta)\hat{I}^{(\mathfrak{s})}_{n,t}(\theta)\hat{I}^{(\mathfrak{s}^\theta)}_{n,t}(\theta) - \frac{1}{n}\sum_{t=1}^n \mathfrak{s}_t(\theta)\hat{I}^{(\mathfrak{s})}_{n,t}(\theta)\hat{I}^{(\mathfrak{s}^\theta)}_{n,t}(\theta)\right),\end{aligned}$$

and a Method of Focused Tail-Trimmed Moments [MFTTM] estimator

$$\hat{\theta}_n := \underset{\theta\in\Theta}{\text{argmin}} \left(\sum_{t=1}^n \hat{m}^*_{n,t}(\theta)\right)' \left(\sum_{t=1}^n \hat{m}^*_{n,t}(\theta)\right).$$

Any positive definite symmetric weight matrix $W \in \mathbb{Q}^{q\times q}$ leads the the same solution $\text{argmin}_{\theta\in\Theta} \sum_{t=1}^n \hat{m}^*_{n,t}(\theta)' \times W \times \sum_{t=1}^n \hat{m}^*_{n,t}(\theta)$. Similarly, any $\Im_{t-1}$-measurable vector $z_t(\theta) \in \mathbb{R}^q$, $q \geq 3$, can be used instead of $\mathfrak{s}_t(\theta)$ for a GMM estimator (Hansen 1982). The scaled volatility derivative $\mathfrak{s}_t(\theta) = h_t^\theta(\theta)/h_t(\theta)$, however, provides an analogue to QML and is uniformly square integrable in many cases, hence the simplifications for trimming $\mathfrak{s}_t(\theta)$ discussed in Section 3 apply.

Now define equations with non-random thresholds

$$\begin{aligned}\ddot{m}^*_{n,t}(\theta) &= \left(\epsilon_t^2(\theta)I^{(\epsilon)}_{n,t}(\theta)I^{(\mathfrak{s})}_{n,t}(\theta) - E\left[\epsilon_t^2(\theta)I^{(\epsilon)}_{n,t}(\theta)I^{(\mathfrak{s})}_{n,t}(\theta)\right]\right) \\ &\quad \times \left(\mathfrak{s}_t(\theta)I^{(\mathfrak{s})}_{n,t}(\theta)I^{(\mathfrak{s}^\theta)}_{n,t}(\theta) - E\left[\mathfrak{s}_t(\theta)I^{(\mathfrak{s})}_{n,t}(\theta)I^{(\mathfrak{s}^\theta)}_{n,t}(\theta)\right]\right).\end{aligned}$$

Since ϵ_t is i.i.d. and has a smooth distribution, and trimming is negligible, it follows for all $n \geq N$ and some large $N \in \mathbb{N}$

$$E\left[\ddot{m}^*_{n,t}(\theta)|\Im_{t-1}\right] = E\left[\ddot{m}^*_{n,t}(\theta)\right] = 0 \textit{ if and only if } \theta = \theta^0.$$

We also center the trimmed $\mathfrak{s}_t(\theta)$ solely to promote a simple scale form: by independence and arguments in Section 3.1 the scale is

$$\ddot{\mathcal{V}}_n = n \times \left(E\left[\epsilon_{n,t}^{*4}\right] - 1\right)^{-1} \times \left\{E\left[\left(\mathfrak{s}_{i,t}^{*} I_{i,n,t}^{(\mathfrak{s})} - E\left[\mathfrak{s}_{i,t}^{*} I_{i,n,t}^{(\mathfrak{s})}\right]\right)\left(\mathfrak{s}_{j,t}^{*\prime} I_{j,n,t}^{(\mathfrak{s})} - E\left[\mathfrak{s}_{j,t}^{*\prime} I_{j,n,t}^{(\mathfrak{s})}\right]\right)\right]\right\}_{i,j=1}^{q},$$

which is smaller than the QMFTTL scale: $\ddot{\mathcal{V}}_n \leq \mathring{\mathcal{V}}_n$. Thus MFTTM assures identification but at a cost of efficiency. If we do not re-center then MFTTM and QMFTTL are of course equivalent. That said, our simulations demonstrate the gain in small sample bias reduction by re-centering outweighs the loss of efficiency. See Section 6.

THEOREM 4.1 (*MFTTM*). *Under Assumptions 1-4 and* $\epsilon_t \overset{i.i.d.}{\sim} (0,1)$ *we have* $\ddot{\mathcal{V}}_n^{1/2}(\hat{\theta}_n - \theta^0) \overset{d}{\rightarrow} N(0, I_q)$.

5. INFERENCE A natural estimator of the QMTTM scale $\mathcal{V}_n = n\mathcal{G}_n'\mathcal{S}_n^{-1}\mathcal{G}_n$ is

$$\hat{\mathcal{V}}_n = \hat{\mathcal{V}}_n(\hat{\theta}_n) = n\widehat{\mathcal{G}}_n'(\hat{\theta}_n)\widehat{\mathcal{S}}_n^{-1}(\hat{\theta}_n)\widehat{\mathcal{G}}_n(\hat{\theta}_n) \text{ where } \widehat{\mathcal{G}}_n(\theta) := \frac{1}{n}\sum_{t=1}^{n} G_t(\theta)\hat{I}_{n,t}^{(G)}(\theta).$$

Unless the trimmed equations $m_{n,t}^{*}$ are uncorrelated, a convenient way to estimate the covariance $\mathcal{S}_n$ uses a kernel weight to ensure positive definiteness with probability one (Newey and West 1987): for some integrable kernel function $\mathcal{K}(x)$ and bandwidth $\gamma_n \rightarrow \infty$ as $n \rightarrow \infty$

$$\widehat{\mathcal{S}}_n(\theta) := \frac{1}{n}\sum_{s,t=1}^{n} \mathcal{K}\left((s-t)/\gamma_n\right)\hat{m}_{n,s}(\theta)\hat{m}_{n,t}(\theta)' \text{ where } \hat{m}_{n,t}(\theta) := m_t(\theta)\hat{I}_{n,t}(\theta).$$

Notice $\widehat{\mathcal{S}}_n(\theta)$ is trimmed by the composite $\hat{I}_{n,t}(\theta)$. We are only able to prove consistency $\widehat{\mathcal{S}}_n\mathcal{S}_n^{-1} \rightarrow I_q$ using a first order expansion of $m_t(\theta)$ that relies on trimming the Jacobian equations $G_t(\theta)$. In principle $\hat{I}_{n,t}(\theta)$ can be replaced with $\hat{I}_{n,t}^{(m)}(\theta)$ but we do not to prove it. Although our proof exploits arguments for kernel estimators in de Jong and Davidson (2000), tail-trimming undoubtedly extends to other covariance estimators (e.g. wavelet estimators in Hong and Lee 1999).

Define Fourier coefficients $\varpi(\xi) := (2\pi)^{-1}\int_{-\infty}^{\infty}\mathcal{K}(x)e^{i\xi x}dx < \infty$.

THEOREM 5.1. *Assume* $\mathcal{K}(\cdot)$ *is continuous at* 0 *and all but a finite number of points,* $\mathcal{K} : \mathbb{R} \rightarrow [-1,1]$, $\mathcal{K}(0) = 1$, $\mathcal{K}(x) = \mathcal{K}(-x)\ \forall x \in \mathbb{R}$, $\int_{-\infty}^{\infty}|\mathcal{K}(x)|dx < \infty$, *and* $\int_{-\infty}^{\infty}|\varpi(\xi)|d\xi < \infty\}$. *Let* $\sum_{s,t=1}^{n}|\mathcal{K}((s-t)/\gamma_n)|$

$= o(n^2)$, $\max_{1\leq s\leq n}\sum_{t=1}^{n}\mathcal{K}((s-t)/\gamma_n) = o(n)$ *and bandwidth* $\gamma_n = o(n)$. *Under Assumptions 1-5* $\hat{\mathcal{V}}_n = \mathcal{V}_n(1+o_p(1))$.

Remark 1: Applicable kernels include Bartlett, Parzen, Quadratic Spectral, Tukey-Hanning and others (de Jong and Davidson 2000).

Remark 2: Notice $\hat{\mathcal{V}}_n = \mathcal{V}_n(1+o_p(1))$ only reduces to $\hat{\mathcal{V}}_n = \mathcal{V}_n + o_p(1)$ in the finite variance case. The former still implies $\mathcal{V}_n^{-1}\hat{\mathcal{V}}_n \xrightarrow{p} I_q$, hence classic inference is available without knowing the true rate of convergence, nor even if trimming is required.

Arguments essentially identical to those used to prove Theorem 5.1 extend to estimators for QMFTTL and MFTTM, while error independence allows for simple covariance estimation. In the case of QMFTTL if we assume $\epsilon_t \overset{i.i.d.}{\sim} (0,1)$ and the additional identification condition (8) holds then we use

$$\widehat{\mathring{\mathcal{V}}}_n = n\left(\frac{1}{n}\sum_{t=1}^{n}\epsilon_t^4(\hat{\theta}_n)\hat{I}_{n,t}^{(\epsilon)}(\hat{\theta}_n) - 1\right)^{-1} \times \left(\left[\frac{1}{n}\sum_{t=1}^{n}\mathfrak{s}_{i,t}(\hat{\theta}_n)\hat{I}_{i,n,t}^{(\mathfrak{s})}(\hat{\theta}_n)\mathfrak{s}_{j,t}(\hat{\theta}_n)\hat{I}_{j,n,t}^{(\mathfrak{s})}(\hat{\theta}_n)\right]_{i,j=1}^{q}\right)$$

with initial values $h_0(\theta) = \theta_1$ and $h_0^{\theta}(\theta) = [1,0,0]'$. In the case of linear strong-GARCH we use

$$\widehat{\mathring{\mathcal{V}}}_n = n\left(\frac{1}{n}\sum_{t=1}^{n}\epsilon_t^4(\hat{\theta}_n)\hat{I}_{n,t}^{(\epsilon)}(\hat{\theta}_n) - 1\right)^{-1} \times \frac{1}{n}\sum_{t=1}^{n}\mathfrak{s}_t(\hat{\theta}_n)\mathfrak{s}_t'(\hat{\theta}_n)\hat{I}_{n,t}^{(y)}. \tag{13}$$

Now center $\widehat{\mathfrak{s}}_{n,i,t}(\theta) := \mathfrak{s}_{i,t}(\theta)\hat{I}_{i,n,t}^{(\mathfrak{s})}(\theta) - 1/n\sum_{t=1}^{n}\mathfrak{s}_{i,t}(\theta)\hat{I}_{i,n,t}^{(\mathfrak{s})}(\theta)$ to construct an estimator for MFTTM:

$$\widehat{\ddot{\mathcal{V}}}_n = n\left(\frac{1}{n}\sum_{t=1}^{n}\epsilon_t^4(\hat{\theta}_n)\hat{I}_{n,t}^{(\epsilon)}(\hat{\theta}_n) - 1\right)^{-1} \times \left[\frac{1}{n}\sum_{t=1}^{n}\widehat{\mathfrak{s}}_{n,i,t}(\hat{\theta}_n)\widehat{\mathfrak{s}}_{n,j,t}(\hat{\theta}_n)\right]_{i,j=1}^{q}. \tag{14}$$

THEOREM 5.2. *Let* $\epsilon_t \overset{i.i.d.}{\sim} (0,1)$ *and Assumptions 1-4 hold. Then* $\widehat{\ddot{\mathcal{V}}}_n = \mathcal{V}_n(1+o_p(1))$, *and if additionally Assumption 5 holds then* $\widehat{\mathring{\mathcal{V}}}_n = \mathcal{V}_n(1+o_p(1))$.

A Wald statistic naturally follows for a test of parameter restrictions $R(\theta^0) = 0$ where $R : \mathbb{R}^q \to \mathbb{R}^J$ and $J \geq 1$. Assume R is differentiable with a gradient $\mathcal{D}(\theta) = (\partial/\partial\theta)R(\theta)$ that is continuous, differentiable and has full column rank. The test statistic with the QMTTLE as a plug-in is

$$\mathcal{W}_n = R(\hat{\theta}_n)'\left(\mathcal{D}(\hat{\theta}_n)\hat{\mathcal{V}}_n^{-1}(\hat{\theta}_n)\mathcal{D}(\hat{\theta}_n)'\right)^{-1}R(\hat{\theta}_n).$$

Use Theorems 2.2 and 5.1 to deduce $\mathcal{W}_n \xrightarrow{d} \chi^2(J)$ under the null, and if $R(\theta^0) \neq 0$ then $\mathcal{W}_n \xrightarrow{p} \infty$.

Similarly, the proof of Theorem 2.1 shows the first order condition is $1/n\sum_{t=1}^{n} m_t(\hat{\theta}_n)I_{n,t}(\hat{\theta}_n) = 0$ *a.s.* which naturally suggests a score test. A Lagrange Multiplier statistic based on the tail-trimmed score $1/n\sum_{t=1}^{n} m_t(\hat{\theta}_n^c)I_{n,t}(\hat{\theta}_n^c)$ and constrained estimator $\hat{\theta}_n^c$ can also be constructed.

6. SIMULATION STUDY We now compare tail-trimmed QML estimators to QML and Log-LAD. Let P_κ denote a symmetric Pareto distribution: if ϵ_t is distributed P_κ then $P(\epsilon_t < -\epsilon) = P(\epsilon_t > \epsilon) = .5(1 + \epsilon)^{-\kappa}$ for $\epsilon > 0$. We draw $2n$ observations for $n \in \{100, 800, 2000\}$ from two GARCH(1,1) models $y_t = \sigma_t\epsilon_t$, we retain the last n observations, and repeat to generate 1000 samples $\{y_t\}_{t=1}^{n}$. The models are linear GARCH $\sigma_t^2 = .3 + .3y_{t-1}^2 + .6\sigma_{t-1}^2$ and GJR-GARCH $\sigma_t^2 = .3 + .2y_{t-1}^2 + .3y_{t-1}^2 I(y_{t-1} \leq 0) + .6\sigma_{t-1}^2$, with a starting value $\sigma_1^2 = .3$. The error ϵ_t is i.i.d. $N(0,1)$, or $P_{2.5}$ distributed and standardized such that $E[\epsilon_t^2] = 1$. Each process therefore has a power-law tail with index κ_y (Basrak et al 2002, Liu 2006), where by numerical computation we find for GARCH $\kappa_y \in \{1.5, 4.1\}$ and GJR-GARCH $\kappa_y \in \{2.2, 2.7\}$.[3]

We compute the focused trimmed estimators QMFTTL and MFTTM, conditional on the first observation. The criteria are QMFTTL: $\sum_{t=2}^{n} l_t(\theta)\hat{I}_{n,t}^{(\epsilon)}(\theta)\hat{I}_{n,t}^{(y)}$ where $\hat{I}_{n,t}^{(y)} = \hat{I}_{1,n,t}^{(y)}\hat{I}_{2,n,t}^{(y)}$, and MFTTM: $\sum_{t=1}^{n}\hat{m}_{n,t}^*(\theta)'\sum_{t=1}^{n}\hat{m}_{n,t}^*(\theta)$ where $\hat{m}_{n,t,i}^*(\theta) = (\epsilon_t^2(\theta)\hat{I}_{n,t}^{(\epsilon)}(\theta) - \hat{\epsilon}_n^{*2}(\theta)) \times (\mathfrak{s}_{i,t}(\theta)\hat{I}_{n,t}^{(y)}(\theta) - \widehat{\mathfrak{s}}_{i,n}^*(\theta))$, $\hat{\epsilon}_n^{*2}(\theta) := 1/n\sum_{t=2}^{n}\epsilon_t^2(\theta)\hat{I}_{n,t}^{(\epsilon)}(\theta)$, and $\widehat{\mathfrak{s}}_{i,n}^*(\theta) = 1/n\sum_{t=2}^{n}\mathfrak{s}_{i,t}(\theta)\hat{I}_{n,t}^{(y)}(\theta)$. In the case of QMFTTL we initialize $h_1(\theta) = \theta_1$, and for MFTTM we use $h_1(\theta) = \theta_1$ and $h_1^\theta(\theta) = [1, 0, 0]'$. The parameter space is $\Theta = [0, 1]^3$.

We use symmetric trimming with fractiles $k_n^{(\epsilon)} = \max\{1, [\lambda_\epsilon n/\ln(n)]\}$ and $k_n^{(y)} = \max\{1, [\lambda_y n/\ln(n)]\}$ where $(\lambda_\epsilon, \lambda_y) = (.05, .01)$. Based on simulations not reported here, if we minimize the bootstrapped mse over a two dimensional grid of $(\lambda_\epsilon, \lambda_y)$ we find values $\lambda_\epsilon \in [.01, .10]$ and $\lambda_y \in [.005, .05]$ lead to sharp results for the sample sizes used here.

Table 1 contains estimator bias, mse, and the Kolmogorov-Smirnov statistic scaled by its 5% critical value. In Table 2 we report t-tests of hypotheses $\theta_3^0 = .6$, $\theta_3^0 = .35$ and $\theta_3^0 = 0$, where the first is true.

Consider the linear GARCH case. Log-LAD performs poorly in small samples, although it is very sharp

[3] Both processes satisfy $P(|y_t| > c) = dc^{-\kappa}(1 + o(1))$ and $E|\epsilon_t^2\mathcal{I}_t(\alpha) + \beta|^{\kappa/2} = 1$, where $\mathcal{I}_t(\alpha)$ is identically α or $\alpha_1 + \alpha_2 I(y_t < 0)$ respectively for GARCH or GJR-GARCH. We draw $R = 10,000$ i.i.d. ϵ_t from $P_{2.5}$ or $N(0,1)$ and report $\hat{\kappa} = \arg\min_{\kappa\in\mathcal{K}} |1/R\sum_{t=1}^{R} |\alpha\epsilon_t^2 I_t + \beta|^{\kappa/2} - 1|$ where $\mathcal{K} = \{.001, .002, ..., 10\}$. See, e.g., Basrak et al (2002).

when $E[\epsilon_t^4] = \infty$ and $n \geq 800$, and is worst overall when $E[\epsilon_t^4] < \infty$. QMFTTL and MFTTM are resilient to heavy tails, in particular they are closest to normal. If the error is Gaussian then all estimators perform somewhat poorly for small n, a well known shortcoming of M-estimators for GARCH models (e.g. Straumann and Mikosch 2006, Ling 2007, Boudt and Croux 2010). Nevertheless, even in thin tailed cases tail-trimming appears to provide a non-negligible improvement in approximate normality for sample sizes $n \leq 800$.

Table 2 contains empirical size and power for t-tests in the GARCH case. The test statistic is $(\hat{\theta}_{n,3} - c)/s_3$ where s_3^2 is the variance of $\hat{\theta}_{n,3}$ across all simulated paths, and $c \in \{0, .35, .6\}$ where $c = .6$ is true. Empirical size and power are sharp for the tail-trimmed estimators, and sharpness improves with sample size and is lower when tails are heavier, both as expected.

In the GJR-GARCH case we draw essentially the same conclusions. Notice QML is substantially biased and non-normal, although both diminish with n.

Finally, re-centering after trimming in the MFTTM estimator leads to lower mean-squared-error in most cases. This estimator is less efficient than QMFTTL, suggesting a strong improvement in small sample bias.

7. EMPIRICAL APPLICATION We now estimate GARCH and GJR-GARCH models for financial returns data. We study daily log-returns for the London Stock Exchange-100 [FTSE], the NASDAQ composite index, and the Hang Seng Index [HSI] over the period is Jan. 1, 2008 - Dec. 31, 2010, representing 757, 757 and 756 daily observations respectively, net of market closures. We use log-returns $y_t = \ln(x_t/x_{t-1})$ where x_t is the daily open/close average.[4]

In order to justify the use of robust methods we first estimate the tail index κ for absolute returns $y_t^{(a)}$ $:= |y_t|$. The case for heavy tails can be made by a plot of the Hill (1975) two-tailed tail index estimator $\hat{\kappa}_n$ $= (1/\tilde{k}_n \sum_{i=1}^{\tilde{k}_n} \ln(y_{(i)}^{(a)}/y_{(\tilde{k}_n+1)}^{(a)}))^{-1}$ over fractiles $\tilde{k}_n \in \{5, 6, ..., 200\}$. As long as $\tilde{k}_n \to \infty$ and $\tilde{k}_n = o(n)$ it is known $\hat{\kappa}_n \xrightarrow{p} \kappa$ and $\tilde{k}_n^{1/2}(\hat{\kappa}_n^{-1} - \kappa^{-1}) \xrightarrow{d} N(0, v^2)$, $v^2 < \infty$, for a broad array of time series, including nonlinear AR-GARCH with hyperbolic or geometric memory. See Hill (2010, 2011b) for theory and references. Hill

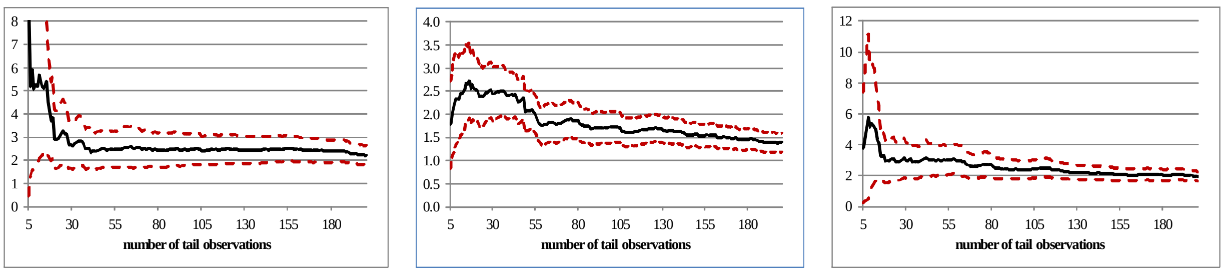
[4] The data were obtained from finance.yahoo.com, and the open/close average is computed using the reported adjusted close values.

(2010) presents a consistent kernel estimator $\hat{v}_n^2$ of the asymptotic variance v^2 of $\hat{\kappa}_n^{-1}$:

$$\hat{v}_n^2 = \frac{1}{n}\sum_{s,t=1}^{n} w_{n,s,t}\left\{\ln\left(y_s^{(a)}/y_{(\tilde{k}_n+1)}^{(a)}\right)_+ - \frac{\tilde{k}_n}{n}\hat{\kappa}_n^{-1}\right\} \times \left\{\ln\left(y_t^{(a)}/y_{(\tilde{k}_n+1)}^{(a)}\right)_+ - \frac{\tilde{k}_n}{n}\hat{\kappa}_n^{-1}\right\}$$

where $w_{n,s,t}$ is a kernel function. We use a Bartlett kernel $w_{n,s,t} = (1 - |s - t|/\gamma_n)_+$ with bandwidth $\gamma_n = n^{.25}$. By the mean-value-theorem the asymptotic 90% confidence band is $\hat{\kappa}_n \pm 1.64\hat{v}_n\hat{\kappa}_n^2/\tilde{k}_n^{1/2}$, which we plot in Figure 1. For each market tail index values of $\kappa \leq 2$ lie in the 90% intervals at nearly every $\tilde{k}_n$ and for the HSI $\kappa > 4$ lie strictly outside the bands.

Figure 1: Tail Index Plots with Kernel 90% Bands



FTSE-100 HSI NASDAQ

Next, we estimate GARCH models by QMFTTL and MFTTM. The latter estimators are computed exactly as in Section 6 with fractiles $k_n^{(\epsilon)} = [.05n/\ln(n)]$ and $k_n^{(y)} = [.2\ln(n)]$.

Since an important question concerns asymmetry in the data generating process, we test each series y_t and its GARCH residuals $\hat{\epsilon}_t$ for symmetry. Let z_t denote either y_t or ϵ_t. Define $\mathcal{I}_t(z) := 2I(z_t > 0) - 1$ and $\widehat{\mathcal{I}}_n(z) := 1/n\sum_{t=1}^n \mathcal{I}_t(z)$, where initial periods are truncated if z_t are residuals. A test of symmetry against positive skew is a test that $E[\mathcal{I}_t(z)] = P(z_t > 0) - P(z_t \leq 0) = 0$ against $E[\mathcal{I}_t] > 0$. We use a t-ratio $n^{1/2}\widehat{\mathcal{I}}_n(z)/\hat{s}_n(z)$ with kernel variance estimator $\hat{s}_n^2(z) = 1/n\sum_{s,t=1}^n \mathcal{K}((s-t)/b_n)\mathcal{I}_s(z)\mathcal{I}_t(z)$, Bartlett kernel $\mathcal{K}$ and bandwidth $b_n = n^{.25}$. Under stationary β-mixing and symmetry, and with a QMTTL plug-in in the case of residuals, $n^{1/2}\widehat{\mathcal{I}}_n(z)/\hat{s}_n \xrightarrow{d} N(0,1)$ is straightforward to prove.[5]

See Table 3 for estimation details where standard errors are computed using (13) for QMFTTL and (14) for MFTTM. In each case a linear GARCH model fits well, although NASDAQ returns y_t exhibit asymmetry. Only the FTSE noise ϵ_t appear asymmetric with a symmetry test p-value at .113. We then fit the returns

[5] Each symmetry test performed is robust to a range of bandwidths $b_n = n^{\zeta}$ for $\zeta \in [.15, .45]$.

series to GJR-GARCH models. In each case the resulting residuals appear symmetric and the likelihood function is closer to 1 than in the linear model, suggesting a reasonable model fit.

8. CONCLUSION We develop tail-trimmed QML and Method of Moments estimators for nonlinear GARCH models with possibly heavy tailed errors. In the latter case we introduce re-centering after trimming to correct for small sample bias induced by trimming. We show by Monte Carlo experiment QMFTTL and MFTTM dominate several existing estimators based on approximate normality and therefore inference, while simply choosing small fractile parameters leads to sharp results. The next stage must involve a complete theoretical development of fractile selection beginning with a formal treatment of the bootstrap mse method. Additional possibilities may incorporate indirect inference which requires an error distribution specification, or choosing the trimming parameter λ by testing the moment condition $E[m^{*}_{n,t}(\theta^0)] = 0$ with a consistent plug-in for θ^0 (e.g. Log-LAD). These must be left for future development.

APPENDIX A: Proofs of Main Results

We first prove stationary solutions $\{h^{*}_t(\theta), h^{*\theta}_{i,t}(\theta), h^{*\theta,\theta}_{i,t}(\theta)\}$ exist due to Lipschitz smoothness of g (cf. Carrasco and Chen 2002, Francq and Zakoïan 2006, MS 2009). Let $a_t(\theta) \in \{h_t(\theta), h^{\theta}_{i,t}(\theta), h^{\theta,\theta}_{i,j,t}(\theta)\}$ and $a^{*}_t(\theta) \in \{h^{*}_t(\theta), h^{*\theta}_{i,t}(\theta), h^{*\theta,\theta}_{i,j,t}(\theta)\}$ be arbitrary, let $I_{n,t}(\theta)$ denote $I^{*}_{n,t}(\theta)$ evaluated with $m_t(\theta)$ and $G_t(\theta)$, and let $w_t(\theta) \in \{m_{i,t}(\theta), G_{i,j,t}(\theta)\}$ and $w^{*}_t(\theta) \in \{m^{*}_{i,t}(\theta), G^{*}_{i,j,t}(\theta)\}$.[6]

PROPOSITION A.1 (*stationary solution*). *Let response smoothness Assumption 1 hold.*

i. A stationary and ergodic solution $a^{*}_t(\theta)$ exists for each $\theta \in \Theta$, it is $\Im_{t-1}$-measurable, and $\inf_{\theta\in\Theta} a^{*}_t(\theta) > 0$ a.s. Further, $h^{*}_t(\theta^0) = \sigma^2_t$ a.s., and $h^{*\theta}_t(\theta) = (\partial/\partial\theta)h^{*}_t(\theta)$ and $h^{*\theta,\theta}_t(\theta) = (\partial/\partial\theta)h^{*\theta}_t(\theta)$ a.s. at each θ;

ii. $E[\sup_{\theta\in\Theta} |a^{*}_t(\theta)|^{\iota}] < \infty$ for some tiny $\iota > 0$;

[6] In order to simplify notation we ignore measurability issues that arise when taking a supremum of a stochastic process. We implicitly assume all functions in this paper satisfy Pollard's (1984) permissibility criteria, the measure space that governs all random variables in this paper is complete, and therefore all majorants are measurable. Cf. Dudley (1978). Probability statements are therefore with respect to *outer probability*, and expectations over majorants are *outer expectations*.

iii. If $a_t(\theta)$ *is any other stationary solution then* $E[(\sup_{\theta\in\Theta}|a_t^*(\theta) - a_t(\theta)|)^\iota] = o(\rho^t)$ *for some* $\rho \in (0,1)$*;*

iv. $E[\sup_{\theta\in\Theta}|w_t^*(\theta) - w_t(\theta)|] = o(\rho^t)$.

PROOF. Claims (i)-(iii) follow from Propositions 1 and 2 of MS (2009) since their Assumptions DGP, C1-C4 and N1-N3 hold under Assumption 1 given stationarity and ergodicity of $\{y_t, \epsilon_t, \sigma_t^2\}$. Claim (iv) follows from response and moment envelope bounds Assumptions 1 and 3.b, and Claims (ii) and (iii). $\mathcal{QED}$.

Our proofs of consistency and asymptotic normality of the QMTTLE are for the infeasible estimator based on minimizing a trimmed $l_t^*(\theta)$ evaluated with $h_t^*(\theta)$. Let $\hat{I}_{n,t}^*(\theta)$ denote the composite trimming indicator constructed with $\{h_t^*(\theta), h_t^{*\theta}(\theta), h_t^{*\theta,\theta}(\theta)\}$, and define

$$\hat{\theta}_n^* = \underset{\theta\in\Theta}{\operatorname{argmin}} \left\{ \frac{1}{n}\sum_{t=1}^n l_t^*(\theta)\hat{I}_{n,t}^*(\theta) \right\}.$$

Similarly, define infeasible stochastically trimmed score equations, their sample Jacobian and HAC:

$$\hat{m}_{n,t}^*(\theta) := m_t^*(\theta)\hat{I}_{n,t}^*(\theta),\ m_{n,t}^*(\theta) := m_t^*(\theta)I_{n,t}^*(\theta),\ \hat{m}_n^*(\theta) := \frac{1}{n}\sum_{t=1}^n \hat{m}_{n,t}^*(\theta),\ m_n^*(\theta) := \frac{1}{n}\sum_{t=1}^n m_{n,t}^*(\theta)$$

$$\widehat{\mathcal{G}}_n^*(\theta) := \frac{1}{n}\sum_{t=1}^n G_t^*(\theta)\hat{I}_{n,t}^*(\theta) \quad \text{and} \quad \widehat{\mathcal{S}}_n^*(\theta) := \frac{1}{n}\sum_{s,t=1}^n \mathcal{K}\left((s-t)/\gamma_n\right)\hat{m}_{n,s}^*(\theta)\hat{m}_{n,t}^*(\theta)',$$

and corresponding variants with non-random thresholds:

$$\Sigma_n(\theta) := E\left[m_{n,t}^*(\theta)m_{n,t}^*(\theta)'\right] \quad \text{and } \mathcal{S}_n(\theta) := \frac{1}{n}\sum_{s,t=1}^n E\left[m_{n,s}^*(\theta)m_{n,t}^*(\theta)'\right]$$

$$G_{n,t}^*(\theta) := G_t^*(\theta)I_{n,t}^*(\theta),\ \ \mathcal{G}_n(\theta) := E\left[G_{n,t}^*(\theta)\right], \ \text{ and } \ \mathcal{V}_n(\theta) = n\mathcal{G}_n(\theta)'\mathcal{S}_n^{-1}(\theta)\mathcal{G}_n(\theta)$$

The following result implies the QMTTL estimator $\hat{\theta}_n$ and its infeasible counterpart $\hat{\theta}_n^*$ are asymptotically equivalent. In view of Proposition A.1 we therefore only need to consider $\hat{\theta}_n^*$, $\hat{m}_{n,t}^*(\theta)$, $\widehat{\mathcal{G}}_n^*(\theta)$ and $\widehat{\mathcal{S}}_n^*(\theta)$ in all that follows. The proof exploits supporting lemmata stated below.

PROPOSITION A.2 (*infeasible estimator*). *Under Assumptions 1-5* $\mathcal{V}_n^{1/2}(\hat{\theta}_n^* - \hat{\theta}_n) \xrightarrow{p} 0$ *where each* $\mathcal{V}_{i,i,n} \to \infty$.

PROOF. The following argument borrows from MS's (2009) proof of their Lemma D.6. By the proof of Theorem 2.1 the first order conditions are $\sum_{t=1}^n \hat{m}_{n,t}^*(\hat{\theta}_n^*) = 0$ *a.s.* and $\sum_{t=1}^n \hat{m}_{n,t}(\hat{\theta}_n) = 0$ *a.s.* Therefore, in lieu of consistency of the infeasible estimator $\hat{\theta}_n^* \xrightarrow{p} \theta^0$ by the proof of Theorem 2.1, the Lemma A.6 asymptotic expansion and Lemma A.8 Jacobian consistency, it follows

$$\begin{aligned}\frac{1}{n^{1/2}}\mathcal{S}_n^{-1/2}\sum_{t=1}^n\left\{\hat{m}_{n,t}(\hat{\theta}_n)-\hat{m}_{n,t}^*(\hat{\theta}_n)\right\} &= \frac{1}{n^{1/2}}\mathcal{S}_n^{-1/2}\sum_{t=1}^n\left\{\hat{m}_{n,t}^*(\hat{\theta}_n^*)-\hat{m}_{n,t}^*(\hat{\theta}_n)\right\} \\ &= n^{1/2}\mathcal{S}_n^{-1/2}\mathcal{G}_n\times\left(\hat{\theta}_n^*-\hat{\theta}_n\right).\end{aligned} \tag{15}$$

Jacobian consistency extends to $1/n\sum_{t=1}^n G_t(\hat{\theta}_n)\hat{I}_{n,t} = \mathcal{G}_n(1 + o_p(1))$ by invoking Proposition A.1 and Lemma A.8. Similarly, asymptotic approximation Lemma A.2.a and expansion Lemma A.6 extend to $\hat{m}_{n,t}$ and $m_{n,t}$. Therefore

$$\begin{aligned}&\frac{1}{n^{1/2}}\mathcal{S}_n^{-1/2}\sum_{t=1}^n\left\{\hat{m}_{n,t}(\hat{\theta}_n)-\hat{m}_{n,t}^*(\hat{\theta}_n)\right\} \\ &\quad=\frac{1}{n^{1/2}}\mathcal{S}_n^{-1/2}\sum_{t=1}^n m_{n,t}\left(1+o_p\left(1\right)\right)+n^{1/2}\mathcal{S}_n^{-1/2}\mathcal{G}_n\times\left(\hat{\theta}_n-\theta^0\right)-\frac{1}{n^{1/2}}\mathcal{S}_n^{-1/2}\sum_{t=1}^n\hat{m}_{n,t}^*(\hat{\theta}_n) \\ &\quad=\frac{1}{n^{1/2}}\mathcal{S}_n^{-1/2}\sum_{t=1}^n\left\{m_{n,t}-m_{n,t}^*\right\}\left(1+o_p\left(1\right)\right)+o_p\left(1\right),\end{aligned} \tag{16}$$

Combine (15) and (16) to obtain $n^{1/2}\mathcal{S}_n^{-1/2}\mathcal{G}_n(\hat{\theta}_n^* - \hat{\theta}_n) = n^{-1/2}\mathcal{S}_n^{-1/2}\sum_{t=1}^n\{m_{n,t} - m_{n,t}^*\}(1 + o_p(1)) + o_p(1)$. By Loève's inequality, non-degeneracy $\liminf_{n\geq N}||\mathcal{S}_n|| > 0$ for some $N > 0$, and Proposition A.1.iv, it follows for tiny $\iota > 0$, $\rho \in (0,1)$, and sufficiently large n and K

$$E\left|\frac{1}{n^{1/2}}\mathcal{S}_n^{-1/2}\sum_{t=1}^n\left\{\hat{m}_{n,t}-\hat{m}_{n,t}^*\right\}\right|^\iota \leq K\frac{1}{n^{\iota/2}}\sum_{t=1}^n E\left|m_{n,t}-m_{n,t}^*\right|^\iota \leq K\frac{1}{n^{\iota/2}}\sum_{t=1}^n\rho^t = o\left(1\right).$$

Therefore $\mathcal{V}_n^{1/2}(\hat{\theta}_n^* - \hat{\theta}_n) = o_p(1)$ by Chebyshev's inequality and the construction $\mathcal{V}_n = n\mathcal{G}_n\mathcal{S}_n^{-1}\mathcal{G}_n$. Finally, $\mathcal{V}_{i,i,n} \to \infty$ given $\liminf_{n\to\infty}||\mathcal{G}_n|| > 0$ by Assumption 4.b and $||n\mathcal{S}_n^{-1}|| \to \infty$ by Lemma A.4.a. $\mathcal{QED}$.

The proofs of consistency and asymptotic normality Theorems 2.1 and 2.2 require supporting Lemmas A.2-A.9. We state them when required, where Assumptions 1-5 implicitly hold. Consistency requires variance bounds, asymptotic bounds on $\hat{m}_{n,t}^*(\theta) - m_{n,t}^*(\theta)$, and laws of large numbers.

LEMMA A.3 (*asymptotic approximation*). *a.* $||n^{-1/2}\mathcal{S}_n^{-1/2}\sum_{t=1}^n\{\hat{m}_{n,t}^* - m_{n,t}^*\}|| = o_p(1)$; *b.* $\sup_{\theta\in\Theta}\{||\hat{m}_n^*(\theta) - m_n^*(\theta)||\} = o_p(\sup_{\theta\in\Theta}||E[m_{n,t}^*(\theta)]||)$; *c.* $\sup_{\theta\in\Theta}\{||\hat{m}_n^*(\theta) - m_n^*(\theta)||/E||m_{n,t}^*(\theta)||\} = o_p(1)$.

LEMMA A.4 (*variance bounds*). *a.* $||\mathcal{S}_n(\theta)|| \leq Kr_n(\theta)||\Sigma_n(\theta)|| = o(n)$ *for some sequence of positive* *numbers* $\{r_n(\theta)\}$*, where in general* $\liminf_{n\rightarrow\infty}\inf_{\theta\in\Theta} r_n(\theta) > 0$ *and* $\sup_{\theta\in\Theta} r_n(\theta) = O(\ln(n))$*, and* $r_n(\theta) \sim K$ *if* $m_t(\theta)$ *is finite dependent or each* $E[m_{i,t}^2(\theta)] < \infty$*; b.* $\Sigma_n(\theta) = o(n||E[m_{n,t}^*(\theta)]||^2)$ *and* $\sup_{\theta\in\Theta}||\Sigma_n(\theta)|| = o(n\sup_{\theta\in\Theta}||E[m_{n,t}^*(\theta)]||^2)$.

LEMMA A.5 (*LLN and ULLN*). *a.* $1/n\sum_{t=1}^n m_{n,t}^*(\theta^0) = o_p(1)$*; b.* $\sup_{\theta\in\Theta}\{||m_n^*(\theta) - E[m_{n,t}^*(\theta)]||\} = o_p(\sup_{\theta\in\Theta}||E[m_{n,t}^*(\theta)]||)$.

Asymptotic normality requires an asymptotic Taylor expansion, a central limit theorem, and Jacobian consistency. Define $\tilde{\mathcal{G}}_n^*(\theta) := 1/n\sum_{t=1}^n G_t^*(\theta)I_{n,t}^*(\theta)$ and $\widehat{\mathcal{G}}_n^*(\theta) := 1/n\sum_{t=1}^n G_t^*(\theta)\hat{I}_{n,t}^*(\theta)$.

LEMMA A.6 (*asymptotic expansion*). *Let* $\theta, \tilde{\theta} \in \Theta$ *be arbitrary, and let* $\delta > 0$ *be arbitrarily large and finite. For* $o_p(1)$ *not a function of* θ*: a.* $m_n^*(\theta) = m_n^*(\tilde{\theta}) + \tilde{\mathcal{G}}_n^*(\theta) \times (\theta - \tilde{\theta}) + n^{-\delta} \times ||\theta - \tilde{\theta}||^{1/\iota} \times o_p(1)$*; and b.* $\hat{m}_n^*(\theta) = \hat{m}_n^*(\tilde{\theta}) + \widehat{\mathcal{G}}_n^*(\theta) \times (\theta - \tilde{\theta}) + n^{-\delta} \times ||\theta - \tilde{\theta}||^{1/\iota} \times o_p(1)$.

LEMMA A.7 (*CLT*). $n^{-1/2}\mathcal{S}_n^{-1/2}\sum_{t=1}^n m_{n,t}^* \xrightarrow{d} N(0, I_q)$.

LEMMA A.8 (*Jacobian consistency*). *a.* $\widehat{\mathcal{G}}_n^*(\hat{\theta}_n) = E[G_t^* I_{n,t}^*] \times (1 + o_p(1))$*; b.* $(\partial/\partial\theta)E[m_{n,t}^*(\theta)]|_{\theta^0} = E[G_t^* I_{n,t}^*] \times (1 + o(1))$.

We are now ready to prove Theorems 2.1 and 2.2.

PROOF OF THEOREM 2.1. Define $\mathcal{M}_n^*(\theta) := E[m_{n,t}^*(\theta)]$ and $\mathfrak{e}_n := \sup_{\theta\in\Theta}||E[m_{n,t}^*(\theta)]||$. We will prove the infeasible $\hat{\theta}_n^* \xrightarrow{p} \theta^0$ by an argument in Pakes and Pollard (1989). The claim $\hat{\theta}_n \xrightarrow{p} \theta^0$ then follows by Proposition A.1 by mimicking arguments in MS (2009: proof of Theorem 1).

Note $\epsilon(\delta) := \liminf_{n\rightarrow\infty}\inf_{\theta\in\Theta:||\theta-\theta^0||>\delta}\{||\mathcal{M}_n^*(\theta)||/\mathfrak{e}_n\} > 0$ for arbitrarily small $\delta > 0$ by moment smoothness Assumption 4.a. Since $P(||\hat{\theta}_n^* - \theta^0|| > \delta) \leq P(||\mathcal{M}_n^*(\hat{\theta}_n^*)||/\mathfrak{e}_n > \epsilon(\delta))$, it suffices to show $||\mathcal{M}_n^*(\hat{\theta}_n^*)||/\mathfrak{e}_n = o_p(1)$ in order to prove $\hat{\theta}_n^* \xrightarrow{p} \theta^0$. By Minkowski's inequality

$$\left\|\mathcal{M}_n^*(\hat{\theta}_n^*)\right\|/\mathfrak{e}_n \leq \left\|\hat{m}_n^*(\hat{\theta}_n^*)\right\|/\mathfrak{e}_n + \left\|\hat{m}_n^*(\hat{\theta}_n^*) - \mathcal{M}_n^*(\hat{\theta}_n^*)\right\|/\mathfrak{e}_n = \mathcal{A}_n(\hat{\theta}_n^*) + \mathcal{B}_n(\hat{\theta}_n^*),$$

say. The proof is complete if we show $\mathcal{A}_n(\hat{\theta}_n^*)$ and $\mathcal{B}_n(\hat{\theta}_n^*)$ are $o_p(1)$.

Consider $\mathcal{A}_n(\hat{\theta}_n^*)$. The QMTTL loss equations $l_t^*(\theta)$ have a smooth distribution under Assumption 3. This implies $\hat{Q}_n^*(\theta) := 1/n\sum_{t=1}^n l_t^*(\theta)\hat{I}_{n,t}^*(\theta)$ is differentiable at $\hat{\theta}_n^*$ with probability one, hence up to a scalar constant $(\partial/\partial\theta)\hat{Q}_n^*(\theta)|_{\hat{\theta}_n^*} = \hat{m}_n^*(\hat{\theta}_n^*)$ $a.s.$ (Čižek 2008: Lemma 2.1). By $\hat{\theta}_n^*$ a minimum $\hat{Q}_n^*(\hat{\theta}_n^*) \leq \hat{Q}_n^*(\theta)$ $\forall\theta \in \Theta$ it follows $||\hat{m}_n^*(\hat{\theta}_n^*)|| = 0$, while $\liminf_{n\to\infty} \mathfrak{e}_n > 0$ by Assumption 4.b, hence $\mathcal{A}_n(\hat{\theta}_n^*) = 0$ $a.s.$

Next, $\mathcal{B}_n(\hat{\theta}_n^*)$. Combine $\sup_{\theta\in\Theta}\{||\hat{m}_n^*(\theta) - m_n^*(\theta)||/\mathfrak{e}_n\} = o_p(1)$ by Lemma A.3.b and $\sup_{\theta\in\Theta}\{||m_n^*(\theta) - \mathcal{M}_n^*(\theta)||/\mathfrak{e}_n\} = o_p(1)$ by ULLN Lemma A.5.b to deduce

$$\sup_{\theta\in\Theta}\{\mathcal{B}_n(\theta)\} \leq \sup_{\theta\in\Theta}\left\{\frac{\|\hat{m}_n^*(\theta) - m_n^*(\theta)\|}{\mathfrak{e}_n}\right\} + \sup_{\theta\in\Theta}\left\{\frac{\|m_n^*(\theta) - \mathcal{M}_n^*(\theta)\|}{\mathfrak{e}_n}\right\} = o_p(1). \ \mathcal{QED}.$$

PROOF OF THEOREM 2.2. Note $1/n\sum_{t=1}^n \hat{m}_{n,t}^*(\hat{\theta}_n^*) = 0$ $a.s.$ by the proof of Theorem 2.1. Apply expansion Lemma A.6.b to deduce for some $\delta > 0$ arbitrarily large and finite,

$$-\widehat{\mathcal{G}}_n^*(\hat{\theta}_n^*)\left(\hat{\theta}_n^* - \theta^0\right) + \frac{1}{n}\sum_{t=1}^n \hat{m}_{n,t}^* + n^{-\delta} \times o_p(1) = 0 \quad a.s. \tag{17}$$

Consistency $\hat{\theta}_n^* \xrightarrow{p} \theta^0$ ensures $\widehat{\mathcal{G}}_n^*(\hat{\theta}_n^*) = \mathcal{G}_n(1 + o_p(1))$ by Lemma A.8.a. Multiply both sides of (17) by $n^{1/2}\mathcal{S}_n^{-1/2}$, rearrange terms and use $\mathcal{V}_n = n\mathcal{G}_n'\mathcal{S}_n^{-1}\mathcal{G}_n$ to deduce

$$\mathcal{V}_n^{1/2}\left(\hat{\theta}_n^* - \theta^0\right) = -\frac{1}{n^{1/2}}\mathcal{S}_n^{-1/2}\sum_{t=1}^n \hat{m}_{n,t}^* + o_p\left(n^{1/2}\left\|\mathcal{S}_n\right\|^{-1/2} n^{-\delta}\right).$$

Since $\delta > 0$ is arbitrary and $\limsup_{n\to\infty} ||\mathcal{S}_n||^{-1/2} > 0$ by Assumption 4.b, the last term $o_p(n^{1/2}||\mathcal{S}_n||^{-1/2}n^{-\delta})$ $= o_p(1)$.

Now use $n^{-1/2}\mathcal{S}_n^{-1/2}\sum_{t=1}^n\{\hat{m}_{n,t}^* - m_{n,t}^*\} = o_p(1)$ by Lemma A.3.a, and $\delta > 0$ arbitrarily large to deduce $\mathcal{V}_n^{1/2}(\hat{\theta}_n^* - \theta^0) = -n^{-1/2}\mathcal{S}_n^{-1/2}\sum_{t=1}^n m_{n,t}^* \times (1 + o(1)) + o_p(1)$. Therefore $\mathcal{V}_n^{1/2}(\hat{\theta}_n^* - \theta^0) \xrightarrow{d} N(0, I_q)$ by CLT Lemma A.7, hence $\mathcal{V}_n^{1/2}(\hat{\theta}_n - \theta^0) \xrightarrow{d} N(0, I_q)$ by Proposition A.2. $\mathcal{QED}$.

PROOF OF THEOREM 3.1. All steps used to prove Theorems 2.1 and 2.2 carry over to show $\mathring{\mathcal{V}}_n^{1/2}(\hat{\theta}_n - \theta^0) \xrightarrow{d} N(0, I_p)$ by straightforward alterations of Lemmas A.3-A.8. $\mathcal{QED}$.

PROOF OF THEOREM 3.3. The GARCH process $\{y_t\}$ is stationary and ergodic, and geometrically β-mixing with power law tails (Nelson 1990, Basrak et al 2002, Carrasco and Chen 2002). Since ϵ_t is independent with power law tails if $E[\epsilon_t^4] = \infty$, it is straightforward to show $m_{i,t}(\theta)$ and $G_{i,j,t}(\theta)$ have power law tails (5). See Cline (1986) and his references. Therefore Assumptions 1-3 hold.

We now verify moment smoothness Assumption 4. Recall $\mathring{m}_{n,t}^*(\theta) = (\epsilon_t^2(\theta) - 1)I_{n,t}^{(\epsilon)}(\theta)\mathfrak{s}_{n,t}^*(\theta)\, I_{n,t}^{(\mathfrak{s}^\theta)}(\theta)$, and by Corollary 3.2 condition (12) implies $\mathring{\mathcal{G}}_n = E[\mathfrak{s}_t^*\mathfrak{s}_t^{*\prime}I_{n,t}^{(y)}]$. By the definition of a derivative and an extension of Lemma A.8.b to $\mathring{m}^*_{n,t}(\theta)$,

$$E\left[\mathring{m}^*_{n,t}(\theta)\right] = \mathring{\mathcal{G}}_n \times \left(\theta - \theta^0\right) \times (1 + o(1)).$$

Let $r \in \mathbb{R}^q$, $r'r = 1$. Distribution smoothness, trimming negligibility and stationarity imply $\liminf_{n\rightarrow\infty} \inf_{r'r=1} r'E[\mathfrak{s}^*_t\mathfrak{s}^{*\prime}_t I^{(y)}_{n,t}]r = \inf_{r'r=1} E(\sum_{i=1}^p r_i\mathfrak{s}^*_{i,t}I^{(y)}_{n,t})^2 > 0$, thus $\mathring{\mathcal{G}}_n$ is non-singular for each $n \geq N$ and some $N \in \mathbb{N}$. Therefore, since $\liminf_{n\rightarrow\infty} ||\mathring{\mathcal{G}}_n|| > 0$ and Θ is compact it follows $\mathfrak{e}_n := \sup_{\theta\in\Theta} ||E[\mathring{m}^*_{n,t}(\theta)]|| \leq K||\mathring{\mathcal{G}}_n|| \times (1 + o(1))$, hence

$$\inf_{||\theta-\theta^0||>\delta} \left\{\mathfrak{e}_n^{-1}\left\|E\left[\mathring{m}^*_{n,t}(\theta)\right]\right\|\right\} \geq K \inf_{||\theta-\theta^0||>\delta} \left\{\left\|\frac{\mathring{\mathcal{G}}_n}{\left\|\mathring{\mathcal{G}}_n\right\|} \times \left(\theta - \theta^0\right)\right\|\right\} \times (1 + o(1)) > 0$$

for every $n \geq N$. This verifies Assumption 4.a. The nondegeneracy moment bounds of Assumption 4.b can be straightforwardly verified from distribution continuity, the linear GARCH data generating process, and the geometric β-mixing property. $\mathcal{QED}$.

PROOF OF THEOREM 4.1. By the proof of Theorem 2.1 $\sum_{t=1}^n \hat{m}^*_{n,t}(\hat{\theta}_n) = 0$ *a.s.* The claim therefore follows by the proof of Theorem 2.2. $\mathcal{QED}$.

LEMMA A.9 (*HAC*). *Under the conditions of Theorem 5.1* $\widehat{\mathcal{S}}^*_n(\hat{\theta}_n) = \mathcal{S}_n(1 + o_p(1))$.

PROOF OF THEOREM 5.1. The claim follows from Jacobian consistency Lemma A.8.a and HAC consistency Lemma A.9. $\mathcal{QED}$.

PROOF OF THEOREM 5.2. The proofs that $1/n\sum_{t=1}^n \epsilon_t^4(\hat{\theta}_n)\hat{I}^{(\epsilon)}_{n,t}(\hat{\theta}_n) = E[\epsilon_t^4 I^{(\epsilon)}_{n,t}] \times (1 + o_p(1))$ and $1/n\sum_{t=1}^n \mathfrak{s}_t(\hat{\theta}_n)\mathfrak{s}'_t(\hat{\theta}_n)\hat{I}^{(y)}_{n,t} = E[\mathfrak{s}_t\mathfrak{s}'_t I^{(y)}_{n,t}] \times (1 + o_p(1))$ mimic the same arguments used to prove Jacobian consistency Lemma A.8. $\mathcal{QED}$.

APPENDIX B: Remaining Proofs

By Proposition A.2 it suffices to treat the infeasible estimator $\hat{\theta}^*_n$ and components $m^*_t(\theta)$ and $G^*_t(\theta)$. We therefore prove all lemmas for $m^*_t(\theta)$, $G^*_t(\theta)$, and so on. Since the notation "$*$" is therefore repetitive we drop it everywhere and write $\hat{\theta}_n$ for $\hat{\theta}^*_n$, $h_t(\theta)$ for $h^*_t(\theta)$, and so on.

Assume *symmetric trimming* throughout to simplify notation. Let $w_t(\theta)$ denote any scalar $m_{i,t}(\theta)$ or $G_{i,j,t}(\theta)$ and let $\{k_n^{(w)}, \mathcal{C}_n^{(w)}(\theta)\}$ be the associated fractile and threshold sequences:

$$w_t(\theta) \in \{m_{i,t}(\theta), G_{i,j,t}(\theta)\}, \;\; P\left(|w_t(\theta)| > \mathcal{C}_n^{(w)}(\theta)\right) = \frac{k_n^{(w)}}{n}.$$

We simply write $\mathcal{C}_n(\theta)$ and k_n whenever $w_t(\theta)$ is understood, and we drop θ^0.

We shorten arguments in lieu of Assumption 3.b by assuming $w_t(\theta)$ have power law tails for *any* θ:

$$\sup_{\theta\in\Theta}\left\{\left|c^{\kappa_w(\theta)}P\left(|w_t(\theta)| > c\right) - d_w(\theta)\right|\right\} \to 0 \;\; as \;\; c \to \infty, \;\; \inf_{\theta\in\Theta}\left\{\kappa_w(\theta), d_w(\theta)\right\} > 0. \tag{18}$$

By construction of $\mathcal{C}_n(\theta)$ and (18) we have the following *fractile and threshold* properties:

$$\inf_{\theta\in\Theta}\frac{\mathcal{C}_n(\theta)}{(n/k_n)^{1/\kappa_w(\theta)}} \to (0,\infty) \quad \text{and} \quad \sup_{\theta\in\Theta}\frac{\mathcal{C}_n(\theta)}{(n/k_n)^{1/\kappa_w(\theta)}} \to (0,\infty)\,. \tag{FT}$$

Applications of Karamata's Theorem therefore gives the following *trimmed moments*:

$$\text{if} \;\; \kappa_w(\theta) < 2: \;\; E\left[w_t^2(\theta)I\left(|w_t(\theta)| \leq \mathcal{C}_n(\theta)\right)\right] \sim K\left(\mathcal{C}_n(\theta)\right)^2 \times (k_n/n) = K(n/k_n)^{2/\kappa_w(\theta)-1}$$

$$\text{if} \;\; \kappa_w(\theta) = 2: \;\; E\left[w_t^2(\theta)I\left(|w_t(\theta)| \leq \mathcal{C}_n(\theta)\right)\right] \sim L\left(n\right).$$

Uniform bounds are similar in lieu of (FT): for example if $\kappa_w(\theta) < 2$ then for finite $K > 0$

$$\sup_{\theta\in\Theta}\left\{\frac{n}{k_n}\frac{\mathcal{C}_n^2(\theta)}{E\left[w_t^2(\theta)I\left(|w_t(\theta)| \leq \mathcal{C}_n(\theta)\right)\right]}\right\} \to K. \tag{TM}$$

The proofs of Lemmas A.3-A.9 require three supporting results which we prove in the supplemental material Hill (2012). Let Assumptions 1-5 hold. First, trimming indicators satisfy a uniform law.

LEMMA B.1 (*uniform indicator law*). *Define* $\mathcal{I}_{n,t}(\theta) := ((n/k_n)^{1/2})\{I(|w_t(\theta)| \leq \mathcal{C}_n(\theta)) - E[I(|w_t(\theta)| \leq \mathcal{C}_n(\theta))]\}$. *Then* $\{n^{-1/2}\sum_{t=1}^n \mathcal{I}_{n,t}(\theta) : \theta \in \Theta\} \Longrightarrow^* \{\mathcal{I}(\theta) : \theta \in \Theta\}$ *and* $E[\sup_{\theta\in\Theta}|n^{-1/2}\sum_{t=1}^n \mathcal{I}_{n,t}(\theta)|] = O(1)$ *where* $\mathcal{I}(\theta)$ *is a Gaussian process with uniformly bounded and uniformly continuous sample paths with respect to* L_2*-norm, and* $\Longrightarrow^*$ *denotes weak convergence on a Polish space.*[7].

Second, intermediate order statistics are uniformly bounded in probability.

[7] See Dudley (1978), Giné and Zinn (1984), and Pollard (1984).

LEMMA B.2 (*uniform order statistic*). *Define* $w_t^{(a)}(\theta) := |w_t(\theta)|$. *Then* $\sup_{\theta\in\Theta} |w_{(k_n)}^{(a)}(\theta)/\mathcal{C}_n(\theta) - 1| = O_p(k_n^{-1/2})$.

Third, we require an approximation for a cross-product sum for HAC asymptotics.

LEMMA B.3 (*cross-product approximation*). *Under the kernel properties of Theorem 5.1* $n^{-1}\mathcal{S}_n^{-1}(\theta)\sum_{s,t=1}^n \mathcal{K}_{n,s,t}\{\hat{m}_{n,s}(\hat{\theta}_n)\hat{m}_{n,t}(\hat{\theta}_n) - m_{n,s}(\theta^0)m_{n,t}(\theta^0)\} = o_p(1)$.

We are now ready to prove Lemmas A.3-A.9. Assume $m_t(\theta)$ and θ are scalars for notational convenience.

PROOF OF LEMMA A.3.

Claim (a): Note by Lemma B.2 $\sup_{\theta\in\Theta} |m_{(k_n)}^{(a)}(\theta)/\mathcal{C}_n(\theta) - 1| = O_p(1/k_n^{1/2})$, while uniform indicator law Lemma B.1 and the threshold construction imply

$$\sup_{\theta\in\Theta}\left\{\left|\frac{1}{k_n^{1/2}}\sum_{t=1}^n\left\{\bar{I}_{n,t}^{(m)}(\theta) - E\left[\bar{I}_{n,t}^{(m)}(\theta)\right]\right\}\right|\right\} = O_p(1), \quad \sup_{\theta\in\Theta}\left\{\left|\frac{n}{k_n}E\left[\bar{I}_{n,t}^{(m)}(\theta)\right] - 1\right|\right\} = 1. \tag{19}$$

Now let $\theta \in \Theta$ be arbitrary, and write $m_t = m_t(\theta)$, $\mathcal{C}_n = \mathcal{C}_n(\theta)$, $\hat{m}_{n,t} = \hat{m}_{n,t}(\theta)$, $m_{n,t} = m_{n,t}(\theta)$, $\bar{I}_{n,t} = 1 - I_{n,t}(\theta)$, $\hat{I}_{n,t} = \hat{I}_{n,t}(\theta)$, and $\mathcal{S}_n = \mathcal{S}_n(\theta)$.

Observe

$$\left|\sum_{t=1}^n\{\hat{m}_{n,t} - m_{n,t}\}\right| \le \max_{1\le t\le n}\left\{\left|m_t\left\{\hat{I}_{n,t} - I_{n,t}\right\}\right|\right\} \times \sum_{t=1}^n\left|\hat{I}_{n,t} - I_{n,t}\right|. \tag{20}$$

By construction $|m_t\{\hat{I}_{n,t} - I_{n,t}\}| \le 2|m_{(k_n)}^{(a)} - \mathcal{C}_n|$, and under the Lemma A.4.a,c bounds it follows $\mathcal{C}_n = o(n^{1/2}||\mathcal{S}_n||^{1/2})$, hence

$$\max_{1\le t\le n}\left\{\left|m_t\left\{\hat{I}_{n,t} - I_{n,t}\right\}\right|\right\} \le 2\left|m_{(k_n)}^{(a)} - \mathcal{C}_n\right| = 2\mathcal{C}_n\left|m_{(k_n)}^{(a)}/\mathcal{C}_n - 1\right| = o_p\left(n^{1/2}\mathcal{S}_n^{1/2}k_n^{-1/2}\right). \tag{21}$$

Next, by construction and the triangle inequality

$$\sum_{t=1}^n\left|\hat{I}_{n,t} - I_{n,t}\right| \le k_n^{1/2}\left|\frac{1}{k_n^{1/2}}\sum_{t=1}^n\left\{\bar{I}_{n,t} - E\left[\bar{I}_{n,t}\right]\right\}\right| + k_n^{1/2}\left|k_n^{1/2}\left(\frac{n}{k_n}E\left[\bar{I}_{n,t}\right] - 1\right)\right| = O_p(k_n^{1/2}) \tag{22}$$

since $(n/k_n)E[\bar{I}_{n,t}] - 1 = 0$ by the threshold construction, and $k_n^{-1/2}\sum_{t=1}^n\{\bar{I}_{n,t} - E\left[\bar{I}_{n,t}\right]\} = O_p(1)$ by (19). Together (20)-(22) imply $\sum_{t=1}^n\{\hat{m}_{n,t} - m_{n,t}\} = o_p(n^{-1/2}\mathcal{S}_n^{1/2})$.

Claim (b): Define

$$\mathcal{M}_n := \max_{1\leq t\leq n}\left\{\sup_{\theta\in\Theta}\left|m_t(\theta)\{\hat{I}_{n,t}(\theta)-I_{n,t}(\theta)\}\right|\right\}.$$

By the Claim (a) argument and (19)

$$\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{t=1}^{n}\{\hat{m}_{n,t}(\theta)-m_{n,t}(\theta)\}\right|\leq\mathcal{M}_n\times\frac{k_n^{1/2}}{n}\sup_{\theta\in\Theta}\left|\frac{1}{k_n^{1/2}}\sum_{t=1}^{n}\left\{\bar{I}_{n,t}(\theta)-E\left[\bar{I}_{n,t}(\theta)\right]\right\}\right|=O_p\left(\mathcal{M}_n k_n^{1/2}/n\right).$$

It remains to prove $\mathcal{M}_n = o_p(\sup_{\theta\in\Theta}|E[m_{n,t}(\theta)]|n/k_n^{1/2})$. Since

$$\left|m_t(\theta)\left\{\hat{I}_{n,t}(\theta)-I_{n,t}(\theta)\right\}\right|\leq 2\mathcal{C}_n(\theta)\left|m_{(k_n)}^{(a)}(\theta)/\mathcal{C}_n(\theta)-1\right|,$$

and $\sup_{\theta\in\Theta}|m_{(k_n)}^{(a)}(\theta)/\mathcal{C}_n(\theta)-1| = O_p(k_n^{-1/2})$, use the Lemma A.4.b bound to deduce

$$\begin{aligned}\mathcal{M}_n &\leq K\sup_{\theta\in\Theta}\{\mathcal{C}_n(\theta)\}\left(\sup_{\theta\in\Theta}m_{(k_n)}^{(a)}(\theta)/\mathcal{C}_n(\theta)-1\right)\\ &\leq o_p\left(\sup_{\theta\in\Theta}\left\{\left(E\left[m_{n,t}^2(\theta)\right]\right)^{1/2}\right\}n^{1/2}/k_n^{1/2}\right)=o_p\left(\sup_{\theta\in\Theta}\left|E\left[m_{n,t}(\theta)\right]\right|n/k_n^{1/2}\right).\end{aligned}$$

Claim (c): By construction $|m_t(\theta)\{\hat{I}_{n,t}(\theta) - I_{n,t}(\theta)\}| \leq 2c_n(\theta)|m_{(k_n)}^{(a)}(\theta)/c_n(\theta) - 1|$ where $\sup_{\theta\in\Theta}|m_{(k_n)}^{(a)}(\theta)/c_n(\theta) - 1| = O_p(1/k_n^{1/2})$, and

$$\sum_{t=1}^{n}\left|\hat{I}_{n,t}(\theta)-I_{n,t}(\theta)\right|\leq k_n^{1/2}\left(\left|\frac{1}{k_n^{1/2}}\sum_{t=1}^{n}\left\{\bar{I}_{n,t}(\theta)-E\left[\bar{I}_{n,t}(\theta)\right]\right\}\right|+\left|k_n^{1/2}\left(\frac{n}{k_n}E\left[\bar{I}_{n,t}(\theta)\right]-1\right)\right|\right),$$

which is $O_p(k_n^{1/2})$ by (19). In view of (20) it therefore follows

$$\sup_{\theta\in\Theta}\left|\frac{1}{nE\left|m_{n,t}(\theta)\right|}\sum_{t=1}^{n}\{\hat{m}_{n,t}(\theta)-m_{n,t}(\theta)\}\right|=O_p\left(\sup_{\theta\in\Theta}\left\{\frac{c_n(\theta)}{n\times E\left|m_{n,t}(\theta)\right|}\right\}\right)=O_p\left(\mathcal{D}_n\right),$$

say. If $\inf_{\theta\in\Theta}k_m(\theta) < 1$ then by power-law tail Assumption 3.b and Karamata's Theorem $\sup_{\theta\in\Theta}\{E||m_{n,t}(\theta)|| /c_n(\theta)\} = K(k_n/n)$ hence $\mathcal{D}_n = O(1/k_n) = o(1)$. If $\inf_{\theta\in\Theta}k_m(\theta) \geq 1$ then by Assumption A.4 we have non-degeneracy $\liminf_{n\to\infty}\inf_{\theta\in\Theta}E|m_{n,t}(\theta)| > 0$, and by power law decay and the threshold construction $\sup_{\theta\in\Theta}\{c_n(\theta)/n\} = K(n/k_n)^{1/\inf_{\theta\in\Theta}k_m(\theta)}/n = o(1)$. $\mathcal{QED}$.

PROOF OF LEMMA A.4. See the supplemental material Hill (2012). $\mathcal{QED}$.

PROOF OF LEMMA A.5.

Claim (a): $1/n\sum_{t=1}^{n}m_{n,t} = o_p(1)$ follows from $E[m_{n,t}]\to 0$, the Lemma A.4.a variance bound $||\mathcal{S}_n|| =$

$o(n)$ and Chebyshev's inequality.

Claims (b): Define for any $i \in \{1, ..., q\}$

$$h_{n,t}(\theta) := \frac{m_{i,n,t}(\theta) - E[m_{i,n,t}(\theta)]}{\sup_{\theta \in \Theta} \|E[m_{n,t}(\theta)]\|}.$$

By Lemma A.4.a,b and Chebyshev's inequality $1/n \sum_{t=1}^{n} h_{n,t}(\theta = o_p(1)$. Further, $h_{n,t}(\theta)$ is uniformly L_1-bounded so it belongs to a separable Banach space, hence the L_1-bracketing numbers satisfy $N_{[\,]}(\varepsilon, \Theta, ||\cdot||_1)$ $< \infty$ (Dudley 1999: Proposition 7.1.7). Now combine the pointwise law and $N_{[\,]}(\varepsilon, \Theta, ||\cdot||_1) < \infty$ to deduce $\sup_{\theta \in \Theta} |1/n \sum_{t=1}^{n} h_{n,t}(\theta)| = o_p(1)$ by Theorem 7.1.5 of Dudley (1999). This proves (b). $\mathcal{QED}$.

PROOF OF LEMMA A.6. We prove (a), the more difficult of the two. Assume θ is a scalar to simplify notation. Choose any $\theta, \tilde{\theta} \in \Theta$, and define $\tilde{\mathcal{G}}_n(\theta) := 1/n \sum_{t=1}^{n} G_t(\theta) I_{n,t}(\theta)$. By the mean value theorem $m_{n,t}(\theta) = \{m_t(\tilde{\theta}) + G_t(\theta_*)(\theta - \tilde{\theta})\} \times I_{n,t}(\theta)$ for some θ_* that satisfies $|\theta_* - \tilde{\theta}| \leq |\theta - \tilde{\theta}|$, hence

$$\begin{aligned} m_n(\theta) - m_n(\tilde{\theta}) &= \tilde{\mathcal{G}}_n(\theta_*) \times (\theta - \tilde{\theta}) + \frac{1}{n}\sum_{t=1}^{n} m_t(\theta) \times \left\{I_{n,t}(\theta) - I_{n,t}(\tilde{\theta})\right\} \\ &\quad + \frac{1}{n}\sum_{t=1}^{n} G_t(\theta_*) \times \{I_{n,t}(\theta) - I_{n,t}(\theta_*)\} \times (\theta - \tilde{\theta}). \end{aligned} \tag{23}$$

We need only show the second and third terms are $o_p(n^{-\delta}) \times |\theta - \tilde{\theta}|^{1/\iota}$ for any $\delta > 0$ and tiny $\iota > 0$.

Consider the second term in (23) and use $I_{n,t}(\theta) - I_{n,t}(\tilde{\theta}) \in \{-1, 0, 1\}$ to bound

$$\begin{aligned} \left|\frac{1}{n}\sum_{t=1}^{n} m_t(\theta)\left\{I_{n,t}(\theta) - I_{n,t}(\tilde{\theta})\right\}\right| &\leq \frac{1}{n^{1/2}}\sum_{t=1}^{n}\left|m_t(\theta)\left\{I_{n,t}(\theta) - I_{n,t}(\tilde{\theta})\right\}\right| \\ &\quad \times \frac{1}{n^{1/2}}\sum_{t=1}^{n}\left|I_{n,t}(\theta) - I_{n,t}(\tilde{\theta})\right| = \mathcal{A}_n(\theta, \tilde{\theta}) \times \mathcal{B}_n(\theta, \tilde{\theta}). \end{aligned}$$

Consider $\mathcal{A}_n(\theta, \tilde{\theta})$, and observe by the threshold $\mathcal{C}_n^{(m)}(\theta)$ and trimming indicator $I_{n,t}(\theta)$ constructions

$$\sup_{\theta, \tilde{\theta} \in \Theta} E\left|I_{n,t}(\theta) - I_{n,t}(\tilde{\theta})\right| \leq 2\sup_{\theta \in \Theta} E\left|1 - I_{n,t}^{(m)}(\theta)\right| \leq 2\sup_{\theta \in \Theta} P\left(|m_t(\theta)| > \mathcal{C}_n(\theta)\right) = O\left(k_n/n\right). \tag{24}$$

By L_ι-boundedness of $m_t(\theta)$ for tiny $\iota > 0$ and the Cauchy-Schwartz inequality

$$\sup_{\theta, \tilde{\theta} \in \Theta}\left(E\left[\mathcal{A}_n(\theta, \tilde{\theta})^\iota\right]\right)^{1/\iota} \leq \sup_{\theta, \tilde{\theta} \in \Theta} n^{1/2}\left[E\left|m_t(\theta)\left\{I_{n,t}(\theta) - I_{n,t}(\tilde{\theta})\right\}\right|^\iota\right]^{1/\iota} = O\left(n^{1/2}\left(k_n/n\right)^{1/\iota}\right).$$

where $O(\cdot)$ is not a function of θ.

Now, by Assumption 3.b $m_t(\theta)$ is uniformly L_ι-bounded with a continuous bounded distribution. Hence by (24), the mean value theorem and the threshold construction $E|I_{n,t}(\theta) - I_{n,t}(\tilde{\theta})|^\iota = O(k_n/n) \times |\theta - \tilde{\theta}|$, where $O(k_n/n)$ can be assumed to not depend on $\theta \in \Theta$ since Θ is bounded.

But this ensures for tiny $\iota > 0$

$$\left(E\left[\mathcal{A}_n(\theta,\tilde{\theta})^\iota\right]\right)^{1/\iota} \leq n^{1/2}\left[E\left|m_t(\theta)\left\{I_{n,t}(\theta) - I_{n,t}(\tilde{\theta})\right\}\right|^\iota\right]^{1/\iota} = O\left(n^{1/2}(k_n/n)^{1/\iota}\right) \times \left|\theta - \tilde{\theta}\right|^{1/\iota}$$

where $O(\cdot)$ is not a function of θ.

Now use $k_n/n \to 0$, the fact that $\iota > 0$ is arbitrarily small, and Markov's inequality to deduce for $\delta > 0$ arbitrarily large that may be different in different places, and $o_p(\cdot)$ that is not a function of θ,

$$\mathcal{A}_n(\theta,\tilde{\theta}) = o_p\left(n^{1/2}(k_n/n)^{1/\iota}\left|\theta - \tilde{\theta}\right|^{1/\iota}\right) = o_p\left(n^{-\delta}\right) \times \left|\theta - \tilde{\theta}\right|^{1/\iota}.$$

Since $\sup_{\theta,\tilde{\theta}\in\Theta}\{\mathcal{B}_n(\theta,\tilde{\theta})\} \leq n^{1/2}$ by construction and $\delta > 0$ is arbitrary we have shown $1/n\sum_{t=1}^n m_t(\theta)\{I_{n,t}(\theta) - I_{n,t}(\tilde{\theta})\} = n^{-\delta} \times o_p(1) \times |\theta - \tilde{\theta}|^{1/\iota}$ where $o_p(1)$ that is not a function of θ.

The same argument applies to the third term in (23) in lieu of the Assumption 3 distribution and moment properties. $\mathcal{QED}$.

PROOF OF LEMMA A.7. See the supplemental material Hill (2012). $\mathcal{QED}$.

PROOF OF LEMMA A.8. See the supplemental material Hill (2012). $\mathcal{QED}$.

PROOF OF LEMMA A.9. Define $\mathcal{A}_{1,n} := \mathcal{S}_n^{-1}\sum_{s,t=1}^n \mathcal{K}_{n,s,t}\{\hat{m}_{n,s}(\hat{\theta}_n)\hat{m}_{n,t}(\hat{\theta}_n) - m_{n,s}m_{n,t}\}$ and $\mathcal{A}_{2,n} := \mathcal{S}_n^{-1}\sum_{s,t=1}^n \mathcal{K}_{n,s,t}m_{n,s}m_{n,t} - 1$. Since $\mathcal{A}_{1,n} \xrightarrow{p} 0$ by Lemma B.3 we need only show $\mathcal{A}_{2,n} \xrightarrow{p} 0$. We will apply Theorem 2.1 of de Jong and Davidson (2000), hence it suffices to verify their Assumptions 1-3. Assumption 1 holds by our kernel and bandwidth assumptions.

Define $\mathcal{Z}_{n,t} := n^{-1/2}\mathcal{S}_n^{-1}m_{n,t}(\theta^0)$. Under geometric β-mixing $\{m_{n,t}(\theta^0), \Im_t\}$ forms a geometric L_2-mixingale with constants $e_{n,t}$ (cf. McLeish 1975: Theorem 2.1)[8]. Therefore $\{\mathcal{Z}_{n,t}, \Im_t\}$ forms a geometric L_2-mixingale with constants $\mathcal{E}_{n,t} := n^{-1/2}\mathcal{S}_n^{-1/2}e_{n,t}$. By inspection of de Jong and Davidson's proof of their

[8] We say $\{y_{n,t}, \Im_t\}$ is a geometric L_2-mixingale with constants $e_{n,t}$ when $||E[y_{n,t}] - E[y_{n,t}|\Im_{t-q}]|| \leq e_{n,t} \times O(\rho^q)$ and $||y_{n,t} - E[y_{n,t}|\Im_{t+q}]|| \leq e_{n,t} \times O(\rho^q)$ for some $\rho \in (0,1)$. See McLeish (1975).

Theorem 2.1 it follows $E(\sum_{t=1}^{n} \mathcal{Z}_{n,t})^2 = 1 \leq K \sum_{t=1}^{n}(1/n^{1/2})^2 = K$ suffices in place of their Assumption 2. Finally, their Assumption 3 is $\gamma_n \times \max_{1 \leq t \leq n}\{\mathcal{E}_{n,t}^2\} = o(1)$. By supposition $\gamma_n = o(n)$, while by construction and Theorem 1.6 in McLeish (1975) $E(\sum_{t=1}^{n} \mathcal{Z}_{n,t})^2 \leq K \sum_{t=1}^{n} \mathcal{E}_{n,t}^2 = Kn^{-1}\mathcal{S}_n^{-1} \sum_{t=1}^{n} e_{n,t}^2$. In view of stationarity and $E(\sum_{t=1}^{n} \mathcal{Z}_{n,t})^2 = 1$ we may therefore assume $e_{n,t} = K\mathcal{S}_n^{1/2}$, hence $\max_{1 \leq t \leq n}\{\mathcal{E}_{n,t}^2\} = n^{-1}$. Thus Assumption 3 holds. $\mathcal{QED}$.

REFERENCES

[1] Agulló, J., C. Croux, and S. Van Aelst (2008). The Multivariate Least-Trimmed Squares Estimator, J. Mult. Anal. 99, 311-338.

[2] Basrak, B., R.A. Davis, and T. Mikosch (2002). Regular Variation of GARCH Processes, Stoch. Proc. Appl. 99, 95-115.

[3] Boudt, K. and C. Croux (2010). Robust M-Estimation of Multivariate GARCH Models, Comp. Stat. Data Anal. 54, 2459 2469.

[4] Carrasco, M. and X. Chen (2002). Mixing and Moment Properties of Various GARCH and Stochastic Volatility Models, Econometric Theory 18, 17-39.

[5] Čižek, P. (2008). General Trimmed Estimation: Robust Approach to Nonlinear and Limited Dependent Variable Models, Econometric Theory 24 , 1500-1529.

[6] Cline, D.B.H. (1986). Convolution Tails, Product Tails and Domains of Attraction, Prob. Theory Related Fields 72, 529–557.

[7] Cline, D.B.H. (2007). Regular Variation of Order 1 Nonlinear AR-ARCH models, Stoch. Proc. Appl. 117, 840-861.

[8] de Jong, R. M. and J. Davidson (2000). Consistency of Kernel Estimators of Heteroscedastic and Autocorrelated Covariance Matrices, Econometrica 68, 407-423.

[9] Dehling, H., M. Denker and W. Phillip (1986). Central Limit Theorems for Mixing Sequences of Random Variables under Minimal Conditions, Ann. Prob. 14, 1359-1370.

[10] Doukhan, P., P. Massart, and E. Rio, 1995, Invariance Principles for Absolutely Regular Empirical Processes. Annal. I. H. P. 31, 393-427.

[11] Dudley, R. M. (1978). Central Limit Theorem for Empirical Processes, Ann. Prob. 6, 899-929.

[12] Dudley, R. M. (1999). Uniform Central Limit Theorems. Cambridge Univ. Press: Cambridge.

[13] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). Modelling Extremal Events for Insurance and Finance. Springer-Verlag: Frankfurt.

[14] Engle, R. and V. Ng (1993). Measuring and Testing the Impact of News on Volatility, J. Fin. 48, 1749-1778.

[15] Engle, R. and J.G. Rangel (2007). The Spline-GARCH Model for Unconditional Volatility and its Global Macroeconomic Causes, Review Fin. Stud. 21, 1187-1222.

[16] Escanciano, J.C. (2009). Quasi-Maximum Likelihood Estimation of Semi-Strong GARCH Models, Econometric Theory 25, 561–570.

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

[18] Francq, C. and J-M. Zakoïan (2006). Mixing Properties of a General Class of GARCH(1,1) Models without Moment Assumptions on the Observed Process, Econometric Theory 22, 815-834.

[19] Francq, C. and J-M. Zakoïan (2010). GARCH Models, Wiley: New York.

[20] Giné, E. and J. Zinn (1984). Some Limit Theorems for Empirical Processes, Ann. Prob. 12, 929-989.

[21] Glosten, L.R., R. Jagannathan, and D.E. Runkle (1993). On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks, J. Fin. 48, 1779-1801.

[22] Hahn, M.G., D.C. Weiner and D.M. Mason (1991). Sums, Trimmed Sums and Extremes, Birkhäuser: Berlin.

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

[24] Hansen, L. (1982). Large Sample Properties of Generalized Method of Moments Estimators, Econometrica 50, 1029-1054.

[25] Hill, B.M. (1975). A Simple General Approach to Inference About the Tail of a Distribution, Ann. Stat. 3, 1163-1174.

[26] Hill, J.B. (2010). On Tail Index Estimation for Dependent, Heterogeneous Data, Econometric Theory 26, 1398-1436.

[27] Hill, J.B. (2011a). Robust M-Estimation for Heavy Tailed Nonlinear AR-GARCH, Working Paper, Dept. of Economics, University of North Carolina-Chapel Hill.

[28] Hill, J.B. (2011b). Tail and Non-Tail Memory with Applications to Extreme Value and Robust Statistics, Econometric Theory 27, 844-884.

[29] Hill, J.B. (2012). Supplemental Material for "Robust Estimation and Inference for Heavy Tailed Nonlinear GARCH" (www.unc.edu/∼jhill/qmttl_supp.pdf).

[30] Hong, Y. and J. Lee (1999). Wavelet-based Estimation of Heteroskedasticity and Autocorrelation Consistent Covariance Matrices, Technical Report, Dept. of Statistical Science, Cornell University.

[31] Leadbetter, M.R., G. Lindgren and H. Rootzén (1983). Extremes and Related Properties of Random Sequences and Processes. Springer-Verlag: New York.

[32] Lee, S-W., and B. E. Hansen (1994). Asymptotic Theory for the GARCH(1,1) Quasi-Maximum Likelihood Estimator, Econometric Theory 10, 29–52.

[33] Ling, S. (2007). Self-Weighted and Local Quasi-Maximum Likelihood Estimators for ARMA-GARCH/IGARCH Models, J. Econometrics 150, 849-873.

[34] Linton, O., J. Pan and H. Wang (2010). Estimation for a Nonstationary Semi-Strong GARCH(1,1) Model with Heavy-Tailed Errors, Econometric Theory 26, 1-28.

[35] Liu, J-C. (2006). On the Tail Behavior of a Family of GARCH Processes, Econometric Theory 22, 852-862.

[36] McLeish, D.L. (1975). Maximal Inequality and Dependent Strong Laws, Ann. Prob. 3, 829-839.

[37] Meitz, M., and P. Saikkonen (2009). Parameter Estimation in Nonlinear AR–GARCH Models, Econometric Theory: in press.

[38] Mikosch, T. and D. Straumann (2002). Whittle Estimation in a Heavy-Tailed GARCH(1,1) Model, Stoch. Proc. Appl. 100, 187-222.

[39] Nelson, D. (1990). Stationarity and Persistence in the GARCH(1,1) Model, Econometric Theory 6, 318-334.

[40] Nelson, D. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach, Econometrica 59, 347-370.

[41] Newey, W.K. and K.D. West (1987). Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, Econometrica 55, 703-708.

[42] Pakes, A. and D. Pollard (1989). Simulation and the Asymptotics of Optimization Estimators, Econometrica 57, 1027-1057.

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

[44] Pollard, D. (1984). Convergence of Stochastic Processes, Springer-Verlang New York.

[45] Ronchetti, E. and F. Trojani (2001). Robust Inference with GMM Estimators, Journal of Econometrics 101, 37-69.

[46] Sawa, T. (1978). Information Criteria for Discriminating Among Alternative Regression Models, Econometrica 46, 1273-1291.

[47] Straumann, D. and T. Mikosch (2006). Quasi-Maximum-Likelihood Estimation in Conditionally

Heteroscedastic Time Series: A Stochastic Recurrence Equations Approach, Ann. Stat. 34, 2449-2495.

[48] White, H. (1982). Maximum Likelihood Estimation of Misspecified Models, Econometrica 50, 1-26.

TABLE 1: Simulations of GARCH and GJR-GARCH

	$n = 100$				$n = 800$				$n = 2000$			
	Bias	RMS[a]	KS[b]	Tr%[c]	Bias	RMS	KS	Tr%	Bias	RMS	KS	Tr%
GARCH: $\theta_3^0 = .6$ and $\epsilon_t \sim \bar{P}_{2.5}$: $\kappa_y = 1.5$[d]												
QMFTTL	.011	.382	.889	.013	.003	.190	.667	.006	-.002	.048	.907	.002
MFTTM	-.001	.230	.759	.019	.008	.104	1.00	.002	-.002	.053	.944	.001
QML	.091	.376	4.67	-	.064	.335	4.65	-	.099	.078	8.74	-
Log-LAD	-074	.420	3.15	-	.003	.265	1.41	-	.002	.043	1.24	-
GARCH: $\theta_3^0 = .6$ and $\epsilon_t \sim N(0,1)$: $\kappa_y = 4.1$												
QMFTTL	-.063	.424	4.02	.014	-.037	.255	2.67	.006	-.019	.057	2.42	.001
MFTTM	.045	.227	2.06	.020	.031	.152	2.26	.002	-.013	.041	2.37	.001
QML	-.098	.425	3.44	-	-.024	.268	2.07	-	-.012	.045	2.76	-
Log-LAD	-.016	.477	5.33	-	-.063	.356	3.56	-	-.280	.079	2.80	-
GJR-GARCH: $\theta_3^0 = .6$ and $\epsilon_t \sim \bar{P}_{2.5}$: $\kappa_y = 2.8$												
QMFTTL	.029	.142	2.35	.018	-.003	.060	.882	.002	-.0046	.040	.795	.0005
MFTTM	.031	.212	2.40	.010	-.001	.069	.923	.001	-.0003	.031	.962	.0004
QML	.087	.140	6.25	-	.078	.089	5.91	-	.0789	.068	5.29	-
Log-LAD	-.098	.156	4.87	-	-.005	.050	.892	-	-.004	.038	.799	-
GJR-GARCH: $\theta_3^0 = .6$ and $\epsilon_t \sim N(0,1)$: $\kappa_y = 4.4$												
QMFTTL	-.094	.177	5.03	.017	-.022	.066	2.62	.001	-.009	.042	1.51	.0003
MFTTM	.048	.185	4.17	.017	.009	.059	1.86	.002	.008	.037	1.09	.0003
QML	-.144	.162	8.22	-	-.033	.062	2.81	-	-.016	.039	2.97	-
Log-LAD	-.174	.204	6.28	-	-.036	.097	3.79	-	-.021	.052	2.14	-

a. The square root of the mean squared error.

b. The Kolmogorov-Smirnov statistic divided by the 5% critical value: KS > 1 indicates rejection of normality at the 5% level.

c. The total sample proportion of the QML loss trimmed for QMTTL where % $\rightarrow 0$ as n $\rightarrow \infty$.

d. Tail index of y_t is κ_y.

TABLE 2 : GARCH and GJR-GARCH, t-tests at 5% level[a]

GARCH: $\epsilon_t \sim \bar{P}_{2.5}$: $\kappa_y = 1.5$

	$n = 100$			$n = 800$			$n = 2000$		
	H_0	H_1^1	H_1^2	H_0	H_1^1	H_1^2	H_0	H_1^1	H_1^2
QMFTTL	.051[b]	.725	.989	.049	1.00	1.00	.060	.997	1.00
MFTTM	.045	.533	.937	.054	.906	.998	.045	1.00	1.00
QML	.061	.796	.993	.064	1.00	1.00	.069	1.00	1.00
Log-LAD	.079	.298	.829	.035	1.00	1.00	.056	.996	1.00

GARCH: $\epsilon_t \sim N(0,1)$: $\kappa_y = 4.1$

	$n = 100$			$n = 800$			$n = 2000$		
	H_0	H_1^1	H_1^2	H_0	H_1^1	H_1^2	H_0	H_1^1	H_1^2
QMFTTL	.058	.167	.622	.052	.985	1.00	.053	.999	1.00
MFTTM	.035	.424	.768	.061	.724	.975	.065	.987	.998
QML	.045	.058	.414	.062	.976	1.00	.053	.998	1.00
Log-LAD	.034	.032	.258	.067	.576	.957	.055	1.00	1.00

GJR-GARCH: $\epsilon_t \sim \bar{P}_{2.5}$: $\kappa_y = 1.5$

	$n = 100$			$n = 800$			$n = 2000$		
	H_0	H_1^1	H_1^2	H_0	H_1^1	H_1^2	H_0	H_1^1	H_1^2
QMFTTL	.054	.768	.988	.048	.998	1.00	.048	1.00	1.00
MFTTM	.041	.467	.851	.046	.997	1.00	.048	1.00	1.00
QML	.052	.803	.990	.070	.998	1.00	.050	1.00	1.00
Log-LAD	.070	.246	.834	.047	1.00	1.00	.048	1.00	1.00

GJR-GARCH: $\epsilon_t \sim N(0,1)$: $\kappa_y = 4.1$

	$n = 100$			$n = 800$			$n = 2000$		
	H_0	H_1^1	H_1^2	H_0	H_1^1	H_1^2	H_0	H_1^1	H_1^2
QMFTTL	.059	.572	.904	.059	.996	1.00	.050	1.00	1.00
MFTTM	.042	.590	.923	.060	.984	1.00	.052	1.00	1.00
QML	.058	.057	.616	.063	.987	1.00	.046	1.00	1.00
Log-LAD	.055	.030	.360	.061	.879	.996	.048	1.00	1.00

a. The hypotheses are H_0: $\theta_3 = \theta_3^0$; H_1^1: $\theta_3 = \theta_3^0 - .25$; and H_1^2: $\theta_3 = 0$.
b. Rejection frequencies at the 5% level.

TABLE 3 : GARCH Model[a] for Returns Data

QMFTTL							
	ω	α	β	|LogLik|[b]	Tr%	Sym(y)	Sym(ϵ_t)
NASDAQ	.028 (.008)[c]	.085 (.017)	.887 (.018)	.6107	.017	.473 [.103][d]	.495 [.413]
HSI	.023 (.010)	.063 (.012)	.913 (.015)	.8355	.025	.513 [.250]	.517 [.193]
FTSE	.019 (.006)	.085 (.015)	.875 (.020)	.3552	.034	.524 [.150]	.520 [.158]

MFTTM							
	ω	α	β	|LogLik|	Tr%	Sym(y)	Sym(ϵ_t)
NASDAQ	.025 (.007)	.159 (.023)	.859 (.017)	.6881	.016	.473 [.103]	.501 [.488]
HSI	.020 (.009)	.078 (.013)	.915 (.012)	.8323	.033	.513 [.250]	.517 [.193]
FTSE	.025 (.004)	.119 (.009)	.922 (.005)	.4742	.040	.524 [.150]	.528 [.113]

a. The model is $\sigma_t^2 = \omega^0 + \alpha^0 y_{t-1}^2 + \beta^0 \sigma_{t-1}^2$.

b. The absolute value of the QML log-likelihood: smaller values indicate a better fit.

c. Standard errors are in parentheses $(\cdot)$.

d. The sample probability of exceeding zero, with the p-value in brackets $[\cdot]$.

TABLE 4 : GJR-GARCH[a] Model for Returns Data

QMFTTL								
	ω	α_1	α_2	β	|LogLik|[b]	Tr%	Sym(y)	Sym(ϵ_t)
NASDAQ	.009 (.005)[c]	.017 (.020)	.125 (.037)	.907 (.020)	.5952	.0026	.473 [.103][d]	.501 [.488]
HSI	.024 (.013)	.058 (.030)	.038 (.136)	.894 (.118)	.8168	.0026	.513 [.250]	.517 [.193]
FTSE	.010 (.004)	.000 (.018)	.079 (.023)	.937 (.017)	.3320	.0023	.524 [.150]	.517 [.226]

MFTTM								
	ω	α_1	α_2	β	|LogLik|	Tr%	Sym(y)	Sym(ϵ_t)
NASDAQ	.028 (.008)	.044 (.022)	.197 (.038)	.898 (.013)	.6579	.0026	.473 [.103]	.498 [.463]
HSI	.023 (.011)	.178 (.049)	.281 (1.43)	.739 (.132)	.7911	.0025	.513 [.250]	.512 [.323]
FTSE	.132 (.035)	.179 (.056)	.188 (.075)	.598 (.070)	.3684	.0027	.524 [.150]	.512 [.323]

a. The model is $\sigma_t^2 = \omega^0 + \alpha_1^0 y_{t-1}^2 + \alpha_2^0 y_{t-1}^2 I(y_t < 0) + \beta^0 \sigma_{t-1}^2$.

b. The absolute value of the QML log-likelihood: smaller values indicate a better fit.

c. Standard errors are in parentheses $(\cdot)$.

d. The sample probability of exceeding zero, with the p-value in brackets $[\cdot]$.