Variance Targeting for Heavy Tailed Time Series

Jonathan B. Hill[1] Eric Renault[2]

University of North Carolina Brown University

(preliminary and incomplete)

May 30, 2012

ABSTRACT

The estimation of GARCH models by QML with variance targeting requires at least a finite unconditional fourth moment in the observed data y_t to ensure Gaussian asymptotics. However, many financial returns series may not have a fourth moment. We robustify the method against heavy tails by exploiting new tail-trimming techniques for both the first step variance estimator *and* the second step criterion. We propose two estimators, the first of which is based on the QML loss when the error $E[\epsilon_t^4] < \infty$ yet $E[y_t^4] = \infty$ is allowed. The second imbeds QML estimating equations with over identifying conditions into GMM criteria allowing for $E[\epsilon_t^4] = \infty$. Both estimators are consistent and asymptotically normal, where different rates of convergence and asymptotic scales arise if $E[y_t^4] = \infty$ and $E[\epsilon_t^4] < \infty$, or $E[\epsilon_t^4] = \infty$. A Monte Carlo study reveals the merits of the new estimators.

1. INTRODUCTION Variance targeting has evolved as a means to reduce the complexity of estimating GARCH models by QML by using a first stage variance plug-in (Engle and Mezrich 1996). By construction a finite variance is assumed on the GARCH process itself, however a finite fourth moment is required for asymptotics (Francq, Horváth and Zakoïan 2010), a severe limitation in lieu of evidence for heavy tails in financial markets (e.g. Embrechts et al 1997, Finkenstädt and Rootzén 2001, Davis 2010, Hill 2011). We alleviate the

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

Key words and phrases: Variance targeting, GARCH, tail trimming, GMM, Value-at-Risk.
AMS subject classifications. Primary 62F35; secondary 62F07.

need to have higher moments by employing tail-trimming techniques recently advanced by Hill and Renault (2010) and Hill (2011).

In the simplest context consider the work-horse GARCH(1,1) model by Bollerslev (1986):

$$y_t = h_t \epsilon_t \text{ where } \epsilon_t \overset{iid}{\sim} (0,1) \tag{1}$$
$$h_t^2 = \omega^0 + \alpha^0 y_{t-1}^2 + \beta^0 h_{t-1}^2 \text{ where } \alpha^0 + \beta^0 < 1.$$

Note $\alpha^0 + \beta^0 < 1$ ensures $E[y_t^2] < \infty$ to justify variance targeting, so define the long-run variance

$$\gamma^0 := E\left[y_t^2\right] = \frac{\omega^0}{1-\alpha^0-\beta^0} := \frac{\omega^0}{\eta^0} < \infty.$$

Variance targeting for QML [Q-VT] exploits a first stage plug-in for the long-run variance γ^0 (Engle and Mezrich 1996):

$$\hat{\gamma}_T := \frac{1}{T}\sum_{t=1}^{T} y_t^2.$$

Conventional asymptotic theory requires $1/T^{1/2}\sum_{t=1}^{T}\{y_t^2 - \gamma^0\}$ to be asymptotically normal, which requires at least a finite fourth moment in the GARCH variable itself $E[y_t^4] < \infty$, let alone in the error $E[\epsilon_t^4] < \infty$. See Francq, Horváth and Zakoïan [FHZ] (2010) for a complete asymptotic theory. Existence of $E[y_t^4]$ allows use of autocorrelations of y_t^2 to check GARCH model fit (e.g. He and Teräsvirta 1999, Kristensen and Linton 2004).

In practice if the higher moments exist then Q-VT allows for consistent estimation of γ^0 in mis-specified GARCH models. Evidence for the non-existence of higher moments in financial, macroeconomic and actuarial time series is, however, substantial. Text book treatments include Embrechts et al (1997) and Finkenstädt and Rootzén (2001), and see Davis (2010). The issue is particularly relevant for those time series with random volatility effects: higher moments must fail to exist if GARCH effects arise (Kesten 1971, Basrak et al 2001).

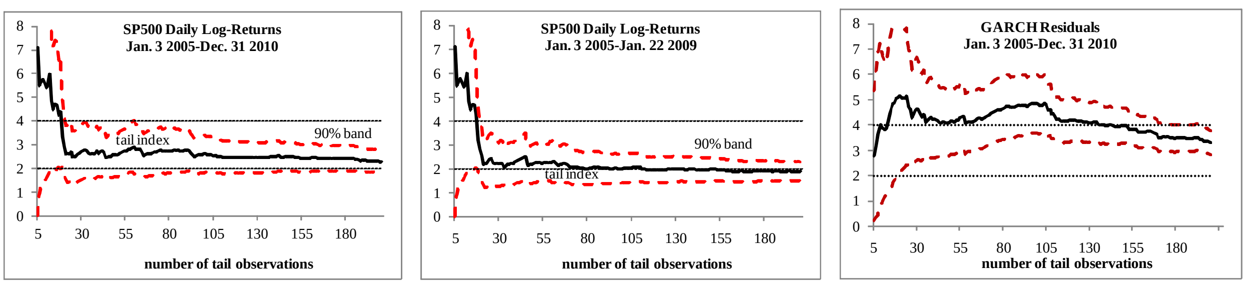
The concern than $E[y_t^2] < \infty$ yet $E[y_t^4] = \infty$ is hardly arcane. In Figure 1 we plot the celebrated Hill (1975) tail index estimator for daily absolute log-returns of the SP500 over the period Jan. 3, 2005 to Dec. 31, 2010[3].

[3] We use the daily open/close average x_t from yahoo.finance.com, and compute $y_t := \ln(x_t) - \ln(x_{t-1})$. There are 1511 trading days net of market closures and holidays.

The plot includes nonparametric 90% confidence bands developed in Hill (2010), and is over the number of included tail observations $\{5, ..., 200\}$. Recall the tail index $\kappa > 0$ is the moment supremum: for Paretian data $E|y_t|^p < \infty$ *if and only if* $p < \kappa$ (Resnick 1987). There is overwhelming evidence for $E[y_t^4] = \infty$: we almost always fail to reject the one sided hypothesis $\kappa \leq 4$ at the 5% level. Even heavier tails are apparent if we use the period Jan. 3, 2005 to Jan. 22, 2009 studied in FHZ (2010), hence FHZ's (2010) assumption of Gaussian asymptotics for QML-VT may be wrong. In fact, the QML residual $|\hat{\epsilon}_t|$ for a GARCH(1,1) model has an index evidently near 4, suggesting a fairly radical departure from the assumption $E[y_t^4] < \infty$.

FIGURE 1

Hill Plots for SP500 Daily Log-Returns



In this paper we escape either restriction $E[y_t^4] < \infty$ or $E[\epsilon_t^4] < \infty$, allow for variance targeting with Gaussian asymptotics, and permit over-identifying restrictions. We achieve this by using a tail-trimmed variance estimator as the variance target plug-in. Under the assumption $E[\epsilon_t^4] < \infty$ as in FHZ (2010), but allowing for $E[y_t^4] = \infty$, we show the QMLE is consistent and asymptotically normal. This alone augments all known results in the variance targeting literature, in particular our estimator of γ^0 is consistent and asymptotically normal for mis-specified models, requiring only $E[y_t^2] < \infty$ and a mixing condition. A tail-trimmed variance, however, must be biased in small samples. We therefore correct the bias by improving on methods developed in Peng (2002).

If the error has an infinite fourth moment $E[\epsilon_t^4] = \infty$ then the QMLE has a non-Gaussian limit, cf. Hall and Yao (2003) and Linton et al (2010). One solution is to tail-trim the QML criterion equations as in Hill (2011), but under tail-trimming there is no present theory supporting asymptotic efficiency relative to GMM (cf. Meddahi and Renault 1998). We therefore exploit and improve upon Hill and Renault's [HR] (2010) robust GMM framework by imbedding the tail-trimmed variance plug-in with QML estimating equations and over identifying restrictions into the Generalized Method of Tail-Trimmed Moments setting. Since GMM with variance targeting has only recently been considered we also treat the non-trimmed case (cf. Prono 2011).

HR's (2010) framework, however, is general while here we work exclusively with QML-type equations, e.g. $(\epsilon_t^2 - 1)x_t$ where x_t contains at least the usual QML weights. We therefore streamline HR's (2010) method by focusing trimming on the precise source of extremes ϵ_t or x_t, as in Hill (2011). As an improvement over both HR (2010) and Hill (2011) we re-center the tail-trimmed errors as in Hill and Aguilar (2010), allowing for an arbitrary choice of how much to trim without affecting small sample bias.

In the case $E[y_t^4] < \infty$ our estimators reduce asymptotically to QML or GMM with variance targeting. This greatly simplifies inference since the analyst does not need to know if tails are heavy. If $E[y_t^4] = \infty$ then we characterize the unique direction along the space of GARCH parameters that leads to $T^{1/2}$-convergence. In any other direct simple rules of thumb for trimming allow the rate to reach $T^{1/2}/L(T)$ for slowly varying $L(T)$ $\rightarrow \infty$. Indeed, if $E[y_t^4] = \infty$ then the optimal amount can always be set to diminish the asymptotic covariance offering arbitrary efficiency improvements, and rendering a rate greater than QML.

We construct the two estimators in Section 2, and present assumptions and limit theory in Section 3. A simulation study is presented in Section 4 with parting comments left for Section 5.

We use the following notation. Partial derivatives evaluated at point are $(\partial/\partial\xi)A(\xi_*) = (\partial/\partial\xi)A(\xi)|_{\xi_*}$. Let $\iota > 0$ be a tiny number whose value may change with the context. $K > 0$ is an arbitrary finite number.

2. TAIL-TRIMMED VARIANCE TARGETING

Assume ϵ_t is symmetrically distributed for ease of discussion, while the following easily extends to the case of asymmetric errors with asymmetric trimming (e.g. HR 2010, Hill 2011). The sample is assumed to be $\{y_0, y_1, ..., y_T\}$ for notational convenience since we will condition on y_0. Define the σ-field induced by the data: $\Im_t := \sigma(y_\tau : \tau \leq t)$.

2.1 TAIL TRIMMED VARIANCES

We first tackle $\hat{\gamma}_T$ by constructing a tail-trimmed version. Define $y_t^{(a)} := |y_t|$ and its order statistics $y_{(1)}^{(a)} \geq y_{(2)}^{(a)} \geq \cdots \geq y_{(T)}^{(a)}$, and let $\{k_T^{(y)}\}$ be an intermediate order sequence: $k_T^{(y)} \rightarrow \infty$ and $k_T^{(y)}/T \rightarrow 0$. See Leadbetter et al (1983) and Hahn et al (1991). The most obvious tail-trimmed sample variance is

$$\hat{\gamma}_T^{(tr)} := \frac{1}{T}\sum_{t=1}^{T} y_t^2 \hat{I}_{T,t}^{(y)} \quad \text{where} \quad \hat{I}_{T,t}^{(y)} := I\left(|y_t| \leq y_{(k_T^{(y)})}^{(a)}\right). \tag{2}$$

Trimming $k_T^{(y)} \to \infty$ promotes Gaussian asymptotics, while negligibility $k_T^{(y)}/T \to 0$ ensures identification $\hat{\gamma}_T^{(tr)} \xrightarrow{p} \gamma^0$ (Hill and Aguilar 2010, HR 2010, Hill 2011). Since $\alpha^0 + \beta^0 < 1$ ensures $E|y_t|^{2+\iota} < \infty$ for tiny $\iota > 0$, the weak limit $\hat{\gamma}_T^{(tr)} \xrightarrow{p} \gamma^0$ can be shown to follow from Theorem 2.2 in Andrews (1988) and consistency of intermediate order statistics $y_{(k_T^{(y)})}^{(a)}$ (Hill 2010). See Section 4 for all regularity conditions and Appendix A for asymptotic theory details.

In small samples, however, $\hat{\gamma}_T^{(tr)}$ is biased by construction since by trimming positive random variables $\hat{\gamma}_T^{(tr)} = \hat{\gamma}_T - 1/T\sum_{t=1}^{T} y_t^2 I(|y_t| > y_{(k_T^{(y)})}^{(a)}) < \hat{\gamma}_T$. In simulations not reported in this paper we find the bias can be substantial, hence a bias correction is needed. We fill the trimmed gap as in Peng (2001) by using an estimator that substitutes for $1/T\sum_{t=1}^{T} y_t^2 I(|y_t| > y_{(k_T^{(y)})}^{(a)})$ yet is asymptotically normal. Contrary to Peng (2001), however, our gap estimator does not influence the limit distribution.

In order to understand Peng's (2001) method, notice $y_{(k_T^{(y)})}^{(a)}$ estimates non-random $k_T^{(y)}/T$-upper quantiles $c_T^{(y)}$ defined by (cf. Leadbetter et al 1983)

$$P\left(|y_t| > c_T^{(y)}\right) = \frac{k_T^{(y)}}{T}.$$

Under a mixing condition the order statistic is consistent $y_{(k_T^{(y)})}^{(a)}/c_T^{(y)} = 1 + O_p(1/(k_T^{(y)})^{1/2})$. See Hill (2010). By construction

$$E\left[y_t^2\right] = E\left[y_t^2 I\left(|y_t| \leq c_T^{(y)}\right)\right] + E\left[y_t^2 I\left(|y_t| > c_T^{(y)}\right)\right] = \gamma_T^{(tr)} + \mathcal{R}_T,$$

and by properties of power-law tails it is straightforward to show (e.g. Peng 2001: p. 257)

$$\mathcal{R}_T \sim \frac{\kappa}{\kappa - 2}\frac{k_T^{(y)}}{T}\left(c_T^{(y)}\right)^2.$$

Each component is easily estimated. We use $y^{(a)}_{(k^{(y)}_T)}$ for $c^{(y)}_T$ and, for example, Hill's (1975) tail index estimator

$$\hat{\kappa}_{k^{(y)}_T} = \left(\frac{1}{k^{(y)}_T} \sum_{i=1}^{k^{(y)}_T - 1} \ln \left(y^{(a)}_{(i)} / y^{(a)}_{(k^{(y)}_T)} \right) \right)^{-1}, \tag{3}$$

hence Peng's proposed plug-in

$$\hat{\mathcal{R}}_T = \frac{\hat{\kappa}_{k^{(y)}_T}}{\hat{\kappa}_{k^{(y)}_T} - 2} \frac{k^{(y)}_T}{T} \left(y^{(a)}_{(k^{(y)}_T)} \right)^2$$

is asymptotically normal under our assumptions below.

Peng (2001) uses the same fractile $k^{(y)}_T$ everywhere to compute $\hat{\gamma}^{(tr)}_T + \hat{\mathcal{R}}_T$, leading to a complicated limit theory because $\hat{\mathcal{R}}_T$ and $\hat{\gamma}^{(tr)}_T$ are asymptotically dependent. In practice, however, very little trimming is required to promote an approximately normal estimator, while little trimming is always optimal to reduce incidental bias caused by trimming (cf. HR 2010). Conversely, in general far more tail observations are required for a sharp estimate of κ, in particular if the data are generated by a random volatility process. See Section 4 for a plot of $\hat{\kappa}_{k^{(y)}_T}$ for GARCH(1,1).

We show in Appendix A that by over-correcting for the gap and augmenting the number of tail observations used to estimate κ, we obtain an asymptotically unbiased and normal estimator with the same and easily estimable normalizing scale as $\hat{\gamma}^{(tr)}_T$. This is achieved by estimating a larger gap with a length determined by an intermediate order sequence $\{\tilde{k}^{(y)}_T\}$ that satisfies

$$\frac{k^{(y)}_T}{\tilde{k}^{(y)}_T} \to 0 \text{ and } \frac{\tilde{k}^{(y)}_T}{T} \to 0, \tag{4}$$

with an associated threshold sequence $\tilde{c}^{(y)}_T$ defined by

$$P\left(|y_t| > \tilde{c}^{(y)}_T \right) = \frac{\tilde{k}^{(y)}_T}{T},$$

and by estimating κ using an intermediate order sequence $\{\check{k}_T^{(y)}\}$ that satisfies

$$\frac{\tilde{k}_T^{(y)}}{\check{k}_T^{(y)}} \rightarrow 0 \text{ and } \frac{\check{k}_T^{(y)}}{T} \rightarrow 0. \tag{5}$$

Our tail-trimmed variance estimator with a negligible *gap* estimator is then

$$\hat{\gamma}_T^{(g)} := \frac{1}{T}\sum_{t=1}^{T} y_t^2 I\left(|y_t| \leq y_{(k_T^{(y)})}^{(a)}\right) + \left(\frac{\hat{\kappa}_{\check{k}_T^{(y)}}}{\hat{\kappa}_{\check{k}_T^{(y)}} - 2}\right)\frac{\tilde{k}_T^{(y)}}{T}\left(y_{(\tilde{k}_T^{(y)})}^{(a)}\right)^2 = \hat{\gamma}_T^{(tr)} + \hat{\mathcal{R}}_T^{(g)}. \tag{6}$$

Notice (4) implies an over-correction for the tail-trimmed gap since $c_T^{(y)}/\tilde{c}_T^{(y)} \rightarrow \infty$ implies asymptotically the gap $[c_T^{(y)}, \infty) \subset [\tilde{c}_T^{(y)}, \infty)$. A comparatively faster $\tilde{k}_T^{(y)}$ promotes faster convergence for $y_{(\tilde{k}_T^{(y)})}^{(a)}$, hence $\hat{\mathcal{R}}_T^{(g)} \xrightarrow{p} 0$ rapidly enough that it does not affect $\hat{\gamma}_T^{(tr)}$ asymptotically. Similarly, an even faster $\check{k}_T^{(y)}$ ensures $\hat{\kappa}_{\check{k}_T^{(y)}}$ does not impact $\hat{\gamma}_T^{(tr)}$.

Clearly (4) and (5) hold for infinitely many sequences $\{k_T^{(y)}, \tilde{k}_T^{(y)}, \check{k}_T^{(y)}\}$: for any $k_T^{(y)}$ just define $\tilde{k}_T^{(y)} \sim (k_T^{(y)})^{1+\iota}$ and $\check{k}_T^{(y)} \sim (k_T^{(y)})^{1+2\iota}$ for any tiny $\iota > 0$. In practice the analyst may want to select $\{\tilde{k}_T^{(y)}, \check{k}_T^{(y)}\}$ such that $\hat{\mathcal{R}}_T^{(g)}$ optimally fills the gap according to some criterion. A natural criterion is to choose $\{\tilde{k}_T^{(y)}, \check{k}_T^{(y)}\}$ such that $\hat{\gamma}_T^{(tr)} + \hat{\mathcal{R}}_T^{(g)}$ is closest to the untrimmed variance $\hat{\gamma}_T$. Since $\hat{\mathcal{R}}_T^{(g)}$ can be assured to be negligible sufficiently fast, we still retain an asymptotically normal tail-trimmed variance estimator.

A convenient solution is to use define intermediate order sequences $\{k_T^{(y)}, \tilde{g}_T^{(y)}, \check{g}_T^{(y)}\}$ that satisfy (4) and (5): $k_T^{(y)}/\tilde{g}_T^{(y)} \rightarrow 0$ and $\tilde{g}_T^{(y)}/\check{g}_T^{(y)} \rightarrow 0$, and define

$$\tilde{k}_T^{(y)} = \tilde{k}_T^{(y)}(\tilde{\lambda}) = \left[\tilde{\lambda}\tilde{g}_T^{(y)}\right] \text{ and } \check{k}_T^{(y)} = \check{k}_T^{(y)}(\check{\lambda}) = \left[\check{\lambda}\check{g}_T^{(y)}\right] \text{ where } \tilde{\lambda}, \check{\lambda} \in \Lambda \text{ for compact } \Lambda \subset (0,\infty)^2. \tag{7}$$

Now solve

$$\left[\tilde{\lambda}^o, \check{\lambda}^o\right] = \underset{\tilde{\lambda},\check{\lambda}\in\Lambda}{\text{arginf}}\left|\hat{\gamma}_T^{(tr)} + \left(\frac{\hat{\kappa}_{\check{k}_T^{(y)}(\check{\lambda})}}{\hat{\kappa}_{\check{k}_T^{(y)}(\check{\lambda})} - 2}\right)\frac{\tilde{k}_T^{(y)}(\tilde{\lambda})}{T}\left(y_{(\tilde{k}_T^{(y)}(\tilde{\lambda}))}^{(a)}\right)^2 - \hat{\gamma}_T\right|$$

and define the tail-trimmed variance estimator with an *optimal* gap estimator

$$\hat{\gamma}_T^{(o)} := \hat{\gamma}_T^{(tr)} + \left(\frac{\hat{\kappa}_{\check{k}_T^{(y)}(\check{\lambda}^o)}}{\hat{\kappa}_{\check{k}_T^{(y)}(\check{\lambda}^o)} - 2}\right)\frac{\tilde{k}_T^{(y)}(\tilde{\lambda}^o)}{T}\left(y_{(\tilde{k}_T^{(y)}(\tilde{\lambda}^o))}^{(a)}\right)^2 = \hat{\gamma}_T^{(tr)} + \hat{\mathcal{R}}_T^{(o)}.$$

2.2 QML WITH TAIL-TRIMMED VARIANCE TARGETING

Throughout $\hat{\gamma}_T^*$ denotes either estimator $\hat{\gamma}_T^{(g)}$ or $\hat{\gamma}_T^{(o)}$:

$$\hat{\gamma}_T^* \in \left\{ \hat{\gamma}_T^{(g)}, \hat{\gamma}_T^{(o)} \right\}.$$

Define parameters from the GARCH model, and the transformed model for variance targeting

$$\theta = [\omega, \alpha, \beta]' \in \Theta \text{ for compact } \Theta \subset \mathbb{R}^3$$

$$\eta = 1 - \alpha - \beta$$

$$\xi = [\alpha, \eta]' \in \Xi \text{ for compact } \Xi \subset (0, 1), \quad \text{and} \quad \varphi = \left[\gamma, \xi'\right]' \in \mathbb{R}_+ \times \Xi.$$

We require iterated conditional variance sequences for estimation since y_t for $t < 0$ is not observed. Define

$$\tilde{\sigma}_0^2(\varphi) = \eta\gamma \text{ and } \tilde{\sigma}_t^2(\gamma, \xi) = \tilde{\sigma}_t^2(\varphi) := \eta\gamma + \alpha y_{t-1}^2 + (1 - \alpha - \eta)\, \tilde{\sigma}_{t-1}^2(\varphi) \text{ for } t = 1, 2, ..., \tag{8}$$

$$\sigma_{T,0}^2(\xi) = \eta\hat{\gamma} \text{ and } \sigma_{T,t}^2(\xi) := \eta\hat{\gamma}_T^* + \alpha y_{t-1}^2 + (1 - \alpha - \eta)\, \sigma_{T,t-1}^2(\xi) \text{ for } t = 1, 2, ... \tag{9}$$

By construction

$$\tilde{\sigma}_t^2(\hat{\gamma}_T^*, \xi) = \sigma_{T,t}^2(\xi).$$

We now drop φ^0 throughout: $\sigma_{T,t}^2 = \sigma_{T,t}^2(\xi^0)$, $\tilde{\sigma}_t^2 = \tilde{\sigma}_t^2(\varphi^0)$. Since $0 \leq 1 - \alpha - \eta = \beta < 1$ we can define a stationary ergodic process (Francq and Zakoïan 2004)

$$\sigma_t^2(\gamma, \xi) = \sigma_t^2(\varphi) := \eta\gamma + \alpha y_{t-1}^2 + (1 - \alpha - \eta)\, \sigma_{t-1}^2(\varphi) \text{ hence } \sigma_t^2(\varphi^0) = h_t^2.$$

Similarly, define error and volatility score functions

$$e_{T,t}(\xi) := \frac{y_t}{\sigma_{T,t}(\xi)} \text{ and } s_{T,t}(\xi) := \frac{\partial}{\partial \xi} \ln \sigma_{T,t}^2(\xi)$$

Note $\hat{\gamma}_T^* \xrightarrow{p} \gamma^0$ implies $\sigma_{T,t}^2(\xi)$ and $\sigma_t^2(\gamma^0, \xi)$ are close in the uniform geometric sense $\sup_{\xi \in \Xi} |\sigma_{T,t}^2(\xi) - \sigma_t^2(\gamma^0, \xi)|$ $= o_p(\rho^t)$ for $\rho \in (0,1)$ by FHZ's (2010: Appendix A.1) argument. We therefore simply use $\sigma_t^2(\varphi)$ in the following to ease notation, and it is always understood that functions of $\sigma_{T,t}^2(\xi)$ have versions based on $\sigma_t^2(\varphi)$: e.g. $e_t(\varphi) = y_t/\sigma_t(\varphi)$ and $s_t(\varphi) := (\partial/\partial\xi) \ln \sigma_t^2(\gamma, \xi)$. By construction, therefore,

$$e_t(\varphi^0) = \frac{y_t}{\sigma_t(\varphi^0)} = \frac{y_t}{h_t} = \epsilon_t.$$

In order to understand where heavy tails affect Q-VT asymptotics, and therefore where tail-trimming must be applied, consider the usual first order expansion. Define

$$\left\{e_t^2(\varphi) - 1\right\} \times s_t(\varphi) := \left\{\frac{y_t^2}{\sigma_t^2(\varphi)} - 1\right\} \times \frac{1}{\sigma_t^2(\varphi)} \frac{\partial}{\partial \xi} \sigma_t^2(\varphi) \in \mathbb{R}^2$$
$$\widehat{\mathcal{J}}_T^{(a,b)}(\varphi) = \frac{1}{T} \sum_{t=1}^{T} \frac{1}{\sigma_t^4(\varphi)} \frac{\partial}{\partial a} \sigma_t^2(\varphi) \frac{\partial}{\partial b'} \sigma_t^2(\varphi) \text{ for any } a, b \in \{\xi, \gamma, \varphi\}.$$

The Q-VT first order condition is

$$\frac{1}{T} \sum_{t=1}^{T} \left\{e_t^2(\hat{\gamma}_T, \hat{\xi}_T) - 1\right\} \times s_t(\hat{\gamma}_T, \hat{\xi}_T) = 0, \tag{10}$$

and under regularity conditions an application of the mean-value-theorem leads to (see Appendix A.2 in FHZ 2010)

$$0 = \frac{1}{T} \sum_{t=1}^{T} \left\{\epsilon_t^2 - 1\right\} \times s_t - \widehat{\mathcal{J}}_T^{(\xi,\xi)} \times \left(\hat{\xi}_T - \xi^0\right) - \widehat{\mathcal{J}}_T^{(\xi,\gamma)} \times \left(\hat{\gamma}_T - \gamma^0\right) + o_p(1) \tag{11}$$

Gaussian asymptotics for Q-VT therefore requires $(\epsilon_t^2 - 1)s_{i,t}$ and y_t^2 to have finite variances, and the outer product averages $\widehat{\mathcal{J}}_T^{(a,b)}$ to have probability limits. If there are GARCH effects $\alpha^0 + \beta^0 > 0$ then the volatility score s_t is $L_{2+\iota}$-bounded (Francq and Zakoïan 2004, 2010) in which case the latter limit is easily satisfied and heavy tails arise only due to ϵ_t. If there are no GARCH effects $\alpha^0 + \beta^0 = 0$ then $s_t = [0, y_{t-1}^2/\gamma^0 - 1]$ is heavy tailed if $E[y_t^4] = \infty$.

There are several approaches to ensure $1/T \sum_{t=1}^{T} \{\epsilon_t^2 - 1\} s_t$ and $\widehat{\mathcal{J}}_T^{(a,b)}$ are robust. The simplest starts with the universal assumption in the GARCH variance targeting literature $E[\epsilon_t^4] < \infty$. Notice $E[\epsilon_t^4] < \infty$ and

$E[y_t^4] = \infty$ can only occur if there are GARCH effects, in which case only the untrimmed $\hat{\gamma}_T$ fails to have a standard limit once standardized: both $1/T\sum_{t=1}^{T}\{\epsilon_t^2 - 1\}s_t$ and $\widehat{\mathcal{J}}_T^{(a,b)}$ satisfy standard asymptotic theory. We can simply use a QML criterion with Tail-Trimmed Variance Targeting [Q-TTVT]:

$$\text{Q-TTVT} \;:\; \hat{Q}_T(\xi) := \sum_{t=1}^{T}\left(\ln\sigma_{T,t}^2(\xi) - \frac{y_t^2}{\sigma_{T,t}^2(\xi)}\right) = \sum_{t=1}^{T}\left(\ln\sigma_t^2(\hat{\gamma}_T^*,\xi) - \frac{y_t^2}{\sigma_t^2(\hat{\gamma}_T^*,\xi)}\right), \tag{12}$$

and the Q-TTVT estimator solves $\hat{\xi}_T = \arg\inf_{\xi\in\Xi}\{\hat{Q}_T(\xi)\}$. This alone extends all existing variance targeting treatments with QML to allow $E[y_t^4] = \infty$, where heavy tails are aligned with parameter values $\{\alpha^0, \beta^0\}$ as opposed to the error ϵ_t.

2.3 GMTTM WITH TAIL-TRIMMED VARIANCE TARGETING

If we allow $E[\epsilon_t^4] = \infty$ then we must use information on extremes in ϵ_t for trimming. Hill (2011) tail–trims the QML criterion equations which here would translate to trimming $\ln\sigma_{T,t}^2(\xi) - y_t^2/\sigma_{T,t}^2(\xi)$ according to large values of ϵ_t. Under tail-trimming, however, it is unknown whether QML is as efficient as GMM with optimal weights (see, e.g., Meddahi and Renault 1998). We therefore adopt and refine the GMTTM method in HR (2010) for variance targeting.

In principle any set of instruments may be entertained, but we naturally exploit information from the volatility score $s_t(\varphi)$ since it is uniformly square integrable if there are GARCH effects. Therefore define the stationary $\Im_{t-1}$-measurable instrument set

$$x_t : \mathbb{R}_+ \times \Xi \to \mathbb{R}^q \text{ where } q \geq 2, \text{ where } [x_{i,t}(\varphi)]_{i=1}^2 = s_t(\varphi) = \frac{1}{\sigma_t^2(\varphi)}\frac{\partial}{\partial\xi}\sigma_t^2(\varphi)$$

and its version with the variance plug-in

$$x_{T,t}(\xi) = \left[s_{T,t}(\xi)',\;\; [x_{i,T,t}(\xi)]_{i=3}^{q\prime}\right]' \in \mathbb{R}^q \text{ where } q \geq 3,$$

and define estimating equations

$$m_{T,t}(\xi) := \left(e^2_{T,t}(\xi) - 1\right) \times x_{T,t}(\xi) \text{ and } m_t(\varphi) := \left(e^2_t(\varphi) - 1\right) \times x_t(\varphi).$$

We assume $x_t(\varphi)$ is continuous and differentiable in φ, thus $m_t(\varphi)$ is continuous, differentiable, and a martingale difference since $e_t(\varphi^0) = \epsilon_t$ is iid and x_t is $\Im_{t-1}$-measurable

$$E[m_t|\Im_{t-1}] = 0.$$

HR (2010) negligibly trim estimating equations in a general environment, and therefore have only the extremes of $m_{i,t}(\varphi)$ itself to gauge when trimming is applied. But here $m_{i,t}(\varphi) = (e^2_t(\varphi) - 1) \times x_{i,t}(\varphi)$, hence we know by a standard expansion and error independence that extremes occur due to $e^2_t(\varphi)$ or $x_{i,t}(\varphi)$. Moreover, by focusing trimming on the source of the extreme in $m_{i,t}(\varphi)$ we may control for artificially introduced bias due to trimming as in Hill and Aguilar (2010).

As before, define $e^{(a)}_t(\varphi) := |e_t(\varphi)|$ with order statistics $e^{(a)}_{(1)}(\varphi) \geq e^{(a)}_{(2)}(\varphi) \ldots$, an intermediate order sequence $\{k^{(\epsilon)}_T\}$, and trimming indicators

$$\hat{I}^{(e)}_{T,t}(\varphi) := I\left(|e_t(\varphi)| \leq e^{(a)}_{(k^{(e)}_T)}(\varphi)\right) \text{ and } \hat{I}^{(e)}_{T,t}(\xi) = \hat{I}^{(e)}_{T,t}(\hat{\gamma}^*_T, \xi) := I\left(|e_{T,t}(\xi)| \leq e^{(a)}_{T,(k^{(e)}_T)}(\xi)\right).$$

Similarly, define $x^{(a)}_{i,t}(\varphi) := |x_{i,t}(\varphi)|$, intermediate order statistics $x^{(a)}_{i,(1)}(\varphi) \geq x^{(a)}_{i,(2)}(\varphi) \ldots$ and indicators

$$\hat{I}^{(x)}_{i,T,t}(\varphi) := I\left(|x_{i,t}(\varphi)| \leq x^{(a)}_{i,(k^{(x)}_{i,T})}(\varphi)\right) \text{ and } \hat{I}^{(x)}_{i,T,t}(\xi) := I\left(|x_{i,T,t}(\xi)| \leq x^{(a)}_{i,T,(k^{(x)}_{i,T})}(\xi)\right)$$

where $\{k_{i,T}^{(x)}\}$ are intermediate order sequences $k_{i,T}^{(x)} \to \infty$ and $k_{i,T}^{(x)} = o(T)$. The tail-trimmed equations are[4]

$$\hat{m}^*_{T,t}(\hat{\gamma}^*_T, \xi) = \hat{m}^*_{T,t}(\xi) = \left(e^2_{T,t}(\xi)\, \hat{I}^{(e)}_{T,t}(\xi) - \frac{1}{T}\sum_{t=1}^{T} e^2_{T,t}(\xi)\, \hat{I}^{(e)}_{T,t}(\xi)\right) \times x_{T,t}(\xi) \prod_{i=1}^{q} \hat{I}^{(x)}_{i,T,t}(\xi) \quad (13)$$

$$= \left(\hat{e}^{*2}_{T,t}(\xi) - \frac{1}{T}\sum_{t=1}^{T} \hat{e}^{*2}_{T,t}(\xi)\right) \times \hat{x}^*_{T,t}(\xi).$$

If $x_{i,t}$ is known to have a finite variance then fix $\hat{I}^{(x)}_{i,t}(\varphi) := 1$. See Section 2.3 for an example. Let $\{\hat{\Upsilon}_T\}$ be a sequence of positive definite, possibly stochastic, weight matrices. The Generalized Method of Tail-Trimmed Moments with Tail-Trimmed Variance Targeting [GTT-TTVT] estimator is

$$\hat{\xi}_T = \underset{\xi \in \Xi}{\text{arginf}} \left\{ \left(\frac{1}{T}\sum_{t=1}^{T} \hat{m}^*_{T,t}(\hat{\gamma}^*_T, \xi)\right)' \times \hat{\Upsilon}_T \times \left(\frac{1}{T}\sum_{t=1}^{T} \hat{m}^*_{T,t}(\hat{\gamma}^*_T, \xi)\right) \right\}.$$

We re-center with the sample mean of $\hat{e}^{*2}_{T,t}(\xi)$ in order to eradicate small sample bias due to trimming. It also has the nice advantage of ensuring the covariance matrix of $\sum_{t=1}^{T} \hat{m}^*_{T,t}$ reduces to $TE[\hat{m}^*_{T,t}\hat{m}^{*\prime}_{T,t}]$ as $T \to \infty$. In the standard GMTTM framework trimming artificially introduces small sample dependence that may not erode fast enough as $T \to \infty$, hence a HAC estimator must be used (see HR 2010). Although in theory and practice it is valid to use the population moment $E[\epsilon_t^2] = 1$ in $\hat{m}^*_{T,t}(\xi) = (\hat{e}^{*2}_{T,t}(\xi) - 1) \times \hat{x}^*_{T,t}(\xi)$, simulation evidence suggests the small sample bias can be substantial when the error tails are very heavy.

We must trim each $x_{i,T,t}(\xi)$ based on extremes in every $x_{i,T,t}(\xi)$ to ensure the sample Jacobian associated with $\hat{m}^*_{T,t}(\xi)$ has a probability limit. HR (2010) show trimming $m_{i,t}(\varphi)$ by its large values suffices to ensure the resulting sample Jacobian converges. In terms of computation, the estimator $\hat{\xi}_T$ is easy to obtain in practice by iterative numerical methods. The benefit of focusing trimming as in (13) arises particularly when over identifying restrictions are based on lags of the $s_{i,t}(\varphi)$, as we discuss below.

[4] The sample size subscript is unavoidable here since trimming induces dependence on T, while the sample plug-in $\hat{\gamma}^*_T$ does as well. We therefore use $*$ to signify trimming.

The untrimmed G-VT estimator is computed with the untrimmed variance plug-in $\hat{\gamma}_T$ to solve

$$\hat{\xi}_T = \underset{\xi \in \Xi}{\text{arginf}} \left\{ \left(\frac{1}{T} \sum_{t=1}^{T} m_t(\hat{\gamma}_T, \xi) \right)' \times \hat{\Upsilon}_T \times \left(\frac{1}{T} \sum_{t=1}^{T} m_t(\hat{\gamma}_T, \xi) \right) \right\}.$$

Note the weight $\hat{\Upsilon}_T$ may be different under non-trimming.

2.4 GMTTM WITH LAGGED OVER IDENTIFYING RESTRICTIONS

Suppose the weights $x_t(\varphi)$ contain only $s_t(\varphi)$ and their lags:

$$x_t(\varphi) = [s_t(\varphi)', s_{t-1}(\varphi)', ..., s_{t-h+1}(\varphi)']' \in \mathbb{R}^q \text{ , } h \geq 1, \text{ where } \; q = 2h.$$

If there are known GARCH effects then each $x_{i,t}(\varphi)$ is uniformly square integrable (Francq and Zakoïan 2004, 2010) so we do not need to trim by them, and we simply use

$$\hat{m}^*_{T,t}(\xi) = \left(\hat{e}^{*2}_{T,t}(\xi) - \frac{1}{T} \sum_{t=1}^{T} \hat{e}^{*2}_{T,t}(\xi) \right) x_{T,t}(\xi).$$

As a special case, if the model is ARCH then $x_t(\varphi)$ is uniformly bounded hence trimming is unnecessary. This is intrinsically easy to implement relative to GMTTM because we have only one fractile value $k_T^{(\epsilon)}$ to choose.

If there are no GARCH effects then it is easy to show $x_{i,t}(\varphi)$ is a constant or a multiple of y^2_{t-i} for $i = 1, ..., h$. Thus, if we allow for the possibility of no GARCH effects then we use trimmed weights $x_{T,t}(\xi) \prod_{i=1}^{h} \hat{I}^{(y)}_{T,t-i}$ for estimation:

$$\hat{m}^*_{T,t}(\xi) = \left(\hat{e}^{*2}_{T,t}(\xi) - \frac{1}{T} \sum_{t=1}^{T} \hat{e}^{*2}_{T,t}(\xi) \right) \times x_{T,t}(\xi) \prod_{i=1}^{h} \hat{I}^{(y)}_{T,t-i} \text{ where } \hat{I}^{(y)}_{T,t} := I\left(|y_t| \leq y^{(a)}_{(k_T^{(y)})} \right).$$

3. ASYMPTOTIC THEORY We present the main asymptotic results for QML and GMM estimators separately. Write

$$\hat{\gamma}^*_T = \hat{\gamma}^{(tr)}_T + \hat{\mathcal{R}}^*_T \text{ where } \hat{\gamma}^*_T, \hat{\mathcal{R}}^*_T \text{ are either } \hat{\gamma}^{(g)}_T, \hat{\mathcal{R}}^{(g)}_T \text{ or } \hat{\gamma}^{(o)}_T, \hat{\mathcal{R}}^{(o)}_T.$$

3.1 MAIN RESULTS FOR Q-TTVT

The asymptotic scales for our estimators require a characterization of the deterministic thresholds which the trimming order statistics $\{y^{(a)}_{(k^{(y)}_T)}, e^{(a)}_{T,(k^{(e)}_T)}(\xi), x^{(a)}_{i,(k^{(x)}_{i,T})}(\xi)\}$ approximate. Write in general

$$w_t(\varphi) \in \{y^{(a)}_t, e^{(a)}_t(\varphi), x^{(a)}_{i,t}(\varphi)\} \text{ and } k^{(w)}_T \in \{k^{(y)}_T, k^{(e)}_T, k^{(x)}_{i,T}\},$$

and define positive two-tailed sequences $c^{(w)}_T(\varphi) \in \{c^{(y)}_T, c^{(\epsilon)}_T(\varphi), c^{(x)}_{i,T}(\varphi)\}$ that satisfy

$$P\left(|w_t(\varphi)| > c^{(w)}_T(\varphi)\right) = \frac{k^{(w)}_T}{T}$$

By construction $k^{(w)}_T/T \to 0$ so $c^{(w)}_T(\varphi) \to \infty$ if $w_t(\varphi)$ has unbounded support (cf. Leadbetter et al 1983). Define

$$w^*_{T,t}(\varphi) := w_t(\varphi) I\left(|w_t(\varphi)| \leq c^{(w)}_T(\varphi)\right) = w_t(\varphi) I^{(w)}_{T,t}(\varphi) \text{ and } w^*_{T,t}(\xi) := w_{T,t}(\xi) I^{(w)}_{T,t}(\xi).$$

Note for some remainder term r_T we have the decomposition (Horváth et al 2006: eq. (37))

$$\hat{\gamma}_T - \gamma^0 = \frac{(1-\beta^0)}{\eta_0}\frac{1}{T}\sum_{t=1}^T h^2_t(\epsilon^2_t - 1) + r_T$$

If $E[y^4_t] < \infty$ then $r_T = o_p(1/T^{1/2})$ and $h^2_t(\epsilon^2_t - 1)$ is a square integrable martingale difference. Under trimming, however, although a similar expansion can be derived for $1/T\sum_{t=1}^T y^2_t I(|y_t| \leq y^{(a)}_{(k^{(y)}_T)})$ we lose the martingale difference property. We therefore work with a long-run variance

$$v^2_{T,\gamma} := E\left(\sum_{t=1}^T \left\{y^2_t I\left(|y_t| \leq c^{(y)}_T\right) - E\left[y^2_t I\left(|y_t| \leq c^{(y)}_T\right)\right]\right\}\right)^2.$$

We require the following assumptions.

A1. $\epsilon_t \overset{iid}{\sim} (0,1)$ *has an absolutely continuous distribution.*

A2. $\gamma^0 = E[y^2_t] < \infty$ *(hence* $\alpha^0 + \beta^0 < 1$*).*

A3. ξ^0 *is an interior point of compact* $\Xi \subset \mathbb{R}^2$.

A4. $Tv_{T,\gamma}^{-1}E[y_t^2 I(\tilde{c}_T^{(y)} \leq |y_t| \leq c_T^{(y)})] \rightarrow 0.$

Remark 1: Assumptions A1 and A2 ensure $\{y_t\}$ is geometrically β-mixing and has regularly varying distribution tails (Basrak et al 202, Francq and Zakoïan 2006).

Remark 2: Assumption A4 is the sole requirement concerns trimming. By an asymptotic expansion the Q-TTVT estimator reduces to a linear function of

$$v_{T,\gamma}^{-1}\sum_{t=1}^{T}\left\{y_t^2 I\left(|y_t| \leq c_T^{(y)}\right) + E\left[y_t^2 I\left(|y_t| \geq \tilde{c}_T^{(y)}\right)\right] - \gamma^0\right\}$$
$$= v_{T,\gamma}^{-1}\sum_{t=1}^{T}\left\{y_t^2 I\left(|y_t| \leq c_T^{(y)}\right) - E\left[y_t^2 I\left(|y_t| \leq c_T^{(y)}\right)\right]\right\} + \frac{T}{v_{T,\gamma}}E\left[y_t^2 I\left(\tilde{c}_T^{(y)} \leq |y_t| \leq c_T^{(y)}\right)\right],$$

hence we require $Tv_{T,\gamma}^{-1}E[y_t^2 I(\tilde{c}_T^{(y)} \leq |y_t| \leq c_T^{(y)})] \rightarrow 0$ to ensure asymptotic unbiasedness. It is non-trivial because we over-correct by using $\tilde{c}_T^{(y)} < c_T^{(y)}$ while $T/v_{T,\gamma} \rightarrow \infty$ under intermediate order trimming (see Appendix A). As in Peng (2001) for iid data we can use $\tilde{c}_T^{(y)} = c_T^{(y)}$ such that A4 trivially holds, but this has the cost of a complicated limit theory and therefore complicated path towards inference.

Under Assumptions A1 and A2 y_t has a Paretian tail (cf. Basrak et al 2010: Theorem 3.1):

$$P\left(|y_t| > y\right) = cy^{-\kappa}\left(1 + o(1)\right),\ c > 0,\ \kappa > 2. \tag{14}$$

In order to ensure either $\hat{\gamma}_T^* = \hat{\gamma}_T^{(tr)} + \hat{\mathcal{R}}_T^*$ has a correction term $\hat{\mathcal{R}}_T^*$ that is negligible sufficiently fast, we require fractile bounds and a sharpening of tail decay, in particular the $o(1)$ term (14).

A5. $P(|y_t| > y) = cy^{-\kappa}(1 + O(y^{-\vartheta}))$ *where* $\vartheta > 0$.

A6. *The fractiles* $\{k_T^{(y)}, \tilde{k}_T^{(y)}, \check{k}_T^{(y)}\}$ *satisfy* $\check{k}_T^{(y)} = o(T^{2\vartheta/(2\vartheta+\kappa)})$, *and* (4) *and* (5).

Remark: Tail decay A5 was explored in Hall (1982) and Haesler and Teugles (1985) for Hill (1975) estimator asymptotics. In particular, under A1-A2, A5 and $\check{k}_T^{(y)} = o(T^{2\vartheta/(2\vartheta+\kappa)})$ asymptotic unbiasedness is assured: $\hat{\kappa}_{\check{k}_T^{(y)}} = \kappa + O_p(1/(\check{k}_T^{(y)})^{1/2})$ and $y_{(\tilde{k}_T^{(y)})}^{(a)}/c_T^{(y)} = 1 + O_p(1/(\tilde{k}_T^{(y)})^{1/2})$ cf. Hill (2010: Theorem 3). Conjoint with the remaining A6 fractile rates, the properties $\check{k}_T^{(y)}/\tilde{k}_T^{(y)} \rightarrow \infty$ and $\tilde{k}_T^{(y)}/k_T^{(y)} \rightarrow \infty$ ensure $\hat{\kappa}_{\check{k}_T^{(y)}} \xrightarrow{p} \kappa$ and $y_{(\tilde{k}_T^{(y)})}^{(a)}/c_T^{(y)} \xrightarrow{p} 1$ so fast that the correction term $\hat{\mathcal{R}}_T^*$ is negligible. See Appendix A. An example of sequences $\{k_T^{(y)}, \tilde{k}_T^{(y)}, \check{k}_T^{(y)}\}$ that satisfy $\check{k}_T^{(y)} = o(T^{2\vartheta/(2\vartheta+\kappa)})$, (4) and (5) is for any tiny $\iota > 0$ and some $\lambda, \tilde{\lambda}, \check{\lambda} > 0$

$$k_T^{(y)} = \left[\lambda \frac{T^{2\vartheta/(2\vartheta+\kappa)}}{(\ln(T))^{3\iota}}\right], \;\; \tilde{k}_T^{(y)} = \left[\tilde{\lambda} \frac{T^{2\vartheta/(2\vartheta+\kappa)}}{(\ln(T))^{2\iota}}\right] \;\; \text{and} \;\; \check{k}_T^{(y)} = \left[\check{\lambda} \frac{T^{2\vartheta/(2\vartheta+\kappa)}}{(\ln(T))^{\iota}}\right].$$

If $\vartheta \geq \kappa$ then we may use $k_T \sim \lambda T^{2/3}/(\ln(T))^{3\iota}$, $\tilde{k}_T \sim \tilde{\lambda} T^{2/3}/(\ln(T))^{2\iota}$ and $\check{k}_T \sim \check{\lambda} T^{2/3}/(\ln(T))^{\iota}$.

Define

$$\hat{\varphi}_T := \left[\hat{\gamma}_T^*, \hat{\xi}_T'\right]' \;\; \text{and} \;\; v_{T,\gamma}^2 := E\left[y_{T,t}^{*4}\right] - (\gamma^0)^2 \tag{15}$$

$$\mathcal{J} := E\left[\frac{1}{\sigma_t^4}\frac{\partial}{\partial \xi}\sigma_t^2 \frac{\partial}{\partial \xi'}\sigma_t^2\right] \;\; \text{and} \;\; \mathcal{K} := E\left[\frac{1}{\sigma_t^4}\frac{\partial}{\partial \xi}\sigma_t^2 \frac{\partial}{\partial \gamma'}\sigma_t^2\right]$$

Note $\mathcal{J}$ and $\mathcal{K}$ exist under Assumptions A1 and A2 as long as $E[\epsilon_t^4] < \infty$ (Francq and Zakoïan 2004, 2010).

See Appendix A for all proofs.

THEOREM 3.1 (Q-VT and Q-TTVT). *Let $E[\epsilon_t^4] < \infty$ and Assumptions A1-A6 hold. Then*

$$\frac{T}{v_{T,\gamma}} \mathcal{V}_T^{-1/2}\left(\hat{\varphi}_T - \varphi^0\right) \xrightarrow{d} N(0, I_3), \tag{16}$$

where

$$\mathcal{V}_T := \begin{bmatrix} 1 & -\mathcal{K}'\mathcal{J}^{-1} \\ -\mathcal{J}^{-1}\mathcal{K} & \left(\frac{E\left[\epsilon_t^4\right]-1}{v_{\gamma,T}^2/T}\right)\mathcal{J}^{-1} + \mathcal{J}^{-1}\mathcal{K}\mathcal{K}'\mathcal{J}^{-1} \end{bmatrix}. \tag{17}$$

In general $T/v_{T,\gamma} \to \infty$ and $\liminf_{T\to\infty} v_{\gamma,T}^2/T > 0$; if $E[y_t^4] = \infty$ then $T/v_{T,\gamma} = o(T^{1/2})$; and if $E[y_t^4] < \infty$ then

$$b := \lim_{T\to\infty} \frac{v_{T,\gamma}^2/T}{\left(E\left[\epsilon_t^4\right]-1\right)} = \left(\frac{1-\beta^0}{1-\alpha^0-\beta^0}\right)^2 \times E\left[h_t^4\right]. \tag{18}$$

Remark 1: The result contains both Q-VT asymptotics contained in Theorem 2.1 in FHZ (2010) for the case $E[y_t^4] < \infty$, and a robust version Q-TTVT for the case $E[y_t^4] = \infty$. If $E[y_t^4] < \infty$ as in FHZ (2010) then trimming does not influence the asymptotic distribution. Use (16)-(18) to deduce

$$\text{if } E\left[y_t^4\right] < \infty \text{ then } T^{1/2}\left(\hat{\varphi}_T - \varphi^0\right) \xrightarrow{d} N\left(0, \left(E[\epsilon_t^4] - 1\right)\mathcal{V}_0\right) \tag{19}$$

where

$$\mathcal{V}_0 = \lim_{T\to\infty} \frac{v_{T,\gamma}^2/T}{\left(E\left[\epsilon_t^4\right] - 1\right)}\mathcal{V}_T = \left[\begin{array}{ll} b & -\, b\mathcal{K}'\mathcal{J}^{-1} \\ -b\mathcal{J}^{-1}\mathcal{K} & \mathcal{J}^{-1} + b\mathcal{J}^{-1}\mathcal{K}\mathcal{K}'\mathcal{J}^{-1} \end{array}\right] \tag{20}$$

is identically FHZ's (2010: eq. (2.12)) asymptotic covariance matrix. If, however, y_t has an unbounded fourth moment $E[y_t^4] = \infty$ then $v_{\gamma,T}^2/T \to \infty$ hence

$$\text{if } E\left[y_t^4\right] < \infty \text{ then } \frac{T}{v_{T,\gamma}}\left(\hat{\varphi}_T - \varphi^0\right) \xrightarrow{d} N\left(0, \mathcal{V}_1\right) \quad \text{where} \quad \mathcal{V}_1 = \lim_{T\to\infty} \mathcal{V}_T = \left[\begin{array}{ll} 1 & -\,\mathcal{K}'\mathcal{J}^{-1} \\ -\mathcal{J}^{-1}\mathcal{K} & \mathcal{J}^{-1}\mathcal{K}\mathcal{K}'\mathcal{J}^{-1} \end{array}\right]$$

and only $T/v_{T,\gamma} = o(T^{1/2})$ convergence is obtained for the vector $\hat{\varphi}_T$.

Remark 2: The rate of convergence of $1/T\sum_{t=1}^T\{\epsilon_t^2 - 1\}s_t$ is exactly $T^{1/2}$ (Francq and Zakoïan 2004). However, heavy tails $E[y_t^4] = \infty$ imply the rate $T/v_{T,\gamma}$ of the plug-in $\hat{\gamma}_T^*$ satisfies both $T/v_{T,\gamma} \to \infty$ and $v_{T,\gamma}/T^{1/2} \to \infty$, and therefore dominates asymptotics given expansion (11). In particular $(Tv_{T,\gamma}^{-1})1/T\sum_{t=1}^T\{\epsilon_t^2 - 1\}s_t = (T^{1/2}/v_{T,\gamma})T^{-1/2}\sum_{t=1}^T\{\epsilon_t^2 - 1\}s_t \xrightarrow{p} 0$ hence $1/T\sum_{t=1}^T\{\epsilon_t^2 - 1\}s_t$ does not play any role in the limit distribution. This leads to the different asymptotic covariance matrix $\mathcal{V}_1$: most notably in the thin tail case $\mathcal{V}_0$ contains an additional Jacobian term $\mathcal{J}^{-1}$ associated with $1/T\sum_{t=1}^T\{\epsilon_t^2 - 1\}s_t$.

Remark 3: Tail-trimmed variance targeting incurs a cost of a diminished rate, with the benefit of standard asymptotics. It is tempting to solve the problem by using fixed quantile trimming, but $\hat{\gamma}_T^*$ would not be consistent for γ^0 since $\hat{\gamma}_T^* \xrightarrow{p} (0, \gamma^0)$ by construction.

Remark 4: FHZ (2010: Theorem 2.1) prove strong consistency $\hat{\varphi}_T \xrightarrow{a.s} \varphi^0$ by exploiting ergodicity. We can only prove weak consistency since we must get around the indicator $\hat{I}_{T,t}^{(y)} = I(|y_t| \leq y_{(k_T^{(y)})}^{(a)})$ to prove $\hat{\gamma}_T^*$ is consistent. This requires showing $1/T\sum_{t=1}^T y_t\{\hat{I}_{T,t}^{(y)} - I_{T,t}^{(y)}\}$ goes to zero and $1/T\sum_{t=1}^T y_t^2 I_{T,t}^{(y)}$ satisfies a LLN. Although a strong law can be applied to the latter, we only have the weak limit $1/T\sum_{t=1}^T y_t^2\{\hat{I}_{T,t}^{(y)} - I_{T,t}^{(y)}\} \xrightarrow{p}$

0 by exploiting $y^{(a)}_{(k^{(y)}_T)}/c^{(y)}_T = 1 + O_p(1/(k^{(y)}_T)^{1/2})$ under α-mixing (cf. HR 2010, Hill 2011).

In the heavy tail case $E[y_t^4] = \infty$ it is important to recognize $T/v_{T,\gamma} = o(T^{1/2})$ is the rate of convergence of the vector $\hat{\varphi}_T - \varphi^0$, but *not of every linear combination* $\zeta'(\hat{\varphi}_T - \varphi^0)$ since $\hat{\xi}_T - \xi^0$ is a linear function of $\hat{\gamma}^*_T - \gamma^0$. Consider the class $\zeta(\lambda) := [\lambda'\mathcal{J}^{-1}\mathcal{K}, \lambda']'$ for any non-zero $\lambda \in \mathbb{R}^2$. It is easy to verify by Theorem 3.1 that $\zeta(\lambda)'(\hat{\varphi}_T - \varphi^0)$ is the unique class of linear combinations that has an asymptotic scale

$$\frac{v^2_{T,\gamma}}{T^2}\zeta(\lambda)'\mathcal{V}_T\zeta(\lambda) = \frac{1}{T}\left(E\left[\epsilon_t^4 - 1\right]\right)\lambda'\mathcal{J}^{-1}\lambda,$$

and is therefore $T^{1/2}$-convergent and asymptotically normal:

$$T^{1/2}\zeta(\lambda)'\left(\hat{\varphi}_T - \varphi^0\right) \xrightarrow{d} N\left(0, \left(E\left[\epsilon_t^4\right] - 1\right) \times \lambda'\mathcal{J}^{-1}\lambda\right).$$

Since non-zero λ is otherwise arbitrary, pick $\lambda = [1, 0]'$ and $\tilde{\lambda} = [0, 1]'$ and take a linear combination of $\zeta(\lambda)'(\hat{\varphi}_T - \varphi^0)$ and $\zeta(\tilde{\lambda})'(\hat{\varphi}_T - \varphi^0)$ to see for any non-zero $d \in \mathbb{R}^2$

$$d'\begin{bmatrix} \zeta(\lambda)'\left(\hat{\varphi}_T - \varphi^0\right) \\ \zeta(\tilde{\lambda})'\left(\hat{\varphi}_T - \varphi^0\right) \end{bmatrix} \xrightarrow{d} N\left(0, \left(E\left[\epsilon_t^4\right] - 1\right) \times d'\mathcal{J}^{-1}d\right).$$

Let $\mathcal{J}^{-1}_{i,\cdot}$ denote the i^{th} row of $\mathcal{J}^{-1}$. Simply choose $d = [-\mathcal{J}^{-1}_{2,\cdot}\mathcal{K}, \mathcal{J}^{-1}_{1,\cdot}\mathcal{K}]'$ to conclude that the two directions $[\zeta(\lambda), \zeta(\tilde{\lambda})]d$ along the parameter space leads to $T^{1/2}$-convergence of a linear function of the GARCH parameters $\hat{\xi}_T$:

$$\begin{aligned} T^{1/2}d'\begin{bmatrix} \zeta(\lambda)' \\ \zeta(\tilde{\lambda})' \end{bmatrix}\left(\hat{\varphi}_T - \varphi^0\right) &= T^{1/2}\left[-\mathcal{J}^{-1}_{2,\cdot}\mathcal{K}, \mathcal{J}^{-1}_{1,\cdot}\mathcal{K}\right] \times \left(\hat{\xi}_T - \xi^0\right) \\ &\xrightarrow{d} N\left(0, \left(E\left[\epsilon_t^4\right] - 1\right) \times \left[-\mathcal{J}^{-1}_{2,\cdot}\mathcal{K}, \mathcal{J}^{-1}_{1,\cdot}\mathcal{K}\right]'\mathcal{J}^{-1}\left[-\mathcal{J}^{-1}_{2,\cdot}\mathcal{K}, \mathcal{J}^{-1}_{1,\cdot}\mathcal{K}\right]\right). \end{aligned}$$

Notice Q-TTVT under heavy tails $E[y_t^4] = \infty$ provides an efficiency improvement over Q-VT in the thin tail case $E[y_t^4] < \infty$. Invoke (19) and (20) to deduce the combination $T^{1/2}[-\mathcal{J}^{-1}_{2,\cdot}\mathcal{K}, \mathcal{J}^{-1}_{1,\cdot}\mathcal{K}] \times (\hat{\xi}_T - \xi^0)$ has an asymptotic scale, up to a multiple constant

$$[-\mathcal{J}_{2,\cdot}^{-1}\mathcal{K},\mathcal{J}_{1,\cdot}^{-1}\mathcal{K}]'\,\mathcal{J}^{-1}\,[-\mathcal{J}_{2,\cdot}^{-1}\mathcal{K},\mathcal{J}_{1,\cdot}^{-1}\mathcal{K}] < [-\mathcal{J}_{2,\cdot}^{-1}\mathcal{K},\mathcal{J}_{1,\cdot}^{-1}\mathcal{K}]'\left(\mathcal{J}^{-1}+b\mathcal{J}^{-1}\mathcal{K}\mathcal{K}'\mathcal{J}^{-1}\right)[-\mathcal{J}_{2,\cdot}^{-1}\mathcal{K},\mathcal{J}_{1,\cdot}^{-1}\mathcal{K}].$$

There are two observations to note. First, $[-\mathcal{J}_{2,\cdot}^{-1}\mathcal{K},\mathcal{J}_{1,\cdot}^{-1}\mathcal{K}]'$ is unique up to a multiple constant. Second, we cannot make the claim that $\lambda'(\hat{\xi}_T-\xi^0)$ is $T^{1/2}$-convergent for *any* $\lambda \in \mathbb{R}^2$ and then conclude by the Cramér-Wold theorem $\hat{\xi}_T=\xi^0+O_p(1/T^{1/2})$. Indeed, by Theorem 3.1 $\hat{\xi}_T$ is in general $T/v_{\gamma,T}$-convergent because it is a linear function of $\hat{\gamma}_T^*-\gamma^0$. Nevertheless, the direction $[-\mathcal{J}_{2,\cdot}^{-1}\mathcal{K},\mathcal{J}_{1,\cdot}^{-1}\mathcal{K}]\times(\hat{\xi}_T-\xi^0)$ *is* $T^{1/2}$ convergent because it is orthogonal to $\hat{\gamma}_T^*-\gamma^0$. See especially Antoine and Renault (2009) for deep theory on optimal rates in related non-standard contexts arising in robust and extreme value estimation contexts.

COROLLARY 3.2. *If $E[\epsilon_t^4]<\infty$ and Assumptions A1-A6 hold then the Q-TTVT estimator $\hat{\xi}_T$ of the GARCH parameters $\xi^0=[\alpha^0,\eta^0]'$ satisfies $T^{1/2}\lambda'(\hat{\xi}_T-\xi^0)\xrightarrow{d}N(0,(E[\epsilon_t^4]-1)\lambda'\mathcal{J}^{-1}\lambda)$ for the unique direction $\lambda=[-\mathcal{J}_{2,\cdot}^{-1}\mathcal{K},\mathcal{J}_{1,\cdot}^{-1}\mathcal{K}]'$ up to a constant scale.*

We have the usual corollary for the GARCH parameter set θ^0. Define $\hat{\theta}_T=\left[\hat{\omega}_T,\hat{\alpha}_T,\hat{\beta}_T\right]'$ where $\hat{\omega}_T=\hat{\gamma}_T^*\hat{\varphi}_{1,T}$, $\hat{\alpha}_T=\hat{\varphi}_{2,T}$, and $\hat{\beta}_T=1-\hat{\varphi}_{2,T}-\hat{\varphi}_{3,T}$, and define the selection matrix

$$\mathcal{L}=\begin{bmatrix} 1-\alpha^0-\beta^0 & 0 & 0 \\ 0 & 1 & -1 \\ \omega^0\left(1-\alpha^0-\beta^0\right)^{-1} & 0 & -1 \end{bmatrix}.$$

COROLLARY 3.3. *Under $E[\epsilon_t^4]<\infty$ and Assumptions A1-A6 $Tv_{T,\gamma}^{-1}(\mathcal{L}'\mathcal{V}_T\mathcal{L})^{-1/2}(\hat{\theta}_T-\theta^0)\xrightarrow{d}N(0,I_3)$.*

Remark: If $E[y_t^4]<\infty$ then by (20) we obtain Corollary 2.2 in FHZ (2010):

$$T^{1/2}(\hat{\theta}_T-\theta^0)\xrightarrow{d}N\left(0,\left(E\left[\epsilon_t^4\right]-1\right)\mathcal{L}'\mathcal{V}_0\mathcal{L}\right) \text{ since } \lim_{T\to\infty}\left(\frac{v_{T,\gamma}^2/T}{\left(E\left[\epsilon_t^4\right]-1\right)}\mathcal{L}'\mathcal{V}_T\mathcal{L}\right)=\mathcal{L}'\mathcal{V}_0\mathcal{L}.$$

3.2 MAIN RESULTS FOR G-VT and GTT-TTVT

The GMM limit distributions merely generalize the QML case. We first tackle the untrimmed G-VT estimator with untrimmed plug-in $\hat{\gamma}_T$ for the case $E[y_t^4] < \infty$ for the sake of reference. The estimator is

$$\hat{\xi}_T := \underset{\xi\in\Xi}{\text{arginf}} \left\{ \left(\frac{1}{T}\sum_{t=1}^{T} m_t(\hat{\gamma}_T, \xi) \right)' \times \hat{\Upsilon}_T \times \left(\frac{1}{T}\sum_{t=1}^{T} m_t(\hat{\gamma}_T, \xi) \right) \right\}.$$

Define the Jacobia

$$J^{<\gamma>} := E\left[x_t \frac{1}{\sigma_t^2} \frac{\partial}{\partial \gamma} \sigma_t^2 \right] \in \mathbb{R}^{q\times 1} \text{ and } J^{<\xi>} := E\left[x_t \frac{1}{\sigma_t^2} \frac{\partial}{\partial \xi'} \sigma_t^2 \right] \in \mathbb{R}^{q\times 2},$$

and notice under exact identification $J^{<\xi>} = \mathcal{J}$ is invertible, and $J^{<\gamma>} = \mathcal{K}$, where $\mathcal{J}$ and $\mathcal{K}$ are defined in (15). Define

$$\mathcal{H}^{<a,b>} := J^{<a>\prime} \times \Upsilon \times J^{<b>}$$

and the scale

$$\mathfrak{V} := \mathcal{A} \begin{bmatrix} b & \left(\frac{1-\beta^0}{\eta^0}\right) E\left[x_t'\sigma_t^2\right] \\ \left(\frac{1-\beta^0}{\eta^0}\right) E\left[x_t\sigma_t^2\right] & E\left[x_t x_t'\right] \end{bmatrix} \mathcal{A}' \in \mathbb{R}^{3\times 3} \tag{21}$$

b is defined by Theorem 3.1 and

$$\mathcal{A} := \begin{bmatrix} 1 & 0 \\ -\left(\mathcal{H}^{<\xi,\xi>}\right)^{-1} \mathcal{H}^{<\xi,\gamma>} & \left(\mathcal{H}^{<\xi,\xi>}\right)^{-1} J^{<\xi>\prime}\Upsilon \end{bmatrix} \in \mathbb{R}^3.$$

We require one assumption specific to GMM.

A7. *There exists a sequence of non-random positive definite matrices $\{\Upsilon_T\}$ that satisfies $\hat{\Upsilon}_T \Upsilon_T^{-1} \xrightarrow{p} I_q$. If $E[y_t^4] < \infty$ then $\Upsilon_T \to \Upsilon$ a finite positive deficit matrix.*

THEOREM 3.4 (G-VT). *Let $E[y_t^4] < \infty$ and Assumptions A1-A7 hold. Then $\hat{\varphi}_T \xrightarrow{p} \varphi^0$ and $T^{1/2}(\hat{\varphi}_T - \varphi^0) \xrightarrow{d} N(0, (E[\epsilon_t^4] - 1) \times \mathfrak{V})$.*

Remark: Under exact identification $x_t = s_t$ the G-VT estimator reduces to Q-VT since it is easily shown

$E[x_t\sigma_t^2]=0$ as in FHZ 2010: eq. (A.13)). Use the identities $J^{<\xi>}=\mathcal{J}=E[x_tx_t']$ and $J^{<\gamma>}=\mathcal{K}$ to deduce

$$\mathfrak{V}=\mathcal{V}_0=\left[\begin{array}{ll} b & -b\mathcal{K}'\mathcal{J}^{-1} \\ -b\mathcal{J}^{-1}\mathcal{K} & \mathcal{J}^{-1}+b\mathcal{J}^{-1}\mathcal{K}\mathcal{K}'\mathcal{J}^{-1} \end{array}\right],$$

identically the Q-VT covariance.

As above, let $w_t(\varphi)\in\{e_t^{(a)}(\varphi),x_{i,t}^{(a)}(\varphi)\}$ and $k_T^{(w)}\in\{k_T^{(e)},k_{i,T}^{(x)},\}$, and define positive two-tailed sequences $c_T^{(w)}(\varphi)\in\{c_T^{(\epsilon)}(\varphi),c_{i,T}^{(x)}(\varphi)\}$ that satisfy

$$P\left(|w_t(\varphi)|>c_T^{(w)}(\varphi)\right)=\frac{k_T^{(w)}}{T}.$$

Define

$$w_{T,t}^*(\varphi):=w_t(\varphi)\,I\left(|w_t(\varphi)|\leq c_T^{(w)}(\varphi)\right)=w_t(\varphi)\,I_{T,t}^{(w)}(\varphi)\quad\text{and}\quad w_{T,t}^*(\xi):=w_{T,t}(\xi)\,I_{T,t}^{(w)}(\xi)\,.$$

and tail-trimmed equations

$$\begin{aligned} m_{T,t}^*(\varphi) &= \left(e_{T,t}^2(\varphi)\,I_{T,t}^{(e)}(\varphi)-E\left[e_{T,t}^2(\varphi)\,I_{T,t}^{(e)}(\varphi)\right]\right)\times x_{T,t}(\varphi)\prod_{i=1}^{q}I_{i,T,t}^{(x)}(\varphi) \\ &= \left(e_{T,t}^{*2}(\varphi)-E\left[e_{T,t}^{*2}(\varphi)\right]\right)\times x_{T,t}^*(\varphi) \end{aligned}$$

$$m_{T,t}^*(\xi)=\left(e_{T,t}^{*2}(\xi)-E\left[e_{T,t}^{*2}(\xi)\right]\right)\times x_{T,t}^*(\xi)$$

If $x_t(\varphi)$ contains only $s_t(\varphi)$ and its lags $s_{t-1}(\varphi),...,s_{t-h+1}(\varphi)$ then we use

$$m_{T,t}^*(\varphi)=\left(e_{T,t}^{*2}(\varphi)-E\left[e_{T,t}^{*2}(\varphi)\right]\right)\times x_{T,t}(\varphi)\text{ if GARCH effects are assumed}$$

or in general

$$m_{T,t}^*(\varphi)=\left(e_{T,t}^{*2}(\varphi)-E\left[e_{T,t}^{*2}(\varphi)\right]\right)\times x_{T,t}(\varphi)\prod_{i=1}^{h+1}I_{T,t-i}^{(y)}$$

By centering with the mean of $e^{*2}_{T,t}(\varphi)$ and by independence of $e_t(\varphi^0) = \epsilon_t$ we ensure $\{m^*_{T,t}, \Im_t\}$ is a martingale difference:

$$E\left[m^*_{T,t}|\Im_{t-1}\right] = x^*_{T,t} \times E\left(e_t I^{(e)}_{T,t} - E\left[e_t I^{(e)}_{T,t}\right]\right) = 0.$$

We therefore only need to define an instantaneous covariance matrix for asymptotics

$$\Sigma_T(\varphi) := E\left[m^*_{T,t}(\varphi)m^*_{T,t}(\varphi)'\right] \text{ and } \Sigma_T = \Sigma_T(\varphi^0) = E\left[\left(e^{*2}_{T,t} - E\left[e^{*2}_{T,t}\right]\right)^2\right] \times E\left[x^*_{T,t}x^{*\prime}_{T,t}\right].$$

This provides a substantial simplification over the GMTTM methodology of HR (2010), cf. Hill and Aguilar (2010). Their approach is necessarily generalistic so they trim m_t by its large values, hence re-centering is not feasible. Their trimmed equations in general are not martingale differences, so a HAC estimator must be used.

Required Jacobian matrices are

$$J^{<\gamma>}_T := E\left[x_t \frac{1}{\sigma^2_t}\frac{\partial}{\partial\gamma}\sigma^2_t \times I^{(x)}_{T,t}\right] \in \mathbb{R}^{q\times 1} \text{ and } J^{<\xi>}_T := E\left[x_t \frac{1}{\sigma^2_t}\frac{\partial}{\partial\xi'}\sigma^2_t \times I^{(x)}_{T,t}\right] \in \mathbb{R}^{q\times 2}$$

with quadratic forms

$$\mathcal{H}^{<a,b>}_T := J^{<a>\prime}_T \times \Upsilon_T \times J^{<b>}_T.$$

Define a long-run covariance matrix (exploiting the mds property)

$$\begin{aligned}\mathcal{C}^{(m,y)}_T &= E\left[\sum_{t=1}^T \left(m^*_{T,t} - E\left[m^*_{T,t}\right]\right)\sum_{t=1}^T \left(y^{*2}_{T,t} - E\left[y^{*2}_{T,t}\right]\right)\right] \\ &= \sum_{r\leq t} E\left[\left(m^*_{T,r} - E\left[m^*_{T,r}\right]\right)\left(y^{*2}_{T,t} - E\left[y^{*2}_{T,t}\right]\right)\right] \in \mathbb{R}^{q\times 1},\end{aligned}$$

and a scale matrix

$$\mathfrak{V}_T := \mathcal{A}_T \begin{bmatrix} 1 & \mathcal{C}^{(m,y)\prime}_T/\upsilon^2_{T,\gamma} \\ \mathcal{C}^{(m,y)}_T/\upsilon^2_{T,\gamma} & T\Sigma_T/\upsilon^2_{T,\gamma} \end{bmatrix} \mathcal{A}'_T \in \mathbb{R}^{3\times 3}$$

where

$$\mathcal{A}_T := \begin{bmatrix} 1 & 0 \\ -\left(\mathcal{H}_T^{<\xi,\xi>}\right)^{-1}\left(\mathcal{H}_T^{<\xi,\gamma>}\right) & \left(\mathcal{H}_T^{<\xi,\xi>}\right)^{-1} J_T^{<\xi>\prime}\Upsilon_T \end{bmatrix} \in \mathbb{R}^{3\times(q+1)}.$$

THEOREM 3.5 (GTT-TTVT). *Under Assumptions A1-A7* $\hat{\varphi}_T \xrightarrow{p} \varphi^0$ *and*

$$\frac{T}{v_{T,\gamma}}\mathfrak{V}_T^{-1/2}\left(\hat{\varphi}_T - \varphi^0\right) \xrightarrow{d} N(0, I_3).$$

Remark: Under exact identification $\mathcal{H}_T^{<\xi,\xi>} = \mathcal{J}_T\Upsilon_T\mathcal{J}_T$ and $\mathcal{H}_T^{<\xi,\gamma>} = \mathcal{J}_T\Upsilon_T\mathcal{K}_T$, where

$$\mathcal{J}_T = E\left[\frac{1}{\sigma_t^4}\frac{\partial}{\partial\xi}\sigma_t^2\frac{\partial}{\partial\xi'}\sigma_t^2 I_{T,t}^{(x)}\right] \text{ and } \mathcal{K}_T = E\left[\frac{1}{\sigma_t^4}\frac{\partial}{\partial\xi}\sigma_t^2\frac{\partial}{\partial\gamma}\sigma_t^2 I_{T,t}^{(x)}\right],$$

hence

$$\mathcal{A}_T := \begin{bmatrix} 1 & 0 \\ -\mathcal{J}_T^{-1}\mathcal{K}_T & \mathcal{J}_T^{-1} \end{bmatrix}$$

and

$$\mathfrak{V}_T = \begin{bmatrix} 1 & -\mathcal{K}_T'\mathcal{J}_T^{-1} + \mathcal{R}_T' \\ -\mathcal{J}_T^{-1}\mathcal{K}_T + \mathcal{R}_T & \mathcal{J}_T^{-1}\mathcal{K}_T\mathcal{K}_T'\mathcal{J}_T^{-1} + \mathcal{J}_T^{-1}T\Sigma_T\ \mathcal{J}_T^{-1}/v_{T,\gamma}^2 - \mathcal{R}_T\mathcal{K}_T\mathcal{J}_T^{-1} - \mathcal{J}_T^{-1}\mathcal{K}_T\mathcal{R}_T' \end{bmatrix}.$$

where $\mathrm{R}_T := \mathcal{J}_T^{-1}\mathcal{C}_T^{(m,y)}/v_{T,\gamma}^2$.

As long as $E[\epsilon_t^4] = \infty$ then the rate of convergence along any direction $\zeta'(\hat{\varphi}_T - \varphi^0)$ is $o_p(T^{1/2})$ due to $\Sigma_{i,i,T} \sim (E[e_{T,t}^{*4}] - 1) \times E\left[s_{i,t}^{*2}\right] \to \infty$. This is best seen by assuming exact identification and GARCH effects without loss of generality (cf. Hill and Renault 2010). Then $\mathcal{J}_T \to \mathcal{J}$ and $\mathcal{K}_T \to \mathcal{K}$ hence

$$\frac{v_{T,\gamma}^2}{T}\zeta'\mathfrak{V}_T\zeta \sim \zeta'\begin{bmatrix} \frac{v_{T,\gamma}^2}{T} & -\mathcal{K}_T\mathcal{J}_T^{-1}\frac{v_{T,\gamma}^2}{T} \\ -\mathcal{J}^{-1}\mathcal{K}\frac{v_{T,\gamma}^2}{T} & \mathcal{J}^{-1}\mathcal{K}\mathcal{K}'\mathcal{J}^{-1}\frac{v_{T,\gamma}^2}{T} + \mathcal{J}^{-1}\left(E[e_{T,t}^{*4}] - 1\right) \end{bmatrix}\zeta.$$

Therefore every direction contains $v_{T,\gamma}^2/T \to \infty$ and/or $E[e_{T,t}^{*4}] \to \infty$, hence $v_{T,\gamma}^2 T^{-1}\zeta'\mathfrak{V}_T\zeta \to \infty$ implies $\zeta'(\hat{\varphi}_T - \varphi^0) = o_p(T^{1/2})$. Otherwise, if $E[\epsilon_t^4] < \infty$ then along the argument of Corollary 3.2 $T^{1/2}$-convergent directions

$\zeta'(\hat{\varphi}_T - \varphi^0)$ can be obtained.

Under exact identification and thin tails $E[\epsilon_t^4] < \infty$ it follows $\mathcal{C}_T^{(m,y)}/v_{T,\gamma}^2 \to 0$ by error independence and $E[s_t\sigma_t^2] = 0$, hence

$$\mathfrak{V}_T = \begin{bmatrix} 1 & -\mathcal{K}_T\mathcal{J}_T^{-1} \\ -\mathcal{J}_T^{-1}\mathcal{K}_T & \mathcal{J}_T^{-1}\mathcal{K}_T\mathcal{K}_T'\mathcal{J}_T^{-1} + \mathcal{J}_T^{-1}T\Sigma_T\mathcal{J}_T^{-1}/v_{T,\gamma}^2 \end{bmatrix}$$

Thus GTT-TTVT reduces asymptotically to QML-TTVT by case $E[y_t^4] = \infty$ or $E[y_t^4] < \infty$. In the case of thin tails $E[y_t^4] < \infty$ trimming does not impact the limit distribution and $\mathfrak{V}_T$ reduces to G-VT.

COROLLARY 3.6. *Let $E[\epsilon_t^4] < \infty$ and Assumptions A1-A7 hold.*

a. If $E[y_t^4] < \infty$ then $\mathfrak{V}_T/T \to (E[\epsilon_t^4] - 1)\mathfrak{V}$ where $\mathfrak{V}$ is defined in (21), hence $T^{1/2}(\hat{\varphi}_T - \varphi^0) \xrightarrow{d} N(0, (E[\epsilon_t^4] - 1) \times \mathfrak{V})$.

b. Under exact identification $x_t = s_t$ with with $\Upsilon_T = \Sigma_T^{-1}$ *it follows $Tv_{T,\gamma}^{-1}\mathcal{V}_T^{-1/2}(\hat{\varphi}_T - \varphi^0) \xrightarrow{d} N(0, I_3)$. In particular, if $E[y_t^4] = \infty$ then $Tv_{T,\gamma}^{-1}(\hat{\varphi}_T - \varphi^0) \xrightarrow{d} N(0, \mathcal{V}_1)$ and if $E[y_t^4] < \infty$ then $T^{1/2}(\hat{\varphi}_T - \varphi^0) \xrightarrow{d} N(0, (E[\epsilon_t^4] - 1) \times \mathcal{V}_0)$.*

4. MONTE CARLO STUDY We now compare standard and robust Variance Targeting estimators for the following GARCH model:

$$y_t = h_t\epsilon_t \text{ and } h_t^2 = 1 + \alpha^0 y_{t-1}^2 + \beta^0 h_{t-1}^2, \text{ where } 0 < \alpha^0, \beta^0 < 1$$

Let $N_{0,1}$ denote a standard normal distribution, and P_κ a symmetric Pareto distribution $P(\epsilon_t \geq \epsilon) = P(\epsilon_t \leq -\epsilon) = .5(1 + \epsilon)^{-\kappa}$. Any $\epsilon_t \sim P_\kappa$ with $\kappa > 2$ is standardized such that $\epsilon_t \sim (0, 1)$. The error tails and parameters (α^0, β^0) both impact the tails of y_t. In Table 1 we present the various combinations of error distributions and (α^0, β^0) used in this study, and the tail index κ_y of y_t computed by simulation[5]. We also denote which estimators are asymptotically normal. Of course, Q-VT is normal is the least number of cases since it requires $E[y_t^4] < \infty$, while Q-TTVT requires $E[\epsilon_t^4] < \infty$ and finally GTT-TTVT [denoted G-TTVT]

[5] The index κ_y of y_t satisfies $E[(\alpha^0\epsilon_t^2 + \beta^0)^{\kappa_y/2}] = 1$ (see, e.g., Basrak et al 2002). The index κ_y is computed as $\hat{\kappa} = \arg\min_{\kappa \in K}\{|1/N\sum_{t=1}^N(\alpha^0\epsilon_t^2 + \beta^0)^{\kappa/2} - 1|\}$ over $K \in \{.01, .02, ..., 10\}$ based on $N = 100,000$ iid random draws ϵ_t from its corresponding distribution. The 1% band half-length is less than .001.

only requires $E[y_t^2] < \infty$. The sample sizes are $T \in \{250, 500, 1000\}$.

TABLE 1

Parameter			Error and Tails			Estimator Normality			
ω^0	α^0	β^0	ϵ_t	κ_ϵ	κ_y	QML	Q-VT	Q-TTVT	G-TTVT
1	.3	.4	$P_{2.5}$	2.5	1.95				*
1	.3	.4	$P_{4.1}$	4.1	3.17	*		*	*
1	.3	.4	$N_{0,1}$	∞	6.25	*	*	*	*
1	.3	.6	$P_{2.5}$	2.5	1.50				*
1	.3	.6	$P_{4.1}$	4.1	2.33	*		*	*
1	.3	.6	$N_{0,1}$	∞	4.10	*	*	*	*

Each case labeled "*" refers to an asymptotically normal estimator[6].

We compute the untrimmed and tail-trimmed estimators $\hat{\gamma}_T = 1/T 1/T \sum_{t=1}^T y_t^2$ and $\hat{\gamma}_T^{(rt)} := 1/T \sum_{t=1}^T y_t^2 I(|y_t| \leq y_{(k_T)}^{(a)})$ for the optimal gap filled estimator $\hat{\gamma}_T^{(o)} = \hat{\gamma}_T^{(tr)} + \hat{\mathcal{R}}_T^{(o)}$, with fractile $k_T = [.05T^{2/3}/(\ln(T))^{3\iota}]$ and $\iota = 10^{-10}$. The optimal gap filler $\hat{\mathcal{R}}_T^{(o)} = (\hat{\kappa}_{\check{k}_T^{(y)}(\check{\lambda}^o)}/(\hat{\kappa}_{\check{k}_T^{(y)}(\check{\lambda}^o)} - 2))(\tilde{k}_T^{(y)}(\tilde{\lambda}^o)/T)(y^{(a)}_{(\tilde{k}_T^{(y)}(\tilde{\lambda}^o))})^2$ is computed using $\tilde{k}_T(\lambda) = [\lambda T^{2/3}/(\ln(T))^{2\iota}]$ and $\check{k}_T^{(y)}(\lambda) = [\lambda T^{2/3}/(\ln(T))^{\iota}]$, thus $k_T = \tilde{k}_T(.05) = \check{k}_T^{(y)}(.05)$ for the sample sizes treated here. We solve

$$\left[\tilde{\lambda}^o, \check{\lambda}^o\right] = \underset{\tilde{\lambda}, \check{\lambda} \in \Lambda_n}{\text{arginf}} \left| \hat{\gamma}_T^{(tr)} + \left(\frac{\hat{\kappa}_{\check{k}_T^{(y)}(\check{\lambda})}}{\hat{\kappa}_{\check{k}_T^{(y)}(\check{\lambda})} - 2} \right) \frac{\tilde{k}_T^{(y)}(\tilde{\lambda})}{T} \left(y^{(a)}_{(\tilde{k}_T^{(y)}(\tilde{\lambda}))} \right)^2 - \hat{\gamma}_T \right|$$

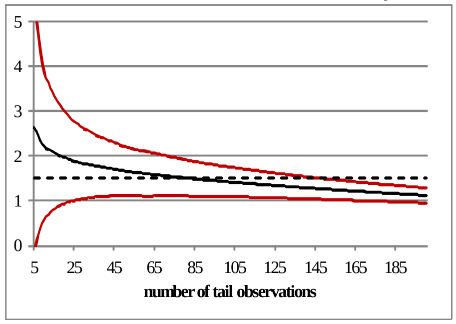
over the grid $\Lambda_n = \{.05 + i20/n : i = 0, ..., n/100\} \times \{1 + i50/n : i = 0, ..., n/100\} \approx [.05, .25] \times [1.0, 1.5]$. Our choice of $\tilde{\lambda} \in [.05, .25]$ follows from the fact that very little trimming leads to a sharp and approximating normal estimator, while $\check{\lambda} \in [1.0, 1.5]$ because $\hat{\kappa}_{\check{k}_T^{(y)}(\check{\lambda})}$ is sharpest for GARCH processes in this range. See Figure 2 for a Hill-plot with nonparametric 90% confidence bands for the case $\beta^0 = .6$, $\epsilon_t \sim P_{2.5}$ and $T = 500$. See Hill (2010) for the construction of the bands. As opposed to the classic case of iid Pareto data (see Embrechts et al 1997), $\hat{\kappa}_{\check{k}_T^{(y)}}$ does not hover around the true $\kappa_y = 1.5$, but crosses κ_y at $\check{k}_T^{(y)} = 85 = [1.35 \times 500^{2/3}]$. In practice

[6] In the remaining cases the limit of a properly scaled estimator is a ratio of stable laws (e.g. Hall and Yao 2003), or the limit is not yet available (e.g. Q-VT with $E[\epsilon_t^4] = \infty$).

it we may be prudent to use a broader grid, and by construction $[\tilde{\lambda}^o, \check{\lambda}^o]$ will promote a sharper estimate. We find here that our choice of graininess and breadth of Λ_n works exceptionally well.

FIGURE 2

Hill Plot for GARCH(1,1): $\kappa_y = 1.5$



In the case of GTT-TTVT we use the same fractile $k_T = [.05T^{2/3}/(\ln(T))^{3\iota}]$ for equation trimming, and re-center the tail-trimmed error. We use two score lags as over identifying restrictions: $x_t = [s_t', s_{t-1}', s_{t-2}']'$. We also use the weight $\hat{\Upsilon}_T = I_q$ since no other weight provides a sharper estimate in simulation experiments (see also HR 2010). See Tables 2-4 for all simulation results: we report simulation averages for $\hat{\theta}_{T,3}$, the simulation mean squared error, the Kolmogorov-Smirnov statistic for a test of standard normality, and rejection frequencies for tests of $\theta_3 = .6$ (true), $\theta_3 = .35$ (false), and $\theta_3 = .0$ (false).

Both QMFTTL and GTT-TTVT estimators work very well, in particular by delivering an estimator closer to normal when tails are heavy. If we do not re-center the tail-trimmed errors in the GMM case then non-negligible small sample bias occurs, ranging from .1 to .3, where heavier tails align with greater bias[7].

Non-robust variance targeting often provides a sharper estimator than QML as measured by the mean-squared-error, in particular in the case of a large signal/noise ratio a la $\beta^0 = .6$ (cf. FHZ 2010). By comparison the tail-trimmed variants provide the best overall estimator. Of particular note is the superlative performance of t-tests by using QMFTTL or GTT-TTVT as a plug-in estimator: empirical size is fairly sharp in all cases and exceptionally sharp in many cases, while power is highest overall for these estimators.

5. CONCLUSION We develop heavy tail robust estimators for variance targeting by negligibly trimming a sample variance and components of QML and GMM estimators. The QML estimator requires a GARCH

[7] These simulation results are available upon request.

error ϵ_t with finite fourth moment, while the GMM estimator allows for heavy tailed errors. We both provide robust estimators under variance targeting and improve upon tail-trimmed GMM estimation for GARCH models by eradicating small sample bias. The new estimators perhaps well in controlled experiments, trumping QML and QML with variance targeting in approximate normality and inference.

Appendix A : Proofs of Main Results

Under A1 and A2 $\{y_t, h_t^2\}$ is strictly stationary, ergodic, and geometrically β-mixing, where a stationarity ergodic solution to $\{\tilde{\sigma}_t^2(\varphi)\}$ in (8) exists, denoted $\{\sigma_t^2(\varphi)\}$. See Francq and Zakoïan (2004, 2006, 2010). Further, under A1 and A2 y_t has an absolutely continuous distribution with a Paretian tail (14). Define loss equations

$$l_t(\varphi) := \ln \sigma_t^2(\varphi) + \frac{y_t^2}{\sigma_t^2(\varphi)} \text{ and } \tilde{l}_t(\varphi) := \ln \tilde{\sigma}_t^2(\varphi) + \frac{y_t^2}{\tilde{\sigma}_t^2(\varphi)}.$$

The core asymptotic theory is developed for $\hat{\gamma}_T^{(tr)}$ and $\hat{\gamma}_T^{(g)} = \hat{\gamma}_T^{(tr)} + \hat{\mathcal{R}}_T^{(g)}$. An empirical process argument shows the proof for the optimal estimator $\hat{\gamma}_T^{(o)} = \hat{\gamma}_T^{(tr)} + \hat{\mathcal{R}}_T^{(o)}$ is identical to the proof for $\hat{\gamma}_T^{(g)}$. See the proof of Lemma A.1 in Appendix B. Let $\hat{\gamma}_T^*, \hat{\mathcal{R}}_T^*$ denote either $\hat{\gamma}_T^{(g)}, \hat{\mathcal{R}}_T^{(g)}$ or $\hat{\gamma}_T^{(o)}, \hat{\mathcal{R}}_T^{(o)}$:

$$\hat{\gamma}_T^*, \hat{\mathcal{R}}_T^* = \hat{\gamma}_T^{(g)}, \hat{\mathcal{R}}_T^{(g)} \text{ or } \hat{\gamma}_T^{(o)}, \hat{\mathcal{R}}_T^{(o)}.$$

We require several preliminary lemmas linking the gap filled estimator $\hat{\gamma}_T^* = \hat{\gamma}_T^{(tr)} + \hat{\mathcal{R}}_T^*$ to $\sum_{t=1}^T y_t^2 I_{T,t}^{(y)}$ for which a Gausssian central limit theorem exists.

LEMMA A.1. *Under A1, A2, A5 and A6* $Tv_{T,\gamma}^{-1}\{\hat{\gamma}_T^{(g)} - \hat{\gamma}_T^{(tr)} - E[y_t^2 I(|y_t| > c_T)]\} \xrightarrow{p} 0$ and $Tv_{T,\gamma}^{-1}\{\hat{\gamma}_T^{(o)} - \hat{\gamma}_T^{(tr)} - E[y_t^2 I(|y_t| > c_T)]\} \xrightarrow{p} 0$.

LEMMA A.2. *Under A1 and A2* $v_{T,\gamma}^{-1} \sum_{t=1}^T \{y_t^2 \hat{I}_{T,t}^{(y)} - y_t^2 I_{T,t}^{(y)}\} \xrightarrow{p} 0$.

LEMMA A.3. *Under A1 and A2* $v_{T,\gamma}^{-1} \sum_{t=1}^T \{y_t^2 I_{T,t}^{(y)} - E[y_t^2 I_{T,t}^{(y)}]\} \xrightarrow{d} N(0,1)$ *where* $v_{T,\gamma}^2 \to \infty$ *and* $v_{T,\gamma}^2 = o(T^2)$. *If* $E[y_t^4] = \infty$ *then* $v_{T,\gamma}^2/T \to \infty$.

Remark: In lieu of the smoothness, mixing and regular variation properties detailed above, Lemmas A2 and A3 are identical to, or follows from, results developed in HR (2010), Hill and Aguilar (2010) and Hill (2011). Consult those source for details. We therefore only prove Lemma A.1 in Appendix B.

We make repeated use of arguments in FHZ (2010), so we simply write FHZ.

PROOF OF THEOREM 3.1. There are two cases: $E[y_t^4] < \infty$ and $E[y_t^4] = \infty$.

Case 1 ($E[y_t^4] < \infty$): Under stationary geometric β-mixing and dominated convergence (cf. Ibragimov 1962)

$$\begin{aligned}
&\frac{1}{T}E\left(\sum_{t=1}^{T}\left(y_{T,t}^{*2} - E\left[y_{T,t}^{*2}\right]\right)\right)^2 \\
&\qquad = E\left(y_{T,t}^{*2} - E\left[y_{T,t}^{*2}\right]\right)^2 + 2\sum_{i=1}^{T-1}\left(1 - \frac{T}{i}\right)E\left[\left(y_{T,1}^{*2} - E\left[y_{T,t}^{*2}\right]\right)\left(y_{T,i+1}^{*2} - E\left[y_{T,t}^{*2}\right]\right)\right] \\
&\qquad \to E\left[y_t^4\right] + 2\sum_{i=1}^{\infty}E\left[\left(y_1^2 - \gamma^0\right)\left(y_{i+1}^2 - \gamma^0\right)\right] = \lim_{T\to\infty}\frac{1}{T}E\left(\sum_{t=1}^{T}\left(y_t^2 - \gamma^0\right)\right)^2 < \infty.
\end{aligned}$$

Now use error independence and derivations in Horváth et al (2006: (37)-(41)) to deduce

$$\lim_{T\to\infty}\frac{1}{T}E\left(\sum_{t=1}^{T}\left(y_t^2 - \gamma^0\right)\right)^2 = \left(E\left[\epsilon_t^4\right] - 1\right) \times \left(\frac{1-\beta^0}{1-\alpha^0-\beta^0}\right)^2 \times E\left[h_t^2\right] = \left(E\left[\epsilon_t^4\right] - 1\right) \times b,$$

hence

$$\begin{aligned}
\lim_{T\to\infty}\left\{\frac{\upsilon_{T,\gamma}^2}{T}\mathcal{V}_T\right\} &= \left(E\left[\epsilon_t^4\right] - 1\right) \times b \begin{bmatrix} 1 & -\mathcal{K}'\mathcal{J}^{-1} \\ -\mathcal{J}^{-1}\mathcal{K} & \frac{1}{b}\mathcal{J}^{-1} + \mathcal{J}^{-1}\mathcal{K}\mathcal{K}'\mathcal{J}^{-1} \end{bmatrix} \\
&= \left(E\left[\epsilon_t^4\right] - 1\right) \times \begin{bmatrix} b & -b\mathcal{K}'\mathcal{J}^{-1} \\ -b\mathcal{J}^{-1}\mathcal{K} & \mathcal{J}^{-1} + b\mathcal{J}^{-1}\mathcal{K}\mathcal{K}'\mathcal{J}^{-1} \end{bmatrix} = \left(E\left[\epsilon_t^4\right] - 1\right)\mathcal{V}_0,
\end{aligned}$$

say.

Trimming has no impact asymptotically. Specifically, it is easy to show all arguments in FHZ (2010: Appendices A.1-A.2) carry over verbatim to prove $T^{1/2}(\hat{\varphi}_T - \varphi^0) \stackrel{d}{\to} N(0, (E[\epsilon_t^4] - 1)\mathcal{V}_0)$. Since $(\upsilon_{T,\gamma}^2/T)\mathcal{V}_T \to (E\left[\epsilon_t^4\right] - 1)\mathcal{V}_0$ the proof is complete.

Case 2 ($E[y_t^4] = \infty$): We prove consistency and asymptotic normality in two steps.

Step 1 (consistency): The limit $\hat{\gamma}_T^* \xrightarrow{p} \gamma^0$ follows from Lemma A.2, and the facts that y_t is $L_{2+\iota}$-bounded and geometrically β-mixing, hence $1/T \sum_{t=1}^T y_t^2 I_{T,t}^{(y)} \xrightarrow{p} \gamma^0$ by Theorem 2 in Andrews (1988).

The remaining proof of $\hat{\xi}_T \xrightarrow{p} \xi^0$ can be adapted from FHZ's proof of their Theorem 2.1. They exploit stationarity and ergodicity of $\{y_t, h_t^2\}$ and strong consistency $\hat{\gamma}_T^* \xrightarrow{a.s} \gamma^0$ to prove strong consistency of $\hat{\xi}_T$. Since their argument does not otherwise involve higher moments we may simply substitute for weak consistency to prove $\hat{\xi}_T \xrightarrow{p} \xi^0$ by their argument verbatim.

Step 2 (normality): Define

$$\mathcal{E}_t(\varphi) := \left\{\epsilon_t^2(\varphi) - 1\right\} s_t(\varphi).$$

The steps required to show asymptotic normality are closely aligned with FHZ. We need only verify their Appendix A.2 claims (i), (ii), and (iv)-(vi) for our setting. Note they use an additional claim (iii) solely to prove (v). The conditions are stated below for ease of reference. Let $\mathcal{N}(\varphi^0)$ be a compact neighborhood of φ^0:

$$E\left\|\frac{\partial}{\partial\varphi}l_t\frac{\partial}{\partial\varphi'}l_t\right\| < \infty \quad \text{and} \quad E\left\|\frac{\partial}{\partial\varphi}\frac{\partial}{\partial\varphi'}l_t\right\| < \infty \tag{i}$$

$$\mathcal{J}^{(a,b)} := E\left[\frac{1}{\sigma_t^4}\frac{\partial}{\partial a}\sigma_t^2\frac{\partial}{\partial b}\sigma_t^2\right] \text{ exists for any sub-vector index } a, b \in \{\gamma, \xi, \varphi\}, \text{ and}$$
$$\mathcal{J}^{(\varphi,\varphi)} \text{ non-singular} \quad \text{and} \quad E\left[\frac{\partial}{\partial\varphi}l_t\frac{\partial}{\partial\varphi'}l_t\right] = E\left[\epsilon_t^4 - 1\right] \times \mathcal{J} \tag{ii}$$

$$\left\|\sum_{t=1}^T\left\{\frac{\partial}{\partial\varphi}l_t - \frac{\partial}{\partial\varphi}\tilde{l}_t\right\}\right\| = O_p(1) \quad \text{and} \quad \sup_{\varphi\in\mathcal{N}(\varphi^0)}\left\|\sum_{t=1}^T\left\{\frac{\partial}{\partial\varphi}\frac{\partial}{\partial\varphi'}l_t(\varphi) - \frac{\partial}{\partial\varphi}\frac{\partial}{\partial\varphi'}\tilde{l}_t(\varphi)\right\}\right\| = O_p(1) \tag{iv}$$

$$\frac{1}{T}\sum_{t=1}^T\frac{\partial}{\partial\varphi}\frac{\partial}{\partial\varphi'}l_t(\varphi_*) \xrightarrow{p} \mathcal{J}^{(\varphi,\varphi)} \quad \text{for any} \quad \left\|\varphi_* - \varphi^0\right\| \leq \left\|\hat{\varphi}_T - \varphi^0\right\| \tag{v}$$

$$\frac{1}{T^{1/2}}\sum_{t=1}^T\mathcal{E}_t = O_p(1) \quad \text{and} \quad \frac{1}{v_{T,\gamma}}\sum_{t=1}^T\left\{y_t^2 I_{T,t}^{(y)} - \gamma^0\right\} \xrightarrow{d} N(0,1) \quad \text{where } v_{T,\gamma}/T^{1/2} \to \infty. \tag{vi}$$

We prove all claims in Step 3 below

The Q-TTVT first order condition is $\sum_{t=1}^{T}\mathcal{E}_t(\hat{\gamma}_T^*, \hat{\xi}_T) = 0$. The mean-value-theorem, coupled with (i), (ii), (iv) and (v) leads to (see equation (A.2) in FHZ)

$$0 = \frac{1}{T}\sum_{t=1}^{T}\mathcal{E}_t - \mathcal{J}^{(\xi,\xi)}\left(\hat{\xi}_T - \xi^0\right) - \mathcal{J}^{(\xi,\gamma)}\left(\hat{\gamma}_T^* - \gamma^0\right)(1 + o_p(1))$$

hence

$$\hat{\xi}_T - \xi^0 \sim \mathcal{J}^{-1}\frac{1}{T}\sum_{t=1}^{T}\mathcal{E}_t - \mathcal{J}^{-1}\mathcal{K}\left(\hat{\gamma}_T^* - \gamma^0\right) \quad \text{where} \quad \mathcal{J} = \mathcal{J}^{(\xi,\xi)} \quad \text{and } \mathcal{K} = \mathcal{J}^{(\xi,\gamma)}.$$

Since by (vi) we have $1/T^{1/2}\sum_{t=1}^{T}\mathcal{E}_t = O_p(1)$ and $v_{T,\gamma}/T^{1/2} \to \infty$ it follows $\sum_{t=1}^{T}\mathcal{E}_t/v_{T,\gamma} = o_p(1)$, hence

$$\frac{T}{v_{T,\gamma}}\left(\hat{\varphi}_T - \varphi^0\right) \sim \begin{bmatrix} 1 \\ -\mathcal{J}^{-1}\mathcal{K} + o_p(1) \end{bmatrix} \frac{T^{1/2}}{v_{T,\gamma}}\left(\hat{\gamma}_T^* - \gamma^0\right).$$

Invoke Lemmas A.1 and A.2 to deduce

$$\begin{aligned} \frac{T^{1/2}}{v_{T,\gamma}}\left(\hat{\gamma}_T^* - \gamma^0\right) &= \frac{T^{1/2}}{v_{T,\gamma}}\left(\frac{1}{T}\sum_{t=1}^{T}\left\{y_t^2 I_{T,t}^{(y)} - E\left[y_t^2 I_{T,t}^{(y)}\right]\right\} + E\left[y_t^2 I_{T,t}^{(y)}\right] + E\left[y_t^2\left(1 - I_{T,t}^{(y)}\right)\right] - \gamma^0\right) \\ &= \frac{T^{1/2}}{v_{T,\gamma}}\left(\frac{1}{T}\sum_{t=1}^{T}\left\{y_t^2 I_{T,t}^{(y)} - E\left[y_t^2 I_{T,t}^{(y)}\right]\right\}\right). \end{aligned}$$

The claim now follows from (vi).

Step 3 (claims i-vi): Claims (i)-(iv) only require $E[\epsilon_t^4] < \infty$ and $E[y_t^2] < \infty$ so arguments in FHZ carry over verbatim.

Limit (v) follows from consistency $\hat{\varphi}_T \xrightarrow{p} \varphi^0$, $E[\epsilon_t^4] < \infty$, and the fact that there are GARCH effects since therefore $(\partial/\partial a)(\partial/\partial b)l_t$ is L_2-bounded and ergodic for $a, b \in \{\gamma, \xi, \varphi\}$.

Finally, for (vi) in view of Assumption A4 we may substitute $1/v_{T,\gamma}\sum_{t=1}^{T}\{y_t^2 I_{T,t}^{(y)} - \gamma^0\}$ for $1/v_{T,\gamma}\sum_{t=1}^{T}\{y_t^2 I_{T,t}^{(y)} - E[y_t^2 I_{T,t}^{(y)}]\}$. Then $v_{T,\gamma}^{-1}\sum_{t=1}^{T}\{y_t^2 I_{T,t}^{(y)} - E[y_t^2 I_{T,t}^{(y)}]\} \xrightarrow{d} N(0,1)$, $T/v_{T,\gamma} \to \infty$ and $T/v_{T,\gamma} = o(T^{1/2})$ follow from Lemma A.3. Further, $\mathcal{J}^{-1/2}T^{-1/2}\sum_{t=1}^{T}\mathcal{E}_t \xrightarrow{d} N(0, I_2)$ by a martingale difference CLT and the fact that $\mathcal{E}_t = (\epsilon_t^2 - 1)s_t$ is stationary and L_2-bounded (see Billingsley 1961, McLeish 1974). Therefore $T^{-1/2}\sum_{t=1}^{T}\mathcal{E}_t =$

$O_p(1)$. $\mathcal{QED}$.

PROOF OF THEOREM 3.4.

Step 1 (consistency): Consistency follows from a straightforward generalization of FHZ's (Appendix A1) argument.

Step 2 (normality): Define

$$\mathcal{H}^{<a,b>} := J^{<a>\prime} \times \Upsilon \times J^{<b>}.$$

Under the present DGP consistency of a sample Jacobian with a consistent plug-in follows from well known arguments (cf. Francq and Zakoïan 2004, 2010, Hill 2011).

Since $\hat{\xi}_T$ optimizes the GMTTM criterion, by Jacobian and GMM weight consistency we have

$$(1 + o_p(1)) \times J^{<\xi>\prime}\Upsilon \frac{1}{T}\sum_{t=1}^{T} m_t(\hat{\gamma}_T^*, \hat{\xi}_T) = 0.$$

Now apply a first order expansion around $\varphi^0 = [\gamma^0, \xi^{0\prime}]'$ and Jacobian consistency to solve for $\hat{\xi}_T - \xi^0$:

$$\hat{\xi}_T - \xi^0 = \left(\left(\mathcal{H}^{<\xi,\xi>}\right)^{-1} J^{<\xi>\prime}\Upsilon \frac{1}{T}\sum_{t=1}^{T} m_t - \left(\mathcal{H}^{<\xi,\xi>}\right)^{-1} \mathcal{H}^{<\xi,\gamma>} \left(\hat{\gamma}_T^* - \gamma^0\right) \right) \times (1 + o_p(1)).$$

The $o_p(1)$ captures Jacobian consistency. In the following we drop it for clarity. Therefore

$$\hat{\varphi}_T - \varphi^0 = \begin{bmatrix} 1 & 0 \\ -\left(\mathcal{H}^{<\xi,\xi>}\right)^{-1}\mathcal{H}^{<\xi,\gamma>} & \left(\mathcal{H}^{<\xi,\xi>}\right)^{-1} J^{<\xi>\prime}\Upsilon \end{bmatrix} \times \begin{bmatrix} \hat{\gamma}_T^* - \gamma^0 \\ \frac{1}{T}\sum_{t=1}^{T} m_t \end{bmatrix} = \mathcal{A} \times \begin{bmatrix} \hat{\gamma}_T^* - \gamma^0 \\ \frac{1}{T}\sum_{t=1}^{T} m_t \end{bmatrix},$$

say. Now use equation (37) in Horváth et al (2006) to write

$$\hat{\gamma}_T^* - \gamma^0 = \left(\frac{1-\beta^0}{\eta^0}\right) \frac{1}{T}\sum_{t=1}^{T} \left(\epsilon_t^2 - 1\right)\sigma_t^2 + o_p(1/T^{1/2}).$$

Since $m_t = (\epsilon_t^2 - 1)x_t$ it therefore follows

$$\hat{\varphi}_T - \varphi^0 = \mathcal{A} \times \frac{1}{T}\sum_{t=1}^{T}\left(\epsilon_t^2 - 1\right)\left[\begin{array}{c} \left(\dfrac{1-\beta^0}{\eta^0}\right)\sigma_t^2 \\ x_t \end{array}\right] + o_p(1/T^{1/2}) = \mathcal{A} \times \frac{1}{T}\sum_{t=1}^{T}\left(\epsilon_t^2 - 1\right)\mathcal{W}_t + o_p(1/T^{1/2}).$$

Since $\{(\epsilon_t^2 - 1)\mathcal{W}_t, \Im_t\}$ is a martingale difference we may apply McLeish's (1974: Theorem 2.3) central limit theorem to deduce as claimed

$$T^{1/2}\left(\hat{\varphi}_T - \varphi^0\right) \overset{d}{\rightarrow} N\left(0, \left(E\left[\epsilon_t^4\right] - 1\right) \times \mathcal{A}\left[\begin{array}{cc} b & \left(\dfrac{1-\beta^0}{\eta^0}\right) E\left[x_t'\sigma_t^2\right] \\ \left(\dfrac{1-\beta^0}{\eta^0}\right) E\left[x_t\sigma_t^2\right] & E\left[x_t x_t'\right] \end{array}\right]\mathcal{A}'\right).$$

$\mathcal{QED}$.

PROOF OF THEOREM 3.5.

Step 1 (consistency): Since $E[y_t^2] < \infty$ the argument is the same as that used in the proof of Theorem 3.1.

Step 2 (normality): Throughout we exploit weight property A5, and asymptotic theory for tail-trimmed estimating and Jacobian equations developed in HR (2010) and Hill (2010). Consult the appendices of those sources.

Stack the estimating equation components

$$\mathcal{Z}_{T,t}^* := \left[y_{T,t}^{*2} - E\left[y_{T,t}^{*2}\right], \left(m_{T,t}^* - E\left[m_{T,t}^*\right]\right)'\right]',$$

define the covariance matrix

$$\Sigma_T = E\left[\left(e_{T,t}^{*2} - E\left[e_{T,t}^{*2}\right]\right)^2\right] \times E\left[x_{T,t}^* x_{T,t}^{*\prime}\right]$$

and define

$$\mathcal{C}_T := E\left(\sum_{t=1}^{T}\mathcal{Z}^*_{T,t}\right)\left(\sum_{t=1}^{T}\mathcal{Z}^*_{T,t}\right)' = \begin{bmatrix} v^2_{T,\gamma} & \mathcal{C}^{(m,y)\prime}_T \\ \mathcal{C}^{(m,y)}_T & T\Sigma_T \end{bmatrix}$$

$$\mathcal{A}_T := \begin{bmatrix} 1 & 0 \\ -\left(\mathcal{H}^{<\xi,\xi>}_T\right)^{-1}\left(\mathcal{H}^{<\xi,\gamma>}_T\right) & \left(\mathcal{H}^{<\xi,\xi>}_T\right)^{-1} J^{<\xi>\prime}_T \Upsilon_T \end{bmatrix}.$$

Since $\hat{\xi}_T$ optimizes the GMTTM criterion, by Jacobian consistency and $\hat{\Upsilon}_T\Upsilon_T^{-1} \xrightarrow{p} I_q$ we have

$$0 = (1 + o_p(1)) \times J^{<\xi>\prime}_T \times \Upsilon_T \times \frac{1}{T}\sum_{t=1}^{T} m^*_{T,t}(\hat{\gamma}^*_T, \hat{\xi}_T)$$

and by an asymptotic first order expansion around $\varphi^0 = [\gamma^0, \xi^{0\prime}]'$,

$$\hat{\xi}_T - \xi^0 = \left(\mathcal{H}^{<\xi,\xi>}_T\right)^{-1} J^{<\xi>\prime}_T \Upsilon_T \frac{1}{T}\sum_{t=1}^{T} m^*_{T,t} - \left(\mathcal{H}^{<\xi,\xi>}_T\right)^{-1} \mathcal{H}^{<\xi,\gamma>}_T\left(\hat{\gamma}^*_T - \gamma^0\right).$$

Hence by the definitions of $\hat{\varphi}_T = [\hat{\gamma}^*_T, \hat{\xi}^{\prime}_T]'$, $\mathcal{Z}^*_{T,t} = [y^{*2}_{T,t} - E[y^{*2}_{T,t}], m^{*\prime}_{T,t}]'$ and $\mathcal{A}_T$

$$T^{1/2}\left(\mathcal{A}_T\mathcal{C}_T/T\mathcal{A}'_T\right)^{-1/2}\left(\hat{\varphi}_T - \varphi^0\right) = \left(\mathcal{A}_T\mathcal{C}^{1/2}_T\mathcal{A}'_T\right)^{-1/2}\mathcal{A}_T\mathcal{C}^{1/2}_T\mathcal{C}^{-1/2}_T\sum_{t=1}^{T}\mathcal{Z}^*_{T,t}.$$

The claim now follows from the definition of $\mathfrak{V}_T$ and a the CLT for tail-trimmed vectors Lemma B.6 in Hill and Aguilar (2010): $\mathcal{C}^{-1/2}_T\sum_{t=1}^{T}\mathcal{Z}^*_{T,t} \xrightarrow{d} N(0, I_{q+1})$. $\mathcal{QED}$.

PROOF OF COROLLARY 3.6. If $E[y^4_t] < \infty$ then the claim follows from Theorem 3.3 and dominated convergence.

Let $E[y^4_t] = \infty$ and assume exact identification $x_t = s_t$ with $\Upsilon_T = \Sigma^{-1}_T$. Notice $E[s_t s'_t] = \mathcal{J}$ by construction. Define

$$\mathcal{H}^{<a,b>}_T := J^{<a>\prime}_T \times \Upsilon_T \times J^{<b>}_T \text{ for any } a, b \in \{\gamma, \xi\},$$

where under exact identification

$$J_T^{<\xi>} = E\left[\frac{1}{\sigma_t^4}\frac{\partial}{\partial \xi}\sigma_t^2 \frac{\partial}{\partial \xi'}\sigma_t^2 I_{T,t}^{(x)}\right] \text{ and } J_T^{<\gamma>} = E\left[\frac{1}{\sigma_t^4}\frac{\partial}{\partial \xi}\sigma_t^2 \frac{\partial}{\partial \gamma}\sigma_t^2 I_{T,t}^{(x)}\right].$$

We maintain $E[\epsilon_t^4] < \infty$ hence $E[m_{i,t}^2] < \infty$ and $1/T^{1/2}\sum_{t=1}^T \{m_{T,t}^* - E[m_{T,t}^*]\} = O_p(1)$. Therefore

$$\Sigma_T \to \Sigma : \left(E\left[e_t^4\right] - 1\right) \times E\left[s_t s_t'\right] = \left(E\left[e_t^4\right] - 1\right) \times \mathcal{J}.$$

$$\mathcal{H}_T^{<a,b>} \to \mathcal{H}^{<a,b>} = \left(E\left[e_t^4\right] - 1\right)^{-1} \times J_T^{<a>\prime}\mathcal{J}^{-1}J_T^{<b>},$$

hence $\mathcal{H}^{<\xi,\xi>} \to (E[\epsilon_t^4] - 1)^{-1} \times \mathcal{J}$ and $\mathcal{H}^{<\xi,\gamma>} \to (E[\epsilon_t^4] - 1) \times \mathcal{K}$ by the definitions of $\mathcal{J}$ and $\mathcal{K}$.

The condition $E[y_t^4] = \infty$ implies $\upsilon_{T,\gamma}^2/T \to \infty$ and $\Sigma_T \sim \Sigma$, hence $\sum_{t=1}^T \{m_{T,t}^* - E[m_{T,t}^*]\} = o_p(\upsilon_{T,\gamma})$ and by dominated convergence

$$\mathcal{C}_T/\upsilon_{T,\gamma}^2 = \begin{bmatrix} 1 & \mathcal{C}_T^{(m,y)\prime}/\upsilon_{T,\gamma}^2 \\ \mathcal{C}_T^{(m,y)}/\upsilon_{T,\gamma}^2 & T\Sigma_T/\upsilon_{T,\gamma}^2 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

Since $\mathcal{H}^{<\xi,\xi>} = J_T^{<\xi>\prime}\Upsilon_T J_T^{<\xi>}$ and $J_T^{<\xi>}$ and Υ_T are invertible, it is then easily verified that

$$\begin{aligned}\frac{T}{\upsilon_{T,\gamma}^2}\mathfrak{V}_T &\sim \begin{bmatrix} 1 & -\mathcal{H}^{<\xi,\gamma>\prime}\left(\mathcal{H}^{<\xi,\xi>}\right)^{-1} \\ -\left(\mathcal{H}^{<\xi,\xi>}\right)^{-1}\mathcal{H}^{<\xi,\gamma>} & \left(\mathcal{H}^{<\xi,\xi>}\right)^{-1}T/\upsilon_{T,\gamma}^2 + \left(\mathcal{H}^{<\xi,\xi>}\right)^{-1}\mathcal{H}^{<\xi,\gamma>}\mathcal{H}^{<\xi,\gamma>\prime}\left(\mathcal{H}^{<\xi,\xi>}\right)^{-1} \end{bmatrix} \\ &\sim \begin{bmatrix} 1 & -\mathcal{K}'\mathcal{J}^{-1} \\ -\mathcal{J}^{-1}\mathcal{K} & \mathcal{J}^{-1}\mathcal{K}\mathcal{K}'\mathcal{J}^{-1} \end{bmatrix}.\end{aligned}$$

Therefore the GTT-TTVT estimator satisfies

$$\frac{T}{\upsilon_{T,\gamma}}\left(\hat{\varphi}_T - \varphi^0\right) \xrightarrow{d} N\left(0, \begin{bmatrix} 1 & -\mathcal{K}'\mathcal{J}^{-1} \\ -\mathcal{J}^{-1}\mathcal{K} & \mathcal{J}^{-1}\mathcal{K}\mathcal{K}'\mathcal{J}^{-1} \end{bmatrix}\right). \ \mathcal{QED}.$$

Appendix B : Proofs of Lemma A.1

We require one supporting result.

LEMMA B.1. *Let Assumptions A1 and A2 hold.*

a. $v_{T,\gamma}^2/T = r_T E[y_{T,t}^{*4}] = o(T)$ *where in general* $\liminf_{T\to\infty} r_T > 0$ *and* $r_T = O(\ln(T))$ *does not depend on* k_T, $r_T \sim 1$ *if* y_t *is iid, and* $r_T \sim K$ *if* $E[y_t^2] < \infty$.

b. $E[y_{T,t}^{*4}]/(v_{T,\gamma}^2/T) = O(1)$.

c. If $\kappa = 2$ *then* $E[y_{T,t}^{*4}] \sim L(T) \to \infty$ *is slowly varying, and if* $\kappa \in (2,4)$ *then* $E[y_{T,t}^{*4}] \sim K(T/k_T^{(y)})^{4/\kappa - 1}$.

d. The thresholds are bounded: $c_T^{(y)} = O((v_{T,\gamma}/k_T^{(y)})^{1/2})$.

PROOF. Write $k_T = k_T^{(y)}$ and $c_T = c_T^{(y)}$, and define $\mathcal{Y}_{T,t}^* := y_{T,t}^{*2} - E[y_{T,t}^{*2}]$. Recall geometric β-mixing implies geometric α-mixing with coefficients $\alpha_h \leq K\rho^h$ for $\rho \in (0,1)$. Assume $\alpha_h = \rho^h$ without loss of generality.

Claim (a). Note $v_{T,\gamma}^2/T \sim E[\mathcal{Y}_{T,t}^{*2}] + 2\sum_{i=1}^{T-1}(1 - i/T)E[\mathcal{Y}_{T,1}^*\mathcal{Y}_{T,i+1}^*]$. If $E[y_t^4] < \infty$ then $v_{T,\gamma}^2/T \sim K$ under geometric a-mixing (c. Ibragimov 1962). If y_t is iid then $v_{T,\gamma}^2/T = E[\mathcal{Y}_{T,t}^{*2}]$ which has a finite limit if $E[y_t^4] < \infty$.

Finally, assume $E[y_t^4] = \infty$ hence $\kappa \in (2,4]$. We will prove $v_{T,\gamma}^2/T = L(T) \times E[\mathcal{Y}_{T,t}^{*2}]$, where $L(T) \times E[\mathcal{Y}_{T,t}^{*2}]$ $= o(T)$ by claim (c).

Define the quantile functions $Q_T(u) = \inf\{y : P(y_{T,t}^{*2} > y) \leq u\}$ and $Q(u) = \inf\{y : P(y_t^2 > y) \leq u\}$ for $u \in [0,1]$. By Theorem 1.1 of Rio (1993)

$$\sum_{i=1}^{T-1} \left|E\left[\mathcal{Y}_{T,1}^*\mathcal{Y}_{T,i+1}^*\right]\right| \leq K\sum_{i=1}^{T-1}\int_0^{\alpha_h} Q_T^2(u)du \leq K\sum_{i=1}^{T-1}\int_0^{\rho^h} Q_T^2(u)du.$$

By tail-trimming $Q_T(u) = 0$ $\forall u \leq k_T/T$ and $Q_T(u) = Q(u)$ otherwise. Further, under tail decay (14) $Q(u) = O(u^{-2/\kappa})$. Therefore

$$\begin{aligned}\sum_{i=1}^{T-1} \left|E\left[\mathcal{Y}_{T,1}^*\mathcal{Y}_{T,i+1}^*\right]\right| &\leq K\sum_{i=1}^{T-1}\int_{k_T/T}^{\rho^h} u^{-4/\kappa}du = K\sum_{i=1}^{T-1}\max\left\{0, (T/k_T)^{(4/\kappa-1)} - \rho^{-i(4/\kappa-1)}\right\} \\ &= K\sum_{i=1}^{\ln(T/k_T)}\left\{(T/k_T)^{(4/\kappa-1)} - \rho^{-i(4/\kappa-1)}\right\}.\end{aligned}$$

Finally, $\sum_{i=1}^{\ln(T/k_T)}\{(T/k_T)^{4/\kappa-1} - \rho^{-i(4/\kappa-1)}\} \leq K\ln(T/k_T) \times (T/k_T)^{4/\kappa-1}$. If $\kappa \in (2,4)$ then by (14) and therefore Karamata's Theorem (Resnick 1987: Theorem 0.6) $E[y_{T,t}^{*4}] \sim Kc_T^4(k_T/T) \sim K(T/k_T)^{4/\kappa-1}$, hence $\sum_{i=1}^{T-1}|E[\mathcal{Y}_{T,1}^*\mathcal{Y}_{T,i+1}^*]| \leq K\ln(T) \times E[y_{T,t}^{*4}]$. Similarly if $\kappa = 4$ then $E[y_{T,t}^{*4}] \sim L(T)$ hence $\sum_{i=1}^{T-1}|E[\mathcal{Y}_{T,1}^*\mathcal{Y}_{T,i+1}^*]|$ $\leq K(\ln(T)/L(T)) \times E[y_{T,t}^{*4}] = L(T) \times E[y_{T,t}^{*4}]$.

Claim (b). The claim follows instantly from (a).

Claim (c)**.** Each moment property is derived above in the claim (a) proof.

Claim (d)**.** See Lemma C.1 in Hill and Aguilar (2010b). $\mathcal{QED}$.

We are now ready to prove Lemma A.1.

PROOF OF LEMMA A.1. We prove the claim for $\hat{\gamma}_T^{(g)}$, and then show $\hat{\gamma}_T^{(o)}$ satisfies the same argument.

Step 1: Note

$$\hat{\gamma}_T^{(g)} = \hat{\gamma}_T^{(tr)} + E\left[y_t^2 I\left(|y_t| \geq c_T\right)\right] + \hat{\mathcal{R}}_T^{(g)} - E\left[y_t^2 I\left(|y_t| \geq \tilde{c}_T\right)\right] + E\left[y_t^2 I\left(\tilde{c}_T < |y_t| < c_T\right)\right].$$

By Assumption A4 $(T/v_{T,\gamma})E[y_t^2 I(\tilde{c}_T < |y_t| < c_T)] \to 0$ hence we need only show $(T/v_{T,\gamma})(\hat{\mathcal{R}}_T^{(g)} - E[y_t^2 I(|y_t| \geq \tilde{c}_T)]) \xrightarrow{p} 0$. Drop all superscripts: $k_T = k_T^{(y)}$, $c_T = c_T^{(y)}$, $y_{(\tilde{k}_T)} = y_{(\tilde{k}_T)}^{(a)}$, etc. Under Assumptions A1, A2, A5 and A6 $\hat{\kappa}_{\check{k}_T} = \kappa + O_p(1/\check{k}_T^{1/2})$ and $\tilde{k}_T^{1/2}((y_{(\tilde{k}_T)}/\tilde{c}_T)^2 - 1) = O_p(1)$ by Theorem 3 in Hill (2010), hence

$$\begin{aligned}
\hat{\mathcal{R}}_T^{(g)} - E\left[y_t^2 I\left(|y_t| \geq \tilde{c}_T\right)\right] &= \frac{\hat{\kappa}_{\check{k}_T}}{\hat{\kappa}_{\check{k}_T} - 2}\frac{\tilde{k}_T}{T}\left(y_{(\tilde{k}_T)}\right)^2 - E\left[y_t^2 I\left(|y_t| \geq \tilde{c}_T\right)\right] \\
&= \frac{\kappa}{\kappa-1}\frac{\tilde{k}_T}{T}\tilde{c}_T^2 - E\left[y_t^2 I\left(|y_t| \geq \tilde{c}_T\right)\right] + \frac{\kappa}{\kappa-1}\frac{\tilde{k}_T^{1/2}}{T}\tilde{c}_T^2 \times \tilde{k}_T^{1/2}\left(\left(y_{(\tilde{k}_T)}/\tilde{c}_T\right)^2 - 1\right) \\
&\quad + \left(\frac{\hat{\kappa}_{\check{k}_T}}{\hat{\kappa}_{\check{k}_T} - 2} - \frac{\kappa}{\kappa-1}\right)\frac{\tilde{k}_T^{1/2}}{T}\tilde{c}_T^2 \times \tilde{k}_T^{1/2}\left(\left(y_{(\tilde{k}_T)}/\tilde{c}_T\right)^2 - 1\right) \\
&\quad + \left(\frac{\hat{\kappa}_{\check{k}_T}}{\hat{\kappa}_{\check{k}_T} - 2} - \frac{\kappa}{\kappa-1}\right)\frac{\tilde{k}_T}{T}\tilde{c}_T^2 \\
&= \left\{\frac{\kappa}{\kappa-1}\frac{\tilde{k}_T}{T}\tilde{c}_T^2 - E\left[y_t^2 I\left(|y_t| \geq \tilde{c}_T\right)\right]\right\} + O_p\left(\frac{\tilde{k}_T^{1/2}}{T}\tilde{c}_T^2\right) + O_p\left(\check{k}_T^{-1/2}\frac{\tilde{k}_T}{T}\tilde{c}_T^2\right) \\
&= \mathcal{A}_{1,T} + \mathcal{A}_{2,T} + \mathcal{A}_{3,T}.
\end{aligned}$$

Under power-law decay A5 and the fact that $T/\upsilon_{T,\gamma} \geq KT^{1/2}/(E[y_{T,t}^{*4}])^{1/2}$ by Lemma B.1, it follows $(T/\upsilon_{T,\gamma})\mathcal{A}_{1,T}$ $\rightarrow 0$. See Peng (2001: p. 259-264).

Consider $\mathcal{A}_{2,T}$ and $\mathcal{A}_{3,T}$. Use power-law decay (14) and the threshold construction to deduce $c_T = K(T/k_T)^{1/\kappa}$. Further, by Lemma B.1 if $\kappa \in (2,4)$ then $\upsilon_{T,\gamma} \leq KT(E[y_{T,t}^{*4}])^{1/2}$ hence

$$\begin{aligned}\frac{T}{\upsilon_{T,\gamma}}\frac{\tilde{k}_T^{1/2}}{T}\tilde{c}_T &\leq \frac{\tilde{k}_T^{1/2}}{T^{1/2}\left(E\left[y_{T,t}^{*4}\right]\right)^{1/2}}\tilde{c}_T^2 \sim \frac{\tilde{k}_T^{1/2}}{T^{1/2}c_T^2\left(k_T/T\right)^{1/2}}\tilde{c}_T^2 = \left(\frac{\tilde{k}_T}{k_T}\right)^{1/2}\left(\frac{\tilde{c}_T}{c_T}\right)^2 \\ &= \left(\frac{\tilde{k}_T}{k_T}\right)^{1/2}\left(\frac{T/\tilde{k}_T}{T/k_T}\right)^{2/\kappa} = \left(\frac{k_T}{\tilde{k}_T}\right)^{2/\kappa-1/2}.\end{aligned}$$

By property (4) under Assumption A6 $k_T/\tilde{k}_T \rightarrow 0$ hence $(T/\upsilon_{T,\gamma})\mathcal{A}_{2,T} \xrightarrow{p} 0$. If $\kappa = 4$ then $(T/\upsilon_{T,\gamma})\mathcal{A}_{2,T}$ $\sim (T/\upsilon_{T,\gamma})\tilde{k}_T^{1/2}\tilde{c}_T/(TL(T)) = K(T/\tilde{k}_T)^{2/\kappa-1/2}/L(T) = 1/L(T) \rightarrow 0$ for slowly varying $L(T) \rightarrow \infty$, and if $\kappa >$ 4 then $(T/\upsilon_{T,\gamma})\mathcal{A}_{2,T} = K(T/\tilde{k}_T)^{2/\kappa-1/2} \xrightarrow{p} 0$.

Similarly if $\kappa \in (2,4)$ then

$$\frac{T}{\upsilon_{T,\gamma}}\check{k}_T^{-1/2}\frac{\tilde{k}_T}{T}\tilde{c}_T^2 = \frac{1}{\upsilon_{T,\gamma}}\frac{\tilde{k}_T}{\check{k}_T^{1/2}}\tilde{c}_T^2 \leq \frac{k_T^{1/2}\tilde{k}_T}{T^{1/2}\check{k}_T^{1/2}}\left(\frac{T/\tilde{k}_T}{T/k_T}\right)^{2/\kappa}.$$

By applying $k_T/\tilde{k}_T \rightarrow 0$ and $\tilde{k}_T/\check{k}_T \rightarrow 0$ under Assumption A6 it can be shown as above $(T/\upsilon_{T,\gamma})\mathcal{A}_{3,T} \xrightarrow{p} 0$, where the cases $\kappa \geq 4$ are similar.

Step 2: Define $\tilde{k}_T := \tilde{k}_T(\lambda)/\lambda$, $\check{k}_T := \check{k}_T(\lambda)/\lambda$ and $\sigma_T^2(\lambda) := E[(\check{k}_T^{1/2}(\lambda)(\hat{\kappa}^{-1}_{\check{k}_T^{1/2}(\lambda)} - \kappa^{-1}))^2]$. Let the threshold sequence $\{\tilde{c}_T(\lambda)\}$ satisfy $P(|y_t| > \tilde{c}_T(\lambda)) = \tilde{k}_T(\lambda)/T$. Let $\Longrightarrow$ denotes weak convergence on $\mathcal{C}[\underline{\lambda}, \bar{\lambda}]$ the space of real continuous functions.

Under the maintained assumptions and Theorem 5.1 in Hill (2009) there exists zero mean Gaussian processes $\{\mathcal{Z}_1(\lambda), \mathcal{Z}_2(\lambda) : \lambda \in [\underline{\lambda}, \bar{\lambda}]\}$ on $\mathcal{C}[\underline{\lambda}, \bar{\lambda}]$ that satisfy

$$\left\{\check{k}_T^{1/2}\left(\hat{\kappa}^{-1}_{\check{k}_T^{1/2}(\lambda)} - \kappa^{-1}\right) : \lambda \in [\underline{\lambda}, \bar{\lambda}]\right\} \Longrightarrow \left\{\mathcal{Z}_1(\lambda) : \lambda \in [\underline{\lambda}, \bar{\lambda}]\right\}$$

$$\left\{\tilde{k}_T^{1/2}\left(\frac{y^{(a)}_{(\tilde{k}_T^{(y)}(\lambda))}}{\tilde{c}_T(\lambda)} - 1\right) : \lambda \in [\underline{\lambda}, \bar{\lambda}]\right\} \Longrightarrow \left\{\mathcal{Z}_2(\lambda) : \lambda \in [\underline{\lambda}, \bar{\lambda}]\right\},$$

hence $\sup_{\lambda \in [\underline{\lambda},\bar{\lambda}]} |y^{(a)}_{(\tilde{k}^{(y)}_T(\lambda))}/\tilde{c}_T(\lambda) - 1| = O_p(1/\tilde{k}_T^{1/2})$ and $\sup_{\lambda \in [\underline{\lambda},\bar{\lambda}]} |\hat{\kappa}^{-1}_{\check{k}_T^{1/2}(\lambda)} - \kappa^{-1}| = O_p(1/\check{k}_T^{1/2})$. In particular since $\tilde{\lambda}^* \in [\underline{\lambda}, \bar{\lambda}]$ we have $\tilde{k}_T^{1/2}(\tilde{\lambda}^*)|y^{(a)}_{(\tilde{k}^{(y)}_T(\tilde{\lambda}^*))}/\tilde{c}_T(\tilde{\lambda}^*) - 1| \leq K\tilde{k}_T^{1/2}|y^{(a)}_{(\tilde{k}^{(y)}_T(\tilde{\lambda}^*))}/\tilde{c}_T(\tilde{\lambda}^*) - 1| = O_p(1)$. The argument used in Step 1 now carries over verbatim. $\mathcal{QED}$.

REFERENCES

[1] Andrews, D.W.K. (1988). Laws of Large Numbers for Dependent Non-Identically Distributed Random Variables, Econometric Theory 4, 458-467.

[2] Antoine, B., and E. Renault (2011). Efficient Minimum Distance Estimation with Multiple Rates of Convergence, Journal of Econometrics: forthcoming.

[3] Basrak, B., R.A. Davis, and T. Mikosch (2002). Regular Variation of GARCH Processes, Stochastic Processes and their Applications 99, 95-115.

[4] Billingsley, P. (1961). The Lindeberg-Levy Theorem for Martingales, Proceedings of the American Mathematical Society 12, 788–792.

[5] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307–327.

[6] Davis, R.A. (2010). Heavy Tails in Financial Time Series, In Cont Rama (ed.), Encyclopedia of Quantitative Finance. Wiley, New York.

[7] Engle, R.F. and J. Mezrich (1996). GARCH for Groups. Risk 9, 36–40.

[7] Finkenstadt, B., and H. Rootzén (2003). Extreme Values in Finance, Telecommunications and the Environment. Chapman and Hall: New York.

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

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

[10] Francq, C. and J-M. Zakoïan (2010). GARCH Models: Structure, Statistical Inference and Financial Applications, Wiley: New York.

[11] Francq, C., L. Horvath and J-M. Zakoïan (2010). Merits and Drawbacks of Variance Targeting in GARCH Models, working paper.

[12] Haeusler, E. and J.L. Teugels (1985). On Asymptotic Normality of Hill's Estimator for the Exponent of Regular Variation, Annals of Statistics 13, 743-756.

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

[14] Hansen, L., J. Heaton, and M. Ogaki (1988). Efficiency Bounds Implied by Multiperiod Conditional Moment Restrictions, Journal of the American Statistical Association 83, 863-871.

[15] Hansen, L.P., J. Heaton, and A. Yaron (1996). Finite-Sample Properties of Some Alternative GMM Estimators, Journal of Business and Economic Statistics 14, 262-280.

[16] He, C. and T. Teräsvirta (1999). Fourth Moment Structure of the GARCH(p, q) Process, Econometric Theory 15, 824-846.

[17] Hill, J.B. (2009). On Functional Central Limit Theorems for Dependent, Heterogeneous Arrays with Applications to Tail Index and Tail Dependence Estimation, Journal of Statistical Planning and Inference 139, 2091-2110.

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

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

[20] Hill, J.B. and M. Aguilar (2010). Moment Condition Tests for Heavy Tailed Time Series, Journal of Econometrics:

Annals Issue on Extreme Value Theory: forthcoming.

[21] Hill, J.B. and E. Renault (2011). Generalized Method of Moments with Tail Trimming, Dept. of Economics, University of North Carolina - Chapel Hill.

[22] Horváth, L., P. Kokoszka, and R. Zitikis (2006). Sample and Implied Volatility in GARCH models, Journal of Financial Econometrics 4, 617–635.

[23] Kesten, H. (1973) Random Difference Equations and Renewal Theory for Products of Random Matrices. Acta Mathemaetica 131, 207–248.

[24] Kristensen, D. and O. Linton (2004). Consistent Standard Errors for Target Variance Approach to GARCH Estimation, Econometric Theory 20, 990-993.

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

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

[27] McLeish, D.L. (1974). Dependent Central Limit Theorems and Invariance Principles, Annals of Probability 2, 620-628.

[28] Meddahi, N. and E. Renault (1998). Quadratic M-Estimators for ARCH-Type Processes, Working Paper 98-29, University of Montreal and Cirano.

[29] Peng, L. (2001). Estimating the Mean of a Heavy Tailed Distribution, Statistics and Probability Letters 52, 255-264.

[30] Prono, T. (2011). Simple GMM Estimation of the Semi-Strong GARCH(1,1) Model, Working Paper, Commodity Futures Trading Commission.

[31] Resnick, S.I. (1987). Extreme Values, Regular Variation and Point Processes. Springer-Verlag: New York.

[32] Rio, E. (1993). Covariance Inequalities for Strongly Mixing Processes, Annals of the Institute of Henri Poincaré 29, 587-597.

[33] Straumann, D. and T. Mikosch (2006). Quasi-Maximum-Likelihood Estimation in Conditionally Heteroscedastic Time Series: A Stochastic Recurrence Equations Approach, Annals of Statistics 34, 2449-2495.

TABLE 2 : $T = 250$

$\epsilon_t \sim P_{2.5}$: $\kappa_\epsilon = 2.5$									
Param	κ_y	Estimator[a]	Norm[b]	Bias	RMSE[c]	KS[d]	Null[e]	ALT1	ALT2
$\beta^0 = .4$	2.71	QML		.017	.162	1.39	.01,.04,.11	.06,.17,.27	.53,.73,.81
		Q-VT		.019	.220	1.46	.00,.02,.08	.01,.12,.25	.30,.50,.59
		Q-TTVT		.026	.206	1.34	.00,.03,.10	.03,.17,.25	.33,.53,.67
		G-TTVT	*	-.005	.202	1.10	.00,.03,.08	.02,.13,.23	.30,.51,.61
$\beta^0 = .6$	2.33	QML		-.016	.258	1.24	.00,.06,.09	.10,.30,.41	.43,.65,.74
		Q-VT		-.037	.224	1.29	.00,.06,.12	.09,.30,.42	.47,.72,.78
		Q-TTVT		-.027	.218	.615	.01,.05,.11	.14,.35,.47	.55,.74,.84
		G-TTVT	*	.007	.149	.881	.00,.05,.11	.42,.66,.76	.90,.98,1.0
$\epsilon_t \sim P_{4.1}$: $\kappa_\epsilon = 4.1$									
Param	κ_y	Estimator	Norm	Bias	RMSE	KS	Null	ALT1	ALT2
$\beta^0 = .4$	3.17	QML	*	.022	.145	2.24	.00,.06,.11	.08,.23,.35	.66,.84,.90
		Q-VT		.014	.201	1.86	.00,.02,.08	.03,.15,.25	.35,.56,.66
		Q-TTVT	*	.003	.202	1.13	.00,.03,.09	.03,.14,.23	.31,.51,.63
		G-TTVT	*	-.009	.187	.689	.00,.03,.11	.02,.12,.21	.34,.57,.65
$\beta^0 = .6$	2.61	QML	*	-.016	.236	1.18	.00,.04,.09	.13,.32,.45	.50,.71,.78
		Q-VT		-.015	.226	1.29	.00,.03,.13	.39,.64,.76	.87,.97,.98
		Q-TTVT	*	-.014	.166	.748	.00,.05,.11	.33,.53,.66	.81,.91,.97
		G-TTVT	*	-.018	.145	1.15	.01,.06,.11	.42,.65,.76	.90,.97,1.0
$\epsilon_t \sim N(0,1)$: $\kappa_\epsilon = \infty$									
Param	κ_y	Estimator	Norm	Bias	RMSE	KS	Null	ALT1	ALT2
$\beta^0 = .4$	6.25	QML	*	.031	.099	3.19	.02,.05,.10	.17,.45,.63	.93,.97,.98
		Q-VT	*	-.030	.162	3.34	.01,.05,.10	.03,.09,.14	.31,.56,.70
		Q-TTVT	*	-.028	.160	2.76	.01,.05,.11	.03,.10,.16	.34,.58,.71
		G-TTVT	*	-.006	.169	1.07	.01,.04,.11	.06,.15,.22	.40,.63,.74
$\beta^0 = .6$	4.10	QML	*	-.037	.199	3.07	.01,.04,.12	.16,.29,.43	.58,.82,.89
		Q-VT	*	-.056	.099	3.78	.00,.04,.12	.52,.74,.83	1.0,1.0,1.0
		Q-TTVT	*	-.048	.087	3.26	.00,.04,.12	.69,.88,.95	1.0,1.0,1.0
		G-TTVT	*	-.019	.104	1.87	.01,.06,.11	.73,.86,.92	1.0,1.0,1.0

a. Q-VT = QML with Variance Targeting; Q-TTVT = QML with Tail-Trimmed Variance Targeting; G-TTVT = GMTTM with Tail-Trimmed Variance Targeting.

b. * denotes whether the estimator is asymptotically normal.

c.. RMSE is the root mean squared error.

d. KS is the Kolmogorov-Smirnov statistic for a test of standard normality. The values are scaled by the 5% critical value: KS ≤ 1 implies we fail to reject standard normality, and KS > 1 we reject at the 5% level.

e. The reported values are rejection frequencies at the 1%, 5% and 10% levels for Null: $\beta = \beta^0$, ALT1: $\beta = \beta^0 - .25$, and ALT1: $\beta = 0$.

TABLE 3 : $T = 500$

$\epsilon_t \sim P_{2.5}$: $\kappa_\epsilon = 2.5$									
Param	κ_y	Estimator[a]	Norm[b]	Bias	RMSE[c]	KS[d]	Null[e]	ALT1	ALT2
$\beta^0 = .4$	2.71	QML		.018	.158	1.65	.00,.04,.08	.03,.18,.26	.56,.72,.81
		Q-VT		.021	.199	1.23	.00,.02,.10	.03,.16,.27	.41,.58,.68
		Q-TTVT		.003	.207	.108	.00,.03,.09	.02,.13,.22	.31,.50,.61
		G-TTVT	*	.001	.195	.788	.00,.04,.11	.02,.13,.21	.34,.55,.66
$\beta^0 = .6$	2.33	QML		-.012	.254	1.34	.00,.07,.12	.10,.29,.42	.44,.67,.76
		Q-VT		-.002	.205	1.10	.01,.04,.09	.19,.45,.58	.68,.82,.87
		Q-TTVT		-.008	.195	.739	.00,.06,.09	.15,.37,.52	.58,.77,.84
		G-TTVT	*	.009	.128	.988	.01,.05,.11	.60,.81,.87	.97,1.0,1.0
$\epsilon_t \sim P_{4.1}$: $\kappa_\epsilon = 4.1$									
Param	κ_y	Estimator	Norm	Bias	RMSE	KS	Null	ALT1	ALT2
$\beta^0 = .4$	3.17	QML	*	-.020	.151	1.42	.00,.08,.12	.03,.10,.20	.52,.73,.81
		Q-VT		.013	.199	1.33	.00,.03,.09	.03,.13,.22	.33,.57,.68
		Q-TTVT	*	-.003	.199	.973	.00,.03,.10	.04,.12,.20	.29,.53,.64
		G-TTVT	*	-.000	.171	.761	.00,.04,.11	.06,.15,.24	.42,.64,.74
$\beta^0 = .6$	2.61	QML	*	-.014	.224	.917	.01,.05,.09	.16,.34,.46	.53,.75,.83
		Q-VT		-.031	1.67	1.21	.01,.06,.11	.23,.50,.63	.78,.88,.93
		Q-TTVT	*	-.008	.161	.643	.01,.07,.11	.34,.62,.74	.86,.92,.95
		G-TTVT	*	.002	.115	.913	.01,.06,.11	.71,.85,.91	1.0,1.0,1.0
$\epsilon_t \sim N(0,1)$: $\kappa_\epsilon = \infty$									
Param	κ_y	Estimator	Norm	Bias	RMSE	KS	Null	ALT1	ALT2
$\beta^0 = .4$	6.25	QML	*	-.003	.118	1.51	.02,.06,.10	.07,.23,.43	.81,.89,.93
		Q-VT	*	-.027	.131	1.59	.02,.05,.10	.03,.14,.23	.58,.80,.88
		Q-TTVT	*	-.025	.121	1.66	.02,.06,.10	.04,.14,.24	.67,.86,.92
		G-TTVT	*	.004	.136	.389	.01,.04,.10	.09,.21,.31	.64,.84,.91
$\beta^0 = .6$	4.10	QML	*	.009	1.72	1.87	.01,.08,.14	.24,.49,.66	.89.95,.98
		Q-VT	*	-.024	.057	2.10	.00,.02,.11	1.0,1.0,1.0	1.0,1.0,1.0
		Q-TTVT	*	-.026	.072	1.92	.00,.06,.12	.97,1.0,1.0	1.0,1.0,1.0
		G-TTVT	*	-.001	.072	.429	.02.06,.10	.98,1.0,1.0	1.0,1.0,1.0

a. Q-VT = QML with Variance Targeting; Q-TTVT = QML with Tail-Trimmed Variance Targeting; G-TTVT = GMTTM with Tail-Trimmed Variance Targeting.

b. * denotes whether the estimator is asymptotically normal.

c.. RMSE is the root mean squared error.

d. KS is the Kolmogorov-Smirnov statistic for a test of standard normality. The values are scaled by the 5% critical value: KS ≤ 1 implies we fail to reject standard normality.

e. The reported values are rejection frequencies at the 1%, 5% and 10% levels for Null: $\beta = \beta^0$, ALT1: $\beta = \beta^0 - .25$, and ALT1: $\beta = 0$.

TABLE 4 : $T = 1000$

$\epsilon_t \sim P_{2.5}$: $\kappa_\epsilon = 2.5$									
Param	κ_y	Estimator[a]	Norm[b]	Bias	RMSE[c]	KS[d]	Null[e]	ALT1	ALT2
$\beta^0 = .4$	2.71	QML		-.016	.161	2.18	.00,.07,.12	.02,.10,.21	.42,.68,.74
		Q-VT		.003	.195	1.16	.00,.03,.09	.02,.14,.23	.34,.54,.66
		Q-TTVT		.003	.198	.966	.00,.04,.10	.02,.14,.23	.33,.54,.67
		G-TTVT	*	.002	.171	.623	.00,.04,.11	.04,.15,.25	.42,.65,.75
$\beta^0 = .6$	2.33	QML		-.011	.234	1.58	.00,.05,.09	.14,.33,.45	.50,.72,.81
		Q-VT		-.006	.195	1.47	.02,.07,.10	.21,.47,.61	.73,.87,.90
		Q-TTVT		-.002	.193	1.37	.03,.06,.10	.22,.49,.63	.75,.87,.92
		G-TTVT	*	.014	.126	1.25	.01,.06,.11	.64,.83,.87	.98,1.0,1.0
$\epsilon_t \sim P_{4.1}$: $\kappa_\epsilon = 4.1$									
Param	κ_y	Estimator	Norm	Bias	RMSE	KS	Null	ALT1	ALT2
$\beta^0 = .4$	3.17	QML	*	-.021	.144	1.31	.00,.06,.12	.03,.11,.22	.58,.76,.81
		Q-VT		-.013	.180	1.42	.00,.04,.09	.05,.13,.20	.35,.55,.68
		Q-TTVT	*	-.012	.181	1.03	.01,.05,.11	.05,.13,.21	.36,.57,.69
		G-TTVT	*	.001	.153	.555	.01,.04,.11	.07,.18,.27	.52,.74,.84
$\beta^0 = .6$	2.61	QML	*	.010	.189	1.03	.01,.05,.11	,24,.48,.62	.75,.90,.94
		Q-VT		.012	.124	1.34	.01,.07,.08	.63,.82,.92	.98,1.0,1.0
		Q-TTVT	*	.007	.119	.473	.01,.05,.10	.67,.82,.90	1.0,1.0,1.0
		G-TTVT	*	-.006	.096	.872	.01,.06,.11	.83,.93,.95	1.0,1.0,1.0
$\epsilon_t \sim N(0,1)$: $\kappa_\epsilon = \infty$									
Param	κ_y	Estimator	Norm	Bias	RMSE	KS	Null	ALT1	ALT2
$\beta^0 = .4$	6.25	QML	*	-.002	.115	2.48	.03,.06,.09	.05,.27,.47	.84,.91,.93
		Q-VT	*	-.013	.095	2.02	.02,.05,.09	.10,.26,.38	.93,.97,.98
		Q-TTVT	*	-.008	.087	1.12	.01,.05,.10	.14,.33,.44	.96,.99,1.0
		G-TTVT	*	-.007	.103	.866	.01,.06,.10	.10,.30,.42	.90,.96,.98
$\beta^0 = .6$	4.10	QML	*	.018	.147	2.90	.02,.11,.14	.32,.72,.87	.97,.99,1.0
		Q-VT	*	-.019	.063	3.15	.02,.06,.09	.98,1.0,1.0	1.0,1.0,1.0
		Q-TTVT	*	-.005	.050	.892	.01,.04,.11	1.0,1.0,1.0	1.0,1.0,1.0
		G-TTVT	*	.003	.054	1.05	.01,.05,.10	1.0,1.0,1.0	1.0,1.0,1.0

a. Q-VT = QML with Variance Targeting; Q-TTVT = QML with Tail-Trimmed Variance Targeting; G-TTVT = GMTTM with Tail-Trimmed Variance Targeting.

b. * denotes whether the estimator is asymptotically normal.

c.. RMSE is the root mean squared error.

d. KS is the Kolmogorov-Smirnov statistic for a test of standard normality. The values are scaled by the 5% critical value: KS ≤ 1 implies we fail to reject standard normality.

e. The reported values are rejection frequencies at the 1%, 5% and 10% levels for Null: $\beta = \beta^0$, ALT1: $\beta = \beta^0 - .25$, and ALT1: $\beta = 0$.