Robust M-Estimation for Heavy Tailed Nonlinear AR-GARCH

Article *in* SSRN Electronic Journal · June 2011

DOI: 10.2139/ssrn.1861125

1 author:

Jonathan B Hill
University of North Carolina at Chapel Hill

Robust M-Estimation for Heavy Tailed Nonlinear AR-GARCH

(with Supplemental Appendix C)

Jonathan B. Hill[1]

Dept. of Economics, University of North Carolina

Nov. 28 2011

ABSTRACT

We develop new tail-trimmed M-estimation methods for heavy tailed Nonlinear AR-GARCH models. Tail-trimming allows both identification of the true parameter and asymptotic normality for nonlinear models with asymmetric errors. In heavy tailed cases the rate of convergence is infinitesimally close to the highest possible amongst M-estimators for a particular loss function, hence super- $n^{1/2}$-convergence can be achieved in nonlinear AR models with infinite variance errors, and arbitrarily near $n^{1/2}$-convergence for GARCH with errors that have an infinite fourth moment. We present a consistent estimator of the covariance matrix that permits classic inference without knowledge of the rate of convergence, and explore asymptotic covariance and bootstrap mean-squared-error methods for selecting trimming parameters. A simulation study shows the estimator trumps existing ones for AR and GARCH models based on sharpness, approximate normality, rate of convergence, and test accuracy. We then use the estimator to study asset returns data.

1. INTRODUCTION It is now widely accepted that log-returns of many macroeconomic and financial time series are heavy tailed, asymmetrically distributed, and exhibit clustering of large values. In broader contexts extremes are encountered in actuarial, meteorological, and telecommunication network data (e.g. Leadbetter et al 1983, Engle and Ng 1993, Glosten et al 1993, Embrechts et al 1997, Finkenstädt and Rootzén 2001). GARCH-type clustering alone implies higher moments do not exist due to Pareto-like distribution tails (e.g. Basrak et al 2002, Cline 2007).

We develop new methods of robust M-estimation for stationary nonlinear AR(k)-GARCH(1,1):

(1) $$y_t = f(x_t, \phi^0) + u_t, \text{ where } u_t = \sigma_t \epsilon_t$$

$$\sigma_t^2 = g\left(u_{t-1}, \sigma_{t-1}^2, \theta^0\right) \text{ where } \theta^0 = \left[\phi^{0\prime}, \beta^{0\prime}\right]'.$$

The response functions f and g are known, the regressors are $x_t = [y_{t-1}, ..., y_{t-k}]'$, and β^0 are parameters unique to the volatility process. We assume there exists a unique θ^0 such that $\{y_t, u_t, \sigma_t^2\}$

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

Key words and phrases: M-estimation; heavy tails; nonlinear AR-GARCH; tail trimming; robust inference.

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

is stationary and ergodic, and θ^0 minimizes our criterion asymptotically. See Section 2 for parameter space specifications and the identification condition. We are particularly interested in heavy tails in the AR error $E[u_t^2] = \infty$, and if we estimate the GARCH parameter β^0 then we allow the volatility error $E[\epsilon_t^4] = \infty$. Our limit theory does not exploit independence in ϵ_t, hence (1) includes weak, semi-strong and strong -GARCH.

Nonlinear GARCH(1,1) models have evolved in the literature as essential representations of the above stylized traits, while retaining parsimony (e.g. Nelson 1991, Engle and Ng 1993, Glosten et al 1993, Engle and Rangle 2007). Examples of models allowed here are Asymmetric-, β-, Quadratic-, Spline-, and Smooth Transition-GARCH. Allowing additional GARCH lags, or regressors other than lagged y_t, is only a matter of notation. Similarly, lags of u_t can be included in f for nonlinear ARMA-GARCH with only added notation (e.g. Lee 2007).

In this paper we tail-trim a smooth loss function to obtain an asymptotically normal Tail-Trimmed M-estimator for θ^0 [TTME]. We treat as special cases Quasi-Maximum Tail-Trimmed Likelihood [QMTTL] which reduces to Nonlinear Least Tail-Trimmed Squares [NLLTTS] and Least Tail-Trimmed Squares [LTTS] depending on the model. See Section 2 for TTME and Section 3 for QMTTL. There are far more cases of (1), trimming strategies and loss types than can be treated here, so we focus on NLLTTS for nonlinear AR, and QMTTL for nonlinear GARCH. The reader can use Sections 2 and 3 to support other loss, trimming approaches and extensions of (1), including non-Gaussian ML and Whittle estimation.

In many cases we can show how model parameters, tail thickness and trimming parameters impact efficiency, while negligibility implies trimming never effects the asymptotic covariance matrix in thin tail cases. Fixed quantile methods by construction influence efficiency irrespective of higher moments, and may cause bias due to asymmetry in the model (e.g. Ronchetti and Trojani 2001, Ling 2005, 2007, Čižek 2008, Boudt and Croux 2010). See Section 4 for comparisons.

In the case of nonlinear AR with an iid infinite variance error u_t we show super-$\sqrt{n}$-convergence can arise where n is the sample size, while the rate is always $o(n^{1/2})$ for pure strong-GARCH models with $E[\epsilon_t^4] = \infty$. In *all* cases, however, if the level or volatility error is iid and heavy tailed the rate is infinitesimally close to the highest possible rate by following simple trimming rules discussed in Section 4, hence QMTTL converges faster than QML (cf. Hall and Yao 2003).

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

Choosing the trimming portion in practice is not transparent, in particular because the literature almost exclusively focuses on fixed quantile methods for outlier robust estimation (see Ollila et al 2002 and Agulló et al 2008 for reviews). In Section 5 we therefore adapt asymptotic covariance and bootstrap mean-squared-error methods to tail-trimming.

A complete theory of QML for nonlinear AR with a semi-strong GARCH error is only recently available. Meitz and Saikkonen (2009), hereafter MS (2009), establish asymptotic normality provided $E|y_t|^{4+\iota}$ for some $\iota > 0$, and $E[\epsilon_t^8] < \infty$. See also Straumann and Mikosch (2006). QML theory for linear strong and semi-strong GARCH is developed in Lee and Hansen (1994), Lumsdaine (1996), Ling (2007), Escanciano (2009), and Linton et al (2010) amongst many others, while Jensen and Rahbek (2004) treat non-stationary GARCH. In all cases ϵ_t has a finite 4^{th} to 16^{th} moment to ensure asymptotic normality See especially Francq and Zakoïan (2004).

Sensitivity of minimum distance estimators to large values is now substantially documented. Consider Huber (1977), Rousseeuw (1984), Ronchetti and Trojani (2001), Rousseeuw et al (2004), Čižek (2008), and Boudt and Croux (2010). In AR(k) models $y_t = \sum_{i=1}^{k} \phi_i^0 y_{t-i} + u_t$ if u_t is iid and $E[u_t^2] = \infty$ then M-estimators are not asymptotically normal, although super-$\sqrt{n}$-convergence is achieved due to a leverage effect. Davis et al (1992) show smooth M-estimators and LAD are

$n^{1/\kappa}L(n)$-convergent for some slowly varying function[2] $L(n)$, e.g. $L(n) = 1$ when ϵ_t exhibits Paretian tail decay with index $\kappa \leq 2$. Conversely, although GARCH models have ARMA representations, error-regressor feedback diminishes QML convergence below $\sqrt{n}$ when $E[\epsilon_t^4] < \infty$ (Hall and Yao 2003, Linton et al 2010). A class of Log-LAD estimators for strong and semi-strong GARCH are $\sqrt{n}$-convergent if $E[\epsilon_t^2] = 1$ and $\ln(\epsilon_t^2)$ is symmetric (Peng and Yao 2003, Linton et al 2010), while Whittle estimation for GARCH requires $E[\epsilon_t^8] < \infty$ (Mikosch and Straumann 2002).

Outlier robust M-estimators like Least Trimmed Squares [LTS] and Maximum Trimmed Likelihood are based on fixed quantile trimming or weighting of criterion equations or the data itself. See Rousseeuw et al (2004) and Čižek (2008) for reviews. In the vast majority of cases trimming is based on residual magnitudes, hence time series data with infinite variance are typically not covered. Ronchetti and Trojani (2001) propose symmetric truncated GMM, but require the true distribution for a simulation step that controls for bias. Peng and Yao (2004) prove asymptotic normality for an M-estimator of a bounded nonparametric response with heavy tailed dependent errors, while boundedness rules out AR-GARCH.

Our approach blends Generalized M-Estimation (see Mallows 1973) with tail-trimming for location estimation (see Hahn et al 1991 and Hill 2011 for reference). In the former vein Boudt and Croux (2010) truncate a non-Gaussian QML criterion by a fixed quantile for multivariate GARCH models with elliptical errors. They only prove consistency and require a specification for the true density. Ling (2005) establishes asymptotic normality for the Least Weighted Absolute Deviations [LWAD] estimator $\arg\min_{\phi\in\Phi}\sum w_t|y_t - \phi' x_t|$ for linear AR(k) models where w_t belongs to a general weight class, although he only exemplifies the theory with a weight w_t based on a truncation function $y_{t-i}^*(c) := |y_{t-i}|$ if $|y_{t-i}| \leq c$ and otherwise $y_{t-i}^*(c) = c > 0$. Only $\sqrt{n}$-convergence can be achieved since the weights work like fixed quantile trimming indicators. Ling (2007) uses the same technique to estimate linear ARMA-GARCH models by Quasi-Maximum Weighted Likelihood [QMWL], but requires $E[\epsilon_t^4] < \infty$ for asymptotic normality since w_t only operates on lagged y_{t-i}. Ling's only detailed weight uses $y_{t-i}^*(c)$, thus only symmetric DGP's are covered[3], and while infinitely many $y_{t-1}^*(c), y_{t-2}^*(c), ...$ are required for GARCH models Ling does not explain how to compute w_t in practice. Further, QMWL is always less efficient than QML (Ling 2007: Section 4).

The QMTTL criterion for GARCH(1,1), by comparison, is based solely on ϵ_t if we know there are GARCH effects, and otherwise on only ϵ_t, y_{t-1} and y_{t-2}. We only require $E[\epsilon_t^2] < \infty$, and if $E[\epsilon_t^4] < \infty$ then QMTTL is asymptotically equivalent to QML. In a simulation study we find QMWL for GARCH is both non-normal when $E[\epsilon_t^4] = \infty$ (supported also by theory), while LWAD for AR and Log-LAD for GARCH are resilient to heavy tails in different degrees. Overall NLLTTS and QMTTL trump each in sharpness, approximate normality and therefore inference.

We use the following notation conventions. The indicator function $I(A) = 1$ if A is true, and otherwise $I(A) = 0$. w_t is an arbitrary random variable. $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote the minimum and maximum eigenvalues of A. The L_r-norm is $||x||_r = (\sum_{i,j} E|x_{i,j}|^r)^{1/r}$, l_1-matrix norm $|x| := \sum_{i,j}|x_{i,j}|$, spectral norm $||A|| = \lambda_{\max}(A'A)^{1/2}$. $(z)_+ := \max\{0, z\}$. K denotes a positive finite constant whose value may change from line to line; $\iota > 0$ is an arbitrarily tiny constant. $\stackrel{p}{\rightarrow}$ and $\stackrel{d}{\rightarrow}$ denote probability and distribution convergence. $x_n \sim a_n$ denotes $x_n/a_n \rightarrow 1$; $x_n = o(a_n)$ denotes $x_n/a_n \rightarrow 0$, and $x_n = o_p(a_n)$ means $x_n/a_n \stackrel{p}{\rightarrow} 0$. $L(x)$ is a slowly varying [s.v.] function that may change with the context. $\epsilon_t \stackrel{iid}{\sim} (0, 1)$ states ϵ_t is iid with zero mean and unit variance.

---

[2]Slowly varying $L(n)$ satisfies $L(\xi n)/L(n) \rightarrow 1$ as $n \rightarrow \infty$ for any $\xi > 0$ (Leadbetter et al 1983). Constants and power transforms of the natural log are classic examples (e.g. $a(\ln(n))^b$, $a > 0$, $b \geq 0$).

[3]Ling's (2005, 2007) general theory clearly extends to nonlinear models like (1). It is not clear, however, that his suggested symmetric weight applies in this case since it should render θ^0 non-identifiable. Tail-trimming, however, ensures identification asymptotically due to negligibility.

2. GENERALIZED TAIL-TRIMMED M-ESTIMATORS Define the total set of variables $z_t = [y_t, x_t']'$, and assume each $w_t \in \{y_t, u_t, \sigma_t^2\}$ is stationary, ergodic, and $E|w_t|^\iota < \infty$ for some tiny $\iota > 0$. The response functions are $f : \mathbb{R}^k \times \Phi \to \mathbb{R}$ and $g : \mathbb{R} \times \mathbb{R}_+ \times \Theta \to \mathbb{R}_+$, where $\Phi \subset \mathbb{R}^p$ and $\Theta \subset \mathbb{R}^q$ are compact parameter spaces, and all parameters are collected as $\theta = [\phi', \beta']' \in \Theta$. Implicitly $\beta \in \mathbb{R}^{q-p}$ are unique to the volatility process. There are $k \geq 0$ regressors $x_t = [y_{t-1}, ..., y_{t-k}]'$, and $q \geq p$ total parameters. The case $p = k = 0$ corresponds to a pure GARCH model $y_t = u_t = \sigma_t \epsilon_t$, hence $\theta = \beta$, while $k = 0$ and $p = 1$ imply $y_t = \phi^0 + \sigma_t \epsilon_t$. Similarly, $q = p$ corresponds to a pure AR model with $\theta = \phi$. We assume at least one parameter $q \geq 1$.

Define the level error

$$u_t(\phi) := y_t - f(x_t, \phi),$$

and an iterated volatility process

$$h_t(\theta) = \omega_0 > 0 \text{ for } t = 0, \text{ and } h_t(\theta) = g(u_{t-1}, h_{t-1}(\theta), \theta) \text{ for } t = 1, 2, ... \tag{2}$$

where ω_0 is not necessarily an element of θ^0. Assume $g(u, h, \theta)$ is twice differentiable in (u, h, θ), and define $h_t^\theta(\theta) := (\partial/\partial\theta) h_t(\theta)$ and $h_t^{\theta,\theta}(\theta) := (\partial/\partial\theta) h_t^\theta(\theta)$. If volatility parameters are estimated we require stationary and ergodic solutions $\{h_t^*(\theta), h_t^{*\theta}(\theta), h_t^{*\theta,\theta}(\theta)\}$ where $h_t^*(\theta)$ solves (2), and $\{h_t^{*\theta}(\theta), h_t^{*\theta,\theta}(\theta)\}$ solve related difference equations in MS (2009: eq.'s (9)-(10)). See Appendix A for sufficient conditions and a proof based on Lipschitz type bounds on f, g and their derivatives (cf. An and Huang 1996, Carrasco and Chen 2002, Francq and Zakoïan 2006, MS 2009).

Denote the loss function $l(z_t, \theta) \geq 0$, assumed measurable, strictly convex and twice continuously differentiable in θ. If volatility parameters β are estimated then $h_t(\theta)$ enters $l(z_t, \theta)$, we assume $l(z_t, \theta)$ can be written $l(z_t, \phi, h_t(\theta))$, and $l(z_t, \phi, h)$ is thrice differentiable in h. The third derivative is required to show asymptotics can be based on derivatives evaluated with $\{h_t^*(\theta), h_t^{*\theta}(\theta), h_t^{*\theta,\theta}(\theta)\}$. If loss is QML this reduces to assuming only twice differentiability of $f(., \phi)$ and $g(u, h, \theta)$ as in MS (2009). See Appendix A.

2.1 TAIL-TRIMMED LOSS AND CRITERION

Asymptotics for smooth M-estimators are based on the first and second derivatives (e.g. Amemiya 1985) which we call *estimating* and *Jacobian* equations:

$$m_t(\theta) := \frac{\partial}{\partial\theta} l(z_t, \theta) \in \mathbb{R}^q \quad \text{and} \quad G_t(\theta) := \frac{\partial}{\partial\theta} m_t(\theta) \in \mathbb{R}^{q \times q}.$$

Hereafter we drop θ^0: $m_t = m_t(\theta^0)$ and so on.

If we negligibly trim $l(z_t, \theta)$ when a large $m_{i,t}(\theta)$ or $G_{i,j,t}(\theta)$ occurs, then a consistent and asymptotically normal estimator of θ^0 is achievable. Since $m_t(\theta)$ and $G_t(\theta)$ can be computed by numerical approximation if an analytic expression does not exist, we assume the analyst has $m_t(\theta)$ and $G_t(\theta)$ in hand. If $m_t(\theta)$ and $G_t(\theta)$ reduce to simple functions of random variables we may focus trimming there, which is the topic of Section 3.

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

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

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

The determination of large $m_{i,t}(\theta)$ and $G_{i,j,t}(\theta)$ is made by intermediate order sequences $\{k_{1,i,n}^{(m)}, k_{2,i,n}^{(m)}\}$ and $\{k_{1,i,j,n}^{(G)}, k_{2,i,j,n}^{(G)}\}$: if $\{k_{1,n}, k_{2,n}\}$ denotes either then $1 \leq k_{1,n} + k_{2,n} < n$, $k_{r,n} \to$

∞ and $k_{r,n} = o(n)$. See Leadbetter et al (1983) and Hahn et al (1991). Define indicator selection functions assuming asymmetry

$$\hat{I}^{(m)}_{i,n,t}(\theta) := I\left(m^{(-)}_{i,(k^{(m)}_{1,i,n})}(\theta) \leq m_{i,t}(\theta) \leq m^{(+)}_{i,(k^{(m)}_{2,i,n})}(\theta)\right)$$

$$\hat{I}^{(G)}_{i,j,n,t}(\theta) := I\left(G^{(-)}_{i,j,(k^{(G)}_{1,i,j,n})}(\theta) \leq G_{i,j,t}(\theta) \leq G^{(+)}_{i,j,(k^{(G)}_{2,i,j,n})}(\theta)\right),$$

and a composite trimming indicator

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

where $\prod_{1\leq i\leq j\leq q} \hat{I}^{(G)}_{i,j,n,t}(\theta)$ removes redundant indicators due to symmetry of G_t.

The sample is $\{y_t\}^n_{t=-k}$ to simplify notation since we condition on the first $k+1$ observations, and the initial value $h_0(\theta) = \omega_0$. The Tail-Trimmed M-estimator [TTME] therefore solves

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

If either $m_{i,t}$ or $G_{i,j,t}$ is known to be symmetric at zero then use, e.g.,

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

If $E[m^2_{i,t}]$ or $E|G_{i,j,t}|$ is finite then it can be dropped from $\hat{I}_{n,t}(\theta)$, including constant $G_{i,j,t}$ for an intercept in (1). Such cases discussed in Section 3 can lead to substantial simplifications for a particular loss or model.

Each $k_{r,n}$ represents the number of trimmed $l(z_t,\theta)$ due to large negative or positive $m_{i,t}(\theta)$ or $G_{i,j,t}(\theta)$. We require $k_{r,n} \to \infty$ for asymptotic normality, while negligibility $k_{r,n}/n \to 0$ ensures identification, and allows super-$\sqrt{n}$-convergence for AR estimation. Yet choosing a complete policy $\{k^{(m)}_{1,i,n}, k^{(m)}_{2,i,n}, k^{(G)}_{1,i,j,n}, k^{(G)}_{2,i,j,n}\}$ in practice is an outstanding challenge with few avenues provided by the literature, although in many cases the problem reduces to choosing one, or very few, sequences $\{k_{r,n}\}$, depending on the model and symmetry (see Sections 3 and 4). We adopt existing covariance and bootstrap methods for fixed quantile problems to our needs in Section 5.

In order to identify θ^0 and characterize the limit distribution of $\hat{\theta}_n$, we need trimming indicators with non-random thresholds. If $h_t(\theta)$ enters into $l(z_t,\theta)$ then let $l^*(z_t,\theta)$ denote the loss evaluated with the stationary solution $h^*_t(\theta)$, and similarly define functions of $\{h^*_t(\theta), h^{*\theta}_t(\theta), h^{*\theta,\theta}_t(\theta)\}$:

$$m^*_t(\theta) := \frac{\partial}{\partial\theta} l^*(z_t,\theta) \in \mathbb{R}^q \;\; \text{and} \;\; G^*_t(\theta) := \frac{\partial}{\partial\theta} m^*_t(\theta) \in \mathbb{R}^{q\times q}.$$

Although we do not express it, technically these derivatives exist $a.s.$ at any point $\{\theta, z_t\}$.

The selection indicators and trimmed equations here are based on positive non-random threshold sequences $\{\mathcal{L}^{(m)}_{i,n}(\theta), \mathcal{U}^{(m)}_{i,n}(\theta)\}$ and $\{\mathcal{L}^{(G)}_{i,j,n}(\theta), \mathcal{U}^{(G)}_{i,j,n}(\theta)\}$:

$$I^{(m^*)}_{i,n,t}(\theta) := I\left(-\mathcal{L}^{(m)}_{i,n}(\theta) \leq m^*_{i,t}(\theta) \leq \mathcal{U}^{(m)}_{i,n}(\theta)\right)$$

$$I^{(G^*)}_{i,j,n,t}(\theta) := I\left(-\mathcal{L}^{(G)}_{i,j,n}(\theta) \leq G^*_{i,j,t}(\theta) \leq \mathcal{U}^{(G)}_{i,j,n}(\theta)\right)$$

and a composite indicator

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

where, e.g., $\mathcal{L}_{i,n}^{(m)}(\theta)$ is the lower $k_{1,i,n}^{(m)}/n$ and $\mathcal{U}_{i,n}^{(m)}(\theta)$ the upper $k_{2,i,n}^{(m)}/n$ quantiles:

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

Under symmetric trimming for $m_{i,t}^{*}(\theta)$, e.g., we use $I_{i,n,t}^{(m^{*})}(\theta):=I(|m_{i,t}^{*}(\theta)|\leq\mathcal{C}_{i,n}^{(m)}(\theta))$ where the positive sequence $\{\mathcal{C}_{i,n}^{(m)}(\theta)\}$ satisfies $P(|m_{i,t}^{*}(\theta)|\geq\mathcal{C}_{i,n}^{(m)}(\theta))=k_{i,n}^{(m)}/n$.

Thresholds $\{\mathcal{L}_{i,n}^{(m)}(\theta),\mathcal{U}_{i,n}^{(m)}(\theta)\}$ and $\{\mathcal{L}_{i,j,n}^{(G)}(\theta),\mathcal{U}_{i,j,n}^{(G)}(\theta)\}$ exist for any choice of fractiles $k_{r,n}$ since we assume $m_{i,t}^{*}(\theta)$ and $G_{i,j,t}^{*}(\theta)$ have smooth distributions, and under a mixing condition the order statistics are consistent for the above thresholds, e.g. $m_{i,(k_{2,i,n}^{(m)})}^{(+)}(\theta)/\mathcal{U}_{i,n}^{(m)}(\theta)\overset{p}{\rightarrow}1$ uniformly on Θ. See Lemma C.2 in the supplementary Appendix C following the main paper below.

Now define deterministically tail-trimmed equations

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

and long run covariance and Jacobian matrices:

$$\mathcal{S}_{n}(\theta):=\frac{1}{n}\sum_{s,t=1}^{n}E\left[m_{n,s}^{*}(\theta)m_{n,t}^{*}(\theta)'\right]\;\text{ and }\;\mathcal{G}_{n}(\theta):=E\left[G_{n,t}^{*}(\theta)\right]. \tag{3}$$

We scale $\mathcal{S}_{n}(\theta)$ by $1/n$ to make clear in Section 4 how the rate of convergence compares to $n^{1/2}$.

The identification condition is

$$E\left[m_{n,t}^{*}(\theta)\right]\rightarrow0\;\textit{ if and only if }\;\theta=\theta^{0}\text{ a unique interior point of }\Theta.$$

Notice we do not explicitly require $E[m_{t}^{*}]=0$, allowing for very heavy tails in some cases. A simple example is a stationary AR(1) $y_{t}=\phi y_{t-1}+u_{t}$ with an iid symmetric u_{t}, and least squares $m_{t}=u_{t}y_{t-1}$: $E[m_{n,t}^{*}]=0$ even if $E|u_{t}|=\infty$. Nevertheless, for a pure GARCH model and QML loss $m_{t}^{*}=(\epsilon_{t}^{2}-1)h_{t}^{\theta*}/h_{t}^{*}$ is in general asymmetric, so we can only say $E[m_{n,t}^{*}]\rightarrow0$. Since by dominated convergence $E[m_{n,t}^{*}]\rightarrow E[m_{t}^{*}]$ at least $E[\epsilon_{t}^{2}]<\infty$ must hold for $E[m_{n,t}^{*}]\rightarrow0$. See Section 3.

2.2 MAIN RESULTS

All assumptions are detailed in Appendix A, covering distribution smoothness and tails (D), bounds for fractiles $k_{j,i,n}^{(m)}$ (F), identification (I), loss smoothness (L), mixing (MX), smoothness of $E[m_{n,t}^{*}(\theta)]$ (MS), non-degeneracy of tail trimmed matrices (N), response bounds and smoothness (RB) and (RS) required for a stationary solution, and stationarity and moments of $\{y_{t},u_{t},\sigma_{t}^{2}\}$ (STM). Proofs of the main results are presented in Appendix B.

In order to show $\hat{\theta}_{n}-\theta^{0}$ is approximated by $n^{-1/2}\mathcal{S}_{n}^{-1/2}\sum_{t=1}^{n}\{m_{n,t}^{*}-E[m_{n,t}^{*}]\}$ for asymptotic normality, we must have Jacobian consistency $1/n\sum_{t=1}^{n}G_{t}(\hat{\theta}_{n})\hat{I}_{n,t}(\hat{\theta}_{n})=\mathcal{G}_{n}\times(1+o_{p}(1))$ which requires $\hat{\theta}_{n}\overset{p}{\rightarrow}\theta^{0}$ from first principles.

THEOREM 2.1 (*consistency*). *Under Assumptions D, F, I, L, MS, MX, N, RB, RS and STM* $\hat{\theta}_{n}\overset{p}{\rightarrow}\theta^{0}$.

Remark 1: Response bounds RB and smoothness under RB are imposed to ensure stationary solutions $\{h_t^*(\theta), h_t^{*\theta}(\theta), h_t^{*\theta,\theta}(\theta)\}$ exist, which we require only if volatility parameters β are estimated.

Remark 2: Consistency requires all assumptions, and in general more assumptions than imposed in Straumann and Mikosch (2006) and MS (2009). This is due to nonlinearities induced by trimming: we require a first order condition under tail-trimming and loss smoothness allows for almost sure criterion differentiability at $\hat{\theta}_n$; we require a UCLT to relate $\hat{I}_{n,t}(\theta)$ to $I_{n,t}(\theta)$ and a ULLN on $m_{n,t}^*(\theta)$ to prove consistency, and both rely on mixing and smoothness properties. It does not appear to be possible to prove consistency under weaker conditions.

THEOREM 2.2 (*normality*). *Under Assumptions D, F, I, L, MS, MX, N, RB, RS and STM* $\mathcal{V}_n^{1/2}(\hat{\theta}_n - \theta^0) \xrightarrow{d} N(0, I_q)$ *where the scale is* $\mathcal{V}_n(\theta) := n\mathcal{G}_n(\theta)'\mathcal{S}_n^{-1}(\theta)\mathcal{G}_n(\theta)$.

Remark 1: Although $\mathcal{V}_n = n\mathcal{G}_n'\mathcal{S}_n^{-1}\mathcal{G}_n$ we cannot conclude the rate is $n^{1/2}$ since in the presence of heavy tails $||\mathcal{S}_n|| \to \infty$ and/or $||\mathcal{G}_n|| \to \infty$ are possible. This implies heterogeneous rates $\mathcal{V}_{i,i,n}^{1/2} \to \infty$ below, at, or above $n^{1/2}$ are possible depending on error-regressor feedback. See Section 4.

Remark 2: The covariance $\mathcal{S}_n$ captures the pure damage to estimation accuracy due to heavy tailed loss directional changes $m_{i,t}$. Under heavy tails $m_t(\theta) := (\partial/\partial\theta)l(z_t, \theta)$ exhibits large fluctuations, making a sharp estimate of θ^0 more difficult. Since m_t is a function of u_t in general, we call $\mathcal{S}_n$ the "*extreme error*" effect.

Remark 3: Although large estimating equations m_t hamper sharp estimation of θ^0, large Jacobian equations $G_t := (\partial/\partial\theta)m_t(\theta)|_{\theta^0}$ may reveal leverage points. Consider least squares [LS] loss for an AR(1) $y_t = \theta^0 y_{t-1} + u_t$ with iid u_t. Then $m_t = u_t y_{t-1}$ and $G_t = -y_{t-1}^2$. Extreme errors u_t and regressors y_{t-1} both show up as extreme m_t. Since a large m_t does not reveal the source of the extreme we have an extreme error effect. The Jacobian equation G_t, however, captures leverage points in the design $[y_1, ..., y_n]'$. This is not always the case: a pure GARCH model estimated with QML only has an error effect due to scaling. Nevertheless, we call $\mathcal{G}_n$ the "*leverage effect*".

Remark 4: If each $E|m_{i,t}|^{2+\iota} < \infty$ and $E|G_{i,j,t}| < \infty$ then by dominated convergence[4]

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

The inverse $\mathcal{V}^{-1}$ is the classic asymptotic covariance matrix, hence trimming has no impact asymptotically on efficiency under thin tails: $n^{1/2}(\hat{\theta}_n - \theta^0) \xrightarrow{d} N(0, \mathcal{V}^{-1})$.

In the case of heavy tails it is impossible to identify *unique* "rates of convergence" and an "asymptotic covariance" from $\mathcal{V}_n$ without first specifying a loss function, error dependence and trimming fractiles. We fully treat this topic in Section 4.

3. QUASI-MAXIMUM TAIL-TRIMMED LIKELIHOOD If we entertain a particular loss $l(z_t, \theta)$ and version of (1) then we can focus trimming on the source of extremes. Indeed, under QML loss trimming can be simplified, and if an error is iid then we can decompose the scale $\mathcal{V}_n$ into components representing an asymptotic covariance matrix and rates of convergence. We therefore restrict attention to $l(z_t, \theta) = \ln h_t(\theta) + (y_t - f(x_t, \phi))^2/h_t(\theta)$, hence Assumption L holds.

[4] Geometric β-mixing and $E|m_{i,t}|^{2+\iota} < \infty$ ensure $||\mathcal{S}|| < \infty$ (cf. Ibragimov 1962). Under independence we only need $E[m_{i,t}^2] < \infty$, in which case $\mathcal{S} = E[m_t m_t']$. See Appendix A for mixing details under Assumption MX.

Throughout, if $w_t(\theta)$ is a symmetric matrix on $\mathbb{R}^{q\times q}$ then $\hat{I}^{(w)}_{n,t}(\theta) = \prod_{1\leq i\leq j\leq q}\hat{I}^{(w)}_{i,j,n,t}(\theta)$ where

$$\hat{I}^{(w)}_{i,j,n,t}(\theta) = I\left(w^{(-)}_{i,j,(k^{(w)}_{1,i,j,n})}(\theta) \leq w_{i,j,t}(\theta) \leq w^{(+)}_{i,j,(k^{(w)}_{2,i,j,n})}(\theta)\right),$$

and if $w_{i,j,t}(\theta)$ is symmetric then $\hat{I}^{(w)}_{i,j,n,t}(\theta) = I(|w_{i,j,t}(\theta)| \leq w^{(a)}_{i,j,(k^{(w)}_{i,j,n})}(\theta))$. If $w^*_t(\theta)$ is the stationary version we write

$$w^*_{n,t}(\theta) := w^*_t(\theta)I^{(w^*)}_{n,t}(\theta) = w^*_t(\theta)\prod_{1\leq i\leq j\leq q} I\left(-\mathcal{L}^{(w)}_{i,j,n}(\theta) \leq w^*_{i,j,t}(\theta) \leq \mathcal{U}^{(w)}_{i,j,n}(\theta)\right).$$

3.1 NONLINEAR AUTOREGRESSIONS

We first treat a nonlinear AR(k) model

$$y_t = f(x_t, \phi^0) + u_t, \text{ where } x_t \in \mathbb{R}^k,\ \phi \in \mathbb{R}^p,\ E\,|u_t| < \infty,\ E|f^{(\phi)}_{i,t}| < \infty \text{ and } E\,[u_t|x_t] = 0.$$

Define $u_t(\phi) = y_t - f_t(\phi)$, $f^{\phi}_t(\phi) := (\partial/\partial\phi)f_t(\phi)$ and $f^{\phi,\phi}_t(\phi) := (\partial^2/\partial\phi\partial\phi')f_t(\phi)$. The TTME minimizes $1/n\sum_{t=1}^n u^2_t(\phi)\hat{I}_{n,t}(\phi)$, but we know extremes in $m_t(\phi) = u_t(\phi)f^{\phi}_t(\phi)$ and $G_t(\phi) = (\partial/\partial\phi)\{u_t(\phi)f^{\phi}_t(\phi)\}$ arise from $u_t(\phi)$, $f^{\phi}_t(\phi)$ and $f^{\phi,\phi}_t(\phi)$. We therefore define the Nonlinear Least Tail-Trimmed Squares [NLLTTS] estimator

$$\hat{\phi}_n = \underset{\phi\in\Phi}{\text{argmin}}\left\{\frac{1}{n}\sum_{t=1}^n u^2_t(\phi)\hat{I}^{(u)}_{n,t}(\phi)\,\hat{I}^{(f^{\phi})}_{n,t}(\phi)\,\hat{I}^{(f^{\phi,\phi})}_{n,t}(\phi)\right\}.$$

In the same sense that TTME asymptotics are based on $m^*_{n,t} = m_t I^*_{n,t}$ and $G^*_{n,t} = G_t I^*_{n,t}$, for NLLTTS we work with

$$\mathring{m}^*_{i,n,t} = m_t I^*_{n,t} = u_t I^{(u)}_{n,t} \times f^{\phi}_t I^{(f^{\phi})}_{n,t} \times I^{(f^{\phi,\phi})}_{n,t} = u^*_{n,t}f^{\phi*}_{i,n,t} \times I^{(f^{\phi,\phi})}_{n,t} \text{ and } \mathring{\mathcal{S}}_n := \frac{1}{n}\sum_{s,t=1}^n E\left[\mathring{m}^*_{n,s}\mathring{m}^{*\prime}_{n,t}\right]$$

$$\mathring{G}^*_{i,j,n,t} := -f^{\phi*}_{i,n,t}f^{\phi*}_{j,n,t} \times I^{(u)}_{n,t}I^{(f^{\phi,\phi})}_{n,t} + u^*_{n,t}f^{\phi,\phi*}_{i,j,n,t} \times I^{(f^{\phi})}_{n,t} \text{ and } \mathring{\mathcal{G}}_n := E\left[\mathring{G}^*_{n,t}\right].$$

Notice $E[\mathring{m}^*_{i,n,t}] \to 0$ by dominated convergence given $E\,[u_t|x_t] = 0$ and $E|f^{(\phi)}_{i,t}| < \infty$. Yet in general $u^*_{n,t}$ is not a martingale difference due to trimming, and $f^{\phi,\phi*}_{i,j,n,t}$ may not be integrable, so $\mathring{\mathcal{G}}_n \sim -E[f^{\phi*}_{n,t}f^{\phi*\prime}_{n,t}]$ need not hold. If u_t is independent of x_t then $E|u_t| < \infty$ and $E|f^{(\phi)}_{i,t}| < \infty$ can be relaxed and both $\mathring{\mathcal{S}}_n$ and $\mathring{\mathcal{G}}_n$ greatly simplify. See Example 1, below.

THEOREM 3.1 (*NLLTTS*). *Under D, F, I, MS, MX, N, RS.i and STM* $\mathring{\mathcal{V}}^{1/2}_n(\hat{\phi}_n - \phi^0) \xrightarrow{d} N(0, I_p)$ *where* $\mathring{\mathcal{V}}_n := n\mathring{\mathcal{G}}'_n\mathring{\mathcal{S}}^{-1}_n\mathring{\mathcal{G}}_n$.

Remark: Relating the TTME scale $\mathcal{V}_n = n\mathcal{G}_n{}'\mathcal{S}^{-1}_n\mathcal{G}_n$ to the NLLTTS scale $\mathring{\mathcal{V}}_n$ is intractable for finite n due to the different trimming mechanisms, and rather technical if we take $n \to \infty$ in the case of heavy tails. We therefore omit any such comparison here.

EXAMPLE 1 (NLLTTS, iid symmetric u_t): If u_t is iid and symmetric at zero then symmetric trimming by u_t is appropriate. Write $u^{(a)}_t(\phi) := |u_t(\phi)|$ and define $\hat{I}^{(u)}_{n,t}(\phi) := I(|u_t(\phi)| \leq u^{(a)}_{(k^{(u)}_n)}(\phi))$ and $I^{(u)}_{n,t}(\phi) := I(|u_t(\phi)| \leq \mathcal{C}^{(u)}_n(\phi))$, where $P(|u_t(\phi)| \geq \mathcal{C}^{(u)}_n(\phi)) = k^{(u)}_n/n$.

Independence implies $E[u^*_{n,t}|\Im_{t-1}] = E[u^*_{n,t}] = 0$ for any $k_n^{(u)}$, hence $E|u_t| = \infty$ and $E|f_{i,t}^{(\phi)}| = \infty$ are allowed. Assumption I therefore holds, and $\mathring{\mathcal{S}}_n = E[u^{*2}_{n,t}] \times E[f^{\phi *}_{n,t} f^{\phi *\prime}_{n,t} I^{(f^{\phi,\phi})}_{n,t}]$ and $\mathring{\mathcal{G}}_n := -E[f^{\phi *}_{n,t} f^{\phi *\prime}_{n,t} I^{(f^{\phi,\phi})}_{n,t}]$. Notice whether $\{\hat{I}^{(f^{\phi})}_{n,t}, \hat{I}^{(f^{\phi,\phi})}_{n,t}\}$ and $\{I^{(f^{\phi})}_{n,t}, I^{(f^{\phi,\phi})}_{n,t}\}$ are based on symmetric or asymmetric trimming is immaterial since $E[\mathring{m}^*_{n,t}] = 0$ due to $E[u^*_{n,t}|\Im_{t-1}] = 0$.

Asymptotics mirror the classic NLLS result (e.g. Amemiya 1985). Under Assumptions D, F, MS, MX, N, and STM

$$n^{1/2}\left(E[u^{*2}_{n,t}]\right)^{-1/2} \times \left(E\left[f^{\phi *}_{n,t} f^{\phi *\prime}_{n,t} I^{(f^{\phi,\phi})}_{n,t}\right]\right)^{1/2} \times \left(\hat{\phi}_n - \phi^0\right) \xrightarrow{d} N(0, I_p).$$

Notice each $f^{\phi}_{i,t} f^{\phi}_{j,t} I^{(f^{\phi})}_{n,t} I^{(f^{\phi,\phi})}_{n,t}$ is trimmed by the same compound $I^{(f^{\phi})}_{n,t} I^{(f^{\phi,\phi})}_{n,t}$, but

$$E\left[f^{\phi 2}_{i,t} I^{(f^{\phi})}_{n,t} I^{(f^{\phi,\phi})}_{n,t}\right] = E\left[f^{\phi 2}_{i,t} I^{(f^{\phi})}_{i,n,t}\right] \times \left(1 - \frac{E\left[f^{\phi 2}_{i,t} I^{(f^{\phi})}_{i,n,t}\left(1 - \prod_{j \neq i} I^{(f^{\phi})}_{j,n,t} I^{(f^{\phi,\phi})}_{n,t}\right)\right]}{E\left[f^{\phi 2}_{i,t} I^{(f^{\phi})}_{i,n,t}\right]}\right) \tag{4}$$

is proportional to $E[f^{\phi 2}_{i,t} I^{(f^{\phi})}_{i,n,t}]$ due to negligibility and dominated convergence. Thus,

$$n^{1/2}\left(E[u^{*2}_{n,t}]\right)^{-1/2} \times \left(\left[E\left(f^{\phi}_{i,t} I^{(f^{\phi})}_{i,n,t} \times f^{\phi}_{j,t} I^{(f^{\phi})}_{j,n,t}\right)\right]_{i,j=1}^{p}\right)^{1/2} \times \left(\hat{\phi}_n - \phi^0\right) \xrightarrow{d} N(0, I_p).$$

Geometric ergodicity and therefore Assumption MX hold under bounds like RB.i if the error has a smooth distribution, covering Logistic AR, Semi-Parametric AR, Threshold AR with a known threshold, and Smooth Transition AR. See An and Huang (1996), Meitz and Saikkonen (2008) and their references.

EXAMPLE 2 (AR, iid symmetric u_t): The AR model is

$$y_t = c^0 + \sum_{i=1}^{p} \zeta^0_i y_{t-i} + u_t = \phi^{0\prime} x_t + u_t \tag{5}$$

with roots outside the unit circle. Let u_t be iid, zero symmetric, with an absolutely continuous distribution, and power-law tail

$$P\left(|u_t| > u\right) = du^{-\kappa}\left(1 + o(1)\right), \; d > 0, \; \kappa \in (0, 2], \tag{6}$$

hence variance is infinite.

Since $f^{\phi}_{1,t}(\phi) = 1$, $f^{\phi}_{1+i,t}(\phi) = y_{t-i}$ is symmetric, and $f^{\phi,\phi}_{i,j,t}(\phi) = 0$, symmetric trimming is based entirely on $u_t(\phi)$ and y_{t-i} with indicators $\hat{I}^{(u)}_{n,t}(\phi) = I(|u_t(\phi)| \leq u^{(a)}_{(k_n^{(u)})}(\phi))$ and $\hat{I}^{(y)}_{i,n,t} := I(|y_{t-i}| \leq y^{(a)}_{(k_n^{(y)})})$. Similarly $I^{(u)}_{n,t}(\phi) := I(|u_t(\phi) \leq \mathcal{C}^{(u)}_n(\phi))$ and $I^{(y)}_{i,n,t} := I(|y_{t-i}| \leq \mathcal{C}^{(y)}_n)$. The Least Tail-Trimmed Squares [LTTS] estimator therefore solves

$$\hat{\phi}_n = \underset{\phi \in \Phi}{\text{argmin}} \left\{\frac{1}{n}\sum_{t=1}^{n}(y_t - \phi' x_t)^2 \hat{I}^{(u)}_{n,t}(\phi)\, \hat{I}^{(y)}_{n,t}\right\}.$$

The conditions in Theorem 3.1 hold, although we must impose fractile bound F.

COROLLARY 3.2 (*LTTS*). *Let (5), (6) and Assumption F hold. Then* $n^{1/2}(E[u^{*2}_{n,t}])^{-1/2}$

$\times\ (E[x^{*}_{n,t}x^{*\prime}_{n,t}])^{1/2} \times (\hat{\phi}_n - \phi^0) \overset{d}{\to} N(0, I_p)$.

3.2 NONLINEAR SEMI-STRONG GARCH

Let $\Im_t := \sigma(y_\tau : \tau \leq t)$. The semi-strong nonlinear GARCH(1,1) model is

(7) $\quad y_t = \sigma_t \epsilon_t, \ \ \sigma_t^2 = g(y_{t-1}, \sigma_{t-1}^2, \beta^0)$, where $\beta \in \mathbb{R}^q$, and $\{\epsilon_t, \Im_t\}$ and $\{\epsilon_t^2 - 1, \Im_t\}$ are *mds*'s.

Write the error $\epsilon_t^2(\beta) := y_t^2/h_t(\beta)$, score $\mathfrak{s}_t(\beta) := h_t^{-1}(\beta) \times h_t^{\beta}(\beta)$, and $\mathfrak{s}_t^{\beta}(\beta) := (\partial/\partial\beta)\mathfrak{s}_t(\beta)$. The equations are

$$l(\beta, z_t) = \ln h_t(\beta) + \epsilon_t^2(\beta)$$

$$m_t(\beta) = \left(\epsilon_t^2(\beta) - 1\right)\mathfrak{s}_t(\beta) \quad \text{and} \quad G_t(\beta) = \left(\epsilon_t^2(\beta) - 1\right)\mathfrak{s}_t^{\beta}(\beta) - \epsilon_t^2(\beta)\,\mathfrak{s}_t(\beta)\,\mathfrak{s}_t(\beta)'.$$

Unless it is known that ϵ_t is symmetric, or $\mathfrak{s}_{i,t}(\beta) \geq 0$ or $\mathfrak{s}_{i,j,t}^{\beta}(\beta) \geq 0$ *a.s.*, then asymmetric trimming is required. The QMTTL estimator [QMTTLE] solves

$$\hat{\beta}_n = \underset{\beta \in \mathcal{B}}{\operatorname{argmin}} \left\{ \sum_{t=1}^{n} \left\{ \ln h_t(\beta) + \epsilon_t^2(\beta) \right\} \hat{I}_{n,t}^{(\epsilon)}(\beta) \times \hat{I}_{n,t}^{(\mathfrak{s})}(\beta) \times \hat{I}_{n,t}^{(\mathfrak{s}^{\beta})}(\beta) \right\},$$

where $\mathcal{B}$ is a compact subset of $\mathbb{R}^q$. Write $u_t(\beta) := \epsilon_t^2(\beta) - 1$, and define the Jacobian and long-run covariance:

$$(8) \qquad \mathring{m}_{n,t}^{*}(\beta) = u_t(\beta) I_{n,t}^{(\epsilon)}(\beta) \times \mathfrak{s}_{n,t}^{*}(\beta) \times I_{n,t}^{(\mathfrak{s}^{\beta})}(\beta), \ \ \mathring{\mathcal{S}}_n(\beta) := \frac{1}{n} \sum_{s,t=1}^{n} E\left[\mathring{m}_{n,s}^{*}(\beta)\, \mathring{m}_{n,t}^{*}(\beta)'\right]$$

$$\mathring{\mathcal{G}}_n(\beta) := E\left[u_t(\beta) I_{n,t}^{(\epsilon)}(\beta) \times \mathfrak{s}_{n,t}^{*\beta}(\beta) \times I_{n,t}^{(\mathfrak{s}^{\beta})}(\beta)\right] - E\left[\epsilon_{n,t}^{*2}(\beta)\, \mathfrak{s}_{n,t}^{*}(\beta)\, \mathfrak{s}_{n,t}^{*\prime}(\beta) \times I_{n,t}^{(\mathfrak{s}^{\beta})}(\beta)\right]'.$$

Similar to Theorem 3.1, unless ϵ_t is iid we cannot in general deduce $\mathring{\mathcal{G}}_n \sim -E[\mathfrak{s}_{n,t}^{*}\mathfrak{s}_{n,t}^{*\prime}]$.

THEOREM 3.3 (*QMTTL*). *Let Assumptions D, F, I, MS, MX, N, RB.ii-iv, RS.ii and STM hold for (7). Then* $\mathring{\mathcal{V}}_n^{1/2}(\hat{\beta}_n - \beta^0) \overset{d}{\to} N(0, I_q)$, *where* $\mathring{\mathcal{V}}_n = n\mathring{\mathcal{G}}_n' \mathring{\mathcal{S}}_n^{-1} \mathring{\mathcal{G}}_n$.

If $\epsilon_t \overset{iid}{\sim} (0,1)$ then we can always assume all fractiles $k_{r,n}$ satisfy $E[u_t I_{n,t}^{(\epsilon)}] = o((\max\{1, ||E[\mathfrak{s}_{n,t}^{*\beta} I_{n,t}^{(\mathfrak{s}^{\beta})}]||\})^{-1})$ and $E[u_t I_{n,t}^{(\epsilon)}] = o(n^{-1/2}||E[\mathfrak{s}_{n,t}^{*}\mathfrak{s}_{n,t}^{*\prime} I_{n,t}^{(\mathfrak{s}^{\beta})}]||^{-1/2})$, similar to Assumption I. Further, trimming negligibility implies $E[u_t^2 I_{n,t}^{(\epsilon)}] \sim E[(\epsilon_{n,t}^{*2} - 1)^2]$, while $E[(\epsilon_{n,t}^{*2} - 1)^2] = E[\epsilon_{n,t}^{*4}] - 2E[\epsilon_{n,t}^{*2}] + 1 \sim E[\epsilon_{n,t}^{*4}] - 1$ since $E[\epsilon_t^2] = 1$. Finally, the logic of (4) applies here: $E[\mathfrak{s}_{i,t}^{*2} I_{n,t}^{(\mathfrak{s})} I_{n,t}^{(\mathfrak{s}^{\beta})}] \sim E[\epsilon_{n,t}^{*2}\mathfrak{s}_{i,t}^{*2} I_{i,n,t}^{(\mathfrak{s})}]$. The QMTTL scale therefore satisfies

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

EXAMPLE 3 (Linear Strong-GARCH): The model is $\sigma_t^2 = \beta_1^0 + \beta_2^0 y_{t-1}^2 + \beta_3^0 \sigma_{t-1}^2$, where $\beta_1^0 > 0$, $\beta_2^0, \beta_3^0 \geq 0$, and $E[\ln(\beta_2^0 + \beta_3^0 \epsilon_t^2)] < 0$ to ensure stationarity (Nelson 1990). Assume $\epsilon_t \overset{iid}{\sim} (0,1)$ is symmetric with an absolutely continuous distribution, and if $E[\epsilon_t^4] = \infty$ then ϵ_t has tail (6) with index $\kappa_\epsilon \in (2,4]$ as in Hall and Yao (2003). In general $E[\mathfrak{s}_{i,t}^{2+\iota}] < \infty$ and $E[(\mathfrak{s}_{i,j,t}^{\beta})^{2+\iota}] < \infty$ for tiny $\iota > 0$ if there are GARCH effects $\beta_2^0 + \beta_3^0 > 0$ (e.g. Francq and Zakoian 2004), in which case we do not need to trim by these elements: $\hat{I}_{n,t}(\beta) = \hat{I}_{n,t}^{(\epsilon)}(\beta)$ and the QMTTLE solves

$$\hat{\beta}_n = \underset{\beta \in \mathcal{B}}{\mathrm{argmin}} \left\{ \sum_{t=1}^{n} \left(\ln h_t(\beta) + \epsilon_t^2(\beta) \right) \hat{I}_{n,t}^{(\epsilon)}(\beta) \right\}.$$

The non-GARCH case $\beta_2^0 + \beta_3^0 = 0$ implies $\mathfrak{s}_{i,t}$ and $\mathfrak{s}_{i,j,t}^{\beta}$ are constants or proportional to powers of ϵ_{t-1} and ϵ_{t-2} (e.g. $\mathfrak{s}_t = (\beta_1^0)^{-1}[1, \epsilon_{t-1}^2, \beta_1^0]$), so we need to trim by ϵ_t, y_{t-1} and y_{t-2}. In this case $\hat{I}_{n,t}(\beta) = \hat{I}_{n,t}^{(\epsilon)}(\beta)\hat{I}_{n,t}^{(y)}$ where $\hat{I}_{n,t}^{(y)} = \hat{I}_{1,n,t}^{(y)}\hat{I}_{2,n,t}^{(y)}$ and $\hat{I}_{i,n,t}^{(y)} := I(|y_{t-i}| \leq y_{(k_n^{(y)})}^{(a)})$. If we trim with $\hat{I}_{n,t}(\beta) = \hat{I}_{n,t}^{(\epsilon)}(\beta)\hat{I}_{n,t}^{(y)}$ and there are GARCH effects $\beta_2^0 + \beta_3^0 > 0$, then by dominated convergence and error independence $\hat{I}_{n,t}^{(y)}$ has no impact on the scale asymptotically.

All conditions are essentially trivial as in the linear AR case.

COROLLARY 3.4 (*Strong-GARCH*). *Assumptions D, I, L, MS, MX, N, RB.ii-iv, RS.ii and STM hold.*

i. Assume there are GARCH effects $\beta_2^0 + \beta_3^0 > 0$. *If* $\hat{I}_{n,t}(\beta) = \hat{I}_{n,t}^{(\epsilon)}(\beta)$ *or* $\hat{I}_{n,t}(\beta) = \hat{I}_{n,t}^{(\epsilon)}(\beta)\hat{I}_{n,t}^{(y)}$ *then* $\mathring{\mathcal{V}}_n \sim n(E[\epsilon_{n,t}^{*4}] - 1)^{-1} \times E[\mathfrak{s}_t\mathfrak{s}_t']$.

ii. Assume there are no GARCH effects $\beta_2^0 + \beta_3^0 = 0$. *If* $\hat{I}_{n,t}(\beta) = \hat{I}_{n,t}^{(\epsilon)}(\beta)\hat{I}_{n,t}^{(y)}$ *then* $\mathring{\mathcal{V}}_n \sim n(E[\epsilon_{n,t}^{*4}] - 1)^{-1} \times E[\mathfrak{s}_{n,t}^{*}\mathfrak{s}_{n,t}^{*\prime}]$.

Ling's (2007) QMWL criterion $\sum_{t=1}^{n} w_t\{\ln h_t(\beta) + y_t^2/h_t(\beta)\}$ has a weight $w_t = w(y_{t-1}, y_{t-2}, ...)$ based on infinitely many variables if the model is GARCH, although $w_t = w(y_{t-1}, ..., y_{t-p})$ in the ARCH(p) case. Ling proposes such a weight to robustify against heavy tails in y_t, but *not* in ϵ_t, hence $E[\epsilon_t^4] < \infty$ must hold for asymptotic normality, and he does not show how to compute $w(y_{t-1}, y_{t-2}, ...)$ in practice. QMTTL then has two advantages: there are at most three objects ϵ_t^2, y_{t-1} and y_{t-2} to base trimming on, and we only require $E[\epsilon_t^2] < \infty$. As it turns out this requires a substantially different weak limit theory because ϵ_t is not observed.

3.4 AR-GARCH

The last case we consider is AR-GARCH: $y_t = c^0 + \sum_{i=1}^{p} \zeta_i^0 y_{t-i} + u_t = \phi^{0\prime} x_t + u_t$ where $u_t = \sigma_t\epsilon_t$, and $\sigma_t^2 = \beta_1^0 + \beta_2^0 u_{t-1}^2 + \beta_3^0 \sigma_{t-1}^2$. Assume the AR roots are outside the unit circle, $\beta_1^0 > 0$, $\beta_2^0, \beta_3^0 \in (0,1)$, and $\epsilon_t \overset{iid}{\sim} (0,1)$ satisfies Example 3. The only source of extremes is ϵ_t and y_{t-i}, exactly as in AR and pure GARCH. The QMTTLE therefore solves

$$\hat{\theta}_n = \underset{\beta \in \mathcal{B}}{\mathrm{argmin}} \left\{ \sum_{t=1}^{n} \left(\ln h_t(\theta) + \frac{(y_t - \phi' x_t)^2}{h_t(\theta)} \right) \hat{I}_{n,t}^{(\epsilon)}(\theta)\hat{I}_{n,t}^{(y)} \right\}.$$

where $\hat{I}_{n,t}^{(y)} = \prod_{i=1}^{p} \hat{I}_{i,n,t}^{(y)}$ and $\hat{I}_{i,n,t}^{(y)} := I(|y_{t-i}| \leq y_{(k_n^{(y)})}^{(a)})$. If we allow for no effects $\beta_2^0 = \beta_3^0 = 0$ then by exploiting the logic from Example 3 it can be shown $\hat{I}_{n,t}^{(y)} = \prod_{i=1}^{p+1} \hat{I}_{i,n,t}^{(y)}$ suffices. Further, although laborious to show the scale $\mathring{\mathcal{V}}_n$ is block diagonal, and the block associated with GARCH parameters β has the same form as in the pure-GARCH case (cf. Francq and Zakoïan 2004). Stationarity is shown in Lee (2007), Ling (2007) and Meitz and Saikkonen (2008) amongst others, while Cline (2007) details power law properties.

4. RATES OF CONVERGENCE AND ASYMPTOTIC COVARIANCE

The focused trimming of NLLTTS and QMTTL lends itself to elegant characterizations of convergence rates and asymptotic covariances. If any diagonal matrix $\mathcal{N}_n \in \mathbb{R}_+^{q\times q}$ with elements $\mathcal{N}_{i,n} \to \infty$ satisfies $\mathring{\mathcal{V}} := \lim_{n\to\infty} \mathring{\mathcal{V}}_n \times \mathcal{N}_n^{-1} \in (0,\infty)$ we call $\mathcal{N}_{i,n}^{1/2}$ the rate of convergence[5] of $\hat{\theta}_{i,n}$, and $\mathring{\mathcal{V}}$ the inverted asymptotic covariance matrix. Characterizing *unique* $\mathcal{N}_{i,n}$ and $\mathring{\mathcal{V}}$, however, is impossible without loss and response function specifications, and even then they depend upon a trimming policy. We therefore focus on nonlinear AR and GARCH, and symmetrically trim to reduce notation.

4.1 NLLTTS RATE OF CONVERGENCE

By Example 1 when $y_t = f(x_t,\phi) + u_t$ with iid u_t then

$$(9)\qquad \mathring{\mathcal{V}}_{i,i,n}^{1/2} \sim n^{1/2}\frac{\left(E\left[f_{i,t}^{\phi 2} I\left(\left|f_{i,t}^{\phi}\right| \le \mathcal{C}_{i,n}^{(f^\phi)}\right)\right]\right)^{1/2}}{\left(E\left[u_t^2 I\left(|u_t| \le \mathcal{C}_n^{(u)}\right)\right]\right)^{1/2}}.$$

Assume u_t and stochastic $f_{i,t}^{\phi}$ have power law tails (6) with indices $\kappa_u, \kappa_i > 0$ and $d_u, d_i > 0$. Karamata's Theorem implies for some s.v. function $L(\cdot)$ each $w_t \in \{u_t, f_{i,t}^{\phi}\}$ satisfies

$$(10)\qquad E\left[w_t^2 I\left(|w_t| \le c\right)\right] \sim \frac{\kappa_w}{2-\kappa_w} c^2 P\left(|w_t| > c\right) \quad \text{if} \quad \kappa_w \in (0,2)$$

$$E\left[w_t^2 I\left(|w_t| \le c\right)\right] \sim L(c) \quad \text{if} \quad \kappa_w = 2,$$

and by construction

$$(11)\qquad \mathcal{C}_{i,n}^{(f^\phi)} = d_i^{1/\kappa_i}(n/k_{i,n}^{(f^\phi)})^{1/\kappa_i} \quad \text{and} \quad \mathcal{C}_n^{(u)} = d_u^{1/\kappa_u}(n/k_n^{(u)})^{1/\kappa_u}.$$

The case of non-stochastic $f_{i,t}^{\phi}$ and stochastic $f_{i,t}^{\phi}$ with $\kappa_i > 2$ are identical, so assume $f_{i,t}^{\phi}$ is stochastic. Together (9)-(11) characterize $\mathring{\mathcal{V}}_{i,i,n}^{1/2}$: if $\kappa_i < 2$ and $\kappa_u < 2$, for example, then

$$(12)\qquad \mathring{\mathcal{V}}_{i,i,n}^{1/2} \sim n^{1/2}\frac{(n/k_{i,n}^{(f^\phi)})^{1/\kappa_i - 1/2}}{(n/k_n^{(u)})^{1/\kappa_u - 1/2}} \times \left(\frac{\kappa_i\left(2-\kappa_u\right)}{\kappa_u\left(2-\kappa_i\right)}\right)^{1/2}\frac{d_i^{1/\kappa_i}}{d_u^{1/\kappa_u}}.$$

Independent errors u_t with an infinite variance have a purely adverse affect, so maximal trimming by u_t should be imposed by optimizing $k_n^{(u)} \to \infty$. Conversely, regressors have a pure leverage effect, so minimal trimming by $f_{i,t}^{\phi}$ is optimal, hence slow $k_{i,n}^{(f^\phi)} \to \infty$. The upper-bound $\mathring{\mathcal{V}}_{i,i,n}^{1/2} < Kn^{1/\kappa_i}$ follows by setting $k_n^{(u)} \sim Kn$ and $k_{i,n}^{(f^\phi)} \sim K$, outside their allowed bounds for $k_n = o(n)$ and $k_n \to \infty$.

Thus, we can never do better than the fastest rate n^{1/κ_i} amongst M-estimators (see Davis et al 1992), but we can come close. Since any $\{k_n^{(u)}, k_{i,n}^{(f^\phi)}\}$ effects the rate *and* covariance, consider a simple class that permits identification of both: for each $w_t \in \{u_t, f_{i,t}^{\phi}\}$ assume

$$k_n^{(w)} \sim \lambda_w b_n^{(w)} \;:\; b_n^{(w)} \text{ is not a function of } \lambda_w \in (0,1)\,,\; b_n^{(w)} \to \infty,\; 1 \le b_n^{(w)} < n,\; b_n^{(w)} = o(n).$$

Thus, λ_w captures scale in $k_n^{(w)} \to \infty$ while $b_n^{(w)} \to \infty$ is scaleless. Examples are $b_n^{(w)} \in \{\ln(n)$, $n/\ln(n)$, $n^{\delta_w}\}$ where $\delta_w \in (0,1)$. The policy $\{b_n^{(u)}, b_{i,n}^{(f^\phi)}\} = \{n/\ln(n), \ln(n)\}$, for example, elevates

[5] Since $\mathring{\mathcal{V}}_{i,i,n}$ may be heterogeneous, linear combinations $\lambda'\hat{\theta}_n$ may have different rates that can be exploited for efficiency gains, an issue we do not treat. See Antoine and Renault (2010).

error and diminishes regressor trimming, and if $\kappa_u < 2$ and $\kappa_i < 2$ then from (12)

$$\mathring{\mathcal{V}}_{i,i,n}^{1/2} \sim \frac{n^{1/\kappa_i}}{L(n)} \times \left\{ \left(\frac{\kappa_i (2-\kappa_u)}{\kappa_u (2-\kappa_i)} \right)^{1/2} \frac{d_i^{1/\kappa_i}}{d_u^{1/\kappa_u}} \frac{\lambda_u^{1/\kappa_u - 1/2}}{\lambda_i^{1/\kappa_i - 1/2}} \right\} = \mathcal{N}_{i,n}^{1/2} \times \mathring{\mathcal{V}}_{i,i}^{1/2}(\kappa, d, \lambda), \tag{13}$$

say. Trimming by u_t at a rate nearly equal to a fixed quantile $\lambda_u n/\ln(n) < \lambda_u n$ and diminishing regressor trimming almost to nothing, i.e. slowly varying $\lambda_y \ln(n)$, ensures the rate $\mathcal{N}_{i,n}^{1/2} = n^{1/\kappa_i}/L(n)$ is infinitesimally close to n^{1/κ_i}. We believe this is a first in the literature: asymptotic normality *and* (nearly) the highest possible rate. Obviously similar results are obtained if any s.v. functions are used in place of $\ln(n)$.

Heavier tailed regressors $\kappa_i \searrow 0$ and/or $d_i \nearrow \infty$ are associated with greater efficiency $\mathring{\mathcal{V}}_{i,i}^{-1}(\kappa, d, \lambda) \searrow 0$. The same applies for thin tailed errors $\kappa_i \nearrow 2$ and/or $d_i \searrow 0$. Minimal regressor trimming by scale $\lambda_i \searrow 0$ or elevated error trimming in the sense of large λ_u optimizes efficiency[6].

EXAMPLE 4 (AR and LTTS Rate): Consider the AR model of Example 2. Since we can write $y_t = \mu + \sum_{i=0}^{\infty} \psi_i u_{t-i}$ where $\psi_i = O(\rho^i)$ and $\rho \in (0,1)$, both u_t and y_t have tail index κ, and y_t has scale $d_y = d_u(\sum_{i=0}^{\infty} |\psi_i|^\kappa)$ (e.g. Brockwell and Cline 1985). Stationary regressors y_{t-i} are all assigned the same fractile $k_n^{(y)}$, and put $k_n^{(u)} = [\lambda_u n/\ln(n)]$ and $k_n^{(y)} = [\lambda_y \ln(n)]$.

The constant term cannot provide leverage by scale formula (9), hence $\hat{\phi}_{1,n} = o_p(n^{1/2})$ when the error has an infinite variance $\kappa \le 2$. However, if $\kappa < 2$ then by (13)

$$\mathring{\mathcal{V}}_{1+i,1+i,n}^{1/2} \sim \frac{n^{1/\kappa}}{L(n)} \times \left\{ \left(\sum_{j=0}^{\infty} |\psi_j|^\kappa \right)^{1/\kappa} \left(\frac{\lambda_u}{\lambda_y} \right)^{1/\kappa - 1/2} \right\} = \mathcal{N}_{1+i,n}^{1/2} \times \mathring{\mathcal{V}}_{1+i,1+i}^{1/2}, \ i = 1, ..., p, \tag{14}$$

where small λ_y and large λ_u are optimal.

4.2 ASYMPTOTIC VARIANCES: LTTS, OLS, LWAD

We now compare LTTS to OLS and LWAD for AR(1) $y_t = \phi y_{t-1} + u_t$, $|\phi| < 1$. Denote by $\mathfrak{V}$ any asymptotic variance, e.g. $\mathfrak{V} := \mathring{\mathcal{V}}^{-1}$. Assume u_t is iid with power law tail (6) and index $\kappa > 0$, and consider $\{k_n^{(u)}, k_n^{(y)}\} \sim \{\lambda_u n/\ln(n), \lambda_y \ln(n)\}$. Then by (14)

$$\mathfrak{V} = \mathfrak{V}_{LTTS} = \mathfrak{V}_{LTTS}(\lambda, \kappa, \phi) := (1 - |\phi|^\kappa)^{2/\kappa} \times (\lambda_y/\lambda_u)^{2/\kappa - 1} \text{ if } \kappa \in (0,2), \text{ and } 1 - \phi^2 \text{ if } \kappa > 2.$$

The case $\kappa = 2$ can similarly be deduced from (9)-(10). Trivially $\mathfrak{V} \searrow 0$ as $|\phi| \nearrow 1$, identical to LS in the finite variance case, and $\mathfrak{V} \searrow 0$ as regressor trimming is diminished $\lambda_y \searrow 0$. Further, efficiency improves $\mathfrak{V} \searrow 0$ for heavier tails $\kappa \searrow 0$ as long as $\lambda_u/\lambda_y > (1 - |\phi|^\kappa)^2 \times |\phi|^{2\kappa|\phi|^\kappa(1-|\phi|^\kappa)^{-1}}$. We already bias trimming toward the error, but since the right hand side is bounded from above by 1 for all $|\phi| < 1$, if we further bias trimming $\lambda_u \ge \lambda_y$ then greater error volatility leads to improved efficiency. This follows since large u_t lead to large future leverage points y_{t+i}.

By comparison, if $\kappa < 2$ the least squares estimator $n^{1/\kappa} \times (\hat{\phi}_n^{LS} - \phi) \xrightarrow{d} (1 - \phi^2) \times (1 - |\phi|^\kappa)^{-1/\kappa} \times S_1/S_0$ where (S_0, S_1) are unit scale independent stable laws with indices $(\kappa/2, \kappa)$, S_1 is symmetric and $S_0 \ge 0$ *a.s.* (Davis and Resnick 1986: Example 5.3). Thus

$$\mathfrak{V}_{LS}(\kappa, \phi) := \left(1 - \phi^2\right)^\kappa / (1 - |\phi|^\kappa) \quad \text{if } \kappa \in (0,2).$$

Since LTTS trimming ensures standard asymptotics, the scale $\mathfrak{V}_{LTTS}$ is more akin to LS for finite variance u_t than to LS for infinite variance u_t.

---

[6] Clearly $\lambda_u \nearrow \infty$ drives $\mathcal{V}_{i,i}^{-1}(\kappa, d, \lambda) \searrow 0$, but this is ruled out since $1 \le k_n^{(u)} \le n$ must hold.

Now consider Ling's (2005) LWAD estimator $\operatorname{argmin}_{\phi\in\Phi}\{\sum_{t=1}^{n} w_t|u_t(\phi)|\}$ and let $\mathfrak{f}(u)$ denote the density of u_t, assumed differentiable and $\mathfrak{f}(0) > 0$ (Ling 2005: p. 383). Then similar to Corollary 1 of Davis and Dunsmuir (1997)

$$\mathfrak{V} = \mathfrak{V}_{LAWD}(c) := \frac{1}{4\mathfrak{f}(0)^2}\frac{E\left[w_t^2(c)\,y_{t-1}^2\right]}{\left(E\left[w_t(c)\,y_{t-1}^2\right]\right)^2}.$$

As trimming is diminished $c \to \infty$ then $\mathfrak{V} = O(1) \times (4\mathfrak{f}(0))^{-2} \times (E[y_{t-1}^2 I(|y_{t-1}| \leq c)] - c^{2-\kappa})^{-1}$. See Ling (2005: p. 384). Larger c are associated with smaller $\mathfrak{V}$, but *only in the limit*. Ling does not characterize the $O(1)$ term, nor $\mathfrak{V}$ for *fixed* c, the only case considered. This highlights another advantage of LTTS: since our thresholds $\mathcal{C}_n \to \infty$ as $n \to \infty$, we obtain an elegant asymptotic variance expression.

4.3 STRONG-GARCH

Assume $\beta_2^0 + \beta_3^0 > 0$; the case of no GARCH effects being similar. By Corollary 3.4 the QMTTL scale is $\mathring{\mathcal{V}}_{i,i,n}^{1/2} \sim n^{1/2}(E[\epsilon_{n,t}^{*4} - 1])^{-1/2}(E[\mathfrak{s}_{i,i,t}^2])^{1/2} = O(n^{1/2})$. The QMTTLE is never super-$n^{1/2}$-convergent due to scaling in the QML criterion, and is $o_p(n^{1/2})$ if $E[\epsilon_t^4] = \infty$, hence maximum error trimming is optimal. By Lyapunov's inequality and (10), if $\kappa_\epsilon \in (2,4)$ then

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

Set $k_n^{(\epsilon)} = [\lambda n/L(n)]$ to obtain $\mathcal{N}_{i,n}^{1/2} = n^{1/2}/L(n)$ and $\mathring{\mathcal{V}}_{i,i}^{-1} = (E[\mathfrak{s}_{i,t}^2])^{-1}(\kappa_\epsilon/(4-\kappa_\epsilon))d_\epsilon^{2/\kappa_\epsilon}\lambda^{-(4/\kappa_\epsilon - 1)}$. By (12) if $\kappa_\epsilon = 4$ then $\mathring{\mathcal{V}}_{i,i,n}^{1/2} \sim n^{1/2}/L(n)$. Hall and Yao (2003) show the QML rate is $n^{1-2/\kappa_\epsilon}/L(n) \leq \mathcal{N}_{i,n}^{1/2}$ for any $\kappa_\epsilon \in (2,4]$, with strict inequality if $\kappa_\epsilon < 4$. The rate extends to Ling's (2007) QMWL since his weighting mechanism does not affect ϵ_t. Thus QMTTL has the plural advantages of allowing $\kappa_\epsilon \in (2,4]$, while obtaining asymptotic normality *and* achieving a greater rate than QML and QMWL when $\kappa_\epsilon \in (2,4)$.

An estimator comparable to QMTTL is Log-LAD since it only requires $E[\epsilon_t^2] = 1$ and symmetric $\ln(\epsilon_t^2)$ for asymptotic normality, and is $n^{1/2}$-convergent (Peng and Yao 2003: Theorem 1).

5. INFERENCE AND FRACTILE SELECTION

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

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

Clearly $\hat{\mathcal{V}}_n$ is used for inference (see Section 5.1), but we also propose a covariance determinant method for selecting the fractiles k_n (see Section 5.2).

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

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

Notice $\widehat{\mathcal{G}}_n(\theta)$ and $\widehat{\mathcal{S}}_n(\theta)$ are trimmed by the composite $\hat{I}_{n,t}(\theta)$. This is required to generate consistent estimators of $\mathcal{G}_n$ and $\mathcal{S}_n$, while the latter involve composite trimming due to the TTME criterion construction $1/n\sum_{t=1}^{n} l(z_t,\theta)\hat{I}_{n,t}(\theta)$.

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

THEOREM 5.1. *Assume $\mathcal{K}(\cdot)$ is continuous at 0 and all but a finite number of points, $\mathcal{K} : \mathbb{R} \to [-1,1]$, $\mathcal{K}(0) = 1$, $\mathcal{K}(x) = \mathcal{K}(-x)$ $\forall x \in \mathbb{R}$, $\int_{-\infty}^{\infty} |\mathcal{K}(x)|dx < \infty$, and $\int_{-\infty}^{\infty} |\varpi(\xi)|d\xi < \infty\}$. Let $\sum_{s,t=1}^{n} |\mathcal{K}((s-t)/\gamma_n)| = o(n^2)$, $\max_{1\leq s\leq n} \sum_{t=1}^{n} \mathcal{K}((s-t)/\gamma_n) = o(n)$ and bandwidth $\gamma_n = o(n)$. Under D, F, I, L, MS, MX, N, RB, RS and STM $\hat{\mathcal{V}}_n = \mathcal{V}_n(1 + o_p(1))$.*

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

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

Remark 3: HAC consistency exploits the Theorem 2.2 rate $\mathcal{V}_n^{1/2}(\hat{\theta}_n - \theta^0) = O_p(1)$ by the proofs of Lemma B.8 and C.3 in Appendix C. Any consistent plug-in $\mathcal{V}_n^{1/2}(\tilde{\theta}_n - \theta^0) = O_p(1)$ can therefore be used, which is helpful for $\hat{\mathcal{V}}_n$-based fractile selection methods: see below.

Now consider QML loss. If u_t is iid in a nonlinear AR, then by Theorem 4.1 and Example 1 a natural estimator of $\mathring{\mathcal{V}}_{i,j,n} \sim nE[f_{i,t}^{\phi}I_{i,n,t}^{(f^{\phi})}f_{i,t}^{\phi}I_{j,n,t}^{(f^{\phi})}]/E[u_{n,t}^{*2}]$ is

$$\hat{\mathcal{V}}_{i,j,n} = \hat{\mathcal{V}}_{i,j,n}(\hat{\phi}_n) = n\left(\frac{1}{n}\sum_{t=1}^{n} f_{i,t}^{\phi}(\hat{\phi}_n)\hat{I}_{j,n,t}^{(f^{\phi})}(\hat{\phi}_n) \times f_{j,t}^{\phi}(\hat{\phi}_n)\hat{I}_{j,n,t}^{(f^{\phi})}(\hat{\phi}_n)\right) \times \left(\frac{1}{n}\sum_{t=1}^{n} u_t^2(\hat{\phi}_n)\hat{I}_{n,t}^{(u)}(\hat{\phi}_n)\right)^{-1}.$$

A linear AR has $f_t^{\phi}(\phi) = x_t := [1, y_{t-1}, ..., y_{t-p}]'$ hence

$$\hat{\mathcal{V}}_{i,j,n} = \hat{\mathcal{V}}_{i,j,n}(\hat{\phi}_n) = n\left(\frac{1}{n}\sum_{t=1}^{n} y_{t-i}\hat{I}_{i,n,t}^{(y)}y_{t-j}\hat{I}_{j,n,t}^{(y)}\right) \times \left(\frac{1}{n}\sum_{t=1}^{n} u_t^2(\hat{\phi}_n)\hat{I}_{n,t}^{(u)}(\hat{\phi}_n)\right)^{-1}. \tag{15}$$

As above, any consistent plug-in $\mathring{\mathcal{V}}_n^{1/2}(\tilde{\phi}_n - \phi^0) = O_p(1)$ may be used.

COROLLARY 5.2. *Let u_t be iid. Under D, F, I, MS, MX, N, RS and STM $\hat{\mathcal{V}}_n = \mathring{\mathcal{V}}_n(1 + o_p(1))$.*

Scale estimation for QMTTL for GARCH is essentially identical in lieu of the form. If $\beta_2^0 + \beta_3^0 > 0$, for example, then by Corollary 4.4 the appropriate estimator is

$$\hat{\mathcal{V}}_n = n\left(\frac{1}{n}\sum_{t=1}^{n} \hat{\mathfrak{s}}_t\hat{\mathfrak{s}}_t'\right) \times \left(\frac{1}{n}\sum_{t=1}^{n} \epsilon_t^4(\hat{\beta}_n)\hat{I}_{n,t}^{(\epsilon)}(\hat{\beta}_n) - 1\right)^{-1}, \tag{16}$$

where $\hat{\mathfrak{s}}_t := h_t^{-1}(\hat{\beta}_n)h_t^{\beta}(\hat{\beta}_n)$, with initial values $h_0(\beta) = \beta_1$ and $h_0^{\beta}(\beta) = [1, 0, 0]'$.

5.1 Inference

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

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

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

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

5.2 Fractile Selection

A variety of covariance methods have been proposed for outlier robust estimation, including Minimum Covariance Determinant (e.g. Rousseeuw 1984, Rousseeuw et al 2004, Agulló et al 2008), sign and rank methods (e.g. Ollila et al 2002), and bootstrap covariance for trimmed means (Léger and Romano 1990). We adopt asymptotic covariance and small sample bootstrap mean-squared-error methods for our purposes, while space constraints force us to leave deeper theory issues aside.

We focus ideas on LTTS for AR with iid u_t and index $\kappa > 0$, and the choice of $\lambda := [\lambda_u, \lambda_y]'$ $\in (0,\infty)^2$ for $\{k_n^{(u)}, k_n^{(y)}\} = \{[\lambda_u n/\ln(n)], [\lambda_y \ln(n)]\}$. Write $\hat{\mathcal{V}}_n$ in (15) as $\hat{\mathcal{V}}_n(\lambda)$ to reflect λ.

Psuedo-MCD Define the normalized covariance determinant on compact $\Lambda \subset (0,\infty)^2$:

$$d_n(\lambda) := \det \hat{\mathcal{V}}_n^{-1}(\lambda)/\max_{\lambda\in\Lambda}\left\{\det \hat{\mathcal{V}}_n^{-1}(\lambda)\right\} \in [0,1].$$

In practice LTTS, OLS and LAD are all valid choices as a plug-in $\tilde{\phi}_n$ for computing $\hat{\mathcal{V}}_n(\lambda)$, in particular since the latter two satisfy $\mathring{\mathcal{V}}_n^{1/2}(\lambda) \times (\tilde{\phi}_n - \phi^0) \xrightarrow{p} 0$ in the infinite variance case.

If $\kappa < 2$ then $\hat{\mathcal{V}}_n^{-1}(\lambda)$ is proportional to $(\lambda_y/\lambda_u)^{2/\kappa-1}$ by (14) and Corollary 5.2, in which case minimizing $d_n(\lambda)$ always leads to a corner solution, a dilemma that applies to all processes studied in Section 4. But the improvement in $d_n(\lambda)$ by increasing λ_u and decreasing λ_y diminishes precipitously, as shown in Figure 1. We plot a typical $d_n(\lambda)$ for an AR(2) $y_t = .2 + .8y_{t-1} - .3y_{t-2}$ $+ u_t$, u_t is iid Pareto with index $\kappa = 1.5$, we use a least squares plug-in for $\hat{\mathcal{V}}_n(\lambda)$, and $n = 100$. In simulation experiments essentially any $\lambda_u \in [.05, .25]$ and $\lambda_y \in [.10, 1.5]$, corresponding to where $d_n(\lambda)$ flattens out, leads to a sharp and approximately normal estimator, even for $n = 100$ (see Section 6). If $\kappa \geq 2$ then $\mathring{\mathcal{V}}_n$ is not affected by λ, so a small sample technique is preferred.

Boostrap-MSE The statistic $\hat{\mathcal{V}}_n(\lambda)$ estimates the asymptotic scale $\mathring{\mathcal{V}}_n(\lambda)$, while the latter does not reveal possible small sample bias due to trimming, and is not affected by λ as $n \to \infty$ when $\kappa \geq 2$. A better method may be to approximate the small sample mean-squared-error $E[(\hat{\phi}_n - \phi^0)(\hat{\phi}_n - \phi^0)']$. Let $\{\hat{\phi}_{r,n}(\lambda)\}_{r=1}^{R_n}$ be a sequence of bootstrap LTTS estimates computed with λ, use any consistent $\tilde{\phi}_n$ as plug-in for ϕ^0, and define $\hat{\mathcal{M}}_n(\lambda) := 1/R_n\sum_{r=1}^{R_n}(\hat{\phi}_{r,n}(\lambda) - \tilde{\phi}_n) \times (\hat{\phi}_{r,n}(\lambda) - \tilde{\phi}_n)'$. See Gonçalves and White (2005) for background theory. Any valid norm may be minimized on Λ, including the matrix norm $||\hat{\mathcal{M}}_n(\lambda)||$. In Figure 2 we plot $||\hat{\mathcal{M}}_n(\lambda)||$ for the same AR(2) sample, based on $R_n = n$ bootstrap draws with replacement, and sub-sample size $n/2 = 50$[7]. The minimizing (λ_u, λ_y) are $\lambda_u = .15$ and $\lambda_y = .45$, roughly aligning with where $d_n(\lambda)$ flattens. In Section 6 we show $||\hat{\mathcal{M}}_n(\lambda)|$ leads to a sharp, and approximately normal estimator.

[7] The r^{th} bootstrapped sample is $\{y_{t_r^*}, y_{t_r^*+1}, ..., y_{t_r^*+[n/2]-1}\}$ where t_r^* is a uniform random draw from $\{1, ..., n - [n/2] + 1\}$.

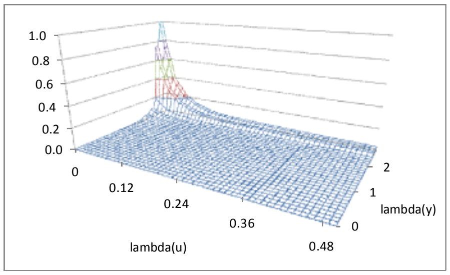

Figure 1: $d_n(\lambda)$ for AR(2) with $\kappa = 1.5$

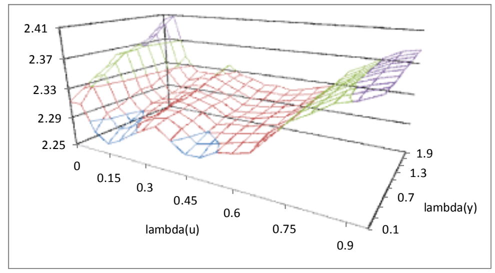

Figure 2: $||\text{mse}(\lambda)||$ for AR(2) with $\kappa = 1.5$

6. SIMULATION STUDY

We now compare LTTS to OLS, LAD and LWAD for linear and nonlinear AR models, and QMTTL to QML, QMWL and Log-LAD for GARCH.

6.1 AR, TAR, LSTAR : LTTS vs. OLS, LAD, and LWAD

We draw $2n$ observations of y_t from three AR models and retain the last $n \in \{100, 800\}$. The models are AR(2) $y_t = .2 + .8y_{t-1} - .3y_{t-2} + u_t$, Threshold AR(1) $y_t = .2 + .6y_{t-1}I(y_{t-1} > 0) + u_t$, or Logistic Smooth Transition AR(1) $y_t = y_{t-1}(1 + \exp\{.25y_{t-1}\})^{-1} + u_t$, each denoted $y_t = f(\phi^0, x_t) + u_t$. The starting value is $y_1 = u_1$, and we retain the last n. We use iid or GARCH errors u_t. The iid errors are zero symmetric Pareto $\bar{P}_\kappa(u) := P(u_t > u) = P(u_t < -u) = .5 \times (1 + u)^{-\kappa}$ with index $\kappa \in \{.75, 1.5, 2.5\}$. The GARCH error is $u_t = \sigma_t \epsilon_t$, $\sigma_t^2 = .3 + .6u_{t-1}^2 + .3\sigma_{t-1}^2$ with iid ϵ_t governed by $\bar{P}_{2.5}$, standardized such that $\epsilon_t \overset{iid}{\sim} (0,1)$, hence $\kappa_y = 1.50$[8]. We simulate 1000 series $\{y_t\}_{t=1}^n$.

The AR and TAR models are estimated by LTTS $\hat{\phi}_n = \text{argmin}_{\phi\in\Phi}\{1/n \sum_{t=k+1}^n (y_t - \phi' x_t)^2 \hat{I}_{n,t}^{(u)}(\phi)\hat{I}_{n,t}^{(y)}\}$. We estimate LSTAR by NLLTS $\hat{\phi}_n = \text{argmin}_{\phi\in\Phi}\{1/n \sum_{t=k+1}^n (y_t - f(x_t, \phi))^2 \hat{I}_{n,t}^{(u)}(\phi)\hat{I}_{n,t}^{(f^\phi)}(\phi)\}$ where $f(x_t, \phi) = y_{t-1}/(1 + \exp\{\phi y_{t-1}\})$ and $f_t^\phi(\phi) = -y_{t-1} \times \exp\{\phi y_{t-1}\}/(1 + \exp\{\phi y_{t-1}\})^2$. Note $k = 1$ for TAR(1) and LSTAR(1), and $k = 2$ for AR(2).

Both are computed with $k_n^{(u)} = [\lambda_u n/\ln(n)]$ and $k_n^{(y)} = [\lambda_y \ln(n)]$. Initially we fix $\{\lambda_u, \lambda_y\} = \{.05, 1.0\}$, where $\lambda_u = .05$ matches Ling's (2005) LWAD setting, discussed below, and since any $\lambda_y \in [.10, 1.5]$ renders an approximately normal estimator we simply choose $\lambda_y = 1$. Clearly in any one sample this amounts to trimming a fixed number of observations, so we investigate below the rate of convergence as n increases.

We then use the bootstrap-mse procedure of Section 5.2, with $R_n = [n]$ sub-samples of size $n/2$ with replacement, and a least squares plug-in. We choose λ by minimizing $||\hat{\mathcal{M}}_n(\lambda)||$ over $\lambda_u \in \Lambda_n(1)$ and $\lambda_y \in \Lambda_n(3)$, where $\Lambda_n(N) := \{1/n, 2/n, ..., N\}$.

The LWAD estimator is $\text{argmin}_{\phi\in\Phi}\{\sum_{t=k+1}^n w_t|u_t(\phi)|\}$ with Ling's (2005) only suggested weight: $w_t = 1$ if $a_t = 0$ and $w_t = (y_{([\lambda n])}^{(a)})^3/a_t^3$ if $a_t \neq 0$, where $a_t := \sum_{i=1}^2 |y_{t-i}|I(|y_{t-i}| \geq y_{([\lambda n])}^{(a)})$. The percentile is $\lambda = .05$ as in Ling (2005).

Both LTTS and LWAD involve an iterative estimation algorithm. Starting values $\hat{\phi}_n^{(0)}$ are randomly selected from a uniform distribution on $\Phi = -[2, 2]^3$, and we update $\hat{\phi}_n$ iteratively subject to $\hat{\phi}_n \in \Phi$. If $\hat{\phi}_n^{(j)}$ is the j^{th} value we cease iterations when $||\hat{\phi}_n^{(j)} - \hat{\phi}_n^{(j-1)}|| \leq .001$.

We perform four experiments concerning estimation, rate of convergence and inference.

[8] The index in the GARCH case satisfies $E[(.3\epsilon_t^2 + .6)^{\kappa_y/2}] = 1$ (e.g. Basrak et al 2002). The index κ_y is computed as $\hat{\kappa} = \arg\min_{\kappa \in K}\{|1/N \sum_{t=1}^N (.3\epsilon_t^2 + .6)^{\kappa/2} - 1|\}$ over $K \in \{.01, .02, ..., 10\}$ based on $N = 100,000$ iid random draws ϵ_t from $P_{2.5}$. The 1% band half-length is less than .001.

6.1.2 Estimation See Tables 1-3 for simulation means, mean-squared-errors, percent of trimmed sample loss $\epsilon_t^2(\phi)$, and Kolmogorov-Smirnov tests of normality for the estimator $\hat{\phi}_{p,n}$ of ϕ_p^0. The KS test is based on the standardization $(\hat{\phi}_{p,n} - \phi_p^0)/s_{p,n}^{1/2}$ where $s_{p,n}^{1/2}$ is the empirical standard deviation of $\hat{\phi}_{p,n}$ based on the 1000 independently drawn samples. We relegate to the supplementary Appendix C redundant results based on iid $\bar{P}_{1.5}$ errors and LAD estimation.

Each estimator for AR is accurate in all cases. Both LTTS and LWAD are roughly normal, even for $n = 100$, while OLS and LAD fail normality tests when variance is infinite, as expected. Each estimator exhibits negative bias for TAR, and LWAD substantially deviates from normality when $\kappa < 1$ like due to symmetric weighting based only on y_{t-i}. LTTS exhibits the least bias, smallest mse, and is closest to normal in most cases, arguably due to its focused treatment of error versus regressor extremes. Finally, for LSTAR each estimator is sharp, but LWAD exhibits the greatest mse, while in most cases NLTTS has the lowest mse and is closest to normal.

The bootstrap-mse method works very well. The simulation average parameters λ in the AR case, for example, have a range $\lambda_u \in (.10, .40)$ and $\lambda_y \in (.45, 2.0)$ across cases, although fixed $(\lambda_u, \lambda_y) = (.05, 1.0)$ work just as well. If the mean is finite $\kappa > 1$ the optimally selected $\lambda's$ are no larger than .5, suggesting remarkably little trimming is required to ensure approximate normality. If variance is finite then for large n the optimal (λ_u, λ_y) are small because trimming cannot improve $\hat{\phi}_n$.

6.1.3 Rate of Convergence In a second experiment we use the fact that the empirical standard deviation $s_{p,n}^{1/2}$ is proportional to the rate of convergence $\mathcal{N}_{p,n}^{1/2}$. We compute $s_{p,n}^{1/2}$ for the AR(2) model with iid $\bar{P}_{1.5}$ error for OLS, LWAD, and LTTS with fixed $(\lambda_u, \lambda_y) = (.05, 1.0)$ over 1000 samples for each size $n \in \{400, 410, 420, ..., 1000\}$. There it little difference amongst the three estimators for $n \leq 400$. Recall $s_{p,n}^{1/2}$ is proportional to $n^{1/2}$ for LWAD, to $n^{1/\kappa}/L(n)$ for LTTS, and to $n^{1/\kappa}$ for OLS. See Figure 3 for plots of $b(n) \in \{n, s_{p,n}\}$, standardized such that $b(400) = 1$. The empirical rates match theory: OLS is fastest, LTTS is above $n^{1/2}$, and LWAD hovers around $n^{1/2}$.

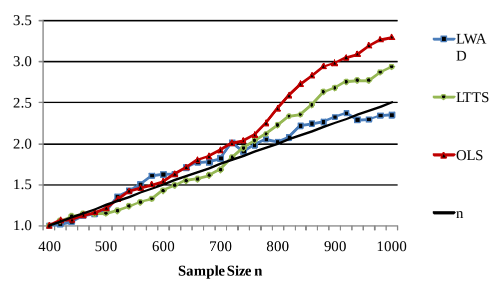

Figure 3: Sample MSE's

6.1.4 Inference: t-Tests We perform three t-tests using simulation standard errors. We test $\phi_p = \phi_p^0$ and $\phi_p = 0$, and $\phi_p = \phi_p^0 + .1$ for AR, and $\phi_p = \phi_p^0 - .2$ for TAR. All tests are asymptotic, performed at the 5%-level. We use the fixed trimming parameters $(\lambda_u, \lambda_y) = (.05, 1.0)$ as a benchmark. See Table 4. In the AR case OLS and LWAD exhibit size distortions, while LTTS performs best under the null overall. This is hardly surprising since LTTS trumps all estimators in terms of approximate normality. Under the two alternatives LTTS exhibits fairly low power when $n = 100$ where LWAD has a clear advantage when the errors are iid, but the advantage vanishes for larger sample sizes.

In the TAR sharp size distortions exist for OLS and LWAD when u_t is iid $\bar{P}_{.75}$ or IGARCH, while LTTS dominates overall in size and power.

6.1.5 Inference: Wald Tests Finally, we simulate three AR(2) models with iid

Pareto errors and $\kappa \in \{1.5, 2.5\}$, and estimate AR(2) models by LTTS with benchmark $(\lambda_u, \lambda_y) = (.05, 1.0)$. We compute Wald statistics $\mathcal{W}_n = (R\hat{\phi}_n - q)'(R\hat{\mathcal{V}}_n^{-1}R')^{-1}(R\hat{\phi}_n - q)$ for three tests of AR(1) against AR(2), hence $R = [0, 0, 1]$ and $q = 0$. We use the covariance estimator $\hat{\mathcal{V}}_n^{-1}$ from (15) with a LTTS $\hat{\phi}_n$ plug-in, perform tests at the 5% level, and present results in Table 5. Empirical sizes are near the nominal level .05, and empirical powers are predominantly above 90%, and near 100% when the alternative is far from the null or n is large, as expected.

6.2 GARCH - QMTTL vs. QML, Log-LAD and QMWL

Finally, we draw $2n$ observations from a GARCH(1,1) $y_t = \sigma_t \epsilon_t$, $\sigma_t^2 = .3 + .3y_{t-1}^2 + .6\sigma_{t-1}^2$, with starting values $\sigma_1^2 = .3$, and retain the last $n \in \{100, 800\}$. The errors ϵ_t are iid symmetric Pareto $\bar{P}_{2.5}$, or iid $N(0,1)$, hence y_t has power-law tails with index $\kappa_y \in \{1.5, 4.1\}$. We simulate 1000 series $\{y_t\}_{t=1}^n$.

The QMTTL estimator solves $\text{argmin}_{\beta \in \mathcal{B}}\{1/n \sum_{t=2}^n l(\beta, z_t)\hat{I}_{n,t}^{(\epsilon)}(\beta)\}$ where $l(\beta, z_t) = \ln h_t(\beta) + y_t^2/h_t(\beta)$ with initial $h_1(\beta) = \beta_1$, and the parameter space is $\mathcal{B} = [0, 1]^3$. We use symmetric trimming with $k_n^{(\epsilon)} = [\lambda n / \ln(n)]$, where either $\lambda = .05$, or λ is selected by bootstrap-mse minimization over λ on $\Lambda_n(1)$, using $R_n = [n]$ sub-samples of size $n/2$, with replacement, and a Log-LAD plug-in.

See Figures 4 and 5 for plots of typical asymptotic covariance determinant $d_n(\lambda)$ and bootstrap-mse $||\hat{\mathcal{M}}_n(\lambda)||$. The error is iid $\bar{P}_{2.5}$, $n = 100$, and we use a Log-LAD plug-in. Notice $d_n(\lambda)$ ceases falling once $\lambda \geq .1$, where $||\hat{\mathcal{M}}_n(\lambda)||$ is roughly minimized.

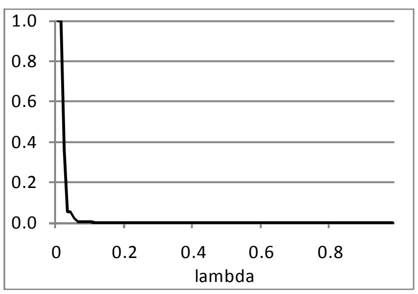

Figure 4: $d_n(\lambda)$ for GARCH

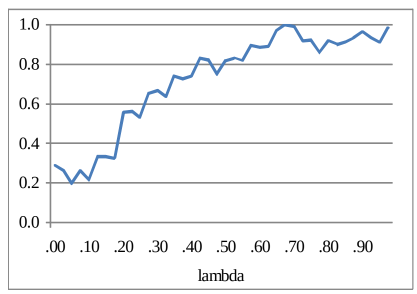

Figure 5: $||\text{mse}(\lambda)||$ for GARCH

We also use QML, Log-LAD $\text{argmin}_{\beta \in \mathcal{B}}\{\sum_{t=2}^n |\ln(y_t^2/h_t(\beta))|\}$ and QMWL $\text{argmin}_{\beta \in \mathcal{B}}\{\sum_{t=2}^n w_t l(\beta, z_t)\}$ with Ling's (2007) only suggested weight: $w_t = 1$ if $a_t = 0$ and $w_t = (y_{([.05n])}^{(a)}/a_t(R))^4$ if $a_t(R) \neq 0$, where $a_t(R) := \sum_{i=1}^R i^{-9}|y_{t-i}|I(|y_{t-i}| \geq y_{([.05n])}^{(a)})$. Ling (2007) requires $R = \infty$ but does not suggest how to compute w_t. We simply use $R = [n^{1/2}]$ with an iteration $a_2(1)$, $a_3(2)$, ..., $a_{R+1}(R)$, ..., $a_n(R)$[9].

Let $\hat{\beta}^{(j)}$ be the j^{th} iteration on $\hat{\beta}$, with initial value $\hat{\beta}^{(0)}$ a uniform draw on $[0, 1]$. We stop iterating when $||\hat{\beta}^{(j)} - \hat{\beta}^{(j-1)}|| \leq .01$. Estimation and t-test results are presented in Tables 6 and 7.

QMWL is biased and non-normal when $E[\epsilon_t^4] = \infty$, with smaller bias for larger n (recall QMWL is not robust to heavy tailed ϵ_t). QMWL leads to the lowest empirical power for t-tests, and slightly dominates QMTTL for thin tailed GARCH when $n = 800$. The latter likely arises due to the smoothness of Ling's weights, while QMTTL removes observations whether it is needed or not. Log-LAD performs poorly in small samples, and roughly the same as QMWL for larger n. In

[9] We obtain similar results if the first $R - 1 = [n^{1/2}] - 1$ observations are dropped, $\text{argmin}_{\beta \in \mathcal{B}}\{\sum_{t=R}^n w_t l(\beta, z_t)\}$, so that $a_t(R)$ is computed for each included t. If we simply fix $R = 1$ the results are similar, although QMWL performs less well in terms of mse and approximate normality.

particular, Log-LAD is very sharp only when $E[\epsilon_t^4] = \infty$ and $n = 800$, and is worst overall when $E[\epsilon_t^4] < \infty$.

QMTTL is resilient to heavy tails, and is closest to normal when tails are heavy. The bootstrap-mse method renders a sharp estimator that is roughly normal for either $n \in \{100, 800\}$, with range $\lambda \in (.05, .13)$ although $\lambda \in (.05, .08)$ was almost always optimal, while the fixed $\lambda = .05$ also worked well. If the error is Gaussian then all estimators perform somewhat poorly for small n, a well known shortcoming of QML for GARCH models (e.g. Straumann and Mikosch 2006, Ling 2007).

7. EMPIRICAL APPLICATION We now analyze linear and nonlinear AR and GARCH models for financial returns data. In order to draw comparisons, we study the same Hang Seng Index [HSI] stock market data Ling (2005) used to demonstrate his LWAD estimator. The period is June 3, 1996 to May 31, 1998 representing 491 daily observations, net of market closures[10]. Consult Ling (2005) for details on the HSI. We compute log-returns $y_t = \ln(x_t/x_{t-1})$ where x_t is the daily closing value on the HSI, and plot y_t in Figure 6.

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

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

where $w_{n,s,t}$ is a kernel function. We use a Bartlett kernel $w_{n,s,t} = (1 - |s - t|/\gamma_n)_+$ with bandwidth[11] $\gamma_n = n^{.225}$. By the mean-value-theorem the asymptotic 90% confidence band $\hat{\kappa}_{y,\tilde{k}_n} \pm 1.64\hat{v}_y\hat{\kappa}_{y,\tilde{k}_n}^2/\tilde{k}_n^{1/2}$, which we plot in Figure 7.

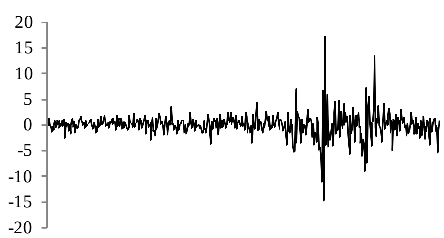

Figure 6: HSI Daily Log-Returns

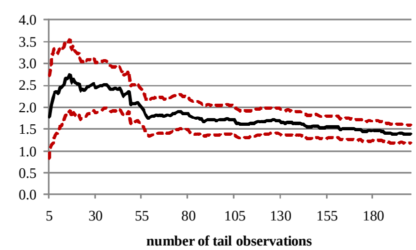

Figure 7: HSI Hill-Plot and Robust 90% Bands

Values of $\kappa_y \leq 2$ lie in the 90% intervals at every $\tilde{k}_n$ while $\kappa_y < 1$ lies in the interval only at $\tilde{k}_n = 5$: we never reject the one-sided hypothesis $\kappa_y \leq 2$ against $\kappa_y > 2$ at the 5% level, and we reject

[10] Our data were taken from finance.yahoo.com, which may be slightly different from Ling's data. Ling reports 497 observations.

[11] Simulation evidence not reported here suggests $\gamma_n \in \{n^{.20}, n^{.25}\}$ is optimal for a large variety of linear and nonlinear processes and sample size sizes n, including AR-GARCH and $50 \leq n \leq 100{,}000$. We simply use the midpoint $\gamma_n = n^{.225}$.

$\kappa_y \leq 1$ in favor of $\kappa_y > 1$ except at $\tilde{k}_n = 5$. The evidence uniformly suggests HSI daily returns have an infinite variance and finite mean, and that non-heavy tail robust methods are inappropriate.

7.1 AR MODEL ESTIMATION

Since a benchmark question is whether asset returns are white noise, we estimate AR(6), AR(8) and AR(12) models with intercepts by LTTS and compute Wald statistics for tests that all slopes are zero over $p \in \{6, 8, 12\}$, denoted $\mathcal{W}_{n,p}$. All AR models in this study include an intercept, and LTTS fractiles are $\{k_n^{(u)}, k_n^{(y)}\} = \{[\lambda_u n/\ln(n)], [\lambda_y \ln(n)]\}$. In Figures 8 and 9 we plot the asymptotic covariance determinant $d_n(\lambda)$ and bootstrap-mse $||\hat{\mathcal{M}}_n(\lambda)||$ for an AR(8), and use least squares to compute $\hat{\mathcal{V}}_n(\lambda)$. In this study the bootstrap is always performed from $R_n = n$ subsamples of size $[n/2]$ with replacement.

Similar to in our simulation study, $d_n(\lambda)$ flattens once $\lambda_u \geq .25$ and $\lambda_y \leq .5$, while the mse is minimized at $\lambda_u = .21$ and $\lambda_y = .10$. In the following we use the minimum bootstrap-mse λ for each model, where in all cases $\lambda_u \in [.15, .50]$ and $\lambda_y \in [.05, .50]$. We also use fixed values below for robustness checks. The statistics are $\mathcal{W}_{n,6} = 45.02$ (.000), $\mathcal{W}_{n,8} = 43.99$ (.000), and $\mathcal{W}_{n,12} = 52.58$ (.000) with p-values in parentheses, hence white noise is rejected.

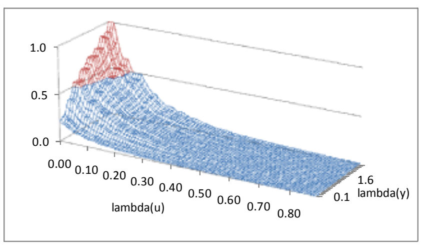

Figure 8: $d_n(\lambda)$ for AR(8) model of HSI

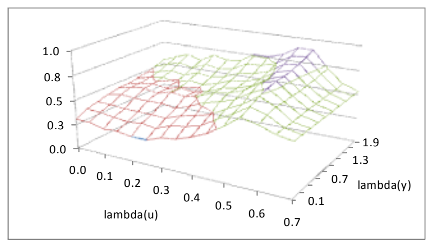

Figure 9: $||\text{mse}(\lambda)||$ for AR(8) model of HSI

We then estimate AR(p) models over $p = 1, ..., 12$ and test the residuals $\hat{u}_t = y_t - \hat{\phi}_n' x_t$ for serial dependence by computing $\mathcal{W}_{n,12}$ for $\hat{u}_t$. Each AR(p) model with $p \geq 3$ has orthogonal residuals, where AR(3) $y_t = \phi_0 + \sum_{i=1}^{3} \phi_i y_{t-i} + u_t$ results in a residual Wald statistic $\mathcal{W}_{n,12} = 15.99$ (.1914), while AR(2) $\mathcal{W}_{n,12} = 26.66$ (.009) and AR(1) $\mathcal{W}_{n,12} = 29.66$ (.003). Wald tests of AR(2) against AR(3) and AR(3) against AR(4) lead to the same conclusion: an AR(3) best describes the data:[12]

$$\begin{array}{rcl} \hat{y}_t & = & -.03 + .14 y_{t-1} - .05 y_{t-2} + .05 y_{t-3}. \\ & & \quad (.07) \quad (.04) \qquad (.04) \qquad (.04) \end{array}$$

The result is robust to a higher order specifications. An AR(7) with $(\lambda_u, \lambda_y) = (.25, 1.0)$, for example, results in

$$\begin{array}{rcl} \hat{y}_t & = & -.02 + .14 y_{t-1} - .05 y_{t-2} + .10 y_{t-3} + .03 y_{t-4} - .06 y_{t-5} - .00 y_{t-6} + .01 y_{t-7}. \\ & & \quad (.06) \quad (.04) \qquad (.04) \qquad (.05) \qquad (.05) \qquad (.05) \qquad (.05) \qquad (.05) \end{array}$$

Similarly, if we simply fix λ we obtain similar results. If $(\lambda_u, \lambda_y) = (.10, .75)$ the best model is again AR(3) $\hat{y}_t = -.07 + .12 y_{t-1} - .03 y_{t-2} + .08 y_{t-3}$ with standard errors (.07, .04, .04.05). If (λ_u, λ_y) $= (.05, 1.0)$ as in our simulation study, then again AR(3) with $\hat{y}_t = -.10 + .25_{t-1} - .02 y_{t-2} +$ $.11 y_{t-3}$ and $(.08, .05, .05, .04)$.

[12]Standard errors in parentheses are $\hat{\mathcal{V}}_{i,i,n}^{-1/2}$ from formula (15). The estimated AR roots of all models in this study lie outside the unit circle.

Finally, in the AR(3) model we test separately whether the first lag y_{t-1}, or second lag y_{t-2}, or both $\{y_{t-1}, y_{t-2}\}$ do not add any predictive power. The resulting Wald statistic values are 28.6 (.000), .165 (.685) and 28.9 (.000) suggesting the appropriate model is $y_t = \phi_2^0 y_{t-1} + \phi_4^0 y_{t-3} + u_t$. Ling's (2005) chosen model by similar Wald tests based on LWAD with $\lambda = .05$ is also AR(3), but with only the third lag $\hat{y}_t = .123 y_{t-3}$ and standard error .04.

7.2 THRESHOLD AR MODEL ESTIMATION

The HSI index reveals significant structure beyond a linear AR. Notice all models above suggest the true intercept is zero hence $E[y_t] = 0$ (recall we concluded $\kappa_y > 1$), but symmetry remains to be justified. Define $\mathcal{I}_t(z) := 2I(z_t > 0) - 1$ and $\widehat{\mathcal{I}}_n(z) := 1/n \sum_{t=1}^n \mathcal{I}_t(z)$, where initial periods are truncated if z_t are residuals. A test of symmetry against positive skew is a test that $E[\mathcal{I}_t(z)] = P(z_t > 0) - P(z_t \leq 0) = 0$ against $E[\mathcal{I}_t] > 0$. We use a t-ratio $n^{1/2}\widehat{\mathcal{I}}_n(z)/\hat{v}_n(z)$ with kernel variance estimator $\hat{v}_n^2(z) = 1/n \sum_{s,t=1}^n \mathcal{K}((s-t)/b_n)\mathcal{I}_s(z)\mathcal{I}_t(z)$, Barlett kernel $\mathcal{K}$ and bandwidth $b_n = n^{.25}$. Under stationary β-mixing and symmetry $n^{1/2}\widehat{\mathcal{I}}_n(z)/\hat{v}_n \xrightarrow{d} N(0,1)$ by a mixing central limit theorem, and HAC consistency (Newey and West 1987)[13].

We find the HSI $\widehat{\mathcal{I}}_n(y) = .132$ (.07), and the AR(3) residual $\widehat{\mathcal{I}}_n(u) = .06$ (.13), where standard errors $\hat{v}_n/n^{1/2}$ are in parentheses. Thus, while a linear AR may reasonably fit y_t since the errors appear to be orthogonal and symmetric, a better model may be nonlinear AR with symmetric error. We therefore fit a two regime TAR(p) with zero threshold $y_t = \zeta_1'\tilde{x}_t I(y_{t-1} < 0) + \zeta_2'\tilde{x}_t I(y_{t-1} \geq 0) + u_t = \phi' x_t + u_t$ where $\zeta_i'\tilde{x}_t := \zeta_{i,0} + \sum_{j=1}^p \zeta_{i,j} y_{t-i}$.

We estimate TAR(p) models for orders $p = 1, ..., 12$ using fractiles $\{k_n^{(u)}, k_n^{(y)}\} = \{[\lambda_u n/\ln(n)], [\lambda_y \ln(n)]\}$ and minimum bootstrap-mse λ, and again find $p = 3$ is the smallest order at which the residuals are orthogonal, where TAR(3) $\mathcal{W}_{n,12} = 13.11$ (.36) and TAR(2) $\mathcal{W}_{n,12} = 25.76$ (.01). The results for the left-tail $\hat{\zeta}_1'\tilde{x}_t$ and right-tail $\hat{\zeta}_2'\tilde{x}_t$ regimes follow:

$$\underset{}{\hat{\zeta}_1' x_t} = \underset{(.09)}{-.03} + \underset{(.05)}{.15} y_{t-1} + \underset{(.05)}{.17} y_{t-2} + \underset{(.06)}{.01} y_{t-3} \qquad \hat{\zeta}_2' x_t = \underset{(.09)}{.01} - \underset{(.06)}{.04} y_{t-1} - \underset{(.06)}{.14} y_{t-2} + \underset{(.05)}{.13} y_{t-3}$$

Clearly the two regimes are different, which alone has important implications for impulse response analysis, a topic almost entirely ignored in the extreme value theory literature (see Hill 2006). We reject regime equivalence $\zeta_1 = \zeta_2$ at the 1% level based on a Wald test, and evidence for error symmetry is even stronger: $\widehat{\mathcal{I}}_n(\hat{u}) = .018$ (.34), suggesting omitted nonlinearity in the AR(3) is simply appearing in the AR error.

7.3 TAR-GARCH MODEL ESTIMATION

Visually the Figure 6 plot of HSI returns suggests volatility clustering, although this is hardly rigorous. Our last task is therefore to fit a TAR(3)-GARCH(1,1) model to the log-returns, where $y_t = \zeta_1'\tilde{x}_t I(y_{t-1} < 0) + \zeta_2'\tilde{x}_t I(y_{t-1} \geq 0) + u_t = \phi' x_t$, $\tilde{x}_t = [1, y_{t-1}, y_{t-2}, y_{t-3}]'$, $u_t = \sigma_t \epsilon_t$, and $\sigma_t^2 = \beta_1 + \beta_2 u_{t-1}^2 + \beta_3 \sigma_{t-1}^2$. After accounting for $k + 1 = 4$ initial periods, the criterion is $\sum_{t=5}^n (\ln h_t(\theta) + (y_t - \phi' x_t)^2/h_t(\theta)) \times \hat{I}_{n,t}^{(\epsilon)}(\theta)\hat{I}_{n,t}^{(y)}(\theta)$ where $\hat{I}_{n,t}^{(y)}(\theta) = \sqcap_{i=1}^4 \hat{I}_{i,n,t}^{(y)}(\theta)$, with the fractiles $\{k_n^{(\epsilon)}, k_n^{(y)}\} = \{[\lambda_\epsilon n/\ln(n)], [\lambda_y \ln(n)]\}$ and the minimum bootstrap-mse λ. As a goodness-of-fit test we fit the residuals $\hat{\epsilon}_t^2 - 1$ to an AR(12) model and compute $\mathcal{W}_{n,12}$, test for symmetry, and compute the tail index.

The estimated GARCH component is $\hat{\sigma}_t^2 = .12 + .10_{t-1}^2 + .82_{t-1}^2$ with standard errors .03, .06, and .08. The latter are computed from the asymptotic variance block $n(1/n \sum_{t=5}^n \widehat{\mathfrak{s}}_{\beta,t}\widehat{\mathfrak{s}}_{\beta,t}')$

[13] All symmetry test results in this paper are robust to a range of bandwidths $b_n = n^\zeta$ for $\zeta \in (.15, .45)$.

$\times$ $(1/n \sum_{t=5}^{n} \epsilon_t^4(\hat{\theta}_n)\hat{I}_{n,t}^{(\epsilon)}(\hat{\theta}_n) - 1)^{-1}$ where we write $\widehat{\mathfrak{s}}_{\beta,t} := h_t^{-1}(\hat{\theta}_n)(\partial/\partial\beta)h_t(\beta)|_{\hat{\theta}_n}$, starting with $h_4(\theta) = \beta_1$ and $h_4^{\theta} = [1,0,0]'$. The white noise statistic for the residuals $\epsilon_t^2(\hat{\theta}_n) - 1$ is $\mathcal{W}_{n,12} = 15.3$ (.225), and the symmetry statistic $\widehat{\mathcal{I}}_n(\epsilon_t) = .002$ (.48) points overwhelmingly to a symmetric GARCH error. Finally, we reject a one-sided t-test for IGARCH $\beta_2 + \beta_3 = 1$ in favor of $\beta_2 + \beta_3 < 1$: evidence for an infinite variance in y_t therefore suggests the GARCH error ϵ_t itself cannot be Gaussian, and has a power-law tail (cf. Hall and Yao 2003, Cline 2007). A Hill-plot of the tail index of $|\hat{\epsilon}_t|$ reveals $\hat{\kappa}_{\epsilon,\tilde{k}_n}$ hovers between 3 and 5 for all $\tilde{k}_n \in \{5,...,200\}$, every $\hat{\kappa}_{\epsilon,\tilde{k}_n} > 2$, nearly every $\hat{\kappa}_{\epsilon,\tilde{k}_n} > 3$, $\hat{\kappa}_{\epsilon,\tilde{k}_n} < 4$ in 84% of the bands, and values $\kappa_\epsilon < 4$ lie in each 90% band. See Figure 10 in Appendix C. Evidence for $E[\epsilon_t^2] < \infty$ is in accord with our assumptions, and the possibility $\kappa_\epsilon < 4$ rules out QML and QMWL for standard inference.

8. CONCLUSION We develop a general class of tail-trimmed M-estimators for nonlinear AR-GARCH(1,1), while added GARCH lags, additional regressors, and bounded forms of non-stationarity are straightforward extensions. Specific estimators easily fall out, and are consistent for the true parameter and asymptotically normal under regularity conditions. If the pure AR or GARCH error is iid we can always choose the fractile in a simple way to obtain nearly the highest possible convergence rate for M-estimators of stationary data. We show by Monte Carlo experiment that LTTS and QMTTL dominate existing estimators based on approximate normality and therefore inference. Asymptotic covariance and bootstrap-mse estimates give a clear picture of how much to trim, in particular the minimum bootstrap-mse values lead to sharp estimates in all cases. Nevertheless, simply choosing small fractiles leads to similar results, suggesting the analyst has some freedom with how much to trim.

The next stage must involve a complete theoretical development of fractile selection, and model specification tests including score tests, tests of GARCH effects, and moment conditions tests, each robust to heavy tails. Additional possibilities for solving the fractile challenge may be by indirect inference, or testing $E[m_{n,t}^*(\theta^0)] = 0$ with a consistent plug-in for θ^0 (e.g. OLS, Log-LAD), but these must be left for future development.

APPENDIX A: Assumptions and Stationarity

We now state all assumptions. Drop θ^0 throughout (e.g. $m_{n,t}^* = m_{n,t}^*(\theta^0)$). First, loss smoothness, response smoothness and bounds, and stationarity ensure stationary solutions $\{h_t^*(\theta), h_{i,t}^{*\theta}(\theta), h_{i,t}^{*\theta,\theta}(\theta)\}$ exist, a requirement if we estimate volatility parameters β.

We require compact notation. We say a matrix function $\xi(u,h,\theta) : \mathbb{R} \times \mathbb{R} \times \Theta \in \mathbb{R}^{a\times b}$, $a,b \geq 1$, is Lipschitz in h if $||\xi(u,h_1,\theta) - \xi(u,h_2,\theta)|| \leq K|h_1 - h_2|$ $\forall h_1,h_2 \in \mathbb{R}$ and $u,\theta \in \mathbb{R}\times\Theta$. Now let $a,b \in \{u,h,\theta\}$ be indices. For $g(u,h,\theta)$ and $f(x,\phi)$ let g_a and f_ϕ denote first partial derivatives, and $g_{a,b}$ and $f_{\phi,\phi}$ second derivatives.

Assumption L (*loss*).

i. $l(z_t,\theta)$ *is continuous and twice continuously differentiable in* θ*;*

ii. If volatility parameters β *are estimated then:* $E[\sup_{\theta\in\Theta} |(\partial/\partial\theta)^i l(z_t,\theta)|^\iota] < \infty$ *for tiny* $\iota > 0$ *and* $i = 0,1,2$*;* $l(z_t,\theta)$ *can be written* $l(z_t,\phi,h_t(\theta))$*; and* $l(z,\phi,h)$ *is thrice differentiable in* h *with* $E[\sup_{\theta\in\Theta} |(\partial/\partial h)^3 l(z,\theta,h)|_{z_t,\phi,h_t(\theta)}|^\iota] < \infty$.

Remark: Assumption L is obviously satisfied for QML loss.

Assumption RS (*response smoothness*). *i.* $f(\cdot,\phi)$ *is twice continuously differentiable on* Φ*; ii.* $g(u,h,\theta)$ *is twice continuously differentiable on* $\mathbb{R}\times\mathbb{R}_+\times\Theta$.

Drop response arguments for compactness (e.g. $f = f(x, \phi)$). The following bounds apply at least to linear and threshold-type models (see MS 2009).

Assumption RB (*response bounds*).

i. $|f|$, $||f_\phi||$, *and* $||f_{\phi,\phi}||$ *are bounded by* $K(1 + \sum_{i=1}^{k} |x_i|)$;

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

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

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

Assumption STM (*stationarity and moments*). *Each* $w_t \in \{y_t, u_t, \sigma_t^2\}$ *is stationary and ergodic, and* $E|w_t|^\iota < \infty$ *for tiny* $\iota > 0$.

A stationary solution follows[14]. Let $a_t(\theta) \in \{h_t(\theta), h_{i,t}^{\theta}(\theta), h_{i,t}^{\theta,\theta}(\theta)\}$ and $a_t^*(\theta) \in \{h_t^*(\theta), h_{i,t}^{*\theta}(\theta), h_{i,t}^{*\theta,\theta}(\theta)\}$ be arbitrary, let $I_{n,t}(\theta)$ denote $I_{n,t}^*(\theta)$ evaluated with $m_t(\theta)$ and $G_t(\theta)$, and let $w_t(\theta) \in \{m_{i,t}(\theta), G_{i,j,t}(\theta)\}$ and $w_t^*(\theta) \in \{m_{i,t}^*(\theta), G_{i,j,t}^*(\theta)\}$.

PROPOSITION A.1 (*stationary solution*). *Under Assumptions RB, RS and STM:*

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

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

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

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

PROOF. Claims (i)-(iii) follow from Propositions 1 and 2 of MS (2009) since their Assumptions DGP, C1-C4 and N1-N3 hold under our Assumptions RB, RS and STM.

We will prove (iv) for $m_{i,t}(\theta)$, the proof for $G_{i,j,t}(\theta)$ being similar given thrice differentiability of the loss under Assumption L.ii. Since we implicitly estimate β, by Assumption L.ii there exist functions $\check{l}_t(\phi, h)$ and $\check{m}_t(\phi, h)$ that satisfy $\check{l}_t(\phi, h_t(\theta)) = l(z_t, \theta)$ and $\check{m}_t(\phi, h_t(\theta)) = m_t(\theta)$, where $\check{m}_t(\phi, h)$ is differentiable in h. Hence, by the mean-value-theorem, the Assumption L.ii envelope properties, the Cauchy-Schwartz inequality, and (iii) it follows as claimed $E[\sup_{\theta \in \Theta} |m_t^*(\theta) - m_t(\theta)|^\iota] \leq KE[\sup_{\theta \in \Theta} |h_t(\theta) - h_t(\theta)|^{\iota/2}] = o(\rho^t)$. $\mathcal{QED}$.

Let $\kappa_w(\theta) \in \{\kappa_{m_i}(\theta), \kappa_{G_{i,j}}(\theta)\}$ be the moment suprema of $w_t(\theta) \in \{m_{i,t}^*(\theta), G_{i,j,t}^*(\theta)\}$: $\kappa_w(\theta) := \sup_{\alpha > 0}\{E|w_t(\theta)|^\alpha < \infty\}$. Notice $\kappa_w(\theta) = \infty$ is possible (e.g. exponential tail decay, or bounded support). Let $\Theta_{2,w} \subseteq \Theta$ denote the set of all θ such that $\kappa_w(\theta) \leq 2$.

Assumption D (*distribution*).

i. Each y_t, $h_t^*(\theta)$, $m_{i,t}^*(\theta)$, and $G_{i,j,t}^*(\theta)$ *have absolutely continuous finite dimensional distributions, and each* $w_t(\theta) \in \{m_{i,t}^*(\theta), G_{i,j,t}^*(\theta)\}$ *has a uniformly bounded density* $\sup_{\theta \in \Theta} \sup_{a \in \mathbb{R}}\{(\partial/\partial\theta) P(w_t(\theta) \leq a)\} < \infty$.

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

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

$$\text{(17)} \qquad \sup_{\theta\in\Theta_{2,w}}\left\{\left|c^{\kappa_w(\theta)}P\left(|w_t(\theta)| > c\right) - d_w(\theta)\right|\right\} \to 0 \text{ as } c \to \infty \text{ where } \inf_{\theta\in\Theta}\{d_w(\theta)\} > 0.$$

Remark: Distribution continuity of the loss $l^*(z_t, \theta)$ permits a unique TTME solution with probability one, and for $w_t(\theta) \in \{m^*_{i,t}(\theta), G^*_{i,j,t}(\theta)\}$ ensures the existence of thresholds $\{\mathcal{L}^{(w)}_n(\theta), \mathcal{U}^{(w)}_n(\theta)\}$ for all θ and any fractiles $\{k_{r,n}\}$. Uniform boundedness of the densities of $w_t(\theta)$ simplifies the verification that $\{I^{(m^*)}_{i,n,t}(\theta), I^{(G^*)}_{i,j,n,t}(\theta)\}$ satisfy a UCLT. We assume heavy tailed $m^*_{i,t}(\theta)$ and $G^*_{i,j,t}(\theta)$ have power-law tails to simplify characterizing $E[m^*_{n,t}m^{*\prime}_{n,t}]$ and $E[G^*_{n,t}]$ and therefore the scale $\mathcal{V}_n$.

Assumption F (*fractiles*). *In general* $\min_{j=1,2}\{k^{(m)}_{j,i,n}\}/L(n) \to \infty$ *for some s.v.* $L(n) \to \infty$. *Further, if* $\kappa_{m_i} < 1$ *then* $\min_{j=1,2}\{k^{(m)}_{j,i,n}\}/n^{2(1-\kappa_{m_i})/(2-\kappa_{m_i})} \to \infty$.

Remark: The property $k^{(m)}_{j,i,n}/L(n) \to \infty$ merely sets a minimal amount of trimming. Consistency, however, requires a uniform law of large numbers for the trimmed equations $m^*_{i,t}(\theta)I^{(m)}_{i,n,t}(\theta)$, even if $m^*_{i,t}(\theta)$ does not have a mean. The far more profound lower bound $k^{(m)}_{j,i,n}/n^{2(1-\kappa_{m_i})/(2-\kappa_{m_i})} \to \infty$ ensures sufficiently many large equations are trimmed for such a limit theory when $\kappa_{m_i} < 1$, and as $\kappa_{m_i} \searrow 0$ we must approach maximal trimming $k^{(m)}_{j,i,n} \nearrow n$. If the mean of $m_{i,t}$ is finite or hairline infinite $\kappa_{m_i} \geq 1$ then there are no further restrictions.

In Appendices B and C we show $\hat{\theta}_n$ obtains the expansion $\mathcal{V}^{1/2}_n(\hat{\theta}_n - \theta^0) \sim n^{-1/2}\mathcal{S}^{-1/2}_n\sum_{t=1}^n m^*_{n,t}$, hence $n^{1/2}\mathcal{S}^{-1/2}_n E[m^*_{n,t}] \to 0$ must hold for asymptotic unbiasedness.

Assumption I (*identification*). $E[m^*_{n,t}(\theta)] \to 0$ if and only if $\theta = \theta^0$, *a unique interior point of compact* $\Theta \subset \mathbb{R}^q$, *in particular* $||n^{1/2}\mathcal{S}^{-1/2}_n E[m^*_{n,t}]|| \to 0$.

Conventional M-estimator asymptotics follows from stationarity, ergodicity and higher moment bounds since then a martingale difference central limit theorem applies (e.g. Straumann and Mikosch 2006, MS 2009). We do not require m_t to be a martingale difference, and even if it were under *tail trimming* $m^*_{n,t}$ is non-stationary and may not be a martingale difference. Further, we require a ULLN and UCLT for non-martingale difference and possibly non-uniformly integrable functions simply to show $\hat{m}_{n,t}(\theta) := m_t(\theta)\hat{I}_{n,t}$ is similar to $m^*_{n,t}(\theta) = [m^*_{i,t}(\theta)I^{(m^*)}_{i,n,t}]^q_{i=1}$ uniformly on θ, which we must have just to prove consistency. Thus, ergodicity and martingale difference arguments do not suffice here. But geometric β-mixing does suffice.

Assumption MX (*mixing*). *Let* $\{y_t\}$ *be geometrically* β*-mixing. If volatility parameters* β *are estimated then* $\{h^*_t(\theta)\}$ *for each* $\theta \in \Theta$ *is geometrically* β*-mixing.*

Remark: Although Proposition A.1 shows $h^*_t(\theta)$ is stationary and ergodic, demonstrating it is mixing on Θ is to date extremely challenging. If we only estimate AR parameters ϕ then Assumption MX can be replaced by sufficient conditions for geometric ergodicity, like continuity of the error distribution and response bounds RB.i (An and Huang 1996, Cline 2007, Meitz and Saikkonen 2008).

Consistency of $\hat{\theta}_n$ and the sample Jacobian require moment smoothness.

Assumption MS (*moment smoothness*). $\liminf_{n\to\infty}\inf_{||\theta-\theta^0||>\delta}\{||E[m^*_{n,t}(\theta)]||/\sup_{\theta\in\Theta}||E[m^*_{n,t}(\theta)]||\} > 0$ *for tiny* $\delta > 0$.

Remark: The property bounds $E[m^*_{n,t}(\theta)]$ from 0 for θ near θ^0, while the scale $\sup_{\theta\in\Theta}||E[m^*_{n,t}(\theta)]||$ is irrelevant if $\sup_{\theta\in\Theta}||E[m_t(\theta)]|| < \infty$. Otherwise the scale ensures a uniform law of large numbers applies to $m^*_{n,t}(\theta)$, which expedites our proof of consistency.

Finally, we must rule out degenerate cases due to over trimming asymptotically.

Assumption N (*non-degeneracy*). *Moment, covariance and Jacobian matrices* $\mathcal{A}_n(\theta) \in \{E|m^*_{i,n,t}(\theta)|^p, \mathcal{S}_n(\theta), \mathcal{G}_n(\theta)\}$ *for any* $p > 0$ *satisfy* $\liminf_{n\to\infty}\inf_\theta\{||\mathcal{A}_n(\theta)||\} > 0$. *Further* $\liminf_{n\to\infty}\inf_\theta\{\lambda_{\min}(\mathcal{S}_n(\theta))\} > 0$ *and* $||E[m^*_{n,t}m^{*\prime}_{n,t}]\mathcal{S}_n^{-1}|| = O(1)$.

Remark: We prove in Lemma B.2 below $||\mathcal{S}_n(E[m^*_{n,t}m^{*\prime}_{n,t}])^{-1}|| = O(\ln(n))$. Conditions similar to $||E[m^*_{n,t}m^{*\prime}_{n,t}]\mathcal{S}_n^{-1}|| = O(1)$ are standard in the central limit theory literature for dependent processes, ruling out degeneracy due to negative correlation since $\mathcal{S}_n \sim E[m^*_{n,t}m^{*\prime}_{n,t}] + 2\sum_{i=1}^{n-1}(1 - i/n)E[m^*_{n,1}m^{*\prime}_{n,i+1}]$. See Dehling et al (1986) for references.

APPENDIX B: Proofs of Main Results

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

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

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

$$\hat{m}^*_{n,t}(\theta) := m^*_t(\theta)\hat{I}^*_{n,t}(\theta), \;\; \hat{m}^*_n(\theta) := \frac{1}{n}\sum_{t=1}^n \hat{m}^*_{n,t}(\theta)$$

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

We also exploit various covariance, Jacobian and envelope equations, some of which are defined in Sections 2 and 5:

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

$$G^*_{n,t}(\theta) := G^*_t(\theta)I^*_{n,t}(\theta) \;\text{ and }\; \mathcal{G}_n(\theta) := E\left[G^*_{n,t}(\theta)\right] \;\text{ and }\; \check{\mathcal{G}}^*_n(\theta) := \frac{\partial}{\partial\theta}E\left[m^*_{n,t}(\theta)\right]$$

$$\tilde{\mathcal{G}}^*_n(\theta) := \frac{1}{n}\sum_{t=1}^n G^*_{n,t}(\theta) \;\text{ and }\; \mathfrak{e}_n := \sup_{\theta\in\Theta}\left\|E\left[m^*_{n,t}(\theta)\right]\right\|.$$

The following result implies the TTME $\hat{\theta}_n$ and its infeasible counterpart $\hat{\theta}^*_n$ are asymptotically equivalent, hence we need only consider $\hat{\theta}^*_n$, $\hat{m}^*_{n,t}(\theta)$, $\widehat{\mathcal{G}}^*_n(\theta)$ and $\widehat{\mathcal{S}}^*_n(\theta)$ in all that follows. We exploit supporting Lemmas B.2-B.7 detailed below.

PROPOSITION B.1 (*infeasible estimator*). *Under Assumptions D, F, I, L, MS, MX, N,* *RB, RS, and STM* $\mathcal{V}_n^{1/2}(\hat{\theta}_n^* - \hat{\theta}_n) \xrightarrow{p} 0$, *where* $||\mathcal{V}_n|| \to \infty$.

PROOF. The following argument borrows from MS's (2009) proof of their Lemma D.6. By the proof of Theorem 2.1 the first order conditions are $\sum_{t=1}^n \hat{m}_{n,t}^*(\hat{\theta}_n^*) = \sum_{t=1}^n \hat{m}_{n,t}(\hat{\theta}_n) = 0$ *a.s.* Therefore, in lieu of consistency of the infeasible estimator $\hat{\theta}_n^* \xrightarrow{p} \theta^0$ by the proof of Theorem 2.1, the Lemma B.5 asymptotic expansion and Lemma B.7 Jacobian consistency, it follows

$$(18)\qquad \frac{1}{n^{1/2}}\mathcal{S}_n^{-1/2}\sum_{t=1}^n\left\{\hat{m}_{n,t}(\hat{\theta}_n) - \hat{m}_{n,t}^*(\hat{\theta}_n)\right\} = \frac{1}{n^{1/2}}\mathcal{S}_n^{-1/2}\sum_{t=1}^n\left\{\hat{m}_{n,t}^*(\hat{\theta}_n^*) - \hat{m}_{n,t}^*(\hat{\theta}_n)\right\} = n^{1/2}\mathcal{S}_n^{-1/2}\mathcal{G}_n \times \left(\hat{\theta}_n^* - \hat{\theta}_n\right).$$

Jacobian consistency extends to $1/n\sum_{t=1}^n G_t(\hat{\theta}_n)\hat{I}_{n,t} = \mathcal{G}_n(1 + o_p(1))$ by invoking Proposition A.1 and replicating the proof of Lemma B.7. Similarly, asymptotic approximation Lemma B.3.a and expansion Lemma B.5 extend to $\hat{m}_{n,t}$ and $m_{n,t}$. Therefore, by Jacobian consistency

$$(19)\frac{1}{n^{1/2}}\mathcal{S}_n^{-1/2}\sum_{t=1}^n\left\{\hat{m}_{n,t}(\hat{\theta}_n) - \hat{m}_{n,t}^*(\hat{\theta}_n)\right\}$$
$$= \frac{1}{n^{1/2}}\mathcal{S}_n^{-1/2}\sum_{t=1}^n m_{n,t}\left(1 + o_p\left(1\right)\right) + n^{1/2}\mathcal{S}_n^{-1/2}\mathcal{G}_n \times \left(\hat{\theta}_n - \theta^0\right) - \frac{1}{n^{1/2}}\mathcal{S}_n^{-1/2}\sum_{t=1}^n \hat{m}_{n,t}^*(\hat{\theta}_n)$$
$$= \frac{1}{n^{1/2}}\mathcal{S}_n^{-1/2}\sum_{t=1}^n\left\{m_{n,t} - m_{n,t}^*\right\}\left(1 + o_p\left(1\right)\right) + o_p\left(1\right),$$

Combine (18) and (19) to obtain $n^{1/2}\mathcal{S}_n^{-1/2}\mathcal{G}_n(\hat{\theta}_n^* - \hat{\theta}_n) = n^{-1/2}\mathcal{S}_n^{-1/2}\sum_{t=1}^n\{m_{n,t} - m_{n,t}^*\}(1 + o_p(1)) + o_p(1)$. By Loève's inequality, non-degeneracy $\liminf_{n\geq N}||\mathcal{S}_n|| > 0$ for some $N > 0$, and Proposition A.1.iv, it follows for tiny $\iota > 0$, $\rho \in (0, 1)$, and sufficiently large n and K

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

Therefore $\mathcal{V}_n^{1/2}(\hat{\theta}_n^* - \hat{\theta}_n) = o_p(1)$ by Chebyshev's inequality and the construction $\mathcal{V}_n = n\mathcal{G}_n\mathcal{S}_n^{-1}\mathcal{G}_n$. $\mathcal{QED}$.

The proofs of the main results require supporting Lemmas B.2-B.7. We state them when required, where Assumptions D, F, MS, MX, N, RB, RS and STM implicitly hold. Proofs are presented in Appendix C.

The proof of consistency Theorem 2.1 requires variance bounds, asymptotic bounds on $\hat{m}_{n,t}^*(\theta) - m_{n,t}^*(\theta)$, and laws of large numbers.

LEMMA B.2 (*variance and threshold bound*). *a.* $||\mathcal{S}_n(\theta)|| \leq K\ln(n)||\Sigma_n(\theta)|| = o(n/\ln(n))$; *b.* $\Sigma_n(\theta) = o(n||E\left[m_{n,t}^*(\theta)\right]||^2)$ *and* $\sup_{\theta\in\Theta}||\Sigma_n(\theta)|| = o(n\sup_{\theta\in\Theta}||E[m_{n,t}^*(\theta)]||^2)$.

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

LEMMA B.4 (*LLN and ULLN*). *a.* $1/n\sum_{t=1}^n m_{n,t}^*(\theta^0) = o_p(1)$; *b.* $\sup_\theta\{||m_n^*(\theta) - E[m_{n,t}^*(\theta)]||\} = o_p(\mathfrak{e}_n)$.

The proof of asymptotic normality Theorem 2.2 requires an asymptotic Taylor expansion for the trimmed equations, a central limit theorem, and Jacobian consistency. Recall $\tilde{\mathcal{G}}_n^*(\theta) := 1/n \sum_{t=1}^n G_t^*(\theta) I_{n,t}^*(\theta)$ and $\widehat{\mathcal{G}}_n^*(\theta) := 1/n \sum_{t=1}^n G_t(\theta) \hat{I}_{n,t}(\theta)$.

LEMMA B.5 (*asymptotic expansion*). *Choose* $\theta, \tilde{\theta} \in \Theta$ *and let* $\delta > 0$ be arbitrarily large and finite. *For* $o_p(1)$ *not a function of* θ*: a.* $m_n^*(\theta) = m_n^*(\tilde{\theta}) + \tilde{\mathcal{G}}_n^*(\theta) \times (\theta - \tilde{\theta}) + n^{-\delta} \times ||\theta - \tilde{\theta}||^{1/\iota} \times o_p(1)$*; and b.* $\hat{m}_n^*(\theta) = \hat{m}_n^*(\tilde{\theta}) + \widehat{\mathcal{G}}_n^*(\theta) \times (\theta - \tilde{\theta}) + n^{-\delta} \times ||\theta - \tilde{\theta}||^{1/\iota} \times o_p(1)$.

LEMMA B.6 (*CLT*). $n^{-1/2} \mathcal{S}_n^{-1/2} \sum_{t=1}^n m_{n,t}^*(\theta^0) \xrightarrow{d} N(0, I_q)$.

LEMMA B.7 (*Jacobian consistency*). *a.* $\widehat{\mathcal{G}}_n^*(\hat{\theta}_n^*) = \mathcal{G}_n(1 + o_p(1))$*; b.* $\check{\mathcal{G}}_n^* = \mathcal{G}_n \times (1 + o(1))$.

We are now ready to prove Theorems 2.1 and 2.2.

PROOF OF THEOREM 2.1. Define $\mathcal{M}_n^*(\theta) := E[m_{n,t}^*(\theta)]$ and $\mathfrak{e}_n := \sup_{\theta \in \Theta} ||\mathcal{M}_n^*(\theta)||$. We establish a uniform approximation property (**??**) below, then prove the infeasible $\hat{\theta}_n^* \xrightarrow{p} \theta^0$ by an argument in Pakes in Pollard (1989), and then prove $\hat{\theta}_n \xrightarrow{p} \theta^0$.

Step 1 (approximation): Combine $\sup_{\theta \in \Theta} ||\hat{m}_n^*(\theta) - m_n^*(\theta)|| = o_p(\mathfrak{e}_n)$ by Lemma B.3.b and $\sup_{\theta \in \Theta} ||m_n^*(\theta) - \mathcal{M}_n^*(\theta)|| = o_p(\mathfrak{e}_n)$ by ULLN Lemma B.4.b to deduce

(20) $$\sup_{\theta \in \Theta} \|\hat{m}_n^*(\theta) - \mathcal{M}_n^*(\theta)\| \leq \sup_{\theta \in \Theta} \|\hat{m}_n^*(\theta) - m_n^*(\theta)\| + \sup_{\theta \in \Theta} \|m_n^*(\theta) - \mathcal{M}_n^*(\theta)\| = o_p(\mathfrak{e}_n).$$

Step 2 ($\hat{\theta}_n^* \xrightarrow{p} \theta^0$): Note $\epsilon(\delta) := \liminf_{n \to \infty} \inf_{\theta \in \Theta, ||\theta - \theta^0|| > \delta} \{\mathfrak{e}_n^{-1} \times ||\mathcal{M}_n^*(\theta)||\} > 0$ for arbitrarily small $\delta > 0$ by moment smoothness Assumption MS. Since $P(||\hat{\theta}_n^* - \theta^0|| > \delta) \leq P(\mathfrak{e}_n^{-1} ||\mathcal{M}_n^*(\hat{\theta}_n^*)|| > \epsilon(\delta))$, it suffices to show $||\mathcal{M}_n^*(\hat{\theta}_n^*)|| = o_p(\mathfrak{e}_n)$ in order to prove $\hat{\theta}_n^* \xrightarrow{p} \theta^0$.

Use Minkowski's inequality and uniform approximation (**??**) to deduce

$$\left\|\mathcal{M}_n^*(\hat{\theta}_n^*)\right\| \leq \left\|\hat{m}_n^*(\hat{\theta}_n^*)\right\| + \left\|\hat{m}_n^*(\hat{\theta}_n^*) - \mathcal{M}_n^*(\hat{\theta}_n^*)\right\| \leq \left\|\hat{m}_n^*(\hat{\theta}_n^*)\right\| + o_p\left(\mathfrak{e}_n\right).$$

The loss has a smooth distribution and is differentiable under Assumptions D.i and L. Therefore TTME criterion $\hat{Q}_n^*(\theta) := 1/n \sum_{t=1}^n l^*(z_t, \theta) \hat{I}_{n,t}^*(\theta)$ is differentiable at $\hat{\theta}_n^*$ with probability one, hence up to a scalar constant $(\partial/\partial\theta) \hat{Q}_n^*(\theta)|_{\hat{\theta}_n^*} = \hat{m}_n^*(\hat{\theta}_n^*)$ *a.s.* (Čižek 2008: Lemma 2.1). In particular, by $\hat{\theta}_n^*$ a minimum $\hat{Q}_n^*(\hat{\theta}_n^*) \leq \hat{Q}_n^*(\theta)$ $\forall \theta \in \Theta$ it follows $0 = ||\hat{m}_n^*(\hat{\theta}_n^*)|| \leq K ||\hat{m}_n^*(\theta^0)||$.

Finally, $||\hat{m}_n^*(\theta^0)|| \leq K ||m_n^*(\theta^0)|| + o_p(||\mathcal{S}_n||^{1/2}/n^{1/2}) = o_p(1)$ by approximation Lemma B.3.a, the Lemma B.4.a LLN, and covariance bound Lemma B.2.a. This proves $||\mathcal{M}_n^*(\hat{\theta}_n^*)|| = o_p(1) + o_p(\mathfrak{e}_n)$. Since Assumption N ensures $\liminf_{n \to \infty} \{\mathfrak{e}_n\} > 0$ we have shown $||\mathcal{M}_n^*(\hat{\theta}_n^*)|| = o_p(\mathfrak{e}_n)$ as required.

Step 3 ($\hat{\theta}_n \xrightarrow{p} \theta^0$): The proof of $\hat{\theta}_n \xrightarrow{p} \theta^0$ follows from Step 2 and Proposition A.1: see MS (2009: Theorem 1). $\mathcal{QED}$.

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

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

Consistency $\hat{\theta}_n^* \xrightarrow{p} \theta^0$ ensures $\widehat{\mathcal{G}}_n^*(\hat{\theta}_n^*) = \mathcal{G}_n(1 + o_p(1))$ by Lemma B.7.a. Multiply both sides of (**??**) by $n^{1/2}\mathcal{S}_n^{-1/2}$, rearrange terms and use $\mathcal{V}_n = n\mathcal{G}_n'\mathcal{S}_n^{-1}\mathcal{G}_n$ to deduce

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

Since $\delta > 0$ is arbitrarily large and $\limsup ||\mathcal{S}_n||^{-1/2} > 0$ by non-degeneracy the last term $o_p(n^{1/2}||\mathcal{S}_n||^{-1/2}n^{-\delta}) = o_p(1)$.

Now use $\sum_{t=1}^n \{\hat{m}_{n,t}^* - m_{n,t}^*\} = o_p(||\mathcal{S}_n||^{1/2}n^{1/2})$ by Lemma B.3 and $\delta > 0$ arbitrarily large to deduce

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

Therefore $\mathcal{V}_n^{1/2}(\hat{\theta}_n^* - \theta^0) \xrightarrow{d} N(0, I_q)$ by CLT Lemma B.6. The claim now follows from Proposition B.1. $\mathcal{QED}$.

PROOF OF THEOREM 3.1. All steps used to proved Theorems 2.1 and 2.2 carry over to show $\mathring{\mathcal{V}}_n^{1/2}(\hat{\phi}_n - \phi^0) \xrightarrow{d} N(0, I_p)$ by straightforward alternations of Lemmas B.2-B.7. $\mathcal{QED}$.

PROOF OF COROLLARY 3.2. The AR process $\{y_t\}$ is stationary, ergodic and geometrically β-mixing (Pham and Tran 1985), hence Assumptions MX and STM hold, and response smoothness RS is automatic. Assumptions D, L and N hold by smoothness, uniform boundedness of $P(u_t \leq u)$ and $(\partial/\partial u)P(u_t \leq u)$, and product convolution tail properties under (6) (see Cline 1986 and his references). Further, $E[\mathring{m}_{i,n,t}^*] = E[u_{n,t}^*] \times E[y_{n,t-i}^*] = 0$ by independence and symmetry, hence I holds.

Lastly, we verify moment smoothness MS. By the definition of a derivative and an extension of Lemma B.7.b to $\mathring{m}_{n,t}^*(\phi)$,

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

By distribution smoothness, $E[y_t^2] = \infty$, trimming negligibility and stationarity $\inf_{r'r=1} r'E[x_{n,t}^* x_{n,t}^{*\prime}]r = \inf_{r'r=1} E(\sum_{i=1}^p r_i x_{i,n,t}^*)^2 \to \infty$, so $\mathring{\mathcal{G}}_n$ is non-singular for each $n \geq N$ and some $N \in \mathbb{N}$. Therefore, since $\liminf_{n\to\infty} ||\mathring{\mathcal{G}}_n|| > 0$ and Φ is compact it follows $\mathfrak{e}_n := \sup_{\phi\in\Phi} ||E[\mathring{m}_{n,t}^*(\phi)]|| \leq K||\mathring{\mathcal{G}}_n|| \times (1 + o(1))$, hence

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

for every $n \geq N$, which verifies MS. $\mathcal{QED}$.

PROOF OF THEOREM 3.3. The same argument as the proof of Theorem 3.1 applies. $\mathcal{QED}$.

PROOF OF COROLLARY 3.4. The process $\{y_t, \sigma_t\}$ is stationary and geometrically β-mixing, and a stationary geometrically β-mixing solution $\{h_t^*(\theta)\}$ exists (Nelson 1990, Carrasco and Chen 2002, Francq and Zakoïan 2006). Further, $\{m_{i,t}(\theta), G_{i,j,t}(\theta)\}$ have power law tails in heavy tailed cases since iid ϵ_t has tail (9). Therefore Assumptions D, MX, and STM hold, while L, N, RB.ii-iv and RS.ii hold by construction, or by arguments identical to those used to prove

Corollary 3.2. Assumption I can always be assumed to hold for some choice of $\{k_{1,n}^{(\epsilon)}, k_{2,n}^{(\epsilon)}\}$, given $E[\epsilon_t^2] = 1$ and $E|\mathfrak{s}_t\mathfrak{s}_t'| < \infty$ (Francq and Zakoïan 2004). Finally, the scale forms follow from (8) and the remarks following Theorem 3.3. $\mathcal{QED}$.

LEMMA B.8 (*HAC*). $\widehat{\mathcal{S}}_n^*(\hat{\theta}_n) = \mathcal{S}_n(1 + o_p(1))$.

PROOF OF THEOREM 5.1. The claim follows from Jacobian consistency Lemma B.7.a and HAC consistency Lemma B.8. $\mathcal{QED}$.

PROOF OF COROLLARY 5.2. Recall $\mathring{\mathcal{V}}_{i,j,n} \sim nE[f_{i,t}^{\phi} I_{i,n,t}^{(f^{\phi})} f_{i,t}^{\phi} I_{j,n,t}^{(f^{\phi})}]/E[u_{n,t}^{*2}]$. The proof of Lemma B.7.a in the supplementary Appendix C, below, can be altered to show $1/n \sum_{t=1}^n f_{i,t}^{\phi}(\hat{\phi}_n)\hat{I}_{j,n,t}^{(f^{\phi})}(\hat{\phi}_n) \times f_{j,t}^{\phi}(\hat{\phi}_n)\hat{I}_{j,n,t}^{(f^{\phi})}(\hat{\phi}_n) = E[f_{i,t}^{\phi} I_{i,n,t}^{(f^{\phi})} f_{i,t}^{\phi} I_{j,n,t}^{(f^{\phi})}] \times (1 + o_p(1))$. Similarly, the proof of Lemma B.8 in Appendix C can be greatly simplified to show $1/n \sum_{t=1}^n u_t^2(\hat{\phi}_n)\hat{I}_{n,t}^{(u)}(\phi) = E[u_{n,t}^{*2}] \times (1 + o_p(1))$. $\mathcal{QED}$.

REFERENCES

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

Amemiya, T. (1985). Advanced Econometrics, Harvard University Press, Cambridge.

An, H.Z. and F.C. Huang (1996). The Geometrical Ergodicity of Nonlinear Autoregressive Models, Stat. Sin. 6, 943–956.

Antoine, B. and E. Renault (2010). Efficient Minimum Distance Estimation with Multiple Rates of Convergence, J. Econometrics: forthcoming.

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

Boudt, K. and C. Croux (2010). Robust M-Estimation of Multivariate GARCH Models, Computational Statistics and Data Analysis 54, 2459 2469.

Brockwell, P.J. and D.B.H. Cline (1985). Linear Prediction of ARMA Processes with Infinite Variance, Stoch. Proc. Appl. 19, 281–296.

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

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

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

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

Davis, R.A. and W.T.M. Dunsmuir (1997). Least Absolute Deviation Estimation for Regression with ARMA Errors, J. Theor. Prob. 10, 481–497.

Davis, R.A. and S.I. Resnick (1986). Limit Theory for the Sample Covariance and Correlation Functions of Moving Averages, Ann. Stat. 14, 533-558.

Davis, R.A., K. Knight and J. Liu (1992). M-Estimation for Autoregressions with Infinite Variance, Stoch. Proc. Appl. 40, 145-180.

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

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

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

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

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

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

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

Finkenstädt, B. and H. Roozén (2001). Extreme Values in Finance, Telecommunications, and the Environment, Chapman and Hall: New York.

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

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

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

Gonçalves, S. and H. White (2005). Bootstrap Standard Error Estimates for Linear Regression, J. Amer. Stat. Assoc. 100, 970-979.

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

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

Hill, B.M. (1975). A Simple General Approach to Inference About the Tail of a Distribution, Annals of Statistics 3, 1163-1174.

Hill, J.B. (2006). Strong Orthogonal Decompositions and Nonlinear Impulse Response Functions for Infinite Variance Processes, Can. J. Stat. 34, 453-473.

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

Hill, J.B. (2011). Tail and Non-Tail Memory with Applications to Extreme Value and Robust Statistics, Econometric Theory: forthcoming.

Ibragimov, I.A. (1962). Some Limit Theorems for Stationary Processes, Theory Prob. Appl. 7, 349–382.

Jensen, S.T. and A. Rahbek (2004). Asymptotic Inference for Nonstationary GARCH, Econometric Theory 20, 1202–1226.

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

Lee, O. (2007). On Strict Stationarity of Nonlinear ARMA Processes with Nonlinear GARCH Innovations, J. Kor. Stat. Soc. 36, 183-200.

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

Léger, C. and J.P. Romano (1990). Bootstrap Adaptive Estimation: The Trimmed-Mean Example, Can. J. Stat. 18, 297-314.

Ling, S. (2005). Self-Weighted LAD Estimation for Infinite Variance Autoregressive Models, J. R. Stat.

Soc. B 67, 381–393.

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

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

Lumsdaine, R. (1996). Consistency and Asymptotic Normality of the Quasi-Maximum Likelihood Estimator in IGARCH(1,1) and Covariance Stationary GARCH(1,1) Models, Econometrica 64, 575-596.

Mallows, C.L. (1973), Influence Functions, Conference on Robust Regression, Cambridge, Mass., National Bureau for Economic Research, unpublished manuscript.

Meitz, M., and P. Saikkonen (2008). Stability of Nonlinear AR-GARCH Models, J. Time Ser. Anal. 29, 453-475.

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

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

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

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

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

Ollila, E., H.O. Thomas, and P. Hettmansperger (2002). Estimates of Regression Coefficients Based on the Sign Covariance Matrix, J. R. Stat. Soc. B 64, 447-466.

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

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

Peng, L. and Q. Yao (2004). Nonparametric Regression under Dependent Errors with Infinite Variance, Ann. Inst. Statist. Math. 56, 73-86.

Pham, T.D. and L.T. Tran (1985). Some Mixing Properties of Time Series Models, Stoch. Proc. Appl. 19, 297-303.

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

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

Rousseeuw, P.J. (1984). Least Median of Squares Regression, J. Amer. Stat. Assoc. 79, 871-880.

Rousseeuw, P.J., S. Van Alest, K. Van Driessen, and J. Agulló (2004). Robust Multivariate Regression, Technometrics 46 293–305.

Straumann, D. and T. Mikosch (2006). Quasi-Maximum-Likelihood Estimation in Conditionally Heteroscedastic Time Series: A Stochastic Recurrence Equations Approach, Ann. Stat. 34, 2449-2495.

TABLE 1 : AR(2) Estimation Results ($\phi_3^0 = -.3$)

	$n = 100$						$n = 800$				
	Mean	MSE	KS[a]	λ_u, λ_y^b	Tr%[c]		Mean	MSE	KS	λ_u, λ_y	Tr%
				$\kappa = .75$ ($u_t \stackrel{iid}{\sim} \bar{P}_\kappa(u)$)							
LTTS-Fix[d]	-.302	.014	.097	.05, 1.0	.08		-.298	.0008	.038	.05, 1.0	.02
LTTS-MSE	-.294	.023	.092	.09, 1.9	.14		-.299	.0007	.041	.42, 1.6	.06
LWAD	-.301	.032	.112	-	.05		-.300	.0006	.043	-	.05
OLS	-.299	.004	.198	-	-		-.300	.0006	.258	-	-
				$\kappa = 2.5$ ($u_t \stackrel{iid}{\sim} \bar{P}_\kappa(u)$)							
LTTS-Fix	-.303	.011	.028	.05, 1.0	.09		-.301	.0007	.020	.05, 1.0	.02
LTTS-MSE	-.298	.007	.039	.30, .48	.04		-.300	.0005	.022	.24, .46	.02
LWAD	-.303	.022	.048	-	.05		-.300	.0001	.024	-	.05
OLS	-.302	.003	.043	-	-		-.301	.0003	.043	-	-
				$\kappa = 1.5$ ($u_t \sim$ GARCH)							
LTTS-Fix	-.296	.020	.036	.05, 1.0	.08		-.297	.002	.036	.05, 1.0	.05
LTTS-MSE	-.292	.022	.051	.18, .47	.03		-.298	.002	.039	.11, .50	.02
LWAD	-.304	.034	.055	-	.05		-.301	.029	.048	-	.05
OLS	-.310	.029	.084	-	-		-.295	.016	.074	-	-

a. Kolmogorov-Smirnov statistic for a test of standard normality on standardized $\hat{\phi}_{n,3}$. The 10%, 5%, 1% critical values based on 1000 iid draws are .043, .038, .034.

b. Fractile parameter in $k_n^{(u)} = [\lambda_u n/\ln(n)]$ and $k_n^{(y)} = [\lambda_y \ln(n)]$. LTTS-Fix uses fixed $(\lambda_u, \lambda_y) = (.05, 1.0)$; LTTS-MSE uses the minimum bootstrap-mse (λ_u, λ_y). The simulation average is shown.

c. The total sample proportion of least squares loss $\epsilon_t^2(\phi)$ trimmed for LTTS. By construction Tr% $\rightarrow 0$ as n $\rightarrow \infty$. For LWAD this is the quantile at which down-weighting takes place.

d. LTTS = Least Tail-Trimmed Squares; LWAD = Least Weighted Absolute Deviations.

TABLE 2 : TAR(1) Estimation Results ($\phi_2^0 = .6$)

	$n = 100$						$n = 800$				
	Mean	MSE	KS	λ_u, λ_y^a	Tr%		Mean	MSE	KS	λ_u, λ_y	Tr%
				$\kappa = .75$ ($u_t \stackrel{iid}{\sim} \bar{P}_\kappa$)							
LTTS-Fix	.589	.019	.116	.05,1.0	.06		.599	.000	.122	.05,1.0	.03
LTTS-MSE	.584	.021	.121	.09,1.7	.13		.602	.001	.119	.38,1.5	.05
LWAD	.596	.019	.241	-	.05		.598	.004	.189	-	.05
OLS	.559	.015	.269	-	-		.574	.008	.373	-	-
				$\kappa = 2.5$ ($u_t \stackrel{iid}{\sim} \bar{P}_\kappa$)							
LTTS-Fix	.591	.017	.067	.05,1.0	.08		.598	.001	.040	.05,1.0	.03
LTTS-MSE	.588	.022	.074	.35,.52	.06		.599	.000	.041	.26,.52	.03
LWAD	.590	.068	.083	-	.05		.600	.018	.038	-	.05
OLS	.579	.014	.097	-	-		.596	.002	.070	-	-
				$\kappa = 1.5$ ($u_t \sim$ GARCH)							
LTTS-Fix	.581	.026	.063	.05,1.0	.07		.594	.004	.041	.05,1.0	.03
LTTS-MSE	.593	.022	.059	.22,.54	.04		.599	.005	.040	.09,.45	.02
LWAD	.588	.086	.079	-	.05		.598	.023	.039	-	.05
OLS	.543	.033	.113	-	-		.568	.019	.109	-	-

a. Fractile parameter in $k_n^{(u)} = [\lambda_u n/\ln(n)]$ and $k_n^{(y)} = [\lambda_y \ln(n)]$.

TABLE 3 : LSTAR(1) Estimation Results ($\phi^0 = .25$)

	$n = 100$						$n = 800$				
	Mean	MSE	KS	λ_u, λ_y^a	Tr%		Mean	MSE	KS	λ_u, λ_y	Tr%
$\kappa = .75$ ($u_t \overset{iid}{\sim} \bar{P}_\kappa$)											
LTTS-Fix	.250	.021	.068	.05,1.0	.16		.249	.020	.071	.05,1.0	.03
LTTS-MSE	.253	.020	.072	.09,1.4	.13		.250	.018	.069	.22,1.3	.06
LWAD	.250	.146	.072	-	.05		.239	.143	.098	-	.05
OLS	.254	.026	.088	-	-		.257	.023	.096	-	-
$\kappa = 2.5$ ($u_t \overset{iid}{\sim} \bar{P}_\kappa$)											
LTTS-Fix	.251	.025	.066	.05,1.0	.04		.254	.003	.070	.05,1.0	.01
LTTS-MSE	.251	.027	.063	.12,1.2	.08		.252	.003	.071	.18,1.3	.03
LWAD	.246	.146	.076	-	.05		.254	.147	.068	-	.05
OLS	.249	.012	.071	-	-		.249	.003	.103	-	-
$\kappa = 1.5$ ($u_t \sim$ GARCH)											
LTTS-Fix	.261	.041	.102	.05,1.0	.06		.253	.071	.061	.05,1.0	.02
LTTS-MSE	.258	.044	.096	.20,1.6	.15		.253	.068	.064	.22,1.5	.04
LWAD	.251	.144	.070	-	.05		.251	.143	.067	-	.05
OLS	.238	.030	.094	-	-		.256	.020	.078	-	-

a. Fractile parameter in $\mathrm{k}_\mathrm{n}^{(u)} = [\lambda_u \mathrm{n}/\ln(\mathrm{n})]$ and $\mathrm{k}_\mathrm{n}^{(y)} = [\lambda_y \ln(\mathrm{n})]$.

TABLE 4 : AR, TAR and LSTAR t-tests at 5% level

	$\kappa = .75$ ($u_t \overset{iid}{\sim} \bar{P}_\kappa$)			$\kappa = 2.5$ ($u_t \overset{iid}{\sim} \bar{P}_\kappa$)			$\kappa = 1.5$ (GARCH)		
	H_0	H_1^1	H_1^2	H_0	H_1^1	H_1^2	H_0	H_1^1	H_1^2
AR[a], $\phi_3^0 = -.3$, $n = 100$									
LTTS-Fix	.058[b]	.376	.758	.051	.453	.784	.061	.186	.644
LWAD	.074	.500	.998	.047	.673	1.00	.037	.224	.948
OLS	.041	.143	.980	.049	.248	.935	.042	.124	.535
AR, $\phi_3^0 = -.3$, $n = 800$									
LTTS-Fix	.053	.995	1.00	.050	1.00	1.00	.050	.894	1.00
LWAD	.058	.791	1.00	.049	1.00	1.00	.041	.837	1.00
OLS	.034	.988	1.00	.051	.826	1.00	.045	.112	.728
TAR[c], $\phi_3^0 = .6$, $n = 100$									
LTTS-Fix	.067	.858	.971	.05	.810	.992	.056	.664	.993
LWAD	.034	.998	1.00	.054	.993	.999	.056	.976	1.00
OLS	.078	.828	.985	.053	.765	.981	.064	.479	.830
TAR, $\phi_3^0 = .6$, $n = 800$									
LTTS-Fix	.060	1.00	1.00	.053	1.00	1.00	.052	.998	1.00
LWAD	.053	1.00	1.00	.056	1.00	1.00	.044	.992	.999
OLS	.049	.960	.980	.050	1.00	1.00	.060	.781	.960
LSTAR[d], $\phi^0 = .25$, $n = 100$									
LTTS-Fix	.020	.421	.852	.030	.407	.852	.009	.381	.626
LWAD	.000	.427	.835	.005	.414	.820	.017	.437	.845
OLS	.002	.375	.750	.082	.644	.969	.074	.305	.649
LSTAR, $\phi^0 = .25$, $n = 800$									
LTTS-Fix	.029	.774	1.00	.053	.703	1.00	.072	.600	.941
LWAD	.012	781	1.00	.023	.719	.837	.037	.627	.977
OLS	.006	.520	.807	.059	.994	1.00	.051	.563	.874

a. H_0: $\phi_3 = -.3$; H_1^1: $\phi_3 = -.2$; and H_1^2: $\phi_3 = 0$. b. Rejection frequencies at the 5% level.
c. H_0: $\phi_2 = .6$; H_1^1: $\phi_2 = .4$; and H_1^2: $\phi_2 = 0$. d. H_0: $\phi = .25$; H_1^1: $\phi = .45$; and H_1^2: $\phi_2 = 0$.

TABLE 5 : LTTS-Fix[a] Wald Test of AR(1) vs. AR(2)[b]

	$\kappa = .75$						$\kappa = 2.5$					
	$n = 100$			$n = 800$			$n = 100$			$n = 800$		
ϕ_3^0	1%[c]	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
.00	.011[d]	.045	.071	.019	.049	.084	.012	.058	.100	.010	.049	.101
-.20	.374	.551	.568	.853	.938	.961	.391	.617	.732	.991	.994	.999
-.30	.553	.768	.806	.971	.985	.992	.750	.903	.938	.996	.998	.999

a. Fixed trimming parameters are $\{\lambda_u, \lambda_y\} = \{.05, 1.0\}$.

b. We simulate AR(2) models $y_t = .2 + .8 y_{t-1} + \phi_3^0 y_{t-2} + u_t$ with $u_t \overset{iid}{\sim} \bar{P}_{2.5}$ and test $\phi_3 = 0$.

c. Nominal test sizes are 1%, 5% and 10%. d. Values are rejection frequencies at the nominal level.

TABLE 6 : GARCH

	$n = 100$						$n = 800$				
	Mean	MSE	KS	λ	Tr%[a]		Mean	MSE	KS	λ	Tr%
$\beta_3^0 = .6$ and $\epsilon_t \sim \bar{P}_{2.5}$: $\kappa_y = 1.5$[b]											
QMTTL-Fix[c]	.611	.146	.048	.050	.01		.603	.036	.036	.050	.01
QMTTL-MSE	.588	.151	.051	.121	.05		.606	.061	.041	.058	.01
QMWL	.685	.147	.236	-	.05		.674	.082	.284	-	.05
QML	.691	.141	.252	-	-		.664	.112	.251	-	-
Log-LAD	.526	.176	.170	-	-		.603	.070	.076	-	-
$\beta_3^0 = .6$ and $\epsilon_t \sim N(0,1)$: $\kappa_y = 4.1$											
QMTTL-Fix	.437	.180	.217	.050	.01		.563	.065	.144	.050	.01
QMTTL-MSE	.438	.181	.233	.074	.02		.572	.065	.136	.063	.01
QMWL	.509	.196	.191	-	.05		.571	.084	.122	-	.05
QML	.502	.181	.186	-	-		.576	.072	.112	-	-
Log-LAD	.416	.228	.288	-	-		.537	.127	.192	-	-

a. The total sample proportion of the QML loss trimmed for QMTTL where % → 0 as n → ∞. For QMWL this is the quantile at which down-weighting begins.

b. Tail index of y_t is κ_y.

c. QMWL = Quasi-Maximum Weighted Likelihood; QMTTL-Fix = Quasi-Maximum Tail-Trimming Likelihood with fixed trimming parameter; and QMTTL–MSE is QMTTL with a trimming parameter chosen by minimizing the bootstrap-mse.

TABLE 7 : GARCH t-tests at 5% level ($\beta_3^0 = .6$)[a]

	$\epsilon_t \sim \bar{P}_{2.5}$: $\kappa_y = 1.5$							$\epsilon_t \sim N(0,1)$: $\kappa_y = 4.1$					
	$n = 100$			$n = 800$				$n = 100$			$n = 800$		
	H_0^a	H_1^1	H_1^2	H_0	H_1^1	H_1^2		H_0	H_1^1	H_1^2	H_0	H_1^1	H_1^2
QMT-Fix	.051[b]	.725	.989	.049	1.00	1.00		.058	.167	.622	.052	.985	1.00
QMT-MSE	.054	.683	.981	.053	.911	.948		.061	.131	.517	.057	.937	.996
QMW	.044	.302	.632	.034	.351	.674		.045	.144	.398	.041	.303	.609
QML	.061	.796	.993	.064	1.00	1.00		.045	.058	.414	.062	.976	1.00
Log-LAD	.079	.298	.829	.035	1.00	1.00		.034	.032	.258	.067	.576	.957

a. The hypotheses are H_0: $\beta_3 = \beta_3^0$; H_1^1: $\beta_3 = \beta_3^0 - .25$; and H_1^2: $\beta_3 = 0$.

b. Rejection frequencies at the 5% level.

Robust M-Estimation for Heavy Tailed Nonlinear AR-GARCH

Supplemental Appendix C

Jonathan B. Hill

Dept. of Economics, University of North Carolina

We present omitted simulation results in Section C.1, the omitted Figure 10 from the empirical application of Section C.2, and the proofs of Lemmas B.2-B.7 in Section C.3. References unique to this appendix are presented last.

C.1 SIMULATION RESULTS In the main paper we omitted results for the iid $\bar{P}_{1.5}$ error for all AR models; and all LAD results for all AR models. Below are all estimation results for AR models.

TABLE 1 : AR(2) Estimation Results ($\phi_3^0 = -.3$)

	$n = 100$					$n = 800$				
	Mean	MSE	KS[a]	λ_u, λ_y[b]	Tr%[c]	Mean	MSE	KS	λ_u, λ_y	Tr%
				$\kappa = .75$ ($u_t \overset{iid}{\sim} \bar{P}_\kappa(u)$)						
LTTS-Fix[d]	-.302	.014	.097	.05, 1.0	.08	-.298	.0008	.038	.05, 1.0	.02
LTTS-MSE	-.294	.023	.092	.09, 1.9	.14	-.299	.0007	.041	.42, 1.6	.06
LWAD	-.301	.032	.112	-	.05	-.300	.0006	.043	-	.05
LAD	-.300	.012	.134	-	-	-.300	.0005	.046	-	-
OLS	-.299	.004	.198	-	-	-.300	.0006	.258	-	-
				$\kappa = 1.5$ ($u_t \overset{iid}{\sim} \bar{P}_\kappa(u)$)						
LTTS-Fix	-.301	.013	.036	.05, 1.0	.088	-.300	.0007	.029	.05, 1.0	.021
LTTS-MSE	-.297	.012	.071	.20, 1.2	.091	-.301	.0005	.042	.26, .55	.040
LWAD	-.302	.023	.064	-	.050	-.300	.0004	.040	-	.050
LAD	-.301	.029	.094	-	-	-.300	.0005	.067	-	-
OLS	-.306	.006	.153	-	-	-.300	.0007	.102	-	-
				$\kappa = 2.5$ ($u_t \overset{iid}{\sim} \bar{P}_\kappa(u)$)						
LTTS-Fix	-.303	.011	.028	.05, 1.0	.09	-.301	.0007	.020	.05, 1.0	.02
LTTS-MSE	-.298	.007	.039	.30, .48	.04	-.300	.0005	.022	.24, .46	.02
LWAD	-.303	.022	.048	-	.05	-.300	.0001	.024	-	.05
LAD	-.303	.046	.072	-	-	-.300	.0003	.034	-	-
OLS	-.302	.003	.043	-	-	-.301	.0003	.043	-	-
				$\kappa = 1.5$ ($u_t \sim$ GARCH)						
LTTS-Fix	-.296	.020	.036	.05, 1.0	.08	-.297	.002	.036	.05, 1.0	.05
LTTS-MSE	-.292	.022	.051	.18, .47	.03	-.298	.002	.039	.11, .50	.02
LWAD	-.304	.034	.055	-	.05	-.301	.029	.048	-	.05
LAD	-2.96	.086	.052	-	-	-.299	.037	.071	-	-
OLS	-.310	.029	.084	-	-	-.295	.016	.074	-	-

a. Kolmogorov-Smirnov statistic for a test of standard normality on standardized $\hat{\phi}_{n,3}$. The 10%, 5%, 1% critical values based on 1000 iid draws are .043, .038, .034.

b. Fractile parameter in $k_n^{(u)} = [\lambda_u n/\ln(n)]$ and $k_n^{(y)} = [\lambda_y \ln(n)]$. LTTS-Fix uses fixed $(\lambda_u, \lambda_y) = (.05,1.0)$; LTTS-MSE uses the minimum bootstrap-mse (λ_u, λ_y). The simulation average is shown.

c. The total sample proportion of least squares loss $\epsilon_t^2(\phi)$ trimmed for LTTS. By construction Tr% $\rightarrow 0$ as n $\rightarrow \infty$. For LWAD this is the quantile at which down-weighting takes place.

d. LTTS = Least Tail-Trimmed Squares; LWAD = Least Weighted Absolute Deviations.

TABLE 2 : TAR(1) Estimation Results ($\phi_2^0 = .6$)

	$n = 100$						$n = 800$				
	Mean	MSE	KS	λ_u, λ_y^a	Tr%		Mean	MSE	KS	λ_u, λ_y	Tr%
$\kappa = .75$ ($u_t \overset{iid}{\sim} \bar{P}_\kappa$)											
LTTS-Fix	.589	.019	.116	.05,1.0	.06		.599	.000	.122	.05,1.0	.03
LTTS-MSE	.584	.021	.121	.09,1.7	.13		.602	.001	.119	.38,1.5	.05
LWAD	.596	.019	.241	-	.05		.598	.004	.189	-	.05
LAD	.594	.034	.335	-	-		.599	.002	.143	-	-
OLS	.559	.015	.269	-	-		.574	.008	.373	-	-
$\kappa = 1.5$ ($u_t \overset{iid}{\sim} \bar{P}_\kappa$)											
LTTS-Fix	.579	.016	.083	.05,1.0	.08		.598	.001	.054	.05,1.0	.03
LTTS-MSE	.581	.019	.087	.16,1.5	.10		.596	.001	.059	.30,1.1	.04
LWAD	.591	.046	.113	-	.05		.601	.009	.067	-	.05
LAD	.593	.045	.137	-	-		.600	.007	.071	-	-
OLS	.579	.017	.123	-	-		.594	.004	.143	-	-
$\kappa = 2.5$ ($u_t \overset{iid}{\sim} \bar{P}_\kappa$)											
LTTS-Fix	.591	.017	.067	.05,1.0	.08		.598	.001	.040	.05,1.0	.03
LTTS-MSE	.588	.022	.074	.35,.52	.06		.599	.000	.041	.26,.52	.03
LWAD	.590	.068	.083	-	.05		.600	.018	.038	-	.05
LAD	.589	.057	.133	-	-		.599	.014	.038	-	-
OLS	.579	.014	.097	-	-		.596	.002	.070	-	-
$\kappa = 1.5$ ($u_t \sim$ GARCH)											
LTTS-Fix	.581	.026	.063	.05,1.0	.07		.594	.004	.041	.05,1.0	.03
LTTS-MSE	.593	.022	.059	.22,.54	.04		.599	.005	.040	.09,.45	.02
LWAD	.588	.086	.079	-	.05		.598	.023	.039	-	.05
LAD	575	.103	.110	-	-		.594	.038	.083	-	-
OLS	.543	.033	.113	-	-		.568	.019	.109	-	-

a. Fractile parameter in $k_n^{(u)} = [\lambda_u n/\ln(n)]$ and $k_n^{(y)} = [\lambda_y \ln(n)]$.

TABLE 3 : LSTAR(1) Estimation Results ($\phi^0 = .25$)

	$n = 100$						$n = 800$				
	Mean	MSE	KS	λ_u, λ_y^a	Tr%		Mean	MSE	KS	λ_u, λ_y	Tr%
$\kappa = .75$ ($u_t \overset{iid}{\sim} \bar{P}_\kappa$)											
LTTS-Fix	.250	.021	.068	.05,1.0	.16		.249	.020	.071	.05,1.0	.03
LTTS-MSE	.253	.020	.072	.09,1.4	.13		.250	.018	.069	.22,1.3	.06
LWAD	.250	.146	.072	-	.05		.239	.143	.098	-	.05
LAD	.248	.141	.078	-	-		.255	.147	.079	-	-
OLS	.254	.026	.088	-	-		.257	.023	.096	-	-
$\kappa = 1.5$ ($u_t \overset{iid}{\sim} \bar{P}_\kappa$)											
LTTS-Fix	.261	.016	.081	.05,1.0	.06		.258	.015	.074	.05,1.0	.04
LTTS-MSE	.260	.016	.080	.15,1.2	.06		.258	.016	.073	.14,1.0	.04
LWAD	.257	.138	.078	-	.05		.259	.140	.073	-	.05
LAD	.256	.142	.067	-	-		.254	.143	.075	-	-
OLS	.277	.020	.000	-	-		.268	.015	.080	-	-
$\kappa = 2.5$ ($u_t \overset{iid}{\sim} \bar{P}_\kappa$)											
LTTS-Fix	.251	.025	.066	.05,1.0	.04		.254	.003	.070	.05,1.0	.01
LTTS-MSE	.251	.027	.063	.12,1.2	.08		.252	.003	.071	.18,1.3	.03
LWAD	.246	.146	.076	-	.05		.254	.147	.068	-	.05
LAD	.248	.140	.061	-	-		.260	.139	.091	-	-
OLS	.249	.012	.071	-	-		.254	.003	.103	-	-
$\kappa = 1.5$ ($u_t \sim$ GARCH)											
LTTS-Fix	.261	.041	.102	.05,1.0	.06		.253	.071	.061	.05,1.0	.02
LTTS-MSE	.258	.044	.096	.20,1.6	.15		.253	.068	.064	.22,1.5	.04
LWAD	.251	.144	.070	-	.05		.251	.143	.067	-	.05
LAD	.253	.141	.069	-	-		.257	.147	.091	-	-
OLS	.238	.030	.094	-	-		.256	.020	.078	-	-

a. Fractile parameter in $\mathrm{k}_\mathrm{n}^{(u)} = [\lambda_u \mathrm{n}/\ln(\mathrm{n})]$ and $\mathrm{k}_\mathrm{n}^{(y)} = [\lambda_y \ln(\mathrm{n})]$.

TABLE 4 : AR, TAR and LSTAR t-tests at 5% level

	$\kappa = .75$ ($u_t \overset{iid}{\sim} \bar{P}_\kappa$)			$\kappa = 1.5$ ($u_t \overset{iid}{\sim} \bar{P}_\kappa$)			$\kappa = 2.5$ ($u_t \overset{iid}{\sim} \bar{P}_\kappa$)			$\kappa = 1.5$ (GARCH)		
	H_0	H_1^1	H_1^2	H_0	H_1^1	H_1^2	H_0	H_1^1	H_1^2	H_0	H_1^1	H_1^2
AR[a], $\phi_3^0 = -.3$, $n = 100$												
LTTS-Fix[b]	.058[c]	.376	.758	.048	.452	.705	.051	.453	.784	.061	.186	.644
LWAD	.074	.500	.998	.068	.914	1.00	.047	.673	1.00	.037	.224	.948
LAD	.054	1.00	1.00	.069	.942	1.00	.060	.634	1.00	.049	.185	.927
OLS	.041	.143	.980	.059	.284	.976	.049	.248	.935	.042	.124	.535
AR, $\phi_3^0 = -.3$, $n = 800$												
LTTS-Fix	.053	.995	1.00	.052	1.00	1.00	.050	1.00	1.00	.050	.894	1.00
LWAD	.058	.791	1.00	.059	1.00	1.00	.049	1.00	1.00	.041	.837	1.00
LAD	.022	1.00	1.00	.055	1.00	1.00	.052	1.00	1.00	.047	.831	.999
OLS	.034	.988	1.00	.069	1.00	1.00	.051	.826	1.00	.045	.112	.728
TAR[d], $\phi_3^0 = .6$, $n = 100$												
LTTS-Fix	.067	.858	.971	.058	.851	.985	.053	.810	.992	.056	.664	.993
LWAD	.034	.998	1.00	.048	.858	.979	.054	.993	.999	.056	.976	1.00
LAD	.017	.997	1.00	.064	1.00	1.00	.064	.998	1.00	.054	.938	.994
OLS	.078	.828	.985	.064	.828	1.00	.053	.765	.981	.064	.479	.830
TAR, $\phi_3^0 = .6$, $n = 800$												
LTTS-Fix	.060	1.00	1.00	.054	1.00	1.00	.053	1.00	1.00	.052	.998	1.00
LWAD	.053	1.00	1.00	.054	1.00	1.00	.056	1.00	1.00	.044	.992	.999
LAD	.023	1.00	1.00	.062	1.00	1.00	.060	1.00	1.00	.056	1.00	1.00
OLS	.049	.960	.980	.052	1.00	1.00	.050	1.00	1.00	.060	.781	.960
LSTAR[e], $\phi^0 = .25$, $n = 100$												
LTTS-Fix	.020	.421	.852	.030	.527	.975	.037	.407	.852	.009	.381	.626
LWAD	.000	.427	.835	.002	.583	.984	.005	.414	.820	.017	.437	.845
LAD	.000	.452	.853	.001	.458	.860	.001	.434	.858	.003	.445	.857
OLS	.002	.375	.750	.024	.737	.922	.082	.644	.969	.074	.305	.649
LSTAR, $\phi^0 = .25$, $n = 800$												
LTTS-Fix	.029	.774	1.00	.055	1.00	.981	.053	.703	1.00	.072	.600	.941
LWAD	.012	781	1.00	.028	1.00	1.00	.023	.719	.837	.037	.627	.977
LAD	.002	.454	.818	.002	.449	.844	.017	.503	.866	.019	.554	.874
OLS	.006	.520	.807	.037	.842	.906	.059	.994	1.00	.051	.563	.874

a. The AR hypotheses are H_0: $\phi_3 = -.3$; H_1^1: $\phi_3 = -.2$; and H_1^2: $\phi_3 = 0$.
b. LTTS with fixed trimming parameters $\{\lambda_u, \lambda_y\} = \{.05, 1\}$.
c. Values are rejection frequencies at the5% level.
d. The TAR hypotheses are H_0: $\phi_2 = .6$; H_1^1: $\phi_2 = .4$; and H_1^2: $\phi_2 = 0$.
e. The LSTAR hypotheses are H_0: $\phi = .25$; H_1^1: $\phi = .45$; and H_1^2: $\phi_2 = 0$.

TABLE 5 : LTTS-Fix[a] Wald-Test of AR(1) vs. AR(2)[b]

	$\kappa = .75$			$\kappa = 1.5$			$\kappa = 2.5$		
	$n = 100$			$n = 100$			$n = 100$		
ϕ_3^0	1%[c]	5%	10%	1%	5%	10%	1%	5%	10%
.00	.011[d]	.045	.071	.008	.048	.086	.012	.058	.100
-.20	.374	.551	.568	.307	.531	.652	.391	.617	.732
-.30	.553	.768	.806	.645	.794	.886	.750	.903	.938
	$n = 800$			$n = 800$			$n = 800$		
ϕ_3^0	1%	5%	10%	1%	5%	10%	1%	5%	10%
.00	.019	.049	.084	.009	.049	.093	.010	.049	.101
-.20	.853	.938	.961	.983	.991	.997	.991	.994	.999
-.30	.971	.985	.992	.997	.998	.999	.996	.998	.999

a. Fixed trimming parameters are $\{\lambda_u, \lambda_y\} = \{.05, 1\}$.

b. We simulate AR(2) models $y_t = .2 + .8y_{t-1} + \phi_3^0 y_{t-2} + u_t$ with $u_t \overset{iid}{\sim} \bar{P}_{2.5}$ and test $\phi_3 = 0$.

c. Nominal test sizes are 10%, 5% and 1%.

d. Values are rejection frequencies at the nominal level.

C.2 OMITTED FIGURES In Section 7 we estimate a TAR(3)-GARCH(1,1) model $y_t = \zeta_1' \tilde{x}_t I(y_{t-1} < 0) + \zeta_2' \tilde{x}_t I(y_{t-1} \geq 0) + u_t = \phi' x_t$, $\tilde{x}_t = [1, y_{t-1}, y_{t-2}, y_{t-3}]'$, $u_t = \sigma_t \epsilon_t$, and $\sigma_t^2 = \beta_1 + \beta_2 u_{t-1}^2 + \beta_3 \sigma_{t-1}^2$. The Hill-plot for the residuals $|\hat{\epsilon}_t|$ is presented in Figure 10, below.

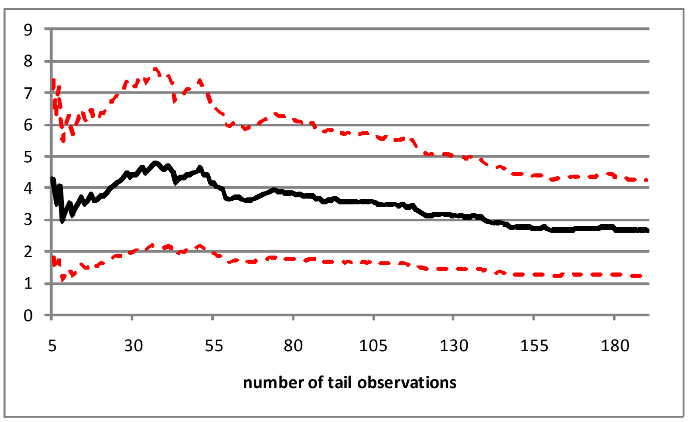

Figure 10: Hill-Plot and Robust 90% Bands for TAR-GARCH Residuals

The Hill-plot of the tail index of $|\hat{\epsilon}_t|$ reveals the Hill (1975) estimator $\hat{\kappa}_{\epsilon,\tilde{k}_n}$ hovers between 3 and 5 for all $\tilde{k}_n \in \{5, ..., 200\}$, every $\hat{\kappa}_{\epsilon,\tilde{k}_n} > 2$, nearly every $\hat{\kappa}_{\epsilon,\tilde{k}_n} > 3$, $\hat{\kappa}_{\epsilon,\tilde{k}_n} < 4$ in 84% of the bands, and values $\kappa_\epsilon < 4$ lie in each 90% band. Evidence for $E[\epsilon_t^2] < \infty$ is in accord with our assumptions, and the possibility $\kappa_\epsilon < 4$ rules out QML and QMWL for standard inference

C.3 PROOFS OF LEMMAS B.2-B.7 By Proposition A.1 stationarity solutions $\{h_t^*(\theta), h_t^{*\theta}(\theta), h_t^{*\theta,\theta}(\theta)\}$ exist. Further, by Proposition B.1 it suffices to treat the infeasible estimator $\hat{\theta}_n^*$ and components $m_t^*(\theta)$ and $G_t^*(\theta)$. We therefore state and prove all lemmas for $m_t^*(\theta)$, $G_t^*(\theta)$, and so on.

Since the notation "$*$" is repetitive we drop it everywhere: we write $\hat{\theta}_n$ for $\hat{\theta}_n^*$, and

$$h_t(\theta) = h_t^*(\theta), \ \ m_t(\theta) = m_t^*(\theta), \ \ m_{n,t}(\theta) = m_{n,t}^*(\theta), \ \ G_t(\theta) = G_t^*(\theta), \ \ G_{n,t}(\theta) = G_{n,t}^*(\theta) = G_t^* I_{n,t}^*(\theta)$$

Recall all Jacobian and covariance matrices:

$$\mathcal{G}_n(\theta) := E\left[G_{n,t}(\theta)\right] \quad \text{and } \check{\mathcal{G}}_n^*(\theta) := \frac{\partial}{\partial\theta}E\left[m_{n,t}^*(\theta)\right]$$

$$\widehat{\mathcal{G}}_n(\theta) := \frac{1}{n}\sum_{t=1}^n G_t(\theta)\hat{I}_{n,t}(\theta) \;\; \text{and} \;\; \tilde{\mathcal{G}}_n(\theta) := \frac{1}{n}\sum_{t=1}^n G_{n,t}(\theta) = \frac{1}{n}\sum_{t=1}^n G_t(\theta)I_{n,t}(\theta)$$

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

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

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

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

In order to shorten arguments, and in lieu of power-law tail Assumption D.ii, we assume $w_t(\theta)$ has power law tails for any θ:

$$\sup_{\theta\in\Theta}\left\{\left|c^{\kappa_w(\theta)}P\left(|w_t(\theta)| > c\right) - d_w(\theta)\right|\right\} \to 0 \;\; \textit{as } c \to \infty \textit{ where } \inf_{\theta\in\Theta}\{d_w(\theta)\} > 0, \tag{22}$$

By construction of $\mathcal{C}_n(\theta)$ and tail (22) we have the following *uniform fractile and threshold* properties: if $\kappa_w(\theta) \leq 2$ then

$$\inf_{\theta\in\Theta}\frac{\mathcal{C}_n(\theta)}{(n/k_n)^{1/\kappa_w(\theta)}} \to K \;\; \text{and} \;\; \sup_{\theta\in\Theta}\frac{\mathcal{C}_n(\theta)}{(n/k_n)^{1/\kappa_w(\theta)}} \to K. \tag{UFT}$$

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

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

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

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

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

The proofs of Lemmas B.2-B.7 require several supporting results. First, intermediate order statistics are uniformly bounded in probability, and trimming indicators satisfy a uniform law.

Notice we require RB, RS and STM throughout to ensure stationary solutions for the volatility process and its derivatives by Proposition A.1. Recall RB is not required if GARCH parameters are not estimated. We do not mention stationarity to reduce repetitive declarations.

LEMMA C.1 (*uniform indicator law*). *Let Assumptions D, MX, RB, RS and STM hold. Let* $\mathcal{I}_{n,t}(\theta)$ *denote* $((n/k_n)^{1/2})\{I(|w_t(\theta)| \leq \mathcal{C}_n(\theta)) - E[I(|w_t(\theta)| \leq \mathcal{C}_n(\theta))]\}$. *Then* $\{n^{-1/2}\sum_{t=1}^n \mathcal{I}_{n,t}(\theta) : \theta \in \Theta\} \Longrightarrow^* \{\mathcal{I}(\theta) : \theta \in \Theta\}$ *and* $E[\sup_{\theta\in\Theta}|n^{-1/2}\sum_{t=1}^n \mathcal{I}_{n,t}(\theta)|] = O(1)$ *where* $\mathcal{I}(\theta)$ *is a Gaussian process with uniformly bounded and uniformly continuous sample paths* *with respect to L_2-norm, and $\Longrightarrow^*$ denotes weak convergence on a Polish space*[15].

PROOF. By construction $\mathcal{I}_{n,t}(\theta)$ is L_2-bounded uniformly on $1 \leq t \leq n$, $n \geq 1$, and Θ, and under MX $\mathcal{I}_{n,t}(\theta)$ is geometrically β-mixing. Further, $\{\mathcal{I}_{n,t}(\theta) : \theta \in \Theta\}$ satisfies the metric entropy with L_2-bracketing bound[16] $\int_0^1 \mathcal{H}_{[\,]}(\varepsilon, \Theta, || \cdot ||_2)d\varepsilon < \infty$. This follows since the finite dimensional distributions of $w_t(\theta)$ are absolutely continuous under D.i, hence the thresholds $\mathcal{C}_n(\theta)$ are continuous. Further, fdd's of $w_t(\theta)$ have bounded densities uniformly on Θ by D.i. Therefore $\mathcal{I}_{n,t}(\theta)$ is L_2-Lipschitz: $E[(\mathcal{I}_{n,t}(\theta) - \mathcal{I}_{n,t}(\tilde{\theta}))^2] \leq K||\theta - \tilde{\theta}||$. Proving the L_2-bracketing numbers satisfy $\int_0^1 \mathcal{H}_{[\,]}(\varepsilon, \Theta, ||\cdot||_2)d\varepsilon < \infty$ is then a classic exercise (Giné and Zinn 1984, Pollard 1984, 1989, van der Vaart and Wellner 1996).

We may therefore extend Doukhan et al's (1995: Theorem 1; eq. (2.17), Application 4) uniform central limit theorem $\{1/n^{1/2} \sum_{t=1}^n \mathcal{I}_{n,t}(\theta) : \theta \in \Theta\} \Longrightarrow^* \{\mathcal{I}(\theta) : \theta \in \Theta\}$, a Gaussian process with a version that has uniformly bounded and uniformly continuous sample paths with respect to $||\cdot||_2$.

Finally, Doukhan et al's (1995: Theorem 2) uniform maximal inequality applies since their required bound (2.10) holds under their (2.17), which $\int_0^1 \mathcal{H}_{[\,]}(\varepsilon, \Theta, ||\cdot||_2)d\varepsilon < \infty$ ensures. Therefore $E[(\sup_\theta |n^{-1/2} \sum_{t=1}^n \mathcal{I}_{n,t}(\theta)|] = O(1)$ which completes the proof. $\mathcal{QED}$.

LEMMA C.2 (*uniform order statistic*). *Define $w_t^{(a)}(\theta) := |w_t(\theta)|$. Under D, MX, RB, RS and STM $\sup_{\theta\in\Theta} |w_{(k_n)}^{(a)}(\theta)/\mathcal{C}_n(\theta) - 1| = O_p(k_n^{-1/2})$.*

PROOF. We first prove the pointwise limit, then the uniform limit.

Step 1 (pointwise): Define $\mathcal{I}_n(u) := 1/k_n \sum_{t=1}^n I(w_t > \mathcal{C}_n e^{u/k_n^{1/2}})$ for arbitrary $u \geq 0$. Under geometric β-mixing and power-tail decay MX and D.ii $\{k_n^{-1/2} I(w_t > \mathcal{C}_n e^u)\}$ satisfies the conditions of Hill's (2009: Theorem 2.1, Lemma 3.1) central limit theorem. Therefore point-wise

$$k_n^{1/2} \{\mathcal{I}_n(u) - E\{\mathcal{I}_n(u)\} \xrightarrow{d} N(0, w^2(u)), \text{ where } w : \mathbb{R}_+ \to \mathbb{R}_+, \text{ and } \sup_{u\geq 0} w^2(u) < \infty. \tag{23}$$

We need only show $k_n^{1/2} \ln(w_{(k_n+1)}/\mathcal{C}_n) \xrightarrow{d} N(0, v^2)$ follows from (23). By construction $k_n^{1/2} \ln(w_{(k_n+1)}/\mathcal{C}_n) \leq u$ sufficiently if $\mathcal{I}_n(u) \leq \rho$ for any $\rho \in [0,1]$ to be chosen below, and $\mathcal{I}_n(u) \leq \rho$ if

$$\begin{aligned} k_n^{1/2}(\mathcal{I}_n(u) - E[\mathcal{I}_n(u)]) &\leq k_n^{1/2}\left(\rho - \frac{n}{k_n} P\left(w_t > \mathcal{C}_n e^{u/k_n^{1/2}}\right)\right) \\ &= k_n^{1/2}\left(\rho - \frac{n}{k_n} P(w_t > \mathcal{C}_n) \frac{P\left(w_t > \mathcal{C}_n e^{u/k_n^{1/2}}\right)}{P(w_t > \mathcal{C}_n)}\right). \end{aligned}$$

Let $\mathfrak{b} : \mathbb{R}_+ \to \mathbb{R}_+$ satisfy $k_n^{1/2}\mathfrak{b}(\mathcal{C}_n) \to 0$. Power-law tail decay implies by a generalization of Hsing's (1991: p. 1553) argument for some $\mathfrak{b}$

$$\frac{n}{k_n} P(w_t > \mathcal{C}_n) = \mathcal{O}(1) \times [1 + \mathcal{O}(\mathfrak{b}(\mathcal{C}_n))] = \mathcal{O}(1) + o\left(1/k_n^{1/2}\right) \text{ as } n \to \infty$$

---

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

[16] The brackets $\{l, u\}$ of an index function class $\mathcal{F}$ satisfy $l \leq f \leq u$ for every member $f \in \mathcal{F}$, where $\{l, u\}$ may not be members of $\mathcal{F}$; an ε-L_2-bracket $\{l, u\}$ satisfies $||l - u|| \leq \varepsilon$; the L_2-bracketing numbers $\mathcal{N}_{[\,]}(\varepsilon, \Phi, ||\cdot||_2)$ are the number of ε-L_2-brackets required to cover $\mathcal{F}$, and metric entropy with L_2-bracketing is $\mathcal{H}_{[\,]}(\varepsilon, \Phi, ||\cdot||_2) = \ln(\mathcal{N}_{[\,]}(\varepsilon, \Phi, ||\cdot||_2))$. See Giné and Zinn (1984) and Pollard (1984). The property $\int_0^1 \mathcal{H}_{[\,]}^{1/2}(\varepsilon, \Phi, ||\cdot||_2)d\varepsilon < \infty$ ensures a required stochastic equicontinuity condition for weak convergence of a partial sum of $\mathcal{I}_{n,t}(\phi)$ (Dudley's 1978).

and
$$\frac{P\left(w_t > \mathcal{C}_n e^u\right)}{P\left(w_t > \mathcal{C}_n\right)} = \mathcal{O}(e^{-u\kappa}) \times \left(1 + \mathcal{O}\left(\mathfrak{b}\left(\mathcal{C}_n\right)\right)\right) = \mathcal{O}(e^{-u\kappa}) \times \left(1 + o\left(1/k_n^{1/2}\right)\right)$$
where $\mathcal{O}(w) \in [0, w]$ is a contraction mapping. Now put $\rho = \mathcal{O}(1) \in [0,1]$ to deduce $k_n^{1/2} \ln(w_{(k_n+1)}/\mathcal{C}_n) \leq u$ sufficiently if

$$\begin{aligned}
&\kappa^{-1} k_n^{1/2} \left(\mathcal{I}_n\left(u\right) - E\left[\mathcal{I}_n\left(u\right)\right]\right) \\
&\qquad \leq \kappa^{-1} k_n^{1/2} \left(\rho - \left(\mathcal{O}(1) + o\left(1/k_n^{1/2}\right)\right) \times \mathcal{O}(e^{-u\kappa}) \times \left(1 + o\left(1/k_n^{1/2}\right)\right)\right) \\
&\qquad \leq \kappa^{-1} k_n^{1/2} \left\{\rho - \mathcal{O}\left(e^{-u\kappa/k_n^{1/2}}\right) \times \left(1 + o\left(1/k_n^{1/2}\right)\right)\right\} \\
&\qquad \leq \kappa^{-1} k_n^{1/2} \left\{u\kappa/k_n^{1/2} + o\left(1/k_n^{1/2}\right)\right\} = u + o\left(1\right).
\end{aligned}$$

Since $\kappa^{-1} k_n^{1/2} \{\mathcal{I}_n(u) - E[\mathcal{I}_n(u)]\} \xrightarrow{d} Z$ a mean-zero normal law with finite variance, it follows

$$\begin{aligned}
&\lim_{n\to\infty} P\left(k_n^{1/2} \ln\left(w_{(k_n+1)}/\mathcal{C}_n\right) \leq u\right) \\
&\qquad = \lim_{n\to\infty} P\left(\kappa^{-1} k_n^{1/2} \left(\mathcal{I}_n\left(u\right) - E\left[\mathcal{I}_n\left(u\right)\right]\right) \leq u + o(1)\right) = P\left(Z \leq u\right).
\end{aligned} \tag{24}$$

Therefore $k_n^{1/2} \ln(w_{(k_n+1)}/\mathcal{C}_n) \xrightarrow{d} N(0, v^2)$ where $v^2 < \infty$, hence $w_t/\mathcal{C}_n = 1 + O_p(k_n^{-1/2})$ by the mean-value-theorem.

Step 2 (uniform): Define $\mathcal{I}_n\left(u, \theta\right) := 1/k_n \sum_{t=1}^n I(w_t^{(a)}(\theta) > \mathcal{C}_n(\theta) e^{u/k_n^{1/2}})$ and

$$\mathcal{Z}_n\left(u, \theta\right) := k_n \left(\frac{n}{k_n}\right)^{1/2} \mathcal{I}_n\left(u, \theta\right).$$

Repeat the argument leading to (24) to obtain

$$P\left(k_n^{1/2} \ln\left(m_{(k_n)}^{(a)}(\theta)/\mathcal{C}_n(\theta)\right) \leq u\right) = P\left(\kappa^{-1} n^{-1/2} \left\{\mathcal{Z}_n\left(u, \theta\right) - E\left[\mathcal{Z}_n\left(u, \theta\right)\right]\right\} \leq u + o(1)\right).$$

Lemma C.1 implies the right–hand-side converges weakly: for any $u \in \mathbb{R}$

$$P\left(\kappa^{-1} n^{-1/2} \left\{\mathcal{Z}_n\left(u, \theta\right) - E\left[\mathcal{Z}_n\left(u, \theta\right)\right]\right\} \leq u + o(1)\right) \Longrightarrow^* P\left(\mathcal{Z}\left(\theta\right) \leq u\right),$$

where $\{\mathcal{Z}(\theta) : \theta \in \Theta\}$ is a Gaussian process with a version that has uniformly bounded and uniformly continuous sample paths with respect to the L_2-norm. Therefore $\sup_\theta |w_{(k_n)}^{(a)}(\theta)/\mathcal{C}_n(\theta) - 1| = O_p(k_n^{-1/2})$ by the continuous mapping theorem. $\mathcal{QED}$.

Finally, we require an expansion of kernel weighted equation cross-products for HAC asymptotics. Write compactly

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

LEMMA C.3 (*cross-product approximation*). *Under D, F, I, MS, MX, N, RB, RS and STM* $\mathfrak{S}_n^{-1} \sum_{s,t=1}^n \mathcal{K}_{n,s,t} \{\hat{m}_{n,s}(\hat{\theta}_n) \hat{m}_{n,t}(\hat{\theta}_n) - m_{n,s}(\theta^0) m_{n,t}(\theta^0)\} = o_p(1)$.

PROOF. Drop θ^0 throughout, and assume $q = 1$ to simplify notation. It suffices to show

(25) $$\frac{1}{\mathfrak{S}_n} \sum_{s,t=1}^{n} \mathcal{K}_{n,s,t} \left\{\hat{m}_{n,s}\hat{m}_{n,t} - m_{n,s}m_{n,t}\right\} = o_p(1)$$

(26) $$\frac{1}{\mathfrak{S}_n} \sum_{s,t=1}^{n} \mathcal{K}_{n,s,t} \left\{\hat{m}_{n,s}(\hat{\theta}_n)\hat{m}_{n,t}(\hat{\theta}_n) - \hat{m}_{n,s}\hat{m}_{n,t}\right\} = o_p(1).$$

Step 1 (bound (25)): By the triangle inequality $|\mathfrak{S}_n^{-1} \sum_{s,t=1}^{n} \mathcal{K}_{n,s,t}\{\hat{m}_{n,s}\hat{m}_{n,t} - m_{n,s}m_{n,t}\}|$ is bounded by

$$K \left| \frac{1}{\mathfrak{S}_n} \sum_{s,t=1}^{n} \mathcal{K}_{n,s,t} m_s \left(\hat{I}_{n,s} - I_{n,s}\right) \left\{m_{n,t} - E\left[m_{n,t}\right]\right\} \right| + K \left| \frac{1}{\mathfrak{S}_n} \sum_{s,t=1}^{n} \mathcal{K}_{n,s,t} m_s \left(\hat{I}_{n,s} - I_{n,s}\right) E\left[m_{n,t}\right] \right|$$
$$+ K \left| \frac{1}{\mathfrak{S}_n} \sum_{s,t=1}^{n} \mathcal{K}_{n,s,t} m_s \left(\hat{I}_{n,s} - I_{n,s}\right) m_t \left(\hat{I}_{n,t} - I_{n,t}\right) \right| = \mathcal{A}_{1,n} + \mathcal{A}_{2,n} + \mathcal{A}_{3,n}.$$

Consider $\mathcal{A}_{1,n}$ and define for any $\delta > 0$

$$\eta_\delta(x) := \frac{1}{\left(2\delta^2\pi\right)^{1/2}} \exp\left\{-x^2\delta^{-2}/2\right\} \quad \text{and} \quad \eta_{\delta,n,j} := \eta_\delta\left(j/\gamma_n\right)$$

$$\mathcal{A}_{1,n,\delta} := \sum_{t=-n+1}^{2n} \left(\frac{1}{\gamma_n^{1/2}} \sum_{l=1-t}^{n-t} \mathcal{K}\left(l/\gamma_n\right) \frac{1}{\mathfrak{S}_n^{1/2}} m_{t+l} \left(\hat{I}_{n,t+l} - I_{n,t+l}\right) I\left(0 \le l \le [\gamma_n/\delta]\right) \right)$$
$$\times \left(\frac{1}{\gamma_n^{1/2}} \sum_{j=1-t}^{n-t} \eta_{\delta,n,j} \frac{1}{\mathfrak{S}_n^{1/2}} m_{n,t+j} I\left(0 \le j \le [\gamma_n/\delta]\right) \right) \times \left(1 + o_p(1)\right).$$

Trivially since $\mathfrak{S}_n = n\mathcal{S}_n$,

(27) $$E\left(\frac{1}{\mathfrak{S}_n^{1/2}} \sum_{t=1}^{n} \left\{m_{n,t} - E\left[m_{n,t}\right]\right\} \right)^2 = 1 = \sum_{t=1}^{n} \left(1/n^{1/2}\right)^2,$$

and by approximation Lemma B.3.a

(28) $$E\left(\frac{1}{\mathfrak{S}_n^{1/2}} \sum_{t=1}^{n} m_t \left(\hat{I}_{n,t} - I_{n,t}\right) \right)^2 = o(1) \le K \sum_{t=1}^{n} \left(1/n^{1/2}\right)^2 = K.$$

Therefore, by Lemmas A.2-A.3 of de Jong and Davidson (2000) we have[17]

(29) $$\lim_{\delta \to 0} \limsup_{n \to \infty} \left\| \mathcal{A}_{1,n} - \mathcal{A}_{1,n,\delta} \times \left(1 + o_p(1)\right) \right\|_1 = 0.$$

[17] de Jong and Davidson (2000: Lemma A.1) invoke a mixingale maximal inequality due to McLeish (1975) solely to ensure partial sum variance bounds. It suffices to replace their Lemma A.1 with (27) and (28) since these duplicate the same bound implied by Theorem 1.6 of McLeish (1975) with mixingale constants $1/n^{1/2}$.

Now consider $\mathcal{A}_{1,n,\delta}$, define $N_n(\delta) := \min\{n, [\gamma_n/\delta] + 1\}$ and note by construction and variance non-degeneracy

$$\limsup_{n\to\infty} \frac{N_n(\delta)}{\gamma_n} \leq K \text{ and } \frac{S^2_{N_n(\delta)}/N_n(\delta)}{\mathfrak{S}_n/n} = O(1).$$

Approximation Lemma B.3.a and CLT Lemma B.6 generalize in a straightforward way to kernel-weighted versions under the suppositions of Theorem 6.1: for any δ

$$\begin{aligned}
&\frac{n^{1/2}}{\gamma_n^{1/2}} \times \max_{-n+1\leq t\leq 2n} \left\| \frac{1}{\mathfrak{S}_n^{1/2}} \sum_{l=1-t}^{n-t} \mathcal{K}\left(l/\gamma_n\right) m_{t+l} \left(\hat{I}_{n,t+l} - I_{n,t+l}\right) I\left(0 \leq l \leq [\gamma_n/\delta]\right) \right\|_2 \\
&\quad \leq \frac{N_n^{1/2}(\delta)}{\gamma_n^{1/2}} \left\{ \frac{S_{N_n(\delta)}/N_n^{1/2}(\delta)}{\mathfrak{S}_n^{1/2}/n^{1/2}} \right\} \times \left\| \frac{1}{S_{N_n(\delta)}} \sum_{t=1}^{N_n(\delta)} \mathcal{K}\left(t/\gamma_n\right) \left\{\hat{m}_{n,t+l} - m_{n,t+l}\right\} \right\|_2 \\
&\quad \to 0 \text{ as } n \to \infty,
\end{aligned}$$

and

$$\begin{aligned}
&\frac{n^{1/2}}{\gamma_n^{1/2}} \max_{-n+1\leq t\leq 2n} \left\| \sum_{j=1-t}^{n-t} \eta_{\delta,n,j} \frac{1}{\mathfrak{S}_n^{1/2}} m_{n,t+j} I\left(0 \leq j \leq [\gamma_n/\delta]\right) \right\|_2 \\
&\quad \leq \frac{N_n^{1/2}(\delta)}{\gamma_n^{1/2}} \left\{ \frac{S_{N_n(\delta)}/N_n^{1/2}(\delta)}{\mathfrak{S}_n^{1/2}/n^{1/2}} \right\} \times \left\| \frac{1}{S_{N_n(\delta)}} \sum_{t=1}^{N_n(\delta)} \eta_{\delta,n,j} m_t \left(\hat{I}_{n,t} - I_{n,t}\right) \right\|_2 \\
&\quad \to 0 \text{ as } n \to \infty.
\end{aligned}$$

Therefore

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \left\|\mathcal{A}_{1,n,\delta}\right\|_1 = 0. \tag{30}$$

Together (29) and (30) imply $\mathcal{A}_{1,n} = o_p(1)$. A similar argument applies to $\mathcal{A}_{3,n}$

Finally, $\mathcal{A}_{2,n}$ and $\mathcal{A}_{3,n}$ follow by the same argument for $m_s(\hat{I}_{n,s} - I_{n,s})$. For $E[m_{n,t}]$ in $\mathcal{A}_{2,n}$ use the Theorem 6.1 supposition $\max_{1\leq s\leq t} \sum_{t=1}^n |\mathcal{K}_{n,s,t}| = o(n)$, identification Assumption I and $\mathfrak{S}_n = n\mathcal{S}_n$, to deduce $o(||n\mathfrak{S}_n^{-1/2} E[m_{n,t}]||) = o(||n^{1/2}\mathcal{S}_n^{-1/2} E[m_{n,t}]||) = o(1)$.

Step 2 (bound (26)): Note $|\mathfrak{S}_n^{-1} \sum_{s,t=1}^n \mathcal{K}_{n,s,t}\{\hat{m}_{n,s}(\hat{\theta}_n)\hat{m}_{n,t}(\hat{\theta}_n) - \hat{m}_{n,s}\hat{m}_{n,t}\}|$ is bounded by

$$2\left|\frac{1}{\mathfrak{S}_n} \sum_{s,t=1}^n \mathcal{K}_{n,s,t} \left\{\hat{m}_{n,s}(\hat{\theta}_n) - \hat{m}_{n,s}\right\} \hat{m}_{n,t}\right| + \left|\frac{1}{\mathfrak{S}_n} \sum_{s,t=1}^n \mathcal{K}_{n,s,t} \left\{\hat{m}_{n,s}(\hat{\theta}_n) - \hat{m}_{n,s}\right\} \left\{\hat{m}_{n,t}(\hat{\theta}_n) - \hat{m}_{n,t}\right\}\right|.$$

Consider the first term, the second being similar. Similar to the proof of Lemma B.5, apply a Taylor expansion and multiple applications of Minkowski's inequality to deduce for some $||\theta_{n,*} -$

$\theta^0|| \leq ||\hat{\theta}_n - \theta^0||$

$$\left|\frac{1}{\mathfrak{S}_n}\sum_{s,t=1}^{n}\mathcal{K}_{n,s,t}\left\{\hat{m}_{n,s}(\hat{\theta}_n) - \hat{m}_{n,s}\right\}\hat{m}_{n,t}\right|$$
$$\leq \left\|\frac{1}{\mathfrak{S}_n}\sum_{s,t=1}^{n}\mathcal{K}_{n,s,t}G_s(\theta_{n,*})\hat{I}_{n,s}(\theta_{n,*})\hat{m}_{n,t}\right\| \times \left\|\hat{\theta}_n - \theta^0\right\|$$
$$+\left\|\frac{1}{\mathfrak{S}_n}\sum_{s,t=1}^{n}\mathcal{K}_{n,s,t}G_s(\theta_{n,*})\left\{\hat{I}_{n,s}(\theta_{n,*}) - \hat{I}_{n,s}\right\}\hat{m}_{n,t}\right\| \times \left\|\hat{\theta}_n - \theta^0\right\|$$
$$+\left\|\frac{1}{\mathfrak{S}_n}\sum_{s,t=1}^{n}\mathcal{K}_{n,s,t}G_s(\theta_{n,*})\left\{\hat{I}_{n,s}(\hat{\theta}_n) - \hat{I}_{n,s}\right\}\hat{m}_{n,t}\right\| \times \left\|\hat{\theta}_n - \theta^0\right\|$$
$$+\left\|\frac{1}{\mathfrak{S}_n}\sum_{s,t=1}^{n}\mathcal{K}_{n,s,t}m_s\left\{\hat{I}_{n,s}(\hat{\theta}_n) - \hat{I}_{n,s}\right\}\hat{m}_{n,t}\right\| = \sum_{i=1}^{4}\mathcal{B}_{i,n}.$$

By Theorem 2.2 $||\theta_{n,*} - \theta^0|| \leq ||\hat{\theta}_n - \theta^0|| = O_p(||\mathcal{V}_n||^{-1/2})$. Now define $i := (-1)^{1/2}$ and apply de Jong and Davidson's (2000: (A.51)) argument under the maintained kernel assumptions to deduce

$$\begin{aligned}\mathcal{B}_{1,n} \quad &\leq \quad K\int_{-\infty}^{\infty}\left(\frac{1}{||\mathcal{G}_n||}\left\|\frac{1}{n}\sum_{s=1}^{n}e^{-ims/\gamma_n}G_{n,s}(\theta_{n,*})\hat{I}_{n,s}(\theta_{n,*})\right\|\left\|\frac{1}{\mathfrak{S}_n^{1/2}}\sum_{t=1}^{n}e^{imt/\gamma_n}\hat{m}_{n,t}\right\|\right)|\varpi(m)|\,dm \\ &= \quad K\int_{-\infty}^{\infty}\mathcal{C}_n\mathcal{D}_n\,|\varpi(m)|\,dm,\end{aligned}$$

where $\varpi(m)$ is defined in Theorem 6.1. We have $\mathcal{C}_n = o_p(1)$ by Jacobian consistency Lemma B.7 in view of $\theta_{n,*} \xrightarrow{p} \theta^0$, and the Theorem 6.1 suppositions $\sum_{s,t=1}^{n}|\mathcal{K}_{n,s,t}| = o(n^2)$, $\max_{1\leq s\leq n}\sum_{t=1}^{n}|\mathcal{K}_{n,s,t}|$ $= o(n)$, and $\gamma_n = o(n)$. Further $\mathcal{D}_n = O_p(1)$ follows from approximation Lemma B.3.a and CLT Lemma B.6. Therefore $\mathcal{B}_{1,n} = o_p(1)$ by dominated convergence given the assumed kernel properties. By exploiting the arguments used in the proof of expansion Lemma B.5 the remaining terms follow. $\mathcal{QED}$.

We are now ready to prove Lemmas B.2-B.7, and implicitly invoke Assumptions D, F, MX, N, RB, RS and STM where ever needed. We present an expanded version of Lemma B.2 to aid in proofs that follow.

LEMMA B.2 (*variance and threshold bounds*). *a.* $||\mathcal{S}_n(\theta)|| = Kr_n(\theta)||\Sigma_n(\theta)|| = o(n/\ln(n))$ *for some sequence of positive numbers* $\{r_n(\theta)\}$, *where in general* $\liminf_{n\to\infty}\inf_{\theta\in\Theta}r_n(\theta)$ > 0 *and* $\sup_{\theta\in\Theta}r_n(\theta) = O(\ln(n))$, *and* $r_n(\theta) \sim K$ *if* $m_t(\theta)$ *is finite dependent or each* $E[m_{i,t}^2(\theta)] < \infty$; *b.* $\Sigma_n(\theta) = o(n||E\left[m_{n,t}^*(\theta)\right]||^2)$ *and* $\sup_{\theta\in\Theta}||\Sigma_n(\theta)|| = o(n\sup_{\theta\in\Theta}||E[m_{n,t}^*(\theta)]||^2)$; *c. for each* $w_t(\theta) \in \{m_{i,t}(\theta), G_{i,j,t}(\theta)\}$ *with index* $\kappa_w(\theta) < 2$ *we have* $\sup_{\theta\in\{\Theta:\kappa_w(\theta)\leq 2\}}\{\mathcal{C}_n(\theta)/(E[w_t^2(\theta)I_{n,t}^{(w)}(\theta)])^{1/2}\} = o((n/k_n)^{1/2})$.

PROOF.

Claim (a): Assume $m_t(\theta)$ is a mean zero scalar to reduce notation, and drop θ everywhere since it does not play a role.

Note $\mathcal{S}_n \sim E[m_{n,t}^{*2}] + 2\sum_{i=1}^{n-1}(1 - i/n)E[m_{n,1}^*m_{n,i+1}^*]$. If $E[m_t^2] < \infty$ then $\mathcal{S}_n \sim K$ under geometric β-mixing (c. Ibragimov 1962). If m_t is finite dependent then $\mathcal{S}_n = KE[m_{n,t}^{*2}]$ has a finite limit if $E[m_t^2] < \infty$.

Finally, assume m_t is non-finite dependent and $E[m_t^2] = \infty$ and recall the Assumption D tail index $\kappa \in (1, 2]$. By Assumption N $\liminf_{n\rightarrow\infty} \mathcal{S}_n/E[m_{n,t}^{*2}] > 0$ so we need only prove $\mathcal{S}_n = K\ln(n) \times E[m_{n,t}^{*2}]$ and $E[m_{n,t}^{*2}] = o(n/\ln(n))$.

Define the quantile functions $Q_n(u) = \inf\{m : P(|m_{n,t}^*| > m) \leq u\}$ and $Q(u) = \inf\{m : P(|m_t| > m) \leq u\}$ for $u \in [0,1]$, recall geometric β-mixing implies α-mixing with coefficients $\alpha_h \leq K\rho^h$ for $\rho \in (0,1)$. Since we assume non-finite dependence, without loss of generality set $\alpha_h = \rho^h$ for notational simplicity. By Theorem 1.1 of Rio (1993)

$$\sum_{i=1}^{n-1} \left|E\left[m_{n,1}^* m_{n,i+1}^*\right]\right| \leq 2\sum_{i=1}^{n-1}\int_0^{\alpha_h} Q_n^2(u)du \leq 2\sum_{i=1}^{n-1}\int_0^{\rho^h} Q_n^2(u)du.$$

By tail-trimming $Q_n(u) = 0\ \forall u \leq k_n/n$ and $Q_n(u) = Q(u)$ otherwise. Further, under tail decay Assumption D $Q(u) = O(u^{-1/\kappa})$. Therefore

$$\begin{aligned}\sum_{i=1}^{n-1} \left|E\left[m_{n,1}^* m_{n,i+1}^*\right]\right| &\leq K\sum_{i=1}^{n-1}\int_{k_n/n}^{\rho^h} u^{-2/\kappa}du = K\sum_{i=1}^{n-1}\max\left\{0, (n/k_n)^{(2/\kappa-1)} - \rho^{-i(2/\kappa-1)}\right\} \\ &= K\sum_{i=1}^{\ln(n/k_n)}\left\{(n/k_n)^{(2/\kappa-1)} - \rho^{-i(2/\kappa-1)}\right\}.\end{aligned}$$

Finally, $\sum_{i=1}^{\ln(n/k_n)}\{(n/k_n)^{2/\kappa-1} - \rho^{-i(2/\kappa-1)}\} = K\ln(n/k_n) \times (n/k_n)^{2/\kappa-1}(1+O(1))$. If $\kappa \in (1,2)$ then by Assumption D and therefore Karamata's Theorem (Resnick 1987: Theorem 0.6) $E[m_{n,t}^{*2}] \sim Kc_n^2(k_n/n) \sim K(n/k_n)^{2/\kappa-1}$, hence $\sum_{i=1}^{n-1}|E[m_{n,1}^* m_{n,i+1}^*]| \leq K\ln(n) \times E[m_{n,t}^{*2}]$. Similarly if $\kappa = 2$ then $E[m_{n,t}^{*2}] \sim L(n)$ hence $\sum_{i=1}^{n-1}|E[m_{n,1}^* m_{n,i+1}^*]| \leq K(\ln(n)/L(n)) \times E[m_{n,t}^{*2}] = L(n) \times E[m_{n,t}^{*2}]$.

By Karamata's Theorem if $\kappa = 2$ then $E[m_{n,t}^{*2}]$ is slowly varying in n hence $E[m_{n,t}^{*2}] = o(n/\ln(n))$. If $\kappa \in (1,2)$ then $E[m_{n,t}^{*2}] \sim Kc_n^2(k_n/n) = K(n/k_n)^{2/\kappa-1}$ where $(n/k_n)^{2/\kappa-1} = o(n/\ln n)$.

Claim (b): Define $\underline{\kappa} := \inf_{\theta\in\Theta}\kappa_{m_i}(\theta)$ where $\kappa_{m_i}(\theta)$ is the moment supremum of $m_{i,t}(\theta)$. By (UFT), (UTM) and the claim (a) argument if $\underline{\kappa} > 1$ then $\sup_{\theta\in\Theta} E[m_{i,n,t}^2(\theta)] = o(n) = o(n|E[m_{i,n,t}(\theta)]|)$. If $\underline{\kappa} \leq 1$ then apply (UTM) to $m_{i,n,t}(\theta)$ to deduce $\sup_{\theta\in\Theta}|E[m_{i,n,t}(\theta)]| \sim (n/k_{i,n})^{1/\underline{\kappa}-1}$

$$\frac{\sup_{\theta\in\Theta} E\left[m_{i,n,t}^2(\theta)\right]}{\sup_{\theta\in\Theta}\left|E\left[m_{i,n,t}(\theta)\right]\right|^2} \sim K\frac{(n/k_{i,n})^{2/\underline{\kappa}-1}}{(n/k_{i,n})^{2(1/\underline{\kappa}-1)}} = (n/k_{i,n}) = o(n).$$

The same argument implies $E[m_{i,n,t}^2(\theta)] = o(n|E[m_{i,n,t}(\theta)]|)$.

Claim (c): The claims follow from (UFT) and (UTM) under Assumption D. $\mathcal{QED}$.

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

PROOF. Assume $m_t(\theta)$ and θ are scalars for notational convenience, hence $\Sigma_n(\theta) = E[m_{n,t}^2(\theta)]$.

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

$$\sup_{\theta\in\Theta}\left\{\left|\frac{1}{k_n^{1/2}}\sum_{t=1}^n\left\{\bar{I}_{n,t}^{(m)}(\theta) - E\left[\bar{I}_{n,t}^{(m)}(\theta)\right]\right\}\right|\right\} = O_p(1), \quad \sup_{\theta\in\Theta}\left\{\left|\frac{n}{k_n}E\left[\bar{I}_{n,t}^{(m)}(\theta)\right] - 1\right|\right\} = 1. \tag{31}$$

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

Bound

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

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

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

Next, by construction and the triangle inequality

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

since $(n/k_n)E[\bar{I}_{n,t}] - 1 = 0$ by the threshold construction, and $k_n^{-1/2}\sum_{t=1}^{n}\{\bar{I}_{n,t} - E\left[\bar{I}_{n,t}\right]\} = O_p(1)$ by (31). Therefore $\sum_{t=1}^{n}\{\hat{m}_{n,t} - m_{n,t}\} = o_p(n^{-1/2}\mathcal{S}_n^{1/2})$.

Claim (b): Define

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

By the Claim (a) argument, $(n/k_n)E[\bar{I}_{n,t}(\theta)] - 1 = 0$ and uniform indicator law Lemma C.1 it follows

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

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

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

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

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

Claim (c): Observe

$$\left|\sum_{t=1}^{n}\{\hat{m}_{n,t}(\theta) - m_{n,t}(\theta)\}\right| \leq \max_{1\leq t\leq n}\left\{\left|m_t(\theta)\left\{\hat{I}_{n,t}(\theta) - I_{n,t}(\theta)\right\}\right|\right\} \times \sum_{t=1}^{n}\left|\hat{I}_{n,t}(\theta) - I_{n,t}(\theta)\right|.$$

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

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

which is $O_p(k_n^{1/2})$ by (31). Hence

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

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

LEMMA B.4 (*LLN and ULLN*). *a.* $1/n\sum_{t=1}^{n} m_{n,t} = o_p(1)$; *b.* $\sup_\theta\{||m_n(\theta) - E[m_{n,t}(\theta)]||\} = o_p(\mathfrak{e}_n)$.

PROOF.

Claim (a): $1/n\sum_{t=1}^{n} m_{n,t}(\theta^0) = o_p(1)$ follows from $E[m_{n,t}(\theta^0)] \to 0$, the Lemma B.2.a,b variance property $||\mathcal{S}_n|| = o(n)$, and Chebyshev's inequality.

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

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

Use variance relations Lemma B.2.a,b to deduce $1/n\sum_{t=1}^{n} h_{n,t}(\theta) = o_p(1)$ by Chebyshev's inequality. Further, $h_{n,t}(\theta)$ is uniformly L_1-bounded so it belongs to a separable Banach space, hence the L_1-bracketing numbers satisfy $N_{[\,]}(\varepsilon,\Theta,||\cdot||_1) < \infty$ (Dudley 1999: Proposition 7.1.7). Now combine the pointwise law and $N_{[\,]}(\varepsilon,\Theta,||\cdot||_1) < \infty$ to deduce $\sup_\theta|1/n\sum_{t=1}^{n} h_{n,t}(\theta)| = o_p(1)$ by, e.g., Theorem 7.1.5 of Dudley (1999). $\mathcal{QED}$.

LEMMA B.5 (*asymptotic expansion*). *Choose $\theta,\tilde{\theta} \in \Theta$ and let finite $\xi > 0$ be arbitrarily large. For $o_p(1)$ not a function of θ: a. $m_n(\theta) = m_n(\tilde{\theta}) + \mathcal{G}_n(\theta) \times (\theta - \tilde{\theta}) + n^{-\xi} \times ||\theta - \tilde{\theta}||^{1/\iota} \times o_p(1)$; and b. $\hat{m}_n(\theta) = \hat{m}_n(\tilde{\theta}) + \widehat{\mathcal{G}}_n(\theta) \times (\theta - \tilde{\theta}) + n^{-\xi} \times ||\theta - \tilde{\theta}||^{1/\iota} \times o_p(1)$.*

PROOF. We prove (a), the more difficult of the two. Assume θ and $m_t(\theta)$ are scalars to reduce notation. Choose $||\theta - \tilde{\theta}|| \le \delta$ for any $\delta > 0$. By Taylor's theorem for some $||\theta_{n,\delta} - \tilde{\theta}|| \le ||\theta - \tilde{\theta}||$

$$m_{n,t}(\theta) = \left\{m_t(\tilde{\theta}) + G_t(\theta_{n,\delta})(\theta - \tilde{\theta})\right\} \times I_{n,t}(\theta),$$

hence

$$(32)\qquad m_n(\theta) - m_n(\tilde{\theta}) = \widehat{\tilde{\mathcal{G}}}_n(\theta_{n,\delta}) \times (\theta - \tilde{\theta}) + \frac{1}{n}\sum_{t=1}^{n} m_t(\theta) \times \left\{I_{n,t}(\theta) - I_{n,t}(\tilde{\theta})\right\} + \frac{1}{n}\sum_{t=1}^{n} G_t(\theta_{n,\delta}) \times \{I_{n,t}(\theta) - I_{n,t}(\theta_{n,\delta})\} \times (\theta - \tilde{\theta}).$$

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

Consider the second term in (32) and use $I_{n,t}(\theta) - I_{n,t}(\tilde{\theta}) \in \{-1, 0, 1\}$ to bound

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

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

$$\text{(33)} \qquad \sup_{\theta,\tilde{\theta}\in\Theta} E\left|I_{n,t}(\theta)-I_{n,t}(\tilde{\theta})\right| \leq \sup_{\theta,\tilde{\theta}\in\Theta} E\left|I_{n,t}^{(m)}(\theta)\right| \leq \sup_{\theta,\tilde{\theta}\in\Theta} P\left(|m_t(\theta)| \leq \mathcal{C}_n(\theta)\right) = O\left(k_n/n\right).$$

Now use stationarity, L_ι-boundedness of $m_t(\theta)$ for some tiny $\iota > 0$ under D.i, compactness of Θ and the Cauchy-Schwartz inequality to deduce

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

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

Now, by D.i $m_t(\theta)$ is uniformly L_ι-bounded with a continuous distribution, hence by (33) and the mean-value-theorem $E[|I_{n,t}(\theta) - I_{n,t}(\tilde{\theta})|^\iota] = O(k_n/n) \times ||\theta - \tilde{\theta}||$, and $m_t(\theta)$ is L_ι-bounded. But this ensures for tiny $\iota > 0$

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

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

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

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

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

The same argument applies to the third term in (32) in lieu of Jacobian distribution and moment properties D.i. $\mathcal{QED}$.

LEMMA B.6 (*CLT*). $\quad n^{-1/2}\mathcal{S}_n^{-1/2}\sum_{t=1}^n m_{n,t} \xrightarrow{d} N(0, I_q)$.

PROOF. We prove the claim for the scalar case under symmetric trimming and $E[m_{n,t}] = 0$, the general case being similar. Write $k_n = k_n^{(m)}$, $\mathcal{C}_n = \mathcal{C}_n^{(m)}$ and $\mathfrak{S}_n^2 := E(\sum_{t=1}^n m_{n,t})^2$.

By the geometric β-mixing property and stationary, we need only verify (2.1) and (2.2) in Pelgrad (1996: Theorem 2.1), stated below:

$$\text{(2.1)} \quad \sup_{n\geq 1}\frac{1}{\mathfrak{S}_n^2}\sum_{t=1}^n E\left[m_{n,t}^2\right] < \infty \qquad \text{(2.2)} \quad \frac{1}{\mathfrak{S}_n^2}\sum_{t=1}^n E\left[m_{n,t}^2 I\left(|m_{n,t}| > \varepsilon\mathfrak{S}_n\right)\right] \to 0.$$

By non-degeneracy Assumption N and stationarity $nE[m_{n,t}^{*2}]/\mathcal{S}_n = O(1)$, hence (2.1).

By stationarity for (2.2) we must how $E[m_{n,t}^2 I(|m_{n,t}| > \varepsilon\mathfrak{S}_n)]/\mathfrak{S}_n^2 \to 0$. Assume $E[m_{n,t}] = 0$ to reduce notation, and observe since $nE[m_{n,t}^2]/\mathfrak{S}_n^2 = O(1)$ it follows

$$\begin{aligned} E\left[m_{n,t}^2 I\left(|m_{n,t}| > \varepsilon\mathfrak{S}_n\right)\right] &= E\left[m_t^2 I\left(|m_t| \leq \mathcal{C}_n\right) I\left(|m_t| I\left(|m_t| \leq \mathcal{C}_n\right) > \varepsilon\mathfrak{S}_n\right)\right] \\ &= E\left[m_t^2 I\left(\varepsilon\mathfrak{S}_n \leq |m_t| \leq \mathcal{C}_n\right)\right] \leq E\left[m_t^2 I\left(\varepsilon n^{1/2}\left(E\left[m_{n,t}^2\right]\right)^{1/2} \leq |m_t| \leq \mathcal{C}_n\right)\right]. \end{aligned}$$

Write $(z)_+ := \max(0, z)$. If $\kappa_m \in (0, 2)$ then by Karamata's Theorem $E[m_{n,t}^2] \sim \mathcal{C}_n^2 P(|m_t| > \mathcal{C}_n)$ $\sim \mathcal{C}_n^2(k_n/n)$. Therefore

$$\frac{n^{1/2}\left(E\left[m_{n,t}^2\right]\right)^{1/2}}{\mathcal{C}_n} = \left(\frac{nE\left[m_{n,t}^2\right]}{\mathcal{C}_n^2}\right)^{1/2} \sim \left(\frac{n\mathcal{C}_n^2\left(k_n/n\right)}{\mathcal{C}_n^2}\right)^{1/2} = k_n^{1/2} \rightarrow \infty.$$

This implies there exists an $N \in \mathbb{N}$ sufficiently large that $\forall n \geq N$ we have $E[m_t^2 I(\varepsilon n^{1/2}(E[m_{n,t}^2])^{1/2}$ $\leq |m_t| \leq \mathcal{C}_n)] = 0$. If $\kappa = 2$ then $E[m_{n,t}^2] = L(n) \rightarrow \infty$ is slowly varying and $\mathcal{C}_n = K(n/k_n)$, hence

$$\frac{n^{1/2}\left(E\left[m_{n,t}^2\right]\right)^{1/2}}{\mathcal{C}_n} \sim \left(\frac{nL(n)}{\mathcal{C}_n^2}\right)^{1/2} = \left(\frac{nL(n)}{n/k_n}\right)^{1/2} = \left(k_n L(n)\right)^{1/2} \rightarrow \infty. \mathcal{QED}.$$

LEMMA B.7 (*Jacobian consistency*). *a.* $\widehat{\mathcal{G}}_n^*(\hat{\theta}_n^*) = \mathcal{G}_n(1 + o_p(1))$; *b.* $\check{\mathcal{G}}_n^* = \mathcal{G}_n \times (1 + o(1))$.

PROOF.

Claim (a): We need only show $\widehat{\mathcal{G}}_n(\hat{\theta}_n) = \tilde{\mathcal{G}}_n(\hat{\theta}_n) \times (1 + o_p(1))$ where $o_p(1)$ is not a function of θ, and $||\tilde{\mathcal{G}}_n(\hat{\theta}_n) - \mathcal{G}_n|| = o_p(||\mathcal{G}_n||)$. The former holds by the same argument used to prove Lemma B.3.c. Drop θ^0 throughout. We have for the latter

$$\begin{aligned}&\left\|\frac{1}{n}\sum_{t=1}^n G_t(\hat{\theta}_n)I_{n,t}(\hat{\theta}_n) - E\left[G_t I_{n,t}\right]\right\| \\ &\qquad \leq \left\|\frac{1}{n}\sum_{t=1}^n \left\{G_t I_{n,t} - E\left[G_t I_{n,t}\right]\right\}\right\| + \left\|\frac{1}{n}\sum_{t=1}^n \left\{G_t(\hat{\theta}_n)I_{n,t}(\hat{\theta}_n) - G_t I_{n,t}\right\}\right\| \\ &\qquad = \mathcal{A}_n + \mathcal{B}_n(\hat{\theta}_n).\end{aligned}$$

If any $G_{i,j,t}$ is $L_{1+\iota}$-bounded (or merely uniformly integrable) then by geometric β-mixing Assumption MX $\{G_{i,j,t}I_{n,t}\}$ satisfies Andrews' (1988) L_1-mixingale LLN: $\mathfrak{G}_{i,j,n} := 1/n\sum_{t=1}^n\{G_{i,j,t}I_{n,t}$ $- E[G_{i,j,t}I_{n,t}]\} \stackrel{p}{\rightarrow} 0$. Otherwise $\mathfrak{G}_n = o_p(||E[G_t I_{n,t}]||)$ can be shown by exploiting geometric β-mixing and using the argument in the proof of ULLN Lemma B.4.b. In both cases, therefore $\mathcal{A}_n = o_p(||E[G_t I_{n,t}]||)$.

Finally, arguments from the proof of Lemma B.5 can be generalized to show $\mathcal{B}_n(\hat{\theta}_n) \leq K||\hat{\theta}_n - \theta^0||^{1/\iota}$ for tiny $\iota > 0$. Therefore $\mathcal{B}_n(\hat{\theta}_n) = o_p(1)$ given consistency $\hat{\theta}_n \stackrel{p}{\rightarrow} \theta^0$ by Theorem 2.1. In lieu of non-degeneracy Assumption N the proof is complete: $||\tilde{\mathcal{G}}_n(\hat{\theta}_n) - \mathcal{G}_n|| = o_p(1) + o_p(||\mathcal{G}_n||)$ $= o_p(||\mathcal{G}_n||)$.

Claim (b): By expansion Lemma B.5.a we have for arbitrarily large finite $\xi > 0$, and tiny $\iota > 0$,

$$\frac{1}{n}\sum_{t=1}^n \left\{m_{n,t}(\theta) - m_{n,t}\right\} = \frac{1}{n}\sum_{t=1}^n G_t I_{n,t} \times \left(\theta - \theta^0\right) + n^{-\xi} \times \left\|\theta - \tilde{\theta}\right\|^{1/\iota} \times o_p(1).$$

Now invoke dominated convergence to deduce

$$\frac{E\left[m_{n,t}\left(\theta\right)\right] - E\left[m_{n,t}\right]}{\left\|\theta - \theta^0\right\|} = E\left[G_t I_{n,t}\right] \times \left(1 + o\left(\left\|\theta - \theta^0\right\|\right)\right) + o\left(n^{-\xi}\right) = \mathcal{G}_n \times \left(1 + o\left(\left\|\theta - \theta^0\right\|\right)\right) + o\left(n^{-\xi}\right).$$

Since $\xi > 0$ is arbitrary and $||\mathcal{G}_n||$ is non-degenerate under Assumption N, it follows

$$\frac{E\left[m_{n,t}\left(\theta\right)\right] - E\left[m_{n,t}\right]}{\left\|\theta - \theta^0\right\|} = \mathcal{G}_n \times \left(1 + o\left(\left\|\theta - \theta^0\right\|\right)\right) + o\left(\left\|\mathcal{G}_n\right\|\right).$$

Identically, by the definition of a derivative $\check{\mathcal{G}}_n^*(\theta) := (\partial/\partial\theta)E[m_{n,t}(\theta)]$:

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

Equate the right hand sides of each identity and take $||\theta - \theta^0|| \rightarrow 0$ to prove the claim: $\check{\mathcal{G}}_n^* = \mathcal{G}_n \times (1 + o(1))$. $\mathcal{QED}$.

LEMMA B.8 (*HAC*). $\widehat{\mathcal{S}}_n(\hat{\theta}_n) = \mathcal{S}_n(1 + o_p(1))$.

PROOF. Assume $m_t(\theta)$ and θ are scalars to simplify notation, drop θ^0, and define

$$\mathcal{A}_{1,n} := \frac{1}{\mathcal{S}_n}\sum_{s,t=1}^{n}\mathcal{K}_{n,s,t}\left\{\hat{m}_{n,s}(\hat{\theta}_n)\hat{m}_{n,t}(\hat{\theta}_n)-m_{n,s}m_{n,t}\right\}\ \text{ and }\ \mathcal{A}_{2,n} := \sum_{s,t=1}^{n}\mathcal{K}_{n,s,t}\frac{m_{n,s}m_{n,t}}{\mathcal{S}_n}-1.$$

Since $\mathcal{A}_{1,n} \stackrel{p}{\rightarrow} 0$ by Lemma C.3, we need only show $\mathcal{A}_{2,n} \stackrel{p}{\rightarrow} 0$ and invoke the triangle inequality to prove the claim. We will apply Theorem 2.1 of de Jong and Davidson (2000), denoted DJ. It suffices to verify their Assumptions 1-3. Assumption 1 holds by our maintained kernel assumptions.

Their Assumptions 2 and 3 impose Near Epoch Dependence and relate the property to bandwidth γ_n. Both conditions are only used to promote partial sum variance bounds for a standardized process by invoking McLeish's (1975: Theorem 1.6) maximal inequality.

Define $\mathcal{Z}_{n,t} : = n^{-1/2}\mathcal{S}_n^{-1}m_{n,t}(\theta^0)$. Under geometric β-mixing $\{m_{n,t}(\theta^0), \Im_t\}$ forms a geometric L_2-mixingale with constants $e_{n,t}$ (cf. McLeish 1975: Theorem 2.1)[18]. Therefore $\{\mathcal{Z}_{n,t}, \Im_t\}$ forms a geometric L_2-mixingale with constants $\mathcal{E}_{n,t} := n^{-1/2}\mathcal{S}_n^{-1/2}e_{n,t}$. By inspection of DJ's proof of their Theorem 2.1 it follows $E(\sum_{t=1}^n \mathcal{Z}_{n,t})^2 = 1 \leq K\sum_{t=1}^n(1/n^{1/2})^2 = K$ suffices in place of their Assumption 2. Finally, their Assumption 3 is $\gamma_n \times \max_{1\leq t\leq n}\{\mathcal{E}_{n,t}^2\} = o(1)$. By supposition $\gamma_n = o(n)$. Theorem 1.6 in McLeish (1975) states $1 = E(\sum_{t=1}^n \mathcal{Z}_{n,t})^2 \leq K\sum_{t=1}^n \mathcal{E}_{n,t}^2 = Kn^{-1}\mathcal{S}_n^{-1}\sum_{t=1}^n e_{n,t}^2$. Therefore, in view of stationarity we we assume $e_{n,t} = K\mathcal{S}_n^{1/2}$, hence $\max_{1\leq t\leq n}\{\mathcal{E}_{n,t}^2\} = n^{-1}$. Thus Assumption 3 holds. $\mathcal{QED}$.

REFERENCES

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

Dedecker, J. and P. Doukhan (2003). A New Covariance Inequality and Applications, Stoch. Proc. Appl. 106, 63-80.

Doukhan, P., P. Massart, and E. Rio (1995). Invariance Principles for Absolutely Regular Empirical Processes, Ann. Inst. H. P. 31, 393-427.

Dudley, R. M. (1999). Uniform Central Limit Theorems. Cambridge University Press: Cambridge.

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

Hill, J.B. (2009). On Functional Central Limit Theorems for Dependent, Heterogeneous Arrays with Applications to Tail Index and Tail Dependence Estimation, J. Statist. Plann. Infer. 139, 2091-2110.

Hsing, T. (1991). On Tail Index Estimation Using Dependent Data, Ann. Stat. 19, 1547-1569.

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

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

Peligrad, M. (1996). On the Asymptotic Normality of Sequences of Weak Dependent Random Variables, J. Theor. Prob. 9, 703-715.

Pollard, D. (1989). Bracketing Methods in Statistics and Econometrics, in Proceedings of the Conference on Nonparametric and Semiparametric Methods in Econometrics and Statistics by W.A. Barnett, J. Powell and G. Tauchen (eds.), Cambridge University Press: Cambridge.

van der Vaart, A. and J. Wellner (1996). Weak Convergence and Empirical Processes, Springer: New York.