C

Technical Appendix for "Tail Index Estimation for a Filtered Dependent Time Series"

Jonathan B. Hill

Dept. of Economics, University of North Carolina

This appendix contains technical proofs. We first present the assumptions and re-state the results for reference.

1 Assumptions

Assumption 1 (Smoothness and Moments).

a. Let $\{\Im_t\}_{t\in\mathbb{Z}}$ *be a sequence of* σ*-fields that do not depend on* θ *and define* $\mathcal{F} := \sigma(\cup_{t\in\mathbb{Z}}\Im_t)$. $x_t(\theta)$ *lies on a complete probability measure space* $(\Omega, \mathcal{F}, P)$ *and is* $\Im_t$*-measurable. All functions of* $x_t(\theta)$ *satisfy Pollard (1984: Appendix C)'s permissibility criteria.*

b. $x_t(\theta)$ *is stationary, ergodic and thrice continuously differentiable with* $\Im_t$*-measurable stationary and ergodic derivatives* $g_t(\theta)$ *and* $h_t(\theta)$.

c. Each $w_t(\theta) \in \{x_t(\theta), g_{i,t}(\theta), h_{i,j,t}(\theta)\}$ *is governed by a non-degenerate distribution that is absolutely continuous with respect to Lebesgue measure, with uniformly bounded derivatives:* $\sup_{\theta\in\Theta}\sup_{a\in\mathbb{R}}||(\partial/\partial\theta)P(w_t(\theta) \leq a)|| < \infty$ *and* $\sup_{\theta\in\Theta}\sup_{a\in\mathbb{R}}\{(\partial/\partial a)P(w_t(\theta) \leq a)\} < \infty$. *Further* $E[\sup_{\theta\in\Theta}|w_t(\theta)|^{\iota}] < \infty$ *for some tiny* $\iota > 0$.

We assume x_t has support $[0, \infty)$ and has for each t a common regularly varying distribution tail with tail index $\kappa > 0$:

$$P(x_t > a) = a^{-\kappa}\mathcal{L}(a) \text{ where } a > 0 \text{ and } \mathcal{L}(a) \text{ is slowly varying.} \qquad (1)$$

Assumption 2 (Regular Variation and Fractile Bound).

a. There exists a neighborhood $\mathcal{N}_0(\delta)$ *such that*

$$\lim_{a\to\infty}\sup_{\theta\in\mathcal{N}_0(\delta)}\left|\frac{a^{\kappa(\theta)}}{\mathcal{L}(a,\theta)}P(x_t(\theta)\ >\ a)-1\right|=0. \tag{2}$$

Note $\mathcal{L}(a,\theta^0)=\mathcal{L}(a)$ *in (1). The tail component* $\mathcal{L}(a,\theta)$ *is slowly varying with remainder in a, uniformly on* Θ, *that is* $\sup_{\theta\in\mathcal{N}_0(\delta)}|\mathcal{L}(\lambda a,\theta)/\mathcal{L}(a,\theta)-1|=O(h(a))$ *as* $a\to\infty$ *for any* $\lambda>0$ *where h is a measurable function on* $(0,\infty)$ *with bounded increase: there exist* $0<D,z_0<\infty$, *and* $\tau\leq 0$ *such that* $h(\vartheta z)/h(z)\leq D\vartheta\tau$ *some for* $\vartheta\geq 1$ *and* $z\geq z_0$ *Goldie and Smith (1987). Further* $m_n^{1/2}h(c_n)\to 0$. *Moreover, the tail index* $\kappa(\theta)$ *is locally bounded* $\inf_{\theta\in\mathcal{N}_0(\delta)}\kappa(\theta)>0$ *and* $\sup_{\theta\in\mathcal{N}_0(\delta)}\kappa(\theta)<\infty$, *and is twice differentiable with locally bounded derivatives and a Lipschitz first derivative:* $||(\partial/\partial\theta)\kappa(\theta)||<\infty$, $||(\partial/\partial\theta)^2\kappa(\theta)||<\infty$, *and* $||(\partial/\partial\theta)\kappa(\theta)-(\partial/\partial\theta)\kappa(\tilde{\theta})||\leq K||\theta-\tilde{\theta}||$ *for each* $\theta,\tilde{\theta}\in\mathcal{N}_0(\delta)$.

b. $m_n\to\infty$ *and* $m_n=o(n/\ln(n))$.

Assumption 3 (mixing). *Let* $\mathcal{N}_0(\delta)$ *be the neighborhood of* θ^0 *defined in Assumption 2.a. Then* $x_t(\theta)$ *is* β*-mixing for each* $\theta\in\mathcal{N}_0(\delta)$ *with summable coefficients. Hence* $\beta_l:=\sup_{\mathcal{A}\subset\Im_{t+l}^{+\infty}}E|P(\mathcal{A}|\Im_{-\infty}^{t})-P(\mathcal{A})|$ *where* $\sum_{l=1}^{\infty}\beta_l<\infty$.

Assumption 4 (Plug-In). *There exists a unique point* $\theta^0\in\Theta$ *such that* $m_n^{1/2}\ln(n)(\hat{\theta}_n-\theta^0)=o_p(1)$.

2 Main Results

The main result Theorem 2.1 is proved in Section 2.1.

Theorem 2.1 *Under Assumptions 1-4* $m_n^{1/2}(\hat{\kappa}_{m_n}^{-1}(\hat{\theta}_n)-\kappa^{-1})/\sigma_{m_n}\overset{d}{\to}N(0,1)$ *where* $\sigma_{m_n}^2:=E(m_n^{1/2}(\hat{\kappa}_{m_n}^{-1}-\kappa^{-1}))^2$.

The main supporting lemma is proved in Section 2.2. Recall

$$\mathcal{I}_{n,t}(\theta):=\left(\frac{n}{m_n}\right)^{1/2}\left\{I\left(|x_t(\theta)|\leq c_n(\theta)\right)-E\left[I\left(|x_t(\theta)|\leq c_n(\theta)\right)\right]\right\}.$$

Lemma 2.2 *Under Assumptions 1-3 there exists a Gaussian process* $\{\mathcal{I}(\theta):\theta\in\mathcal{N}_0(\delta)\}$ *with uniformly bounded and uniformly continuous sample paths with* *respect to* $||\cdot||_2$ *such that:*

a. $\{n^{-1/2}\sum_{t=1}^{n}\mathcal{I}_{n,t}(\theta):\theta\in\mathcal{N}_0(\delta)\}\Longrightarrow^{*}\{\mathcal{I}(\theta):\theta\in\mathcal{N}_0(\delta)\}$.

b. $\sup_{\theta\in\mathcal{N}_0(\delta)}|m_n^{1/2}\ln(x_{(m_n+1)}(\theta)/c_n(\theta))-\kappa^{-1}n^{-1/2}\sum_{t=1}^{n}\mathcal{I}_{n,t}(\theta)|\overset{p}{\to}0$.

c. $\{m_n^{1/2}\ln(x_{(m_n+1)}(\theta)/c_n(\theta)):\theta\in\mathcal{N}_0(\delta)\}\Longrightarrow^{*}\{\kappa^{-1}\mathcal{I}(\theta):\theta\in\mathcal{N}_0(\delta)\}$.

2.1 Proof of Theorem 2.1

In order to prove Theorem 2.1 we require several preliminary results. Drop θ^0 and write $x_t=x_t(\theta^0)$, $c_n=c_n(\theta^0)$, $\hat{\kappa}_{m_n}=\hat{\kappa}_{m_n}(\theta^0)$. Throughout $\mathcal{N}_0(\delta)$ denotes the neighborhood of $\theta^0\in\Theta\subset\mathbb{R}^k$ defined by Assumption 2.a. Also, for two sequences of real numbers $\{a_n\}$ and $\{b_n\}$ we write $a_n\sim b_n$ to imply $a_n/b_n\to 1$ (or $a_n\to 0$ if $b_n=0\ \forall n$).

Define indicator functions

$$\hat{I}_{n,t}(\theta):=I\left(x_t(\theta)\geq x_{(m_n+1)}(\theta)\right)\quad\text{and}\quad I_{n,t}(\theta):=I\left(x_t(\theta)\geq c_n(\theta)\right)$$

and sample and population Jacobia

$$\widehat{\mathcal{J}}_n(\theta):=\frac{1}{m_n}\sum_{t=1}^{n}\frac{\partial}{\partial\theta}\ln\left(x_t(\theta)\right)\times\hat{I}_{n,t}(\theta)$$

$$\mathcal{J}_n(\theta):=\frac{n}{m_n}E\left[\frac{\partial}{\partial\theta}\ln\left(x_t(\theta)\right)\times I_{n,t}(\theta)\right].$$

Lemma 2.3 *Let* $\theta,\tilde{\theta}\in\mathcal{N}_0(\delta)$ *be arbitrary.*

a. $1/m_n^{1/2}\sum_{t=1}^{n}\{\ln(x_t(\theta))\hat{I}_{n,t}(\theta)-\ln(x_t(\tilde{\theta}))\hat{I}_{n,t}(\tilde{\theta})\}=m_n^{1/2}\widehat{\mathcal{J}}_n(\theta_*)'(\theta-\tilde{\theta})+o_p(1)$ *where* θ_* *satisfies* $||\theta_*-\tilde{\theta}||\leq||\theta-\tilde{\theta}||$.

b. $1/m_n^{1/2}\sum_{t=1}^{n}\{I_{n,t}(\theta)-I_{n,t}(\tilde{\theta})\}=o_p(m_n^{1/2}||\theta-\tilde{\theta}||)$.

Proof.

Claim (a): Let $\theta,\tilde{\theta}\in\mathcal{N}_0(\delta)$, and define $y_t(\theta):=\ln(x_t(\theta))$, $\hat{y}_{n,t}(\theta):=y_t(\theta)\hat{I}_{n,t}(\theta)$, $y_{n,t}(\theta):=y_t(\theta)I_{n,t}(\theta)$, and $J_t(\theta)=[J_{i,t}(\theta)]:=(\partial/\partial\theta)y_t(\theta)$. By power law Assumption 2.a $y_t(\theta)$ is uniformly L_1-bounded, and by Corollary 2.6, below, $J_{i,t}(\theta)$ is uniformly L_1-bounded.

By the Mean Value Theorem $\hat{y}_{n,t}(\theta) = \{y_t(\tilde{\theta}) + J_t(\theta_*)'(\theta - \tilde{\theta})\} \times \hat{I}_{n,t}(\theta)$ for some θ_* that satisfies $||\theta_* - \tilde{\theta}|| \leq ||\theta - \tilde{\theta}||$, hence

$$\frac{1}{m_n^{1/2}} \sum_{t=1}^{n} \hat{y}_{n,t}(\theta) - \frac{1}{m_n^{1/2}} \sum_{t=1}^{n} \hat{y}_{n,t}(\tilde{\theta})$$

$$= \frac{1}{m_n^{1/2}} \sum_{t=1}^{n} J_t(\theta_*)' \times \hat{I}_{n,t}(\theta_*) \times \left(\theta - \tilde{\theta}\right)$$

$$+ \frac{1}{m_n^{1/2}} \sum_{t=1}^{n} y_t(\theta) \times \left\{\hat{I}_{n,t}(\theta) - \hat{I}_{n,t}(\tilde{\theta})\right\}$$

$$+ \frac{1}{m_n^{1/2}} \sum_{t=1}^{n} J_t(\theta_*)' \times \left\{\hat{I}_{n,t}(\theta) - \hat{I}_{n,t}(\theta_*)\right\} \times (\theta - \tilde{\theta})$$

$$= \frac{1}{m_n^{1/2}} \sum_{t=1}^{n} J_t(\theta_*)' \times \hat{I}_{n,t}(\theta_*) \times (\theta - \tilde{\theta}) + \mathcal{E}_{1,n}(\theta, \tilde{\theta}) + \mathcal{E}_{2,n}(\theta, \tilde{\theta}).$$

It suffices to show $\mathcal{E}_{1,n}(\theta, \tilde{\theta}) = o_p(1)$ since a similar argument extends to $\mathcal{E}_{2,n}(\theta, \tilde{\theta})$.

The indicator function $I(u) := I(u \geq 0)$ can be approximated by a smooth regular sequence $\{\mathfrak{I}_n(u)\}$, cf. Lighthill (1958). Let $\{\mathcal{N}_n\}$ be a sequence of finite positive numbers, $\mathcal{N}_n \to \infty$, the rate to be chosen below. Define $\mathfrak{I}_n(u) := \int_{-\infty}^{\infty} I(\varpi)\mathrm{S}(\mathcal{N}_n(\varpi - u))\mathcal{N}_n e^{-\varpi^2/\mathcal{N}_n^2} d\varpi$ where $\mathrm{S}(\xi) = e^{-1/(1-\xi^2)} / \int_{-1}^{1} e^{-1/(1-w^2)} dw$ if $|\xi| < 1$ and $\mathrm{S}(\xi) = 0$ if $|\xi| \geq 1$. Note $\mathfrak{I}_n(u)$ is uniformly bounded in u, and continuous and differentiable. $I(u)$ is differentiable, except at 0, with derivative $\delta(u) = (\partial/\partial u)I(u) = 0 \; \forall u \neq 0$, the Dirac delta function, hence $\delta(u)$ has a regular sequence $\mathfrak{D}_n(u) := (\mathcal{N}_n/\pi)^{1/2} \exp\{-\mathcal{N}_n u^2\}$. See Lighthill (1958: p. 22).

Now define $\hat{\mathcal{X}}_{n,t}(\theta) := x_t(\theta) - x_{(m_n+1)}(\theta)$, hence by definition $\hat{I}_{n,t}(\theta) = I(\hat{\mathcal{X}}_{n,t}(\theta))$. Since the rate $\mathcal{N}_n \to \infty$ can be made to be as fast as we chose, it can be set to ensure

$$\frac{1}{m_n^{1/2}} \sum_{t=1}^{n} y_t(\theta) \times \left\{\hat{I}_{n,t}(\theta) - \hat{I}_{n,t}(\tilde{\theta})\right\}$$

$$= \frac{1}{m_n^{1/2}} \sum_{t=1}^{n} y_t(\theta) \times \left\{\mathfrak{I}_n(\hat{\mathcal{X}}_{n,t}(\theta)) - \mathfrak{I}_n(\hat{\mathcal{X}}_{n,t}(\tilde{\theta}))\right\} + o_p(1).$$

In view of $\sup_{\theta \in \mathcal{N}_0(\delta)} |x_{(m_n+1)}(\theta)/c_n(\theta) - 1| = O_p(1/m_n^{1/2})$ be Lemma 2.2 and the Mean Value Theorem, we may similarly write for some θ^*, $||\theta^* - \theta|| \leq ||\theta - \tilde{\theta}||$

$$\frac{1}{m_n^{1/2}} \sum_{t=1}^{n} y_t(\theta) \times \left\{ \mathfrak{I}_n(\hat{\mathcal{X}}_{n,t}(\theta)) - \mathfrak{I}_n(\hat{\mathcal{X}}_{n,t}(\tilde{\theta})) \right\}$$

$$= \frac{1}{m_n^{1/2}} \sum_{t=1}^{n} y_t(\theta) \times \mathfrak{D}_n(\hat{\mathcal{X}}_{n,t}(\theta^*)) \times \left(x_t(\theta) - x_t(\tilde{\theta}) \right)$$

$$+ O_p\left(\frac{1}{m_n} \sum_{t=1}^{n} y_t(\theta) \times \mathfrak{D}_n(\hat{\mathcal{X}}_{n,t}(\theta^*)) \times c_n(\theta) \right)$$

$$+ O_p\left(\frac{1}{m_n} \sum_{t=1}^{n} y_t(\theta) \times \mathfrak{D}_n(\hat{\mathcal{X}}_{n,t}(\theta^*)) \times c_n(\tilde{\theta}) \right)$$

$$- \frac{1}{m_n^{1/2}} \sum_{t=1}^{n} y_t(\theta) \times \mathfrak{D}_n(\hat{\mathcal{X}}_{n,t}(\theta^*)) \times \left(c_n(\theta) - c_n(\tilde{\theta}) \right) + o_p(1)$$

$$= \sum_{i=1}^{4} \mathcal{A}_{i,n}(\theta, \theta^*, \tilde{\theta}) + o_p(1).$$

Distribution continuity implies $\hat{\mathcal{X}}_{n,t}(\theta) \neq 0$ $a.s.$ for any $\theta \in \Theta$. Hence by construction the rate at which $\mathfrak{D}_n(\hat{\mathcal{X}}_{n,t}(\theta^*)) \xrightarrow{p} 0$ can be made so fast by choice of $\{\mathcal{N}_n\}$ that $E|\mathfrak{D}_n(\hat{\mathcal{X}}_{n,t}(\theta^*))|^{\iota} \rightarrow 0$ as fast as we choose for tiny $\iota > 0$ by dominated convergence. Therefore by Loéve and Hölder inequalities and dominated convergence, each $E|\mathcal{A}_{i,n}(\theta, \theta^*, \tilde{\theta})|^{\iota} \rightarrow 0$ for tiny $\iota > 0$, hence $\mathcal{A}_{i,n}(\theta, \theta^*, \tilde{\theta}) = o_p(1)$ by Markov's inequality.

Claim (b): Define $\mathcal{X}_{n,t}(\theta) := x_t(\theta) - c_n(\theta)$. Throughout θ^* satisfies $||\theta^* - \theta|| \leq ||\theta - \tilde{\theta}||$ and may be different in different places. Repeat the above argument to obtain for some θ^*

$$\frac{1}{m_n^{1/2}} \sum_{t=1}^{n} \left\{ I_{n,t}(\theta) - I_{n,t}(\tilde{\theta}) \right\}$$

$$= \frac{1}{m_n^{1/2}} \sum_{t=1}^{n} \mathfrak{D}_n(\mathcal{X}_{n,t}(\theta^*)) \times \left(x_t(\theta) - x_t(\tilde{\theta}) \right)$$

$$- \frac{1}{m_n^{1/2}} \sum_{t=1}^{n} \mathfrak{D}_n(\mathcal{X}_{n,t}(\theta^*)) \times \left(c_n(\theta) - c_n(\tilde{\theta}) \right) + o_p(1).$$

By smoothness Assumption 1.b,c and the mean-value theorem: $x_t(\theta) - x_t(\tilde{\theta}) = g_t(\theta^*)'(\theta - \tilde{\theta})$ where $g_t(\theta)$ is L_ι-bounded for tiny $\iota > 0$, and $c_n(\theta) - c_n(\tilde{\theta}) = d_n(\theta^*)'(\theta - \tilde{\theta})$ for some sequence of finite vectors $\{d_n(\theta^*)\}$ in $\mathbb{R}^k$. Hence

$$\left|\frac{1}{m_n^{1/2}}\sum_{t=1}^{n}\left\{I_{n,t}(\theta) - I_{n,t}(\tilde{\theta})\right\}\right|$$

$$\leq \frac{1}{m_n}\sum_{t=1}^{n}|\mathfrak{D}_n(\mathcal{X}_{n,t}(\theta^*))| \times \|g_t(\theta^*) - d_n(\theta^*)\| \times m_n^{1/2}\left\|\theta - \tilde{\theta}\right\| + o_p(1).$$

By the Claim (a) argument we can chose $\{\mathcal{N}_n\}$ to allow $\mathfrak{D}_n(\mathcal{X}_{n,t}(\theta^*)) \xrightarrow{p} 0$ so fast that $1/m_n\sum_{t=1}^{n}|\mathfrak{D}_n(\mathcal{X}_{n,t}(\theta^*))| \times ||g_t(\theta^*) - d_n(\theta^*)|| \xrightarrow{p} 0$, hence as claimed $1/m_n^{1/2}\sum_{t=1}^{n}\{I_{n,t}(\theta) - I_{n,t}(\tilde{\theta})\} = o_p(m_n^{1/2}||\theta - \tilde{\theta}||)$. $\mathcal{QED}$.

Lemma 2.4 $1/m_n^{1/2}\sum_{t=1}^{n}\ln(x_t)\{\hat{I}_{n,t} - I_{n,t}\} = o_p(1)$.

Proof. The proof follows from the same arguments used to prove Lemma 2.3. $\mathcal{QED}$.

Lemma 2.5 *Let $\{i_1, ..., i_d\}$ be an arbitrary set of $d \in \{1, 2, 3\}$ integers $i_j \in \{1, ..., k\}$, and let $\theta \in \mathcal{N}_0(\delta)$ be arbitrary.*

a. If $(\partial/\partial\theta)\kappa = 0$ then $\widehat{\mathcal{J}}_n(\hat{\theta}_n) \xrightarrow{p} 0$, and otherwise $\liminf_{n\to\infty}|\mathcal{J}_{i,n}| > 0$ and $\widehat{\mathcal{J}}_{i,n}(\hat{\theta}_n)/\mathcal{J}_{i,n} \xrightarrow{p} 1$ for each $i = 1.,,,.k$;

b. $\mathcal{J}_n(\theta) = -\ln(n)(\partial/\partial\theta)\ln(\kappa(\theta)) \times (1 + o(1))$ and $(\partial/\partial\theta_{i_1}\cdots\partial\theta_{i_d})\mathcal{J}_n(\theta) = -\ln(n)(\partial/\partial\theta_{i_1}\cdots\partial\theta_{i_d})\ln(\kappa(\theta)) \times (1 + o(1))$;

c. $\sup_{\theta\in\mathcal{N}_0(\delta)}||\mathcal{J}_n(\theta)|| \sim \ln(n) \times ||(\partial/\partial\theta)\ln\kappa(\theta)||$ and $\sup_{\theta\in\mathcal{N}_0(\delta)}|(\partial/\partial\theta_{i_1}\cdots\partial\theta_{i_d})\mathcal{J}_n(\theta)| \sim \ln(n) \times |(\partial/\partial\theta_{i_1}\cdots\partial\theta_{i_d})\ln(\kappa(\theta))|$.

Proof.

Claim (a): We exploit notation and arguments from the proof of Lemma 2.3, in particular $\mathfrak{I}_n(\cdot)$ and $\mathcal{N}_n$. Recall $\mathcal{J}_n(\theta) := (n/m_n)E[(\partial/\partial\theta)\ln(x_t(\theta))I_{n,t}(\theta)]$.

We have

$$\begin{aligned}
&\widehat{\mathcal{J}}_n(\hat{\theta}_n) - \mathcal{J}_n \\
&= \left(\frac{1}{n} \sum_{t=1}^{n} \frac{n}{m_n} \frac{\partial}{\partial \theta} \ln(x_t(\hat{\theta}_n)) \times I_{n,t}(\hat{\theta}_n) - \frac{n}{m_n} E\left[\frac{\partial}{\partial \theta} \ln(x_t) \times I_{n,t} \right] \right) \\
&\quad + \frac{1}{n} \sum_{t=1}^{n} \frac{n}{m_n} \frac{\partial}{\partial \theta} \ln(x_t(\hat{\theta}_n)) \times \left\{ \hat{I}_{n,t}(\hat{\theta}_n) - I_{n,t}(\hat{\theta}_n) \right\} \\
&= \mathcal{A}_n + \mathcal{B}_n.
\end{aligned}$$

The arguments used to prove Lemmas 2.3 and 2.4 can be straightforwardly generalized to show the second term is $o_p(1)$ in view of Assumption 4 $m_n^{1/2} \ln(n)(\hat{\theta}_n - \theta^0) = o_p(1)$, and the Lemma 2.2 result $\sup_{\theta \in \mathcal{N}_0(\delta)} |x_{(m_n+1)}(\theta)/c_n(\theta) - 1| = O_p(1/m_n^{1/2})$.

Consider the first term. We have:

$$\begin{aligned}
&\frac{1}{n} \sum_{t=1}^{n} \frac{n}{m_n} \left\{ \frac{\partial}{\partial \theta} \ln(x_t(\hat{\theta}_n)) \times I_{n,t}(\hat{\theta}_n) - E\left[\frac{\partial}{\partial \theta} \ln(x_t) \times I_{n,t} \right] \right\} \\
&= \frac{1}{m_n} \sum_{t=1}^{n} \left\{ \frac{\partial}{\partial \theta} \ln(x_t(\hat{\theta}_n)) \times \mathfrak{I}_n(\mathcal{X}_{n,t}(\hat{\theta}_n)) - E\left[\frac{\partial}{\partial \theta} \ln(x_t) \times \mathfrak{I}_n(\mathcal{X}_{n,t}) \right] \right\} \\
&\quad + o_p(1),
\end{aligned}$$

where $\mathcal{X}_{n,t}(\theta) := x_t(\theta) - c_n(\theta)$. Define

$$\mathcal{H}_{n,t}(\theta) = [\mathcal{H}_{i,j,n,t}(\theta)] := \left[\frac{n}{m_n} \frac{\partial}{\partial \theta_i} \frac{\partial}{\partial \theta_j} \ln(x_t(\theta)) \times \mathfrak{I}_n(\mathcal{X}_{n,t}(\theta)) \right]$$

$$\mathsf{J}_{n,t} = [\mathsf{J}_{i,n,t}] := \left[\frac{n}{m_n} \frac{\partial}{\partial \theta_i} \ln(x_t) \times \mathfrak{I}_n(\mathcal{X}_{n,t}) \right],$$

and note $||E[\mathsf{J}_{n,t}] - \mathcal{J}_n|| \to 0$ as fast as we choose by choice of $\{\mathcal{N}_n\}$. Now expand around θ^0. Use $m_n^{1/2} \ln(n)(\hat{\theta}_n - \theta^0) = o_p(1)$, and $(\partial/\partial\theta)\mathfrak{I}_n(\mathcal{X}_{n,t}(\theta^*)) \xrightarrow{p} 0$ as fast as we choose by choice of the sequence $\{\mathcal{N}_n\}$, to deduce for some θ^*, $||\theta^* - \theta^0||$

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

$$\frac{1}{m_n}\sum_{t=1}^{n}\frac{\partial}{\partial\theta}\ln(x_t(\hat{\theta}_n)) \times \Im_n(\mathcal{X}_{n,t}(\hat{\theta}_n)) \tag{3}$$

$$= \frac{1}{n}\sum_{t=1}^{n}\mathsf{J}_{n,t} + o_p\left(\left\|\frac{1}{m_n^{1/2}\ln(n)} \times \frac{1}{n}\sum_{t=1}^{n}\mathcal{H}_{n,t}(\theta^*)\right\|\right).$$

Consider the first term in (3). Suppose $(\partial/\partial\theta)\kappa = 0$. By (b) $\mathcal{J}_n = 0$ hence $E[\mathsf{J}_{n,t}] \rightarrow 0$ as fast as we choose, therefore $1/n\sum_{t=1}^{n}\mathsf{J}_{n,t} = o_p(1)$ in view of stationarity and ergodicity of x_t. Conversely, if $(\partial/\partial\theta)\kappa \neq 0$ then $|\mathcal{J}_{i,n}|/\ln(n)$ $\rightarrow (0,\infty)$ by (b). In view of $E[\mathsf{J}_{i,n,t}]/\mathcal{J}_{i,n} \rightarrow 1$ and the fact that $\mathsf{J}_{i,n,t}$ is β-mixing with summable coefficients, it follows $1/n\sum_{t=1}^{n}\mathsf{J}_{i,n,t}/\mathcal{J}_{i,n} \overset{p}{\rightarrow} 1$ by Theorem 2 and Example 4 in Andrews (1988).

It remains to prove the second term in (3) is $o_p(1)$. By (b) and twice differentiability of $\kappa(\theta)$ it follows for any $\theta \in \mathcal{N}_0(\delta)$:

$$E\left[\mathcal{H}_{n,t}(\theta)\right] = \frac{n}{m_n}E\left[\frac{\partial}{\partial\theta}\frac{\partial}{\partial\theta'}\ln\left(x_t(\theta)\right) \times \Im_n(\mathcal{X}_{n,t}(\theta))\right] \sim -\ln(n) \times \frac{\partial}{\partial\theta}\frac{\partial}{\partial\theta'}\kappa.$$

Hence pointwise $(n\ln(n))^{-1}\sum_{t=1}^{n}\{\mathcal{H}_{n,t}(\theta) - E[\mathcal{H}_{n,t}(\theta)]\} \overset{p}{\rightarrow} 0$ by Theorem 2 in Andrews (1988) given integrability and stationary β-mixing of $\mathcal{H}_{n,t}(\theta)$ on $\mathcal{N}_0(\delta)$. Further, since $\mathcal{H}_{n,t}(\theta)/\ln(n)$ is uniformly L_1-bounded by (c) 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)). Therefore $\sup_{\theta\in\Theta}|(n\ln(n))^{-1}\sum_{t=1}^{n}\{\mathcal{H}_{n,t}(\theta)$ $- E[\mathcal{H}_{n,t}(\theta)]\}| \overset{p}{\rightarrow} 0$ by Theorem 7.1.5 of Dudley (1999), hence the second term in (3) is $o_p(1)$.

Claim (b): Under measure space Assumption 1.a it follows by Leibniz's theorem

$$E\left[\frac{\partial}{\partial\theta}\ln\left(x_t(\theta)\right) \times I_{n,t}(\theta)\right] = \frac{\partial}{\partial\theta}E\left[\ln\left(x_t(\theta)/c_n(\theta)\right) \times I_{n,t}(\theta)\right] + \frac{\partial}{\partial\theta}\left\{\ln\left(c_n(\theta)\right) \times P\left(x_t(\theta) \geq c_n(\theta)\right)\right\}.$$

By construction $P(x_t(\theta) \geq c_n(\theta)) = m_n/n$ and by uniform regular variation Assumption 2.a $E[\ln(x_t(\theta)/c_n(\theta))I_{n,t}(\theta)] = (m_n/n)\kappa(\theta)^{-1} \times (1 + o(1))$ where

$o(1)$ is not a function of θ (Hsing (1991: eq. (1.5))). Moreover, $(\partial/\partial\theta)\kappa(\theta)$ exists for each $\theta \in \mathcal{N}_0(\delta)$. We will prove below for each $\theta \in \mathcal{N}_0(\delta)$

$$\frac{\partial}{\partial\theta}\ln\left(c_n(\theta)\right) = -\frac{\partial}{\partial\theta}\ln\kappa(\theta) \times \ln\left(n\right) \times \left(1 + o\left(1\right)\right). \tag{4}$$

Therefore, as claimed, for each $\theta \in \mathcal{N}_0(\delta)$:

$$\begin{aligned} \mathcal{J}_n(\theta) &= \frac{n}{m_n}E\left[\frac{\partial}{\partial\theta}\ln\left(x_t(\theta)\right) \times I_{n,t}(\theta)\right] \\ &= \frac{\partial}{\partial\theta}\kappa(\theta)^{-1} + \frac{\partial}{\partial\theta}\ln\left(c_n(\theta)\right) = -\ln\left(n\right) \times \frac{\partial}{\partial\theta}\ln\kappa(\theta) \times \left(1 + o\left(1\right)\right). \end{aligned}$$

By the same argument it can similarly be shown $(\partial/\partial\theta_{i_1} \cdots \partial\theta_{i_d})\mathcal{J}_n(\theta) = -\ln(n)(\partial/\partial\theta_{i_1} \cdots \partial\theta_{i_d})\ln(\kappa(\theta)) \times (1 + o(1))$ for any set $\{i_1, ..., i_d\}$ of integers $i_j \in \{1, ..., k\}$ with index $d \in \{1, 2, 3\}$, where $o(1)$ is not a function of θ.

We now prove (4). Note by Assumption 2.a $P(x_t(\theta) \geq c_n(\theta)) = c_n^{-\kappa(\theta)}(\theta)\mathcal{L}\left(c_n(\theta), \theta\right)$ and by construction $m_n/n = P(x_t(\theta) \geq c_n(\theta))$ and $c_n(\theta) = \{(n/m_n)\mathcal{L}\left(c_n(\theta), \theta\right)\}^{1/\kappa(\theta)}$. Thus

$$\begin{aligned} 0 &= \frac{\partial}{\partial\theta}P\left(x_t(\theta) \geq c_n(\theta)\right) \\ &= \frac{\partial}{\partial\theta}c_n^{-\kappa(\theta)}(\theta) \times \mathcal{L}\left(c_n(\theta), \theta\right) + c_n^{-\kappa(\theta)}(\theta)\frac{\partial}{\partial\theta}\mathcal{L}\left(c_n(\theta), \theta\right) \\ &= \frac{m_n}{n}\left(\frac{\partial}{\partial\theta}\frac{1}{c_n^{\kappa(\theta)}(\theta)}\right) \times c_n^{\kappa(\theta)}(\theta) + \frac{m_n}{n}\frac{1}{c_n^{\kappa(\theta)}(\theta)}\left(\frac{\partial}{\partial\theta}c_n^{\kappa(\theta)}(\theta)\right) \\ &= -\frac{m_n}{n}\frac{1}{c_n(\theta)} \times \frac{\partial}{\partial\theta}\kappa(\theta) + \frac{m_n}{n}\left(\frac{\partial}{\partial\theta}\kappa(\theta) \times \ln c_n(\theta) + \kappa(\theta)\frac{\partial}{\partial\theta}\ln c_n(\theta)\right) \end{aligned}$$

hence

$$\frac{\partial}{\partial\theta}\ln\left(c_n(\theta)\right) = -\ln\left(c_n(\theta)\right) \times \left(1 + o\left(1\right)\right) \times \frac{\partial}{\partial\theta}\ln\kappa(\theta).$$

Finally, since $\mathcal{L}(c_n(\theta), \cdot)$ is slowly varying it follows $\mathcal{L}(c_n(\theta), \theta) = o(c_n(\theta))$. Now use $c_n(\theta) = \{(n/m_n)\mathcal{L}\left(c_n(\theta), \theta\right)\}^{1/\kappa(\theta)}$ to deduce

$$\ln\left(\frac{n}{m_n}\right) = \kappa(\theta)\ln c_n(\theta) - \ln\mathcal{L}\left(c_n(\theta), \theta\right) = \kappa(\theta)\ln c_n(\theta) - o\left(\ln c_n(\theta)\right),$$

hence

$$\ln c_n(\theta) = \ln\left(\frac{n}{m_n}\right) \times (1 + o(1)) = \ln(n) \times (1 + o(1)).$$

This completes the proof of (4).

Claim (c): In view of uniform tail properties Assumption 2.a the above arguments can be easily generalized to prove $\sup_{\theta\in\mathcal{N}_0(\delta)} ||\mathcal{J}_n(\theta)|| \sim \ln(n)||(\partial/\partial\theta)\ln\kappa(\theta)||$ and $\sup_{\theta\in\mathcal{N}_0(\delta)} |(\partial/\partial\theta_{i_1}\cdots\partial\theta_{i_d})\mathcal{J}_n(\theta)| \sim \ln(n)|(\partial/\partial\theta_{i_1}\cdots\partial\theta_{i_d})\ln(\kappa(\theta))|$. $\mathcal{QED}$.

By Lemma 2.5.b it follows $\sup_{\theta\in\mathcal{N}_0(\delta)} E|(\partial/\partial\theta_i)\ln(x_t(\theta))I_{n,t}(\theta)| = O(m_n\ln(n)/n)$. In view of $m_n\ln(n)/n = o(1)$ by Assumption 2.b and $I_{n,t}(\theta) \stackrel{p}{\to} 1$ by construction, it follows by dominated convergence $(\partial/\partial\theta)\ln(x_t(\theta))$ is uniformly L_1-bounded on $\mathcal{N}_0(\delta)$. This in turn helps us prove expansion Lemma 2.3 above.

Corollary 2.6 $\sup_{\theta\in\mathcal{N}_0(\delta)} E|(\partial/\partial\theta_i)\ln(x_t(\theta))| < \infty$ *for each* $i = 1, ..., k$.

We now prove Theorem 2.1. We need only show $m_n^{1/2}(\hat{\kappa}_{m_n}^{-1}(\hat{\theta}_n) - \hat{\kappa}_{m_n}^{-1}) \stackrel{p}{\to} 0$ since under Assumptions 2 and 3 $m_n^{1/2}(\hat{\kappa}_{m_n}^{-1} - \kappa^{-1})/\sigma_{m_n} \stackrel{d}{\to} N(0,1)$ by Theorem 2 in Hill (2010), where $\sigma_{m_n}^2 := E(m_n^{1/2}(\hat{\kappa}_{m_n}^{-1} - \kappa^{-1}))^2$.

Decompose

$$\begin{aligned} m_n^{1/2}\left(\hat{\kappa}_{m_n}^{-1}(\hat{\theta}_n) - \hat{\kappa}_{m_n}^{-1}\right) &= \frac{1}{m_n^{1/2}}\sum_{t=1}^{n}\left\{\ln(x_t(\hat{\theta}_n))\hat{I}_{n,t}(\hat{\theta}_n) - \ln(x_t)\hat{I}_{n,t}\right\} \\ &\quad - m_n^{1/2}\left(\ln\left(x_{(m_n+1)}(\hat{\theta}_n)\right) - \ln\left(x_{(m_n+1)}\right)\right) \\ &= \mathcal{A}_n + \mathcal{B}_n. \end{aligned}$$

By Assumption 4 $\hat{\theta}_n \stackrel{p}{\to} \theta^0$ hence $\hat{\theta}_n \in \mathcal{N}_0(\delta)$ for any $\delta > 0$ with probability approaching one as $n \to \infty$. Now use expansion Lemma 2.3.a, approximation Lemma 2.4 and Jacobian limit Lemma 2.5.a to deduce $\mathcal{A}_n = m_n^{1/2}\mathcal{J}_n'(\hat{\theta}_n - \theta^0)(1 + o_p(1))$. In view of the Lemma 2.5.b Jacobian bound $\mathcal{J}_n = O(\ln(n))$, and $m_n^{1/2}\ln(n)(\hat{\theta}_n - \theta^0) = o_p(1)$ by Assumption 4, it follows $\mathcal{A}_n = o_p(1)$.

Next, for $\mathcal{B}_n$ apply uniform order statistic property Lemma 2.2.b to $\ln(x_{(m_n+1)}(\theta))$

and $\ln(x_{(m_n+1)})$ to deduce

$$\mathcal{B}_n = \kappa^{-1}\left(\frac{1}{m_n^{1/2}}\sum_{t=1}^{n}\left\{I_{n,t}(\hat{\theta}_n) - I_{n,t}\right\}\right) - \kappa^{-1}m_n^{1/2}\frac{n}{m_n}\left\{P\left(x_t(\hat{\theta}_n) \geq c_n(\hat{\theta}_n)\right) - P\left(x_t \geq c_n\right)\right\} + o_p(1)$$

$$= \mathcal{C}_{1,n} - \mathcal{C}_{2,n} + o_p(1),$$

say. By construction $P(x_t(\hat{\theta}_n) \geq c_n(\hat{\theta}_n)) = P(x_t > c_n) = m_n/n$ hence $\mathcal{C}_{2,n} = 0$. Finally, combine Lemma 2.3.b with $m_n^{1/2}\ln(n)(\hat{\theta}_n - \theta^0) = o_p(1)$ to deduce $\mathcal{C}_{1,n}$ $= O_p(m_n^{1/2}||\hat{\theta}_n - \theta^0||) = o_p(1)$, hence $\mathcal{B}_n = o_p(1)$. This proves $m_n^{1/2}(\hat{\kappa}_{m_n}^{-1}(\hat{\theta}_n) - \hat{\kappa}_{m_n}^{-1}) \overset{p}{\to} 0$ which completes the proof. $\mathcal{QED}$.

2.2 Proof of Lemma 2.2

Claim (a): By construction and Assumption 3, $\mathcal{I}_{n,t}(\theta)$ is L_2-bounded uniformly on $1 \leq t \leq n$, $n \geq 1$, and Θ, and is geometrically β-mixing on a compact subset $\mathcal{N}_0(\delta)$ of θ^0. Further, $\{\mathcal{I}_{n,t}(\theta) : \theta \in \Theta\}$ satisfies the metric entropy with L_2-bracketing bound $\int_0^1 \ln(\mathcal{N}_{[\,]}(\varepsilon, \Theta, ||\cdot||_2))d\varepsilon < \infty$, where $\mathcal{N}_{[\,]}(\varepsilon, \Theta, ||\cdot||_2)$ are the L_2-bracketing numbers. This follows since, by Assumption 1, $x_t(\theta)$ has an absolutely continuous and uniformly bounded distribution $\sup_{\theta\in\Theta}\sup_{a\in\mathbb{R}}||(\partial/\partial\theta)P(x_t(\theta) \leq a)|| < \infty$, and by continuity $c_n(\theta)$ is continuous. Therefore $\mathcal{I}_{n,t}(\theta)$ is L_2-Lipschitz on Θ: $E[(\mathcal{I}_{n,t}(\theta) - \mathcal{I}_{n,t}(\tilde{\theta}))^2] \leq K||\theta - \tilde{\theta}||$. Proving $\int_0^1 \ln(\mathcal{N}_{[\,]}(\varepsilon, \Theta, ||\cdot||_2))d\varepsilon < \infty$ is then a classic exercise (e.g. Pollard (1984)).

We may therefore apply Doukhan, Massart and Rio's (1995: Theorem 1, eq. (2.17), Application 4) uniform central limit theorem to deduce $\{1/n^{1/2}\sum_{t=1}^{n}\mathcal{I}_{n,t}(\theta)$ $: \theta \in \mathcal{N}_0(\delta)\} \Longrightarrow^* \{\mathcal{I}(\theta) : \theta \in \mathcal{N}_0(\delta)\}$, a Gaussian process with a version that has uniformly bounded and uniformly continuous sample paths with respect to $||\cdot||_2$.

Claim (b): Write for arbitrary $u \in \mathbb{R}$:

$$\mathcal{I}_n(u,\theta) := \frac{1}{m_n}\sum_{t=1}^{n} I\left(x_t(\theta) > c_n(\theta)e^{u/m_n^{1/2}}\right).$$

The following borrows arguments in Hsing (1991: p. 1553). By construction $m_n^{1/2}\ln(x_{(m_n+1)}(\theta)/c_n(\theta)) \leq u$ for $u \in \mathbb{R}$ iff (*if and only if*) $\mathcal{I}_n(u,\theta) \leq 1$ hence iff

$$\begin{aligned} m_n^{1/2}\left(\mathcal{I}_n\left(u,\theta\right) - E\left[\mathcal{I}_n\left(u,\theta\right)\right]\right) &\leq m_n^{1/2}\left(1 - \frac{n}{m_n}P\left(x_t(\theta) > c_n(\theta)e^{u/m_n^{1/2}}\right)\right) \\ &= m_n^{1/2}\left(1 - \frac{P\left(x_t(\theta) > c_n(\theta)e^{u/m_n^{1/2}}\right)}{P\left(x_t(\theta) > c_n(\theta)\right)}\right), \end{aligned}$$

since $(n/m_n)P(x_t(\theta) > c_n(\theta)) = 1$. Exploit the uniform second order regular variation Assumption 2.a to deduce

$$\frac{P\left(x_t(\theta) > c_n(\theta)e^{u/m_n^{1/2}}\right)}{P\left(x_t(\theta) > c_n(\theta)\right)} = e^{-\kappa u/m_n^{1/2}}\left(1 + \frac{1}{m_n^{1/2}} \times o(1)\right),$$

where by uniformity the term $a(1)$ is not a function of θ. Hence by the Mean Value Theorem

$$\begin{aligned} m_n^{1/2}\left(\mathcal{I}_n\left(u,\theta\right) - E\left[\mathcal{I}_n\left(u,\theta\right)\right]\right) &\leq m_n^{1/2}\left(1 - e^{-\kappa u/m_n^{1/2}}\left(1 + o(1/m_n^{1/2}))\right)\right) \\ &= m_n^{1/2}\left(1 - e^{-\kappa u/m_n^{1/2}} + o\left(1/m_n^{1/2}\right)\right) \\ &= \kappa u + o(1) \end{aligned}$$

where the $o(1)$ term is a non-random function that does not depend on θ. Therefore $\mathcal{I}_n(u,\theta) \leq 1$ iff $\kappa^{-1}m_n^{1/2}(\mathcal{I}_n\left(u,\theta\right) - E[\mathcal{I}_n\left(u,\theta\right)]) = u + o\left(1\right)$ hence

$$\begin{aligned} &P\left(m_n^{1/2}\left|\ln\left(x_{(m_n+1)}(\theta)/c_n(\theta)\right)\right| \leq u\right) \quad (5) \\ &= P\left(\kappa^{-1}m_n^{1/2}\left|\mathcal{I}_n\left(u,\theta\right) - E\left[\mathcal{I}_n\left(u,\theta\right)\right]\right| \leq u + o(1)\right). \end{aligned}$$

Claim (b) therefore follows since $o(1)$ does not depend on θ.

Claim (c): Use (5), uniform indicator law Claim (a) and the mapping theorem to prove the claim. $\mathcal{QED}$.

3 Proofs of Lemmas for Examples

Write $\mathcal{N}_0(\delta) = \mathcal{N}_0$ since the value of $\delta > 0$ is not exploited in the arguments below.

Throughout ϵ_t is an i.i.d. random variable with an absolutely continuous distribution that is positive on $\mathbb{R}$, and bounded $\sup_{a \in \mathbb{R}}(\partial/\partial a)P(\epsilon_t \leq a) < \infty$. In each example below we impose the following second order tail expansion for ϵ_t (or a similar error) for brevity (cf. Hall (1982), Haeusler and Teugels (1985)):

$$P\left(|\epsilon_t| > a\right) = da^{-\kappa}\left(1 + ca^{-\beta}\right), \ \beta, c, d, \kappa \in (0, \infty). \tag{6}$$

Let the fractile sequence $\{m_n\}$ satisfy

$$m_n \to \infty \ \text{ and } \ m_n = o(n^{2\tilde{\beta}/(2\tilde{\beta}+\kappa)}) \ \text{ where } \ \tilde{\beta} := \min\{\beta/2, 2\}. \tag{7}$$

Lemma 3.1 *Assumptions 1-3 hold. Further, in the general ARMA case, estimators in Davis (1996), Mikosch, Gadrich, Klüppelberg, and Alder (1995) and Zhu and Ling (2012) satisfy Assumption 4. Additionally, in the AR case estimators in Hill (2013) and Davis, Knight, and Liu (1992) satisfy Assumption 4.*

Proof. By construction and the fact that ϵ_t has finite moments of order less than κ, Assumption 1 is easily verified. Let L be the backshift operator: $L^p y_t = y_{t-p}$. Since y_t is geometrically β-mixing so are the finite and infinite lags $a(L)y_t = b(L)\epsilon_t(\theta)$ and $y_t = a(L)^{-1}b(L)\epsilon_t(\theta)$, and therefore so is $\epsilon_t(\theta)$. See, e.g., Mokkadem (1988: Theorem 1'), cf. Doukhan (1994: p. 99). Geometric mixing implies mixing in the ergodic hence, and therefore ergodicity (see, e.g., Petersen (1983)). Hence Assumption 3 holds.

Since we can write $a(L)y_t = \sum_{i=0}^{\infty} \tilde{\psi}_i(a)\epsilon_{t-i}$ for some $\tilde{\psi}_i : \mathcal{A} \to \mathbb{R}$, $\tilde{\psi}_i(a) = O(\rho^i)$, it follows by invertibility $\epsilon_t(\theta) = b(L)^{-1}a(L)y_t = \sum_{i=0}^{\infty} \mathring{\psi}_i(\theta)\epsilon_{t-i}$ where $\mathring{\psi}_i : \Theta \to \mathbb{R}$ is continuous and differentiable with a uniformly bounded derivative, and compactness of the parameter space ensures $\sup_{\theta \in \mathcal{N}_0} |\mathring{\psi}_i(\theta)| = O(\rho^i)$. Therefore $\epsilon_t(\theta)$ satisfies the second order power law property (6) with the same tail indices κ and β and some tail scales $c(\theta), d(\theta) > 0$ (Geluk, de Haan, Resnick, and Stărică (1997: Theorem 3.2)), and by construction $c(\theta^0) = c$ and $d(\theta^0) = d$. By properties of regularly varying functions it must be the case $d(\theta) = \sum_{i=0}^{\infty} |\mathring{\psi}_i(\theta)|^{\kappa}$, cf.

BrockCline (1985).

In order to see that the filter $x_t(\theta) = (\epsilon_t^2(\theta) + \varepsilon)^{1/2}$ for small $\varepsilon > 0$ satisfies Assumption 2.a, use the fact that $\epsilon_t(\theta)$ has tail (6) to obtain for large a:

$$P\left(x_t(\theta) \geq a\right) = P\left(\left|\epsilon_t(\theta)\right| \geq a\left(1 - \varepsilon/a^2\right)^{1/2}\right)$$

$$= d(\theta)a^{-\kappa} \times \left(\left(1 - \varepsilon/a^2\right)^{-\kappa/2} + c\left(\theta\right)a^{-\beta/2}\left(1 - \varepsilon/a^2\right)^{-\kappa/2-\beta/2}\right).$$

By the Mean Value Theorem, for some $\varepsilon_* \in [0, \varepsilon]$:

$$P\left(x_t(\theta) \geq a\right)$$

$$= d(\theta)a^{-\kappa}\left\{1 + \frac{c\left(\theta\right)}{a^{\beta/2}} + \frac{\kappa c\left(\theta\right)}{2a^2}\varepsilon\left(1 - \frac{\varepsilon_*}{a^2}\right)^{-\kappa/2-1} + \left(\frac{\kappa+\beta}{2a^2}\right)\frac{c\left(\theta\right)}{a^{\beta/2}}\varepsilon\left(1 - \frac{\varepsilon_*}{a^2}\right)^{-\kappa/2-\beta/2-1}\right\}$$

$$= d(\theta)a^{-\kappa}\left(1 + \tilde{c}\left(\theta\right)a^{-\tilde{\beta}}\right)$$

for some $\tilde{c}(\theta) > 0$ and $\tilde{\beta} := \min\{\beta/2, 2\}$. This tail class with fractile bound $m_n = o(n^{2\tilde{\beta}/(2\tilde{\beta}+\kappa)})$ by (7) satisfies Assumption 2.a (Hall (1982), Haeusler and Teugels (1985: Section 5)), while $m_n \to \infty$ and $m_n = o(n^{2\tilde{\beta}/(2\tilde{\beta}+\kappa)})$ imply Assumption 2.b.

Now consider the Assumption 4 plug-in requirement $m_n^{1/2}\ln(n)(\hat{\theta}_n - \theta^0) = o_p(1)$. Since $m_n = o(n/\ln(n))$ holds under (7), any $n^{1/2}$-convergent plug-in is valid. See Mikosch, Gadrich, Klüppelberg, and Alder (1995) and Davis (1996) for various estimators. If the model is a pure AR then a large class of smooth M-estimators (e.g. OLS) and LAD are also valid Davis, Knight, and Liu (1992). Finally, Zhu and Ling (2012)'s weighted LAD estimator is $n^{1/2}$-convergent and Hill (2013)'s Least Tail-Trimmed Squares estimator is at least $n^{1/2}$-convergent under our stated error properties. $\mathcal{QED}$.

Lemma 3.2 *Assumptions 1-3 hold, and the OLS estimator satisfies Assumption*

4.

Proof. Since $y_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}$ is geometrically β-mixing by assumption and $v_t(\theta) := y_t - \sum_{i-1}^{p} \theta_i y_{t-i}$ is an infinite order lag function of i.i.d. ϵ_t, the arguments used to prove Lemma 3.1 carry over verbatim to prove Assumptions 1-3 hold.

Now consider plug-in Assumption 4, assume an AR(1) model $y_t = \theta^0 y_{t-1} + v_t$ for notational economy, and define $\theta^0 := \sum_{i=0}^{\infty} \psi_i \psi_{i-1} / \sum_{i=0}^{\infty} \psi_i^2$. Since the least squares estimator $\hat{\theta}_n = \sum_{t=2}^{n} y_t y_{t-1} / \sum_{t=2}^{n} y_{t-1}^2$ is identically the first order sample autocorrelation, Theorem 4.4 of Davis and Resnick (1986) applies: $(n/\ln(n))^{1/\kappa}(\hat{\theta}_n - \theta^0) = O_p(1)$ if $\kappa \in (1,2]$, and if $\kappa > 2$ then $n^{1/2}(\hat{\theta}_n - \theta^0) = O_p(1)$. Assumption 4 therefore holds since $\kappa > 1$. $\mathcal{QED}$.

Lemma 3.3 *Assumptions 1-3 hold, and Log-LAD, Quasi-Maximum Tail-Trimmed Likelihood and weighted Laplace QML satisfy Assumption 4. QML satisfies Assumption 4 when $\kappa > 4$, and when $\kappa \in (2,4]$ provided $m_n = o(n^{2-4/\kappa})$.*

Proof. Define $S_t(\theta) := \sigma_t/\sigma_t(\theta)$ hence $\epsilon_t(\theta) = \epsilon_t S_t(\theta)$. By the mixing property of $\{y_t, \sigma_t\}$ it follows $\sigma_t^2(\theta)$ is stationary geometrically β-mixing on some neighborhood $\mathcal{N}_0$ of θ^0 (see Doukhan (1994: Chapter 2.4)), hence $\epsilon_t(\theta) = y_t/\sigma_t(\theta)$ is stationary geometrically β-mixing on $\mathcal{N}_0$. This verifies Assumption 3.

Further, under the stated parameter restrictions $(\omega^0, \omega) > 0$ and $(\alpha^0, \alpha, \beta^0, \beta) \in (0,1)$ it follows $E|S_t(\theta)|^p < \infty$ for each $p > 0$ and some compact neighborhood $\mathcal{N}_0$ of θ^0 that may depend on p (Francq and Zakoïan (2004: eq. (4.25))). Therefore $E[S_t(\theta)^\kappa] < \infty$ on $\mathcal{N}_0$. But since ϵ_t has tail (6) it follows $\epsilon_t(\theta) = \epsilon_t S_t(\theta)$ also satisfies (6) for each θ since by iterated expectations and independence

$$\begin{aligned} P\left(\epsilon_t S_t(\theta) > a\right) &= E\left[P\left(\epsilon_t > a/S_t(\theta)|S_t(\theta)\right)\right] \\ &= a^{-\kappa} E\left[S_t(\theta)^\kappa\right]\left(1 + c a^{\beta} E\left[S_t(\theta)^{\kappa-\beta}\right]/E\left[S_t(\theta)^\kappa\right]\right) \\ &= d(\theta)\, a^{-\kappa}\left(1 + c(\theta)\, a^{\beta}\right). \end{aligned}$$

Francq and Zakoïan (2004: eq. (4.25)) prove $E|S_t(\theta)| < \infty$ on some neighborhood of θ^0. Their argument generalizes to $E|S_t(\theta)|^p < \infty$ for any $p > 0$ and neighborhoods $\mathcal{N}_0$ of θ^0 that may depend on p: if $p \in (0,1)$ then use their (4.25) with Jensen's inequality, and if $p \geq 1$ then use Minkowski's inequality.

See also Breiman (1965: Proposition 3). Since $E[S_t(\theta)^p]$ is bounded on $\mathcal{N}_0$, it follows $\epsilon_t(\theta)$ satisfies Assumption 2.a. Therefore $x_t(\theta) := (\epsilon_t^2(\theta) + \varepsilon)^{1/2}$ satisfies Assumption 2.a by the proof of Lemma 3.1. Assumption 2.b holds in view of (7). Hence Assumption 1 holds by construction and the existence of a moment or order less than κ.

Now consider Assumption 4. By assumption $m_n = o(n^{1-\iota})$ necessarily holds for tiny $\iota > 0$ hence any $n^{1/2-\iota/2}\ln(n)$-convergent plug-in satisfies Assumption 4. QML, Log-LAD, Quasi-Maximum Tail-Trimmed Likelihood and weighted Laplace QML have rate $n^{1/2}$ if the error tail index $\kappa > 4$ (cf. Francq and Zakoïan (2004), Peng and Yao (2003), Zhu and Ling (2011), Hill (2014a)).

If $\kappa \in (2,4]$ then Log-LAD and weighted Laplace QML have rate $n^{1/2}$, and Quasi-Maximum Tail-Trimmed Likelihood has rate $n^{1/2}/g_n$ where $\{g_n\}$ is any sequence of positive numbers satisfying $g_n \to \infty$ as slow as desired based on the chosen number of trimmed GARCH errors for each n (see Hill (2014a)). Hence these three satisfy Assumption 4. The QML rate is $n^{1-2/\kappa}/\mathcal{L}(n) \leq n^{1/2}/\mathcal{L}(n)$ for some slowly varying $\mathcal{L}(n) \to \infty$ (Hall and Yao (2003)), so Assumption 4 holds if $m_n = o(n^{2-4/\kappa})$ provided also the bound on m_n in (7) holds. $\mathcal{QED}$.

Lemma 3.4 *Assumptions 1-3 hold and weighted Laplace QML satisfies Assumption 4.*

Proof. Define the subvector $\psi := [\omega, \alpha]'$ of $\theta = [\phi, \omega, \alpha]'$, where ψ lies in Ψ a compact subset of $(0,\infty) \times (0,1)$. Both ϵ_t and y_t are geometrically β-mixing (Cline (2007)), hence the finite lag function

$$\epsilon_t(\theta) = \left(\epsilon_t + \frac{\left(\phi^0 - \phi\right) y_{t-1}}{\left(\omega^0 + \alpha^0 y_{t-1}^2\right)^{1/2}}\right)\left(\frac{\omega^0 + \alpha^0 y_{t-1}^2}{\omega + \alpha y_{t-1}^2}\right)^{1/2} = \left(\epsilon_t + A_t(\phi)\right) B_t(\psi)$$

is for any $\theta \in \Theta$ geometrically β-mixing. This verifies Assumption 3. The weighted Laplace QML estimator of θ^0 is $n^{1/2}$-convergent, hence it satisfies Assumption 4 (Zhu and Ling (2011)).

It remains to verify Assumption 2. Let $K > 0$ be a finite constant that may be different in different places. By assumption ϵ_t is independent of $A_t(\phi)$ and $B_t(\psi)$, $B_t(\psi)$ is bounded from below $\inf_{\psi\in\Psi} B_t(\psi) \geq K$, and both are bounded from above $\sup_{\phi\in(-1,1)} |A_t(\phi)| \leq K$ and $\sup_{\psi\in\Psi} B_t(\psi) \leq K$ *a.s.* In view of the assumption that i.i.d. ϵ_t has tail (6) and is independent of bonded $A_t(\phi)$, it is easy to show $\epsilon_t + A_t(\phi)$ also satisfies (6) by exploiting the identity $P(|\epsilon_t + A_t(\phi)| > a) = E[P(\epsilon_t > a - A_t(\phi)|A_t(\phi)] + E[P(\epsilon_t < -a - A_t(\phi)|A_t(\phi)]$ and the proof of Lemma 3.1. Moreover, since $B_t(\psi)$ is bounded from below and above and is independent of ϵ_t, it follows $\epsilon_t(\theta) = (\epsilon_t + A_t(\phi))B_t(\psi)$ also satisfies (6) by the proof of Lemma 3.1. Hence $x_t(\theta) := (\epsilon_t^2(\theta) + \varepsilon)^{1/2}$ satisfies Assumption 2.a by the proof of Lemma 3.1, and Assumption 2.b holds by (7). Hence Assumption 1 holds. $\mathcal{QED}$.

Lemma 3.5 *Assumptions 1 and 3 hold, and x_t has tail $P(|x_t| > a) = da^{-\kappa}(1 + o(1))$. If ϵ_t has a symmetric distribution then Hill (2014b))'s Generalized Empirical Likelihood estimator satisfies Assumption 4.*

Proof . The AR error $u_t = \sigma_t\epsilon_t$ has a power law tail $P(|u_t| \geq a) = da^{-\kappa}(1 + o(1))$ under the assumed GARCH error properties (Basrak, Davis, and Mikosck (2002: Theorem 3.2)) and $\{u_t, \sigma_t^2\}$ is stationary and geometrically β-mixing (Meitz and Saikkonen (2008: Theorem 1)). Finally, if the i.i.d. GARCH error ϵ_t has a symmetric distribution then Hill (2014b)'s class of GEL estimators is $n^{1/2}$-convergent hence Assumption 4 holds. $\mathcal{QED}$.

References

Andrews, D. W. K. (1988). Laws of Large Numbers for Dependent Non-Identically Distributed Random Variables. *Economet. Th.* **4**, 458-467.

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

Breiman, L. (1965). On Some Limit Theorems Similar to the Arc-Sine Law. *Th. Prob. Appl.* **10**, 323-331.

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

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

Davis R. A. (1996). Gauss-Newton and M-Estimation for ARMA Processes with Infinite Variance. *Stoch. Proc. Appl.* **63**, 75-95.

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

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

Doukhan, P. (1994). *Mixing Properties and Examples.* Springer-Verlag: New York.

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

Dudley, R. M. (1999). *Uniform Central Limit Theorems.* Cambridge Univ. Press: Cambridge, UK.

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

Geluk, J., de Haan, L., Resnick, S., and Stărică, C. (1997). Second-Order Regular Variation, Convolution and the Central Limit Theorem. *Stoch. Proc. Appl.* **69**, 139 I59.

Goldie, C. M., and Smith, R. L. (1987). Slow Variation with Remainder: Theory and Applications. *Q. J. Math.* **38**, 45-71.

Haeusler, E., and Teugels, J. L. (1985). On Asymptotic Normality of Hill's Estimator for the Exponent of Regular Variation. *Ann. Stat.* **13**, 743-756.

Hall, P. (1982). On Some Simple Estimates of an Exponent of Regular Variation. *J. R. Stat. Soc. B* **44**, 37-42.

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

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

Hill, J. B. (2013). Least Tail-Trimmed Squares for Infinite Variance Autoregressions. *J. Time Ser. Anal.* **34**, 168-186.

Hill, J. B. (2014a). Robust Estimation for Heavy Tailed GARCH. *Bernoulli*, forthcoming.

Hill, J. B. (2014b). Robust Generalized Empirical Likelihood for Heavy Tailed Autoregressions with Conditionally Heteroscedastic Errors. Working Paper, Dept. of Economics, University of North Carolina.

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

Lighthill, M. (1958). *Introduction to Fourier Analysis and Generalized Functions*. Cambridge Univ.Press, Cambridge.

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

Mikosch, T., Gadrich, T., Klüppelberg, C., and Alder, R. J. (1995). Parameter Estimation for ARMA Models with Infinite Variance Innovations. *Ann. Stat.* **23**, 305-326.

Mokkadem, A. (1988). Mixing Properties of ARMA Processes. *Stoch. Proc. Appl.* **29**, 309-315.

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

Petersen, K. (1983). *Ergodic Theory*. Cambridge Univ. Press.

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

Zhu, K., and Ling, S. (2011). Global Self-Weighted and Local Quasi-Maximum Exponential Likelihood Estimators for ARMA-GARCH/IGARCH Models. *Ann. Stat.* **39**, 2131-2163.

Zhu, K., and Ling, S. (2012). The Global Weighted LAD Estimators for Finite/Infinite Variance ARMA(p,q) Models. *Economet. Th.* **28**, 1065-1086.

Dept. of Economics, University of North Carolina
E-mail: (jbhill@email.unc.edu)