Supplementary Material for
"Tail Index Estimation for a Filtered Dependent Time Series"

Jonathan B. Hill

Dept. of Economics, University of North Carolina

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

1 Assumptions

Assumption 1 (Smoothness and Moments).

a. Let \( \{\mathcal{F}_t\}_{t \in \mathbb{Z}} \) be a sequence of \( \sigma \)-fields that do not depend on \( \theta \) and define \( \mathcal{F} := \sigma(\bigcup_{t \in \mathbb{Z}} \mathcal{F}_t) \). \( x_t(\theta) \) lies on a complete probability measure space \( (\Omega, \mathcal{F}, P) \) and is \( \mathcal{F}_t \)-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 \( \mathcal{F}_t \)-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 \( \kappa > 0 \):

\[
P(x_t > a) = a^{-\kappa} L(a) \quad \text{where } a > 0 \text{ and } L(a) \text{ is slowly varying.} \quad (1)
\]

Assumption 2 (Regular Variation and Fractile Bound).
a. There exists a neighborhood $N_0(\delta)$ such that

$$
\lim_{a \to \infty} \sup_{\theta \in N_0(\delta)} \left| \frac{a^{\kappa(\theta)}}{\mathcal{L}(a, \theta)} P(x_t(\theta) > a) - 1 \right| = 0.
$$

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 $\Theta$, that is $\sup_{\theta \in N_0(\delta)} |\mathcal{L}(\lambda a, \theta)/\mathcal{L}(a, \theta) - 1| = O(h(a))$ 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 $h(dz)/h(z) \leq D\tau$ some for $\theta \geq 1$ 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 $\inf_{\theta \in N_0(\delta)} \kappa(\theta) > 0$ and $\sup_{\theta \in N_0(\delta)} \kappa(\theta) < \infty$, 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 $\|((\partial/\partial \theta)\kappa(\theta) - (\partial/\partial \theta)\kappa(\theta))\| \leq K\|\theta - \hat{\theta}\|$ for each $\theta, \hat{\theta} \in N_0(\delta)$.

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

Assumption 3 (mixing). Let $N_0(\delta)$ be the neighborhood of $\theta^0$ defined in Assumption 2.a. Then $x_t(\theta)$ is $\beta$-mixing for each $\theta \in N_0(\delta)$ with summable coefficients. Hence $\beta_l := \sup_{\mathcal{A} \subset \mathcal{A}^0} \mathbb{E}|P(\mathcal{A}^0|\mathcal{A}) - P(\mathcal{A})|$ 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 $m_n^{1/2} (\hat{\kappa}_m^{-1}(\hat{\kappa}_m^{-1}(\hat{\theta}_n) - \kappa^{-1})(\hat{\theta}_n) - \kappa^{-1})/\sigma_m \overset{d}{\to} N(0, 1)$ where $\sigma_m^2 := E(m_n^{1/2}((\hat{\kappa}_m^{-1} - \kappa^{-1}))^2)$.

The main supporting lemma is proved in Section 2.2. Recall

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

Lemma 2.2 Under Assumptions 1-3 there exists a Gaussian process $\{\mathcal{I}(\theta) : \theta \in N_0(\delta)\}$ with uniformly bounded and uniformly continuous sample paths with
Supplementary Material: Tail Index Estimation for a Filtered Series

3

respect to $|| \cdot ||_2$ such that:

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

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

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

2.1 Proof of Theorem 2.1

In order to prove Theorem 2.1 we require several preliminary results. Drop $\theta^0$ 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

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

and sample and population Jacobia

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

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

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

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

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 $y_t(\theta) := \ln(x_t(\theta)), \tilde{y}_{n,t}(\theta) := y_t(\theta)\hat{I}_{n,t}(\theta), y_{n,t}(\theta) := y_t(\theta)I_{n,t}(\theta),$ 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 \( \dot{y}_{n,t}(\theta) = \{y_t(\tilde{\theta}) + J_t(\theta_*)'(\theta - \tilde{\theta})\} \times \hat{I}_{n,t}(\theta) \) for some \( \theta_* \) that satisfies \( ||\theta_* - \tilde{\theta}|| \leq ||\theta - \tilde{\theta}|| \), hence

\[
\frac{1}{m_n^{1/2}} \sum_{t=1}^{n} \dot{y}_{n,t}(\theta) - \frac{1}{m_n^{1/2}} \sum_{t=1}^{n} \dot{y}_{n,t}(\tilde{\theta}) = \frac{1}{m_n^{1/2}} \sum_{t=1}^{n} J_t(\theta_*)' \times \hat{I}_{n,t}(\theta_*) \times (\theta - \tilde{\theta}) + \frac{1}{m_n^{1/2}} \sum_{t=1}^{n} y_t(\theta) \times \{\hat{I}_{n,t}(\theta) - \hat{I}_{n,t}(\tilde{\theta})\} + \frac{1}{m_n^{1/2}} \sum_{t=1}^{n} J_t(\theta_*)' \times \{\hat{I}_{n,t}(\theta) - \hat{I}_{n,t}(\theta_*)\} \times (\theta - \tilde{\theta}) + \frac{1}{m_n^{1/2}} \sum_{t=1}^{n} J_t(\theta_*)' \times \hat{I}_{n,t}(\theta_*) \times (\theta - \tilde{\theta}) + E_{1,n}(\theta, \tilde{\theta}) + E_{2,n}(\theta, \tilde{\theta}).
\]

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

The indicator function \( I(u) := I(u \geq 0) \) can be approximated by a smooth regular sequence \( \{J_n(u)\} \), cf. Lighthill (1958). Let \( \{N_n\} \) be a sequence of finite positive numbers, \( N_n \to \infty \), the rate to be chosen below. Define \( J_n(u) := \int_{-\infty}^{\infty} I(\varpi) S(N_n(\varpi - u)N_n^{-1/2}d\varpi \text{ where } S(\xi) = e^{-1/(1-\xi^2)}/\int_{-1}^{1} e^{-1/(1-w^2)}dw \text{ if } |\xi| < 1 \text{ and } S(\xi) = 0 \text{ if } |\xi| \geq 1. \) Note \( J_n(u) \) 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 \( D_n(u) := (N_n/\pi)^{1/2} \exp\{-N_nu^2\}. \) See Lighthill (1958: p. 22, eqs. (23) and (24)).

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

\[
\frac{1}{m_n^{1/2}} \sum_{t=1}^{n} y_t(\theta) \times \{\hat{I}_{n,t}(\theta) - \hat{I}_{n,t}(\tilde{\theta})\} = \frac{1}{m_n^{1/2}} \sum_{t=1}^{n} y_t(\theta) \times \{J_n(\hat{x}_{n,t}(\theta)) - J_n(\hat{x}_{n,t}(\tilde{\theta}))\} + o_p(1).
\]
In view of $\sup_{\theta \in \mathcal{N}_n(\theta)} |x_{(m_n + 1)}(\theta)/c_n(\theta) - 1| = O_p(1/m_n^{1/2})$ be Lemma 2.2 and the Mean Value Theorem, we may similarly write for some $\theta^*$, $||\theta^* - \theta|| \leq ||\theta - \hat{\theta}||$

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

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

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

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

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

$$= \sum_{i=1}^{4} A_{i,n}(\theta, \theta^*, \hat{\theta}) + o_p(1).$$

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

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

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

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

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

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 $L_1$-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

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

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

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

**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. \quad \text{QED}.

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

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

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

**Proof.**

**Claim (a):** We exploit notation and arguments from the proof of Lemma 2.3, in particular \( J_{n}(\cdot) \) and \( N_{n}. \) Recall \( J_{n}(\theta) := \binom{n/m}{n} E[(\partial/\partial \theta) \ln(x_t(\theta))] I_{n,t}(\theta). \)

We have

\[
\hat{J}_{n}(\hat{\theta}_n) - J_{n} = \left( \frac{1}{n} \sum_{t=1}^{n} \frac{n}{m} \frac{\partial}{\partial \theta} \ln(x_{t}(\hat{\theta}_n)) \times I_{n,t}(\hat{\theta}_n) \right) - \frac{n}{m} \left( \frac{\partial}{\partial \theta} \ln(x_{t}) \times I_{n,t} \right) \\
+ \frac{1}{n} \sum_{t=1}^{n} \left\{ \frac{\partial}{\partial \theta} \ln(x_{t}(\hat{\theta}_n)) \times \left\{ I_{n,t}(\hat{\theta}_n) - I_{n,t}(\hat{\theta}_n) \right\} \right\} \\
= A_n + B_n.
\]

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 \( m^{1/2} \ln(n) (\hat{\theta}_n - \theta^0) = o_p(1) \), and the Lemma 2.2 result \( \sup_{\theta \in \mathcal{N}_0(\delta)} |x_{(m_n+1)}(\theta)/c_n(\theta) - 1| = O_p(1/m^{1/2}) \).

Consider the first term. We have:

\[
\frac{1}{n} \sum_{t=1}^{n} \left\{ \frac{\partial}{\partial \theta} \ln(x_{t}(\hat{\theta}_n)) \times I_{n,t}(\hat{\theta}_n) \right\} - E \left( \frac{\partial}{\partial \theta} \ln(x_{t}) \times I_{n,t} \right) \\
= \frac{1}{m} \sum_{t=1}^{n} \left\{ \frac{\partial}{\partial \theta} \ln(x_{t}(\hat{\theta}_n)) \times J_{n,t}(X_{n,t}(\hat{\theta}_n)) - E \left( \frac{\partial}{\partial \theta} \ln(x_{t}) \times J_{n,t}(X_{n,t}) \right) \right\} \\
+ o_p(1),
\]

where \( X_{n,t}(\theta) := x_{t}(\theta) - c_{n}(\theta). \) Define

\[
H_{n,t}(\theta) = \left[ H_{i,j,n,t}(\theta) \right] := \left[ \frac{n}{m} \frac{\partial}{\partial \theta_i} \frac{\partial}{\partial \theta_j} \ln(x_t(\theta)) \times J_{n,t}(X_{n,t}(\theta)) \right] \\
J_{n,t} = \left[ J_{i,n,t} \right] := \left[ \frac{n}{m} \frac{\partial}{\partial \theta_i} \ln(x_t) \times J_{n,t}(X_{n,t}) \right],
\]
and note \( ||E[J_{n,t}] - J_n|| \to 0 \) as fast as we choose by choice of \( \{N_n\} \). Now expand around \( \theta^0 \). Use \( m_n^{1/2} \ln(n)(\hat{\theta}_n - \theta^0) = o_p(1) \), and \( (\partial/\partial \theta) \mathcal{J}_n(\mathcal{X}_{n,t}(\theta^*)) \overset{p}{\to} 0 \) as fast as we choose by choice of the sequence \( \{N_n\} \), to deduce for some \( \theta^* \), \( ||\theta^* - \theta^0|| \leq ||\hat{\theta}_n - \theta^0|| \):\[
\frac{1}{m_n} \sum_{t=1}^{n} \frac{\partial}{\partial \theta} \ln(x_t(\hat{\theta}_n)) \times \mathcal{J}_n(\mathcal{X}_{n,t}(\hat{\theta}_n)) = \frac{1}{n} \sum_{t=1}^{n} J_{n,t} + o_p \left( \frac{1}{m_n^{1/2} \ln(n)} \times \frac{1}{n} \sum_{t=1}^{n} \mathcal{H}_{n,t}(\theta^*) \right). \quad (3)\]

Consider the first term in (3). Suppose \( (\partial/\partial \theta) \kappa = 0 \). By (b) \( J_n = 0 \) hence \( E[J_{n,t}] \to 0 \) as fast as we choose, therefore \( 1/n \sum_{t=1}^{n} J_{n,t} = o_p(1) \) in view of stationarity and ergodicity of \( x_t \). Conversely, if \( (\partial/\partial \theta) \kappa \neq 0 \) then \( \mathcal{J}_{J_n}/\ln(n) \to (0, \infty) \) by (b). In view of \( E[J_{n,t}]/\mathcal{J}_{J_n} \to 1 \) and the fact that \( \mathcal{J}_{J_n,t} \) is \( \beta \)-mixing with summable coefficients, it follows \( 1/n \sum_{t=1}^{n} J_{n,t}/\mathcal{J}_{J_n} \overset{p}{\to} 1 \) 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 \( \theta \in \mathcal{N}_0(\delta) \):

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

Hence pointwise \( (n \ln(n))^{-1} \sum_{t=1}^{n} \{\mathcal{H}_{n,t}(\theta) - E[\mathcal{H}_{n,t}(\theta)]\} \overset{p}{\to} 0 \) by Theorem 2 in Andrews (1988) given integrability and stationary \( \beta \)-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 \( N_{\phi}(\delta) \leq c \delta < \infty \) (Dudley (1999: Proposition 7.1.7)). Therefore \( \sup_{\theta \in \Theta} (n \ln(n))^{-1} \sum_{t=1}^{n} \{\mathcal{H}_{n,t}(\theta) - E[\mathcal{H}_{n,t}(\theta)]\} \overset{L_1}{\to} 0 \) by Theorem 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

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

\[
+ \frac{\partial}{\partial \theta} \{\ln(c_n(\theta)) \times P(x_t(\theta) \geq c_n(\theta))\}.
\]
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 $\theta$ (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 $\theta \in \mathcal{N}_0(\delta)$

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

(4)

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

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

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

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 $\{i_1, \ldots, i_d\}$ of integers $i_j \in \{1, \ldots, k\}$ with index $d \in \{1, 2, 3\}$, where $o(1)$ is not a function of $\theta$.

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

Thus

$$
0 = \frac{\partial}{\partial \theta} P(x_t(\theta) \geq c_n(\theta))
$$

$$
= \frac{\partial}{\partial \theta} c_n^{-\kappa(\theta)}(\theta) \times \mathcal{L}(c_n(\theta), \theta) + c_n^{-\kappa(\theta)}(\theta) \frac{\partial}{\partial \theta} \mathcal{L}(c_n(\theta), \theta)
$$

$$
= \frac{m_n}{n} \left( \frac{\partial}{\partial \theta} \frac{1}{c_n^{\kappa(\theta)}(\theta)} \right) \times c_n^{\kappa(\theta)}(\theta) + \frac{m_n}{n} \frac{1}{c_n^{\kappa(\theta)}(\theta)} \left( \frac{\partial}{\partial \theta} c_n^{\kappa(\theta)}(\theta) \right)
$$

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

hence

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

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 \( c_n(\theta) = \{(n/m_n)\mathcal{L}(c_n(\theta), \theta)\}^{1/\kappa(\theta)} \) to deduce

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

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 \( \sup_{\theta \in \mathcal{N}_0(\delta)} \| J_n(\theta) \| \sim \ln(n) \| (\partial/\partial \theta) \ln \kappa(\theta) \| \) and \( \sup_{\theta \in \mathcal{N}_0(\delta)} |(\partial/\partial \theta_1 \cdots \partial \theta_u) J_n(\theta) | \sim \ln(n) \| (\partial/\partial \theta_1 \cdots \partial \theta_u) \ln(\kappa(\theta)) | \). QED.

By Lemma 2.5.b it follows \( \sup_{\theta \in \mathcal{N}_0(\delta)} E[|(\partial/\partial \theta_1) \ln(x_1(\theta)) I_{n,t}(\theta)|] = O(m_n \ln(n)/n) \).

In view of \( m_n \ln(n)/n = o(1) \) by Assumption 2.b and \( I_{n,t}(\theta) \overset{p}{\to} 1 \) by construction, it follows by dominated convergence \( (\partial/\partial \theta) \ln(x_1(\theta)) \) is uniformly \( L_1 \)-bounded on \( \mathcal{N}_0(\delta) \). This in turn helps us prove expansion Lemma 2.3 above.

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

We now prove Theorem 2.1. We need only show \( m_n^{1/2}(\hat{\kappa}_{m_n}^{-1} - \hat{\kappa}_{m_n}^{-1}) \overset{p}{\to} 0 \) since under Assumptions 2 and 3 \( m_n^{1/2}(\hat{\kappa}_{m_n}^{-1} - \kappa^{-1})/\sigma_{m_n} \overset{d}{\to} N(0,1) \) 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

\[
m_n^{1/2}(\hat{\kappa}_{m_n}^{-1} - \hat{\kappa}_{m_n}^{-1}) = \frac{1}{m_n^{1/2}} \sum_{t=1}^n \left\{ \ln(x_t(\hat{\theta}_n)) I_{n,t}(\hat{\theta}_n) - \ln(x_t) I_{n,t} \right\}
- m_n^{1/2} \left( \ln(x_{(m_n+1)}(\hat{\theta}_n)) - \ln(x_{(m_n+1)}) \right)
= A_n + B_n.
\]

By Assumption 4 \( \hat{\theta}_n \overset{p}{\to} \theta^0 \) 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 \( A_n = m_n^{1/2} J_n'(\hat{\theta}_n - \theta^0)(1 + o_p(1)) \). In view of the Lemma 2.5.b Jacobian bound \( J_n = O(\ln(n)) \), and \( m_n^{1/2} \ln(n)(\hat{\theta}_n - \theta^0) = o_p(1) \) by Assumption 4, it follows \( A_n = o_p(1) \).
Next, for $B_n$ apply uniform order statistic property Lemma 2.2.b to $\ln(x(m_n+1)(\theta))$ and $\ln(x(m_n+1))$ to deduce

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

$$- \kappa^{-1} m_n^{1/2} \frac{n}{m_n} \left\{ P \left( x_t(\hat{\theta}_n) \geq c_n(\hat{\theta}_n) \right) - P \left( x_t \geq c_n \right) \right\} + o_p(1)$$

$$= C_{1,n} - C_{2,n} + o_p(1),$$

say. By construction $P(x_t(\hat{\theta}_n) \geq c_n(\hat{\theta}_n)) = P(x_t > c_n) = m_n/n$ hence $C_{2,n} = 0$.

Finally, combine Lemma 2.3.b with $m_n^{1/2} \ln(n)(\hat{\theta}_n - \theta^0) = o_p(1)$ to deduce $C_{1,n} = O_p(m_n^{1/2}||\hat{\theta}_n - \theta^0||) = o_p(1)$, hence $B_n = o_p(1)$. This proves $m_n^{1/2}(\hat{\kappa}^{-1}_{m_n}(\hat{\theta}_n) - \hat{\kappa}^{-1}_{m_n}) \overset{L_2}{\rightarrow} 0$ which completes the proof. QED.

### 2.2 Proof of Lemma 2.2

**Claim (a):** By construction and Assumption 3, $I_{n,t}(\theta)$ is $L_2$-bounded uniformly on $1 \leq t \leq n$, $n \geq 1$, and $\Theta$, and is geometrically $\beta$-mixing on a compact subset $\mathcal{N}_0(\delta)$ of $\theta^0$. Further, $\{I_{n,t}(\theta) : \theta \in \Theta\}$ satisfies the metric entropy with $L_2$-bracketing bound $\int_0^1 \ln(\mathcal{N}_1(\delta, \Theta, ||\cdot||_2))d\varepsilon < \infty$, where $\mathcal{N}_1(\delta, \Theta, ||\cdot||_2)$ 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 $I_{n,t}(\theta)$ is $L_2$-Lipschitz on $\Theta$: $E[(I_{n,t}(\theta) - I_{n,t}(\bar{\theta}))^2] \leq K||\theta - \bar{\theta}||$. Proving $\int_0^1 \ln(\mathcal{N}_1(\varepsilon, \Theta, ||\cdot||_2))d\varepsilon < \infty$ 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 $\{1/n^{1/2} \sum_{t=1}^{n} I_{n,t}(\theta) : \theta \in \mathcal{N}_0(\delta)\} \rightsquigarrow \{I(\theta) : \theta \in \mathcal{N}_0(\delta)\}$, 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 $u \in \mathbb{R}$:

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

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

\[
= m_n^{1/2} \left( 1 - \frac{P \left( x_t(\theta) > c_n(\theta)e^{u/m_n^{1/2}} \right)}{P (x_t(\theta) > c_n(\theta))} \right),
\]

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

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

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

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

\[
= m_n^{1/2} \left( 1 - e^{-\kappa u/m_n^{1/2}} + o \left( 1/m_n^{1/2} \right) \right)
\]

\[
= \kappa u + o(1)
\]

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

\[
P \left( m_n^{1/2} \left| \ln \left( \frac{x_{(m_n+1)}(\theta)}{c_n(\theta)} \right) \right| \leq u \right)
\]

\[
= P \left( \kappa^{-1} m_n^{1/2} \left| \mathcal{I}_n (u, \theta) - E [\mathcal{I}_n (u, \theta)] \right| \leq u + o(1) \right)
\]

Claim (b) therefore follows since \( o(1) \) does not depend on \( \theta \).

**Claim (c):** Use (5), uniform indicator law Claim (a) and the mapping theorem to prove the claim. 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 $\epsilon_t$ 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 $\epsilon_t$ (or a similar error) for brevity (cf. Hall (1982), Haeusler and Teugels (1985)):

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

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

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


**Proof.** By construction and the fact that $\epsilon_t$ has finite moments of order less than $\kappa$, Assumption 1 is easily verified. Let $L$ be the backshift operator: $L^p y_t = y_{t-p}$. Since $y_t$ is geometrically $\beta$-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 $\tilde{\psi}_i : \mathbb{A} \to \mathbb{R}$, $\tilde{\psi}_i(a) = O(\rho^i)$, it follows by invertibility $\epsilon_t(\theta) = b(L)^{-1}a(L)y_t = \sum_{i=0}^{\infty} \tilde{\psi}_i(\theta)\epsilon_{t-i}$ where $\tilde{\psi}_i : \Theta \to \mathbb{R}$ is continuous and differentiable with a uniformly bounded derivative, and compactness of the parameter space ensures $\sup_{\theta \in \mathcal{N}_0} |\tilde{\psi}_i(\theta)| = O(\rho^i)$. Therefore $\epsilon_t(\theta)$ satisfies the second order power law property (6) with the same tail indices $\kappa$ and $\beta$ and some tail scales $c(\theta), d(\theta) > 0$ (Geluk, de Haan, Resnick, and Stărică (1997: Theorem 3.2)), and by construction $c(\theta^2) = c$ and $d(\theta^2) = d$. By properties of regularly varying functions it must be the case $d(\theta) = \sum_{i=0}^{\infty} |\tilde{\psi}_i(\theta)|^c$, 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 $a$:

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

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

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

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

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

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

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, Gadjrich, 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 at least $n^{1/2}$-convergent and Hill (2013)’s Least Tail-Trimmed Squares estimator is at least $n^{1/2}$-convergent under our stated error properties. QED.

**Lemma 3.2** Assumptions 1-3 hold, and the OLS estimator satisfies Assumption
Proof. \quad Since \( y_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i} \) is geometrically \( \beta \)-mixing by assumption and \( \epsilon_t(\theta) := y_t - \sum_{i=1}^{\infty} \psi_i y_{t-i} \) is an infinite order lag function of i.i.d. \( \epsilon_t \), 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 \( y_t = \theta^0 y_{t-1} + \nu_t \) for notational economy, and define \( \theta^0 := \frac{\sum_{i=0}^{\infty} \psi_i \psi_{i-1}}{\sum_{i=0}^{\infty} \psi_i^2} \). Since the least squares estimator \( \hat{\theta}_n = \frac{\sum_{i=2}^{n} y_i y_{i-1}}{\sum_{i=2}^{n} y_i^2} \) 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 \( \kappa > 1 \). \( \Omega \Xi \Delta \).

Lemma 3.3 \quad 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. \quad 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 \( \beta \)-mixing on some neighborhood \( N_0 \) of \( \theta^0 \) (see Doukhan (1994: Chapter 2.4)), hence \( \epsilon_t(\theta) = y_t / \sigma_t(\theta) \) is stationary geometrically \( \beta \)-mixing on \( N_0 \). 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 \( N_0 \) of \( \theta^0 \) that may depend on \( p \) (Francq and Zakoian (2004: eq. (4.25))). Therefore \( E[S_t(\theta)^{\kappa}] < \infty \) on \( N_0 \). But since \( \epsilon_t \) has tail (6) it follows \( \epsilon_t(\theta) = \epsilon_t S_t(\theta) \) also satisfies (6) for each \( \theta \) since by iterated expectations and independence

\[
P(\epsilon_t S_t(\theta) > a) = E \left[ P(\epsilon_t > a/S_t(\theta)|S_t(\theta)) \right]
\]

\[
= a^{-\kappa} E \left[ S_t(\theta)^{\kappa} \right] \left( 1 + \kappa a^\beta E \left[ S_t(\theta)^{\kappa-\beta} \right] / E \left[ S_t(\theta)^{\kappa} \right] \right)
\]

\[
= d(\theta) a^{-\kappa} \left( 1 + c(\theta) a^\beta \right).
\]

Francq and Zakoian (2004: eq. (4.25)) prove \( E|S_t(\theta)|^p < \infty \) on some neighborhood of \( \theta^0 \). Their argument generalizes to \( E|S_t(\theta)|^p < \infty \) for any \( p > 0 \) and neighborhoods \( N_0 \) of \( \theta^0 \) 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 $x_t(\theta) := (\epsilon_t^2(\theta) + \epsilon)^{1/2}$ 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 $\kappa$.

Now consider Assumption 4. By assumption $m_n = o(n^{1-\epsilon})$ necessarily holds for tiny $\epsilon > 0$ hence any $n^{1/2-\epsilon/2}\ln(n)$-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 $n^{1/2}/\mathcal{L}(n) \leq n^{1/2}/\mathcal{L}(n)$ 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. 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 $\psi$ lies in $\Psi$ a compact subset of $(0, \infty) \times (0, 1)$. Both $\epsilon_t$ and $y_t$ are geometrically $\beta$-mixing (Cline (2007)), hence the finite lag function

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

is for any $\theta \in \Theta$ geometrically $\beta$-mixing. This verifies Assumption 3. The weighted Laplace QML estimator of $\theta^0$ 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 $\epsilon_t$ 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. \( \epsilon_t \) 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 \( \epsilon_t \), 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. \( \Box \).

**Lemma 3.5** Assumptions 1 and 3 hold, and \( x_t \) has tail \( P(|x_t| > a) = da^{-\kappa}(1 + o(1)) \). If \( \epsilon_t \) 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 Mikosch (2002: Theorem 3.2)) and \( \{u_t, \sigma_t^2\} \) is stationary and geometrically \( \beta \)-mixing (Meitz and Saikkonen (2008: Theorem 1)). Finally, if the i.i.d. GARCH error \( \epsilon_t \) has a symmetric distribution then Hill (2014b)'s class of GEL estimators is \( n^{1/2} \)-convergent hence Assumption 4 holds. \( \Box \).

**References**


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