The University of Sydney Business School
The University of Sydney

BUSINESS ANALYTICS WORKING PAPER SERIES

Supplemental Material for GEL Estimation for Heavy-Tailed GARCH Models with Robust Empirical Likelihood Inference

Jonathan B. Hill
University of North Carolina Chapel Hill

Artem Prokhorov
University of Sydney

September 8, 2015

The following supplemental material contains an omitted simulation experiment, and omitted proofs of theorems and preliminary lemmata. Section S contains simulation results, and Section A contains an appendix with omitted proofs.

BA Working Paper No: BAWP-2015-04

http://sydney.edu.au/business/business_analytics/research/working_papers

Supplemental Material for GEL Estimation for Heavy-Tailed GARCH Models with Robust Empirical Likelihood Inference

Jonathan B. Hill*
University of North Carolina – Chapel Hill

Artem Prokhorov†
University of Sydney

September 8, 2015

The following supplemental material contains an omitted simulation experiment, and omitted proofs of theorems and preliminary lemmata. Section S contains simulation results, and Section A contains an appendix with omitted proofs.

S Simulation : Trimming Variations

In the main paper we reported GELITT simulation bias over a grid of trimming fractiles $\{k_n^{(\epsilon)}, k_n^{(y)}\}$. We now repreat the simulation and fix either $k_n^{(\epsilon)}$ or $k_n^{(y)}$, and report bias, mse, and test statistics.

We use $k_n^{(\epsilon)} \sim \lambda n/\ln(n)$, $\lambda n^{1/2}$ and $\lambda \ln(n)$ each with $k_n^{(y)} \sim .2\ln(n)$, and $k_n^{(y)} \sim \lambda n/\ln(n)$, $\lambda n^{1/2}$ and $\lambda \ln(n)$ each with $k_n^{(\epsilon)} \sim .05n/\ln(n)$. We summarize the various λ's and actual fractile values $\{k_n^{(\epsilon)}, k_n^{(y)}\}$ for $n = \{100, 250\}$ in the table below.

*Corresponding author. Dept. of Economics, University of North Carolina, Chapel Hill, NC; http://www.unc.edu/∼jbhill; jbhill@email.unc.edu

†Business School & CIREQ, University of Sydney; http://sydney.edu.au/business/staff/artemp; artem.prokhorov@sydney.edu.au.

Alternative Fractiles for $n = \{100, 250\}$

$k_n^{(\epsilon)}$			$k_n^{(y)}$	
$.01n/\ln(n)$	$\{1,1\}^a$		0	$\{0,0\}$
$.1n/\ln(n)$	$\{2,5\}$		$.1n/\ln(n)$	$\{2,5\}$
$.2n/\ln(n)$	$\{4,9\}$		$.2n/\ln(n)$	$\{4,9\}$
$.5n/\ln(n)$	$\{11,23\}$		$.5n/\ln(n)$	$\{11,23\}$
$.4n^{1/2}$	$\{4,6\}$		$1.75n^{1/2}$	$\{17,28\}$
$1.5\ln(n)$	$\{7,8\}$		$6\ln(n)$	$\{28,33\}$

a. Values are $k_n^{(\epsilon)}$ for $n = \{100, 250\}$.

See Tables A.1 and A.2 for simulation results. We find that many fractile values lead to roughly similar results. Overall, setting the error fractile $k_n^{(\epsilon)}$ to be small for each n is optimal, where greater bias and therefore t-test distortions arise when $k_n^{(\epsilon)}$ is larger. If we do not trim by y_t such that $k_n^{(y)} = 0$ then again there is bias. Furthermore, somewhat suprisingly trimming by a larger number of y_t extremes leads to better results than trimming by few values. That may arise since $k_n^{(\epsilon)}$ is small, as the next experiment demonstrates.

TABLE A.1: Trimming Variations : TT-CUE Results for $\theta_3^0 = .6$

$\epsilon_t \sim \bar{P}_{2.5}$ and $\kappa_y = 1.5$									
		$n = 100$				$n = 250$			
$k_n^{(\epsilon)}$	$k_n^{(y)}$	$\{k_n^{(\epsilon)}, k_n^{(y)}\}^a$	Bias	RMS	KS	$\{k_n^{(\epsilon)}, k_n^{(y)}\}$	Bias	RMS	KS
$\frac{.01n}{\ln(n)}$	$\mathbf{.2}\ln(\boldsymbol{n})$	$\{1,1\}$	.010	.165	1.73	$\{1,1\}$	.011	.149	1.12
$\boldsymbol{\frac{.05n}{\ln(n)}}$ [b]		$\boldsymbol{\{1,1\}}$	**.002**	**.169**	**1.03**	$\boldsymbol{\{2,1\}}$	**.001**	**.140**	**.895**
$\frac{.1n}{\ln(n)}$		$\{2,1\}$	.001	.166	1.01	$\{5,1\}$	.005	.134	1.03
$\frac{.2n}{\ln(n)}$		$\{4,1\}$	-.019	.177	1.22	$\{9,1\}$	.016	.148	1.75
$\frac{.5n}{\ln(n)}$		$\{11,1\}$	-.017	.176	1.11	$\{23,1\}$	.011	.134	1.54
$.8n^{1/2}$		$\{8,1\}$	-.016	.175	1.14	$\{13,1\}$	.020	.152	1.72
$1.5\ln(n)$		$\{7,1\}$	-.019	.179	1.35	$\{8,1\}$	.014	.140	1.34
$\boldsymbol{\frac{.05n}{\ln(n)}}$ [b]	0	$\{1,0\}$	.014	.167	1.74	$\{2,0\}$	.012	.140	1.23
	$\frac{.1n}{\ln(n)}$	$\{1,2\}$	-.005	.176	1.04	$\{2,5\}$	.004	.138	.995
	$\frac{.2n}{\ln(n)}$	$\{1,4\}$	-.007	.172	1.04	$\{2,9\}$	.009	.134	1.27
	$\frac{.5n}{\ln(n)}$	$\{1,11\}$	-.007	.161	.995	$\{2,23\}$	-.011	.150	1.16
	$1.75n^{1/2}$	$\{1,17\}$	-.009	.178	1.15	$\{2,28\}$	.012	.141	1.08
	$\mathbf{.2\ln(\boldsymbol{n})}$	$\boldsymbol{\{1,1\}}$	**.002**	**.169**	**1.03**	$\boldsymbol{\{2,1\}}$	**.001**	**.140**	**.895**
	$6\ln(n)$	$\{1,28\}$	.010	.165	1.64	$\{2,33\}$	.009	.160	1.52

$\epsilon_t \sim N(0,1)$ and $\kappa_y = 4.1$									
		$n = 100$				$n = 250$			
$k_n^{(\epsilon)}$	$k_n^{(y)}$	$\{k_n^{(\epsilon)}, k_n^{(y)}\}$	Bias	RMS	KS	$\{k_n^{(\epsilon)}, k_n^{(y)}\}$	Bias	RMS	KS
$\frac{.01n}{\ln(n)}$	$\mathbf{.2}\ln(\boldsymbol{n})$	$\{1,1\}$	-.012	.145	1.25	$\{1,1\}$	.004	.081	.801
$\boldsymbol{\frac{.05n}{\ln(n)}}$		$\boldsymbol{\{1,1\}}$	**-.004**	**.101**	**.987**	$\boldsymbol{\{2,1\}}$	**.002**	**.080**	**.687**
$\frac{.1n}{\ln(n)}$		$\{2,1\}$	-.009	.105	1.12	$\{5,1\}$	.008	.087	.772
$\frac{.2n}{\ln(n)}$		$\{4,1\}$	-.003	.102	.821	$\{9,1\}$	-.008	.073	.457
$\frac{.5n}{\ln(n)}$		$\{11,1\}$	-.022	.098	1.47	$\{23,1\}$	.009	.074	.845
$.8n^{1/2}$		$\{8,1\}$	-.009	.107	.969	$\{13,1\}$	.009	.088	1.04
$1.5\ln(n)$		$\{7,1\}$	-.006	.106	.985	$\{8,1\}$	.006	.077	.948
$\boldsymbol{\frac{.05n}{\ln(n)}}$	0	$\{1,0\}$	-.011	.106	1.22	$\{2,0\}$	.007	.082	.841
	$\frac{.1n}{\ln(n)}$	$\{1,2\}$	.003	.105	.774	$\{2,5\}$	.003	.076	.633
	$\frac{.2n}{\ln(n)}$	$\{1,4\}$	-.003	.110	.911	$\{2,9\}$	.001	.078	.649
	$\frac{.5n}{\ln(n)}$	$\{1,11\}$	-.006	.105	1.11	$\{2,23\}$	.008	.082	1.18
	$1.75n^{1/2}$	$\{1,17\}$	-.008	.102	1.20	$\{2,28\}$	-.006	.078	.737
	$\mathbf{.2\ln(\boldsymbol{n})}$	$\boldsymbol{\{1,1\}}$	**-.004**	**.101**	**.987**	$\boldsymbol{\{2,1\}}$	**.002**	**.080**	**.687**
	$6\ln(n)$	$\{1,28\}$	-.008	.112	1.17	$\{2,33\}$	-.010	.077	1.24

a. Displayed values for $k_e^\epsilon = k$ are $max\{1, k\}$.

b. The base-case is $\{k_e^{(\epsilon)}, k_n^{(y)}\} = \{.05n/\ln(n), .2\ln(n)\}$.

TABLE A.2 : Trimming Variations : TT-CUE t-tests[a] at 5% level for θ_3^0

$\epsilon_t \sim \bar{P}_{2.5}$ and $\kappa_y = 1.5$										
		$n = 100$				$n = 250$				
$k_n^{(\epsilon)}$	$k_n^{(y)}$	$\{k_n^{(\epsilon)}, k_n^{(y)}\}^b$	H_0	H_1^1	H_1^2	H_1^3	$\{k_n^{(\epsilon)}, k_n^{(y)}\}$	H_0	H_1^1	H_1^2	H_1^3
$\frac{.01n}{\ln(n)}$		$\{1,1\}$	.083[c]	.602	.817	.926	$\{1,1\}$	.067	.779	.945	.990
$\boldsymbol{\frac{.05n}{\ln(n)}}$ [b]		$\boldsymbol{\{1,1\}}$	**.084**	**.589**	**.821**	**.930**	$\boldsymbol{\{2,1\}}$	**.091**	**.863**	**.989**	**1.00**
$\frac{.1n}{\ln(n)}$		$\{2,1\}$	.089	.602	.829	.940	$\{5,1\}$	.098	.857	.973	1.00
$\frac{.2n}{\ln(n)}$	$\boldsymbol{.2\ln(n)}$	$\{4,1\}$	.082	.630	.815	.933	$\{9,1\}$	.086	.786	.958	1.00
$\frac{.5n}{\ln(n)}$		$\{11,1\}$	.095	.627	.847	.911	$\{23,1\}$	.087	.834	.968	1.00
$.8n^{1/2}$		$\{8,1\}$	.087	.624	.832	.902	$\{13,1\}$	.077	.753	.923	.981
$1.5\ln(n)$		$\{7,1\}$	.075	.669	.849	.933	$\{8,1\}$	.102	.833	.969	1.00
	0	$\{1,0\}$	.067	.696	.893	.964	$\{2,0\}$	.092	.814	.970	1.00
	$\frac{.1n}{\ln(n)}$	$\{1,2\}$	.097	.632	.840	.923	$\{2,5\}$	.087	.833	.956	1.00
$\boldsymbol{\frac{.05n}{\ln(n)}}$ [b]	$\frac{.2n}{\ln(n)}$	$\{1,4\}$	.078	.665	.865	.940	$\{2,9\}$	.057	.800	.987	1.00
	$\frac{.5n}{\ln(n)}$	$\{1,11\}$	.087	.714	.899	.966	$\{2,23\}$	.078	.718	.927	.982
	$1.75n^{1/2}$	$\{1,17\}$	.063	.526	.807	.937	$\{2,28\}$	.103	.789	.959	1.00
	$\boldsymbol{.2\ln(n)}$	$\boldsymbol{\{1,1\}}$	**.084**	**.589**	**.821**	**.930**	$\boldsymbol{\{2,1\}}$	**.091**	**.863**	**.989**	**1.00**
	$6\ln(n)$	$\{1,28\}$	.064	.674	.902	.958	$\{2,33\}$	.082	.756	.900	.956

$\epsilon_t \sim N(0,1)$ and $\kappa_y = 4.1$											
			$n = 100$					$n = 250$			
$k_n^{(\epsilon)}$	$k_n^{(y)}$	$\{k_n^{(\epsilon)}, k_n^{(y)}\}$	H_0	H_1^1	H_1^2	H_1^3	$\{k_n^{(\epsilon)}, k_n^{(y)}\}$	H_0	H_1^1	H_1^2	H_1^3
$\frac{.01n}{\ln(n)}$		$\{1,1\}$	.094	.830	.977	1.00	$\{1,1\}$	.101	1.00	1.00	1.00
$\boldsymbol{\frac{.05n}{\ln(n)}}$		$\boldsymbol{\{1,1\}}$	**.094**	**.834**	**.979**	**1.00**	$\boldsymbol{\{2,1\}}$	**.106**	**1.00**	**1.00**	**1.00**
$\frac{.1n}{\ln(n)}$		$\{2,1\}$	.090	.929	1.00	1.00	$\{5,1\}$	.121	1.00	1.00	1.00
$\frac{.2n}{\ln(n)}$	$\boldsymbol{.2\ln(n)}$	$\{4,1\}$	.106	.949	1.00	1.00	$\{9,1\}$	.108	1.00	1.00	1.00
$\frac{.5n}{\ln(n)}$		$\{11,1\}$	.081	.966	1.00	1.00	$\{23,1\}$	.114	1.00	1.00	1.00
$.8n^{1/2}$		$\{8,1\}$	.131	.925	1.00	1.00	$\{13,1\}$	.111	1.00	1.00	1.00
$1.5\ln(n)$		$\{7,1\}$	.094	.915	1.00	1.00	$\{8,1\}$	.095	1.00	1..00	1.00
	0	$\{1,0\}$	.093	.940	1.00	1.00	$\{2,0\}$	.111	1.00	1.00	1.00
	$\frac{.1n}{\ln(n)}$	$\{1,2\}$	.092	.953	1.00	1.00	$\{2,5\}$	.103	1.00	1.00	1.00
$\boldsymbol{\frac{.05n}{\ln(n)}}$	$\frac{.2n}{\ln(n)}$	$\{1,4\}$	.079	.823	1.00	1.00	$\{2,9\}$	.102	1.00	1.00	1.00
	$\frac{.5n}{\ln(n)}$	$\{1,11\}$	.095	.944	1.00	1.00	$\{2,23\}$	.091	1.00	1.00	1.00
	$1.75n^{1/2}$	$\{1,17\}$	.089	.960	1.00	1.00	$\{2,28\}$	.095	1.00	1.00	1.00
	$\boldsymbol{.2\ln(n)}$	$\boldsymbol{\{1,1\}}$	**.094**	**.834**	**.979**	**1.00**	$\boldsymbol{\{2,1\}}$	**.106**	**1.00**	**1.00**	**1.00**
	$6\ln(n)$	$\{1,28\}$	.079	.910	1.00	1.00	$\{2,33\}$	.109	1.00	1.00	1.00

a. The true $\theta_3^0 = .6$. The hypotheses are H_0: $\theta_3 = .6$; H_1^1: $\theta_3 = .5$; H_1^2: $\theta_3 = .35$; and H_1^3: $\theta_3 = 0$.
b. Displayed values for $k_e^{\epsilon} = k$ are $max\{1, k\}$.
c. Rejection frequencies at the 5% level.
d. The base-case is $\{k_e^{(\epsilon)}, k_n^{(y)}\} = \{.05n/\ln(n), .2\ln(n)\}$.

A Appendix: Omitted Proofs

A.1 Notation and Assumptions

Throughout $o_p(1)$ does not depend on θ and λ, unless otherwise specified. "*w.p.a.1*" means "*with probability approaching one*".

Recall

$$\Theta \subseteq \left\{\theta \in (0,\infty) \times (0,1) \times (0,1) : E\left[\ln\left(\alpha + \beta\epsilon_t^2\right)\right] < \infty\right\} \tag{A.1}$$

and

$$P\left(|\epsilon_t(\theta)| \geq c_n^{(\epsilon)}(\theta)\right) = \frac{k_n^{(\epsilon)}}{n} \text{ and } P\left(|w_{i,t}(\theta)| \geq c_{i,n}^{(w)}(\theta)\right) = \frac{k_{i,n}^{(w)}}{n} \tag{A.2}$$

and

$$\hat{\Lambda}_n(\theta) = \left\{\lambda : \lambda'\hat{m}_{n,t}^*(\theta) \in \mathcal{D},\ t = 1,2,...,n\right\} \text{ and } \Lambda_n = \left\{\lambda : \sup_{\theta\in\Theta}\left\|\lambda'\Sigma_n^{1/2}(\theta)\right\| \leq Kn^{-1/2}\right\}.$$

We require a criterion and moments based on the trimmed equations $m_{n,t}^*(\theta)$ that use non-stochastic thresholds:

$$\hat{Q}_n(\theta,\lambda) \equiv \frac{1}{n}\sum_{t=1}^n \rho\left(\lambda'\hat{m}_{n,t}^*(\theta)\right) \text{ and } \tilde{Q}_n(\theta,\lambda) \equiv \frac{1}{n}\sum_{t=1}^n \rho\left(\lambda' m_{n,t}^*(\theta)\right)$$

$$\Lambda_n = \left\{\lambda : \sup_{\theta\in\Theta}\left\|\lambda'\Sigma_n^{1/2}(\theta)\right\| \leq Kn^{-1/2}\right\}$$

$$m_n^*(\theta) \equiv \frac{1}{n}\sum_{t=1}^n m_{n,t}^*(\theta) \text{ and } \hat{m}_n^*(\theta) \equiv \frac{1}{n}\sum_{t=1}^n \hat{m}_{n,t}^*(\theta) \text{ and } \mathfrak{m}_n \equiv \sup_{\theta\in\Theta}\left\|E\left[m_{n,t}^*(\theta)\right]\right\|.$$

Asymptotic arguments require covariance and Jacobian components for tail-trimmed equations:

$$\hat{\Sigma}_n(\theta) \equiv \frac{1}{n}\sum_{t=1}^n \hat{m}_{n,t}^*(\theta)\hat{m}_{n,t}^*(\theta)' \text{ and } \tilde{\Sigma}_n(\theta) \equiv \frac{1}{n}\sum_{t=1}^n m_{n,t}^*(\theta)m_{n,t}^*(\theta)' \tag{A.3}$$

$$\begin{aligned}\widehat{\mathcal{J}}_{n,t}(\theta) \equiv &\left(\frac{\partial}{\partial\theta}\epsilon_t^2(\theta) \times \hat{I}_{n,t}^{(\epsilon)}(\theta) - \frac{1}{n}\sum_{t=1}^n \frac{\partial}{\partial\theta}\epsilon_t^2(\theta) \times \hat{I}_{n,t}^{(\epsilon)}(\theta)\right) x_t(\theta) \\ &+ \left(\epsilon_t^2(\theta)\hat{I}_{n,t}^{(\epsilon)}(\theta) - \frac{1}{n}\sum_{t=1}^n \epsilon_t^2(\theta)\hat{I}_{n,t}^{(\epsilon)}(\theta)\right)\frac{\partial}{\partial\theta}x_t(\theta)\end{aligned}$$

Non-negligible trimming, and distribution continuity and non-degeneracy, ensure

$$\liminf_{n\to\infty} \|\mathfrak{m}_n\| > 0 \text{ and } \liminf_{n\to\infty} \|\Sigma_n\| > 0, \text{ and } \Sigma_n^{-1} \text{ exists as } n \to \infty.$$

In order to reduce the number of cases and to keep notation simple, we assume wherever useful that we have exact identification:

$$x_{n,t}^* = s_t.$$

The proofs below extend to the over-identification case where w_t contains lags of s_t, and can be easily generalized to allow for other $\Im_{t-1}$-measurable w_t that require trimming. Similarly, we augment Assumption A.2 and impose power law tails on ϵ_t in general:

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

We compactly write throughout:

$$d = d_\epsilon, \;\; \kappa = \kappa_\epsilon, \;\; k_n = k_n^{(\epsilon)} \;\;\text{and}\;\; c_n = c_n^{(\epsilon)}.$$

Assumption A holds throughout. Then $\{y_t, \sigma_t^2(\theta)\}$ on Θ are stationary, ergodic, and geometrically β-mixing on Θ by (A.1), cf. Nelson (1990) and Carrasco and Chen (2002). Therefore, $w_t(\theta)$ is geometrically β-mixing since it is $\Im_{t-1}$-measurable, and $\epsilon_t(\theta) = \epsilon_t\sigma_t/\sigma_t(\theta)$ is stationary and ergodic.

Since $E(\sup_{\theta\in\Theta} |\sigma_t^2/\sigma_t^2(\theta)|)^p < \infty$ for any $p > 0$, cf. Francq and Zakoïan (2004, eq. (4.25)), it follows the product convolution $\epsilon_t(\theta) = \epsilon_t\sigma_t/\sigma_t(\theta)$ has a power law tail with the same index κ > 2 (Breiman, 1965):

$$P\left(|\epsilon_t(\theta)| > a\right) = d(\theta)a^{-\kappa}\left(1 + o\left(1\right)\right) \tag{A.5}$$

$$\text{where} \quad \inf_{\theta\in\Theta} d(\theta) \in (0,\infty) \;\text{ and }\; o\left(1\right) \;\text{ does not depend on } \theta.$$

By construction of $c_n(\theta)$ in (A.2), therefore,

$$c_n(\theta) = d(\theta)^{1/\kappa}\left(n/k_n\right)^{1/\kappa}. \tag{A.6}$$

Similarly $\sup_{\theta\in\mathcal{N}_0} |s_{i,t}(\theta)|$ is L_p-bounded for any $p > 2$ and some compact subset $\mathcal{N}_0 \subseteq \Theta$ containing θ^0. This follows by a trivial generalization of arguments in Francq and Zakoïan (2004, Section 4.2). Therefore, in the exact identification case by independence $m_{i,t}(\theta) = (\epsilon_t^2(\theta) - 1)s_{i,t}(\theta) = (\epsilon_t^2\sigma_t^2/\sigma_t^2(\theta) - 1)s_{i,t}(\theta)$ has a power-law tail with index $\kappa/2$ (see, e.g., Breiman, 1965):

$$P\left(|m_{i,t}(\theta)| > a\right) = d_i(\theta)a^{-\kappa/2}\left(1 + o\left(1\right)\right) \tag{A.7}$$

$$\text{where } \inf_{\theta\in\Theta} d_i(\theta) \in (0,\infty) \text{ and } o\left(1\right) \text{ does not depend on } \theta.$$

The trimmed moment $\mathfrak{E}_n(\theta) \equiv E[\epsilon_t^4(\theta)I(|\epsilon_t(\theta)| \leq c_n(\theta))]$ can be characterized by case by invoking (A.5), (A.6) and Karamata's Theorem (cf. Theorem 0.6 in Resnick, 1987):

$$\text{if } \kappa = 4: \ (0,\infty) \leftarrow \inf_{\theta\in\Theta}\left\{\frac{\mathfrak{E}_n(\theta)}{\ln(n)}\right\} \leq \sup_{\theta\in\Theta}\left\{\frac{\mathfrak{E}_n(\theta)}{\ln(n)}\right\} \to (0,\infty) \tag{A.8}$$

$$\text{if } \kappa < 4: \ (0,\infty) \leftarrow \inf_{\theta\in\Theta}\left\{\frac{\mathfrak{E}_n(\theta)}{c_n^4(\theta)(k_n/n)}\right\} \leq \sup_{\theta\in\Theta}\left\{\frac{\mathfrak{E}_n(\theta)}{c_n^4(\theta)(k_n/n)}\right\} \to (0,\infty)\,.$$

Similarly, by (A.7) and Karamata's Theorem, $\mathfrak{M}_{i,j,n}(\theta) \equiv E[m_{i,n,t}^*(\theta)m_{j,n,t}^*(\theta)]$ satisfies

$$\text{if } \kappa = 4: \ (0,\infty) \leftarrow \inf_{\theta\in\Theta}\left\{\frac{\mathfrak{M}_{i,j,n}(\theta)}{\ln(n)}\right\} \leq \sup_{\theta\in\Theta}\left\{\frac{\mathfrak{M}_{i,j,n}(\theta)}{\ln(n)}\right\} \to (0,\infty) \tag{A.9}$$

$$\text{if } \kappa < 4: \ (0,\infty) \leftarrow \inf_{\theta\in\Theta}\left\{\frac{\mathfrak{M}_{i,j,n}(\theta)}{c_n^4(\theta)(k_n/n)}\right\} \leq \sup_{\theta\in\Theta}\left\{\frac{\mathfrak{M}_{i,j,n}\theta)}{c_n^4(\theta)(k_n/n)}\right\} \to (0,\infty)\,.$$

Assumption A.

1. $z_t(\theta) \in \{\epsilon_t(\theta), w_{i,t}(\theta)\}$ *have for each* $\theta \in \Theta$ *strictly stationary, ergodic, and absolutely continuous non-degenerate finite dimensional distributions that are uniformly bounded:*

$$\sup_{a\in\mathbb{R},\theta\in\Theta}\left\{\frac{\partial}{\partial a}P\left(z_t(\theta) \leq a\right)\right\} < \infty \text{ and } \sup_{a\in\mathbb{R},\theta\in\Theta}\left\{\frac{\partial}{\partial \theta}P\left(z_t(\theta) \leq a\right)\right\} < \infty.$$

2. $\kappa_i > 1$ *and* $\kappa_\epsilon > 2$*. If* $\kappa_\epsilon \leq 4$ *then* $P(|\epsilon_t| > a) = da^{-\kappa}(1 + o\left(1\right))$ *where* $d \in (0,\infty)$*. If* $\Theta_{1,i}$ *is not empty such that* $\kappa_i(\theta) \leq 1$ *for some* θ*, then*

$$P\left(|w_{i,t}(\theta)| > c\right) = d_i(\theta)c^{-\kappa_i(\theta)}(1 + o(1)),$$

where $\inf_{\theta\in\Theta_{1,i}} d_i(\theta) > 0$, $\inf_{\theta\in\Theta_{1,i}} \kappa_i(\theta) > 0$ *and* $o(1)$ *is not a function of* θ*.*

3. $w_t(\theta)$ *is* $\Im_{t-1}$*-measurable, continuous, differentiable, and* $E[\sup_{\theta\in\Theta}|w_{i,t}(\theta)|^\iota] < \infty$ *for some tiny* $\iota > 0$*.*

4. $k_n/n^{\iota} \to \infty$ *for some tiny* $\iota > 0$.

A.2 Theorem 2.5 (higher order expansion)

Let $\{z^*_{n,t}\}$ be tail-trimmed randoms variable and write $\tilde{z}_n \equiv 1/n^{1/2}\sum_{t=1}^n z^*_{n,t}$. Let $z^*_{n,t}(\theta) \equiv z_t(\theta)I_{n,t}(\theta)$ where $z_t(\theta)$ is differentiable, $I_{n,t}(\theta) \in \{0,1\}$ and $\inf_{\theta\in\Theta} I_{n,t}(\theta) \stackrel{p}{\to} 1$, and define

$$\frac{\mathring{\partial}}{\mathring{\partial}\theta} z^*_{n,t}(\theta) \equiv \left(\frac{\partial}{\partial\theta} z_t(\theta)\right) \times I_{n,t}(\theta).$$

Define

$$\mathfrak{M}^*_{n,t}(\beta) \equiv \rho^{(1)}\left(\lambda' m^*_{n,t}(\theta)\right) \times \left[\begin{array}{l} \frac{\mathring{\partial}}{\mathring{\partial}\theta} m^*_{n,t}(\theta)'\lambda \\ m^*_{n,t}(\theta) \end{array}\right]$$

$$\mathfrak{G}^*_n(\beta) \equiv E\left[\frac{\mathring{\partial}}{\mathring{\partial}\beta}\mathfrak{M}^*_{n,t}(\beta)\right], \; \mathfrak{G}^*_{j,n}(\beta) \equiv E\left[\frac{\mathring{\partial}^2}{\mathring{\partial}\beta_j\mathring{\partial}\beta}\mathfrak{M}^*_{n,t}(\beta)\right], \; \mathfrak{G}^*_{j,k,n}(\beta) \equiv E\left[\frac{\mathring{\partial}^3}{\mathring{\partial}\beta_j\mathring{\partial}\beta_k\mathring{\partial}\beta}\mathfrak{M}^*_{n,t}(\beta)\right]$$

$$\mathfrak{A}^*_{n,t} \equiv \frac{\mathring{\partial}}{\mathring{\partial}\beta}\mathfrak{M}^*_{n,t} - \mathfrak{G}^*_n, \; \mathfrak{B}^*_{j,n,t} \equiv \frac{\mathring{\partial}^2}{\mathring{\partial}\beta_j\mathring{\partial}\beta}\mathfrak{M}^*_{n,t} - \mathfrak{G}^*_{j,n} \text{ and } \psi^*_{n,t} \equiv -\mathfrak{G}^{*-1}_n\mathfrak{M}^*_{n,t}.$$

Recall we assume over-identifying restrictions are square integrable (e.g. they lags of $s_t(\theta)$) and therefore need not be trimmed:

$$m^*_{n,t}(\theta) = \left(\epsilon^{*2}_{n,t}(\theta) - E\left[\epsilon^{*2}_{n,t}(\theta)\right]\right)\left(x_t(\theta) - E\left[x_t(\theta)\right]\right) \quad \text{where} \quad \epsilon^*_{n,t}(\theta) \equiv \epsilon_t(\theta) I^{(\epsilon)}_{n,t}(\theta). \tag{A.10}$$

Recall Paretian tails ensures by Karamata theory (cf. Resnick, 1987, Theorem 0.6)

$$p > \kappa_\epsilon : E\left|\epsilon^*_{n,t}\right|^p \sim \frac{p}{p-\kappa_\epsilon}\left(c^{(\epsilon)}_n\right)^p P\left(|\epsilon_t| > c^{(\epsilon)}_n\right) = \frac{p}{p-\kappa_\epsilon} d^{p/\kappa_\epsilon}\left(\frac{n}{k^{(\epsilon)}_n}\right)^{4/\kappa_\epsilon - 1} \tag{A.11}$$

$$p = \kappa_\epsilon : E\left|\epsilon^*_{n,t}\right|^p \sim d\ln(n).$$

Theorem 2.5. *Under Assumption A and* $||E[w_t w_t']|| < \infty$*:*

$$\hat{\beta}_n - \beta^0 = \frac{1}{n^{1/2}}\tilde{\psi}^*_n + \frac{1}{n}Q_1\left(\tilde{\psi}^*_n\right) + \frac{1}{n^{3/2}}Q_2\left(\tilde{\psi}^*_n\right) + O_p\left(\frac{\left(E\left[\epsilon^{*4}_{n,t}\right]\right)^2}{n^2}\right), \tag{A.12}$$

where

$$Q_1\left(\tilde{\psi}_n^*\right) \equiv -\mathfrak{G}_n^{*-1}\left\{\widetilde{\mathfrak{A}}_n^*\tilde{\psi}_n^* + \frac{1}{2}\sum_{i=1}^{q+3}\tilde{\psi}_{i,n}^*\mathfrak{G}_{i,n}^*\tilde{\psi}_n^*\right\}$$

$$Q_2\left(\tilde{\psi}_n^*\right) \equiv -\mathfrak{G}_n^{*-1}\mathfrak{Q}_n$$

$$\mathfrak{Q}_n \equiv \widetilde{\mathfrak{A}}_n^* Q_1\left(\tilde{\psi}_n^*\right) + \frac{1}{2}\sum_{i=1}^{q+3}\left\{\tilde{\psi}_{i,n}^*\mathfrak{G}_{i,n}^* Q_1\left(\tilde{\psi}_n^*\right) + Q_{i,1}\left(\tilde{\psi}_n^*\right)\mathfrak{G}_{i,n}^*\tilde{\psi}_n^* + \tilde{\psi}_{i,n}^*\mathfrak{G}_{i,n}^*\tilde{\psi}_n^*\right\}$$
$$+ \frac{1}{6}\sum_{i,j=1}^{q+3}\tilde{\psi}_{i,n}^*\tilde{\psi}_{j,n}^*\mathfrak{G}_{i,j,n}^*\tilde{\psi}.$$

If $k_n^{(\epsilon)} \sim n/L(n)$ for some slowly varying $L(n) \to \infty$ then for any $\kappa_\epsilon > 2$:

$$\hat{\beta}_n - \beta^0 = \frac{1}{n^{1/2}}\tilde{\psi}_n^* + \frac{1}{n}Q_1\left(\tilde{\psi}_n^*\right) + O_p\left(\frac{L(n)}{n^{3/2}}\right) \text{ for slowly varying } L(n) \to \infty. \tag{A.13}$$

Proof. Observe that

$$\frac{\mathring{\partial}}{\mathring{\partial}\theta}m_{n,t}^* = \frac{\mathring{\partial}}{\mathring{\partial}\theta}\left(\epsilon_{n,t}^{*2} - E\left[\epsilon_{n,t}^{*2}\right]\right) \times \left(x_t - E\left[x_t\right]\right) + \left(\epsilon_{n,t}^{*2} - E\left[\epsilon_{n,t}^{*2}\right]\right) \times \frac{\partial}{\partial\theta}\left(x_t - E\left[x_t\right]\right)$$

$$\frac{\mathring{\partial}^2}{\mathring{\partial}\theta_i\mathring{\partial}\theta}m_{n,t}^*(\theta) = \frac{\mathring{\partial}^2}{\mathring{\partial}\theta_i\mathring{\partial}\theta}\left(\epsilon_{n,t}^{*2} - E\left[\epsilon_{n,t}^{*2}\right]\right) \times \left(x_t - E\left[x_t\right]\right) + \frac{\mathring{\partial}}{\mathring{\partial}\theta}\left(\epsilon_{n,t}^{*2} - E\left[\epsilon_{n,t}^{*2}\right]\right) \times \frac{\partial}{\partial\theta_i}\left(x_t - E\left[x_t\right]\right)$$

and so on for $(\mathring{\partial}^3/\mathring{\partial}\theta_i\mathring{\partial}\theta_j\mathring{\partial}\theta)m_{n,t}^*$ and $(\mathring{\partial}^4/\mathring{\partial}\theta_i\mathring{\partial}\theta_j\mathring{\partial}\theta_k\mathring{\partial}\theta)m_{n,t}^*$. Hence by the asymptotic theory developed in the appendices:

$$\frac{1}{n}\sum_{t=1}^{n}\mathfrak{M}_{n,t}^* = O_p\left(\left(E\left[\epsilon_{n,t}^{*4}\right]/n\right)^{1/2}\right)$$

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\mathring{\partial}}{\mathring{\partial}\beta}\mathfrak{M}_{n,t}^* = \mathfrak{G}_n^* \times \left(1 + O_p\left(\left(E\left[\epsilon_{n,t}^{*4}\right]/n\right)^{1/2}\right)\right)$$

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\mathring{\partial}^2}{\mathring{\partial}\beta_i\mathring{\partial}\beta}\mathfrak{M}_{n,t}^* = \mathfrak{G}_{j,n}^* \times \left(1 + O_p\left(\left(E\left[\epsilon_{n,t}^{*4}\right]/n\right)^{1/2}\right)\right)$$

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\mathring{\partial}^3}{\mathring{\partial}\beta_i\mathring{\partial}\beta_j\mathring{\partial}\beta}\mathfrak{M}_{n,t}^* = \mathfrak{G}_{j,k,n}^* \times \left(1 + O_p\left(\left(E\left[\epsilon_{n,t}^{*4}\right]/n\right)^{1/2}\right)\right).$$

Further, $(\mathring{\partial}/\mathring{\partial}\beta)\mathfrak{M}^*_{n,t}$ has elements either 0 or $E[(\mathring{\partial}/\mathring{\partial}\theta)m^*_{n,t}]$ hence

$$\|\mathfrak{G}^*_n\| = K \times E\left[\frac{\mathring{\partial}}{\mathring{\partial}\theta}m^*_{n,t}\right]\left(1 + O_p\left(\left(E\left[\epsilon^{*4}_{n,t}\right]/n\right)^{1/2}\right)\right) \sim K.$$

Expand:

$$\begin{aligned} 0 \;=\; & \frac{1}{n}\sum_{t=1}^{n}\mathfrak{M}^*_{n,t}\left(1+o_p(1)\right) + \frac{1}{n}\sum_{t=1}^{n}\frac{\mathring{\partial}}{\mathring{\partial}\beta}\mathfrak{M}^*_{n,t}\left(\hat{\beta}_n - \beta^0\right)\left(1+o_p(1)\right) \\ & + \sum_{i=1}^{q+3}\left(\hat{\beta}_{i,n} - \beta^0_i\right)\frac{1}{2n}\sum_{t=1}^{n}\frac{\mathring{\partial}^2}{\mathring{\partial}\beta_i\mathring{\partial}\beta}\mathfrak{M}^*_{n,t}\left(\hat{\beta}_n - \beta^0\right)\left(1+o_p(1)\right) \\ & + \sum_{i,j=1}^{q+3}\left(\hat{\beta}_{i,n} - \beta^0_i\right)\left(\hat{\beta}_{j,n} - \beta^0_j\right)\frac{1}{6n}\sum_{t=1}^{n}\frac{\mathring{\partial}^3}{\mathring{\partial}\beta_i\mathring{\partial}\beta_j\mathring{\partial}\beta}\mathfrak{M}^*_{n,t}(\tilde{\beta}_n)\left(\hat{\beta}_n - \beta^0\right)\left(1+o_p(1)\right). \end{aligned} \tag{A.14}$$

The derivatives are valid asymptotically with probability approaching one, as fast as we choose. This follows from arguments used to prove Theorem 2.2, cf. Cizek (2008, Appendices), but also from indicator smoothing arguments used in similar proofs in Hill (2012, 2015b, 2013, 2015a). In the following we drop $o_p(1)$ in reduce notation: all subsequent equalities resulting from (A.14) hold with probability approaching one.

By Remark 6 of Theorem 2.2 and the equation form (A.10), it follows $||\tilde{\beta}_n - \beta^0|| = O_p\left(||\mathcal{A}_n||^{-1/2}\right)$ $= O_p\left((E\left[\epsilon^{*4}_{n,t}\right]/n)^{1/2}\right)$, hence:

$$\begin{aligned} & \left\|\frac{1}{n}\sum_{t=1}^{n}\frac{\mathring{\partial}^3}{\mathring{\partial}\beta_i\mathring{\partial}\beta_j\mathring{\partial}\beta}\mathfrak{M}^*_{n,t}(\tilde{\beta}_n) - \mathfrak{G}^*_{j,k,n}\right\| \\ & \qquad \leq \left\|\frac{1}{n}\sum_{t=1}^{n}\frac{\mathring{\partial}^3}{\mathring{\partial}\beta_i\mathring{\partial}\beta_j\mathring{\partial}\beta}\mathfrak{M}^*_{n,t} - \mathfrak{G}^*_{j,k,n}\right\| + \frac{1}{n}\sum_{t=1}^{n}\left\|\frac{\mathring{\partial}^3}{\mathring{\partial}\beta_i\mathring{\partial}\beta_j\mathring{\partial}\beta}\left\{\mathfrak{M}^*_{n,t}(\tilde{\beta}_n) - \mathfrak{M}^*_{n,t}\right\}\right\| \\ & \qquad \leq \left\|\frac{1}{n}\sum_{t=1}^{n}\frac{\mathring{\partial}^3}{\mathring{\partial}\beta_i\mathring{\partial}\beta_j\mathring{\partial}\beta}\mathfrak{M}^*_{n,t} - \mathfrak{G}^*_{j,k,n}\right\| + K\left\|\tilde{\beta}_n - \beta^0\right\| \\ & \qquad = O_p\left(\left(\frac{E\left[\epsilon^{*4}_{n,t}\right]}{n}\right)^{1/2}\right). \end{aligned}$$

Therefore

$$\hat{\beta}_n - \beta^0 \;=\; \frac{1}{n^{1/2}}\tilde{\psi}^*_n \tag{A.15}$$

$$-\mathfrak{G}_n^{*-1}\left\{\frac{1}{n^{1/2}}\widetilde{\mathfrak{A}}_n^*\left(\hat{\beta}_n-\beta^0\right)+\frac{1}{2}\sum_{i=1}^{q+3}\left(\hat{\beta}_{i,n}-\beta_i^0\right)\mathfrak{G}_{j,n}^*\left(\hat{\beta}_n-\beta^0\right)\right\}$$
$$-\mathfrak{G}_n^{*-1}\frac{1}{2}\sum_{i=1}^{q+3}\left(\hat{\beta}_{i,n}-\beta_i^0\right)\frac{1}{n^{1/2}}\widetilde{\mathfrak{B}}_{j,n}^*\left(\hat{\beta}_n-\beta^0\right)$$
$$-\mathfrak{G}_n^{*-1}\frac{1}{6}\sum_{i,j=1}^{q+3}\left(\hat{\beta}_{i,n}-\beta_i^0\right)\left(\hat{\beta}_{j,n}-\beta_j^0\right)\mathfrak{G}_{i,j,n}^*\left(\hat{\beta}_n-\beta^0\right)+O_p\left(\frac{\left(E\left[\epsilon_{n,t}^{*4}\right]\right)^2}{n^2}\right).$$

All terms except $\tilde{\psi}_n^*/n^{1/2}$ are $O_p(E\left[\epsilon_{n,t}^{*4}\right]/n)$ hence $\hat{\beta}_n-\beta^0=\tilde{\psi}_n^*/n^{1/2}+O_p(E\left[\epsilon_{n,t}^{*4}\right]/n)$. The last three terms are $O_p((E\left[\epsilon_{n,t}^{*4}\right]/n)^{3/2})$, and replacing $\hat{\beta}_n-\beta^0$ with $\tilde{\psi}_n^*/n^{1/2}$ in the second and third terms of (A.15) generates an error of order $O_p((E\left[\epsilon_{n,t}^{*4}\right]/n)^{3/2})$. Therefore

$$\begin{aligned}\hat{\beta}_n-\beta^0 &= \frac{1}{n^{1/2}}\tilde{\psi}_n^*-\mathfrak{G}_n^{*-1}\frac{1}{n}\left\{\widetilde{\mathfrak{A}}_n^*\tilde{\psi}_n^*+\frac{1}{2}\sum_{i=1}^{q+3}\tilde{\psi}_{i,n}^*\mathfrak{G}_{i,n}^*\tilde{\psi}_n^*\right\}+O_p\left(\left(\frac{E\left[\epsilon_{n,t}^{*4}\right]}{n}\right)^{3/2}\right)\\ &= \frac{1}{n^{1/2}}\tilde{\psi}_n^*+\frac{1}{n}Q_1\left(\tilde{\psi}_n^*,\widetilde{\mathfrak{A}}_n^*\right)+O_p\left(\left(\frac{E\left[\epsilon_{n,t}^{*4}\right]}{n}\right)^{3/2}\right).\end{aligned}$$

Now, in (A.15) replace $\hat{\beta}_n-\beta^0$ with $\tilde{\psi}_n^*/n^{1/2}+Q_1(\tilde{\psi}_n^*,\widetilde{\mathfrak{A}}_n^*)$ in the second and third terms, and with $\tilde{\psi}_n^*/n^{1/2}$ in the fourth and fith terms, to deduce (A.16).

Finally, consider (A.13). By construction

$$\hat{\beta}_n-\beta^0=\frac{1}{n^{1/2}}\tilde{\psi}_n^*+\frac{1}{n}Q_1\left(\tilde{\psi}_n^*\right)+O_p\left(\max\left\{\frac{\left(E\left[\epsilon_{n,t}^{*4}\right]\right)^2}{n^2},\frac{E\left[\epsilon_{n,t}^{*10}\right]}{n^{3/2}}\right\}\right), \tag{A.16}$$

by independence of the errors $E[\tilde{\psi}_n^*]=0$, and if $E[\epsilon_t^4]=\infty$ then by (A.11) we have:

$$\begin{aligned}\max\left\{\frac{\left(E\left[\epsilon_{n,t}^{*4}\right]\right)^2}{n^2},\frac{E\left[\epsilon_{n,t}^{*10}\right]}{n^{3/2}}\right\} &= K\max\left\{\frac{\left(\left(n/k_n^{(\epsilon)}\right)^{4/\kappa_\epsilon-1}\right)^2}{n^2},\frac{\left(n/k_n^{(\epsilon)}\right)^{10/\kappa_\epsilon-1}}{n^{3/2}}\right\}\\ &= K\frac{\left(n/k_n^{(\epsilon)}\right)^{10/\kappa_\epsilon-1}}{n^{3/2}}.\end{aligned}$$

Hence, irrespective of heavy tails in ϵ_t, the asymptotic higher order bias is

$$E\left[\hat{\beta}_n-\beta^0\right]=\frac{1}{n}E\left[Q_1\left(\tilde{\psi}_n^*\right)\right]$$

only if the remaining term in (A.16) vanishes rapidly enough. This holds for any $\kappa_\epsilon > 2$, for example, whenever $k_n^{(\epsilon)} \sim n/L(n)$ for some slowly varying $L(n) \to \infty$ since

$$\hat{\beta}_n - \beta^0 = \frac{1}{n^{1/2}}\tilde{\psi}_n^* + \frac{1}{n}Q_1\left(\tilde{\psi}_n^*\right) + O_p\left(\frac{L(n)}{n^{3/2}}\right) \text{ for slowly varying } L(n) \to \infty.$$

This completes the proof. $\mathcal{QED}$.

A.3 Theorem 2.6 (higher order bias)

Theorem 2.6. *Write $X_t \equiv x_t - E[x_t]$ and $S_t \equiv s_t - E[s_t]$, and define*

$$\mathcal{E}_n^{(1)} \equiv E\left[\epsilon_{n,t}^{*2}\right],\ \mathcal{E}_n^{(i)} \equiv E\left[\left(\epsilon_{n,t}^{*2} - E\left[\epsilon_{n,t}^{*2}\right]\right)^i\right] \text{ for } i = 2, 3$$

$$\mathcal{J} = -E[X_t'S_t],\ \Sigma_x \equiv E[X_tX_t'],\ \mathcal{H} \equiv \left(\mathcal{J}'\Sigma_x^{-1}\mathcal{J}\right)^{-1}\mathcal{J}'\Sigma_x^{-1} \in \mathbb{R}^{3\times q}$$

$$\mathcal{P} \equiv \Sigma_x^{-1} - \Sigma_x^{-1}\mathcal{J}(\mathcal{J}'\Sigma_x^{-1}\mathcal{J})^{-1}\mathcal{J}'\Sigma_x^{-1},$$

and $a = [a_j]_{j=1}^q$ where

$$a_j \equiv \frac{1}{2}tr\left\{\left(\mathcal{J}'\Sigma_x^{-1}\mathcal{J}\right)^{-1} \times E\left[\frac{\partial^2}{\partial\theta\partial\theta'}\left(\epsilon_t^2 - 1\right)X_{j,t}\right]\right\}.$$

Under Assumption A and $||E[w_tw_t']|| < \infty$:

$$Bias\left(\hat{\beta}_n\right) = \frac{1}{n}\left[\begin{array}{c} \frac{1}{\mathcal{E}_n^{(1)}}\mathcal{H}\left\{\frac{\mathcal{E}_n^{(2)}}{\mathcal{E}_n^{(1)}}\left(-a + E\left[S_tX_t'\mathcal{H}X_t\right]\right) + \frac{\mathcal{E}_n^{(3)}}{\mathcal{E}_n^{(2)}}\left(1 + \frac{\rho_3}{2}\right)E\left[X_t'X_t\mathcal{P}X_t\right]\right\} \\ \\ \frac{1}{\mathcal{E}_n^{(2)}}\mathcal{P}\left\{\frac{\mathcal{E}_n^{(2)}}{\mathcal{E}_n^{(1)}}\left(-a + E\left[S_tX_t'\mathcal{H}X_t\right]\right) + \frac{\mathcal{E}_n^{(3)}}{\mathcal{E}_n^{(2)}}\left(1 + \frac{\rho_3}{2}\right)E\left[X_t'X_t\mathcal{P}X_t\right]\right\} \end{array}\right].$$

Proof. Let e_i be the unit vector, write

$$X_t(\theta) \equiv x_t(\theta) - E\left[x_t(\theta)\right] \quad \text{and} \quad S_t(\theta) \equiv s_t(\theta) - E\left[s_t(\theta)\right]$$

and define

$$\Sigma_x \equiv E\left[X_tX_t'\right] \quad \text{and} \quad \tilde{\mathcal{J}}_n \equiv E\left[\frac{\mathring{\partial}}{\mathring{\partial}\theta}m_{n,t}^*\right] \quad \text{and} \quad \mathcal{J} = -E\left[\left(x_t - E\left[x_t\right]\right)\left(s_t - E\left[s_t\right]\right)'\right]$$

$$\tilde{\mathcal{H}}_n \equiv \left(\tilde{\mathcal{J}}_n' \Sigma_n^{-1} \tilde{\mathcal{J}}_n\right)^{-1} \tilde{\mathcal{J}}_n' \Sigma_n^{-1} \text{ and } \mathcal{H} \equiv \left(\mathcal{J}' \Sigma_x^{-1} \mathcal{J}\right)^{-1} \mathcal{J} \Sigma_x^{-1}$$

$$\tilde{\mathcal{P}}_n \equiv \Sigma_n^{-1} - \Sigma_n^{-1} \tilde{\mathcal{J}}_n \left(\tilde{\mathcal{J}}_n' \Sigma_n^{-1} \tilde{\mathcal{J}}_n\right)^{-1} \tilde{\mathcal{J}}_n \Sigma_n^{-1} \text{ and } \mathcal{P} = \Sigma_x^{-1} - \Sigma_x^{-1} \mathcal{J} \left(\mathcal{J}' \Sigma_x^{-1} \mathcal{J}\right)^{-1} \mathcal{J}' \Sigma_x^{-1}$$

$$\mathcal{E}_n^{(1)} \equiv E\left[\epsilon_{n,t}^{*2}\right] \text{ and } \mathcal{E}_n^{(i)} \equiv E\left[\left(\epsilon_{n,t}^{*2} - E\left[\epsilon_{n,t}^{*2}\right]\right)^i\right] \text{ for } i = 2, 3$$

$$a_{n,j} \equiv \frac{1}{2} \times tr\left\{\left(\tilde{\mathcal{J}}_n' \Sigma_n^{-1} \tilde{\mathcal{J}}_n\right)^{-1} \times E\left[\frac{\mathring{\partial}^2}{\mathring{\partial}\theta \mathring{\partial}\theta'} m_{j,n,t}^*\right]\right\} \text{ and } a_n \equiv \left[a_{n,j}\right]_{j=1}^q .$$

Hence:[1]

$$\Sigma_n = \mathcal{E}_n^{(2)} \Sigma_x \text{ and } \tilde{\mathcal{J}}_n = -\mathcal{E}_n^{(1)} \times E\left[\left(x_t - E\left[x_t\right]\right)\left(s_t - E\left[s_t\right]\right)'\right] = \mathcal{E}_n^{(1)} \mathcal{J} \tag{A.17}$$

$$\tilde{\mathcal{H}}_n = \frac{1}{\mathcal{E}_n^{(1)}} \mathcal{H} \text{ and } \tilde{\mathcal{P}}_n = \frac{1}{\mathcal{E}_n^{(2)}} \mathcal{P}$$

$$a_{n,j} = \frac{\mathcal{E}_n^{(2)}}{\mathcal{E}_n^{(1)}} \times \frac{1}{2} tr\left\{\left(\mathcal{J}' \Sigma_x^{-1} \mathcal{J}\right)^{-1} \times E\left[\frac{\partial^2}{\partial\theta \partial\theta'}\left(\epsilon_t^2 - 1\right) X_{j,t}\right]\right\} = \frac{\mathcal{E}_n^{(2)}}{\mathcal{E}_n^{(1)}} \times a_j$$

$$\mathfrak{G}_n^* = -\begin{bmatrix} 0 & \tilde{\mathcal{J}}_n' \\ \tilde{\mathcal{J}}_n & \Sigma_n \end{bmatrix} \text{ and } \mathfrak{G}_n^{*-1} = -\begin{bmatrix} -\Sigma_n & \tilde{\mathcal{H}}_n \\ \tilde{\mathcal{H}}_n' & \tilde{\mathcal{P}}_n \end{bmatrix} = -\begin{bmatrix} -\mathcal{E}_n^{(2)} \Sigma_x & \frac{1}{\mathcal{E}_n^{(1)}} \mathcal{H}' \\ \frac{1}{\mathcal{E}_n^{(1)}} \mathcal{H}' & \frac{1}{\mathcal{E}_n^{(2)}} \mathcal{P} \end{bmatrix}$$

and if $\beta_j = \theta_j$ $(j = 1, 2, 3)$:

$$\mathfrak{G}_{j,n}^* = -\begin{bmatrix} 0 & E\left[\frac{\mathring{\partial}^2 m_{n,t}^*}{\mathring{\partial}\theta_j \mathring{\partial}\theta'}\right] \\ E\left[\frac{\mathring{\partial}^2 m_{n,t}^*}{\mathring{\partial}\theta_j \mathring{\partial}\theta}\right] & E\left[\frac{\mathring{\partial} m_{n,t}^*}{\mathring{\partial}\theta_j} m_{n,t}^{*\prime}\right] + E\left[m_{n,t}^* \frac{\mathring{\partial} m_{n,t}^{*\prime}}{\mathring{\partial}\theta_j}\right] \end{bmatrix}$$

or if $\beta_j = \lambda_j$ $(j > 3)$:

$$\mathfrak{G}_{j,n}^* = -\begin{bmatrix} E\left[\frac{\mathring{\partial}^2 m_{j,n,t}^*}{\mathring{\partial}\theta \mathring{\partial}\theta'}\right] & E\left[\frac{\mathring{\partial} m_{j,n,t}^*}{\mathring{\partial}\theta} m_{n,t}^{*\prime}\right] + E\left[m_{j,n,t}^* \frac{\mathring{\partial} m_{n,t}^*}{\mathring{\partial}\theta'}\right] \\ E\left[m_{n,t}^* \frac{\mathring{\partial} m_{j,n,t}^*}{\mathring{\partial}\theta}\right] + E\left[m_{j,n,t}^* \frac{\mathring{\partial} m_{n,t}^*}{\mathring{\partial}\theta}\right] & -\rho_3 \times E\left[m_{j,n,t}^* m_{n,t}^* m_{n,t}^{*\prime}\right] \end{bmatrix}$$

[1]Notice asymptotically $\tilde{\mathcal{J}}_n = \mathcal{J}_n(1 + o(1))$ hence $\tilde{\mathcal{H}}_n = \mathcal{H}_n(1 + o(1))$ and $\tilde{\mathcal{P}}_n = \mathcal{P}_n(1 + o(1))$, but this requires replacing $E[\epsilon_{n,t}^{*2}]$ with 1, and clearly $E[\epsilon_{n,t}^{*2}] < 1$ places a role in the higher order bias.

By the martingale difference property it follows:

$$
\begin{aligned}
E\left[Q_1\left(\tilde{\psi}_n^*\right)\right] &= -\mathfrak{G}_n^{*-1}\left\{E\left[\mathfrak{A}_{n,t}^*\psi_{n,t}^*\right]+\frac{1}{2}\sum_{i=1}^{q+3}\mathfrak{G}_{i,n}^*E\left[\psi_{n,t}^*\psi_{n,t}^{*\prime}\right]e_i\right\} \\
&= -\mathfrak{G}_n^{*-1}\left\{E\left[\frac{\mathring{\partial}}{\mathring{\partial}\beta}\mathfrak{M}_{n,t}^*\mathfrak{G}_n^{*-1}\mathfrak{M}_{n,t}^*\right]+\frac{1}{2}\sum_{i=1}^{q+3}\mathfrak{G}_{i,n}^*\mathfrak{G}_n^{*-1}E\left[\mathfrak{M}_{n,t}^*\mathfrak{M}_{n,t}^{*\prime}\right]\mathfrak{G}_n^{*-1}e_i\right\}
\end{aligned}
$$

Then since $\mathfrak{M}_{n,t}^* = -[0, m_{n,t}^{*\prime}]' = -[0, (\epsilon_{n,t}^{*2} - E\left[\epsilon_{n,t}^{*2}\right])(x_t - E[x_t])']'$ it is easily verified that

$$
\begin{aligned}
E\left[\frac{\mathring{\partial}}{\mathring{\partial}\beta}\mathfrak{M}_{n,t}^*\times\mathfrak{G}_n^{*-1}\times\mathfrak{M}_{n,t}^*\right] &= \begin{bmatrix} E\left[\frac{\mathring{\partial}}{\mathring{\partial}\theta'}m_{n,t}^*\tilde{\mathcal{P}}_n m_{n,t}^*\right] \\ E\left[\frac{\mathring{\partial}}{\mathring{\partial}\theta}m_{n,t}^*\tilde{\mathcal{H}}_n m_{n,t}^* + m_{n,t}^*m_{n,t}^{*\prime}\tilde{\mathcal{P}}_n m_{n,t}^*\right] \end{bmatrix} \\
&= \begin{bmatrix} \mathcal{E}_n^{(2)}\times E\left[X_t'X_t\tilde{\mathcal{P}}_nX_t\right] \\ \mathcal{E}_n^{(2)}\times E\left[S_tX_t'\tilde{\mathcal{H}}_nX_t\right]+\mathcal{E}_n^{(3)}\times E\left[X_t'X_t\tilde{\mathcal{P}}_nX_t\right] \end{bmatrix} \\
&= \begin{bmatrix} E\left[X_t'X_t\mathcal{P}X_t\right] \\ \frac{\mathcal{E}_n^{(2)}}{\mathcal{E}_n^{(1)}}\times E\left[S_tX_t'\mathcal{H}X_t\right]+\frac{\mathcal{E}_n^{(3)}}{\mathcal{E}_n^{(2)}}\times E\left[X_t'X_t\mathcal{P}X_t\right] \end{bmatrix}.
\end{aligned}
$$

Further:

$$
E\left[\mathfrak{M}_{n,t}^*\mathfrak{M}_{n,t}^{*\prime}\right] = \begin{bmatrix} 0 & 0 \\ 0 & \Sigma_n \end{bmatrix}
$$

hence

$$
\begin{aligned}
&\frac{1}{2}\sum_{i=1}^{q+3}\mathfrak{G}_{i,n}^*\mathfrak{G}_n^{*-1}E\left[\mathfrak{M}_{n,t}^*\mathfrak{M}_{n,t}^{*\prime}\right]\mathfrak{G}_n^{*-1}e_i \\
&\qquad = \frac{1}{2}\sum_{i=1}^{3}\mathfrak{G}_{i,n}^*\begin{bmatrix} \Sigma_n & 0 \\ 0 & \tilde{\mathcal{P}}_n \end{bmatrix}e_i + \frac{1}{2}\sum_{i=4}^{q+3}\mathfrak{G}_{i,n}^*\begin{bmatrix} \Sigma_n & 0 \\ 0 & \tilde{\mathcal{P}}_n \end{bmatrix}e_i \\
&\qquad = -\sum_{i=1}^{3}\begin{bmatrix} 0 \\ \frac{1}{2}E\left[\frac{\mathring{\partial}}{\mathring{\partial}\theta_j}\frac{\mathring{\partial}}{\mathring{\partial}\theta}m_{n,t}^*\right]\Sigma_n e_i \end{bmatrix}
\end{aligned}
$$

$$
-\sum_{i=4}^{q+3}\left[\begin{array}{c}\frac{1}{2}\left(E\left[\frac{\mathring{\partial}}{\mathring{\partial}\theta}m^*_{i,n,t}m^{*\prime}_{n,t}\right]+E\left[m^*_{i,n,t}\frac{\mathring{\partial}}{\mathring{\partial}\theta'}m^*_{n,t}\right]\right)e_i\tilde{\mathcal{P}}_n \\ -\frac{1}{2}\rho_3 E\left[m^*_{j,n,t}m^*_{n,t}m^{*\prime}_{n,t}\right]\tilde{\mathcal{P}}_n e_i\end{array}\right]
$$

$$
=\left[\begin{array}{c}-E\left[\frac{\mathring{\partial}}{\mathring{\partial}\theta}m^*_{n,t}\tilde{\mathcal{P}}_n m^{*\prime}_{n,t}\right] \\ -a_n+\frac{\rho_3}{2}E\left[m^*_{n,t}m^{*\prime}_{n,t}\tilde{\mathcal{P}}_n m^*_{n,t}\right]\end{array}\right]
$$

$$
=\left[\begin{array}{c}-E\left[X_t'X_t\mathcal{P}X_t\right] \\ -\frac{\mathcal{E}_n^{(2)}}{\mathcal{E}_n^{(1)}}a+\frac{\rho_3}{2}\frac{\mathcal{E}_n^{(3)}}{\mathcal{E}_n^{(2)}}E\left[X_t'X_t\mathcal{P}X_t\right]\end{array}\right].
$$

Therefore:

$$
E\left[\hat{\beta}_n-\beta^0\right]
$$

$$
=-\frac{1}{n}\mathfrak{G}_n^{*-1}\left[\begin{array}{c}0 \\ -a_n+E\left[\frac{\mathring{\partial}}{\mathring{\partial}\theta}m^*_{n,t}\tilde{\mathcal{H}}_n m^*_{n,t}+m^*_{n,t}m^{*\prime}_{n,t}\tilde{\mathcal{P}}_n m^*_{n,t}\right]+\frac{\rho_3}{2}E\left[m^*_{n,t}m^{*\prime}_{n,t}\tilde{\mathcal{P}}_n m^*_{n,t}\right]\end{array}\right]
$$

$$
=-\frac{1}{n}\mathfrak{G}_n^{*-1}\left[\begin{array}{c}0 \\ -a_n+E\left[\frac{\mathring{\partial}}{\mathring{\partial}\theta}m^*_{n,t}\tilde{\mathcal{H}}_n m^*_{n,t}\right]+\left(1+\frac{\rho_3}{2}\right)E\left[m^*_{n,t}m^{*\prime}_{n,t}\tilde{\mathcal{P}}_n m^*_{n,t}\right]\end{array}\right]
$$

$$
=-\frac{1}{n}\mathfrak{G}_n^{*-1}\left[\begin{array}{c}0 \\ \frac{\mathcal{E}_n^{(2)}}{\mathcal{E}_n^{(1)}}\left(-a+E\left[S_tX_t'\mathcal{H}X_t\right]\right)+\frac{\mathcal{E}_n^{(3)}}{\mathcal{E}_n^{(2)}}\left(1+\frac{\rho_3}{2}\right)E\left[X_t'X_t\mathcal{P}X_t\right]\end{array}\right]
$$

Now use expressions for (A.17) to conclude

$$
E\left[\hat{\beta}_n-\beta^0\right]
$$

$$
=\frac{1}{n}\left[\begin{array}{cc}-\mathcal{E}_n^{(2)}\Sigma_x & \frac{1}{\mathcal{E}_n^{(1)}}\mathcal{H} \\ \frac{1}{\mathcal{E}_n^{(1)}}\mathcal{H}' & \frac{1}{\mathcal{E}_n^{(2)}}\mathcal{P}\end{array}\right]\times\begin{array}{c}0 \\ \frac{\mathcal{E}_n^{(2)}}{\mathcal{E}_n^{(1)}}\left(-a+E\left[S_tX_t'\mathcal{H}X_t\right]\right)+\frac{\mathcal{E}_n^{(3)}}{\mathcal{E}_n^{(2)}}\left(1+\frac{\rho_3}{2}\right)E\left[X_t'X_t\mathcal{P}X_t\right]\end{array}
$$

$$= \frac{1}{n}\left[\begin{array}{c} \frac{1}{\mathcal{E}_n^{(1)}}\mathcal{H}\left\{\frac{\mathcal{E}_n^{(2)}}{\mathcal{E}_n^{(1)}}\left(-a+E\left[S_tX_t'\mathcal{H}X_t\right]\right)+\frac{\mathcal{E}_n^{(3)}}{\mathcal{E}_n^{(2)}}\left(1+\frac{\rho_3}{2}\right)E\left[X_t'X_t\mathcal{P}X_t\right]\right\} \\ \frac{1}{\mathcal{E}_n^{(2)}}\mathcal{P}\left\{\frac{\mathcal{E}_n^{(2)}}{\mathcal{E}_n^{(1)}}\left(-a+E\left[S_tX_t'\mathcal{H}X_t\right]\right)+\frac{\mathcal{E}_n^{(3)}}{\mathcal{E}_n^{(2)}}\left(1+\frac{\rho_3}{2}\right)E\left[X_t'X_t\mathcal{P}X_t\right]\right\} \end{array}\right].$$

The proof is therefore complete. $\mathcal{QED}$.

A.4 Proofs of Supporting Lemmas

Lemma A.1 (threshold bound). $\sup_{\theta\in\Theta}\{c_n^4(\theta)/||\Sigma_n(\theta)||\} = o(n)$.

Proof. Use (A.6) and (A.9) to deduce the claim. $\mathcal{QED}$.

Lemma A.2 (covariance bound). $\sup_{\theta\in\Theta}||\Sigma_n(\theta)|| = o(n)$.

Proof. Let $g : \Theta \to (0,\infty)$ be a bounded function, $0 < \inf_{\theta\in\Theta} g(\theta) \leq \sup_{\theta\in\Theta} g(\theta) < \infty$, that may be different in different places. Similarly $o(1)$ does not depend on θ and may be different in different places. By (A.6) and (A.8) we can express $\mathfrak{M}_{i,n}(\theta)$ as $g(\theta)(1+o(1))$, $\ln(n)g(\theta)(1+o(1))$ and $(n/k_n)^{4/\kappa-1}g(\theta)(1+o(1))$ respectively if $\kappa > 4$, $\kappa = 4$ or $\kappa < 4$. The proof is complete since $\sup_{\theta\in\Theta}\{\mathfrak{M}_{i,n}(\theta)\} = o(n)$ in each case. $\mathcal{QED}$.

Lemma A.3 (uniform threshold law). $\sup_{\theta\in\Theta}|\epsilon_{(k_n)}^{(a)}(\theta)/c_n(\theta) - 1| = O_p(1/k_n^{1/2})$.

Proof. In view of the stationary geometric β-mixing property, the claim follows from Lemma B.2 in Hill (2015a). $\mathcal{QED}$.

Lemma A.4 (generic ULLN). *Let $\{z_t(\theta)\}$ be a strictly stationary geometrically β-mixing process, with Paretian tail*

$$P\left(|z_t(\theta)| > z\right) = d(\theta)z^{-\kappa(\theta)}(1+o(1)), \text{ where } (d(\theta),\kappa(\theta)) \in (0,\infty).$$

Define the tail trimmed version $z_{n,t}^*(\theta) \equiv z_t(\theta)I(|z_t(\theta)| \leq c_n(\theta))$, where $P(|z_t(\theta)| > c_n(\theta)) = k_n/n = o(1)$, and $k_n \to \infty$. Let $k_n/n^\iota \to \infty$ for some tiny $\iota > 0$. Then

$$\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{t=1}^n\left\{z_{n,t}^*(\theta) - E\left[z_{n,t}^*(\theta)\right]\right\}\times(1+o_p(1))\right| \xrightarrow{p} 0$$

where $o_p(1)$ may be a functions of θ.

Proof. We first prove a pointwise LLN $1/n\sum_{t=1}^{n} z_{n,t}^{*}(\theta)/E[z_{n,t}^{*}(\theta)] \stackrel{p}{\rightarrow} 1$ when $E[z_{n,t}^{*}(\theta)] \neq 0$. We then prove the required ULLN.

Step 1 (LLN): Let $E[z_{n,t}^{*}(\theta)] \neq 0$. If $\kappa(\theta) > 1$ then $z_{n,t}^{*}(\theta)/E[z_{n,t}^{*}(\theta)]$ is uniformly integrable, hence $1/n\sum_{t=1}^{n} z_{n,t}^{*}(\theta)/E[z_{n,t}^{*}(\theta)] \stackrel{p}{\rightarrow} 1$ by Theorem 2 in Andrews (1988). Now assume $\kappa(\theta) \in (0,1]$, write $w_{n,t}^{*} \equiv z_{n,t}^{*}(\theta)$ and $\kappa = \kappa(\theta)$, and define

$$\rho_n(h) \equiv \frac{E\left[\left\{w_{n,1}^{*} - E\left[w_{n,1}^{*}\right]\right\}\left\{w_{n,h+1}^{*} - E\left[w_{n,h+1}^{*}\right]\right\}\right]}{E\left[\left\{w_{n,1}^{*} - E\left[w_{n,1}^{*}\right]\right\}^{2}\right]}.$$

Note $E(1/n\sum_{t=1}^{n}\{w_{n,t}^{*}/E\left[w_{n,t}^{*}\right] - 1\})^{2}$ is bounded from above by

$$2\frac{1}{n}\frac{E\left[w_{n,t}^{*}\right]}{\left(E\left[w_{n,t}^{*}\right]\right)^{2}} + 2\frac{1}{n}\frac{E\left[\left\{w_{n,1} - E\left[w_{n,1}\right]\right\}^{2}\right]}{\left(E\left[w_{n,t}^{*}\right]\right)^{2}}\sum_{h=1}^{n}\left|\rho_n(h)\right|. \tag{A.18}$$

Recall by Karamata's Theorem $E|w_{n,t}^{*}|^{q} \sim Kc_n^{q}(k_n/n) = K(n/k_n)^{q/\kappa-1}$ for any $q > \kappa$. If $\kappa \in (0,1)$ then

$$\begin{aligned} E\left(\frac{1}{n}\sum_{t=1}^{n}\left\{\frac{w_{n,t}^{*}}{E\left[w_{n,t}^{*}\right]} - 1\right\}\right)^{2} &\leq K\frac{1}{n}\frac{(n/k_n)^{2/\kappa-1}}{(n/k_n)^{2/\kappa-2}} + K\frac{1}{n}\frac{(n/k_n)^{2/\kappa-1}}{(n/k_n)^{2/\kappa-2}}\sum_{h=1}^{n}\left|\rho_n(h)\right| \\ &= K\left(\frac{1}{k_n} + \frac{1}{k_n}\sum_{h=1}^{n}\left|\rho_n(h)\right|\right). \end{aligned}$$

Since $k_n \rightarrow \infty$ the proof follows by Chebyshev's inequality if $1/k_n\sum_{i=1}^{n}|\rho_n(i)| \rightarrow 0$. The latter holds by noting under geometric β-mixing, for some $\rho \in (0,1)$ and tiny $\delta > 0$ (Ibragimov, 1962)

$$\sum_{h=1}^{n}\left|\rho_n(h)\right| \leq K\left|\frac{\left\|w_{n,1}^{*} - E\left[w_{n,1}^{*}\right]\right\|_{2+\delta}}{\left\|w_{n,1}^{*} - E\left[w_{n,1}^{*}\right]\right\|_{2}}\right|\sum_{h=1}^{\infty}\rho^{h} = K\left|\frac{\left\|w_{n,1}^{*} - E\left[w_{n,1}^{*}\right]\right\|_{2+\delta}}{\left\|w_{n,1}^{*} - E\left[w_{n,1}^{*}\right]\right\|_{2}}\right|, \tag{A.19}$$

hence by Karamata's Theorem

$$\sum_{h=1}^{n}\left|\rho_n(h)\right| \leq K\frac{\left(n/k_n\right)^{1/\kappa-1/(2+\delta)}}{\left(n/k_n\right)^{1/\kappa-1/2}} = \left(n/k_n\right)^{1/2-1/(2+\delta)}.$$

Therefore $1/k_n\sum_{i=1}^{n}|\rho_n(i)| \rightarrow 0$ for any sequence $\{k_n\}$ that satisfies $(n/k_n)^{1/2-1/(2+\delta)}/k_n \rightarrow 0$. It is easy to check that $k_n/n^{\iota} \rightarrow \infty$ for some infinitessimal $\iota > 0$ by supposition ensures $(n/k_n)^{1/2-1/(2+\delta)}/k_n \rightarrow 0$ since we can take $\delta > 0$ to be arbitrarily small.

If $\kappa = 1$ then by Karamata's Theorem for some slowly varying $L(n) \to \infty$

$$E\left(\frac{1}{n}\sum_{t=1}^{n}\left\{\frac{w_{n,t}^{*}}{E\left[w_{n,t}^{*}\right]}-1\right\}\right)^{2} \leq \frac{1}{n}\frac{(n/k_n)^{2/\kappa-1}}{L(n)}\left(1+\sum_{h=1}^{n}|\rho_n(h)|\right) = \frac{1}{k_n}\frac{1}{L(n)}\left(1+\sum_{h=1}^{n}|\rho_n(h)|\right).$$

Therefore $\sum_{h=1}^{n}|\rho_n(h)| \leq (n/k_n)^{1/2-1/(2+\delta)} = o(k_n)$ exactly as above.

Step 2 (ULLN): We proceed by proving two preliminary ULLN's. First, define $\mu_{n,t}^{*}(\theta) \equiv |z_{n,t}^{*}(\theta)|/\sup_{\theta\in\Theta}\{E|z_{n,t}^{*}(\theta)|\}$. Since by construction $\mu_{n,t}^{*}(\theta)$ is uniformly L_1-bounded on compact $\Theta \times \Gamma$, it belongs to a separable Banach space. Therefore the L_1-bracketing numbers satisfy $N_{[\,]}(\varepsilon, \Theta \times \Gamma, ||\cdot||_1) < \infty$ (e.g. Dudley, 1999, Proposition 7.1.7). In view of the Step 1 pointwise law $1/n\sum_{t=1}^{n}(\mu_{n,t}^{*}(\theta) - E\left[\mu_{n,t}^{*}(\theta)\right]) = o_p(1)$, we have by Theorem 7.1.5 of Dudley (1999):

$$\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{t=1}^{n}\left\{\mu_{n,t}^{*}(\theta) - E\left[\mu_{n,t}^{*}(\theta)\right]\right\}\right| = o_p(1). \tag{A.20}$$

Second, replace $z_{n,t}^{*}(\theta)$ with $g_{n,t}^{*}(\theta) \equiv |z_{n,t}^{*}(\theta)|/E|z_{n,t}^{*}(\theta)|$ and invoke (A.20) to obtain:

$$\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{t=1}^{n}\left\{g_{n,t}^{*}(\theta) - E\left[g_{n,t}^{*}(\theta)\right]\right\}\right| = o_p\left(\sup_{\theta\in\Theta}\left|E\left[g_{n,t}^{*}(\theta)\right]\right|\right) = o_p(1). \tag{A.21}$$

Finally, for any $\delta > 0$ define

$$r_n(\theta,\delta) \equiv \frac{1}{n}\sum_{t=1}^{n}\left(\left\{\frac{z_{n,t}^{*}(\theta) - E\left[z_{n,t}^{*}(\theta)\right]}{\left|E\left[z_{n,t}^{*}(\theta)\right]\right| + \delta}\right\} - \frac{1}{\left|E\left[z_{n,t}^{*}(\theta)\right]\right| + \delta}\left\{\frac{z_{n,t}^{*}(\theta) - E\left[z_{n,t}^{*}(\theta)\right]}{\left|E\left[z_{n,t}^{*}(\theta)\right]\right| + \delta}\right\}\right).$$

Note that $\sup_{\theta\in\Theta}|r_n(\theta,\delta)| = o_p(1)$ by a generalization of the second ULLN. In particular

$$\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{t=1}^{n}\left\{z_{n,t}^{*}(\theta) - E\left[z_{n,t}^{*}(\theta)\right] - r_n(\theta,\delta)\times\left(\left|E\left[z_{n,t}^{*}(\theta)\right]\right| + \delta\right)\right\}\right|$$

$$= \sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{t=1}^{n}\left\{\frac{z_{n,t}^{*}(\theta) - E\left[z_{n,t}^{*}(\theta)\right]}{\left|E\left[z_{n,t}^{*}(\theta)\right]\right| + \delta}\right\}\right|$$

is $o_p(1)$ by (A.21). Now use $\sup_{\theta\in\Theta}|r_n(\theta,\delta)| = o_p(1)$ to conclude $\sup_{\theta\in\Theta}|1/n\sum_{t=1}^{n}\{z_{n,t}^{*}(\theta) - E[z_{n,t}^{*}(\theta)] \times (1 - o_p(1))\}| = o_p(1)$ as claimed. $\mathcal{QED}$.

Lemma A.5 (approximation)**.** $\sup_{\theta\in\Theta}||n^{-1/2}\Sigma_n^{-1/2}(\theta)\sum_{t=1}^{n}\{\hat{m}_{n,t}^{*}(\theta) - m_{n,t}^{*}(\theta)\}|| = o_p(1)$.

Proof. Define $\mathcal{E}_{n,t}(\theta) \equiv |\epsilon_t(\theta)| - c_n(\theta)$ and $\widehat{\mathcal{E}}_{n,t}(\theta) \equiv |\epsilon_t(\theta)| - \epsilon_{(k_n)}^{(a)}(\theta)$. We exploit arguments developed in Hill (2012, 2015a, 2013) to prove

$$\frac{1}{\|\Sigma_n(\theta)\|^{1/2} n^{1/2}} \sum_{t=1}^{n} \left\{ \hat{m}^*_{i,n,t}(\theta) - m^*_{i,n,t}(\theta) \right\} = o_p(1).$$

Throughout $c^*_n(\theta)$ satisfies $|c^*_n(\theta) - c_n(\theta)| < |\epsilon^{(a)}_{(k_n)}(\theta) - c_n(\theta)|$ $a.s.$ and may be different in different places.

The indicator function $I(u) \equiv I(u \leq 0)$ can be approximated by a smooth regular sequence $\{\mathfrak{I}_n(u)\}_{n\geq 1}$. Define

$$\mathfrak{I}_n(u) \equiv \int_{-\infty}^{\infty} I(\varpi) \mathrm{S}(\mathcal{N}_n(\varpi - u)) \mathcal{N}_n e^{-\varpi^2/\mathcal{N}_n^2} d\varpi$$

where

$$\mathrm{S}(\xi) = \begin{cases} e^{-1/(1-\xi^2)} / \int_{-1}^{1} e^{-1/(1-w^2)} dw & \text{if } |\xi| < 1 \\ 0 & \text{if } |\xi| \geq 1 \end{cases}$$

and let $\{\mathcal{N}_n\}$ be a sequence of finite positive numbers that satisfies $\mathcal{N}_n \to \infty$. Observe $\mathfrak{I}_n(u)$ is uniformly bounded in u, and continuous and differentiable. Also, $I(u)$ is differentiable except at 0, with derivative $\delta(u) = (\partial/\partial u)I(u)$ the Dirac delta function. Therefore $\delta(u)$ has a regular sequence $\mathfrak{D}_n(u) \equiv (\mathcal{N}_n/\pi)^{1/2} \exp\{-\mathcal{N}_n u^2\}$. Lighthill (1958, p. 22).

Note that $\mathcal{N}_n \to \infty$ be made as fast as we choose. Hence, for some $o_p(1)$ that is not a function of θ,

$$\begin{aligned} &\frac{1}{\|\Sigma_n(\theta)\|^{1/2} n^{1/2}} \sum_{t=1}^{n} \left\{ \hat{m}^*_{i,n,t}(\theta) - m^*_{i,n,t}(\theta) \right\} \qquad \text{(A.22)} \\ &= \frac{1}{\|\Sigma_n(\theta)\|^{1/2} n^{1/2}} \sum_{t=1}^{n} \epsilon_t^2(\theta) \left(\mathfrak{I}_n(\widehat{\mathcal{E}}_{n,t}(\theta)) - \mathfrak{I}_n\left(\mathcal{E}_{n,t}(\theta)\right) \right) \times \left(s_{i,t}(\theta) - E\left[s_{i,t}(\theta)\right]\right) \\ &\quad - \frac{E\left[\epsilon_t^2(\theta)\mathfrak{I}_n\left(\mathcal{E}_{n,t}(\theta)\right)\right]}{\|\Sigma_n(\theta)\|^{1/2}} \left(\frac{1/n \sum_{t=1}^{n} \epsilon_t^2(\theta) \mathfrak{I}_n(\widehat{\mathcal{E}}_{n,t}(\theta))}{E\left[\epsilon_t^2(\theta)\mathfrak{I}_n\left(\mathcal{E}_{n,t}(\theta)\right)\right]} - 1 \right) \\ &\quad \times \frac{1}{n^{1/2}} \sum_{t=1}^{n} \left(s_{i,t}(\theta) - E\left[s_{i,t}(\theta)\right]\right) + o_p(1), \end{aligned}$$

Consider $1/n \sum_{t=1}^{n} \epsilon_t^2(\theta)\mathfrak{I}_n(\widehat{\mathcal{E}}_{n,t}(\theta))$. By differentiability of $\mathfrak{I}_n(\cdot)$ we have by the mean-value-theorem

$$\frac{1}{n} \sum_{t=1}^{n} \epsilon_t^2(\theta)\mathfrak{I}_n(\widehat{\mathcal{E}}_{n,t}(\theta)) = \frac{1}{n} \sum_{t=1}^{n} \epsilon_t^2(\theta)\mathfrak{I}_n\left(\mathcal{E}_{n,t}(\theta)\right)$$

$$+\frac{1}{n}\sum_{t=1}^{n}\epsilon_t^2(\theta)\times\frac{\mathcal{N}_n^{1/2}/\pi^{1/2}}{\exp\left\{\mathcal{N}_n\left(|\epsilon_t|-c_n^*(\theta)\right)^2u^2\right\}}c_n(\theta)\times\left(\frac{\epsilon_{(k_n)}^{(a)}(\theta)}{c_n(\theta)}-1\right).$$

By distribution continuity $\inf_{\theta\in\Theta}||\epsilon_t(\theta)|-c_n^*(\theta)|>0$ $a.s.$; by Lemma A.3 $\sup_{\theta\in\Theta}|\epsilon_{(k_n)}^{(a)}(\theta)/c_n(\theta)$ $-1|=O_p(1/k_n^{1/2})$; and use Lemmas A.1 and A.2 to deduce $\sup_{\theta\in\Theta}\{c_n(\theta)\}=o(n^{1/2})$. Therefore

$$\sup_{\theta\in\Theta}\left|c_n(\theta)\times\left(\frac{\epsilon_{(k_n)}^{(a)}(\theta)}{c_n(\theta)}-1\right)\right|=o_p\left(n/k_n^{1/2}\right).$$

Since $\mathcal{N}_n\rightarrow\infty$ is arbitrary, and $\epsilon_t(\theta)$ is L_2-bounded, $\mathcal{N}_n$ can therefore be set to satisfy

$$\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{t=1}^{n}\epsilon_t^2(\theta)\left\{\mathfrak{I}_n(\widehat{\mathcal{E}}_{n,t}(\theta))-\mathfrak{I}_n\left(\mathcal{E}_{n,t}(\theta)\right)\right\}\right|=o_p\left(1\right). \tag{A.23}$$

By the same argument:

$$\sup_{\theta\in\Theta}\left\{\frac{1}{\left\|\Sigma_n(\theta)\right\|^{1/2}n^{1/2}}\left|\sum_{t=1}^{n}\epsilon_t^2(\theta)\left(\mathfrak{I}_n(\widehat{\mathcal{E}}_{n,t}(\theta))-\mathfrak{I}_n\left(\mathcal{E}_{n,t}(\theta)\right)\right)\times\left(s_{i,t}(\theta)-E\left[s_{i,t}(\theta)\right]\right)\right|\right\} \tag{A.24}$$

$$=\sup_{\theta\in\Theta}\left\{\frac{1}{\left\|\Sigma_n(\theta)\right\|^{1/2}n^{1/2}}\left|\sum_{t=1}^{n}\epsilon_t^2(\theta)\mathfrak{D}_{n,t}^{(\epsilon)}\left(c_n^*(\theta)\right)\times\left(s_{i,t}(\theta)-E\left[s_{i,t}(\theta)\right]\right)\right|\right\}=o_p(1).$$

Moreover, $s_t(\theta)$ is stationary, continuous, with a $L_{2+\iota}$-bounded enveloped, and an $L_{2+\iota}$-bounded gradient enveloped (Francq and Zakoïan, 2004). Hence by Theorem 1 in Doukhan, Massart, and Rio (1995)

$$\sup_{\theta\in\Theta}\left\{\frac{1}{n^{1/2}}\left|\sum_{t=1}^{n}\left(s_{i,t}(\theta)-E\left[s_{i,t}(\theta)\right]\right)\right|\right\}=O_p(1). \tag{A.25}$$

Combine (A.22)-(A.25) to conclude

$$\sup_{\theta\in\Theta}\left\{\frac{1}{\left\|\Sigma_n(\theta)\right\|^{1/2}n^{1/2}}\sum_{t=1}^{n}\left\{\hat{m}_{i,n,t}^*(\theta)-m_{i,n,t}^*(\theta)\right\}\right\}=o_p\left(\sup_{\theta\in\Theta}\left\{\frac{E\left[\epsilon_t^2(\theta)\mathfrak{I}_n\left(\mathcal{E}_{n,t}(\theta)\right)\right]}{\left\|\Sigma_n(\theta)\right\|^{1/2}}\right\}\right).$$

Finally, by construction $\limsup_{n\rightarrow\infty}\sup_{1\leq t\leq n}\mathfrak{I}_n(\mathcal{E}_{n,t}(\theta))\leq K$, by tail property (A.5) $\sup_{\theta\in\Theta}E[\epsilon_t^2(\theta)]$ $<\infty$, and by non-degeneracy $\liminf_{n\rightarrow\infty}||\Sigma_n(\theta)||>0$. Hence

$$\sup_{\theta\in\Theta}\left\{\frac{E\left[\epsilon_t^2(\theta)\mathfrak{I}_n\left(\mathcal{E}_{n,t}(\theta)\right)\right]}{\left\|\Sigma_n(\theta)\right\|^{1/2}}\right\}=O(1).$$

This completes the proof. $\mathcal{QED}$.

Lemma A.6 (covariance consistency)**.** *Recall* $\tilde{\Sigma}_n$ *and* $\hat{\Sigma}_n$ *in (A.3), and assume* $\tilde{\theta}_n \xrightarrow{p} \theta^0$*. a.* $\tilde{\Sigma}_n(\tilde{\theta}_n) = \Sigma_n(1 + o_p(1))$*; and b.* $\hat{\Sigma}_n(\tilde{\theta}_n) = \Sigma_n(1 + o_p(1))$.

Proof.

Claim (a). By an application of ULLN Lemma A.4: $\sup_{\theta\in\Theta} ||\Sigma_n^{-1}(\theta)n^{-1}\sum_{t=1}^n m_{n,t}^*(\theta)m_{n,t}^*(\theta)' - I_q|| \xrightarrow{p} 0$. Furthermore, by continuity and the definition of a derivative, any $\{i,j\}$ element satisfies $\Sigma_{n,i,j}(\tilde{\theta}_n) - \Sigma_{n,i,j} = (\partial/\partial\theta)\Sigma_{n,i,j}(\theta)|_{\theta^0} \times (\tilde{\theta}_n - \theta^0) \times (1 + o_p(1))$. We will show below that

$$\left\| \frac{\partial}{\partial\theta}\Sigma_{n,i,j}(\theta)|_{\theta^0} \right\| = O\left(\|\Sigma_n\|\right). \tag{A.26}$$

Since $\tilde{\theta}_n \xrightarrow{p} \theta^0$ hence $\Sigma_n^{-1}\Sigma_n(\tilde{\theta}_n) \xrightarrow{p} I_q$, thus as claimed:

$$\Sigma_n^{-1}\frac{1}{n}\sum_{t=1}^n m_{n,t}^*(\tilde{\theta}_n)m_{n,t}^*(\tilde{\theta}_n)' \xrightarrow{p} I_q.$$

Now consider (A.26). Since trimming is negligible we can use Lemma A.6.c in Hill (2015a) to deduce:

$$\frac{\partial}{\partial\theta}E[\epsilon_{n,t}^{*2}(\theta)]|_{\theta^0} = -E\left[\epsilon_t^2 s_t I_{n,t}^{(\epsilon)}\right] \times (1 + o(1)) = -E\left[\epsilon_{n,t}^{*2}\right] \times E\left[s_t\right] \times (1 + o(1)) \sim -E\left[s_t\right].$$

Similarly, by independence

$$\begin{aligned} \frac{\partial}{\partial\theta}\Sigma_{n,i,j}(\theta)|_{\theta^0} &= \frac{\partial}{\partial\theta}\left\{E\left[\left(\epsilon_{n,t}^{*2}(\theta) - E\left[\epsilon_{n,t}^{*2}(\theta)\right]\right)^2\right] \times E\left[s_{i,t}(\theta)s_{j,t}(\theta)\right]\right\}|_{\theta^0} \\ &= \left\{E\left[\epsilon_{n,t}^{*4}\right] - 1\right\} \times E\left[s_t\right] \times E\left[s_{i,t}s_{j,t}\right] \times (1 + o(1)) \\ &\quad + \left(E\left[\epsilon_{n,t}^{*4}\right] - 1\right) \times \frac{\partial}{\partial\theta}E\left[s_{i,t}(\theta)s_{j,t}(\theta)\right]|_{\theta^0} \times (1 + o(1)). \end{aligned}$$

Observe $||(\partial/\partial\theta)E[s_t(\theta)s_t(\theta)']|_{\theta^0}|| < \infty$, $||E[s_t]|| < \infty$ and $||E[s_t s_t]|| < \infty$. By dominated convergence and independence $\Sigma_n = \{E[\epsilon_{n,t}^{*4}] - 1\}E[s_t s_t'] \times (1 + o(1))$, hence (A.26) follows.

Claim (b). Considering (a), it suffices to prove

$$\left\|\frac{1}{n}\sum_{t=1}^n\left\{m_{n,t}^*(\tilde{\theta}_n)m_{n,t}^*(\tilde{\theta}_n)' - \hat{m}_{n,t}^*(\tilde{\theta}_n)\hat{m}_{n,t}^*(\tilde{\theta}_n)'\right\}\right\| = o_p\left(1\right).$$

The property can be shown by imitating the proof of Lemma A.5. $\mathcal{QED}$.

Lemma A.7 (Jacobian consistency). $1/n\sum_{t=1}^{n}\widehat{\mathcal{J}}_{n,t}(\tilde{\theta}_n) = \mathcal{J}_n \times (1 + o_p(1))$ *for any* $\tilde{\theta}_n \stackrel{p}{\to} \theta^0$.

Proof. Define

$$\mathcal{J}^*_{n,t}(\theta) \equiv \left(\frac{\partial}{\partial\theta}\epsilon^2_t(\theta) \times I^{(\epsilon)}_{n,t}(\theta) - \frac{1}{n}\sum_{t=1}^{n}\frac{\partial}{\partial\theta}\epsilon^2_t(\theta) \times I^{(\epsilon)}_{n,t}(\theta)\right) x_t(\theta) + \left(\epsilon^2_t(\theta)I^{(\epsilon)}_{n,t}(\theta) - \frac{1}{n}\sum_{t=1}^{n}\epsilon^2_t(\theta)I^{(\epsilon)}_{n,t}(\theta)\right)\frac{\partial}{\partial\theta}x_t(\theta)$$

By the same argument used to prove Lemma A.5

$$\sup_{\theta\in\Theta}\left\|\sum_{t=1}^{n}\{\mathcal{J}^*_{n,t}(\theta) - \widehat{\mathcal{J}}_{n,t}(\theta)\}\right\| = o_p(1).$$

Further:

$$\mathcal{J}^*_{n,t} \equiv -x_t\left(\epsilon^2_t I^{(\epsilon)}_{n,t}s_t - \frac{1}{n}\sum_{t=1}^{n}\epsilon^2_t I^{(\epsilon)}_{n,t}s_t\right)' + \left(\epsilon^2_t I^{(\epsilon)}_{n,t} - \frac{1}{n}\sum_{t=1}^{n}\epsilon^2_t I^{(\epsilon)}_{n,t}\right)\frac{\partial}{\partial\theta}x_t.$$

By stationary geometric mixing $1/n\sum_{t=1}^{n}s_t \stackrel{p}{\to} E[s_t]$ (Andrews, 1988), hence by independence, stationarity and $E[\epsilon^2_t I^{(\epsilon)}_{n,t}] \to 1$:

$$E\left[\mathcal{J}^*_{n,t}\right] = -E\left[x_t\left(s_t - E\left[s_t\right]\right)'\right] \times (1 + o(1)) = \mathcal{J}_n \times (1 + o(1)).$$

The remaining proof that $1/n\sum_{t=1}^{n}\mathcal{J}^*_{n,t}(\tilde{\theta}_n) = \mathcal{J}_n \times (1 + o_p(1))$ is essentially identical to the proof of Lemma A.6. See also the proof of Lemma A.5 in Hill (2015a). $\mathcal{QED}$.

Lemma A.8 (CLT). $n^{-1/2}\Sigma_n^{-1/2}\sum_{t=1}^{n}m^*_{n,t} \stackrel{d}{\to} N(0, I_q)$.

Proof. Write $z^*_{n,t}(\xi) \equiv n^{-1/2}\xi'\Sigma_n^{-1/2}m^*_{n,t}$ for any $\xi \in \mathbb{R}^q$, $\xi'\xi = 1$. Write $\mathfrak{S}_t = [\mathfrak{S}_{i,t}]_{i=1}^3 \equiv s_t - E[s_t]$, $\mathcal{E}^*_{n,t} \equiv \epsilon^{*2}_{n,t} - E[\epsilon^{*2}_{n,t}]$ and $\mathfrak{S}_{\xi,t} \equiv \xi'\left(E[\mathfrak{S}_t\mathfrak{S}_t']\right)^{-1/2}\mathfrak{S}_t$. Hence $E[\mathfrak{S}^2_{\xi,t}] = 1$; under Assumption A $\inf_{t\geq 1}\inf_{\xi'\xi=1}|\mathfrak{S}_{\xi,t}| > \iota$ *almost surely* for some tiny $\iota > 0$; Francq and Zakoïan (2004, p. 619)'s arguments carry over to show $E|\mathfrak{S}_{\xi,t}|^{4+\delta} < \infty$ for tiny $\delta > 0$; and by independence

$$z^*_{n,t}(\xi) = \frac{1}{n^{1/2}}\frac{\mathcal{E}^*_{n,t}}{\left(E\left[\mathcal{E}^{*2}_{n,t}\right]\right)^{1/2}}\mathfrak{S}_{\xi,t} \text{ where } \liminf_{n\to\infty} E[\mathcal{E}^{*2}_{n,t}] > 0.$$

$\{z^*_{n,t}(\xi)\}$ is a martingale difference array with respect to $\Im_t \equiv \sigma(\{y_\tau\} : \tau \leq t)$. We will verify equations (1.1), (1.2) and (2.14) in McLeish (1974, Corollary 2.13) to prove $\sum_{t=1}^{n}z^*_{n,t}(\xi)$

$\stackrel{d}{\to} N(0,1)$, and invoke the Cramér-Wold Theorem. The equality

$$E\left[\left(\sum_{t=1}^{n} z_{n,t}^{*}(\xi)\right)^{2}\right] = \sum_{t=1}^{n} E\left[z_{n,t}^{*}(\xi)^{2}\right] = 1,$$

yields (1.1).

Lindeberg condition (1.2) holds:

$$\sum_{t=1}^{n} E\left[z_{n,t}^{*2}(\xi) I\left(\left|z_{n,t}^{*}(\xi)\right| > \varepsilon\right)\right] \to 0 \; \forall \varepsilon > 0.$$

This follows by first noting by stationarity

$$\begin{aligned}
&\sum_{t=1}^{n} E\left[z_{n,t}^{*2}(\xi) I\left(\left|z_{n,t}^{*}(\xi)\right| > \varepsilon\right)\right] \\
&\quad = \frac{1}{nE\left[\mathcal{E}_{n,t}^{*2}\right]} E\left[\mathcal{E}_{n,t}^{*2}\mathfrak{S}_{\xi,t}^{2} I\left(\mathcal{E}_{n,t}^{*2}\mathfrak{S}_{\xi,t}^{2} > nE\left[\mathcal{E}_{n,t}^{*2}\right]\varepsilon^{2}\right)\right] \\
&\quad = \frac{1}{E\left[\mathcal{E}_{n,t}^{*2}\right]} E\left(\mathfrak{S}_{\xi,t}^{2} E\left[\mathcal{E}_{n,t}^{*2} I\left(\mathcal{E}_{n,t}^{*2} > \frac{nE\left[\mathcal{E}_{n,t}^{*2}\right]\varepsilon^{2}}{\mathfrak{S}_{\xi,t}^{2}}\right) |\Im_{t-1}\right]\right) \\
&\quad = \frac{1}{E\left[\mathcal{E}_{n,t}^{*2}\right]} E\left(\mathfrak{S}_{\xi,t}^{2} \int_{nE[\mathcal{E}_{n,t}^{*2}]\mathfrak{S}_{\xi,t}^{-2}\varepsilon^{2}} P\left(\mathcal{E}_{n,t}^{*2} > u\right) du\right) \\
&\quad = \frac{1}{E\left[\mathcal{E}_{n,t}^{*2}\right]} E\left(\mathfrak{S}_{\xi,t}^{2} \int_{nE[\mathcal{E}_{n,t}^{*2}]\mathfrak{S}_{\xi,t}^{-2}\varepsilon^{2}}^{c_n^4} P\left(\left(\epsilon_{n,t}^{*2} - E\left[\epsilon_{n,t}^{*2}\right]\right)^{2} > u\right) du\right) &\text{(A.27)} \\
&\quad \leq \frac{1}{E\left[\mathcal{E}_{n,t}^{*2}\right]} E\left(\mathfrak{S}_{\xi,t}^{2} \int_{nE[\mathcal{E}_{n,t}^{*2}]\iota^{-1}\varepsilon^{2}}^{c_n^4} P\left(\left(\epsilon_{n,t}^{*2} - E\left[\epsilon_{n,t}^{*2}\right]\right)^{2} > u\right) du\right) &\text{(A.28)} \\
&\quad = \frac{1}{E\left[\mathcal{E}_{n,t}^{*2}\right]} \int_{nE[\mathcal{E}_{n,t}^{*2}]\iota^{-1}\varepsilon^{2}}^{c_n^4} P\left(\left(\epsilon_{n,t}^{*2} - E\left[\epsilon_{n,t}^{*2}\right]\right)^{2} > u\right) du. &\text{(A.29)}
\end{aligned}$$

Equality (A.27) exploits $\mathcal{E}_{n,t}^{*2} \leq c_n^4$ in view of trimming; (A.28) uses $\inf_{t\geq 1}\inf_{\xi'\xi=1}|\mathfrak{S}_{\xi,t}| > \iota$; and (A.29) follows from independence of ϵ_t and $E[\mathfrak{S}_{\xi,t}^2] = 1$.

By (A.4), ϵ_t^2 has a power law tail with index $\kappa/2 > 1$, and $E[\epsilon_{n,t}^{*2}] \to 1$. Hence on the interval $[nE[\mathcal{E}_{n,t}^{*2}]\iota^{-1}\varepsilon^2, c_n^4]$ we have for some positive sequence $\{d_n\}$:

$$P\left(\left(\epsilon_{n,t}^{*2} - E\left[\epsilon_{n,t}^{*2}\right]\right)^{2} > u\right) = P\left(\left(\epsilon_{t}^{2} - E\left[\epsilon_{n,t}^{*2}\right]\right)^{2} > u\right) = d_n u^{-\kappa/4}\left(1 + o(1)\right), \text{ as } u \to \infty.$$

It is easily verified that $d_n \to d$, the tail scale in (A.4). Therefore

$$E\left[z_{n,t}^{*2}(\xi) I\left(\left|z_{n,t}^{*}(\xi)\right| > \varepsilon\right)\right] \leq K \frac{1}{nE\left[\mathcal{E}_{n,t}^{*2}\right]} E\left(\int_{nE[\mathcal{E}_{n,t}^{*2}]\iota^{-1}\varepsilon^2}^{c_n^4} u^{-\kappa/4} du\right)(1 + o(1)). \tag{A.30}$$

If $\kappa > 4$ then $\int_a^\infty u^{-\kappa/4} du < \infty$ for any $a > 0$ hence $\int_{nE[\mathcal{E}_{n,t}^{*2}]\iota^{-1}\varepsilon^2}^{c_n^4} u^{-\kappa/4} du \to 0$ *a.s.*, thus by dominated convergence $E(\int_{nE[\mathcal{E}_{n,t}^{*2}]\iota^{-1}\varepsilon^2}^{c_n^4} u^{-\kappa/4} du) \to 0$. This implies the Cesàro mean converges:

$$\frac{1}{n}\sum_{t=1}^{n} E(\int_{nE[\mathcal{E}_{n,t}^{*2}]\iota^{-1}\varepsilon^2}^{c_n^4} u^{-\kappa/4} du) \to 0.$$

In view of (A.30) and $\liminf_{\to\infty} E[\mathcal{E}_{n,t}^{*2}] > 0$, the Lindeberg condition therefore follows.

If $\kappa = 4$ then use (A.6) and (A.8) to deduce $4\ln(c_n) - \ln(nE[\mathcal{E}_{n,t}^{*2}]) \sim \ln(n/k_n) - \ln(n) - \ln(n) < 0$ for all $n \geq N$ and finite $N \geq 1$. Therefore $\int_{nE[\mathcal{E}_{n,t}^{*2}]\iota^{-1}\varepsilon^2}^{c_n^4} u^{-\kappa/4} du = 0 \; \forall n \geq N$. Repeat the above argument to deduce the Lindeberg condition.

Finally, if $\kappa \in (2,4)$ then from (A.6) and (A.8) it follows $c_n^4/(nE[\mathcal{E}_{n,t}^{*2}]) \sim K/[n(k_n/n)] = K/k_n \to 0$. Hence $\int_{nE[\mathcal{E}_{n,t}^{*2}]\iota^{-1}\varepsilon^2}^{c_n^4} u^{-\kappa/4} du = 0 \; \forall n \geq N$ which again proves the Lindeberg condition.

Finally, for McLeish (1974)'s (2.14) we must show $\limsup_{n\to\infty} \sum_{s\neq t} E[z_{n,s}^{*2}(\xi) z_{n,t}^{*2}(\xi)] \leq 1$. By independence of ϵ_t, $\limsup_{n\to\infty} \sum_{s\neq t} E[z_{n,s}^{*2}(\xi) z_{n,t}^{*2}(\xi)]$ is exactly

$$2\limsup_{n\to\infty} \sum_{s<t} E\left[\frac{1}{n^2} \frac{\mathcal{E}_{n,s}^{*2}\mathcal{E}_{n,t}^{*2}}{\left(E\left[\mathcal{E}_{n,t}^{*2}\right]\right)^2} \mathfrak{S}_{\xi,s}^2 \mathfrak{S}_{\xi,t}^2\right] = 2\limsup_{n\to\infty} \frac{1}{n^2}\sum_{s<t} E\left[\mathfrak{S}_{\xi,s}^2 \mathfrak{S}_{\xi,t}^2\right].$$

Write $\boldsymbol{S}_n \equiv E(1/n^{1/2}\sum_{t=1}\{\mathfrak{S}_{\xi,t}^2 - E[\mathfrak{S}_{\xi,t}^2]\})^2$. Invoke the β-mixing property and $E|\mathfrak{S}_{\xi,t}|^{4+\delta} < \infty$ to deduce by Theorem 1.7 in Ibragimov (1962)

$$\begin{aligned} E\left(\frac{1}{n}\sum_{t=1} \mathfrak{S}_{\xi,t}^2\right)^2 &= \frac{1}{n} E\left(\frac{1}{n^{1/2}}\sum_{t=1}\left\{\mathfrak{S}_{\xi,t}^2 - E\left[\mathfrak{S}_{\xi,t}^2\right]\right\}\right)^2 + \frac{1}{n^2}\sum_{s,t=1}^{n} E\left[\mathfrak{S}_{\xi,s}^2\right] E\left[\mathfrak{S}_{\xi,t}^2\right] \\ &= \frac{1}{n}\boldsymbol{S}_n + 1 = O(1/n) = o(1). \end{aligned}$$

Therefore $\limsup_{n\to\infty} \sum_{s\neq t} E[z_{n,s}^{*2}(\xi) z_{n,t}^{*2}(\xi)] \leq 1$. This completes the proof. $\mathcal{QED}$.

Lemma A.9 (uniform GEL argument)**.** $\sup_{\theta\in\Theta,\lambda\in\Lambda_n}\{\max_{1\leq t\leq n} |\lambda' m_{n,t}^*(\theta)|\} \xrightarrow{p} 0$, $\sup_{\theta\in\Theta,\lambda\in\Lambda_n}$ $\{\max_{1\leq t\leq n} |\lambda' \hat{m}_{n,t}^*(\theta)|\} \xrightarrow{p} 0$ *and* $\Lambda_n \subseteq \hat{\Lambda}_n(\theta)$ *w.p.a.1.* $\forall\theta \in \Theta$. *In particular* $\sup_{\theta\in\Theta,\lambda\in\Lambda_n}$ $\{\max_{1\leq t\leq n} |\lambda'\{\hat{m}_{n,t}^*(\theta) - m_{n,t}^*(\theta)\}|\} \xrightarrow{p} 0$.

Proof. By threshold bound Lemma A.1

$$\sup_{\theta\in\Theta}\max_{1\leq t\leq n}\left\|\Sigma_n^{-1/2}(\theta)m_{n,t}^*(\theta)\right\| = o_p\left(n^{1/2}\right).$$

The first claim now follows by the construction of Λ_n:

$$\begin{aligned}\sup_{\theta\in\Theta,\lambda\in\Lambda_n}\max_{1\leq t\leq n}\left|\lambda' m_{n,t}^*(\theta)\right| &= n^{-1/2}\sup_{\theta\in\Theta,\lambda\in\Lambda_n}\max_{1\leq t\leq n}\left|\lambda'\Sigma_n^{1/2}(\theta)n^{1/2}\Sigma_n^{-1/2}(\theta)m_{n,t}^*(\theta)\right| \\ &\leq n^{-1/2}\sup_{\theta\in\Theta}\max_{1\leq t\leq n}\left\|\Sigma_n^{-1/2}(\theta)m_{n,t}^*(\theta)\right\| \xrightarrow{p} 0.\end{aligned}$$

Apply uniform threshold law Lemma A.3 to obtain

$$\sup_{\theta\in\Theta,\lambda\in\Lambda_n}\left\{\max_{1\leq t\leq n}\left|\lambda'\hat{m}_{n,t}^*(\theta)\right|\right\} \xrightarrow{p} 0.$$

Next, $\Lambda_n \subseteq \hat{\Lambda}_n(\theta)$ *w.p.a.1.* $\forall\theta\in\Theta$ follows from the second claim and $0\in\mathcal{D}$ since $\lambda'\hat{m}_{n,t}^*(\theta)\in\mathcal{D}$ *w.p.a.1* for all $\theta\in\Theta$ and any λ: $||\lambda'\Sigma_n^{1/2}(\theta)|| \leq n^{-1/2}$. The last claim follows from the first and second. $\mathcal{QED}$.

Lemma A.10 (constrained GEL)**.** *Consider any sequence* $\{\tilde{\theta}_n\}$, $\tilde{\theta}_n\in\Theta$, $\tilde{\theta}_n \xrightarrow{p} \theta^0$, *such that* $||m_n^*(\tilde{\theta}_n)|| = O_p(||\Sigma_n||^{1/2}/n^{1/2})$. *Then* $\bar{\lambda}_n \equiv \arg\max_{\lambda\in\hat{\Lambda}_n(\tilde{\theta}_n)}\{\hat{Q}_n(\tilde{\theta}_n,\lambda)\}$ *exists w.p.a.1,*

$$\bar{\lambda}_n = O_p(||\tilde{\Sigma}_n(\tilde{\theta}_n)||^{-1/2}n^{-1/2}) = o_p(1),$$

and

$$\sup_{\lambda\in\hat{\Lambda}_n(\tilde{\theta}_n)}\left\{\hat{Q}_n(\tilde{\theta}_n,\lambda)\right\} \leq \rho^{(0)} + O_p\left(\frac{1}{\left\|\tilde{\Sigma}_n(\tilde{\theta}_n)\right\| n}\right).$$

Proof. We prove the following below:

$$\sup_{\theta\in\Theta,\lambda\in\Lambda_n}\left|\hat{Q}_n(\theta,\lambda) - \tilde{Q}_n(\theta,\lambda)\right| = o_p(1). \tag{A.31}$$

Hence, it suffices to work with $\tilde{Q}_n(\theta,\lambda)$. We assume $\tilde{\theta}_n \xrightarrow{p} \theta^0$, and smoothness of ρ coupled with the uniform GEL argument Lemma A.9 ensure $\tilde{\lambda}_n = \arg\max_{\lambda\in\Lambda_n}\{\tilde{Q}_n(\tilde{\theta}_n,\lambda)\}$ exists *w.p.a.1*, where $\tilde{\lambda}_n\in\Lambda_n$ satisfies by construction $\tilde{\lambda}_n = O_p(||\tilde{\Sigma}_n(\tilde{\theta}_n)||^{-1/2}n^{-1/2})$. We may therefore apply Newey and Smith (2004, Lemma A.2, p. 239) argument to prove each claim.

Now consider (A.31). We need only show

$$\mathcal{P}_n \equiv \sup_{\theta\in\Theta,\lambda\in\Lambda_n} \max_{1\leq t\leq n} \left|\rho(\lambda'\hat{m}^*_{n,t}(\theta)) - \rho(\lambda' m^*_{n,t}(\theta))\right| \xrightarrow{p} 0.$$

By the definition of a derivative and the triangle inequality

$$\begin{aligned}\mathcal{P}_n \leq & \sup_{\theta\in\Theta,\lambda\in\Lambda_n} \max_{1\leq t\leq n} \left|\rho^{(1)}\left(\lambda' m^*_{n,t}(\theta)\right)\right| \\ & \times \sup_{\theta\in\Theta,\lambda\in\Lambda_n} \max_{1\leq t\leq n} \left|\lambda'\hat{m}^*_{n,t}(\theta) - \lambda' m^*_{n,t}(\theta)\right| \times (1 + o_p(1)).\end{aligned} \tag{A.32}$$

Let μ denote Lebesgue measure on $\mathbb{R}$. Apply Lemma A.9 to deduce there exists a sequence of neighborhoods $\{\tilde{\mathcal{D}}_n\}$ in $\mathcal{D}$, with $\lim_{n\to\infty}\mu(\tilde{\mathcal{D}}_n) = 0$, such that asymptotically *w.p.a.1*

$$\sup_{\theta\in\Theta,\lambda\in\Lambda_n} \max_{1\leq t\leq n} \left|\lambda' m^*_{n,t}(\theta)\right| \in \tilde{\mathcal{D}}_n \text{ and } \sup_{\theta\in\Theta,\lambda\in\Lambda_n} \max_{1\leq t\leq n} \left|\lambda'\hat{m}^*_{n,t}(\theta)\right| \in \tilde{\mathcal{D}}_n.$$

Further, $\rho(\cdot)$ is twice differentiable, and $\rho^{(1)}(0) = -1$. Hence, some sequence of positive numbers $\{\delta_n\}$, $\delta_n \to 0$,

$$\sup_{\theta\in\Theta,\lambda\in\Lambda_n} \max_{1\leq t\leq n} \left|\rho^{(1)}\left(\lambda' m^*_{n,t}(\theta)\right) + 1\right| \in [-\delta_n, \delta_n] \quad \textit{w.p.a.1.} \tag{A.33}$$

Therefore

$$\sup_{\theta\in\Theta,\lambda\in\Lambda_n} \max_{1\leq t\leq n} \left|\rho^{(1)}(\lambda' m^*_{n,t}(\theta)) + 1\right| \xrightarrow{p} 0.$$

Lemma A.9 with (A.32) and (A.33) prove $\mathcal{P}_n \xrightarrow{p} 0$. $\mathcal{QED}$.

Lemma A.11 (equation limit)**.** $m^*_n(\hat{\theta}_n) = O_p(||\Sigma_n||^{1/2}/n^{1/2}) = o_p(1)$.

Proof. Lemma A.10 trivially holds for $\tilde{\theta}_n = \theta^0$. Now combine that with ULLN Lemma A.4, CLT Lemma A.8, and uniform GEL argument Lemma A.9 to deduce $m^*_n(\hat{\theta}_n) = O_p(||\Sigma_n||^{1/2}/n^{1/2})$ by the same proof Newey and Smith (2004) use for their Lemma A3. Now invoke Lemma A.2 for $||\Sigma_n||/n \to 0$ hence $m^*_n(\hat{\theta}_n) = o_p(1)$. $\mathcal{QED}$.

Lemma A.12 (profile weight)**.** *Let*

$$\tilde{\pi}^*_{n,t}(\theta) \equiv \frac{\rho^{(1)}(\tilde{\lambda}_n'\hat{m}^*_{n,t}(\theta))}{\sum_{t=1}^n \rho^{(1)}(\tilde{\lambda}_n'\hat{m}^*_{n,t}(\theta))}.$$

If $\tilde{\lambda}_n = O_p(||\Sigma_n||^{-1/2} n^{-1/2})$ *where* $O_p(\cdot)$ *is not a function of* θ*, then*

$$\sup_{\theta \in \Theta} \max_{1 \le t \le n} \left| \hat{\pi}^*_{n,t}(\theta) - \frac{1}{n} \right| = O_p\left(\frac{1}{\|\Sigma_n\|^{1/2} n^{3/2}} \right).$$

Proof. Expand $\rho^{(1)}(\tilde{\lambda}_n' \hat{m}^*_{n,t}(\theta))$ around $\lambda = 0$: for some $||\lambda_{n,*}|| \le ||\tilde{\lambda}_n|| = O_p(||\Sigma_n||^{-1/2} n^{-1/2})$:

$$\begin{aligned} \rho^{(1)}\left(\tilde{\lambda}_n' \hat{m}^*_{n,t}(\theta)\right) &= -1 + \rho^{(2)}\left(\lambda_{n,*}' \hat{m}^*_{n,t}(\theta)\right) \times \tilde{\lambda}_n \\ &= -1 + \rho^{(2)}\left(\lambda_{n,*}' \hat{m}^*_{n,t}(\theta)\right) \times O_p\left(\|\Sigma_n\|^{-1/2} n^{-1/2}\right). \end{aligned}$$

Further, Lemma A.9, twice differentiability of ρ, and $\rho^{(2)}(0) = -1$ ensure $\sup_{\theta \in \Theta, \lambda \in \Lambda_n} \max_{1 \le t \le n} |\rho^{(2)}(\lambda' m^*_{n,t}(\theta))| \xrightarrow{p} 1$. Hence

$$\sup_{\theta \in \Theta} \max_{1 \le t \le n} \left| \rho^{(1)}\left(\tilde{\lambda}_n' \hat{m}^*_{n,t}(\theta)\right) + 1 + O_p\left(\|\Sigma_n\|^{-1/2} n^{-1/2}\right) \right| = 0,$$

which proves

$$\begin{aligned} \sup_{\theta \in \Theta} \max_{1 \le t \le n} \left| \hat{\pi}^*_{n,t}(\theta) - \frac{1}{n} \right| &= \sup_{\theta \in \Theta} \max_{1 \le t \le n} \left| \frac{\rho^{(1)}\left(\tilde{\lambda}_n' \hat{m}^*_{n,t}(\theta)\right)}{\sum_{t=1}^n \rho^{(1)}\left(\tilde{\lambda}_n' \hat{m}^*_{n,t}(\theta)\right)} + \frac{1}{n} \right| \\ &= \left| \frac{1 + O_p\left(\|\Sigma_n\|^{-1/2} n^{-1/2}\right)}{n\left(1 + O_p\left(\|\Sigma_n\|^{-1/2} n^{-1/2}\right)\right)} - \frac{1}{n} \right| \le O_p\left(\|\Sigma_n\|^{-1/2} n^{-3/2}\right). \end{aligned}$$

This completes the proof. $\mathcal{QED}$.

References

Andrews, D. W. K. (1988): "Laws of Large Numbers for Dependent Non-Identically Distributed Random Variables," *Econometric Theory*, 14, 458–467.

Breiman, L. (1965): "On Some Limit Theorems Similar to the Arc-Sin Law," *Theory of Probability and its Applications*, 10, 323–331.

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

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

Doukhan, P., P. Massart, and E. Rio (1995): "Invariance Principles for Absolutely Regular Empirical Processes," *Annales de l'Institut Henri Poincar*, 31, 393–427.

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

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

Hill, J. B. (2012): "Heavy-Tail and Plug-In Robust Consistent Conditional Moment Tests of Functional Form," in *Festschrift in Honor of Hal White*, ed. by X. Chen, and N. Swanson, pp. 241–274. Springer: New York.

——— (2013): "Least Tail-Trimmed Squares for Infinite Variance Autoregressions," *Journal of Time Series Analysis*, 34, 168–186.

——— (2015a): "Robust Estimation and Inference for Heavy Tailed GARCH," *Bernoulli*, 21, 1629–1669.

——— (2015b): "Robust Expected Shorfall Estimation for Infinite Variance Time Series," *Journal of Financial Econometrics*, 13, 1–44.

Ibragimov, I. (1962): "Some Limit Theorems for Stationary Processes," *Theory of Probability and its Applications*, 7, 349–382.

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

McLeish, D. L. (1974): "Dependent Central Limit Theorems and Invariance Principles," *Annals of Probability*, 2, 620–628.

Nelson, D. B. (1990): "Stationarity and Persistence in the GARCH(1,1) Model," *Econometric Theory*, 6, 318–334.

Newey, W. K., and R. J. Smith (2004): "Higher Order Properties of GMM and Generalized Empirical Likelihood estimators," *Econometrica*, 72, 219–255.

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