# Heavy Tail Robust Frequency Domain Estimation

Jonathan B. Hill* and Adam McCloskey†
Dept. of Economics, University of North Carolina
Dept. of Economics, Brown University

September 12, 2014

**Abstract**

We develop heavy tail robust frequency domain estimators for covariance stationary time series with a parametric spectrum, including ARMA, GARCH and stochastic volatility. We use robust techniques to reduce the moment requirement down to only a finite variance. In particular, we negligibly trim the data, permitting both identification of the parameter for the candidate model, and asymptotically normal frequency domain estimators, while leading to a classic limit theory when the data have a finite fourth moment. The transform itself can lead to asymptotic bias in the limit distribution of our estimators when the fourth moment does not exist, hence we correct the bias using extreme value theory that applies whether tails decay according to a power law or not. In the case of symmetrically distributed data, we compute the mean-squared-error of our biased estimator and characterize the mean-squared-error minimization number of sample extremes. A simulation experiment shows our QML estimator works well and in general has lower bias than the standard estimator, even when the process is Gaussian, suggesting robust methods have merit even for thin tailed processes.

*Key words and phrases*: Frequency domain QML, Whittle estimation, heavy tails, robust estimation.

*JEL classifications* : C13, C22, C49.

*AMS subject classifications* : 62M15, 62F35.

## 1 Introduction

Let $\{y_t\}_{t\in\mathbb{Z}}$ be a stationary ergodic process with $E[y_t^2] < \infty$. We consider those processes with a parametric spectrum $f(\lambda, \theta)$ at frequency $\lambda$, where $\theta$ is a $k \times 1$ vector of unknown parameters. We assume there exists a unique point $\theta_0$ in the interior of a compact subset $\Theta \subset \mathbb{R}^k$ such that $f(\lambda, \theta_0)$ is the spectrum of $y_t$. Specifically, $f(\lambda, \theta) > 0$ for all points $(\lambda, \theta) \in [-\pi, \pi] \times \Theta$, and

*First, and corresponding, author. Dept. of Economics, University of North Carolina, Chapel Hill; www.unc.edu/∼jbhill; jbhill@email.unc.edu.

†Dept. of Economics, Brown University, adam_mccloskey@brown.edu.

$f(\cdot, \theta)$ is twice continuously differentiable. This applies to ARMA processes, squares of GARCH processes, and log-squares of stochastic volatility processes, each stationary with a finite second moment. We present frequency domain [FD] estimators of $\theta_0$ that are robust to heavy tails: our estimators are asymptotically normal and unbiased under mild regularity conditions, as long as $E[y_t^2] < \infty$. Frequency domain methods are useful for estimation at a given set of time series periodicities, especially at economic business cycle frequencies (see Granger, 1966; Qu and Tkachenko, 2012), and recently for estimation robust to trends and level shifts (McCloskey, 2013; McCloskey and Perron, 2013; McCloskey and Hill, 2014).

The observed sample is $\{y_t\}_{t=1}^T$ with sample size $T \geq 1$. The discrete Fourier transform and periodogram of $\{y_t\}$ at frequency $\lambda \in [-\pi, \pi]$ are respectively defined as follows:

$$w_T(\lambda) \equiv \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} y_t e^{-i\lambda t} \text{ and } \mathcal{I}_T(\lambda) \equiv |w_T(\lambda)|^2.$$

Define Fourier frequencies $\lambda_j \equiv 2\pi j/T$ for $j = -[T/2] + 1, \ldots, [T/2] - 1, [T/2]$. If $E[y_t^4] = \infty$ then Whittle (1953)'s estimator, and the FD-QML estimator (e.g. Whittle, 1953)

$$\underset{\theta \in \Theta}{\arg\min} \sum_{j \in \mathbb{F}} \left( \ln f(\lambda_j, \theta) + \frac{\mathcal{I}_T(\lambda_j)}{f(\lambda_j, \theta)} \right) \quad \text{where } \mathbb{F} \equiv (-T/2, T/2] \cap \mathbb{Z} \setminus \{0\},$$

are not known to be asymptotically normal. See Hannan (1973a,b), Dunsmuir and Hannan (1976), Dunsmuir (1979), and Hosoya and Taniguchi (1982).

We exploit recent developments in the heavy tail robust estimation literature to create new robust FD-QML and Whittle estimators. Tail-trimming in the time domain has been used to develop, amongst others, robust estimators for autoregressions and GARCH models with possibly heavy tailed errors (Hill, 2012b, 2014a); robust moment condition tests and tests of volatility spillover (Hill, 2012a; Hill and Aguilar, 2013; Aguilar and Hill, 2014); and robust moment estimators (see, e.g., Khan and Tamer, 2010; Hill, 2013; Chaudhuri and Hill, 2014).

We use transformed data $y_t y_{t-h} I(|y_t y_{t-h} - \gamma| < c)$ for estimating $E[y_t y_{t-h}]$, where $\gamma$ is a value used for centering. Such transforms have a long history for robustness against so called outliers and sample extremes. See, for example, Andrews, Bickel, Hampel, Huber, Rogers, and Tukey (1972), Huber (1964), Hampel (1974), and Hampel, Ronchetti, Rousseeuw, and Stahel (1986). An intermediate order statistic of $|y_t y_{t-h} - \gamma|$ is used for $c$ such that $c \to \infty$ as $T \to \infty$, which ensures we identify $\theta_0$, and allows us to control the amount of trimming (see, e.g., Hill, 2012a,b, 2013, 2014a; Hill and Aguilar, 2013; Chaudhuri and Hill, 2014). Centering with $\gamma = E[y_t y_{t-h}]$ allows us to identify $E[y_t y_{t-h}]$ when $y_t y_{t-h}$ has a symmetric distribution, diminishing small sample bias of our estimator. Thus, we center for trimming when $y_t y_{t-h}$ can possibly have a symmetric distribution. See Section 2.

Other transforms can be used, including smoothed versions of $y_t y_{t-h} I(|y_t y_{t-h} - \gamma| < c)$ like Tukey's bisquare (cf Andrews, Bickel, Hampel, Huber, Rogers, and Tukey, 1972; Hampel, Ronchetti, Rousseeuw, and Stahel, 1986). However, our bias correction exploits Karamata theory for tail-trimmed moments, and this places severe restrictions on the available transforms, and excludes conventional transforms like Tukey's. Thus, we focus on simple trimming.

We do not employ truncation $sign(x) \times \min\{|x|, c\}$ because in the present context either it results in estimator bias that cannot be reduced by our bias correction methods when $c$ is fixed, or it does not lead to an asymptotically normal estimator when $c \to \infty$. In general, truncation or trimming with a bounded threshold $c \to (0, \infty)$, will promote a bounded influence function, and therefore lead to infinitessimal robustness (Hampel, 1974). The cost, however, is asymptotic bias that is not corrected by our methods, although such bias in principle can be corrected by using indirect inference and an assumed error distribution (e.g., Genton and Ronchetti, 2003; Mancini, Ronchetti, and Trojani, 2005). See, e.g., Martin and Thomson (1982) for data contamination robust spectral density estimation. This literature focuses on practice at the expense of theory details, and it does not appear to be extended to parameter estimation with asymptotic theory.

Even with negligibility due to $c \to \infty$, the transform $y_t y_{t-h} I(|y_t y_{t-h} - \gamma| < c)$ can lead to asymptotic bias in the limit distribution of our estimators due to potential asymmetry in the distribution of $y_t y_{t-h}$, and necessarily due to asymmetry in the distribution of $y_t^2$. This bias may not vanish fast enough asymptotically when $y_t$ has an infinite fourth moment, leading to bias in the limit distribution of our estimators of $\theta_0$. We therefore estimate and remove the bias using extreme value theory developed in Hill (2013), cf. Peng (2001), and show that power laws are not actually required for our bias estimators to work in practice.

A necessary trade-off arises from robustifying sample correlations against heavy tails. We implicitly assume $y_t$ has a finite variance, and when $E[y_t^4] = \infty$ then our estimator has a sub-$T^{1/2}$ convergence rate. This applies to estimators of ARMA and stochastic volatility models, which are often estimated in the frequency domain. Further, in the GARCH case our methods cover the square $y_t = x_t^2$ of GARCH $x_t$, hence $x_t$ must have a finite fourth moment. Nevertheless, we show by simulation experiment that our tail-trimmed FD-QML estimator performs better than the conventional FD-QML estimator in thin and thick tail cases: trimming just a few large values and using a bias correction strongly improves small sample bias, efficiency, and test performance, whether tails are thin or thick.

It is also interesting to note that, in principle, we do not need covariance stationarity for our robust Whittle estimator to be valid. For example, an AR(1) $y_t = \phi_0 y_{t-1} + \epsilon_t$ with $|\phi_0| < 1$, iid $\epsilon_t$ and $\sigma_0^2 = E[\epsilon_t^2] < \infty$, has a spectrum $\sigma_0^2 (2\pi)^{-1} \omega(\lambda, \phi_0)$ where $\omega(\lambda, \phi_0) = 1/(1 + \phi_0^2 - 2\phi_0 \cos(\lambda))$. Even when $E[\epsilon_t^2] = \infty$, we may evidently use $\omega(\lambda, \phi)$ to identify $\phi_0$ based on our tail-trimmed Whittle estimator, similar to Mikosch, Gadrich, Kluppelberg, and Adler (1995). However, we conjecture that our robust Whittle estimator is still asymptotically normal, contrary to Mikosch, Gadrich, Kluppelberg, and Adler (1995). We do not tackle the theory here in order to focus ideas, and therefore always maintain covariance stationarity.

There are various efforts to study frequency domain estimators for heavy tailed data. Mikosch, Gadrich, Kluppelberg, and Adler (1995) characterize the Whittle estimator for infinite variance ARMA models, and derive its non-standard limit distribution. Li (2008, 2010) develops Laplace and $L_p$-moment frequency domain estimators by replacing the usual periodogram with alternative ones for linear models $y_t = \theta' x_t + \epsilon_t$, where $\epsilon_t$ is finite dependent or a linear function of iid random variables. Our methods, by comparison, focus on the Fourier based spectrum by exploiting a robust sample version of $E[y_t y_{t-h}]$, and our spectrum class and weak dependence assumptions discussed in Section 2 cover a far larger class of processes than allowed in Li (2008, 2010). See also Shao and Wu (2007) for recent work on spectral estimation theory for nonlinear processes with more than a fourth moment under a geometric moment contraction assumption (as opposed to a mixing condition). We require a finite second moment and a positive continuous spectrum, allowing for geometric or hyperbolic memory decay in the form of a mixing condition.

Spectral analysis has been extended to indicator transforms, quantiles and copulas, all of which are inherently robust to extreme values (Dette, Hallin, Kley, and Volgushev, 2011; Lee and Rao, 2011; Hagemann, 2012), or are explicitly constructed for extreme values (Mikosch and Zhao, 2014). In these cases, spectral density estimators are not proposed to estimate model parameters, but to describe underlying dependence in a time series. Indeed, little is known about how these various spectra can be used to identify model parameters.

The remaining sections of this paper are as follows. In Section 2 we present the robust estimator of $E[y_t y_{t-h}]$ and FD-QML estimator, and tackle bias correction in Section 3. We discuss choosing the amount of trimming in Section 4, a simulation study follows in Section 5, and parting comments are left for Section 6. Since theory for our Whittle estimator is similar to our FD-QML estimator, we present those details in Section C of the supplemental material Hill and McCloskey (2014).

We use the following notation conventions. Drop $\theta_0$ and write $f(\lambda) = f(\lambda, \theta_0)$. All random variables lie on a common probability measure space $(\Omega, \mathcal{F}, \mathcal{P})$. $||A||$ is the spectral norm of matrix $A$; $[z]$ rounds a scalar $z$ to the nearest integer. $\mathcal{L}(c)$ denotes a slowly varying function: $\lim_{c\to\infty} \mathcal{L}(\lambda c)/\mathcal{L}(c) = 1$ $\forall \lambda > 0$ (e.g. $a(\ln(c))^b$ for finite $a > 0$ and $b \geq 0$). $K > 0$ is a finite constant that may change from place to place. $I(\cdot)$ is the indicator function: $I(A) = 1$ if $A$ is true, else $I(A) = 0$. Unless otherwise specified, all limits are as $T \to \infty$.

# 2 Robust Frequency Domain-QML

## 2.1 Robust FD-QML

Let $x$ be an arbitrary $\mathcal{F}$-measurable random variable. Centering with the mean $\gamma = E[x]$ ensures $E[xI(|x - \gamma| < c)]$ identifies $E[x]$ when $x$ has a symmetric distribution since, in this case, $E[xI(|x - \gamma| < c)] = E[x] \times E[I(|x - E[x]| < c)]$. If $x = y_t y_{t-h} \geq 0$ $a.s.$ then mean-centering is irrelevant, hence $\gamma = 0$.[1] Examples include $h = 0$, or $y_t$ is squared GARCH, or log-squared stochastic volatility with returns in $[-1, 1]$.

[1]Our weak dependence assumptions rule out $P(y_t y_{t-h} < 0) = 1$, hence $P(y_t y_{t-h} \geq 0) > 0$.

#### 2.1.1 Sequences for centering and trimming

We require sequences for centering $\{\tilde{\gamma}_{T,h}\}$, for determining the amount of trimming $\{k_{T,h}\}$, and for controlling the number of lags $\{b_T\}$ of cross-product moment estimators used in constructing the criterion function. First, $\{\tilde{\gamma}_{T,h}\}$. It is preferrable to ensure the transformed sample moments are unbiased whenever possible, and it is possible to do so when $y_t y_{t-h}$ has a symmetric distribution. We may not know whether $y_t y_{t-h}$ has a symmetric distribution, but it *cannot* when the distribution is non-degenerate with support $[0, \infty)$. Thus, the quantity we recommend for centering in practice is:

$$\tilde{\gamma}_{T,h} \equiv \begin{cases} \frac{1}{T-h} \sum_{t=h+1}^{T} y_t y_{t-h} & \text{if } P\left(y_t y_{t-h} > 0\right) < 1 \\ 0 & \text{if } P\left(y_t y_{t-h} > 0\right) = 1 \end{cases} \quad \text{for each exploited lag } h \geq 0.$$

Since mean-centering is only helpful when $y_t y_{t-h}$ has a symmetric distribution, if we know $y_t y_{t-h}$ has an asymmetric distribution then implicitly $\tilde{\gamma}_{T,h} = 0$. Thus, $\tilde{\gamma}_{T,0} = 0$ since $y_t^2 > 0$ $a.s.$, and when $y_t$ is a squared GARCH process then $\tilde{\gamma}_{T,h} = 0$ for each $h$.

It is important to understand that centering does not completely remove bias in our estimator since the transformed $y_t y_{t-h}$ at $h = 0$ will be biased, which adds to the small sample (and possibly limit distribution) bias of our FD estimators. This is irrelevant because we remove all bias by characterizing and estimating it in Section 3. Mean centering, however, does lead to an improved FD estimator in controlled experiments when $y_t y_{t-h}$ has a symmetric distribution.[2]

Second, $\{k_{T,h}\}$. We use $\{y_t y_{t-h} - \tilde{\gamma}_{T,h}\}_{t=1}^{T-h}$ to determine which $y_t y_{t-h}$ are trimmed, so define:

$$\widehat{\mathcal{Y}}_{h,t}^{(0)} \equiv |y_t y_{t-h} - \tilde{\gamma}_{T,h}| \text{ with order statistics } \widehat{\mathcal{Y}}_{h,(1)}^{(0)} \geq \widehat{\mathcal{Y}}_{h,(2)}^{(0)} \geq \cdots \widehat{\mathcal{Y}}_{h,(T-h)}^{(0)}.$$

The chosen threshold is the order statistic $\widehat{\mathcal{Y}}_{h,(k_{T,h})}^{(0)}$, where $\{k_{T,h}\}$ is an intermediate order sequence: $k_{T,h} \in \{1, ..., T-h\}$ and

$$k_{T,h} \to \infty \text{ and } \frac{k_{T,h}}{T-h} \to 0 \text{ for each exploited lag } h \geq 0. \tag{1}$$

The transformation we use is then

$$y_t y_{t-h} I\left(|y_t y_{t-h} - \tilde{\gamma}_{T,h}| < \widehat{\mathcal{Y}}_{h,(k_{T,h})}^{(0)}\right),$$

while $k_{T,h}/(T-h) \to 0$ ensures negligibility $y_t y_{t-h} I(|y_t y_{t-h} - \tilde{\gamma}_{T,h}| < \widehat{\mathcal{Y}}_{h,(k_{T,h})}^{(0)}) \xrightarrow{p} y_t y_{t-h}$.

Finally, $\{b_T\}$. Trimming negligibility (1) naturally places a restriction on the number of usable lags $h$. Consider that $h = T - a$ for constant $a \geq 1$ implies $k_{T,T-a} \in \{1, ..., a\}$ $\forall T$, which contradicts

[2] In the Section 5 simulation study we use mean-centering for AR model estimation and $h \geq 1$. In experiments not reported here we find that our FD-QML estimator has slightly higher small sample bias when mean-centering is not used.

the need for $k_{T,h} \to \infty$ to ensure robustness to heavy tails. Usable lags are therefore

$$h = \{0, 1, ..., b_T\}$$

for a sequence of bandwidths $\{b_T\}$ that satisfy

$$b_T \leq T - 1, \quad b_T \to \infty, \quad \text{and } T - b_T \to \infty.$$

Examples of $\{k_{T,h}, b_T\}$ include $b_T = [\delta_b(T-1)]^{a_b}$ and $k_{T,h} = [\delta_{k,h}(\ln(T))^{a_{k,h}}]$ for any $h$, and some $a_b \in (0,1]$, $\delta_b \in (0,1)$, and $\delta_{k,h}, a_{k,h} > 0$.[3]

### 2.1.2 Robust Periodogram and FD-QML

The periodogram has the well known expansion

$$\mathcal{I}_T(\lambda) \equiv \left| \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} y_t e^{-i\lambda t} \right|^2 = \frac{1}{2\pi} \left\{ \frac{1}{T} \sum_{t=1}^{T} y_t^2 + 2 \sum_{h=1}^{T-1} \frac{1}{T} \sum_{t=h+1}^{T} y_t y_{t-h} \cos(\lambda h) \right\}, \tag{2}$$

hence it is a function of unbiased estimators $1/T \sum_{t=h+1}^{T} y_t y_{t-h}$ of $(T-h)T^{-1}E[y_t y_{t-h}]$. Our robust estimator of $(T-h)T^{-1}E[y_t y_{t-h}]$ is then:

$$\widehat{\hat{\gamma}}^*_{T,h}(c) \equiv \begin{cases} \frac{T-h}{T-h-k_{T,h}} \frac{1}{T} \sum\limits_{t=h+1}^{T} y_t y_{t-h} I\left( \left| y_t y_{t-h} - \frac{1}{T-h} \sum\limits_{t=h+1}^{T} y_t y_{t-h} \right| < c \right) & \text{if } P\left(y_t y_{t-h} > 0\right) < 1 \\ \frac{1}{T} \sum\limits_{t=h+1}^{T} y_t y_{t-h} I\left( |y_t y_{t-h}| < c \right) & \text{if } P\left(y_t y_{t-h} > 0\right) = 1 \end{cases} \tag{3}$$

In view of (2), we simply use $\widehat{\hat{\gamma}}^*_{T,h}(\widehat{\mathcal{Y}}^{(0)}_{h,(k_{T,h})})$ in place of $1/T \sum_{t=h+1}^{T} y_t y_{t-h}$:

$$\widehat{\mathcal{I}}^*_T(\lambda) \equiv \frac{1}{2\pi} \left( \widehat{\hat{\gamma}}^*_{T,0}(\widehat{\mathcal{Y}}^{(0)}_{0,(k_{T,h})}) + 2 \sum_{h=1}^{b_T} \widehat{\hat{\gamma}}^*_{T,h}(\widehat{\mathcal{Y}}^{(0)}_{h,(k_{T,h})}) \times \cos(\lambda h) \right), \tag{4}$$

hence our negligibly transformed FD-QML estimator of $\theta_0$ is

$$\hat{\theta}^*_T = \underset{\theta \in \Theta}{\arg\min} \sum_{j \in \mathbb{F}} \left( \ln f(\lambda_j, \theta) + \frac{\widehat{\mathcal{I}}^*_T(\lambda_j)}{f(\lambda_j, \theta)} \right) \quad \text{where} \;\; \mathbb{F} \equiv (-T/2, T/2] \cap \mathbb{Z} \setminus \{0\}.$$

The construction of $\widehat{\hat{\gamma}}^*_{T,h}(c)$ ensures identification of $(T-h)T^{-1}E[y_t y_{t-h}]$ when $y_t y_{t-h}$ has a symmetric distribution. Simply note that for all asymptotic arguments we can replace the centering

[3] In practice we obviously want $b_T$ to be close to $T-1$, but our bias correction is based on tail exponent estimators, and for large $h$ there are very few extreme values to work with from the sample $\{y_t y_{t-h} - \tilde{\gamma}_{T,h}\}_{t=1}^{T-h}$. We find in our simulation experiments that $b_T = T^{.95}$ works well for any $T \geq 100$. Notice $b_T \approx 80$ when $T = 100$, while larger values do not necessarily permit viable bias estimates for $h$ near $b_T$.

$1/(T-h)\sum_{t=h+1}^{T} y_t y_{t-h}$ with $E[y_t y_{t-h}]$, and the threshold $\widehat{\mathcal{Y}}_{0,(k_{T,h})}^{(0)}$ with the non-random sequence $\{c_{T,h}\}$ that satisfies $P(|y_t y_{t-h} - E[y_t y_{t-h}]| > c_{T,h}) = k_{T,h}/(T-h)$. See Section 2.3, and see Lemma A.7 in Appendix A.2. Then, as per the discussion on mean-centering at the top of this section:

$$\frac{T-h}{T-h-k_{T,h}} \frac{T-h}{T} E\left[y_t y_{t-h} I\left(|y_t y_{t-h} - E\left[y_t y_{t-h}\right]| < c_{T,h}\right)\right]$$

$$= \frac{T-h}{T-h-k_{T,h}} P\left(|y_t y_{t-h} - [E y_t y_{t-h}]| < c_{T,h}\right) \times \frac{T-h}{T} E\left[y_t y_{t-h}\right]$$

$$= \frac{T-h}{T-h-k_{T,h}} \left(1 - \frac{k_{T,h}}{T-h}\right) \times \frac{T-h}{T} E\left[y_t y_{t-h}\right] = \frac{T-h}{T} E\left[y_t y_{t-h}\right].$$

Notice $\widehat{\mathcal{I}}_T^*(\lambda)$ can in principle be negative for a finite sample because the sequence $\{\widehat{\gamma}_{T,h}^*(\widehat{\mathcal{Y}}_{h,(k_{T,h})}^{(0)})\}_{h=0}^{b_T}$ need not be non-negative definite simply due to trimming ($k_{T,h} \geq 1$) and lag truncation ($b_T \leq T-1$). Hence, strictly speaking $\widehat{\mathcal{I}}_T^*(\lambda)$ is not a periodogram. This is irrelevant here since we only need to estimate $\theta_0$, and $\widehat{\mathcal{I}}_T^*(\lambda)$ can be used for a consistent and asymptotically normal estimator.

### 2.2 Assumptions

We present first the assumptions and then the main results. Asymptotic theory requires a sequence $\{c_{T,h}\}$ of non-random numbers $c_{T,h} \geq 0$ that the order statistic $\widehat{\mathcal{Y}}_{h,(k_{T,h})}^{(0)}$ of $|y_t y_{t-h} - \tilde{\gamma}_{T,h}|$ approximates. Since $\tilde{\gamma}_{T,h}$ exists for lags $h = 0, ..., b_T$, and $T - b_T \to \infty$, it follows that under mean-centering $\tilde{\gamma}_{T,h} = (T-h)^{-1}\sum_{t=h+1}^{T} y_t y_{t-h}$ is consistent for $E[y_t y_{t-h}]$ under regularity assumptions detailed below. Define the population centering parameter:

$$\tilde{\gamma}_h \equiv \begin{cases} E\left[y_t y_{t-h}\right] & \text{if } P\left(y_t y_{t-h} > 0\right) < 1 \\ 0 & \text{if } P\left(y_t y_{t-h} > 0\right) = 1 \end{cases} \quad \text{for } h \geq 0. \tag{5}$$

Now let $\{c_{T,h}\}$ be any (possibly non-unique) sequence that satisfies:

$$P\left(|y_t y_{t-h} - \tilde{\gamma}_h| \geq c_{T,h}\right) = \frac{k_{T,h}}{T-h} \text{ for } h \geq 0. \tag{6}$$

We need to impose bounds on dependence coefficients. Define $\Im_s^t \equiv \sigma(y_\tau : s \leq \tau \leq t)$, a measure space $(\Omega, \sigma(\cup_{t\in\mathbb{Z}}\Im_{-\infty}^t), \mathcal{P})$, and $\alpha$-mixing coefficients $\alpha_h \equiv \sup_{\mathcal{A}\subset\Im_{-\infty}^{t-h}, \mathcal{B}\subset\Im_t^\infty} |\mathcal{P}(\mathcal{A}\cap\mathcal{B}) - \mathcal{P}(\mathcal{A})P(\mathcal{B})|$. Next, let $\mathcal{L}_2(\mathcal{A}) \equiv \mathcal{L}_2(\Omega, \mathcal{A}, \mathcal{P})$ be the space of $\mathcal{A}$-measurable $L_2$-bounded random variables for $\mathcal{A} \subseteq \sigma(\cup_{t\in\mathbb{Z}}\Im_{-\infty}^t)$. Define $\rho(\mathcal{A}, \mathcal{B}) \equiv \sup_{f\in\mathcal{L}_2(\mathcal{A}), g\in\mathcal{L}_2(\mathcal{B})} |corr(f,g)|$ and let $\mathfrak{S}_h$ and $\mathfrak{T}_h$ be non-empty subsets of $\mathbb{N}$ with $\inf_{s\in\mathfrak{S}_h, t\in\mathfrak{T}_h}\{|s-t|\} \geq h$. Then the *interlaced maximal correlation coefficient* is $\rho_h^* \equiv \sup_{\mathfrak{S}_h, \mathfrak{T}_h} \rho(\sigma(y_t : t \in \mathfrak{S}_h), \sigma(y_s : s \in \mathfrak{T}_h))$, where the supremum is taken over all $\mathfrak{S}_h$ and $\mathfrak{T}_h$ (see Bradley, 1993).

**Assumption A (data generating process).**

1. *$\{y_t y_{t-h}\}$ is a stationary, $L_p$-bounded process, $p > 1$, with an absolutely continuous non-degenerate distribution with unbounded support. Further, $y_t$ has absolutely summable covariances, and is $\alpha$-mixing $\alpha_h = O(h^{-p/(p-2)})$ with $\rho_1^* < 1$.*

2. *$y_t$ has spectrum $f(\lambda, \theta_0)$ for unique $\theta_0$ in the interior of compact $\Theta \subset \mathbb{R}^k$ with properties:*

   *(i) if $\theta_0 \neq \theta$ then $f(\lambda; \theta) \neq f(\lambda; \theta_0)$;*

   *(ii) $0 < f(\lambda, \theta) \leq K < \infty$ for each $\lambda \in [-\pi, \pi]$ and $\theta \in \Theta$;*

   *(iii) $f(\lambda, \theta)$ is twice continuously differentiable in $\theta$, with derivatives $(\partial/\partial\theta)^i f(\lambda, \theta)$ for $i = 1, 2$ uniformly bounded on $[-\pi, \pi] \times \Theta$;*

   *(iv) $h(\lambda, \theta) \in \{f(\lambda, \theta), (\partial/\partial\theta) f(\lambda, \theta)\}$ are uniformly Hölder continuous of degree $\alpha \in (1/2, 1]$ in $\lambda$: $\sup_{\theta \in \Theta} ||h(\lambda, \theta) - h(\omega, \theta)|| \leq K|\lambda - \omega|^\alpha$ for all $\lambda, \omega \in [-\pi, \pi]$ and some $K > 0$.*

3. $\inf_{\theta \in \Theta} ||\mathcal{H}(\theta)|| > 0$.

**Remark 1** Distribution continuity A.1 simplifies order statistic asymptotics. The assumption can always be assured by adding a small iid noise with a continuous distribution to $y_t$. Geometric $\alpha$-mixing arises in stationary ARMA, squared GARCH, and log-squared autoregressive stochastic volatility, as well as a variety of nonlinear processes. See, e.g., Doukhan (1994), Carrasco and Chen (2002) and Meitz and Saikkonen (2008).

**Remark 2** Assumption A.2 is essentially C2.1, C2.2 and C2.4 in Dunsmuir (1979) since $\theta_0$ in the interior of compact $\Theta$ is presumed by Dunsmuir (1979). We relax Dunsmuir (1979)'s C2.3 since that assumes $\{y_t\}$ has a finite fourth moment. Hölder continuity Assumption A.2.iv with any degree $\alpha \in (1/2, 1]$ applies to ARMA, squared GARCH, and log-squared autoregressive stochastic volatility on some compact $\Theta$ containing $\theta_0$.

**Remark 3** The mixing rate $\alpha_h = O(h^{-p/(p-2)})$ ensures $k_{T,h}^{1/2}$-convergence for the thresholds $\widehat{\mathcal{Y}}_{h,(k_{T,h})}^{(0)}$, cf. Hill (2010, 2014b). Standardized partial sums of the transformed variables, however, may only have a second moment asymptotically, hence we must exploit a dependence property other than $\alpha$-mixing for a central limit theory to apply. Indeed, it will not help to further restrict the rate $\alpha_h \to 0$ since most central limit theorems for dependent arrays require more than a second moment (see, e.g., Ibragimov, 1975; Bradley, 1992). Further, in order to reduce the number of cases in proofs, we desire partial sum variances to be strictly positive, a well known challenge (see Ibragimov, 1962, 1975; Dehling, Denker, and Phillip, 1986; Peligrad, 1996). A simple way to ensure a Gaussian limit theory, and partial sum variance positivity, is to impose $\alpha_h \to 0$ and $\rho_1^* < 1$. Of course, $\alpha_h \to 0$ and the existence of a continuous positive spectral density suffice as well (cf. Ibragimov, 1962) since that implies $\rho_1^* < 1$ (Bradley, 1992), but it is unknown whether any of the triangular arrays in this paper that arise from the sample-size specific data transformation have positive spectra at frequency zero.

**Remark 4** We use A.3 to show that $\hat{\theta}_T^*$ is the same asymptotically if we replace the stochastic threshold $\widehat{\mathcal{Y}}_{h,(k_{T,h})}^{(0)}$ with the deterministic one $c_{T,h}$. See the proof of consistency Theorem 2.1.

We assume $y_t y_{t-h}$ have regularly varying probability tails in order to allow for heavy tails. Recall $\tilde{\gamma}_h = E[y_t y_{t-h}]$ if $P(y_t y_{t-h} > 0) < 1$, else $\tilde{\gamma}_h \equiv 0$.

**Assumption B (regularly varying tails).** *$P(y_t y_{t-h} - \tilde{\gamma}_h \leq -c) = \mathcal{L}_{h,1}(c)c^{-\kappa_{h,1}}$ and $P(y_t y_{t-h} - \tilde{\gamma}_h \geq c) = \mathcal{L}_{h,2}(c)c^{-\kappa_{h,2}}$ where $\mathcal{L}_{h,i}(c)$ are slowly varying, and $\kappa_{h,i} > 1$.*

**Remark 5** The tail index $\kappa_{h,0} \equiv \min\{\kappa_{h,1}, \kappa_{h,2}\}$ is identically the moment supremum $\kappa_{h,0} = \arg\sup\{\alpha > 0 : E|y_t y_{t-h}|^{\alpha} < \infty\} > 1$, hence $\kappa_{0,0} = \kappa/2$ and

$$P\left(|y_t y_{t-h} - \tilde{\gamma}_h| \geq c\right) = \mathcal{L}_{h,0}(c)c^{-\kappa_{h,0}}(1 + o(1)) \text{ where } \mathcal{L}_{h,0}(c) = \mathcal{L}_{h,1}(c) + \mathcal{L}_{h,2}(c). \tag{7}$$

See Resnick (1987). Since $y_t^2 \geq 0$ only has a right tail, notice $\kappa_{0,0} = \kappa_{0,2} = \kappa/2 > 1$ and $\mathcal{L}_{0,0}(c) = \mathcal{L}_{0,2}(c)$. In general the moment supremum $\kappa_{h,0} \in [\kappa/2, \kappa]$ hence $\kappa_{h,0} \geq \kappa_{0,0}$. Power law tails in linear and random volatility processes is extensively treated (e.g. Brockwell and Cline, 1985; Basrak, Davis, and Mikosch, 2002; Davis and Mikosch, 2009), while their extensions two products is tackled, e.g., in Cline (1986) and Mikosch and Starica (2000).

Finally, under mean-centering we require that $(T - h)^{-1}\sum_{t=h+1}^{T} y_t y_{t-h}$ does not asymptotically impact the order statistic $\widehat{\mathcal{Y}}_{h,(k_{T,h})}^{(0)}$. By exploiting a maximal inequality in Hansen (1991), we use the non-sharp bound $(T - h)^{-1}\sum_{t=h+1}^{T} y_t y_{t-h} = E[y_t y_{t-h}] + O_p(1/T^{\iota})$ for tiny $\iota > 0$ in our proof of $\widehat{\mathcal{Y}}_{h,(k_{T,h})}^{(0)}/c_{T,h} = 1 + O_p(1/k_{T,h}^{1/2})$, and $O_p(1/T^{\iota}) = o_p(1/k_{T,h}^{1/2})$ when $k_{T,h} \to \infty$ at most at a slowly varying rate. A sharper bound can be obtained if we know the rate of convergence of $(T - h)^{-1}\sum_{t=h+1}^{T} y_t y_{t-h}$, but this requires knowledge of the rate of tail decay of $y_t$, and in general requires additional information about probability tails since $y_t y_{t-h}$ is weakly dependent (e.g. Bartkiewicz, Jakubowski, Mikosch, and Wintenberger, 2010).

**Assumption C (trimming and bandwidth rates).**

1. *In general $k_{T,h} \to \infty$, $k_{T,h}/(T - h) \to 0$, and $k_{T,h} \sim K_{h,\tilde{h}} k_{T,\tilde{h}}$ for some $K_{h,\tilde{h}} > 0$ and each $\tilde{h}, h \in \{0, ..., b_T\}$. If mean-centering at lag $h$ is used then $k_{T,h} \to \infty$ at most at a slowly varying rate.*

2. *Let $b_T \leq T - 1$, $b_T \to \infty$, and $T - b_T \to \infty$. Further $b_T/T^{1/(2\alpha)} \to \infty$ where $\alpha \in (1/2, 1]$ is the Hölder continuity degree in Assumption A.2.iv.*

**Remark 6** $k_{T,h+1} \sim K k_{T,h}$ allows us to give a simple characterization of the rate of convergence of $\hat{\theta}_T^*$. See the proofs of Theorem 2.2, and a key central limit theorem Lemma A.9 in Appendix A.2. The restriction, though, implies when mean-centering is used for *some* $h$ that $k_{T,h} \to \infty$ at most at a slowly varying rate for *each* $h$. We discuss in Section 4 that a slowly increasing $k_{T,h} \to \infty$ for all $h$ both reduces small sample bias, and improves the accuracy of our bias estimator.

**Remark 7** $b_T/T^{1/(2\alpha)} \to \infty$ is required since we use Hölder continuity to show feasible and infeasible FD-QML estimators are asymptotically equivalent, and for a central limit theorem, similar to Dunsmuir (1979, proof of Theorenm 2.1, Corollary 2.2). In our case, however, there are $T - 1$ Fourier frequencies $\{\lambda_j\}_{j\in\mathbb{F}}$ but only $b_T = o(T)$ lags. If Assumption A.2.iv holds for *any* $\alpha \in (1/2, 1]$ then take $\alpha = 1$ and assume $b_T/T^{1/2} \to \infty$. This applies to ARMA, squared GARCH and log-squared autoregressive stochastic volatility.

### 2.3 Main Results

Let $\kappa$ be the moment supremum of $y_t$,

$$\kappa \equiv \arg\sup\left\{\alpha > 0 : E\left|y_t\right|^{\alpha} < \infty\right\},$$

and define the scaled gradient

$$\varpi(\lambda,\theta) = [\varpi_i(\lambda,\theta)]_{i=1}^{k} \equiv -\frac{1}{f(\lambda,\theta)}\frac{\partial}{\partial\theta}\ln f(\lambda,\theta) \quad \text{and} \quad \varpi(\lambda) = \varpi(\lambda,\theta_0)$$

and Hessian:

$$\mathcal{H}(\theta) = -\frac{1}{2\pi}\int_{-\pi}^{\pi}\varpi(\lambda,\theta)\frac{\partial}{\partial\theta'}f(\lambda,\theta)d\lambda - \frac{1}{2\pi}\int_{-\pi}^{\pi}\left\{f(\lambda,\theta) - f(\lambda)\right\}\frac{\partial}{\partial\theta}\varpi(\lambda,\theta)d\lambda.$$

Calling $\mathcal{H}(\theta)$ a Hessian is justified since asymptotically with probability approaching one $\mathcal{H}(\theta) = (\partial/\partial\theta)^2\sum_{j\in\mathbb{F}}\{\ln f(\lambda_j,\theta) + \widehat{\mathcal{I}}_T^*(\lambda_j)/f(\lambda_j,\theta)\}$. See the proof of Theorem 2.1 in Appendix A.3.

The first main result follows. Proofs are given in Appendix A.

**Theorem 2.1** *Under Assumptions A-C we have* $\hat{\theta}_T^* \xrightarrow{p} \theta_0$.

**Remark 8** We cannot prove consistency using less structure on the spectrum. In particular, to handle the order statistic $\widehat{\mathcal{Y}}_{h,(k_{T,h})}^{(a)}$ in the trimming indicator, we exploit spectrum differentiability and Hölder continuity to allow a Taylor expansion argument.

The asymptotic distribution of $\hat{\theta}_T^*$ closely follows classic arguments. Define the robust moment estimator with population mean centering:

$$\hat{\gamma}_{T,h}^*(c) \equiv \begin{cases} \frac{T-h}{T-h-k_{T,h}}\frac{1}{T}\sum\limits_{t=h+1}^{T} y_t y_{t-h} I\left(\left|y_t y_{t-h} - E\left[y_t y_{t-h}\right]\right| < c\right) & \text{if } P\left(y_t y_{t-h} > 0\right) < 1 \\ \frac{1}{T}\sum\limits_{t=h+1}^{T} y_t y_{t-h} I\left(\left|y_t y_{t-h}\right| < c\right) & \text{if } P\left(y_t y_{t-h} > 0\right) = 1 \end{cases} \tag{8}$$

define a spectral density estimator

$$\mathcal{I}_T^*(\lambda) \equiv \frac{1}{2\pi}\left(\hat{\gamma}_{T,0}^*(c_{T,0}) + 2\sum_{h=1}^{b_T}\hat{\gamma}_{T,h}^*(c_{T,h})\cos(\lambda h)\right),$$

and construct Hessian, covariance, and scale matrices:

$$\Omega \equiv \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\partial}{\partial\theta}\ln f(\lambda;\theta_0)\frac{\partial}{\partial\theta'}\ln f(\lambda;\theta_0)d\lambda \tag{9}$$
$$\mathcal{S}_T \equiv T\times E\left[\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\mathcal{I}_T^*(\lambda)-E\left[\mathcal{I}_T^*(\lambda)\right]\right)\varpi(\lambda)d\lambda\right\}\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\mathcal{I}_T^*(\lambda)-E\left[\mathcal{I}_T^*(\lambda)\right]\right)\varpi(\lambda)d\lambda\right\}'\right]$$
$$\mathcal{V}_T = \Omega^{-1}\mathcal{S}_T\Omega^{-1}.$$

Throughout $\varpi_h$ denotes the $h^{th}$ Fourier coefficient of $\varpi(\lambda) \equiv -(f(\lambda))^{-1}(\partial/\partial\theta)\ln f(\lambda)$.

**Theorem 2.2** *Under Assumptions A-C* $T^{1/2}\mathcal{V}_T^{-1/2}(\hat{\theta}_T^* - \theta_0 + \mathcal{B}_T) \overset{d}{\to} N(0, I_k)$ *where*

$$\mathcal{B}_T \equiv \Omega^{-1}\frac{1}{(2\pi)^2}\sum_{h=-b_T}^{b_T}\varpi_h\left\{E\left[\hat{\gamma}_{T,|h|}^*(c_{T,|h|})\right] - E\left[y_t y_{t-h}\right]\right\}.$$

*If* $y_t y_{t-h}$ *is symmetrically distributed for each* $h > 0$ *then* $\mathcal{B}_T = \Omega^{-1}(2\pi)^{-2}\varpi_0\{E[y_t^2 I(y_t^2 < c_{T,0})] - E[y_t^2]\}$. *In general* $||\mathcal{V}_T|| = K||\mathcal{S}_T|| \sim KE[y_t^4 I(y_t^2 < c_{T,0})]$, $\liminf_{T\to\infty}||\mathcal{V}_T|| > 0$ *and* $T^{1/2}/||\mathcal{V}_T||^{1/2} \to \infty$. *If* $\kappa > 4$ *then* $||\mathcal{V}_T|| \sim K$, *if* $\kappa = 4$ *then* $||\mathcal{V}_T|| \to \infty$ *is slowly varying, and if* $\kappa \in (2,4)$ *then* $||\mathcal{V}_T|| \sim K(T/k_{T,0})^{4/\kappa-1}$.

**Remark 9** $||\mathcal{S}_T||$ is proportional to $E[y_t^4 I(y_t^2 < c_{T,0})]$ because the trimming fractiles are proportional $k_{T,h} \sim K_{h,\tilde{h}} k_{T,\tilde{h}}$ for $K_{h,\tilde{h}} > 0$ under Assumption C.1. If fractile proportionality is relaxed, then in general we cannot show $||\mathcal{S}_T||$ is proportional to a particular trimmed moment, and this complicates deducing the rate of convergence of $\hat{\theta}_T^*$ .

**Remark 10** The rate of convergence $T^{1/2}/||\mathcal{V}_T||^{1/2}$ depends on tail decay, and if $\kappa < 4$ then the rate is $T^{1/2}/(T/k_{T,0})^{2/\kappa-1/2}$. By Assumption C.1, if mean-centering is not used then we are free to choose all $k_{T,h}$ hence $k_{T,0} \to \infty$ provided $k_{T,0}/T = o(1)$. By setting $k_{T,0} = T/\mathcal{L}(T)$ for slowly increasing $\mathcal{L}(T) \to \infty$ we achieve a rate arbitrarily close to $T^{1/2}$ when $\kappa < 4$. This is similar to heavy tail robust estimators in Hill (2012b, 2013, 2014a). If mean-centering is used then $k_{T,h} \to \infty$ no faster than a slowly varying function, while $k_{T,h}$ are all proportional, hence $T^{1/2}/(T/k_{T,0})^{2/\kappa-1/2}$ cannot be made arbitrarily close to $T^{1/2}$. See Section 2.2 for a discussion of the assumptions.

**Remark 11** $\mathcal{S}_T$ can be estimated by standard methods (e.g. Chiu, 1988).

The next result summarizes when bias vanishes such that $\hat{\theta}_T^*$ is asymptotically unbiased in its limit distribution. The proof is presented in Hill and McCloskey (2014) since it is based on well known moment properties by Karamata theory given regular variation Assumption B.

**Theorem 2.3** *Under Assumptions A-C:*

*i.* $T^{1/2}\mathcal{V}_T^{-1/2}(\hat{\theta}_T^* - \theta_0) \overset{d}{\to} N(0, I_k)$ *if either* $\kappa > 4$; *or* $\kappa = 4$, $k_{T,h} = o(\ln(T))$ *and* $P(|y_t y_{t-h} - \tilde{\gamma}_h| \geq c) = d_{h,0}c^{-\kappa_{h,0}}(1 + o(1))$ *where* $\kappa_{0,0} = 2 \leq \kappa_{h,0}$;

*ii.* $T^{1/2}\mathcal{V}_T^{-1/2}(\hat{\theta}_T^* - \theta_0 + \mathcal{B}_T) \xrightarrow{d} N(0, I_k)$ *if* $\kappa \in (2,4)$*, where* $T^{1/2}||\mathcal{V}_T^{-1/2}\mathcal{B}_T|| \sim K k_{T,h}^{1/2} \to \infty$ *if* $y_t y_{t-h}$ *has a symmetric distribution, else* $\lim_{T\to\infty} T^{1/2}||\mathcal{V}_T^{-1/2}\mathcal{B}_T||/k_{T,h}^{1/2} \in [0,\infty)$.

**Remark 12** Case (i) can be expanded to include other "thin tail" cases, for example when $y_t$ has exponential tails. Although we do not provide details here, if $\lim_{c\to\infty} c^\kappa P(|y_t| > c) = 0$ for any finite $\kappa > 0$, then (i) holds provided $k_{T,h}/\ln(T) \to 0$. In general, in the hairline infinite fourth moment case $\kappa = 4$, as long as $k_{T,h} \to \infty$ sufficiently slow, then bias is small and vanishes rapidly enough.

**Remark 13** Case (ii) is similar to the far simpler case of a tail-trimmed mean of an invariance process, noted in Csörgő, Horváth, and Mason (1986).

If $E[y_t^4] < \infty$ then negligibility ensures the data transformation does not affect the asymptotic variance, and $\hat{\theta}_T^*$ is asymptotically unbiased in its limit distribution by Theorem 2.3.i. The asymptotic variance has a classic structure if $y_t$ is linear,

$$y_t = \sum_{i=0}^{\infty} \xi_i(\theta_0)\, \epsilon_{t-i} \text{ where } \sum_{i=0}^{\infty} \xi_i^2(\theta_0) < \infty,\ \xi_0(\theta_0) = 1 \tag{10}$$
$$\sigma^2(\theta_0) \equiv E[\epsilon_t^2] < \infty \text{ and } E[\epsilon_s \epsilon_t] = 0\ \forall s \neq t,$$

where $\epsilon_t$ is a homoscedastic martingale difference. The following is a consequence of Theorems 2.2 and 2.3, negligibility, dominated convergence, and arguments in Dunsmuir (1979, proof of Theorem 2.1). Notice $E[y_t^4] < \infty$ ensures under mean-centering and mixing Assumption A.1 $(T-h)^{-1}\sum_{t=h+1}^T y_t y_{t-h}$ $= E[y_t y_{t-h}] + O_p(1/T^{1/2})$, hence Assumption C.1 is superfluous by the discussion preceding it in Section 2.2. See also Remark 17 following Lemma A.4 in Appendix A.1. Define the $\sigma$-field $\Im_t \equiv \sigma(y_\tau : \tau \leq t)$.

**Corollary 2.4** *In addition to Assumptions A and C.2, let* $y_t$ *satisfy (10), and assume* $E[\epsilon_t|\Im_{t-1}] = 0$ *a.s.,* $E[\epsilon_t^2|\Im_{t-1}] = \sigma^2$ *a.s.,* $E[\epsilon_t^3|\Im_{t-1}] = s$ *a.s., and* $E[\epsilon_t^4] = \mathcal{K} < \infty$*. Let* $k_{T,h} \to \infty$ *and* $k_{T,h}/(T - h) = o(1)$*. Then* $T^{1/2}(\hat{\theta}_T^* - \theta_0) \xrightarrow{d} N(0, \mathcal{V})$ *where* $\mathcal{V} = \lim_{T\to\infty} \mathcal{V}_T = \Omega^{-1}(2\Omega + \Pi)\Omega^{-1}$ *and*

$$\Omega \equiv \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\partial \ln f(\lambda;\theta_0)}{\partial\theta}\frac{\partial \ln f(\lambda;\theta_0)}{\partial\theta'}d\lambda \quad \textit{and} \quad \Pi \equiv \frac{\left(\mathcal{K} - 3\sigma^4\right)}{\sigma^8(\theta_0)}\frac{\partial\sigma^2(\theta_0)}{\partial\theta}\frac{\partial\sigma^2(\theta_0)}{\partial\theta'}. \tag{11}$$

# 3 Bias Corrected FD-QML

We need a bias-corrected version of $\widehat{\gamma}_{T,h}^*(\widehat{\mathcal{Y}}_{h,(k_{T,h})}^{(0)})$ to ensure asymptotically unbiased FD-QML estimation. If $y_t y_{t-h}$ is known to have a symmetric distribution for $h \geq 1$, then in theory we need only treat $\widehat{\gamma}_{T,0}^*(\widehat{\mathcal{Y}}_{0,(k_{T,0})}^{(0)})$. Our simulation experiments, however, show that for heavy tailed processes bias estimation and removal is helpful even when true bias is zero. This occurs since $T - h$ may be very small, and a random draw $\{y_t y_{t-h}\}_{t=h+1}^T$ may appear asymmetrically distributed. Symmetric trimming can, therefore, add bias in small samples where none in theory exists.

### 3.1 Bias Correction : Power Law Tails

We must first specify the slowly varying components $\mathcal{L}_{h,i}(c)$ in Assumption B in order to parametrically characterize, and therefore estimate, bias in $\widehat{\hat{\gamma}}^{*}_{T,h}(\widehat{\mathfrak{Y}}^{(0)}_{h,(k_{T,h})})$. We assume for convenience:

$$P\left(y_t y_{t-h} - \tilde{\gamma}_h \leq -c\right) = d_{h,1} c^{-\kappa_{h,1}} \left(1 + o(1)\right) \text{ and } P\left(y_t y_{t-h} - \tilde{\gamma}_h \geq c\right) = d_{h,2} c^{-\kappa_{h,2}} \left(1 + o(1)\right), \quad (12)$$

where $\tilde{\gamma}_h = E\left[y_t y_{t-h}\right]$ if $P(y_t y_{t-h} > 0) < 1$, else $\tilde{\gamma}_h \equiv 0$. We exploit a second order version of (12) under Assumption B$'$, below, so that arguments for bias approximation sharpness in Peng (2001) and Hill (2013) apply. Other tails can in principle be considered, with corrections to arguments in Peng (2001, p. 259-263) and the following bias formulas.

We can replace $\widehat{\hat{\gamma}}^{*}_{T,h}(\widehat{\mathfrak{Y}}^{(0)}_{h,(k_{T,h})})$ based on (3) with $\hat{\gamma}^{*}_{T,h}(c_{h,T})$ in (8) for all asymptotic arguments: see Lemma A.7 in Appendix A.2. Thus, we need an expression for $E[\hat{\gamma}^{*}_{T,h}(c_{T,h})]$. If $P(y_t y_{t-h} > 0) = 1$, then by (12) with $\kappa_{h,2} = \kappa_{h,0}$, and Karamata's Theorem (see, e.g., (A.3) in Appendix A.2, and see Resnick, 1987, Theorem 0.6), for $h \geq 0$:

$$\begin{aligned} E\left[\hat{\gamma}^{*}_{T,h}(c_{T,h})\right] &= \frac{T-h}{T}\left\{E\left[y_t y_{t-h}\right] - E\left[y_t y_{t-h} I\left(|y_t y_{t-h}| \geq c_{T,h}\right)\right]\right\} \\ &\sim \frac{T-h}{T}\left\{E\left[y_t y_{t-h}\right] - \frac{1}{\kappa_{h,0}-1}\frac{k_{T,h}}{T-h}c_{T,h}\right\} \\ &= \frac{T-h}{T}E\left[y_t y_{t-h}\right] - \frac{1}{\kappa_{h,0}-1}\frac{k_{T,h}}{T}c_{T,h}. \end{aligned} \quad (13)$$

Now suppose $P(y_t y_{t-h} > 0) < 1$. Since by (12) each $c_{T,h} \sim d_{h,0}^{1/\kappa_{h,0}}(T/k_{T,h})^{1/\kappa_{h,0}}$, hence for $h \geq 0$:

$$\begin{aligned} &E\left[\left(y_t y_{t-h} - E[y_t y_{t-h}]\right) I\left(|y_t y_{t-h} - E[y_t y_{t-h}]| \geq c_{T,h}\right)\right] \\ &\quad = \int_{c_{T,h}}^{\infty}\left\{P\left(y_t y_{t-h} - E[y_t y_{t-h}] \geq u\right) - P\left(y_t y_{t-h} - E[y_t y_{t-h}] \leq -u\right)\right\} du \\ &\quad = \int_{c_{T,h}}^{\infty}\left\{d_{h,2}u^{-\kappa_{h,2}} - d_{h,1}u^{-\kappa_{h,1}}\right\} du \sim \frac{d_{h,2}c_{T,h}^{1-\kappa_{h,2}}}{\kappa_{h,2}-1} - \frac{d_{h,1}c_{T,h}^{1-\kappa_{h,1}}}{\kappa_{h,1}-1}. \end{aligned}$$

Therefore:

$$\begin{aligned} E\left[\hat{\gamma}^{*}_{T,h}(c_{T,h})\right] &= \frac{T-h}{T-h-k_{h,T}}\frac{T-h}{T}E\left[y_t y_{t-h} I\left(|y_t y_{t-h} - E[y_t y_{t-h}]| < c_{T,h}\right)\right] \\ &= \frac{T-h}{T}\left\{E[y_t y_{t-h}] - \frac{T-h}{T-h-k_{h,T}}E\left[\left(y_t y_{t-h} - E[y_t y_{t-h}]\right) I\left(|y_t y_{t-h} - E[y_t y_{t-h}]| \geq c_{T,h}\right)\right]\right\} \\ &\sim \frac{T-h}{T}E[y_t y_{t-h}] - \frac{T-h}{T}\frac{T-h}{T-h-k_{h,T}}\left(\frac{d_{h,2}c_{T,h}^{1-\kappa_{h,2}}}{\kappa_{h,2}-1} - \frac{d_{h,1}c_{T,h}^{1-\kappa_{h,1}}}{\kappa_{h,1}-1}\right). \end{aligned} \quad (14)$$

Asymptotic approximations (13) and (14) lead to asymptotically negligible approximation errors under second order power law Assumption B$'$, below, cf. Peng (2001) and Hill (2013). Thus, (13) and

(14) implicitly define asymptotic approximations of biases $-(E[\hat{\gamma}^*_{T,h}(c_{T,h})] - (T-h)T^{-1}E[y_t y_{t-h}])$:

$$\mathcal{R}_{T,h} \equiv \begin{cases} \frac{1}{\kappa_{h,0}-1}\frac{k_{T,h}}{T}c_{T,h} & \text{if } P\left(y_t y_{t-h} > 0\right) = 1 \text{ (e.g. } h=0) \\ \frac{T-h}{T}\frac{T-h}{T-h-k_{h,T}}\left(\frac{d_{h,2}c_{T,h}^{1-\kappa_{h,2}}}{\kappa_{h,2}-1} - \frac{d_{h,1}c_{T,h}^{1-\kappa_{h,1}}}{\kappa_{h,1}-1}\right) & \text{else} \end{cases}$$

The bias terms $\mathcal{R}_{T,h}$ are easily estimated. First, $c_{T,h}$ is estimated with $\widehat{\mathcal{Y}}^{(0)}_{h,(k_{T,h})}$. Second, for tail exponents $d_{h,i}$ and $\kappa_{h,i}$, define left, right and two-tailed observations

$$\widehat{\mathcal{Y}}^{(0)}_{h,t} \equiv |y_t y_{t-h} - \tilde{\gamma}_{T,h}|$$
$$\widehat{\mathcal{Y}}^{(1)}_{h,t} \equiv -\left(y_t y_{t-h} - \tilde{\gamma}_{T,h}\right) I\left(y_t y_{t-h} - \tilde{\gamma}_{T,h} < 0\right) \text{ and } \widehat{\mathcal{Y}}^{(2)}_{h,t} \equiv \left(y_t y_{t-h} - \tilde{\gamma}_{T,h}\right) I\left(y_t y_{t-h} - \tilde{\gamma}_{T,h} \geq 0\right),$$

and let $\{m_{T,h}\}$ be an intermediate order sequence:

$$m_{T,h} \in \{1, ..., T-h\}, \quad m_{T,h} \to \infty \text{ and } m_{T,h}/(T-h) \to \infty.$$

Many estimators are possible, but we use Hill (1975)'s estimator of $\kappa_{h,i}$ and Hall (1982)'s estimator of $d_{h,i}$ due to their popularity, and available limit theory (see Hill, 2009, 2010, 2014b):

$$\hat{\kappa}_{h,i,m_{T,h}} \equiv \left(\frac{1}{m_{T,h}}\sum_{j=1}^{m_{T,h}} \ln\left(\frac{\widehat{\mathcal{Y}}^{(i)}_{h,(j)}}{\widehat{\mathcal{Y}}^{(i)}_{h,(m_{T,h}+1)}}\right)\right)^{-1} \text{ and } \hat{d}_{h,i,m_{T,h}} \equiv \frac{m_{T,h}}{T-h}\left(\widehat{\mathcal{Y}}^{(i)}_{h,(m_{T,h})}\right)^{\hat{\kappa}_{h,i,m_{T,h}}} \text{ for } i = 0,1,2.$$

The bias-corrected tail-trimmed estimators are therefore

$$\hat{\hat{\gamma}}^{(bc)}_{T,h}(\widehat{\mathcal{Y}}^{(0)}_{h,(k_{T,h})}) \equiv \hat{\hat{\gamma}}^{*}_{T,h}(\widehat{\mathcal{Y}}^{(0)}_{h,(k_{T,h})}) + \hat{\mathcal{R}}_{T,h} \text{ for } h \geq 0, \tag{15}$$

where:

$$\hat{\mathcal{R}}_{T,h} = \frac{1}{\hat{\kappa}_{h,0,m_{T,h}} - 1}\frac{k_{T,h}}{T}\widehat{\mathcal{Y}}^{(0)}_{h,(k_{T,h})} \text{ if } P\left(y_t y_{t-h} > 0\right) = 1 \tag{16}$$

$$\hat{\mathcal{R}}_{T,h} = \frac{T-h}{T}\frac{T-h}{T-h-k_{T,h}}\left(\frac{\hat{d}_{h,2,m_{T,h}}\left(\widehat{\mathcal{Y}}^{(0)}_{h,(k_{T,h})}\right)^{1-\hat{\kappa}_{h,2,m_{T,h}}}}{\hat{\kappa}_{h,2,m_{T,h}} - 1} - \frac{\hat{d}_{h,1,m_{T,h}}\left(\widehat{\mathcal{Y}}^{(0)}_{h,(k_{T,h})}\right)^{1-\hat{\kappa}_{h,1,m_{T,h}}}}{\hat{\kappa}_{h,1,m_{T,h}} - 1}\right) \text{ else} \tag{17}$$

Finally, if the tail indices are known to be equivalent $\kappa_{h,0} = \kappa_{h,1} = \kappa_{h,2}$, then we can replace tail-specific $\hat{\kappa}_{h,i,m_{T,h}}$ with the two-tailed $\hat{\kappa}_{h,0,m_{T,h}}$, since the latter is computed from a larger sample and therefore will be sharper with higher probability. We then have:

$$\hat{\mathcal{R}}_{T,h} = \frac{T-h}{T}\frac{T-h}{T-h-k_{T,h}}\left(\frac{\hat{d}_{h,2,m_{T,h}} - \hat{d}_{h,1,m_{T,h}}}{\hat{\kappa}_{h,0,m_{T,h}} - 1}\right)\left(\widehat{\mathcal{Y}}^{(0)}_{h,(k_{T,h})}\right)^{1-\hat{\kappa}_{h,0,m_{T,h}}} \text{ if } P\left(y_t y_{t-h} > 0\right) < 1. \tag{18}$$

### 3.2 Optimal Bias Correction

As discussed in Hill (2013), a shortcoming of a bias-corrected estimator like $\widehat{\hat{\gamma}}_{T,h}^{(bc)}(\widehat{\mathfrak{Y}}_{h,(k_{T,h})}^{(0)})$ is its reliance on estimates $\hat{\kappa}_{h,i,m_{T,h}}$ and $\hat{d}_{h,i,m_{T,h}}$ based on one fractile $m_{T,h}$, while a variety of $m'_{T,h}s$ can be used. Further, $\hat{\mathcal{R}}_{T,h}$ is well defined only when $\hat{\kappa}_{h,i,m_{T,h}} > 1$, and it seems desirable to choose $m_{T,h}$ such that $\widehat{\hat{\gamma}}_{T,h}^{(bc)}(\widehat{\mathfrak{Y}}_{h,(k_{T,h})}^{(0)})$ is close to the untrimmed estimator, e.g. $1/T\sum_{t=h+1}^{T} y_t y_{t-h}$.

A simple technique for generating a window of fractiles is to pick some $0 < \underline{\xi} < \bar{\xi} < \infty$ and generate a fractile function:

$$m_{T,h}(\xi) = [\xi m_{T,h}] \text{ where } \xi \in \Upsilon = [\underline{\xi}, \bar{\xi}].$$

Let $\hat{\mathcal{R}}_{T,h}(\xi)$ be the bias estimator computed with $m_{T,h}(\xi)$. The new bias corrected estimator is

$$\widehat{\hat{\gamma}}_{T,h}^{**}(\widehat{\mathfrak{Y}}_{h,(k_{T,h})}^{(0)}) \equiv \widehat{\hat{\gamma}}_{T,h}^{*}(\widehat{\mathfrak{Y}}_{h,(k_{T,h})}^{(0)}) + \hat{\mathcal{R}}_{T,h}(\hat{\xi}_{T,h}) \tag{19}$$

where $\hat{\xi}_{T,h}$ optimally places $\widehat{\hat{\gamma}}_{T,h}^{**}(\widehat{\mathfrak{Y}}_{h,(k_{T,h})}^{(0)})$ near $1/T\sum_{t=h+1}^{T} y_t y_{t-h}$ such that $\hat{\kappa}_{h,i,m_{T,h}} > 1$:

$$\hat{\xi}_{T,h} = \underset{\xi\in\Upsilon_h^*}{\arg\min} \left| \widehat{\hat{\gamma}}_{T,h}^{*}(\widehat{\mathfrak{Y}}_{h,(k_{T,h})}^{(0)}) + \hat{\mathcal{R}}_{T,h}(\xi) - \frac{1}{T}\sum_{t=h+1}^{T} y_t y_{t-h} \right| \tag{20}$$

$$\text{where } \Upsilon_0^* \equiv \left\{\xi \in \Upsilon : \hat{\kappa}_{0,0,m_{T,0}(\xi)} > 1\right\} \quad \text{and} \quad \Upsilon_h^* \equiv \left\{\xi \in \Upsilon : \hat{\kappa}_{h,i,m_{T,h}(\xi)} > 1 \text{ for } i = 1,2\right\}. \tag{21}$$

The set of minimizing $\hat{\xi}_{T,h}$ may contain more than one value, in which case $\hat{\xi}_{T,h}$ is one such element. This is irrelevant since $\hat{\kappa}_{h,i,m_{T,h}(\xi)}$ and $\hat{d}_{h,i,m_{T,h}(\xi)}$ will not affect our bias estimator uniformly on $\Upsilon$, asymptotically with probability one, as long as $m_{T,h}/k_{T,h} \to \infty$. Conversely, if $\hat{\kappa}_{h,i,m_{T,h}(\xi)} \leq 1$ on $\Upsilon$ then a solution does not exist. This may occur when the assumption $E[y_t^2] < \infty$ fails, or $E[y_t^2] < \infty$ holds but $h$ is large enough that there are too few tail observations of $\{y_t y_{t-h}\}_{t=h+1}^{T}$ from which to obtain a good estimate of $\kappa_{h,i}$.[4] We find in simulation experiments that if $T = 100$ then $h > 80$ can lead to highly volatile estimates $\hat{\kappa}_{h,i,m_{T,h}(\xi)}$, in particular $\hat{\kappa}_{h,i,m_{T,h}(\xi)} \leq 1$ over a large window of $\xi$ when $\kappa_{h,i}$ is close to, but larger than, 1.

In practice, sampling error can render $\widehat{\hat{\gamma}}_{T,h}^{**}(\widehat{\mathfrak{Y}}_{h,(k_{T,h})}^{(0)})$ farther from $1/T\sum_{t=h+1}^{T} y_t y_{t-h}$ than the uncorrected estimator $\widehat{\hat{\gamma}}_{T,h}^{*}(\widehat{\mathfrak{Y}}_{h,(k_{T,h})}^{(0)})$, especially when $\widehat{\hat{\gamma}}_{T,h}^{*}(\widehat{\mathfrak{Y}}_{h,(k_{T,h})}^{(0)})$ already has negligible small sample bias (e.g. $y_t y_{t-h}$ has a symmetric distribution). We suggest using whichever estimator, $\widehat{\hat{\gamma}}_{T,h}^{*}(\widehat{\mathfrak{Y}}_{h,(k_{T,h})}^{(0)})$ or $\widehat{\hat{\gamma}}_{T,h}^{**}(\widehat{\mathfrak{Y}}_{h,(k_{T,h})}^{(0)})$, is closest to $1/T\sum_{t=h+1}^{T} y_t y_{t-h}$:

$$\begin{aligned} &\widehat{\hat{\gamma}}_{T,h}^{(obc)}(\widehat{\mathfrak{Y}}_{h,(k_{T,h})}^{(0)}) \\ &\quad = \widehat{\hat{\gamma}}_{T,h}^{**}(\widehat{\mathfrak{Y}}_{h,(k_{T,h})}^{(0)}) \times I\left(\left|\widehat{\hat{\gamma}}_{T,h}^{**}(\widehat{\mathfrak{Y}}_{h,(k_{T,h})}^{(0)}) - \frac{1}{T}\sum_{t=h+1}^{T} y_t y_{t-h}\right| < \left|\widehat{\hat{\gamma}}_{T,h}^{*}(\widehat{\mathfrak{Y}}_{h,(k_{T,h})}^{(0)}) - \frac{1}{T}\sum_{t=h+1}^{T} y_t y_{t-h}\right|\right) \end{aligned} \tag{22}$$

[4] The problem of tail inference based on point estimates like $\hat{\kappa}_{h,i,m_{T,h}}$ is well documented. This accounts for the use of so-called window plots, or Hill-plots over a window of $m'_{T,h}s$, e.g. Drees, de Haan, and Resnick (2000).

$$+\hat{\tilde{\gamma}}^{*}_{T,h}(\widehat{\mathcal{Y}}^{(0)}_{h,(k_{T,h})}) \times I\left(\left|\hat{\tilde{\gamma}}^{**}_{T,h}(\widehat{\mathcal{Y}}^{(0)}_{h,(k_{T,h})}) - \frac{1}{T}\sum_{t=h+1}^{T} y_t y_{t-h}\right| \geq \left|\hat{\tilde{\gamma}}^{*}_{T,h}(\widehat{\mathcal{Y}}^{(0)}_{h,(k_{T,h})}) - \frac{1}{T}\sum_{t=h+1}^{T} y_t y_{t-h}\right|\right).$$

This is a merely small sample correction since, e.g., $\hat{\tilde{\gamma}}^{(obc)}_{T,0}(\widehat{\mathcal{Y}}^{(0)}_{0,(k_{T,h})}) = \hat{\tilde{\gamma}}^{**}_{T,0}(\widehat{\mathcal{Y}}^{(0)}_{0,(k_{T,h})})$ as $T \to \infty$ with probability approaching one due to negative bias. The optimal bias-corrected FD-QML estimator is

$$\hat{\theta}^{(obc)}_{T} = \arg\min_{\theta\in\Theta} \sum_{j\in\mathbb{F}} \left(\ln f(\lambda_j, \theta) + \frac{\widehat{\mathcal{I}}^{(obc)}_{T}(\lambda_j)}{f(\lambda_j, \theta)}\right)$$

where

$$\widehat{\mathcal{I}}^{(obc)}_{T}(\lambda) = \frac{1}{2\pi}\left(\hat{\tilde{\gamma}}^{(obc)}_{T,0}(\widehat{\mathcal{Y}}^{(0)}_{0,(k_{T,h})}) + 2\sum_{h=1}^{b_T} \hat{\tilde{\gamma}}^{(obc)}_{T,h}(\widehat{\mathcal{Y}}^{(0)}_{h,(k_{T,h})}) \times \cos(\lambda h)\right). \tag{23}$$

As long as $\hat{\kappa}_{h,i,m_{T,h}}$ and $\hat{d}_{h,i,m_{T,h}}$ in $\hat{\mathcal{R}}_{T,h}$ are $m^{1/2}_{T,h}$-convergent, and we use more tail observations for tail index estimation in the sense $m_{T,h}/k_{T,h} \to \infty$, then $\hat{\kappa}_{h,i,m_{T,h}}$ and $\hat{d}_{h,i,m_{T,h}}$ do not affect the limiting distribution of $\hat{\tilde{\gamma}}^{**}_{T,h}(\widehat{\mathcal{Y}}^{(0)}_{h,(k_{T,h})})$. Indeed, since $m_{T,h}(\xi) = [\xi m_{T,h}]$ also satisfies $m_{T,h}(\xi)/k_{T,h} \to \infty$ $\forall \xi \in \Upsilon$, we can always choose $\xi$ from from any subset $\Upsilon^* \subseteq \Upsilon$ and still the tail index estimators will not impact asymptotics. See Hill (2013, Theorem 2.2) for discussion.

The following summarizes the required second order power law property, and bounds for $m_{T,h}$.

**Assumption B′ (second order power law and fractile rates).** *$P(y_t y_{t-h} - \tilde{\gamma}_h \leq -c) = d_{h,1}c^{-\kappa_{h,1}}(1 + O(r_1(c))$ and $P(y_t y_{t-h} - \tilde{\gamma}_h \geq c) = d_{h,2}c^{-\kappa_{h,2}}(1 + O(r_2(c)))$, where $d_{h,i} > 0$, $\kappa_{h,i} > 1$, and $r_i$ are measurable functions. Let $e_{h,i} > 0$, $e_{h,0} \equiv \min\{e_{h,1}, e_{h,2}\}$ and $\kappa_{h,0} \equiv \min\{\kappa_{h,1}, \kappa_{h,2}\}$. Then $m_{T,h} \in \{1, ..., T-h\}$ and $m_{T,h} \to \infty$, and either $r_i(c) = c^{-e_{h,i}}$ and $m_{T,h} = o((T-h)^{2e_{h,0}/(2e_{h,0}+\kappa_{h,0})})$, or $r_i(c) = \ln(c)^{-e_{h,i}}$ and $m_{T,h} = o(\ln(T-h)^{2e_{h,0}})$. Finally, $m_{T,h}/k_{T,h} \to \infty$.*

**Remark 14** The two cases $r_i(c) = c^{-e_{h,i}}$ and $r_i(c) = \ln(c)^{-e_{h,i}}$ fall under second order properties SR1 and SR2 in Goldie and Smith (1987), while other SR1 and SR2 cases are possible here. See also Haeusler and Teugels (1985), Hsing (1991) and Hill (2010). It is well known that as tails deviate from a Pareto law, $m^{1/2}_{T,h}$-convergence of Hill (1975)'s estimator requires observations to come from farther out in the tails, hence $m_{T,h}$ must grow slower. See especially Haeusler and Teugels (1985, Section 5). Fractiles that satisfy Assumptions B′ and C for either case, any exponents $\{e_h, \kappa_{h,0}\}$ and any lag $h = 0, ..., b_T$, include $m_{T,h} = [\delta_{m,h}\ln(\ln(T))]$ and $k_{T,h} = [\delta_{k,h}(\ln(\ln(T)))^{1-\iota}]$, where $\iota > 0$ is tiny and $\delta_{m,h}, \delta_{k,h} > 0$.

Define

$$\mathfrak{I}^{*}_{T,h,t} \equiv I\left(|y_t y_{t-h} - \tilde{\gamma}_h| \geq c_{T,h}\right) - P\left(|y_t y_{t-h} - \tilde{\gamma}_h| \geq c_{T,h}\right) \text{ and } \mathfrak{I}^{*}_{T,h} \equiv \frac{1}{\kappa_{h,0}}\frac{1}{k^{1/2}_{T,h}}\sum_{t=1}^{T}\mathfrak{I}^{*}_{T,h,t}$$

$$\mathcal{D}_{T,0} \equiv \frac{1}{\kappa_{0,0}-1}\frac{k^{1/2}_{T,0}}{T-k_{T,0}}c_{T,0} \text{ and } \mathcal{D}_{T,h} \equiv \frac{1}{k^{1/2}_{T,h}}\left(d_{h,1}c^{1-\kappa_{h,1}}_{T,h} - d_{h,2}c^{1-\kappa_{h,2}}_{T,h}\right) \text{ for } h \neq 0$$

and

$$\tilde{\mathcal{Z}}_T \equiv T^{1/2}\frac{1}{2\pi}\left(\left\{\int_{-\pi}^{\pi}\left(\mathcal{I}_T^*(\lambda) - E\left[\mathcal{I}_T^*(\lambda)\right]\right)\varpi(\lambda)d\lambda\right\} + \frac{1}{2\pi}\sum_{h=-b_T}^{b_T}\varpi_h\mathcal{D}_{T,h}\mathfrak{I}_{T,h}^*\right)$$

$$\tilde{\mathcal{S}}_T \equiv E\left[\tilde{\mathcal{Z}}_T\tilde{\mathcal{Z}}_T'\right] \text{ and } \tilde{\mathcal{V}}_T = \Omega^{-1}\tilde{\mathcal{S}}_T\Omega^{-1}.$$

**Theorem 3.1** *Under Assumptions A, B′ and C,* $T^{1/2}\tilde{\mathcal{V}}_T^{-1/2}(\hat{\theta}_T^{(obc)} - \theta_0) \stackrel{d}{\rightarrow} N(0, I_k)$ *where* $\tilde{\mathcal{V}}_T = \mathcal{V}_T(1 + o(1))$ *when* $\kappa \geq 4$ *and* $\tilde{\mathcal{V}}_T = \mathcal{V}_T(1 + O(1))$ *when* $\kappa \in (2, 4)$.

**Remark 15** In order to estimate $\tilde{\mathcal{S}}_T$, a sample counterpart to $\sum_{h=-b_T}^{b_T}\varpi_h\mathcal{D}_{T,h}\mathfrak{I}_{T,h}^*$ is available by truncating the sum and using estimators of $c_{T,h}$ and $\kappa_{h,i}$. Conversely, a variety of subsampling techniques may be valid, although we do not present a formal theory here. See, for example, Kirch and Politis (2011) and Kreiss and Paparoditis (2012) for theory and references.

**Remark 16** The order statistic $\widehat{\mathcal{Y}}_{h,(k_{T,h})}^{(0)}$ is used both for trimming in $\widehat{\hat{\gamma}}_{T,h}^*(\widehat{\mathcal{Y}}_{h,(k_{T,h})}^{(0)})$ and for the bias estimator $\hat{\mathcal{R}}_{T,h}$, hence the FD-QML scale $\tilde{\mathcal{V}}_T$ is no longer $\mathcal{V}_T$ in the heavy tail case $\kappa < 4$. We show in Appendix A that, for asymptotic analysis, we can replace $\widehat{\hat{\gamma}}_{T,h}^*(\widehat{\mathcal{Y}}_{h,(k_{T,h})}^{(0)})$ with $\hat{\gamma}_{T,h}^*(c_{T,h})$ and write $\hat{\mathcal{R}}_{T,h} = \mathcal{R}_{T,h} + \mathcal{D}_{T,h} \times k_{T,h}^{1/2}(\widehat{\mathcal{Y}}_{h,(k_{T,h})}^{(0)}/c_{T,h} - 1) \times (1 + o_p(1))$, while also $k_{T,h}^{1/2}(\widehat{\mathcal{Y}}_{h,(k_{T,h})}^{(0)}/c_{T,h} - 1) = \mathfrak{I}_{T,h}^*(1 + o_p(1))$. Hence $\hat{\gamma}_{T,h}^*(c_{T,h}) = K_T\sum_{t=h+1}^{T} y_t y_{t-h} I(|y_t y_{t-h} - \tilde{\gamma}_h| < c_{T,h})$ for some function $K_T$ of $T$, and $\hat{\mathcal{R}}_{T,h} - \mathcal{R}_{T,h} = \mathcal{D}_{T,h} \times \mathfrak{I}_{T,h}^* \times (1 + o_p(1))$, are both partial sum functions of tail indicators $I(|y_t y_{t-h} - \tilde{\gamma}_h| < c_{T,h})$ and $I(|y_t y_{t-h} - \tilde{\gamma}_h| \geq c_{T,h})$. Since $I(|y_t y_{t-h} - \tilde{\gamma}_h| \geq c_{T,h})$ is degenerate as $T \rightarrow \infty$ with a relatively fast rate when $\kappa \geq 4$, and $y_t y_{t-h} I(|y_t y_{t-h} - \tilde{\gamma}_h| < c_{T,h})$ is nondegenerate and relatively thin tailed when $\kappa \geq 4$, it can be shown $\sum_{h=-b_T}^{b_T}\varpi_h\mathcal{D}_{T,h}\mathfrak{I}_{T,h}^*$ does not contribute to the variance of $\tilde{\mathcal{Z}}_T$ when $\kappa \geq 4$, thus $\tilde{\mathcal{S}}_T \sim \mathcal{S}_T$, cf. (cf. Hill, 2013, proof of Theorem 2.2). Finally, $\tilde{\mathcal{V}}_T = \mathcal{V}_T(1 + O(1))$ implies $\hat{\theta}_T^{(obc)}$ and $\hat{\theta}_T^*$ have the same convergence rates.

The next result mirrors Corollary 2.4 in the linear thin tail case. Here we impose second order power law Assumption B′ to safely handle tail exponent estimator asymptotics. See Section 3.3 for removal of the power law assumption.

**Corollary 3.2** *In addition to Assumptions A, B′, and C.2 let* $y_t$ *satisfy (10), let* $k_{T,h} \rightarrow \infty$ *and* $k_{T,h}/(T - h) = o(1)$, *and assume* $E[\epsilon_t|\Im_{t-1}] = 0$ *a.s.,* $E[\epsilon_t^2|\Im_{t-1}] = \sigma^2$ *a.s.,* $E[\epsilon_t^3|\Im_{t-1}] = s$ *a.s., and* $E[\epsilon_t^4] = \mathcal{K} < \infty$. *Then* $T^{1/2}(\hat{\theta}_T^{(obc)} - \theta_0) \stackrel{d}{\rightarrow} N(0, \mathcal{V})$ *where* $\mathcal{V} = \Omega^{-1}(2\Omega + \Pi)\Omega^{-1}$ *and* $\Omega$ *and* $\Pi$ *are defined in (11).*

### 3.3 Bias-Correction: Thin Tails

Bias $\mathcal{R}_{T,h}$ in (13) and (14) follows from Karamata theory under power law Assumption B. If tails decay faster than a power law function, for example exponentially fast, then Karamata's Theorem does not hold, and (13) and (14) are not valid. We now show how $\hat{\mathcal{R}}_{T,h}$ behave as $T$ increases for exponential tails:

$$P(|y_t y_{t-h} - \tilde{\gamma}_h| \geq c) = \vartheta_h \exp\{-\zeta_h c^{\delta_h}\} \text{ where } \vartheta_h, \zeta_h, \delta_h > 0. \tag{24}$$

The following extends to other thin tailed distributions with obvious changes to the derivations.

Dominated convergence and (24) imply we do not have an asymptotic bias problem if $T^{1/2}\hat{\mathcal{R}}_{T,h} \xrightarrow{p} 0$. Logically, the intermediate order statistic $\widehat{\mathcal{Y}}^{(0)}_{h,(k_{T,h})}$ in $\hat{\mathcal{R}}_{T,h}$ is well behaved, in particular $\widehat{\mathcal{Y}}^{(0)}_{h,(k_{T,h})}/c_{T,h} = 1 + O_p(1/k^{1/2}_{T,h})$ provided in the exponential case $k_{T,h}/\ln(T) \to 0$, while Hill (1975)'s estimator $\hat{\kappa}_{h,i,m_{T,h}}$ can be easily shown to diverge in probability. Finally, we only need $m_{T,h} \to \infty$ and $m_{T,h}/(T - h) \to 0$ because we no longer impose power law Assumption B$'$, and we no longer need $m_{T,h}/k_{T,h} \to \infty$ because tails decay so rapidly that the tail estimators in the bias terms do not impact asymptotics. We prove the next result in Hill and McCloskey (2014) since it is similar to Hill (2013, Theorem 2.3).

**Theorem 3.3** *Let Assumption A hold, assume (24) applies for each $h$, and let $k_{T,h} \to \infty$, $k_{T,h}/\ln(T) \to 0$, $m_{T,h} \to \infty$, and $m_{T,h}/(T - h) \to 0$. Then $T^{1/2}\hat{\mathcal{R}}_{T,h} \xrightarrow{p} 0$ hence $T^{1/2}\mathcal{V}_T^{-1/2}(\hat{\theta}_T^{(obc)} - \theta_0) \xrightarrow{d} N(0, I_k)$. Further, if the conditions of Corollary 3.2 hold then $T^{1/2}(\hat{\theta}_T^{(obc)} - \theta_0) \xrightarrow{d} N(0, \mathcal{V})$ where $\mathcal{V} = \Omega^{-1}(2\Omega + \Pi)\Omega^{-1}$ and $\Omega$ and $\Pi$ are defined in (11).*

# 4 Fractile Selection

We now discuss fractile selection from both theoretical and practical perspectives.

## 4.1 Mean-Squared-Error Minimization

In Section B of the supplemental material, Hill and McCloskey (2014), we characterize the minimum mean-squared-error [mse] trimming fractile $k_{T,h}$ for the special case where $y_t y_{t-h}$ has a symmetric distribution for $h \neq 0$. In this case FD-QML bias reduces to just $\mathcal{B}_T = \Omega^{-1}(2\pi)^{-1}\varpi_0\{E[y_t^2 I(y_t^2 < c_{T,0})] - E[y_t^2]\}$, which is easily analyzed. The general case for $\mathcal{B}_T$ is far more complicated, and possibly intractable, without a specific model and known distribution for $y_t y_{t-h}$.

By Theorem B.1 in Hill and McCloskey (2014), $k_{T,h} = a$ for some constant $a \geq 0$ minimizes mse as $T \to \infty$, where $a = 0$ if $\kappa = 4$. Bounded $k_{T,h}$ follows from the fact that bias dominates the variance in their contribution to the mse, and bias $E[y_t^2 I(y_t^2 < c_{T,0})] - E[y_t^2]$ is reduced with a smaller amount of trimming. Obviously a constant $k_{T,h}$ is not allowed here precisely to ensure asymptotic normality when tails are heavy. But this does clearly suggest $k_{T,h} \to \infty$ very slowly will keep the bias in $\hat{\theta}^*_{i,T}$ very small; and as a practical matter that we discuss below, it is easier to correct for small bias.

The conclusions of Theorem B.1 in Hill and McCloskey (2014) are qualitatively similar to those in Chaudhuri and Hill (2014) for higher order bias minimization of a bias corrected tail-trimmed mean. Chaudhuri and Hill (2014)'s bias representation exploits independence, and higher order bias is smaller when trimming is reduced and more tail observations are used for bias estimation. Bias in the tail-trimmed FD-QML estimator is merely a weighted average of biases from tail-trimmed cross-moments, and in the iid case higher order bias for each bias-corrected tail-trimmed cross-moment has the same structure as in Chaudhuri and Hill (2014). Bias derivations for dependent data will be similar under the Assumption A dependence properties.

### 4.2 Practical Concerns for Fractile Selection

In small samples $k_{T,h}$ plays three roles in estimator bias and bias correction. First, the trimmed second moment $\widehat{\hat{\gamma}}_{T,0}^{*}(\widehat{\mathcal{Y}}_{0,(k_{T,h})}^{(0)})$ and possibly cross-moments $\widehat{\hat{\gamma}}_{T,h}^{*}(\widehat{\mathcal{Y}}_{h,(k_{T,h})}^{(0)})$ for $h \geq 1$ are biased, where the bias in general is augmented for larger $k_{T,h}$ (cf. Chaudhuri and Hill, 2014). Second, $\widehat{\mathcal{I}}_{T}^{(obc)}(\lambda)$ uses sample moments $\widehat{\hat{\gamma}}_{T,h}^{*}(\widehat{\mathcal{Y}}_{h,(k_{T,h})}^{(0)})$ for only $h = 0, ..., b_T$, where $T - b_T \to \infty$ since robustness requires $k_{T,h} \in \{1, ..., T-h\}$ *and* $k_{T,h} \to \infty$ for all $h$. The restriction $k_{T,b_T} \in \{1, ..., T - b_T\}$ implies in general $b_T$ must be small to accomodate a large $k_{T,h}$ for each $h$, while small $b_T < T - 1$ can distort FD-QML estimation. In both cases, a large $k_{T,h}$ can generate substantial bias in $\hat{\theta}_T^{*}$. Third, the trimmed portion $|E[y_t y_{t-h} I(|y_t y_{t-h} - \tilde{\gamma}_h| < c_{T,h})] - E[y_t y_{t-h}]|$ is larger when $k_{T,h}$ is large, provided bias exists, and a large gap is more difficult to approximate and therefore estimate using Karamata theory. Thus, when $k_{T,h}$ is large the non-bias corrected $\widehat{\hat{\gamma}}_{T,h}^{*}(\widehat{\mathcal{Y}}_{h,(k_{T,h})}^{(0)})$ *and* bias corrected $\widehat{\hat{\gamma}}_{T,h}^{**}(\widehat{\mathcal{Y}}_{h,(k_{T,h})}^{(0)})$ can be poor approximations to $E[y_t y_{t-h}]$, rendering $\hat{\theta}_T^{(obc)}$ less sharp. In general, then, a small and slowly increasing $k_{T,h}$ will improve small sample performance, and be valid based on Theorems 3.1 and 3.3. Moreover, recall that $k_{T,h} \to \infty$ at a slowly varying rate is required to ensure mean-centering does not impact asymptotics when $y_t$ is heavy tailed.

The conclusion of slow $k_{T,h} \to \infty$ is contrary to the finding of Theorem 2.2, that if $\kappa < 4$ then the rate of convergence is higher when $k_{T,h} \to \infty$ faster. This obviously does not concern small and large sample bias. In terms of inference on $\theta_0$, clearly small bias is preferred, hence slow $k_{T,h} \to \infty$.

## 5 Simulation Study

We now investigate the small sample properties of the standard and robust FD-QML estimators. We draw samples $\{y_t\}_{t=1}^{T}$ of size $T \in \{100, 250, 500, 1000\}$ from AR(1) and GARCH(1,1) models. In each case $10,000$ samples are of size $2T$ are drawn, and we retain the last $T$ for analysis. The AR(1) models are $y_t = \phi_0 y_{t-1} + \epsilon_t$ with $\phi_0 \in \{.00, .75, .90\}$, where $\epsilon_t$ is iid standard normal, or symmetric Pareto with distribution $P(\epsilon_t < -c) = P(\epsilon_t > c) = .5(1 + c)^{-\kappa}$ and tail index $\kappa \in \{2.25, 4.5\}$ standardized such that $\sigma_0^2 \equiv E[\epsilon_t^2] = 1$. In this setup, it is well known that $y_t$ belongs to the same distribution class as $\epsilon_t$ (e.g. Brockwell and Cline, 1985). We initialize the draw with $y_1 = \epsilon_1$. The spectrum is $f(\lambda, \theta_0) = (2\pi)^{-1}\sigma_0^2|1 - \phi_0 e^{-i\lambda}|^{-2}$ where $\theta_0 = [\phi_0, \sigma_0^2]'$, and we optimize the FD-QML criteria on $\Theta \equiv [-.999, .999] \times [0, 100]$.[5]

[5]We use Matlab R2014a, where optimization is performed by the *fmincon* routine.

The GARCH model is $x_t = \sigma_t \epsilon_t$ where $\sigma_t^2 = \omega_0 + \alpha_0 x_{t-1}^2 + \beta_0 \sigma_{t-1}^2$, $\omega_0 = 1$, and $[\alpha_0, \beta_0]$ is $[.3, .6]$, $[.2, .7]$ or $[.05, .9]$, and $\epsilon_t$ is iid standard normal. We initialize the draw with $\sigma_1^2 = \omega_0$. We use $y_t \equiv x_t^2$ which has an ARMA(1,1) representation $y_t = \omega_0 + (\alpha_0 + \beta_0) y_{t-1} - \beta_0 u_{t-1} + u_t$, where $u_t = \sigma_t^2(\epsilon_t^2 - 1)$ is a martingale difference, and $\sigma_0^2 \equiv E[u_t^2]$. Thus, the spectrum for $y_t$ is $f(\lambda, \theta_0) = (2\pi)^{-1}\sigma_0^2 |1 - \beta_0 e^{-i\lambda}|^2 \times |1 - (\alpha_0 + \beta_0) e^{-i\lambda}|^{-2}$ where $\theta_0 = [\alpha_0, \beta_0, \sigma_0^2]'$. Note that $P(|y_t| > c) \sim dc^{-\kappa_y}$ with tail index $\kappa_y \approx 2.05$ when $[\alpha_0, \beta_0] = [.3, .6]$, $\kappa_y \approx 3.02$ when $[\alpha_0, \beta_0] = [.2, .7]$, and $\kappa_y \approx 4.05$ when $[\alpha_0, \beta_0] = [.05, .9]$.[6] Define $\tilde{\Theta} \equiv [10^{-10}, 100] \times [0, .999] \times [0, .999]$. We optimize the FD-QML criteria on the subset $\Theta \equiv \{\theta \in \tilde{\Theta} : \alpha + \beta \leq .999\}$.

We use optimal bias correction since, by simulation experiments not reported here, we otherwise find that tail-trimmed FD-QML estimates are strongly negatively biased. In the AR case we use centering with $\tilde{\gamma}_{T,0} = 0$ and $\tilde{\gamma}_{T,h} = (T - h)^{-1} \sum_{t=h+1}^{T} y_t y_{t-h}$ at lags $h \geq 1$, and in the GARCH case $\tilde{\gamma}_{T,h} = 0 \; \forall h \geq 0$.

Let $\iota = 10^{-10}$. We use a trimming fractile $k_{T,h} = [.2(\ln(T))^{1-\iota}]$ for each lag $h = 0, ..., b_T$ with bandwidth $b_T = [T^{.95}] \in \{79, 190, 366, 708\}$, and bias estimation fractile $m_{T,h} = [8\ln(T)]$ and fractile function $m_{T,h}(\xi) = [\xi m_{T,h}]$ over $\xi \in [.1, 2]$. Using lags $b_T = [T^{\delta}]$ for $\delta > .95$ can lead to a breakdown in bias estimation for $T \leq 250$ and $h \approx b_T$, because the sample $\{y_t y_{t-h} - \tilde{\gamma}_{T,h}\}_{t=h+1}^{T}$ may not have a tail index estimate $\hat{\kappa}_{h,0,m_{T,h}} > 1$ for *any* $m_{T,h}(\xi)$ due to the sample's small size $T - h$. As long as second order power law Assumption B$'$ holds then $\{k_{T,h}, m_{T,h}\}$ is valid for asymptotically unbiased estimation in all cases by Theorems 3.1 and 3.3.

We compute the optimally bias-corrected $\widehat{\tilde{\gamma}}_{T,h}^{(obc)}(\widehat{\mathcal{Y}}_{h,(k_{T,h})}^{(0)})$ in (22) based on the potentially biased $\widehat{\tilde{\gamma}}_{T,h}^{*}(\widehat{\mathcal{Y}}_{h,(k_{T,h})}^{(0)})$, and the bias-corrected $\widehat{\tilde{\gamma}}_{T,h}^{**}(\widehat{\mathcal{Y}}_{h,(k_{T,h})}^{(0)}) = \widehat{\tilde{\gamma}}_{T,h}^{*}(\widehat{\mathcal{Y}}_{h,(k_{T,h})}^{(0)}) + \hat{\mathcal{R}}_{T,h}(\hat{\xi}_{T,h})$ in (19). In this study, $y_t y_{t-h}$ from the AR model have symmetric tails for $h > 0$, hence $\hat{\mathcal{R}}_{T,0}(\hat{\xi}_{T,0})$ is computed from (16) and all other $\hat{\mathcal{R}}_{T,h}(\hat{\xi}_{T,h})$ from (18). We compute all $\hat{\mathcal{R}}_{T,h}(\hat{\xi}_{T,h})$ from (16) for the GARCH model. The optimal tuning parameter $\hat{\xi}_{T,h}$ is computed as in (20) with the set restriction (21).

Trimming very few $y_t y_{t-h}$ promotes small trimmed moment bias that is easily corrected if many tail observations are used for bias estimation, and suffices to generate a robust FD-QML estimator that is approximately normally distributed. In this study, for example, for $T \in \{100, 250\}$ the computed fractiles are $k_{T,h} = \{1, 1\}$ and $m_{T,h} = \{37, 44\}$, with ranges $m_{T,h}(\xi) \in \{[4, 74], [4, 88]\}$. Trimming just one observation in samples of size $T \leq 250$ corrects for heavy tails, while having available substantially more observations for tail exponent estimation (e.g. $m_{T,h}(\xi) \in \{[4, 74]\}$) leads to sharp bias correction.

Let $\hat{\theta}_{T,i}^{(r)}$ denote the $r^{th}$ sample estimate of $\theta_{0,i}$ for any estimator. We report the simulation median, bias $1/R \sum_{i=1}^{R} \hat{\theta}_{T,i}^{(r)} - \theta_{0,i}$, and root-mean-squared-error $\{1/R \sum_{i=1}^{R} (\hat{\theta}_{T,i}^{(r)} - \theta_{0,i})^2\}^{1/2}$, where $R = 10,000$. We then compute the standardized variable $z_{T,i}^{(r)} \equiv (\hat{\theta}_{T,i}^{(r)} - \theta_{0,i})/s_{T,i}$ with simulation variance $s_{T,i}^2 \equiv 1/R \sum_{i=1}^{R} (\hat{\theta}_{T,i}^{(r)} - 1/R \sum_{j=1}^{R} \hat{\theta}_{j,T}^{(r)})^2$ and perform a Kolmogorov-Smirnov test of standard

[6] The GARCH process $\{x_t\}$ satisfies $P(|x_t| > c) = dc^{-\kappa_x}(1 + o(1))$ and $E|\alpha_0 \epsilon_t^2 + \beta_0|^{\kappa_x/2} = 1$ (cf. Basrak, Davis, and Mikosch, 2002), hence $P(|y_t| > c) \sim dc^{-\kappa_x/2}$. We draw $R = 100,000$ iid standard normal $\epsilon_t$ and compute $\hat{\kappa}_x \equiv \arg\min_{\kappa \in \mathcal{K}} |1/R \sum_{t=1}^{R} |\alpha_0 \epsilon_t^2 + \beta_0|^{\kappa/2} - 1|$, where $\mathcal{K} = \{.001, .002, ..., 10\}$. We repeat this 10,000 times and find the median and mean $\hat{\kappa}_x$ are roughly 4.1 when $[\alpha_0, \beta_0] = [.3, .6]$, roughly 6.04 when $[\alpha_0, \beta_0] = [.2, .7]$, and roughly 8.1 when $[\alpha_0, \beta_0] = [.05, .9]$.

normality on the simulation sample $\{z_{T,i}^{(r)}\}_{r=1}^{R}$. We report the KS statistic divided by its 5% critical value: values greater than unity are evidence against normality.

First, we demonstrate the accuracy of the optimally bias corrected tail-trimmed covariances for an AR(1) with $\phi_0 = .9$, and for the GARCH models, in both cases with $T = 100$. In Figures 1 and 2 we plot the trimmed and untrimmed ratios $\widehat{\hat{\gamma}}_{T,h}^{(obc)}/\widehat{\hat{\gamma}}_{T,0}^{(obc)}$ and $\hat{\gamma}_{T,h}/\hat{\gamma}_{T,0}$, where $\hat{\gamma}_{T,h} \equiv 1/T\sum_{t=h+1}^{T} y_t y_{t-h}$, over lags $h = 1, ..., 80$. The plots are the 2.5%, 50% and 97.5% quantiles of $\widehat{\hat{\gamma}}_{T,h}^{(obc)}/\widehat{\hat{\gamma}}_{T,0}^{(obc)}$, and the average difference $\widehat{\hat{\gamma}}_{T,h}^{(obc)}/\widehat{\hat{\gamma}}_{T,0}^{(obc)} - \hat{\gamma}_{T,h}/\hat{\gamma}_{T,0}$, over the 10,000 samples. Clearly $\widehat{\hat{\gamma}}_{T,h}^{(obc)}/\widehat{\hat{\gamma}}_{T,0}^{(obc)}$ matches $\hat{\gamma}_{T,h}/\hat{\gamma}_{T,0}$ exceptionally well, which demonstrates the sharpness of the optimally fitted bias estimator. Indeed, we plot $\widehat{\hat{\gamma}}_{T,h}^{(obc)}/\widehat{\hat{\gamma}}_{T,0}^{(obc)}$ and the difference separately because $\widehat{\hat{\gamma}}_{T,h}^{(obc)}/\widehat{\hat{\gamma}}_{T,0}^{(obc)}$ and $\hat{\gamma}_{T,h}/\hat{\gamma}_{T,0}$ cannot be visually distinguished apart for most $h$. The difference, however, is larger when tails are heavier.

Second, FD-QML estimates for the AR case $[\phi_0, \sigma_0^2] = [.90, 1.00]$ are presented in Table 1. Since estimation results for $\phi_0 \in \{.00, .75\}$ are qualitatively similar, they are presented in the supplemental material Hill and McCloskey (2014, Tables E.1-E.2). The untrimmed estimator works well, but in most cases it is more biased and "less normal" than the trimmed estimator. In particular, the untrimmed estimator is non-normal in heavy tailed cases, while the trimmed estimator of $\phi_0$ is closer to normal for even small $T$, and roughly normal when $T \geq 250$. The robust and non-robust estimators of $\sigma_0^2$ tend to be farther from normal than the estimators of $\phi_0$, especially in the heavy tailed case $\kappa = 2.25$, but the trimmed estimator is generally closer to normal, and has smaller bias in most cases.

Third, estimates for the GARCH model are presented in Table 2. It is known that time domain non-robust M-estimators for GARCH parameters, like QML, appear biased in small samples, while tail-trimming reduces the negative impact of larger errors on bias, and improves approximate normality. See Hill (2014a) for evidence, and references. The same essential patterns arise in the frequency domain: tail-trimmed FD-QML has smaller bias and is closer to normal in most cases when $T \geq 250$. In general, samples sizes $T \geq 2000$ are required to achieve the bias and KS statistic values that we obtain in AR model estimation. See Hill and McCloskey (2014, Table E.3) for the case $T = 2000$.

# 6 Conclusion

We extend recent advances in the literature on heavy tail estimation to the frequency domain. We deliver asymptotically normal and unbiased estimators by negligibly transforming the data, and using optimally fitted bias correction for transformed sample cross-moments. A simulation study reveals that the tail-trimmed FD-QML estimator works well compared to the non-robust conventional estimator: our estimator is closer to normal and in most cases exhibits smaller bias and dispersion; while our robust moment estimators are quite sharp, showing that the optimally fitted bias estimator developed in Hill (2013) works very well.

In principle our robust Whittle estimator is asymptotically normal even when the stationary sequence is not covariance stationary, and therefore does not have a spectral density. This possibility will allow for an extension of our methods to arbitrarily heavy tailed data as in Mikosch, Gadrich, Kluppelberg, and Adler (1995) for ARMA processes, and suggests a useful future line of research. Finally, a higher order bias representation for the tail-trimmed moments computed in this paper will be useful for determining a trimming policy. Such expansions, however, are tricky under data dependence, and for FD-QML bias involves many tail-trimmed cross-moments with a complicated structure. This, too, is beyond the scope of this paper.

# A Appendix: Proofs of Main Results

The proofs of Theorems 2.1, 2.2 and 3.1 exploit several preliminary results. We first present asymptotic theory for tail trimmed and tail event processes. These are used to then prove lemmas that directly support the proofs of the main results. Finally, we prove the main theorems.

In order to avoid confusion, it is understood that $h \geq 0$ unless otherwise noted. In preliminary lemmata we reduce the number of required cases by mean-centering $I(|y_t y_{t-h} - (T-h)^{-1}\sum_{t=h+1}^{T} y_t y_{t-h}| < c)$ for *all* lags $h$. Thus:

$$\widehat{\mathcal{Y}}_{h,t}^{(0)} \equiv \left| y_t y_{t-h} - \frac{1}{T-h}\sum_{t=h+1}^{T} y_t y_{t-h} \right| \text{ for } h = 0, ..., b_T, \text{ and } \widehat{\mathcal{Y}}_{h,(1)}^{(0)} \geq \widehat{\mathcal{Y}}_{h,(2)}^{(0)} \geq \cdots \widehat{\mathcal{Y}}_{h,(T-h)}^{(0)}$$

$$\widehat{\hat{\gamma}}_{T,h}^{*}(c) \equiv \frac{T}{T-k_T}\frac{1}{T}\sum_{t=h+1}^{T} y_t y_{t-h} I\left( \left| y_t y_{t-h} - \frac{1}{T-h}\sum_{t=h+1}^{T} y_t y_{t-h} \right| < c \right)$$

$$\hat{\gamma}_{T,h}^{*}(c) \equiv \frac{T}{T-k_T}\frac{1}{T}\sum_{t=h+1}^{T} y_t y_{t-h} I\left( \left| y_t y_{t-h} - E\left[ y_t y_{t-h} \right] \right| < c \right)$$

This does not reduce generality since $\hat{\gamma}_{T,h}^{*}(c_{T,h})$ is still biased at lag $h = 0$, and unbiased at lag $h > 0$ when $y_t y_{t-h}$ has a symmetric distribution.

Further, since most proof details do not rely on $k_{T,h}$ being a function of $h$, unless otherwise stated:

$$k_{T,h} = k_T.$$

Since $k_{T,h} \in \{1, ..., T-h\}$, the restriction $k_{T,h} = k_T$ is synonymous to imposing $k_{T,h} = k_T \in \{1, ..., T-b_T\}$.

## A.1 Asymptotic Theory for Tail and Tail-Trimmed Processes

We repeatedly use the following central limit theorem for an arbitrary triangular array $\{g_{T,t} : 1 \leq t \leq T\}_{T \geq 1}$. Write $\mathcal{F}_{T,s}^{t} \equiv \sigma(g_{T,t} : s \leq \tau \leq t)$, and define mixing coefficients $\tilde{\alpha}_h \equiv \sup_{T \in \mathbb{N}} \sup_{1 \leq t \leq T} \sup_{\mathcal{A} \subset \mathcal{F}_{T,-\infty}^{t-h}, \mathcal{B} \subset \mathcal{F}_{T,t}^{\infty}} |P(\mathcal{A} \cap \mathcal{B}) - P(\mathcal{A})P(\mathcal{B})|$ and *interlaced maximal correlation coefficients* $\tilde{\rho}_h^{*} \equiv \sup_{T \in \mathbb{N}} \sup_{\mathfrak{S}_{T,h}, \mathfrak{T}_{T,h}} \rho(\sigma(g_{T,t} : t \in \mathfrak{S}_h), \sigma(g_{T,s} : s \in \mathfrak{T}_h))$ where $\mathfrak{S}_{T,h}$ and $\mathfrak{T}_{T,h}$ are non-empty subsets of $\{1, ..., T\}$ with $\inf_{s \in \mathfrak{S}_{T,h}, t \in \mathfrak{T}_{T,h}} \{|s-t|\} \geq h$. By convention $\sup_{\mathfrak{S}_{T,h}, \mathfrak{T}_{T,h}} \rho(\sigma(g_{T,t} : t \in \mathfrak{S}_h), \sigma(g_{T,s} : s \in \mathfrak{T}_h)) = 0$ if $h \geq T$.

**Lemma A.1** *Let Assumption A hold. Let* $\{g_{T,t} : 1 \leq t \leq T\}_{T \geq 1}$ *be a* $\sigma(y_t, y_{t-1}, ..., y_{t-i})$*-measurable triangular array for some finite* $i$*, where* $E[g_{T,t}] = 0$ *and* $E[g_{T,t}^2] < \infty$ *for each* $1 \leq t \leq T$ *and* $T \geq 1$*. Then* (*a*) $\upsilon_T^2 \equiv E(\sum_{t=1}^T g_{T,t})^2 \sim K \sum_{t=1}^T E[g_{T,t}^2]$*; and* (*b*) $1/\upsilon_T \sum_{t=1}^T g_{T,t} \stackrel{d}{\rightarrow} N(0,1)$ *provided the Lindeberg condition holds:*

$$\frac{1}{\upsilon_T^2} \sum_{t=1}^T E\left[g_{T,t}^2 I\left(|g_{T,t}| > \upsilon_T \times \epsilon\right)\right] \rightarrow 0 \text{ for every } \epsilon > 0. \tag{A.1}$$

**Proof.** For ($a$), by Assumption A.1 and measurability we have $\tilde{\alpha}_h \rightarrow 0$ and $\tilde{\rho}_1^* < 1$, hence $\upsilon_T^2 \sim K \sum_{t=1}^T E[g_{T,t}^2]$ by an application of Lemma 1 in Bradley (1992). For ($b$) note $\tilde{\alpha}_h \rightarrow 0$, $\upsilon_T^2 \sim K \sum_{t=1}^T E[g_{T,t}^2]$ and (A.1) imply $1/\upsilon_T \sum_{t=1}^T g_{T,t} \stackrel{d}{\rightarrow} N(0,1)$ by Theorem 2.2 in Peligrad (1996). $\mathcal{QED}$.

The next result allows us to use tail-trimmed covariance estimators with non-random thresholds.

**Lemma A.2** *Let Assumptions A and C hold, and let* $h \in \{0, ..., b_T\}$*. Then* (*a*) $|\hat{\gamma}_{T,h}^*(c_{T,h}) - E[\hat{\gamma}_{T,h}^*(c_{T,h})]| \stackrel{p}{\rightarrow} 0$*; and* (*b*) *if additionally Assumption B holds then* $(T/||\mathcal{S}_T||)^{1/2} \times |\widehat{\hat{\gamma}}_{T,h}^*(\widehat{\mathcal{Y}}_{h,(k_T)}^{(0)}) - \hat{\gamma}_{T,h}^*(c_{T,h})| \stackrel{p}{\rightarrow} 0$.

**Proof.**

**Claim (a):** Define $\Im_t \equiv \sigma(y_\tau : \tau \leq t)$ and $\psi_{h,t}(c) \equiv y_t y_{t-h} I(|y_t y_{t-h} - E[y_t y_{t-h}]| < c)$, hence $\hat{\gamma}_{T,h}^*(c) \equiv (T - h)(T - h - k_{T,h})^{-1} T^{-1} \sum_{t=h+1}^T \psi_{h,t}(c)$. Under Assumptions A and C.1 $\psi_{h,t}(c_{T,h})$ is uniformly $L_{1+\iota}$-bounded for tiny $\iota > 0$, and $\alpha$-mixing, hence $\{\psi_{h,t}(c_{T,h}), \Im_t\}$ forms an $L_1$-mixingale array (McLeish, 1975, Lemma 2.1). Since $h \leq b_T$ and $T - b_T \rightarrow \infty$ by Assumption C.2, it follows $1/T \sum_{t=h+1}^T \{\psi_{h,t}(c_{T,h}) - E[\psi_{h,t}(c_{T,h})]\} \stackrel{p}{\rightarrow} 0$ by Theorem 2 of Andrews (1988). The claim follows instantly.

**Claim (b):** Write $\mathfrak{S}_T^2 \equiv E[\psi_{0,t}^2(c_{T,0})]$, $\tilde{\gamma}_{T,h} \equiv (T - h)^{-1} \sum_{t=h+1}^T y_t y_{t-h}$ and $\tilde{\gamma}_h \equiv E[y_t y_{t-h}]$. By Lemma A.9.b $||\mathcal{S}_T|| \sim K \mathfrak{S}_T^2$, hence we need only prove

$$\mathcal{A}_T \equiv \frac{1}{T^{1/2} \mathfrak{S}_T} \sum_{t=h+1}^T y_t y_{t-h} \left\{ I(|y_t y_{t-h} - \tilde{\gamma}_{T,h}| < \widehat{\mathcal{Y}}_{h,(k_T)}^{(0)}) - I\left(|y_t y_{t-h} - \tilde{\gamma}_h| < c_{T,h}\right) \right\} = o_p(1).$$

We have

$$\begin{aligned} \mathcal{A}_T = &\frac{1}{T^{1/2} \mathfrak{S}_T} \sum_{t=h+1}^T y_t y_{t-h} \left\{ I\left(|y_t y_{t-h} - \tilde{\gamma}_{T,h}| - c_{T,h} < 0\right) - I\left(|y_t y_{t-h} - \tilde{\gamma}_h| - c_{T,h} < 0\right) \right\} \\ &+ \frac{1}{T^{1/2} \mathfrak{S}_T} \sum_{t=h+1}^T y_t y_{t-h} \left\{ I\left(|y_t y_{t-h} - \tilde{\gamma}_{T,h}| - \widehat{\mathcal{Y}}_{h,(k_T)}^{(0)} < 0\right) - I\left(|y_t y_{t-h} - \tilde{\gamma}_{T,h}| - c_{T,h} < 0\right) \right\} \\ &= \mathfrak{B}_T + \mathfrak{C}_T. \end{aligned}$$

Consider $\mathfrak{B}_T$. The trimming function $I(u) \equiv I(u < 0)$ can be approximated by a continuous, differentiable, uniformly bounded positive function $\mathcal{J}_{\mathcal{N}}(u)$ that has a uniformly bounded derivative $\mathcal{D}_{\mathcal{N}}(u)$

$\equiv (\partial/\partial u)\mathcal{J}_{\mathcal{N}}(u)$, where $\mathcal{N} \in \mathbb{N}$ guides the approximation: $\lim_{\mathcal{N}\rightarrow\infty}\sup_{u\in\mathbb{R}}|\mathcal{J}_{\mathcal{N}}(u) - I(u)| = 0$ and $\lim_{\mathcal{N}\rightarrow\infty}\sup_{u\in\mathcal{U}_0}|\mathcal{D}_{\mathcal{N}}(u)| = 0$ for any compact subsets $\mathcal{U} \subset \mathbb{R}$ and $\mathcal{U}_0 \subset \mathbb{R}/0$. See Hill (2012b, proof of Lemma A.1) and Hill (2013, proof of Lemma A.2), cf. Lighthill (1958, p. 22). In particular, we can always set $\mathcal{N} = \mathcal{N}_T \rightarrow \infty$ as $T \rightarrow \infty$ fast enough to ensure

$$\mathfrak{B}_T = \frac{1}{T^{1/2}\mathfrak{S}_T}\sum_{t=h+1}^{T} y_t y_{t-h}\left\{\mathcal{J}_{\mathcal{N}_T}\left((y_t y_{t-h} - \tilde{\gamma}_{T,h})^2 - c_{T,h}^2\right) - \mathcal{J}_{\mathcal{N}_T}\left((y_t y_{t-h} - \tilde{\gamma}_h)^2 - c_{T,h}^2\right)\right\} + o_p(1).$$

Hence, by the mean-value-theorem, for some $\mathring{\gamma}_{T,h}$, $|\mathring{\gamma}_{T,h} - \tilde{\gamma}_h| \leq |\tilde{\gamma}_{T,h} - \tilde{\gamma}_h|$:

$$\mathfrak{B}_T = -2\frac{1}{T^{1/2}\mathfrak{S}_T}\sum_{t=h+1}^{T} y_t y_{t-h} \times \mathcal{D}_{\mathcal{N}_T}\left((y_t y_{t-h} - \mathring{\gamma}_{T,h})^2 - c_{T,h}^2\right) \times (y_t y_{t-h} - \tilde{\gamma}_{T,h})(\tilde{\gamma}_{T,h} - \tilde{\gamma}_h) + o_p(1).$$

Mixing Assumption A.1 implies mixing in the ergodic sense, and therefore ergodicity (see, e.g., Petersen, 1983). Stationarity, ergodicity, $L_p$-boundedness for $p > 2$, and $h \leq b_T$ with $T - b_T \rightarrow \infty$ imply $\tilde{\gamma}_{T,h} \overset{p}{\rightarrow} \tilde{\gamma}_h = E[y_t y_{t-h}]$, hence $\mathring{\gamma}_{T,h} \overset{p}{\rightarrow} E[y_t y_{t-h}]$. Further, $\max_{1\leq t\leq T}|y_t| = O_p(T^{1/2})$ by stationarity and square integrability, and $|y_t y_{t-h} - \mathring{\gamma}_{T,h}| \neq c_{T,h}$ *a.s.* by distribution continuity. Thus, since $\max_{1\leq t\leq T}|\mathcal{D}_{\mathcal{N}_T}((y_t y_{t-h} - \mathring{\gamma}_{T,h})^2 - c_{T,h}^2)| \overset{p}{\rightarrow} 0$ as fast as we choose by setting $\mathcal{N}_T \rightarrow \infty$ as $T \rightarrow \infty$ fast enough, it follows $\mathfrak{B}_T \overset{p}{\rightarrow} 0$.

The same type of argument applies to $\mathfrak{C}_T$ where we use $\tilde{\gamma}_{T,h} \overset{p}{\rightarrow} E[y_t y_{t-h}]$, and $\widehat{\mathcal{Y}}_{h,(k_T)}^{(0)}/c_{T,h} = 1 + O_p(1/k_T^{1/2})$ by Lemma A.3. $\mathcal{QED}$.

Finally, we characterize order statistics and tail exponent estimators.

**Lemma A.3 (order statistic)** *Define* $\mathfrak{I}_{T,h,t}^{*} \equiv I(|y_t y_{t-h} - \tilde{\gamma}_h| \geq c_{T,h}) - P(|y_t y_{t-h} - \tilde{\gamma}_h| \geq c_{T,h})$. *Under Assumptions A.1, B and C* $k_{T,h}^{1/2}(\widehat{\mathcal{Y}}_{h,(k_T)}^{(0)}/c_{T,h} - 1) = \kappa_{h,0}^{-1}k_{T,h}^{-1/2}\sum_{t=1}^{T}\mathfrak{I}_{T,h,t}^{*} = O_p(1/k_T^{1/2})$ $\forall h \in \{0, ..., b_T\}$.

**Proof.** Define $\mathcal{Y}_{h,t}^{(0)} \equiv |y_t y_{t-h} - \tilde{\gamma}_h|$. We show that $k_{T,h}^{1/2}(\widehat{\mathcal{Y}}_{h,(k_T)}^{(0)}/c_{T,h} - 1)$ has the same limit distribution as $k_{T,h}^{1/2}(\mathcal{Y}_{h,(k_T)}^{(0)}/c_{T,h} - 1)$. The claim then follows by applying Lemma A.3 in Hill (2013) to $k_{T,h}^{1/2}(\mathcal{Y}_{h,(k_T)}^{(0)}/c_{T,h} - 1)$.

Observe that $E|y_t y_{t-h}|^{1+\iota} < \infty$ for some $\iota > 0$ by Assumption A, and the Assumption A.1 $\alpha$-mixing property implies $y_t y_{t-h}$ is an adapted $L_{1+\iota}$-mixingale. Further, $h \leq b_T$ with $T - b_T \rightarrow \infty$ under Assumption C.2. Hence $(T - h)^{-1}\sum_{t=h+1}^{T} y_t y_{t-h} = E[y_t y_{t-h}] + O_p(1/T^{\iota})$ since $E|\sum_{t=h+1}^{T} y_t y_{t-h}|^{1+\iota}$ $= O(T - h)$ by stationarity, $L_{1+\iota}$-boundedness, and an application of Lemma 2 in Hansen (1991, 1992). Since $k_T$ is slowly varying by mean-centering and Assumption C.1, it follows $k_T = o(T^a)$ $\forall a$ $> 0$, hence $(T - h)^{-1}\sum_{t=h+1}^{T} y_t y_{t-h} = E[y_t y_{t-h}] + o_p(1/k_T^{1/2})$. Therefore $k_{T,h}^{1/2}(\widehat{\mathcal{Y}}_{h,(k_T)}^{(0)}/c_{T,h} - 1)$ and $k_{T,h}^{1/2}(\mathcal{Y}_{h,(k_T)}^{(0)}/c_{T,h} - 1)$ have the same limit distribution in view of mixing Assumption A.1, tail decay Assumption B, and arguments presented in Hill (2010, Lemma 3) and Hill (2014b, proof of Theorem

2.1, Lemma 2.2).[7] $\mathcal{QED}$.

**Lemma A.4 (tail exponents)** *Define* $\hat{\kappa}^{-1}_{h,m_{T,h}} \equiv 1/m_{T,h}\sum_{j=1}^{m_{T,h}}\ln(\widehat{\mathcal{Y}}^{(0)}_{h,(j)}/\widehat{\mathcal{Y}}^{(0)}_{h,(m_{T,h}+1)})$ *and* $\hat{d}_{h,m_{T,h}} \equiv (m_{T,h}/(T-h))(\widehat{\mathcal{Y}}^{(0)}_{h,(m_{T,h})})^{\hat{\kappa}_{h,m_{T,h}}}$. *Let Assumptions A.1, B′ and C hold. Then* $\hat{\kappa}_{h,m_{T,h}} = \kappa_{h,0} + O_p(1/m^{1/2}_{T,h})$ *and* $\hat{d}_{h,m_{T,h}} = d_{h,0} + O_p(1/m^{1/2}_{T,h})$.

**Proof.** Note $(T-h)^{-1}\sum_{t=h+1}^{T} y_t y_{t-h} = E[y_t y_{t-h}] + O_p(1/T^{\iota})$ and $k_T = o(T^a)\ \forall a > 0$. Then $\hat{\kappa}_{h,m_{T,h}} = \kappa_{h,0} + O_p(1/m^{1/2}_{T,h})$ by Theorem 2.1 and Lemma 2.2 in Hill (2014b). Simply note that Assumption B′ falls under second order regular variation categories SR1 and SR2 in Goldie and Smith (1987). Hill (2010, 2014b) uses SR1, and exploits arguments developed in Hsing (1991), to prove $\hat{\kappa}_{h,m_{T,h}} = \kappa_{h,0} + O_p(1/m^{1/2}_{T,h})$ for a filtered dependent sequence. It is easily verified that Hill (2010, 2014b)'s asymptotic theory covers SR2 with the appropriate restriction to $m_{T,h}$, in particular that Assumption B′ suffices. $\hat{d}_{h,m_{T,h}} = d_{h,0} + O_p(1/m^{1/2}_{T,h})$ follows from Lemma A.3, $\hat{\kappa}_{h,m_{T,h}} = \kappa_{h,0} + O_p(1/m^{1/2}_{T,h})$, and the mapping theorem. $\mathcal{QED}$.

**Remark 17** If mean-centering is not used, e.g. if $y_t y_{t-h} \geq 0$, then Lemmas A.3 and A.4 respectively follow from Lemma 3 and Theorem 2 in Hill (2010). We only need the Assumption C.1 properties $k_T \to \infty$ and $k_T/T = o(1)$ since the limit properties of $(T-h)^{-1}\sum_{t=h+1}^{T} y_t y_{t-h}$ are then irrelevant. Similarly, if $E[y_t^4] < \infty$ then by Assumption A.1 and, e.g., Lemma 2.1 in McLeish (1975), $(T-h)^{-1}\sum_{t=h+1}^{T} y_t y_{t-h} = E[y_t y_{t-h}] + O_p(1/T^{1/2}) = E[y_t y_{t-h}] + o_p(1/k_T^{1/2})$, hence Lemmas A.3 and A.4 follow from arguments in Hill (2010, Lemma 3) and Hill (2014b, proof of Theorem 2.1, Lemma 2.2).

## A.2 Supporting Lemmas for Robust FD-QML

Throughout $\{r_{M,T}\}_{M,T\in\mathbb{N}}$ is a double array of finite non-random numbers with $\sup_{T\in\mathbb{N}}|r_{M,T}| \to 0$ as $M \to \infty$, and may be different in different places. Recall

$$\varpi(\lambda,\theta) \equiv -\frac{1}{f(\lambda,\theta)}\frac{\partial}{\partial\theta}\ln f(\lambda,\theta),$$

and recall moment and spectral density estimators

$$\widehat{\hat{\gamma}}^{*}_{T,h}(c) \equiv \frac{T-h}{T-h-k_T}\frac{1}{T}\sum_{t=h+1}^{T} y_t y_{t-h} I\left(\left|y_t y_{t-h} - \frac{1}{T-h}\sum_{t=h+1}^{T} y_t y_{t-h}\right| < c\right) : h \in \{0,...,b_T\}$$

$$\hat{\gamma}^{*}_{T,h}(c) \equiv \frac{T-h}{T-h-k_T}\frac{1}{T}\sum_{t=h+1}^{T} y_t y_{t-h} I\left(\left|y_t y_{t-h} - E\left[y_t y_{t-h}\right]\right| < c\right) : h \in \{0,...,b_T\}$$

[7] Hill (2014b) proves Hill (1975)'s tail index estimator is asymptotically normal for filtered weakly dependent sequences, e.g. $\widehat{\mathcal{Y}}^{(0)}_{h,t}$, by extending arguments in Hill (2010, Theorem 2) for distribution tails that satisfy a second order regular variation property. In particular, under conditions presented here, the plug-in does not affect the limit distribution of the Hill (1975)'s estimator. The same set of arguments extend to intermediate order sequences $\widehat{\mathcal{Y}}^{(0)}_{h,(k_T)}$, while asymptotics for $\widehat{\mathcal{Y}}^{(0)}_{h,(k_T)}$ does not require a second order regular variation property (see Hill, 2010, Lemma 3).

$$\widehat{\mathcal{I}}_T^*(\lambda) \equiv \frac{1}{2\pi}\left(\widehat{\hat{\gamma}}_{T,0}^*(\widehat{\mathcal{Y}}_{0,(k_T)}^{(0)}) + 2\sum_{h=1}^{b_T}\widehat{\hat{\gamma}}_{T,h}^*(\widehat{\mathcal{Y}}_{h,(k_T)}^{(0)})\cos(\lambda h)\right)$$
$$\mathcal{I}_T^*(\lambda) \equiv \frac{1}{2\pi}\left(\hat{\gamma}_{T,0}^*(c_{T,0}) + 2\sum_{h=1}^{b_T}\hat{\gamma}_{T,h}^*(c_{T,h})\cos(\lambda h)\right)$$
$$f_T^*(\lambda) \equiv (2\pi)^{-1}\sum_{h=-b_T}^{b_T} E\left[\hat{\gamma}_{T,|h|}^*(c_{T,|h|})\right]e^{-i\lambda h}.$$

Notice $f_T^*(\lambda)$ is defined here for the first time. Define criteria and estimators:

$$\tilde{\theta}_T^* = \arg\min_{\theta\in\Theta}\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\ln f(\lambda,\theta) + \frac{\widehat{\mathcal{I}}_T^*(\lambda)}{f(\lambda,\theta)}\right)d\lambda\right\} = \arg\min_{\theta\in\Theta}\left\{\tilde{\mathcal{Q}}_T^*(\theta)\right\} \tag{A.2}$$
$$\hat{\theta}_T^* = \left\{\arg\min_{\theta\in\Theta}\frac{1}{T}\sum_{j\in\mathbb{F}}\left(\ln f(\lambda_j,\theta) + \frac{\widehat{\mathcal{I}}_T^*(\lambda_j)}{f(\lambda_j,\theta)}\right)\right\} = \arg\min_{\theta\in\Theta}\left\{\hat{\mathcal{Q}}_T^*(\theta)\right\},$$

where $\mathbb{F} \equiv (-T/2, T/2] \cap \mathbb{Z} \setminus \{0\}$, and recall $\Omega$, $\mathcal{S}_T$ and $\mathcal{V}_T$ defined in (9). Recall also:

$$\kappa = \arg\sup\left\{\alpha > 0 : E\left|y_t\right|^\alpha < \infty\right\}.$$

We exploit the following implications of Karamata's Theorem under Assumption B[8]:

$$\kappa \in (2,4) : E\left[y_t^4 I\left(y_t^2 < c_{T,0}\right)\right] \sim \frac{4}{4-\kappa}\frac{k_T}{T}c_{T,0}^2 \tag{A.3}$$

$$\kappa = 4 : E\left[y_t^4 I\left(y_t^2 < c_{T,0}\right)\right] \sim \tilde{\mathcal{L}}_4(T) \text{ for some slowly varying } \tilde{\mathcal{L}}_4(T).$$

In the Paretian case of Assumption B$'$, by integration is it easily verified that $\tilde{\mathcal{L}}_4(T) = d_{0,0}\ln(T)$.

First, we can always work with $\tilde{\theta}_T^*$.

**Lemma A.5 (equivalent estimators)** *Under Assumptions A-C* $T^{1/2}\mathcal{V}_T^{-1/2}(\hat{\theta}_T^* - \tilde{\theta}_T^*) \xrightarrow{p} 0$ *where* $T/||\mathcal{V}_T|| \to \infty$.

**Proof.** The property $T/||\mathcal{V}_T|| \to \infty$ follows from $\mathcal{V}_T = \Omega^{-1}\mathcal{S}_T\Omega^{-1}$ and Lemma A.9. We now repeatedly use $\liminf_{T\to\infty}||\mathcal{S}_T|| > 0$ by Lemma A.9. Lemma A.6 can be easily modified to apply to $\hat{\mathcal{Q}}_T^*(\mathring{\theta}_T)$ to show $(\partial^2/\partial\theta\partial\theta')\hat{\mathcal{Q}}_T^*(\mathring{\theta}_T) \xrightarrow{p} \Omega$ for any $\mathring{\theta}_T \xrightarrow{p} \theta_0$. Similarly, Lemma A.7 can be modified to prove $T^{-1/2}\mathcal{S}_T^{-1/2}\sum_{j\in\mathbb{F}}(\widehat{\mathcal{I}}_T^*(\lambda_j) - \mathcal{I}_T^*(\lambda_j))\varpi(\lambda_j) \xrightarrow{p} 0$. Further $\mathcal{V}_T^{-1/2} = \mathcal{S}_T^{-1/2}\Omega$ by the definition of a square root matrix. Hence, by the optimization problems leading to $\{\hat{\theta}_T^*, \tilde{\theta}_T^*\}$ and first order

[8] The case $\kappa = 4$ is shown as follows. Let $\tilde{\mathcal{L}}(c)$ be a slowly varying function that may be different in different places. By Assumption B, a change of variables, and Karamata's Theorem: $\mathcal{E}_T \equiv E[(y_t^2 - E[y_t^2])^2 I(|y_t^2 - E[y_t^2]| < c_{T,0})] = K\int_0^{c_{T,0}^2}\mathcal{L}(u)u^{-1}du = \tilde{\mathcal{L}}(c_{T,0}^2)$. By Assumption B, $\kappa_{0,0} = \kappa/2$, and properties of regularly varying functions: $c_{T,0} \sim (T/k_T)^{2/\kappa}\tilde{\mathcal{L}}(T/k_T)$. Further, $\tilde{\mathcal{L}}(T/k_T) \sim \tilde{\mathcal{L}}(T)$ because $\tilde{\mathcal{L}}$ is slowly varying and $k_T > 0$. Hence, $c_{T,0} \sim (T/k_T)^{2/\kappa}\tilde{\mathcal{L}}(T)$. Therefore $\mathcal{E}_T = \tilde{\mathcal{L}}(c_{T,0}^2) = \tilde{\mathcal{L}}((T/k_T)^{2/\kappa}\tilde{\mathcal{L}}(T))$, where $\tilde{\mathcal{L}}((\xi T/k_T)^{2/\kappa}\tilde{\mathcal{L}}(\xi T))/\tilde{\mathcal{L}}((T/k_T)^{2/\kappa}\tilde{\mathcal{L}}(T)) \to 1$ by slow variation, hence $\mathcal{E}_T$ is a slowly varying function of $T$.

expansions around $\theta_0$, we have:

$$\begin{aligned}
&T^{1/2}\mathcal{V}_T^{-1/2}\left(\hat{\theta}_T^* - \tilde{\theta}_T^*\right) \\
&\quad = -T^{1/2}\mathcal{S}_T^{-1/2}\left\{\frac{1}{T}\sum_{j\in\mathbb{F}}\left(\mathcal{I}_T^*(\lambda_j) - f(\lambda_j)\right)\varpi(\lambda_j) - \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\mathcal{I}_T^*(\lambda) - f(\lambda)\right)\varpi(\lambda)d\lambda\right\} + o_p(1) \\
&\quad = -\mathcal{S}_T^{-1/2}\frac{1}{T^{1/2}}\sum_{j\in\mathbb{F}}\left(f_{(b_T)}(\lambda_j) - f(\lambda_j)\right)\varpi(\lambda_j) \\
&\qquad - T^{1/2}\mathcal{S}_T^{-1/2}\left\{\frac{1}{T}\sum_{j\in\mathbb{F}}\left(\mathcal{I}_T^*(\lambda_j) - f_{(b_T)}(\lambda_j)\right)\varpi(\lambda_j) - \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\mathcal{I}_T^*(\lambda) - f(\lambda)\right)\varpi(\lambda)d\lambda\right\} + o_p(1) \\
&\quad = -\mathcal{A}_{T,1} - \mathcal{A}_{T,2} + o_p(1).
\end{aligned}$$

where $f_{(M)}(\lambda)$ is the $M^{th}$ Cesáro sum of the Fourier series of $f(\lambda)$. Hölder continuity Assumption A.2.iv, bandwidth Assumption C.2, and boundedness of $\varpi(\lambda)$ imply $\mathcal{A}_{T,1} = O(T^{1/2}/b_T^{\alpha}) = o(1)$. See Zygmund (2002, Vol. I, Chapt. 2).

Next, letting $f_{T,(M)}^*(\lambda)$ be the $M^{th}$ Cesáro sum of the Fourier series of $f_T^*(\lambda) \equiv (2\pi)^{-1}\sum_{h=-b_T}^{b_T} E[\hat{\gamma}_{T,|h|}^*(c_{T,|h|})]e^{-i\lambda h}$, and noting $E[\mathcal{I}_T^*(\lambda)] = f_{T,(b_T)}^*(\lambda)$, we may write:

$$\begin{aligned}
\mathcal{A}_{T,2} &= T^{1/2}\mathcal{S}_T^{-1/2}\left\{\frac{1}{T}\sum_{j\in\mathbb{F}}\left(\mathcal{I}_T^*(\lambda_j) - E\left[\mathcal{I}_T^*(\lambda_j)\right]\right)\varpi(\lambda_j) - \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\mathcal{I}_T^*(\lambda) - E\left[\mathcal{I}_T^*(\lambda)\right]\right)\varpi(\lambda)d\lambda\right\} \\
&\quad + T^{1/2}\mathcal{S}_T^{-1/2}\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} f(\lambda)\varpi(\lambda)d\lambda - \frac{1}{T}\sum_{j\in\mathbb{F}} f(\lambda_j)\varpi(\lambda_j)\right) \\
&\quad - T^{1/2}\mathcal{S}_T^{-1/2}\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} f_T^*(\lambda)\varpi(\lambda)d\lambda - \frac{1}{T}\sum_{j\in\mathbb{F}} f_T^*(\lambda_j)\varpi(\lambda_j)\right) + \mathcal{S}_T^{-1/2}\frac{1}{T^{1/2}}\sum_{j\in\mathbb{F}}\left(f(\lambda_j) - f_{(b_T)}(\lambda_j)\right)\varpi(\lambda_j) \\
&\quad - \mathcal{S}_T^{-1/2}\frac{1}{T^{1/2}}\sum_{j\in\mathbb{F}}\left(f_T^*(\lambda_j) - f_{T,(b_T)}^*(\lambda_j)\right)\varpi(\lambda_j) + T^{1/2}\mathcal{S}_T^{-1/2}\frac{1}{2\pi}\int_{-\pi}^{\pi}\left(f_T^*(\lambda) - f_{T,(b_T)}^*(\lambda)\right)\varpi(\lambda)d\lambda \\
&= \mathcal{C}_{1,T} + \mathcal{C}_{2,T} - \mathcal{C}_{3,T} + \mathcal{C}_{4,T} - \mathcal{C}_{5,T} + \mathcal{C}_{6,T}.
\end{aligned}$$

By boundedness of $\varpi(\lambda)$, Hölder continuity Assumption A.2.iv, and bandwidth Assumption C.2, $\mathcal{C}_{i,T} = O(T^{1/2}/b_T^{\alpha}) = o(1)$ for $i = 4, 5, 6$.

Further, by Hölder continuity Assumption A.2.iv, and arguments in Dunsmuir (1979, p. 497):

$$\|\mathcal{C}_{2,T}\| \leq K\left(\frac{T}{\|\mathcal{S}_T\|}\right)^{1/2}\sup_{j\in\mathbb{F}}\sup_{|\lambda-\lambda_j|\leq 2\pi/T}\|f(\lambda)\varpi(\lambda) - f(\lambda_j)\varpi(\lambda_j)\| = O(T^{1/2-\alpha}) = o(1),$$

and likewise $\mathcal{C}_{3,T} = o(1)$. The remaining term $\mathcal{C}_{1,T}$ is $o_p(1)$ by replicating Dunsmuir (1979, p. 498)'s argument. $\mathcal{QED}$.

Next, standard asymptotic results apply under a negligible data transform.

**Lemma A.6 (Hessian)** *Let $\{\theta_T^*\}$ be any stochastic sequence $\{\theta_T^*\}$ that satisfies $\theta_T^* \xrightarrow{p} \theta_0$. Under Assumptions A-C* $(\partial^2/\partial\theta\partial\theta')\tilde{\mathcal{Q}}_T^*(\theta_T^*) \xrightarrow{p} \Omega$.

**Proof.** By Minkowski's inequality and approximation Lemma A.7:

$$\begin{aligned}
&\left\| \frac{\partial^2}{\partial\theta\partial\theta'} \tilde{\mathcal{Q}}_T^* \left(\theta_T^*\right) - \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\partial}{\partial\theta} \ln f(\lambda) \frac{\partial}{\partial\theta} \ln f(\lambda)' d\lambda \right\| \\
&\quad \le \left\| \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\partial}{\partial\theta} \ln f(\lambda) \frac{\partial}{\partial\theta} \ln f(\lambda)' d\lambda - \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\partial}{\partial\theta} \ln f(\lambda, \theta_T^*) \frac{\partial}{\partial\theta} \ln f(\lambda, \theta_T^*)' d\lambda \right\| \\
&\qquad + \left\| \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\mathcal{I}_T^*(\lambda) - f(\lambda)\right) \frac{\partial}{\partial\theta} \varpi(\lambda) d\lambda \right\| \\
&\qquad + \left\| \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\mathcal{I}_T^*(\lambda) - f(\lambda)\right) \left( \frac{\partial}{\partial\theta} \varpi(\lambda, \theta_T^*) - \frac{\partial}{\partial\theta} \varpi(\lambda) \right) d\lambda \right\| \\
&\qquad + \left\| \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(f(\lambda, \theta_T^*) - f(\lambda)\right) \frac{\partial}{\partial\theta} \varpi(\lambda, \theta_T^*) d\lambda \right\| + o_p\left(1\right) \\
&\quad = \mathcal{A}_1(\theta_T^*) + \mathcal{A}_{2,T} + \mathcal{A}_{3,T}(\theta_T^*) + \mathcal{A}_{4,T}(\theta_T^*) + o_p\left(1\right).
\end{aligned}$$

The $o_p(1)$ term follows from Lemma A.7. The terms $\mathcal{A}_{2,T}$ and $\mathcal{A}_{3,T}(\theta_T^*)$ are $o_p(1)$ by Lemma A.8.

Next, consider $\mathcal{A}_1(\theta_T^*)$ and let $a_{i,j}(\theta)$ be the $(i,j)^{th}$ element of the matrix $(2\pi)^{-1}\int_{-\pi}^{\pi}(\partial/\partial\theta)\ln f(\lambda,\theta) \times (\partial/\partial\theta)\ln f(\lambda,\theta)'d\lambda$. Assumption A.2 spectrum boundedness and smoothness properties imply $\sup_{\theta\in\Theta}||(\partial/\partial\theta)a_{i,j}(\theta)|| \le K$. By a first order expansion $|a_{i,j}(\theta_T^*) - a_{i,j}(\theta_0)| \le \sup_{\theta\in\Theta}||(\partial/\partial\theta)a_{i,j}(\theta)|| \times ||\theta_T^* - \theta_0|| \le K||\theta_T^* - \theta_0|| \xrightarrow{p} 0$ hence $\mathcal{A}_1(\theta_T^*) \xrightarrow{p} 0$.

Lastly, $\mathcal{A}_{4,T}(\theta_T^*)$. Assumption A, a first order expansion and consistency imply: $|f(\lambda,\theta_T^*) - f(\lambda)| \times ||(\partial/\partial\theta)\varpi(\lambda,\theta_T^*)|| \le K|f(\lambda,\theta_T^*) - f(\lambda)| \le K||\theta_T^* - \theta_0|| \xrightarrow{p} 0$. Hence $\mathcal{A}_{4,T}(\theta_T^*) \le K\int_{-\pi}^{\pi}|f(\lambda,\theta_T^*) - f(\lambda)|d\lambda \le K||\theta_T^* - \theta_0|| \xrightarrow{p} 0$. $\mathcal{QED}$.

**Lemma A.7 (periodogram approximation)** *Let $\omega(\lambda,\theta)$ be an $\mathbb{R}$-valued uniformly bounded mapping on $[-\pi,\pi]\times\Theta$. Under Assumptions A-C* $\sup_{\theta\in\Theta}|(T/||\mathcal{S}_T||)^{1/2}\int_{-\pi}^{\pi}(\widehat{\mathcal{I}}_T^*(\lambda) - \mathcal{I}_T^*(\lambda))\omega(\lambda,\theta)d\lambda| \xrightarrow{p} 0$ *where* $T/||\mathcal{S}_T|| \to \infty$.

**Proof.** The property $T/||\mathcal{S}_T|| \to \infty$ follows from Lemma A.9. Let $\omega_{(M)}(\lambda,\theta)$ be the $M^{th}$ Cesáro sum of the Fourier series of uniformly bounded $\omega(\lambda,\theta)$, and let $\omega_h(\theta)$ be the $h^{th}$ Fourier coefficient of $\omega(\lambda,\theta)$. Then by Fejer's theorem and the construction of $\omega_{(M)}(\lambda,\theta)$, for $M \in \mathbb{N}$:

$$\begin{aligned}
&\left(\frac{T}{\|\mathcal{S}_T\|}\right)^{1/2} \int_{-\pi}^{\pi} \left(\widehat{\mathcal{I}}_T^*(\lambda) - \mathcal{I}_T^*(\lambda)\right) \omega(\lambda,\theta) d\lambda \\
&\quad = \left(\frac{T}{\|\mathcal{S}_T\|}\right)^{1/2} \int_{-\pi}^{\pi} \left(\widehat{\mathcal{I}}_T^*(\lambda) - \mathcal{I}_T^*(\lambda)\right) \omega_{(M)}(\lambda,\theta) d\lambda + o_p(1) \\
&\quad = \frac{1}{2\pi} \sum_{h=-M}^{M} \left(1 - \frac{|h|}{M}\right) \omega_h(\theta) \left(\frac{T}{\|\mathcal{S}_T\|}\right)^{1/2} \left\{ \widehat{\hat{\gamma}}_{T,h}^*(\widehat{\mathcal{Y}}_{h,(k_T)}^{(0)}) - \hat{\gamma}_{T,h}^*(c_{T,h}) \right\} + o_p(r_{M,T}) \\
&\quad = \mathcal{A}_{M,T}(\theta) + o_p(r_{M,T}).
\end{aligned}$$

Observe $\lim_{M\to\infty}\sup_{\theta\in\Theta}|\sum_{h=-M}^{M}(1-|h|/M)\omega_h(\theta)| = \sup_{\theta\in\Theta}|2\pi\int_{-\pi}^{\pi}\omega(\lambda,\theta)d\lambda| < \infty$ by construction and uniform boundedness, and $(T/||\mathcal{S}_T||)^{1/2}\times|\widehat{\hat{\gamma}}_{T,h}^{*}(\widehat{\mathcal{Y}}_{h,(k_T)}^{(0)}) - \hat{\gamma}_{T,h}^{*}(c_{T,h})| \xrightarrow{p} 0$ by Lemma A.2.b. Hence $\lim_{M\to\infty}\sup_{\theta\in\Theta}|\mathcal{A}_{M,T}(\theta)| \xrightarrow{p} 0$ as $T\to\infty$. The claim now follows from $r_{M,T}\to 0$ as $M\to\infty$ for any $T$. $\mathcal{QED}$.

**Lemma A.8 (LLN)** *Let $\omega(\lambda,\theta)$ be an $\mathbb{R}$ -valued uniformly bounded mapping on $[-\pi,\pi]\times\Theta$ with integrable envelope $\sup_{\theta\in\Theta}|\omega(\lambda,\theta)|$. Under Assumption A and C* $\sup_{\theta\in\Theta}|\int_{-\pi}^{\pi}\{\mathcal{I}_T^{*}(\lambda)-f(\lambda)\}\omega(\lambda,\theta)d\lambda| \xrightarrow{p} 0$.

**Proof.** Let $\omega_{(M)}(\lambda,\theta)$ and $\omega_h(\theta)$ be as in the proof of Lemma A.7. By Lemma A.2.a, negligibility $k_T/T\to 0$, and dominated convergence $\int_{-\pi}^{\pi}\mathcal{I}_T^{*}(\lambda)d\lambda = 1/T\sum_{t=1}^{T}y_t^2 I(|y_t^2 - E[y_t^2]| + o_p(1) < c_{T,0}) \xrightarrow{p} E[y_t^2] < \infty$, and by supposition $\int_{-\pi}^{\pi}\sup_{\theta\in\Theta}|\omega(\lambda,\theta)|d\lambda < \infty$. Hence, by Fejer's theorem:

$$\begin{aligned}\int_{-\pi}^{\pi}\{\mathcal{I}_T^{*}(\lambda)-f(\lambda)\}\omega(\lambda,\theta)d\lambda &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\{\mathcal{I}_T^{*}(\lambda)-f(\lambda)\}\,\omega_{(M)}(\lambda,\theta)d\lambda \\ &\quad + \frac{1}{2\pi}\int_{-\pi}^{\pi}\{\mathcal{I}_T^{*}(\lambda)-f(\lambda)\}\left(\omega_{(M)}(\lambda,\theta)-\omega(\lambda,\theta)\right)d\lambda \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\{\mathcal{I}_T^{*}(\lambda)-f(\lambda)\}\,\omega_{(M)}(\lambda,\theta)d\lambda + o_p(r_{M,T}).\end{aligned}$$

By construction $(2\pi)^{-1}\int_{-\pi}^{\pi}\{\mathcal{I}_T^{*}(\lambda)-f(\lambda)\}\omega_{(M)}(\lambda,\theta)d\lambda = \mathcal{A}_{1,T}(M,\theta) + \mathcal{A}_{2,T}(M,\theta)$ where

$$\begin{aligned}\mathcal{A}_{1,T}(M,\theta) &= \frac{1}{(2\pi)^2}\sum_{h=-M}^{M}\left(1-\frac{|h|}{M}\right)\omega_h(\theta)\left\{E\left[\hat{\gamma}_{T,|h|}^{*}(c_{T,|h|})\right]-E[y_t y_{t-h}]\right\} \\ \mathcal{A}_{2,T}(M,\theta) &= \frac{1}{(2\pi)^2}\sum_{h=-M}^{M}\left(1-\frac{|h|}{M}\right)\omega_h(\theta)\left\{\hat{\gamma}_{T,|h|}^{*}(c_{T,|h|})-E\left[\hat{\gamma}_{T,h}^{*}(c_{T,h})\right]\right\}.\end{aligned}$$

Observe $\lim_{M\to\infty}\sup_{\theta\in\Theta}|\sum_{-M}^{M}(1-|h|/M)\omega_h(\theta)| = 2\pi\sup_{\theta\in\Theta}|\int_{-\pi}^{\pi}\omega(\lambda,\theta)d\lambda| < \infty$ by uniform boundedness and integrability, $E[\hat{\gamma}_{T,|h|}^{*}(c_{T,|h|})]\to E[y_t y_{t-h}]$ by negligibility and dominated convergence, and $|\hat{\gamma}_{T,|h|}^{*}(c_{T,|h|}) - E[\hat{\gamma}_{T,|h|}^{*}(c_{T,|h|})]| \xrightarrow{p} 0$ by Lemma A.2.a. Therefore, $\lim_{M\to\infty}\sup_{\theta\in\Theta}|\mathcal{A}_{1,T}(M,\theta)| \to 0$ and $\lim_{M\to\infty}\sup_{\theta\in\Theta}|\mathcal{A}_{2,T}(M,\theta)| \xrightarrow{p} 0$, hence $\sup_{\theta\in\Theta}|\int_{-\pi}^{\pi}\{\mathcal{I}_T^{*}(\lambda)-f(\lambda)\}\omega(\lambda,\theta)d\lambda| \xrightarrow{p} 0$. $\mathcal{QED}$.

We require central limit theorems for both the tail-trimmed spectral density estimator, and a combination of tail-trimmed and tail event random variables, used respectively for the non-bias corrected and bias corrected estimators. Let $\omega_h$ be the $h^{th}$ Fourier coefficient of $\omega(\lambda)$, and define

$$\tilde{\mathcal{B}}_T \equiv \frac{1}{2\pi}\sum_{h=-b_T}^{b_T}\varpi_h\left\{E\left[\hat{\gamma}_{T,|h|}^{*}(c_{T,|h|})\right]-E\left[y_t y_{t-h}\right]\right\}.$$

**Lemma A.9 (CLT for the tail-trimmed periodogram)** *Let Assumptions A-C hold (in particular, let the trimming fractiles satisfy $k_{T,h}\sim K_{h,\tilde{h}}k_{T,\tilde{h}}$ for some finite $K_{h,\tilde{h}}>0$ and each $\tilde{h},h\in$*

$\{0, ..., b_T\}$).

*a.* $(2\pi)^{-1}T^{1/2}\mathcal{S}_T^{-1/2}\{\int_{-\pi}^{\pi}(\mathcal{I}_T^*(\lambda) - f(\lambda))\varpi(\lambda)d\lambda - \tilde{\mathcal{B}}_T\} \xrightarrow{d} N(0, I_k)$. *If* $y_t y_{t-h}$ *has a symmetric distribution then* $\tilde{\mathcal{B}}_T = (2\pi)^{-1}\varpi_0\{E[\hat{\gamma}_{T,0}^*(c_{T,0})] - E[y_t^2]\}$.

*b.* $||\mathcal{S}_T|| \sim KE[y_t^4 I(y_t^2 \leq c_{T,0})]$ *hence* $\liminf_{T\to\infty} ||\mathcal{S}_T|| > 0$.

*c. If* $\kappa > 4$ *then* $||\mathcal{S}_T|| \sim K$, *if* $\kappa = 4$ *then* $||\mathcal{S}_T|| \sim K\tilde{\mathcal{L}}_4(T)$ *for slowly varying* $\tilde{\mathcal{L}}_4(T)$ *in (A.3), and if* $\kappa \in (2, 4)$ *then* $||\mathcal{S}_T|| \sim K(T/k_T)^{4/\kappa - 1}$, *hence* $T/||\mathcal{S}_T|| \to \infty$ $\forall \kappa > 2$.

**Proof.** Write $\psi_{h,t}(c) \equiv y_t y_{t-h} I(|y_t y_{t-h} - E[y_t y_{t-h}]| < c)$, where $\psi_{h,t}(c) = 0$ for $t \notin \{1 + h, ..., T\}$ and $h > b_T$.

**Claim (a).** Let

$$\mathcal{S}_{M,T} \equiv E\left[\mathfrak{G}_{T,M}\mathfrak{G}_{T,M}'\right] \text{ where } \mathfrak{G}_{T,M} \equiv T^{1/2}\frac{1}{(2\pi)^2}\sum_{h=-M}^{M}\left(1 - \frac{|h|}{M}\right)\varpi_h\left\{\hat{\gamma}_{T,|h|}^*(c_{T,|h|}) - E\left[\hat{\gamma}_{T,|h|}^*(c_{T,|h|})\right]\right\}.$$

Recall $f_T^*(\lambda) \equiv (2\pi)^{-1}\sum_{h=-b_T}^{b_T} E[\hat{\gamma}_{T,|h|}^*(c_{T,|h|})]e^{-i\lambda h}$, let $\omega_{(M)}(\lambda)$ and $\omega_h$ be as in the proof of Lemma A.7, and $f_{(M)}(\lambda)$ and $f_{T,(M)}^*(\lambda)$ respectively be the $M^{th}$ Cesáro sum of the Fourier series of $f(\lambda)$ and $f_T^*(\lambda)$. By construction:

$$f_{T,(b_T)}^*(\lambda) = E[\mathcal{I}_T^*(\lambda)] = \frac{1}{2\pi}\sum_{h=-b_T}^{b_T} E[\hat{\gamma}_{T,|h|}^*(c_{T,|h|})]e^{-ih\lambda} \text{ and } f_{(b_T)}(\lambda) = \frac{1}{2\pi}\sum_{h=-b_T}^{b_T} E[y_t y_{t-h}]e^{-ih\lambda}.$$

Now use $\liminf_{T\to\infty} ||\mathcal{S}_T|| > 0$ by Claim (c), Hölder continuity Assumption A.2.iv, bandwidth Assumption C.2 $b_T^{\alpha}/T^{1/2} \to \infty$, and the construction of the Fourier coefficients $\omega_h$ to deduce

$$\begin{aligned}
T^{1/2}\mathcal{S}_T^{-1/2}\int_{-\pi}^{\pi}\varpi(\lambda)\{f_T^*(\lambda) - f(\lambda)\}d\lambda &= T^{1/2}\mathcal{S}_T^{-1/2}\int_{-\pi}^{\pi}\varpi(\lambda)\{f_{T,(b_T)}^*(\lambda) - f_{(b_T)}(\lambda)\}d\lambda + O(T^{1/2}/b_T^{\alpha}) \\
&= T^{1/2}\mathcal{S}_T^{-1/2}\frac{1}{2\pi}\sum_{h=-b_T}^{b_T}\left(1 - \frac{|h|}{b_T}\right)\varpi_h\left\{E\left[\hat{\gamma}_{T,|h|}^*(c_{T,|h|})\right] - E\left[y_t y_{t-h}\right]\right\} + o(1) \\
&= T^{1/2}\mathcal{S}_T^{-1/2}\frac{1}{2\pi}\sum_{h=-b_T}^{b_T}\varpi_h\left\{E\left[\hat{\gamma}_{T,|h|}^*(c_{T,|h|})\right] - E\left[y_t y_{t-h}\right]\right\} + o(1).
\end{aligned}$$

This implies bias $\tilde{\mathcal{B}}_T$ can be expressed as:

$$\tilde{\mathcal{B}}_T = \int_{-\pi}^{\pi}\varpi(\lambda)\{f_T^*(\lambda) - f(\lambda)\}d\lambda + o\left(\|\mathcal{S}_T\|^{1/2}/T^{1/2}\right).$$

Hence, by the same Hölder continuity argument:

$$\begin{aligned}
&T^{1/2}\mathcal{S}_T^{-1/2}\frac{1}{2\pi}\left(\int_{-\pi}^{\pi}\left(\mathcal{I}_T^*(\lambda) - f(\lambda)\right)\varpi(\lambda)d\lambda - \tilde{\mathcal{B}}_T\right) \qquad \text{(A.4)} \\
&\quad = T^{1/2}\mathcal{S}_T^{-1/2}\frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\mathcal{I}_T^*(\lambda) - f_T^*(\lambda)\right)\varpi(\lambda)d\lambda + o(1)
\end{aligned}$$

$$= T^{1/2}\mathcal{S}_T^{-1/2}\frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\mathcal{I}_T^*(\lambda) - E\left[\mathcal{I}_T^*(\lambda)\right]\right)\varpi(\lambda)d\lambda + o\left(1\right).$$

Now use the same Cesáro sum argument exploited in the previous proofs, and (A.4), to obtain:

$$\begin{aligned}
&T^{1/2}\mathcal{S}_T^{-1/2}\frac{1}{2\pi}\left(\int_{-\pi}^{\pi}\left(\mathcal{I}_T^*(\lambda) - f(\lambda)\right)\varpi(\lambda)d\lambda - \tilde{\mathcal{B}}_T\right) \qquad \text{(A.5)}\\
&\quad = T^{1/2}\mathcal{S}_T^{-1/2}\frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\mathcal{I}_T^*(\lambda) - E\left[\mathcal{I}_T^*(\lambda)\right]\right)\varpi(\lambda)d\lambda + o\left(1\right)\\
&\quad = \left(\mathcal{S}_T^{-1/2}\mathcal{S}_{M,T}^{1/2}\right)\mathcal{S}_{M,T}^{-1/2}T^{1/2}\frac{1}{(2\pi)^2}\sum_{h=-M}^{M}\left(1-\frac{|h|}{M}\right)\varpi_h\left\{\hat{\gamma}^*_{T,|h|}(c_{T,|h|}) - E\left[\hat{\gamma}^*_{T,|h|}(c_{T,|h|})\right]\right\} + o_p(r_{M,T}) + o\left(1\right)\\
&\quad = \left(\mathcal{S}_T^{-1/2}\mathcal{S}_{M,T}^{1/2}\right)\times\mathcal{Z}_{M,T} + o_p(r_{M,T}) + o\left(1\right),
\end{aligned}$$

say. By construction of the Cesáro sums and $\mathcal{S}_T$ we have $\lim_{M\to\infty}\mathcal{S}_T^{-1}\mathcal{S}_{M,T} = I_k$. The claim therefore follows if we show $\mathcal{Z}_{M,T}\xrightarrow{d}N(0,I_k)$ as $T\to\infty$ for any $M$ since we may then take $M$ to be arbitrarily large to complete the proof. By the Cramér-Wold Theorem we need only prove $\xi'\mathcal{Z}_{M,T}\xrightarrow{d}N(0,1)$ as $T\to\infty$ for any $\xi\in\mathbb{R}^k$, $\xi'\xi = 1$.

Write $\tilde{\varpi}_h \equiv \varpi_{-h} + \varpi_h$ for $h\neq 0$ and $\tilde{\varpi}_0\equiv\varpi_0$. Define

$$\begin{aligned}
\zeta_{M,T,h}(\xi) &\equiv \begin{cases}\frac{1}{(2\pi)^2}\left(1-\frac{h}{M}\right)\xi'\mathcal{S}_{M,T}^{-1/2}\tilde{\varpi}_h\times\left(\xi'\mathcal{S}_{M,T}\xi\right)^{1/2} & \text{for } h\in\{0,...,M\}\\ 0 \ \text{ for } \ h>M\end{cases} \qquad \text{(A.6)}\\
\mathcal{E}_{T,h,t}(\xi) &\equiv \frac{1}{\left(\xi'\mathcal{S}_{M,T}\xi\right)^{1/2}}\frac{T-h}{T-h-k_T}\left\{\psi_{h,t}(c_{T,h}) - E\left[\psi_{h,t}(c_{T,h})\right]\right\}\\
\mathcal{X}_{M,T,t}(\xi) &\equiv \sum_{h=0}^{\infty}\zeta_{h,M,T}(\xi)\mathcal{E}_{T,h,t}(\xi) = \sum_{h=0}^{M}\zeta_{h,M,T}(\xi)\mathcal{E}_{T,h,t}(\xi).
\end{aligned}$$

Notice by construction $\mathcal{E}_{T,h,t}(\xi) = 0\ \forall t\notin\{h+1,...,T\}$ and $h>b_T$. Then

$$\xi'\mathcal{Z}_{M,T} = \frac{1}{T^{1/2}}\sum_{t=1}^{T}\mathcal{X}_{M,T,t}(\xi)\times(1+o_p(1)). \qquad \text{(A.7)}$$

It remains to prove $1/T^{1/2}\sum_{t=1}^{T}\mathcal{X}_{M,T,t}(\xi)\xrightarrow{d}N(0,1)$. We now drop $\xi$ to reduce notation.

Central limit theorem Lemma A.1 applies to $\mathcal{X}_{M,T,t}$ by measurability and the stated assumptions on $y_t$. First, by construction and Lemma A.1.a $1 = E(1/T^{1/2}\sum_{t=1}^{T}\mathcal{X}_{M,T,t})^2 \sim KE[\mathcal{X}_{M,T,t}^2]$, hence $E[\mathcal{X}_{M,T,t}^2]\sim K$. Notice $K$ depends on $M$: we do not show this since $K$ is positive and finite for each $M$.

Next, by Lemma A.1.b $1/T^{1/2}\sum_{t=1}^{T}\mathcal{X}_{M,T,t}\xrightarrow{d}N(0,1)$ provided Lindeberg condition (A.1) holds for $\mathcal{X}_{M,T,t}$. By stationarity and $E(1/T^{1/2}\sum_{t=1}^{T}\mathcal{X}_{M,T,t})^2\sim K$ it suffices to show $E[\mathcal{X}_{M,T,t}^2 I(\mathcal{X}_{M,T,t}^2 > T\epsilon^2)]\to 0\ \forall\epsilon>0$. Use the triangle inequality, $\lim_{M\to\infty}\mathcal{S}_T^{-1}\mathcal{S}_{M,T} = I_k$, $||\mathcal{S}_T||\sim KE[\psi_{0,t}^2(c_{T,0})]$ by Claim (b), $\liminf_{T\to\infty}E[\psi_{0,t}^2(c_{T,0})] > 0$ by negligibility and distribution non-degenerateness, and

$|\psi_{h,t}(c_{T,h})| \leq c_{T,h}$ by construction, to deduce for some $\vartheta \in \mathbb{R}^k$, $\vartheta'\vartheta = 1$:

$$|\mathcal{X}_{M,T,t}| \leq \sum_{h=0}^{M} |\zeta_{h,M,T}\mathcal{E}_{T,h,t}| \leq K\sum_{h=0}^{M}\left|\left(1-\frac{h}{M}\right)(\vartheta'\tilde{\varpi}_h) \times \frac{c_{T,h}}{\left(E\left[\psi_{0,t}^2(c_{T,0})\right]\right)^{1/2}}\right|.$$

Under power law Assumption B $c_{T,h} \sim \mathcal{L}_{h,0}(T)^{1/\kappa_{h,0}}(T/k_T)^{1/\kappa_{h,0}}$, and by construction $\kappa_{h,0} \geq \kappa_{0,0}$. Hence $c_{T,h} \leq \mathcal{L}_{h,0}(T)^{1/\kappa_{0,0}}(T/k_T)^{1/\kappa_{0,0}}$. Since under Assumption A $\limsup_{M\to\infty} K\sum_{h=0}^{M}|(1-h/M)(\vartheta'\tilde{\varpi}_h)| \leq K$ we therefore have $|\mathcal{X}_{M,T,t}| \leq K\mathcal{L}_{h,0}(T)^{1/\kappa_{0,0}} \times (T/k_T)^{1/\kappa_{0,0}}/(E[\psi_{0,t}^2(c_{T,0})])^{1/2} \equiv \mathfrak{A}_T$, say.

If we prove $\mathfrak{A}_T = o(T^{1/2})$, then $\mathcal{X}_{M,T,t}^2 < T\epsilon^2$ *a.s.* for all $T \geq N_\epsilon$ and some $N_\epsilon \geq 1$, hence $E[\mathcal{X}_{M,T,t}^2 I(\mathcal{X}_{M,T,t}^2 > T\epsilon^2)] = 0\ \forall T \geq N_\epsilon$ which completes the proof. If $\kappa \geq 4$ then $\kappa_{0,0} \geq 2$ hence $\mathfrak{A}_T \leq K\mathcal{L}_{h,0}(T)^{1/\kappa_{0,0}}(T/k_T)^{1/2} = o(T^{1/2})$. Conversely, if $\kappa \in (2,4)$ then $\kappa_{0,0} \in (1,2)$ hence by Karamata theory (A.3) $E[\psi_{0,t}^2(c_{T,0})] \sim K(T/k_T)^{2/\kappa_{0,0}-1}$. Hence $\mathfrak{A}_T \sim K\mathcal{L}_{h,0}(T)^{1/\kappa_{0,0}} \times (T/k_T)^{1/\kappa_{0,0}}/(T/k_T)^{1/\kappa_{0,0}-1/2}$ $= \mathcal{L}_{h,0}(T)^{1/\kappa_{0,0}}(T/k_T)^{1/2} = o(T^{1/2})$.

**Claim (b).** Define $\Psi_{h,t} \equiv \psi_{h,t}(c_{T,h}) - E[\psi_{h,t}(c_{T,h})]$, $\eta_{M,h} \equiv 1 - h/M$ and $\vartheta_T(\xi) \equiv \mathcal{S}_T^{-1/2}\xi/(\xi'\mathcal{S}_T^{-1}\xi)^{1/2}$. By the arguments above $\mathcal{S}_{M,T}\mathcal{S}_T^{-1} \sim I_k$ and $1 = E(1/T^{1/2}\sum_{t=1}^{T}\mathcal{X}_{M,T,t}(\xi))^2 \sim KE[\mathcal{X}_{M,T,t}^2(\xi)]$, and by another application of Lemma A.1.a: $E[\mathcal{X}_{M,T,t}^2(\xi)] \sim K\sum_{h=0}^{M}\zeta_{h,M,T}^2(\xi)E[\mathcal{E}_{T,h,t}^2(\xi)]$. Therefore

$$E\left[\mathcal{X}_{M,T,t}^2(\xi)\right] \sim K\sum_{h=0}^{M}\zeta_{h,M,T}^2(\xi)E\left[\mathcal{E}_{T,h,t}^2(\xi)\right] \sim K\sum_{h=0}^{M}\eta_{M,h}^2\left(\xi'\mathcal{S}_T^{-1/2}\tilde{\varpi}_h\tilde{\varpi}_h'\mathcal{S}_T^{-1/2}\xi\right)E\left[\Psi_{h,t}^2\right] \sim K$$

hence

$$\begin{aligned}\left(\xi'\mathcal{S}_T^{-1}\xi\right)^{-1} &\sim K\sum_{h=0}^{M}\eta_{M,h}^2 \times \left(\vartheta_T(\xi)'\tilde{\varpi}_h\tilde{\varpi}_h'\vartheta_T(\xi)\right) \times E\left[\Psi_{h,t}^2\right] \\ &= KE\left[\Psi_{0,t}^2\right]\left(\left\{\vartheta_T(\xi)'\tilde{\varpi}_0\tilde{\varpi}_0'\vartheta_T(\xi)\right\} + \sum_{h=1}^{M}\eta_{M,h}^2\left\{\vartheta_T(\xi)'\tilde{\varpi}_h\tilde{\varpi}_h'\vartheta_T(\xi)\right\}\frac{E\left[\Psi_{h,t}^2\right]}{E\left[\Psi_{0,t}^2\right]}\right).\end{aligned}$$

Notice $\vartheta_T(\xi)'\vartheta_T(\xi) = 1$ for each $T$ and $\xi$. Further, $\kappa_{h,0} \geq \kappa_{0,0}$ and $k_{T,h} \sim K_{h,\tilde{h}}k_{T,\tilde{h}}$ for $K_{h,\tilde{h}} > 0$ imply $\max_{0\leq h\leq M} E[\Psi_{h,t}^2] \sim KE[\Psi_{0,t}^2]$ for any $M$. Moreover, $\limsup_{M\to\infty}\sup_{\vartheta'\vartheta=1}\sum_{h=0}^{M}\eta_{M,h}^2(\vartheta'\tilde{\varpi}_h\tilde{\varpi}_h'\vartheta) < \infty$ by boundedness properties Assumption A. Therefore $\sup_{\xi'\xi=1}(\xi'\mathcal{S}_T^{-1}\xi)^{-1} \sim KE[\Psi_{0,t}^2]$ and $\inf_{\xi'\xi=1}(\xi'\mathcal{S}_T^{-1}\xi)^{-1}$ $\sim KE[\Psi_{0,t}^2]$. This proves by the definition of the spectral norm $||\mathcal{S}_T^{-1}||^{-1} \sim KE[\Psi_{0,t}^2]$ hence $||\mathcal{S}_T|| \sim KE[\Psi_{0,t}^2] \sim KE[\psi_{0,t}^2(c_{T,0})]$. Finally, by negligibility $E[\psi_{0,t}^2(c_{T,0})] \sim E[y_t^4 I(y_t^2 < c_{T,0})]$, and negligibility and a non-degenerate distribution imply $\liminf_{T\to\infty} E[y_t^4 I(y_t^2 < c_{T,0})] > 0$.

**Claim (c).** If $\kappa > 4$ then by dominated convergence $E[(y_t y_{t-h})^2 I(|y_t y_{t-h} - E[y_t y_{t-h}]| < c_{T,h})]$ $\sim E[(y_t y_{t-h})^2] \in (0,\infty)$, and if $\kappa \leq 4$ then by Karamata theory $E[(y_t y_{t-h} - E[y_t y_{t-h}])^2 I(|y_t y_{t-h} - E[y_t y_{t-h}]| < c_{T,h})]$ is shown in (A.3). Coupled with Claim (b) this proves the orders for $||\mathcal{S}_T||$. $\mathcal{QED}$.

Define $\kappa_{h,0} \equiv \min\{\kappa_{h,1}, \kappa_{h,2}\}$ and:

$$\mathfrak{I}^*_{T,h,t} \equiv I\left(|y_t y_{t-h} - E\left[y_t y_{t-h}\right]| \geq c_{T,h}\right) - P\left(|y_t y_{t-h} - E\left[y_t y_{t-h}\right]| \geq c_{T,h}\right) \tag{A.8}$$
$$\mathfrak{I}^*_{T,h} \equiv \frac{1}{\kappa_{h,0}} \frac{1}{k_T^{1/2}} \sum_{t=h+1}^{T} \mathfrak{I}^*_{T,h,t}$$
$$\mathcal{D}_{T,0} = \frac{1}{\kappa_{0,0} - 1} \frac{k_T^{1/2}}{T - k_T} c_{T,0} \text{ and } \mathcal{D}_{T,h} = \frac{1}{k_T^{1/2}} \left(d_{h,1} c_{T,h}^{1-\kappa_{h,1}} - d_{h,2} c_{T,h}^{1-\kappa_{h,2}}\right) \text{ for } h \neq 0.$$

**Lemma A.10 (CLT for tail events and tail-trimmed periodogram)** *Let Assumptions A, B′ and C hold, let $M \in \mathbb{N}$ be arbitrary, and define*

$$\tilde{\mathcal{Z}}_{M,T} \equiv T^{1/2} \frac{1}{2\pi} \left( \int_{-\pi}^{\pi} \left(\mathcal{I}^*_T(\lambda) - f(\lambda)\right) \varpi(\lambda) d\lambda - \tilde{\mathcal{B}}_T + \frac{1}{2\pi} \sum_{h=-M}^{M} \left(1 - \frac{|h|}{M}\right) \varpi_h \mathcal{D}_{T,h} \mathfrak{I}^*_{T,h} \right)$$

*and $\tilde{\mathcal{S}}_{M,T} \equiv E[\tilde{\mathcal{Z}}_{M,T} \tilde{\mathcal{Z}}'_{M,T}]$. Then $\tilde{\mathcal{S}}_{M,T}^{-1/2} \tilde{\mathcal{Z}}_{M,T} \xrightarrow{d} N(0, I_k)$. If $\kappa > 4$ then $\tilde{\mathcal{S}}_T = \lim_{M\to\infty} \tilde{\mathcal{S}}_{M,T} = \mathcal{S}_T(1 + o(1))$ and if $\kappa \in (2, 4]$ then $\tilde{\mathcal{S}}_T = \lim_{M\to\infty} \tilde{\mathcal{S}}_{M,T} = \mathcal{S}_T(1 + O(1))$.*

**Proof.** Write $\psi_{h,t}(c) \equiv y_t y_{t-h} I(|y_t y_{t-h} - E[y_t y_{t-h}]| < c)]$. Let $\omega_{(M)}(\lambda)$ and $\omega_h$ be as in the proof of Lemma A.7, and define $\tilde{\varpi}_0 \equiv \varpi_0$ and $\tilde{\varpi}_h \equiv \varpi_{-h} + \varpi_h$ for $h \neq 0$. Define

$$\mathcal{A}_{T,h,t} \equiv \kappa_{h,0}^{-1} \mathcal{D}_{T,h} T k_T^{-1/2} \mathfrak{I}^*_{T,h,t}.$$

By the same arguments leading to (A.4)-(A.7), it can be shown $\tilde{\mathcal{S}}_{M,T}^{-1/2} \tilde{\mathcal{Z}}_{M,T}$ equals:

$$T^{1/2} \tilde{\mathcal{S}}_{M,T}^{-1/2} \left( \frac{1}{2\pi} \left\{ \int_{-\pi}^{\pi} \left(\mathcal{I}^*_T(\lambda) - f(\lambda)\right) \varpi(\lambda) d\lambda - \tilde{\mathcal{B}}_T \right\} + \frac{1}{(2\pi)^2} \sum_{h=-M}^{M} \left(1 - \frac{|h|}{M}\right) \mathcal{D}_{T,h} \mathfrak{I}^*_{T,h} \right) \tag{A.9}$$
$$= T^{1/2} \tilde{\mathcal{S}}_{M,T}^{-1/2} \left( \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\mathcal{I}^*_T(\lambda) - E\left[\mathcal{I}^*_T(\lambda)\right]\right) \varpi(\lambda) d\lambda + \frac{1}{(2\pi)^2} \sum_{h=-M}^{M} \left(1 - \frac{|h|}{M}\right) \mathcal{D}_{T,h} \mathfrak{I}^*_{T,h} \right) + o(1)$$
$$= \frac{1}{T^{1/2}} \sum_{t=1}^{T} \frac{1}{(2\pi)^2} \sum_{h=0}^{M} \left(1 - \frac{|h|}{M}\right) \mathring{\mathcal{S}}_{M,T}^{-1/2} \tilde{\varpi}_h \left\{\psi_{h,t}(c_{T,h}) - E\left[\psi_{h,t}(c_{T,h})\right] + \mathcal{A}_{T,h,t}\right\} (1 + o_p(1)) + o_p(r_{M,T}) + o(1)$$
$$= \frac{1}{T^{1/2}} \sum_{t=1}^{T} \tilde{\mathcal{X}}_{M,T,t} + o_p(r_{M,T}) + o(1),$$

say, where $\mathring{\mathcal{S}}_{M,T}$ satisfies $E(1/T^{1/2} \sum_{t=1}^{T} \tilde{\mathcal{X}}_{M,T,t})^2 = 1$. By construction and (A.9), $\lim_{M\to\infty} \mathring{\mathcal{S}}_{M,T} \tilde{\mathcal{S}}_{M,T}^{-1} = I_k$ and $\tilde{\mathcal{S}}_T = \lim_{M\to\infty} \tilde{\mathcal{S}}_{M,T}$ for each $T$.

In view of measurability, as in the proof of Lemma A.9 we need only use Lemma A.1 to prove the claim. We simplify notation by assuming $\theta_0$ is a scalar, but the following easily extends to the general case by use of the Cramér-Wold theorem.

First, consider how $\lim_{M\to\infty} \tilde{\mathcal{S}}_{M,T}$ relates to $\mathcal{S}_T$. By two applications of Lemma A.1.a, $E[(\tilde{\mathcal{S}}_{M,T}^{-1/2} \tilde{\mathcal{Z}}_{M,T})^2] = 1$ by construction, and $\lim_{M\to\infty} \mathring{\mathcal{S}}_{M,T} \tilde{\mathcal{S}}_{M,T}^{-1} = 1$, we have $1 = E(T^{-1/2} \sum_{t=1}^{T} \tilde{\mathcal{X}}_{M,T,t})^2 \sim K E[\tilde{\mathcal{X}}^2_{M,T,t}]$, and

$$\tilde{\mathcal{S}}_{M,T} \sim K\frac{1}{T}\sum_{t=1}^{T}\sum_{h=0}^{M}\left(1-\frac{|h|}{M}\right)^2 \tilde{\varpi}_h^2 E\left(\{\psi_{h,t}(c_{T,h}) - E\left[\psi_{h,t}(c_{T,h})\right] + \mathcal{A}_{T,h,t}\}\right)^2.$$

By degenerateness $I(|y_t y_{t-h} - E[y_t y_{t-h}]| \geq c_{T,h}) \to 0$ $a.s.$, the construction $I(|y_t y_{t-h} - E[y_t y_{t-h}]| < c_{T,h}) = 1 - I(|y_t y_{t-h} - E[y_t y_{t-h}]| \geq c_{T,h})$, and negligibility, the proof that $E(\{\psi_{h,t}(c_{T,h}) - E[\psi_{h,t}(c_{T,h})] + \mathcal{A}_{T,h,t}\})^2 \sim KE(\psi_{h,t}(c_{T,h}) - E[\psi_{h,t}(c_{T,h})])^2$, where $K = 1$ if $\kappa \geq 4$, is identical to the corresponding proof for Theorem 2.1 in Hill (2013). The claims $\lim_{M\to\infty} \tilde{\mathcal{S}}_{M,T} = \mathcal{S}_T(1 + o(1))$ if $\kappa > 4$ and $\lim_{M\to\infty} \tilde{\mathcal{S}}_{M,T} = \mathcal{S}_T(1 + O(1))$ if $\kappa \leq 4$ now follow from $\mathcal{S}_T \sim \lim_{M\to\infty} KT^{-1}\sum_{t=1}^{T}\sum_{h=0}^{M}(1 - |h|/M)^2\tilde{\varpi}_h^2 E(\psi_{h,t}(c_{T,h}) - E[\psi_{h,t}(c_{T,h})])^2$ by the proof of Lemma A.9.

Finally, by Lemma A.1.b we need only demonstrate the Lindeberg condition. Since $1 = E(T^{-1/2}\sum_{t=1}^{T}\tilde{\mathcal{X}}_{M,T,t})^2 \sim KE[\tilde{\mathcal{X}}_{M,T,t}^2]$ it suffices to check $E[\tilde{\mathcal{X}}_{M,T,t}^2 I(\tilde{\mathcal{X}}_{M,T,t}^2 > T\epsilon^2)] \to 0$ $\forall \epsilon > 0$. Observe that $1 < \kappa_{0,0} \leq \kappa_{h,0} = \min\{\kappa_{h,1}, \kappa_{h,2}\}$, and under power law Assumption B$'$ $c_{T,h} \sim K(T/k_T)^{1/\kappa_{h,0}}$, hence $\mathcal{D}_{T,0} \leq K(k_T^{1/2}/T)(T/k_T)^{1/\kappa_{0,0}}$ and $|\mathcal{D}_{T,h}| \leq Kc_{T,h}^{1-\kappa_{h,0}}/k_T^{1/2} = K(k_T^{1/2}/T)(T/k_T)^{1/\kappa_{h,0}} \leq K(k_T^{1/2}/T)(T/k_T)^{1/\kappa_{0,0}}$. Therefore

$$|\mathcal{A}_{T,h,t}| \leq K\left(\frac{T}{k_T}\right)^{1/\kappa_{0,0}} |I\left(|y_t y_{t-h}| \geq c_{T,h}\right) - P\left(|y_t y_{t-h}| \geq c_{T,h}\right)| \leq K\left(\frac{T}{k_T}\right)^{1/\kappa_{0,0}}.$$

Similarly, $\psi_{h,t}(c_{T,h}) \leq Kc_{T,h} \sim K(T/k_T)^{1/\kappa_{h,0}} \leq K(T/k_T)^{1/\kappa_{0,0}}$. By the construction of $\tilde{\mathcal{X}}_{M,T,t}$, the remaining steps for showing $E[\tilde{\mathcal{X}}_{M,T,t}^2 I(\tilde{\mathcal{X}}_{M,T,t}^2 > T\epsilon^2)] \to 0$ follow from the line of proof of Lemma A.9.a given $\lim_{M\to\infty} \mathring{\mathcal{S}}_{M,T}\tilde{\mathcal{S}}_{M,T}^{-1} = 1$, $\lim_{M\to\infty} \tilde{\mathcal{S}}_{M,T} = \mathcal{S}_T(1 + o(1))$, and $||\mathcal{S}_T|| \sim KE[\psi_{0,t}^2(c_{T,0})]$ by Lemma A.9.b. $\mathcal{QED}$.

## A.3 Proofs of Theorems

We now prove Theorems 2.1, 2.2 and 3.1. Let $\omega_h$ be the $h^{th}$ Fourier coefficient of $\varpi(\lambda) \equiv -(f(\lambda))^{-1}(\partial/\partial\theta)\ln f(\lambda)$.

**Proof of Theorem 2.1**. Define criteria

$$\hat{\mathcal{Q}}_T^*(\theta) \equiv \sum_{j\in\mathbb{F}}\left\{\ln f(\lambda_j,\theta) + \frac{\widehat{\mathcal{I}}_T^*(\lambda_j)}{f(\lambda_j,\theta)}\right\} \text{ and } \mathcal{Q}_T^*(\theta) \equiv \sum_{j\in\mathbb{F}}\left\{\ln f(\lambda_j,\theta) + \frac{\mathcal{I}_T^*(\lambda_j)}{f(\lambda_j,\theta)}\right\},$$

and write sample gradients and Hessians as

$$\widehat{\mathcal{G}}_T(\theta) \equiv \frac{\partial}{\partial\theta}\hat{\mathcal{Q}}_T^*(\theta),\ \mathcal{G}_T(\theta) \equiv \frac{\partial}{\partial\theta}\mathcal{Q}_T^*(\theta),\ \widehat{\mathcal{H}}_T(\theta) \equiv \frac{\partial}{\partial\theta'}\widehat{\mathcal{G}}_T(\theta) \text{ and } \mathcal{H}_T(\theta) \equiv \frac{\partial}{\partial\theta'}\mathcal{G}_T(\theta).$$

Recall the population Hessian

$$\mathcal{H}(\theta) \equiv -\frac{1}{2\pi}\int_{-\pi}^{\pi}\varpi(\lambda,\theta)\frac{\partial}{\partial\theta'}f(\lambda,\theta)d\lambda - \frac{1}{2\pi}\int_{-\pi}^{\pi}\left\{f(\lambda,\theta) - f(\lambda)\right\}\frac{\partial}{\partial\theta}\varpi(\lambda,\theta)d\lambda.$$

Define $\theta_T^* \equiv \arg\min_{\theta\in\Theta}\{\mathcal{Q}_T^*(\theta)\}$. We prove $||\hat{\theta}_T^* - \theta_T^*|| \xrightarrow{p} 0$ and $\theta_T^* \xrightarrow{p} \theta_0$.

**Step 1 ($||\hat{\theta}_T^* - \theta_T^*|| \xrightarrow{p} 0$).** Approximation Lemma A.7 and LLN Lemma A.8, combined with spectrum boundedness and Hölder continuity Assumption A.2.iv imply:

$$\sup_{\theta\in\Theta}\left\|\widehat{\mathcal{G}}_T(\theta) - \mathcal{G}_T(\theta)\right\| \xrightarrow{p} 0 \text{ and } \sup_{\theta\in\Theta}\left\|\widehat{\mathcal{H}}_T(\theta) - \mathcal{H}_T(\theta)\right\| \xrightarrow{p} 0. \tag{A.10}$$

See the proofs of Lemmas A.5 and A.7 for relating the integrated robust spectral density estimator in $\{\mathcal{G}_T(\theta), \mathcal{H}_T(\theta)\}$ to the weighted partial sums of the robust estimator embodying $\{\widehat{\mathcal{G}}_T(\theta), \widehat{\mathcal{H}}_T(\theta)\}$, cf. Dunsmuir (1979) and Hannan (1973a). We omit the added steps here to conserve space.

Taylor expansions show by the FD-QML first order conditions $\widehat{\mathcal{G}}_T = \widehat{\mathcal{H}}_T(\hat{\tilde{\theta}}_T) \times (\hat{\theta}_T^* - \theta_0)$ and $\mathcal{G}_T = \mathcal{H}_T(\tilde{\theta}_T) \times (\theta_T^* - \theta_0)$ for sequences $\{\hat{\tilde{\theta}}_T, \tilde{\theta}_T\}$ satisfying $||\hat{\tilde{\theta}}_T - \theta_0|| \leq ||\hat{\theta}_T^* - \theta_0||$ and $||\tilde{\theta}_T - \theta_0|| \leq ||\theta_T^* - \theta_0||$. Now use (A.10), spectrum boundedness, and multiple applications of Minkowski's inequality to deduce

$$\begin{aligned}
\left\|\hat{\theta}_T^* - \theta_T^*\right\| &= \left\|\left(\hat{\theta}_T^* - \theta_0\right) - (\theta_T^* - \theta_0)\right\| \\
&\leq \left\|\widehat{\mathcal{H}}_T(\hat{\tilde{\theta}}_T)\right\|^{-1} \times \left\|\widehat{\mathcal{G}}_T - \mathcal{G}_T\right\| + \left\|\widehat{\mathcal{H}}_T(\hat{\tilde{\theta}}_T) - \mathcal{H}_T(\tilde{\theta}_T)\right\| \times \left\|\widehat{\mathcal{H}}_T(\hat{\tilde{\theta}}_T)\right\|^{-1} \times \left\|\mathcal{H}(\tilde{\theta}_T)\right\|^{-1} \times \|\mathcal{G}_T\| \\
&= o_p\left(\left\|\widehat{\mathcal{H}}_T(\hat{\tilde{\theta}}_T)\right\|^{-1}\left(1 + \left\|\mathcal{H}(\tilde{\theta}_T)\right\|^{-1} \times \|\mathcal{G}_T\|\right)\right).
\end{aligned}$$

If $\sup_{\theta\in\Theta}\{||\mathcal{H}_T(\theta)||^{-1}\} = O_p(1)$ then by (A.10) we have $\sup_{\theta\in\Theta}||\widehat{\mathcal{H}}_T(\theta)\mathcal{H}_T^{-1}(\theta) - I_k|| \xrightarrow{p} 0$ hence

$$\sup_{\theta\in\Theta}\left\|\widehat{\mathcal{H}}_T(\theta)\right\|^{-1} \leq O_p(1) \times \sup_{\theta\in\Theta}\left\|\widehat{\mathcal{H}}_T(\theta)\mathcal{H}_T^{-1}(\theta) - I_k\right\|^{-1} + O_p(1) = O_p(1).$$

Therefore, if we show $||\mathcal{G}_T|| = O_p(1)$ and $\sup_{\theta\in\Theta}||\mathcal{H}_T(\theta)||^{-1} = O_p(1)$ then the proof that $||\hat{\theta}_T^* - \theta_T^*|| \xrightarrow{p} 0$ is complete.

By Hölder continuity Assumption A.2.iv (see Hannan, 1973a):

$$\mathcal{G}_T = \sum_{j\in\mathbb{F}}\{\mathcal{I}_T^*(\lambda_j) - f(\lambda_j)\}\varpi(\lambda_j) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\{\mathcal{I}_T^*(\lambda) - f(\lambda)\}\varpi(\lambda)d\lambda + o_p(1).$$

Spectrum boundedness and LLN Lemma A.8 therefore imply $||\mathcal{G}_T|| = o_p(1)$.

Next, recall $\mathcal{I}_T(\lambda)$ is the periodogram of $y_t$. By construction $\mathcal{H}_T(\theta) = \mathcal{H}(\theta) + r_T(\theta)$ where

$$\begin{aligned}
r_T(\theta) &= -\left(\sum_{j\in\mathbb{F}}\varpi(\lambda_j,\theta)\frac{\partial}{\partial\theta'}f(\lambda_j,\theta) - \frac{1}{2\pi}\int_{-\pi}^{\pi}\varpi(\lambda,\theta)\frac{\partial}{\partial\theta'}f(\lambda,\theta)d\lambda\right) \\
&\quad -\left(\sum_{j\in\mathbb{F}}\{f(\lambda_j,\theta) - f(\lambda_j)\}\frac{\partial}{\partial\theta}\varpi(\lambda,\theta) - \frac{1}{2\pi}\int_{-\pi}^{\pi}\{f(\lambda,\theta) - f(\lambda)\}\frac{\partial}{\partial\theta}\varpi(\lambda,\theta)d\lambda\right)
\end{aligned}$$

$$+\sum_{j\in\mathbb{F}}\left\{\mathcal{I}_T(\lambda_j)-f(\lambda_j)\right\}\frac{\partial}{\partial\theta}\varpi(\lambda_j,\theta)+\sum_{j\in\mathbb{F}}\left\{\mathcal{I}_T^*(\lambda_j)-\mathcal{I}_T(\lambda_j)\right\}\frac{\partial}{\partial\theta}\varpi(\lambda_j,\theta)$$
$$= \mathcal{A}_{1,T}(\theta)+\mathcal{A}_{2,T}(\theta)+\mathcal{A}_{3,T}(\theta)+\mathcal{A}_{4,T}(\theta).$$

Spectrum boundedness and Hölder continuity Assumption A.2.iv imply $||\mathcal{A}_{1,T}(\theta)||$, $||\mathcal{A}_{2,T}(\theta)||$ and $||\mathcal{A}_{3,T}(\theta)||$ are $o_p(1)$ uniformly on $\Theta$. See the proof of Lemma A.5 for related arguments, and see Dunsmuir (1979, p. 497-498) for classic arguments.

Consider $\mathcal{A}_{4,T}(\theta)$ and let $\eta:[-\pi,\pi]\times\Theta\rightarrow\mathbb{R}$ be any mapping that satisfies $|\eta(\lambda;\theta)|\leq K$ for all $(\lambda,\theta)\in[-\pi,\pi]\times\Theta$. Write $\hat{\gamma}_{T,h}\equiv 1/T\sum_{t=h+1}^{T}y_ty_{t-h}$. Square integrability and distribution continuity ensure $E|y_ty_{t-h}|^{1+\iota}<\infty$ for some infinitessimal $\iota>0$. Hence, by triangle and Hölder inequalities, stationarity and negligibility:

$$E\left|\hat{\gamma}_{T,h}^*(c_{T,h})-\hat{\gamma}_{T,h}\right|\leq\frac{1}{T}\sum_{t=h+1}^{T}\left(E\left|y_ty_{t-h}\right|^{1+\iota}\right)^{1/(1+\iota)}P\left(\left|y_ty_{t-h}-E\left[y_ty_{t-h}\right]\right|>c_{T,h}\right)^{\iota/(1+\iota)}=K\left(\frac{k_T}{T}\right)^{\iota/(1+\iota)}\rightarrow 0.$$

Therefore $|\hat{\gamma}_{T,h}^*(c_{T,h})-\hat{\gamma}_{T,h}|\overset{p}{\rightarrow}0$ by Markov's inequality. Now use boundedness of $\eta(\lambda;\theta)$ to deduce

$$\sup_{\theta\in\Theta}\left|\int_{-\pi}^{\pi}\left\{\mathcal{I}_T^*(\lambda)-\mathcal{I}_T(\lambda)\right\}\eta(\lambda;\theta)d\lambda\right|\overset{p}{\rightarrow}0. \tag{A.11}$$

Combine boundedness Assumption A.2 and (A.11) to deduce $\sup_{\theta\in\Theta}||\mathcal{A}_{4,T}(\theta)||\overset{p}{\rightarrow}0$. Therefore $\sup_{\theta\in\Theta}||\mathcal{H}_T(\theta)-\mathcal{H}(\theta)||\overset{p}{\rightarrow}0$. Finally, Assumption A.3 states $\inf_{\theta\in\Theta}||\mathcal{H}(\theta)||>0$ hence $\sup_{\theta\in\Theta}||\mathcal{H}_T(\theta)||^{-1}$ $=O_p(1)$.

**Step 2 ($\theta_T^*\overset{p}{\rightarrow}\theta_0$).** In view of (A.11) we can use $\mathcal{I}_T(\lambda)$ in place of $\mathcal{I}_T^*(\lambda)$. The proof of $\theta_T^*\overset{p}{\rightarrow}$ $\theta_0$ therefore follows by arguments in Dunsmuir and Hannan (1976), or by the proof of Theorem 1 in McCloskey and Hill (2014). $\mathcal{QED}$.

**Proof of Theorem 2.2**. Recall the solution $\tilde{\theta}_T^*$ defined by (A.2). We prove $T^{1/2}\mathcal{V}_T^{-1/2}(\tilde{\theta}_T^*-\theta_0$ $+\mathcal{B}_T)\overset{d}{\rightarrow}N(0,I_k)$ hence $T^{1/2}\mathcal{V}_T^{-1/2}(\hat{\theta}_T^*-\theta_0+\mathcal{B}_T)\overset{d}{\rightarrow}N(0,I_k)$ by Lemma A.5.

By optimization problem (A.2), and a first order expansion around $\theta_0$, for some $\mathring{\theta}_T$, $||\mathring{\theta}_T-\theta_0||\leq$ $||\tilde{\theta}_T^*-\theta_0||$:

$$T^{1/2}\mathcal{V}_T^{-1/2}\left(\tilde{\theta}_T^*-\theta_0\right)=-T^{1/2}\mathcal{V}_T^{-1/2}\left(\frac{\partial^2}{\partial\theta\partial\theta'}\tilde{\mathcal{Q}}_T^*(\mathring{\theta}_T)\right)^{-1}\frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\tilde{\mathcal{I}}_T^*(\lambda)-f(\lambda)\right)\varpi(\lambda)d\lambda.$$

Hessian consistency Lemma A.6 applies with $\mathring{\theta}_T$ since $||\mathring{\theta}_T-\theta_0||\leq||\tilde{\theta}_T^*-\theta_0||\overset{p}{\rightarrow}0$ by equivalency Lemma A.5 and consistency Theorem 2.1. Therefore by the definition $\mathcal{V}_T=\Omega^{-1}\mathcal{S}_T\Omega^{-1}$, and approximation Lemma A.7:

$$T^{1/2}\mathcal{V}_T^{-1/2}\left(\tilde{\theta}_T^*-\theta_0\right) = -T^{1/2}\mathcal{V}_T^{-1/2}\Omega^{-1}\frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\widehat{\mathcal{I}}_T^*(\lambda)-f(\lambda)\right)\varpi(\lambda)d\lambda\times(1+o_p(1))$$

$$= \quad -T^{1/2}\mathcal{S}_T^{-1/2}\frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\widehat{\mathcal{I}}_T^*(\lambda) - f(\lambda)\right)\varpi(\lambda)d\lambda \times (1 + o_p(1))$$

$$= \quad -T^{1/2}\mathcal{S}_T^{-1/2}\frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\mathcal{I}_T^*(\lambda) - f(\lambda)\right)\varpi(\lambda)d\lambda \times (1 + o_p(1)) . \quad \text{(A.12)}$$

Now define bias terms

$$\tilde{\mathcal{B}}_T \equiv \frac{1}{2\pi}\sum_{h=-b_T}^{b_T}\varpi_h\left\{E\left[\hat{\gamma}_{T,h}^*(c_{T,h})\right] - E[y_t y_{t-h}]\right\} \in \mathbb{R}^k \text{ and } \mathcal{B}_T \equiv \Omega^{-1}\frac{1}{2\pi}\tilde{\mathcal{B}}_T. \quad \text{(A.13)}$$

Use $\mathcal{V}_T^{1/2}\mathcal{S}_T^{-1/2} = \Omega^{-1}$ and apply Lemma A.9.a to deduce

$$\begin{aligned}
&T^{1/2}\mathcal{V}_T^{-1/2}\left(\tilde{\theta}_T^* - \theta_0 + \mathcal{B}_T\right) \\
&= T^{1/2}\mathcal{V}_T^{-1/2}\left(\tilde{\theta}_T^* - \theta_0\right) + T^{1/2}\mathcal{S}_T^{-1/2}\frac{1}{2\pi}\tilde{\mathcal{B}}_T \\
&= -\frac{T^{1/2}}{2\pi}\mathcal{S}_T^{-1/2}\int_{-\pi}^{\pi}\left(\mathcal{I}_T^*(\lambda) - f(\lambda)\right)\varpi(\lambda)d\lambda \times (1 + o_p(1)) + T^{1/2}\mathcal{S}_T^{-1/2}\frac{1}{2\pi}\tilde{\mathcal{B}}_T \\
&= -\frac{T^{1/2}}{2\pi}\mathcal{S}_T^{-1/2}\left\{\int_{-\pi}^{\pi}\left(\mathcal{I}_T^*(\lambda) - f(\lambda)\right)\varpi(\lambda)d\lambda - \tilde{\mathcal{B}}_T\right\}(1 + o_p(1)) + o_p(1) \xrightarrow{d} N(0, I_k) . \quad \text{(A.14)}
\end{aligned}$$

Finally, the claimed properties of $\mathcal{V}_T = \Omega^{-1}\mathcal{S}_T\Omega^{-1}$ hold by Lemma A.9.b,c. $\mathcal{QED}$.

**Proof of Theorem 3.1**. Define bias corrected spectral density and FD-QML estimators

$$\begin{aligned}
\widehat{\mathcal{I}}_T^{(bc)}(\lambda) &= \frac{1}{2\pi}\left(\widehat{\hat{\gamma}}_{T,0}^{(bc)}(\widehat{\mathcal{Y}}_{0,(k_T)}^{(0)}) + 2\sum_{h=1}^{b_T}\widehat{\hat{\gamma}}_{T,h}^{(bc)}(\widehat{\mathcal{Y}}_{h,(k_T)}^{(0)}) \times \cos(\lambda h)\right) \\
\tilde{\theta}_T^{(bc)} &\equiv \underset{\theta\in\Theta}{\arg\min}\frac{1}{2\pi}\int_{-\pi}^{\pi}\left\{\ln f(\lambda,\theta) + \widehat{\mathcal{I}}_T^{(bc)}(\lambda)/f(\lambda,\theta)\right\}d\lambda \\
\hat{\theta}_T^{(bc)} &\equiv \underset{\theta\in\Theta}{\arg\min}\sum_{j\in\mathbb{F}}\left\{\ln f(\lambda_j,\theta) + \widehat{\mathcal{I}}_T^{(bc)}(\lambda_j)/f(\lambda_j,\theta)\right\}.
\end{aligned}$$

We will prove $T^{1/2}\tilde{\mathcal{V}}_T^{-1/2}(\tilde{\theta}_T^{(bc)} - \theta_0) \xrightarrow{d} N(0, I_k)$. Then $T^{1/2}\tilde{\mathcal{V}}_T^{-1/2}(\hat{\theta}_T^{(bc)} - \theta_0) \xrightarrow{d} N(0, I_k)$ follows by the same argument used to prove Lemma A.5. The proof of $T^{1/2}\tilde{\mathcal{V}}_T^{-1/2}(\hat{\theta}_T^{(obc)} - \theta_0) \xrightarrow{d} N(0, I_k)$ is identical since $m_{T,h}(\xi) = [\xi m_{T,h}]$ for $0 < \underline{\xi} \leq \xi \leq \bar{\xi} < \infty$ and $m_{T,h}/k_T \to \infty$ imply $m_{T,h}(\xi)/k_T \sim \xi m_{T,h}/k_T \to \infty$, while also $\hat{\kappa}_{h,i,m_{T,h}(\xi)} = \kappa_{h,i} + O_p(1/m_{T,h}^{1/2})$ and $\hat{d}_{h,i,m_{T,h}(\xi)} = d_{h,i} + O_p(1/m_{T,h}^{1/2})$ by Lemma A.4 hence $\{\hat{\kappa}_{h,i,m_{T,h}(\xi)}, \hat{d}_{h,i,m_{T,h}(\xi)}\}$ have the same $m_{T,h}^{1/2}$-rate of convergence as $\{\hat{\kappa}_{h,i,m_{T,h}}, \hat{d}_{h,i,m_{T,h}}\}$. In particular, Hill (2013, proof of Theorem 2.2) shows $m_{T,h}/k_T \to \infty$ ensures $\{\hat{\kappa}_{h,i,m_{T,h}(\xi)}, \hat{d}_{h,i,m_{T,h}(\xi)}\}$ do not effect $T^{1/2}\tilde{\mathcal{V}}_T^{-1/2}(\hat{\theta}_T^{(obc)} - \theta_0)$ asymptotically uniformly in $\xi \in [\underline{\xi}, \bar{\xi}]$, hence $\{\hat{\kappa}_{h,i,m_{T,h}(\hat{\xi}_{T,h})}, \hat{d}_{h,i,m_{T,h}(\hat{\xi}_{T,h})}\}$ do not effect the limit distribution of $T^{1/2}\tilde{\mathcal{V}}_T^{-1/2}(\hat{\theta}_T^{(obc)} - \theta_0)$. The claims $\tilde{\mathcal{V}}_T = \mathcal{V}_T(1 + o(1))$ if $\kappa < 4$ and $\tilde{\mathcal{V}}_T = \mathcal{V}_T(1 + O(1))$ if $\kappa = 4$ follow from $\tilde{\mathcal{V}}_T = \Omega^{-1}\tilde{\mathcal{S}}_T\Omega^{-1}$ and the Lemma A.10 results that if

$\kappa > 4$ then $\tilde{\mathcal{S}}_T = \lim_{M\to\infty} \tilde{\mathcal{S}}_{M,T} = \mathcal{S}_T(1 + o(1))$ and if $\kappa \in (2,4]$ then $\tilde{\mathcal{S}}_T = \lim_{M\to\infty} \tilde{\mathcal{S}}_{M,T} = \mathcal{S}_T(1 + O(1))$, where $\tilde{\mathcal{S}}_{M,T}$ is defined in Lemma A.10.

**Step 1 ($\tilde{\theta}_T^{(bc)}$).** Let $\{\hat{\mathcal{R}}_{T,h}\}_{h=1}^{b_T}$ be defined as in (16)-(17), and define a bias estimator:

$$\widehat{\tilde{\mathcal{B}}}_T \equiv -\frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\hat{\mathcal{R}}_{T,0} + 2\sum_{h=1}^{b_T}\hat{\mathcal{R}}_{T,h} \times \cos(\lambda h)\right)\varpi(\lambda)d\lambda. \tag{A.15}$$

Let $\{r_{M,T}\}_{M,T\in\mathbb{N}}$ be a non-random double array of finite constants that satisfies $\sup_{T\in\mathbb{N}}|r_{M,T}| \to 0$ as $M \to \infty$, and may be different in different places.

We need the process that governs $\tilde{\theta}_T^{(bc)}$. Take $\mathfrak{I}_{T,h}^*$, $\kappa_{h,0}$ and $\mathcal{D}_{T,h}$ defined in (A.8), and define for arbitrary $M \in \mathbb{N}$:

$$\tilde{\mathcal{Z}}_{M,T} \equiv T^{1/2}\frac{1}{2\pi}\left(\left\{\int_{-\pi}^{\pi}\left(\mathcal{I}_T^*(\lambda) - f(\lambda)\right)\varpi(\lambda)d\lambda - \tilde{\mathcal{B}}_T\right\} + \frac{1}{2\pi}\sum_{h=-M}^{M}\left(1-\frac{|h|}{M}\right)\varpi_h\mathcal{D}_{T,h}\mathfrak{I}_{T,h}^*\right)$$

$$\tilde{\mathcal{S}}_{M,T} \equiv E\left[\tilde{\mathcal{Z}}_{M,T}\tilde{\mathcal{Z}}_{M,T}'\right] \text{ and } \tilde{\mathcal{V}}_T = \Omega^{-1}\tilde{\mathcal{S}}_T\Omega^{-1}.$$

By construction $\lim_{M\to\infty}\tilde{\mathcal{Z}}_{M,T} = \tilde{\mathcal{Z}}_T$ and $\lim_{M\to\infty}\tilde{\mathcal{S}}_{M,T} = \tilde{\mathcal{S}}_T$.

In view of bias definitions (A.13) and (A.15), and arguments leading to (A.12) and (A.14), use $\tilde{\mathcal{V}}_T = \Omega^{-1}\tilde{\mathcal{S}}_T\Omega^{-1}$ and approximation Lemma A.7 to deduce:

$$\begin{aligned} T^{1/2}\tilde{\mathcal{V}}_T^{-1/2}\left(\tilde{\theta}_T^{(bc)} - \theta_0\right) &= -T^{1/2}\tilde{\mathcal{S}}_T^{-1/2}\frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\widehat{\mathcal{I}}_T^{(bc)}(\lambda) - f(\lambda)\right)\varpi(\lambda)d\lambda \times (1+o_p(1)) \\ &= -T^{1/2}\tilde{\mathcal{S}}_T^{-1/2}\frac{1}{2\pi}\left(\int_{-\pi}^{\pi}\left(\mathcal{I}_T^*(\lambda) - f(\lambda)\right)\varpi(\lambda)d\lambda - \tilde{\mathcal{B}}_T\right) \times (1+o_p(1)) \\ &\quad + T^{1/2}\tilde{\mathcal{S}}_T^{-1/2}\frac{1}{2\pi}\left(\widehat{\tilde{\mathcal{B}}}_T - \tilde{\mathcal{B}}_T\right) \times (1+o_p(1)). \end{aligned} \tag{A.16}$$

In Step 2 we show $\widehat{\tilde{\mathcal{B}}}_T$ is, asymptotically, a linear function of the indicator partial sum $\mathfrak{I}_{T,h}^*$:

$$\begin{aligned} &T^{1/2}\tilde{\mathcal{S}}_T^{-1/2}\left(\widehat{\tilde{\mathcal{B}}}_T - \tilde{\mathcal{B}}_T\right) \\ &\quad = -\frac{1}{2\pi}T^{1/2}\tilde{\mathcal{S}}_T^{-1/2}\sum_{h=-M}^{M}\left(1-\frac{|h|}{M}\right)\varpi_h\mathcal{D}_{T,h} \times \mathfrak{I}_{T,h}^* \times (1+o_p(1)) + o_p(r_{M,T}). \end{aligned} \tag{A.17}$$

Combine (A.16) and (A.17) to obtain by the construction of $\tilde{\mathcal{Z}}_{M,T}$:

$$\begin{aligned} T^{1/2}\tilde{\mathcal{V}}_T^{-1/2}\left(\tilde{\theta}_T^{(bc)} - \theta_0\right) &= -T^{1/2}\tilde{\mathcal{S}}_T^{-1/2}\frac{1}{2\pi}\left(\int_{-\pi}^{\pi}\left(\mathcal{I}_T^*(\lambda) - f(\lambda)\right)\varpi(\lambda)d\lambda - \tilde{\mathcal{B}}_T\right) \times (1+o_p(1)) \\ &\quad -\frac{1}{(2\pi)^2}T^{1/2}\tilde{\mathcal{S}}_T^{-1/2}\sum_{h=-M}^{M}\left(1-\frac{|h|}{M}\right)\varpi_h\mathcal{D}_{T,h} \times \mathfrak{I}_{T,h}^* \times (1+o_p(1)) + o_p(r_{M,T}) \\ &= -\left(\tilde{\mathcal{S}}_T^{-1/2}\tilde{\mathcal{S}}_{M,T}^{1/2}\right) \times \tilde{\mathcal{S}}_{M,T}^{-1/2}\tilde{\mathcal{Z}}_{M,T} \times (1+o_p(1)) + o_p(r_{M,T}). \end{aligned}$$

By Lemma A.10: $\tilde{\mathcal{S}}_{M,T}^{-1/2}\tilde{\mathcal{Z}}_{M,T} \xrightarrow{d} N(0, I_k)$. Since $M$ can be made arbitrarily large we can always impose $M \to \infty$ as $T \to \infty$ such that $\lim_{T\to\infty} \tilde{\mathcal{S}}_T^{-1/2}\tilde{\mathcal{S}}_{M,T}^{1/2} = I_k$ and $\lim_{T\to\infty} r_{M,T} = 0$.

**Step 2 ($\hat{\tilde{\mathcal{B}}}_T$).** We now verify (A.17). Define $\tilde{\mathcal{R}}_{T,h} \equiv -(E[\hat{\gamma}^*_{T,h}(c_{T,h})] - E[y_t y_{t-h}])$, and recall by constructions (13) and (14) that $\mathcal{R}_{T,h} \sim \tilde{\mathcal{R}}_{T,h}$. Then, by Fejer's theorem, we can write:

$$
\begin{aligned}
\tilde{\mathcal{B}}_T &= \frac{1}{2\pi}\sum_{h=-M}^{M}\left(1 - \frac{|h|}{M}\right)\varpi_h\{E\left[\hat{\gamma}^*_{T,h}(c_{T,h})\right] - E[y_t y_{t-h}]\} + o_p(r_{M,T}) \\
&= -\frac{1}{2\pi}\sum_{h=-M}^{M}\left(1 - \frac{|h|}{M}\right)\varpi_h\tilde{\mathcal{R}}_{T,h} + o_p(r_{M,T}) \\
\hat{\tilde{\mathcal{B}}}_T &= -\frac{1}{2\pi}\sum_{h=-M}^{M}\left(1 - \frac{|h|}{M}\right)\varpi_h\hat{\mathcal{R}}_{T,h} + o_p(r_{M,T}).
\end{aligned}
$$

By arguments in Peng (2001, p. 259-263) for second order Paretian tails, and Assumption B$'$, we have $T^{1/2}\sum_{h=-M}^{M}\varpi_h\{\tilde{\mathcal{R}}_{T,h} - \mathcal{R}_{T,h}\} = o(1)$, hence with $\tilde{\mathcal{B}}_T$ and $\hat{\tilde{\mathcal{B}}}_T$ above it follows:[9]

$$
T^{1/2}\tilde{\mathcal{S}}_T^{-1/2}\left(\hat{\tilde{\mathcal{B}}}_T - \tilde{\mathcal{B}}_T\right) = -\frac{1}{2\pi}T^{1/2}\tilde{\mathcal{S}}_T^{-1/2}\sum_{h=-M}^{M}\left(1 - \frac{|h|}{M}\right)\varpi_h\left\{\hat{\mathcal{R}}_{T,h} - \mathcal{R}_{T,h}\right\} + o_p(r_{M,T}). \quad \text{(A.18)}
$$

Next, with $\mathcal{D}_{T,h}$ defined in (A.8), we show:

$$
\hat{\mathcal{R}}_{T,h} - \mathcal{R}_{T,h} = \mathcal{D}_{T,h} \times k_T^{1/2}\left(\widehat{\mathcal{Y}}^{(0)}_{h,(k_T)}/c_{T,h} - 1\right) \times (1 + o_p(1)). \quad \text{(A.19)}
$$

By imitating arguments in Hill (2013: proof of Theorem 2.1), $m_{T,h}/k_T \to \infty$ can be shown to imply the tail index estimators $\hat{\kappa}_{h,i,m_{T,h}}$ and $\hat{d}_{h,i,m_{T,h}}$ in $\hat{\mathcal{R}}_{T,m}$ do not affect asymptotics since they are $m_{T,h}^{1/2}$-consistent under Assumptions A and B$'$ by Lemma A.4, while by Lemma A.3 $\widehat{\mathcal{Y}}^{(0)}_{h,(k_T)}$ is $k_T^{1/2}$-consistent under Assumptions A.1, B and C, and Assumption B$'$ implies Assumption B. We therefore simply write, without loss of generality,

$$
\begin{aligned}
\hat{\mathcal{R}}_{T,0} &= \frac{1}{\kappa_{0,0} - 1}\left(\frac{k_T}{T}\right)\widehat{\mathcal{Y}}^{(0)}_{0,(k_T)} = \mathcal{R}_{T,0} + \frac{1}{\kappa_{0,0} - 1}\frac{k_T^{1/2}}{T}c_{T,0} \times k_T^{1/2}\left(\frac{\widehat{\mathcal{Y}}^{(0)}_{0,(k_T)}}{c_{T,0}} - 1\right) \\
\hat{\mathcal{R}}_{T,h} &= \frac{(T-h)^2}{T(T-h-k_T)}\left(\frac{d_{h,2}\left(\widehat{\mathcal{Y}}^{(0)}_{h,(k_T)}\right)^{1-\kappa_{h,2}}}{\kappa_{h,2} - 1} - \frac{d_{h,1}\left(\widehat{\mathcal{Y}}^{(0)}_{h,(k_T)}\right)^{1-\kappa_{h,1}}}{\kappa_{h,1} - 1}\right) \text{ for } h > 0.
\end{aligned}
$$

A first order expansions around $c_{T,h}$ implies for some sequence of positive random numbers $\{c^*_{T,h}\}$

[9] Peng (2001) assumes $P(|y_t y_{t-h} - \tilde{\gamma}_h| \geq c) = d_{h,0}c^{-\kappa_{h,0}}(1 + O(c^{-e_{h,0}}))$ with $m_{T,h} = o(T^{2e_{h,0}/(2e_{h,0}+\kappa_{h,0})})$, but Peng (2001, p. 259-263)'s arguments extend to the other Assumption B$'$ class $P(|y_t y_{t-h} - \tilde{\gamma}_h| \geq c) = d_{h,0}c^{-\kappa_{h,0}}(1 + O(\ln(c)^{-e_{h,0}})$ with $m_{T,h} = o(\ln(T)^{2e_{h,0}})$, cf. Haeusler and Teugels (1985, Section 5).

where $|\widehat{\mathcal{Y}}^{(0)}_{h,(k_T)} - c^*_{T,h}| \leq |\widehat{\mathcal{Y}}^{(0)}_{h,(k_T)} - c_{T,h}|$:

$$\hat{\mathcal{R}}_{T,h} = \mathcal{R}_{T,h} + \frac{(T-h)^2/k_T^{1/2}}{T(T-h-k_T)} \left( d_{h,1} c_{T,h}^{1-\kappa_{h,1}} \left( \frac{c_{T,h}}{c^*_{T,h}} \right)^{\kappa_{h,1}} - d_{h,2} c_{T,h}^{1-\kappa_{h,2}} \left( \frac{c_{T,h}}{c^*_{T,h}} \right)^{\kappa_{h,2}} \right) k_T^{1/2} \left( \frac{\widehat{\mathcal{Y}}^{(0)}_{h,(k_T)}}{c_{T,h}} - 1 \right)$$
$$= \mathcal{R}_{T,h} + \frac{1}{k_T^{1/2}} \left( d_{h,1} c_{T,h}^{1-\kappa_{h,1}} - d_{h,2} c_{T,h}^{1-\kappa_{h,2}} \right) \times k_T^{1/2} \left( \frac{\widehat{\mathcal{Y}}^{(0)}_{h,(k_T)}}{c_{T,h}} - 1 \right) \times (1 + o_p(1)) .$$

The last equality exploits $c^*_{T,h}/c_{T,h} \xrightarrow{p} 1$ given $|\widehat{\mathcal{Y}}^{(0)}_{h,(k_T)} - c^*_{T,h}| \leq |\widehat{\mathcal{Y}}^{(0)}_{h,(k_T)} - c_{T,h}|$ and $\widehat{\mathcal{Y}}^{(0)}_{h,(k_T)}/c_{T,h} \xrightarrow{p} 1$ by Lemma 1 in Hill (2010). Hence we have shown (A.19).

Finally, combine (A.18) and (A.19), with $k_T^{1/2}(\widehat{\mathcal{Y}}^{(0)}_{h,(k_T)}/c_{T,h} - 1) = \mathfrak{I}^*_{T,h}(1 + o_p(1))$ by Lemma A.3, to deduce (A.17). $\mathcal{QED}$.

## References


Aguilar, M., and J. B. Hill (2014): “Robust Score and Portmanteau Tests of Volatility Spillover,” *Journal of Econometrics*, forthcoming.

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

Andrews, D., P. Bickel, F. Hampel, P. Huber, W. Rogers, and J. Tukey (1972): *Robust Estimates of Location.* Princeton University Press.

Bartkiewicz, K., A. Jakubowski, T. Mikosch, and O. Wintenberger (2010): “Stable limits for sums of dependent infinite variance random variables,” *Probability Theory and Related Fields*, 150, 337–372.

Basrak, B., R. Davis, and T. Mikosch (2002): “Regular Variation of GARCH Processes,” *Stochastic Processes and their Applications*, 99, 95–115.

Bradley, R. (1992): “On the Spectral Density and Asymptotic Normality of Weakly Dependent Random Fields,” *Journal of Theoretical Probability*, 5, 355–373.

——— (1993): “Equivalent Mixing Conditions for Random Fields,” *Annals of Probability*, 21, 1921–1926.

Brockwell, P., and D. Cline (1985): “Linear Prediction of ARMA Processes with Infinite Variance,” *Stochastic Processes and their Applications*, 19, 281–296.

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

Chaudhuri, S., and J. B. Hill (2014): “Heavy Tail Robust Estimation and Inference for Average Treatment Effects,” *Working Paper, University of North Carolina, Dept. of Economics.*

Chiu, S. T. (1988): “Weighted least squares estimators on the frequency domain for the parameters of a time series,” *Annals of Statistics*, 16, 1315–1326.

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

Csörgő, S., L. Horváth, and D. Mason (1986): "What Portion of the Sample Makes a Partial Sum Asymptotically Stable or Normal?," *Probability Theory and Related Fields*, 72, 1–16.

Davis, R. A., and T. Mikosch (2009): "Probabilistic Properties of Stochastic Volatility Models," in *Handbook of Financial Time Series*, ed. by t. Mikosch, J.-P. Kreiss, R. A. Davis, and T. G. Andersen. Springer.

Dehling, H., M. Denker, and W. Phillip (1986): "Central Limit Theorems for Mixing Sequences of Random Variables Under Minimal Conditions," *Annals of Probability*, 14, 1359–1370.

Dette, H., M. Hallin, T. Kley, and S. Volgushev (2011): "Of Copulas, Quantiles, Ranks and Spectra: an L1-Approach to Spectral Analysis," *Technical Report: Ruhr-University at Bochum.*

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

Drees, H., L. de Haan, and S. Resnick (2000): "How to Make a Hill Plot," *Annals of Statistics*, 28, 254–274.

Dunsmuir, W. (1979): "A Central Limit Theorem for Parameter Estimation in Stationary Vector Time Series and its Application to Models for a Signal Observed with Noise," *Annals of Statistics*, 7, 490–506.

Dunsmuir, W., and E. J. Hannan (1976): "Vector Linear Time Series Models," *Advances in Applied Probability*, 8, 339–364.

Genton, M. G., and E. Ronchetti (2003): "Robust Indirect Inference," *Journal of the American Statistical Association*, 98, 67–76.

Goldie, C. M., and R. L. Smith (1987): "Slow Variation with Remainder: Theory and Applications," *Quarterly Journal of Mathematics*, 38, 45–71.

Granger, C. W. J. (1966): "The Typical Spectral Shape of an Economic Variable," *Econometrica*, 34, 150–161.

Haeusler, E., and J. Teugels (1985): "On Asymptotic Normality of Hill's Estimator for the Exponent of Regular Variation," *Annals of Statistics*, 13, 743–756.

Hagemann, A. (2012): "Robust Spectral Analysis," *Working Paper, Notre Dame University, Dept. of Economics.*

Hall, P. (1982): "On Some Simple Estimates of an Exponent of Regular Variation," *Journal of the Royal Statistical Societ Series B*, 44, 37–42.

Hampel, F. (1974): "The Influence Curve and Its Role in Robust Estimation," *Journal of the American Statistical Association*, 69, 383–393.

Hampel, F., E. Ronchetti, P. Rousseeuw, and W. Stahel (1986): *Robust Statistics: The Approach Based on Influence Functions.* Wiley, New York.

Hannan, E. (1973a): "The Asymptotic Theory of Linear Time-Series Models," *Journal of Applied Probability*, 10, 130–145.

——— (1973b): "The Estimation of Frequency," *Journal of Applied Probability*, 10, 510–519.

Hansen, B., E. (1991): "Strong Laws for Dependent Heterogeneous Processes," *Econometric Theory*, 7, 213–221.

——— (1992): "Erratum: Strong Laws for Dependent Heterogeneous Processes," *Econometric Theory*, 8, 421–422.

Hill, B. M. (1975): "A Simple General Approach to Inference about the Tail of a Distribution," *Annals of Statistics*, 3(5), 1163–1174.

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

——— (2010): "On Tail Index Estimation for Dependent, Heterogeneous Data," *Econometric Theory*, 26, 1398–1436.

——— (2012a): "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.

——— (2012b): "Least Tail-Trimmed Squares for Infinite Variance Autoregressions," *Journal of Time Series Analysis*, 34, 168–186.

——— (2013): "Expected Shortfall Estimation and Gaussian Inference for Infinite Variance Time Series," *Journal of Financial Econometrics*, forthcoming.

——— (2014a): "Robust Estimation and Inference for Heavy-Tailed GARCH," *Bernoulli*, forthcoming.

——— (2014b): "Tail Index Estimation for a Filtered Dependent Time Series," *Statistica Sinica*, forthcoming.

Hill, J. B., and M. Aguilar (2013): "Moment Condition Tests for Heavy Tailed Time Series," *Journal of Econometrics*, 172, 255–274.

Hill, J. B., and A. McCloskey (2014): "Supplemental Material for "Heavy Tail Robust Frequency Domain Estimation"," *Dept. of Economics, University of North Carolina, Chapel Hill.*

Hosoya, Y., and M. Taniguchi (1982): "A Central Limit Theorem for Stationary Processes and the Parameter Estimation of Linear Processes," *Annals of Statistics*, 10, 132–153.

Hsing, T. (1991): "On Tail Index Estimation Using Dependent Data," *Annals of Statistics*, 19, 1547–1569.

Huber, P. (1964): "Robust Estimation of a Location Parameter," *Annals of Mathematical Statistics*, 35, 73–101.

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

——— (1975): "A Note on the Central Limit Theorem for Dependent Random Variables," *Theory of Probability and its Applications*, 20, 135–141.

Khan, S., and E. Tamer (2010): "Irregular Identification, Support Conditions, and Inverse Weight Estimation," *Econometrica*, 78, 2021–2042.

Kirch, C., and D. N. Politis (2011): "TFT-Bootstrap: Resampling Time Series in the Frequency Domain to Obtain Replicates in the Time Domain," *Annals of Statistics*, 31, 47–59.

Kreiss, J.-P., and E. Paparoditis (2012): "The Hybrid Wild Bootstrap for Time Series," *Journal of the American Statistical Association*, 107, 1073–1084.

Lee, J., and S. S. Rao (2011): "The Quantile Spectral Density and Comparison Based Tests for Nonlinear Time Series," *Technical Report : Dept. of Statistics, Texas A&M Univ.*

Li, T.-H. (2008): "Laplace Periodogram for Time Series Analysis," *Journal of the American Statistical Association*, 103, 757–768.

——— (2010): "A Nonlinear Method for Robust Spectral Analysis," *IEEE Transactions on Signal Processing*, 58, 2466–2474.

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

Mancini, L., E. Ronchetti, and F. Trojani (2005): "Optimal Conditionally Unbiased Bounded-Influence Inference in Dynamic Location and Scale Models," *Journal of the American Statistical Association*, 100, 628–641.

Martin, R. D., and D. J. Thomson (1982): "Robust-Resistant Spectrum Estimation," *Proceedings of the IEEE*, 70, 1097–1115.

McCloskey, A. (2013): "Estimation of the Long-Memory Stochastic Volatility Model that is Robust to Level Shifts and Deterministic Trends," *Journal of Time Series Analysis*, 34, 285–301.

McCloskey, A., and J. B. Hill (2014): "Parameter Estimation Robust to Low-Frequency Contamination," *Working Paper, Brown University & University of North Carolina.*

McCloskey, A., and P. Perron (2013): "Memory Parameter Estimation in the Presence of Level Shifts and Deterministic Trends," Forthcoming in Econometric Theory.

McLeish, D. (1975): "A Maximal Inequality and Dependent Strong Laws," *Annals of Probability*, 3, 829–839.

Meitz, M., and P. Saikkonen (2008): "Stability of Nonlinear AR-GARCH Models," *Journal of Time Series Analysis*, 29, 453–475.

Mikosch, T., T. Gadrich, C. Kluppelberg, and R. Adler (1995): "Parameter Estimation for ARMA Models with Infinite Variance Innovations," *Annals of Statistics*, 23, 305–326.

Mikosch, T., and C. Starica (2000): "Limit Theory for the Sample Autocorrelations and Extremes of a GARCH(1,1) Process," *Annals of Statistics*, 28, 1427–1451.

Mikosch, T., and Y. Zhao (2014): "A Fourier Analysis of Extreme Events," *Bernoulli*, 20, 803–845.

Peligrad, M. (1996): "On the Asymptotic Normality of Sequences of Weak Dependent Random Variables," *Journal of Theoretical Probability*, 9, 703–715.

Peng, L. (2001): "Estimating the Mean of a Heavy Tailed Distribution," *Statistics and Probability Letters*, 52, 255–264.

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

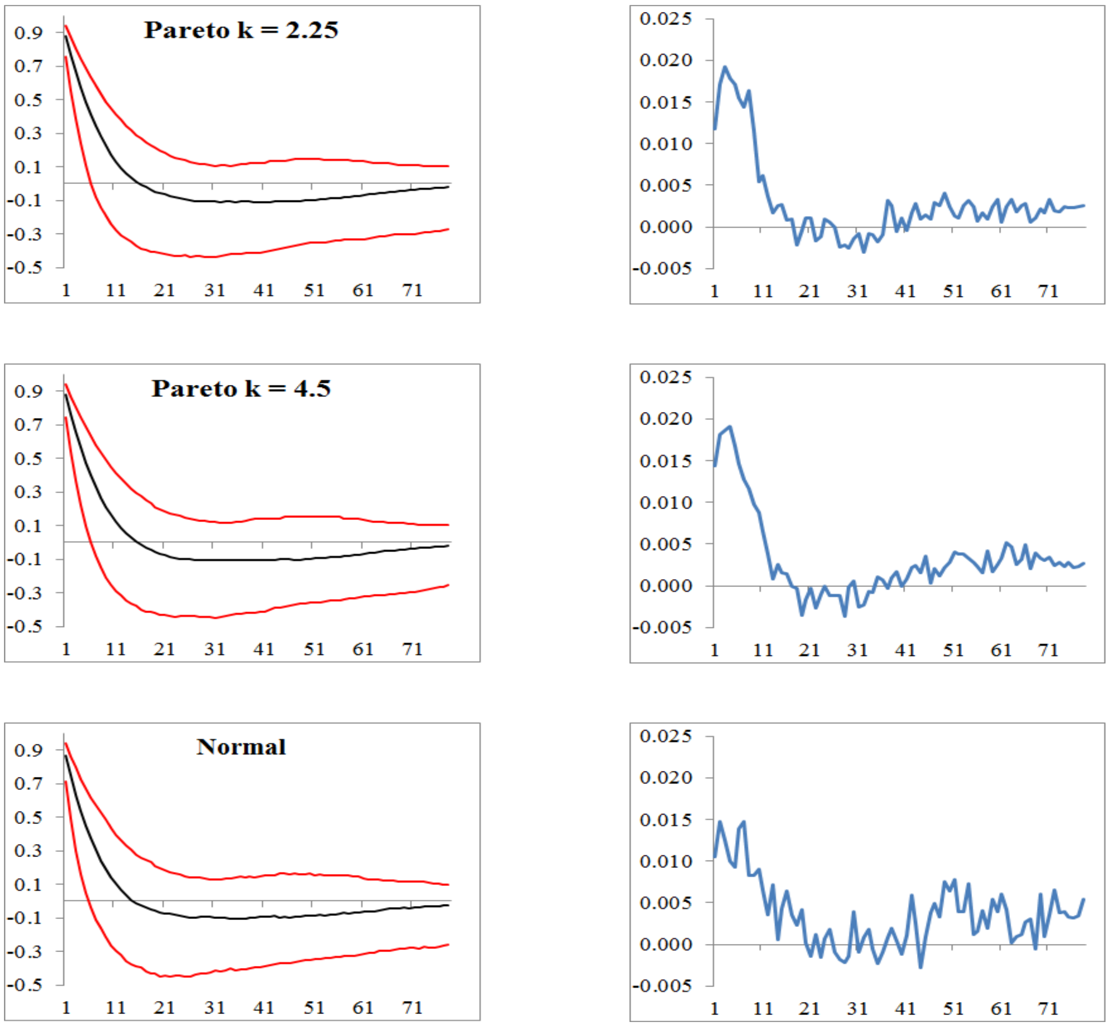
Qu, Z., and D. Tkachenko (2012): "Identification and Frequency Domain Quasi-Maximum Likelihood Estimation of Linearized Dynamic Stochastic General Equilibrium Models," *Quantitative Economics*, 3, 95–132.

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

Shao, X., and W. Wu (2007): "Asymptotic Spectral Theory for Nonlinear Time Series," *Annals of Statistics*, 35, 1773–1801.

Whittle, P. (1953): "Estimation and Information in Stationary Time Series," *Arkiv for Matematik*, 2, 423–434.

Zygmund, A. (2002): *Trigonometric Series, Vol.'s I and II.* Cambridge Univ. Press, Cambridge.




(a) (b)

Figure 1: Tail-trimmed correlations for AR(1). The process is $y_t = .9y_{t-1} + \epsilon_t$, $\epsilon$ is iid Pareto with tail index $\kappa = 2.5$, the sample size is $T = 100$, and the number of samples is $10,000$. Panel (a) contains optimally bias-corrected tail-trimmed correlations: simulation 2.5%, 50% and 97.5% quantiles (bottom, middle, top lines). Panel (b) contains the simulation average difference between optimally bias-corrected tail-trimmed and untrimmed correlations.

**Table 1 : FD-QML for AR(1)**

$\phi_0 = .90$ (where $\sigma_0^2 = 1$)[a]

**T = 100**

| | Bias | Med[b] | RMSE | $KS^c_{.05}$ |
|---|---|---|---|---|
| Pareto Error: $\kappa = 2.25$ | | | | |
| no trim[d] | -.0286 | .8801 | .0690 | 1.593 |
| trim-obc | -.0212 | .8850 | .0697 | 1.108 |
| Pareto Error: $\kappa = 4.50$ | | | | |
| no trim | -.0224 | .8862 | .0653 | 1.638 |
| trim-obc | -.0193 | .8913 | .0657 | 1.213 |
| Normal Error | | | | |
| no trim | -.0289 | .8756 | .0674 | 1.075 |
| trim-obc | -.0272 | .8890 | .0672 | 1.117 |

**T = 250**

| | Bias | Med | RMSE | $KS_{.05}$ |
|---|---|---|---|---|
| Pareto Error: $\kappa = 2.25$ | | | | |
| no trim | -.0100 | .8956 | .0337 | 1.901 |
| trim-obc | -.0106 | .8918 | .0358 | 1.055 |
| Pareto Error: $\kappa = 4.50$ | | | | |
| no trim | -.0109 | .8925 | .0333 | 1.423 |
| trim-obc | -.0103 | .8938 | .0344 | 1.232 |
| Normal Error | | | | |
| no trim | -.0093 | .8942 | .0331 | 1.296 |
| trim-obc | -.0094 | .8952 | .0328 | 1.174 |

**T = 500**

| | Bias | Med | RMSE | $KS_{.05}$ |
|---|---|---|---|---|
| Pareto Error: $\kappa = 2.25$ | | | | |
| no trim | -.0053 | .8960 | .0211 | 1.892 |
| trim-obc | -.0051 | .8964 | .0240 | .7139 |
| Pareto Error: $\kappa = 4.50$ | | | | |
| no trim | -.0052 | .8970 | .0219 | 1.256 |
| trim-obc | -.0049 | .8967 | .0210 | .9391 |
| Normal Error | | | | |
| no trim | -.0063 | .8954 | .0216 | .9492 |
| trim-obc | -.0058 | .8954 | .0214 | .6528 |

**T = 1000**

| | Bias | Med | RMSE | $KS_{.05}$ |
|---|---|---|---|---|
| Pareto Error: $\kappa = 2.25$ | | | | |
| no trim | -.0021 | .8982 | .0141 | 1.835 |
| trim-obc | -.0024 | .8981 | .0153 | .7857 |
| Pareto Error: $\kappa = 4.50$ | | | | |
| no trim | -.0027 | .8982 | .0144 | .9266 |
| trim-obc | -.0025 | .8976 | .0142 | .5915 |
| Normal Error | | | | |
| no trim | -.0023 | .8982 | .0143 | .6356 |
| trim-obc | -.0023 | .8987 | .0147 | .5647 |

$\sigma_0^2 = 1$ (where $\phi_0 = .90$)

**T = 100**

| | Bias | Med[b] | RMSE | $KS^c_{.05}$ |
|---|---|---|---|---|
| Pareto Error: $\kappa = 2.25$ | | | | |
| no trim[d] | .0164 | 1.012 | .2682 | 1.817 |
| trim-obc | .0137 | .9967 | .2954 | 1.532 |
| Pareto Error: $\kappa = 4.50$ | | | | |
| no trim | .0332 | 1.022 | .1098 | 1.638 |
| trim-obc | .0278 | 1.019 | .1408 | 1.335 |
| Normal Error | | | | |
| no trim | .0449 | 1.037 | .1725 | 1.543 |
| trim-obc | .0399 | 1.018 | .2134 | 1.346 |

**T = 250**

| | Bias | Med | RMSE | $KS_{.05}$ |
|---|---|---|---|---|
| Pareto Error: $\kappa = 2.25$ | | | | |
| no trim | .0195 | 1.016 | .2132 | 1.991 |
| trim-obc | .0129 | 1.013 | .2469 | 1.352 |
| Pareto Error: $\kappa = 4.50$ | | | | |
| no trim | .0111 | 1.009 | .0541 | 1.106 |
| trim-obc | .0121 | 1.010 | .0601 | 1.216 |
| Normal Error | | | | |
| no trim | .0140 | 1.011 | .0960 | .4722 |
| trim-obc | .0137 | 1.012 | .1019 | .5627 |

**T = 500**

| | Bias | Med | RMSE | $KS_{.05}$ |
|---|---|---|---|---|
| Pareto Error: $\kappa = 2.25$ | | | | |
| no trim | .0022 | 1.007 | .1932 | 1.892 |
| trim-obc | -.0087 | 1.005 | .2063 | 1.267 |
| Pareto Error: $\kappa = 4.50$ | | | | |
| no trim | .0067 | 1.005 | .0364 | 1.085 |
| trim-obc | .0066 | 1.004 | .0382 | 1.002 |
| Normal Error | | | | |
| no trim | .0077 | 1.006 | .0646 | .9492 |
| trim-obc | .0057 | 1.005 | .0452 | .6755 |

**T = 1000**

| | Bias | Med | RMSE | $KS_{.05}$ |
|---|---|---|---|---|
| Pareto Error: $\kappa = 2.25$ | | | | |
| no trim | .0024 | 1.002 | .1906 | 2.143 |
| trim-obc | .0034 | .9959 | .2152 | 1.214 |
| Pareto Error: $\kappa = 4.50$ | | | | |
| no trim | .0032 | 1.002 | .0277 | .8934 |
| trim-obc | .0035 | 1.003 | .0311 | .7862 |
| Normal Error | | | | |
| no trim | .0052 | 1.005 | .0466 | .6356 |
| trim-obc | .0017 | 1.002 | .0432 | .7653 |

a. The model is $y_t = \phi_0 y_{t-1} + \epsilon_t$ where $\epsilon_t$ is iid Pareto or Normal, $E[\epsilon_t] = 0$ and $\sigma_0^2 \equiv E[\epsilon_t^2]$. b. "Med" is the median, and "RMSE" is the root-mean-squared-error.
c. "$KS_{.05}$" is the ratio of Kolmogorov-Smirnov statistic divided by its 5% critcal value. Values *greater* than one suggest non-normality at the 5% level.
d. "no-trim" is standard FD-QML. "trim-obc" is the optimal bias corrected tail-trimmed FD-QML

**Table 2 : FD-QML for GARCH(1,1)[a]**

$[\alpha_0, \beta_0] = [.3, .6]$ : $\kappa \approx 2.05$

| $\alpha_0 = .3$ | | | | | | | | | | $\beta_0 = .6$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **T = 100** | | | | | **T = 250** | | | | | **T = 100** | | | | | **T = 250** | | | | |
| | Bias | Med[b] | RMSE | $KS^c_{.05}$ | | Bias | Med | RMSE | $KS_{.05}$ | | Bias | Med | RMSE | $KS_{.05}$ | | Bias | Med | RMSE | $KS_{.05}$ |
| no trim[d] | -.0542 | .2116 | .1386 | 2.546 | no trim | -.0561 | .2202 | .1202 | 2.395 | no trim | .0174 | .6532 | .2321 | 1.832 | no trim | -.0142 | .6304 | .1880 | 1.867 |
| trim-obc | .0112 | .2461 | .2022 | 1.752 | trim-obc | -.0260 | .2232 | .1728 | 1.673 | trim-obc | .0267 | .6538 | .2152 | 1.452 | trim-obc | .0131 | .6655 | .1879 | 1.411 |
| **T = 500** | | | | | **T = 1000** | | | | | **T = 500** | | | | | **T = 1000** | | | | |
| | Bias | Med | RMSE | $KS_{.05}$ | | Bias | Med | RMSE | $KS_{.05}$ | | Bias | Med | RMSE | $KS_{.05}$ | | Bias | Med | RMSE | $KS_{.05}$ |
| no trim | -.0528 | .2233 | .1087 | 2.883 | no trim | -.0378 | .2402 | .0977 | 2.727 | no trim | .0036 | .6391 | .1629 | 2.883 | no trim | -.0028 | .6277 | .1465 | 2.160 |
| trim-obc | .0004 | .2449 | .1688 | 1.241 | trim-obc | -.0047 | .2445 | .1561 | 1.572 | trim-obc | .0014 | .6482 | .1725 | 1.346 | trim-obc | .0015 | .6445 | .1435 | 1.172 |

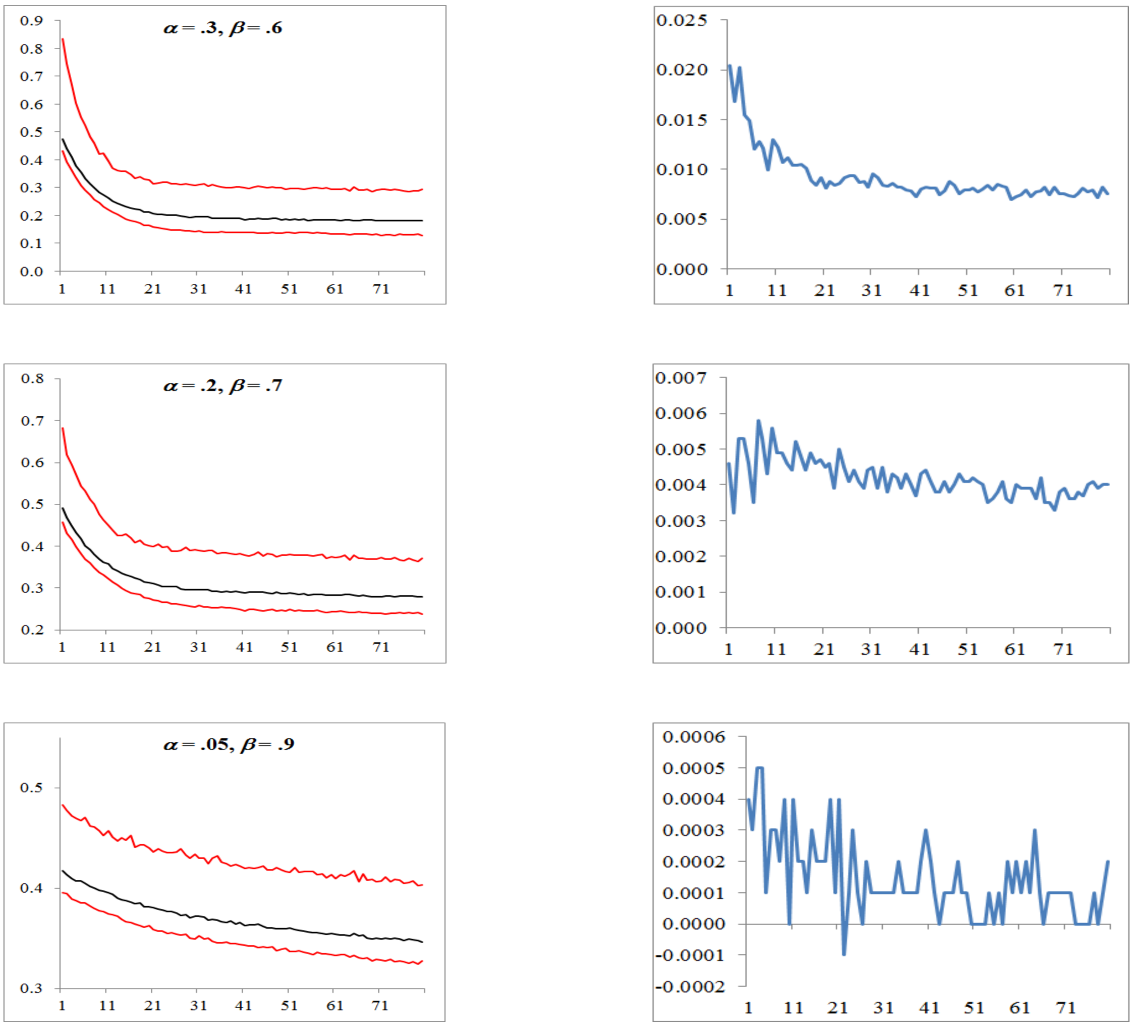
$[\alpha_0, \beta_0] = [.2, .7]$ : $\kappa \approx 3.02$

| $\alpha_0 = .2$ | | | | | | | | | | $\beta_0 = .7$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **T = 100** | | | | | **T = 250** | | | | | **T = 100** | | | | | **T = 250** | | | | |
| | Bias | Med | RMSE | $KS_{.05}$ | | Bias | Med | RMSE | $KS_{.05}$ | | Bias | Med | RMSE | $KS_{.05}$ | | Bias | Med | RMSE | $KS_{.05}$ |
| no trim | .0644 | .2162 | .2182 | 2.505 | no trim | .0597 | .2158 | .1983 | 2.331 | no trim | -.0646 | .7082 | .3210 | 3.321 | no trim | -.0005 | .7234 | .2219 | 2.708 |
| trim-obc | .0622 | .2201 | .2432 | 2.142 | trim-obc | .0602 | .2150 | .2208 | 1.978 | trim-obc | -.0770 | .6924 | .3165 | 2.856 | trim-obc | .0010 | .7306 | .2366 | 2.154 |
| **T = 500** | | | | | **T = 1000** | | | | | **T = 500** | | | | | **T = 1000** | | | | |
| | Bias | Med | RMSE | $KS_{.05}$ | | Bias | Med | RMSE | $KS_{.05}$ | | Bias | Med | RMSE | $KS_{.05}$ | | Bias | Med | RMSE | $KS_{.05}$ |
| no trim | .0620 | .2159 | .1877 | 2.568 | no trim | .0619 | .2038 | .1857 | 3.137 | no trim | -.0058 | .7104 | .1936 | 2.706 | no trim | .0123 | .7183 | .1573 | 3.077 |
| trim-obc | .0502 | .2094 | .1969 | 1.946 | trim-obc | .0499 | .2078 | .1867 | 2.442 | trim-obc | .0095 | .7220 | .1969 | 2.111 | trim-obc | .0159 | .7149 | .1670 | 2.206 |

$[\alpha_0, \beta_0] = [.05, .9]$ : $\kappa \approx 4.05$

| $\alpha_0 = .05$ | | | | | | | | | | $\beta_0 = .9$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **T = 100** | | | | | **T = 250** | | | | | **T = 100** | | | | | **T = 250** | | | | |
| | Bias | Med | RMSE | $KS_{.05}$ | | Bias | Med | RMSE | $KS_{.05}$ | | Bias | Med | RMSE | $KS_{.05}$ | | Bias | Med | RMSE | $KS_{.05}$ |
| no trim | .1032 | .1002 | .1889 | 4.361 | no trim | .0624 | .0677 | .1449 | 4.377 | no trim | -.2062 | .8170 | .4031 | 4.221 | no trim | -.1663 | .8587 | .3741 | 5.032 |
| trim-obc | .1175 | .1256 | .2242 | 2.790 | trim-obc | .0663 | .1027 | .1533 | 2.167 | trim-obc | -.2076 | .7893 | .3884 | 2.560 | trim-obc | -.1123 | .8626 | .2856 | 2.530 |
| **T = 500** | | | | | **T = 1000** | | | | | **T = 500** | | | | | **T = 1000** | | | | |
| | Bias | Med | RMSE | $KS_{.05}$ | | Bias | Med | RMSE | $KS_{.05}$ | | Bias | Med | RMSE | $KS_{.05}$ | | Bias | Med | RMSE | $KS_{.05}$ |
| no trim | .0501 | .0674 | .1212 | 3.960 | no trim | .0408 | .0632 | .1022 | 4.475 | no trim | -.0951 | .8858 | .2835 | 5.598 | no trim | -.0371 | .8955 | .1827 | 5.552 |
| trim-obc | .0478 | .0890 | .1064 | 1.732 | trim-obc | .0266 | .0758 | .0688 | 1.378 | trim-obc | -.0348 | .8990 | .1812 | 1.795 | trim-obc | .0110 | .9188 | .0903 | 1.104 |

a. The model is $y_t = x_t^2 = \sigma_t^2 \epsilon_t^2$ where $\sigma_t^2 = 1 + \alpha_0 x_{t-1}^2 + \beta_0 \sigma_{t-2}^2$ where $\epsilon_t$ is iid standard normal. b. "Med" is the median, and "RMSE" is the root-mean-squared-error.
c. "$KS_{.05}$" is the ratio of Kolmogorov-Smirnov statistic divided by its 5% critical value. Values *greater* than one suggest non-normality at the 5% level.
d. "no-trim" is standard FD-QML. "trim-obc" is the optimal bias corrected tail-trimmed FD-QML.




(a) (b)

Figure 2: Tail-trimmed correlations for GARCH(1,1). The process is $y_t = x_t^2$ where $x_t = .9\sigma_{t-1}\epsilon_t$, with $\sigma_{t-1}^2 = 1 + \alpha_0 x_{t-1}^2 + \beta_0 \sigma_{t-1}^2$, and $\epsilon$ is iid $N(0,1)$, the sample size is $T = 100$, and the number of samples is $10,000$. Panel (a) contains optimally bias-corrected tail-trimmed correlations: simulation 2.5%, 50% and 97.5% quantiles (bottom, middle, top lines). Panel (b) contains the simulation average difference between optimally bias-corrected tail-trimmed and untrimmed correlations.