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.

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 \( f(\cdot, \theta) \) 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 \( E[y_t^4] < \infty \). 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:

\[
w_T(\lambda) \equiv \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^T y_t e^{-i\lambda t} \quad \text{and} \quad I_T(\lambda) \equiv |w_T(\lambda)|^2.
\]

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)

\[
\arg \min_{\theta \in \Theta} \sum_{j \in F} \left( \ln f(\lambda_j, \theta) + \frac{I_T(\lambda_j)}{f(\lambda_j, \theta)} \right) \quad \text{where} \quad F \equiv (-T/2, T/2] \cap \mathbb{Z} \setminus \{0\},
\]

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); and robust moment estimators (see, e.g., Khan and Tamer, 2010; Hill, 2013; Chaudhuri and Hill, 2014).

We use transformed data \( y_t; y_t - h I(|y_t - y_{t-h}| \leq c) \) for computing \( E[y_t; y_{t-h}] \), 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 \( y_t; y_{t-h} \) is used for \( c \) such that \( c \rightarrow \infty \) as \( T \rightarrow \infty \), 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. See Section 2.

Other transforms can be used, including smoothed versions of \( y_t; y_{t-h} I(|y_t - y_{t-h}| \leq c) \) 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.

We do not employ truncation $\text{sign}(x) \times \min\{|x|, c\}$ 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 $c \rightarrow \infty$. In general, truncation or trimming with a bounded threshold $c \rightarrow (0, \infty)$, will promote a bounded influence function, and therefore lead to infinitesimal 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 $c \rightarrow \infty$, the transform $y_t y_{t-h} I(|y_t y_{t-h} - \gamma| \leq c)$ 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. See, for example, Hagemann (2012) for discussion on the quantile spectrum for a Quantile AR model.

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. $L(c)$ denotes a slowly varying function: $\lim_{c \to \infty} L(\lambda c)/L(c) = 1 \forall \lambda > 0$ (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 $E[x I(|x - \gamma| < c)]$ identifies $E[x]$ when $x$ has a symmetric distribution since, in this case, $E[x I(|x - \gamma| < c)] = E[x] \times E[I(|x - E[x]| \leq c)]$. If $x = y_t y_{t-h} \geq 0$ a.s. then mean-centering is irrelevant, hence $\gamma = 0$. Examples include $h = 0$, or $y_t$ is squared GARCH, or log-stochastic volatility on $[0, 1]$.

The construction of our robust spectral density estimator requires sequences for centering $\tilde{\gamma}_{T,h}$, for determining the amount of trimming $k_{T,h}$, and for controlling the number of lags $b_T$. In
view of the above discussion, we want to ensure the transformed sample moments are unbiased when \( y_t y_{t-h} \) has a symmetric distribution. In practice we may not know whether \( y_t y_{t-h} \) has a symmetric distribution, but clearly it cannot have a non-degenerate symmetric distribution with unbounded support when \( P(y_t y_{t-h} > 0) = 1 \). Thus, the quantity used for centering in practice is \(^1\):

\[
\tilde{\gamma}_{T,h} \equiv \begin{cases} 
\frac{1}{T-h} \sum_{t=h+1}^{T} y_t y_{t-h} & \text{if } P(y_t y_{t-h} > 0) < 1 \\
0 & \text{if } P(y_t y_{t-h} > 0) = 1
\end{cases}
\]

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 \( y_t^2 \) \( \gtrsim \) 0 a.s., and when \( y_t \) is a squared GARCH process then \( \tilde{\gamma}_{T,h} = 0 \) for each \( h \). Otherwise, mean-centering is recommended as long as \( y_t y_{t-h} \) can possibly have a symmetric distribution. We assume symmetry for all lag functions, without additional notation, hence \( \tilde{\gamma}_{T,-h} \) is identically \( \tilde{\gamma}_{T,h} \) for \( h \geq 0 \).\(^2\)

We use the centered sample \( \{y_t y_{t-h} - \tilde{\gamma}_{T,h}\}_{t=1}^{T-h} \) to determine which \( y_t y_{t-h} \) are trimmed:

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

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

\[
k_{T,h} \rightarrow \infty \quad \text{and} \quad \frac{k_{T,h}}{T-h} \rightarrow 0 \quad \text{for each exploited lag } h \geq 0.
\]

If the lag \( h < 0 \), as we require for asymptotic arguments, then implicitly \( k_{T,h} / (T - |h|) \rightarrow 0 \).

The transformation we use is then

\[
y_t y_{t-h} I \left( |y_t y_{t-h} - \tilde{\gamma}_{T,h}| \leq \hat{Y}_{h,(k_{T,h})}^{(0)} \right),
\]

while \( k_{T,h} / (T - h) \rightarrow 0 \) ensures negligibility \( y_t y_{t-h} I (|y_t y_{t-h} - \tilde{\gamma}_{T,h}| \leq \hat{Y}_{h,(k_{T,h})}^{(0)}) \overset{P}{\rightarrow} y_t y_{t-h} \).

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

\[
h = \{0, 1, ..., b_T\}
\]

\(^1\)Our weak dependence assumptions rule out \( P(y_t y_{t-h} < 0) = 1 \), hence the only cases are \( P(y_t y_{t-h} > 0) < 1 \) and \( P(y_t y_{t-h} > 0) = 1 \).

\(^2\)Alternatively, we can write \( \tilde{\gamma}_{T,h} = 1 / (T - |h|) \sum_{t=|h|+1}^{T} y_t y_{t-h} |y_{t-h}| \) if \( P(y_t y_{t-h} > 0) < 1 \) for any viable \( h \).
for a sequence of bandwidths \( \{b_T\} \) that satisfy

\[
b_T \leq T - 1, \quad b_T \rightarrow \infty, \quad \text{and} \quad T - b_T \rightarrow \infty.
\]

There are many examples of viable sequences \( \{k_{T,h}, b_T\} \), including \( b_T = [\delta_b(T - 1)]^{a_b} \) 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 \( \delta_{k,h}, a_{k,h} > 0 \). 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 \( \{y_t y_{t-h} - \tilde{\gamma}_{T,h}\}_{t=1}^{T-h} \).

The periodogram has the well known expansion

\[
\left| \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} y_t e^{-i\lambda t} \right|^2 = \frac{1}{2\pi} \left\{ \frac{1}{T} \sum_{t=1}^{T} y_t^2 + 2 \sum_{h=1}^{T-1} \frac{1}{T} \sum_{t=h+1}^{T} y_t y_{t-h} \cos(\lambda h) \right\},
\]

hence it is a function of sample moments \( 1/T \sum_{t=h+1}^{T} y_t y_{t-h} \) that are unbiased estimators 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:

\[
\tilde{\gamma}_{T,h}^*(c) \equiv \left\{ \begin{array}{ll}
\frac{T-h}{T-h-k_{T,h}} \frac{1}{T} \sum_{t=h+1}^{T} y_t y_{t-h} I\left( y_t y_{t-h} - \frac{T-h}{T-h} \sum_{t=h+1}^{T} y_t y_{t-h} \right) < c \right) & \text{if} \quad P(y_t y_{t-h} > 0) < 1 \\
\frac{1}{T} \sum_{t=h+1}^{T} y_t y_{t-h} I\left( |y_t y_{t-h}| < c \right) & \text{if} \quad P(y_t y_{t-h} > 0) = 1
\end{array} \right.
\]

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

\[
\tilde{T}_T(\lambda) = \frac{1}{2\pi} \left( \tilde{\gamma}_{T,0}^*(\tilde{\gamma}_{0,(k_{T,h})}^{(0)}) + 2 \sum_{h=1}^{b_T} \tilde{\gamma}_{T,h}^*(\tilde{\gamma}_{h,(k_{T,h})}^{(0)}) \times \cos(\lambda h) \right),
\]

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

\[
\hat{\theta}_T = \arg \min_{\theta \in \Theta} \sum_{j \in \mathcal{F}} \left( \ln f(\lambda_j, \theta) + \frac{\tilde{T}_T(\lambda_j)}{f(\lambda_j, \theta)} \right).
\]

The construction of \( \tilde{\gamma}_{T,h}^*(c) \) 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 \( 1/(T - h) \sum_{t=h+1}^{T} y_t y_{t-h} \) with \( E[y_t y_{t-h}] \), and the threshold \( \tilde{\gamma}_{h,(k_{T,h})}^{(0)} \) 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 2.2, and see Lemma

\footnote{We find in our simulation experiments that \( b_T = T^{0.95} \) 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 \( b_T \).}
A.7 in Appendix A.2. Then, as per the discussion at the top of this section:

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

Notice \( \hat{I}^*_T(\lambda) \) can in principle be negative for a finite sample because the sequence \( \{ \hat{\gamma}^{(0)}_{T,h} \}_{h=0}^{b_T} \) 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 \( \hat{I}^*_T(\lambda) \) is not a periodogram. This is irrelevant here since we only need to estimate \( \theta_0 \), and \( \hat{I}^*_T(\lambda) \) can be used for a consistent and asymptotically normal estimator.

### 2.2 Assumptions and Main Result

Asymptotic theory requires a sequence \( \{ c_{T,h} \} \) of non-random numbers \( c_{T,h} \geq 0 \) that the order statistic \( \hat{y}_{T,h}^{(0)} \) approximates. Since \( \hat{\gamma}_{T,h} \) exists for lags \( h = 0, ..., b_T \), and \( T - b_T \to \infty \), it follows that under mean-centering \( \hat{\gamma}_{T,h} = (T-h)^{-1} \sum_{t=h+1}^{T} y_{t-h} \) is consistent for \( E[y_{t-h}] \) under regularity assumptions detailed below. Define the population centering parameter:

\[
\hat{\gamma}_h = \begin{cases} 
E[y_{t-h}] & \text{if } P(y_{t-h} > 0) < 1 \\
0 & \text{if } P(y_{t-h} > 0) = 1 
\end{cases} \quad \text{for } h \geq 0. \tag{4}
\]

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

\[
P( |y_{t-h} - \hat{\gamma}_h| \geq c_{T,h} ) = \frac{k_{T,h}}{T-h} \quad \text{for } h \geq 0. \tag{5}
\]

If \( h < 0 \) then implicitly \( P( |y_{t-h} - \hat{\gamma}_h| \geq c_{T,h} ) = k_{T,h}/(T - |h|) \) where \( k_{T,h}/(T - |h|) \to 0 \).

Let \( \kappa \) be the moment supremum of \( y_t \),

\[
\kappa \equiv \arg\sup \{ \alpha > 0 : E[|y_t|^\alpha] < \infty \},
\]

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 by asymptotically with probability approaching one \( \mathcal{H}(\theta) = (\partial/\partial \theta)^2 \sum_{j \in F} \{ \ln f(\lambda_j, \theta) + \hat{T}_T(\lambda_j)/f(\lambda_j, \theta) \} \). See the proof of Theorem 2.1 in Appendix A.3.

We need to impose bounds on mixing and interlaced correlation coefficients. Define \( \Omega \equiv \sigma(y_t : s \leq \tau \leq t) \) and \( \Omega \equiv \Omega_{\infty} \), a measure space \( (\mathbb{F}, \mathcal{F}, \mathbb{P}) \), and \( \alpha \)-mixing coefficients \( \alpha_h \equiv \sup_{j \leq j' \leq h} \mathbb{P}(A_{j} \cap B_{j'}) \) for all finite \( h \). Then the interlaced maximal correlation coefficient is \( \rho_h^* \equiv \sup_{\mathcal{G}_h, \mathcal{T}_h} \rho(\sigma(y_t : t \in \mathcal{G}_h), \sigma(y_{s-t} : s \in \mathcal{T}_h)) \), where the supremum is taken over all \( \mathcal{G}_h \) and \( \mathcal{T}_h \). 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 \( \rho_h^* < 1 \).

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 \( y_t \). If \( y_t \) is stationary ARMA or GARCH, or is stationary autoregressive stochastic volatility, in each case with iid innovations that have a continuous bounded distribution, then \( y_t \) is \( \alpha \)-mixing with geometric decay, hence \( \alpha_h = O(h^{-p/(p-2)}) \) \( \forall p > 0 \). See, e.g., Doukhan (1994) and Carrasco and Chen (2002).

**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). Dunsmuir (1979)'s C2.3 assumes \( \{y_t\} \) has a finite fourth moment, and is a linear function of a homoscedastic martingale difference, each to ensure Gaussian asymptotics and to render a characterizable asymptotic variance. We allow an infinite fourth
Remark 5 The mixing rate $\alpha_h = O(h^{-p/(p-2)})$ ensures $k_T^{1/2}$-convergence for the thresholds $\hat{Y}_{h,(k_T,h)}^{(0)}$, 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 3 We use A.3 to show that $\hat{Y}_{T,h}$ is the same asymptotically if we replace a stochastic threshold $\hat{Y}_{h,(k_T,h)}^{(0)}$ with the deterministic one $c_{T,h}$. See the proof of consistency Theorem 2.1.

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

Assumption B (regularly varying tails). $P(y_{t-h} - \tilde{\gamma}_h \leq -c) = \mathcal{L}_{h,1}(c) c^{-\kappa_{h,1}}$ and $P(y_{t-h} - \tilde{\gamma}_h \geq c) = \mathcal{L}_{h,2}(c) c^{-\kappa_{h,2}}$ 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 $\kappa_{h,0} = \arg \sup\{\alpha > 0 : E[|y_{t-h}|^{\alpha}] < \infty\} > 1$, hence $\kappa_{h,0} = \kappa/2$ and

$$P(\{y_{t-h} - \tilde{\gamma}_h| \geq c\} = \mathcal{L}_{h,0}(c) c^{-\kappa_{h,0}}(1 + o(1))$$

where $\mathcal{L}_{h,0}(c) = \mathcal{L}_{h,1}(c) + \mathcal{L}_{h,2}(c)$. (6)

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 $\kappa_{h,0} \geq 0$.

If mean-centering is used then we require that $(T - h)^{-1}\sum_{t=h+1}^{T} y_{t-h}$ does not asymptotically impact the order statistic $\hat{Y}_{h,(k_T,h)}^{(0)}$ of $y_{t,h} - (T - h)^{-1}\sum_{t=h+1}^{T} y_{t-h}$. By exploiting a maximal inequality in Hansen (1991), we use the non-sharp bound $(T - h)^{-1}\sum_{t=h+1}^{T} y_{t-h} = E[y_{t-h}] + O_p(1/T)$ for tiny $\ell > 0$ in our proof of $\hat{Y}_{h,(k_T,h)}^{(0)}/c_{T,h} = 1 + O_p(1/k^{1/2}_{T,h})$. Thus, it suffices to restrict $k_{T,h} \to \infty$ at most at a slowly varying rate to ensure $(T - h)^{-1}\sum_{t=h+1}^{T} y_{t-h} = E[y_{t-h}] + o_p(1/k^{1/2}_{T,h})$. A sharper bound can be obtained if we know the rate of convergence of $(T - h)^{-1}\sum_{t=h+1}^{T} y_{t-h}$, 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). If mean-centering is not used, e.g. $h = 0$ or a GARCH model is estimated, then we are free to choose $k_{T,h}$.

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

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

2. $b_T \leq T - 1$, $b_T \to \infty$, and $T - b_T \to \infty$.

**Remark 6** Proportionality of the fractiles $k_{T,h+1} \sim Kk_{T,h}$ allows us to give a simple characterization of the rate of convergence of our estimators. See especially the proofs of Theorem 2.2, and a key central limit theorem Lemma A.9 in Appendix A.2.

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

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

**Remark 7** We cannot prove consistency using less structure on the spectrum. In particular, to handle the order statistic $\hat{\gamma}_{T,h}^{(a)}$ 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:

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

for $h \geq 0$.

Define a spectral density estimator:

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

and construct Hessian, covariance, and scale matrices:

$$
\Omega \equiv \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\partial}{\partial \theta} \ln f(\lambda; \theta_0) \frac{\partial}{\partial \theta} \ln f(\lambda; \theta_0) d\lambda
$$

$$
S_T \equiv T \times E \left[ \left( \frac{1}{2\pi} \int_{-\pi}^{\pi} (I_T^*(\lambda) - f(\lambda)) \varpi(\lambda) d\lambda \right) \left( \frac{1}{2\pi} \int_{-\pi}^{\pi} (I_T^*(\lambda) - f(\lambda)) \varpi(\lambda) d\lambda \right) \right]
$$
The rate

\[ \mathcal{V}_T = \Omega^{-1} \mathcal{S}_T \Omega^{-1}. \]

We also need the population transformed moments:

\[ \gamma_{T,h}^n(c) \equiv \begin{cases} 
\frac{T-h}{T-h-kT,h}E[y_t y_{t-h} \mathbb{1}(|y_t y_{t-h}| < c)] & \text{if } P(y_t y_{t-h} > 0 < 1) \\
E[y_t y_{t-h} \mathbb{1}(|y_t y_{t-h}| < c)] & \text{if } P(y_t y_{t-h} > 0) = 1 \\
0 & \text{if } |h| > b_T
\end{cases} \]

\[ f_T^r(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma_{T,h}^n(c_{T,h}) e^{-i\lambda h}. \]

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 \( T^{1/2} \mathcal{V}_T^{-1/2}(\hat{\theta}_T^* - \theta_0 + B_T) \overset{d}{\rightarrow} N(0, I_k) \) where

\[ B_T = \Omega^{-1} \int_{-\pi}^{\pi} \varpi(\lambda) \left\{ f_T^r(\lambda) - f(\lambda) \right\} d\lambda = \Omega^{-1/2} \sum_{h=-\infty}^{\infty} \varpi_h \left\{ \gamma_{T,h}^n(c_{T,h}) - E[y_t y_{t-h}] \right\}. \] (8)

If \( y_t y_{t-h} \) is symmetrically distributed for each \( h > 0 \) then \( \gamma_{T,h}^n(c_{T,h}) = E[y_t y_{t-h}] \), hence \( B_T = \Omega^{-1}(2\pi)^{-1} \varpi_0 \{ \gamma_0^n(c_{T,0}) - E[y_t^2] \}. \) In general \( ||\mathcal{V}_T|| = K ||\mathcal{S}_T|| \sim K E[y_t^4 I(y_t^2 < c_{T,0})] \), \( \lim_{T \rightarrow \infty} ||\mathcal{V}_T|| > 0 \) and \( T^{1/2} ||\mathcal{V}_T||_{1/2} \rightarrow \infty. \) If \( \kappa > 4 \) then \( ||\mathcal{V}_T|| \sim K \), if \( \kappa = 4 \) then \( ||\mathcal{V}_T|| \rightarrow \infty \) is slowly varying, and if \( \kappa \in (2, 4) \) then \( ||\mathcal{V}_T|| \sim K(T/k_{T,0})^{4/\kappa-1}. \)

**Remark 8** \( ||\mathcal{S}_T|| \sim K E[y_t^4 I(y_t^2 < c_{T,0})] \) because the trimming fractiles are proportional \( k_{T,h} \sim Kk_{T,h} \) 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.

**Remark 9** \( \mathcal{S}_T \) can be estimated by standard methods (e.g. Chiu, 1988).

The rate of convergence \( T^{1/2} ||\mathcal{V}_T||_{1/2} \) depends on tail decay. If \( y_t \) has a fourth moment \( \kappa > 4 \) then we have \( T^{1/2} \)-convergence, and if \( \kappa = 4 \) then the rate is \( T^{1/2}/\mathcal{L}(T) \) for some slowly varying \( \mathcal{L}(T) \rightarrow \infty \) that depends on the slowly varying component in Assumption B. Otherwise, if \( \kappa < 4 \) then the rate is \( T^{1/2}/(T/k_{T,0})^{2/\kappa-1/2}. \) Since by Assumption C \( k_{T,0} \rightarrow \infty \) is unrestricted beyond \( k_{T,0} = o(T) \), the rate \( T^{1/2}/(T/k_{T,0})^{2/\kappa-1/2} \) can be made arbitrarily close to \( T^{1/2} \) by setting \( k_{T,0} \sim T/\varsigma_T \) for \( \varsigma_T \rightarrow \infty \) very slowly, e.g. \( k_{T,0} \sim \lambda T/\ln(T) \) or \( \lambda T/\ln(\ln(T)) \) for some \( \lambda > 0. \) This is similar to heavy tail robust estimators in Hill (2012b, 2013, 2014a), in particular for GARCH models (Hill, 2014a).

If \( y_t y_{t-h} \) has a symmetric distribution then \( \gamma_0^0(c_{T,0}) = E[y_t^2], \) and \( \gamma_{T,h}^n(c_{T,h}) = E[y_t y_{t-h}] \) for each \( h \neq 0, \) and \( \hat{\theta}_T^* \) is necessarily asymptotically biased in its limit distribution when \( \kappa < 4. \) This was noted in Csörgő, Horváth, and Mason (1986) for location estimation. However, since \( k_{T,h} \rightarrow \infty \) is slowly varying, if \( \kappa \geq 4 \) then bias vanishes rapidly enough that \( \hat{\theta}_T^* \) is unbiased in its limit distribution.

If \( y_t y_{t-h} \) has an asymmetric distribution then asymptotic unbiasedness is possible even when \( \kappa < 4 \)
because the transformed moment biases $\gamma_{T,h}^*(c_{T,h}) - E[y_t y_{t-h}]$ can, in principle, cancel out. A proof of the following is presented in Hill and McCloskey (2014) since it is based on well known moment properties by Karamata theory in view of regular variation Assumption B.

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

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

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

**Remark 10** Case (i) can be expanded to include other “thin tail” cases, including when $y_t$ has exponential tails. Although we do not provide details here, if $\lim_{c \rightarrow \infty} P(|y_t| > c) = 0$ for any finite $\kappa > 0$, then it can be shown that (i) holds provided $k_{T,h}/\ln(T) \rightarrow 0.$

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,

$$y_t = \sum_{i=0}^{\infty} \xi_i \epsilon_{t-i} \quad \text{where} \quad \sum_{i=0}^{\infty} \xi_i^2 (\theta_0) < \infty, \xi_0 (\theta_0) = 1, \quad (9)$$

$$\sigma^2 (\theta_0) \equiv E[\xi_t^2] < \infty \quad \text{and} \quad E[\epsilon_t \epsilon_s] = 0 \forall s \neq t,$$

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 is superfluous by the discussion preceding it. See also Remark 14 following Lemma A.4 in Appendix A.1. Define the $\sigma$-field $\mathcal{I}_t \equiv \sigma(y_{\tau}: \tau \leq t)$.

**Corollary 2.4** In addition to Assumption A, let $y_t$ satisfy (9), and assume $E[\epsilon_t | \mathcal{I}_{t-1}] = 0 \ a.s., E[\epsilon_t^2 | \mathcal{I}_{t-1}] = \sigma^2 \ a.s., E[\epsilon_t^2 | \mathcal{I}_{t-1}] = s \ a.s., \text{ and } E[\epsilon_t^4] = K < \infty$. Let $k_{T,h} \rightarrow \infty \text{ and } k_{T,h}/(T - h) = o(1)$. Then $T^{1/2}(\hat{\theta}_T^* - \theta_0) \xrightarrow{d} N(0, \mathcal{V})$ where $\mathcal{V} = \lim_{T \rightarrow \infty} \mathcal{V}_T = \Omega^{-1}(2\Omega + \Pi)\Omega^{-1}$ and

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

**3 Bias Corrected FD-QML**

We need a bias-corrected version of $\hat{\gamma}_{T,h}^*(c)$ 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 we need only treat $\hat{\gamma}_{T,0}^*(c)$. 


3.1 Bias Correction : Power Law Tails

In order to parametrically characterize, and therefore estimate, bias in our estimators $\hat{\gamma}_{T,h}(\hat{\mathcal{Y}}_{h,(k_T,h)})$ of $(T-h)T^{-1}E[y_t y_{t-h}]$, we must specify the slowly varying components $L_{h,i}(c)$ in Assumption B. We assume for convenience:

$$P(y_t y_{t-h} - \hat{\gamma}_h \leq -c) = d_{h,1} c^{-\kappa_{h,1}} (1 + o(1)) \quad \text{and} \quad P(y_t y_{t-h} - \hat{\gamma}_h \geq c) = d_{h,2} c^{-\kappa_{h,2}} (1 + o(1)),$$  \hspace{1cm} (11)

where $\hat{\gamma}_h = E[y_t y_{t-h}]$ if $P(y_t y_{t-h} > 0) = 1$, else $\hat{\gamma}_h \equiv 0$. We exploit a second order version of (11) 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.

Since $\hat{\gamma}_{T,h}^*(\hat{\mathcal{Y}}_{h,(k_T,h)})$ is, with probability approaching one, equivalent to $\hat{\gamma}_{T,h}^*(c_{h,T})$ such that the latter suffices for asymptotic analysis (see Lemma A.7 in Appendix A.2), we need an expression for $E[\hat{\gamma}_{T,h}^*(c_{T,h})]$. If $P(y_t y_{t-h} > 0) = 1$, then by (11) 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 $h \geq 0$:

$$E[\hat{\gamma}_{T,h}^*(c_{T,h})] = \frac{T-h}{T} \{ E[y_t y_{t-h}] - E[y_t y_{t-h}]I(|y_t y_{t-h}| \geq c_{T,h}) \} \quad (12)$$

$$\sim \frac{T-h}{T} \left\{ E[y_t y_{t-h}] - \frac{1}{\kappa_{h,0}} \frac{k_{T,h}}{1 - T-h} c_{T,h} \right\}$$

$$= \frac{T-h}{T} E[y_t y_{t-h}] - \frac{1}{\kappa_{h,0}} \frac{k_{T,h}}{1 - T-h} c_{T,h}$$

Thus, for example, if $h = 0$ then

$$E[\hat{\gamma}_{T,0}^*(c_{T,0})] \sim E[y_t^2] - \frac{1}{\kappa_{0,0}} \frac{k_{T,0}}{1 - T} c_{T,0}. \quad (13)$$

Next, suppose $P(y_t y_{t-h} > 0) < 1$. Since by (11) each $c_{T,h} \sim d_{h,0}^{1/\kappa_{h,0}}(T/k_{T,h})^{1/\kappa_{h,0}}$, it follows for $h \geq 0$:

$$E[(y_t y_{t-h} - E[y_t y_{t-h}]) I(|y_t y_{t-h} - E[y_t y_{t-h}]| \geq c_{T,h})]$$

$$= \int_{c_{T,h}}^{\infty} \{ P(y_t y_{t-h} - E[y_t y_{t-h}] \geq u) - P(y_t y_{t-h} - E[y_t y_{t-h}] \leq -u) \} du$$

$$= \int_{c_{T,h}}^{\infty} \{ d_{h,2} u^{-\kappa_{h,2}} - d_{h,1} u^{-\kappa_{h,1}} \} du \sim \frac{d_{h,2} c_{T,h}^{1-\kappa_{h,2}}}{\kappa_{h,2} - 1} - \frac{d_{h,1} c_{T,h}^{1-\kappa_{h,1}}}{\kappa_{h,1} - 1}.$$

Therefore, for $h \geq 0$:

$$E[\hat{\gamma}_{T,h}^*(c_{T,h})] = \frac{T-h}{T-h-k_{h,T}} \frac{T-h}{T} E[y_t y_{t-h} I(|y_t y_{t-h} - E[y_t y_{t-h}]| < c_{T,h})] \quad (14)$$
Many estimators are possible, but we use Hill (1975)’s estimator of \( \kappa \): 

\[
E[y_t y_{t-h}] = \frac{E[y_t y_{t-h} - E[y_t y_{t-h}] I (|y_t y_{t-h} - E[y_t y_{t-h}]| \geq c_T, h)]}{E[y_t y_{t-h} - E[y_t y_{t-h}] I (|y_t y_{t-h} - E[y_t y_{t-h}]| \geq c_T, h)]}
\]

Asymptotic approximations (12) and (14) lead to asymptotically negligible approximation errors under second order power law Assumption B’, below, cf. Peng (2001) and Hill (2013). Thus, (12) and (14) implicitly define asymptotic approximations of biases 

\[
- \left( E[\hat{\gamma}_{T, h}^* (c_T, h)] - (T - h)T^{-1} E[y_t y_{t-h}] \right)
\]

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

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

\[
\hat{\gamma}_{h, t}^{(0)} \equiv |y_t y_{t-h} - \hat{\gamma}_{T, h}|
\]

\[
\hat{\gamma}_{h, t}^{(1)} \equiv - (y_t y_{t-h} - \hat{\gamma}_{T, h}) I (y_t y_{t-h} - \hat{\gamma}_{T, h} < 0) \quad \text{and} \quad \hat{\gamma}_{h, t}^{(2)} \equiv (y_t y_{t-h} - \hat{\gamma}_{T, h}) I (y_t y_{t-h} - \hat{\gamma}_{T, h} \geq 0),
\]

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

\[
m_{T, h} \in \{ 1, ..., T-h \}, \quad m_{T, h} \rightarrow \infty \quad \text{and} \quad m_{T, h}/(T-h) \rightarrow \infty.
\]

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 for a wide range of dependent processes (see Hill, 2009, 2010, 2014b):

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

In order to minimize estimated bias, and because there are \( T - h - 1 \) observations of \( y_t y_{t-h} \), it is helpful to use a potentially different \( m_{T, h} \) for each lag \( h \geq 0 \).

The bias-corrected tail-trimmed estimators are therefore

\[
\hat{\gamma}_{T, h}^{(bc)} \equiv \hat{\gamma}_{T, h} (\hat{\gamma}_{h, (k_{T, h})}) + \mathcal{R}_{T, h} \quad \text{for } h \geq 0,
\]

\[
(15)
\]
Finally, if the tail indices are known to be equivalent $\kappa_{0,h} = \kappa_{1,h} = \kappa_{2,h}$, then we can replace tail-specific $\hat{\kappa}_{h,i,mT,h}$ with the two-tailed $\hat{\kappa}_{0,mT,h}$, since the latter is computed from a larger sample and therefore will be sharper with higher probability. We then have:

$$\hat{\kappa}_{T,h} = \frac{T-h}{T-h-kT,h} \left( \frac{\tilde{d}_{h,2,mT,h}(\tilde{y}_{h,(kT,h)})^{1-\hat{\kappa}_{h,2,mT,h}} - \tilde{d}_{h,1,mT,h}(\tilde{y}_{h,(kT,h)})^{1-\hat{\kappa}_{h,1,mT,h}}}{\hat{\kappa}_{h,0,mT,h} - 1} \right)$$

if $P(yTy_{-h} > 0) < 1$  

(16)

$$\frac{1}{\hat{\kappa}_{h,0,mT,h}} \sum_{t=h+1}^{T} yTy_{-h}$$

(17)

$$\hat{\kappa}_{T,h} = \frac{T-h}{T-h-kT,h} \left( \frac{\tilde{d}_{h,2,mT,h}(\tilde{y}_{h,(kT,h)})^{1-\hat{\kappa}_{h,2,mT,h}} - \tilde{d}_{h,1,mT,h}(\tilde{y}_{h,(kT,h)})^{1-\hat{\kappa}_{h,1,mT,h}}}{\hat{\kappa}_{h,0,mT,h} - 1} \right)$$

if $P(yTy_{-h} > 0) = 1$  

(18)

### 3.2 Optimal Bias Correction

As discussed in Hill (2013), a shortcoming of a bias-corrected estimator like $\hat{\gamma}_{T,h}(\tilde{y}_{h,(kT,h)})$ is its reliance on estimates $\hat{\kappa}_{h,i,mT,h}$ and $\tilde{d}_{h,i,mT,h}$ based on one fractile $mT,h$, while a variety of $mT,h$ values can be used for estimation. Further, $\hat{\kappa}_{T,h}$ is well defined only when $\hat{\kappa}_{h,i,mT,h} > 1$, and it seems desirable to choose $mT,h$ such that $\hat{\gamma}_{T,h}(\tilde{y}_{h,(kT,h)})$ is close to the untrimmed estimator, e.g. $1/T \sum_{t=h+1}^{T} yTy_{-h}$.

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

$$mT,h(\xi) = [\xi mT,h] \text{ where } \xi \in Y = [\xi, \bar{\xi}].$$

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

$$\hat{\gamma}^{**}_{T,h}(\tilde{y}_{h,(kT,h)}) = \hat{\gamma}^{*}_{T,h}(\tilde{y}_{h,(kT,h)}) + \hat{\kappa}_{T,h}(\hat{\xi}_{T,h})$$

(19)

where $\hat{\xi}_{T,h}$ optimally places $\hat{\gamma}^{**}_{T,h}(\tilde{y}_{h,(kT,h)})$ near $1/T \sum_{t=h+1}^{T} yTy_{-h}$ such that $\hat{\kappa}_{h,i,mT,h} > 1$. Specifically:

$$\hat{\xi}_{T,h} = \arg \min_{\xi \in Y_{h}} \left| \hat{\gamma}^{*}_{T,h}(\tilde{y}_{h,(kT,h)}) + \hat{\kappa}_{T,h}(\xi) - \frac{1}{T} \sum_{t=h+1}^{T} yTy_{-h} \right|$$

(20)

where $Y_{h} \equiv \{ \xi \in Y : \hat{\kappa}_{0,0,T,0}(\xi) > 1 \}$ and $Y_{h}^{*} \equiv \{ \xi \in Y : \hat{\kappa}_{h,i,mT,h}(\xi) > 1 \text{ for } i = 1, 2 \}$.

(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 $\hat{\kappa}_{h,i,mT,h}(\xi)$ and $\tilde{d}_{h,i,mT,h}(\xi)$ will not affect our bias estimator uniformly on $Y$, asymptotically with probability one, as long as $mT,h/kT,h \rightarrow \infty$. See Hill (2013, Section 2.2), and see
the discussion below. 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 \( E[y_\xi^2] < \infty \) fails, or \( E[y_\xi^2] < \infty \) holds but \( h \) is large enough that there are too few tail observations of \( \{yt_{t-h}\}_{t=h+1}^T \) from which to obtain a good estimate of \( \kappa_{h,i} \). 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 \( \xi \).

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

\[
\frac{\gamma_{T,h}^{(0)}}{\hat{\gamma}_{h,i(m,T,h)}^{(0)}} = \frac{\tilde{\gamma}_{T,h}^{**}(\hat{\gamma}_{h,i(m,T,h)}^{(0)}) \times I \left( \left| \tilde{\gamma}_{T,h}^{**}(\hat{\gamma}_{h,i(m,T,h)}^{(0)}) - \frac{1}{T} \sum_{t=h+1}^{T} y_t y_{t-h} \right| < \left| \tilde{\gamma}_{T,h}^{*}(\hat{\gamma}_{h,i(m,T,h)}^{(0)}) - \frac{1}{T} \sum_{t=h+1}^{T} y_t y_{t-h} \right| \right)}{\tilde{\gamma}_{T,h}^{*}(\hat{\gamma}_{h,i(m,T,h)}^{(0)}) \times I \left( \left| \tilde{\gamma}_{T,h}^{**}(\hat{\gamma}_{h,i(m,T,h)}^{(0)}) - \frac{1}{T} \sum_{t=h+1}^{T} y_t y_{t-h} \right| \geq \left| \tilde{\gamma}_{T,h}^{*}(\hat{\gamma}_{h,i(m,T,h)}^{(0)}) - \frac{1}{T} \sum_{t=h+1}^{T} y_t y_{t-h} \right| \right)}.
\] (22)

This is a small sample correction since, e.g., \( \tilde{\gamma}_{T,0}^{**}(\hat{\gamma}_{0,0}^{(0)}) = \tilde{\gamma}_{T,0}^{**}(\hat{\gamma}_{0,0}^{(0)}, \xi_{T,0}) \) as \( T \to \infty \) with probability approaching one due to negative bias. In small samples, however, this correction matters since it leads to a sharper FD-QML estimator than if we merely use \( \tilde{\gamma}_{T,h}^{**}(\hat{\gamma}_{h,i(m,T,h)}^{(0)}) \) due to sampling error in our bias estimator. The optimal bias-corrected FD-QML estimator is

\[
\hat{\theta}_{T}^{(obc)} = \arg \min_{\theta \in \Theta} \sum_{j \in \mathcal{F}} \left( \ln f(\lambda_j; \theta) + \frac{\tilde{I}_{T}^{(obc)}(\lambda_j)}{f(\lambda_j; \theta)} \right)
\]

where

\[
\tilde{I}_{T}^{(obc)}(\lambda) = \frac{1}{2\pi} \left( \gamma_{T,h}^{(0)}(\hat{\gamma}_{0,0}^{(0)}) + 2 \sum_{h=1}^{b_T} \gamma_{T,h}^{(0)}(\hat{\gamma}_{h,i(m,T,h)}^{(0)}) \times \cos(\lambda h) \right).
\] (23)

As long as \( \hat{\kappa}_{h,i,m,T,h} \) and \( \hat{d}_{h,i,m,T,h} \) in \( \hat{R}_{T,h} \) are \( m_{T,h}^{1/2} \)-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 \( \tilde{\gamma}_{T,h}^{**}(\hat{\gamma}_{h,i(m,T,h)}^{(0)}) \). 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 any subset \( \Upsilon^* \subset \Upsilon \) 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).**

\[
P(y_t y_{t-h} - \hat{\gamma}_h \leq -c) = d_{h,1} c^{-\kappa_{h,1}}(1 + O(r_1(c))) \quad \text{and} \quad P(y_t y_{t-h} - \hat{\gamma}_h \geq c) = d_{h,2} c^{-\kappa_{h,2}}(1 + O(r_2(c))),
\]

where \( d_{h,i} > 0, \kappa_{h,i} > 1 \),

\footnote{The problem of tail inference based on point estimates like \( \hat{\kappa}_{h,i,m,T,h} \) is well known. 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).}
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

$$m_{T,h} \in \{1, \ldots, T - h\} \text{ and } m_{T,h} \to \infty, \text{ and either } r_i(c) = c^{-e_{h,i}} \text{ and } m_{T,h} = o((T - h)^{2e_{h,0} / (2e_{h,0} + \kappa_{h,0}))},$$

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 11** 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 cases are possible. See also Haeusler and Teugels (1985), Hsing (1991) and Hill (2010). It well known that as tails deviate from a Pareto law, $m_{T,h}$-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 $h = 0, \ldots, b_T$, 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 $\delta_{m,h}, \delta_{k,h} > 0$.

Recall $f_T^*(\lambda) \equiv (2\pi)^{-1} \sum_{h=-\infty}^{\infty} \gamma_{T,h}^*(c_{T,h}) e^{-i\lambda h}$, and define

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

$$J_{T,h,t} \equiv I(|y_t y_{t-h}| \geq c_{T,h}) - P(|y_t y_{t-h}| \geq c_{T,h}) \text{ and } J_{T,h} \equiv \frac{1}{\kappa_{h,0}} k_{T,h}^{1/2} \sum_{t=1}^{T} J_{T,h,t}$$

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

and

$$\hat{Z}_T \equiv T^{1/2} \frac{1}{2\pi} \left( \int_{-\infty}^{\infty} J_T^*(\lambda) - f_T^*(\lambda)) \varpi(\lambda)d\lambda \right) + \frac{1}{2\pi} \sum_{h=-b_T}^{b_T} \varpi_h D_{T,h} J_{T,h}^*$$

$$\hat{S}_T \equiv E[\hat{Z}_T \hat{Z}_T^*] \text{ and } \hat{V}_T = \Omega^{-1} \hat{S}_T \Omega^{-1}.$$  

**Theorem 3.1** Under Assumptions A, B’ and C, $T^{1/2} \hat{V}_T^{-1/2}(\hat{\theta}_T^{(obc)} - \theta_0) \overset{d}{\to} N(0, I_k)$ where $\hat{V}_T = V_T(1 + o(1))$ when $\kappa \geq 4$ and $\hat{V}_T = V_T(1 + O(1))$ when $\kappa \in (2, 4)$.

**Remark 12** In order to estimate $\hat{S}_T$, a sample counterpart to $\sum_{h=-b_T}^{b_T} \varpi_h D_{T,h} J_{T,h}^*$ is available by truncating the sum and using estimators of $c_{T,h}$ and $\kappa_{h,i}$. Conversely, we may subsample $\hat{Z}_T^{(obc)}(\lambda_j)/f(\lambda_j, \hat{\theta}_T^{(obc)})$ standardized to have mean one. Since subsampling with heavy tailed data has known complications (see Hall, 1990), we do not treat the topic here.

**Remark 13** The order statistic $\hat{Y}_{h, (k,T,h)}^{(0)}$ is used both for trimming in $\hat{\gamma}_{T,h}^*(\hat{Y}_{h, (k,T,h)}^{(0)})$ and for the bias estimator $\hat{R}_{T,h}$, hence the FD-QML scale $\hat{V}_T$ is no longer $V_T$ in the heavy tail case $\kappa < 4$. We show in Appendix A that, for asymptotic analysis, we can replace $\hat{\gamma}_{T,h}^*(\hat{Y}_{h, (k,T,h)}^{(0)})$ with $\hat{\gamma}_{T,h}^*(c_{T,h})$ and write
\( \hat{R}_{T,h} = R_{T,h} + D_{T,h} \times k_{T,h}^{1/2} (\hat{y}_{h,(kT,h)}^{(0)}/c_{T,h} - 1) \times (1 + o_p(1)) \), while also \( k_{T,h}^{1/2} (\hat{y}_{h,(kT,h)}^{(0)}/c_{T,h} - 1) = 2^h (1 + o_p(1)) \). Hence \( \hat{y}_{h,(kT,h)}^{(0)}/c_{T,h} - 1 \). Finally, \( \hat{T} \contribute to the variance of \( \hat{T} \). 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 \( \hat{T} \) behaves as \( T \) increases for exponential tails:

\[ P(|y_T \gamma_T - \gamma_T| \geq c) = \vartheta \exp\{-\zeta_T c^{\delta_T} \} \] where \( \vartheta, \zeta_T, \delta_T > 0 \).

The following extends to other thin tailed distributions with obvious changes to the derivations that utilize (24).

Dominated convergence and (24) imply we do not have an asymptotic bias problem if \( T^{1/2} \hat{R}_{T,h} \). Logically, the intermediate order statistic \( \hat{y}_{h,(kT,h)}^{(0)}/c_{T,h} = 1 + O_p(1/k_{T,h}^{1/2}) \) provided in the exponential case \( k_{T,h}/\ln(T) \to 0 \), while Hill (1975)'s estimator can be easily shown to diverge in probability. Finally, we only need \( m_{T,h} \to \infty \) and \( m_{T,h}/T \to 0 \) 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 \( m_{T,h}/T \to 0 \). Then \( T^{1/2} \hat{R}_{T,h} \overset{d}{\to} 0 \) hence \( T^{1/2} \hat{V}_{T}^{(0)} - \theta_0 \overset{d}{\to} N(0,I_k) \). Further, if the conditions of Corollary 3.2 hold then \( T^{1/2} (\hat{\theta}_{T}^{(0)} \gamma_{T} - \theta_0) \overset{d}{\to} N(0,V) \) where \( V = \Omega^{-1}(2\Omega + \Pi)\Omega^{-1} \) and \( \Omega \) and \( \Pi \) are defined in (10).
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] fractile 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 \( B_T = \Omega^{-1} (2\pi)^{-1} \varpi_0 \{ \gamma_0^* (c_{T,0}) - E[y_t^2] \} \), which is easily analyzed. The general case is far more complicated, and possibly intractable, without a specific model and known distribution for \( y_t y_{t-h} \).

If \( \kappa \neq 4 \) then \( k_{T,h} = a \) for some constant \( a \geq 0 \) minimizes the mse as \( T \to \infty \), and if \( \kappa = 4 \) then \( k_{T,h} = 0 \) minimizes the mse as \( T \to \infty \). See Theorem B.1 in Hill and McCloskey (2014). A bounded \( k_{T,h} \) follows from the fact that bias dominates the variance in their contribution to the mse, and bias \( \gamma_0^* (c_{T,0}) - E[y_t^2] \) is reduced with a smaller amount of trimming. Obviously a constant \( k_{T,h} \) is not allowed here precisely to ensure asymptotic normality and unbiasedness 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, as we discuss below it is then much 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, provided deeper conditions like the interlaced correlation coefficient bound in Assumption A are imposed. This follows since otherwise required partial sum variances of tail-trimmed processes can only be bounded\(^5\), in which case a bias expansion may not lead to useful results.

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 \( \hat{\gamma}_{T,0}^* (\hat{y}_{0,(k_{T,h})}) \) and possibly cross-moments \( \hat{\gamma}_{T,h}^* (\hat{y}_{h,(k_{T,h})}) \) for \( h \geq 1 \) are biased, where the bias in general is augmented for larger \( k_{T,h} \). Second, \( \hat{Z}_{T}^{(obe)} (\lambda) \) uses sample moments \( \hat{\gamma}_{T,h}^* (\hat{y}_{h,(k_{T,h})}) \) for only \( h = 0, \ldots, b_T \), where \( T - b_T \to \infty \). The latter must hold since robustness requires \( k_{T,h} \in \{1, \ldots, T-h\} \) and \( k_{T,h} \to \infty \) for all \( h \), hence we cannot have \( b_T \sim T-a \) for constant \( a \), while in general

\(^5\)See Lemma A.1 in Hill (2013) and Lemma A.2 in Hill (2014a). An exact asymptotic approximation of a partial sum variance is available if the dependent stationary process has a positive spectrum at frequency zero (Ibragimov, 1962). The tail-trimmed processes in question, however, may not have a bounded spectrum (see Remark 3). Conversely, under Assumption A we can use Lemma 1 in Bradley (1992) to achieve a partial sum variance approximation this is sharp enough for higher order bias details. See Lemma A.1 in Appendix A.1 for the variance approximation.
Thus, when \( k \) exists, and a large gap is more difficult to approximate and therefore estimate using Karamata theory. Hence, when \( \hat{\theta}_T \) is heavy tailed.

The conclusion of slow \( k_{T,h} \rightarrow \infty \) is contrary to the finding of Theorem 2.2, that if \( \kappa < 4 \) then the rate of convergence is higher when \( k_{T,h} \rightarrow \infty \) 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 \( k_{T,h} \rightarrow \infty \).

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 drawn of size \( 2T \) drawn, and we detail the last \( T \) observations for analysis. The AR(1) models are:

\[
y_t = \phi_0 y_{t-1} + \epsilon_t \quad \text{with} \quad \phi_0 \in \{.00, .75, .90\},
\]

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 case it 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 \( \Theta \equiv [-.999, .999] \times [0, 100] \).\(^6\)

The GARCH model is \( x_t = \sigma_t \epsilon_t \) where \( \sigma_t^2 = \omega_0 + \alpha_0 \sigma_{t-1}^2 + \beta_0 \sigma_{t-1}^2 \), \( \omega_0 = 1 \), and \( \alpha_0, \beta_0 = [3, 6] \) or \([2, 7]\), and \( \epsilon_t \) is iid standard normal. It is easily verified that \( x_t \) has a power law tail with index \( \kappa_x \approx 4.1 \) when \( [\alpha_0, \beta_0] = [3, 6] \) and \( \kappa_x \approx 6.05 \) when \( [\alpha_0, \beta_0] = [2, 7] \).\(^7\) 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 \( \sigma_t^2 \equiv E[u_t^2] \). 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 \( y_t \) has tail index \( \kappa_y \approx 2.05 \) when \( [\alpha_0, \beta_0] = [3, 6] \) and \( \kappa_y \approx 2.03 \) when \( [\alpha_0, \beta_0] = [2, 7] \). Define \( \tilde{\Theta} \equiv [10^{-10}, 100] \times [0, .999] \times [0, .999] \). We optimize the FD-QML criteria on the subset \( \Theta \equiv \{ \theta \in \tilde{\Theta} : \alpha_c \neq \alpha \} \).

\(^6\)We use Matlab R2014a, where optimization is performed by the \textit{fmincon} routine.

\(^7\)The GARCH process \( \{x_t\} \) satisfies \( P(|x_t| > c) = dc^{-\kappa c}(1 + o(1)) \) and \( E[\alpha_0 \epsilon_t^2 + \beta_0 |\alpha|^{\kappa/2}] = 1 \). See, e.g., Basrak, Davis, and Mikosch (2002). We draw \( R = 100,000 \) iid standard normal \( \epsilon_t \) and compute \( \hat{\kappa}_x \equiv \arg \min_{\kappa \in \mathbb{K}} |1/R \sum_{i=1}^R \alpha_0 \epsilon_t^2 + \beta_0 |^{\kappa/2} - 1 \), where \( \mathbb{K} = \{.001, .002, \ldots, .10\} \). We repeat this 10,000 times and find the median and mean \( \hat{\kappa}_x \) are roughly 4.1 when \( [\alpha_0, \beta_0] = [3, 6] \) and roughly 6.05 when \( [\alpha_0, \beta_0] = [2, 7] \).
+ β ≤ .999).

We use optimal bias correction since, by simulation experiments not reported here, we find that tail-trimmed FD-QML estimates are strongly negatively biased if bias correction is not used. 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 \( h ≥ 1 \), and in the GARCH case \( \tilde{\gamma}_{T,h} = 0 \) ∀\( h ≥ 0 \).

Let \( t = 10^{-10} \). We use a trimming fractile \( k_{T,h} = \lfloor 2(\ln(T))^{1-\epsilon} \rfloor \) for each lag \( h = 0, ..., b_T \) with bandwidth \( b_T = [T^{.95}] \), and bias estimation fractile \( m_{T,h} = [8\ln(T)] \) and fractile function \( m_{T,h}(\xi) = [\xi m_{T,h}] \) over \( \xi ∈ [1, 2] \). Using lags \( b_T = [T^\delta] \) for \( \delta > .95 \) can lead to a breakdown in bias estimation for \( T ≤ 250 \) and \( h ≈ b_T \), 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 \( m_{T,h} \) due to the sample’s small size \( T - h \). The tail shape \( P(|\epsilon_t| > c) = (1 + c)^{-\xi} = O(c^{-\xi}) \) and fractiles \( \{k_{T,h}, m_{T,h}\} \) satisfy Assumption B’, in particular \( \{k_{T,h}, m_{T,h}\} \) is valid for asymptotically unbiased estimation in all cases by Theorems 3.1 and 3.3.

We compute the optimally bias-corrected \( \hat{\gamma}_{T,h}(\hat{Y}_{h,0}(k_{T,h})) \) in (22) based on the potentially biased tail-trimmed estimator \( \tilde{\gamma}_{T,h}(\hat{Y}_{h,0}(k_{T,h})) \), and the bias-corrected \( \tilde{\gamma}_{T,h}(\hat{Y}_{h,0}(k_{T,h})) = \tilde{\gamma}_{T,h}(\hat{Y}_{h,0}(k_{T,h})) + \hat{R}_{T,h}(\hat{\xi}_{T,h}) \) in (19). In this study, \( y_t y_{t-h} \) has symmetric tails for all \( h > 0 \), hence we compute \( \hat{R}_{T,0}(\hat{\xi}_{T,0}) \) in (17) and all other \( \hat{R}_{T,h}(\hat{\xi}_{T,h}) \) in (18). 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 \( T ∈ \{100, 250\} \) the computed fractiles are \( k_{T,h} = \{1, 1\} \) and \( m_{T,h} = \{37, 44\} \), with ranges \( m_{T,h}(\xi) ∈ \{[4, 74], [4, 88]\} \). Trimming just one observation in samples of size \( T ≤ 250 \) corrects for heavy tails, while using substantially more observations for tail index estimation (e.g. \( m_{T,h}(\xi) ∈ \{[4, 74]\} \) leads to a 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 \( 1/R \sum_{i=1}^{R} (\hat{\theta}_{T,i}^{(r)} - \theta_{0,i})^2 \) for all estimators, where \( R = 10,000 \). We then compute the standardized variable \( z_{T,i}^{(r)} = (\hat{\theta}_{T,i}^{(r)} - \theta_{0,i}) / s_{T,i} \) with simulation variance \( s_{T,i}^2 = 1/R \sum_{i=1}^{R} (\hat{\theta}_{T,i}^{(r)} - 1/R \sum_{j=1}^{R} \hat{\theta}_{T,i}^{(r)})^2 \) and perform a Kolmogorov-Smirnov test of standard normality on the sampling distribution \( \{z_{T,i}^{(r)}\}_{r=1}^{R} \). 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 sample size \( T = 100 \). In Figure 1 we plot the trimmed and untrimmed ratios \( \frac{\tilde{\gamma}_{T,0}^{(obc)}}{\tilde{\gamma}_{T,0}} \) and \( \frac{\tilde{\gamma}_{T,h}^{(obc)}}{\tilde{\gamma}_{T,h}} \), where \( \tilde{\gamma}_{T,h} ≡ 1/T \sum_{t=h+1}^{T} y_t y_{t-h} \), over lags \( h = 1, ..., 80 \). The plots are the 2.5%, 50% and 97.5% quantiles of \( \frac{\tilde{\gamma}_{T,0}^{(obc)}}{\tilde{\gamma}_{T,0}} \), and the average difference \( \frac{\tilde{\gamma}_{T,0}^{(obc)}}{\tilde{\gamma}_{T,0}} - \tilde{\gamma}_{T,0}^{(obc)} \), over the 10,000 samples. Clearly \( \frac{\tilde{\gamma}_{T,0}^{(obc)}}{\tilde{\gamma}_{T,0}} \) matches \( \tilde{\gamma}_{T,0}^{(obc)} \), which demonstrates the sharpness of the optimally fitted bias estimator. Indeed, we plot \( \frac{\tilde{\gamma}_{T,h}^{(obc)}}{\tilde{\gamma}_{T,h}} \) and the difference separately because \( \frac{\tilde{\gamma}_{T,0}^{(obc)}}{\tilde{\gamma}_{T,0}} \) and \( \tilde{\gamma}_{T,0}^{(obc)} \) cannot be visually distinguished apart.

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 $\phi_0 \in \{0.00, 0.75\}$ 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 well 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.

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. We demonstrate by a simulation study 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.

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. Since most proof
We use \( k_{T,h} = k_T \) in the simulation study and show it works quite well.

In order to reduce the number of required cases, we assume mean-centering \( I(|y_t y_{t-h} - (T - h)^{-1} \sum_{t=1}^{T} y_t y_{t-h}| < c) \) for all lags \( h \), while asymptotic arguments are simpler without centering. Thus:

\[
\hat{\gamma}_{h,t}^{(0)} = \left| \frac{y_t y_{t-h}}{T-h} \sum_{t=h+1}^{T} y_t y_{t-h} \right| \quad \text{for } h = 0, ..., b_T, \text{ and } \ \hat{\gamma}_{h,(1)}^{(0)} \geq \hat{\gamma}_{h,(2)}^{(0)} \geq \cdots \hat{\gamma}_{h,(T-h)}^{(0)}.
\]

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

We repeatedly use the following central limit theorem for an arbitrary triangular array \( \{g_{t,s} : 1 \leq t \leq T\} \). Write \( F_{t,s} \equiv \sigma(g_{t,s} : s \leq t) \), and define mixing coefficients \( \hat{\alpha}_h \equiv \sup_{T \in \mathbb{N}} \sup_{1 \leq t \leq T} \sup_{A \subset F_{t,s}, A \subset F_{t,s}^c} |P(A \cap B) - P(A)P(B)| \) and interlaced maximal correlation coefficients \( \hat{\rho}_h \equiv \sup_{T \in \mathbb{N}} \sup_{A \subset T, h, A \subset T, h} \rho(\sigma(g_{t,s} : t \in \Xi_h), \sigma(g_{t,s} : s \in \Xi_h)) \) where \( \Xi_{T,h} \) and \( \Xi_{T,\cdot} \) are non-empty subsets of \( \{1, ..., T\} \) with inf \( s \in \Xi_{T,h}, t \in \Xi_{T,\cdot} \{ |s-t| \} \geq h \). By convention \( \sup_{\Xi_{T,h}, \Xi_{T,\cdot}} \rho(\sigma(g_{t,s} : t \in \Xi_h), \sigma(g_{t,s} : s \in \Xi_h)) = 0 \) if \( h \geq T \).

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

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

**Proof.** By Assumption A.1 and measurability we have \( \hat{\alpha}_h \to 0 \) and \( \hat{\rho}_h^* \to 0 \), hence \( v_T^2 \sim K \sum_{t=1}^{T} E[g_{t,s}^2] \) by an application of Lemma 1 in Bradley (1992). This proves (a). For (b) note \( \hat{\alpha}_h \to 0 \), \( v_T^2 \sim K \sum_{t=1}^{T} E[g_{t,s}^2] \) and (A.1) imply \( 1/v_T \sum_{t=1}^{T} g_{t,s} \xrightarrow{d} N(0,1) \) by Theorem 2.2 in Peligrad (1996). 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) \( |\hat{\gamma}_{T,h}^{(0)}(c_{T,h}) - E[\hat{\gamma}_{T,h}^{(0)}(c_{T,h})]| \xrightarrow{p} 0 \); and (b) if additionally Assumption B holds then \( (T/||S_T||)^{1/2} \times \hat{\gamma}_{T,h}(\hat{\gamma}_{h,(k_T)}) - \hat{\gamma}_{T,h}(c_{T,h}) \xrightarrow{p} 0 \).

**Proof.**
Claim (a): Define $\Theta_t \equiv \sigma(y_t : t \leq t)$ and $\psi_{h,t}(c) \equiv y_{t_0}y_{t-h}I(|y_{t_0}y_{t-h} - E[y_{t_0}y_{t-h}] < c)$. Under Assumption A $\psi_{h,t}(ct,h)$ is uniformly $L_{1+\varepsilon}$-bounded for tiny $\varepsilon > 0$, and $\alpha$-mixing. Hence $\{\psi_{h,t}(ct,h), \Theta_t\}$ forms an $L_1$-mixing array (Lemma 2.1 in McLeish, 1975; Example 3.4 in Andrews, 1988). In view of $h \leq b_T$ and $T - b_T \to \infty$ by Assumption C, it follows $1/T \sum_{t=1}^{T} \psi_{h,t}(ct,h) - E[\psi_{h,t}(ct,h)] \xrightarrow{p} 0$ by Theorem 2 of Andrews (1988). The claim follows instantly.

Claim (b): Write $\Theta^2_T \equiv E[\psi^2_{0,t}(ct,0)]$, $\gamma_{T,h} \equiv (T-h)^{-1} \sum_{t=h+1}^{T} y_{t}y_{t-h}$ and $\gamma_h \equiv E[y_{t_0}y_{t-h}]$. By Lemma A.9.b $||S_T|| \sim K\Theta^2_T$, hence we need only prove

$$A_T \equiv \frac{1}{T^{1/2} \Theta_T} \sum_{t=h+1}^{T} y_{t}y_{t-h} \left\{ I(|y_{t}y_{t-h} - \gamma_T,h| < \gamma^0_{h,(k_T)} - I(|y_{t}y_{t-h} - \gamma_h| < c_{T,h}) \right\} = o_p(1).$$

We have

$$A_T = \frac{1}{T^{1/2} \Theta_T} \sum_{t=h+1}^{T} y_{t}y_{t-h} \left\{ I(|y_{t}y_{t-h} - \gamma_T,h| - c_{T,h} \leq 0) - I(|y_{t}y_{t-h} - \gamma_h| - c_{T,h} \leq 0) \right\}$$

$$+ \frac{1}{T^{1/2} \Theta_T} \sum_{t=h+1}^{T} y_{t}y_{t-h} \left\{ I \left( \gamma_T,h \right) - \gamma^0_{h,(k_T)} \right\} - I \left( \gamma_T,h \right) = B_T + C_T.$$ 

Consider $B_T$. The trimming function $I(u) \equiv I(u < 0)$ can be approximated by a continuous, differentiable, uniformly bounded positive function $J_N(u)$ that has a uniformly bounded derivative $D_N(u) \equiv (\partial/\partial u)J_N(u)$, where $N \in \mathbb{N}$ guides the approximation: $\lim_{N \to \infty} \sup_{u \in \mathbb{R}} |J_N(u) - I(u)| = 0$ and $\lim_{N \to \infty} \sup_{u \in U} |D_N(u)| = 0$ for any compact subset $U$ of $\mathbb{R}$. 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 $N = N_T \to \infty$ as $T \to \infty$ fast enough to ensure

$$B_T = \frac{1}{T^{1/2} \Theta_T} \sum_{t=h+1}^{T} y_{t}y_{t-h} \left\{ J_N \left( \left( y_{t}y_{t-h} - \gamma_T,h \right)^2 - c_{T,h}^2 \right) - J_N \left( \left( y_{t}y_{t-h} - \gamma_h \right)^2 - c_{T,h}^2 \right) \right\} + o_p(1).$$

Hence, by the mean-value-theorem for some $\gamma^*_T,h, |\gamma^*_T,h - \gamma_h| \leq |\gamma_T,h - \gamma_h|:

$$B_T = -2 \frac{1}{T^{1/2} \Theta_T} \sum_{t=h+1}^{T} y_{t}y_{t-h} \times D_N \left( \left( y_{t}y_{t-h} - \gamma^*_T,h \right)^2 - c_{T,h}^2 \right) \times \left( y_{t}y_{t-h} - \gamma_T,h \right) \left( \gamma_T,h - \gamma_h \right) + o_p(1).$$

Mixing Assumption A.1 implies mixing in the ergodic since, and therefore ergodicity (see, e.g., Petersen, 1983). Stationarity, ergodicity, $L_p$-boundedness for $p > 2$, and $h \leq b_T$ with $T - b_T \to \infty$ imply $\gamma_T,h \xrightarrow{p} \gamma_h = E[y_{t_0}y_{t-h}]$, hence $\gamma^*_T,h \xrightarrow{p} E[y_{t_0}y_{t-h}]$. Further, max$_{1 \leq t \leq T} |y_t| = O_p(T^{1/2})$ by stationarity and square integrability, and $|y_{t_0}y_{t-h} - \gamma^*_T,h| \neq c_{T,h}$ a.s. by distribution continuity. Thus, since max$_{1 \leq t \leq T} |D_N((y_{t_0}y_{t-h} - \gamma^*_T,h)^2 - c_{T,h}^2)| \xrightarrow{p} 0$ as fast as we choose by setting $N_T \to \infty$ as $T \to \infty$.
fast enough, it follows $\mathfrak{B}_T \overset{p}{\to} 0$.

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

Finally, we characterize order statistics and tail exponent estimators.

**Lemma A.3 (order statistic)** Under Assumptions A.1, B and C $\hat{\gamma}_{h, (k_T)}^{(0)}/C_{T,h} = 1 + O_p(1/k_T^{1/2})$ \( \forall h \in \{0, ..., b_T\} \).

**Proof.** $E[|y_t y_{t-h}|^1] < \infty$ 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 $T - b_T \to \infty$ under Assumption C. 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 $E[\sum_{t=h+1}^{T} y_t y_{t-h}]^{1+\iota} = O(T)$ by stationarity, $L_{1+\iota}$-boundedness, and an application of Lemma 2 in Hansen (1991, 1992). Since by Assumption C $k_T$ is slowly varying, it follows $k_T = o(T^a) \forall a > 0$ (see Resnick, 1987, Chapter 0), hence $(T - h)^{-1} \sum_{t=h+1}^{T} y_t y_{t-h} = E[|y_t y_{t-h}|] + O_p(1/k_T^{1/2})$. The claim now follows from 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). QED.

**Lemma A.4 (tail exponents)** Define $\hat{\kappa}_{h,m T,h}^{-1} = 1/m_{T,h} \sum_{j=1}^{m_{T,h}} \ln (\hat{\gamma}_{h,(j)}^{(0)}/\hat{\gamma}_{h,(m_{T,h} + 1)}^{(0)})$ and $\hat{d}_{h,m_{T,h}} = (m_{T,h}/T)(\hat{\gamma}_{h,(m_{T,h})})^{\hat{\kappa}_{h,m_{T,h}}^{-1}}$. Let Assumptions A.1, B' and C hold. Then $\hat{\kappa}_{h,m_{T,h}} = \kappa_{h,0} + O_p(1/m_{T,h}^{1/2})$ and $\hat{d}_{h,m_{T,h}} = d_{h,0} + O_p(1/m_{T,h}^{1/2})$.

**Proof.** The claim therefore follows from Theorem 2.1 and Lemma 2.2 in Hill (2014b). QED.

**Remark 14** 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 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 $E[y_t^2] < \infty$ then by Assumption A.1 and, e.g., Lemma 2.1 in McLeish (1975), $(T - h)^{-1} \sum_{t=h+1}^{T} y_t y_{t-h} = E[y_t y_{t-h}] + O_p(1/T^{1/2}) = E[y_t y_{t-h}] + o_p(1/K^{1/2})$, 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), and footnotes 8 and 9.

---

8 Hill (2014b) proves Hill (1975)'s tail index estimator is asymptotically normal for filtered weakly dependent sequences, e.g. $\hat{\gamma}_{h,(t)}$, by extending arguments in Hill (2010, Theorem 2) for distribution tails that satisfy a second order regular variation property. The same set of arguments extend to intermediate order sequences $\hat{\gamma}_{h,(k_T)}^{(0)}$, while asymptotics for $\hat{\gamma}_{h,(k_T)}^{(0)}$ does not require a second order regular variation property (see Hill, 2010, Lemma 3).

9 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 $k_T^{1/2}$-Gaussian asymptotics for Hill (1975)'s estimator 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. The claim for $\hat{d}_{h,m_{T,h}}^{-1}$ then follows instantly from $k_T^{1/2}$-asymptotics for $\hat{\gamma}_{h,(m_{T,h})}$ and $\hat{\kappa}_{h,m_{T,h}}^{-1}$, and the mapping theorem.
A.2 Supporting Lemmas for Robust FD-QML

Throughout \( \{r_{M,T}\}_{M,T \in \mathbb{N}} \) is a non-random double array of finite constants 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 the sample moments, spectrum and spectral density estimators

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

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

\[
\gamma^*_T, h(c) \equiv \begin{cases} \frac{T-h}{T-h-k_T} E[y_t y_{t-h} I (|y_t y_{t-h} - E[y_t y_{t-h}]| < c)] & \text{if } P(y_t y_{t-h} > 0) < 1 \\ E[y_t y_{t-h} I (|y_t y_{t-h}| < c)] & \text{if } P(y_t y_{t-h} > 0) = 1 \\ 0 & \text{if } |h| > b_T \end{cases},
\]

\[
f^*_T(\lambda) \equiv \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma^*_T, h(ct_T, h)e^{-i\lambda h},
\]

\[
\hat{I}^*_T(\lambda) \equiv \frac{1}{2\pi} \left( \hat{\gamma}^*_T, 0(\hat{\gamma}^{(0)}_0, (k_T)) + 2 \sum_{h=1}^{b_T} \hat{\gamma}^*_T, h(\hat{\gamma}^{(0)}_{h,(k_T)}) \cos(\lambda h) \right),
\]

\[
I^*_T(\lambda) \equiv \frac{1}{2\pi} \left( \hat{\gamma}^*_T, 0(ct_T, 0) + 2 \sum_{h=1}^{b_T} \hat{\gamma}^*_T, h(ct_T, h) \cos(\lambda h) \right).
\]

Define criteria and estimators:

\[
\hat{\theta}^*_T = \text{arg min}_{\theta \in \hat{\Omega}} \left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} \left( \ln f(\lambda, \theta) + \frac{\hat{I}^*_T(\lambda)}{f(\lambda, \theta)} \right) d\lambda \right\} = \text{arg min}_{\theta \in \hat{\Omega}} \left\{ \hat{Q}^*_T(\theta) \right\} \quad (A.2)
\]

\[
\hat{\theta}^*_T = \text{arg min}_{\theta \in \hat{\Omega}} \frac{1}{T} \sum_{j \in \mathcal{F}} \left( \ln f(\lambda_j, \theta) + \frac{\hat{I}^*_T(\lambda_j)}{f(\lambda_j, \theta)} \right) = \text{arg min}_{\theta \in \hat{\Omega}} \left\{ \hat{Q}^*_T(\theta) \right\},
\]

where \( \mathcal{F} \equiv (-T/2, T/2] \cap \mathbb{Z} \setminus \{0\} \), and recall \( \Omega, S_T \) and \( \Psi_T \) defined in (7). Recall also:

\[
\kappa = \text{arg sup} \{ \alpha > 0 : E|y_t|^\alpha < \infty \}.
\]
We exploit the following implications of Karamata’s Theorem under Assumption B:

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

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

In the Paretian case of Assumption B', by integration is it easily verified that:

\[ \hat{\mathcal{L}}_4(T) = d_{0,0} \ln(T). \]

First, we can always work with \( \hat{\theta}_T^* \).

**Lemma A.5 (equivalent estimators)** Under Assumptions A-C \( T^{1/2} \mathcal{V}_{-1/2}^{-1} (\hat{\theta}_T^* - \hat{\theta}_T) \xrightarrow{p} 0 \) 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 \( \lim \inf_{T \to \infty} ||\mathcal{S}_T|| > 0 \) by Lemma A.9. Lemma A.6 can be easily generalized to \( \hat{Q}_T^*(\hat{\theta}_T) \) to show \( \langle \partial^2/\partial \theta \partial \theta' \rangle \hat{Q}_T^*(\hat{\theta}_T) \xrightarrow{p} \Omega \) for any \( \hat{\theta}_T \xrightarrow{p} \theta_0 \). Similarly, Lemma A.7 can be generalized to prove \( T^{-1/2} \mathcal{S}_T^{-1/2} \sum_{j \in F} (\hat{\mathcal{I}}_T^*(\lambda_j) - \mathcal{I}_T^*(\lambda_j)) \mathcal{W}(\lambda_j) \xrightarrow{p} 0 \). 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^*, \hat{\theta}_T\} and first order expansions around \( \theta_0 \), we have:

\[
T^{1/2} \mathcal{V}_T^{-1/2} \left( \hat{\theta}_T^* - \hat{\theta}_T \right) \nonumber
= -T^{1/2} \mathcal{S}_T^{-1/2} \left\{ \frac{1}{T} \sum_{j \in F} (\mathcal{I}_T^*(\lambda_j) - f(\lambda_j)) \mathcal{W}(\lambda_j) - \frac{1}{2\pi} \int_{-\pi}^{\pi} (\mathcal{I}_T^*(\lambda) - f(\lambda)) \mathcal{W}(\lambda) d\lambda \right\} + o_p(1) \\
= -T^{1/2} \mathcal{S}_T^{-1/2} \left\{ \frac{1}{T} \sum_{j \in F} (f(T)(\lambda_j) - f(\lambda_j)) \mathcal{W}(\lambda_j) \right\} \\
- T^{1/2} \mathcal{S}_T^{-1/2} \left\{ \frac{1}{T} \sum_{j \in F} (\mathcal{I}_T^*(\lambda_j) - f(T)(\lambda_j)) \mathcal{W}(\lambda_j) - \frac{1}{2\pi} \int_{-\pi}^{\pi} (\mathcal{I}_T^*(\lambda) - f(\lambda)) \mathcal{W}(\lambda) d\lambda \right\} + o_p(1) \\
= -A_{T,1} - A_{T,2} + o_p(1),
\]

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

---

\(^{10}\)The case \( \kappa = 4 \) is shown as follows. Let \( \hat{\mathcal{L}}(c) \) be a slowly varying function that may be different in different places. By Assumption B, a change of variables, and Karamata’s Theorem: \( \hat{E}_T \equiv E[(y_t^4 - E[y_t^4])^2 I(|y_t^4 - E[y_t^4]| < c_{T,0})] = K \int c_{T,0}^{2/\alpha} \mathcal{L}(u) u^{-1} du = \hat{\mathcal{L}}(c_{T,0}^2) \). By Assumption B, \( \kappa_{0,0} = \kappa/2 \), and properties of regularly varying functions: \( c_{T,0} \sim (T/k_T)^{2/\alpha} \hat{\mathcal{L}}(T/k_T) \). Further, \( \hat{\mathcal{L}}(T/k_T) \sim \hat{\mathcal{L}}(T) \) because \( \hat{\mathcal{L}} \) is slowly varying and \( k_T > 0 \). Hence, \( c_{T,0} \sim (T/k_T)^{2/\alpha} \hat{\mathcal{L}}(T) \). Therefore \( \hat{E}_T = \hat{\mathcal{L}}(c_{T,0}^2) = \hat{\mathcal{L}}((T/k_T)^{2/\alpha} \hat{\mathcal{L}}(T)) \), where \( \hat{\mathcal{L}}((T/k_T)^{2/\alpha} \hat{\mathcal{L}}(T)) / \hat{\mathcal{L}}((T/k_T)^{2/\alpha} \hat{\mathcal{L}}(T)) \to 1 \) by slow variation, hence \( \hat{E}_T \) is a slowly varying function of \( T \).
Next, letting \( f^*_T(M)(\lambda) \) be the \( M \)-th Cesáro sum of the Fourier series of \( f^*_T(\lambda) \) and noting \( E[I^*_T(\lambda)] = f^*_T(\lambda) \), we may write:

\[
\mathcal{A}_{T,2} = T^{1/2}S_T^{-1/2} \left\{ \frac{1}{T} \sum_{j \in \mathcal{F}} (I^*_T(\lambda_j) - E[I^*_T(\lambda_j)]) \varpi(\lambda_j) - \frac{1}{2\pi} \int_{-\pi}^{\pi} (I^*_T(\lambda) - E[I^*_T(\lambda)]) \varpi(\lambda)d\lambda \right\} 
+ T^{1/2}S_T^{-1/2} \left( \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\lambda) \varpi(\lambda)d\lambda - \frac{1}{T} \sum_{j \in \mathcal{F}} f(\lambda_j) \varpi(\lambda_j) \right) 
- T^{1/2}S_T^{-1/2} \left( \frac{1}{2\pi} \int_{-\pi}^{\pi} f^*_T(\lambda) \varpi(\lambda)d\lambda - \frac{1}{T} \sum_{j \in \mathcal{F}} f^*_T(\lambda_j) \varpi(\lambda_j) \right) 
+ T^{1/2}S_T^{-1/2} \left( \frac{1}{T} \sum_{j \in \mathcal{F}} (f^*_T(\lambda_j) - f^*_T(\lambda)) \varpi(\lambda_j) \right) + T^{1/2}S_T^{-1/2} \left( \frac{1}{T} \sum_{j \in \mathcal{F}} (f^*_T(\lambda) - f^*_T(\lambda)) \varpi(\lambda) \right) 
= C_{1,T} + C_{2,T} - C_{3,T} + C_{4,T} - C_{5,T} + C_{6,T}.
\]

By Hölder continuity and boundedness of \( \varpi(\lambda) \), \( C_{4,T} = O(T^{1/2-\alpha}) = o(1) \). Similarly, in view of \( \lim_{T \to \infty} \sup_{\lambda \in [-\pi, \pi]} |f^*_T(\lambda) - f(\lambda)| = 0 \) by dominated convergence, and by the construction of the Cesáro sum approximation, it follows \( C_{5,T} = o(1) \) and \( C_{6,T} = o(1) \).

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

\[
\|C_{2,T}\| \leq K \left( \frac{T}{\|S_T\|} \right)^{1/2} \sup_{j \in \mathcal{F}} \sup_{|\lambda - \lambda_j| \leq 2\pi/T} \|f(\lambda)\varpi(\lambda) - f(\lambda_j)\varpi(\lambda_j)\| = O(T^{1/2-\alpha}) = o(1),
\]

and likewise \( C_{3,T} = o(1) \). The remaining term \( C_{1,T} \) 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 \( \theta^*_T \overset{p}{\to} \theta_0 \). Under Assumptions A-C \((\partial^2 / \partial \theta \partial \theta') \hat{Q}^*_T(\theta^*_T) \overset{p}{\to} \Omega\).

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

\[
\left\| \frac{\partial^2}{\partial \theta \partial \theta'} \hat{Q}^*_T(\theta^*_T) - \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\partial}{\partial \theta} \ln f(\lambda) \frac{\partial}{\partial \theta} \ln f(\lambda)d\lambda \right\| 
\leq \left\| \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\partial}{\partial \theta} \ln f(\lambda) \frac{\partial}{\partial \theta} \ln f(\lambda) d\lambda - \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\partial}{\partial \theta} \ln f(\lambda, \theta^*_T) \frac{\partial}{\partial \theta} \ln f(\lambda, \theta^*_T) d\lambda \right\| 
+ \left\| \frac{1}{2\pi} \int_{-\pi}^{\pi} (I^*_T(\lambda) - f(\lambda)) \frac{\partial}{\partial \theta} \varpi(\lambda)d\lambda \right\| 
+ \left\| \frac{1}{2\pi} \int_{-\pi}^{\pi} (I^*_T(\lambda) - f(\lambda)) \left( \frac{\partial}{\partial \theta} \varpi(\lambda, \theta^*_T) - \frac{\partial}{\partial \theta} \varpi(\lambda) \right)d\lambda \right\|
\]
sum of the Fourier series of uniformly bounded

The property  

Proof.

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 \( \sup_{\theta \in \Theta} |(T/||S_T||)^{1/2} \int_{-\pi}^{\pi} (\hat{I}_T^*(\lambda) - \hat{I}_T(\lambda)) \omega(\lambda, \theta) d\lambda| \xrightarrow{p} 0 \) where \( T/||S_T|| \to \infty \).

Proof. The property \( T/||S_T|| \to \infty \) 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 \( M \in \mathbb{N} \):

\[
\left( \frac{T}{||S_T||} \right)^{1/2} \int_{-\pi}^{\pi} \left( \hat{I}_T^*(\lambda) - \hat{I}_T(\lambda) \right) \omega(\lambda, \theta) d\lambda
= \left( \frac{T}{||S_T||} \right)^{1/2} \int_{-\pi}^{\pi} \left( \hat{I}_T^*(\lambda) - \hat{I}_T(\lambda) \right) \omega(M)(\lambda, \theta) d\lambda + o_p(1)
= \frac{1}{2\pi} \sum_{h=-M}^{M} \left( 1 - \frac{|h|}{M} \right) \omega_h(\theta) \left( \frac{T}{||S_T||} \right)^{1/2} \left\{ \hat{\gamma}^*_T(\hat{\gamma}^{(0)}_{T,h(k_T)}) - \hat{\gamma}^*_T(c_{T,h}) \right\} + o_p(r_{M,T})
= \mathcal{A}_{M,T}(\theta) + o_p(r_{M,T}).
\]

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 \( (T/||S_T||)^{1/2} \times |\hat{\gamma}^*_T(\hat{\gamma}^{(0)}_{T,h(k_T)}) - \hat{\gamma}^*_T(c_{T,h})| \xrightarrow{p} 0 \) 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 \( \sup_{\theta \in \Theta} |\int_{-\pi}^{\pi} (\hat{I}_T^*(\lambda) - f(\lambda)) \omega(\lambda, \theta) d\lambda| \xrightarrow{p} 0 \).
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 \( \int_{-\pi}^{\pi} I_T^*(\lambda) d\lambda = 1/T \sum_{t=1}^{T} y_t^2 I(y_t^2 - E[y_t^2]) < c_{T,0} \xrightarrow{p} E[y_t^2], \) and by supposition $\int_{-\pi}^{\pi} \sup_{\theta \in \Theta} |\omega(\lambda, \theta)| d\lambda < \infty$. Hence, by Fejer's theorem:

\[
\int_{-\pi}^{\pi} \{ I_T^*(\lambda) - f(\lambda) \} \omega(\lambda, \theta) d\lambda = \frac{1}{2\pi} \int_{-\pi}^{\pi} \{ I_T^*(\lambda) - f(\lambda) \} \omega_{(M)}(\lambda, \theta) d\lambda
\]

\[
+ \frac{1}{2\pi} \int_{-\pi}^{\pi} \{ I_T^*(\lambda) - f(\lambda) \} (\omega_{(M)}(\lambda, \theta) - \omega(\lambda, \theta)) d\lambda
\]

\[= \frac{1}{2\pi} \int_{-\pi}^{\pi} \{ I_T^*(\lambda) - f(\lambda) \} \omega_{(M)}(\lambda, \theta) d\lambda + o_p(r_{M,T}). \]

By construction $(2\pi)^{-1} \int_{-\pi}^{\pi} (I_T^*(\lambda) - f(\lambda)) \omega_{(M)}(\lambda, \theta) d\lambda = A_{1,T}(M, \theta) + A_{2,T}(M, \theta)$ where

\[
A_{1,T}(M, \theta) = \frac{1}{(2\pi)^2} \sum_{h=-M}^{M} \left( 1 - \frac{|h|}{M} \right) \omega_h(\theta) \{ \gamma^\ast_{T,h}(c_{T,h}) - E[y_t y_{t-h}] \}
\]

\[
A_{2,T}(M, \theta) = \frac{1}{(2\pi)^2} \sum_{h=-M}^{M} \left( 1 - \frac{|h|}{M} \right) \omega_h(\theta) \{ \gamma^\ast_{T,h}(c_{T,h}) - \gamma^\ast_{T,h}(c_{T,h}) \}.
\]

Observe $\lim_{M \to \infty} \sup_{\theta \in \Theta} |\sum_{-M}^{M} (1 - |h|/M) \omega_h(\theta)| = 2\pi \int_{-\pi}^{\pi} \sup_{\theta \in \Theta} |\omega(\lambda, \theta)| d\lambda < \infty$ by uniform boundedness and integrability, $\gamma^\ast_{T,h}(c_{T,h}) \to E[y_t y_{t-h}]$ by negligibility and dominated convergence, and $|\gamma^\ast_{T,h}(c_{T,h}) - \gamma^\ast_{T,h}(c_{T,h})| \xrightarrow{p} 0$ by Lemma A.2.a. Therefore, $\lim_{M \to \infty} \sup_{\theta \in \Theta} |A_{1,T}(M, \theta)| \to 0$ and $\lim_{M \to \infty} \sup_{\theta \in \Theta} |A_{2,T}(M, \theta)| \xrightarrow{p} 0$, hence $\sup_{\theta \in \Theta} |\int_{-\pi}^{\pi} (I_T^*(\lambda) - f(\lambda)) \omega(\lambda, \theta) d\lambda| \xrightarrow{p} 0$. 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. Recall $f_T^*(\lambda) \equiv (2\pi)^{-1} \sum_{h=-\infty}^{\infty} \gamma^\ast_{T,h}(c_{T,h}) e^{-i\lambda h}$, let $\omega_h$ be the $h^{th}$ Fourier coefficient of $\omega(\lambda)$, and define

\[
\tilde{B}_T \equiv \int_{-\pi}^{\pi} \omega(\lambda) \{ f_T^*(\lambda) - f(\lambda) \} d\lambda = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \omega_h \{ \gamma^\ast_{T,h}(c_{T,h}) - E[y_t y_{t-h}] \}.
\]

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

a. $(2\pi)^{-1/2} S_{T}^{-1/2} (\int_{-\pi}^{\pi} (I_T^*(\lambda) - f(\lambda)) \tilde{\omega}(\lambda) d\lambda - \tilde{B}_T) \sim (2\pi)^{-1/2} S_{T}^{-1/2} (\int_{-\pi}^{\pi} (I_T^*(\lambda) - f_T^*(\lambda)) \tilde{\omega}(\lambda) d\lambda - \tilde{B}_T) = N(0, I_k)$.

b. $||S_T|| \sim KE[y_t^4 I(y_t^2 \leq c_{T,0})]$ hence $\inf_{T \to \infty} ||S_T|| > 0$.

c. If $k > 4$ then $||S_T|| \sim K\tilde{L}_d(T)$ for slowly varying $\tilde{L}_d(T)$ in (A.3), and if $k \in (2, 2.4)$ then $||S_T|| \sim K(T/k_T)^{4/k-1}$, hence $T/||S_T|| \to \infty \forall k > 2$.

Proof. Write $\psi_{h,t}(c) \equiv y_t y_{t-h} I(|y_t y_{t-h} - E[y_t y_{t-h}]| \leq c)$ where $\psi_{h,t}(c) = 0$ for $t \leq h$. 

30
Claim (a). Let $\omega_{(M)}(\lambda, \theta)$ and $\omega_h(\theta)$ be as in the proof of Lemma A.7. Let $S_{M,T}$ denote the covariance matrix of

$$T^{1/2} \frac{1}{(2\pi)^T} \sum_{h=-M}^{M} \left(1 - \frac{|h|}{M}\right) \omega_h \{\hat{\gamma}^*_T(h, c_{T,h}) - \gamma^*_T(h, c_{T,h})\}.$$ 

By the same Cesàro sum argument exploited above, note:

$$T^{1/2} S_{T}^{-1/2} \frac{1}{2\pi} \left( \int_{-\pi}^{\pi} (I_T^*(\lambda) - f(\lambda)) \varpi(\lambda) d\lambda - \tilde{B}_T \right) = T^{1/2} S_{T}^{-1/2} \frac{1}{2\pi} \int_{-\pi}^{\pi} (I_T^*(\lambda) - f_T^*(\lambda)) \varpi_{(M)}(\lambda)d\lambda + o_p(r_{M,T})$$

$$= (S_{T}^{-1/2} S_{M,T}^{1/2}) \times S_{M,T}^{-1/2} \frac{1}{(2\pi)^T} \sum_{h=-M}^{M} \left(1 - \frac{|h|}{M}\right) \omega_h \{\hat{\gamma}^*_T(h, c_{T,h}) - \gamma^*_T(h, c_{T,h})\}$$

$$= (S_{T}^{-1/2} S_{M,T}^{1/2}) \times Z_{M,T} + o_p(r_{M,T}),$$

say. By construction of the Cesàro sums and $S_T$ we have $\lim_{M \to \infty} S_{T}^{-1} S_{M,T} = I_k$. The claim therefore follows if we show $Z_{M,T} \xrightarrow{d} N(0, I_k)$ 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 $\xi' Z_{M,T} \xrightarrow{d} N(0, 1)$ as $T \to \infty$ for any $\xi \in \mathbb{R}^k$, $\xi' \xi = 1$.

By construction $\hat{\gamma}^*_T(h, c_{T,h}) = \hat{\gamma}^*_T(-h, c_{T,h})$, $\gamma^*_T(h, c_{T,h}) = \gamma^*_T(-h, c_{T,h})$, and by convention $\hat{\gamma}^*_T(h, c_{T,h}) = \gamma^*_T(h, c_{T,h}) = 0$ if $|h| > b_T$. Write $\tilde{w}_h = w_{-h} = w_h$ for $h \neq 0$ and $\tilde{w}_0 = w_0$. Define

$$\zeta_{M,T,h}(\xi) =\begin{cases} \frac{1}{(2\pi)^T} \left(1 - \frac{h}{M}\right) (\xi')^{1/2} S_{M,T}^{-1/2} \tilde{w}_h \times (\xi')^{1/2} S_{M,T} \xi & \text{for } h \in \{0, ..., M\} \\ 0 & \text{for } h > M \end{cases}$$

$$e_{T,h,t}(\xi) = \frac{1}{(\xi')^{1/2}} \left\{ \psi_{h,t}(c_{T,h}) - E[\psi_{h,t}(c_{T,h})] \right\}$$

$$\chi_{M,T,t}(\xi) = \sum_{h=0}^{\infty} \zeta_{h,M,T}(\xi) e_{T,h,t}(\xi) = \sum_{h=0}^{M} \zeta_{h,M,T}(\xi) e_{T,h,t}(\xi)$$

Then

$$\xi' Z_{M,T} = \frac{1}{T^{1/2}} \sum_{t=1}^{T} \chi_{M,T,t}(\xi) \times (1 + o_p(1)).$$

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

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

31
Next, by Lemma A.1.b $1/T^{1/2} \sum_{t=1}^{T} X_{M,T,t} \xrightarrow{d} N(0,1)$ provided Lindeberg condition (A.1) holds for $X_{M,T,t}$. By stationarity and $E(1/T^{1/2} \sum_{t=1}^{T} X_{M,T,t})^2 \sim K$ it suffices to show $E[X_{M,T,t} I(X_{M,T,t} > T^2)] \to 0 \forall \epsilon > 0$. Use the triangle inequality, $\lim_{M \to \infty} S_{M,T}^{-1} S_{M,T} = I_k, ||S_T|| \sim KE[\psi_{0,t}^2(c_T,0)]$ by Claim (b), $\liminf_{T \to \infty} E[\psi_{0,t}^2(c_T,0)] > 0$ by negligibility and distribution non-degenerateness, and $|\psi_{t}(c_T,0)| \leq c_T$, by construction, to deduce for some $\vartheta \in \mathbb{R}^k$, $\vartheta' \vartheta = 1$:

$$|X_{M,T,t}| \leq \sum_{h=0}^{M} \xi_{h,M,T} \xi_{T,h,t} | \leq K \sum_{h=0}^{M} \left( 1 - \frac{h}{M} \right) (\vartheta' \vartheta_h) \times \frac{c_T^2}{E[\psi_{0,t}^2(c_T,0)]} \frac{1}{2}$$

Under power law Assumption B $c_T \leq L(0)(T)^{1/\kappa_0} (T/k^T)^{1/\kappa_0}$, and by construction $\kappa_0 \geq \kappa_0$. Hence $c_T \leq L(0)(T)^{1/\kappa_0} (T/k^T)^{1/\kappa_0}$. Since under Assumption A $\limsup_{M \to \infty} K \sum_{h=0}^{M} \left( 1 - \frac{h}{M} \right) (\vartheta' \vartheta_h)$

$$\leq K \sum_{h=0}^{M} \left( 1 - \frac{h}{M} \right) (\vartheta' \vartheta_h) \times \frac{c_T^2}{E[\psi_{0,t}^2(c_T,0)]} \frac{1}{2}$$

If we prove $\mathfrak{A}_T = o(T^{1/2})$, then $X_{M,T,t} < T^2$ a.s. for all $T \geq N_\epsilon$ and some $N_\epsilon < 1$, hence $E[X_{M,T,t} I(X_{M,T,t} > T^2)] = 0 \forall T \geq N_\epsilon$ which completes the proof. If $\kappa > 3$ then $\kappa_0 \geq \kappa_0$, hence $\mathfrak{A}_T \leq K \mathcal{L}(T)^{1/\kappa_0} (T/k^T)^{1/\kappa_0}$. Conversely, if $\kappa < 2$ then $\kappa_0 < 2$ hence by Karamata theory (A.3) $E[\psi_{0,t}^2(c_T,0)] \sim K(T/k^T)^{2/\kappa_0-1}$. Hence $\mathfrak{A}_T \sim K \mathcal{L}(T)^{1/\kappa_0} (T/k^T)^{1/\kappa_0-1} = \mathcal{L}(T)^{1/\kappa_0} (T/k^T)^{1/\kappa_0-1} = o(T^{1/2})$.

Claim (b). Define $\Psi_{T,h,t} = \psi_{t}(c_T,0) - E[\psi_{t}(c_T,0)], \eta_{T,h} = 1 - h/M$ and $\vartheta_T(\xi) = \xi' S_t^{-1/2} (\xi' S_t^{-1/2})^{-1/2}$. By the arguments above $S_{M,T} S_{T}^{-1} \sim I_k$ and $1 = E(1/T^{1/2} \sum_{t=1}^{T} X_{M,T,t}(\xi))^2 \sim KE[X_{M,T,t}(\xi)]$, and by another application of Lemma A.1.a: $E[X_{M,T,t}(\xi)] \sim \sum_{h=0}^{M} \xi_{h,M,T} \xi_{T,h,t} |E[\xi_{T,h,t}(\xi)]$. Therefore

$$E[X_{M,T,t}(\xi)] \sim \sum_{h=0}^{M} \xi_{h,M,T} \xi_{T,h,t} |E[\xi_{T,h,t}(\xi)] \sim \sum_{h=0}^{M} \eta_{T,h} |E[\xi_{T,h,t}(\xi)] \sim K$$

hence

$$(\xi' S_t^{-1} \xi)^{-1} \sim K \sum_{h=0}^{M} \eta_{T,h} (\vartheta_T(\xi)) \xi_{h,M,T} \vartheta_T(\xi) \times E[\psi_{2,T,h,t}^2]$$

$$= KE[\psi_{2,T,h,t}^2] \left( (\vartheta_T(\xi)) \xi_{h,M,T} \vartheta_T(\xi) + \sum_{h=1}^{M} \eta_{T,h} (\vartheta_T(\xi)) \xi_{h,M,T} \vartheta_T(\xi) \right)$$

Notice $\vartheta_T(\xi) \vartheta_T(\xi) = 1$ for each $T$ and $\xi$. Further, $\kappa_0 \geq \kappa_0$, and $\kappa_0 / \kappa_0 \sim K k T^k h$ imply $\lim_{h \to M} E[\psi_{2,T,h,t}^2] \sim KE[\psi_{2,T,h,t}^2]$ for any $M$. Moreover, $\limsup_{M \to \infty} \inf_{\vartheta' \vartheta = 1} \sum_{h=0}^{M} \eta_{T,h} (\vartheta' \vartheta_h \vartheta_h \vartheta) > 0$ and $\limsup_{M \to \infty} \sup_{\vartheta' \vartheta = 1} \sum_{h=0}^{M} \eta_{T,h} (\vartheta' \vartheta_h \vartheta_h \vartheta) < \infty$ by boundedness properties Assumption A. Therefore $\sup_{\xi} (\xi' S_t^{-1} \xi)^{-1} \sim KE[\psi_{2,T,h,t}^2]$ for any $M$. This proves by the definition of the spectral norm $||S_t^{-1}||^{-1} \sim KE[\psi_{2,T,h,t}^2]$ hence $||S_T|| \sim KE[\psi_{2,T,h,t}^2]$ for any $M$. Finally, by negligibility and distribution non-degenerateness $\liminf_{T \to \infty} E[\psi_{0,t}^2(c_T,0)] > 0$. 32
Claim (c). If \( \kappa > 4 \) then by dominated convergence \( E[(y_t y_t - E[y_t y_t])^2 I(\|y_t y_t - E[y_t y_t]\| \leq c_{T,h})] \approx E[(y_t y_t - E[y_t y_t])^2 I(\|y_t y_t - E[y_t y_t]\| \leq c_{T,h})] \) is shown in (A.3). Coupled with Claim (b) this proves the orders for \( \|\mathcal{S}_T\| \). \( \mathcal{Q} \in \mathcal{D} \).

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

\[
\mathcal{J}_{T,h,t}^* \equiv I(\|y_t y_t - E[y_t y_t]\| \geq c_{T,h}) - P(\|y_t y_t - E[y_t y_t]\| \geq c_{T,h}) \quad \text{and} \quad \mathcal{J}_{T,h}^* \equiv \frac{1}{\kappa_{h,0}} \frac{1}{k_{T}^{1/2}} \sum_{t-h+1}^{T} \mathcal{J}_{T,h,t}^*
\]

\[
\mathcal{D}_{T,0} = \frac{1}{\kappa_{0,0} - 1} \frac{k_{T}^{1/2}}{T - k_{T}} - c_{T,0} \quad \text{and} \quad \mathcal{D}_{T,h} = \frac{1}{k_{T}^{1/2}} \left( d_{h,1} c_{T,h}^{1 - \kappa_{h,1}} - d_{h,2} c_{T,h}^{1 - \kappa_{h,2}} \right) \quad \text{for} \ h \neq 0.
\]

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

\[
\tilde{Z}_{M,T} \equiv T^{1/2} \frac{1}{2\pi} \left( \int_{-\pi}^{\pi} (I_{T}^{*}(\lambda) - f(\lambda)) \varpi(\lambda) d\lambda - \tilde{B}_{T} + \frac{1}{2\pi} \sum_{h=M}^{M} \left( 1 - \frac{|h|}{M} \right) \varpi_{h} \mathcal{D}_{T,h} \mathcal{J}_{T,h}^{*} \right)
\]

and \( \tilde{S}_{M,T} \equiv E[\tilde{Z}_{M,T} \tilde{Z}_{M,T}^{*}] \). Then \( \tilde{S}_{M,T}^{-1/2} \tilde{Z}_{M,T} \xrightarrow{d} N(0, I_k) \). If \( \kappa > 4 \) then \( \tilde{S}_{T} = \lim_{M \to \infty} \tilde{S}_{M,T} = \mathcal{S}_{T}(1 + o(1)) \) and if \( \kappa \in (2, 4]\) then \( \tilde{S}_{T} = \lim_{M \to \infty} \tilde{S}_{M,T} = \mathcal{S}_{T}(1 + O(1)) \).

Proof. Define \( \tilde{\varpi}_{h} = \varpi_{h} + \varpi_{0} \) and \( \tilde{\varpi}_{0} = \varpi_{0} \), and

\[
\tilde{A}_{T,h,t} \equiv \kappa_{h,0}^{-1} \mathcal{D}_{T,h} T k_{T}^{-1/2} \mathcal{J}_{T,h,t}^{*}
\]

By the same arguments leading to (A.4)-(A.6), it can be shown:

\[
\tilde{S}_{M,T}^{-1/2} \tilde{Z}_{M,T} = T^{1/2} \tilde{S}_{M,T}^{-1/2} \left( \frac{1}{2\pi} \int_{-\pi}^{\pi} (I_{T}^{*}(\lambda) - f(\lambda)) \varpi(\lambda) d\lambda - \tilde{B}_{T} \right) + \frac{1}{(2\pi)^{2}} \sum_{h=M}^{M} \left( 1 - \frac{|h|}{M} \right) \varpi_{h} \mathcal{D}_{T,h} \mathcal{J}_{T,h}^{*} \left( \tilde{\varpi}_{h,t}(c_{T,h}) - E[\tilde{\varpi}_{h,t}(c_{T,h})] + \tilde{A}_{T,h,t} \right) + o_{p}(r_{M,T})
\]

say, where \( \tilde{S}_{M,T} \) satisfies \( E(1/T^{1/2} \sum_{t=1}^{T} \tilde{X}_{M,T,t})^{2} = 1 \). By construction \( \lim_{M \to \infty} \tilde{S}_{M,T} \tilde{S}_{M,T}^{-1} = I_k \) and \( \tilde{S}_{T} = \lim_{M \to \infty} \tilde{S}_{M,T} \) 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 hence \( k = 1 \).

First, consider how \( \lim_{M \to \infty} \tilde{S}_{M,T} \) relates to \( \mathcal{S}_{T} \). By two applications of Lemma A.1.a, \( E[(\tilde{S}_{M,T}^{-1/2} \tilde{Z}_{M,T})^2] = 1 \) by construction, and \( \lim_{M \to \infty} \tilde{S}_{M,T} \tilde{S}_{M,T}^{-1} = 1 \), we have: \( 1 = E(T^{-1/2} \sum_{t=1}^{T} \tilde{X}_{M,T,t})^{2} \sim KE[\tilde{X}_{M,T,t}^{2}] \).
and
\[ \hat{S}_{M,T} \sim K \frac{1}{T} \sum_{t=1}^{T} \sum_{h=0}^{M} \left( 1 - \frac{|h|}{M} \right)^2 \hat{\varepsilon}_h^2 E \left( \{ \psi_{h,t}(c_{T,h}) - E[\psi_{h,t}(c_{T,h})] + A_{T,h,t} \}^2 \right). \]

In view of asymptotic degenerateness \( I(|y_t y_{t-h}| \geq c_{T,h}) \to 0 \text{ a.s.}, \) and \( I(|y_t y_{t-h} - E[y_t y_{t-h}]| < c_{T,h}) = 1 - I(|y_t y_{t-h}| \geq c_{T,h}) \text{ a.s. as } T \to \infty, \) the proof that
\[
E \left( \{ \psi_{h,t}(c_{T,h}) - E[\psi_{h,t}(c_{T,h})] + A_{T,h,t} \}^2 \right) \sim E \left( \psi_{h,t}(c_{T,h}) - E[\psi_{h,t}(c_{T,h})] \right)^2 \text{ if } \kappa \geq 4
\]
\[
E \left( \{ \psi_{h,t}(c_{T,h}) - E[\psi_{h,t}(c_{T,h})] + A_{T,h,t} \}^2 \right) \sim KE \left( \psi_{h,t}(c_{T,h}) - E[\psi_{h,t}(c_{T,h})] \right)^2 \text{ if } \kappa < 4
\]
is identical to the corresponding proof for Theorem 2.1 in Hill (2013). The claims \( \lim_{M \to \infty} \hat{S}_{M,T} = \mathcal{S}_T(1 + o(1)) \) if \( \kappa > 4 \) and \( \lim_{M \to \infty} \hat{S}_{M,T} = \mathcal{S}_T(1 + O(1)) \) if \( \kappa \leq 4 \) now follow from \( \mathcal{S}_T \sim \lim_{M \to \infty} KT^{-1} \sum_{t=1}^{T} \sum_{h=0}^{M} (1 - |h|/M)^2 \bar{\varepsilon}_h^2 E(\psi_{h,t}(c_{T,h}) - E[\psi_{h,t}(c_{T,h})]) \) by the proof of Lemma A.9.

Finally, by Lemma A.1.b we need only demonstrate the Lindeberg condition. Since \( 1 = E(T^{-1/2} \sum_{t=1}^{T} \hat{X}_{M,T,t})^2 \sim KE[\hat{X}_{M,T,t}^2] \) it suffices to check \( E[\hat{X}_{M,T,t}^2 I(\hat{X}_{M,T,t} > T^2)] \to 0 \forall \epsilon > 0. \) 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
\[
\mathcal{D}_{T,0} \leq K \frac{k_T^{1/2}}{T} \left( \frac{T}{k_T} \right)^{1/\kappa_{0,0}} \text{ and } |\mathcal{D}_{T,h}| \leq K \frac{k_T^{1/2} \kappa_{h,0}}{k_T^{1/2}} \left( \frac{T}{k_T} \right)^{1/\kappa_{h,0}} \leq K \frac{k_T^{1/2}}{T} \left( \frac{T}{k_T} \right)^{1/\kappa_{0,0}}.
\]
Therefore
\[
|A_{T,h,t}| \leq K \left( \frac{T}{k_T} \right)^{1/\kappa_{0,0}} \left| I \left( \{|y_t y_{t-h}| \geq c_{T,h}\} - P \left( \{|y_t y_{t-h}| \geq c_{T,h}\} \right) \right| \leq K \left( \frac{T}{k_T} \right)^{1/\kappa_{0,0}}
\]
Similarly, \( \psi_{h,t}(c_{T,h}) \leq c_{T,h} \sim K(T/k_T)^{1/\kappa_{h,0}} \leq K(T/k_T)^{1/\kappa_{0,0}}. \) By the construction of \( \hat{X}_{M,T,t}, \) the remaining steps for showing \( E[\hat{X}_{M,T,t}^2 I(\hat{X}_{M,T,t} > T^2)] \to 0 \) follow from the line of proof of Lemma A.9.a given \( \lim_{M \to \infty} \hat{S}_{M,T} \hat{S}_{M,T}^{-1} = 1, \lim_{M \to \infty} \hat{S}_{M,T} = \mathcal{S}_T(1 + O(1)) \) and \( ||\mathcal{S}_T|| \sim KE[\psi_{0,t}(c_{T,0})] \) by Lemma A.9.b. 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 \( \varpi(\lambda) \equiv -(f(\lambda))^{-1} (\partial \rho/\partial \theta) \log f(\lambda). \)

**Proof of Theorem 2.1.** Define criteria
\[
\hat{\mathcal{Q}}_T^\prime(\theta) \equiv \sum_{j \in \mathcal{F}} \left\{ \ln f(\lambda_j, \theta) + \frac{T_T^\prime(\lambda_j)}{f(\lambda_j, \theta)} \right\} \text{ and } \mathcal{Q}_T(\theta) \equiv \sum_{j \in \mathcal{F}} \left\{ \ln f(\lambda_j, \theta) + \frac{T_T^\prime(\lambda_j)}{f(\lambda_j, \theta)} \right\},
\]

34
and write sample gradients and Hessians as

\[ \tilde{G}_T(\theta) \equiv \frac{\partial}{\partial \theta} \hat{Q}_T(\theta), \quad G_T(\theta) \equiv \frac{\partial}{\partial \theta} Q_T^r(\theta), \quad \hat{H}_T(\theta) \equiv \frac{\partial}{\partial \theta'} \hat{G}_T(\theta) \text{ and } H_T(\theta) \equiv \frac{\partial}{\partial \theta'} G_T(\theta). \]

Recall the population Hessian

\[ H(\theta) \equiv -\frac{1}{2\pi} \int_{-\pi}^{\pi} \varpi(\lambda, \theta) \frac{\partial}{\partial \theta'} f(\lambda, \theta) d\lambda - \frac{1}{2\pi} \int_{-\pi}^{\pi} \{ f(\lambda, \theta) - f(\lambda) \} \frac{\partial}{\partial \theta} \varpi(\lambda, \theta) d\lambda + o_p(1). \]

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

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

\[ \sup_{\theta \in \Theta} \| \tilde{G}_T(\theta) - G_T(\theta) \| \overset{p}{\rightarrow} 0 \quad \text{and} \quad \sup_{\theta \in \Theta} \| \hat{H}_T(\theta) - H_T(\theta) \| \overset{p}{\rightarrow} 0. \tag{A.8} \]

Taylor expansions show by the FD-QML first order conditions \( \tilde{G}_T = \hat{H}_T(\hat{\theta}_T) \times (\hat{\theta}_T^* - \theta_0) \) and \( G_T = H_T(\hat{\theta}_T) \times (\theta^*_T - \theta_0) \) for sequences \( \{ \hat{\theta}_T, \hat{\theta}_T \} \) satisfying \( \| \hat{\theta}_T - \theta_0 \| \leq \| \theta^*_T - \theta_0 \| \) and \( \| \hat{\theta}_T - \theta_0 \| \leq \| \theta^*_T - \theta_0 \| \). Now use (A.8), spectrum boundedness, and multiple applications of Minkowski’s inequality to deduce

\[
\| \hat{\theta}_T^* - \theta_T^* \| = \left\| (\hat{\theta}_T^* - \theta_0) - (\theta^*_T - \theta_0) \right\| \\
\leq \| \hat{H}_T(\hat{\theta}_T) \|^{-1} \times \| \tilde{G}_T - G_T \| + \| \hat{H}_T(\hat{\theta}_T) - H_T(\hat{\theta}_T) \| \times \| \hat{H}_T(\hat{\theta}_T) \|^{-1} \times \| H_T(\hat{\theta}_T) \|^{-1} \times \| G_T \| \\
= o_p \left( \| \hat{H}_T(\hat{\theta}_T) \|^{-1} \left( 1 + \| H_T(\hat{\theta}_T) \|^{-1} \times \| G_T \| \right) \right).
\]

If \( \sup_{\theta \in \Theta} \{ \| H_T(\theta) \|^{-1} \} = O_p(1) \) then by (A.8) we have \( \sup_{\theta \in \Theta} \| \hat{H}_T(\theta) H_T^{-1}(\theta) - I_k \| \overset{p}{\rightarrow} 0 \) hence

\[
\sup_{\theta \in \Theta} \| \hat{H}_T(\theta) \|^{-1} \leq O_p(1) \times \sup_{\theta \in \Theta} \| \hat{H}_T(\theta) H_T^{-1}(\theta) - I_k \|^{-1} + O_p(1) = O_p(1).
\]

Therefore, if we show \( \| G_T \| = O_p(1) \) and \( \sup_{\theta \in \Theta} \| H_T(\theta) \|^{-1} = O_p(1) \) then the proof that \( \| \hat{\theta}_T^* - \theta^*_T \| \overset{p}{\rightarrow} 0 \) is complete.

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

\[ G_T = \sum_{j \in \mathcal{F}} \{ T^*_T(\lambda_j) - f(\lambda_j) \} \varpi(\lambda_j) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \{ T^*_T(\lambda) - f(\lambda) \} \varpi(\lambda) d\lambda + o_p(1). \]

Spectrum boundedness and LLN Lemma A.8 therefore imply \( \| G_T \| = o_p(1) \).
Next, recall \( I_T(\lambda) \) is the periodogram of \( y_t \). By construction \( H_T(\theta) = H(\theta) + r_T(\theta) \) where

\[
r_T(\theta) = - \left( \sum_{j \in F} \varpi(\lambda_j, \theta) \frac{\partial}{\partial \lambda_j} f(\lambda_j, \theta) - \frac{1}{2\pi} \right) \int_{-\pi}^{\pi} \varpi(\lambda, \theta) \frac{\partial}{\partial \theta} f(\lambda, \theta) d\lambda \]

\[
- \left( \sum_{j \in F} \{ f(\lambda_j, \theta) - f(\lambda_j) \} \frac{\partial}{\partial \theta} \varpi(\lambda, \theta) - \frac{1}{2\pi} \right) \int_{-\pi}^{\pi} \{ f(\lambda, \theta) - f(\lambda) \} \frac{\partial}{\partial \theta} \varpi(\lambda, \theta) d\lambda 
\]

\[
+ \sum_{j \in F} \{ I_T(\lambda_j) - f(\lambda_j) \} \frac{\partial}{\partial \lambda_j} \varpi(\lambda, \theta) + \sum_{j \in F} \{ I_T^*(\lambda_j) - I_T(\lambda_j) \} \frac{\partial}{\partial \lambda_j} \varpi(\lambda, \theta) 
\]

\[
= A_{1,T}(\theta) + A_{2,T}(\theta) + A_{3,T}(\theta) + A_{4,T}(\theta). 
\]

Spectrum boundedness and Hölder continuity Assumption A.2.iv imply \( ||A_{4,T}(\theta)|| \) and \( ||A_{3,T}(\theta)|| \) 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 \( A_{4,T}(\theta) \) 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 \( \hat{\gamma}_{T,h} \equiv 1/T \sum_{t=h+1}^{T} y_t y_{t-h} \). Square integrability and distribution continuity ensure \( E|\hat{\gamma}_{T,h}|^{1+\iota} < \infty \) for some infinitesimal \( \iota > 0 \). Hence, by triangle and Hölder inequalities, stationarity and negligibility:

\[
E \left| \hat{\gamma}_{T,h}(cT,h) - \hat{\gamma}_{T,h} \right|^2 \leq \frac{1}{T} \sum_{t=h+1}^{T} \left( E|\hat{\gamma}_t cT,h|^{1+\iota} \right) \frac{1}{1+\iota} P \left( |\hat{\gamma}_t cT,h|^{1+\iota} \right) \leq K \left( \frac{h}{T} \right)^{1/(1+\iota)} \rightarrow 0. 
\]

Therefore \( \hat{\gamma}_{T,h}(cT,h) \rightarrow \hat{\gamma}_{T,h} \) in probability. Now use boundedness of \( \eta(\lambda; \theta) \) to deduce

\[
\sup_{\theta \in \Theta} \left| \int_{-\pi}^{\pi} \{ I_T^*(\lambda) - I_T(\lambda) \} \eta(\lambda; \theta) d\lambda \right| \rightarrow 0. \tag{A.9} 
\]

Combine boundedness Assumption A.2 and (A.9) to deduce \( \sup_{\theta \in \Theta} ||A_{4,T}(\theta)|| \rightarrow 0 \). Therefore \( \sup_{\theta \in \Theta} ||H_T(\theta) - H(\theta)|| \rightarrow 0 \). Finally, Assumption A.3 states \( \inf_{\theta \in \Theta} ||H(\theta)|| > 0 \) hence \( \sup_{\theta \in \Theta} ||H_T(\theta)||^{-1} \rightarrow 0 \).

**Step 2** \( (\theta_T^* \rightarrow \theta_0) \). In view of (A.9) we can use \( I_T(\lambda) \) in place of \( I_T^*(\lambda) \). The proof of \( \theta_T^* \rightarrow \theta_0 \) 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 \( \hat{\theta}_T^* \) defined by (A.2). We prove \( T^{1/2} V_T^{-1/2} (\hat{\theta}_T^* - \theta_0 + B_T) \xrightarrow{d} N(0, I_k) \) hence \( T^{1/2} V_T^{-1/2} (\hat{\theta}_T^* - \theta_0 + B_T) \xrightarrow{d} N(0, I_k) \) by Lemma A.5.

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

\[
T^{1/2} V_T^{-1/2} \left( \hat{\theta}_T^* - \theta_0 \right) = -T^{1/2} V_T^{-1/2} \left( \frac{\partial^2}{\partial \theta \partial \theta} \hat{Q}_T^*(\theta_T) \right)^{-1} \frac{1}{2\pi} \int_{-\pi}^{\pi} \left( \hat{I}_T^*(\lambda) - f(\lambda) \right) \varpi(\lambda) d\lambda. 
\]
Proof of Theorem 3.1. Define bias corrected spectral density and FD-QML estimators

\[ \hat{\mathcal{Z}}^{(bc)}_T(\lambda) = \frac{1}{2\pi} \left( \hat{\mathcal{Z}}^{(bc)}_{T,0}(\hat{\mathcal{Y}}^{(0)}_{0,(kT)}) + 2 \sum_{h=1}^{b_T} \hat{\mathcal{Z}}^{(bc)}_{T,h}(\hat{\mathcal{Y}}^{(0)}_{h,(kT)}) \times \cos(\lambda h) \right) \]

\[ \hat{\theta}^{(bc)}_T \equiv \arg \min_{\theta \in \Theta} \frac{1}{2\pi} \int_{-\pi}^{\pi} \left\{ \ln f(\lambda, \theta) + \frac{\hat{\mathcal{Z}}^{(bc)}_T(\lambda)}{f(\lambda, \theta)} \right\} d\lambda \]

\[ \hat{\theta}^{(bc)}_T \equiv \arg \min_{\theta \in \Theta} \sum_{j \in \mathcal{F}} \left\{ \ln f(\lambda_j, \theta) + \frac{\hat{\mathcal{Y}}^{(bc)}_T(\lambda_j)}{f(\lambda_j, \theta)} \right\} . \]
We will prove $T^{1/2}\hat{V}_T^{-1/2}(\hat{\theta}_T^{(bc)} - \theta_0) \xrightarrow{d} N(0, I_k)$. Then $T^{1/2}\hat{V}_T^{-1/2}(\hat{\theta}_T^{(bc)} - \theta_0) \xrightarrow{d} N(0, I_k)$ follows by the same argument used to prove Lemma A.5. The proof of $T^{1/2}\hat{V}_T^{-1/2}(\hat{\theta}_T^{(obc)} - \theta_0) \xrightarrow{d} N(0, I_k)$ is identical since $m_{T,h}(\xi) = [\xi m_{T,h}]$ for $0 < \xi \leq \xi \leq \tilde{\xi} < \infty$ 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 $\hat{\kappa}_{h,i,m_{T,h}(\xi)} = \kappa_{h,i} + O_p(1/m_{T,h})$ and $\hat{d}_{h,i,m_{T,h}(\xi)} = d_{h,i} + O_p(1/m_{T,h})$ by Lemma A.4 hence $\{\hat{\kappa}_{h,i,m_{T,h}(\xi)}, \hat{d}_{h,i,m_{T,h}(\xi)}\}$ have the same $m_{T,h}$-rate of convergence as $\{\kappa_{h,i,m_{T,h}}, d_{h,i,m_{T,h}}\}$. 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 $T^{1/2}\hat{V}_T^{-1/2}(\hat{\theta}_T^{(obc)} - \theta_0)$ asymptotically uniformly in $\xi \in [\xi, \tilde{\xi}]$, hence $\{\hat{\kappa}_{h,i,m_{T,h}(\xi)}, \hat{d}_{h,i,m_{T,h}(\xi)}\}$ do not effect the limit distribution of $T^{1/2}\hat{V}_T^{-1/2}(\hat{\theta}_T^{(obc)} - \theta_0)$. The claims $\hat{V}_T = V_T(1 + o(1))$ if $\kappa < 4$ and $\hat{V}_T = V_T(1 + O(1))$ if $\kappa = 4$ follow from $\hat{V}_T = \Omega^{-1}\hat{S}_T\Omega^{-1}$ and the Lemma A.10 results that if $\kappa > 4$ then $\hat{S}_T = \lim_{M \to \infty} \hat{S}_{M,T} = S_T(1 + o(1))$ and if $\kappa \in (2, 4]$ then $\hat{S}_T = \lim_{M \to \infty} \hat{S}_{M,T} = S_T(1 + O(1))$, where $\hat{S}_{M,T}$ is defined in Lemma A.10.

**Step 1** ($\hat{\theta}_T^{(bc)}$). Let $\{\hat{\theta}_{T,h}\}_{h \in \mathbb{N}}$ be defined as in (17) for $h = 0$ and in (16) or (18) for $h = 1, \ldots, b_T$, and satisfy $\hat{\theta}_{T,h} = 0 \forall h > b_T$. Define a bias estimator:

$$\hat{B}_T = -\frac{1}{2\pi} \int_{-\pi}^{\pi} \left( \hat{R}_{T,0} + 2 \sum_{h=1}^{\infty} \hat{R}_{T,h} \times \cos(\lambda h) \right) \varpi(\lambda) d\lambda.$$ 

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 differ in different places.

We need the process that governs $\hat{\theta}_T^{(bc)}$. Take $\hat{V}_T$, $\kappa_{h,0}$ and $D_{T,h}$ defined in (A.7), and define for arbitrary $M \in \mathbb{N}$:

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

$$\hat{S}_{M,T} \equiv E \left[ \hat{Z}_{M,T} \right] \text{ and } \hat{V}_T = \Omega^{-1}\hat{S}_T\Omega^{-1}.$$ 

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

In view of bias definition (A.11), and arguments leading to (A.10) and (A.12), use $\hat{V}_T = \Omega^{-1}\hat{S}_T\Omega^{-1}$ and approximation Lemma A.7 to deduce:

$$T^{1/2}\hat{V}_T^{-1/2}(\hat{\theta}_T^{(bc)} - \theta_0) = -T^{1/2}\hat{S}_T^{-1/2} \frac{1}{2\pi} \int_{-\pi}^{\pi} \left( \hat{I}_T^{(bc)}(\lambda) - f(\lambda) \right) \varpi(\lambda) d\lambda \times (1 + o_p(1)) \quad (A.13)$$

$$= -T^{1/2}\hat{S}_T^{-1/2} \frac{1}{2\pi} \left( \int_{-\pi}^{\pi} (I_T(\lambda) - f(\lambda)) \varpi(\lambda) d\lambda - \hat{B}_T \right) \times (1 + o_p(1))$$

$$+ T^{1/2}\hat{S}_T^{-1/2} \frac{1}{2\pi} \left( \hat{B}_T - \hat{B}_T \right) \times (1 + o_p(1)).$$
In Step 2 we show $\hat{\beta}_T$ is, asymptotically, a linear function of the indicator partial sum $Z_{T,h}$:

$$T^{1/2} \hat{\beta}_T = - \frac{1}{2\pi} T^{1/2} \hat{\beta}_T$$

where $\hat{\beta}_T$ is the tail index estimator.

Combine (A.13) and (A.14) to obtain by the construction of $\tilde{Z}_{M,T}$:

$$T^{1/2} \tilde{Z}_{M,T} = - \frac{1}{2\pi} T^{1/2} \tilde{Z}_{M,T}$$

We now verify (A.14). Define $\tilde{Z}_{M,T}$, we show:

By Lemma A.10: $\tilde{Z}_{M,T} \rightarrow N(0,I_k)$. Since $M$ can be made arbitrarily large we can always impose $M \rightarrow \infty$ as $T \rightarrow \infty$ such that $\lim_{T \rightarrow \infty} \tilde{Z}_{M,T} = I_k$ and $\lim_{T \rightarrow \infty} r_{M,T} = 0$.

Step 2 ($\hat{\beta}_T$). We now verify (A.14). Define $\hat{\beta}_T := - \gamma_{T,h}(c_{T,h}) - E[y_{y_{T,h}} - h]$, and recall by constructions (13) and (14) that $\hat{\beta}_T$ is defined in (A.7), we show:

By arguments in Peng (2001, p. 259-263) for second order Pareto tails, and Assumption B', we have $T^{1/2} \sum_{h=-M}^M \varpi \{ \hat{\beta}_T - \beta_T \} = o(1)$, hence with $\hat{\beta}_T$ and $\tilde{\beta}_T$ above it follows:

Next, with $D_{T,h}$ defined in (A.7), we show:

By imitating arguments in Hill (2013: proof of Theorem 2.1), $m_{T,h}/k_T \rightarrow \infty$ can be shown to imply the tail index estimators $\hat{\beta}_{T,h}$ and $\tilde{\beta}_{T,h}$ in $\hat{\beta}_{T,m}$ do not affect asymptotics since they are $m_{T,h}$-
consistent under Assumptions A and B’ by Lemma A.4, while by Lemma A.3 $\hat{Y}_{h,(kT)}^{(0)}$ 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,

$$\hat{R}_{T,0} = \frac{1}{\kappa_{0,0} - 1} \left( \frac{k_T}{T - k_T} \right) \hat{Y}_{0,(kT)}^{(0)} = R_{T,0} + \frac{1}{\kappa_{0,0} - 1} \frac{k_T^{1/2}}{T - k_T} c_{T,0} \times k_T^{1/2} \left( \frac{\hat{Y}_{0,(kT)}^{(0)}}{c_{T,0}} - 1 \right)$$

$$\hat{R}_{T,h} = \frac{T}{T - k_T} \left( \frac{d_{h,2} \left( \hat{Y}_{h,(kT)}^{(0)} \right)^{1-\kappa_{h,2}}}{\kappa_{h,2} - 1} - \frac{d_{h,1} \left( \hat{Y}_{h,(kT)}^{(0)} \right)^{1-\kappa_{h,1}}}{\kappa_{h,1} - 1} \right) \text{ for } h > 0.$$

A first order expansions around $c_{T,h}$ implies for some sequence of positive random numbers $\{c_{T,h}^*\}$ where $|\hat{Y}_{h,(kT)}^{(0)} - c_{T,h}^*| \leq |\hat{Y}_{h,(kT)}^{(0)} - c_{T,h}|$:

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

The last equality exploits $c_{T,h}^*/c_{T,h} \xrightarrow{p} 1$ given $|\hat{Y}_{h,(kT)}^{(0)} - c_{T,h}^*| \leq |\hat{Y}_{h,(kT)}^{(0)} - c_{T,h}|$ and $\hat{Y}_{h,(kT)}^{(0)}/c_{T,h} \xrightarrow{p} 1$ by Lemma 1 in Hill (2010). Hence we have shown (A.16).

Finally, combine (A.15) and (A.16), with $k_T^{1/2} (\hat{Y}_{h,(kT)}^{(0)}/c_{T,h} - 1) = \hat{Y}_{T,h}^* (1 + o_p(1))$ by Lemma A.4, to deduce (A.14). \textit{QED}

References


Figure 1: The process is AR(1) $y_t = 0.9y_{t-1} + \epsilon_t$, $\epsilon$ 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)

<table>
<thead>
<tr>
<th></th>
<th>(\phi_0 = .90) (where (\sigma^2_0 = 1))</th>
<th>(\sigma^2_0 = 1) (where (\phi_0 = .90))</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n = 100)</td>
<td>(n = 250)</td>
</tr>
<tr>
<td></td>
<td>Bias</td>
<td>Med</td>
</tr>
<tr>
<td>Pareto Error: (\kappa = 2.25)</td>
<td>n = 500</td>
<td>n = 1000</td>
</tr>
<tr>
<td>no trim</td>
<td>-0.0286</td>
<td>.8801</td>
</tr>
<tr>
<td>trim-obc</td>
<td>-0.0212</td>
<td>.8850</td>
</tr>
<tr>
<td>Pareto Error: (\kappa = 4.50)</td>
<td>n = 500</td>
<td>n = 1000</td>
</tr>
<tr>
<td>no trim</td>
<td>-0.0224</td>
<td>.8862</td>
</tr>
<tr>
<td>trim-obc</td>
<td>-0.0193</td>
<td>.8913</td>
</tr>
<tr>
<td>Pareto Error: (\kappa = 2.25)</td>
<td>n = 500</td>
<td>n = 1000</td>
</tr>
<tr>
<td>no trim</td>
<td>-0.0503</td>
<td>.8960</td>
</tr>
<tr>
<td>trim-obc</td>
<td>-0.0514</td>
<td>.8964</td>
</tr>
<tr>
<td>Pareto Error: (\kappa = 4.50)</td>
<td>n = 500</td>
<td>n = 1000</td>
</tr>
<tr>
<td>no trim</td>
<td>-0.052</td>
<td>.8970</td>
</tr>
<tr>
<td>trim-obc</td>
<td>-0.049</td>
<td>.8967</td>
</tr>
<tr>
<td>Normal Error</td>
<td>n = 500</td>
<td>n = 1000</td>
</tr>
<tr>
<td>no trim</td>
<td>-0.0063</td>
<td>.8954</td>
</tr>
<tr>
<td>trim-obc</td>
<td>-0.058</td>
<td>.8954</td>
</tr>
</tbody>
</table>

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 \(\sigma^2_0 \equiv E[\epsilon^2].\) b. “Med” is the median, and “RMSE” is the root-mean-squared-error. c. “KS_{0.05}” 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.
Table 2: FD-QML for GARCH(1,1)\textsuperscript{a}

<table>
<thead>
<tr>
<th>n = 100</th>
<th>n = 250</th>
<th>n = 100</th>
<th>n = 250</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias</td>
<td>Med\textsuperscript{b}</td>
<td>RMSE</td>
<td>$K_{0.05}$</td>
</tr>
<tr>
<td>no trim</td>
<td>-0.0542</td>
<td>2.116</td>
<td>0.1386</td>
</tr>
<tr>
<td>trim-obc</td>
<td>0.0112</td>
<td>2.461</td>
<td>0.2022</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>n = 500</th>
<th>n = 1000</th>
<th>n = 500</th>
<th>n = 1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias</td>
<td>Med\textsuperscript{b}</td>
<td>RMSE</td>
<td>$K_{0.05}$</td>
</tr>
<tr>
<td>no trim</td>
<td>-0.0528</td>
<td>2.333</td>
<td>0.1087</td>
</tr>
<tr>
<td>trim-obc</td>
<td>0.0044</td>
<td>2.449</td>
<td>0.1688</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>n = 500</th>
<th>n = 1000</th>
<th>n = 500</th>
<th>n = 1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias</td>
<td>Med\textsuperscript{b}</td>
<td>RMSE</td>
<td>$K_{0.05}$</td>
</tr>
<tr>
<td>no trim</td>
<td>0.0174</td>
<td>6.532</td>
<td>0.2321</td>
</tr>
<tr>
<td>trim-obc</td>
<td>0.0267</td>
<td>6.538</td>
<td>0.2152</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>n = 500</th>
<th>n = 1000</th>
<th>n = 500</th>
<th>n = 1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias</td>
<td>Med\textsuperscript{b}</td>
<td>RMSE</td>
<td>$K_{0.05}$</td>
</tr>
<tr>
<td>no trim</td>
<td>0.0036</td>
<td>6.391</td>
<td>0.1629</td>
</tr>
<tr>
<td>trim-obc</td>
<td>0.0014</td>
<td>6.682</td>
<td>0.1725</td>
</tr>
</tbody>
</table>

a. The model is $y_t = x_t^2 + \alpha_0 x_t^2 + \beta_0 x_{t-1}^2 + \epsilon_t$ where $\epsilon_t$ is iid standard normal. b. “Med” is the median, and “RMSE” is the root-mean-squared-error.
c. “$K_{0.05}$” 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.