

Parameter Estimation Robust to Low-Frequency Contamination

Adam McCloskey & Jonathan B. Hill

To cite this article: Adam McCloskey & Jonathan B. Hill (2017) Parameter Estimation Robust to Low-Frequency Contamination, Journal of Business & Economic Statistics, 35:4, 598-610, DOI: [10.1080/07350015.2015.1093948](https://doi.org/10.1080/07350015.2015.1093948)

To link to this article: <http://dx.doi.org/10.1080/07350015.2015.1093948>



View supplementary material [↗](#)



Accepted author version posted online: 20 Oct 2015.
Published online: 20 Apr 2017.



Submit your article to this journal [↗](#)



Article views: 84



View Crossmark data [↗](#)



Citing articles: 2 View citing articles [↗](#)

Parameter Estimation Robust to Low-Frequency Contamination

Adam McCLOSKEY

Department of Economics, Brown University, Providence, RI 02912 (adam_mccloskey@brown.edu)

Jonathan B. HILL

Department of Economics, University of North Carolina, Chapel Hill, NC 27599 (jbhill@email.unc.edu)

We provide methods to robustly estimate the parameters of stationary ergodic short-memory time series models in the potential presence of additive low-frequency contamination. The types of contamination covered include level shifts (changes in mean) and monotone or smooth time trends, both of which have been shown to bias parameter estimates toward regions of persistence in a variety of contexts. The estimators presented here minimize trimmed frequency domain quasi-maximum likelihood (FDQML) objective functions without requiring specification of the low-frequency contaminating component. When proper sample size-dependent trimmings are used, the FDQML estimators are consistent and asymptotically normal, asymptotically eliminating the presence of any spurious persistence. These asymptotic results also hold in the absence of additive low-frequency contamination, enabling the practitioner to robustly estimate model parameters without prior knowledge of whether contamination is present. Popular time series models that fit into the framework of this article include autoregressive moving average (ARMA), stochastic volatility, generalized autoregressive conditional heteroscedasticity (GARCH), and autoregressive conditional heteroscedasticity (ARCH) models. We explore the finite sample properties of the trimmed FDQML estimators of the parameters of some of these models, providing practical guidance on trimming choice. Empirical estimation results suggest that a large portion of the apparent persistence in certain volatility time series may indeed be spurious. Supplementary materials for this article are available online.

KEY WORDS: Deterministic trends; Frequency domain estimation; Level shifts; Robust estimation; Spurious persistence.

1. INTRODUCTION

Empirical evidence of level shifts (changes in mean) or other deterministic trends has been recognized for many years. Of the numerous and varied examples, Garcia and Perron (1996) found the presence of large level shifts in U.S. real interest rate series; Qu (2011) rejected the null hypothesis of stationary short or long-memory against the alternative of level shifts or deterministic trends for U.S. inflation rate series; Eraker, Johannes, and Polson (2003), Stărică and Granger (2005), and Qu and Perron (2013) found that incorporating jumps into volatility models provides a substantial improvement for capturing the dynamics of stock market returns volatility; McCloskey and Perron (2013) provided evidence that contaminating components bias standard memory parameter estimates upward for daily stock market returns volatility. The literature also indicates that the presence of level shifts or trends leads to the phenomenon of “spurious persistence,” in which the econometrician is misled into believing a process is more persistent than it actually is. For example, Perron (1990) demonstrated that the presence of abrupt changes in mean often induces spurious nonrejection of the unit root hypothesis while Bhattacharya, Gupta, and Waymire (1983) demonstrated that certain deterministic trends can induce the spurious presence of long-memory features in the data. Perron (1990) also showed that level shifts asymptotically bias autoregressive coefficient estimates toward one. Similarly, through simulation evidence, Lamoureux and Lastrapes (1990), and analytically, Mikosch and Stărică (2004) and Hillebrand (2005), have shown that when the mean of a GARCH process changes, the sum of its estimated autoregressive parameters converge to unity. Much

recent attention has focused on the phenomenon that changes in mean also lead to the presence of spurious long-memory. For example, Diebold and Inoue (2001), Granger and Hyung (2004), Mikosch and Stărică (2004), and Perron and Qu (2010) have shown through simulation and theory that level shifts induce hyperbolically decaying autocorrelations and spectral density estimates that approach infinity at the null frequency. Thus, the data indicate that an estimation method that is robust to these contaminating components would be quite useful in practice. The current study provides such a method.

This article presents a robust estimation technique that exploits the difference in stochastic orders of magnitude between the periodograms of an additive low-frequency contaminating process and a weakly dependent contaminated process whose parameters we wish to estimate. Depending on the frequency range under study, one of these periodograms comes to asymptotically dominate the other. The estimators fit the periodogram of the observed process to the spectral density function of a candidate model within certain (higher) frequency ranges for which the periodogram of the contaminated process dominates that of the contaminating process. The robust estimators we focus on are trimmed versions of the frequency domain quasi-maximum likelihood (FDQML) estimator for which the trimming depends on the sample size. The asymptotic properties of the (untrimmed) FDQML estimator have been extensively

studied in standard (uncontaminated) linear process contexts by Dunsmuir and Hannan (1976), Dunsmuir (1979), and Hosoya and Taniguchi (1982), among others. We obtain the asymptotic properties of the trimmed version of this estimator in the presence of additive low-frequency contamination.

To overcome the issue of spurious persistence caused by level shifts in the context of fully parametric estimation, a substantial amount of work has been done to explicitly estimate mean changes and change points within a sample (see, among many others, Bai and Perron 1998). After accounting for mean changes by estimating them, one may subsequently estimate the other parameters of a given process. Similarly, one may attempt to explicitly model an underlying trend function that could be causing the presence of spurious persistence. However, in practice one may not have a good idea of the form a contaminating component takes or whether contamination is present at all. Our robust estimators perform well in both the presence and absence of a wide variety of contaminating components. The types of contamination we are interested in are not amenable to high-pass filtering because they cause the observed periodogram to diverge in a *shrinking* band of frequencies surrounding the origin. Moreover, an approximating high-pass filter, as detailed by Baxter and King (1999), for example, is unnecessary in this context since estimation can be performed directly in the frequency domain. (Not only is the use of such filtering unnecessary and not designed for the problem at hand, but also preliminary simulation results indicate that the performance of such a method in this context is clearly inferior to the technique advanced here.) In a separate regression context, Engle (1974) also noted that ignoring a band of low frequencies could help to improve estimation performance when the model being estimated is only valid within a higher frequency range. Yet his analysis was similarly concerned with the properties of estimators that ignore a *fixed* band of frequencies, which, in the present context is both unnecessary and wasteful of valuable information on the parameters of the contaminated process. The methodology presented here complements the work of Müller and Watson (2008), who focus upon modeling the *low*-frequency variability of a time series by examining a shrinking band (but fixed number) of low-frequency ordinates.

The fully parametric estimation methodology we introduce in this article is related to the robust semiparametric memory parameter estimators of Iacone (2010), McCloskey and Perron (2013), and Hou and Perron (2014). However, the estimators in these articles do not specify the other, short-run dynamics of the time series and hence require a bandwidth parameter that limits the highest frequency ordinate examined. In contrast, our fully parametric approach does not require this upper bound, resulting in an estimator that is consistent at the fully parametric rate. On the other hand, the types of contaminated processes considered in this article are restricted to have absolutely summable autocovariances. In another related article, McCloskey (2013) showed how to use the trimmed FDQML estimator presented here to estimate the fully parametric long-memory stochastic volatility model robust to low-frequency contamination.

Our trimmed FDQML estimators can be used to consistently estimate the parameters of a large class of models including ARMA, stochastic volatility (SV), ARCH, and GARCH, in the presence of a wide variety of additive low-frequency contami-

nating components. They are asymptotically normal under mild conditions. The asymptotic variances of these estimators are the same as those in the uncontaminated and untrimmed case. Depending upon the model being estimated, the trimming can be chosen to balance the competing finite sample biases arising from the potential presence of contaminating components and ignoring lower frequency ordinates when fitting the model. We provide a simulation study that addresses the issue of trimming parameter choice for popular ARMA, SV, and GARCH models. We subsequently use high-frequency exchange rate volatility and daily stock market volatility data to robustly estimate the parameters of particular specifications of these models, finding that robust estimation significantly reduces the persistence of the fitted models.

The rest of this article proceeds as follows. In Section 2, we detail the robust estimation technique and provide the reasoning behind the use of trimming. Here, we also discuss some of the contaminating processes of interest. Section 3 supplies the asymptotic properties of the trimmed FDQML estimators under general assumptions on the contaminated and contaminating processes and the class of models being estimated. The finite sample properties of the trimmed FDQML estimator are analyzed for three different popular time series models in Section 4 while Section 5 comprises the empirical application of the robust estimation technique to financial volatility data. Section 6 concludes. Proofs of the main results are contained in a supplemental mathematical appendix available online.

In what follows, \mathbb{R} and \mathbb{Z} denote the sets of real numbers and integers; $\lfloor K \rfloor$ denotes the largest integer value below any generic $K \in \mathbb{R}$; $O(\cdot)$, $o(\cdot)$, $O_p(\cdot)$, and $o_p(\cdot)$ denote the usual (stochastic) orders of magnitude; $a_T \sim b_T$ implies $a_T/b_T \rightarrow 1$; $I(\cdot)$ is the indicator function; “ \xrightarrow{p} ” and “ \xrightarrow{d} ” indicate convergence in probability and distribution; A_{hk} denotes the entry of the h th row and k th column of the generic matrix A . All convergence concepts are taken to mean as the sample size goes to infinity.

2. ROBUST ESTIMATION TECHNIQUE

Let $\{y_t\}$ denote a generic covariance stationary process. The discrete Fourier transform and periodogram of $\{y_t\}$ at frequency $\lambda \in [-\pi, \pi]$ are, respectively, defined as follows:

$$w_y(\lambda) \equiv \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^T y_t e^{-i\lambda t} \quad \text{and} \quad I_y(\lambda) = |w_y(\lambda)|^2,$$

where T is the sample size. The estimators we examine involve evaluating the observed time series' periodogram at a subset of the Fourier frequencies $\lambda_j \equiv 2\pi j/T$ for $j = -\lfloor T/2 \rfloor + 1, \dots, \lfloor T/2 \rfloor - 1, \lfloor T/2 \rfloor$. We wish to estimate the model parameters of a contaminated process $\{v_t\}$, but we observe the process

$$x_t = \mu + v_t + u_t,$$

where μ is a constant, $E[I_v(\lambda_j)] = O(1)$, $E[I_u(\lambda_j)] = O(T/j^2)$ and $\{v_t\}$ and $\{u_t\}$ are independent. Henceforth, “contaminated process” refers to $\{v_t\}$ and “contaminating process” refers to $\{u_t\}$.

2.1 Contaminating Processes of Interest

Many contaminating processes satisfy $E[I_u(\lambda_j)] = O(T/j^2)$, a subset of which we provide here.

Random Level Shifts (RLS)

$$u_t = \sum_{j=1}^t \delta_{T,j}, \quad \delta_{T,t} = \pi_{T,t} \eta_t,$$

where η_t is iid with $E[\eta_t] = 0$ and $E[\eta_t^2] = \sigma_\eta^2 > 0$ and $\pi_{T,t}$ is iid Bernoulli taking the value 1 with probability p/T for some $p \geq 0$. The components $\pi_{T,t}$ and η_t are mutually independent.

Deterministic Level Shifts (DLS)

$$u_t = \sum_{i=1}^B c_i I(T_{i-1} < t \leq T_i),$$

where B is a fixed positive integer (the number of breaks plus one), $|c_i| < \infty$ for $i = 1, \dots, B$, $T_0 = 0$, $T_B = T$, $T_0 < T_1 < \dots < T_{B-1} < T_B$, and $T_i/T \rightarrow \tau_i \in (0, 1)$ for $i = 1, \dots, B$.

Deterministic Trends (DT)

$$u_t = h(t/T),$$

where $h(\cdot)$ is a deterministic nonconstant function on $[0, 1]$ that is either Lipschitz continuous or monotone with $h(1) = 0$. (This includes all cases for which $h(\cdot)$ is monotonic and bounded since we can simply subtract $h(1)$ from $h(\cdot)$ and add $h(1)$ to μ to have the same DGP.)

Fractional Trends (FT)

$$u_t = O((t+1)^{\phi-1/2}) \quad \text{with} \quad u_0 = 0, \quad |u_{t+1} - u_t| = O(|u_t|/t),$$

where $\phi \in (-1/2, 1/2)$.

Outliers

$$u_t = \sum_{i=1}^M m_i I(t = T_i),$$

where M is a fixed positive integer (the number of outliers), $|m_i| < \infty$ for $i = 1, \dots, M$, and $0 < T_1 < \dots < T_{M-1} < T_M \leq T$.

The Bernoulli probability of the RLS process is sample size-dependent so that the level shifts are rare. Otherwise, $\{u_t\}$ would be better construed as a random walk. For $p = 0$, the RLS process nests the no level shift, no trend case. For a discussion of how $E[I_u(\lambda_j)] = O(T/j^2)$ is satisfied for the above processes, we refer the interested reader to McCloskey (2013) and references therein.

2.2 Trimmed FDQML Estimation

Frequency domain estimation is designed to select the parameters that provide the best fit of a spectral density function within a given class of models to the periodogram of the observed time series. The FDQML estimator (sometimes referred to as a Whittle estimator) minimizes the negative logarithm of the frequency domain approximation to the Gaussian likelihood function:

$$L_T(\theta) \equiv T^{-1} \sum_{j \in \mathcal{F}_1} \left\{ \log f(\lambda_j; \theta) + \frac{I_x(\lambda_j)}{f(\lambda_j; \theta)} \right\},$$

where $\mathcal{F}_1 \equiv (-T/2, T/2) \cap \mathbb{Z} \setminus \{0\}$ and $f(\lambda; \theta)$ is the spectral density function of the candidate model with parameters θ , evaluated at frequency λ . For the contaminated processes we examine, the contaminating process component of the observed process $\{x_t\}$ may bias this minimizer away from its true value when frequencies close to zero enter the objective function. Letting $l \in (0, T/2)$ denote the trimming parameter, this leads to the trimmed FDQML estimator: $\hat{\theta}_T \equiv \text{argmin}_{\theta \in \Theta} L_{T,l}(\theta)$ for some appropriate parameter set Θ , where

$$L_{T,l}(\theta) \equiv T^{-1} \sum_{j \in \mathcal{F}_l} \left\{ \log f(\lambda_j; \theta) + \frac{I_x(\lambda_j)}{f(\lambda_j; \theta)} \right\} \quad (1)$$

with $\mathcal{F}_l \equiv (-T/2, T/2) \cap \mathbb{Z} \setminus [-l+1, l-1]$.

3. ASYMPTOTIC PROPERTIES OF ROBUST ESTIMATORS

In what follows, θ_0 denotes the true parameter vector.

3.1 Consistency and Asymptotic Normality of the Trimmed FDQML Estimator

We begin with sufficient conditions for consistency of the trimmed FDQML estimator of a linear process.

Assumption 1. The process $\{x_t\}$ is generated by the following data-generating process (DGP): $x_t = \mu + v_t + u_t$, where μ is a finite constant and v_t and u_t are independent at all leads and lags.

- $\{v_t\}$ is strictly stationary and ergodic and $v_t = \sum_{j=0}^{\infty} c(j; \theta_0) e_{t-j}$, where $c(0; \theta_0) = 1$, $E[e_t] = 0$, $E[e_t e_s] = I(t=s) \sigma^2(\theta_0)$, and $\sum_{j=0}^{\infty} c(j; \theta_0)^2 < \infty$.
- $\Theta \subset \mathbb{R}^s$ is compact and the following properties hold over Θ : (i) $g(\lambda; \theta) \equiv |\sum_{j=0}^{\infty} c(j; \theta) e^{-i\lambda j}|^2$ is continuous in $(\lambda, \theta) \in [-\pi, \pi] \times \Theta$; (ii) $\sigma^2(\theta)$ is continuous and strictly greater than zero over Θ ; (iii) $g(\lambda, \theta) > 0$ for all $(\lambda, \theta) \in [-\pi, \pi] \times \Theta$; (iv) If $\theta_1, \theta_2 \in \Theta$, if $\theta_1 \neq \theta_2$, then $f(\lambda; \theta_1) \neq f(\lambda; \theta_2)$ on a subset of $[-\pi, \pi]$ that is of positive Lebesgue measure; (v) $\theta_0 \in \Theta$.
- $E[I_u(\lambda_j)] = O(T/j^2)$ for $j \in \mathcal{F}_1$.

Note that Assumption 1(a) implies that the spectrum of $\{v_t\}$ takes the form

$$f(\lambda; \theta_0) = \frac{\sigma^2(\theta_0)}{2\pi} g(\lambda; \theta_0).$$

Parts (a) and (b) are variants of standard assumptions for frequency domain estimation of models with linear representations (see, e.g., Condition B of Dunsmuir and Hannan 1976). Part (a) limits the dependence of the contaminated process and part (b) is composed of identification conditions. Assumption 1 contrasts with the assumptions imposed in the related articles of McCloskey and Perron (2013) and McCloskey (2013) in that it applies to short-memory processes with a fully parametric spectrum that is not necessarily derived from a stochastic volatility

model. Part (c) is a high-level assumption that encompasses contamination by many processes of interest.

Remark 1. The methods of this article are useful for estimating the parameters of a short-memory process that is potentially contaminated by a process $\{u_t\}$ with $E[I_u(\lambda_j)] = O(T/j^2)$. If the user is instead interested in semiparametrically estimating the memory parameter of a long-memory process that is robust to the same form of contamination, one may use the methods McCloskey and Perron (2013) or Hou and Perron (2014). If the user is interested in parametrically estimating all of the parameters of a long-memory stochastic volatility model (advanced by Breidt, Crato, and de Lima 1998; Harvey 1998) that is robust to the same form of contamination, one may use the methods of McCloskey (2013). The literature does not appear to contain results for parametric estimation of long-memory models robust to low-frequency contamination outside of this long-memory stochastic volatility class. This would likely require a completely different treatment by building upon the work of Hosoya (1997) and is beyond the scope of the present article.

For consistency, the trimming parameter must grow fast enough to asymptotically rid the FDQML objective function's dependence on the frequencies for which the contaminating component dominates the observed periodogram:

Assumption 2. $l/T + \log^4 T/l \rightarrow 0$.

This is quite a weak assumption since we only require trimming to grow $l \rightarrow \infty$ faster than the slowly growing function $\log^4 T$ but slower than the sample size T . We then have the following consistency result.

Theorem 1. Under Assumptions 1 and 2, $\hat{\theta}_T \xrightarrow{p} \theta_0$.

To establish asymptotic normality of the trimmed FDQML estimator, we strengthen the moment conditions for the innovations in the linear representation of the process, the coefficients in this representation and smoothness conditions on the spectral density function.

Assumption 3. For the process $\{v_t\}$ in Assumption 1, the following hold for its spectrum and innovations:

- (i) $f(\lambda; \theta)$ is twice continuously differentiable in $\theta \in \Theta$. For $\theta \in \Theta$, $\partial f(\lambda; \theta)/\partial \theta$ and $\partial^2 f(\lambda; \theta)/\partial \theta \partial \theta'$ are continuous in $\lambda \in [-\pi, \pi]$; (ii) $\theta_0 \in \text{interior}(\Theta)$.
- $f(\lambda; \theta_0)$ is Hölder continuous of maximal degree $\alpha \in (1/2, 1]$ in λ , that is, there is a constant H such that $|f(\lambda; \theta_0) - f(\omega; \theta_0)| \leq H|\lambda - \omega|^\alpha$ for all $\lambda, \omega \in [-\pi, \pi]$.
- Let \mathcal{I}_t denote the σ -field generated by $e_s, s \leq t$. (i) $E[e_t | \mathcal{I}_{t-1}] = 0$ a.s.; (ii) $E[e_t^2 | \mathcal{I}_{t-1}] = \sigma^2(\theta_0)$ a.s.; (iii) $E[e_t^3 | \mathcal{I}_{t-1}] = \mu_3$ a.s.; (iv) $E[e_t^4] = \mu_4 < \infty$.
- $\sum_{j=0}^{\infty} |c(j; \theta_0)| < \infty$.

Parts (a)–(c) are standard (see, e.g., Condition C of Dunsmuir 1979). Parts (a) and (b) impose smoothness conditions on the spectral density function of the contaminated process. In this case, as in Dunsmuir (1979), we impose linearity (c) with innovations that are third-order martingale differences solely to render a parametric asymptotic variance form. This may be restrictive in some applications where the second- and third-

order martingale difference properties are unknown or higher moments may not exist. Nevertheless, simulation evidence in Section 4 indicates that the trimmed FDQML estimator continues to perform well in the presence of heavy-tailed data. Part (d) requires the contaminated process to be weakly dependent.

Asymptotic normality requires we also impose a stronger negligible trimming condition. In particular $l \rightarrow \infty$ faster than $T^{1/2}$ but slower than T ensures identification and $T^{1/2}$ -asymptotics follow from the periodogram I_v of the stationary process $\{v_t\}$.

Assumption 4. $l/T + T^{1/2}/l \rightarrow 0$.

Theorem 2. Under Assumptions 1, 3, and 4, $T^{1/2}(\hat{\theta}_T - \theta_0)$ is asymptotically multivariate normal with zero mean and covariance matrix given by $\Omega^{-1}(2\Omega + \Pi)\Omega^{-1}$, where

$$\Omega_{hk} \equiv \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\partial \log f(\lambda; \theta_0)}{\partial \theta_h} \frac{\partial \log f(\lambda; \theta_0)}{\partial \theta_k} d\lambda$$

and

$$\Pi_{hk} \equiv \frac{\kappa_e(\theta_0)}{(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \frac{\partial \log f(\lambda; \theta_0)}{\partial \theta_h} \frac{\partial \log f(\omega; \theta_0)}{\partial \theta_k} d\lambda d\omega$$

with $\kappa_e(\theta_0) \equiv \mu_4/\sigma^4(\theta_0) - 3$ (the kurtosis of e_t).

It is worth reiterating here that consistency (Theorem 1) and asymptotic normality (Theorem 2) cannot be obtained by estimation in the time domain unless one takes a parametric approach to the form of contamination u_t . Assumption 1(c) allows for a much more robust approach in allowing for a wide range of contamination, or even the absence of contamination altogether.

Remark 2. The asymptotic covariance matrix of Theorem 2 is standard in the literature on FDQML estimation (see Dunsmuir 1979). Hence, neither the sample size-dependent trimming nor the presence of low-frequency contamination reduces the asymptotic efficiency of the estimator.

Remark 3. If $E[e_t^2] = \sigma^2$ is separately parameterized from θ , one may alternatively estimate θ by minimizing a simplified objective function (see, e.g., Hannan 1973 for linear processes and Giraitis and Robinson 2001 and Mikosch and Straumann 2002 for GARCH processes): $(2\pi/T) \sum_{j \in \mathcal{F}_l} I_x(\lambda_j)/g(\lambda_j; \theta)$. In the absence of contaminating components, this simplified estimator is asymptotically equivalent to the untrimmed FDQML estimator. However, this equivalence breaks down when a trimming large enough to make the simplified estimator robust to low-frequency contamination is used. In this case, the estimator becomes asymptotically biased with a degenerate limiting distribution and slower than standard parametric rate of convergence. (These results are omitted for the sake of brevity and available from the authors upon request.) We therefore recommend against trimming this simplified objective function for robust estimation.

Remark 4. Rather than contamination by a process $\{u_t\}$ with $E[I_u(\lambda_j)] = O(T/j^2)$, one may wonder whether it is possible to estimate $\{v_t\}$ satisfying Assumptions 1(a)–(b) and 3 when it is contaminated by a long-memory process. If $\{u_t\}$ were a long-memory process with memory parameter d , it would instead satisfy $E[I_u(\lambda_j)] = O(T^{2d}/j^{2d})$. A periodogram decomposition similar to that used by McCloskey and Perron (2013) would imply that $I_x(\lambda_j) = I_v(\lambda_j) + I_u(\lambda_j) + 2I_{vu}(\lambda_j)$, where $I_v(\lambda_j) = O_p(1)$, $I_u(\lambda_j) = O_p(T^{2d}/j^{2d})$, and $I_{vu} = O_p(T^d/j^d)$.

When $d > 0$, the typical range of interest in economic applications, (i) for all λ_j such that $\lim_{T \rightarrow \infty} j/T = 0$, $I_u(\lambda_j)$ would asymptotically dominate the periodogram of the observed process and (ii) for all λ_j such that $\lim_{T \rightarrow \infty} |j|/T \in (0, 1]$, neither $I_u(\lambda_j)$ nor $I_v(\lambda_j)$ would dominate. Thus, it does not appear possible to robustly estimate the parameters of a short-memory process $\{v_t\}$ when it is contaminated by a long-memory process $\{u_t\}$ by using the frequency domain trimming employed in this article. However, by parametrically specifying the spectral density function of $\{x_t\}$ to include a short-memory component and a separate long-memory component, existing frequency domain estimation methods would allow one to form consistent and asymptotically normal estimates of the parameters of both components. See, for example, Hosoya (1997) and references therein.

3.2 Examples of Processes Satisfying the Assumptions

We now discuss a few popular models for the contaminated process $\{v_t\}$.

3.2.1 ARMA Models. Stationary ergodic ARMA(p, q) processes satisfy Assumption 1(a). If the parameter set Θ is restricted to models for which (i) the roots of the autoregressive and moving average polynomials, $A(z; \theta)$ and $B(z; \theta)$ say, lie outside of the unit circle, (ii) $A(z; \theta)$ and $B(z; \theta)$ have no common roots, and (iii) the highest order AR or MA parameters are nonzero, then these ARMA(p, q) models satisfy Assumptions 1(b) and 3(a)–(b).

3.2.2 Stochastic Volatility Models. The standard short-memory SV model with an additive low-frequency component contaminating the volatility is given by a returns process $\{r_t\}$ that follows

$$r_t = \sigma_t \varepsilon_t, \quad \sigma_t = \sigma \exp(h_t/2), \quad (2)$$

where $\varepsilon_t \sim \text{iid}(0, 1)$, $\{h_t\}$ is independent of $\{\varepsilon_t\}$ with

$$h_t = u_t + \tilde{v}_t, \quad A(L; \theta)\tilde{v}_t = B(L; \theta)\xi_t, \quad \xi_t \sim \text{iid}(0, \sigma_\xi^2), \quad (3)$$

the polynomials $A(z; \theta)$ and $B(z; \theta)$ satisfying the same conditions as for the above ARMA(p, q) case, and $\{u_t\}$ satisfying Assumption 1(c). Qu and Perron (2013) studied a special case of a contaminated AR(1) version of this model, providing a Bayesian procedure for estimation and inference. In contrast, our frequentist methods apply to a much larger class of contaminated and contaminating processes.

For this contaminated SV model, the log-squared returns have the following decomposition:

$$x_t \equiv \log(r_t^2) = \log \sigma^2 + E[\log(\varepsilon_t^2)] + u_t + \tilde{v}_t + (\log(\varepsilon_t^2) - E[\log(\varepsilon_t^2)]) \equiv \mu + u_t + v_t,$$

where $\mu = \log \sigma^2 + E[\log(\varepsilon_t^2)]$ and $v_t = \tilde{v}_t + \zeta_t$ with $\zeta_t \equiv \log(\varepsilon_t^2) - E[\log(\varepsilon_t^2)]$. Given that $\{h_t\}$ is independent of $\{\varepsilon_t\}$, the spectral density function of $\{v_t\}$ is

$$f(\lambda; \theta) = \frac{\sigma_\xi^2 |B(e^{-i\lambda}; \theta)|^2}{2\pi |A(e^{-i\lambda}; \theta)|^2} + \frac{\sigma_\zeta^2}{2\pi},$$

where $\theta_1 = \sigma_\xi^2$ and $\theta_2 = \sigma_\zeta^2$ are the first two entries of the parameter vector. Under the same assumptions imposed on the ARMA component $\{\tilde{v}_t\}$ as those imposed above for the pure

ARMA model, $\{v_t\}$ is weakly stationary and therefore has the linear representation required by Assumption 1(a). If the ARMA component $\{\tilde{v}_t\}$ is purely autoregressive of finite order and all roots of $A(z; \theta)$ lie outside of the unit circle for $\theta \in \Theta$, Assumptions 1(b) and 3(a)–(b) follow from results in Pagano (1974).

3.2.3 (G)ARCH Models. Finally, we turn to contaminated (G)ARCH models for the returns process $\{r_t\}$:

$$x_t \equiv r_t^2 = \tilde{v}_t + u_t, \quad \text{where } \tilde{v}_t = h_t \varepsilon_t^2, \quad E[\varepsilon_t | \mathcal{I}_{t-1}^\varepsilon] = 0, \\ E[\varepsilon_t^2 | \mathcal{I}_{t-1}^\varepsilon] = 1, \quad (4)$$

$$h_t = E[\tilde{v}_t | \mathcal{I}_{t-1}^\varepsilon] = \psi_0(\theta) + \sum_{j=1}^{\infty} \psi_j(\theta) \tilde{v}_{t-j},$$

$\{u_t\}$ satisfies Assumption 1(c) and $\mathcal{I}_t^\varepsilon$ is the σ -field of events generated by $\varepsilon_s, s \leq t$. From (4), several specifications for the returns process $\{r_t\}$ may arise, perhaps the most natural being

$$r_t = \sqrt{\tilde{v}_t + u_t} (I(\varepsilon_t \geq 0) - I(\varepsilon_t < 0)),$$

which nests the standard (G)ARCH framework when $u_t = 0$. One may view the process $\{\tilde{v}_t\}$ as the squares of a latent (G)ARCH process whose parameter vector θ we wish to estimate. For this contaminated (G)ARCH model,

$$x_t = \tilde{v}_t + u_t \equiv \mu + v_t + u_t \quad \text{with} \\ \mu = \psi_0(\theta) / \left(1 - \sum_{j=1}^{\infty} \psi_j(\theta) \right) \quad \text{and} \quad v_t = \tilde{v}_t - \mu$$

since \tilde{v}_t has the representation

$$\tilde{v}_t = \psi_0(\theta) + \sum_{j=1}^{\infty} \psi_j(\theta) \tilde{v}_{t-j} + \xi_t, \quad (5)$$

where $\xi_t \equiv \tilde{v}_t - h_t$ are martingale differences. If $\{\tilde{v}_t\}$ is second-order stationary, $\{v_t\}$ has the spectral density function

$$f(\lambda; \theta) = \frac{\sigma_\xi^2}{2\pi |1 - \sum_{j=1}^{\infty} \psi_j(\theta) e^{-ij\lambda}|^2}. \quad (6)$$

The GARCH(p, q) model similarly has an ARMA ($\max(p, q), q$) representation and an invertible GARCH(p, q) model is a subcase of the model for \tilde{v}_t in (5). The functional form for the spectral density function (6) resembles that given in Assumption 1(a) but in this contaminated (G)ARCH case, the sequence $\{\xi_t\}$ is not independent in second moments, violating Assumption 3(c). Hence, a separate proof of asymptotic normality for the trimmed FDQML estimator is required.

3.3 Consistency and Asymptotic Normality for Contaminated (G)ARCH Models

We now provide conditions under which the trimmed FDQML estimator of the parameters of a contaminated (G)ARCH process given by (4) is consistent and asymptotically normal.

Assumption ARCH-J. $\tilde{v}_t = h_t \varepsilon_t^2$, where $\{\varepsilon_t\}$ is strictly stationary and ergodic and (i) $E[\varepsilon_t | \mathcal{I}_{t-1}^\varepsilon] = 0$ a.s.; (ii) $E[\varepsilon_t^2 | \mathcal{I}_{t-1}^\varepsilon] = 1$ a.s.; (iii) $E[\varepsilon_t^{2j} | \mathcal{I}_{t-1}^\varepsilon] \equiv \beta_{2j} < \infty$ (constants)

for $j = 2, \dots, J/2$ a.s.; (iv) $h_t \equiv E[\tilde{v}_t | \mathcal{I}_{t-1}^e] = \psi_0(\theta_0) + \sum_{j=1}^{\infty} \psi_j(\theta_0) \tilde{v}_{t-j}$ with $\psi_0(\theta_0) > 0$, $\psi_j(\theta_0) \geq 0$ for $j \geq 1$, and $|\beta_j|^{2/J} \sum_{j=1}^{\infty} \psi_j(\theta_0) < 1$.

This assumption comes from Giraitis and Robinson (2001) who imposed it to obtain consistency and asymptotic normality of a standard (untrimmed) simplified FDQML estimator of a (G)ARCH process when no contaminating components are present. It is a simultaneous restriction on the dependence and tail behavior of $\{\tilde{v}_t\}$, allowing for heavier tails when the process exhibits less serial dependence. Assumptions ARCH-4 and 1(b) imply Assumption 1(a), giving the following corollary to Theorem 1:

Corollary 1. For the (G)ARCH process given by (4), under Assumption ARCH-4, suppose Assumption 1(b) holds for

$$g(\lambda; \theta) = |1 - \sum_{j=1}^{\infty} \psi_j(\theta) e^{ij\lambda}|^{-2},$$

and $f(\lambda; \theta) = (\sigma^2/2\pi)g(\lambda; \theta)$ with $\sigma^2 = E[v_t^2] - E[h_t^2]$. Then, if Assumptions 1(c) and 2 also hold, $\hat{\theta}_T \xrightarrow{p} \theta_0$.

Remark 5. Assumption 1(b)(i) holds when $\{\tilde{v}_t\}$ is modeled as the squares of a stationary ARCH(p) or GARCH(p,q) process, which follows from its implicit ARMA(max(p, q), q) representation. The remainder of Assumption 1(b) is implied by Assumptions ARCH-4 and 2 of Giraitis and Robinson (2001).

Theorem 3. Under Assumption ARCH-8, suppose Assumptions 1(b) and 3(a)–(b) hold for

$$g(\lambda; \theta) = |1 - \sum_{j=1}^{\infty} \psi_j(\theta) e^{ij\lambda}|^{-2},$$

and $f(\lambda; \theta) = (\sigma^2/2\pi)g(\lambda; \theta)$ with $\sigma^2 = E[v_t^2] - E[h_t^2]$. Then, if Assumptions 1(c) and 4 also hold and $\partial f(\lambda; \theta_0)^{-1}/\partial \theta_k$ is Hölder continuous of degree greater than 1/2 in λ for all $k = 1, \dots, s$, $T^{1/2}(\hat{\theta}_T - \theta_0)$ is asymptotically multivariate normal with zero mean and covariance matrix given by $\Omega^{-1}(2\Omega + \Pi)\Omega^{-1}$, where

$$\Omega_{hk} \equiv \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\partial \log f(\lambda; \theta_0)}{\partial \theta_h} \frac{\partial \log f(\lambda; \theta_0)}{\partial \theta_k} d\lambda,$$

$$\Pi_{hk} \equiv \frac{2\pi}{\sigma^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \frac{\partial f(\lambda; \theta_0)^{-1}}{\partial \theta_h} \frac{\partial f(\omega; \theta_0)^{-1}}{\partial \theta_k} f_4(\lambda, -\omega, \omega; \theta_0) d\lambda d\omega,$$

$$f_4(\lambda, \omega, \nu; \theta_0) \equiv \frac{1}{(2\pi)^3} \sum_{j,k,m=-\infty}^{\infty} e^{-i\lambda j - i\omega k - i\nu l}$$

$$\text{Cum}(\tilde{v}_0, \tilde{v}_j, \tilde{v}_k, \tilde{v}_m)$$

(the fourth cumulant spectrum of $\{\tilde{v}_t\}$) and $\text{Cum}(w, x, y, z)$ is the joint cumulant of generic random variables w, x, y , and z .

Remark 6. As noted by previous authors (see Giraitis and Robinson 2001; Mikosch and Straumann 2002), ARCH-8 may be a strong assumption in certain contexts. For example, it implies the existence of the eighth moment of the ARCH innovations ε_t , whereas standard time domain methods typically impose the

existence of the fourth moment. Nevertheless, Corollary 1 shows that consistency is still attainable under the weaker ARCH-4 Assumption.

Remark 7. The limiting covariance matrices of Theorems 2 and 3 can be difficult to estimate in practice, particularly in the (G)ARCH framework (see Remark 2.2 of Giraitis and Robinson 2001). This challenge, which is also present for standard frequency domain estimators in the absence of data contamination, is exacerbated by the potential presence of data contamination. Estimates of these variances and/or valid bootstrap procedures may still be feasible. The works of Taniguchi (1982) and Chiu (1988) are likely to provide useful starting points for this direction of research.

4. FINITE SAMPLE PROPERTIES OF ROBUST FDQML ESTIMATORS

We now study the finite sample properties of our trimmed FDQML estimators in the presence and absence of low-frequency contamination, and under thin and heavy tails. We chose to examine the models below for their popularity in applied work and ease in interpreting what the “persistence parameters” of the model are. The values reported in this section all result from 1000 Monte Carlo replications.

4.1 The AR(1) Model

We begin by fitting an AR(1) time series in the absence of contamination:

$$x_t = v_t = a v_{t-1} + e_t, \quad e_t \sim \text{iid}(0, \sigma_e^2). \quad (7)$$

We use $\sigma_{e0}^2 = 1$, and $a_0 = 0, 0.5, 0.95$ as they are representative of the range of persistent values for a_0 . We estimate (a_0, σ_{e0}^2) on $[0, 0.99] \times [0.01, 10]$. (Given that the minimization is conducted over $a \in [0, 0.99]$, Assumption 3(a)(ii) does not technically hold when $a_0 = 0$. However, if asymptotic normality is required, one can easily expand the search space.) We compute the standard FDQML and three trimmed FDQML estimators with trimmings set as $l = \lfloor T^{0.4} \rfloor, \lfloor T^{0.51} \rfloor, \lfloor T^{0.6} \rfloor$ to illustrate the broad picture. We use sample sizes $T = 1000, 2000$, and 4000, which align with sizes available for financial returns data. The finite sample biases and root mean squared errors (RMSEs) for a_0 for the uncontaminated processes are given in the top three panels of Tables 1 and 2. The first column shows that for Gaussian innovations, $e_t \sim \text{iid } \mathcal{N}(0, 1)$, all four estimators of a_0 perform well with very little bias or RMSE differences except when using the largest trimming $l = \lfloor T^{0.6} \rfloor$. We can see from the third column of Tables 1 and 2 that changing the innovation distribution to be heavy-tailed hardly changes the results. Here, e_t has a symmetric Pareto distribution such that $P(e_t < -c) = P(e_t > c) = 0.5(1+c)^{-\kappa}$ with tail index $\kappa = 2.5$, standardized so that $\sigma_e^2 = E[e_t^2] = 1$. In this case, v_t has an infinite third moment, in violation of Assumption 3(c).

Next, we contaminate the AR(1) model by adding u_t to v_t in (7), where $\{u_t\}$ is the RLS process described in Section 2.1. Again, $\sigma_{e0}^2 = 1$ but now the variance of a level shift σ_η^2 is set equal to the variance of v_t . We examine a range of specifications for p and results for the representative value $p = 10$ are

Table 1. Bias for AR(1) processes with(out) RLS

T	$l =$	Trimming; $\mathcal{N}(0, 1)$				Trimming; Par(2.5)			
		1	$T^{0.40}$	$T^{0.51}$	$T^{0.60}$	1	$T^{0.40}$	$T^{0.51}$	$T^{0.60}$
$a_0 = 0, p = 0$									
1000		0.000	-0.001	0.000	-0.001	0.001	0.000	0.000	0.000
2000		-0.001	-0.001	0.000	-0.001	0.001	-0.001	0.001	0.001
4000		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$a_0 = 0.5, p = 0$									
1000		-0.001	-0.001	-0.001	-0.001	-0.001	-0.002	-0.004	-0.001
2000		-0.001	-0.001	0.000	-0.001	-0.001	-0.002	-0.001	0.001
4000		-0.001	-0.001	0.000	0.000	-0.001	0.000	0.000	-0.001
$a_0 = 0.95, p = 0$									
1000		-0.003	-0.003	-0.012	-0.041	-0.003	-0.004	-0.016	-0.041
2000		-0.001	0.000	0.000	-0.012	-0.002	-0.001	-0.002	-0.019
4000		-0.001	-0.001	0.001	-0.002	-0.001	-0.001	0.002	-0.004
$a_0 = 0, p = 10$									
1000		0.504	0.057	0.026	0.012	0.518	0.062	0.028	0.013
2000		0.512	0.046	0.019	0.009	0.511	0.047	0.021	0.009
4000		0.505	0.032	0.015	0.006	0.519	0.035	0.015	0.006
$a_0 = 0.5, p = 10$									
1000		0.301	0.042	0.020	0.010	0.297	0.042	0.018	0.010
2000		0.300	0.033	0.014	0.007	0.297	0.033	0.013	0.007
4000		0.298	0.024	0.009	0.004	0.302	0.026	0.010	0.004
$a_0 = 0.95, p = 10$									
1000		0.028	0.003	-0.014	-0.042	0.028	0.002	-0.012	-0.040
2000		0.030	0.005	-0.002	-0.019	0.030	0.005	0.002	-0.013
4000		0.032	0.003	0.003	-0.004	0.031	0.003	0.004	-0.002

NOTE: In this table, a_0 refers to the true autoregressive parameter and p equals the Bernoulli probability of a random level shift in the (contaminated) AR(1) model described in Section 4.1. No contamination is implied by $p = 0$. The columns corresponding to the $\mathcal{N}(0, 1)$ block indicate $e_t \sim \text{iid}\mathcal{N}(0, 1)$ while those corresponding to the Par(2.5) block indicate e_t are iid draws from a symmetric Paretian distribution with tail index $\kappa = 2.5$. The trimmed set of Fourier frequency indices is $(-T/2, T/2) \cap \mathbb{Z} \setminus [-l + 1, l - 1]$ for $l = 1, T^{0.40}, T^{0.51}, T^{0.60}$. No trimming occurs when $l = 1$.

presented here. The finite sample biases and RMSEs for estimators of a_0 for these processes are displayed in the final three panels of Tables 1 and 2. Beginning with the case of contaminated white noise (i.e., $a_0 = 0$) with Gaussian or heavy-tailed innovations, we can see that the level shifts induce a very large upward bias in the untrimmed estimator of a_0 that appears to stabilize around 0.51. We can also see that trimming removes most of this bias, resulting in *substantially* lower RMSE. Adding short-run dynamics to the $\{v_t\}$ process by setting $a_0 = 0.5$ or 0.95 does not change the general pattern except for decreasing the finite sample biases. Too high a trimming (e.g., $l = \lfloor T^{0.6} \rfloor$) can induce a *downward* bias that overwhelms the upward bias due to the level shifts but, as in the uncontaminated case, this problem can be avoided by using a more moderately sized trimming.

Remark 8. The trimmed FDQML estimator continues to perform quite well in the presence of heavy-tailed data that violates Assumption 3(c). This is also the case for the autoregressive SV model of the following section. These simulation results appear to complement the theoretical results of Mikosch et al. (1995) who found that frequency domain estimators continue to perform well even in the presence of an infinite second moment.

4.2 The Autoregressive SV Model

The (contaminated) autoregressive SV (ARSV) model we examine is specified in (2) and (3) with $A(L; \theta) = 1 - aL$,

$B(L) = 1$, $\varepsilon_t \sim \text{iid}\mathcal{N}(0, 1)$ and $\{u_t\}$ being an RLS process. Our thin-tailed specification sets $\xi_t \sim \text{iid}\mathcal{N}(0, 1)$, while our heavy-tailed one uses ξ_t with a symmetric Paretian distribution with tail index equal to 2.5, as defined in the previous subsection. FDQML estimation of $\theta = (a, \sigma_\xi^2, \sigma_\zeta^2)'$ is performed using the periodogram of the logarithm of the squares of r_t . We examine $a_0 = 0.5$ and 0.95. Using the same sample sizes and trimmings and the analogous parameter space as in Section 4.1, the finite sample biases and RMSEs for estimators of a_0 are provided in Table 3. Beginning with the case of no contamination (for which $p = 0$) and $a_0 = 0.5$, we again see good performance and little differences across estimators in the presence of thin or heavy tails. For the case of high persistence, $a_0 = 0.95$, we encounter the same phenomenon as in the AR(1) model: a moderate trimming should be used.

Adding an RLS contaminating component by setting $p = 10$ and the variance of a level shift σ_η^2 to the variance of \tilde{v}_t , it is first interesting to note that, unlike for the AR(1) models, $E[\hat{a}]$ does not appear to stabilize at some value less than one. Instead, it seems that $E[\hat{a}] \rightarrow 1$. For the case of $a_0 = 0.5$, trimming removes the vast majority of the bias induced by level shifts and the trimming of $l = \lfloor T^{0.51} \rfloor$ seems to perform the best. We again see that competing biases arise from too high of a trimming that may discard valuable information about v_t , and too low of a trimming that subjects the estimator to upward bias from u_t . This bias competition is quite pronounced in the cases for which $a_0 = 0.95$ and we can see its presence in the RMSE values. Often

Table 2. RMSE for AR(1) processes with(out) RLS

T	l =	Trimming; $\mathcal{N}(0, 1)$				Trimming; Par(2.5)			
		1	$T^{0.40}$	$T^{0.51}$	$T^{0.60}$	1	$T^{0.40}$	$T^{0.51}$	$T^{0.60}$
$a_0 = 0, p = 0$									
1000		0.032	0.033	0.035	0.037	0.031	0.031	0.034	0.035
2000		0.023	0.023	0.023	0.025	0.022	0.022	0.024	0.025
4000		0.016	0.016	0.017	0.017	0.016	0.016	0.016	0.017
$a_0 = 0.5, p = 0$									
1000		0.032	0.035	0.038	0.051	0.028	0.029	0.034	0.046
2000		0.025	0.026	0.027	0.032	0.020	0.020	0.022	0.028
4000		0.021	0.021	0.022	0.024	0.014	0.014	0.015	0.017
$a_0 = 0.95, p = 0$									
1000		0.032	0.045	0.065	0.105	0.011	0.034	0.056	0.097
2000		0.031	0.036	0.048	0.068	0.008	0.020	0.038	0.066
4000		0.030	0.032	0.039	0.052	0.005	0.010	0.024	0.042
$a_0 = 0, p = 10$									
1000		0.541	0.079	0.048	0.041	0.556	0.087	0.050	0.041
2000		0.553	0.064	0.034	0.028	0.552	0.066	0.034	0.028
4000		0.546	0.045	0.025	0.019	0.559	0.049	0.025	0.019
$a_0 = 0.5, p = 10$									
1000		0.319	0.062	0.046	0.052	0.315	0.060	0.042	0.047
2000		0.317	0.050	0.034	0.034	0.313	0.047	0.029	0.029
4000		0.315	0.038	0.026	0.025	0.318	0.037	0.019	0.019
$a_0 = 0.95, p = 10$									
1000		0.043	0.045	0.066	0.104	0.031	0.033	0.056	0.097
2000		0.044	0.037	0.049	0.073	0.032	0.020	0.037	0.061
4000		0.045	0.032	0.039	0.053	0.032	0.010	0.026	0.042

NOTE: In this table, a_0 refers to the true autoregressive parameter and p equals the Bernoulli probability of a random level shift in the (contaminated) AR(1) model described in Section 4.1. No contamination is implied by $p = 0$. The columns corresponding to the $\mathcal{N}(0, 1)$ block indicate $e_t \sim \text{iid}\mathcal{N}(0, 1)$ while those corresponding to the Par(2.5) block indicate e_t are iid draws from a symmetric Paretian distribution with tail index $\kappa = 2.5$. The trimmed set of Fourier frequency indices is $(-T/2, T/2] \cap \mathbb{Z} \setminus [-l + 1, l - 1]$ for $l = 1, T^{0.40}, T^{0.51}, T^{0.60}$. No trimming occurs when $l = 1$.

to a lesser degree, this bias competition is in fact a generic feature of the trimmed FDQML estimation procedure in the presence of low-frequency contamination, manifesting itself to varying extents in certain models and parameter configurations, the present one being a particularly dramatic case. Figures 1 and 2 more clearly illustrate this phenomenon, displaying the absolute bias and RMSE of \hat{a} as a function of the trimming exponent for the three sample sizes when $a_0 = 0.95$ and $p = 10$.

Remark 9. There is an important parallel between the competing biases present in this context and the context of semi-parametric frequency domain estimation of the memory parameter that is also robust to low-frequency contamination, as in McCloskey and Perron (2013). In this latter case, too high of a trimming induces bias in the estimate of the memory parameter in the presence of unmodeled noise or short-run dynamics.

4.3 The GARCH(1,1) Model

The final process we study is GARCH(1,1), with uncontaminated version

$$r_t = h_t^{1/2} \varepsilon_t \text{ with } h_t = c + ar_{t-1}^2 + bh_{t-1} \text{ and } \varepsilon_t \sim \text{iid}\mathcal{N}(0, 1).$$

FDQML estimation of $\theta = (a, b, \sigma_\varepsilon^2)'$ is performed on the periodogram of the squares of r_t . We examine the values of

$(a_0, b_0) = (0.05, 0.3), (0.05, 0.6), (0.05, 0.9)$. (ARCH-8 does not hold for the latter specification but ARCH-4 and the other assumptions required for consistency does). The parameter space Θ is the subset of $[0.01, 0.99] \times [0.01, 0.99] \times [0.1, 10]$ for which $a + b$ is restricted to be no larger than 0.99. We use the same trimmings as above but look at sample sizes of 1000, 4000, and 16,000 since accurate FDQML estimation of the GARCH(1,1) model requires somewhat larger samples. The finite sample biases and RMSEs for estimators of (a_0, b_0) corresponding to these uncontaminated GARCH(1,1) processes are recorded in the top three panels of Tables 4 and 5, where “QML” corresponds here to the standard time domain quasi-maximum likelihood estimator, included for comparison. When $(a_0, b_0) = (0.05, 0.3)$ or $(0.05, 0.6)$, the (trimmed) FDQML estimators exhibit almost no bias when moderate trimmings are used. The standard time domain estimator RMSE dominates the FDQML estimators but not to a major extent. For the case of highest persistence, $(a_0, b_0) = (0.05, 0.9)$, the FDQML estimators run into some problems even under moderate trimming, with substantial downward biases, though estimates using a small trimming $l = \lfloor T^{0.4} \rfloor$ perform fairly well in larger samples.

We now add a DLS component to the GARCH(1,1) model:

$$r_t = \sqrt{\tilde{v}_t + u_t}(I(\varepsilon_t \geq 0) - I(\varepsilon_t < 0)), \text{ where } \tilde{v}_t = h_t \varepsilon_t^2, \\ h_t = E[\tilde{v}_t | \mathcal{I}_{t-1}^\varepsilon] = c + a\tilde{v}_{t-1} + bh_{t-1}$$

Table 3. Bias and RMSE for ARSV processes with(out) RLS

T	$l =$	Trimming; $\mathcal{N}(0, 1)$				Trimming; Par(2.5)			
		1	$T^{0.40}$	$T^{0.51}$	$T^{0.60}$	1	$T^{0.40}$	$T^{0.51}$	$T^{0.60}$
Bias: $a_0 = 0.5, p = 0$									
1000		-0.004	-0.003	-0.006	0.050	-0.017	-0.006	0.001	0.000
2000		-0.002	-0.001	-0.001	0.001	-0.006	-0.006	-0.004	0.002
4000		0.001	0.001	0.000	0.001	-0.004	-0.005	-0.001	-0.002
Bias: $a_0 = 0.95, p = 0$									
1000		-0.003	-0.025	-0.136	-0.333	-0.003	-0.003	-0.033	-0.114
2000		-0.002	-0.003	-0.039	-0.196	-0.002	-0.001	-0.006	-0.048
4000		-0.001	-0.001	-0.008	-0.074	-0.001	-0.001	0.003	-0.015
Bias: $a_0 = 0.5, p = 10$									
1000		0.458	0.087	0.038	0.043	0.447	0.097	0.034	0.028
2000		0.464	0.077	0.027	0.012	0.453	0.082	0.025	0.007
4000		0.474	0.066	0.024	0.002	0.466	0.070	0.024	0.008
Bias: $a_0 = 0.95, p = 10$									
1000		0.030	-0.014	-0.107	-0.300	0.008	-0.004	-0.030	-0.105
2000		0.033	0.003	-0.039	-0.188	0.0010	0.001	-0.009	-0.041
4000		0.034	0.003	-0.004	-0.069	0.010	0.000	0.002	-0.013
RMSE: $a_0 = 0.5, p = 0$									
1000		0.127	0.144	0.176	0.265	0.131	0.137	0.163	0.235
2000		0.090	0.098	0.114	0.149	0.087	0.094	0.108	0.142
4000		0.061	0.063	0.073	0.086	0.063	0.065	0.069	0.086
RMSE: $a_0 = 0.95, p = 0$									
1000		0.035	0.088	0.284	0.515	0.013	0.039	0.084	0.208
2000		0.032	0.046	0.109	0.360	0.009	0.022	0.048	0.110
4000		0.031	0.035	0.056	0.173	0.007	0.011	0.029	0.060
RMSE: $a_0 = 0.5, p = 10$									
1000		0.464	0.177	0.173	0.267	0.455	0.174	0.168	0.243
2000		0.470	0.138	0.117	0.154	0.460	0.137	0.109	0.140
4000		0.477	0.111	0.079	0.092	0.471	0.108	0.078	0.081
RMSE: $a_0 = 0.95, p = 10$									
1000		0.045	0.068	0.229	0.477	0.016	0.038	0.084	0.204
2000		0.046	0.045	0.112	0.339	0.014	0.023	0.049	0.104
4000		0.047	0.035	0.055	0.156	0.014	0.012	0.029	0.059

NOTE: In this table, a_0 refers to the true autoregressive parameter and p equals the Bernoulli probability of a random level shift in the (contaminated) ARSV model described in Section 4.2. No contamination is implied by $p = 0$. The columns corresponding to the $\mathcal{N}(0, 1)$ block indicate $\xi_t \sim \text{iid}\mathcal{N}(0, 1)$ while those corresponding to the Par(2.5) block indicate ξ_t are iid draws from a symmetric Paretian distribution with tail index $\kappa = 2.5$. The trimmed set of Fourier frequency indices is $(-T/2, T/2] \cap \mathbb{Z} \setminus [-l+1, l-1]$ for $l = 1, T^{0.40}, T^{0.51}, T^{0.60}$. No trimming occurs when $l = 1$.

and $\varepsilon_t \sim \text{iid}\mathcal{N}(0, 1)$. Letting $\text{var}(\tilde{v}_t) = \sigma_v^2$,

$$u_t = \sigma_v^2 \{0.24I(T/4 < t \leq T/2) + 0.48I(T/2 < t \leq 3T/4) + 0.12I(3T/4 < t \leq T)\}$$

is a DLS process with three level shifts. We now wish to estimate the parameters of the latent $\{\tilde{v}_t\}$ process. The finite sample biases and RMSEs of the estimators are given in the lower three panels of Tables 4 and 5. Starting with the case of least persistence, the trimmed FDQML estimators of a and b clearly outperform both the standard FDQML and time domain estimators. As the sample size grows, it appears that the level shifts cause the standard estimators to have $E[\hat{a}] + E[\hat{b}] \rightarrow 1$, consistent with the results of Lamoureux and Lastrapes (1990) and Hillebrand (2005). On the other hand, the trimmed FDQML estimators remove very substantial portions of the biases in \hat{a} and \hat{b} with moderate trimmings typically yielding the lowest bias. For the case $(a_0, b_0) = (0.05, 0.6)$, the overall pattern is quite similar. The results become more nuanced in the highly persistent case

though the trimmed FDQML estimators often have lower biases than the standard time domain estimator.

The Monte Carlo results of this section show that if the potential for additive low-frequency contamination is present, trimmed FDQML estimators can provide *major* improvements in terms of bias and RMSE when care is taken regarding the choice of trimming. A moderate trimming, for example, $T = \lfloor T^{0.51} \rfloor$, performs quite well in most scenarios.

5. EMPIRICAL APPLICATIONS

We now demonstrate the practical differences between low-frequency contamination-robust and nonrobust estimation by estimating the parameters of two popular time series models.

5.1 Autoregressive SV Estimation of High-Frequency Exchange Rate Returns

For this application, we follow Deo, Hurvich, and Lu (2006) in modeling high-frequency log returns via an SV model

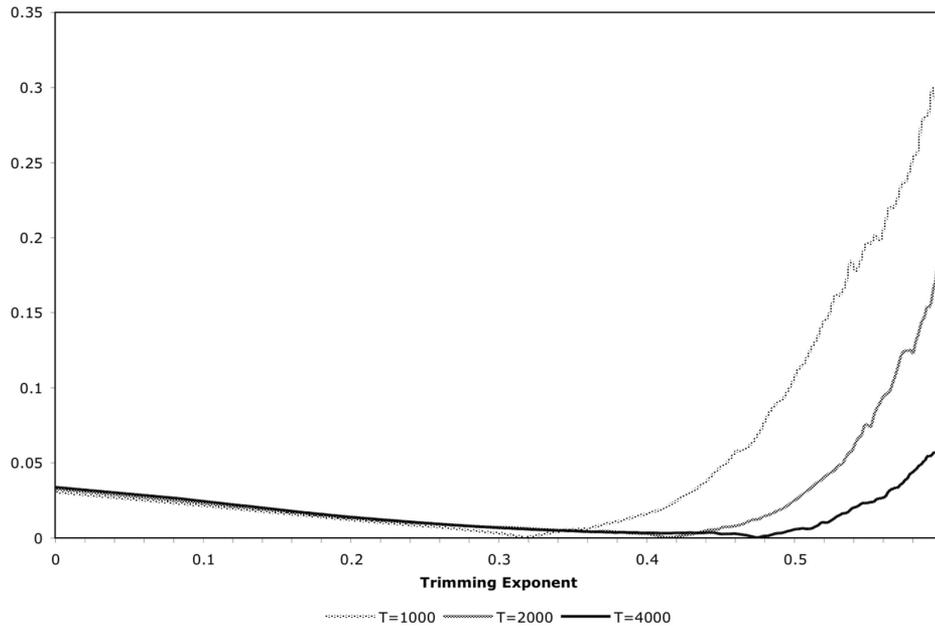


Figure 1. Absolute bias, $a_0 = 0.95$.

although we do not allow for long-memory. As we will see, allowing for long-memory in the model may be unnecessary when low-frequency contamination is accounted for. The data are 30 min log returns of the Japanese Yen per U.S. dollar exchange rate from 10:30 pm on December 12, 1986, to 10:00 pm on June 29, 1999 (148,416 observations). (The cleaned data and Andersen and Bollerslev (1997) deseasonalization code were kindly provided by Denis Tkachenko.) The choice of 30 min sampling frequency is used to reduce microstructure noise (see Deo, Hurvich, and Lu 2006).

We use two different deseasonalization approaches to rid the data of intraday periodicity (see Andersen and Bollerslev 1997).

The first is the flexible Fourier form approach of Andersen and Bollerslev (1997) (see their Appendix B). We deseasonalize 5 min return data using six trigonometric terms and then aggregate the 5 min deseasonalized returns to the 30 min sampling frequency. (In the notation of (A.3) of Andersen and Bollerslev (1997), we set $J = 0$, $D = 0$ and $P = 6$.) The second deseasonalization approach is due to Deo, Hurvich, and Lu (2006). It is conducted directly in the frequency domain by ignoring certain Fourier frequencies of the periodogram that are fixed with the sample size, allowing the deseasonalization to occur concurrently with the (trimmed) FDQML parameter estimation on the 30 min data. We ignore both the Fourier frequencies λ_j

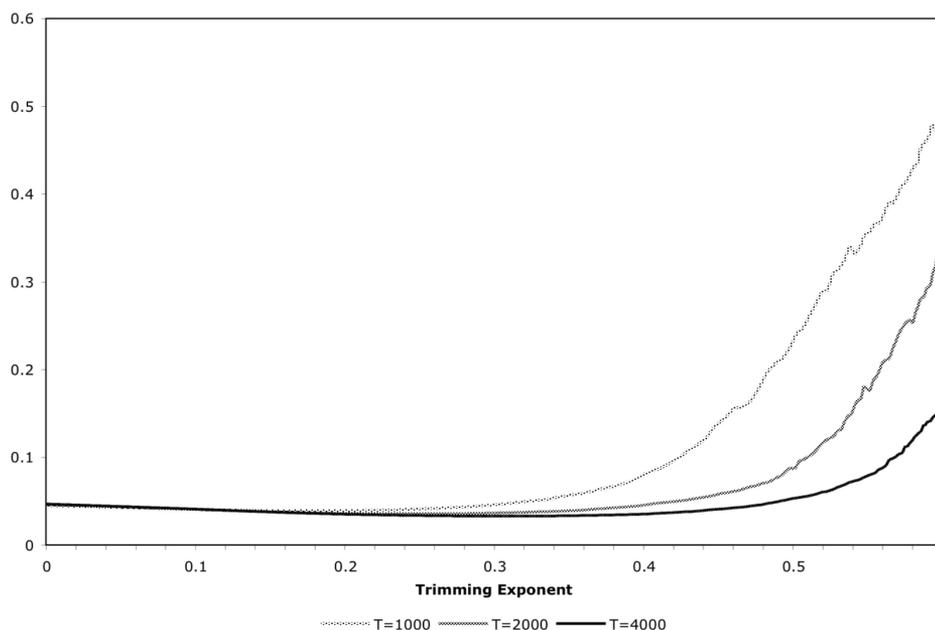


Figure 2. RMSE, $a_0 = 0.95$.

Table 4. Bias for GARCH(1,1) processes with(out) DLS

T	l =	Trimming: \hat{a}					Trimming: \hat{b}				
		1	$T^{0.40}$	$T^{0.51}$	$T^{0.60}$	QML	1	$T^{0.40}$	$T^{0.51}$	$T^{0.60}$	QML
$a_0 = 0.05, b_0 = 0.3$											
1000		-0.007	0.002	0.011	0.025	0.001	0.161	0.134	0.214	0.218	-0.019
4000		-0.006	-0.003	0.000	0.006	0.000	0.108	0.089	0.078	0.061	-0.011
16,000		-0.004	-0.004	-0.001	0.000	0.000	0.053	0.001	0.004	0.004	-0.001
$a_0 = 0.05, b_0 = 0.6$											
1000		-0.003	0.006	0.016	0.023	0.004	-0.025	-0.047	-0.001	0.010	-0.126
4000		0.000	0.002	0.004	0.012	0.002	-0.024	-0.040	-0.036	-0.042	-0.049
16,000		-0.001	0.000	-0.001	-0.002	0.000	-0.004	-0.009	-0.012	-0.014	-0.009
$a_0 = 0.05, b_0 = 0.9$											
1000		0.004	-0.005	-0.010	-0.005	0.002	-0.093	-0.207	-0.195	-0.264	-0.032
4000		0.000	0.000	-0.009	-0.015	0.000	-0.009	-0.023	-0.193	-0.196	-0.004
16,000		0.000	0.000	-0.002	-0.012	0.000	-0.004	-0.009	-0.038	-0.165	-0.002
$a_0 = 0.05, b_0 = 0.3, \text{DLS}$											
1000		-0.004	0.004	0.013	0.024	-0.004	0.599	0.165	0.238	0.227	0.613
4000		-0.015	-0.001	0.001	0.007	-0.035	0.629	0.105	0.069	0.067	0.682
16,000		-0.017	-0.006	0.000	0.002	-0.040	0.631	0.047	0.016	0.017	0.688
$a_0 = 0.05, b_0 = 0.6, \text{DLS}$											
1000		0.000	0.010	0.018	0.024	0.003	0.293	-0.054	-0.019	0.004	0.304
4000		-0.010	0.003	0.004	0.013	-0.033	0.321	-0.031	-0.044	-0.050	0.379
16,000		-0.012	0.000	-0.002	-0.003	-0.042	0.325	0.003	-0.003	-0.007	0.392
$a_0 = 0.05, b_0 = 0.9, \text{DLS}$											
1000		0.008	-0.003	-0.009	-0.007	0.010	-0.029	-0.211	-0.203	-0.274	0.003
4000		0.005	0.001	-0.007	-0.014	0.003	-0.005	-0.030	-0.179	-0.205	0.023
16,000		0.004	0.000	-0.001	-0.007	0.002	-0.004	-0.003	-0.010	-0.074	0.028

NOTE: In this table, a_0 and b_0 refer to the true GARCH persistence parameters in the (contaminated) GARCH(1,1) model described in Section 4.3. "DLS" indicates contamination by deterministic level shifts. The trimmed set of Fourier frequency indices is $(-T/2, T/2] \cap \mathbb{Z} \setminus [-l + 1, l - 1]$ for $l = 1, T^{0.40}, T^{0.51}, T^{0.6}$. No trimming occurs when $l = 1$. "QML" denotes the standard time domain QML estimator.

Table 5. RMSE for GARCH(1,1) processes with(out) DLS

T	l =	Trimming: \hat{a}					Trimming: \hat{b}				
		1	$T^{0.40}$	$T^{0.51}$	$T^{0.60}$	QML	1	$T^{0.40}$	$T^{0.51}$	$T^{0.60}$	QML
$a_0 = 0.05, b_0 = 0.3$											
1000		0.038	0.037	0.042	0.062	0.033	0.423	0.385	0.419	0.426	0.289
4000		0.026	0.023	0.021	0.024	0.018	0.352	0.320	0.305	0.296	0.237
16,000		0.016	0.017	0.010	0.010	0.009	0.232	0.187	0.157	0.165	0.145
$a_0 = 0.05, b_0 = 0.6$											
1000		0.036	0.034	0.043	0.062	0.032	0.333	0.320	0.322	0.341	0.315
4000		0.021	0.019	0.019	0.028	0.017	0.220	0.216	0.233	0.271	0.201
16,000		0.010	0.009	0.012	0.014	0.009	0.097	0.092	0.119	0.135	0.087
$a_0 = 0.05, b_0 = 0.9$											
1000		0.026	0.033	0.040	0.053	0.021	0.245	0.375	0.375	0.422	0.102
4000		0.012	0.012	0.022	0.033	0.010	0.057	0.081	0.351	0.361	0.038
16,000		0.008	0.008	0.010	0.023	0.007	0.033	0.037	0.109	0.311	0.032
$a_0 = 0.05, b_0 = 0.3, \text{DLS}$											
1000		0.025	0.036	0.042	0.062	0.041	0.609	0.393	0.431	0.433	0.632
4000		0.016	0.022	0.020	0.023	0.035	0.630	0.323	0.297	0.304	0.683
16,000		0.017	0.022	0.014	0.014	0.040	0.633	0.250	0.217	0.231	0.689
$a_0 = 0.05, b_0 = 0.6, \text{DLS}$											
1000		0.025	0.036	0.044	0.063	0.040	0.308	0.327	0.329	0.350	0.328
4000		0.013	0.017	0.018	0.027	0.034	0.322	0.209	0.225	0.267	0.380
16,000		0.013	0.009	0.013	0.016	0.042	0.326	0.094	0.120	0.147	0.393
$a_0 = 0.05, b_0 = 0.9, \text{DLS}$											
1000		0.023	0.031	0.041	0.053	0.025	0.130	0.380	0.381	0.433	0.071
4000		0.011	0.012	0.021	0.032	0.012	0.033	0.099	0.329	0.369	0.042
16,000		0.007	0.006	0.007	0.014	0.006	0.030	0.031	0.037	0.167	0.041

NOTE: In this table, a_0 and b_0 refer to the true GARCH persistence parameters in the (contaminated) GARCH(1,1) model described in Section 4.3. "DLS" indicates contamination by deterministic level shifts. The trimmed set of Fourier frequency indices is $(-T/2, T/2] \cap \mathbb{Z} \setminus [-l + 1, l - 1]$ for $l = 1, T^{0.40}, T^{0.51}, T^{0.6}$. No trimming occurs when $l = 1$. "QML" denotes the standard time domain QML estimator.

Table 6. ARSV estimates for high frequency FX returns

Trimming	\hat{a}	$\hat{\sigma}_{\xi}^2$	$\hat{\sigma}_{\zeta}^2$	Half-life	HL reduction
DHL deseasonalization					
$l = 1$	0.981	0.037	6.413	35.582	—
$l = \lfloor T^{0.4} \rfloor$	0.849	0.266	5.979	4.241	88.1%
$l = \lfloor T^{0.51} \rfloor$	0.785	0.372	5.870	2.865	91.9%
AB deseasonalization					
$l = 1$	0.982	0.030	5.458	38.346	—
$l = \lfloor T^{0.4} \rfloor$	0.849	0.227	5.082	4.231	89.0%
$l = \lfloor T^{0.51} \rfloor$	0.776	0.331	4.974	2.739	92.9%

associated with daily periodicity (integer multiples of $T/48$) and surrounding frequencies in a fixed neighborhood (the 30 Fourier frequencies to the left and right) so that the deseasonalization is robust to periodicity that changes slowly over time (see Section 4 of Deo, Hurvich, and Lu 2006 for details). (The choice of 30 was used by Deo, Hurvich, and Lu (2006) but our estimation results are insensitive to this choice.) As noted by Deo, Hurvich, and Lu (2006), excluding these frequencies from the objective function has no effect on the asymptotic properties of the parameter estimates while making them robust to the presence of intraday periodicity.

The FDQML parameter estimates of the model are reported in Table 6 for the two trimmings $\lfloor T^{0.4} \rfloor$ and $\lfloor T^{0.51} \rfloor$ and the two different deseasonalization procedures. We report the estimated half-life of a shock to the short-memory component of log-squared returns and the percentage reduction in the estimated half-life when moving from standard to robust estimation. The results are striking: robustly estimating the parameters via trimmed FDQML reduces the implied half-life of a shock by roughly 90% relative to standard estimation. This result is insensitive to the trimming or deseasonalization procedure.

5.2 GARCH(1,1) Estimation of Daily Stock Market Returns

Finally, we examine low-frequency robust estimation of the GARCH(1,1) model for two standard daily stock market returns time series: the S&P 500 and Dow Jones Industrial Average (DJIA). The S&P 500 series consists of daily returns from January 8, 1926, to March 25, 2004 (20,327 observations) and

Table 7. GARCH(1,1) estimates for daily stock market returns

Estimator	\hat{a}	\hat{b}	$\hat{\sigma}_{\xi}^2 \times 10^6$	Half-life	HL reduction
S&P 500					
TDQML	0.084	0.913	—	244.701	—
FDQML, $l = \lfloor T^{0.4} \rfloor$	0.132	0.735	0.290	4.841	98.0%
FDQML, $l = \lfloor T^{0.51} \rfloor$	0.130	0.720	0.289	4.264	98.3%
DJIA					
TDQML	0.088	0.905	—	91.200	—
FDQML, $l = \lfloor T^{0.4} \rfloor$	0.114	0.707	0.438	3.518	96.1%
FDQML, $l = \lfloor T^{0.51} \rfloor$	0.113	0.694	0.437	3.248	96.4%

the DJIA series consists of daily returns from May 27, 1896 to March 30, 2007 (27,766 observations). Table 7 reports the results for two trimmed FDQML estimators and their standard time domain counterparts. The reported half-lives of shocks are for shocks to the ARMA representation of squared returns. The results are again quite notable. The standard estimates of the model parameters produce the usual result of (nearly) integrated GARCH (IGARCH). However, the robust estimators clearly do not, reducing the implied half-lives of shocks by 96%–98%. These results are again insensitive to the trimming used.

6. CONCLUSION

This article addresses the well-documented issue of spurious persistence from an estimation standpoint. We introduce trimmed FDQML estimation methods that are robust to low-frequency contamination, yielding consistent and asymptotically normal estimators. In the potential presence of low-frequency contaminating components, we have shown that trimmed FDQML estimation outperforms existing estimation techniques, removing large biases, and decreasing RMSEs for a wide variety of models. The empirical applications of this article provide evidence that an estimation methodology that is robust to low-frequency contamination is practically important for fitting time series models to economic and financial data.

SUPPLEMENTARY MATERIALS

The supplemental appendix provides technical proofs for the main results of the article. These proofs make use of several supporting lemmas which are also stated and proved in the supplemental appendix.

ACKNOWLEDGMENTS

A previous version of this article circulated under the title “Parameter Estimation Robust to Low-Frequency Contamination with Applications to ARMA, GARCH and Stochastic Volatility Models.” The authors are very grateful to Pierre Perron and Zhongjun Qu for numerous helpful discussions and to Nickolai Riabov for research assistance. This article also benefited from the comments of Ivan Fernandez-Val, Hiroaki Kaido, Rasmus Varneskov, the editors, and two anonymous referees. Thanks are due to Denis Tkachenko for kindly sharing cleaned high-frequency foreign exchange data and deseasonalization code.

[Received October 2014. Revised July 2015.]

REFERENCES

- Andersen, T. G., and Bollerslev, T. (1997), “Intraday Periodicity and Volatility Persistence in Financial Markets,” *Journal of Empirical Finance*, 4, 115–158. [607]
- Bai, J., and Perron, P. (1998), “Estimating and Testing Linear Models With Multiple Structural Changes,” *Econometrica*, 66, 47–78. [599]
- Baxter, M., and King, R. (1999), “Measuring Business Cycles: Approximating Band-pass Filters for Economic Time Series,” *The Review of Economics and Statistics*, 81, 575–593. [599]
- Bhattacharya, R., Gupta, V., and Waymire, E. (1983), “The Hurst Effect Under Trends,” *Journal of Applied Probability*, 20, 649–662. [598]
- Breidt, F. J., Crato, N., and de Lima, P. (1998), “The Detection and Estimation of Long Memory in Stochastic Volatility,” *Journal of Econometrics*, 83, 325–348. [601]
- Chiu, S. (1988), “Weighted Least Squares Estimators on the Frequency Domain for the Parameters of a Time Series,” *The Annals of Statistics*, 16, 1315–1326. [603]

- Deo, R., Hurvich, C., and Lu, Y. (2006), "Forecasting Realized Volatility Using a Long-Memory Stochastic Volatility Model: Estimation, Prediction and Seasonal Adjustment," *Journal of Econometrics*, 131, 29–58. [606,607]
- Diebold, F., and Inoue, A. (2001), "Long Memory and Regime Switching," *Journal of Econometrics*, 105, 131–159. [598]
- Dunsmuir, W. (1979), "A Central Limit Theorem for Parameter Estimation in Stationary Vector Time Series and its Application to Models for a Signal Observed With Noise," *The Annals of Statistics*, 7, 490–506. [599,601]
- Dunsmuir, W., and Hannan, E. J. (1976), "Vector Linear Time Series Models," *Advances in Applied Probability*, 8, 339–364. [599,600]
- Engle, R. F. (1974), "Band Spectrum Regression," *International Economic Review*, 15, 1–11. [599]
- Eraker, B., Johannes, M., and Polson, N. (2003), "The Impact of Jumps in Volatility and Returns," *The Journal of Finance*, 58, 1269–1300. [598]
- Garcia, R., and Perron, P. (1996), "An Analysis of the Real Interest Rate Under Regime Shifts," *The Review of Economics and Statistics*, 78, 111–125. [598]
- Giraitis, L., and Robinson, P. M. (2001), "Whittle Estimation of ARCH Models," *Econometric Theory*, 17, 608–631. [601,603]
- Granger, C. W. J., and Hyung, N. (2004), "Occasional Structural Breaks and Long Memory With an Application to the S&P 500 Absolute Stock Returns," *Journal of Empirical Finance*, 11, 399–421. [598]
- Hannan, E. J. (1973), "The Asymptotic Theory of Linear Time-Series Models," *Journal of Applied Probability*, 10, 130–145. [601]
- Harvey, A. C. (1998), "Long-Memory in Stochastic Volatility," in *Forecasting Volatility in Financial Markets*, eds. J. Knight and S. E. Satchell, London: Butterworth-Heinemann, pp. 307–320. [601]
- Hillebrand, E. (2005), "Neglecting Parameter Changes in GARCH Models," *Journal of Econometrics*, 129, 121–138. [598,606]
- Hosoya, Y. (1997), "A Limit Theory for Long-Range Dependence and Statistical Inference on Related Models," *Annals of Statistics*, 25, 105–136. [601]
- Hosoya, Y., and Taniguchi, M. (1982), "A Central Limit Theorem for Stationary Processes and the Parameter Estimation of Linear Processes," *The Annals of Statistics*, 10, 132–153. [599]
- Hou, J., and Perron, P. (2014), "Modified Local Whittle Estimator for Long Memory Processes in The Presence of Low Frequency (and other) Contaminations," *Journal of Econometrics*, 182, 309–328. [599,601]
- Iacone, F. (2010), "Local Whittle Estimation of the Memory Parameter in Presence of Deterministic Components," *Journal of Time Series Analysis*, 31, 37–49. [599]
- Lamoureux, C. G., and Lastrapes, W. D. (1990), "Persistence in Variance, Structural Change, and the GARCH Model," *Journal of Business and Economic Statistics*, 8, 225–234. [598,606]
- McCloskey, A. (2013), "Estimation of the Long-Memory Stochastic Volatility Model Parameters that is Robust to Level Shifts and Deterministic Trends," *Journal of Time Series Analysis*, 34, 285–301. [599,600,601]
- McCloskey, A., and Perron, P. (2013), "Memory Parameter Estimation in the Presence of Level Shifts and Deterministic Trends," *Econometric Theory*, 29, 1196–1237. [598,599,600,601,605]
- Mikosch, T., Gadrich, T., Kluppelberg, C., and Adler, R. (1995), "Parameter Estimation for ARMA Models With Infinite Variance Innovations," *Annals of Statistics*, 23, 305–326. [604]
- Mikosch, T., and Stărică, C. (2004), "Nonstationarities in Financial Time Series, the Long-Range Dependence, and the IGARCH Effects," *Review of Economics and Statistics*, 86, 378–390. [598]
- Mikosch, T., and Straumann, D. (2002), "Whittle Estimation in a Heavy-Tailed GARCH(1, 1) Model," *Stochastic Processes and Their Applications*, 100, 187–222. [601,603]
- Müller, U. K., and Watson, M. W. (2008), "Testing Models of Low-Frequency Variability," *Econometrica*, 76, 979–1016. [599]
- Pagano, M. (1974), "Estimation of Models of Auto-Regressive Signal Plus White Noise," *The Annals of Statistics*, 2, 99–108. [602]
- Perron, P. (1990), "Testing for a Unit Root in a Time Series With a Changing Mean," *Journal of Business and Economic Statistics*, 8, 153–162. [598]
- Perron, P., and Qu, Z. (2010), "Long-Memory and Level Shifts in the Volatility of Stock Market Return Indices," *Journal of Business and Economic Statistics*, 28, 275–290. [598]
- Qu, Z. (2011), "A Test Against Spurious Long Memory," *Journal of Business and Economic Statistics*, 29, 423–438. [598]
- Qu, Z., and Perron, P. (2013), "A Stochastic Volatility Model With Random Level Shifts: Theory and Applications to S&P 500 and NASDAQ Returns," *Econometrics Journal*, 16, 309–339. [598,602]
- Stărică, C., and Granger, C. (2005), "Nonstationarities in Stock Returns," *Review of Economics and Statistics*, 87, 503–522. [598]
- Taniguchi, M. (1982), "On Estimation of the Integrals of the Fourth Cumulant Spectral Density," *Biometrika*, 69, 117–122. [603]