Estimation of the Long-Memory Stochastic Volatility Model Parameters that is Robust to Level Shifts and Deterministic Trends[*]

Adam McCloskey[†]

Brown University

December 2010; This Version: October 2012

Abstract

I provide conditions under which the trimmed FDQML estimator, advanced by McCloskey (2010) in the context of fully parametric short-memory models, can be used to estimate the long-memory stochastic volatility model parameters in the presence of additive low-frequency contamination in log-squared returns. The types of low-frequency contamination covered include level shifts as well as deterministic trends. I establish consistency and asymptotic normality in the presence or absence of such low-frequency contamination under certain conditions on the growth rate of the trimming parameter. I also provide theoretical guidance on the choice of trimming parameter by heuristically obtaining its asymptotic MSE-optimal rate under certain types of low-frequency contamination. A simulation study examines the finite sample properties of the robust estimator, showing substantial gains from its use in the presence of level shifts. The finite sample analysis also explores how different levels of trimming affect the parameter estimates in the presence and absence of low-frequency contamination and long-memory.

JEL Classification Numbers: C58, C22, C13, C18

Keywords: stochastic volatility, frequency domain estimation, robust estimation, spurious persistence, long-memory, level shifts, structural change, deterministic trends

1 Introduction

The empirical features of volatility measures, such as slowly decaying correlograms and large spectral density estimates near the zero frequency, have spurred considerable interest in modeling volatility as a long-memory process over the past twenty years. Of the fully parametric methods for long-memory modeling, the long-memory stochastic volatility (LMSV) model of Breidt et al. (1998) and Harvey (1998) has proven quite popular. In large part this is because the LMSV model can simultaneously account for the empirical presence of noise while allowing for both long and short memory dynamics in log-squared returns, one of the foremost measures of volatility. The LMSV model has also proven useful for forecasting volatility, see Deo et al. (2006) for example. Nonetheless, recent work has shown that the appearance of long-memory in some semiparametric memory parameter estimates may be spuriously caused, or at least exaggerated, by level shifts or deterministic trends in the mean of the log-squared returns series. See, e.g., Diebold and Inoue (2001), Granger and Hyung (2004), Mikosch and Stărică (2004), Perron and Qu (2010) and McCloskey and Perron (2010). As we will see, the standard fully parametric technique for estimating the parameters of LMSV models also suffers this upward biased in the memory parameter estimate in the presence of "additive low-frequency contamination", including level shifts and deterministic trends. The potential presence of low-frequency contamination in volatility measures is not merely a theoretical curiosity. Among others, Granger and Hyung (2004), Mikosch and Stărică (2004), Perron and Qu (2010) and McCloskey and Perron (2010) have also provided evidence that this is in fact a very real and salient feature of volatility data. This study aims to fill the gap in current estimation methodology by providing a robust estimator of the LMSV model parameters in the presence or absence of low-frequency contamination.

There are two basic methodologies for estimating the memory parameter: fully and semiparametric. Breidt et al. (1998) and Harvey (1998) were the first to propose fully parametric estimates of the LMSV model parameters. Many refinements of standard semiparametric memory parameter estimates (e.g., the log-periodogram regression estimator of Geweke and Porter-Hudak, 1983, and the local Whittle estimator of Künsch, 1987) have also been advanced to reduce the bias that may arise from the short-memory dynamics and noise present in the log-squared returns of a LMSV process; see Deo and Hurvich (2001), Sun and Phillips (2003), Andrews and Sun (2004), Hurvich et al. (2005) and Frederiksen and Nielsen (2008), among others. However, these estimators are not robust to low-frequency contamination. Focusing on a separate problem, McCloskey and Perron (2010) and Iacone (2010) have also

proposed semiparametric estimation techniques for estimating the memory parameter of a process that is subject to low-frequency contamination. However, their estimators are sensitive to the presence of short-memory dynamics and noise, which can be quite substantial in the data. An estimator that simultaneously accounts for noise, short-memory dynamics and long-memory dynamics while remaining robust to low-frequency contamination is therefore desirable.

To the author's knowledge, the trimmed frequency domain quasi-maximum likelihood (FDQML) estimator of the LMSV model parameters, studied in this article, is the first to do just that. However, as is always the case in fully parametric estimation, the price one must pay for these desirable features is that the LMSV model must be correctly specified. Nevertheless, a distinct advantage of the fully parametric approach provided here is that it makes forecasting feasible. The trimmed FDQML methodology also has the appeal of allowing the practitioner to remain agnostic about the form that the low-frequency contamination may take (e.g., level shifts, smooth trends or fractional trends) or whether contamination is indeed present at all. For an overview of (trimmed) FDQML estimation, see McCloskey (2010) and the references therein.

Breidt et al. (1998) and Harvey (1998) originally proposed estimation of the LMSV model parameters in the frequency domain but did not explore the implications of trimming and/or low-frequency contamination. McCloskey (2010) advanced the trimmed FDQML methodology in the context of low-frequency contaminated short-memory models, ruling out the specification of this article. This article establishes the consistency and asymptotic normality of the trimmed FDQML estimator in the long-memory context of the LMSV model in the potential presence of low-frequency contamination and provides theoretical guidance on the choice of trimming by heuristically obtaining the asymptotic mean-squared error (MSE)-minimizing rate for the trimming parameter to grow with the sample size. A Monte Carlo study confirms the potential usefulness of this "optimal" rate for the trimming parameter and assesses the tradeoffs involved in the use of various trimmings in the presence and absence of low-frequency contamination.

The remainder of this article is composed as follows. Section 2 details the (potentially) contaminated LMSV model. Section 3 introduces the robust estimation strategy and discusses the assumptions imposed for consistency and asymptotic normality of the trimmed FDQML estimator. Section 4 develops theoretical guidance for choosing the trimming parameter based on minimizing the asymptotic MSE of the estimator. Section 5 provides a simulation study of the finite sample properties of the estimator for calibrated data generat-

ing processes, discussing the choice of trimming parameter in finite samples. All proofs are contained in the mathematical appendix. In what follows, $\mathbb{R}$ and $\mathbb{Z}$ denote the sets of real numbers and integers; $\mathbb{I}(\cdot)$ denotes the indicator function; $\lfloor K \rfloor$ 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; "$\xrightarrow{d}$", "$\xrightarrow{p}$" and "$\xrightarrow{a.s.}$" indicate weak convergence, convergence in probability and almost sure convergence. For two function of T, $A(T)$ and $B(T)$, $A(T) \simeq B(T)$ means that $A(T)/B(T)$ converges to a constant as $T \to \infty$. For a generic vector or matrix M, $M = O_p(A(T))/O(A(T))/o_p(A(T))/o(A(T))$ indicates that each element of M has the corresponding order of magnitude. All convergence concepts are taken to mean as the sample size grows to infinity.

2 The Contaminated LMSV Model

We begin with some preliminary terminology and quantities of interest. For a generic process $\{y_t\}$, the discrete Fourier transform and periodogram of $\{y_t\}$ 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 and $\lambda \in [-\pi, \pi]$ denotes the frequency. The estimators examined here will involve evaluating the observed time series' periodogram at a subset of the Fourier frequencies $\lambda_j \equiv 2\pi j/T$ for $j = -\lfloor T/2 \rfloor + 1, \ldots, \lfloor T/2 \rfloor - 1, \lfloor T/2 \rfloor$. The periodogram is the natural method of moments estimator of the population spectral density function of a process.

Let us turn to a description of the contaminated LMSV model which extends the standard LMSV model to allow for abrupt or smoothly varying changes in the mean of the volatility process, among other forms of low-frequency contamination.

Assumption 1. *The data generating process (DGP) follows*

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

where $\sigma > 0$ is some constant, $e_t \sim i.i.d.(0,1)$,

$$h_t = u_t + v_t, \quad (1-L)^d A(L) v_t = B(L)\eta_t, \quad \eta_t \sim i.i.d.N(0, \sigma_\eta^2),$$

$\{u_t\}$, $\{e_t\}$ and $\{v_t\}$ are mutually independent, $E[I_u(\lambda_j)] = O(T/j^2)$ uniformly for all nonzero Fourier frequencies λ_j and $A(L)$ and $B(L)$ are polynomials in the lag operator L of finite

3

orders m and q with coefficients $\{a_i\}_{i=1}^m$ and $\{b_j\}_{j=1}^q$: $A(L) = 1 - \sum_{i=1}^m a_i L^i$ and $B(L) = 1 + \sum_{j=1}^q b_j L^j$.

Remark 1. *The proof of consistency of the trimmed FDQML estimator relies on the ergodicity of $\{v_t\}$. When $\{\eta_t\}$ is Gaussian, the ergodicity of $\{v_t\}$ is implied by the other conditions of Assumption 1 (see the proof of Theorem 1 in Breidt et al., 1998). One may instead explicitly impose ergodicity of $\{v_t\}$ to avoid the Gaussianity assumption on $\{\eta_t\}$. However, in the case that $\{\eta_t\}$ is not Gaussian, the asymptotic variance in Theorem 2 needs to be modified to include a component arising from the excess kurtosis of η_t. See Remark 3 for details.*

The process $\{r_t\}$ is specified as the returns series of some underlying asset, typically measured as the first difference of the logarithm of the asset price. Apart from the component u_t, this is the original LMSV model proposed by Breidt et al. (1998) and Harvey (1998). The high-level assumption on $I_u(\cdot)$ is identical to that imposed by McCloskey (2010). This assumption is more appealing than explicit specification of the process $\{u_t\}$ when the practitioner would like to take an agnostic view on the form of contamination that may or may not be present, as should be the case in many practical applications. It allows the process $\{u_t\}$ to take a variety of forms, some of which are provided below.

Random Level Shifts (RLS)

$$u_t = \sum_{j=1}^t \delta_{T,j}, \qquad \delta_{T,t} = \pi_{T,t}\xi_t,$$

where $\xi_t \sim i.i.d.\left(0, \sigma_\xi^2\right)$ and $\pi_{T,t} \sim i.i.d.Bernoulli(p/T, 1)$ for some $p \geq 0$. The components $\pi_{T,t}$ and ξ_t are mutually independent.

Deterministic Level Shifts (DLS)

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

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

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 $h(1) = 0$.[1]

---

[1] This includes all cases for which $h(\cdot)$ is monotonic and bounded since we can simply subtract $h(1)$ from $h(\cdot)$ and multiply σ by $\exp(h(1)/2)$ to have the same DGP.

Fractional Trends (FT)

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

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

Outliers

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

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 specified to be sample size-dependent so that level shifts are rare, with the expected number being constant in the sample size. If this were not the case, $\{u_t\}$ would be better construed as a random walk. It is important to note that p can equal zero so that the RLS process nests the no level shift, no trend case. Perron and Qu (2010) showed that RLS satisfy the periodogram assumption $E[I_u(\lambda_j)] = O(T/j^2)$ uniformly for all nonzero λ_j, while McCloskey and Perron (2010) and Mikosch and Stărică (2004) showed the same for DLS; Qu (2011) and Künsch (1986) did so for DT and Iacone (2010) did for FT. A slight generalization of results in Iacone (2010) provides the result for outliers. Although not listed above, Robinson (1997) showed that a class of nonparametric mean functions also satisfies the high-level assumption.

In relation to Iacone (2010), here we are imposing the weakest version of his Assumption 2, that with $\phi = 1/2$, in order to accommodate the widest range of contaminating processes. Although $E[I_u(\lambda_j)] \simeq T/j^2$ in the RLS case for which $p \geq 1$ and the DLS case for which $c_i \neq 0$ for some i (see the above references), $\lim_{T\to\infty}(j^2/T)E[I_u(\lambda_j)] = 0$ for some DT cases, any FT cases or when $j/T \to 0$ for outliers. However, assuming an exact order for some of these cases would rule out contamination by some of the others. For example, Theorem 1 of Iacone (2010) provides that $E[I_u(\lambda_j)] = O(T^{2\widetilde{\phi}}/j^{1+2\widetilde{\phi}})$ uniformly in nonzero λ_j for all FT cases with $\phi \leq \widetilde{\phi}$. However, if we were to replace the assumption $E[I_u(\lambda_j)] = O(T/j^2)$ with $E[I_u(\lambda_j)] = O(T^{2\widetilde{\phi}}/j^{1+2\widetilde{\phi}})$ for some $\widetilde{\phi} \in (-1/2, 1/2)$, we would be ruling out contamination by RLS or DLS.

It is also worth discussing a couple of the processes that the assumption on $E[I_u(\cdot)]$ rules out. Assumption 1 does not allow $\{u_t\}$ to be a long-memory process since, for a long-memory parameter $d < 1$, the expectation of the periodogram of a long-memory process evaluated at nonzero frequency λ_j is of order $O(T^{2d}/j^{2d})$ when $j/T \to 0$ (see Theorem 1 of Velasco, 1999). Hence, such expectations do not converge to zero when evaluated at nonzero Fourier

frequencies for which $\sqrt{T}/j \to 0$, as required by Assumption 1. Similarly, in the unit root case, for which $d = 1$, $E[I_u(\lambda_j)] = O(T^2/j^2)$ for all nonzero λ_j (see Theorem 1 of Hurvich and Ray, 1995). On the other hand, most short-memory processes, such as white noise or ARMA processes, have bounded but strictly positive periodograms at all frequencies and similarly violate Assumption 1 as candidates for $\{u_t\}$.

3   The Trimmed FDQML Estimator

Under Assumption 1, we have the following decomposition of the log-squared returns process:

$$x_t \equiv \log(r_t^2) = \log \sigma^2 + E[\log(e_t^2)] + u_t + v_t + \left(\log(e_t^2) - E[\log(e_t^2)]\right) \equiv \mu + u_t + v_t + \varepsilon_t,$$

where $\mu \equiv \log \sigma^2 + E[\log(e_t^2)]$ and $\varepsilon_t \equiv (\log(e_t^2) - E[\log(e_t^2)])$. Assuming stationarity, let $f(\lambda; \theta)$ denote the spectral density function of $\{v_t + \varepsilon_t\}$ evaluated at frequency $\lambda \in [-\pi, \pi]$, where θ denotes a finite-dimensional parameter vector. Then,

$$f(\lambda; \theta) = \frac{\sigma_\eta^2 |B(e^{-i\lambda})|^2}{2\pi |1 - e^{-i\lambda}|^{2d} |A(e^{-i\lambda})|^2} + \frac{\sigma_\varepsilon^2}{2\pi}, \tag{1}$$

where

$$\theta \equiv (d, \sigma_\eta^2, \sigma_\varepsilon^2, a_1, \ldots, a_m, b_1, \ldots, b_q)$$

is the parameter vector of interest with $\sigma_\varepsilon^2 \equiv \mathrm{Var}(\varepsilon_t)$. The following assumption imposes specific conditions on the parameters that ensure the existence of the spectral density function. It also employs conditions that guarantee the identification of θ in the (second-order) spectral domain.

Assumption 2. *The parameter vector θ is an element of a compact parameter space Θ and for $\theta^A, \theta^B \in \Theta$, $f(\cdot; \theta^A) = f(\cdot; \theta^B)$ implies $\theta^A = \theta^B$. The parameter space Θ is contained in a space such that $\sigma_\varepsilon^2 \in (0, \infty)$, $A(x)B(x) \neq 0$ for $|x| \leq 1$ and $A(\cdot)$ and $B(\cdot)$ have no common roots. The true parameter vector, θ^0 lies in Θ with $d^0 \in [0, 1/2)$.*

Assumption 2 is quite standard for estimation in the spectral domain. It rules out unidentifiable cases that would occur if $d = 0$ and $A(z), B(z) = 1$ or $\sigma_\eta^2 = 0$, for example. It also ensures that the spectral density function is positive everywhere. When e_t is Gaussian, it is known that $(\sigma_\varepsilon^0)^2 = \pi^2/2$ and this may be imposed in the estimation if one is willing to make this additional assumption.

We will perform estimation of the LMSV model parameters via the periodogram of the log-squared returns process $\{x_t\}$. More specifically, the trimmed FDQML objective function is defined as

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

where $\mathcal{F}_l \equiv (-T/2, T/2] \cap \mathbb{Z} \setminus [-l+1, l-1]$, and the trimmed FDQML estimator is defined as $\hat{\theta}_T \equiv \operatorname{argmin}_{\theta \in \Theta} L_{T,l}(\theta)$ for a parameter space Θ satisfying Assumption 2. The objective function selects model parameters to fit a candidate spectral density function to the observed periodogram. Let $\widetilde{v}_t \equiv v_t + \varepsilon_t$. Then the periodogram of $\{x_t\}$ has the following decomposition:

$$I_x(\lambda) = I_{\widetilde{v}}(\lambda) + I_u(\lambda) + w_{\widetilde{v}}(\lambda)w_u(-\lambda) + w_{\widetilde{v}}(-\lambda)w_u(\lambda). \tag{2}$$

The standard FDQML objective function sets $l = 1$ but here we wish to asymptotically rid the objective function's dependence on the contaminating process $\{u_t\}$ by using a large enough trimming so that the influence of $I_u(\lambda_j)$, $w_{\widetilde{v}}(\lambda)w_u(-\lambda)$ and $w_{\widetilde{v}}(-\lambda)w_u(\lambda)$ in (2) are asymptotically negligible. The next assumption makes this necessary trimming precise.

Assumption 3.
$$\frac{\log T}{l} + \frac{l \log T}{T} \to 0 \ \text{as} \ T \to \infty.$$

The imposed trimming is quite weak and is basically the same as that imposed for consistency by McCloskey (2010). This is not surprising in light of the fact that we are dealing with the same objective function but with a process $\{\widetilde{v}_t\}$ that asymptotically dominates the periodogram $I_x(\cdot)$ for at least as many Fourier frequencies as the processes he considered (and more if $d^0 > 0$).

Remark 2. *Though he addresses semiparametric estimation, Iacone (2010) uses a similar objective function to the trimmed FDQML objective function to estimate the memory parameter. In fact the trimmed local Whittle objective function he uses approximates the trimmed FDQML objective function in a frequency band local to zero. Hence, it is interesting to compare the trimming he requires to establish consistency in the semiparametric context with that imposed by Assumption 3. The comparison cannot be made directly since the bandwidth parameter in Iacone (2010) must grow slower than the sample size. Nevertheless, as the the bandwidth m approaches its admissible boundary of T in the semiparametric context, the trimming Iacone (2010) requires when allowing for the broadest range of low-frequency contamination (i.e., $\phi = 1/2$ in his Assumption 3) approximately approaches that which is required by Assumption 3 above. That is, for any $d^0 \in [0, 1/2)$, as m approaches T, the permissible range for the growth rate of the trimming is the rate T^α for any $\alpha \in (0, 1)$.*

With these assumptions in hand, we may establish consistency.

Theorem 1. *Under Assumptions 1-3, $\hat{\theta} \overset{p}{\longrightarrow} \theta^0$.*

Asymptotic normality requires a stronger trimming condition in order to asymptotically rid the influence of the low-frequency contamination on the scaled bias and variance of the estimator.

Assumption 3*.
$$\frac{T^{1/2}\log T}{l} + \frac{l^5 \log^2 T}{T^4} \to 0 \ as \ T \to \infty.$$

The first condition of Assumption 3*, which enforces a lower bound on the growth rate of the trimming, is also imposed in the short-memory context of McCloskey (2010). As was the case for the trimming required for consistency, this is again unsurprising. The second condition, which is *not* imposed by McCloskey (2010), bounds the trimming's growth rate from above. This upper bound is used in the long-memory context of the present paper to ensure that the low-frequency ordinates containing relevant information on the long-memory behavior the of the contaminated process are not entirely trimmed away.

Remark 3. *Unlike the trimming assumption used for consistency (Assumption 3), it does not make sense to compare Assumption 3* to the trimming required by Iacone (2010) for asymptotic normality of his semiparametric estimator of d^0. For asymptotic normality, Iacone (2010) would require the bandwidth to grow slower than $T^{4d^0/(1+4d^0)} = o(T^{1/3})$ in the context of the LMSV model. Hence, we cannot examine the trimming his results would require for m approaching T.*

With this stronger trimming condition in hand, we may now present the asymptotic normality result.

Theorem 2. *Under Assumptions 1, 2 and 3*, with the additions that θ^0 is in the interior of Θ, $E[\varepsilon_t^4] < \infty$ and for all $\theta \in \Theta$, $d = \theta_1 \geq \underline{d}$ for some $\underline{d} > 0$,*
$$\sqrt{T}(\hat{\theta} - \theta^0) \overset{d}{\longrightarrow} \mathcal{N}(0, \Omega^{-1}(2\Omega + \Pi)\Omega^{-1}),$$

where
$$\Omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\partial \log f(\lambda; \theta^0)}{\partial \theta} \frac{\partial \log f(\lambda; \theta^0)}{\partial \theta'} d\lambda$$

and
$$\Pi = \frac{E[\varepsilon_t^4] - 3E[\varepsilon_t^2]}{(2\pi)^4} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \frac{\partial f(\lambda_1; \theta^0)^{-1}}{\partial \theta} \frac{\partial f(\lambda_2; \theta^0)^{-1}}{\partial \theta'} d\lambda_1 d\lambda_2.$$

8

Remark 4. *In the case that the Gaussianity assumption on $\{\eta_t\}$ is replaced by an ergodicity condition on $\{v_t\}$ (as discussed in Remark 1), the quantity Π in the above asymptotic variance expression would need to be modified to equal*

$$\Pi = \frac{(E[\eta_t^4] - 3E[\eta_t^2]) + (E[\varepsilon_t^4] - 3E[\varepsilon_t^2])}{(2\pi)^4} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \frac{\partial f(\lambda_1; \theta^0)^{-1}}{\partial \theta} \frac{\partial f(\lambda_2; \theta^0)^{-1}}{\partial \theta'} d\lambda_1 d\lambda_2,$$

assuming that $E[\eta_t^4] < \infty$ (see the proof of Theorem 2 and Hosoya, 1997).

Remark 5. *Finite sample evidence indicates that a normal approximation would lead to poor inference on the model parameters, even in the absence of contamination (see Pérez and Ruiz, 2001). Moreover, the limiting variance of the trimmed FDQML estimator appears to be exceedingly difficult to estimate. It stands to reason that these two problems would be further compounded in the presence of low-frequency contamination. Thus, I suggest using a bootstrap procedure that is valid for stationary long-range dependent observations for inference on the model parameters. A careful implementation of the block or sieve bootstrap for long-memory processes (e.g., Kapetanios and Papailias, 2011 or Poskitt, 2007) would likely suffice. However, establishing the validity of such a procedure for the DGPs examined here is beyond the scope of this paper.*

Remark 6. *We must impose the condition $d \geq \underline{d} > 0$ for all $\theta \in \Theta$, and consequently $d^0 > \underline{d} > 0$ by Assumption 2, due to the discontinuous behavior of $f(\lambda; \theta)$ at $d = \theta_1 = 0$. More specifically, the continuity and integrability properties of derivatives such as $\partial f(\lambda; \theta)^{-1}/\partial \theta$ and $\partial^2 [f(\lambda; \theta)^{-1}]/\partial \theta \partial \theta'$ crucial to establishing asymptotic normality do not hold when $d = 0$. A review of the literature on fully parametric estimation of long-memory models shows that this type of assumption is crucial to obtaining the asymptotic properties of estimators in various long-memory contexts. See, e.g., Fox and Taqqu (1986), Dahlhaus (1989) and Hosoya (1997).*

4 Asymptotically Optimal Trimming Rate

As with all methods that require a user-chosen tuning parameter, a theoretically-driven means for choosing the trimming parameter is desirable in the trimmed FDQML estimation context. In this section, I provide a heuristic argument for the asymptotic MSE-minimizing rate for the trimming parameter.

To begin the analysis, assume that Assumptions 1-3 hold, θ^0 is in the interior of Θ and

for all $\theta \in \Theta$, $d = \theta_1 \geq \underline{d}$ for some $\underline{d} > 0$. Note that

$$\hat{\theta} - \theta^0 = -\left[\frac{\partial^2 L_{T,l}(\widetilde{\theta})}{\partial\theta\partial\theta'}\right]^{-1}\frac{\partial L_{T,l}(\theta^0)}{\partial\theta},$$

for some $\widetilde{\theta}$ such that $\|\widetilde{\theta} - \theta^0\| \leq \|\hat{\theta} - \theta^0\|$.[2] If $l\log^2 T/T \to 0$, then

$$\frac{\partial^2 L_{T,l}(\widetilde{\theta})}{\partial\theta\partial\theta'} \xrightarrow{p} \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\partial\log f(\lambda;\theta^0)}{\partial\theta}\frac{\partial\log f(\lambda;\theta^0)}{\partial\theta'},$$

an invertible deterministic matrix (see the arguments leading up to (A.15) in the proof of Theorem 2). Hence, assuming that they exist, the order of magnitude of the bias and variance of $\hat{\theta}$ are equal to those of $\partial L_{T,l}(\theta^0)/\partial\theta$. Suppose in addition that $l^5\log^2 T/T^4 \to 0$ and $E[\varepsilon_t^4] < \infty$. Using the decomposition (A.16)-(A.19) in the proof of Theorem 2, we can see

$$\text{Bias}(\hat{\theta}) = O\left(\frac{1}{T^{1/2}} + \frac{l^{1/2}\log T}{T} + \frac{\log T}{l}\right)$$

by results derived in the proof of Theorem 2 and the fact that $E[w_{\widetilde{v}}(\lambda_j)w_u(-\lambda_j)] = 0$. Similarly, results in the proof of Theorem 2 establish that the variance of terms (A.16) and (A.19) are $O(1)$ and $O(T^{2d^0}l^{-1-2d^0}\log^3 T)$. Assuming it exists, the variance of term (A.17) is $O(l\log^2 T/T)$ (see Lemma 3) while Lemma A.3 of McCloskey and Perron (2010) provides that the variance of (A.18) is $O(T\log^2 T/l^2)$, at least when the low-frequency contamination takes the form of RLS, DLS, DT or FT. Finally, the Cauchy-Schwartz inequality shows that covariances of pairs of the (A.16)-(A.19) terms are all of lower order than the variance terms if $l^{1-2d^0}\log T/T^{1-2d^0} \to 0$. Hence, under this additional condition, we have

$$\text{Var}(\hat{\theta}) = O\left(\frac{1}{T} + \frac{l\log^2 T}{T^2} + \frac{\log^2 T}{l^2} + \frac{\log^3 T}{T^{1-2d^0}l^{1+2d^0}}\right)$$

and

$$\text{MSE}(\hat{\theta}) = O\left(\frac{1}{T} + \frac{l^{1/2}\log T}{T^{3/2}} + \frac{\log T}{T^{1/2}l} + \frac{l\log^2 T}{T^2} + \frac{\log^2 T}{Tl^{1/2}} + \frac{\log^2 T}{l^2} + \frac{\log^3 T}{T^{1-2d^0}l^{1+2d^0}}\right).$$

Choosing l to minimize the order of this expression while satisfying the constraint $l^5\log^2 T/T^4 \to 0$ yields

$$l^{\text{OPT}} \simeq T^{1/2}\log T.$$

---

[2]Strictly speaking, this is a slight abuse of notation since $\widetilde{\theta}$ differs on elements of the vector equation as the Mean Value Theorem only applies to individual elements of partial derivatives.

This trimming rate satisfies the condition $l^{1-2d^0} \log T / T^{1-2d^0} \to 0$ so that the variance expression given above holds for all trimming rates within a neighborhood of l^{OPT}.

It is interesting to note the relationship between l^{OPT} and the MSE-minimizing rate for the bandwidth of semiparametric estimators of the memory parameter in the absence of low-frequency contamination. For standard semiparametric estimators (e.g., log-periodogram and local Whittle), the optimal bandwidth rate is equal to the boundary that determines whether asymptotic normality can be achieved, namely $T^{4/5}$ (see, e.g., Hurvich et al., 1998 and Henry and Robinson, 1996). Quite similarly, in the fully parametric context of this paper, the optimal *trimming* rate is equal to the boundary at which asymptotic normality can be achieved (see Assumption 3*). In the standard semiparametric estimation context (with no low-frequency contamination), a higher bandwidth allows higher frequency components of the time series to contaminate the estimate. Conversely, in the context of this paper, a lower trimming allows lower frequency components to contaminate the estimate.

Remark 7. *Though we can obtain the optimal asymptotic order of the trimming parameter, we cannot obtain the proportionality constant in the expression of l^{OPT} without imposing more explicit assumptions on the form of the contaminating process. For example, if we knew that $\{u_t\}$ was indeed described by the RLS model, the proportionality constant would depend on the parameters p and σ_ξ^2 of this model. We could then "plug in" estimates of these parameters to estimate l^{OPT}. However, such a methodology would seemingly defeat one of the main draws of the trimmed FDQML methodology: allowing the practitioner to be agnostic regarding the particular form the contaminating component may take.*

Remark 8. *In cases for which $E[I_u(\lambda_j)] = o(T/j^2)$, the true asymptotically optimal trimming grows slower than the order given by l^{OPT}. If one wished to obtain the exact optimal rate for l in one of these cases, e.g., FT models, the arguments used to find l^{OPT} would carry through with appropriate modification. Contaminating processes for which $E[I_u(\lambda_j)] \simeq T/j^2$ include both RLS and DLS processes. See the references in Section 2 for details.*

Remark 9. *As can be seen from the above expressions, there is no clear bias-variance tradeoff in the choice of trimming parameter. Although increasing the trimming decreases the bias due to the shift or trend component, it simultaneously increases the bias that comes from ignoring lower frequencies that contain information on the latent long-memory process. Similarly, an increased trimming has opposing effects on the variance: ignoring more lower frequency ordinates decreases estimation variability arising from the low-frequency contamination while*

it increases variability arising from discarding "observations" on the contaminated long-memory process.

5   Finite Sample Performance of the Robust Estimator

With the consistency and asymptotic normality of the trimmed FDQML estimator of the LMSV model parameters established, let us now investigate how well the estimator performs in data-calibrated, finite sample scenarios. For this finite sample experiment, I examined four DGPs whose parameters were calibrated from two different high-frequency returns series.[3] The first two sets of parameter values come from McCloskey's (2010) trimmed FDQML estimation of the standard autoregressive stochastic volatility (ARSV) model for deseasonalized high-frequency Japanese Yen to US dollar exchange rate returns.[4] McCloskey's (2010) estimation methodology restricts $d = 0$ but is robust to the presence of low-frequency contamination. On the other hand, Deo et al. (2006) estimated the parameters of an autoregressive LMSV model for deseasonalized high-frequency S&P 500 stock market returns (mentioned on p. 50). They used the standard FDQML estimator which is *not* robust to low-frequency contamination but allows for long-memory. By examining specifications based upon these two data calibrations, we can hope to obtain an idea of how well the trimmed FDQML estimator works in practically relevant contexts in which low-frequency contamination and/or long-memory may be present. Since the standard ARSV model is nested in the autoregressive LMSV specification, we are free to estimate all of the model parameters of the latter, including d. In order to make fair comparisons, the parameter search space Θ was set to be equal across DGPs and estimators while satisfying Assumption 2. Specifically, I set $\Theta = \mathcal{D} \times \Sigma_\eta \times \Sigma_\varepsilon \times \mathcal{A} = [-0.1, 0.7] \times [0.1, 100] \times [0.1, 100] \times [0.1, 0.99]$. Finally, I found the bias and root mean-squared error (RMSE) of the standard FDQML estimator and seven different trimmed FDQML estimators in order to assess how trimming affects the estimation, which trimming works best for which type of DGP and to see if the order of l^{OPT} can provide some practical guidance for trimming choice in finite samples. All values were calculated from 1000 Monte Carlo draws for sample sizes of $T = 4000$, 8000 and 16000.

In terms of the DGP of Assumption 1, McCloskey (2010) estimated the parameter vector θ for the model with $d = 0$, $q = 0$ and $m = 1$, the standard ARSV model. The parameter

---

[3]These calibrations are not exact because Gaussianity of e_t is imposed in the simulations in order to make bias and root MSE calculations for estimates of σ_ε^2 feasible. Recall that when e_t is Gaussian, $(\sigma_\varepsilon^0)^2 = \pi^2/2$. Nevertheless, the sets of calibrated values for θ used here entail estimates of σ_ε^2 very close to $\pi^2/2$.

[4]Specifically, I use his estimates for "AB" deseasonalized data with trimming $l = \lfloor T^{0.51} \rfloor$, see Table 12 in McCloskey (2010).

estimates he obtained correspond to $a^0 = 0.78$ (the autoregressive parameter) and $(\sigma_\eta^0)^2 = 0.33$ with d^0 restricted to equal zero ($(\sigma_\varepsilon^0)^2$ is set equal to $\pi^2/2$ here, see footnote 2). Since his estimation technique is robust to low-frequency contamination, the u_t component of Assumption 1 is free to take a variety of forms. For the purposes of this Monte Carlo exercise, I specified $\{u_t\}$ as a RLS process since the literature indicates level shifts are particularly relevant in this volatility context. In terms of the parameters described on p. 4, I set the variance of the level shifts σ_ξ^2 equal to the variance of v_t and the average number of level shifts per sample p equal to 10.

The bias and RMSE values for the average of 10 RLS case are reported in Table 1 with the numbers under "trimming" corresponding to the trimming exponent α where $l = \lfloor T^\alpha \rfloor$. Notice that only the trimmings with $\alpha = 0.51$ and 0.6 satisfy Assumption 3*, required for asymptotic normality, while all but that with $\alpha = 0$ saitisfy Assumption 3, required for consistency. The first feature to note from this table is that the standard (untrimmed) FDQML estimators of the persistence parameters (d and a) are largely biased. The estimates of d are upward biased, in agreement with similar findings for semiparametric memory parameter estimates in the presence of level shifts (see the references in the introduction). Interestingly, in the untrimmed context, the estimates of a exhibit large negative biases. It appears that low-frequency contamination induces the standard estimator to confuse some of the short-memory persistence of the time series with long-memory persistence. To the best of my knowledge, this is the first time such a feature has been uncovered. In line with the asymptotic results, we can see that for all sample sizes, trimming removes substantial portions of the bias in the persistence parameter estimates. This bias reduction is nonuniform across trimmings. Of the trimmings studied, the trimming exponents of 0.4 and 0.45 seem to provide the largest bias reductions for the persistence parameters. The extent of these bias reductions is manifest in major RMSE reductions over the standard estimator, oftentimes exceeding 50%. Nevertheless, very large trimming is detrimental because it induces bias by ignoring too many periodogram ordinates for which the LMSV process $\{\widetilde{v}_t\}$ dominates the level shift contamination $\{u_t\}$. In this long-memory context, there appears to be a very similar bias tradeoff to that explored rather extensively by McCloskey (2010) in the short-memory context. The presence of level shifts does not seem to induce much bias in the estimators of the variance parameters of the model (σ_η^2 and σ_ε^2). Moreover, the variance parameter estimates perform quite similarly across trimmings, excluding perhaps the very highest. In terms of RMSE, the trimming exponent of 0.45 appears to work best all-around for this DGP. This is broadly in line with the rate that would be recommended by

l^{OPT} if the analysis of Section 4 extended to the boundary case where $d^0 = 0$. Furthermore, it appears that the finite sample RMSE-minimizing rate approaches l^{OPT} from below as the sample size grows.

Directing our attention now to cases for which low-frequency contamination is absent and the standard FDQML estimator has been shown to be consistent and asymptotically normal, Table 2 details the performance of the various FDQML estimators for the same calibrated DGP as in Table 1 but with no level shifts. A perhaps surprising feature of this table is that the standard estimator does not always dominate the trimmed one in terms of bias and RMSE even though trimming is unnecessary in this context. For example, a trimming exponent of 0.3 or 0.35 performs best in terms of bias for $\hat{d}$ when the sample size is 16000. However, the differences in performance across trimmings are rather minor in comparison to the case with level shifts as long as the trimming is not excessive. We can see that when evaluated at a trimming exponent that works well in the presence of level shifts and that is roughly in line with the theoretically-driven choice l^{OPT}, say 0.45, the difference in RMSE between the trimmed and untrimmed estimators in the uncontaminated case is quite small in comparison to this difference in the presence of low-frequency contamination (i.e., Table 1). In fact, the difference in RMSE between these two estimators appears to shrink in the sample size in the absence of low-frequency contamination while it appears to grow in its presence. Of course, in the absence of contamination, l^{OPT} is no longer the asymptotic MSE-minimizing trimming rate. As noted by McCloskey (2010) in the short-memory context, too high of a trimming induces bias by discarding valuable information on the contaminated long-memory process.

Next we examine a DGP that entails both long-memory and low-frequency contamination. Deo et al. (2006) estimated the parameter vector θ for the model with $q = 0$ and $m = 1$, the autoregressive LMSV model. The parameter estimates they obtained correspond to $d^0 = 0.37$, $a^0 = 0.35$ and $(\sigma_\eta^0)^2 = 0.27$. Since the estimation methodology used by these authors is not robust to low-frequency contamination, these parameter estimates may be biased. Nevertheless, by adding low-frequency contamination to these calibrated values, we can obtain a good idea of how it may affect parameter estimates when $\{\widetilde{v}_t\}$ is actually a long-memory process. In this vein, I added the same RLS process to the LMSV process as in Table 1, adjusting the variance of the level shifts σ_ξ^2 to be equal to the variance of v_t in this context. The results are provided in Table 3. As in Table 1, the parameter estimates exhibit large biases although, at least for the persistence parameters, these biases are not as large as in the $d^0 = 0$ case since $d^0 = 0.37$ is approaching the upper bound of admissible

values 0.5. This is in fact a manifestation of a more general feature: low-frequency contamination induces more bias and higher MSE in standard estimators for lower values of d^0. The direction of the biases in Table 3 is also in line with those in Table 1. We can again see that trimming removes most, and in some cases all, of the bias induced by the level shifts with the trimming exponent of 0.3 seeming to provide the largest bias reductions. As in Table 1, we again see major reductions in the RMSE of the persistence parameters over the standard FDQML estimator when the trimming used is not too high. For this DGP the trimming exponent of 0.3 appears to work best all-around. This rate is substantially smaller than that given by l^{OPT}, which may be partially due to the fact that we are ignoring the proportionality factor in l^{OPT}. Nevertheless, as in Table 1, we can see that the finite sample RMSE-minimizing trimming rate increases towards l^{OPT} as the sample size grows.[5]

Turning to the final DGP under scrutiny in this Monte Carlo exercise, we remove the low-frequency contamination from the long-memory process to study the calibrated pure LMSV model of Deo et al. (2006). The results for this DGP, provided in Table 4, are broadly in line with those of Table 2: a high trimming induces bias and RMSE-growth but the differences in performance across trimmings are not as large as the differences when the process is contaminated by level shifts (as in Table 3) as long as the trimming is not too high. Here, using a trimming exponent based on the theoretically guided choice l^{OPT} can substantially reduce the performance of the estimator.

In summary, this simulation experiment shows that substantial gains can be made from using the trimmed FDQML estimator to estimate the parameters of the LMSV model when the presence of low-frequency contamination is a potential concern. This is indeed the case for returns volatility series, the very series LMSV models are designed to model. We have seen that using l^{OPT} as a rough guide for trimming choice can lead to much improved performance of the FDQML estimator in the presence of low-frequency contamination. However, simply using the rate given by l^{OPT} in the absence of contamination has the potential to reduce the performance of the estimator. A pre-test with power against low-frequency contamination, such as that of Qu (2011), would thus provide some useful direction for trimming choice.

---

[5]Strictly speaking, the parameter space Θ used in this simulation study does not satisfy the extra assumption $\underline{d} > 0$ imposed in Theorem 2. However, the simulation results for this DGP do not perceptibly change when we change Θ to bound θ_1 below by some number arbitrarily close to zero.

6   Mathematical Appendix

The proofs of the main results are followed by auxiliary lemmas. For the sake of brevity, we will say "uniformly in θ" to mean uniformly in θ over Θ. Similarly, "uniformly in λ" should be taken to mean uniformly in λ over $[-\pi, \pi]$ and "uniformly in j" should be taken to mean uniformly in j over $\mathcal{F}_1$.

Proof of Theorem 1: Begin by noting $f(\lambda; \theta)^{-1} \leq 2\pi/\widetilde{\sigma}_\varepsilon^2 < \infty$ for all $\lambda \in [-\pi, \pi]$ and $\theta \in \Theta$, where $\widetilde{\sigma}_\varepsilon^2 = \min_{\theta \in \Theta} \theta_3$, which exists by the compactness of Θ imposed in Assumption 2. Hence given Assumption 1,

$$\sup_{\theta \in \Theta} T^{-1} \sum_{j \in \mathcal{F}_l} \frac{I_u(\lambda_j)}{f(\lambda_j; \theta)} \leq \frac{2\pi}{\widetilde{\sigma}_\varepsilon^2} T^{-1} \sum_{j \in \mathcal{F}_l} I_u(\lambda_j) = O_p\left(T^{-1} \sum_{j=l}^{T} \frac{T}{j^2}\right) = O_p\left(\frac{\log T}{l}\right) = o_p(1) \tag{A.1}$$

by Assumption 3. Also,

$$\sup_{\theta \in \Theta} \left| T^{-1} \sum_{j \in \mathcal{F}_l} \frac{w_{\widetilde{v}}(\lambda_j) w_u(-\lambda_j)}{f(\lambda_j; \theta)} \right| \leq \frac{2\pi}{\widetilde{\sigma}_\varepsilon^2} T^{-1} \sum_{j \in \mathcal{F}_l} |w_{\widetilde{v}}(\lambda_j) w_u(-\lambda_j)| = O_p\left(T^{-1} \sum_{j=l}^{T} \frac{T^{1/2+d^0}}{j^{1+d^0}}\right)$$

$$= O_p\left(T^{d^0-1/2} \sum_{j=l}^{T} \frac{1}{j^{1+d^0}}\right) = O_p(T^{d^0-1/2}) = o_p(1), \tag{A.2}$$

using Lemma 1. Thus, by (A.1), (A.2) and the decomposition (2),

$$L_{T,l}(\theta) = T^{-1} \sum_{j \in \mathcal{F}_l} \left\{ \log\left(f(\lambda_j; \theta)\right) + \frac{I_{\widetilde{v}}(\lambda_j)}{f(\lambda_j; \theta)} \right\} + o_p(1) \tag{A.3}$$

uniformly in θ. Now note that by Lemma 1 and Assumption 3,

$$\sup_{\theta \in \Theta} \left| T^{-1} \sum_{j \in \mathcal{F}_1} \left\{ \log f(\lambda_j; \theta) + \frac{I_{\widetilde{v}}(\lambda_j)}{f(\lambda_j; \theta)} \right\} - T^{-1} \sum_{j \in \mathcal{F}_l} \left\{ \log\left(f(\lambda_j; \theta)\right) + \frac{I_{\widetilde{v}}(\lambda_j)}{f(\lambda_j; \theta)} \right\} \right|$$

$$= \sup_{\theta \in \Theta} \left| 2 T^{-1} \sum_{j=1}^{l} \left\{ \log f(\lambda_j; \theta) + \frac{I_{\widetilde{v}}(\lambda_j)}{f(\lambda_j; \theta)} \right\} \right| = O\left(T^{-1} \sum_{j=1}^{l} \log \lambda_j^{-2\widetilde{d}}\right) + O_p\left(T^{-1} \sum_{j=1}^{l} \lambda_j^{-2d^0}\right)$$

$$= O_p\left(\frac{l \log T}{T}\right) + O_p\left(\left(\frac{l}{T}\right)^{1-2d^0}\right) = o_p(1) \tag{A.4}$$

where $\widetilde{d} = \max_{\theta \in \Theta} \theta_1$. Hence, (A.3) and (A.4) provide that

$$L_{T,l}(\theta) = T^{-1} \sum_{j \in \mathcal{F}_1} \left\{ \log\left(f(\lambda_j; \theta)\right) + \frac{I_{\widetilde{v}}(\lambda_j)}{f(\lambda_j; \theta)} \right\} + o_p(1)$$

uniformly in θ. Then, the proof of Theorem 1 in Breidt et al. (1998), with the above result, shows that

$$L_{T,l}(\theta) \overset{p}{\longrightarrow} \frac{1}{2\pi} \int_{-\pi}^{\pi} \left\{ \log f(\lambda; \theta) + \frac{f(\lambda; \theta^0)}{f(\lambda; \theta)} \right\} d\lambda \tag{A.5}$$

A.1

uniformly in θ. The remainder of the proof then follows since the right hand side of (A.5) is uniquely minimized at $\theta = \theta^0$ and Θ is compact (see the proof of Theorem 1 in Breidt et al., 1998 for details). ∎

Proof of Theorem 2: Using the Mean Value Theorem, we have

$$\sqrt{T}\left(\hat{\theta} - \theta^0\right) = -\left[\frac{\partial^2 L_{T,l}(\widetilde{\theta})}{\partial\theta\partial\theta'}\right]^{-1}\sqrt{T}\frac{\partial L_{T,l}(\theta^0)}{\partial\theta} \tag{A.6}$$

for some $\widetilde{\theta}$ such that $||\widetilde{\theta} - \theta^0|| \leq ||\hat{\theta} - \theta^0||$ since θ^0 is in the interior of Θ. (Strictly speaking, this is a slight abuse of notation. See footnote 2.) Using the decomposition (2),

$$\frac{\partial^2 L_{T,l}(\theta)}{\partial\theta\partial\theta'} = \frac{1}{T}\sum_{j\in\mathcal{F}_l}\{I_x(\lambda_j) - f(\lambda_j;\theta)\}\frac{\partial^2[f(\lambda_j;\theta)^{-1}]}{\partial\theta\partial\theta'} - \frac{1}{T}\sum_{j\in\mathcal{F}_l}\frac{\partial f(\lambda_j;\theta)}{\partial\theta}\frac{\partial[f(\lambda_j;\theta)^{-1}]}{\partial\theta'}$$

$$= \frac{1}{T}\sum_{j\in\mathcal{F}_1}\{I_{\widetilde{v}}(\lambda_j) - f(\lambda_j;\theta)\}\frac{\partial^2[f(\lambda_j;\theta)^{-1}]}{\partial\theta\partial\theta'} \tag{A.7}$$

$$- \frac{1}{T}\sum_{j=-l,j\neq0}^{l}\{I_{\widetilde{v}}(\lambda_j) - f(\lambda_j;\theta)\}\frac{\partial^2[f(\lambda_j;\theta)^{-1}]}{\partial\theta\partial\theta'} \tag{A.8}$$

$$+ \frac{1}{T}\sum_{j\in\mathcal{F}_l}I_u(\lambda_j)\frac{\partial^2[f(\lambda_j;\theta)^{-1}]}{\partial\theta\partial\theta'} \tag{A.9}$$

$$+ \frac{1}{T}\sum_{j\in\mathcal{F}_l}\{w_{\widetilde{v}}(\lambda_j)w_u(-\lambda_j) + w_{\widetilde{v}}(-\lambda_j)w_u(\lambda_j)\}\frac{\partial^2[f(\lambda_j;\theta)^{-1}]}{\partial\theta\partial\theta'} \tag{A.10}$$

$$+ \frac{1}{T}\sum_{j\in\mathcal{F}_1}\frac{\partial\log f(\lambda_j;\theta)}{\partial\theta}\frac{\partial\log f(\lambda_j;\theta)}{\partial\theta'} \tag{A.11}$$

$$- \frac{1}{T}\sum_{j=-l,j\neq0}^{l}\frac{\partial\log f(\lambda_j;\theta)}{\partial\theta}\frac{\partial\log f(\lambda_j;\theta)}{\partial\theta'} \tag{A.12}$$

for all $\theta \in \Theta$. Starting with (A.7), note that $\partial^2[f(\lambda;\theta)^{-1}]/\partial\theta\partial\theta'$ is uniformly continuous over $[-\pi,\pi]\times\Theta$. To see this, note for example that

$$\frac{\partial^2[f(\lambda;\theta)^{-1}]}{\partial d^2} = \frac{8\pi\sigma_\eta^4\sigma_\varepsilon^2(2\sin|\lambda/2|)^{4d}\log^2(2\sin|\lambda/2|)|B(e^{-i\lambda})|^2|A(e^{-i\lambda})|^4}{(\sigma_\eta^2|B(e^{-i\lambda})|^2 + \sigma_\varepsilon^2(2\sin|\lambda/2|)^{2d}|A(e^{-i\lambda})|^2)^3}. \tag{A.13}$$

(In this and the following expositions, I focus on derivatives with respect to d since they diverge the fastest/converge the slowest as $\lambda \to 0$.) Also, $\{\widetilde{v}_t\}$ is ergodic so that straightforward modification of the proof of Lemma 1 of Hannan (1973) (also see the proof of Theorem 1 in Breidt et al., 1998) provides that

$$\frac{1}{T}\sum_{j\in\mathcal{F}_1}I_{\widetilde{v}}(\lambda_j)\frac{\partial^2[f(\lambda_j;\theta)^{-1}]}{\partial\theta\partial\theta'} \xrightarrow{a.s.} \frac{1}{2\pi}\int_{-\pi}^{\pi}f(\lambda;\theta^0)\frac{\partial^2[f(\lambda;\theta)^{-1}]}{\partial\theta\partial\theta'}d\lambda$$

A.2

uniformly in θ. Moreover, $f(\lambda;\theta)\partial^2[f(\lambda;\theta)^{-1}]/\partial\theta\partial\theta'$ is integrable in λ with a continuous integral uniformly in θ. This can be seen from inspection of quantities such as (A.13) multiplied by $f(\lambda;\theta)$. Hence,

$$\frac{1}{T}\sum_{j\in\mathcal{F}_1}f(\lambda_j;\theta)\frac{\partial^2[f(\lambda_j;\theta)^{-1}]}{\partial\theta\partial\theta'}\longrightarrow\frac{1}{2\pi}\int_{-\pi}^{\pi}f(\lambda;\theta)\frac{\partial^2[f(\lambda;\theta)^{-1}]}{\partial\theta\partial\theta'}d\lambda$$

uniformly in θ and

$$\frac{1}{T}\sum_{j\in\mathcal{F}_1}\{I_{\widetilde{v}}(\lambda_j)-f(\lambda_j;\widetilde{\theta})\}\frac{\partial^2[f(\lambda_j;\widetilde{\theta})^{-1}]}{\partial\theta\partial\theta'}\xrightarrow{p}0$$

since $\widetilde{\theta}\xrightarrow{p}\theta^0$ by Theorem 1. Moving now to (A.8), note that $f(\lambda;\theta)\partial^2[f(\lambda;\theta)^{-1}]/\partial\theta\partial\theta'$ is also uniformly continuous over $[-\pi,\pi]\times\Theta$, which follows similarly. Using this result, Lemma 1 and the fact that $\partial^2[f(\lambda;\theta)^{-1}]/\partial\theta\partial\theta'$ is uniformly bounded over $[-\pi,\pi]\times\Theta$, (A.8) and can be given the following order uniformly in θ:

$$O_p\left(\frac{1}{T}\sum_{j=1}^{l}\left\{\lambda_j^{-2d^0}+1\right\}\right)=O_p\left(\left(\frac{l}{T}\right)^{1-2d^0}+\frac{l}{T}\right)=o_p(1).$$

Using the same boundedness properties of $\partial^2[f(\lambda_j;\theta)^{-1}]/\partial\theta\partial\theta'$ as above and the order of $E[I_u(\lambda_j)]$, we can establish that (A.9) is $O_p(T^{-1}\sum_{j\in\mathcal{F}_l}T/j^2)=O_p(\log T/l)=o_p(1)$ uniformly in θ. Similarly, with the help of Lemma 1, we obtain that (A.10) is

$$O_p\left(\frac{1}{T}\sum_{j\in\mathcal{F}_l}T^{-1/2}\lambda_j^{-(1+d^0)}\right)=O_p(T^{d^0-1/2})=o_p(1)$$

uniformly in θ. Next, $(\partial\log f(\lambda_j;\theta)/\partial\theta)(\partial\log f(\lambda_j;\theta)/\partial\theta')$ is integrable in λ with a continuous integral uniformly in θ. To see this, note that, e.g.,

$$\frac{\partial\log f(\lambda;\theta)}{\partial d}=-\frac{2\log(2\sin|\lambda/2|)|B(e^{-i\lambda})|^2}{\sigma_\eta^2|B(e^{-i\lambda})|^2+\sigma_\varepsilon^2(2\sin|\lambda/2|)^{2d}|A(e^{-i\lambda})|^2}.\tag{A.14}$$

Thus, (A.11) is equal to

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\partial\log f(\lambda;\theta)}{\partial\theta}\frac{\partial\log f(\lambda;\theta)}{\partial\theta'}d\lambda+o(1)$$

uniformly in θ. Finally, expressions like (A.14) show that

$$\frac{\partial\log f(\lambda_j;\theta)}{\partial\theta}\frac{\partial\log f(\lambda_j;\theta)}{\partial\theta'}=O(\log^2|\lambda_j|)$$

uniformly in j and θ so that (A.12) is

$$O\left(\frac{1}{T}\sum_{j=1}^{l}\log^2\lambda_j\right)=O(l\log^2 T/T)=o_p(1)$$

A.3

uniformly in θ. Putting all of these results together and applying Theorem 1, we obtain

$$\frac{\partial^2 L_{T,l}(\widetilde{\theta})}{\partial \theta \partial \theta'} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\partial \log f(\lambda; \theta^0)}{\partial \theta} \frac{\partial \log f(\lambda; \theta^0)}{\partial \theta'} d\lambda + o_p(1). \tag{A.15}$$

A similar decomposition to that used for $\partial^2 L_{T,l}(\theta)/\partial\theta\partial\theta'$ yields

$$\sqrt{T}\frac{\partial L_{T,l}(\theta^0)}{\partial \theta} = \frac{1}{\sqrt{T}} \sum_{j \in \mathcal{F}_1} \{I_{\widetilde{v}}(\lambda_j) - f(\lambda_j; \theta^0)\}\frac{\partial[f(\lambda_j; \theta^0)^{-1}]}{\partial \theta} \tag{A.16}$$

$$- \frac{1}{\sqrt{T}} \sum_{j=-l, j\neq 0}^{l} \{I_{\widetilde{v}}(\lambda_j) - f(\lambda_j; \theta^0)\}\frac{\partial[f(\lambda_j; \theta^0)^{-1}]}{\partial \theta} \tag{A.17}$$

$$+ \frac{1}{\sqrt{T}} \sum_{j \in \mathcal{F}_l} I_u(\lambda_j)\frac{\partial[f(\lambda_j; \theta^0)^{-1}]}{\partial \theta} \tag{A.18}$$

$$+ \frac{1}{\sqrt{T}} \sum_{j \in \mathcal{F}_l} \{w_{\widetilde{v}}(\lambda_j)w_u(-\lambda_j) + w_{\widetilde{v}}(-\lambda_j)w_u(\lambda_j)\}\frac{\partial[f(\lambda_j; \theta^0)^{-1}]}{\partial \theta}. \tag{A.19}$$

By Theorem 1.2 of Hosoya (1997), (A.16) is asymptotically normally distributed with mean zero and covariance matrix whose $(j,k)^{th}$ element is equal to

$$4\pi \int_{-\pi}^{\pi} \frac{\partial \log f(\lambda; \theta^0)}{\partial \theta_j} \frac{\partial \log f(\lambda; \theta^0)}{\partial \theta_k} d\lambda + \frac{E[\varepsilon_t^4] - 3E[\varepsilon_t^2]}{(2\pi)^2} \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \frac{\partial[f(\lambda_1; \theta^0)^{-1}]}{\partial \theta_j}\frac{\partial[f(\lambda_2; \theta^0)^{-1}]}{\partial \theta_k} d\lambda_1 d\lambda_2.$$

By Lemma 3, (A.17) is $O_p(l^{1/2}\log T/T^{1/2}) = o_p(1)$. Similar arguments to those used to find the order of (A.9) yield that (A.18) is $O_p(T^{1/2}\log T/l) = o_p(1)$ by Assumption 3* since $\partial[f(\lambda; \theta^0)^{-1}]/\partial\theta$ is uniformly bounded, as can be seen from quantities like

$$\frac{\partial[f(\lambda; \theta)^{-1}]}{\partial d} = \frac{4\pi\sigma_\eta^2(2\sin|\lambda/2|)^{2d}\log(2\sin|\lambda/2|)|B(e^{-i\lambda})|^2}{\sigma_\eta^2|B(e^{-i\lambda})|^2 + \sigma_\varepsilon^2(2\sin|\lambda/2|)^{2d}|A(e^{-i\lambda})|^2}.$$

For (A.19), using the same uniform boundedness property, note that

$$E\left[\left(\frac{1}{\sqrt{T}} \sum_{j \in \mathcal{F}_l} w_{\widetilde{v}}(\lambda_j)w_u(-\lambda_j)\frac{\partial[f(\lambda_j; \theta^0)^{-1}]}{\partial \theta}\right)\left(\frac{1}{\sqrt{T}} \sum_{j \in \mathcal{F}_l} w_{\widetilde{v}}(\lambda_j)w_u(-\lambda_j)\frac{\partial f(\lambda_j; \theta^0)^{-1}}{\partial \theta'}\right)\right]$$

$$= \frac{1}{T} \sum_{j \in \mathcal{F}_l}\sum_{k \in \mathcal{F}_l} E[w_{\widetilde{v}}(\lambda_j)w_{\widetilde{v}}(-\lambda_k)]E[w_u(\lambda_j)w_u(-\lambda_k)]\frac{\partial[f(\lambda_j; \theta^0)^{-1}]}{\partial \theta}\frac{\partial f(\lambda_k; \theta^0)^{-1}}{\partial \theta'}$$

$$= O\left(T^{-1}\sum_{j=l}^{T}\sum_{k=l}^{T} E[w_{\widetilde{v}}(\lambda_j)w_{\widetilde{v}}(-\lambda_k)]E[w_u(-\lambda_j)w_u(\lambda_k)]\right)$$

$$= O\left(T^{-1}\sum_{j=l}^{T} E[I_{\widetilde{v}}(\lambda_j)]E[I_u(\lambda_j)]\right) + O\left(T^{-1}\sum_{j=l}^{T}\sum_{k=j+1}^{T} E[w_{\widetilde{v}}(\lambda_j)w_{\widetilde{v}}(-\lambda_k)]E[w_u(-\lambda_j)w_u(\lambda_k)]\right). \tag{A.20}$$

A.4

Assumption 1 and the Cauchy-Schwartz inequality provide that $E[I_u(\lambda_j)] = O(T/j^2)$ uniformly in j and $E[w_u(-\lambda_j)w_u(\lambda_k)] = O(T/jk)$ uniformly in $j < k$ while Lemma 2 of this paper and Theorem 2 of Robinson (1995) provide that

$$E[w_{\widetilde{v}}(\lambda_j)w_{\widetilde{v}}(-\lambda_k)] = O\left(\lambda_j^{-d^0}\lambda_k^{-d^0}\frac{\log k}{j}\right) = O\left(\frac{T^{2d^0}\log k}{j^{1+d^0}k^{d^0}}\right)$$

uniformly in $j < k$. Hence, (A.20) is

$$O\left(\frac{1}{T}\sum_{j=l}^{T}\frac{T^{2d^0}}{j^{2d^0}}\frac{T}{j^2}\right) + O\left(\frac{1}{T}\sum_{j=l}^{T}\sum_{k=j+1}^{T}\frac{T^{2d^0}\log k}{j^{1+d^0}k^{d^0}}\frac{T}{jk}\right) = O\left(\frac{T^{2d^0}\log^3 T}{l^{1+2d^0}}\right) = o_p(1).$$

Combining these results with (A.6) and (A.15), we obtain the statement of the theorem. ■

Lemma 1. *Under Assumptions 1 and 2,*

$$E[I_{\widetilde{v}}(\lambda_j)] = O(\lambda_j^{-2d^0}) \quad and \quad E|w_{\widetilde{v}}(\lambda_j)w_u(-\lambda_j)| = O(T^{-1/2}\lambda_j^{-(1+d^0)})$$

uniformly for all nonzero Fourier frequencies λ_j.

Proof: For the first result, note that $I_{\widetilde{v}}(\lambda_j)$ has an analogous decomposition to (2) in terms of v and ε quantities. Denote the first term on the right hand side of (1) by $f_1(\lambda; \theta)$. Then, by well-known results, uniformly for all nonzero λ_j and $\theta^0 \in \Theta$,

$$E[I_v(\lambda_j)] = O(f_1(\lambda_j; \theta^0)) = O(\lambda_j^{-2d^0})$$

(see Robinson, 1995 and Hurvich and Beltrao, 1993 for example). Also, $E[I_\varepsilon(\lambda_j)] = O(1)$ uniformly for all λ_j. So, by the Cauchy-Schwartz inequality, we obtain the first result:

$$E|w_v(\lambda_j)w_\varepsilon(-\lambda_j)| \leq (E|I_v(\lambda_j)|)^{1/2}(E|I_\varepsilon(\lambda_j)|)^{1/2} = O(\lambda_j^{-d^0}) = O(\lambda_j^{-2d^0}).$$

The second result follows from the Cauchy-Schwartz inequality and the first result:

$$E|w_{\widetilde{v}}(\lambda_j)w_u(-\lambda_j)| \leq (E|I_{\widetilde{v}}(\lambda_j)|)^{1/2}(E|I_u(\lambda_j)|)^{1/2} = O(\lambda_j^{-d^0})O(T^{-1/2}\lambda_j^{-1}). \quad ■$$

Lemma 2. *Under Assumptions 1 and 2, for any sequence of positive integers, $j = j(T)$ and $k = k(T)$ such that $j > k$ and $\lim_{T\to\infty} j/T \in (0, \infty)$,*

$$E[w_{\widetilde{v}}(\lambda_j)w_{\widetilde{v}}(-\lambda_k)] = O\left(\frac{\log j}{k\lambda_j^{d^0}\lambda_k^{d^0}}\right).$$

A.5

Proof: Using the expansion on p. 1063 of Robinson (1995),

$$E[w_{\tilde{v}}(\lambda_j)w_{\tilde{v}}(-\lambda_k)] = \int_{(\lambda_j+\lambda_k)/2}^{\pi} \{f(\lambda;\theta^0) - f(\lambda_j;\theta^0)\}E_{jk}(\lambda)d\lambda \tag{A.21}$$

$$+ \int_{\lambda_k/2}^{(\lambda_j+\lambda_k)/2} \{f(\lambda;\theta^0) - f(\lambda_k;\theta^0)\}E_{jk}(\lambda)d\lambda \tag{A.22}$$

$$- \{f(\lambda_j;\theta^0) - f(\lambda_k;\theta^0)\} \int_{\lambda_k/2}^{(\lambda_j+\lambda_k)/2} E_{jk}(\lambda)d\lambda \tag{A.23}$$

$$+ \int_{-\pi}^{\lambda_k/2} \{f(\lambda;\theta^0) - f(\lambda_j;\theta^0)\}E_{jk}(\lambda)d\lambda, \tag{A.24}$$

where $E_{jk}(\lambda) = (2\pi T)^{-1}D(\lambda_j - \lambda)D(\lambda - \lambda_k)$ with $D(\lambda) \equiv \sum_{t=1}^{T} e^{it\lambda}$ being Dirichlet's kernel. By the Mean Value Theorem and the fact that

$$|D(\lambda)| \leq 2|\lambda|^{-1} \text{ for } 0 < |\lambda| < \pi \tag{A.25}$$

(see, pp. 49-51 of Zygmund, 1977), (A.21) is bounded by

$$\left\{ \max_{(\lambda_j+\lambda_k)/2 \leq \lambda \leq 2\lambda_j} |f'(\lambda;\theta^0)| \right\} (\pi T)^{-1} \int_{(\lambda_j+\lambda_k)/2}^{\pi} |D(\lambda - \lambda_k)|d\lambda = O\left(\frac{\log j}{T}\right) = O\left(\frac{\log j}{k\lambda_j^{d^0}\lambda_k^{d^0}}\right),$$

with $f'(\lambda;\theta^0) \equiv \partial f(\lambda;\theta^0)/\partial\lambda$, where the first equality follows from the facts that $|f'(\lambda;\theta^0)| < \infty$ for all $0 < \lambda \leq \pi$ and

$$\int_{-C\lambda_j}^{C\lambda_j} |D(\lambda)|d\lambda = O(\log j) \tag{A.26}$$

for $C < \infty$ (see p. 67 of Zygmund, 1977). By nearly identical arguments, (A.22) is bounded by

$$\left\{ \max_{\lambda_k/2 \leq \lambda \leq (\lambda_j+\lambda_k)/2} |f'(\lambda;\theta^0)| \right\} (\pi T)^{-1} \int_{\lambda_k/2}^{(\lambda_j+\lambda_k)/2} |D(\lambda_j - \lambda)|d\lambda = O\left(\frac{\log j}{T\lambda_k^{1+2d^0}}\right) = O\left(\frac{\log j}{k\lambda_j^{d^0}\lambda_k^{d^0}}\right),$$

where the first equality follows from the fact that $|f'(\lambda;\theta^0)| = O(\lambda^{-1-2d^0})$ as $\lambda \to 0+$. For $k \geq j/2$, (A.23) is bounded by

$$(\lambda_j - \lambda_k)\{ \max_{\lambda_k \leq \lambda \leq \lambda_j} |f'(\lambda;\theta^0)|\} \int_{\lambda_k/2}^{(\lambda_j+\lambda_k)/2} |E_{jk}(\lambda)|d\lambda$$

$$= O\left((\lambda_j - \lambda_k)\lambda_k^{-1-2d^0}T^{-1} \max_{\lambda_k/2 \leq \lambda \leq (\lambda_j+\lambda_k)/2} |\lambda_j - \lambda|^{-1} \int_{\lambda_k/2}^{(\lambda_j+\lambda_k)/2} |D(\lambda - \lambda_k)|d\lambda \right)$$

$$= O\left(\frac{\log j}{T\lambda_k^{1+2d^0}}\right) = O\left(\frac{\log j}{k\lambda_j^{d^0}\lambda_k^{d^0}}\right),$$

A.6

where the first equality uses (A.25) and the second uses (A.26). Similarly, if $k < j/2$, (A.23) is bounded by

$$\{|f(\lambda_j; \theta^0)| + |f(\lambda_k; \theta^0)|\} \int_{\lambda_k/2}^{(\lambda_j + \lambda_k)/2} |E_{jk}(\lambda)| d\lambda$$

$$= O\left(\lambda_k^{-2d^0} T^{-1} \max_{\lambda_k/2 \le \lambda \le (\lambda_j + \lambda_k)/2} |\lambda_j - \lambda|^{-1} \int_{\lambda_k/2}^{(\lambda_j + \lambda_k)/2} |D(\lambda - \lambda_k)| d\lambda\right)$$

$$= O\left(\lambda_k^{-2d^0} T^{-1} \lambda_j^{-1} \log j\right) = O\left(\frac{\log j}{k \lambda_j^{d^0} \lambda_k^{d^0}}\right)$$

since $f(\lambda; \theta^0) = O(\lambda^{-2d^0})$ as $\lambda \to 0+$. Finally, (A.24) is bounded by

$$\int_{-\pi}^{\lambda_k/2} \{|f(\lambda; \theta^0)| + |f(\lambda_j; \theta^0)|\} |E_{jk}(\lambda)| d\lambda$$

$$= O\left(T^{-1} \max_{-\pi \le \lambda \le \lambda_k/2} |\lambda_j - \lambda|^{-1} \max_{-\pi \le \lambda \le \lambda_k/2} |\lambda - \lambda_k|^{-1} \int_{-\pi}^{\lambda_k/2} \{|f(\lambda; \theta^0)| + |f(\lambda_j; \theta^0)|\} d\lambda\right)$$

$$= O\left(T^{-1} \lambda_j^{-1} \lambda_k^{-1} (1 + \lambda_j^{-2d^0})\right) = O\left(\frac{\log j}{k \lambda_j^{d^0} \lambda_k^{d^0}}\right). \quad \blacksquare$$

Lemma 3. *Under Assumptions 1 and 2, if $d^0 \ge \underline{d} > 0$, $E[\varepsilon_t^4] < \infty$ and*

$$\frac{1}{l} + \frac{l^5 \log^2 T}{T^4} \to 0,$$

then

$$E\left|\sum_{j=1}^{l} \{I_{\tilde{v}}(\lambda_j) - f(\lambda_j; \theta^0)\} \frac{\partial [f(\lambda_j; \theta^0)^{-1}]}{\partial \theta}\right| = O(l^{1/2} \log T).$$

Proof: To reduce notational clutter, let

$$g(\lambda; \theta^0) = \lambda^{-2d^0} \frac{(\sigma_\eta^0)^2 |B(1)|^2}{2\pi |A(1)|^2} \left\{1 + \lambda^{2d^0} \frac{(\sigma_\varepsilon^0)^2 |A(1)|^2}{(\sigma_\eta^0)^2 |B(1)|^2}\right\}$$

and $I_j = I_{\tilde{v}}(\lambda_j)$, $g_j = g(\lambda_j; \theta^0)$ and $f_j = f(\lambda_j; \theta^0)$ for $j = 1, \ldots, l$. Then, we have

$$E\left|\sum_{j=1}^{l} \{I_{\tilde{v}}(\lambda_j) - f(\lambda_j)\} \frac{\partial [f(\lambda_j; \theta^0)^{-1}]}{\partial \theta}\right| = E\left|\sum_{j=1}^{l} \left\{\frac{I_j}{f_j} - 1\right\} \frac{\partial \log f(\lambda_j; \theta^0)}{\partial \theta}\right|$$

$$\le E\left|\sum_{j=1}^{l} \left\{\frac{I_j}{g_j} - 1\right\} \frac{\partial \log f(\lambda_j; \theta^0)}{\partial \theta}\right| \quad (A.27)$$

A.7

$$+ E\left|\sum_{j=1}^{l}\left\{\frac{I_j}{f_j} - \frac{I_j}{g_j}\right\}\frac{\partial \log f(\lambda_j; \theta^0)}{\partial \theta}\right|. \qquad (A.28)$$

By Proposition A.1 of Hurvich et al. (2005), (A.27) is bounded by $Cl^{1/2}\log T$ for some finite constant C since for $i = 1, \ldots, m + q + 3$,

$$\frac{1}{l \log T}\left|\frac{\partial \log f(\lambda_j; \theta^0)}{\partial \theta_i} - \frac{\partial \log f(\lambda_{j+1}; \theta^0)}{\partial \theta_i}\right| \leq K_1 \frac{1}{l \log T}$$

for $j = 1, \ldots, l - 1$ and

$$\frac{1}{l \log T}\left|\frac{\partial \log f(\lambda_j; \theta^0)}{\partial \theta_i}\right| \leq K_2 \frac{1}{l}$$

for some generic finite constants K_1 and K_2. This can be seen by examining quantities such as (A.14). To bound (A.28), note that as $\lambda \to 0+$,

$$f(\lambda; \theta^0) \simeq \lambda^{-2d^0}\left\{\frac{(\sigma_\eta^0)^2|B(1)|^2}{2\pi|A(1)|^2} + O(\lambda^2) + \lambda^{2d^0}\frac{(\sigma_\varepsilon^0)^2}{2\pi}\right\} = g(\lambda; \theta^0)(1 + O(\lambda^2)).$$

Hence, for $i = 1, \ldots, m + q + 3$, (A.28) is bounded by

$$E\left[\sum_{j=1}^{l}\left|\frac{I_j}{f_j} - \frac{I_j}{g_j}\right|\left|\frac{\partial \log f(\lambda_j; \theta^0)}{\partial \theta_i}\right|\right] \leq K_3 \log T E\left[\sum_{j=1}^{l}\left|\frac{I_j}{g_j}\left(\frac{g_j}{f_j} - 1\right)\right|\right]$$

$$\leq K_4 \log T E\left[\sum_{j=1}^{l}\left|\frac{g_j}{f_j} - 1\right|\right]$$

$$= O\left(\log T \sum_{j=1}^{l}\lambda_j^2\right) = O\left(\frac{l^3 \log T}{T^2}\right) = O(l^{1/2}\log T),$$

for some generic finite constants K_3 and K_4, where the second inequality uses the fact that $E[I_j/g_j] = O(1)$ (see Theorem 2 of Robinson, 1995 for example).

References

Andrews, D. W. K., Sun, Y., 2004. Adaptive local polynomial Whittle estimation of long-range dependence. Econometrica 72, 569–614.

Breidt, F. J., Crato, N., de Lima, P., 1998. The detection and estimation of long memory in stochastic volatility. Journal of Econometrics 83, 325–348.

Dahlhaus, R., 1989. Efficient parameter estimation for self-similar processes. Annals of Statistics 17, 1749–1766.

Deo, R., Hurvich, C., Lu, Y., 2006. Forecasting realized volatility using a long-memory stochastic volatility model: Estimation, prediction and seasonal adjustment. Journal of Econometrics 131, 29–58.

Deo, R. S., Hurvich, C. M., 2001. On the log periodogram regression estimator of the memory parameter in long memory stochastic volatility models. Econometric Theory 17, 686–710.

Diebold, F., Inoue, A., 2001. Long memory and regime switching. Journal of Econometrics 105, 131–159.

Fox, R., Taqqu, M. S., 1986. Large-sample properties of parameter estimates for strongly dependent stationary gaussian time series. The Annals of Statistics 14, 517–532.

Frederiksen, P., Nielsen, M. Ø., 2008. Bias-reduced estimation of long-memory stochastic volatility. Journal of Financial Econometrics 6, 496–512.

Geweke, J., Porter-Hudak, S., 1983. The estimation and application of long memory time series models. Journal of Time Series Analysis 4, 221–238.

Granger, C. W. J., 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.

Hannan, E. J., 1973. The asymptotic theory of linear time-series models. Journal of Applied Probability 10, 130–145.

Harvey, A. C., 1998. Long-memory in stochastic volatility. In: Knight, J., Satchell, S. E. (Eds.), Forecasting Volatility in Financial Markets. Butterwoth-Heinemann, London, pp. 307–320.

Henry, M., Robinson, P., 1996. Bandwidth choice in Gaussian semiparametric estimation of long range dependence. In: Robinson, P., Rosenblatt, M. (Eds.), Athens Conference on

Applied Probability and Time Series Analysis. Vol. II: Time Series Analysis, In Memory of E. J. Hannan. Springer, New York, pp. 220–232.

Hosoya, Y., 1997. A limit theory for long-range dependence and statistical inference on related models. The Annals of Statistics 25, 105–137.

Hurvich, C., Deo, R., Brodsky, J., 1998. The mean squared error of Geweke and Porter-Hudak's estimator of the memory parameter of a long memory time series. Journal of Time Series Analysis 19, 19–46.

Hurvich, C. M., Beltrao, K., 1993. Asymptotics for the low-frequency ordinates of the periodogram of a long-memory time series. Journal of Time Series Analysis 14, 455–472.

Hurvich, C. M., Moulines, E., Soulier, P., 2005. Estimating long memory in volatility. Econometrica 73, 1283–1328.

Hurvich, C. M., Ray, B. K., 1995. Estimation of the memory parameter for nonstationary of noninvertible fractionally integrated processes. Journal of Time Series Analysis 16, 17–41.

Iacone, F., 2010. Local Whittle estimation of the memory parameter in presence of deterministic components. Journal of Time Series Analysis 31, 37–49.

Kapetanios, G., Papailias, F., 2011. Block bootstrap and long memory, Queen Mary University of London, School of Economics and Finance Working Paper No. 679.

Künsch, H., 1986. Discriminating between monotonic trends and long-range dependence. Journal of Applied Probability 23, 1025–1030.

Künsch, H., 1987. Statistical aspects of self-similar processes. In: Prohorov, Y., Sazarov, V. (Eds.), Proceedings of the First World Congress of the Bernoulli Society. Vol. 1. VNU Science Press, Utrecht, pp. 67–74.

McCloskey, A., 2010. Parameter estimation robust to low-frequency contamination with applications to ARMA, GARCH and stochastic volatility models, Unpublished Manuscript, Department of Economics, Boston University.

McCloskey, A., Perron, P., 2010. Memory parameter estimation in the presence of level shifts and deterministic trends, Unpublished Manuscript, Department of Economics, Boston University.

Mikosch, T., Stărică, C., 2004. Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects. Review of Economics and Statistics 86, 378–390.

Pérez, A., Ruiz, E., 2001. Finite sample properties of a qml estimator of stochastic volatility models with long memory. Economics Letters 70, 157–164.

Perron, P., 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.

Poskitt, D. S., 2007. Properties of the sieve bootstrap for fractionally integrated and non-invertible processes. Journal of Time Series Analysis, 224–250.

Qu, Z., 2011. A test against spurious long memory. Journal of Business and Economic Statistics 29, 423–438.

Robinson, P., 1995. Log-periodogram regression of time series with long range dependence. The Annals of Statistics 23, 1048–1072.

Robinson, P., 1997. Large sample inference for nonparametric regression with dependent errors. Annals of Statistics 25, 2054–2083.

Sun, Y., Phillips, P., 2003. Nonlinear log-periodogram regression for perturbed fractional processes. Journal of Econometrics 115, 355–389.

Velasco, C., 1999. Non-stationary log-periodogram regression. Journal of Econometrics 91, 325–371.

Zygmund, A., 1977. Trigonometric Series. Cambridge University Press.

Table 1: Calibrated ARSV Process with RLS ($p = 10$)

Trimming	$T; \hat{d}$			$T; \hat{a}$			$T; \hat{\sigma}^2_\eta$			$T; \hat{\sigma}^2_\varepsilon$		
	4000	8000	16000	4000	8000	16000	4000	8000	16000	4000	8000	16000
Bias												
0	0.486	0.452	0.427	-0.450	-0.436	-0.414	0.000	-0.006	-0.011	0.053	0.059	0.062
0.2	0.357	0.317	0.309	-0.316	-0.278	-0.279	0.078	0.036	0.024	-0.052	-0.020	-0.012
0.3	0.238	0.220	0.204	-0.213	-0.177	-0.163	0.093	0.033	0.018	-0.072	-0.026	-0.019
0.35	0.184	0.156	0.155	-0.170	-0.129	-0.120	0.095	0.038	0.018	-0.074	-0.029	-0.020
0.4	0.166	0.126	0.108	-0.170	-0.112	-0.087	0.108	0.048	0.022	-0.079	-0.037	-0.018
0.45	0.158	0.108	0.073	-0.175	-0.111	-0.059	0.120	0.063	0.016	-0.087	-0.048	-0.013
0.51	0.180	0.134	0.077	-0.181	-0.146	-0.069	0.117	0.076	0.024	-0.087	-0.057	-0.019
0.6	0.269	0.253	0.207	-0.189	-0.219	-0.180	0.137	0.071	0.037	-0.113	-0.052	-0.027
RMSE												
0	0.517	0.476	0.447	0.500	0.477	0.455	0.189	0.138	0.116	0.318	0.262	0.241
0.2	0.386	0.340	0.330	0.382	0.329	0.321	0.296	0.186	0.149	0.350	0.260	0.226
0.3	0.296	0.257	0.235	0.316	0.248	0.221	0.332	0.208	0.171	0.366	0.271	0.230
0.35	0.262	0.215	0.195	0.290	0.218	0.190	0.332	0.209	0.179	0.362	0.271	0.231
0.4	0.256	0.200	0.161	0.293	0.205	0.164	0.312	0.199	0.157	0.359	0.265	0.224
0.45	0.270	0.204	0.148	0.307	0.221	0.134	0.315	0.200	0.106	0.358	0.265	0.207
0.51	0.297	0.248	0.173	0.311	0.269	0.156	0.343	0.198	0.109	0.380	0.264	0.206
0.6	0.381	0.362	0.312	0.332	0.343	0.290	0.508	0.235	0.148	0.515	0.281	0.219

Table 2: Calibrated ARSV Process

Trimming	$T;\ \hat{d}$			$T;\ \hat{a}$			$T;\ \hat{\sigma}^2_\eta$			$T;\ \hat{\sigma}^2_\varepsilon$		
	4000	8000	16000	4000	8000	16000	4000	8000	16000	4000	8000	16000
Bias												
0	-0.007	-0.009	-0.007	-0.030	-0.013	-0.006	0.083	0.040	0.024	-0.065	-0.031	-0.020
0.2	0.013	-0.005	-0.007	-0.063	-0.025	-0.007	0.117	0.058	0.025	-0.094	-0.045	-0.021
0.3	0.028	0.001	-0.003	-0.080	-0.033	-0.019	0.119	0.060	0.040	-0.095	-0.047	-0.033
0.35	0.039	0.019	-0.002	-0.087	-0.059	-0.017	0.116	0.085	0.034	-0.092	-0.067	-0.028
0.4	0.061	0.023	0.006	-0.102	-0.048	-0.023	0.106	0.055	0.030	-0.094	-0.049	-0.027
0.45	0.099	0.047	0.015	-0.137	-0.073	-0.032	0.113	0.064	0.031	-0.098	-0.056	-0.028
0.51	0.172	0.099	0.045	-0.181	-0.115	-0.055	0.105	0.066	0.032	-0.092	-0.057	-0.029
0.6	0.278	0.260	0.204	-0.175	-0.206	-0.177	0.085	0.044	0.032	-0.083	-0.038	-0.027
RMSE												
0	0.090	0.073	0.056	0.141	0.098	0.081	0.289	0.184	0.143	0.350	0.263	0.220
0.2	0.115	0.089	0.063	0.192	0.132	0.083	0.343	0.214	0.140	0.381	0.275	0.222
0.3	0.140	0.102	0.078	0.209	0.142	0.116	0.326	0.205	0.176	0.374	0.269	0.237
0.35	0.154	0.124	0.084	0.215	0.184	0.108	0.314	0.258	0.142	0.363	0.299	0.220
0.4	0.181	0.130	0.097	0.226	0.151	0.113	0.285	0.199	0.125	0.355	0.261	0.208
0.45	0.222	0.162	0.116	0.268	0.188	0.125	0.294	0.210	0.119	0.357	0.266	0.209
0.51	0.291	0.221	0.155	0.309	0.240	0.151	0.301	0.212	0.123	0.363	0.271	0.210
0.6	0.396	0.368	0.312	0.309	0.328	0.292	0.397	0.222	0.157	0.436	0.274	0.222

Table 3: Calibrated ARLMSV Process with RLS ($p = 10$)

Trimming	$T;\hat{d}$			$T;\hat{a}$			$T;\hat{\sigma}^2_\eta$			$T;\hat{\sigma}^2_\varepsilon$		
	4000	8000	16000	4000	8000	16000	4000	8000	16000	4000	8000	16000
Bias												
0	0.142	0.142	0.101	-0.028	-0.060	-0.041	-0.103	-0.104	-0.091	0.137	0.137	0.125
0.2	0.051	0.063	0.056	0.029	-0.035	-0.061	-0.030	-0.019	-0.007	0.054	0.040	0.029
0.3	0.008	0.029	0.029	0.029	-0.006	-0.043	0.031	0.012	0.031	-0.009	0.005	-0.013
0.35	-0.044	0.004	0.015	0.074	0.008	-0.028	0.046	0.028	0.038	-0.025	-0.012	-0.021
0.4	-0.071	-0.038	-0.004	0.083	0.048	-0.010	0.072	0.038	0.045	-0.050	-0.022	-0.028
0.45	-0.121	-0.078	-0.032	0.116	0.064	0.018	0.089	0.062	0.048	-0.068	-0.043	-0.032
0.51	-0.203	-0.123	-0.072	0.211	0.098	0.045	0.124	0.075	0.054	-0.104	-0.058	-0.039
0.6	-0.187	-0.209	-0.192	0.227	0.174	0.134	0.239	0.132	0.073	-0.228	-0.117	-0.059
RMSE												
0	0.258	0.222	0.185	0.290	0.255	0.255	0.165	0.142	0.132	0.320	0.277	0.253
0.2	0.214	0.167	0.135	0.291	0.238	0.206	0.203	0.149	0.123	0.318	0.254	0.219
0.3	0.193	0.138	0.094	0.285	0.241	0.198	0.264	0.180	0.157	0.339	0.259	0.224
0.35	0.222	0.144	0.087	0.308	0.249	0.202	0.301	0.195	0.167	0.362	0.264	0.226
0.4	0.240	0.177	0.103	0.311	0.274	0.216	0.338	0.229	0.182	0.383	0.281	0.233
0.45	0.279	0.211	0.131	0.323	0.290	0.235	0.374	0.255	0.198	0.410	0.295	0.243
0.51	0.312	0.262	0.182	0.366	0.303	0.255	0.495	0.302	0.202	0.512	0.330	0.246
0.6	0.342	0.315	0.294	0.368	0.327	0.303	0.748	0.455	0.236	0.750	0.464	0.269

Table 4: Calibrated ARLMSV Process

Trimming	$T; \hat{d}$ 4000	8000	16000	$T; \hat{a}$ 4000	8000	16000	$T; \hat{\sigma}_\eta^2$ 4000	8000	16000	$T; \hat{\sigma}_\varepsilon^2$ 4000	8000	16000
Bias												
0	-0.056	-0.036	-0.021	0.096	0.052	0.017	0.028	0.031	0.036	-0.015	-0.021	-0.029
0.2	-0.080	-0.036	-0.019	0.100	0.040	0.007	0.047	0.052	0.047	-0.034	-0.042	-0.039
0.3	-0.103	-0.052	-0.028	0.095	0.052	0.023	0.072	0.051	0.043	-0.059	-0.042	-0.037
0.35	-0.123	-0.070	-0.030	0.118	0.061	0.020	0.076	0.055	0.045	-0.064	-0.046	-0.040
0.4	-0.128	-0.093	-0.038	0.109	0.083	0.023	0.098	0.054	0.049	-0.079	-0.039	-0.034
0.45	-0.150	-0.114	-0.056	0.113	0.089	0.039	0.121	0.065	0.048	-0.101	-0.050	-0.034
0.51	-0.231	-0.148	-0.086	0.220	0.108	0.052	0.131	0.084	0.059	-0.114	-0.069	-0.044
0.6	-0.190	-0.212	-0.195	0.221	0.172	0.133	0.263	0.132	0.074	-0.256	-0.120	-0.059
RMSE												
0	0.138	0.108	0.083	0.296	0.258	0.217	0.265	0.216	0.184	0.359	0.285	0.240
0.2	0.175	0.097	0.073	0.307	0.249	0.211	0.288	0.237	0.191	0.368	0.297	0.241
0.3	0.216	0.127	0.089	0.311	0.260	0.221	0.329	0.238	0.198	0.394	0.296	0.245
0.35	0.236	0.159	0.091	0.320	0.275	0.222	0.347	0.244	0.195	0.410	0.299	0.245
0.4	0.257	0.197	0.109	0.317	0.290	0.227	0.391	0.256	0.195	0.421	0.297	0.240
0.45	0.282	0.227	0.143	0.318	0.296	0.242	0.432	0.268	0.203	0.453	0.306	0.245
0.51	0.324	0.271	0.188	0.368	0.308	0.258	0.509	0.314	0.209	0.528	0.339	0.250
0.6	0.344	0.315	0.295	0.364	0.323	0.303	0.813	0.461	0.238	0.814	0.468	0.276