Econometric Theory, **31**, 2015, 1382–1402.
doi:10.1017/S0266466614000735

SIGNAL EXTRACTION IN LONG MEMORY STOCHASTIC VOLATILITY

JOSU ARTECHE
University of the Basque Country UPV/EHU

Long memory in stochastic volatility (LMSV) models are flexible tools for the modeling of persistent dynamic volatility, which is a typical characteristic of financial time series. However, their empirical applicability is limited because of the complications inherent in the estimation of the model and in the extraction of the volatility component. This paper proposes a new technique for volatility extraction, based on a semiparametric version of the optimal Wiener–Kolmogorov filter in the frequency domain. Its main characteristics are its simplicity and generality, because no parametric specification is needed for the volatility component and it remains valid for both stationary and nonstationary signals. The applicability of the proposal is shown in a Monte Carlo and in a daily series of returns from the Dow Jones Industrial index.

1. INTRODUCTION

The modeling of long memory in volatility has sparked great interest since the last few years of the 20th century, because of the strong persistence found in proxies of the volatility of many financial series. These efforts have mainly concentrated on two approaches: Stochastic volatility (SV) models where volatility and levels are driven by two different series of innovations, and Autoregressive Conditional Heteroskedasticity (ARCH) based models where only one innovation drives both levels and volatility. Measures of volatility are simple to obtain in ARCH based models because the conditional variances are exact functions of past observations, but estimating volatility in long memory SV (LMSV) models is far more complicated. Based on a previous work by Hidalgo and Yajima (2002), we propose a Wiener–Kolmogorov filter in the frequency domain which is optimal in a MSE (mean square error) sense, simple to implement, robust to misspecification (since only a partial specification of the (pseudo) spectral density function of the volatility component is required while the short memory behavior is left unrestricted) and generally applicable, since it is valid with stationary and nonstationary (but mean reverting) volatilities. It requires an estimator of the (pseudo) spectral density function of the series observed that needs to be consistent for the whole band

Research supported by the Spanish Ministry of Science and Innovation and ERDF grants ECO2010-15332 and ECO2013-40935-P, UPV/EHU Econometrics Research Group, Basque Government grant IT-642-13 and UPV/EHU UFI 11/03 Sustainable Economics & Welfare. The author is especially grateful to the referees and the Co-editor for their detailed comments and numerous suggestions, which helped to improve the paper. Address correspondence to Josu Arteche, Dept. of Econometrics and Statistics, University of the Basque Country UPV/EHU, Bilbao 48015, Spain; e-mail: josu.arteche@ehu.es.

 © Cambridge University Press 2014

https://doi.org/10.1017/S0266466614000735 Published online by Cambridge University Press

of Fourier frequencies in stationary and nonstationary contexts. To that end we use a proposal by Hidalgo and Yajima (2002) and show its consistency at every Fourier frequency between zero and π, no matter how close to or far from the spectral pole it is. This is a significant improvement on other, more traditional estimators, which lack consistency at frequencies close to the origin (Velasco, 2003).

The rest of the paper is structured as follows. Section 2 describes the type of models that we are dealing with. Section 3 describes the optimal Wiener–Kolmogorov filter in the frequency domain and Section 4 proposes a feasible plug-in version based on replacing the unknowns in the weights of the filter by consistent estimates. Section 5 concludes. The technical details are relegated to the Appendix.

A Monte Carlo analysis comparing the performance of our proposal with versions of the Kalman filter and Harvey's (1998) strategy adapted to our semiparametric setup is included in the supplementary material associated with this article, which is available at Cambridge Journals Online (journals.cambridge.org/ect). It also includes an illustrative application in a daily series of returns from the Dow Jones Industrial index.

2. LONG MEMORY IN STOCHASTIC VOLATILITY (LMSV)

We start by describing the theoretical framework in which we are working. Our interest is focused on zero mean LMSV processes defined as

$$z_t = \sigma \sigma_t \epsilon_t, \tag{1}$$

for σ a positive constant scale factor, ϵ_t are $iid(0,1)$ and $\sigma_t = \exp(x_t/2)$, with the volatility component x_t being a long memory process that satisfies the following assumptions.

A.1: Let d_0 denote the memory parameter of x_t. If $0 < d_0 < 1/2$, then $x_t = v_t$ and for $1/2 \leq d_0 < 1$, then $x_t = x_0 + \sum_{s=1}^{t} v_s$, where x_0 is a random variable not depending on t and $v_t = \sum_{j=0}^{\infty} b_j \varepsilon_{t-j}$, $\sum_{j=0}^{\infty} b_j^2 < \infty$ where $E(\varepsilon_t | F_{t-1}) = 0$, $E(\varepsilon_t^2 | F_{t-1}) = 1$, $E(\varepsilon_t^3) < \infty$, $E(\varepsilon_t^4) < \infty$ where F_{t-1} is the σ-field of events generated by ε_s, $s \leq t-1$.

A.2: The spectral density of v_t is

$$f_v(\lambda) = \lambda^{-2d_v} g_v(\lambda), \quad 0 < \lambda \leq \pi,$$

where $0 < d_v < 1/2$ in the stationary case and $-1/2 \leq d_v < 0$ for a nonstationary x_t and $g_v(\lambda)$ is a function which is positive, finite, symmetric around the origin and twice continuously differentiable.

Assumption A.1 avoids the restriction of Gaussianity and only imposes linearity of v_t with bounded fourth moments of the innovations. The nonstationary case is included as in Velasco (1999a, 1999b) and Hurvich, Moulines, and Soulier (2005) as Type I long memory. The pseudospectral density function of x_t is then $f_x(\lambda) = |1 - \exp(i\lambda)|^{-2} f_v(\lambda)$ such that in both the stationary and nonstationary cases the spectral or pseudospectral density function of the signal x_t satisfies

https://doi.org/10.1017/S0266466614000735 Published online by Cambridge University Press

$f_x(\lambda) = g_v(0)\lambda^{-2d_0}(1 + O(\lambda^2))$ as $\lambda \to 0$ for $d_0 = d_v$ in the stationary case and $d_0 = d_v + 1$ if x_t is nonstationary, entailing $d_0 \in (0, 1)$. Assumption A.2 constrains the possibility of seasonal or cyclical long memory and only allows for standard long memory at frequency zero. The analysis could be extended to cover other types of long memory where the spectral density diverges at a positive frequency as in Arteche and Robinson (2000) but is constrained here to the empirically more popular case of standard long memory with positive d_0.

These models, with x_t being a stationary fractional ARIMA, were originally proposed by Breidt, Crato, and de Lima (1998) and Harvey (1998). We generalize their original definition in two ways: First, we avoid parametric restrictions by only imposing a local behavior on the spectral density function of the volatility component around the origin, and second we allow for the possibility of a nonstationary x_t.

The statistical properties of this stochastic volatility class of models are easier to determine than those of the different extensions of ARCH, and they permit a wide range of characteristics typical of financial time series to be modeled. For example, covariance stationarity is guaranteed as long as $d < 1/2$, a positive d entailing long memory in the squares, a characteristic ruled out in ARCH based models (Giraitis, Kokoska, and Leipus, 2000; Zaffaroni, 2004). Also, in contrast to ARCH-type models, the two different innovations allow for more flexible modeling and some statistical characteristic of the series, such as excess kurtosis, can be easily adapted without affecting the memory structure of the volatility.

The nonlinear LMSV can be linearized by the transformation $y_t = \log z_t^2$, such that

$$y_t = \mu + x_t + u_t, \tag{2}$$

where $\mu = \log\sigma^2 + E\log\epsilon_t^2$ and $u_t = \log\epsilon_t^2 - E\log\epsilon_t^2$ is a mean zero white noise process with finite variance σ_u^2. The estimation of the volatility component x_t is then a particular case of signal extraction in a long memory signal plus white noise process. For that purpose the following assumptions are also considered:

A.3: u_t is mean zero i.i.d. (independent and identically distributed) with finite fourth moment.

A.4: u_t and ε_s are uncorrelated at all leads and lags.

A.5: For a finite constant k, $cum(\varepsilon_t, \varepsilon_s, u_l, u_m) = k$ if $t = s = l = m$ and 0 otherwise.

Assumption A.3 restricts the distributional properties of the added noise as in the traditional literature on stochastic volatility. It is satisfied for a Gaussian ϵ_t but Gaussianity is not necessary and any degree of finite kurtosis is allowed for. The absence of serial correlation in u_t is needed to correctly identify its spectral density over the whole Fourier frequency band $(0, \pi]$, since in that case it is constant and can be consistently estimated with the information given by the observables y_t. Assumption A.4 entails a null cross-spectral density function, which has been found to be necessary to identify the weights of the filter.

https://doi.org/10.1017/S0266466614000735 Published online by Cambridge University Press

Note that this does not preclude the existence of leverage in the form of correlation between ϵ and x, which may be allowed on a noncontemporaneous basis to maintain the martingale difference character of z_t. In fact, as long as the joint distribution of ϵ and x is symmetric around the origin (e.g., Gaussian, Student's t, or a generalized error distribution among others) the possible correlation between the two does not preclude the fulfillment of hypothesis A.4 (see Harvey, Ruiz, and Shephard, 1994). Higher order dependencies between x and u are also allowed. They are only limited by Assumption A.5, which is also imposed in Assumption H.5 of Hurvich et al. (2005). Of course, Assumptions A.4 and A.5 with $k = 0$ hold for independent signal and noise.

3. OPTIMAL SIGNAL EXTRACTION IN THE FREQUENCY DOMAIN

Optimal (in a MSE sense) estimation of x_t can be achieved by means of the linear Wiener–Kolmogorov filter applied to the observables y_t. This strategy was proposed by Harvey (1998) in fully parametric LMSV models with the weights of the filter obtained in the time domain as functions of variances and covariances. In addition to requiring a complete parametric specification of x_t, the implementation of this filter suffers from other drawbacks. From a theoretical perspective, its main restriction is that it is only valid for a stationary x_t. From an applied point of view it suffers from the disadvantage of being computationally intensive because it requires the inversion of the variance covariance matrix of the observables, which can be of large dimension with long series such as those found in finance (for further details see the supplementary material).

As an alternative, the optimal linear estimator of x_t based on Wiener–Kolmogorov filter in the frequency domain is defined as

$$\tilde{x}_{t|\infty} = \sum_{j=-\infty}^{\infty} \psi_j (y_{t-j} - \mu), \tag{3}$$

where, with $f_{xy}(\lambda)$ denoting the cross-spectral density function of x and y, the weights are

$$\psi_j = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{f_{xy}(\lambda)}{f_y(\lambda)} e^{ij\lambda} \mathrm{d}\lambda \tag{4}$$

which under Assumption A.4 become

$$\psi_j = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{f_x(\lambda)}{f_y(\lambda)} e^{ij\lambda} \mathrm{d}\lambda \tag{5}$$

because $f_{xu}(\lambda) = 0$. The filter proposed in (3) and (5) is obviously symmetric avoiding time shifts, which is a desirable property because the signal and observables are synchronized. In contrast to Harvey's method, the ψ_j in (5) are computationally simple to obtain even for large sample sizes and remain valid

https://doi.org/10.1017/S0266466614000735 Published online by Cambridge University Press

for nonstationary series ($1/2 \leq d < 1$) where pseudospectral density functions are then considered. This is justified by the fact that a stationary environment is achieved after first differencing, and first differences of the signal can be optimally extracted with the filter

$$\psi_j = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{f_{\Delta x}(\lambda)}{f_{\Delta y}(\lambda)} e^{ij\lambda} \mathrm{d}\lambda = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{f_x(\lambda)}{f_y(\lambda)} e^{ij\lambda} \mathrm{d}\lambda$$

with f_x and f_y pseudospectral density functions based on transfer functions (see also Bell, 1984).

Hidalgo and Yajima (2002) propose a semiparametric version of ψ_j in the case of a stationary ARFIMA(0,d,0) x_t and u_t linear weak dependent with no parametric restrictions. However, we prefer to impose the parametric restriction on u_t, which is white noise in SV models, leaving the volatility component only partially specified. In that sense, the framework described in the previous section is more general because it does not parametrically restrict the volatility component x_t, but imposes only the local requirement in Assumption A.2 and it allows for a nonstationary x_t. By contrast the noise u_t, not restricted in Hidalgo and Yajima (2002), is assumed to be white noise, which in any case is a natural restriction in SV models. The spectral density function of u_t is then constant $f_u(\lambda) = \sigma_u^2/2\pi = \theta$ and the weights ψ_j in (5) can be written in this context as

$$\psi_j = \frac{1}{2\pi}\int_{-\pi}^{\pi} \left(1 - \frac{\theta}{f_y(\lambda)}\right) e^{ij\lambda} \mathrm{d}\lambda = \mathbf{1}_{j=0} - \frac{1}{\pi}\int_{0}^{\pi} \frac{\theta}{f_y(\lambda)} \cos(j\lambda)\mathrm{d}\lambda, \tag{6}$$

where the indicator function $\mathbf{1}_A = 1$ if A holds and 0 otherwise, and the last equality is due to spectral symmetry.

4. SEMIPARAMETRIC ESTIMATION OF VOLATILITY

The feasible version of the ψ_j in (6) is achieved after two transformations: the integral is replaced by summation over Fourier frequencies and consistent estimators of the unknowns θ and f_y are plugged into ψ_j. With these modifications the feasible version of ψ_j is $\hat{\psi}_j = \hat{\psi}_j(\hat{\theta}, \hat{f}_y)$ where

$$\hat{\psi}_j(\hat{\theta}, \hat{f}_y) = \mathbf{1}_{j=0} - \frac{1}{n^*}\sum_{p=1}^{n^*} \frac{\hat{\theta}}{\hat{f}_y(\lambda_p)} \cos(j\lambda_p) \tag{7}$$

for $n^* = [n/2]$, [] denoting "the integer part of", n is the sample size such that one observes data $y_1, y_2, ..., y_n$ and $\lambda_p = 2\pi p/n$ are Fourier frequencies. Assuming that such consistent estimators exist, the following result emerges.

THEOREM 1. *Let $\hat{\theta}$ be a consistent estimator of θ and $\hat{f}_y(\lambda_p)$ consistently estimate $f_y(\lambda_p)$ uniformly over $p = 1, ..., n^*$. Then as $n \to \infty$*

https://doi.org/10.1017/S0266466614000735 Published online by Cambridge University Press

$$\hat{\psi}_j = \tilde{\psi}_j(1 + o_p(1)) \tag{8}$$

$$\tilde{\psi}_j - \psi_j = O\left(\frac{1}{n}\right) \tag{9}$$

for $\tilde{\psi}_j = \hat{\psi}_j(\theta, f_y)$.

Note that what $\hat{\psi}_j$ actually estimates consistently is $\tilde{\psi}_j$ and not ψ_j. Theorem 1 however gives a bound of the difference between $\tilde{\psi}_j$ and ψ_j. Since ψ_j is proportional to $|j|^{-2d-1}$ then consistency of $\hat{\psi}_j$ as an estimator of ψ_j is only rigorously proved for j satisfying $|j|^{2d+1}/n \to 0$ as $n \to \infty$. This implies that not all the observations can be used to get the estimated signal, only observations not too far from the point of estimation should be used.

With all these considerations, the final estimator of the volatility component is

$$\hat{x}_{t|n} = \sum_{j=\max(t-n,-M)}^{\min(t-1,M)} \hat{\psi}_j(y_{t-j} - \hat{\mu}) \tag{10}$$

for M satisfying $M^{-1} + n^{-1}M^{2d+1} \to 0$ as $n \to \infty$. Restricting the inclusion of values of the observable data to a maximum of M lags has two effects. First, consistency of the $\hat{\psi}_j$ used in the filter is guaranteed. Second, relatively low values of M enable a symmetric feasible filter to be applied for a large number of intermediate observations, thus avoiding undesirable phase shifts. Note also that, since $\psi_j = O(j^{-1-2d})$, there is no need to consider weights for larger lags that are virtually zero.

An estimate of the constant μ is also needed for feasible signal extraction. Under stationarity the sample mean $\hat{\mu} = \bar{y} = \sum_{t=1}^{n} y_t/n$ is a good option due to its consistency. However, if $d \geq 1/2$ the sample mean is not a consistent estimator of μ. To solve this problem we follow Shimotsu (2010) and use as an estimate of μ the first observation $\hat{\mu} = y_1$ or the average of a finite number of initial observations, which is $O_p(1)$ under Assumption A.1. The validity of this procedure is based on the fact that the variance of a nonstationary x_t is infinite and consequently the signal asymptotically dominates the influence of the constant μ, such that an estimate of it is only needed in finite samples for location purposes.

Consistent estimation of the signal extraction weights ψ_j calls for consistent estimation of the constant spectral density of the noise θ and of $f_y(\lambda_p)$ uniformly over $p = 1, ..., n^*$. Under the assumptions in the previous section θ can be consistently estimated by means of the local Whittle estimator of Hurvich et al. (2005) (another alternative, requiring Gaussianity, is the augmented log periodogram regression (ALPE) of Arteche, 2006). The estimation of $f_y(\lambda_p)$ is more demanding since it needs to be consistent uniformly over $p = 1, ..., n^*$. This entails consistent estimation for constant frequencies ($p = O(n)$ meaning exact rate of proportionality) and Fourier frequencies collapsing to zero ($p = o(n)$) in stationary and nonstationary contexts. To that end we use the estimator proposed by Hidalgo and Yajima (2002) and show that it satisfies these requirements.

https://doi.org/10.1017/S0266466614000735 Published online by Cambridge University Press

At this point it is important to stress the relevance of Assumption A.4, which entails a null cross-spectral density of signal and noise. Under A.4 the weights ψ_j have the form in (6) and both θ and f_y can be consistently estimated. However, if Assumption A.4 is not satisfied the weights ψ_j are

$$\psi_j = \mathbf{1}_{j=0} - \frac{1}{\pi}\int_0^\pi \frac{\theta}{f_y(\lambda)}\cos(j\lambda)\mathrm{d}\lambda - \frac{1}{\pi}\int_0^\pi \frac{f_{ux}^R(\lambda)}{f_y(\lambda)}\cos(j\lambda)\mathrm{d}\lambda + \frac{1}{\pi}\int_0^\pi \frac{f_{ux}^I(\lambda)}{f_y(\lambda)}\sin(j\lambda)\mathrm{d}\lambda,$$

where f_{ux}^R and f_{ux}^I represent the real and imaginary parts of f_{ux}. Practical application of the filter requires estimation of both quantities over the whole band of Fourier frequencies with only the information provided by the observables y_t. But $f_y(\lambda) = f_x(\lambda) + \theta + 2f_{ux}^R$ and it contains no information on f_{ux}^I, precluding its estimation even under more restrictive assumptions than those required in this paper. Thus, we find that Assumption A.4 is necessary for a feasible plug-in application of the filter with weights ψ_j.

4.1. Consistent Estimation of θ

Various existing techniques can be used to estimate θ, such as the local Whittle estimation proposed by Hurvich et al. (2005) and the ALPE by Arteche (2006, 2012). The local Whittle is chosen here for its greater efficiency in an uncorrelated signal and noise context and because it does not require Gaussianity in either the signal or the noise for consistency and asymptotic normality.

Under conditions A.1–A.4 the (pseudo) spectral density function of the observables y_t behaves as $\lambda \to 0$ as

$$f_y(\lambda) = f_x(\lambda) + \theta = g_v(0)\lambda^{-2d_0}(1 + \theta_{10}\lambda^{2d_0} + O(\lambda^2))$$

for $\theta_{10} = \theta/g_v(0)$ as the long-run noise to signal ratio. This expression motivates the local Whittle estimation of θ_{10} (together with d_0) defined as

$$(\hat{d}, \hat{\theta}_1) = \arg\min_{(d,\theta_1)\in\Delta\times\Theta}\left\{\log\hat{C}(d,\theta_1) + \frac{1}{m}\sum_{k=1}^m \log(\lambda_k^{-2d}(1+\theta_1\lambda_k^{2d}))\right\}$$

for Fourier frequencies $\lambda_k = 2\pi k/n$, $\Delta = [\Delta_1, \Delta_2]$, $0 < \Delta_1 < \Delta_2 < 1$, $\Theta = (0, \Theta_0)$, $\Theta_0 < \infty$ and

$$\hat{C}(d,\theta_1) = \frac{1}{m}\sum_{k=1}^m \frac{\lambda_k^{2d} I_y(\lambda_k)}{1+\theta_1\lambda_k^{2d}},$$

where the periodogram of y_t, $t = 1, 2, ..., n$ is

$$I_y(\lambda) = |W_y(\lambda)|^2 \quad W_y(\lambda) = \frac{1}{\sqrt{2\pi n}}\sum_{t=1}^n y_t \exp(-it\lambda)$$

https://doi.org/10.1017/S0266466614000735 Published online by Cambridge University Press

with $W_y(\lambda)$ being the discrete Fourier transform, and m the bandwidth parameter that indicates the number of Fourier frequencies used in the estimation. This bandwidth is restricted to satisfy the following assumption:

A.6: As $n \to \infty$

$$\frac{n^{4d_0}}{m^{1+4d_0}} + \frac{m^5}{n^4}\log^2 m \to 0.$$

Consistency of $\hat{d}$ requires only $m = o(n)$. Assumption A.6 is more restrictive to avoid asymptotic flatness of the objective function as a function of θ_1 and to permit consistent estimation of θ_{10}. Under these conditions θ can be estimated as

$$\hat{\theta} = \hat{\theta}_1 \hat{C}(\hat{d}, \hat{\theta}_1).$$

PROPOSITION 1. *Under* **A.1–A.6**, *as* $n \to \infty$

$$\hat{\theta} = \theta + o_p(1).$$

Proof. By Hurvich et al. (2005) $\hat{\theta}_1 - \theta_{10} = o_p(1)$ and $\hat{C}(\hat{d}, \hat{\theta}_1)$ estimates consistently $g_\upsilon(0)$. ■

4.2. Consistent Estimation of the (Pseudo) Spectral Density f_y

The estimation of $f_y(\lambda_p)$ is more demanding because it needs to be uniformly consistent over $p = 1, ..., n^*$, which implies consistency for fixed frequencies and for all Fourier frequencies shrinking to zero, including those around the spectral pole. The main problem here is to get consistency at frequencies close to zero due to the divergence of the spectral density at the origin (see Velasco, 2003). Moreover, the estimator has to be consistent even in the nonstationary case, where $f_y(\lambda_p)$ is no longer a spectral density function. With all these considerations, we propose the use of the frequency domain estimator introduced by Hidalgo and Yajima (2002) defined as

$$\hat{f}_y(\lambda_\upsilon) = \frac{|\lambda_\upsilon|^{-2\hat{d}}}{2m^* + \mathbf{1}_{\upsilon > m^*}} \sum_{j=-m^*, \neq -\upsilon}^{m^*} |\lambda_\upsilon + \lambda_j|^{2\hat{d}} I_y(\lambda_\upsilon + \lambda_j), \tag{11}$$

where $\lambda_\upsilon = 2\pi\upsilon/n$, $\upsilon = 1, ..., n^*$, $\hat{d}$ is a consistent estimator of d_0 satisfying

A.7: $(\hat{d} - d_0) = o_p(\log^{-1} m^*)$ as $n \to \infty$,

and m^* is a bandwidth number such that

A.8: as $n \to \infty$

$$\frac{1}{m^*} + \frac{m^*}{n} \to 0.$$

In this expression, the periodogram is damped around frequency zero with the term $|\lambda_\upsilon + \lambda_j|^{2\hat{d}}$ prior to the usual averaging, which can be considered as

https://doi.org/10.1017/S0266466614000735 Published online by Cambridge University Press

prewhitening in the frequency domain since it neutralizes the spectral pole at the origin and makes $\hat{f}_y$ consistent at all Fourier frequencies between zero and π after recoloring by multiplying by $|\lambda_\upsilon|^{-2\hat{d}}$. Note also that the periodogram at frequency zero is not used in the construction of $\hat{f}_y(\lambda_\upsilon)$ due to the spectral divergency at the origin.

Hidalgo and Yajima (2002) proposed this estimator for stationary long memory processes ($d \in (0, 1/2)$) and $\hat{f}_y$ evaluated at Fourier frequencies λ_{2mj}, such that frequencies close to the origin are not covered because $m \to \infty$. Under these conditions consistency of $\hat{f}_y$ can be inferred from the results in their Lemmas 1–3. We extend these results and show in the theorem below that $\hat{f}_y$ is consistent at any Fourier frequency λ_υ for $\upsilon = 1, ..., n^*$, no matter how close to or far from the spectral pole it is, also covering the perturbed long memory case and allowing for nonstationary series as long as $d \in (0, 1)$.

THEOREM 2. *Under Assumptions A.1–A.5, A.7, and A.8 as* $n \to \infty$,

$$\hat{f}_y(\lambda_\upsilon) = f_y(\lambda_\upsilon)\left(1 + o_p(1)\right)$$

uniformly for $\upsilon = 1, ..., n^*$.

Theorem 2 establishes the consistency of $\hat{f}_y(\lambda_\upsilon)$ for $\lambda_\upsilon = O(1)$ and $\lambda_\upsilon = o(1)$. For $\upsilon = o(n)$ such that $\lambda_\upsilon \to 0$ it holds that $f_y(\lambda_\upsilon) = f_x(\lambda_\upsilon)(1 + o(1))$ and $\hat{f}_y(\lambda_\upsilon)$ basically estimates $f_x(\lambda_\upsilon)$ since $f_u(\lambda_\upsilon)$ is of a lower order. However, if $\lambda_\upsilon > 0$ is $O(1)$ (implying $\upsilon = O(n)$), then $f_x(\lambda_\upsilon)$ and $f_u(\lambda_\upsilon)$ are both $O(1)$ and Theorem 2 shows that $f_y(\lambda_\upsilon) = f_x(\lambda_\upsilon) + f_u(\lambda_\upsilon)$ is estimated consistently by $\hat{f}_y(\lambda_\upsilon)$. For the estimation of d the local Whittle estimator is suggested, which has a $\sqrt{m}$-convergence for m as in A.6. Then, in view of Assumption A.7, both bandwidths m and m^* are required to satisfy the mild condition $m^{-1/2}\log m^* = o(1)$ as $n \to \infty$, which is naturally satisfied if both sequences are powers of n.

There are other ways of estimating the (pseudo) spectral density function in a long memory context that do not rely on prior estimation of d, such as the kernel based proposal by Velasco (2003). Its consistency in our signal-plus-noise setting can be easily proved for frequencies sufficiently far from the origin, but it lacks consistency at frequencies around the origin, which precludes the use of those frequencies in the construction of $\hat{\psi}_j$, when in fact those are precisely the frequencies which contain the most valuable information on the behavior of the signal. Thus, even in an unperturbed long memory series the consistency of $\hat{f}_y(\lambda_\upsilon)$ is an improvement over those techniques that lack consistency at frequencies close to the spectral pole.

4.3. Feasible Signal Extraction

With all the above considerations and results the frequency domain strategy for the estimation of the signal comprises the following steps:

https://doi.org/10.1017/S0266466614000735 Published online by Cambridge University Press

Step 1: Estimate d_0 and θ by local Whittle estimation. To that end a bandwidth m has to be selected. A simple and effective option is to choose m from among those bandwidths that lead to a stable region of estimates of d_0 and θ, as suggested by Taqqu and Teverovsky (1996).

Step 2: With the estimates in Step 1, use (11) to estimate $f_y(\lambda_p)$ for $p = 1, 2, ..., n^*$. As in other density estimation issues, the bandwidth m^* can be chosen based on the smoothness of the spectral density at frequencies far from the origin. The greater the smoothness the larger the m^* that should be selected. In practice we suggest trying different sizes of m^*, for example, large, small, and medium, and if the estimates with all these m^* do not show a significant structure at frequencies far from the origin a large m^* should be chosen. Otherwise, a low m^* better reflects a rough spectral density with a relevant short memory component.

Step 3: With the estimates in Steps 1 and 2, construct the weights $\hat{\psi}_j$ in (7). Choose M as the lowest value such that if $j > M$, then $|\hat{\psi}_j| < \eta$ for a prespecified small $\eta > 0$.

Step 4: With this M extract the signal as in (10) with $\hat{\mu}$ as the sample mean if the estimate of d_0 in Step 1 is lower than 0.5 and $\hat{\mu}$ as the average of a small number of initial observations otherwise.

Step 5: Validation. The suitability of the estimation of the signal obtained in Step 4 can be assessed by a validating strategy based on the original form of the LMSV model in (1). If the signal extraction is correct the standardized residuals, defined as $\hat{\varepsilon}_{t|n} = z_t \exp(-\hat{x}_{t|n}/2)$, should be close to $i.i.d$ (not necessarily normal). In particular, their squares should be uncorrelated and a portmanteau test in the squared standardized residuals can be used to check this property. Failure to satisfy this condition may indicate wrong signal extraction, due perhaps to a wrong choice of bandwidths, and the procedure returns to Step 1. If no improvement is achieved by changing the bandwidths, the local LMSV should be rejected as inappropriate for the series under consideration (e.g., due to correlation between signal and noise as in FIEGARCH models, to mistaking structural breaks for long memory, etc.). However, evidence of no autocorrelation in the squared standardized residuals supports the existence of LMSV against other alternatives that may give rise to strong persistent behavior in volatility, such as FIEGARCH or level shifts in volatility.

The conditional variances can then be estimated as

$$\hat{\sigma}_t^2 = \hat{\sigma}^2 \exp(\hat{x}_{t|n}),$$

where $\hat{\sigma}^2 = \sum_{t=1}^{n} z_t^2 \exp(-\hat{x}_{t|n})/n$ as suggested by Harvey (1998).

https://doi.org/10.1017/S0266466614000735 Published online by Cambridge University Press

5. CONCLUSION AND DIRECTIONS FOR FURTHER RESEARCH

One of the main reasons for the low level of interest of practitioners in LMSV models is the difficulty of extracting appropriate estimates of volatility. We propose a solution to this problem involving a feasible Wiener–Kolmogorov signal extraction technique based on a local spectral specification of the volatility component. We prove the consistency of the estimated weights in stationary and nonstationary (but mean reverting) long memory contexts. Its performance in finite samples is better in all cases than a feasible version of the time domain proposal by Harvey (1998), and in general better than smoothing via the Kalman filter based on an AR approximation (see the supplementary material associated with this article, available at Cambridge Journals Online (journals.cambridge.org/ect)). The validating step, based on the $i.i.d.$ character of the standardized residuals, can be used to assess the adequacy of LMSV and to discard other possibilities such as FIEGARCH (which implies correlation between signal and noise) or level shifts (which can be mistaken for long memory) in volatility.

Estimation of the weights of the filter requires consistent estimation of the (pseudo) spectral density function over the whole band of Fourier frequencies. To that end we use a new estimator originally proposed by Hidalgo and Yajima (2002) and show that it is consistent at fixed and shrinking-to-zero Fourier frequencies, no matter how close to or far from the spectral pole they are, as long as $0 < d < 1$. This is an improvement over other proposals, which lack consistency at frequencies close to the spectral pole.

Although our focus of interest is LMSV, the technique proposed can be used in a much more general context. There is a broad range of situations which adapt to a long memory signal plus noise model in economics and finance (see for example, Arteche, 2012). Just a few examples are economic mechanisms with short-run and long-run behavior depending on different factors, measurement errors, rational expectation models with a persistent ex ante variable, realized volatility contaminated by microstructure noise, etc. In all these cases, estimation of the signal is of great interest.

The main drawback of the technique proposed is that it is not fully automatic but depends on bandwidth parameters that have to be chosen by the user. However, we suggest some simple criteria for selecting them all based on the information provided by the sample, and in that sense they can be considered as data-driven (see the empirical application in the supplementary material as an illustrative example).

REFERENCES

Arteche, J. (2006) Semiparametric estimation in perturbed long memory series. *Computational Statistics and Data Analysis* 51, 2118–2141.

Arteche, J. (2012) Semiparametric inference in correlated long memory signal plus noise models. *Econometric Reviews* 31, 440–474.

https://doi.org/10.1017/S0266466614000735 Published online by Cambridge University Press

Arteche, J. & P.M. Robinson (2000) Semiparametric inference in seasonal and cyclical long memory processes. *Journal of Time Series Analysis* 21, 1–25.
Bell, W. (1984) Signal extraction for nonstationary time series. *Annals of Statistics* 12, 646–664.
Breidt, F.J., N. Crato, & P. de Lima (1998) The detection and estimation of long memory in stochastic volatility. *Journal of Econometrics* 83, 325–348.
Brillinger, D.R. (1975) *Time Series: Data Analysis and Theory*. Holt, Rinehart and Winston, Inc.
Brockwell, P.J. & R.A. Davies (2006) *Time Series: Theory and Methods*. Springer.
Giraitis, L., P. Kokoska, & R. Leipus (2000) Stationary ARCH models: Dependence structure and central limit theorem. *Econometric Theory* 16, 3–22.
Harvey, A.C. (1998) Long memory in stochastic volatility. In J. Knight & S. Satchell (eds.), *Forecasting Volatility in Financial Markets*, pp. 307–320. Butterworth-Heinemann.
Harvey, A.C., E. Ruiz, & N. Shephard (1994) Multivariate stochastic variance models. *Review of Economic Studies* 61, 247–264.
Hidalgo, J. & Y. Yajima (2002) Prediction and signal extraction of strongly dependent processes in the frequency domain. *Econometric Theory* 18, 584–624.
Hurvich, C.M., E. Moulines, & P. Soulier (2005) Estimating long memory in volatility. *Econometrica* 73, 1283–1328.
Robinson, P.M. (1995) Gaussian semiparametric estimation of long-range dependence. *Annals of Statistics* 23, 1630–1661.
Shimotsu, K. (2010) Exact local Whittle estimation of fractional integration with unknown mean and time trend. *Econometric Theory* 26, 501–540.
Taqqu, M.S. & V. Teverovsky (1996) *Semi-parametric Graphical Estimation Techniques for Long-Memory Data*. *Lecture Notes in Statistics*, vol. 115, pp. 420–432.
Velasco, C. (1999a) Non-stationary log-periodogram regression. *Journal of Econometrics* 91, 325–371.
Velasco, C. (1999b) Gaussian semiparametric estimation of non-stationary time series. *Journal of Time Series Analysis* 20, 87–127.
Velasco, C. (2003) Nonparametric frequency domain analysis of nonstationary multivariate time series. *Journal of Statistical Planning and Inference* 116, 209–247.
Zaffaroni, P. (2004) Stationarity and memory of ARCH(∞) models. *Econometric Theory* 20, 147–160.

APPENDIX

A. Technical Details

Proof of Theorem 1. (9) is proved using the approximation of integrals by sums in Brillinger (1975, p. 15). For (8) note that for $j \neq 0$ $\hat{\psi}_j - \tilde{\psi}_j$ is

$$\frac{\theta - \hat{\theta}}{\theta} \tilde{\psi}_j + \frac{\hat{\theta}}{n^*} \sum_{p=1}^{n^*} \frac{\hat{f}_y(\lambda_p) - f_y(\lambda_p)}{\hat{f}_y(\lambda_p)} \frac{\cos(j\lambda_p)}{f_y(\lambda_p)} = o_p(\tilde{\psi}_j)$$

by the consistency of $\hat{\theta}$ and $\hat{f}_y(\lambda_p)$ uniformly over $p = 1, ..., n^*$. For $j = 0$ it holds similarly that $\hat{\psi}_0 - \tilde{\psi}_0 = o_p(1)$, which completes the proof since $\tilde{\psi}_0 = O_p(1)$ as $n \to \infty$. ■

Proof of Theorem 2. Denote $\check{f}_y(\lambda_v)$ as $\hat{f}_y(\lambda_v)$ with d_0 instead of $\hat{d}$. Then

$$\hat{f}_y(\lambda_v) = \hat{f}_x(\lambda_v) + \hat{f}_u(\lambda_v) + \hat{f}_{xu}(\lambda_v) + \hat{f}_{ux}(\lambda_v)$$

https://doi.org/10.1017/S0266466614000735 Published online by Cambridge University Press

and similarly for $\check{f}_y(\lambda_v)$, where the subindex indicates the variable of which the periodogram or cross-periodogram is calculated. With this notation

$$\check{f}_x(\lambda_v) - f_x(\lambda_v) = \frac{\lambda_v^{-2d_0}}{2m^* + \mathbf{1}_{v>m^*}} \sum_{j=-m^*,\neq -v}^{m^*} |\lambda_v + \lambda_j|^{2d_0} \times \left(I_x(\lambda_v + \lambda_j) - f_x(\lambda_v + \lambda_j)\right) \quad \textbf{(A.1)}$$

$$+ \frac{\lambda_v^{-2d_0}}{2m^* + \mathbf{1}_{v>m^*}} \sum_{j=-m^*,\neq -v}^{m^*} |\lambda_v + \lambda_j|^{2d_0} \times f_x(\lambda_v + \lambda_j) - f_x(\lambda_v). \quad \textbf{(A.2)}$$

Now (A.2) is by Assumptions A.1 and A.2

$$\begin{aligned}
&\frac{\lambda_v^{-2d_0}}{2m^* + \mathbf{1}_{v>m^*}} \sum_{j=-m^*,\neq -v}^{m^*} g_x(\lambda_v + \lambda_j) - f_x(\lambda_v) \\
&\leq \frac{\lambda_v^{-2d_0}}{2m^* + \mathbf{1}_{v>m^*}} \sum_{j=-m^*,\neq -v}^{m^*} \left(g_x(\lambda_v) + \frac{\lambda_j^2}{2} \max_{\lambda_v \leq \lambda \leq \lambda_v + \lambda_j} |g_x''(\lambda)| \right) - f_x(\lambda_v) \\
&= \frac{\lambda_v^{-2d_0}}{2m^* + \mathbf{1}_{v>m^*}} g_x(\lambda_v) O\left(g_x^{-1}(\lambda_v) \sum_{j=-m^*,\neq -v}^{m^*} \lambda_j^2 \right) \\
&= O\left(f_x(\lambda_v) \left[\frac{m^*}{n} \right]^2 \right) = o\left(f_x(\lambda_v)\right),
\end{aligned}$$

where $g_x(\lambda) = g_v(\lambda)$ in the stationary case and $g_x(\lambda) = g_v(\lambda)|2\lambda^{-1}\sin(\lambda/2)|^{-2}$ in the nonstationary case, which also satisfies the requirements for g_v in Assumption A.2.

Also (A.1) is

$$\frac{\lambda_v^{-2d_0}}{2m^* + 1} \sum_{j=-m^*}^{m^*} |\lambda_v + \lambda_j|^{2d_0} \left(I_x(\lambda_v + \lambda_j) - |\alpha_{v+j}|^2 I_\varepsilon(\lambda_v + \lambda_j) \right) \quad \textbf{(A.3)}$$

$$+ \frac{\lambda_v^{-2d_0}}{2m^* + 1} \sum_{j=-m^*}^{m^*} |\lambda_v + \lambda_j|^{2d_0} \left(|\alpha_{v+j}|^2 I_\varepsilon(\lambda_v + \lambda_j) - f_x(\lambda_v + \lambda_j) \right) \quad \textbf{(A.4)}$$

for $\alpha_j = \alpha(\lambda_j)$ where $\alpha(\lambda) = \sum_{k=0}^{\infty} b_k \exp(ik\lambda)$ under stationarity and $\alpha(\lambda) = (1 - \exp(\lambda)) \sum_{k=0}^{\infty} b_k \exp(ik\lambda)$ in the nonstationary case.

The proof of the consistency requires a different treatment of frequencies close to and far from the origin. Hereafter the $O()$ and $o()$ terms are uniform over those v considered. Let us examine first those frequencies such that $m^*/v \to 0$ as n$\to \infty$ (including $v = O(n)$ corresponding to fixed frequencies λ_v). Using formulas (3.17) in Robinson (1995) and (A.2) in Velasco (1999b) (see also Velasco, 2003), (A.3) is bounded in probability by

https://doi.org/10.1017/S0266466614000735 Published online by Cambridge University Press

$$O_p\left(\frac{\lambda_\upsilon^{-2d_0}}{2m^*+1}\sum_{j=-m^*}^{m^*}|\lambda_\upsilon+\lambda_j|^{2d_0}f_x(\lambda_\upsilon+\lambda_j)\log^{1/2}|\upsilon+j|\left[|\upsilon+j|^{-1/2}+|\upsilon+j|^{d_0-1}\right]\right)$$
$$=O_p\left(\frac{\lambda_\upsilon^{-2d_0}}{2m^*+1}\sum_{j=-m^*}^{m^*}g_x(\lambda_\upsilon+\lambda_j)\log^{1/2}|\upsilon+j|\left[|\upsilon+j|^{-1/2}+|\upsilon+j|^{d_0-1}\right]\right)$$
$$=O_p\left(f_x(\lambda_\upsilon)\frac{\log^{1/2}(\upsilon+m^*)}{(\upsilon-m^*)^{1/2\wedge(1-d_0)}}\right)$$
$$=O_p\left(f_x(\lambda_\upsilon)\frac{(\log\upsilon+\log(1+m^*/\upsilon))^{1/2}}{(\upsilon(1-m^*/\upsilon))^{1/2\wedge(1-d_0)}}\right)$$
$$=O_p\left(f_x(\lambda_\upsilon)\frac{\log^{1/2}(\upsilon)}{\upsilon^{1/2\wedge(1-d_0)}}\right)=o_p(f_x(\lambda_\upsilon)).$$

Now (A.4) is

$$\frac{\lambda_\upsilon^{-2d_0}}{2m^*+1}\sum_{j=-m^*}^{m^*}|\lambda_\upsilon+\lambda_j|^{2d_0}f_x(\lambda_\upsilon+\lambda_j)\left(2\pi I_\varepsilon(\lambda_\upsilon+\lambda_j)-1\right)$$
$$=\frac{\lambda_\upsilon^{-2d_0}}{2m^*+1}\sum_{j=-m^*}^{m^*}g_x(\lambda_\upsilon+\lambda_j)\left(2\pi I_\varepsilon(\lambda_\upsilon+\lambda_j)-1\right)$$

which has mean zero and variance

$$\frac{\lambda_\upsilon^{-4d_0}}{(2m^*+1)^2}\sum_{j=-m^*}^{m^*}g_x^2(\lambda_\upsilon+\lambda_j)(2\pi)^2\upsilon ar(I_\varepsilon(\lambda_\upsilon+\lambda_j))$$
$$+\frac{\lambda_\upsilon^{-4d_0}}{(2m^*+1)^2}\sum_{j=-m^*}^{m^*}\sum_{k\neq j}g_x(\lambda_\upsilon+\lambda_j)g_x(\lambda_\upsilon+\lambda_k)(2\pi)^2cov(I_\varepsilon(\lambda_\upsilon+\lambda_j),$$
$$\times I_\varepsilon(\lambda_\upsilon+\lambda_k)=O\left(f_x(\lambda_\upsilon)^2\left[\frac{1}{m^*}+\frac{1}{n}\right]\right)=o(f_x(\lambda_\upsilon)^2)$$

by for example, Proposition 10.3.2 in Brockwell and Davis (2006). Thus

$$\check{f}_x(\lambda_\upsilon)=f_x(\lambda_\upsilon)(1+o_p(1)+o(1))$$

uniformly for those υ such that $m^*/\upsilon\to 0$.

Similarly

$$\check{f}_u(\lambda_\upsilon)-\theta=\frac{\lambda_\upsilon^{-2d_0}}{2m^*+1}\sum_{j=-m^*}^{m^*}|\lambda_\upsilon+\lambda_j|^{2d_0}I_u(\lambda_\upsilon+\lambda_j)-\theta. \tag{A.5}$$

The mean of (A.5) is

$$\theta\left(\frac{1}{2m^*+1}\sum\left|1+\frac{j}{\upsilon}\right|^{2d_0}-1\right)=O\left(\frac{m^*}{\upsilon}\right)=o(1),$$

https://doi.org/10.1017/S0266466614000735 Published online by Cambridge University Press

since $|1+j/v|^{2d_0} = 1+O(j/v)$ for $j \le m^*$ and $m^* = o(v)$. Now the variance is

$$\begin{aligned}
&\frac{\lambda_v^{-4d_0}}{(2m^*+1)^2}\sum_{j=-m^*}^{m^*}|\lambda_v+\lambda_j|^{4d_0}var(I_u(\lambda_v+\lambda_j)) \\
&\quad+\frac{\lambda_v^{-4d_0}}{(2m^*+1)^2}\sum_{j=-m^*}^{m^*}\sum_{k\neq j}|\lambda_v+\lambda_j|^{2d_0}|\lambda_v+\lambda_k|^{2d_0}cov(I_u(\lambda_v+\lambda_j),I_u(\lambda_v+\lambda_k)) \\
&= O\left(\frac{1}{(2m^*+1)^2}\sum_{j=-m^*}^{m^*}\left|1+\frac{j}{v}\right|^{4d_0}\right) \\
&\quad+O\left(\frac{1}{(2m^*+1)^2}\sum_{j=-m^*}^{m^*}\sum_{k\neq j}\left|1+\frac{j}{v}\right|^{2d_0}\left|1+\frac{k}{v}\right|^{2d_0}\frac{1}{n}\right) \\
&= O\left(\frac{1}{m^*}+\frac{1}{n}\right) = O\left(\frac{1}{m^*}\right).
\end{aligned}$$

Then $\check{f}_u(\lambda_v)-\theta = O_p(1/\sqrt{m^*}+m^*/v) = o_p(1)$.

With respect to the (pseudo) cross-spectral density function, $\check{f}_{xu}(\lambda_v)$ has mean zero under Assumption A.4 and its variance is

$$\frac{\lambda_v^{-4d_0}}{(2m^*+1)^2}\sum_{j=-m^*,\neq -v}^{m^*}|\lambda_v+\lambda_j|^{4d_0}var(I_{xu}(\lambda_v+\lambda_j)) \tag{A.6}$$

$$\begin{aligned}
&+\frac{\lambda_v^{-4d_0}}{(2m^*+1)^2}\sum_{j}\sum_{k\neq j}|\lambda_v+\lambda_j|^{2d_0}|\lambda_v+\lambda_k|^{2d_0} \\
&\times cov(I_{xu}(\lambda_v+\lambda_j),I_{xu}(\lambda_v+\lambda_k))
\end{aligned} \tag{A.7}$$

Now, using Lemma 2, (A.6) is

$$\begin{aligned}
&\frac{\lambda_v^{-4d_0}}{(2m^*+1)^2}\sum_{j=-m^*}^{m^*}|\lambda_v+\lambda_j|^{4d_0}\frac{\sigma_u^2}{2\pi}f_x(\lambda_v+\lambda_j)\left(1+O\left[\frac{\log|v+j|}{|v+j|^{1\wedge 2(1-d_0)}}+\frac{1}{\sqrt{n}}\right]\right) \\
&= \frac{\lambda_v^{-4d_0}}{(2m^*+1)^2}\frac{\sigma_u^2}{2\pi}g_x(\lambda_v)\sum_{j=-m^*}^{m^*}|\lambda_v+\lambda_j|^{2d_0}\frac{g_x(\lambda_v+\lambda_j)}{g_x(\lambda_v)} \\
&\quad\times\left(1+O\left[\frac{\log|v+j|}{|v+j|^{1\wedge 2(1-d_0)}}+\frac{1}{\sqrt{n}}\right]\right) \\
&= f_u(\lambda_v)f_x(\lambda_v)\frac{1}{(2m^*+1)^2}\sum_{j=-m^*}^{m^*}\left(1+\frac{j}{v}\right)^{2d_0}\frac{g_x(\lambda_v+\lambda_j)}{g_x(\lambda_v)} \\
&\quad\times\left(1+O\left[\frac{\log|v+j|}{|v+j|^{1\wedge 2(1-d_0)}}+\frac{1}{\sqrt{n}}\right]\right) \\
&= f_u(\lambda_v)f_x(\lambda_v)O\left(\frac{1}{m^*}\right).
\end{aligned}$$

https://doi.org/10.1017/S0266466614000735 Published online by Cambridge University Press

Similarly (A.7) is bounded by

$$O\left(\frac{\lambda_v^{-4d_0}}{\sqrt{n}(2m^*+1)^2}\sum_j\sum_{k\neq j}|\lambda_v+\lambda_j|^{2d_0}|\lambda_v+\lambda_k|^{2d_0}f_x^{1/2}(\lambda_j)f_x^{1/2}(\lambda_k)\right)$$
$$=O\left(\frac{\lambda_v^{-4d_0}}{\sqrt{n}(2m^*+1)^2}\sum_j\sum_{k\neq j}|\lambda_v+\lambda_j|^{d_0}|\lambda_v+\lambda_k|^{d_0}\right)$$
$$=O\left(\frac{\lambda_v^{-2d_0}}{\sqrt{n}(2m^*+1)^2}\sum_j\sum_{k\neq j}\left|1+\frac{j}{v}\right|^{d_0}\left|1+\frac{k}{v}\right|^{d_0}\right)$$
$$=O\left(\frac{\lambda_v^{-2d_0}(m^*)^2}{\sqrt{n}(2m^*+1)^2}\right)=O\left(\frac{f_x(\lambda_v)}{\sqrt{n}}\right).$$

In consequence $\check{f}_{xu}(\lambda_v)=O_p(f_x^{1/2}(\lambda_v)[m^{*-1/2}+n^{-1/4}])=o_p(f_x(\lambda_v))$ and

$$\check{f}_y(\lambda_v)=f_y(\lambda_v)+o_p(f_x(\lambda_v))+o_p\left(\frac{1}{\sqrt{m^*}}+\frac{m^*}{v}\right).$$

Now

$$\hat{f}_y(\lambda_v)-\check{f}_y(\lambda_v)=\frac{1}{2m^*+1}\sum\left|1+\frac{j}{v}\right|^{2d_0}I_y(\lambda_v+\lambda_j)\left\{\left|1+\frac{j}{v}\right|^{2(\hat{d}-d_0)}-1\right\}$$

is $o_p(f_x(\lambda_v))$ since, as in Lemma 3 of Hidalgo and Yajima (2002)

$$\left|\left|1+\frac{j}{v}\right|^{2(\hat{d}-d_0)}-1\right|\leq constant\times|\hat{d}-d_0|\log\left|1+\frac{j}{v}\right|\leq constant\times|\hat{d}-d_0|$$

and $|\hat{d}-d_0|=o_p(1)$. In consequence, for λ_v such that $m^*/v\to 0$ we have that

$$\hat{f}_y(\lambda_v)=f_y(\lambda_v)+o_p(f_x(\lambda_v))+o_p(1)$$

which implies consistency at shrinking frequencies where $f_y(\lambda_v)=f_x(\lambda_v)(1+o(1))$, and also at constant frequencies ($v=O(n)$) where $f_x(\lambda_v)=O(1)$ and $f_y(\lambda_v)=f_x(\lambda_v)+f_u(\lambda_v)$.

Consider now frequencies close to zero such that $v=O(m^*)$, the $O()$ bound meaning here exact rate and also including the case $v/m^*\to 0$ as $n\to\infty$. In that range of frequencies $f_y(\lambda_v)=f_x(\lambda_v)(1+o(1))$ and thus consistent estimation of $f_y(\lambda_v)$ implies consistent estimation of $f_x(\lambda_v)$. Assume also that $m^*>v$, the proof for other v being similar and easier because there is no need to split the sums into negative and positive frequencies. Partition now the sum in (A.1) in two parts, one from $j=-m^*$ to $j=-v-1$ and other from $j=-v+1$ to $j=m^*$. The analysis of both expressions is similar taking into account the symmetry of the (pseudo) spectral density and periodogram around the origin, and thus we only show here the results corresponding to the second sum. For that interval of frequencies (A.1) can be written

https://doi.org/10.1017/S0266466614000735 Published online by Cambridge University Press

$$\frac{1}{2m^*}\sum_{j=-\upsilon+1}^{m^*}\left|1+\frac{j}{\upsilon}\right|^{2d_0}(I_x(\lambda_\upsilon+\lambda_j)-f_x(\lambda_\upsilon+\lambda_j))$$

$$=\frac{1}{2m^*}\sum_{j=1}^{m^*+\upsilon}\left|\frac{j}{\upsilon}\right|^{2d_0}(I_x(\lambda_j)-f_x(\lambda_j))$$

$$=\frac{1}{2m^*}\sum_{j=1}^{m^*+\upsilon}\left(\frac{j}{\upsilon}\right)^{2d_0}(I_x(\lambda_j)-|\alpha_j|^2I_\varepsilon(\lambda_j)) \tag{A.8}$$

$$+\frac{1}{2m^*}\sum_{j=1}^{m^*+\upsilon}\left(\frac{j}{\upsilon}\right)^{2d_0}(|\alpha_j|^2I_\varepsilon(\lambda_j)-f_x(\lambda_j)). \tag{A.9}$$

Using again formulas (3.17) in Robinson (1995) and (A.2) in Velasco (1999b), (A.8) is bounded in probability by

$$O_p\left(\frac{1}{2m^*}\sum_{j=1}^{m^*+\upsilon}\left(\frac{j}{\upsilon}\right)^{2d_0}\lambda_j^{-2d_0}\left[\frac{1}{j^{1/2\wedge(1-d_0)}}\right]\log^{1/2}j\right)$$

$$=O_p\left(\frac{\log^{1/2}m^*}{m^*}\lambda_\upsilon^{-2d_0}(m^*+\upsilon)^{1/2\vee d_0}\right)=o_p(f_x(\lambda_\upsilon))$$

since $d_0<1$ and $\upsilon=O(m^*)$.

The bound for (A.9) is proved similarly to that of (A.4). Then (A.9) is

$$O_p\left(\frac{1}{m^*}\left[\sum_{j=1}^{m^*+\upsilon}\left(\frac{j}{\upsilon}\right)^{4d_0}\lambda_j^{-4d_0}\right]^{1/2}\right)$$

$$=O_p\left(\frac{(m^*+\upsilon)^{1/2}}{m^*}f_x(\lambda_\upsilon)\right)=o_p\left(f_x(\lambda_\upsilon)\right).$$

Now, with respect to the added noise, the sum in (A.5) can be similarly split into two parts, one from $j=-m^*$ t $j=-\upsilon-1$ and other from $j=-\upsilon+1$ to $j=m^*$. Since the analysis of both is similar we focus again on the latter. In that interval the sum in (A.5) is

$$\frac{1}{2m^*}\sum_{j=1}^{m^*+\upsilon}\left(\frac{j}{\upsilon}\right)^{2d_0}I_u(\lambda_j)-\theta, \tag{A.10}$$

which has mean

$$O\left(\frac{1}{m^*}\sum_{j=1}^{m^*+\upsilon}\left(\frac{j}{\upsilon}\right)^{2d_0}\right)=O\left(\frac{(m^*+\upsilon)^{2d_0+1}}{m\upsilon^{2d_0}}\right)=O\left(\left[\frac{m^*}{n}\right]^{2d_0}\lambda_\upsilon^{-2d_0}\right)=o_p(f_x(\lambda_\upsilon))$$

and variance

$$O\left(\frac{1}{m^{*2}}\sum_{j=1}^{m^*+\upsilon}\left(\frac{j}{\upsilon}\right)^{4d_0}\right)=O\left(\frac{(m^*+\upsilon)^{4d_0+1}}{m^{*2}\upsilon^{4d_0}}\right)=O\left(f_x(\lambda_\upsilon)^2\frac{m^{*4d_0-1}}{n^{4d_0}}\right).$$

Thus (A.10) is $o_p(f_x(\lambda_\upsilon))$ for $\upsilon=O(m^*)$.

https://doi.org/10.1017/S0266466614000735 Published online by Cambridge University Press

The bounds for the cross-spectral density are similarly obtained. We need to bound (A.6) and (A.7). First, the sums are splitted in two parts as before, in both of which the analysis is equal. Consider, for example, the sum over $j = -m^*, ..., -\upsilon - 1$. By Lemma 2, this part of the sum in (A.6) is bounded by

$$O\left(\frac{\lambda_\upsilon^{-4d_0}}{(m^*)^2}\sum_{j=1}^{m-\upsilon}\lambda_j^{4d_0} f_x(\lambda_j)\right) = O\left(\frac{\lambda_\upsilon^{-4d_0}}{(m^*)^2}\sum_{j=1}^{m-\upsilon}\left(\frac{j}{n}\right)^{2d_0}\right)$$
$$= O\left(f_x^2(\lambda_\upsilon)\left(1-\frac{\upsilon}{m^*}\right)^{1+2d_0}\left(\frac{m^*}{n}\right)^{2d_0}\frac{1}{m^*}\right) = o(f_x^2(\lambda_\upsilon))$$

for $\upsilon = O(m^*)$. Similarly, (A.7) is bounded by

$$O\left(\frac{\lambda_\upsilon^{-4d_0}}{\sqrt{n}(m^*)^2}\left[\left\{\sum_{j=-m^*}^{-\upsilon-1} + \sum_{j=-\upsilon+1}^{m^*}\right\}|\lambda_\upsilon + \lambda_j|^{d_0}\right]^2\right)$$
$$= O\left(\frac{\lambda_\upsilon^{-4d_0}}{\sqrt{n}(m^*)^2}\left[\left\{\sum_{j=1}^{m^*-\upsilon} + \sum_{j=1}^{m^*+\upsilon}\right\}\lambda_j^{d_0}\right]^2\right)$$
$$= O\left(\frac{\lambda_\upsilon^{-4d_0}}{\sqrt{n}}\left(\frac{m^*}{n}\right)^{2d_0}\left[\left(1-\frac{\upsilon}{m^*}\right)^{1+d_0} + \left(1+\frac{\upsilon}{m^*}\right)^{1+d_0}\right]^2\right)$$
$$= O\left(f_x^2(\lambda_\upsilon)\left(\frac{m^*}{n}\right)^{2d_0}\frac{1}{\sqrt{n}}\right) = o(f_x^2(\lambda_\upsilon)),$$

which implies $\check{f}_{xu}(\lambda_\upsilon) = o_p(f_x(\lambda_\upsilon))$ also for $\upsilon = O(m^*)$ and then $\check{f}_y(\lambda_\upsilon) = f_x(\lambda_\upsilon)(1 + o_p(1))$ for $\upsilon = O(m^*)$.

Finally $\hat{f}_y(\lambda_\upsilon) - \check{f}_y(\lambda_\upsilon)$ is

$$\frac{1}{2m^*}\left\{\sum_{j=1}^{m^*-\upsilon} + \sum_{j=1}^{m^*+\upsilon}\right\} I_y(\lambda_j)\left|\frac{j}{\upsilon}\right|^{2d_0}\left[1-\left|\frac{j}{\upsilon}\right|^{2(\hat{d}-d_0)}\right] = o_p(\check{f}_y(\lambda_\upsilon)) = o_p(f_x(\lambda_\upsilon))$$

because

$$\left[1-\left|\frac{j}{\upsilon}\right|^{2(\hat{d}-d_0)}\right] \le constant \times (\hat{d}-d_0)\log(j/\upsilon) = o_p(1)$$

by Assumption A.7. ■

LEMMA 1. *Under Assumptions A.1 and A.2, uniformly in integer* $j \le [n/2]$,

$$\int_{-\pi}^{\pi} f_x(\lambda)K(\lambda+\lambda_j)d\lambda = f_x(\lambda_j)\begin{cases} 1+O(j^{-1}+j^{2d_0-2}\log j) & \text{if } j = o(n) \\ 1+O(l^{-1}+j^{2d_0-2}\log j) & \text{if } j \text{ proportional to } n \end{cases}$$

for l *satisfying* $l < j$, $l = o(n)$ *and* $K(\lambda) = |D(\lambda)|^2/2\pi n$ *with* $D(\lambda) = \sum_{t=1}^{n} e^{it\lambda}$.

https://doi.org/10.1017/S0266466614000735 Published online by Cambridge University Press

Proof. Split the integral

$$\int_{-\pi}^{\pi} f_x(\lambda)K(\lambda+\lambda_j)\mathrm{d}\lambda - f_x(\lambda_j) = \int_{-\pi}^{\pi} [f_x(\lambda)-f_x(\lambda_j)]K(\lambda+\lambda_j)\mathrm{d}\lambda$$

into the following parts

$$\int_{-\pi}^{-\varepsilon} + \int_{-\varepsilon}^{-3\lambda_j/2} + \int_{-3\lambda_j/2}^{-\lambda_j/2} + \int_{-\lambda_j/2}^{\lambda_j/2} + \int_{\lambda_j/2}^{\varepsilon} + \int_{\varepsilon}^{\pi}$$

for $\varepsilon > 3\lambda_j/2$. Taking into account that $K(\lambda) \leq constant \times \min(n, n^{-1}\lambda^{-2})$, the integral over $[-\pi,-\varepsilon]$ is bounded in absolute value by

$$\max_{[-\pi,-\varepsilon]} K(\lambda+\lambda_j)\left[\int_{-\pi}^{-\varepsilon} f_x(\lambda)\mathrm{d}\lambda + f_x(\lambda_j)\right] = O\left(\frac{1}{n} + \frac{f_x(\lambda_j)}{n}\right) = O\left(\frac{f_x(\lambda_j)}{n}\right).$$

The same bound is similarly obtained for the integral over $[\varepsilon,\pi)$. Now

$$\begin{aligned}
&\left|\int_{-\varepsilon}^{-3\lambda_j/2}\right| \\
&\quad = O\left(n^{-1}\int_{-\varepsilon}^{-3\lambda_j/2} |\lambda|^{-2d_0}|\lambda+\lambda_j|^{-2}\mathrm{d}\lambda + n^{-1}f_x(\lambda_j)\int_{-\varepsilon}^{-3\lambda_j/2}|\lambda+\lambda_j|^{-2}\mathrm{d}\lambda\right) \\
&\quad = O\left(n^{-1}\lambda_j^{-2d_0-1}\right) = O\left(j^{-1}f_x(\lambda_j)\right).
\end{aligned}$$

The bound for the interval $[\lambda_j/2,\varepsilon)$ is similarly obtained. Next,

$$\begin{aligned}
\left|\int_{-\lambda_j/2}^{\lambda_j/2}\right| &= O\left(\max_{[-\lambda_j/2,-\lambda_j/2]} K(\lambda+\lambda_j)\left[\int_{-\lambda_j/2}^{\lambda_j/2} f_x(\lambda)\mathrm{d}\lambda + \lambda_j f_x(\lambda_j)\right]\right) \\
&= (n^{-1}\lambda_j^{-2}\lambda_j^{1-2d_0}) = O\left(j^{-1}f_x(\lambda_j)\right)
\end{aligned}$$

for $d_0 < 1/2$. If $d_0 \geq 1/2$ we follow the steps in the bounds of formula (A.4) in Velasco (1999a) in order to bound

$$\int_{-\lambda_j/2}^{\lambda_j/2} K(\lambda+\lambda_j)f_x(\lambda)\mathrm{d}\lambda. \tag{A.11}$$

Dividing (A.11) by $f_x(\lambda_j)$ we get

$$\begin{aligned}
&\frac{1}{2\pi n}\int_{-\lambda_j/2}^{\lambda_j/2} \frac{\sin^2\left(\frac{n(\lambda+\lambda_j)}{2}\right)}{\sin^2\left(\frac{\lambda+\lambda_j}{2}\right)} \frac{|\lambda|^{-2d_0}g_x(\lambda)}{\lambda_j^{-2d_0}g_x(\lambda_j)}\mathrm{d}\lambda \\
&\quad = O\left(\frac{1}{2\pi n^2}\int_{-\pi j}^{\pi j} \frac{\sin^2\left(\frac{\lambda+2\pi j}{2}\right)}{\sin^2\left(\frac{\lambda+2\pi j}{2n}\right)} \frac{|\lambda|^{-2d_0}}{(2\pi j)^{-2d_0)}}\mathrm{d}\lambda\right) \\
&\quad = O\left(\frac{1}{2\pi n^2}\int_{-\pi j}^{\pi j} \frac{\sin^2\left(\frac{\lambda}{2}\right)}{\left(\frac{\lambda+2\pi j}{2n}\right)^2} \frac{|\lambda|^{-2d_0}}{(2\pi j)^{-2d_0)}}\mathrm{d}\lambda\right)
\end{aligned}$$

https://doi.org/10.1017/S0266466614000735 Published online by Cambridge University Press

$$= O\left(j^{2d_0-2} \int_{-\pi j}^{\pi j} \sin^2\left(\frac{\lambda}{2}\right) |\lambda|^{-2d_0} \mathrm{d}\lambda \right)$$
$$= \left(j^{2_0-2} \left[\int_0^1 \lambda^{2-2d_0} \mathrm{d}\lambda + \int_1^{\pi j} \lambda^{-2d_0} \mathrm{d}\lambda \right] \right)$$
$$= O(j^{2d_0-2} \log j)$$

for $j \leq n/3$ since $|\sin x| > |x|/2$ for $|x| \leq \pi/2$ and $1/2 \leq d_0 < 1$. If $j > n/3$, then $\sin^2\left(\frac{\lambda+2\pi j}{2n}\right) \neq 0$ for the limits of the integral and (A.11) is bounded by

$$O\left(f_x(\lambda_j) j^{2d_0} n^{-2} \int_{-\pi j}^{\pi j} \sin^2\left(\frac{\lambda}{2}\right) |\lambda|^{-2d_0} \mathrm{d}\lambda \right)$$
$$= O(f_x(\lambda_j) j^{2d_0} n^{-2} \log j) = O(f_x(\lambda_j) j^{2d_0-2} \log j).$$

Finally, for the bound between $-3\lambda_j/2$ and $-\lambda_j/2$ we first consider shrinking to zero frequencies such that $\lambda_j \to 0$, implying $j = o(n)$. For those Fourier frequencies

$$\left| \int_{-3\lambda_j/2}^{-\lambda_j/2} \right| = \left| \int_{-\lambda_j/2}^{\lambda_j/2} \left[f_x(\lambda - \lambda_j) - f_x(\lambda_j) \right] K(\lambda) \mathrm{d}\lambda \right| \tag{A.12}$$
$$= O\left(\int_{-\lambda_j/2}^{\lambda_j/2} \left[\lambda f_x'(\lambda_j) + \lambda^2 f_x''(\lambda_j) \right] K(\lambda) \mathrm{d}\lambda \right)$$
$$= O\left(\lambda_j^{-2-2d_0} \int_{-\lambda_j/2}^{\lambda_j/2} \lambda^2 K(\lambda) \mathrm{d}\lambda \right)$$
$$= O\left(\lambda_j^{-2-2d_0} \left[\int_0^{n^{-1}} n\lambda^2 \mathrm{d}\lambda + \int_{n^{-1}}^{\lambda_j/2} n^{-1} \mathrm{d}\lambda \right] \right)$$
$$= O(f_x(\lambda_j) j^{-1})$$

since $K(\lambda)$ is even and positive. For fixed λ_j such that j is proportional to n, we split the integral in (A.12) into

$$\int_{-\lambda_j/2}^{\lambda_j/2} = \int_{-\lambda_j/2}^{-\lambda_l/2} + \int_{-\lambda_l/2}^{\lambda_l/2} + \int_{\lambda_l/2}^{\lambda_j/2}$$

for $l = o(j)$ (implying $l = o(n)$). As before, the second integral in the sum is $O(f_x(\lambda_j) l^{-1})$ and

$$\left| \int_{\lambda_l/2}^{\lambda_j/2} \right| = \left(\lambda_j^{-2d_0} \int_{\lambda_l/2}^{\lambda_j/2} n^{-1} \lambda^{-2} \mathrm{d}\lambda \right) = O(f_x(\lambda_j) l^{-1}). \quad \blacksquare$$

LEMMA 2. *Under Assumptions A1–A5, uniformly in integer* $j, k \leq [n/2]$,

$$cov(I_{xu}(\lambda_j), I_{xu}(\lambda_k)) = \begin{cases} \frac{\sigma_u^2}{2\pi} f_x(\lambda_j) \left(1 + O\left(\frac{\log j}{j^{1\wedge 2(1-d_0)}} + \frac{1}{\sqrt{n}} \right) \right) & \text{if } k = j \\ O\left(\frac{1}{\sqrt{n}} f_x^{1/2}(\lambda_j) f_x^{1/2}(\lambda_k) \right) & \text{if } k \neq j \end{cases}$$

https://doi.org/10.1017/S0266466614000735 Published online by Cambridge University Press

Proof. The covariance between $I_{xu}(\lambda_j)$ and $I_{xu}(\lambda_k)$ is

$$E[W_x(\lambda_j)\overline{W_x(\lambda_k)}]E[W_u(\lambda_k)\overline{W_u(\lambda_j)}] \tag{A.13}$$

$$+cum[W_x(\lambda_j)\overline{W_x(\lambda_k)}W_u(\lambda_k)\overline{W_u(\lambda_j)}]. \tag{A.14}$$

for uncorrelation between x and u. Now, (A.13) is

$$\frac{\sigma_u^2}{(2\pi n)^2}\sum_t\sum_m E(x_t x_m)\exp(-it\lambda_j)\exp(im\lambda_k)\sum_s \exp(-is(\lambda_k-\lambda_j))$$

$$= \begin{cases} 0 & \text{if } k \neq j \\ \frac{\sigma_u^2}{2\pi} E I_x(\lambda_j) = \frac{\sigma_u^2}{2\pi} f_x(\lambda_j)\left(1+O\left(\frac{\log j}{j^{1\wedge 2(1-d_0)}}\right)\right) & \text{for } k = j \end{cases} \tag{A.15}$$

by Theorem 1 in Velasco (1999a).

Using formulas (2.6.3), (2.10.1), and (2.10.3) in Brillinger (1975), (A.14) is under Assumption A.5

$$\frac{k}{(2\pi)^4 n^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\alpha(\lambda)\alpha(\mu)D(\lambda-\lambda_j)D(\mu+\lambda_k)D(\lambda_j-\lambda_k-\lambda-\mu)\mathrm{d}\lambda\mathrm{d}\mu$$

which by Schwarz inequality is bounded in absolute value by

$$\frac{k}{(2\pi)^4 n^2}\left[\int\int|\alpha(\lambda)\alpha(\mu)D(\lambda-\lambda_j)D(\mu+\lambda_k)|^2\mathrm{d}\lambda\mathrm{d}\mu\right]^{1/2}$$

$$\times\left[\int\int|D(\lambda_j-\lambda_k-\lambda-\mu)|^2\mathrm{d}\lambda\mathrm{d}\mu\right]^{1/2}$$

$$= \frac{k}{(2\pi)^2(2\pi n)^{1/2}}\left[\int|\alpha(\lambda)|^2K(\lambda-\lambda_j)\mathrm{d}\lambda\int|\alpha(\mu)|^2K(\mu+\lambda_k)\mathrm{d}\mu\right]^{1/2}$$

$$\times\left[\int\int K(\lambda_j-\lambda_k-\lambda-\mu)\mathrm{d}\lambda\mathrm{d}\mu\right]^{1/2}$$

$$= \frac{k}{(2\pi)n^{1/2}}\left[\int f_x(\lambda)K(\lambda-\lambda_j)\mathrm{d}\lambda\int f_x(\mu)K(\mu+\lambda_k)\mathrm{d}\mu\right]^{1/2}$$

$$= O\left(n^{-1/2}f_x^{1/2}(\lambda_j)f_x^{1/2}(\lambda_k)\right)$$

by Lemma 1 and because $K(\lambda)$ is periodic and $\int_{-\pi}^{\pi}K(\lambda)\mathrm{d}\lambda = 1$. ■

https://doi.org/10.1017/S0266466614000735 Published online by Cambridge University Press