Fractional cointegration in stochastic volatility models

Afonso Gonçalves da Silva[*], Peter M. Robinson[†]

Department of Economics, London School of Economics and Political Science,

Houghton Street, London WC2A 2AE, UK

February 14, 2005

Abstract

We consider narrow band semiparametric estimation of the factor loading on a bivariate factor model, where the common factor and idiosyncratic errors are serially uncorrelated but have stationary long memory in conditional higher moments. The special case of stochastic volatility models for the latent variables is considered, since it has direct application to asset pricing and other financial models. Assuming the underlying memory is higher in the factor than in the errors, a fractional cointegrating relationship can be recovered by a suitable transformation of the data. The estimate is shown to be consistent and its small sample properties are investigated in a Monte Carlo experiment.

JEL classification: C22

Keywords: Fractional cointegration; stochastic volatility; narrow band least squares; semiparametric analysis.

1 Introduction

Financial time series, such as asset returns, are commonly found to be approximately uncorrelated but not independent across time. Much of this dependence can be traced to the fact that volatilities are time dependent, with highly volatile observations grouped in some periods, and relatively low volatilities elsewhere. A great deal of attention has

[*]Research supported by FCT grant SFRH/BD/4783/2001 and ESRC grant R000239936. Tel: +44 (0)20 7955 7721. E-mail: a.m.goncalves-da-silva@lse.ac.uk.

[†]Research supported by ESRC grant R000239936.

focused on modelling the consequent conditional heteroskedasticity. Influential early contributions were the ARCH model of Engle (1982) (applied there to inflation data), the GARCH extension of Bollerslev (1986), and along a different line, the stochastic volatility (SV) model of Taylor (1986). Empirical evidence has suggested a higher degree of persistence than these models entail, leading to Engle and Bollerslev's (1986) introduction of the IGARCH model. However, the persistence implied by this model (and other unit root based ones, such as IEGARCH) seems too extreme. On the one hand, the absence of mean reversion in the second moments implies permanent shifts to long term volatility forecasts, which is theoretically implausible. On the other, empirical investigation of volatility measures, such as absolute values and squares of observations, suggested they are better explained as stationary processes with long memory, indicating the need for a more flexible model of volatility persistence; see, for example, Whistler (1990), Ding, Granger, and Engle (1993), Ding and Granger (1996), Andersen and Bollerslev (1997).

Several parametric models for this phenomenon have been proposed. Robinson (1991) extended the GARCH framework to an ARCH(∞) model that can explain greater persistence. Other models within this framework include Ding and Granger (1996), Baillie, Bollerslev, and Mikkelsen (1996), Bollerslev and Mikkelsen (1996). Other authors have extended Taylor's (1986) SV model to explain long memory in squares, e.g. Andersen and Bollerslev (1997), Harvey (1998), Breidt, Crato, and de Lima (1998).

In a parallel line of research, asset pricing and other models predict the existence of one or more unobservable common factors explaining a multivariate time series. The classical CAPM decomposes asset returns into a single factor, interpreted as the market return, and an idiosyncratic component. In the present paper, we consider a model in which two observable scalar time series, y_t and x_t, $t = 0, \pm 1, \ldots$, whose squares or other powers have long memory, are generated by

$$y_t = \beta\zeta_t + \varepsilon_t, \tag{1.1}$$

$$x_t = \zeta_t + \delta_t, \tag{1.2}$$

where β is unknown and ζ_t, ε_t, δ_t are unobservable stationary processes.

Due to the measurement error δ_t in ζ_t, the ordinary least squares (OLS) estimate of y_t on x_t is inconsistent. Indeed, our assumptions will imply that ζ_t, ε_t and δ_t are white noise sequences (i.e. have zero autocorrelations at all lags), so in no meaningful sense can (1.1), (1.2) be described as a cointegrating relation. However, ζ_t, ε_t, δ_t are supposed to be generated by SV models and so are not serially independent, but exhibit persistence in higher moments. In particular, for some integer $p > 1$, our assumptions imply that x_t^p and y_t^p are cointegrated long memory $I(d)$ processes, $0 < d < 1/2$, with cointegrating coefficient $\theta = \beta^p$ and cointegrating errors are $I(d^*)$ for $0 \leq d^* < d$. Still, x_t^p and y_t^p are stationary, so the OLS estimate is inconsistent for θ, unlike under the traditional assumption of $I(1)$ observables and $I(0)$ cointegrating errors.

When the spectral density of stationary regressors dominates that of cointegrating errors at low frequencies, Robinson (1994a) showed that a narrow band least squares (NBLS) estimate can be consistent. His observable sequences were linear processes in conditionally homoscedastic martingale difference (md) innovations, which is manifestly not the case in our intrinsically nonlinear framework. Nevertheless, NBLS has been applied to financial data (see e.g. Christensen and Nielsen, 2004; Bandi and Perron, 2004), so it would be desirable to establish consistency under more relevant assumptions. The present paper fills this gap in the theoretical properties of the NBLS estimate of θ, allowing the latent variables in (1.1), (1.2) to be quite general SV processes. The NBLS estimate converges at a slow (nonparametric) rate, but in long financial series adequate precision may be achievable. Better estimates of θ are possible, though they would require at least estimating the memory parameters, being computationally more intensive and more complicated to handle theoretically. Even for the relatively simple NBLS estimate, our proof of consistency is extremely lengthy.

A key component of the proof is an approximation for expectations of products of nonlinear functions of Gaussian processes, which is presented in the following section, with proof in Appendix A. Section 3 describes the SV setting. Section 4 introduces the NBLS estimate and our consistency result, which is proved in a series of propositions stated and proved in Appendix B, using lemmas in Appendix C. Sections 5 and 6 consist of a Monte Carlo study of finite sample performance and a discussion of further directions for research.

2 Approximating cross-moments of nonlinear functions of Gaussian variables

With the objective of examining the memory of SV models similar to those introduced in the following section, Robinson (2001) established an asymptotic expansion for the covariance between nonlinear functions of multivariate normal random vectors. Here we need an extension to cross-moments of more than two real functions.

Let $\phi(\cdot)$ denote the standard normal density and $H_j(\cdot)$ the j-th Hermite polynomial, for $j \geq 0$, defined by

$$H_j(x)\phi(x) = (-1)^j \frac{\partial^j}{\partial x^j}\phi(x). \tag{2.1}$$

For a function $f(\cdot)$ satisfying $\int_{\mathbb{R}} f^2(x)\phi(x)\mathrm{d}x < \infty$, define the j-th Hermite coefficient $G_j = \int_{\mathbb{R}} f(x)H_j(x)\phi(x)\mathrm{d}x$ and the Hermite rank $r = \min\{j \geq 0 : G_j \neq 0\}$. Define $P_q = \{i \in \mathbb{N} : i \leq q\}$, where $\mathbb{N} = \{1, 2, \ldots\}$, $Q_q = \{(i,j) \in P_q^2 : i < j\}$ and $R_{q,k} = \{(i,j) \in Q_q : i = k \vee j = k\}$ for $k \in P_q$.

Theorem 1 *For integer $J > 1$, let μ_j, $j \in P_J$ be jointly normally distributed with zero mean, unit variance and covariances $\rho_{jk} = \mathrm{Cov}(\mu_j, \mu_k)$, $j \neq k$; let $f_j = f_j(\mu_j)$ be a function such that $E(f_j^2) < \infty$, with k-th Hermite coefficient $G_{j,k}$ and Hermite rank r_j. Then*

$$E\left(\prod_{j \in P_J} f_j\right) = \sum_{q=0}^{\infty} a_q, \tag{2.2}$$

where

$$a_q = \sum_{\substack{v_\alpha \geq 0: \\ \Sigma v_\alpha = q, \\ \alpha \in Q_J}} \prod_{j \in P_J} G_{j,w_j} \prod_{\alpha \in Q_J} \frac{\rho_\alpha^{v_\alpha}}{v_\alpha !}, \qquad w_j = \sum_{\alpha \in R_{j,J}} v_\alpha. \tag{2.3}$$

If in addition $\tau = 2\sum_{\alpha \in Q_J} |\rho_\alpha| < 1$, then

$$a_q = 0, \qquad 2q < r, \tag{2.4}$$

$$|a_q| \leq \sigma \prod_{j \in P_J} \left(\sum_{\alpha \in R_{j,J}} |\rho_\alpha|\right)^{\frac{r_j}{2}} \tau^{q - \frac{r}{2}}, \qquad 2q \geq r, \tag{2.5}$$

$$\sum_{i=q}^{\infty} |a_i| \leq \sigma \prod_{j \in P_J} \left(\sum_{\alpha \in R_{j,J}} |\rho_\alpha|\right)^{\frac{r_j}{2}} \frac{\tau^{q - \frac{r}{2}}}{1 - \tau}, \qquad 2q \geq r, \tag{2.6}$$

where $r = \sum_{j \in P_J} r_j$ *and* $\sigma = \{\prod_{j \in P_J} E(f_j^2)\}^{1/2}$.

The bounds (2.4), (2.5) reflect the individual, possibly differing, Hermite ranks r_j of the f_j. The weakest version of Theorem 1 arises when $r_j \equiv 0$ (i.e. when $E(f_j) \neq 0$ for all j), and because this would be relevant also when the r_j are unknown, we present it in the following Corollary, whose proof follows from the inequality $\sum_{\alpha \in R_{j,J}} |\rho_\alpha| \leq \tau$.

Corollary 1

$$|a_q| \leq \sigma \tau^q, \qquad \sum_{i=q}^{\infty} |a_i| \leq \sigma \frac{\tau^q}{1-\tau}.$$

As in Robinson (2001) in case $J = 2$, Theorem 1 provides a valid asymptotic expansion. Robinson (1994a) established consistency of the NBLS estimate using L_1 arguments enabled by linear process (in md innovations) assumptions. Since those are unavailable to us, we use L_2 arguments. These were also employed by Robinson (1994b) in studying the mean squared error of the averaged periodogram, but in case of Gaussian and linear (in independent and identically distributed, iid, innovations) assumptions. In the SV setting introduced in the following section, matters are considerably more complicated, and we are led to consider various cross-moments of nonlinear functions of Gaussian processes. Theorem 1 is crucial in obtaining sufficiently sharp bounds on these cross-moments to establish consistency.

3 Long memory stochastic volatility framework

To describe the structure of the latent processes ζ_t, ε_t, δ_t in (1.1), (1.2), we first introduce a technical definition of $I(d)$ processes. We say z_t is $I(d)$, with memory parameter $d \in [0, 1/2)$, if it is stationary with zero mean and finite variance, and it has autocovariance function $\rho_j = \mathrm{Cov}(z_0, z_j)$ satisfying

$$\sum_{j=0}^{\infty} |\rho_j| < \infty, \tag{3.1}$$

if $d = 0$, and

$$\rho_j \sim C_\rho j^{2d-1} \text{ as } j \to \infty, \text{ for } C_\rho > 0, \tag{3.2}$$

$$|\rho_j - \rho_{j+1}| \leq K \frac{|\rho_{j+1}|}{j}, \; j \geq 0, \tag{3.3}$$

if $0 < d < 1/2$, where K throughout denotes a generic, arbitrarily large finite constant, and the symbol " $\sim$ " indicates that the ratio of left- and right-hand sides tends to one. We say an $I(0)$ process has short memory, and an $I(d)$ process, for $0 < d < 1/2$, has long memory. We can deduce from (3.1) or (3.2), (3.3) properties of the spectral density $f(\lambda)$ of z_t, which satisfies $\rho_j = \int_{-\pi}^{\pi} f(\lambda) \cos(j\lambda) \mathrm{d}\lambda$. For $d = 0$, $f(\lambda)$ is continuous for all λ, whereas for for $0 < d < 1/2$, Theorem III-12 of Yong (1974) indicates that

$$f(\lambda) \sim C_f \lambda^{-2d} \text{ as } \lambda \to 0^+, \tag{3.4}$$

where

$$C_f = \pi^{-1} \Gamma(2d) \sin\left\{(1-2d)\frac{\pi}{2}\right\} C_\rho,$$

so that $f(\lambda)$ diverges at $\lambda = 0$. Stationary autoregressive moving average (ARMA) processes satisfy (3.1), and stationary fractionally integrated ARMA (ARFIMA) processes satisfy (3.2), (3.3).

Assumption 1 *For $t = 0, \pm 1, \ldots,$*

$$\zeta_t = \eta_{1t} g_t, \qquad \delta_t = \nu_{1t} h_t, \qquad \varepsilon_t = \xi_{1t} l_t, \tag{3.5}$$

where for real-valued functions g, h, l,

$$g_t = g(\eta_{2t}), \qquad h_t = h(\nu_{2t}), \qquad l_t = l(\xi_{2t}), \tag{3.6}$$

and
(i) $\{\eta_{1t}\}$, $\{\nu_{1t}\}$, $\{\xi_{1t}\}$ are jointly iid processes with zero mean;
(ii) $\{\eta_{2t}\}$ is $I(d)$, $\{\nu_{2t}\}$ is $I(d')$ and $\{\xi_{2t}\}$ is $I(d'')$, for $d' \geq 0$, $d'' \geq 0$, $\max\{d', d''\} < d < 1/2$;
(iii) $\{\eta_{2t}\}$, $\{\nu_{2t}\}$, $\{\xi_{2t}\}$ are standard Gaussian processes, independent of each other and of $\{\eta_{1t}\}$, $\{\nu_{1t}\}$, $\{\xi_{1t}\}$;

(iv) For some integer $p > 1$,

$$E(\eta_{1t}^p)E\left\{g^p(\eta_{2t})\eta_{2t}\right\} \neq 0, \tag{3.7}$$

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

$$E(\eta_{1t}^j \nu_{1t}^{p-j})E\left\{g^j(\eta_{2t})\eta_{2t}\right\} = E(\eta_{1t}^j \xi_{1t}^{p-j})E\left\{g^j(\eta_{2t})\eta_{2t}\right\} = 0. \tag{3.8}$$

(v) $\{\eta_{1t}\}$, $\{\nu_{1t}\}$, $\{\xi_{1t}\}$, $\{g_t\}$, $\{h_t\}$, $\{l_t\}$ have finite $4p$-th moments.

It follows that ζ_t, ε_t, δ_t, described by SV models in (3.5), are serially uncorrelated but not serially independent. In particular, x_t^p is autocorrelated with long memory, due to $d > 0$ and (3.7), which entails $E(\eta_{1t}^p) \neq 0$ and $g_t^p - E(g_t^p)$ having Hermite rank one. Condition (3.8) ensures a valid cointegrating relationship between x_t^p and y_t^p, since it implies that the cointegrating error has memory smaller than d. If η_{1t} is independent of ν_{1t}, ξ_{1t}, the smallest integer satisfying (3.7) will also satisfy (3.8). It is assumed that p is known, which imposes some restrictions on g; in practice it may be reasonable to suppose that $p = 2$. The most notable exception would occur if g is a symmetric function, e.g. $g_t = |\eta_{2t}|^a$, $\alpha > 0$, but then no finite p satisfies (3.7). This does not rule out a cointegrating relationship of the type that we study below, but the associated conditions would be extremely complex, involving the magnitudes of d, d', d'' and the Hermite ranks of each centered power of g_t, h_t, l_t. Note that for $\gamma \neq 0$, $g_t = |\gamma + \eta_{2t}|^a$ gives $p = 2$. Further discussion concerning the Hermite rank for functional forms in SV models with long memory can be found in Robinson (2001).

An advantage of a low p is that the moment conditions in part (v) of Assumption 1 are then minimal. Even for $p = 2$, the 8-th moment condition that is required seems stringent for most financial data, though it is never possible to establish the existence or non-existence of moments from real data. Other parts of Assumption 1 might be relaxed at cost of substantial lengthening of the proof, in particular the mutual independence assumptions of (iii). The Gaussianity assumptions on η_{2t}, ν_{2t}, ξ_{2t} are mitigated by allowing g, h, l to be quite general functions, and without them the details would be considerably more complex; of course Gaussianity frequently plays a role in short memory SV models also.

4 Consistency of the Narrow Band Least Squares estimate

We transform (1.1), (1.2) to

$$Y_t = \theta X_t + U_t, \tag{4.1}$$

where

$$Y_t = y_t^p = \sum_{j=0}^{p} \binom{p}{j} \beta^j \zeta_t^j \varepsilon_t^{p-j}, \qquad X_t = x_t^p = \sum_{j=0}^{p} \binom{p}{j} \zeta_t^j \delta_t^{p-j}, \qquad \theta = \beta^p,$$

$$U_t = y_t^p - \beta^p x_t^p = \sum_{j=0}^{p} \binom{p}{j} (\beta^j \zeta_t^j \varepsilon_t^{p-j} - \beta^p \zeta_t^j \delta_t^{p-j}) = \sum_{j=0}^{p-1} \binom{p}{j} \zeta_t^j (\beta^j \varepsilon_t^{p-j} - \beta^p \delta_t^{p-j}).$$

It will follow from (4.1) and Assumption 1 that Y_t and X_t are cointegrated $I(d)$ processes.

Given observations x_t, y_t, $t = 1, \ldots, n$, the NBLS estimate of Robinson (1994a) for θ (and thus, at least up to unknown sign, β) is given by

$$\widehat{\theta}_m = \frac{\mathrm{Re}\left\{\widehat{F}_{XY}(\lambda_m)\right\}}{\widehat{F}_{XX}(\lambda_m)}, \qquad 1 \leq m \leq \frac{n}{2}, \tag{4.2}$$

where $\lambda_j = 2\pi j/n$ are the Fourier frequencies, and for generic scalar sequences a_t, b_t, $t = 1, \ldots, n$, we define the discretely averaged (cross-) periodogram

$$\widehat{F}_{ab}(\lambda_m) = \frac{2\pi}{n} \sum_{j=1}^{m} I_{ab}(\lambda_j),$$

where $I_{ab}(\lambda) = w_a(\lambda)\overline{w_b(\lambda)}$ is the (cross-) periodogram and $w_a(\lambda) = \sum_{t=1}^{n} a_t e^{it\lambda}/\sqrt{2\pi n}$ is the discrete Fourier transform of $a_1, \ldots, a_n$.

For $m = [n/2]$, where $[\cdot]$ denotes integer part, (4.2) reduces to OLS, but for consistency we require, on the contrary:

Assumption 2 *The bandwidth sequence $m = m(n)$ satisfies*

$$\frac{1}{m} + \left(\frac{m}{n}\right)^{\epsilon} \log n \to 0 \text{ as } n \to \infty, \tag{4.3}$$

for all $\epsilon > 0$.

This assumption is slightly stronger than that of Robinson (1994a,b), namely

$$\frac{1}{m} + \frac{m}{n} \to 0 \text{ as } n \to \infty. \tag{4.4}$$

We need (4.3) over (4.4) only in order to handle powers of g_t, h_t, l_t with particular combinations of memory parameters and Hermite ranks, notably for $d = 1/4$. This case presents no special problems with the method of proof in Robinson (1994a), and is excluded in Robinson (1994b).

For integers $j \in [1, p-1]$ and $k \in [0, p-1]$, denote the Hermite rank of centered g^j, h^{p-k}, l^{p-k} by r_{gj}, r_{hk}, r_{lk} respectively, and introduce the sets

$$\begin{aligned}
S_g &= \left\{j : \beta^j E(\eta_{1t}^j \varepsilon_t^{p-j}) \neq \beta^p E(\eta_{1t}^j \delta_t^{p-j}), \quad 0 < j < p\right\}, \\
S_h &= \left\{k : E(\nu_{1t}^{p-k} \zeta_t^k) \neq 0, \quad 0 \leq k < p\right\}, \\
S_l &= \left\{k : E(\xi_{1t}^{p-k} \zeta_t^k) \neq 0, \quad 0 \leq k < p\right\}, \\
S_{gh} &= \left\{j : E(\eta_{1t}^j \nu_{1t}^{p-j}) \neq 0, \quad 0 < j < p\right\}, \\
S_{gl} &= \left\{j : E(\eta_{1t}^j \xi_{1t}^{p-j}) \neq 0, \quad 0 < j < p\right\}.
\end{aligned}$$

Using the convention that the maximum over an empty set is $-\infty$, define

$$d_g^* = \max_{j \in S_g} \left\{\frac{1}{2} - \left(\frac{1}{2} - d\right) r_{gj}\right\}, \tag{4.5}$$

$$d_h^* = \max_{k \in S_h} \left\{\frac{1}{2} - \left(\frac{1}{2} - d'\right) r_{hk}\right\} - 1(d' = 0), \tag{4.6}$$

$$d_l^* = \max_{k \in S_l} \left\{\frac{1}{2} - \left(\frac{1}{2} - d''\right) r_{lk}\right\} - 1(d'' = 0), \tag{4.7}$$

$$d_{gh}^* = \max_{j \in S_{gh}} \left\{\frac{1}{2} - r_{gj}\left(\frac{1}{2} - d\right) - r_{hj}\left(\frac{1}{2} - d'\right)\right\}, \tag{4.8}$$

$$d_{gl}^* = \max_{j \in S_{gl}} \left\{\frac{1}{2} - r_{gj}\left(\frac{1}{2} - d\right) - r_{lj}\left(\frac{1}{2} - d''\right)\right\}, \tag{4.9}$$

where $1(\cdot)$ throughout denotes the identity function, and

$$d^* = \max\{d_g^*, d_h^*, d_l^*, d_{gh}^*, d_{gl}^*\}. \tag{4.10}$$

Theorem 2 *Under Assumptions 1 and 2, as $n \to \infty$*

$$\widehat{\theta}_m - \theta = O_p\left(\left(\frac{m}{n}\right)^{d-d_u}\right), \tag{4.11}$$

where $d_u = d^* 1(d^* > 0) + \epsilon 1(d^* = 0)$, for any $\epsilon > 0$.

Proof. As in Section 5.3 of Robinson (1994a),

$$|\widehat{\theta}_m - \theta| \le \left\{ \frac{\widehat{F}_{UU}(\lambda_m)}{\widehat{F}_{XX}(\lambda_m)} \right\}^{\frac{1}{2}}.$$

By Proposition 2,

$$\left(\frac{m}{n}\right)^{2d_u - 1} \widehat{F}_{UU}(\lambda_m) = O_p(1),$$

while by Propositions 1, 3 and Slutsky's Theorem,

$$\frac{\left(\frac{m}{n}\right)^{1-2d}}{\widehat{F}_{XX}(\lambda_m)} \xrightarrow{p} \frac{1}{C^*} < \infty. \; \square$$

Since ϵ is arbitrarily small and $d^* < d$, it follows that $\widehat{\theta}_m$ is consistent for θ. Moreover, when $d^* > 0$, we can write $d - d_u = d - d^*$, which is the difference between the integration orders of X_t and U_t, where the rate in (4.11) corresponds to that of Robinson and Marinucci (2003).

5 Simulations

We now present a Monte Carlo study of finite sample performance. For linear processes, Robinson and Marinucci (2003) reported simulation experiments of NBLS with $I(1)$ observables and $I(0)$ cointegrating errors, while Marinucci and Robinson (2001) explored different cases of $I(d_x)$ nonstationary observables and $I(d_e)$ stationary errors. Bandi and Perron (2004) examined NBLS for the regression between realized and implied volatility, generating the data from a discretized continuous time SV model. We employ 50,000 replications of series of various lengths n generated by (1.1), (1.2), (3.5), (3.6), setting $\beta = 1$. All basic processes in (3.5), (3.6) are independent of each other and standard Gaussian. Processes η_{1t}, ν_{1t}, ξ_{1t} in (3.5) were generated as iid, while for the ones in (3.6) the Davies and Harte (1987) algorithm was used to generate η_{2t} as ARFIMA$(0, d, 0)$ and ν_{2t}, ξ_{2t} as ARFIMA$(0, d', 0)$. In most cases h and l are constant functions and ν_{2t}, ξ_{2t} are not required. For all functions g considered, $p = 2$ holds.

We compare the performance of NBLS (4.2) with OLS estimates obtained either from levels,

$$\widetilde{\beta} = \frac{\sum(x_t - \overline{x})(y_t - \overline{y})}{\sum(x_t - \overline{x})^2}, \tag{5.1}$$

or squares,

$$\widetilde{\theta} = \frac{\sum(X_t - \overline{X})(Y_t - \overline{Y})}{\sum(X_t - \overline{X})^2}. \tag{5.2}$$

To make the three estimates comparable, we take $\widetilde{\beta}^2$ as an estimate of θ. We report the bias, standard deviation (SD) and root mean squared error (RMSE) for each estimate. On occasion, relative quantities are reported, meaning the ratio between the corresponding quantity for NBLS and (5.2), which dominates (5.1) in every experiment.

Bandwidth choice

Theorem 2 highlights the relationship between bandwidth m and rate of convergence. In the first experiment, we present the evolution of relative bias, SD and RMSE for different m and memory parameters. We set $n = 256$, $d = 0.1$, 0.2, 0.3, 0.4, $g(x) = \exp(kx)$, with k chosen to satisfy $\mathrm{Var}(\zeta_t) = 2$, and $h(x) = l(x) = 1$. We chose this value for $\mathrm{Var}(\zeta_t)$ in several experiments in order to balance the contributions of bias and SD to RMSE; the impact of the signal to noise ratio is explored later.

Figure 1 shows the bias reduction achieved by NBLS relative to OLS. Not surprisingly, it is greater for small m and large d. It is only around frequency zero that the spectral density of X_t dominates that of U_t; frequencies further from the origin are more contaminated by the correlation between X_t and U_t and contribute more to bias. Also, a higher d indicates a stronger cointegrating relationship, increasing the spectral density of X_t around the origin and thus the averaged periodogram.

The increase in SD of NBLS relative to OLS, displayed in Figure 2, is a consequence of discarding high frequency information, and is decreasing in m. The influence of d on SD appears to be small, specially if compared to Figure 1.

The different profiles of bias and SD give rise to the traditional trade-off in bandwidth

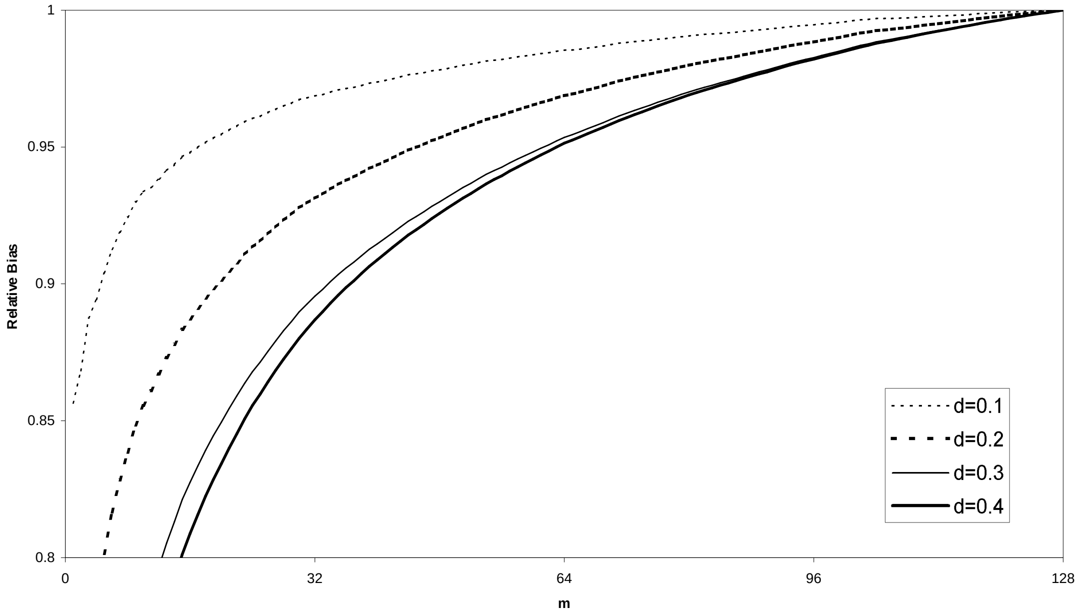

Figure 1: Relative bias of NBLS versus OLS, for varying m and d.

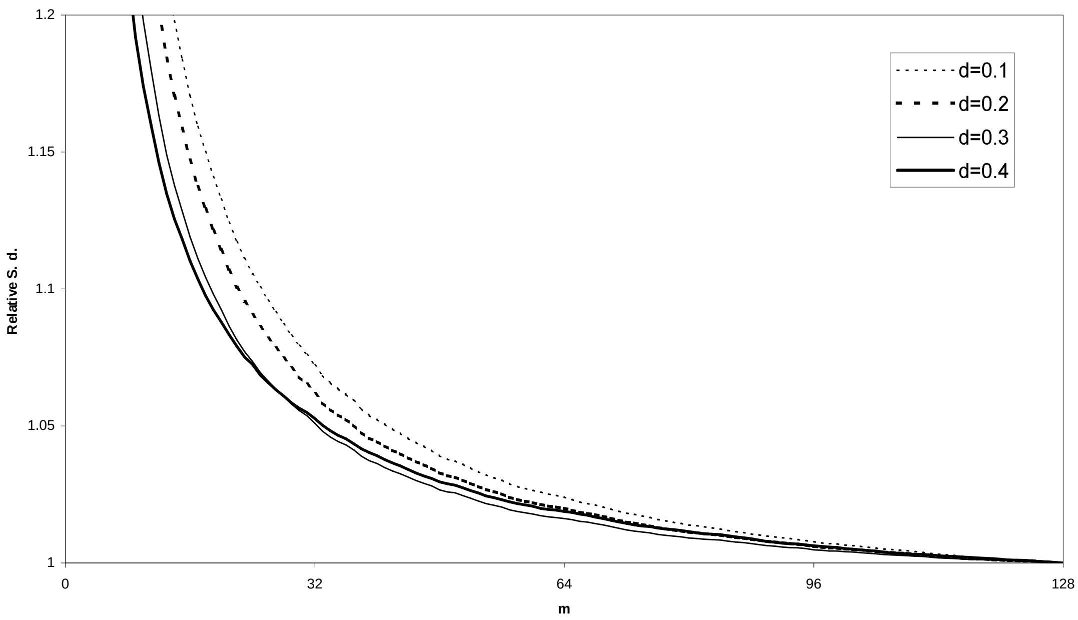

Figure 2: Relative SD of NBLS versus OLS, for varying m and d.

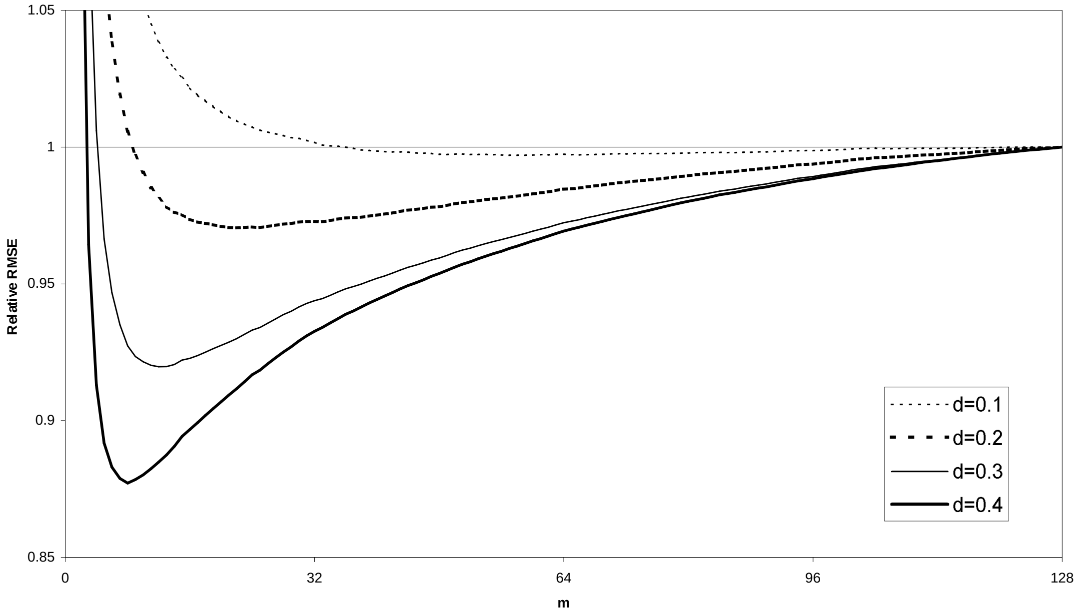

Figure 3: Relative RMSE of NBLS versus OLS, for varying m and d.

choice. Figure 3 presents the relative RMSE of NBLS. For most m, NBLS dominates OLS. For this particular n, a low d does not provide enough information for NBLS to work, due to the modest bias reductions displayed in Figure 1, making the improvement over OLS negligible. The RMSE is essentially a flat function of m, implying that any m above a certain threshold, thereby taking in OLS, attains similar RMSE. However, note that an increase in n should have a similar effect to an increase in d on RMSE, although it will be minimized at a different m. This effect is explored in the next subsection. Higher d lead to very low values for the optimal m, and more significant improvements in RMSE. For $d = 0.4$, a noticeable reduction is already achieved, of over 10% for a number of different m. It should also be noted that if the bandwidth selection is larger than optimal, it is still possible to considerably reduce RMSE, while choosing too small an m can lead to an undesirably large SD.

Memory in signal

We now investigate the influence of n and d on the performance of the three estimates. We consider $n = 256$, 512, 1024, 2048 and $d = 0.1$, 0.2, 0.3, 0.4. As before, $g(x) =$

$\exp(kx)$, with k chosen to satisfy $\mathrm{Var}(\zeta_t) = 2$, and $h(x) = l(x) = 1$. In this experiment and in the following ones, we evaluate NBLS at the bandwidth m^* that minimizes RMSE. Although this is not a feasible choice in the usual sense, it gives an indication of potential gains. Table 1 summarizes the results.

As expected, the RMSE of all estimates improves with n. For even moderate n, NBLS has the lowest RMSE, being clearly less biased than OLS; while OLS in levels attains the lowest SD, especially for small n, its larger bias makes it the worst.

Both bias and SD of OLS increase with d. Both also decrease with n, but while SD seems to be rapidly converging to zero, bias decreases rather slowly and appears to stabilize at some substantial non-zero value. For small n, changes in d similarly affect NBLS, but for larger n, the small m^* makes bias decrease with d.

The bias reduction of NBLS becomes quite large with n, while the variance penalty is always of small magnitude. In fact, when $d = 0.4$ and $n = 2048$, NBLS actually dominates OLS in both SD and bias. The improvement in performance for high d and the rate of decay of RMSE seem compatible with the asymptotic result of Theorem 2.

While Figure 3 and Table 1 both illustrate the high sensitivity of m^* to d, caused by the different scope for bias reduction in each case, m^* does not appear to grow with n. This is surely a small sample effect, as NBLS is only consistent if $m \to \infty$. As a consequence, m^* will diverge when bias becomes negligible compared to SD, a situation which does not occur in the sample sizes considered. Since Theorem 2 shows that convergence of $\widehat{\theta}_m$ is faster the slower m grows, this phenomenon is not entirely surprising.

Memory in signal and noise

In Table 2, d is kept constant, while we introduce long memory in the errors. We set $g(x) = \exp(k_1 x)$ and $h(x) = l(x) = \exp(k_2 x)$, with k_1, k_2 chosen to satisfy $\mathrm{Var}(\zeta_t) = 10$ and $\mathrm{Var}(\delta_t) = \mathrm{Var}(\varepsilon_t) = 2$. These values were again chosen to balance contributions of bias and SD to RMSE. We consider $n = 256$, 512, 1024, 2048, $d = 0.4$ and $d' = 0$, 0.1, 0.2, 0.3.

The results are very similar to the previous experiment, but here $d - d'$ takes the role of d. As before, RMSE improves with n for all estimates. OLS displays similar patterns of bias and SD across $d - d'$ and n, with the exception that SD decays much more slowly with n. The bias of NBLS decreases with $d - d'$ for all n; for $n > 256$ even

	n	256				512				1024				2048			
d		m^*	Bias	SD	RMSE	m^*	Bias	SD	RMSE	m^*	Bias	SD	RMSE	m^*	Bias	SD	RMSE
	$\widetilde{\beta}^2$	—	-0.560	0.085	0.567	—	-0.558	0.062	0.561	—	-0.557	0.045	0.559	—	-0.556	0.033	0.557
0.1	$\widetilde{\theta}$	—	-0.269	0.179	0.323	—	-0.244	0.141	0.282	—	-0.226	0.111	0.252	—	-0.213	0.087	0.230
	$\widehat{\theta}_{m^*}$	58	-0.264	0.184	0.322	53	-0.234	0.149	0.278	50	-0.212	0.123	0.245	64	-0.197	0.097	0.220
	$\widetilde{\beta}^2$	—	-0.564	0.098	0.573	—	-0.561	0.075	0.566	—	-0.559	0.057	0.562	—	-0.558	0.044	0.559
0.2	$\widetilde{\theta}$	—	-0.279	0.184	0.334	—	-0.251	0.145	0.290	—	-0.231	0.115	0.258	—	-0.216	0.090	0.234
	$\widehat{\theta}_{m^*}$	22	-0.253	0.203	0.324	23	-0.215	0.164	0.270	22	-0.184	0.136	0.228	24	-0.160	0.111	0.194
	$\widetilde{\beta}^2$	—	-0.577	0.127	0.591	—	-0.571	0.107	0.581	—	-0.568	0.090	0.575	—	-0.564	0.076	0.570
0.3	$\widetilde{\theta}$	—	-0.309	0.199	0.367	—	-0.276	0.160	0.319	—	-0.252	0.129	0.283	—	-0.233	0.102	0.254
	$\widehat{\theta}_{m^*}$	12	-0.246	0.232	0.338	12	-0.191	0.188	0.268	14	-0.152	0.147	0.212	13	-0.113	0.121	0.165
	$\widetilde{\beta}^2$	—	-0.611	0.181	0.637	—	-0.604	0.168	0.627	—	-0.599	0.156	0.619	—	-0.594	0.146	0.611
0.4	$\widetilde{\theta}$	—	-0.404	0.239	0.469	—	-0.368	0.208	0.423	—	-0.339	0.183	0.385	—	-0.314	0.159	0.352
	$\widehat{\theta}_{m^*}$	8	-0.291	0.290	0.411	7	-0.205	0.247	0.321	8	-0.148	0.192	0.242	8	-0.096	0.147	0.175

Table 1: Monte Carlo bias, SD, RMSE for varying n and d.

	n	256				512				1024				2048			
d'		m^*	Bias	SD	RMSE	m^*	Bias	SD	RMSE	m^*	Bias	SD	RMSE	m^*	Bias	SD	RMSE
	$\widetilde{\beta}^2$	—	-0.520	0.276	0.588	—	-0.498	0.261	0.562	—	-0.476	0.246	0.536	—	-0.455	0.232	0.511
0.0	$\widetilde{\theta}$	—	-0.281	0.318	0.424	—	-0.225	0.277	0.357	—	-0.176	0.232	0.291	—	-0.133	0.189	0.231
	$\widehat{\theta}_{m^*}$	12	-0.228	0.327	0.398	11	-0.156	0.266	0.308	9	-0.097	0.205	0.227	8	-0.054	0.145	0.155
	$\widetilde{\beta}^2$	—	-0.519	0.276	0.588	—	-0.497	0.261	0.561	—	-0.475	0.247	0.536	—	-0.455	0.232	0.511
0.1	$\widetilde{\theta}$	—	-0.280	0.318	0.423	—	-0.224	0.276	0.356	—	-0.175	0.232	0.291	—	-0.133	0.189	0.231
	$\widehat{\theta}_{m^*}$	14	-0.236	0.327	0.403	11	-0.162	0.271	0.316	9	-0.104	0.211	0.235	9	-0.062	0.152	0.164
	$\widetilde{\beta}^2$	—	-0.516	0.278	0.586	—	-0.495	0.262	0.560	—	-0.474	0.247	0.535	—	-0.454	0.233	0.510
0.2	$\widetilde{\theta}$	—	-0.276	0.317	0.421	—	-0.222	0.276	0.354	—	-0.174	0.232	0.290	—	-0.132	0.189	0.231
	$\widehat{\theta}_{m^*}$	19	-0.245	0.326	0.408	13	-0.173	0.277	0.327	13	-0.120	0.219	0.250	13	-0.078	0.164	0.181
	$\widetilde{\beta}^2$	—	-0.506	0.282	0.579	—	-0.487	0.266	0.555	—	-0.468	0.251	0.530	—	-0.449	0.236	0.508
0.3	$\widetilde{\theta}$	—	-0.266	0.316	0.413	—	-0.215	0.274	0.348	—	-0.169	0.231	0.286	—	-0.129	0.189	0.229
	$\widehat{\theta}_{m^*}$	35	-0.251	0.323	0.409	26	-0.189	0.278	0.336	27	-0.141	0.227	0.267	35	-0.102	0.180	0.206

Table 2: Monte Carlo bias, SD, RMSE for varying n and d', with $d = 0.4$.

n	256			512			1024			2048		
S2N	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
0.5	0.994	1.280	0.998	0.981	1.960	0.992	0.961	2.924	0.980	0.927	4.141	0.956
1	0.910	1.406	0.959	0.818	1.682	0.902	0.737	1.773	0.829	0.621	1.955	0.731
2	0.797	1.163	0.920	0.691	1.175	0.839	0.604	1.144	0.749	0.485	1.181	0.650
4	0.834	1.050	0.969	0.738	1.057	0.922	0.652	1.048	0.857	0.557	1.059	0.781
8	0.996	1.000	1.000	0.879	1.013	0.989	0.773	1.019	0.965	0.683	1.016	0.922
		Bias / SD			Bias / SD			Bias / SD			Bias / SD	
S2N	m^*	OLS	NBLS	m^*	OLS	NBLS	m^*	OLS	NBLS	m^*	OLS	NBLS
0.5	52	-9.29	-7.22	27	-11.42	-5.72	17	-14.19	-4.66	12	-17.26	-3.86
1	12	-3.38	-2.19	8	-3.72	-1.81	8	-4.14	-1.72	7	-4.69	-1.49
2	12	-1.55	-1.06	12	-1.73	-1.02	14	-1.95	-1.03	13	-2.28	-0.94
4	26	-0.82	-0.65	28	-0.93	-0.65	33	-1.09	-0.68	35	-1.31	-0.69
8	121	-0.43	-0.43	87	-0.49	-0.42	86	-0.57	-0.43	90	-0.69	-0.47

Table 3: Monte Carlo relative bias, SD, RMSE of NBLS versus OLS, for varying n and S2N, with $d = 0.3$.

SD decreases with $d - d'$. A surprising fact in this case is related to the variance/bias trade-off of NBLS. While this can be found in small samples, as n increases it starts dominating OLS in both bias and variance. The evolution of m^* is also similar to the previous section.

Signal to noise ratio

This experiment investigates the influence of the signal to noise (S2N) ratio on the performance of NBLS. We use $g(x) = \exp(kx)$, such that $\operatorname{Var}(\zeta_t) = 2$, and $h(x) = l(x) = \sigma$, so that $\operatorname{Var}(\delta_t) = \operatorname{Var}(\varepsilon_t) = \sigma^2$, for $\sigma^2 = 0.25$, 0.5, 1, 2, 4. The results obtained for different d were qualitatively similar, so we report results only for $d = 0.3$ and $n = 256$, 512, 1024, 2048. Since it is unreasonable to compare absolute performance for different S2N ratios, in Table 3 we focus on relative performance only. We also report the ratio between bias and SD. Although we refer to $\operatorname{Var}(\zeta_t)/\operatorname{Var}(\delta_t)$ as the S2N ratio, for simplicity, it is only an accurate description for the regression in levels. For $x_t^2 = \zeta_t^2 + 2\zeta_t\delta_t + \delta_t^2$, the dominant term is ζ_t^2, and even there η_{1t}^2 could be considered a multiplicative noise. Hence, the definition of the "true" S2N ratio would be ambiguous, but it would be arguably smaller than the one in levels.

NBLS performs best when bias and SD are balanced. The regressor X_t consists of two parts: a long memory component containing a dominating pole at frequency zero, and a component with less memory not orthogonal to U_t. In this case, it is actually short memory, since δ_t is iid. If the S2N ratio is very large, the first component will dominate the second even at frequencies distant from zero. As a result, any large enough m will perform well, and even with OLS, bias will contribute very little to RMSE and gains from NBLS will be small. On the other hand, for very small S2N, the second component will be relatively large, dominating the signal even at frequencies close to zero. In small samples, an attempt to reduce bias by only choosing informative frequencies would imply the use of very small m, which would force SD to be too high (see Figure 2). In this case, NBLS would also provide little gains, as the cost (in terms of SD) of reducing bias is too high for RMSE.

With OLS the ratio between bias and SD increases with n. This is expected, since OLS still converges in probability to a constant. In NBLS, the ratio is very close to that of OLS in small samples. From that point, it increases with n if it was originally small, but decreases if it was originally large. It appears that this ratio will stabilize at some value close to unity for large enough n, and from that point on NBLS will have a noticeable RMSE improvement over OLS.

Nonlinearity

To investigate the influence of nonlinearity on NBLS, Table 4 reports its performance in three different settings, for $n = 256$, 512, 1024, 2048 and $d = 0.1$, 0.2, 0.3, 0.4. The nonlinear setting (NL), already used in the first two subsections, has $g(x) = \exp(kx)$, with k chosen to satisfy $\mathrm{Var}(\zeta_t) = 2$, and $h(x) = l(x) = 1$. In the other two we deviate from (1.1), (1.2), (3.5), (3.6), using instead $Y_t = X_t + u_t$, $X_t = f_t + v_t$, where u_t, v_t are generated as iid mean zero Gaussian with $\mathrm{Var}(u_t) = 20$, $\mathrm{Var}(v_t) = 6$, and $\mathrm{Cov}(u_t, v_t) = -10$. In a fully linear setting (L), we generate f_t as a Gaussian mean zero ARFIMA$(0, d, 0)$, with $\mathrm{Var}(f_t) = 44$. In a linear setting with a multiplicative noise (MN), we set $f_t = \eta_{1t}^2 z_t$, where η_{1t} is iid standard Gaussian while z_t is independently generated as a Gaussian ARFIMA$(0, d, 0)$, with $E(z_t) = 2$ and $\mathrm{Var}(z_t) = 12$. The chosen moments replicate those of corresponding processes in the nonlinear setting.

Both OLS and NBLS perform much better under L than NL, while performance under MN falls in the middle. A similar ordering is found in relative performance (not shown), since a relatively stable, large bias of OLS estimates throughout makes variations

	n	256				512				1024				2048			
d		m^*	Bias	SD	RMSE	m^*	Bias	SD	RMSE	m^*	Bias	SD	RMSE	m^*	Bias	SD	RMSE
	L	20	-0.151	0.086	0.174	19	-0.132	0.083	0.156	23	-0.120	0.072	0.140	24	-0.106	0.066	0.125
0.1	MN	41	-0.207	0.088	0.225	37	-0.190	0.079	0.206	41	-0.179	0.069	0.192	34	-0.167	0.070	0.181
	NL	58	-0.264	0.184	0.322	53	-0.234	0.149	0.278	50	-0.212	0.123	0.245	64	-0.197	0.097	0.220
	L	13	-0.101	0.090	0.136	16	-0.083	0.073	0.111	17	-0.065	0.063	0.091	22	-0.054	0.050	0.074
0.2	MN	17	-0.178	0.111	0.210	17	-0.152	0.098	0.181	19	-0.133	0.085	0.158	20	-0.114	0.077	0.138
	NL	22	-0.253	0.203	0.324	23	-0.215	0.164	0.270	22	-0.184	0.136	0.228	24	-0.160	0.111	0.194
	L	11	-0.072	0.084	0.111	14	-0.054	0.064	0.084	17	-0.040	0.050	0.063	22	-0.030	0.038	0.048
0.3	MN	12	-0.151	0.121	0.193	13	-0.118	0.099	0.154	15	-0.094	0.081	0.124	15	-0.069	0.070	0.098
	NL	12	-0.246	0.232	0.338	12	-0.191	0.188	0.268	14	-0.152	0.147	0.212	13	-0.113	0.121	0.165
	L	9	-0.061	0.087	0.106	12	-0.042	0.062	0.075	15	-0.028	0.045	0.053	20	-0.019	0.032	0.037
0.4	MN	8	-0.133	0.140	0.193	10	-0.098	0.105	0.143	11	-0.067	0.082	0.106	13	-0.046	0.062	0.077
	NL	8	-0.291	0.290	0.411	7	-0.205	0.247	0.321	8	-0.148	0.192	0.242	8	-0.096	0.147	0.175

Table 4: Monte Carlo bias, SD, RMSE of NBLS in settings L, MN, NL, for varying n and d.

d	g	$\exp(x)$				$(1+x)^2$				$\lvert 1+x \rvert$			
		m^*	Bias	SD	RMSE	m^*	Bias	SD	RMSE	m^*	Bias	SD	RMSE
	$\widetilde{\beta}^2$	—	-0.558	0.062	0.561	—	-0.557	0.059	0.560	—	-0.556	0.056	0.559
0.1	$\widetilde{\theta}$	—	-0.244	0.141	0.282	—	-0.262	0.122	0.289	—	-0.291	0.112	0.312
	$\widehat{\theta}_{m^*}$	53	-0.234	0.149	0.278	47	-0.250	0.136	0.284	52	-0.280	0.128	0.308
	$\widetilde{\beta}^2$	—	-0.561	0.075	0.566	—	-0.560	0.072	0.564	—	-0.559	0.066	0.563
0.2	$\widetilde{\theta}$	—	-0.251	0.145	0.290	—	-0.267	0.125	0.295	—	-0.295	0.114	0.316
	$\widehat{\theta}_{m^*}$	23	-0.215	0.164	0.270	17	-0.219	0.159	0.271	23	-0.255	0.148	0.295
	$\widetilde{\beta}^2$	—	-0.571	0.107	0.581	—	-0.570	0.105	0.580	—	-0.566	0.094	0.574
0.3	$\widetilde{\theta}$	—	-0.276	0.160	0.319	—	-0.285	0.136	0.316	—	-0.308	0.122	0.332
	$\widehat{\theta}_{m^*}$	12	-0.191	0.188	0.268	12	-0.190	0.174	0.258	13	-0.220	0.174	0.280
	$\widetilde{\beta}^2$	—	-0.604	0.168	0.627	—	-0.603	0.172	0.627	—	-0.593	0.152	0.612
0.4	$\widetilde{\theta}$	—	-0.368	0.208	0.423	—	-0.358	0.181	0.401	—	-0.359	0.150	0.390
	$\widehat{\theta}_{m^*}$	7	-0.205	0.247	0.321	8	-0.190	0.219	0.290	8	-0.205	0.219	0.299

Table 5: Monte Carlo bias, SD, RMSE for varying $g(x)$ and d, with $n = 512$.

in RMSE smaller than for NBLS. Although some of the gap in performance should be a consequence of nonlinearity, significant excess kurtosis in NL and MN is arguably the dominant factor, since it directly affects the variance of the periodogram. In MN, the kurtosis of f_t is around 77, while in NL it is around 3523 for f_t, 36 for v_t and 30 for u_t.

Volatility function

Finally, we explore the impact of the functional form of the volatility function g. considering $g(x) = \exp(kx)$, $(1 + kx)^2$, $|1 + kx|$, with k chosen in each case so that $\mathrm{Var}(\zeta_t) = 2$. We set $h(x) = l(x) = 1$ and $d = 0.1, 0.2, 0.3, 0.4$. Table 5 presents the results for $n = 512$, where the properties of each estimate seem robust to the choice of volatility function. Normalizing $\mathrm{Var}(\zeta_t)$ appears to be sufficient to capture most of the differences across functions. Results for other n are similar and available from the authors upon request.

6 Final comments

To our knowledge this paper represents the first treatment of fractional cointegration in the context of nonlinear processes. The stationary environment, the SV models employed, and the NBLS estimate seem well motivated by applications in finance. Our model is semiparametric both in the sense that only assumptions about low frequency behavior are required, and the volatility functions are nonparametric. While the nonlinear setting necessitates a considerably more complex proof of consistency of NBLS than earlier ones, a comparable result is obtained, with rate of convergence depending essentially on the strength of the cointegrating relation, namely the gap between integration orders of observables and cointegrating error. Monte Carlo results show encouraging performances in moderate sample sizes across a variety of specifications.

As always, consistency results are reassuring only in very large data sets. Though these do exist in finance, one would like a limit distributional result that could be used in statistical inference. Christensen and Nielsen (2004) have achieved this in a simpler setting, indeed with regressor and disturbance assumed incoherent at frequency zero and linear process (on conditionally homoscedastic md innovations) assumptions. In general, not only is the proof likely to be much more complicated than even our proof of Theorem 2, but the limit distribution is likely to be non-standard for various combinations of memory parameters, though a bootstrap procedure might be investigated. By analogy with experience in $I(1)/I(0)$ cointegrated models (e.g. Johansen, 1991; Phillips, 1991), and nonstationary fractionally cointegrated models (e.g. Jeganathan, 1999; Robinson and Hualde, 2003), it may be possible to obtain estimates with nicer asymptotic distributional properties, in particular leading to Wald statistics with null limiting χ^2 distributions. However, in our nonlinear setting it is not immediately obvious that the sort of transformations used in those references to achieve the necessary "whitening" will be successful, the estimates would require preliminary estimation of memory parameters, and proofs would be significantly more complicated. Nevertheless, those wishing to embark on limit distributional proofs for NBLS or other estimates in our SV setting should find techniques described in the present paper relevant.

Though our Monte Carlo study addressed the choice of bandwidth m, it would evidently be desirable to develop a feasible rule for bandwidth selection. In a Gaussian or linear setting, Robinson (1994b) developed formulae for minimum-MSE bandwidth with respect to the basic averaged periodogram statistic, and these were further analyzed by Delgado and Robinson (1996). In principle these could be extended to the NBLS estimate, though the formulae will be highly complex, and feasible versions would require estimating memory parameters and other quantities. As in other circumstances, sensitivity to choice of m can be assessed by a "window-closing" approach, computing NBLS over a sensibly chosen grid of m values; since discrete Fourier transforms at all Fourier frequencies can be obtained simultaneously by the Fast Fourier Transform, and NBLS is algebraically simple, this can cheaply be achieved, indeed a simple recursion deals with unit or other increases in m.

The bulk of the fractional and non-fractional cointegration literature assumes nonstationary observables. The motivation usually comes from macroeconomics, but nonstationarity can often appear in financial time series also. The modelling of nonstationary series via analogues of (4.1) is itself a somewhat open topic, but given that X_t has a kind of $I(d)$ property, for $d \geq 1/2$, some of the arguments of Robinson and Marinucci (2001) should be relevant in establishing rates of convergence of NBLS. Indeed, these authors, following Stock (1987) in the $I(1)/I(0)$ case, found OLS also to be consistent here, though in some circumstances NBLS has bias of smaller order. The nonstationary X_t case is in some respects technically easier than the stationary one, because consistency of OLS follows from the domination of sums of squares of U_t by those of X_t.

Other directions of research could extend (1.1), (1.2) to more than two observables, and then possibly to allow more than one common factor, i.e. more than one cointegrating relation. On the one hand, cointegrating relations between a potentially large number of asset returns can be of interest, while on the other Fama and French (1993) argued for the inclusion of additional factors in the CAPM. It should be possible to determine a form of multivariate regression linking the observables, analogous to (4.1), and then a multivariate extension for NBLS (4.2). Its consistency, subject to identifiability conditions, can then be established under an analogue of Assumption 1, using Theorem 1 and techniques employed in the proof of Theorem 2, though of course the details would be even more complicated. The issue of determining cointegrating rank is of more pressing concern than in our simple model (1.1), (1.2), but procedures such as those of Robinson and Yajima (2002) might be employed in practice, though again their theoretical justification in our setting would require considerable further work.

Appendix A: Proof of Theorem 1

Throughout the proof, we denote $P = P_J$, $Q = Q_J$ and $R_j = R_{J,j}$, $j \in P$. Furthermore, all sums and products run over P unless otherwise stated. We have

$$E\left(\prod_j f_j\right) = \int_{\mathbb{R}^J} \prod_j f_j \phi_J(\mu;\Omega)\mathrm{d}\mu, \tag{A.1}$$

where $\phi_J(\mu;\Omega)$ denotes the density function of $\mu = (\mu_1, \ldots, \mu_J)'$ and $\Omega = E(\mu\mu')$. From (22) of Slepian (1972) and (2.1), $\phi_J(\mu;\Omega)$ is

$$\begin{aligned}\sum_{v_\alpha=0:\alpha\in Q}^{\infty} \prod_{\alpha\in Q} \frac{\rho_\alpha^{v_\alpha}}{v_\alpha!} \prod_j \left\{\left(\frac{\partial}{\partial\mu_j}\right)^{w_j} \phi(\mu_j)\right\} &= \sum_{v_\alpha=0:\alpha\in Q}^{\infty} \prod_{\alpha\in Q} \frac{\rho_\alpha^{v_\alpha}}{v_\alpha!} \prod_j \left\{(-1)^{w_j} H_{w_j}(\mu_j)\phi(\mu_j)\right\} \\ &= \sum_{v_\alpha=0:\alpha\in Q}^{\infty} \prod_{\alpha\in Q} \frac{\rho_\alpha^{v_\alpha}}{v_\alpha!} \prod_j \left\{H_{w_j}(\mu_j)\phi(\mu_j)\right\}, \end{aligned}\tag{A.2}$$

since $\sum_j w_j = 2\sum_{\alpha\in Q} v_\alpha$ is even. Using (A.2) in (A.1), $E(\prod_j f_j)$ is

$$\begin{aligned}&\int_{\mathbb{R}^J} \prod_j f_j(\mu_j) \sum_{v_\alpha=0:\alpha\in Q}^{\infty} \prod_{\alpha\in Q} \frac{\rho_\alpha^{v_\alpha}}{v_\alpha!} \prod_j \left\{H_{w_j}(\mu_j)\phi(\mu_j)\right\} \mathrm{d}\mu \\ &= \sum_{v_\alpha=0:\alpha\in Q}^{\infty} \prod_{\alpha\in Q} \frac{\rho_\alpha^{v_\alpha}}{v_\alpha!} \int_{\mathbb{R}^J} \prod_j \left\{f_j(\mu_j) H_{w_j}(\mu_j)\phi(\mu_j)\right\} \mathrm{d}\mu \\ &= \sum_{v_\alpha=0:\alpha\in Q}^{\infty} \prod_{\alpha\in Q} \frac{\rho_\alpha^{v_\alpha}}{v_\alpha!} \prod_j E\left\{f_j(\mu_j) H_{w_j}(\mu_j)\right\} = \sum_{v_\alpha=0:\alpha\in Q}^{\infty} \prod_j G_{j,w_j} \prod_{\alpha\in Q} \frac{\rho_\alpha^{v_\alpha}}{v_\alpha!}.\end{aligned}$$

This proves (2.3). For the remainder of the proof, we use the Cauchy-Schwarz inequality in

$$\begin{aligned}|a_q| &\le \sum_{\substack{v_\alpha\ge 0: \\ \Sigma v_\alpha = q, \alpha\in Q}} \left|\prod_j G_{j,w_j} \prod_{\alpha\in Q} \frac{\rho_\alpha^{v_\alpha}}{v_\alpha!}\right| \\ &\le \sum_{\substack{v_\alpha\ge 0: \\ \Sigma v_\alpha = q, \alpha\in Q}} \prod_j \frac{|G_{j,w_j}|}{\sqrt{w_j!}} \prod_j \sqrt{w_j!} \prod_{\alpha\in Q} \frac{|\rho_\alpha|^{v_\alpha}}{v_\alpha!} \le (A_q B_q)^{\frac{1}{2}}, \end{aligned}\tag{A.3}$$

where

$$A_q = \sum_{\substack{v_\alpha \geq 0: \\ \Sigma v_\alpha = q, \alpha \in Q \\ w_j \geq r_j, j \in P}} \prod_j \frac{G_{j,w_j}^2}{w_j!}, \qquad B_q = \sum_{\substack{v_\alpha \geq 0: \\ \Sigma v_\alpha = q, \alpha \in Q \\ w_j \geq r_j, j \in P}} \prod_j \left(w_j! \prod_{\alpha \in R_j} \frac{|\rho_\alpha|^{v_\alpha}}{v_\alpha!} \right).$$

The A_q term is bounded since

$$A_q \leq \prod_j \sum_{w_j = r_j}^{\infty} \frac{G_{j,w_j}^2}{w_j!} \leq \prod_j E(f_j^2) \leq \sigma^2. \tag{A.4}$$

If $2q < r$, there always exists a j in (2.3) such that $w_j < r_j$, implying (2.4). For $2q \geq r$, the multinomial theorem yields

$$\begin{aligned}
B_q &\leq \sum_{\substack{w_j \geq r_j: \\ \Sigma w_j = 2q, j \in P}} \sum_{\substack{v_\alpha \geq 0: \Sigma v_\alpha = w_j \\ \alpha \in R_j, j \in P}} \prod_j \left(w_j! \prod_{\alpha \in R_j} \frac{|\rho_\alpha|^{v_\alpha}}{v_\alpha!} \right) \\
&\leq \sum_{\substack{w_j \geq r_j: \\ \Sigma w_j = 2q, j \in P}} \prod_j \sum_{\substack{v_\alpha \geq 0: \\ \Sigma v_\alpha = w_j, \alpha \in R_j}} w_j! \prod_{\alpha \in R_j} \frac{|\rho_\alpha|^{v_\alpha}}{v_\alpha!} \leq \sum_{\substack{w_j \geq r_j: \\ \Sigma w_j = 2q, j \in P}} \prod_j \left(\sum_{\alpha \in R_j} |\rho_\alpha| \right)^{w_j} \\
&\leq \prod_j \left(\sum_{\alpha \in R_j} |\rho_\alpha| \right)^{r_j} \sum_{\substack{w_j \geq 0: \\ \Sigma w_j = 2q - r, j \in P}} \prod_j \left(\sum_{\alpha \in R_j} |\rho_\alpha| \right)^{w_j} \\
&\leq \prod_j \left(\sum_{\alpha \in R_j} |\rho_\alpha| \right)^{r_j} \sum_{\substack{w_j \geq 0: \\ \Sigma w_j = 2q - r, j \in P}} (2q - r)! \prod_j \frac{\left(\sum_{\alpha \in R_j} |\rho_\alpha| \right)^{w_j}}{w_j!} \\
&\leq \prod_j \left(\sum_{\alpha \in R_j} |\rho_\alpha| \right)^{r_j} \left(\sum_{j,k: j \neq k} |\rho_{jk}| \right)^{2q - r} \leq \prod_j \left(\sum_{\alpha \in R_j} |\rho_\alpha| \right)^{r_j} \tau^{2q - r}.
\end{aligned} \tag{A.5}$$

Using (A.4), (A.5) in (A.3) gives (2.5). Then (2.6) follows from $\tau < 1$. □

Appendix B: Propositions for Theorem 2

We denote the Dirichlet kernel by $D_m(\lambda) = \sum_{j=1}^{m} e^{ij\lambda}$, for $m \geq 1$, and will use the fact that

$$D_n(\lambda_j) = n1(j = 0, \mathrm{mod}\, n). \tag{B.1}$$

We also use the abbreviating notation

$$S_m(a,b) = E\left\{\widehat{F}_{ab}(\lambda_m)\right\} = \frac{1}{n^2}\sum_{s,t=1}^{n} \mathrm{Cov}(a_s, b_t) D_m(\lambda_{t-s}),$$

from (B.1), and

$$S'_m(a,b;a',b') = \frac{1}{n^2}\sum_{s,t=1}^{n} \mathrm{Cov}(a_s, b_t)\,\mathrm{Cov}(a'_s, b'_t) D_m(\lambda_{t-s}),$$

where a_t, b_t, a'_t, b'_t, $t = 1, \ldots, n$ are scalar sequences with finite second moments.

Proposition 1 *Under (1.2) and Assumptions 1 and 2,*

$$\left(\frac{m}{n}\right)^{2d-1} E\left\{\widehat{F}_{XX}(\lambda_m)\right\} \to C^* \text{ as } n \to \infty,$$

where

$$\begin{aligned} C^* &= 2\frac{(2\pi)^{-2d}\Gamma(2d)}{1-2d}\sin\left\{(1-2d)\frac{\pi}{2}\right\} E(\eta_1^p)^2 E\left\{g^p(\eta_{2t})\eta_{2t}\right\}^2 C_\rho > 0, \\ C_\rho &= \lim_{j\to\infty} E(\eta_{20}\eta_{2j}) j^{1-2d}. \end{aligned}$$

Proof. Write

$$X_t = \sum_{j=0}^{p} A_{jt}B_{jt}, \qquad A_{jt} = \binom{p}{j}\eta_{1t}^j \nu_{1t}^{p-j}, \qquad B_{jt} = g_t^j h_t^{p-j}.$$

Using Lemma 2, since $\{A_{jt}\}$ is independent of $\{B_{kt}\}$, for any j and k,

$$\begin{aligned} \mathrm{Cov}(X_s, X_t) &= \sum \mathrm{Cov}(A_{js}B_{js}, A_{kt}B_{kt}) \\ &= \sum \left\{E(A_{js})E(A_{kt})\,\mathrm{Cov}(B_{js}, B_{kt}) + \mathrm{Cov}(A_{js}, A_{kt})E(B_{js}B_{kt})\right\}, \end{aligned}$$

where $\sum$ denotes $\sum_{j,k=0}^{p}$ throughout the proof.

Now define $a_j = E(A_{jt})$, $b_{g,j} = a_j E(h_t^{p-j})$ and $b_{h,j} = a_j E(g_t^j)$. Since $\{A_{jt}\}$ is iid, using Lemma 2 again, for $s \neq t$, $\mathrm{Cov}(X_s, X_t)$ is

$$\sum a_j a_k \,\mathrm{Cov}(B_{js}, B_{kt}) = \sum \Big\{ b_{g,j} b_{g,k} \,\mathrm{Cov}(g_s^j, g_t^k) + b_{h,j} b_{h,k} \,\mathrm{Cov}(h_s^{p-j}, h_t^{p-k}) + a_j a_k \,\mathrm{Cov}(g_s^j, g_t^k)\,\mathrm{Cov}(h_s^{p-j}, h_t^{p-k}) \Big\}. \tag{B.2}$$

For $s = t$, denote by Λ the difference between $\mathrm{Var}(X_t)$ and (B.2). It follows that $E\{\widehat{F}_{XX}(\lambda_m)\}$ is

$$\sum \left\{ b_{g,j} b_{g,k} S_m(g^j, g^k) + b_{h,j} b_{h,k} S_m(h^{p-j}, h^{p-k}) + a_j a_k S'_m(g^j, g^k; h^{p-j}, h^{p-k}) \right\} + \frac{m}{n}\Lambda.$$

From (3.7), (3.8) and Lemma 4, $b_{g,j} b_{g,k} S_m(g^j, g^k) = o((m/n)^{1-2d})$, if either $j < p$ or $k < p$, while $b_{g,p}^2 S_m(g^p, g^p) = C^* \,(m/n)^{1-2d} + o((m/n)^{1-2d})$. Lemma 4 and $d' < d$ imply that $b_{g,j} b_{g,k} S_m(h^{p-j}, h^{p-k}) = o((m/n)^{1-2d})$, and by Lemma 5, $a_j a_k S'_m(g^j, g^k; h^{p-j}, h^{p-k}) = o((m/n)^{1-2d})$. $\square$

Proposition 2 *Under (1.1), (1.2) and Assumptions 1 and 2,*

$$E\left\{ \widehat{F}_{UU}(\lambda_m) \right\} = O\left(\left(\frac{m}{n}\right)^{1-2d_u} \right).$$

Proof. Write

$$U_t = \sum_{j=0}^{p-1} A_{\varepsilon,jt} B_{\varepsilon,jt} - A_{\delta,jt} B_{\delta,jt},$$

where

$$A_{\varepsilon,jt} = \binom{p}{j} \beta^j \eta_{1t}^j \xi_{1t}^{p-j}, \qquad A_{\delta,jt} = \binom{p}{j} \beta^p \eta_{1t}^j \nu_{1t}^{p-j}, \qquad B_{\varepsilon,jt} = g_t^j l_t^{p-j}, \qquad B_{\delta,jt} = g_t^j h_t^{p-j}.$$

Using Lemma 2 repeatedly, since $\{A_{\varepsilon,jt}\}$, $\{A_{\delta,jt}\}$ are independent of $\{B_{\varepsilon,kt}\}$, $\{B_{\delta,kt}\}$, for any j and k,

$$\begin{aligned}
\mathrm{Cov}(U_s, U_t) &= \sum \{ \mathrm{Cov}(A_{\varepsilon,js} B_{\varepsilon,js}, A_{\varepsilon,kt} B_{\varepsilon,kt}) + \mathrm{Cov}(A_{\delta,js} B_{\delta,js}, A_{\delta,kt} B_{\delta,kt}) \\
&\qquad - \mathrm{Cov}(A_{\varepsilon,js} B_{\varepsilon,js}, A_{\delta,kt} B_{\delta,kt}) - \mathrm{Cov}(A_{\delta,js} B_{\delta,js}, A_{\varepsilon,kt} B_{\varepsilon,kt}) \} \\
&= \sum \{ E(A_{\varepsilon,js}) E(A_{\varepsilon,kt}) \,\mathrm{Cov}(B_{\varepsilon,js}, B_{\varepsilon,kt}) + \mathrm{Cov}(A_{\varepsilon,js}, A_{\varepsilon,kt}) E(B_{\varepsilon,js} B_{\varepsilon,kt})
\end{aligned}$$

$$
\begin{aligned}
&+ E(A_{\delta,js})E(A_{\delta,kt})\operatorname{Cov}(B_{\delta,js},B_{\delta,kt}) + \operatorname{Cov}(A_{\delta,js},A_{\delta,kt})E(B_{\delta,js}B_{\delta,kt})\\
&- E(A_{\varepsilon,js})E(A_{\delta,kt})\operatorname{Cov}(B_{\varepsilon,js},B_{\delta,kt}) - \operatorname{Cov}(A_{\varepsilon,js},A_{\delta,kt})E(B_{\varepsilon,js}B_{\delta,kt})\\
&-E(A_{\delta,js})E(A_{\varepsilon,kt})\operatorname{Cov}(B_{\delta,js},B_{\varepsilon,kt}) - \operatorname{Cov}(A_{\delta,js},A_{\varepsilon,kt})E(B_{\delta,js}B_{\varepsilon,kt})\} .
\end{aligned}
$$

where $\sum$ denotes $\sum_{j,k=0}^{p-1}$ throughout the proof.

Now define $a_{\varepsilon j} = E(A_{\varepsilon,jt})$, $a_{\delta j} = E(A_{\delta,jt})$, $b_{gj} = a_{\varepsilon j}E(l_t^{p-j}) - a_{\delta j}E(h_t^{p-j})$, $b_{hj} = a_{\delta j}E(g_t^j)$ and $b_{lj} = a_{\varepsilon j}E(g_t^j)$. Since $\{A_{\varepsilon,jt}\}$, $\{A_{\delta,jt}\}$ are jointly iid, using Lemma 2 again, for $s \neq t$, $\operatorname{Cov}(U_s,U_t)$ is

$$
\begin{aligned}
&\sum \{a_{\varepsilon j}a_{\varepsilon k}\operatorname{Cov}(B_{\varepsilon,js},B_{\varepsilon,kt}) + a_{\delta j}a_{\delta k}\operatorname{Cov}(B_{\delta,js},B_{\delta,kt})\\
&\quad -a_{\varepsilon j}a_{\delta k}\operatorname{Cov}(B_{\varepsilon,js},B_{\delta,kt}) - a_{\delta j}a_{\varepsilon k}\operatorname{Cov}(B_{\delta,js},B_{\varepsilon,kt})\}\\
=&\sum \Big\{b_{gj}b_{gk}\operatorname{Cov}(g_s^j,g_t^k) + b_{hj}b_{hk}\operatorname{Cov}(h_s^{p-j},h_t^{p-k}) + b_{lj}b_{lk}\operatorname{Cov}(l_s^{p-j},l_t^{p-k})\\
&\quad +a_{\delta j}a_{\delta k}\operatorname{Cov}(g_s^j,g_t^k)\operatorname{Cov}(h_s^{p-j},h_t^{p-k}) + a_{\varepsilon j}a_{\varepsilon k}\operatorname{Cov}(g_s^j,g_t^k)\operatorname{Cov}(l_s^{p-j},l_t^{p-k})\Big\} . \quad \text{(B.3)}
\end{aligned}
$$

For $s = t$, denote by Λ the difference between $\operatorname{Var}(U_t)$ and (B.3). It follows that $E\{\widehat{F}_{UU}(\lambda_m)\}$ is

$$
\begin{aligned}
&\sum \{b_{gj}b_{gk}S_m(g^j,g^k) + b_{hj}b_{hk}S_m(h^{p-j},h^{p-k}) + b_{lj}b_{lk}S_m(l^{p-j},l^{p-k})\\
&\quad +a_{\delta j}a_{\delta k}S'_m(g^j,g^k;h^{p-j},h^{p-k}) + a_{\varepsilon j}a_{\varepsilon k}S'_m(g^j,g^k;l^{p-j},l^{p-k})\} + \frac{m}{n}\Lambda.
\end{aligned}
$$

By (4.5), applying Lemma 4 to each (j,k) pair with non-zero coefficient,

$$
b_{gj}b_{gk}S_m(g^j,g^k) = O\left((m/n)^{1-2\max\{d_g^*,0\}}(\log n)^{1(d_g^*=0)}\right).
$$

Similarly, by (4.6) and (4.7), Lemma 4 yields

$$
\begin{aligned}
b_{hj}b_{hk}S_m(h^{p-j},h^{p-k}) &= O\left((m/n)^{1-2\max\{d_h^*,0\}}(\log n)^{1(d_h^*=0)}\right),\\
b_{lj}b_{lk}S_m(l^{p-j},l^{p-k}) &= O\left((m/n)^{1-2\max\{d_l^*,0\}}(\log n)^{1(d_l^*=0)}\right).
\end{aligned}
$$

Finally, Lemma 5, (4.8) and (4.9) give

$$
\begin{aligned}
a_{\delta j}a_{\delta k}S'_m(g^j,g^k;h^{p-j},h^{p-k}) &= O\left((m/n)^{1-2\max\{d_{gh}^*,0\}}(\log n)^{1(d_{gh}^*=0)}\right),\\
a_{\varepsilon j}a_{\varepsilon k}S'_m(g^j,g^k;l^{p-j},l^{p-k}) &= O\left((m/n)^{1-2\max\{d_{gl}^*,0\}}(\log n)^{1(d_{gl}^*=0)}\right).
\end{aligned}
$$

By (3.8), $d_g^* < d$. Since d_h^* and d_{gh}^* are bounded by $d' < d$ while d_l^* and d_{gl}^* are bounded

by $d'' < d$, we have $d^* < d$. The bound for $d^* = 0$ follows from Assumption 2. □

Proposition 3 *Under (1.2) and Assumptions 1 and 2,*

$$\operatorname{Var}\left\{\widehat{F}_{XX}(\lambda_m)\right\} = o\left(\left(\frac{m}{n}\right)^{2-4d}\right).$$

Proof. Define $\rho_t = E(\eta_{20}, \eta_{2t})$; wherever time indexes t_i, $i = 1, \ldots, 4$ are used, it will be convenient to write also $\gamma_{ij} = \rho_{t_j - t_i}$. Denoting $Z_t = X_t - E(X_t)$, there exists a Gaussian $I(d)$ process V_t such that the bounds in Lemma 6 hold. Lemmas 7 and 10 in Robinson (1994b) and Lemma 7 imply that

$$\operatorname{Var}\left\{\widehat{F}_{VV}(\lambda_m)\right\} = o\left(\left(\frac{m}{n}\right)^{2-4d}\right),$$

so we need to show that the approximation error satisfies

$$A = \operatorname{Var}\left\{\widehat{F}_{XX}(\lambda_m)\right\} - \operatorname{Var}\left\{\widehat{F}_{VV}(\lambda_m)\right\} = o\left(\left(\frac{m}{n}\right)^{2-4d}\right). \tag{B.4}$$

Since $n^2[\widehat{F}_{XX}(\lambda_m) - E\{\widehat{F}_{XX}(\lambda_m)\}]$ can be written

$$\sum_{t_1,t_2=1}^{n} \{X_{t_1}X_{t_2} - E(X_{t_1}X_{t_2})\} D_m(\lambda_{t_2-t_1}) = \sum_{t_1,t_2=1}^{n} \{Z_{t_1}Z_{t_2} - E(Z_{t_1}Z_{t_2})\} D_m(\lambda_{t_2-t_1}),$$

by (B.1), we have

$$\operatorname{Var}\left\{\widehat{F}_{XX}(\lambda_m)\right\} = \frac{1}{n^4} \sum_{t_1,t_2,t_3,t_4=1}^{n} \operatorname{Cov}(Z_{t_1}Z_{t_2}, Z_{t_3}Z_{t_4}) D_m(\lambda_{t_2-t_1}) \overline{D_m(\lambda_{t_4-t_3})},$$

and therefore

$$A = \frac{1}{n^4} \sum_{t_1,t_2,t_3,t_4=1}^{n} \{\operatorname{Cov}(Z_{t_1}Z_{t_2}, Z_{t_3}Z_{t_4}) - \operatorname{Cov}(V_{t_1}V_{t_2}, V_{t_3}V_{t_4})\} D_m(\lambda_{t_2-t_1}) \overline{D_m(\lambda_{t_4-t_3})}.$$

We now decompose A into sums where the time indexes conform to cases (a) to (g) in Lemma 6. Using Lemmas 3 and 6 repeatedly, the approximation error for each case is bounded by:

(a)

$$\sum_{\alpha_1,\alpha_2\in Q_4:\alpha_1\neq\alpha_2} \frac{K}{n^4} \sum_{t_1,t_2,t_3,t_4=1}^{n} \gamma_{\alpha_1}^2 |\gamma_{\alpha_2}||D_m(\lambda_{t_2-t_1})||D_m(\lambda_{t_4-t_3})|.$$

If α_1 is either $(1,2)$ or $(3,4)$, each element in the first summation is bounded by

$$\begin{aligned}&\frac{K}{n^2}\sum_{j=1}^{n}\rho_j^2|D_m(\lambda_j)|\sum_{j=1}^{n}|\rho_j||D_m(\lambda_j)| + \frac{K}{n^3}\sum_{j=1}^{n}\rho_j^2|D_m(\lambda_j)|\sum_{j=1}^{n}|\rho_j|\sum_{j=1}^{n}|D_m(\lambda_j)|\\ &\leq K\left(\frac{m}{n}\right)^{2-2d}\left(1+\frac{\log m}{m^{1-2d}}\right)\left\{1+(\log n)1\left(d=\frac{1}{4}\right)+\left(\frac{n}{m}\right)^{4d-1}1\left(d>\frac{1}{4}\right)\right\},\end{aligned}$$

while if α_1 is not equal to $(1,2)$ or to $(3,4)$, we have a bound

$$\begin{aligned}&\frac{K}{n^3}\sum_{j=1}^{n}\rho_j^2\sum_{j=1}^{n}|\rho_j||D_m(\lambda_j)|\sum_{j=1}^{n}|D_m(\lambda_j)| + \frac{K}{n^4}\sum_{j=1}^{n}\rho_j^2\sum_{j=1}^{n}|\rho_j|\left\{\sum_{j=1}^{n}|D_m(\lambda_j)|\right\}^2\\ &\leq K\left(\frac{m}{n}\right)^{2-2d}\frac{\log m}{m}\left(1+\frac{\log m}{m^{1-2d}}\right)\left\{1+(\log n)1\left(d=\frac{1}{4}\right)+n^{4d-1}1\left(d>\frac{1}{4}\right)\right\}.\end{aligned}$$

(b)

$$\begin{aligned}&\frac{K}{n^4}\sum_{t_1,t_2,t_3=1}^{n}(\gamma_{12}^2+\gamma_{13}^2+\gamma_{23}^2)|D_m(\lambda_{t_2-t_1})||D_m(\lambda_{t_3-t_1})|\\ \leq&\frac{K}{n^3}\sum_{j=1}^{n}\rho_j^2|D_m(\lambda_j)|\sum_{j=1}^{n}|D_m(\lambda_j)| + \frac{K}{n^4}\sum_{j=1}^{n}\rho_j^2\left\{\sum_{j=1}^{n}|D_m(\lambda_j)|\right\}^2\\ \leq&K\left(\frac{m}{n}\right)^2\frac{\log m}{m}\left[\left(1+\frac{\log m}{m}\right)\left\{1+(\log n)1\left(d=\frac{1}{4}\right)\right\}\right.\\ &\left.+\left\{\left(\frac{n}{m}\right)^{4d-1}+n^{4d-1}\frac{\log m}{m}\right\}1\left(d>\frac{1}{4}\right)\right].\end{aligned}$$

(c)

$$\begin{aligned}&\frac{K}{n^4}\sum_{t_1,t_2,t_3=1}^{n}(\gamma_{12}^2+\gamma_{13}^2+\gamma_{23}^2)|D_m(0)||D_m(\lambda_{t_3-t_2})|\\ \leq&K\frac{m}{n^2}\sum_{j=1}^{n}\rho_j^2|D_m(\lambda_j)| + K\frac{m}{n^3}\sum_{j=1}^{n}\rho_j^2\sum_{j=1}^{n}|D_m(\lambda_j)|\end{aligned}$$

$$\leq K\left(\frac{m}{n}\right)^2\left[\left(1+\frac{\log m}{n}\right)\left\{1+(\log n)1\left(d=\frac{1}{4}\right)\right\}\right.$$
$$\left.+\left\{\left(\frac{n}{m}\right)^{4d-1}+n^{4d-1}\frac{\log m}{n}\right\}1\left(d>\frac{1}{4}\right)\right].$$

(d), (e), (f) For any $a = a(t_1, t_2)$ and $b = b(t_1, t_2)$,

$$\frac{K}{n^4}\sum_{t_1,t_2=1}^{n}|\gamma_{12}||D_m(a)||D_m(b)| \leq K\frac{m^2}{n^3}\sum_{j=1}^{n}|\rho_j| \leq K\left(\frac{m}{n}\right)^2 n^{2d-1}.$$

(g)

$$\frac{K}{n^4}\sum_{t_1=1}^{n}|D_m(0)|^2 \leq K\left(\frac{m}{n}\right)^2 n^{-1}.$$

Since cases (a) to (g) satisfy (B.4), the proof is complete. □

Appendix C: Technical lemmas for Appendix B

Lemma 1 *Let $|\rho_j - \rho_{j+1}| \leq K|\rho_{j+1}|/j$ and $|\gamma_j - \gamma_{j+1}| \leq K|\gamma_{j+1}|/j$, for all $j \geq 1$. Then, for any positive integers r, s and j,*

$$|\rho_j^r - \rho_{j+1}^r| \leq K\frac{|\rho_{j+1}^r|}{j}, \tag{C.1}$$

$$|\rho_j^r\gamma_j^s - \rho_{j+1}^r\gamma_{j+1}^s| \leq K\frac{|\rho_{j+1}^r\gamma_{j+1}^s|}{j}. \tag{C.2}$$

Proof. First note that

$$(a-b)^k = \sum_{i=0}^{k}\binom{k}{i}a^i(-b)^{k-i} = \sum_{i=0}^{k}\binom{k}{i}(a^i - b^i)(-b)^{k-i},$$

since

$$\sum_{i=0}^{k}\binom{k}{i}b^i(-b)^{k-i} = (b-b)^k = 0.$$

Hence,

$$\begin{aligned}|a^k - b^k| &= \left|(a-b)^k - \sum_{i=1}^{k-1}\binom{k}{i}(a^i - b^i)(-b)^{k-i}\right| \\ &\le |a-b|^k + \sum_{i=1}^{k-1}\binom{k}{i}|b|^{k-i}|a^i - b^i|.\end{aligned}$$

Proceeding by induction, suppose (C.1) holds for $r = 1, 2, \ldots, k-1$. Then

$$\begin{aligned}|\rho_j^k - \rho_{j+1}^k| &\le |\rho_j - \rho_{j+1}|^k + \sum_{i=1}^{k-1}\binom{k}{i}|\rho_{j+1}^{k-i}||\rho_j^i - \rho_{j+1}^i| \\ &\le K\frac{|\rho_{j+1}^k|}{j^k} + K\sum_{i=1}^{k-1}|\rho_{j+1}^{k-i}|\frac{|\rho_{j+1}^i|}{j} \le K\frac{|\rho_{j+1}^k|}{j},\end{aligned}$$

proving (C.1). To prove (C.2) we use (C.1):

$$\begin{aligned}|\rho_j^r\gamma_j^s - \rho_{j+1}^r\gamma_{j+1}^s| &= |(\rho_j^r - \rho_{j+1}^r)(\gamma_j^s - \gamma_{j+1}^s) + \gamma_{j+1}^s(\rho_j^r - \rho_{j+1}^r) + \rho_{j+1}^r(\gamma_j^s - \gamma_{j+1}^s)| \\ &\le K\frac{|\rho_{j+1}^r|}{j}\frac{|\gamma_{j+1}^s|}{j} + K|\gamma_{j+1}^s|\frac{|\rho_{j+1}^r|}{j} + K|\rho_{j+1}^r|\frac{|\gamma_{j+1}^s|}{j} \le K\frac{|\rho_{j+1}^r\gamma_{j+1}^s|}{j}. \square\end{aligned}$$

Lemma 2 *If (a_1, b_1) is independent of (a_2, b_2) and $E(a_i^2 + b_i^2) < \infty$,*

$$\begin{aligned}\operatorname{Cov}(a_1a_2, b_1b_2) &= \operatorname{Cov}(a_1, b_1)E(a_2)E(b_2) + E(a_1b_1)\operatorname{Cov}(a_2, b_2) \\ &= \operatorname{Cov}(a_1, b_1)E(a_2)E(b_2) + E(a_1)E(b_1)\operatorname{Cov}(a_2, b_2) + \operatorname{Cov}(a_1, b_1)\operatorname{Cov}(a_2, b_2).\end{aligned}$$

Proof. Straightforward. $\square$

Lemma 3 *Let $\rho_j = O(j^{2d-1})$, $a > 0$, $b \ge 1$, $m \le n/2$, and $d^+ = (a-1)/2a$. Then,*

$$\sum_{j=1}^{n}|\rho_j|^a = O\left(1 + (\log n)1(d = d^+) + n^{a(2d-1)+1}1(d > d^+)\right),$$

$$\sum_{j=1}^{n}|D_m(\lambda_j)|^b = O\left(n\left\{\log m + m^{b-1}1(b > 1)\right\}\right),$$

$$\sum_{j=1}^{n} |\rho_j|^a |D_m(\lambda_j)|^b = O\left(m^b \left\{1 + (\log n)1(d = d^+) + \left(\frac{n}{m}\right)^{a(2d-1)+1} 1(d > d^+)\right\}\right).$$

Proof. From, e.g. Zygmund (1977, p. 11) and an elementary inequality,

$$|D_m(\lambda_j)| \leq K \min\left\{m, \frac{n}{|j|}\right\}, \qquad |j| \leq \frac{n}{2}. \tag{C.3}$$

The remainder of the proof is straightforward. □

Lemma 4 *For $j = 1, 2$, define $g_{j,t} = g_j(\mu_t)$, where μ_t is a standard Gaussian $I(d)$ process and $\rho_t = E(\mu_0 \mu_t)$. Assume $E(g_{j,t}^2) < \infty$. Denote by $G_{j,k}$ the k-th Hermite coefficient of $g_j(\cdot)$ and let*

$$r = \min\{k \in \mathbb{N} : G_{1,k} G_{2,k} \neq 0\}. \tag{C.4}$$

If $d > 0$, define

$$d^* = \frac{1}{2} - r\left(\frac{1}{2} - d\right), \qquad C_\rho = \lim_{j \to \infty} \rho_j j^{1-2d}.$$

Let $A = S_m(g_1, g_2)$, where m satisfies Assumption 2 if $d^* = 1/(2r+2)$ or just (4.4) otherwise. Then,

$$\begin{aligned} A =& O\left(\frac{m}{n}\{1(d = 0) + 1(d^* < 0) + (\log n)1(d^* = 0)\}\right) \\ &+ \left\{C^* \left(\frac{m}{n}\right)^{1-2d^*} + o\left(\left(\frac{m}{n}\right)^{1-2d^*}\right)\right\} 1(d^* > 0), \end{aligned} \tag{C.5}$$

where

$$C^* = 2\frac{(2\pi)^{-2d^*}\Gamma(2d^*)}{1-2d^*} \frac{G_{1,r}G_{2,r}}{r!} \sin\left\{(1-2d^*)\frac{\pi}{2}\right\} C_\rho^r \neq 0.$$

Proof. Let $\gamma_t = \mathrm{Cov}(g_{1,0}; g_{2,t})$. Then

$$A = \frac{1}{n^2} \sum_{s,t=1}^{n} \gamma_{t-s} D_m(\lambda_{t-s}) = \frac{1}{n} \sum_{u=1-n}^{n-1} \left(1 - \frac{|u|}{n}\right) \gamma_u D_m(\lambda_u). \tag{C.6}$$

We will make repeated use of (C.3) and of $\rho_u^r = O(u^{2d^*-1})$. By Theorem 1 and (C.4),

$$\gamma_u = \sum_{k=1}^{\infty} \frac{G_{1,k}G_{2,k}}{k!} \rho_u^k = C\rho_u^r + O(|\rho_u^{r+1}|),$$

where $C = G_{1,r}G_{2,r}/r!$.

(a) If $d = 0$, then $\gamma_u = O(|\rho_u^r|)$ are summable. Similarly, if $d^* < 0$, then $\gamma_u = O(|\rho_u^r|) = O(u^{2d^*-1})$ are summable. In either case,

$$A \le \frac{K}{n} \sum_{u=1-n}^{n-1} \left(1 - \frac{|u|}{n}\right) |\gamma_u||D_m(\lambda_u)| \le K\frac{m}{n} \sum_{u=1-n}^{n-1} |\gamma_u| = O\left(\frac{m}{n}\right). \tag{C.7}$$

(b) If $d^* = 0$, $\gamma_u = O(|\rho_u^r|) = O(u^{-1})$, hence

$$A \le K\frac{m}{n} \sum_{u=1-n}^{n-1} |\gamma_u| = O\left(\frac{m}{n} \log n\right). \tag{C.8}$$

(c) If $d^* > 0$,

$$|\gamma_u - C\rho_u^r| \le K|\rho_u^{r+1}| \le K|\rho_u^r|^{1+\omega},$$

where $\omega = r^{-1}$. Defining

$$B_1 = \frac{1}{n} \sum_{u=1-n}^{n-1} \left(1 - \frac{|u|}{n}\right) \rho_u^r D_m(\lambda_u),$$

we get

$$\begin{aligned}|A - CB_1| &\le \frac{1}{n} \sum_{u=1-n}^{n-1} \left(1 - \frac{|u|}{n}\right) |\gamma_u - C\rho_u^r||D_m(\lambda_u)| \\ &\le \frac{K}{n} \sum_{u=1-n}^{n-1} |\rho_u^r|^{1+\omega}|D_m(\lambda_u)| \le K\frac{m}{n} + \frac{K}{n} \sum_{u=1}^{n} |\rho_u^r|^{1+\omega}|D_m(\lambda_u)|.\end{aligned}$$

Therefore, setting $d^+ = \omega/(2 + 2\omega)$ in Lemma 3,

$$\begin{aligned}|A - CB_1| &= O\left(\frac{m}{n}\left\{1 + (\log n)1(d^* = d^+) + \left(\frac{m}{n}\right)^{(1+\omega)(1-2d^*)} 1(d^* > d^+)\right\}\right) \\ &= o\left(\left(\frac{m}{n}\right)^{1-2d^*}\right),\end{aligned}$$

choosing $0 < \epsilon < 2d^*$ in Assumption 2 if $d^* = d^+$. Now, write

$$\begin{aligned} B_1 &= \frac{1}{n}\sum_{|u|<n}\left(1 - \frac{|u|}{n}\right)\rho_u^r D_m(\lambda_u) = \frac{1}{n}\sum_{|u|<n}\rho_u^r D_m(\lambda_u) - \frac{1}{n^2}\sum_{|u|<n}|u|\rho_u^r D_m(\lambda_u) \\ &= \frac{1}{n}\sum_{u=-\infty}^{\infty}\rho_u^r D_m(\lambda_u) - \frac{1}{n}\sum_{|u|\geq n}\rho_u^r D_m(\lambda_u) - \frac{1}{n^2}\sum_{|u|<n}|u|\rho_u^r D_m(\lambda_u) = B_2 + B_3 + B_4, \end{aligned}$$

where

$$B_2 = \frac{2\pi}{n}\sum_{j=1}^{m} f(\lambda_j), \qquad f(\lambda_j) = \frac{1}{2\pi}\sum_{u=-\infty}^{\infty}\rho_u^r e^{iu\lambda_j}, \tag{C.9}$$

$$B_3 = -\frac{1}{n}\sum_{u=n}^{\infty}\rho_u^r\left\{D_m(\lambda_u) + \overline{D_m(\lambda_u)}\right\}, \tag{C.10}$$

$$B_4 = -\frac{1}{n^2}\sum_{u=1}^{n-1}u\rho_u^r\left\{D_m(\lambda_u) + \overline{D_m(\lambda_u)}\right\}. \tag{C.11}$$

Then

$$|A - CB_2| = |A - C(B_1 - B_3 - B_4)| \leq |A - CB_1| + C|B_3| + C|B_4|. \tag{C.12}$$

Note that, for any u,

$$\left|\sum_{k=1}^{u} D_m(\lambda_k)\right| = \left|\sum_{k=1}^{u}\sum_{j=1}^{m} e^{ij\lambda_k}\right| \leq \sum_{j=1}^{m}|D_u(\lambda_j)| \leq K\sum_{j=1}^{m}\frac{n}{j} \leq Kn\log m. \tag{C.13}$$

Using summation by parts, (C.13) and Lemma 1 in (C.11),

$$\begin{aligned} |B_4| &\leq \frac{K}{n^2}\sum_{u=1}^{n-1}\left\{|u\rho_u^r - (u+1)\rho_{u+1}^r|\left|\sum_{k=1}^{u} D_m(\lambda_k)\right|\right\} + \frac{K}{n^2}\left|\sum_{j=1}^{n-1} D_m(\lambda_j)\right| n|\rho_n^r| \\ &\leq \frac{K}{n^2}\sum_{u=1}^{n-1}(u|\rho_u^r - \rho_{u+1}^r| + |\rho_{u+1}^r|)n\log m + Kn^{2d^*-1}\log m \\ &\leq K\frac{\log m}{n}\sum_{u=1}^{n-1}u^{2d^*-1} + Kn^{2d^*-1}\log m \leq Kn^{2d^*-1}\log m = o\left(\left(\frac{m}{n}\right)^{1-2d^*}\right). \end{aligned} \tag{C.14}$$

Using the partial summation formula for infinite sums, (C.13) and Lemma 1 in (C.10),

$$
\begin{aligned}
|B_3| &\leq \frac{K}{n}\sum_{u=n}^{\infty}|\rho_u^r-\rho_{u+1}^r|\left|\sum_{k=1}^{u}D_m(\lambda_k)\right|+\frac{K}{n}\left|\sum_{j=1}^{n-1}D_m(\lambda_j)\right||\rho_n^r| \\
&\leq K\log m\sum_{u=n}^{\infty}u^{2d^*-2}+Kn^{2d^*-1}\log m\leq Kn^{2d^*-1}\log m=o\left(\left(\frac{m}{n}\right)^{1-2d^*}\right). \quad \text{(C.15)}
\end{aligned}
$$

Lemma 1 implies that $f(\lambda)\sim C_f\lambda^{-2d^*}$ as $\lambda\to 0^+$, where

$$
C_f=\pi^{-1}\Gamma(2d^*)\sin\left\{(1-2d^*)\frac{\pi}{2}\right\}C_\rho^r.
$$

Thus, by Proposition 1 in Robinson (1994a),

$$
B_2=\frac{2\pi}{n}\sum_{j=1}^{m}f(\lambda_j)\sim\int_0^{\lambda_m}f(t)\mathrm{d}t\sim C_f\frac{\lambda_m^{1-2d^*}}{1-2d^*}\sim C_f\frac{(2\pi)^{1-2d^*}}{1-2d^*}\left(\frac{m}{n}\right)^{1-2d^*},
$$

which together with (C.12), (C.14), (C.15) gives (C.5). □

Lemma 5 *For $i,j=1,2$, define $g_{ij,t}=g_{ij}(\mu_{it})$, where μ_{it} is a standard Gaussian $I(d_i)$ process and $\rho_{i,t}=E(\mu_{i0}\mu_{it})$. Assume $E(g_{ij,t}^2)<\infty$. Denote by $G_{ij,k}$ the k-th Hermite coefficients of $g_{ij}(\cdot)$, with*

$$
r_i=\min\{k>0:G_{i1,k}G_{i2,k}\neq 0\}. \quad \text{(C.16)}
$$

Let $d_1\geq d_2$ without loss of generality, and define

$$
d^*=\frac{1}{2}-r_1\left(\frac{1}{2}-d_1\right)-r_2\left(\frac{1}{2}-d_2\right), \qquad C_{i\rho}=\lim_{j\to\infty}\rho_{ij}j^{1-2d_i}.
$$

Let $A=S_m'(g_{11},g_{12};g_{21},g_{22})$, where m satisfies Assumption 2 if $d^*+d_1=1/2$ or just (4.4) otherwise. Then,

$$
\begin{aligned}
A=&O\left(\frac{m}{n}\{1+(\log n)1(d^*=0)\}\right) \\
&+\left\{C^*\left(\frac{m}{n}\right)^{1-2d^*}+o\left(\left(\frac{m}{n}\right)^{1-2d^*}\right)\right\}1(d^*>0), \quad \text{(C.17)}
\end{aligned}
$$

where

$$C^* = 2\frac{(2\pi)^{-2d^*}\Gamma(2d^*)}{1-2d^*}\frac{G_{11,r_1}G_{12,r_1}}{r_1!}\frac{G_{21,r_2}G_{22,r_2}}{r_2!}\sin\left\{(1-2d^*)\frac{\pi}{2}\right\}C_{1\rho}^{r_1}C_{2\rho}^{r_2} \neq 0.$$

Proof. Let $\gamma_{i,t} = \mathrm{Cov}(g_{i1,0}; g_{i2,t})$. Then, similarly to (C.6),

$$A = \frac{1}{n}\sum_{u=1-n}^{n-1}\left(1-\frac{|u|}{n}\right)\gamma_{1u}\gamma_{2u}D_m(\lambda_u).$$

By Theorem 1 and (C.16),

$$\gamma_{iu} = \sum_{k=1}^{\infty}\frac{G_{i1,k}G_{i2,k}}{k!}\rho_{iu}^k = C_i\rho_{iu}^{r_i} + O(|\rho_{iu}^{r_i+1}|),$$

where $C_i = G_{i1,r_i}G_{i2,r_i}/r_i!$.

(a) If $d_1d_2 = 0$, then $\gamma_{1u}\gamma_{2u} = O(|\rho_{1u}^{r_1}\rho_{2u}^{r_2}|)$ are summable. Similarly, if $d_1d_2 > 0$ but $d^* < 0$, then $\gamma_{1u}\gamma_{2u} = O(|\rho_{1u}^{r_1}\rho_{2u}^{r_2}|) = O(j^{2d^*-1})$ are summable. In either case, writing $\gamma_{1u}\gamma_{2u}$ instead of γ_u in (C.7) yields $A = O(m/n)$.

(b) If $d^* = 0$, $\gamma_{1u}\gamma_{2u} = O(j^{-1})$, hence (C.8) holds for $\gamma_{1u}\gamma_{2u}$.

(c) If $d^* > 0$,

$$\begin{aligned}|\gamma_{1u}\gamma_{2u} - C_1C_2\rho_{2u}^{r_1}\rho_{2u}^{r_2}| &\leq |\gamma_{1u}||\gamma_{2u} - C_2\rho_{2u}^{r_2}| + C_2|\rho_{2u}^{r_2}||\gamma_{1u} - C_1\rho_{1u}^{r_1}| \\ &\leq K|\rho_{1u}^{r_1}\rho_{2u}^{r_2+1}| + K|\rho_{2u}^{r_2}\rho_{1u}^{r_1+1}| \leq K|\rho_{1u}^{r_1+1}\rho_{1u}^{r_2}| \leq K|\rho_{1u}^{r_1}\rho_{2u}^{r_2}|^{1+\omega},\end{aligned}$$

where $\omega = (1-2d_1)/(1-2d^*)$, since $d_1 \geq d_2$. Then (C.17) follows from the proof of case (c) of Lemma 4, writing $\rho_{1u}^{r_1}\rho_{2u}^{r_2}$ instead of ρ_u^r. $\square$

Lemma 6 *Under (1.2) and Assumption 1, let* $Z_t = X_t - E(X_t)$. *For* t_1, t_2, t_3, t_4 *distinct, define* $\rho_{ij} = E(\eta_{2t_i}\eta_{2t_j})$ *and* $\rho'_{ij} = E(\nu_{2t_i}\nu_{2t_j})$. *Then there exists a mean-zero Gaussian* $I(d)$ *process* V_t *such that:*

(a) $\mathrm{Cov}(Z_{t_1}Z_{t_2}, Z_{t_3}Z_{t_4}) - \mathrm{Cov}(V_{t_1}V_{t_2}, V_{t_3}V_{t_4}) = O(\sum_{\alpha_1,\alpha_2\in Q_4:\alpha_1\neq\alpha_2}\rho_{\alpha_1}^2|\rho_{\alpha_2}|)$;

(b) $\mathrm{Cov}(Z_{t_1}Z_{t_2}, Z_{t_1}Z_{t_3}) - \mathrm{Cov}(V_{t_1}V_{t_2}, V_{t_1}V_{t_3}) = O(\rho_{12}^2 + \rho_{13}^2 + \rho_{23}^2)$;

(c) $\mathrm{Cov}(Z_{t_1}^2, Z_{t_2}Z_{t_3}) - \mathrm{Cov}(V_{t_1}^2, V_{t_2}V_{t_3}) = O(\rho_{12}^2 + \rho_{13}^2 + \rho_{23}^2)$;

(d) $\mathrm{Cov}(Z_{t_1}Z_{t_2}, Z_{t_1}Z_{t_2}) - \mathrm{Cov}(V_{t_1}V_{t_2}, V_{t_1}V_{t_2}) = O(|\rho_{12}|)$;

(e) $\mathrm{Cov}(Z_{t_1}^2, Z_{t_2}^2) - \mathrm{Cov}(V_{t_1}^2, V_{t_2}^2) = O(|\rho_{12}|)$;

(f) $\mathrm{Cov}(Z_{t_1}^2, Z_{t_1}Z_{t_2}) - \mathrm{Cov}(V_{t_1}^2, V_{t_1}V_{t_2}) = O(|\rho_{12}|)$*;*
(g) $\mathrm{Cov}(Z_{t_1}^2, Z_{t_1}^2) - \mathrm{Cov}(V_{t_1}^2, V_{t_1}^2) = O(1)$.

Proof. All Z_t covariances in (a) to (g) can be written as linear combinations of

$$E\left(\prod_{i=1}^{4} Z_{t_i}^{k_i}\right) = E\left\{E\left(\prod_{i=1}^{4} Z_{t_i}^{k_i}\middle| g_{t_j}, h_{t_j},\ j=1,\dots,4\right)\right\}$$
$$= E\left\{\prod_{i=1}^{4} E(Z_{t_i}^{k_i}|g_{t_j}, h_{t_j},\ j=1,\dots,4)\right\} = E\left\{\prod_{i=1}^{4} E(Z_{t_i}^{k_i}|g_{t_i}, h_{t_i})\right\}, \qquad \text{(C.18)}$$

where, conditionally on g_s and h_s, Z_s is independent of Z_t, g_t and h_t, for any $t \neq s$.

In what follows, let s_i, $i = 1, \dots, 4$ denote (not necessarily distinct) elements of $\{t_1, t_2, t_3, t_4\}$. Wherever u_i and s_i are both defined, let

$$A_i = \binom{p}{u_i} \eta_{1s_i}^{u_i} \nu_{1s_i}^{p-u_i}, \qquad B_i = g_{s_i}^{u_i} h_{s_i}^{p-u_i}$$

and define $c_i = E(A_i)$, $c_{ij} = E(\widetilde{A}_i\widetilde{A}_j)$, $c_{ijk} = E(\widetilde{A}_i\widetilde{A}_j\widetilde{A}_k)$, where throughout the proof $\widetilde{z}_t^u = z_t^u - E(z_t^u)$. We first compute $E(Z_t^k|g_t, h_t)$ for $k = 1, \dots, 3$. Setting $s_1 = s_2 = s_3 = t$, but omitting time subscripts for convenience,

$$Z = \sum_{u_1=0}^{p} \{A_1B_1 - E(A_1B_1)\} = \sum_{u_1=0}^{p} (\widetilde{A}_1B_1 + c_1\widetilde{B}_1).$$

Therefore, independence of A_1 and B_1 yields

$$E(Z|g,h) = \sum_{u_1=0}^{p} c_1\widetilde{B}_1. \qquad \text{(C.19)}$$

Similarly,

$$Z^k = \prod_{i=1}^{k} \sum_{u_i=0}^{p} (\widetilde{A}_iB_i + c_i\widetilde{B}_i),$$

so by independence of A_i and B_j, for all i, j

$$E(Z^2|g,h) = \sum_{u_1,u_2=0}^{p} (c_1c_2\widetilde{B}_1\widetilde{B}_2 + c_{12}B_1B_2), \qquad \text{(C.20)}$$

$$E(Z^3|g,h) = \sum_{u_1,u_2,u_3=0}^{p} (c_1c_2c_3\widetilde{B}_1\widetilde{B}_2\widetilde{B}_3 + c_{123}B_1B_2B_3$$
$$+ c_{12}c_3B_1B_2\widetilde{B}_3 + c_{13}c_2B_1B_3\widetilde{B}_2 + c_{23}c_1B_2B_3\widetilde{B}_1). \qquad \text{(C.21)}$$

Unless otherwise noted, we will use $\sum$ to mean $\sum_{u_1,u_2,u_3,u_4=0}^{p}$ for the remainder of the proof. Using (C.19), (C.20), (C.21) in (C.18), we can write $E(\prod_{i=1}^{4} Z_{s_i})$ as follows:
(i) if $s_i = t_i$, $i = 1, \ldots, 4$,

$$E(Z_{t_1}Z_{t_2}Z_{t_3}Z_{t_4}) = \sum c_1c_2c_3c_4E(\widetilde{B}_1\widetilde{B}_2\widetilde{B}_3\widetilde{B}_4); \qquad \text{(C.22)}$$

(ii) if $s_1 = s_2 = t_1$, $s_3 = t_2$, $s_4 = t_3$,

$$E(Z_{t_1}^2 Z_{t_2}Z_{t_3}) = \sum \left\{c_1c_2c_3c_4E(\widetilde{B}_1\widetilde{B}_2\widetilde{B}_3\widetilde{B}_4) + c_{12}c_3c_4E(B_1B_2\widetilde{B}_3\widetilde{B}_4)\right\}; \qquad \text{(C.23)}$$

(iii) if $s_1 = s_2 = t_1$, $s_3 = s_4 = t_2$,

$$E(Z_{t_1}^2 Z_{t_2}^2) = \sum \left\{c_1c_2c_3c_4E(\widetilde{B}_1\widetilde{B}_2\widetilde{B}_3\widetilde{B}_4) + c_{12}c_{34}E(B_1B_2B_3B_4)\right.$$
$$\left. + c_{12}c_3c_4E(B_1B_2\widetilde{B}_3\widetilde{B}_4) + c_1c_2c_{34}E(\widetilde{B}_1\widetilde{B}_2B_3B_4)\right\}; \qquad \text{(C.24)}$$

(iv) if $s_1 = s_2 = s_3 = t_1$, $s_4 = t_2$,

$$E(Z_{t_1}^3 Z_{t_2}) = \sum \left\{c_1c_2c_3c_4E(\widetilde{B}_1\widetilde{B}_2\widetilde{B}_3\widetilde{B}_4) + c_{123}c_4E(B_1B_2B_3\widetilde{B}_4) + c_{12}c_3c_4E(B_1B_2\widetilde{B}_3\widetilde{B}_4)\right.$$
$$\left. + c_{13}c_2c_4E(B_1B_3\widetilde{B}_2\widetilde{B}_4) + c_{23}c_1c_4E(B_2B_3\widetilde{B}_1\widetilde{B}_4)\right\}. \qquad \text{(C.25)}$$

We can also write $E(Z_{s_1}Z_{s_2})$ as follows:
(v) if $s_1 = t_i$, $s_2 = t_j$, $i \neq j$,

$$E(Z_{t_i}Z_{t_j}) = \sum_{u_1,u_2=0}^{p} c_1c_2E(\widetilde{B}_1\widetilde{B}_2); \qquad \text{(C.26)}$$

(vi) if $s_1 = s_2 = t_i$,

$$E(Z_{t_i}^2) = \sum_{u_1,u_2=0}^{p} \left\{c_1c_2E(\widetilde{B}_1\widetilde{B}_2) + c_{12}E(B_1B_2)\right\}. \qquad \text{(C.27)}$$

We now proceed to expand B_i and $\widetilde{B}_i$ in terms of g and h. Wherever u_i and s_i are both defined, we use the following notation: $\chi_i = \widetilde{g}_{s_i}^{u_i}$, $\psi_i = \widetilde{h}_{s_i}^{p-u_i}$; $\chi_{ij} = g_{s_i}^{u_i+u_j}$, $\psi_{ij} = h_{s_i}^{2p-u_i-u_j}$; $\chi_{123} = g_{s_1}^{u_1+u_2+u_3}$, $\psi_{123} = h_{s_1}^{3p-u_1-u_2-u_3}$; $\chi_{12,34} = g_{s_1}^{u_1+u_2} g_{s_3}^{u_3+u_4}$, $\psi_{12,34} = h_{s_1}^{2p-u_1-u_2} h_{s_3}^{2p-u_3-u_4}$. Note that

$$\widetilde{B}_i = c_{h,i}\chi_i + \chi_i\psi_i + c_{g,i}\psi_i, \tag{C.28}$$

where $c_{h,i} = E(h_{s_i}^{p-u_i})$, $c_{g,i} = E(g_{s_i}^{u_i})$. Four forms of expectations need to be accounted for in (C.22) to (C.25).

1. $E(\widetilde{B}_1\widetilde{B}_2\widetilde{B}_3\widetilde{B}_4)$ will be a linear combination of 81 terms, all of them expectations of products of χ_i and ψ_i, $i = 1,\ldots,4$. Denoting by $\langle 1\rangle$, $\langle 2\rangle$, $\langle 3\rangle$, $\langle 4\rangle$ any permutation of P_4, those terms can be separated into the following categories: terms that vanish due to $E(\chi_i) = E(\psi_i) = 0$, namely $E(\chi_{\langle 1\rangle}\psi_{\langle 2\rangle}\psi_{\langle 3\rangle}\psi_{\langle 4\rangle})$, $E(\psi_{\langle 1\rangle}\chi_{\langle 2\rangle}\chi_{\langle 3\rangle}\chi_{\langle 4\rangle})$, $E(\chi_{\langle 1\rangle}\psi_1\psi_2\psi_3\psi_4)$ and $E(\psi_{\langle 1\rangle}\chi_1\chi_2\chi_3\chi_4)$; non-vanishing terms with four factors, namely $E(\chi_1\chi_2\chi_3\chi_4)$, $E(\psi_1\psi_2\psi_3\psi_4)$ and $E(\chi_{\langle 1\rangle}\chi_{\langle 2\rangle}\psi_{\langle 3\rangle}\psi_{\langle 4\rangle})$; non-vanishing terms with five factors, namely $E(\chi_{\langle 1\rangle}\chi_{\langle 2\rangle}\psi_{\langle 1\rangle}\psi_{\langle 3\rangle}\psi_{\langle 4\rangle})$ and $E(\psi_{\langle 1\rangle}\psi_{\langle 2\rangle}\chi_{\langle 1\rangle}\chi_{\langle 3\rangle}\chi_{\langle 4\rangle})$; terms with six factors, namely $E(\chi_{\langle 1\rangle}\chi_{\langle 2\rangle}\psi_1\psi_2\psi_3\psi_4)$, $E(\psi_{\langle 1\rangle}\psi_{\langle 2\rangle}\chi_1\chi_2\chi_3\chi_4)$ and $E(\chi_{\langle 1\rangle}\chi_{\langle 2\rangle}\chi_{\langle 3\rangle}\psi_{\langle 1\rangle}\psi_{\langle 2\rangle}\psi_{\langle 4\rangle})$; terms with seven factors, namely $E(\chi_{\langle 1\rangle}\chi_{\langle 2\rangle}\chi_{\langle 3\rangle}\psi_1\psi_2\psi_3\psi_4)$ and $E(\psi_{\langle 1\rangle}\psi_{\langle 2\rangle}\psi_{\langle 3\rangle}\chi_1\chi_2\chi_3\chi_4)$; a term with eight factors, namely $E(\chi_1\chi_2\chi_3\chi_4\psi_1\psi_2\psi_3\psi_4)$. It can be seen from (C.28) that, for each $i = 1,\ldots,4$, the corresponding coefficient will include a factor $c_{h,i}$ if only χ_i is present or $c_{g,i}$ if only ψ_i is present.

2. $E(B_1B_2\widetilde{B}_3\widetilde{B}_4)$ will be a linear combination of 9 terms. Denoting by $\langle 3\rangle$, $\langle 4\rangle$ any permutation of $\{3,4\}$, these can be grouped in the following categories: $E(\chi_{12}\psi_{12}\psi_3\psi_4)$, $E(\psi_{12}\chi_{12}\chi_3\chi_4)$, $E(\chi_{12}\chi_{\langle 3\rangle}\psi_{12}\psi_{\langle 4\rangle})$, $E(\chi_{12}\chi_{\langle 3\rangle}\psi_{12}\psi_3\psi_4)$, $E(\psi_{12}\psi_{\langle 3\rangle}\chi_{12}\chi_3\chi_4)$ and $E(\chi_{12}\chi_3\chi_4\psi_{12}\psi_3\psi_4)$. As in the previous case, (C.28) implies that, for each $i = 3,4$, the corresponding coefficient will include a factor $c_{h,i}$ if only χ_i is present or $c_{g,i}$ if only ψ_i is present.

3. $E(B_1B_2B_3\widetilde{B}_4) = c_{h,4}E(\chi_{123}\chi_4\psi_{123}) + c_{g,4}E(\chi_{123}\psi_{123}\psi_4) + E(\chi_{123}\chi_4\psi_{123}\psi_4)$.

4. $E(B_1B_2B_3B_4) = E(\chi_{12,34}\psi_{12,34})$.

In (C.26) and (C.27), the relevant expectations are $E(\widetilde{B}_1\widetilde{B}_2) = c_{h,1}c_{h,2}E(\chi_1\chi_2) + c_{g,1}c_{g,2}E(\psi_1\psi_2) + E(\chi_1\chi_2\psi_1\psi_2)$ and $E(B_1B_2) = E(\chi_{12}\psi_{12})$.

We can now use Theorem 1 to expand each these expectations as $\sum_{q=0}^{\infty} a_q$. Let μ_t represent either η_{2t} or ν_{2t}, with $\gamma_{ij} = E(\mu_{t_i}\mu_{t_j})$, and define $f_{i,t} = f_i(\mu_t)$, $f_{ij,t}^* = f_{ij}^*(\mu_t)$ such that $E(f_{i,t}) = 0$. Denote by $G_{i;q}$, $G_{ij;q}$, $G_{ijk;q}$, $G_{ij;q}^*$ the q-th Hermite coefficient of

$f_{i,t}$, $f_{i,t}f_{j,t}$, $f_{i,t}f_{j,t}f_{k,t}$, $f^*_{ij,t}$ respectively.

For $E(f_{1,t_1}f_{2,t_2}f_{3,t_3}f_{4,t_4})$, we have

$$a_q = \sum_{\substack{v_\alpha \geq 0: \\ \Sigma v_\alpha = q, \alpha \in Q_4}} \prod_{i=1}^4 G_{i;w_i} \prod_{\alpha \in Q_4} \frac{\gamma_\alpha^{v_\alpha}}{v_\alpha!}, \qquad w_k = \sum_{\alpha \in R_{4,k}} v_\alpha,$$

$$a_0 = a_1 = 0, \qquad a_2 = G_{1;1}G_{2;1}G_{3;1}G_{4;1}(\gamma_{12}\gamma_{34} + \gamma_{13}\gamma_{24} + \gamma_{14}\gamma_{23}).$$

Since $G_{i;0} = 0$, Theorem 1 yields

$$\sum_{q=3}^\infty |a_q| \leq K \prod_{i=1}^4 \left(\sum_{\alpha \in R_{4,i}} |\gamma_\alpha| \right)^{\frac{1}{2}} \sum_{\alpha \in Q_4} |\gamma_\alpha|.$$

Label the elements of P_4 as $\langle 1 \rangle$, $\langle 2 \rangle$, $\langle 3 \rangle$, $\langle 4 \rangle$, such that $|\gamma_{\langle 1 \rangle \langle 2 \rangle}|$ is the largest absolute correlation. Then $\sum_{\alpha \in Q_4} |\gamma_\alpha| \leq K|\gamma_{\langle 1 \rangle \langle 2 \rangle}|$ and

$$\prod_{i=1}^4 \sum_{\alpha \in R_{4,i}} |\gamma_\alpha| \leq K\gamma^2_{\langle 1 \rangle \langle 2 \rangle}(|\gamma_{\langle 1 \rangle \langle 3 \rangle}| + |\gamma_{\langle 2 \rangle \langle 3 \rangle}| + |\gamma_{\langle 3 \rangle \langle 4 \rangle}|)(|\gamma_{\langle 1 \rangle \langle 4 \rangle}| + |\gamma_{\langle 2 \rangle \langle 4 \rangle}| + |\gamma_{\langle 3 \rangle \langle 4 \rangle}|).$$

Choosing the second largest absolute correlation, we have a bound of the form

$$\sum_{q=4}^\infty |a_q| \leq K\gamma^2_{\langle 1 \rangle \langle 2 \rangle}|\gamma_{\langle 3 \rangle \langle 4 \rangle}| \qquad \text{or} \qquad \sum_{q=4}^\infty |a_q| \leq K\gamma^2_{\langle 1 \rangle \langle 2 \rangle}|\gamma_{\langle 1 \rangle \langle 3 \rangle}|.$$

Therefore, taking all possible permutations for $\langle 1 \rangle$, $\langle 2 \rangle$, $\langle 3 \rangle$, $\langle 4 \rangle$, $\sum_{q=3}^\infty |a_q| = O(e_3)$, where $e_3 = \sum_{\alpha_1, \alpha_2 \in Q_4 : \alpha_1 \neq \alpha_2} \gamma^2_{\alpha_1}|\gamma_{\alpha_2}|$, yielding

$$E(f_{1,t_1}f_{2,t_2}f_{3,t_3}f_{4,t_4}) = \prod_{i=1}^4 G_{i;1}(\gamma_{12}\gamma_{34} + \gamma_{13}\gamma_{24} + \gamma_{14}\gamma_{23}) + O(e_3). \tag{C.29}$$

Again from Theorem 1 but using Corollary 1, defining $e_2 = \gamma_{12}^2 + \gamma_{13}^2 + \gamma_{23}^2$,

$$E(f_{1,t_1}f_{2,t_1}f_{3,t_2}f_{4,t_3}) = G_{12;0}G_{3;1}G_{4;1}\gamma_{23} + O(e_2), \tag{C.30}$$

$$E(f_{1,t_1}f_{2,t_1}f_{3,t_2}f_{4,t_2}) = G_{12;0}G_{34;0} + O(|\gamma_{12}|), \tag{C.31}$$

$$E(f_{1,t_1}f_{2,t_1}f_{3,t_1}f_{4,t_2}) = O(|\gamma_{12}|), \tag{C.32}$$

$$
\begin{aligned}
E(f_{1,t_1} f_{2,t_2} f_{3,t_3}) &= O(e_2), && \text{(C.33)}\\
E(f_{1,t_1} f_{2,t_1} f_{3,t_2}) &= G_{12;1} G_{3;1} \gamma_{12} + O(\gamma_{12}^2), && \text{(C.34)}\\
E(f_{1,t_1} f_{2,t_1} f_{3,t_1}) &= G_{123;0}, && \text{(C.35)}\\
E(f_{1,t_1} f_{2,t_2}) &= G_{1;1} G_{2;1} \gamma_{12} + O(\gamma_{12}^2), && \text{(C.36)}\\
E(f_{1,t_1} f_{2,t_1}) &= G_{12;0}, && \text{(C.37)}\\
E(f^*_{12,t_1} f_{3,t_2} f_{4,t_3}) &= G^*_{12;0} G_{3;1} G_{4;1} \gamma_{23} + O(e_2), && \text{(C.38)}\\
E(f^*_{12,t_1} f_{3,t_2} f_{4,t_2}) &= G^*_{12;0} G_{34;0} + O(|\gamma_{12}|), && \text{(C.39)}\\
E(f^*_{12,t_1} f_{3,t_2}) &= G^*_{12;1} G_{3;1} \gamma_{12} + O(\gamma_{12}^2), && \text{(C.40)}\\
E(f^*_{12,t_1}) &= G^*_{12;0}, && \text{(C.41)}\\
E({}^*_{12,t_1} f^*_{34,t_2}) &= G^*_{12;0} G^*_{34;0} + O(|\gamma_{12}|). && \text{(C.42)}
\end{aligned}
$$

Now let the G and G^* coefficients in (C.29) to (C.42) apply to the case

$$
f_{i,t} = \widetilde{g}_t^{u_i}, \qquad f^*_{ij,t} = g_t^{u_i+u_j}, \tag{C.43}
$$

while corresponding G' and $G^{*\prime}$ coefficients apply to

$$
f_{i,t} = \widetilde{h}_t^{p-u_i}, \qquad f^*_{ij,t} = h_t^{2p-u_i-u_j}. \tag{C.44}
$$

We can approximate each term in the expansion of (C.22) to (C.27) using (C.29) to (C.42):

(i) $E(Z_{t_1} Z_{t_2} Z_{t_3} Z_{t_4})$. Denote by $\langle 1\rangle$, $\langle 2\rangle$, $\langle 3\rangle$, $\langle 4\rangle$ any permutation of P_4. Using (C.29), (C.33), (C.36), the only terms that are not $O(e_3)$ are:

$$
\begin{aligned}
E(\chi_1\chi_2\chi_3\chi_4) &= G_{1;1} G_{2;1} G_{3;1} G_{4;1} (\rho_{12}\rho_{34} + \rho_{13}\rho_{24} + \rho_{14}\rho_{23}) + O(e_3),\\
E(\psi_1\psi_2\psi_3\psi_4) &= G'_{1;1} G'_{2;1} G'_{3;1} G'_{4;1} (\rho'_{12}\rho'_{34} + \rho'_{13}\rho'_{24} + \rho'_{14}\rho'_{23}) + O(e_3),\\
E(\chi_{\langle 1\rangle}\chi_{\langle 2\rangle}) E(\psi_{\langle 3\rangle}\psi_{\langle 4\rangle}) &= G_{\langle 1\rangle;1} G_{\langle 2\rangle;1} G'_{\langle 3\rangle;1} G'_{\langle 4\rangle;1} \rho_{\langle 1\rangle\langle 2\rangle} \rho'_{\langle 3\rangle\langle 4\rangle} + O(e_3).
\end{aligned}
$$

(ii) $E(Z^2_{t_1} Z_{t_2} Z_{t_3})$. Using (C.30), (C.33), (C.34), (C.36), (C.37), the only terms in $E(\widetilde{B}_1\widetilde{B}_2\widetilde{B}_3\widetilde{B}_4)$ that are not $O(e_2)$ are:

$$
\begin{aligned}
E(\chi_1\chi_2\chi_3\chi_4) &= G_{12;0} G_{3;1} G_{4;1} \rho_{23} + O(e_2),\\
E(\psi_1\psi_2\psi_3\psi_4) &= G'_{12;0} G'_{3;1} G'_{4;1} \rho'_{23} + O(e_2),
\end{aligned}
$$

$$
\begin{aligned}
E(\chi_1\chi_2)E(\psi_3\psi_4) &= G_{12;0}G'_{3;1}G'_{4;1}\rho'_{23} + O(e_2),\\
E(\psi_1\psi_2)E(\chi_3\chi_4) &= G'_{12;0}G_{3;1}G_{4;1}\rho_{23} + O(e_2),\\
E(\chi_1\chi_2)E(\psi_1\psi_2\psi_3\psi_4) &= G_{12;0}G'_{12;0}G'_{3;1}G'_{4;1}\rho'_{23} + O(e_2),\\
E(\psi_1\psi_2)E(\chi_1\chi_2\chi_3\chi_4) &= G'_{12;0}G_{12;0}G_{3;1}G_{4;1}\rho_{23} + O(e_2).
\end{aligned}
$$

Using (C.38), (C.40), (C.41) the only terms in $E(B_1B_2\widetilde{B}_3\widetilde{B}_4)$ that are not $O(e_2)$ are:

$$
\begin{aligned}
E(\chi_{12})E(\psi_{12}\psi_3\psi_4) &= G^*_{12,0}G^{*\prime}_{12;0}G'_{3;1}G'_{4;1}\rho'_{23} + O(e_2),\\
E(\psi_{12})E(\chi_{12}\chi_3\chi_4) &= G^{*\prime}_{12,0}G^*_{12;0}G_{3;1}G_{4;1}\rho_{23} + O(e_2).
\end{aligned}
$$

(iii) $E(Z^2_{t_1}Z^2_{t_2})$. From (C.31), (C.34), (C.36), (C.37) it follows that all terms in $E(\widetilde{B}_1\widetilde{B}_2\widetilde{B}_3\widetilde{B}_4)$ will be $O(|\rho_{12}|)$ except the ones only involving (C.31) and (C.37):

$$
\begin{aligned}
E(\chi_1\chi_2\chi_3\chi_4) &= G_{12;0}G_{34;0} + O(|\rho_{12}|),\\
E(\psi_1\psi_2\psi_3\psi_4) &= G'_{12;0}G'_{34;0} + O(|\rho_{12}|),\\
E(\chi_1\chi_2)E(\psi_3\psi_4) &= G_{12;0}G'_{34;0},\\
E(\psi_1\psi_2)E(\chi_3\chi_4) &= G'_{12;0}G_{34;0},\\
E(\chi_1\chi_2)E(\psi_1\psi_2\psi_3\psi_4) &= G_{12;0}G'_{12;0}G'_{34;0} + O(|\rho_{12}|),\\
E(\chi_3\chi_4)E(\psi_1\psi_2\psi_3\psi_4) &= G_{34;0}G'_{12;0}G'_{34;0} + O(|\rho_{12}|),\\
E(\psi_1\psi_2)E(\chi_1\chi_2\chi_3\chi_4) &= G'_{12;0}G_{12;0}G_{34;0} + O(|\rho_{12}|),\\
E(\psi_3\psi_4)E(\chi_1\chi_2\chi_3\chi_4) &= G'_{34;0}G_{12;0}G_{34;0} + O(|\rho_{12}|),\\
E(\chi_1\chi_2\chi_3\chi_4)E(\psi_1\psi_2\psi_3\psi_4) &= G_{12;0}G_{34;0}G'_{12;0}G'_{34;0} + O(|\rho_{12}|).
\end{aligned}
$$

Similarly, from (C.39), (C.40), (C.41), (C.42), all terms in $E(B_1B_2\widetilde{B}_3\widetilde{B}_4)$, $E(\widetilde{B}_1\widetilde{B}_2B_3B_4)$ and $E(B_1B_2B_3B_4)$ will be $O(|\rho_{12}|)$ except the following:

$$
\begin{aligned}
E(\chi_{12})E(\psi_{12}\psi_3\psi_4) &= G^*_{12;0}G^{*\prime}_{12;0}G'_{34;0} + O(|\rho_{12}|),\\
E(\chi_{34})E(\psi_{34}\psi_1\psi_2) &= G^*_{12;0}G^{*\prime}_{12;0}G'_{34;0} + O(|\rho_{12}|),\\
E(\psi_{12})E(\chi_{12}\chi_3\chi_4) &= G^{*\prime}_{12;0}G^*_{12;0}G_{34;0} + O(|\rho_{12}|),\\
E(\psi_{34})E(\chi_{34}\chi_1\chi_2) &= G^{*\prime}_{12;0}G^*_{12;0}G_{34;0} + O(|\rho_{12}|),\\
E(\chi_{12}\chi_3\chi_4)E(\psi_{12}\psi_3\psi_4) &= G^*_{12;0}G_{34;0}G^{*\prime}_{12;0}G'_{34;0} + O(|\rho_{12}|),\\
E(\chi_{34}\chi_1\chi_2)E(\psi_{34}\psi_1\psi_2) &= G^*_{12;0}G_{34;0}G^{*\prime}_{12;0}G'_{34;0} + O(|\rho_{12}|),
\end{aligned}
$$

$$E(\chi_{12,34})E(\psi_{12,34}) = G^*_{12;0}G^*_{34;0}G^{*\prime}_{12;0}G^{*\prime}_{34;0} + O(|\rho_{12}|).$$

(iv) $E(Z^3_{t_1}Z_{t_2})$. Using (C.32), (C.34), (C.35), (C.36), (C.37) in $E(\widetilde{B}_1\widetilde{B}_2\widetilde{B}_3\widetilde{B}_4)$ note that at least one factor in each term necessarily involves t_2. Therefore, one of (C.32), (C.34), (C.36) will apply, making all terms $O(|\rho_{12}|)$.

Similarly, in $E(B_1B_2\widetilde{B}_3\widetilde{B}_4)$, $E(B_1\widetilde{B}_2B_3\widetilde{B}_4)$, $E(\widetilde{B}_1B_2B_3\widetilde{B}_4)$ and $E(B_1B_2B_3\widetilde{B}_4)$, at least one factor in each term necessarily involves t_2. Thus, (C.40) will apply for some function $f^*_{ij,t}$, not necessarily one given in (C.43) or (C.44), making all terms $O(|\rho_{12}|)$.

(v) $E(Z_{t_i}Z_{t_j})$. Using (C.36), the following are not $O(\rho^2_{ij})$:

$$E(\chi_1\chi_2) = G_{1;1}G_{2;1}\rho_{ij} + O(\rho^2_{ij}), \qquad E(\psi_1\psi_2) = G'_{1;1}G'_{2;1}\rho'_{ij} + O(\rho^2_{ij}).$$

(vi) $E(Z^2_{t_i})$. Using (C.37), $E(\widetilde{B}_1\widetilde{B}_2)$ and $E(B_1B_2)$ include the following terms:

$$\begin{aligned} E(\chi_1\chi_2) &= G_{12;0}, & E(\chi_1\chi_2)E(\psi_1\psi_2) &= G_{12;0}G'_{12;0},\\ E(\psi_1\psi_2) &= G'_{12;0}, & E(\chi_{12})E(\psi_{12}) &= G^*_{12;0}G^{*\prime}_{12;0}. \end{aligned}$$

We now compute the coefficients of the leading terms listed above. Define

$$\begin{aligned} L_i &= c_ic_{h,i}G_{i;1}, & \overline{L_1} &= \sum_{u_i=0}^{p} L_i, & L'_i &= c_ic_{g,i}G'_{i;1}, & \overline{L'_1} &= \sum_{u_i=0}^{p} L'_i;\\ L_{ij} &= c_ic_{h,i}c_jc_{h,j}G_{ij;0}, & \overline{L_2} &= \sum_{u_i,u_j=0}^{p} L_{ij}, & L'_{ij} &= c_ic_{g,i}c_jc_{g,j}G'_{ij;0}, & \overline{L'_2} &= \sum_{u_i,u_j=0}^{p} L'_{ij};\\ L^*_{ij} &= c_ic_jG_{ij;0}G'_{ij;0}, & \overline{L^*_2} &= \sum_{u_i,u_j=0}^{p} L^*_{ij}, & L^{**}_{ij} &= c_{ij}G^*_{ij;0}G^{*\prime}_{ij;0}, & \overline{L^{**}_2} &= \sum_{u_i,u_j=0}^{p} L^{**}_{ij}. \end{aligned}$$

Note that $L_p = c_pc_{h,p}G_{p;1} = E(\eta^p_{1t})E\{g^p(\eta_{2t})H_1(\eta_{2t})\} \neq 0$ by assumption, but $L_i = 0$ for any $i < p$. Hence $\overline{L_1} = L_p \neq 0$. The contributions of the non-negligible terms will be:

(i) $E(Z_{t_1}Z_{t_2}Z_{t_3}Z_{t_4})$

$$\begin{aligned} E(\chi_1\chi_2\chi_3\chi_4) &: L_1L_2L_3L_4(\rho_{12}\rho_{34} + \rho_{13}\rho_{24} + \rho_{14}\rho_{23}) + O(e_3);\\ E(\psi_1\psi_2\psi_3\psi_4) &: L'_1L'_2L'_3L'_4(\rho'_{12}\rho'_{34} + \rho'_{13}\rho'_{24} + \rho'_{14}\rho'_{23}) + O(e_3);\\ E(\chi_{\langle 1\rangle}\chi_{\langle 2\rangle})E(\psi_{\langle 3\rangle}\psi_{\langle 4\rangle}) &: L_{\langle 1\rangle}L_{\langle 2\rangle}L'_{\langle 3\rangle}L'_{\langle 4\rangle}\rho_{\langle 1\rangle\langle 2\rangle}\rho'_{\langle 3\rangle\langle 4\rangle} + O(e_3). \end{aligned}$$

Thus, $E(Z_{t_1}Z_{t_2}Z_{t_3}Z_{t_4})$ is

$$
\begin{aligned}
&\sum\{L_1L_2L_3L_4(\rho_{12}\rho_{34}+\rho_{13}\rho_{24}+\rho_{14}\rho_{23})+L_1'L_2'L_3'L_4'(\rho_{12}'\rho_{34}'+\rho_{13}'\rho_{24}'+\rho_{14}'\rho_{23}')\\
&+L_1L_2L_3'L_4'\rho_{12}\rho_{34}'+L_1'L_2'L_3L_4\rho_{12}'\rho_{34}+L_1L_2'L_3L_4'\rho_{13}\rho_{24}'+L_1'L_2L_3'L_4\rho_{13}'\rho_{24}\\
&+L_1L_2'L_3'L_4\rho_{14}\rho_{23}'+L_1'L_2L_3L_4'\rho_{14}'\rho_{23}\}+O(e_3)\\
=&\sum\{(L_1L_2\rho_{12}+L_1'L_2'\rho_{12}')(L_3L_4\rho_{34}+L_3'L_4'\rho_{34}')\\
&+(L_1L_3\rho_{13}+L_1'L_3'\rho_{13}')(L_2L_4\rho_{24}+L_2'L_4'\rho_{24}')\\
&+(L_1L_4\rho_{14}+L_1'L_4'\rho_{14}')(L_2L_3\rho_{23}+L_2'L_3'\rho_{23}')\}+O(e_3)\\
=&\left(\overline{L_1}^2\rho_{12}+\overline{L_1'}^2\rho_{12}'\right)\left(\overline{L_1}^2\rho_{34}+\overline{L_1'}^2\rho_{34}'\right)+\left(\overline{L_1}^2\rho_{13}+\overline{L_1'}^2\rho_{13}'\right)\left(\overline{L_1}^2\rho_{24}+\overline{L_1'}^2\rho_{24}'\right)\\
&+\left(\overline{L_1}^2\rho_{14}+\overline{L_1'}^2\rho_{14}'\right)\left(\overline{L_1}^2\rho_{23}+\overline{L_1'}^2\rho_{23}'\right)+O(e_3).
\end{aligned}
\tag{C.45}
$$

(ii) $E(Z_{t_1}^2Z_{t_2}Z_{t_3})$

$$
\begin{aligned}
E(\chi_1\chi_2\chi_3\chi_4)&:L_{12}L_3L_4\rho_{23}+O(e_2); & E(\chi_1\chi_2)E(\psi_1\psi_2\psi_3\psi_4)&:L_{12}^*L_3'L_4'\rho_{23}'+O(e_2);\\
E(\psi_1\psi_2\psi_3\psi_4)&:L_{12}'L_3'L_4'\rho_{23}'+O(e_2); & E(\psi_1\psi_2)E(\chi_1\chi_2\chi_3\chi_4)&:L_{12}^*L_3L_4\rho_{23}+O(e_2);\\
E(\chi_1\chi_2)E(\psi_3\psi_4)&:L_{12}L_3'L_4'\rho_{23}'+O(e_2); & E(\chi_{12})E(\psi_{12}\psi_3\psi_4)&:L_{12}^{**}L_3'L_4'\rho_{23}'+O(e_2);\\
E(\psi_1\psi_2)E(\chi_3\chi_4)&:L_{12}'L_3L_4\rho_{23}+O(e_2); & E(\psi_{12})E(\chi_{12}\chi_3\chi_4)&:L_{12}^{**}L_3L_4\rho_{23}+O(e_2).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
E(Z_{t_1}^2Z_{t_2}Z_{t_3})=&\sum(L_{12}L_3L_4\rho_{23}+L_{12}'L_3'L_4'\rho_{23}'+L_{12}L_3'L_4'\rho_{23}'+L_{12}'L_3L_4\rho_{23}\\
&+L_{12}^*L_3'L_4'\rho_{23}'+L_{12}^*L_3L_4\rho_{23}+L_{12}^{**}L_3'L_4'\rho_{23}'+L_{12}^{**}L_3L_4\rho_{23})+O(e_2)\\
=&\sum(L_{12}+L_{12}'+L_{12}^*+L_{12}^{**})(L_3L_4\rho_{23}+L_3'L_4'\rho_{23}')+O(e_2)\\
=&\left(\overline{L_2}+\overline{L_2'}+\overline{L_2^*}+\overline{L_2^{**}}\right)\left(\overline{L_1}^2\rho_{23}+\overline{L_1'}^2\rho_{23}'\right)+O(e_2).
\end{aligned}
\tag{C.46}
$$

(iii) $E(Z_{t_1}^2Z_{t_2}^2)$

$$
\begin{aligned}
E(\chi_1\chi_2\chi_3\chi_4)&:L_{12}L_{34}+O(|\rho_{12}|); & E(\chi_1\chi_2)E(\psi_1\psi_2\psi_3\psi_4)&:L_{12}^*L_{34}'+O(|\rho_{12}|);\\
E(\psi_1\psi_2\psi_3\psi_4)&:L_{12}'L_{34}'+O(|\rho_{12}|); & E(\chi_3\chi_4)E(\psi_1\psi_2\psi_3\psi_4)&:L_{12}'L_{34}^*+O(|\rho_{12}|);\\
E(\chi_1\chi_2)E(\psi_3\psi_4)&:L_{12}L_{34}'; & E(\psi_1\psi_2)E(\chi_1\chi_2\chi_3\chi_4)&:L_{12}^*L_{34}+O(|\rho_{12}|);\\
E(\psi_1\psi_2)E(\chi_3\chi_4)&:L_{12}'L_{34}; & E(\psi_3\psi_4)E(\chi_1\chi_2\chi_3\chi_4)&:L_{12}L_{34}^*+O(|\rho_{12}|);
\end{aligned}
$$

$$E(\chi_{12})E(\psi_{12}\psi_3\psi_4) : L_{12}^{**}L_{34}' + O(|\rho_{12}|); \qquad E(\chi_{12}\chi_3\chi_4)E(\psi_{12}\psi_3\psi_4) : L_{12}^{**}L_{34}^{*} + O(|\rho_{12}|);$$
$$E(\chi_{34})E(\psi_{34}\psi_1\psi_2) : L_{12}'L_{34}^{**} + O(|\rho_{12}|); \qquad E(\chi_{34}\chi_1\chi_2)E(\psi_{34}\psi_1\psi_2) : L_{12}^{*}L_{34}^{**} + O(|\rho_{12}|);$$
$$E(\psi_{12})E(\chi_{12}\chi_3\chi_4) : L_{12}^{**}L_{34} + O(|\rho_{12}|); \quad E(\chi_1\chi_2\chi_3\chi_4)E(\psi_1\psi_2\psi_3\psi_4) : L_{12}^{*}L_{34}^{*} + O(|\rho_{12}|);$$
$$E(\psi_{34})E(\chi_{34}\chi_1\chi_2) : L_{12}L_{34}^{**} + O(|\rho_{12}|); \qquad E(\chi_{12,34})E(\psi_{12,34}) : L_{12}^{**}L_{34}^{**} + O(|\rho_{12}|).$$

Thus,

$$\begin{aligned}
E(Z_{t_1}^2 Z_{t_2}^2) &= \sum (L_{12}L_{34} + L_{12}'L_{34}' + L_{12}L_{34}' + L_{12}'L_{34} + L_{12}^{*}L_{34}' \\
&\quad + L_{12}'L_{34}^{*} + L_{12}^{*}L_{34} + L_{12}L_{34}^{*} + L_{12}^{*}L_{34}^{*} + L_{12}^{**}L_{34}' + L_{12}^{**}L_{34} \\
&\quad + L_{12}^{**}L_{34}^{*} + L_{12}'L_{34}^{**} + L_{12}L_{34}^{**} + L_{12}^{*}L_{34}^{**} + L_{12}^{**}L_{34}^{**}) + O(|\rho_{12}|) \\
&= \sum (L_{12} + L_{12}' + L_{12}^{*} + L_{12}^{**})(L_{34} + L_{34}' + L_{34}^{*} + L_{34}^{**}) + O(|\rho_{12}|) \\
&= \left(\overline{L_2} + \overline{L_2'} + \overline{L_2^{*}} + \overline{L_2^{**}}\right)^2 + O(|\rho_{12}|).
\end{aligned} \tag{C.47}$$

(iv) $E(Z_{t_1}^3 Z_{t_2})$

$$E(Z_{t_1}^3 Z_{t_2}) = O(|\rho_{12}|). \tag{C.48}$$

(v) $E(Z_{t_i} Z_{t_j})$, $i \neq j$

$$E(\chi_1\chi_2) : L_1L_2\rho_{ij} + O(\rho_{ij}^2); \qquad E(\psi_1\psi_2) : L_1'L_2'\rho_{ij}' + O(\rho_{ij}^2).$$

Thus,

$$E(Z_{t_i} Z_{t_j}) = \sum_{u_1,u_2=0}^{p} (L_1L_2\rho_{ij} + L_1'L_2'\rho_{ij}') + O(\rho_{ij}^2) = \overline{L_1}^2\rho_{ij} + \overline{L_1'}^2\rho_{ij}' + O(\rho_{ij}^2). \tag{C.49}$$

(vi) $E(Z_{t_i}^2)$

$$E(\chi_1\chi_2) : L_{12}; \qquad E(\chi_1\chi_2)E(\psi_1\psi_2) : L_{12}^{*};$$
$$E(\psi_1\psi_2) : L_{12}'; \qquad E(\chi_{12})E(\psi_{12}) : L_{12}^{**}.$$

Thus,

$$E(Z_{t_i}^2) = \sum_{u_1,u_2=0}^{p} (L_{12} + L'_{12} + L^*_{12} + L^{**}_{12}) = \overline{L_2} + \overline{L'_2} + \overline{L^*_2} + \overline{L^{**}_2}. \tag{C.50}$$

Define V_t as a mean-zero Gaussian $I(d)$ process with $E(V_t^2) = \overline{L_2} + \overline{L'_2} + \overline{L^*_2} + \overline{L^{**}_2}$ and $E(V_{t_i}V_{t_j}) = \overline{L_1}^2 \rho_{ij} + \overline{L'_1}^2 \rho'_{ij}$, for $i \neq j$. Using equations (C.45) to (C.50) to compute the covariances of interest in each case, they are easily shown to be identical to

$$\mathrm{Cov}(V_{s_1}V_{s_2}, V_{s_3}V_{s_4}) = E(V_{s_1}V_{s_3})E(V_{s_2}V_{s_4}) + E(V_{s_1}V_{s_4})E(V_{s_2}V_{s_3}),$$

up to the desired approximation errors. □

Lemma 7 *If V_t is Gaussian $I(1/4)$, under (4.4),*

$$\mathrm{Var}\left\{\widehat{F}_{VV}(\lambda_m)\right\} = O\left(\frac{\log m}{n}\right).$$

Proof. Let $\rho_j = E(V_0V_j)$ and assume $\rho_0 = 1$, without loss of generality. By assumption, $|\rho_j| \leq Kj^{-1/2}$. We will use similar methods to the proof of Lemma 10 in Robinson (1994b), including the decomposition

$$\begin{aligned}
\mathrm{Var}\left\{\widehat{F}_{VV}(\lambda_m)\right\} &= \frac{1}{n^4}\sum_{s,t,u,v=1}^{n} \mathrm{Cov}(V_sV_t, V_uV_v) D_m(\lambda_{t-s})\overline{D_m(\lambda_{v-u})} \\
&= \frac{1}{n^4}\sum_{s,t,u,v=1}^{n} (\rho_{u-s}\rho_{v-t} + \rho_{v-s}\rho_{u-t}) D_m(\lambda_{t-s})\overline{D_m(\lambda_{v-u})} \\
&= \frac{1}{n^4}\sum_{j,k=1}^{m} (W_{j,k-j}W_{k,j-k} + W_{j,-j-k}W_{-k,j+k}),
\end{aligned} \tag{C.51}$$

where

$$W_{j,k} = \sum_{u=1-n}^{n-1} \rho_u e^{i\lambda_j u} T_k(u), \qquad T_k(u) = \sum_{t=1+u^+}^{n-u^-} e^{it\lambda_k},$$

denoting the positive and negative parts of u by $u^+ = (|u| + u)/2$ and $u^- = (|u| - u)/2$ respectively. Note that Robinson (1994b) has a typo in this decomposition, using k instead of $-k$ in the first index of the last W. However, the correct expression is used in the remainder of his proof.

To bound $W_{j,0}$, for $j = 1, \ldots, m$, note that $T_0(u) = n - |u|$. Summation by parts gives

$$
\begin{aligned}
W_{j,0} &= \sum_{u=1-n}^{n-1} (n-|u|)\rho_u e^{i\lambda_j u} = n + \sum_{u=1}^{n-1}(n-u)\rho_u(e^{i\lambda_j u} + e^{-i\lambda_j u}) \\
&= n + \sum_{u=1}^{n-1} \left\{(n-u)\rho_u - (n-u-1)\rho_{u+1}\right\} \left\{D_u(\lambda_j) + \overline{D_u(\lambda_j)}\right\},
\end{aligned}
$$

so using (C.3) we get

$$
\begin{aligned}
|W_{j,0}| &\le n + K\sum_{u=1}^{n-1} \left\{(n-u)|\rho_u - \rho_{u+1}| + |\rho_{u+1}|\right\} |D_u(\lambda_j)| \\
&\le n + K\sum_{u=1}^{n-1} \frac{n}{u}|\rho_{u+1}||D_u(\lambda_j)| \le n + Kn\sum_{u=1}^{[n/j]} u^{-\frac{1}{2}} + K\frac{n^2}{j}\sum_{u=[n/j]}^{n} u^{-\frac{3}{2}} \\
&\le n + Kn\left(\frac{n}{j}\right)^{\frac{1}{2}} + K\frac{n^2}{j}\left(\frac{n}{j}\right)^{-\frac{1}{2}} \le K\frac{n^{\frac{3}{2}}}{j^{-\frac{1}{2}}}.
\end{aligned}
\tag{C.52}
$$

For $k \neq 0$ and $u > 0$, (B.1) implies that $T_k(0) = 0$,

$$
\begin{aligned}
T_k(u) &= \sum_{t=1+u}^{n} e^{it\lambda_k} = \sum_{t=1}^{n} e^{it\lambda_k} - \sum_{t=1}^{u} e^{it\lambda_k} = -D_u(\lambda_k), \\
T_k(-u) &= \sum_{t=1}^{n-u} e^{it\lambda_k} = \sum_{t=1}^{n} e^{it\lambda_k} - e^{i\lambda_k}\sum_{t=n-u}^{n} e^{it\lambda_k} = -e^{i\lambda_k}\overline{D_u(\lambda_k)}.
\end{aligned}
$$

Therefore, using summation by parts,

$$
\begin{aligned}
W_{j,k} &= \sum_{u=1}^{n-1} \rho_u \left\{e^{i\lambda_j u}T_k(u) + e^{-i\lambda_j u}T_k(-u)\right\} \\
&= \sum_{u=1}^{n-1}(\rho_u - \rho_{u+1})\sum_{q=1}^{u} \left\{-e^{i\lambda_j q}D_q(\lambda_k) - e^{i(\lambda_k - \lambda_j q)}\overline{D_q(\lambda_k)}\right\} \\
&\quad + \rho_n \sum_{q=1}^{n-1} \left\{-e^{i\lambda_j q}D_q(\lambda_k) - e^{i(\lambda_k - \lambda_j q)}\overline{D_q(\lambda_k)}\right\},
\end{aligned}
$$

implying

$$|W_{j,k}| \leq K \sum_{u=1}^{n-1} |\rho_u - \rho_{u+1}| \left| \sum_{q=1}^{u} e^{i\lambda_j q} D_q(\lambda_k) \right| + K|\rho_n| \left| \sum_{q=1}^{n-1} e^{i\lambda_j q} D_q(\lambda_k) \right|.$$

Since

$$\begin{aligned} \sum_{q=1}^{u} e^{i\lambda_j q} D_q(\lambda_k) &= \sum_{q=1}^{u} e^{i\lambda_j q} \sum_{t=1}^{q} e^{it\lambda_k} = \sum_{q=1}^{u} e^{i\lambda_j q} \frac{1 - e^{iq\lambda_k}}{e^{-i\lambda_k} - 1} \\ &= \frac{1}{e^{-i\lambda_k} - 1} \sum_{q=1}^{u} (e^{i\lambda_j q} - e^{i\lambda_{j+k} q}) = \frac{D_u(\lambda_j) - D_u(\lambda_{j+k})}{e^{-i\lambda_k} - 1} \end{aligned}$$

and $|e^{-i\lambda_k} - 1| \sim |\lambda|$ as $\lambda \to 0$, we have

$$\left| \sum_{q=1}^{u} e^{i\lambda_j q} D_q(\lambda_k) \right| \leq K \frac{n}{|k|} \left\{ |D_u(\lambda_j)| + |D_u(\lambda_{j+k})| \right\}.$$

So, using (C.3) and for $a = \min\{|j|, |j+k|\}$,

$$\begin{aligned} |W_{j,k}| &\leq K \frac{n}{|k|} \sum_{u=1}^{[n/a]} |\rho_{u+1}| + K \frac{n}{|k|} \sum_{u=[n/a]+1}^{n-1} \frac{|\rho_{u+1}|}{u} \frac{n}{a} + K|\rho_n| \frac{n^2}{|k|a} \\ &\leq K \frac{n}{|k|} \sum_{u=1}^{[n/a]} u^{-\frac{1}{2}} + K \frac{n^2}{|k|a} \sum_{u=[n/a]+1}^{n-1} u^{-\frac{3}{2}} + K \frac{n^{\frac{3}{2}}}{|k|a} \\ &\leq K \frac{n}{|k|} \left(\frac{n}{a} \right)^{\frac{1}{2}} + K \frac{n^2}{|k|a} \left(\frac{n}{a} \right)^{-\frac{1}{2}} + K \frac{n^{\frac{3}{2}}}{|k|a} \leq K \frac{n^{\frac{3}{2}}}{|k|a^{\frac{1}{2}}}, \end{aligned}$$

yielding

$$|W_{j,k-j} W_{k,j-k}| \leq K \frac{n^3}{(j-k)^2 \min\{j,k\}}, \qquad 1 \leq j,k \leq m, \; j \neq k, \tag{C.53}$$

$$|W_{j,-j-k} W_{-k,j+k}| \leq K \frac{n^3}{(j+k)^2 \min\{j,k\}}, \qquad 1 \leq j,k \leq m. \tag{C.54}$$

Thus, using (C.52), (C.53), (C.54) in (C.51),

$$
\begin{aligned}
\operatorname{Var}\left\{\widehat{F}_{VV}(\lambda_m)\right\} &\leq \frac{K}{n^4}\sum_{j=1}^{m}\frac{n^3}{j} + \frac{K}{n^4}\sum_{\substack{j,k=1\\ j\neq k}}^{m}\frac{n^3}{(j-k)^2\min\{j,k\}} + \frac{K}{n^4}\sum_{j,k=1}^{m}\frac{n^3}{(j+k)^2\min\{j,k\}} \\
&\leq K\frac{\log m}{n} + \frac{K}{n^4}\sum_{\substack{j,k=1\\ j<k}}^{m}\frac{n^3}{(j-k)^2 j} + \frac{K}{n^4}\sum_{\substack{j,k=1\\ j\leq k}}^{m}\frac{n^3}{(j+k)^2 j} \\
&\leq K\frac{\log m}{n} + \frac{K}{n}\sum_{j=1}^{m-1} j^{-1}\sum_{a=1}^{m-j} a^{-2} + \frac{K}{n}\sum_{j=1}^{m} j^{-1}\sum_{a=2j}^{m+j} a^{-2} \leq K\frac{\log m}{n}. \;\square
\end{aligned}
$$

References

Andersen, T. G., and T. Bollerslev (1997): "Intraday periodicity and volatility persistence in financial markets," *Journal of Empirical Finance*, 4(2), 115–158.

Baillie, R. T., T. Bollerslev, and H. O. Mikkelsen (1996): "Fractionally integrated generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, 74(1), 3–30.

Bandi, F. M., and B. Perron (2004): "Long memory and the relation between implied and realized volatility," Economics Working Paper Archive at WUSTL, Econometrics.

Bollerslev, T. (1986): "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, 31(3), 302–327.

Bollerslev, T., and H. O. Mikkelsen (1996): "Modeling and pricing long memory in stock market volatility," *Journal of Econometrics*, 73(1), 151–184.

Breidt, F. J., N. Crato, and P. de Lima (1998): "The detection and estimation of long memory in stochastic volatility," *Journal of Econometrics*, 83(1-2), 325–348.

Christensen, B. J., and M. Ø. Nielsen (2004): "Asymptotic normality of narrow-band least squares in the stationary fractional cointegration model and volatility forecasting," forthcoming in Journal of Econometrics.

Davies, R. B., and D. S. Harte (1987): "Tests for Hurst effect," *Biometrika*, 74(1), 95–101.

Delgado, M. A., and P. M. Robinson (1996): "Optimal spectral bandwidth for long memory," *Statistica Sinica*, 6(1), 97–112.

Ding, Z., and C. W. J. Granger (1996): "Modeling volatility persistence of speculative returns: A new approach," *Journal of Econometrics*, 73(1), 185–215.

Ding, Z., C. W. J. Granger, and R. F. Engle (1993): "A long memory property of stock market returns and a new model," *Journal of Empirical Finance*, 1(1), 83–106.

Engle, R. F. (1982): "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica*, 50(4), 987–1008.

Engle, R. F., and T. Bollerslev (1986): "Modelling the persistence of conditional variances," *Econometric Reviews*, 5(1), 1–50, 81–87.

Fama, E. F., and K. R. French (1993): "Common risk factors in the returns on stocks and bonds," *Journal of Financial Economics*, 33(1), 3–56.

Harvey, A. C. (1998): "Long memory in stochastic volatility," in *Forecasting volatility in financial markets*, ed. by J. L. Knight, and S. E. Satchell, Quantitative Finance Series, chap. 12, pp. 307–320. Butterworth-Heinemann, London.

Jeganathan, P. (1999): "On asymptotic inference in cointegrated time series with fractionally integrated errors," *Econometric Theory*, 15(4), 583–621.

Johansen, S. (1991): "Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models," *Econometrica*, 59(6), 1551–1580.

Marinucci, D., and P. M. Robinson (2001): "Semiparametric fractional cointegration analysis," *Journal of Econometrics*, 105(1), 225–247.

Phillips, P. C. B. (1991): "Optimal inference in cointegrated systems," *Econometrica*, 59(2), 283–306.

Robinson, P. M. (1991): "Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression," *Journal of Econometrics*, 47(1), 67–84.

——— (1994a): "Semiparametric analysis of long-memory time series," *Annals of Statistics*, 22(1), 515–539.

——— (1994b): "Rates of convergence and optimal spectral bandwidth for long range dependence," *Probability Theory and Related Fields*, 99, 443–473.

——— (2001): "The memory of stochastic volatility models," *Journal of Econometrics*, 101(2), 195–218.

Robinson, P. M., and J. Hualde (2003): "Cointegration in fractional systems with unknown integration orders," *Econometrica*, 71(6), 1727–1766.

Robinson, P. M., and D. Marinucci (2001): "Narrow band analysis of nonstationary processes," *Annals of Statistics*, 29(4), 947–986.

——— (2003): "Semiparametric frequency domain analysis of fractional cointegration," in *Time series with long memory*, ed. by P. M. Robinson, Advanced Texts in Econometrics, chap. 14, pp. 334–373. Oxford University Press, Oxford.

Robinson, P. M., and Y. Yajima (2002): "Determination of cointegrating rank in fractional systems," *Journal of Econometrics*, 106(2), 217–241.

Slepian, D. (1972): "On the symmetrized Kronecker power of a matrix and extensions of Mehler's formula for Hermite polynomials," *SIAM Journal on Mathematical Analysis*, 3, 606–616.

Stock, J. H. (1987): "Asymptotic properties of least squares estimators of cointegrating vectors," *Econometrica*, 55(5), 1035–1056.

Taylor, S. J. (1986): *Modelling financial time series*. John Wiley and Sons, Chichester.

Whistler, D. E. N. (1990): "Semiparametric models of daily and intradaily exchange rate volatility," PhD thesis, University of London, London School of Economics and Political Science, Department of Economics.

Yong, C. H. (1974): *Asymptotic behaviour of trigonometric series*. Chinese University of Hong Kong, Hong Kong.

Zygmund, A. (1977): *Trigonometric series*. Cambridge University Press, Cambridge.