Queen's Economics Department Working Paper No. 1218

Local polynomial Whittle estimation of perturbed fractional processes

Per Frederiksen
Nordea Markets

Frank S. Nielsen
DONG Energy and CREATES

Morten Ørregaard Nielsen
Queen's University and CREATES

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

9-2009

Local polynomial Whittle estimation of perturbed fractional processes*

Per Frederiksen
Nordea Markets

Frank S. Nielsen†
DONG Energy and CREATES

Morten Ørregaard Nielsen
Queen's University and CREATES

April 12, 2010

Abstract

We propose a semiparametric local polynomial Whittle with noise estimator of the memory parameter in long memory time series perturbed by a noise term which may be serially correlated. The estimator approximates the log-spectrum of the short-memory component of the signal as well as that of the perturbation by two separate polynomials. Including these polynomials we obtain a reduction in the order of magnitude of the bias, but also inflate the asymptotic variance of the long memory estimator by a multiplicative constant. We show that the estimator is consistent for $d \in (0, 1)$, asymptotically normal for $d \in (0, 3/4)$, and if the spectral density is sufficiently smooth near frequency zero, the rate of convergence can become arbitrarily close to the parametric rate, $\sqrt{n}$. A Monte Carlo study reveals that the proposed estimator performs well in the presence of a serially correlated perturbation term. Furthermore, an empirical investigation of the 30 DJIA stocks shows that this estimator indicates stronger persistence in volatility than the standard local Whittle (with noise) estimator.

JEL Classifications: C22.

Keywords: Bias reduction, local Whittle, long memory, perturbed fractional process, semiparametric estimation, stochastic volatility.

1 Introduction

We are interested in estimation of the memory parameter in a so-called perturbed fractional process,

$$z_t = y_t + w_t, \tag{1}$$

i.e. a signal-plus-noise model where the signal process y_t is a long memory process with memory parameter d which is perturbed by the additive noise term w_t. These processes are a version of the

*We are grateful to the editors, two anonymous referees, Torben G. Andersen, Jörg Breitung, Niels Haldrup, Esben Høg, Asger Lunde, and the participants at SETA 2008 for valuable suggestions and comments. This work was partly done while P. Frederiksen was visiting Northwestern University and F. S. Nielsen was visiting Cornell University; their hospitality is gratefully acknowledged. We are grateful for financial support from the Social Sciences and Humanities Research Council of Canada (SSHRC grant no. 410-2009-0183), the Danish Social Sciences Research Council (FSE grant no. 275-05-0220), and the Center for Research in Econometric Analysis of Time Series (CREATES, funded by the Danish National Research Foundation).

†Please address correspondence to: Frank S. Nielsen, CREATES, School of Economics and Management, Aarhus University, Universitetsparken building 1322, 8000 Aarhus C, Denmark; e-mail: fnielsen@creates.au.dk

random walk plus noise or local level unobserved components model, e.g. Harvey (1989), except the signal is a long memory process rather than a random walk. They have found extensive use in modeling the long memory characteristics of many observed time series, in particular financial volatility.

Another motivation for the perturbed fractional process is the version of the long memory stochastic volatility (LMSV) model for financial returns proposed by Bollerslev & Jubinski (1999),

$$r_t = \kappa e^{(y_t + x_t)/2} u_t, \tag{2}$$

where r_t denotes the return, y_t is the (long memory component of) log-volatility of the returns, x_t is a short-memory process, and y_t, x_t, and u_t are independent. Under this model r_t is a martingale difference sequence if u_t is a martingale difference sequence. Model (2) generalizes the usual LMSV model introduced by Breidt, Crato & de Lima (1998) and Harvey (1998),

$$r_t = \kappa e^{y_t/2} u_t. \tag{3}$$

The reasoning behind (2) is that allowing for different short-lived news impacts, while imposing a common long memory component, may provide a better characterization of the joint volume-volatility relationship in the context of the Mixture of Distributions Hypothesis, which asserts that stock returns and trading volumes are jointly dependent on the same underlying latent information arrival process. The formulation in (2) allows the volatility to be affected by both long and short-lived news impacts, which is also consistent with the findings of Liesenfeld (2001). It therefore seems natural that an estimator of the memory in $\log r_t^2$ should be able to incorporate both (2) and (3).

The LMSV models (2) and (3) imply that a logarithmic transformation of the squared returns series, $\log r_t^2$, becomes a long memory signal-plus-noise process (1) where the signal y_t corresponds to (the long memory component of) the log-volatility of the original returns series and w_t is an additive noise term. Specifically, in model (2), $\log r_t^2 = y_t + x_t + \log \kappa^2 + \log u_t^2$, i.e. $w_t = x_t + \log \kappa^2 + \log u_t^2$. In the context of the LMSV model (3), w_t is usually assumed to be *i.i.d.*, but to allow for short-memory persistence in w_t as implied by (2) we will not make that restriction here. In general, when w_t is not assumed to be *i.i.d.*, z_t is referred to as a perturbed fractional process.[1] For reviews of fractionally integrated processes and some applications, see Baillie (1996), Henry & Zaffaroni (2003), or Robinson (2003). In particular, long memory in volatility has received considerable interest recently.[2]

If we assume that the log-volatility process $\{y_t\}$ and the noise process $\{w_t\}$ are independent, the spectral density of z_t can be written as

$$f_z(\lambda) = \lambda^{-2d} \phi_y(\lambda) + \phi_w(\lambda), \tag{4}$$

where $f_y(\lambda) = \lambda^{-2d} \phi_y(\lambda)$ is the spectrum of the signal y_t, $\phi_w(\lambda)$ is the spectrum of the noise term w_t, and d is the degree of long memory in y_t (or equivalently in z_t).

The assumption of independence between the processes $\{y_t\}$ and $\{w_t\}$ rules out the so-called

[1] In the following we use the terms "long memory process" and "fractionally integrated process" or just "fractional process" synonymously, although strictly speaking a fractional process is a particular form of a long memory process.

[2] See, e.g., Ding, Granger & Engle (1993), Baillie, Bollerslev & Mikkelsen (1996), Comte & Renault (1998), Ray & Tsay (2000), Andersen, Bollerslev, Diebold & Ebens (2001), Andersen, Bollerslev, Diebold & Labys (2001, 2003), Wright (2002), Hurvich & Ray (2003), and Arteche (2004) among others.

leverage effect. This assumption is common in the random walk plus noise unobserved components models, and has also been imposed by Breidt et al. (1998), Deo & Hurvich (2001), and Arteche (2004), among others, in the LMSV model. To accommodate the leverage effect, we could allow contemporaneous correlation, while the return process remains a martingale difference sequence by replacing y_t with y_{t-1} in (3). An additional assumption of distributional symmetry around $(0, 0)$ would imply that the spectral density decomposition in (4) holds, see Harvey, Ruiz & Shephard (1994). Alternatively, the model could be modified along the lines of model (P2) of Hurvich, Moulines & Soulier (2005).

In semiparametric spectral estimation of long memory models, the spectrum (4) is typically approximated using the periodogram of the data near the zero frequency, i.e. for frequencies up to $\lambda_m = 2\pi m/n$ only, where n is the sample size and m is a user-chosen bandwidth number which tends to infinity slower than n such that $\lambda_m \to 0$. Although the popular log-periodogram regression estimator of Geweke & Porter-Hudak (1983) and Robinson (1995*b*) and the local Whittle (LW) estimator of Künsch (1987) and Robinson (1995*a*) both preserve consistency and asymptotic normality when applied to perturbed fractional processes, as shown recently by Deo & Hurvich (2001) and Arteche (2004), these estimators can be severely biased since they do not take the perturbation into account. Indeed, for non-perturbed processes (where $\phi_w(\lambda) = 0$) the bias of the standard semiparametric frequency domain estimators is of order $O(\lambda_m^2)$, whereas the leading bias term when $\phi_w(\lambda) \neq 0$ is of order $O(\lambda_m^{2d})$ (in both cases assuming sufficient smoothness of the spectral density). As shown in Deo & Hurvich (2001) and Arteche (2004), this bias is typically negative and can be very large (note that $d < 1$). Therefore, estimating long memory in perturbed time series can be challenging, and calls for an estimator which explicitly accounts for the perturbation.

Sun & Phillips (2003), Hurvich & Ray (2003), Hurvich et al. (2005), and Arteche (2006), among others, have proposed such estimators with $\phi_y(\lambda)$ and $\phi_w(\lambda)$ locally approximated by constants as $\lambda \to 0$, see section 2 below. On the other hand, we propose an estimator where we allow the logarithms of both the spectrum of the short-memory component of the signal and the spectrum of the perturbation, i.e. $\log \phi_y(\lambda)$ and $\log \phi_w(\lambda)$, to be approximated by polynomials $h_y(\boldsymbol{\theta}_y, \lambda)$ and $h_w(\boldsymbol{\theta}_w, \lambda)$ of (finite and even) orders $2R_y$ and $2R_w$ near the zero frequency, thereby obtaining a bias reduction depending on the smoothness of $\phi_y(\lambda)$ and $\phi_w(\lambda)$ near the origin. The approach taken here in modeling the short-run dynamics by a polynomial was introduced by Andrews & Guggenberger (2003) and Andrews & Sun (2004) for non-perturbed processes, but is novel in the context of perturbed fractional processes. To maintain generality, $\phi_y(\lambda)$ and $\phi_w(\lambda)$ are only characterized by regularity conditions near frequency zero instead of imposing specific functional forms.

The LMSV model (3) often assumes that the noise term is *i.i.d.* in which case $\phi_w(\lambda) = \sigma_w^2/(2\pi)$ is a constant. This case is of independent interest and is considered in simulations and in an empirical study in Frederiksen & Nielsen (2008). In that paper $\phi_y(\lambda)$ itself is approximated by a polynomial and $\phi_w(\lambda)$ by a constant as $\lambda \to 0$ thus focusing on exactly the LMSV model (3). However, the theory for their estimator is developed in the present paper.[3]

[3]Note that their specification involves modeling $\phi_y(\lambda)$ rather than $\log \phi_y(\lambda)$ by a polynomial. The two approaches provide the same degree of local approximation to the spectral density, so our proofs are valid with little change for their specification as well. However, the approach taken here is preferred in practice since the estimate of $\phi_y(\lambda)$ is guaranteed to be positive. Note also that

Thus, to allow serial dependence in the noise as in (2) above we include both polynomials, $h_y(\boldsymbol{\theta}_y, \lambda)$ and $h_w(\boldsymbol{\theta}_w, \lambda)$, and call the resulting estimator the local polynomial Whittle with noise (LPWN) estimator. Furthermore, empirical studies have typically found that the noise term has much higher (long-run) variance than the short-memory component of the signal. Indeed, Breidt et al. (1998) and Hurvich & Ray (2003) find that the noise term may be as much as 10 or 20 times as variable as the short-memory component of the signal. Thus, careful modeling of the noise term is important and this consideration has lead us to approximate the log-spectrum of the noise term by a polynomial instead of a constant as $\lambda \to 0$.

Our results show that introducing $h_y(\boldsymbol{\theta}_y, \lambda)$ and $h_w(\boldsymbol{\theta}_w, \lambda)$ inflates the asymptotic variance of the long memory estimator, $\hat{d}$, by a multiplicative constant which depends on the true long memory parameter, d, and decreases when d increases. However, we obtain a reduction in the order of magnitude of the bias if $\phi_y(\lambda)$ and $\phi_w(\lambda)$ are sufficiently smooth near frequency zero. We show that the estimator is consistent for $d \in (0, 1)$, asymptotically normal for $d \in (0, 3/4)$, and if $\phi_y(\lambda)$ and $\phi_w(\lambda)$ are infinitely smooth near frequency zero, the rate of convergence can become arbitrary close to the parametric rate, $n^{1/2}$. This constitutes a rate of convergence improvement relative to Sun & Phillips (2003), Hurvich & Ray (2003), and Hurvich et al. (2005) who are only able to obtain a semiparametric rate of convergence, $m^{1/2}$, which is much slower than the parametric rate due to their requirement that $m^{1+2\beta}/n^{2\beta} \to 0$ for some $\beta \in (0, 2]$.

We present the results of a Monte Carlo study which illustrates the usefulness of the proposed LPWN estimator. Compared to standard estimators, such as Hurvich & Ray's (2003) local Whittle with noise (LWN) estimator, the LPWN estimator is able to achieve considerable bias reductions in practice, especially in cases with short-run dynamics in both the signal and noise components. We also include an empirical application to daily log-squared returns series of the 30 DJIA stocks where the LPWN estimator indicates stronger persistence in volatility than the standard estimators, and for most of the stocks produce estimates of d in the nonstationary region.

The remainder of the paper is organized as follows. In the next section we discuss semiparametric spectral estimation of long memory for perturbed processes and formally define the proposed local Whittle estimator. In section 3 we establish consistency and asymptotic normality of the estimator. Section 4 investigates the finite sample performance in simulations, and section 5 presents the empirical application. Section 6 concludes. The proofs of our theorems are gathered in the appendix.

2 Local Whittle estimation of perturbed fractional processes

Semiparametric frequency domain estimators for non-perturbed fractional processes are based on the local approximation

$$f_z(\lambda) \sim G\lambda^{-2d} \text{ as } \lambda \to 0, \tag{5}$$

where $G = \phi_y(0) > 0$ is a constant and the symbol "$\sim$" means that the ratio of the left- and right-hand sides tends to one in the limit. Thus, the estimators enjoy robustness to short-run dynamics, since they use only information from periodogram ordinates in the vicinity of the origin.

The local Whittle (LW) estimation method of Künsch (1987) and Robinson (1995*a*) has become popular because of its likelihood interpretation, nice asymptotic properties, and mild assumptions.

It is defined as the minimizer of the (negative) local Whittle likelihood function

$$Q(G,d)=\frac{1}{m}\sum_{j=1}^{m}\left[\log\left(G\lambda_j^{-2d}\right)+\frac{I_z(\lambda_j)}{G\lambda_j^{-2d}}\right], \tag{6}$$

where $m=m(n)$ is a bandwidth number which tends to infinity as $n\to\infty$ but at a slower rate than n, $\lambda_j=2\pi j/n$ are the Fourier frequencies, and $I_z(\lambda)=(2\pi n)^{-1}|\sum_{t=1}^{n}z_te^{it\lambda}|^2$ is the periodogram of z_t. Note that the estimator is invariant to a non-zero mean since $j=0$ is left out of the minimization. Concentrating (6) with respect to G, the estimator of d is

$$\hat{d}_{LW}=\arg\min_{d}\left[\log\hat{G}(d)-2d\frac{1}{m}\sum_{j=1}^{m}\log\lambda_j\right],\quad \hat{G}(d)=\frac{1}{m}\sum_{j=1}^{m}\lambda_j^{2d}I_z(\lambda_j).$$

It was shown by Robinson (1995*a*) that

$$\sqrt{m}(\hat{d}_{LW}-d)\xrightarrow{d}N(0,1/4), \tag{7}$$

and later by Velasco (1999) that the range of consistency is $d\in(-1/2,1]$ and the range of asymptotic normality is $d\in(-1/2,3/4)$.

To reduce the asymptotic bias of the LW estimator, Andrews & Sun (2004) suggested to replace the constant, $\log G$, in (6) by the polynomial $\log G-\sum_{r=1}^{R}\theta_r\lambda_j^{2r}$. That is, to model the logarithm of the spectral density of the short-memory component by a polynomial instead of a constant in the vicinity of the origin. This leads to the following (negative) local Whittle likelihood function,

$$Q(G,d,\boldsymbol{\theta})=\frac{1}{m}\sum_{j=1}^{m}\left[\log\left(\lambda_j^{-2d}G\exp\left(-\sum_{r=1}^{R}\theta_r\lambda_j^{2r}\right)\right)+\frac{I_z(\lambda_j)}{\lambda_j^{-2d}G\exp\left(-\sum_{r=1}^{R}\theta_r\lambda_j^{2r}\right)}\right],$$

such that

$$\begin{aligned}(\hat{d}_{LPW},\hat{\boldsymbol{\theta}}) &= \arg\min_{d\in(-1/2,1/2),\boldsymbol{\theta}\in\Theta}\left[\log\hat{G}(d,\boldsymbol{\theta})-2d\frac{1}{m}\sum_{j=1}^{m}\log\lambda_j-\frac{1}{m}\sum_{j=1}^{m}\sum_{r=1}^{R}\theta_r\lambda_j^{2r}\right],\\ \hat{G}(d,\boldsymbol{\theta}) &= \frac{1}{m}\sum_{j=1}^{m}\lambda_j^{2d}\exp\left(\sum_{r=1}^{R}\theta_r\lambda_j^{2r}\right)I_z(\lambda_j),\end{aligned}$$

where Θ is a compact and convex set in $\mathbb{R}^R$. As shown by Andrews & Sun (2004), this method does, however, increase the asymptotic variance of $\hat{d}$ in (7) by a multiplicative constant.

For non-perturbed fractional processes, the asymptotic bias of $\hat{d}_{LW}$ and $\hat{d}_{LPW}$ is of order $O(\lambda_m^{\min\{s,2\}})$ and $O(\lambda_m^{\min\{s,2+2R\}})$, respectively, where s is a measure of the smoothness of the spectral density near frequency zero, see below. However, for perturbed fractional processes the bias is of order $O(\lambda_m^{\min\{s,2d\}})$ and, as shown by e.g. Hurvich & Ray (2003) and Arteche (2004), this bias is typically negative and can be very severe.

For perturbed fractional processes we have the spectral representation (4), which implies $f_z(\lambda)\sim G\lambda^{-2d}+G\theta_\rho$ as $\lambda\to 0$, where the constant $\theta_\rho=\phi_w(0)/\phi_y(0)>0$ is the long-run noise-to-signal ratio. There are two main consequences: first, the extra additive term θ_ρ needs to be taken into account to avoid serious asymptotic bias as mentioned above, and second the rate of convergence of the estimators is reduced if the extra term is not modeled. The latter follows because the choice of bandwidth parameter is severely constrained for perturbed fractional processes when the perturbation term in (4) is not modeled. Thus, for non-perturbed processes (with $s \geq 2$) the bandwidth requirement is typically $m = o(n^{4/5})$, whereas for perturbed processes it is $m = o(n^{2d/(1+2d)})$ (apart from logarithmic terms). Since the estimator is $\sqrt{m}-$consistent and $d \leq 1$ this is a serious constraint.

To allow for (moderate) nonstationarity in volatility we generalize (1) as

$$z_t = \begin{cases} y_t + w_t & \text{if } d \in (0, 1/2), \\ \sum_{s=1}^{t} x_s + w_t & \text{if } d \in [1/2, 1), \end{cases} \tag{8}$$

where, if $d \in [1/2, 1)$, x_t has spectrum of the form $f_x(\lambda) = \lambda^{-2d_x} \phi_x(\lambda)$ with memory parameter $d_x = d - 1$. Defining $y_t = \sum_{s=1}^{t} x_s$ if $d \in [1/2, 1)$, this approach allows $z_t = y_t + w_t$ to possibly be nonstationary with memory parameter $d \in (0, 1)$. Velasco (1999), Hurvich & Ray (2003), and Hurvich et al. (2005) also assume this type of process. Since $\{\sum_{s=1}^{t} x_s\}$ is nonstationary z_t does not have a spectral density if $d \in [1/2, 1)$ but it has a pseudo spectral density, see e.g. Hurvich & Ray (1995) and Velasco (1999). Thus, we may define

$$\begin{aligned} f_z(\lambda) &= \begin{cases} f_y(\lambda) + f_w(\lambda) & \text{if } d \in (0, 1/2) \\ \left|1 - e^{i\lambda}\right|^{-2} f_x(\lambda) + f_w(\lambda) & \text{if } d \in [1/2, 1) \end{cases} \\ &= \lambda^{-2d} \left(\phi_y(\lambda) + \lambda^{2d} \phi_w(\lambda) \right), \end{aligned} \tag{9}$$

where we maintain the assumption of independence between $\{y_t\}$ and $\{w_t\}$.

The functions $\phi_y(\lambda)$ and $\phi_w(\lambda)$ are both positive constants at $\lambda = 0$, so we set $\phi_y(0) = G$ and $\phi_w(0) = G\theta_\rho$. Since both functions are everywhere positive we model their logarithms. Specifically, we propose to approximate the logarithms of $\phi_y(\lambda)$ and $\phi_w(\lambda)$ by

$$\log \phi_y(\lambda) \simeq \log G + h_y(\boldsymbol{\theta}_y, \lambda) \text{ and } \log \phi_w(\lambda) \simeq \log G + \log \theta_\rho + h_w(\boldsymbol{\theta}_w, \lambda), \tag{10}$$

respectively, where $G > 0$ and $h_a(\boldsymbol{\theta}_a, \lambda) = \sum_{r=1}^{R_a} \theta_{a,r} \lambda^{2r}, a = y, w$.[4] If $R_a = 0$ we set $h_a(\boldsymbol{\theta}_a, \lambda) = 0$. Note that the parameter θ_ρ is the long-run noise-to-signal ratio because $\phi_w(0)/\phi_y(0) = \theta_\rho$, and thus we assume $\theta_\rho > 0$. Defining also the function

$$h(d, \boldsymbol{\theta}, \lambda) = \exp(h_y(\boldsymbol{\theta}_y, \lambda)) + \theta_\rho \lambda^{2d} \exp(h_w(\boldsymbol{\theta}_w, \lambda)) \tag{11}$$

with $\boldsymbol{\theta} = (\boldsymbol{\theta}'_y, \theta_\rho, \boldsymbol{\theta}'_w)'$, we approximate (9) or equivalently (4) locally near the zero frequency by

$$f_z(\lambda) = \lambda^{-2d} \exp(\log \phi_y(\lambda)) + \exp(\log \phi_w(\lambda)) \simeq G\lambda^{-2d} h(d, \boldsymbol{\theta}, \lambda),$$

which yields the (concentrated) local Whittle log-likelihood

$$Q(d, \boldsymbol{\theta}) = \log \hat{G}(d, \boldsymbol{\theta}) - \frac{2d}{m} \sum_{j=1}^{m} \log \lambda_j + \frac{1}{m} \sum_{j=1}^{m} \log h(d, \boldsymbol{\theta}, \lambda_j), \tag{12}$$

$$\hat{G}(d, \boldsymbol{\theta}) = \frac{1}{m} \sum_{j=1}^{m} \frac{\lambda_j^{2d} I_z(\lambda_j)}{h(d, \boldsymbol{\theta}, \lambda_j)}. \tag{13}$$

Thus, we propose to minimize (12) over the admissible set $D \times \Theta$,

$$(\hat{d}, \hat{\boldsymbol{\theta}}) = \underset{(d,\boldsymbol{\theta}) \in D \times \Theta}{\arg\min} \, Q(d, \boldsymbol{\theta}),$$

[4] Note that $\log \phi_y(\lambda)$ and $\log \phi_w(\lambda)$ are symmetric around $\lambda = 0$ and are therefore approximated by even polynomials.

where Θ is a compact and convex set in $\mathbb{R}^{R_y} \times (0,\infty) \times \mathbb{R}^{R_w}$ and $D = [d_1, d_2]$ with $0 < d_1 < d_2 < 1$. This defines the LPWN estimator.

Note that $h(d, \boldsymbol{\theta}, \lambda) = 1$ is the standard local Whittle specification in (6), which does not explicitly account for the perturbation. For $R_y = R_w = 0$ we get $h(d, \boldsymbol{\theta}, \lambda) = 1 + \theta_\rho \lambda^{2d}$, where $\phi_y(\lambda)$ and $\phi_w(\lambda)$ in (4) are both modeled locally by constants, which is the LWN estimator of Hurvich & Ray (2003) and Hurvich et al. (2005) (parameterization (P1)). Thus, our model parameterization includes the standard LW estimator and the LWN estimator as special cases. Furthermore, the model with $R_w = 0$, where the noise is modeled by a constant near the zero frequency, is analyzed empirically and in simulations by Frederiksen & Nielsen (2008), using the asymptotic theory provided here.

3 Asymptotic properties

In this section we first introduce the assumptions needed to establish consistency and asymptotic normality of the proposed estimator for the perturbed fractional process, and consequently we present the main results in two theorems. In the following, true values of the parameters are denoted by subscript zero and $\lfloor x \rfloor$ denotes the integer part of a real number x. We also define a function $\phi(\lambda)$ to be smooth of order s at $\lambda = 0$ if, in a neighborhood of $\lambda = 0$, $\phi(\lambda)$ is $\lfloor s \rfloor$ times continuously differentiable with $\lfloor s \rfloor$ −derivative, $\phi^{(\lfloor s \rfloor)}$, satisfying $|\phi^{(\lfloor s \rfloor)}(\lambda) - \phi^{(\lfloor s \rfloor)}(0)| \leq C |\lambda|^{s - \lfloor s \rfloor}$ for some constant $C < \infty$.

A1 The noise process $\{w_t\}$ is independent of the signal process $\{y_t\}$.

A2 The spectral density of z_t is $f_z(\lambda) = \lambda^{-2d_0} G_0 \frac{\phi_y(\lambda)}{\phi_y(0)} + G_0 \theta_{0\rho} \frac{\phi_w(\lambda)}{\phi_w(0)}$, where $\phi_y(\lambda)$ and $\phi_w(\lambda)$ are real, even, positive, continuous functions on $[-\pi, \pi)$, and $d_0 \in D = [d_1, d_2]$ with $0 < d_1 < d_2 < 1$.

A3 The functions $\phi_y(\lambda)$ and $\phi_w(\lambda)$ are smooth of orders s_y and s_w at $\lambda = 0$, where $s_y > 2R_y$, $s_w > 2R_w$, and $s_y, s_w \geq 1$.

Assumption A1 is the independence assumption used above to write the spectral density of z_t as the sum of the (pseudo) spectral densities of y_t and w_t. Assumption A3 is a smoothness condition on the functions $\phi_y(\lambda)$ and $\phi_w(\lambda)$ similar to that applied by Andrews & Sun (2004). Note that Assumption A3 holds for all $s_y < \infty$ when, e.g., y_t is a finite order ARFIMA process, and for all $s_w < \infty$ when, e.g., w_t is a finite order ARMA process. Under Assumption A3 we establish the following Taylor series expansions of $\log \phi_y(\lambda)$ and $\log \phi_w(\lambda)$ around $\lambda = 0$ (recall that odd order derivatives of even functions are zero at frequency zero),

$$\log \phi_y(\lambda) = \log \phi_y(0) + \sum_{r=1}^{\lfloor s_y/2 \rfloor} \theta_{y,r} \lambda^{2r} + O(\lambda^{s_y}) \text{ as } \lambda \to 0$$

and

$$\log \phi_w(\lambda) = \log \phi_y(0) + \log \frac{\phi_w(0)}{\phi_y(0)} + \sum_{r=1}^{\lfloor s_w/2 \rfloor} \theta_{w,r} \lambda^{2r} + O(\lambda^{s_w}) \text{ as } \lambda \to 0.$$

With these expansions we can now state (10) more precisely as

$$\begin{aligned}\log \phi_y(\lambda) &= \log G_0 + h_y(\boldsymbol{\theta}_{0y},\lambda) + O(\lambda^{\min\{s_y,2+2R_y\}}) \text{ as } \lambda \to 0,\\ \log \phi_w(\lambda) &= \log G_0 + \log\theta_{0\rho} + h_w(\boldsymbol{\theta}_{0w},\lambda) + O(\lambda^{\min\{s_w,2+2R_w\}}) \text{ as } \lambda \to 0,\end{aligned}$$

where the true values G_0 and $\boldsymbol{\theta}_0 = (\boldsymbol{\theta}'_{0y}, \theta_{0\rho}, \boldsymbol{\theta}'_{0w})'$ are thus $G_0 = \phi_y(0)$, $(\boldsymbol{\theta}_{0a})_r = \frac{1}{(2r)!}\frac{\partial^{2r}}{\partial\lambda^{2r}} \log\phi_a(\lambda)|_{\lambda=0}$, $r = 1,\dots,R_a$, $a = y, w$, and $\theta_{0\rho} = \phi_w(0)/\phi_y(0)$. Hence, defining the function

$$g(d,\boldsymbol{\theta},\lambda) = G_0\lambda^{-2d}h(d,\boldsymbol{\theta},\lambda), \tag{14}$$

the approximation $g(d_0,\boldsymbol{\theta}_0,\lambda)$ to $f_z(\lambda)$ satisfies

$$\begin{aligned}\frac{f_z(\lambda)}{g(d_0,\boldsymbol{\theta}_0,\lambda)} &= 1 + \frac{\exp(h_y(\boldsymbol{\theta}_{0y},\lambda))\left(\exp\left(O(\lambda^{\min\{s_y,2+2R_y\}})\right)-1\right)}{h(d_0,\boldsymbol{\theta}_0,\lambda)}\\ &\quad + \lambda^{2d_0}\theta_{0\rho}\frac{\exp(h_w(\boldsymbol{\theta}_{0w},\lambda))\left(\exp\left(O(\lambda^{\min\{s_w,2+2R_w\}})\right)-1\right)}{h(d_0,\boldsymbol{\theta}_0,\lambda)}\\ &= 1 + O(\lambda^{\min\{s_y,2+2R_y\}}) + \lambda^{2d_0}O(\lambda^{\min\{s_w,2+2R_w\}}) \text{ as } \lambda \to 0 \end{aligned} \tag{15}$$

using the well known expansion $\exp(x) = 1 + x + \frac{x^2}{2!} + \dots$ of the exponential function.

A4 (a) The signal y_t has zero mean and admits an infinite order moving average representation $y_t = \sum_{j=0}^{\infty}\beta_{yj}\varepsilon_{t-j}$ (stationary case) or $\Delta y_t = x_t = \sum_{j=0}^{\infty}\beta_{yj}\varepsilon_{t-j}$ (nonstationary case), where $\sum_{j=0}^{\infty}\beta_{yj}^2 < \infty$ and ε_t satisfies, for all t, $E(\varepsilon_t|\mathcal{F}_{t-1}) = 0$, $E\left(\varepsilon_t^2\middle|\mathcal{F}_{t-1}\right) = 1$, $E\left(\varepsilon_t^3\middle|\mathcal{F}_{t-1}\right) = \mu_{\varepsilon 3} < \infty$, and $E\left(\varepsilon_t^4\middle|\mathcal{F}_{t-1}\right) = \mu_{\varepsilon 4} < \infty$ almost surely, where $\mathcal{F}_{t-1}$ is the σ-field generated by $\{\varepsilon_s, s < t\}$.

(b) There exists a random variable ε with $E(\varepsilon^2) < \infty$ such that for all $\tau > 0$ and some $K > 0$, $P(|\varepsilon_t| > \tau) < KP(|\varepsilon| > \tau)$.

(c) For $\beta_y(\lambda) = \sum_{k=0}^{\infty}\beta_{yk}e^{ik\lambda}$, the derivative satisfies $\frac{\partial}{\partial\lambda}\beta_y(\lambda) = O\left(|\beta_y(\lambda)|/\lambda\right)$ as $\lambda \to 0$.

A5 (a) The noise w_t has zero mean and admits an infinite order moving average representation $w_t = \sum_{j=0}^{\infty}\beta_{wj}\eta_{t-j}$, where $\sum_{j=0}^{\infty}\beta_{wj}^2 < \infty$ and η_t satisfies, for all t, $E(\eta_t|\mathcal{F}_{t-1}) = 0$, $E\left(\eta_t^2\middle|\mathcal{F}_{t-1}\right) = 1$, $E\left(\eta_t^3\middle|\mathcal{F}_{t-1}\right) = \mu_{\eta 3} < \infty$, and $E\left(\eta_t^4\middle|\mathcal{F}_{t-1}\right) = \mu_{\eta 4} < \infty$ almost surely, where $\mathcal{F}_{t-1}$ is the σ-field generated by $\{\eta_s, s < t\}$.

(b) There exists a random variable η with $E(\eta^2) < \infty$ such that for all $\tau > 0$ and some $K > 0$, $P(|\eta_t| > \tau) < KP(|\eta| > \tau)$.

(c) For $\beta_w(\lambda) = \sum_{k=0}^{\infty}\beta_{wk}e^{ik\lambda}$, the derivative satisfies $\frac{\partial}{\partial\lambda}\beta_w(\lambda) = O(|\beta_w(\lambda)|/\lambda)$ as $\lambda \to 0$.

We assume that the signal process y_t has zero mean. Since our estimator is a function of the periodogram at nonzero frequencies only, this is without loss of generality in the stationary case. In the nonstationary case the zero mean assumption implies that z_t is free of linear trends which does entail a loss of generality in that case. Importantly, Assumptions A4 and A5 allow for non-Gaussian processes. Note that Assumptions A1-A4 plus the assumption that w_t is white noise with finite fourth moment imply the assumptions needed on y_t and w_t to prove consistency and asymptotic normality (if, in addition, $d_2 < 3/4$) of the LWN estimator of Hurvich & Ray (2003).

It follows from Theorems 1 and 2 below that their results for the LWN estimator are also valid for our more general assumptions on w_t in Assumption A5.

A6 Θ is a compact and convex subset of $\mathbb{R}^{R_y} \times (0,\infty) \times \mathbb{R}^{R_w}$ and $\boldsymbol{\theta}_0$ is in the interior of Θ.

We are now ready to prove consistency of our estimator. Some assumptions could be relaxed somewhat to prove this theorem, see e.g. Hurvich et al. (2005), but for simplicity we have preferred to list only one set of assumptions which will be used also for the proof of asymptotic normality below. The proofs of both theorems are given in the appendix.

Theorem 1 *If Assumptions A1-A6 hold and the bandwidth $m = m(n)$ is such that*

$$\frac{1}{m} + \frac{m}{n} \to 0, \tag{16}$$

then $\hat{d} - d_0 = o_P((\log n)^{-5})$.

Note that the theorem proves consistency only for the estimator of the memory parameter (at logarithmic rate). There is no proof of consistency for the estimators of the polynomial parameters $\boldsymbol{\theta}$. The strategy of proof in Hurvich et al. (2005) would require next a separate proof of consistency for the polynomial parameters. However, we follow instead the method of proof in Andrews & Sun (2004) which does not require an intermediate result on the consistency of $\hat{\boldsymbol{\theta}}$. Thus, we present next the joint asymptotic normality[5] result for $\hat{d}$ and $\hat{\boldsymbol{\theta}}$. Let $\mathbb{I}\{A\}$ be the indicator function of the event A.

Theorem 2 *Let Assumptions A1-A6 hold with d_0 in the interior of $D = [d_1, d_2]$, $0 < d_1 < d_2 < 3/4$, and suppose the bandwidth $m = m(n)$ is such that*

$$\frac{m^{1+4R_y}}{n^{4R_y}} + \frac{m^{1+4(d_0+R_w)}}{n^{4(d_0+R_w)}} \to \infty \text{ and } \frac{m^{2\varphi_y+1}}{n^{2\varphi_y}} + \frac{m^{2\varphi_w+4d_0+1}}{n^{2\varphi_w+4d_0}} \to 0, \tag{17}$$

where $\varphi_a = \min\{s_a, 2+2R_a\}, a = y, w$. Then $\hat{d}$ and $\hat{\boldsymbol{\theta}}$ are both consistent and

$$\mathbf{B}_n \begin{pmatrix} \hat{d} - d_0 \\ \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \end{pmatrix} \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Omega}^{-1}_{R_y,R_w}), \quad \boldsymbol{\Omega}_{R_y,R_w} = \begin{pmatrix} 4 & \boldsymbol{\mu}'_{R_y} & \boldsymbol{\nu}'_{R_w} \\ \boldsymbol{\mu}_{R_y} & \boldsymbol{\Gamma}_{R_y} & \boldsymbol{\psi}'_{R_y,R_w} \\ \boldsymbol{\nu}_{R_w} & \boldsymbol{\psi}_{R_w,R_y} & \boldsymbol{\Psi}_{R_w} \end{pmatrix},$$

where $\mathbf{B}_n = \mathbf{B}_n(d_0)$ is the $(R_y + R_w + 2) \times (R_y + R_w + 2)$ deterministic diagonal matrix with diagonal elements

$$(\mathbf{B}_n)_{11} = \sqrt{m}, \ (\mathbf{B}_n)_{k+1,k+1} = \sqrt{m}\lambda_m^{2k} \text{ for } k = 1, \ldots, R_y,$$
$$\text{and } (\mathbf{B}_n)_{k+R_y+2,k+R_y+2} = \sqrt{m}\lambda_m^{2d_0+2k} \text{ for } k = 0, \ldots, R_w,$$

$\boldsymbol{\mu}_{R_y}$ and $\boldsymbol{\nu}_{R_w} = \boldsymbol{\nu}_{R_w}(d_0, \theta_{0\rho})$ are the vectors

$$(\boldsymbol{\mu}_{R_y})_k = \frac{-4k}{(1+2k)^2} \text{ for } k = 1, \ldots, R_y \text{ and } (\boldsymbol{\nu}_{R_w})_{k+1} = \frac{-4(d_0+k)\theta_{0\rho}^{\mathbb{I}\{k\geq 1\}}}{(1+2d_0+2k)^2} \text{ for } k = 0, \ldots, R_w,$$

[5] Note that asymptotic normality is obtained when $d_0 \in (0, 3/4)$. Phillips & Shimotsu (2004) showed that the asymptotic distribution of the local Whittle estimator is non-normal for $d_0 \in [3/4, 1]$, and we conjecture that a similar result holds for the LPWN estimator.

$\mathbf{\Gamma}_{R_y}$ *and* $\mathbf{\Psi}_{R_w} = \mathbf{\Psi}_{R_w}(d_0, \theta_{0\rho})$ *are the* $R_y \times R_y$ *and* $(R_w + 1) \times (R_w + 1)$ *matrices*

$$\left(\mathbf{\Gamma}_{R_y}\right)_{ik} = \frac{4ik}{(1+2i+2k)(1+2i)(1+2k)} \text{ for } i,k = 1,\ldots,R_y,$$

$$(\mathbf{\Psi}_{R_w})_{i+1,k+1} = \frac{4(d_0+i)(d_0+k)\theta_{0\rho}^{\mathbb{I}\{k\geq 1\}+\mathbb{I}\{i\geq 1\}}}{(1+2i+2k+4d_0)(1+2i+2d_0)(1+2k+2d_0)} \text{ for } i,k = 0,\ldots,R_w,$$

and $\boldsymbol{\psi}_{R_w,R_y} = \boldsymbol{\psi}_{R_w,R_y}(d_0, \theta_{0\rho})$ *is the* $(R_w+1) \times R_y$ *matrix*

$$(\boldsymbol{\psi}_{R_w,R_y})_{i+1,k} = \frac{4k(d_0+i)\theta_{0\rho}^{\mathbb{I}\{i\geq 1\}}}{(1+2d_0+2k+2i)(1+2d_0+2i)(1+2k)} \text{ for } i = 0,\ldots R_w, k = 1,\ldots,R_y.$$

If $R_y = 0$ *define* $\mathbf{\Omega}_{0,R_w} = \begin{pmatrix} 4 & \boldsymbol{\nu}'_{R_w} \\ \boldsymbol{\nu}_{R_w} & \mathbf{\Psi}_{R_w} \end{pmatrix}$.

First of all, we note that by setting $R_y = R_w = 0$ we obtain as a special case the results for the LWN estimator of Hurvich & Ray (2003). Secondly, the leading $(R_y+1) \times (R_y+1)$ submatrix of $\mathbf{\Omega}_{R_y,R_w}$ is the same as that obtained by Andrews & Sun (2004). Third, we note that the asymptotic variance of $\sqrt{m}(\hat{d} - d_0)$ is free of the parameters $\boldsymbol{\theta}_0$, including the noise-to-signal ratio[6] $\theta_{0\rho}$, but it depends on d_0. In fact, the use of the polynomials $h_y(\boldsymbol{\theta}_y, \lambda)$ and $h_w(\boldsymbol{\theta}_w, \lambda)$ increases the asymptotic variance of $\hat{d}$ by a multiplicative constant compared to the LWN estimator of Hurvich & Ray (2003). For example, by use of the partitioned matrix inverse formula we note that the (1,1) element of $\mathbf{\Omega}_{0,0}^{-1}$, i.e. the asymptotic variance of the LWN estimator, with $d_0 = 0.4$ is approximately 1.27, and the corresponding elements of $\mathbf{\Omega}_{1,0}^{-1}$, $\mathbf{\Omega}_{0,1}^{-1}$, and $\mathbf{\Omega}_{1,1}^{-1}$ are approximately 2.85, 2.33, and 5.24, respectively. Andrews & Sun (2004) obtain a similar result for their LPW estimator in a non-perturbed model. In particular, the variance of the LPWN estimator with $R_y = 1, R_w = 0$ is 2.25 times that of the LWN estimator, which is exactly the same multiplicative constant found by Andrews & Sun (2004).

The first condition in (17) guarantees that all the elements of the scaling matrix $\mathbf{B}_n$ diverge as $n \to \infty$, which is a minimal condition for consistency. The second condition restricts the expansion rate of the bandwidth to control bias and ensures that the estimator uses only information from periodogram ordinates sufficiently near the zero frequency. Alternatively, we can view the bandwidth conditions in (17) separately for the signal process and the noise process. In this way we would write the conditions as

$$\frac{m^{1+4R_y}}{n^{4R_y}} \to \infty, \frac{m^{2\varphi_y+1}}{n^{2\varphi_y}} \to 0 \text{ and } \frac{m^{1+4(d_0+R_w)}}{n^{4(d_0+R_w)}} \to \infty, \frac{m^{2\varphi_w+4d_0+1}}{n^{2\varphi_w+4d_0}} \to 0.$$

It is now easy to see that the bandwidth conditions for both the signal process and the noise process are always compatible because $s_y > 2R_y$ and $s_w > 2R_w$, respectively, by Assumption A3.

Note that the second condition in (17) implies that if $\phi_y(\lambda)$ and $\phi_w(\lambda)$ are infinitely smooth near frequency zero, i.e. if they are smooth of any orders $s_y < \infty$ and $s_w < \infty$, then any (R_y, R_w) can be chosen and the estimator is $n^{1/2-\delta}$ consistent for all $\delta > 0$. Hence, in that case, the rate of convergence is arbitrarily close to the parametric rate. Thus, the condition (17) allows the bandwidth m to be much larger than for the LWN estimator and the standard LW estimator, which (assuming $s_y \geq 2, s_w \geq 2$) require that $m^5 n^{-4} \to 0$ and $m^{4d_0+1}n^{-4d_0} \to 0$, respectively, see Hurvich & Ray (2003) and Arteche (2004). Therefore, Theorem 2 provides an improvement in the rate

[6] This fact can be seen using the formula for the inverse of a partitioned matrix.

of convergence relative to existing estimators of the memory parameter for perturbed fractional processes. This comes at the cost of an increase in the asymptotic variance by a multiplicative constant, but that is clearly more than off-set by the faster rate of convergence, at least asymptotically. For example, in the empirically relevant case of $d_0 = 0.4$, which is a typical value of d_0 for financial volatility series, the LW estimator is at most $n^{0.31}$-consistent and the LWN estimator is at most $n^{0.4}$-consistent, whereas our estimator can be arbitrarily close to $n^{0.5}$-consistent if the spectral density is sufficiently smooth near the zero frequency.

As in Andrews & Sun (2004) we could calculate the asymptotic bias which is of order $O((m/n)^{\varphi_y} + (m/n)^{2d_0+\varphi_w})$, see the proof of Lemma 1(e) in the appendix. This is in contrast to the orders $O((m/n)^2)$ and $O((m/n)^{2d_0})$ (assuming sufficient smoothness) for the LWN and LW estimators, respectively, see Hurvich & Ray (2003) and Arteche (2004). Thus, as in Andrews & Sun (2004) for the pure long memory case, the order of magnitude of the asymptotic bias is smaller when modeling the (smooth) spectral density of the short-memory component locally by a polynomial instead of a constant. Furthermore, a data-dependent adaptive procedure to select R_y, R_w, and m could be derived in the same manner as in Andrews & Sun (2004). Then R_y, R_w, and m would adapt to the smoothness of $\phi_y(\lambda)$ and $\phi_w(\lambda)$ and therefore depend on $s_a \in [s_a^l, s_a^u]$, where $1 \leq s_a^l \leq s_a^u < \infty$ for $a = y, w$.

Finally, with the asymptotic distribution in Theorem 2, it is possible to conduct inference also on $\boldsymbol{\theta}$. In particular, it is possible to test for the existence of the short-run components, possibly as a prior tool to select the most efficient estimator. However, note that the rate of convergence of the polynomial parameters is $\sqrt{m}\lambda_m^{2r}$ (for $\theta_{y,r}$) and $\sqrt{m}\lambda_m^{2d_0+2r}$ (for $\theta_{w,r}$) which can be quite slow.

4 Finite sample comparisons

In this section we present simulation results to examine the finite sample bias and root mean squared error (RMSE) of our LPWN estimator. The LPWN estimator is implemented with (R_y, R_w) equal to $(1,0)$, $(0,1)$, and $(1,1)$, denoted LPWN(Ry, R_w), and is compared with the LWN estimator. From Hurvich & Ray (2003) we know that the LWN estimator is superior to the LW estimator in terms of bias and RMSE in the context of the standard LMSV model. Furthermore, Hurvich et al. (2005) show that the polynomial log-periodogram regression estimator of Andrews & Guggenberger (2003) suffers from severe bias in the case of perturbed fractional processes and the LPW estimator is expected to perform similarly. Therefore, to conserve space we only present the results for the LWN and LPWN estimators. Results for the LW estimator are available from the authors upon request.

4.1 Monte Carlo setup

We simulate model (1), i.e.

$$z_t = y_t + w_t, \tag{18}$$

where $\{y_t\}$ is the signal process and $\{w_t\}$ is the perturbation process. We model $\{w_t\}$ as an ARMA process and $\{y_t\}$ as an ARFIMA process, and consider five different DGPs for these processes. The setup for $\{y_t\}$ and $\{w_t\}$ is

$$(1-\alpha_y L)(1-L)^d y_t = \left(1+\beta_y L\right)\varepsilon_t, \quad \varepsilon_t \sim NID(0,\sigma_\varepsilon^2), \tag{19}$$

$$(1-\alpha_w L) w_t = (1+\beta_w L)\eta_t, \quad \eta_t \sim NID(0,1), \tag{20}$$

with parameter configurations

$$\begin{aligned}
\text{Model I} &: (\alpha_y, \beta_y, \alpha_w, \beta_w) = (0, 0, 0, 0), \\
\text{Model II} &: (\alpha_y, \beta_y, \alpha_w, \beta_w) = (0.8, 0, 0, 0), \\
\text{Model III} &: (\alpha_y, \beta_y, \alpha_w, \beta_w) = (0, 0, 0, 0.8), \\
\text{Model IV} &: (\alpha_y, \beta_y, \alpha_w, \beta_w) = (0, 0, 0.8, 0), \\
\text{Model V} &: (\alpha_y, \beta_y, \alpha_w, \beta_w) = (0.8, 0, -0.8, 0).
\end{aligned}$$

We remark that in all the models the noise-to-signal ratio is given as

$$nsr = \frac{f_w(0)}{f_{(1-L)^d y_t}(0)} = \frac{\frac{(1+\beta_w)^2}{(1-\alpha_w)^2}}{\sigma_\varepsilon^2 \frac{(1+\beta_y)^2}{(1-\alpha_y)^2}}. \tag{21}$$

For each Monte Carlo DGP we generated 10,000 time series with a sample size of 2048, 4096, or 8192.[7] For all estimators we set the bandwidth as $m = \lfloor an^{0.8} \rfloor$, where $a \in \{3/4, 1, 5/4\}$. The parameter of interest, d, is set equal to 0.4. For the noise-to-signal ratio, we choose $nsr \in \{5, 10, 15\}$, and the variance σ_ε^2 is set as a function of $(\alpha_y, \beta_y, \alpha_w, \beta_w)$ such that the nsr has the desired value. The values of d, nsr, (α_y, β_y), (α_w, β_w), and the sample sizes are chosen to reflect empirical findings on long memory in volatility (see the references in the introduction for some examples). The chosen parameter values for the short-run contamination in the signal and the noise are also inspired by the results from the empirical (parametric) analysis of the DJIA stocks in section 5 below.

The signal $\{y_t\}$ was generated by the circulant embedding method as described in Davies & Harte (1987), see also Beran (1994, pp. 215-217). Numerical optimization was carried out in Ox using the SQP constrained optimization algorithm, see Doornik (2006). We used $D = [0.01, 0.99]$ and constrained the long-run noise-to-signal ratio θ_ρ to be in the interval $[10^{-6}, 10^6]$. The initial values were set as follows. For the LWN estimator we used $d = 0.25$ and $\theta_\rho = 1$. As starting value for the LPWN estimators we used the LWN estimate of (d, θ_ρ) if it was in the interior of the set $[0.01, 0.99] \times [0.01, 100]$, c.f. Assumption A2. Otherwise, the starting value of (d, θ_ρ) was set equal to $(0.25, 1)$.[8] As starting values for the polynomial parameters we used 0.

4.2 Monte Carlo results

Tables 1-5 display the results of the simulation study and show how the two different sources of bias, i.e. the additive noise term and the contamination from the short-memory dynamics in both the signal and the noise, affect the estimators.

[Table 1 about here]

In the case where there is no contamination by short-run dynamics in the signal or noise, i.e. Model I with results displayed in Table 1, the bias is small for all estimators. The theoretical inflation of the variances from $h(\lambda, d, \boldsymbol{\theta})$ is also noticeable in the RMSEs. Additionally, the RMSE decreases as either the sample size or bandwidth increase.

[7] The number of observations is chosen as a power of two in order to use the fast Fourier transform in calculating the periodogram. This speeds up the simulations considerably compared to using the discrete Fourier transform.

[8] We tried different starting values for d in these cases and the results were indistinguishable.

[Table 2 about here]

In Table 2 we consider model II, i.e. the signal is an ARFIMA process with $(\alpha_y, \beta_y) = (0.8, 0)$ and the noise is serially uncorrelated. Here we would presume that the LPWN(1,0) and LPWN(1,1) estimators are the better choices, at least for higher bandwidths. We see that we are able to obtain a reduction in bias relative to the LWN estimator. Overall, the two LPWN estimators modeling the short-run part of the signal with a polynomial, i.e. LPWN(1,0) and LPWN(1,1), outperform the LWN estimator in terms of bias, and for $nsr = 5$ the LPWN(1,0) estimator is mostly also superior in terms of RMSE. When nsr is 10 or higher, the noise component appears to be the most important term and modeling the signal with a polynomial does not improve the RMSE compared to the LWN estimator.

[Table 3 about here]

We consider next Model III, i.e. where there is MA contamination in the additive noise term, with results presented in Table 3. The results for the LPWN(0,1) and LPWN(1,1) estimators are similar to those in Table 1. That is, there is essentially no bias in those estimators. On the other hand, there is rather severe negative bias in the LWN estimator. Indeed, for the two higher bandwidth choices in Panels B and C the LPWN(0,1) (and sometimes the LPWN(1,1)) estimators generally have the lowest RMSE compared to the LWN.

[Table 4 about here]

Table 4 contains results for Model IV, where there is AR contamination in the additive noise term. The LWN estimator now suffers from moderate positive bias and the LPWN(0,1) estimator is, as expected, able to almost eliminate this bias and is nearly unbiased in most cases. Similarly to Table 3, the LPWN(0,1) estimator clearly outperforms the other estimators in terms of RMSE for the two higher bandwidth choices in Panels B and C.

[Table 5 about here]

Results for Model V where $(\alpha_y, \beta_y) = (0.8, 0)$ and $(\alpha_w, \beta_w) = (-0.8, 0)$ are shown in Table 5. The LWN estimator suffers from severe positive bias, and its RMSE is also higher than for the previous models. On the other hand, the LPWN estimators have relatively low biases, and in particular the LPWN(1,1) estimator appears essentially unbiased. When compared in terms of RMSE the LPWN estimators are clearly superior as well. Thus, we have a considerable reduction in bias for all LPWN estimators compared to the LWN estimator, and we also have quite a remarkable reduction in RMSE.

To sum up, the Monte Carlo study shows the usefulness of estimators that explicitly take the short-run dynamics in the perturbation into account, i.e. the LPWN estimators where $(R_y, R_w) = (0, 1)$ and $(R_y, R_w) = (1, 1)$, although the LPWN estimator with $(R_y, R_w) = (1, 0)$ also performs well. All three estimators generally have much smaller biases than the LWN estimator and are fairly robust to serial correlation in the perturbation and to contamination from short-memory dynamics in the signal.

5 Long memory in DJIA stock volatility

This section analyzes the long memory in daily log-squared returns series of the 30 DJIA stocks corrected for the effects of stock splits and dividends from 1 January 1990 to 31 March 2008, for a sample of $n = 4753$. To avoid the problem of taking logarithm of zero we based the analysis on adjusted log-squared returns using the method of Fuller (1996, pp. 495-496), i.e. we analyze

$$\log \tilde{r}_t^2 = \log \left(r_t^2 + \alpha \right) - \frac{\alpha}{r_t^2 + \alpha},$$

where $\alpha = \frac{0.02}{n} \sum_{t=1}^{n} r_t^2$. We estimate the long memory in $\log \tilde{r}_t^2$ using the proposed LPWN estimator. We implement the estimator with (R_y, R_w) equal to $(1, 0)$, $(0, 1)$, and $(1, 1)$, and with starting values as in the Monte Carlo study above. For comparison we also report standard LW, LPW, and LWN estimates. For all estimators we set the bandwidth as $m = \lfloor an^{0.8} \rfloor$, where $a \in \{3/4, 1, 5/4\}$.

[**Table 6 about here**]

Table 6 presents the results for the LW, LPW, and LWN estimators. As expected from theory, the LW and LPW estimators are decreasing in the bandwidth and appear downward biased. For the LWN estimator the memory estimates of some of the stocks are in the stationary region, but for the most part they are in the nonstationary region.

[**Table 7 about here**]

In Table 7 we present the results for the three variants of the LPWN estimator, i.e. for (R_y, R_w) equal to $(1, 0)$, $(0, 1)$, and $(1, 1)$. First of all, as expected from theory and the simulations above, it is clear that this estimator does not suffer from the downward bias that increases with bandwidth as is present in the LW and LPW estimators. Second, we note that the three different implementations of the estimator agree with each other for most stocks and bandwidth choices. Thirdly, the LPWN estimates are similar to the LWN estimates, although the LPWN(0,1) and LPWN(1,1) estimates are slightly higher on average.

To emphasize the importance of the polynomial approximation of the signal process $\{y_t\}$ and the pertubation process $\{w_t\}$, we also fitted an extended parametric LMSV-ARFIMA(1,d,1)-(1,1) model, where the extension is that the noise is modeled by an ARMA process. That is, we model the periodogram of $\log \tilde{r}_t^2$ using the Whittle likelihood framework of Fox & Taqqu (1986) and Breidt et al. (1998), where the fitted model has spectral density

$$f_z(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi} \left(2 \sin \frac{\lambda}{2} \right)^{-2d} \frac{\left(1 + 2\beta_y \cos \lambda + \beta_y^2\right)}{\left(1 - 2\alpha_y \cos \lambda + \alpha_y^2\right)} + \frac{\sigma_\eta^2}{2\pi} \frac{\left(1 + 2\beta_w \cos \lambda + \beta_w^2\right)}{\left(1 - 2\alpha_w \cos \lambda + \alpha_w^2\right)}. \tag{22}$$

In Table 8 the resulting estimates are reported, where we have removed insignificant ARMA terms from both the signal and the noise.

[**Insert Table 8 about here**]

The estimated values of d from the parametric results are in line with those from the LWN and LPWN estimators in Tables 6 and 7. Furthermore, there is significant (at 10% level) short-run dynamics in the signal (19 out of 30 cases), in the noise (16 out of 30 cases), and in both the signal and noise (13 out of 30 cases). The estimated (long-run) nsr can be calculated from the parameter estimates as in (21), and for most of the stocks it is in the vicinity of $10-30$, although there are cases where the estimated nsr is very high because the estimate of σ_ε^2 is very small. Taking the high nsr's and significant short-run dynamics in both the signal and the noise into consideration stresses the importance of the LPWN estimators.

6 Concluding remarks

In this paper we have proposed a semiparametric local polynomial Whittle with noise estimator of the degree of long memory, d, in fractionally integrated time series perturbed by additive short-run noise. The estimator models the log-spectrum of the the short-memory component of the signal and that of the perturbation by finite even polynomials instead of constants near the zero frequency. This is shown to yield a bias reduction depending on the smoothness of the spectra. However, including the polynomials inflates the asymptotic variance of $\hat{d}$ by a multiplicative constant which depends on the true long memory parameter, d.

We have shown that the estimator is consistent for $d \in (0,1)$, asymptotically normal for $d \in (0,3/4)$, and if the spectral density is sufficiently smooth near frequency zero the rate of convergence becomes arbitrary close to the parametric rate, $\sqrt{n}$.

Our Monte Carlo study shows that the proposed local polynomial Whittle with noise estimator is able to achieve considerable bias reductions in practice compared to standard (e.g., local Whittle with noise) estimators, especially in cases with short-run dynamics in both the signal and noise components. In an empirical investigation of the 30 DJIA stocks the local polynomial Whittle with noise estimator indicated stronger persistence in volatility than standard estimators, and for most of the stocks produced estimates of d in the nonstationary region.

Appendix A: Proof of Theorem 1

This proof follows the proofs of Theorem 3.1 and Lemma C.2 of Hurvich et al. (2005). As in their proofs, to show consistency of $\hat{d}$ we need to separately prove that $\lim_{n\to\infty} P(\hat{d} \in D_1) = 0$ and that $(\hat{d} - d_0)\mathbb{I}\{\hat{d} \in D_2\} \xrightarrow{P} 0$, where $D_1 = (-\infty, d_0 - 1/2 + \epsilon) \cap D$, $D_2 = [d_0 - 1/2 + \epsilon, +\infty) \cap D$, $\epsilon < 1/4$, and $\mathbb{I}\{A\}$ is the indicator function of the event A.

Let $\alpha_k(d,\boldsymbol{\theta}) = \frac{h_k(d_0,\boldsymbol{\theta}_0)}{h_k(d,\boldsymbol{\theta})}$. Then the proof that $(\hat{d} - d_0)\mathbb{I}\{\hat{d} \in D_2\} \xrightarrow{P} 0$ proceeds as in Hurvich et al. (2005, pp. 1303-1305) by showing that

$$Z_m = \sum_{k=1}^{m} \frac{k^{2(d-d_0)}\alpha_k(d,\boldsymbol{\theta})}{\sum_{j=1}^m j^{2(d-d_0)}\alpha_j(d,\boldsymbol{\theta})} \left(\frac{I_z(\lambda_k)}{f_z(\lambda_k)} - 1 \right) = o_P(1) \tag{23}$$

uniformly on $(d,\boldsymbol{\theta}) \in D_2 \times \Theta$ and that

$$R_m = \log\left(1 + \frac{\sum_{k=1}^m k^{2(d-d_0)}\left(\alpha_k(d,\boldsymbol{\theta}) - 1\right)}{\sum_{j=1}^m j^{2(d-d_0)}}\right) - \frac{1}{m}\sum_{k=1}^m \log\left(1 + \left(\alpha_k(d,\boldsymbol{\theta}) - 1\right)\right) = o(1) \tag{24}$$

uniformly on $(d,\boldsymbol{\theta}) \in D \times \Theta$.

Note that, from Lemma 2(iii)-(v), there exists a constant $C > 0$ such that

$$\sup_{(d,\boldsymbol{\theta})\in D\times\Theta, k=1,\ldots,m} \left|\alpha_k(d,\boldsymbol{\theta}) - 1\right| = \sup_{(d,\boldsymbol{\theta})\in D\times\Theta, k=1,\ldots,m} \left| \frac{h_k(d_0,\boldsymbol{\theta}_0) - h_k(d,\boldsymbol{\theta})}{h_k(d,\boldsymbol{\theta})} \right| \le C\,(m/n)^{2d_1},$$

since $d \geq d_1 > 0$. Now we use that $\log(1+x) = x + O(x^2)$ as $x \to 0$ to obtain

$$\sup_{(d,\boldsymbol{\theta}) \in D \times \Theta} |R_m| \leq C \sup_{(d,\boldsymbol{\theta}) \in D \times \Theta} \sup_{k=1,\ldots,m} |\alpha_k(d,\boldsymbol{\theta}) - 1| \leq C (m/n)^{2d_1} = o(1).$$

To show (23) we apply Proposition A.1 of Hurvich et al. (2005), which holds here since our Assumptions A1-A6 imply their Assumptions (H1)-(H3) with the exception that we allow serially correlated peturbation terms. It is, however, easily shown that replacing their Assumption (H2) with our Assumption A5, their Proposition A.1 still holds. The only other change is that the term $(k/n)^{\min(\beta,d_0)}$ in their eq. (F.15) should be replaced by $(k/n)^{\min(\varphi_y,\varphi_w)}$ due to the more accurate approximation of $f_z(\lambda)$ offered by the included polynomials in our function $h(d,\boldsymbol{\theta},\lambda)$ in (11), see also Lemma 3 below. Thus, from their Proposition A.1, letting

$$c_k = \frac{k^{2(d-d_0)}\alpha_k(d,\boldsymbol{\theta})}{\sum_{j=1}^m j^{2(d-d_0)}\alpha_j(d,\boldsymbol{\theta})},$$

then for $K \in (0,\infty)$ and all $k \in \{1,\ldots,m-1\}$, we need to show that $|c_k - c_{k+1}| \leq K m^{-\epsilon} k^{\epsilon-2}$ and $|c_m| \leq K m^{-1}$ uniformly on $(d,\boldsymbol{\theta}) \in D_2 \times \Theta$, which implies (23).

Note that, uniformly on $(d,\boldsymbol{\theta}) \in D_2 \times \Theta$, we have that $\sum_{j=1}^m j^{2(d-d_0)}\alpha_j(d,\boldsymbol{\theta}) \geq C m^{2(d-d_0)+1}$ and $|k^{2(d-d_0)}\alpha_k(d,\boldsymbol{\theta}) - (k+1)^{2(d-d_0)}\alpha_{k+1}(d,\boldsymbol{\theta})| \leq C k^{2(d-d_0)-1}$ using the mean value theorem and Lemma 2, see also Hurvich et al. (2005, p. 1305). It follows that

$$\begin{aligned}
\sup_{(d,\boldsymbol{\theta}) \in D_2 \times \Theta} \left| \frac{k^{2(d-d_0)}\alpha_k(d,\boldsymbol{\theta}) - (k+1)^{2(d-d_0)}\alpha_{k+1}(d,\boldsymbol{\theta})}{\sum_{j=1}^m j^{2(d-d_0)}\alpha_j(d,\boldsymbol{\theta})} \right| &\leq \sup_{(d,\boldsymbol{\theta}) \in D_2 \times \Theta} C \left| \frac{k^{2(d-d_0)-1}}{m^{2(d-d_0)+1}} \right| \leq C k^{2\epsilon-2} m^{-2\epsilon}, \\
\sup_{(d,\boldsymbol{\theta}) \in D_2 \times \Theta} \left| \frac{m^{2(d-d_0)}\alpha_m(d,\boldsymbol{\theta})}{\sum_{j=1}^m j^{2(d-d_0)}\alpha_j(d,\boldsymbol{\theta})} \right| &\leq C m^{-1},
\end{aligned}$$

which proves (23).

The proof that $\lim_{n\to\infty} P(\hat{d} \in D_1) = 0$ follows exactly as in Hurvich et al. (2005, pp. 1305-1306) since their Proposition A.1 holds in our case as well. Hence $\hat{d} \xrightarrow{P} d_0$. To strengthen this result to $\hat{d} - d_0 = o_P((\log n)^{-5})$ we use the proof of Lemma C.2 of Hurvich et al. (2005) without change.

Appendix B: Proof of Theorem 2

For the proof of Theorem 2 we need the score and Hessian (both multiplied by m) of (12):

$$\begin{aligned}
\mathbf{S}_n(d,\boldsymbol{\theta}) &= \hat{G}(d,\boldsymbol{\theta})^{-1} \sum_{j=1}^m \left(\frac{\lambda_j^{2d} I_z(\lambda_j)}{h_j(d,\boldsymbol{\theta})} - \frac{1}{m} \sum_{k=1}^m \frac{\lambda_k^{2d} I_z(\lambda_k)}{h_k(d,\boldsymbol{\theta})} \right) \mathbf{X}_j, \\
\mathbf{H}_n(d,\boldsymbol{\theta}) &= \mathbf{H}_{1n}(d,\boldsymbol{\theta}) + \mathbf{H}_{2n}(d,\boldsymbol{\theta}), \\
\mathbf{H}_{1n}(d,\boldsymbol{\theta}) &= \hat{G}(d,\boldsymbol{\theta})^{-2} \left(\hat{G}(d,\boldsymbol{\theta}) \sum_{j=1}^m \frac{\lambda_j^{2d} I_z(\lambda_j)}{h_j(d,\boldsymbol{\theta})} \mathbf{X}_j \mathbf{X}_j' - m \left(\frac{1}{m} \sum_{j=1}^m \frac{\lambda_j^{2d} I_z(\lambda_j)}{h_j(d,\boldsymbol{\theta})} \mathbf{X}_j \right) \left(\frac{1}{m} \sum_{j=1}^m \frac{\lambda_j^{2d} I_z(\lambda_j)}{h_j(d,\boldsymbol{\theta})} \mathbf{X}_j \right)' \right), \\
\mathbf{H}_{2n}(d,\boldsymbol{\theta}) &= \hat{G}(d,\boldsymbol{\theta})^{-1} \sum_{j=1}^m \left(\frac{\lambda_j^{2d} I_z(\lambda_j)}{h_j(d,\boldsymbol{\theta})} - \frac{1}{m} \sum_{k=1}^m \frac{\lambda_k^{2d} I_z(\lambda_k)}{h_k(d,\boldsymbol{\theta})} \right) \frac{\partial \mathbf{X}_j}{\partial(d,\boldsymbol{\theta}')},
\end{aligned}$$

where $\mathbf{X}_j = (X_{1j}, \mathbf{X}'_{2j}, \mathbf{X}'_{3j})'$ with

$$X_{1j} = \frac{2\exp(h_{yj}(\boldsymbol{\theta}_y))\log\lambda_j}{h_j(d,\boldsymbol{\theta})}, \mathbf{X}_{2j} = \left(\frac{-\lambda_j^2\exp(h_{yj}(\boldsymbol{\theta}_y))}{h_j(d,\boldsymbol{\theta})}, \ldots, \frac{-\lambda_j^{2R_y}\exp(h_{yj}(\boldsymbol{\theta}_y))}{h_j(d,\boldsymbol{\theta})}\right)',$$

$$\mathbf{X}_{3j} = \left(\frac{-\lambda_j^{2d}\exp(h_{wj}(\boldsymbol{\theta}_w))}{h_j(d,\boldsymbol{\theta})}, \frac{-\lambda_j^{2d+2}\theta_\rho\exp(h_{wj}(\boldsymbol{\theta}_w))}{h_j(d,\boldsymbol{\theta})}, \ldots, \frac{-\lambda_j^{2d+2R_w}\theta_\rho\exp(h_{wj}(\boldsymbol{\theta}_w))}{h_j(d,\boldsymbol{\theta})}\right)',$$

$h_j(d,\boldsymbol{\theta}) = h(d,\boldsymbol{\theta},\lambda_j)$, $h_{aj}(\boldsymbol{\theta}_a) = h_a(\boldsymbol{\theta}_a,\lambda_j), a = y, w$, and $D_n(\eta) = \{d \in D : (\log n)^5 |d - d_0| < \eta\}$ for $\eta > 0$. Note that $\mathbf{X}_j$ is the vector of partial derivatives of $-\log(\lambda_j^{-2d}h_j(d,\boldsymbol{\theta}))$. Also define

$$\mathbf{J}_n = \sum_{j=1}^m \left(\mathbf{X}_j - \frac{1}{m}\sum_{k=1}^m \mathbf{X}_k\right)\left(\mathbf{X}_j - \frac{1}{m}\sum_{k=1}^m \mathbf{X}_k\right)'.$$

We next state a lemma adapted from Lemma 2 of Andrews & Sun (2004), henceforth abbreviated AS. The proof is given in the next section.

Lemma 1 *Under the assumptions of Theorem 2 we have, as $n \to \infty$,*

(a) $\mathbf{B}_n^{-1}\mathbf{J}_n\mathbf{B}_n^{-1} \to \boldsymbol{\Omega}_{R_y,R_w}$,

(b) $\left\|\mathbf{B}_n^{-1}(\mathbf{H}_{1n}(d_0,\boldsymbol{\theta}_0) - \mathbf{J}_n)\mathbf{B}_n^{-1}\right\| = o_P(1)$ *and* $\left\|\mathbf{B}_n^{-1}\mathbf{H}_{2n}(d_0,\boldsymbol{\theta}_0)\mathbf{B}_n^{-1}\right\| = o_P(1)$,

(c) $\sup_{\boldsymbol{\theta}\in\Theta}\left\|\mathbf{B}_n^{-1}(\mathbf{H}_{kn}(d_0,\boldsymbol{\theta}) - \mathbf{H}_{kn}(d_0,\boldsymbol{\theta}_0))\mathbf{B}_n^{-1}\right\| = o_P(1)$, $k = 1, 2$,

(d) $\sup_{d\in D_n(\eta_n),\boldsymbol{\theta}\in\Theta}\left\|\mathbf{B}_n^{-1}(\mathbf{H}_{kn}(d,\boldsymbol{\theta}) - \mathbf{H}_{kn}(d_0,\boldsymbol{\theta}))\mathbf{B}_n^{-1}\right\| = o_P(1)$, $k = 1, 2$, *for all sequences of constants $\{\eta_n\}_{n\geq 1}$ for which $\eta_n = o(1)$,*

(e) $\mathbf{B}_n^{-1}\mathbf{S}_n(d_0,\boldsymbol{\theta}_0) \xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{\Omega}_{R_y,R_w}\right)$.

Since the LPWN likelihood (12) is a continuous function on a compact set the LPWN estimator exists. From Lemma 1 we know by Lemma 1 of AS that there exists a solution to the first order conditions with probability tending to one, and that the solution satisfies the result in Theorem 2. If the (negative) likelihood function is strictly convex and twice differentiable then the solution to the first order conditions is unique and minimizes (12) and hence equals the LPWN estimator.

Thus, all that remains is to show that the Hessian is positive definite which proves convexity. The positive definiteness of $\mathbf{H}_{1n}$ follows as in eq. (5.1) of AS. Compared to AS we have the additional term $\mathbf{H}_{2n}$, for which we know that $\left\|\mathbf{B}_n^{-1}\mathbf{H}_{2n}(d,\boldsymbol{\theta})\mathbf{B}_n^{-1}\right\| = o_P(1)$ uniformly on $(d,\boldsymbol{\theta}) \in D_n(\eta_n) \times \Theta$ by Lemma 1(b)-(d) and the triangle inequality. Since $\hat{d} \in D_n(\eta_n)$ with probability tending to one by Theorem 1, this shows that $\mathbf{H}_n$ is positive definite with probability tending to one, which concludes the proof.

Appendix C: Proof of Lemma 1

We now turn to the proof of Lemma 1, which follows the method of proof for Lemma 2 of AS, with modifications to allow $d \geq 1/2$ (following Velasco (1999)) and to accommodate the additive noise term in the spectral density (see Lemma 3), and with an additional proof for each of (b), (c), and (d) of negligibility of the term $\mathbf{H}_{2n}$.

C.1 Proof of (a)

This follows by approximating sums by integrals, see Lemma 2 of Andrews & Guggenberger (2003).

C.2 Proof of (b), first statement

This proof roughly follows that of Lemma 2(b) in AS. Corresponding to their (A.6), define

$$\begin{aligned}\tilde{G}_{a,b,c,c_1,c_2}(d,\boldsymbol{\theta}) &= m^{-1}\sum_{j=1}^{m}\frac{\lambda_j^{2d}I_z(\lambda_j)}{h_j(d,\boldsymbol{\theta})^{c_1+c_2+1}}\left(\frac{2\exp(h_{yj}(\boldsymbol{\theta}_y))\log\lambda_j}{h_j\,(d,\boldsymbol{\theta})}\right)^a (j/m)^{2b}\,\theta_\rho^c \\ &\quad \times\exp(c_1 h_{yj}(\boldsymbol{\theta}_y)+c_2 h_{wj}(\boldsymbol{\theta}_w)), \\ \hat{G}_{a,b,c}(d,\boldsymbol{\theta}) &= m^{-1}\sum_{j=1}^{m}\frac{\lambda_j^{2d}I_z(\lambda_j)}{h_j(d,\boldsymbol{\theta})}\,(2\log\lambda_j)^a\,(j/m)^{2b}\,\theta_\rho^c, \\ J_{a,b,c} &= G_0 m^{-1}\sum_{j=1}^{m}(2\log\lambda_j)^a\,(j/m)^{2b}\,\theta_{0\rho}^c,\end{aligned}$$

for $a,c,c_1,c_2=0,1,2$ and $b=0,1,\ldots,2R_y,d,d+1,\ldots,d+R_w+R_y,2d,2d+1,\ldots,2d+2R_w$. The elements of $\mathbf{B}_n^{-1}\mathbf{H}_{1n}\,(d,\boldsymbol{\theta})\,\mathbf{B}_n^{-1}$ are (omitting the arguments for brevity)

$$\begin{aligned}(1,1) &: \tilde{G}_{0,0,0,0,0}^{-2}\left(\tilde{G}_{0,0,0,0,0}\tilde{G}_{2,0,0,0,0}-\tilde{G}_{1,0,0,0,0}^2\right), \\ (1,1+k) &: \tilde{G}_{0,0,0,0,0}^{-2}\left(\tilde{G}_{0,0,0,0,0}\tilde{G}_{1,k,0,1,0}-\tilde{G}_{1,0,0,0,0}\tilde{G}_{0,k,0,1,0}\right), \\ (1,2+R_y+i) &: \tilde{G}_{0,0,0,0,0}^{-2}\left(\tilde{G}_{0,0,0,0,0}\tilde{G}_{1,i+d,\mathbb{I}\{i\geq1\},0,1}-\tilde{G}_{1,0,0,0,0}\tilde{G}_{0,i+d,\mathbb{I}\{i\geq1\},0,1}\right), \\ (1+k_1,1+k_2) &: \tilde{G}_{0,0,0,0,0}^{-2}\left(\tilde{G}_{0,0,0,0,0}\tilde{G}_{0,k_1+k_2,0,2,0}-\tilde{G}_{0,k_1,0,1,0}\tilde{G}_{0,k_2,0,1,0}\right), \\ (1+k,2+R_y+i) &: \tilde{G}_{0,0,0,0,0}^{-2}\left(\tilde{G}_{0,0,0,0,0}\tilde{G}_{0,k+i+d,\mathbb{I}\{i\geq1\},1,1}-\tilde{G}_{0,k,0,1,0}\tilde{G}_{0,i+d,\mathbb{I}\{i\geq1\},0,1}\right), \\ (2+R_y+i_1,2+R_y+i_2) &: \tilde{G}_{0,0,0,0,0}^{-2}\left(\tilde{G}_{0,0,0,0,0}\tilde{G}_{0,i_1+i_2+2d,\mathbb{I}\{i_1\geq1\}+\mathbb{I}\{i_2\geq1\},0,2}-\tilde{G}_{0,i_1+d,\mathbb{I}\{i_1\geq1\},0,1}\tilde{G}_{0,i_2+d,\mathbb{I}\{i_2\geq1\},0,1}\right),\end{aligned}$$

for $k,k_1,k_2=1,\ldots,R_y$ and $i,i_1,i_2=0,\ldots,R_w$. The corresponding elements of $\mathbf{B}_n^{-1}\mathbf{J}_n\mathbf{B}_n^{-1}$ are given by the same expressions with $\tilde{G}_{a,b,c,c_1,c_2}$ replaced by $J_{a,b,c}$. To prove the first statement of Lemma 1(b) it suffices to show that (since b is a function of d, we distinguish between b and b_0)

$$\Delta_{a,b_0,c} = \left|\hat{G}_{a,b_0,c}(d_0,\boldsymbol{\theta}_0)-J_{a,b_0,c}\right| = o_P((\log m)^{-2}), \tag{25}$$

$$\tilde{\Delta}_{a,b_0,c,c_1,c_2} = \left|\tilde{G}_{a,b_0,c,c_1,c_2}(d_0,\boldsymbol{\theta}_0)-\hat{G}_{a,b_0,c}(d_0,\boldsymbol{\theta}_0)\right| = o_P((\log n)^2(m/n)^{2d_0}). \tag{26}$$

Note that, because $\mathbf{B}_n^{-1}\mathbf{J}_n\mathbf{B}_n^{-1}=O((\log m)^2)$, $\tilde{\Delta}_{a,b_0,c,c_1,c_2}=o_P((\log m)^{-2})$ would be sufficient to prove part (b) for $\mathbf{H}_{1n}$, but we show the stronger version since it will be useful in the proof of part (c).

In view of Lemma 3 below, the proof of (A.9) in AS pp. 598-599 works also for our eq. (25), where we find that (with $\xi_{k,n}(d)$ defined in Lemma 3)

$$\begin{aligned}\Delta_{a,b_0,c} = O_P\Big(&(\log m)^a m^{-1}\xi_{m,n}(d_0)+(\log m)^a m^{\varphi_y}n^{-\varphi_y}+(\log m)^a m^{d_0+\varphi_w}n^{-d_0-\varphi_w} \\ &+(\log m)^a m^{2d_0}n^{-2d_0}+(\log m)^{a+1}m^{2d_0-1}n^{-d_0}+(\log m)^a m^{-1/2}\Big),\end{aligned}$$

which is

$$\begin{aligned}O_P\Big(&(\log m)^{a+2/3}m^{-2/3}+(\log m)^a m^{-1/2}n^{-1/4}+(\log m)^a(m/n)^{\min(\varphi_y,d_0+\varphi_w,2d_0)} \\ &+(\log m)^{a+1}m^{2d_0-1}n^{-d_0}+(\log m)^a m^{-1/2}\Big)\end{aligned}$$

in the stationary case and

$$O_P\left((\log m)^{a+2/(5-4d_0)}m^{1/(5-4d_0)-1}+(\log m)^{a+1}m^{2d_0-2}+(\log m)^a m^{(d_0-1)/2}n^{-1/2}(\log n)^{5/4}\right.$$
$$\left.+(\log m)^{a+1/2}n^{-1/4}m^{d_0-1}+(\log m)^a(m/n)^{\min(\varphi_y,d_0+\varphi_w,2d_0)}+(\log m)^{a+1}m^{2d_0-1}n^{-d_0}+(\log m)^a m^{-1/2}\right)$$

in the nonstationary case. By (17) and $d_0<d_2<3/4$, clearly $\Delta_{a,b_0,c}=o_P((\log m)^{-2})$ in both cases.

To prove (26) we write $\tilde{G}_{a,b_0,c,c_1,c_2}(d_0,\boldsymbol{\theta}_0)-\hat{G}_{a,b_0,c}(d_0,\boldsymbol{\theta}_0)$ as

$$\frac{\theta_{0\rho}^c}{m}\sum_{j=1}^m\frac{\lambda_j^{2d_0}I_z(\lambda_j)}{h_j(d_0,\boldsymbol{\theta}_0)}\left[\frac{\exp(c_1h_{yj}(\boldsymbol{\theta}_{0y})+c_2h_{wj}(\boldsymbol{\theta}_{0w}))}{h_j(d_0,\boldsymbol{\theta}_0)^{c_1+c_2}}\left(\frac{2\exp(h_{yj}(\boldsymbol{\theta}_{0y}))\log\lambda_j}{h_j\left(d_0,\boldsymbol{\theta}_0\right)}\right)^a-(2\log\lambda_j)^a\right]\left(\frac{j}{m}\right)^{2b_0}$$
$$=\frac{\theta_{0\rho}^c}{m}\sum_{j=1}^m\frac{\lambda_j^{2d_0}I_z(\lambda_j)}{h_j(d_0,\boldsymbol{\theta}_0)}\left[\frac{(1+O((j/n)^{2d_0}))(1+O((j/n)^2))^a}{(1+O((j/n)^{2d_0}))^{c_1+c_2+a}}(2\log\lambda_j)^a-(2\log\lambda_j)^a\right]\left(\frac{j}{m}\right)^{2b_0}$$

by Lemma 2(i)-(ii). Using the definition of $\hat{G}_{a,b,c}(d,\boldsymbol{\theta})$ and the fact that $\hat{G}_{a,b_0,c}(d_0,\boldsymbol{\theta}_0)=O_P((\log m)^a)$ by (25), we have

$$\begin{aligned}\tilde{\Delta}_{a,b,c,c_1,c_2} &= \frac{\theta_{0\rho}^c}{m}\sum_{j=1}^m\frac{\lambda_j^{2d_0}I_z(\lambda_j)}{h_j(d_0,\boldsymbol{\theta}_0)}\left[(1+O((j/n)^{2d_0}))\,(2\log\lambda_j)^a-(2\log\lambda_j)^a\right]\left(\frac{j}{m}\right)^{2b_0}\\ &= O_P((m/n)^{2d_0}(\log n)^a\hat{G}_{0,b_0,c}(d_0,\boldsymbol{\theta}_0))=O_P\left((m/n)^{2d_0}(\log n)^a\right).\end{aligned}$$

C.3 Proof of (e)

We now prove part (e) since it will be useful in the proof of the remaining statements. By (25) and (26) with $a=b=c=c_1=c_2=0$ we get that $\hat{G}\left(d_0,\boldsymbol{\theta}_0\right)\xrightarrow{P}G_0$, so that, apart from smaller order terms,

$$\begin{aligned}\mathbf{B}_n^{-1}\mathbf{S}_n\left(d_0,\boldsymbol{\theta}_0\right) &= m^{-1/2}\sum_{j=1}^m\left(\frac{I_z\left(\lambda_j\right)}{g_j\left(d_0,\boldsymbol{\theta}_0\right)}-\frac{1}{m}\sum_{k=1}^m\frac{I_z\left(\lambda_k\right)}{g_k\left(d_0,\boldsymbol{\theta}_0\right)}\right)\tilde{\mathbf{X}}_{0,j}\\ &= m^{-1/2}\sum_{j=1}^m\left(\frac{I_z\left(\lambda_j\right)}{g_j\left(d_0,\boldsymbol{\theta}_0\right)}-1\right)\left(\tilde{\mathbf{X}}_{0,j}-\frac{1}{m}\sum_{k=1}^m\tilde{\mathbf{X}}_{0,k}\right),\end{aligned}\qquad(27)$$

where $g_j\left(d,\boldsymbol{\theta}\right)=g\left(d,\boldsymbol{\theta},\lambda_j\right)$ and $\tilde{\mathbf{X}}_j=(X_{1,j},\tilde{\mathbf{X}}_{2,j}',\tilde{\mathbf{X}}_{3,j}')'$ with

$$(\tilde{\mathbf{X}}_{2,j})_k=-\frac{\exp(h_{yj}(\boldsymbol{\theta}_y))(j/m)^{2k}}{h_j\left(d,\boldsymbol{\theta}\right)},(\tilde{\mathbf{X}}_{3,j})_{i+1}=-\frac{\theta_\rho^{\mathbb{I}\{i\geq1\}}\exp(h_{wj}(\boldsymbol{\theta}_w))(j/m)^{2d+2i}}{h_j\left(d,\boldsymbol{\theta}\right)}$$

for $k=1,\ldots,R_y$ and $i=0,\ldots,R_w$, and $\tilde{\mathbf{X}}_{0,j}$ is $\tilde{\mathbf{X}}_j$ evaluated at $(d_0,\boldsymbol{\theta}_0)$.

As in AS p. 601 we write the right-hand side of (27) as $T_{1,n}+T_{2,n}+T_{3,n}+T_{4,n}$, where

$$\begin{aligned}T_{1,n} &= m^{-1/2}\sum_{j=1}^m\left(\frac{I_z\left(\lambda_j\right)}{g_j\left(d_0,\boldsymbol{\theta}_0\right)}-2\pi I_\varepsilon\left(\lambda_j\right)-E\left(\frac{I_z\left(\lambda_j\right)}{g_j\left(d_0,\boldsymbol{\theta}_0\right)}-2\pi I_\varepsilon\left(\lambda_j\right)\right)\right)\left(\tilde{\mathbf{X}}_{0,j}-\frac{1}{m}\sum_{k=1}^m\tilde{\mathbf{X}}_{0,k}\right),\\ T_{2,n} &= m^{-1/2}\sum_{j=1}^m\left(\frac{EI_z\left(\lambda_j\right)}{f_z\left(\lambda_j\right)}-1\right)\frac{f_z\left(\lambda_j\right)}{g_j\left(d_0,\boldsymbol{\theta}_0\right)}\left(\tilde{\mathbf{X}}_{0,j}-\frac{1}{m}\sum_{k=1}^m\tilde{\mathbf{X}}_{0,k}\right),\end{aligned}$$

$$T_{3,n} = m^{-1/2}\sum_{j=1}^{m}\left(2\pi I_{\varepsilon}(\lambda_j)-1\right)\left(\tilde{\mathbf{X}}_{0,j}-\frac{1}{m}\sum_{k=1}^{m}\tilde{\mathbf{X}}_{0,k}\right),$$
$$T_{4,n} = m^{-1/2}\sum_{j=1}^{m}\left(\frac{f_z(\lambda_j)}{g_j(d_0,\boldsymbol{\theta}_0)}-1\right)\left(\tilde{\mathbf{X}}_{0,j}-\frac{1}{m}\sum_{k=1}^{m}\tilde{\mathbf{X}}_{0,k}\right).$$

Then we show that $T_{3,n}\xrightarrow{d}N\left(\mathbf{0},\boldsymbol{\Omega}_{R_y,R_w}\right)$ while $T_{i,n}=o_P(1)$ for $i=1,2,4$.

Clearly the proof for $T_{3,n}$ of AS works here as well. We just have to verify that $m^{-1}\sum_{j=1}^{m}\zeta_j^2\to\boldsymbol{\beta}'\boldsymbol{\Omega}_{R_y,R_w}\boldsymbol{\beta}$, where

$$\zeta_j=\boldsymbol{\beta}'(\tilde{\mathbf{X}}_{0,j}-\frac{1}{m}\sum_{k=1}^{m}\tilde{\mathbf{X}}_{0,k}) \text{ and } \boldsymbol{\Omega}_{R_y,R_w}=\begin{pmatrix}4 & \boldsymbol{\mu}'_{R_y} & \boldsymbol{\nu}'_{R_w}\\ \boldsymbol{\mu}_{R_y} & \boldsymbol{\Gamma}_{R_y} & \boldsymbol{\psi}'_{R_y,R_w}\\ \boldsymbol{\nu}_{R_w} & \boldsymbol{\psi}_{R_w,R_y} & \boldsymbol{\Psi}_{R_w}\end{pmatrix},$$

which follows from part (a) of the lemma.

To show the result for $T_{1,n}$ we use summation by parts:

$$\begin{aligned}
T_{1,n} &= m^{-1/2}\sum_{k=1}^{m-1}\left(\tilde{\mathbf{X}}_{0,k}-\tilde{\mathbf{X}}_{0,k+1}\right)\sum_{j=1}^{k}\left(\frac{I_z(\lambda_j)}{g_j(d_0,\boldsymbol{\theta}_0)}-2\pi I_{\varepsilon}(\lambda_j)-E\left(\frac{I_z(\lambda_j)}{g_j(d_0,\boldsymbol{\theta}_0)}-2\pi I_{\varepsilon}(\lambda_j)\right)\right)\\
&\quad+\left(\tilde{\mathbf{X}}_{0,m}-\frac{1}{m}\sum_{k=1}^{m}\tilde{\mathbf{X}}_{0,k}\right)m^{-1/2}\sum_{j=1}^{m}\left(\frac{I_z(\lambda_j)}{g_j(d_0,\boldsymbol{\theta}_0)}-2\pi I_{\varepsilon}(\lambda_j)-E\left(\frac{I_z(\lambda_j)}{g_j(d_0,\boldsymbol{\theta}_0)}-2\pi I_{\varepsilon}(\lambda_j)\right)\right)\\
&= m^{-1/2}\sum_{k=1}^{m-1}O(k^{-1})O_P(\xi_{k,n}(d_0)+k^{\varphi_y+1/2}n^{-\varphi_y}+k^{1/2+2d_0}n^{-2d_0})\\
&\quad+O(1)m^{-1/2}O_P(\xi_{m,n}(d_0)+m^{\varphi_y+1/2}n^{-\varphi_y}+m^{1/2+2d_0}n^{-2d_0})\\
&= O_P(m^{-1/2}(\log m)\xi_{m,n}(d_0)+(m/n)^{\min(\varphi_y,2d_0)}),
\end{aligned}$$

where $\xi_{k,n}(d)$ is defined in Lemma 3. The second equality applies Lemma 3 and that $\tilde{\mathbf{X}}_{0,k}-\tilde{\mathbf{X}}_{0,k+1}=O(k^{-1})$ uniformly in $k=1,\ldots,m$ and $\tilde{\mathbf{X}}_{0,m}-\frac{1}{m}\sum_{k=1}^{m}\tilde{\mathbf{X}}_{0,k}=O(1)$, follows by approximating sums by integrals, see also AS p. 602. Thus $T_{1,n}=O_P((\log m)^{5/3}m^{-1/6}+(\log m)n^{-1/4}+(m/n)^{\min(\varphi_y,2d_0)})$ in the stationary case and $T_{1,n}=O_P((\log m)^{1+2/(5-4d_0)}m^{-(3-4d_0)/(10-8d_0)}+(\log m)^2m^{2d_0-3/2}+(\log m)(\log n)^{5/4}n^{-1/2}m^{d_0/2}+(\log m)^{3/2}n^{-1/4}m^{d_0-1/2}+(m/n)^{\min(\varphi_y,2d_0)})$ in the nonstationary case. Since d_0 belongs to the interior of the parameter space it follows that $T_{1,n}=o_P(1)$.

To prove the result for $T_{2,n}$ we use Robinson's (1995*b*) Theorem 2, i.e., that $EI_y(\lambda_j)/f_y(\lambda_j)=1+O(j^{-1}(\log j))$ uniformly in $j=1,\ldots,m$ in the stationary case, as well as Velasco's (1999) Theorem 1, which shows that $EI_y(\lambda_j)/f_y(\lambda_j)=1+O(j^{2d_0-2}(\log j))$ uniformly in $j=1,\ldots,m$ in the nonstationary case. Note that, as in AS, the remainder terms are different from those of Robinson (1995*b*) and Velasco (1999) because of the normalization by $f_y(\lambda_j)$ rather than by $G_0\lambda_j^{-2d_0}$. Thus, as in the proof of Lemma 3 we can write

$$\begin{aligned}
\frac{EI_z(\lambda_j)}{f_z(\lambda_j)}-1 &= \frac{f_y(\lambda_j)-f_z(\lambda_j)}{f_z(\lambda_j)}\left(\frac{EI_y(\lambda_j)}{f_y(\lambda_j)}-1\right)+\left(\frac{EI_y(\lambda_j)}{f_y(\lambda_j)}-1\right)\\
&\quad+\frac{2\sqrt{f_y(\lambda_j)}}{f_z(\lambda_j)}\frac{E\,\mathrm{Re}\left(I_{yw}(\lambda_j)\right)}{\sqrt{f_y(\lambda_j)}}+\frac{EI_w(\lambda_j)+f_y(\lambda_j)-f_z(\lambda_j)}{f_z(\lambda_j)}.
\end{aligned}$$

Because $EI_w(\lambda_j)=f_w(\lambda_j)+O(j^{-1}(\log j))$ and $f_z(\lambda_j)-f_y(\lambda_j)=f_w(\lambda_j)$, the last term is

$O(j^{-1}(\log j)\lambda_j^{2d_0})$. By the same reasoning and by independence of $\{y_t\}$ and $\{w_t\}$, the second to last term is $O_P(\lambda_j^{d_0}j^{-1}(\log j))$ in the stationary case and $O_P(\lambda_j^{d_0}j^{2d_0-2}(\log j))$ in the nonstationary case, see also the proof of Lemma 3 below and the second to last equation on p. 108 of Velasco (1999). We thus obtain the bounds $EI_z(\lambda_j)/f_z(\lambda_j)-1=O(j^{-1}(\log j))$ for the stationary case and $EI_z(\lambda_j)/f_z(\lambda_j)-1=O(j^{2d_0-2}(\log j))$ for the nonstationary case, for all $j=1,\ldots,m$. We also have that $f_z(\lambda_j)/g_j(d_0,\boldsymbol{\theta}_0)-1=O((j/n)^{\varphi_y}+(j/n)^{2d_0+\varphi_w})$ for all $j=1,\ldots,m$ by (15). Therefore, in the stationary case, $T_{2,n}$ can be bounded similarly to (A.24) of AS,

$$T_{2,n}=m^{-1/2}\sum_{j=1}^{m}O(j^{-1}(\log j))O(\log m)=O((\log m)^3m^{-1/2}),$$

using also that $|\tilde{\mathbf{X}}_{0,j}-\frac{1}{m}\sum_{k=1}^m\tilde{\mathbf{X}}_{0,k}|=O(\log m)$ uniformly in $j=1,\ldots,m$. In the nonstationary case we find in the same way that

$$T_{2,n}=m^{-1/2}\sum_{j=1}^{m}O(j^{2d_0-2}(\log j))O(\log m)=O((\log m)^3m^{2d_0-3/2}).$$

In both the stationary and nonstationary cases, $T_{2,n}$ is $o(1)$ since $d_0<d_2<3/4$.

The proof for $T_{4,n}$ follows from summation by parts and the approximation $f_z(\lambda_j)/g_j(d_0,\boldsymbol{\theta}_0)-1=O((j/n)^{\varphi_y}+(j/n)^{2d_0+\varphi_w})$ for all $j=1,\ldots,m$, which implies that

$$\begin{aligned}T_{4,n} &= \frac{1}{\sqrt{m}}\sum_{k=1}^{m-1}\left(\tilde{\mathbf{X}}_{0,k}-\tilde{\mathbf{X}}_{0,k+1}\right)\sum_{j=1}^{k}\left(\frac{f_z(\lambda_j)}{g_j(d_0,\boldsymbol{\theta}_0)}-1\right)+\left(\tilde{\mathbf{X}}_{0,m}-\frac{1}{m}\sum_{k=1}^{m}\tilde{\mathbf{X}}_{0,k}\right)\frac{1}{\sqrt{m}}\sum_{j=1}^{m}\left(\frac{f_z(\lambda_j)}{g_j(d_0,\boldsymbol{\theta}_0)}-1\right)\\ &= \frac{1}{\sqrt{m}}\sum_{k=1}^{m-1}O(k^{-1})\sum_{j=1}^{k}O((j/n)^{\varphi_y}+(j/n)^{2d_0+\varphi_w})+O(1)\frac{1}{\sqrt{m}}\sum_{j=1}^{m}O((j/n)^{\varphi_y}+(j/n)^{2d_0+\varphi_w})\\ &= O(m^{1/2+\varphi_y}n^{-\varphi_y}+m^{1/2+2d_0+\varphi_w}n^{-2d_0-\varphi_w}).\end{aligned}$$

Condition (17) shows that this is $o_P(1)$.

C.4 Proof of (b), second statement

The matrix $\mathbf{B}_n^{-1}\mathbf{H}_{2n}(d,\boldsymbol{\theta})\mathbf{B}_n^{-1}$ is symmetric and has (i,l)'th and (l,i)'th elements

$$-\hat{G}(d,\boldsymbol{\theta})^{-1}\frac{1}{m}\sum_{j=1}^{m}\left(\frac{\lambda_j^{2d}I_z(\lambda_j)}{h_j(d,\boldsymbol{\theta})}-\frac{1}{m}\sum_{k=1}^{m}\frac{\lambda_k^{2d}I_z(\lambda_k)}{h_k(d,\boldsymbol{\theta})}\right)(\tilde{\mathbf{X}}_j)_i\frac{\lambda_j^{2d}\exp(h_{wj}(\boldsymbol{\theta}_w))(2\log\lambda_j)}{h_j(d,\boldsymbol{\theta})},\ i=1,\ldots,R_y+1,l=1,$$

$$\hat{G}(d,\boldsymbol{\theta})^{-1}\frac{1}{m}\sum_{j=1}^{m}\left(\frac{\lambda_j^{2d}I_z(\lambda_j)}{h_j(d,\boldsymbol{\theta})}-\frac{1}{m}\sum_{k=1}^{m}\frac{\lambda_k^{2d}I_z(\lambda_k)}{h_k(d,\boldsymbol{\theta})}\right)X_{1j}(\tilde{\mathbf{X}}_j)_i,\ i=R_y+2,\ldots,R+2,l=1,$$

$$-\hat{G}(d,\boldsymbol{\theta})^{-1}\frac{1}{m}\sum_{j=1}^{m}\left(\frac{\lambda_j^{2d}I_z(\lambda_j)}{h_j(d,\boldsymbol{\theta})}-\frac{1}{m}\sum_{k=1}^{m}\frac{\lambda_k^{2d}I_z(\lambda_k)}{h_k(d,\boldsymbol{\theta})}\right)(\tilde{\mathbf{X}}_j)_i((\tilde{\mathbf{X}}_j)_l+(j/m)^{2(l-1)}),\ i,l=2,\ldots,R_y+1,$$

$$-\hat{G}(d,\boldsymbol{\theta})^{-1}\frac{1}{m}\sum_{j=1}^{m}\left(\frac{\lambda_j^{2d}I_z(\lambda_j)}{h_j(d,\boldsymbol{\theta})}-\frac{1}{m}\sum_{k=1}^{m}\frac{\lambda_k^{2d}I_z(\lambda_k)}{h_k(d,\boldsymbol{\theta})}\right)(\tilde{\mathbf{X}}_j)_i((\tilde{\mathbf{X}}_j)_l+(j/m)^{2(l-R_y-2)}),\ i,l=R_y+2,\ldots,R+2,$$

$$\hat{G}(d,\boldsymbol{\theta})^{-1}\frac{1}{m}\sum_{j=1}^{m}\left(\frac{\lambda_j^{2d}I_z(\lambda_j)}{h_j(d,\boldsymbol{\theta})}-\frac{1}{m}\sum_{k=1}^{m}\frac{\lambda_k^{2d}I_z(\lambda_k)}{h_k(d,\boldsymbol{\theta})}\right)(\tilde{\mathbf{X}}_j)_i(\tilde{\mathbf{X}}_j)_l,\ i=2,\ldots,R_y+1,l=R_y+2,\ldots,R+2,$$

with $R = R_y + R_w$. To prove the second statement of Lemma 1(b) we have to show that these are all negligible when evaluated at $(d_0, \boldsymbol{\theta}_0)$. It suffices to prove the result for the generic term

$$V_n(d,\boldsymbol{\theta}) = \hat{G}(d,\boldsymbol{\theta})^{-1}\frac{1}{m}\sum_{j=1}^{m}\left(\frac{\lambda_j^{2d}I_z(\lambda_j)}{h_j(d,\boldsymbol{\theta})} - \frac{1}{m}\sum_{k=1}^{m}\frac{\lambda_k^{2d}I_z(\lambda_k)}{h_k(d,\boldsymbol{\theta})}\right)(\tilde{\mathbf{X}}_j)_{R_y+2}q_j(d,\boldsymbol{\theta}), \qquad (28)$$

where $q_j(d_0,\boldsymbol{\theta}_0)$ depends on j but is at most of order $O\left((\log n)^2\right)$ and satisfies $q_{j+1}(d_0,\boldsymbol{\theta}_0) - q_j(d_0,\boldsymbol{\theta}_0) = O(j^{-1}(\log n))$ uniformly in $j = 1,\ldots,m$. Summation by parts on $V_n(d_0,\boldsymbol{\theta}_0)$ yields

$$\begin{aligned}
&\hat{G}(d_0,\boldsymbol{\theta}_0)^{-1}q_m(d_0,\boldsymbol{\theta}_0)\frac{1}{m}\sum_{j=1}^{m}\left(\frac{\lambda_j^{2d}I_z(\lambda_j)}{h_j(d_0,\boldsymbol{\theta}_0)} - \frac{1}{m}\sum_{k=1}^{m}\frac{\lambda_k^{2d}I_z(\lambda_k)}{h_k(d_0,\boldsymbol{\theta}_0)}\right)(\tilde{\mathbf{X}}_{0,j})_{R_y+2}\\
&+\hat{G}(d_0,\boldsymbol{\theta}_0)^{-1}\frac{1}{m}\sum_{l=1}^{m-1}(q_l(d_0,\boldsymbol{\theta}_0)-q_{l+1}(d_0,\boldsymbol{\theta}_0))\sum_{j=1}^{l}\left(\frac{\lambda_j^{2d}I_z(\lambda_j)}{h_j(d_0,\boldsymbol{\theta}_0)} - \frac{1}{m}\sum_{k=1}^{m}\frac{\lambda_k^{2d}I_z(\lambda_k)}{h_k(d_0,\boldsymbol{\theta}_0)}\right)(\tilde{\mathbf{X}}_{0,j})_{R_y+2}\\
=\;&\frac{1}{m}q_m(d_0,\boldsymbol{\theta}_0)O_P(m^{1/2}) + \frac{1}{m}\sum_{l=1}^{m-1}(q_l(d_0,\boldsymbol{\theta}_0)-q_{l+1}(d_0,\boldsymbol{\theta}_0))O_P(l^{1/2}) = O_P\left(m^{-1/2}(\log n)^2\right),
\end{aligned}$$

where the first equality follows from part (e) of the lemma.

C.5 Proof of (c)

First we prove the result for $\mathbf{H}_{1n}$, where we need to show that

$$\sup_{\boldsymbol{\theta}\in\Theta}\left|\tilde{G}_{a,b_0,c,c_1,c_2}(d_0,\boldsymbol{\theta}) - \tilde{G}_{a,b_0,c,c_1,c_2}(d_0,\boldsymbol{\theta}_0)\right| = o_P((\log n)^2(m/n)^{2d_0})$$

for $a, c, c_1, c_2 = 0, 1, 2$ and $b = 0, 1, \ldots, 2R_y, d, d+1, \ldots, d+R_w+R_y, 2d, 2d+1, \ldots, 2d+2R_w$. Again, showing that the difference is $o_P((\log m)^{-2})$ informly in $\boldsymbol{\theta}\in\Theta$ would be sufficient to prove part (c) for $\mathbf{H}_{1n}$, but the stronger version listed in the above equation will be useful in the proof for $\mathbf{H}_{2n}$. By the triangle inequality and (26) it suffices to show that

$$\sup_{\boldsymbol{\theta}\in\Theta}\left|\tilde{G}_{a,b_0,c,c_1,c_2}(d_0,\boldsymbol{\theta}) - \hat{G}_{a,b_0,c}(d_0,\boldsymbol{\theta})\right| + \sup_{\boldsymbol{\theta}\in\Theta}\left|\hat{G}_{a,b_0,c}(d_0,\boldsymbol{\theta}) - \hat{G}_{a,b_0,c}(d_0,\boldsymbol{\theta}_0)\right| = o_P((\log n)^2(m/n)^{2d_0}). \qquad (29)$$

The first term on the left-hand side of (29) is can be bounded in exactly the same way as (26), and the result follows using Lemma 2(i)-(ii). The second term on the left-hand side of (29) is

$$\begin{aligned}
&\sup_{\boldsymbol{\theta}\in\Theta}\left|m^{-1}\sum_{j=1}^{m}\frac{\lambda_j^{2d_0}I_z(\lambda_j)}{h_j(d_0,\boldsymbol{\theta}_0)}\left(\frac{h_j(d_0,\boldsymbol{\theta}_0)}{h_j(d_0,\boldsymbol{\theta})} - 1\right)(2\log\lambda_j)^a\left(\frac{j}{m}\right)^{2b_0}\right|\\
=\;&\sup_{\boldsymbol{\theta}\in\Theta,j=1,\ldots,m}\left|\frac{h_j(d_0,\boldsymbol{\theta}_0)}{h_j(d_0,\boldsymbol{\theta})} - 1\right| m^{-1}\sum_{j=1}^{m}\frac{I_z(\lambda_j)}{h_j(d_0,\boldsymbol{\theta}_0)}(2\log\lambda_j)^a\left(\frac{j}{m}\right)^{2b_0}\\
=\;&\hat{G}_{a,b_0,c}(d_0,\boldsymbol{\theta}_0)\sup_{\boldsymbol{\theta}\in\Theta,j=1,\ldots,m}\left|\frac{h_j(d_0,\boldsymbol{\theta}_0) - h_j(d_0,\boldsymbol{\theta})}{h_j(d_0,\boldsymbol{\theta})}\right|,
\end{aligned}$$

noting that all the terms inside the summation on the right-hand side of the first equality are positive. From Lemma 2(ii),(iv) and $\hat{G}_{a,b_0,c}(d_0,\boldsymbol{\theta}_0) = O_P((\log m)^a)$, it follows that the second term on the left-hand side of (29) is $O_P((\log m)^a(1+o(1))^{-1}\lambda_m^{2d_0})$, which proves (29).

Next we prove the result for $\mathbf{H}_{2n}$. Again, it suffices to show the result for the generic term $V_n(d,\boldsymbol{\theta})$ defined in (28), i.e. we must show that $\sup_{\boldsymbol{\theta}\in\Theta}|V_n(d_0,\boldsymbol{\theta}) - V_n(d_0,\boldsymbol{\theta}_0)| = o_P(1)$. By (25) and (29) we have that

$$\sup_{\boldsymbol{\theta}\in\Theta}\hat{G}(d_0,\boldsymbol{\theta})\xrightarrow{P}G_0, \tag{30}$$

and $\sup_{\boldsymbol{\theta}\in\Theta}|V_n(d_0,\boldsymbol{\theta})-V_n(d_0,\boldsymbol{\theta}_0)|$ is, apart from a term that is negligible uniformly in $\boldsymbol{\theta}$,

$$\sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{1}{m}\sum_{j=1}^{m}\left(\frac{I_z(\lambda_j)}{g_j(d_0,\boldsymbol{\theta})}-\frac{1}{m}\sum_{k=1}^{m}\frac{I_z(\lambda_k)}{g_k(d_0,\boldsymbol{\theta})}\right)\frac{\exp(h_{wj}(\boldsymbol{\theta}_w))(j/m)^{2d_0}}{h_j(d_0,\boldsymbol{\theta})}q_j(d_0,\boldsymbol{\theta})\right.$$

$$\left.-\frac{1}{m}\sum_{j=1}^{m}\left(\frac{I_z(\lambda_j)}{g_j(d_0,\boldsymbol{\theta}_0)}-\frac{1}{m}\sum_{k=1}^{m}\frac{I_z(\lambda_k)}{g_k(d_0,\boldsymbol{\theta}_0)}\right)\frac{\exp(h_{wj}(\boldsymbol{\theta}_{0w}))(j/m)^{2d_0}}{h_j(d_0,\boldsymbol{\theta}_0)}q_j(d_0,\boldsymbol{\theta}_0)\right|$$

$$\leq \sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{1}{m}\sum_{j=1}^{m}\left(\frac{I_z(\lambda_j)}{g_j(d_0,\boldsymbol{\theta})}\frac{\exp(h_{wj}(\boldsymbol{\theta}_w))}{h_j(d_0,\boldsymbol{\theta})}q_j(d_0,\boldsymbol{\theta})-\frac{I_z(\lambda_j)}{g_j(d_0,\boldsymbol{\theta}_0)}\frac{\exp(h_{wj}(\boldsymbol{\theta}_{0w}))}{h_j(d_0,\boldsymbol{\theta}_0)}q_j(d_0,\boldsymbol{\theta}_0)\right)\left(\frac{j}{m}\right)^{2d_0}\right| \tag{31}$$

$$+\sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{1}{m}\sum_{j=1}^{m}\frac{1}{m}\sum_{k=1}^{m}\left(\frac{I_z(\lambda_k)}{g_k(d_0,\boldsymbol{\theta})}\frac{\exp(h_{wj}(\boldsymbol{\theta}_w))}{h_j(d_0,\boldsymbol{\theta})}q_j(d_0,\boldsymbol{\theta})-\frac{I_z(\lambda_k)}{g_k(d_0,\boldsymbol{\theta}_0)}\frac{\exp(h_{wj}(\boldsymbol{\theta}_{0w}))}{h_j(d_0,\boldsymbol{\theta}_0)}q_j(d_0,\boldsymbol{\theta}_0)\right)\left(\frac{j}{m}\right)^{2d_0}\right| \tag{32}$$

By the triangle inequality, (31) is bounded by

$$\sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{1}{m}\sum_{j=1}^{m}\frac{I_z(\lambda_j)}{g_j(d_0,\boldsymbol{\theta}_0)}\left(\frac{j}{m}\right)^{2d_0}\left(\frac{\exp(h_{wj}(\boldsymbol{\theta}_w))}{h_j(d_0,\boldsymbol{\theta})}q_j(d_0,\boldsymbol{\theta})-\frac{\exp(h_{wj}(\boldsymbol{\theta}_{0w}))}{h_j(d_0,\boldsymbol{\theta}_0)}q_j(d_0,\boldsymbol{\theta}_0)\right)\right| \tag{33}$$

$$+\sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{1}{m}\sum_{j=1}^{m}\left(\frac{j}{m}\right)^{2d_0}\left(\frac{I_z(\lambda_j)}{g_j(d_0,\boldsymbol{\theta})}-\frac{I_z(\lambda_j)}{g_j(d_0,\boldsymbol{\theta}_0)}\right)\frac{\exp(h_{wj}(\boldsymbol{\theta}_w))}{h_j(d_0,\boldsymbol{\theta})}q_j(d_0,\boldsymbol{\theta})\right|. \tag{34}$$

Note that, by inspection of the definition of $q_j(d,\boldsymbol{\theta})$ in (28) and application of Lemma 2(i)-(ii), it holds that

$$\sup_{\boldsymbol{\theta}\in\Theta,j=1,\ldots,m}\left|\frac{\exp(h_{wj}(\boldsymbol{\theta}_w))}{h_j(d_0,\boldsymbol{\theta})}q_j(d_0,\boldsymbol{\theta})-\frac{\exp(h_{wj}(\boldsymbol{\theta}_{0w}))}{h_j(d_0,\boldsymbol{\theta}_0)}q_j(d_0,\boldsymbol{\theta}_0)\right|=O(\lambda_m^{2d_0}(\log n)^2), \tag{35}$$

such that (33) is

$$\begin{aligned}(33) &= O_P\left(\sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{1}{m}\sum_{j=1}^{m}\frac{I_z(\lambda_j)}{g_j(d_0,\boldsymbol{\theta}_0)}\left(\frac{j}{m}\right)^{2d_0}\lambda_m^{2d_0}(\log n)^2\right|\right)\\ &= O_P\left(\hat{G}_{0,d_0,0}(d_0,\boldsymbol{\theta}_0)\lambda_m^{2d_0}(\log n)^2\right)=O_P(\lambda_m^{2d_0}(\log n)^2).\end{aligned}$$

Applying summation by parts to (34) we get the bound

$$\sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{\exp(h_{wm}(\boldsymbol{\theta}_w))}{h_m(d_0,\boldsymbol{\theta})}q_m(d_0,\boldsymbol{\theta})\frac{1}{m}\sum_{j=1}^{m}\left(\frac{j}{m}\right)^{2d_0}\left(\frac{I_z(\lambda_j)}{g_j(d_0,\boldsymbol{\theta})}-\frac{I_z(\lambda_j)}{g_j(d_0,\boldsymbol{\theta}_0)}\right)\right|$$

$$+\sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{1}{m}\sum_{k=1}^{m-1}\left(\frac{\exp(h_{wk}(\boldsymbol{\theta}_w))}{h_k(d_0,\boldsymbol{\theta})}q_k(d_0,\boldsymbol{\theta})-\frac{\exp(h_{wk+1}(\boldsymbol{\theta}_w))}{h_{k+1}(d_0,\boldsymbol{\theta})}q_{k+1}(d_0,\boldsymbol{\theta})\right)\right.$$

$$\left.\times\sum_{j=1}^{k}\left(\frac{j}{m}\right)^{2d_0}\left(\frac{I_z(\lambda_j)}{g_j(d_0,\boldsymbol{\theta})}-\frac{I_z(\lambda_j)}{g_j(d_0,\boldsymbol{\theta}_0)}\right)\right|.$$

The first term is

$$\sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{\exp(h_{wm}(\boldsymbol{\theta}_w))}{h_m\left(d_0,\boldsymbol{\theta}\right)}q_m(d_0,\boldsymbol{\theta})\frac{1}{G_0}\left(\hat{G}_{0,d_0,c}(d_0,\boldsymbol{\theta})-\hat{G}_{0,d_0,c}(d_0,\boldsymbol{\theta}_0)\right)\right| = o_P\left((\log n)^2(\log n)^2(m/n)^{2d_0}\right)$$

by (29), Lemma 2(i)-(ii), and $\sup_{\boldsymbol{\theta}\in\Theta} q_m(d_0,\boldsymbol{\theta}) = O((\log n)^2)$. The second term is

$$\begin{aligned}
& O_P\left(\sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{1}{m}\sum_{k=2}^{m-1}\left(\frac{\exp(h_{wk}(\boldsymbol{\theta}_w))}{h_k\left(d_0,\boldsymbol{\theta}\right)}q_k(d_0,\boldsymbol{\theta})-\frac{\exp(h_{wk+1}(\boldsymbol{\theta}_w))}{h_{k+1}\left(d_0,\boldsymbol{\theta}\right)}q_{k+1}(d_0,\boldsymbol{\theta})\right)\left(\frac{k}{m}\right)^{2d_0}k(\log n)^2\left(\frac{k}{n}\right)^{2d_0}\right|\right) \\
= \; & O_P\left(\sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{1}{m}\sum_{k=2}^{m-1}\frac{\exp(h_{wk}(\boldsymbol{\theta}_w))}{h_k\left(d_0,\boldsymbol{\theta}\right)}\left(q_k(d_0,\boldsymbol{\theta})-q_{k+1}(d_0,\boldsymbol{\theta})\right)\left(\frac{k}{m}\right)^{2d_0}k(\log n)^2\left(\frac{k}{n}\right)^{2d_0}\right|\right) \\
& +O_P\left(\sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{1}{m}\sum_{k=2}^{m-1}q_{k+1}(d_0,\boldsymbol{\theta})\left(\frac{\exp(h_{wk}(\boldsymbol{\theta}_w))}{h_k\left(d_0,\boldsymbol{\theta}\right)}-\frac{\exp(h_{wk+1}(\boldsymbol{\theta}_w))}{h_{k+1}\left(d_0,\boldsymbol{\theta}\right)}\right)\left(\frac{k}{m}\right)^{2d_0}k(\log n)^2\left(\frac{k}{n}\right)^{2d_0}\right|\right),
\end{aligned}$$

which, using $\sup_{\boldsymbol{\theta}\in\Theta}|q_k(d_0,\boldsymbol{\theta})-q_{k+1}(d_0,\boldsymbol{\theta})| = O(k^{-1}\log n)$ and $\sup_{\boldsymbol{\theta}\in\Theta}|q_{k+1}(d_0,\boldsymbol{\theta})| = O((\log n)^2)$, is

$$\begin{aligned}
& O_P\left(\frac{1}{m}\sum_{k=2}^{m-1}\left(\frac{k}{m}\right)^{2d_0}(\log n)^3\left(\frac{k}{n}\right)^{2d_0}+\frac{1}{m}\sum_{k=1}^{m-1}(\log n)^4\left(\lambda_{k+1}^{2d_0}-\lambda_k^{2d_0}\right)\left(\frac{k}{m}\right)^{2d_0}k\left(\frac{k}{n}\right)^{2d_0}\right) \\
= \; & O_P\left(\frac{1}{m}\sum_{k=2}^{m-1}\left(\frac{k}{m}\right)^{2d_0}(\log n)^3\left(\frac{k}{n}\right)^{2d_0}+(\log n)^4\lambda_m^{2d_0}\frac{1}{m}\sum_{k=1}^{m-1}\left(\frac{k}{m}\right)^{2d_0}\left(\frac{k}{n}\right)^{2d_0}\right) \\
= \; & O_P\left((\log n)^3(m/n)^{2d_0}+(\log n)^4(m/n)^{4d_0}\right).
\end{aligned}$$

Thus both terms of (34) are $o_P(1)$ by (17) because $d_0 > d_1 > 0$.

Along the same lines we rewrite (32) as

$$\begin{aligned}
& \sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{1}{m}\sum_{j=1}^{m}\left(\frac{\exp(h_{wj}(\boldsymbol{\theta}_w))}{h_j\left(d_0,\boldsymbol{\theta}\right)}q_j(d_0,\boldsymbol{\theta})-\frac{\exp(h_{wj}(\boldsymbol{\theta}_{0w}))}{h_j\left(d_0,\boldsymbol{\theta}_0\right)}q_j(d_0,\boldsymbol{\theta}_0)\right)\left(\frac{j}{m}\right)^{2d_0}\frac{1}{m}\sum_{k=1}^{m}\frac{I_z\left(\lambda_k\right)}{g_k\left(d_0,\boldsymbol{\theta}_0\right)}\right| \\
& +\sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{1}{m}\sum_{j=1}^{m}\frac{\exp(h_{wj}(\boldsymbol{\theta}_w))}{h_j\left(d_0,\boldsymbol{\theta}\right)}q_j(d_0,\boldsymbol{\theta})\left(\frac{j}{m}\right)^{2d_0}\frac{1}{m}\sum_{k=1}^{m}\left(\frac{I_z\left(\lambda_k\right)}{g_k\left(d_0,\boldsymbol{\theta}\right)}-\frac{I_z\left(\lambda_k\right)}{g_k\left(d_0,\boldsymbol{\theta}_0\right)}\right)\right|
\end{aligned}$$

and, using the definition of $\hat{G}_{a,b,c}(d,\boldsymbol{\theta})$, this is equal to

$$\begin{aligned}
& \sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{\hat{G}_{0,0,0}(d_0,\boldsymbol{\theta}_0)}{G_0}\frac{1}{m}\sum_{j=1}^{m}\left(\frac{\exp(h_{wj}(\boldsymbol{\theta}_w))}{h_j\left(d_0,\boldsymbol{\theta}\right)}q_j(d_0,\boldsymbol{\theta})-\frac{\exp(h_{wj}(\boldsymbol{\theta}_{0w}))}{h_j\left(d_0,\boldsymbol{\theta}_0\right)}q_j(d_0,\boldsymbol{\theta}_0)\right)\left(\frac{j}{m}\right)^{2d_0}\right| \\
& +\sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{1}{G_0}\left(\hat{G}_{0,0,0}(d_0,\boldsymbol{\theta})-\hat{G}_{0,0,0}(d_0,\boldsymbol{\theta}_0)\right)\frac{1}{m}\sum_{j=1}^{m}\frac{\exp(h_{wj}(\boldsymbol{\theta}_w))}{h_j\left(d_0,\boldsymbol{\theta}\right)}q_j(d_0,\boldsymbol{\theta})\left(\frac{j}{m}\right)^{2d_0}\right| \\
= \; & O_P\left(\sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{1}{m}\sum_{j=1}^{m}\left(\frac{\exp(h_{wj}(\boldsymbol{\theta}_w))}{h_j\left(d_0,\boldsymbol{\theta}\right)}q_j(d_0,\boldsymbol{\theta})-\frac{\exp(h_{wj}(\boldsymbol{\theta}_{0w}))}{h_j\left(d_0,\boldsymbol{\theta}_0\right)}q_j(d_0,\boldsymbol{\theta}_0)\right)\left(\frac{j}{m}\right)^{2d_0}\right|\right) \\
& +o_P\left(\sup_{\boldsymbol{\theta}\in\Theta}\left|(\log n)^2\left(\frac{m}{n}\right)^{2d_0}\frac{1}{m}\sum_{j=1}^{m}\frac{\exp(h_{wj}(\boldsymbol{\theta}_w))}{h_j\left(d_0,\boldsymbol{\theta}\right)}q_j(d_0,\boldsymbol{\theta})\left(\frac{j}{m}\right)^{2d_0}\right|\right),
\end{aligned}$$

where the second term is easily seen to be $o_P((\log n)^4(m/n)^{2d_0}) = o_P(1)$. By (35), the first term is

$$O_P\left(\sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{1}{m}\sum_{j=1}^{m}\left(\frac{m}{n}\right)^{2d_0}(\log n)^2\left(\frac{j}{m}\right)^{2d_0}\right|\right) = O_P((\log n)^2(m/n)^{2d_0}).$$

This is $o_P(1)$ which proves part (c).

C.6 Proof of (d)

Again, we first prove the result for $\mathbf{H}_{1n}$ which follows if

$$\sup_{d\in D_n(\eta_n),\boldsymbol{\theta}\in\Theta}\left|\tilde{G}_{a,b,c,c_1,c_2}(d,\boldsymbol{\theta}) - \tilde{G}_{a,b_0,c,c_1,c_2}(d_0,\boldsymbol{\theta})\right| = o_P((\log m)^{-2}) \tag{36}$$

for $a, c, c_1, c_2 = 0, 1, 2$ and $b = 0, 1, \ldots, 2R_y, d, d+1, \ldots, d+R_w+R_y, 2d, 2d+1, \ldots, 2d+2R_w$. Defining

$$\begin{aligned}\tilde{E}_{a,b,c,c_1,c_2}(d,\boldsymbol{\theta}) &= \frac{1}{m}\sum_{j=1}^{m}\frac{j^{2d}I_z(\lambda_j)}{h_j(d,\boldsymbol{\theta})^{c_1+c_2+1}}\left(\frac{2\exp(h_{yj}(\boldsymbol{\theta}_y))\log\lambda_j}{h_j(d,\boldsymbol{\theta})}\right)^a (j/m)^{2b}\,\theta_\rho^c \\ &\quad \times \exp(c_1 h_{yj}(\boldsymbol{\theta}_y) + c_2 h_{wj}(\boldsymbol{\theta}_w)),\\ \hat{E}_{a,b,c}(d,\boldsymbol{\theta}) &= \frac{1}{m}\sum_{j=1}^{m}\frac{j^{2d}I_z(\lambda_j)}{h_j(d,\boldsymbol{\theta})}(2\log\lambda_j)^a (j/m)^{2b}\,\theta_\rho^c,\end{aligned}$$

we need to show that, for all $a, c, c_1, c_2 = 0, 1, 2$ and $b = 0, 1, \ldots, 2R_y, d, d+1, \ldots, d+R_w+R_y, 2d, 2d+1, \ldots, 2d+2R_w$,

$$Z_{a,b,c,c_1,c_2}(\eta_n) := \sup_{d\in D_n(\eta_n),\boldsymbol{\theta}\in\Theta}\left|\tilde{E}_{a,b,c,c_1,c_2}(d,\boldsymbol{\theta}) - \tilde{E}_{a,b_0,c,c_1,c_2}(d_0,\boldsymbol{\theta})\right| = o_P(n^{2d_0}(\log m)^{-2}),$$

see also AS p. 600. Since b is a function of d, we distinguish between b and b_0 which are obviously the same in case $b = 0, 1, \ldots, 2R_y$. By the triangle inequality it is sufficient to show that

$$\begin{aligned}&\sup_{d\in D_n(\eta_n),\boldsymbol{\theta}\in\Theta}\left|\tilde{E}_{a,b,c,c_1,c_2}(d,\boldsymbol{\theta}) - \hat{E}_{a,b,c}(d,\boldsymbol{\theta})\right| + \sup_{d\in D_n(\eta_n),\boldsymbol{\theta}\in\Theta}\left|\hat{E}_{a,b,c}(d,\boldsymbol{\theta}) - \hat{E}_{a,b_0,c}(d_0,\boldsymbol{\theta})\right| \\ &\qquad + \sup_{\boldsymbol{\theta}\in\Theta}\left|\hat{E}_{a,b_0,c}(d_0,\boldsymbol{\theta}) - \tilde{E}_{a,b_0,c,c_1,c_2}(d_0,\boldsymbol{\theta})\right| \\ &\qquad =: Z_{1,a,b,c,c_1,c_2}(\eta_n) + Z_{2,a,b,c}(\eta_n) + Z_{3,a,b_0,c,c_1,c_2}(\eta_n) = o_P(n^{2d_0}(\log m)^{-2}).\end{aligned}$$

The result for $Z_{3,a,b_0,c,c_1,c_2}(\eta_n)$ follows from part (c) of the lemma since it does not depend on d.

For $Z_{2,a,b,c}(\eta_n)$ we find that

$$\begin{aligned}&Z_{2,a,b,c}(\eta_n)\\ &= \sup_{d\in D_n(\eta_n),\boldsymbol{\theta}\in\Theta}\left|\frac{\theta_\rho^c}{m}\sum_{j=1}^{m}\left[\frac{j^{2d}I_z(\lambda_j)}{h_j(d,\boldsymbol{\theta})}\left(\frac{j}{m}\right)^{2b} - \frac{j^{2d_0}I_z(\lambda_j)}{h_j(d_0,\boldsymbol{\theta})}\left(\frac{j}{m}\right)^{2b_0}\right](2\log\lambda_j)^a\right| \\ &= \sup_{d\in D_n(\eta_n),\boldsymbol{\theta}\in\Theta}\left|\frac{\theta_\rho^c}{m}\sum_{j=1}^{m}\left(j^{2d} - j^{2d_0}\right)I_z(\lambda_j)\frac{1}{h_j(d,\boldsymbol{\theta})}\left(\frac{j}{m}\right)^{2b}(2\log\lambda_j)^a\right| \\ &\quad + \sup_{d\in D_n(\eta_n),\boldsymbol{\theta}\in\Theta}\left|\frac{\theta_\rho^c}{m}\sum_{j=1}^{m}j^{2d_0}I_z(\lambda_j)\left(1 - \frac{h_j(d,\boldsymbol{\theta})}{h_j(d_0,\boldsymbol{\theta})}\left(\frac{j}{m}\right)^{2b_0-2b}\right)\frac{1}{h_j(d,\boldsymbol{\theta})}\left(\frac{j}{m}\right)^{2b}(2\log\lambda_j)^a\right|.\end{aligned}$$

By Lemma 2(iii), the first term of $Z_{2,a,b,c}(\eta_n)$ is bounded by

$$\sup_{d\in D_n(\eta_n)}\left|\frac{1}{m}\sum_{j=1}^{m} j^{2d_0} I_z(\lambda_j)(1+o(1))^{-1}(2\log\lambda_j)^a \left|j^{2d-2d_0}-1\right|\right|,$$

which is $o_P(n^{2d_0}(\log m)^{-2})$ as in (A.18) of AS.

The second term of $Z_{2,a,b,c}(\eta_n)$ is bounded by

$$\sup_{d\in D_n(\eta_n),\boldsymbol{\theta}\in\Theta}\left|\frac{\theta_\rho^c}{m}\sum_{j=1}^{m} j^{2d_0} I_z(\lambda_j)\left(1-\frac{h_j(d,\boldsymbol{\theta})}{h_j(d_0,\boldsymbol{\theta})}\right)\frac{1}{h_j(d,\boldsymbol{\theta})}\left(\frac{j}{m}\right)^{2b}(2\log\lambda_j)^a\right| \tag{37}$$

$$+\sup_{d\in D_n(\eta_n),\boldsymbol{\theta}\in\Theta}\left|\frac{\theta_\rho^c}{m}\sum_{j=1}^{m} j^{2d_0} I_z(\lambda_j)\frac{1}{h_j(d_0,\boldsymbol{\theta})}\left(\left(\frac{j}{m}\right)^{2b_0}-\left(\frac{j}{m}\right)^{2b}\right)(2\log\lambda_j)^a\right|, \tag{38}$$

and using Lemma 2(iii),(v) we find that (37) is

$$O_P\left(\frac{1}{m}\sum_{j=1}^{m} j^{2d_0} I_z(\lambda_j)\lambda_m^{2d_1}(\log n)^a\right)=O_P\left(\frac{1}{m}\sum_{j=1}^{m}\lambda_j^{2d_0} I_z(\lambda_j)\left(\frac{n}{2\pi}\right)^{2d_0}\lambda_m^{2d_1}(\log n)^a\right).$$

Noting that $m^{-1}\sum_{j=1}^{m}\lambda_j^{2d_0}I_z(\lambda_j)=\hat{G}_{0,0,0}(d_0,(\mathbf{0}',1,\mathbf{0}')')=G_0(1+o_P(1))$, (37) is $o_P(n^{2d_0}(\log m)^{-2})$. By the mean value theorem, $x^a=x^b+(a-b)x^{a^*}(\log x)$ for $a\le a^*\le b$ which implies that

$$\sup_{d\in D_n(\eta_n),j=1,\ldots,m}\left|\left(\frac{j}{m}\right)^{2b_0}-\left(\frac{j}{m}\right)^{2b}\right|=O\left(\sup_{d\in D_n(\eta_n)}(b_0-b)(\log m)\right)=O\left((\log n)^{-5}\eta_n(\log m)\right).$$

Thus, applying also Lemma 2(ii), (38) is

$$O_P\left(\frac{1}{m}\sum_{j=1}^{m} j^{2d_0} I_z(\lambda_j)(\log n)^{a-5}(\log m)\eta_n\right)=O_P\left(\eta_n n^{2d_0}(\log n)^{a-5}(\log m)\hat{G}_{0,0}(d_0,(\mathbf{0}',1,\mathbf{0}')')\right),$$

which is $o_P\left(n^{2d_0}(\log m)^{-2}\right)$ because $\eta_n=o(1)$ and $a\le 2$.

Next, $Z_{1,a,b,c,c_1,c_2}(\eta_n)$ is

$$\sup_{\substack{d\in D_n(\eta_n)\\ \boldsymbol{\theta}\in\Theta}}\left|\frac{\theta_\rho^c}{m}\sum_{j=1}^{m}\frac{j^{2d_0}I_z(\lambda_j)}{h_j(d,\boldsymbol{\theta})}j^{2d-2d_0}\left[\frac{\exp(c_1h_{yj}(\boldsymbol{\theta}_y)+c_2h_{wj}(\boldsymbol{\theta}_w))}{h_j(d,\boldsymbol{\theta})^{c_1+c_2}}\left(\frac{2\exp(h_{yj}(\boldsymbol{\theta}_y))\log\lambda_j}{h_j(d,\boldsymbol{\theta})}\right)^a-(2\log\lambda_j)^a\right]\left(\frac{j}{m}\right)^{2b}\right|.$$

Using the arguments applied to (26) and using Lemma 2(i),(iii), the result for $Z_{1,a,b,c,c_1,c_2}(\eta_n)$ follows.

We proceed to show that $\sup_{d\in D_n(\eta_n),\boldsymbol{\theta}\in\Theta}\mathbf{B}_n^{-1}\left\|\mathbf{H}_{2n}(d,\boldsymbol{\theta})-\mathbf{H}_{2n}(d_0,\boldsymbol{\theta})\right\|\mathbf{B}_n^{-1}=o_P(1)$ or equivalently that $\sup_{d\in D_n(\eta_n),\boldsymbol{\theta}\in\Theta}|V_n(d,\boldsymbol{\theta})-V_n(d_0,\boldsymbol{\theta})|=o_P(1)$. Since we have shown (36) we have that $\hat{G}(d,\boldsymbol{\theta})\xrightarrow{P}G$ uniformly in $d\in D_n(\eta_n)$ and $\boldsymbol{\theta}\in\Theta$, so we need to show that the following is $o_P(1)$:

$$\sup_{d\in D_n(\eta_n),\boldsymbol{\theta}\in\Theta}\left|\frac{1}{m}\sum_{j=1}^{m}\left(\frac{I_z(\lambda_j)}{g_j(d,\boldsymbol{\theta})}-\frac{1}{m}\sum_{k=1}^{m}\frac{I_z(\lambda_k)}{g_k(d,\boldsymbol{\theta})}\right)\frac{\exp(h_{wj}(\boldsymbol{\theta}_w))(j/m)^{2d}}{h_j(d,\boldsymbol{\theta})}q_j(d,\boldsymbol{\theta})\right.$$

$$\left.-\frac{1}{m}\sum_{j=1}^{m}\left(\frac{I_z(\lambda_j)}{g_j(d_0,\boldsymbol{\theta})}-\frac{1}{m}\sum_{k=1}^{m}\frac{I_z(\lambda_k)}{g_k(d_0,\boldsymbol{\theta})}\right)\frac{\exp(h_{wj}(\boldsymbol{\theta}_w))(j/m)^{2d_0}}{h_j(d_0,\boldsymbol{\theta})}q_j(d_0,\boldsymbol{\theta})\right|,$$

which is bounded by

$$\sup_{d\in D_n(\eta_n),\boldsymbol{\theta}\in\Theta}\left|\frac{1}{m}\sum_{j=1}^{m}\exp(h_{wj}(\boldsymbol{\theta}_w))\left(\frac{I_z(\lambda_j)}{g_j(d,\boldsymbol{\theta})}\frac{q_j(d,\boldsymbol{\theta})}{h_j(d,\boldsymbol{\theta})}\left(\frac{j}{m}\right)^{2d}-\frac{I_z(\lambda_j)}{g_j(d_0,\boldsymbol{\theta})}\frac{q_j(d_0,\boldsymbol{\theta})}{h_j(d_0,\boldsymbol{\theta})}\left(\frac{j}{m}\right)^{2d_0}\right)\right| \tag{39}$$

$$+\sup_{d\in D_n(\eta_n),\boldsymbol{\theta}\in\Theta}\left|\frac{1}{m}\sum_{j=1}^{m}\frac{\exp(h_{wj}(\boldsymbol{\theta}_w))}{m}\sum_{k=1}^{m}\left(\frac{I_z(\lambda_k)}{g_k(d_0,\boldsymbol{\theta})}\frac{q_j(d_0,\boldsymbol{\theta})}{h_j(d_0,\boldsymbol{\theta})}\left(\frac{j}{m}\right)^{2d_0}-\frac{I_z(\lambda_k)}{g_k(d,\boldsymbol{\theta})}\frac{q_j(d,\boldsymbol{\theta})}{h_j(d,\boldsymbol{\theta})}\left(\frac{j}{m}\right)^{2d}\right)\right| \tag{40}$$

By the triangle inequality we get the bounds

$$(39)\quad\leq\quad\sup_{d\in D_n(\eta_n),\boldsymbol{\theta}\in\Theta}\left|\frac{1}{m}\sum_{j=1}^{m}\frac{\exp(h_{wj}(\boldsymbol{\theta}_w))I_z(\lambda_j)}{g_j(d_0,\boldsymbol{\theta})}\left(\left(\frac{j}{m}\right)^{2d}-\left(\frac{j}{m}\right)^{2d_0}\right)\frac{q_j(d_0,\boldsymbol{\theta})}{h_j(d_0,\boldsymbol{\theta})}\right| \tag{41}$$

$$+\sup_{d\in D_n(\eta_n),\boldsymbol{\theta}\in\Theta}\left|\frac{1}{m}\sum_{j=1}^{m}\frac{\exp(h_{wj}(\boldsymbol{\theta}_w))I_z(\lambda_j)}{g_j(d_0,\boldsymbol{\theta})}\left(\frac{j}{m}\right)^{2d}\left(\frac{g_j(d_0,\boldsymbol{\theta})}{g_j(d,\boldsymbol{\theta})}-1\right)\frac{q_j(d_0,\boldsymbol{\theta})}{h_j(d_0,\boldsymbol{\theta})}\right| \tag{42}$$

$$+\sup_{d\in D_n(\eta_n),\boldsymbol{\theta}\in\Theta}\left|\frac{1}{m}\sum_{j=1}^{m}\frac{\exp(h_{wj}(\boldsymbol{\theta}_w))I_z(\lambda_j)}{g_j(d,\boldsymbol{\theta})}\left(\frac{j}{m}\right)^{2d}\left(\frac{q_j(d,\boldsymbol{\theta})}{h_j(d,\boldsymbol{\theta})}-\frac{q_j(d_0,\boldsymbol{\theta})}{h_j(d_0,\boldsymbol{\theta})}\right)\right| \tag{43}$$

and

$$(40)\quad\leq\quad\sup_{d\in D_n(\eta_n),\boldsymbol{\theta}\in\Theta}\left|\frac{1}{m}\sum_{j=1}^{m}\left(\left(\frac{j}{m}\right)^{2d_0}-\left(\frac{j}{m}\right)^{2d}\right)\frac{q_j(d_0,\boldsymbol{\theta})}{h_j(d_0,\boldsymbol{\theta})}\frac{\exp(h_{wj}(\boldsymbol{\theta}_w))}{m}\sum_{k=1}^{m}\frac{I_z(\lambda_k)}{g_k(d_0,\boldsymbol{\theta})}\right| \tag{44}$$

$$+\sup_{d\in D_n(\eta_n),\boldsymbol{\theta}\in\Theta}\left|\frac{1}{m}\sum_{j=1}^{m}\left(\frac{j}{m}\right)^{2d}\frac{q_j(d_0,\boldsymbol{\theta})}{h_j(d_0,\boldsymbol{\theta})}\frac{\exp(h_{wj}(\boldsymbol{\theta}_w))}{m}\sum_{k=1}^{m}\frac{I_z(\lambda_k)}{g_k(d_0,\boldsymbol{\theta})}\left(1-\frac{g_k(d_0,\boldsymbol{\theta})}{g_k(d,\boldsymbol{\theta})}\right)\right| \tag{45}$$

$$+\sup_{d\in D_n(\eta_n),\boldsymbol{\theta}\in\Theta}\left|\frac{1}{m}\sum_{j=1}^{m}\left(\frac{j}{m}\right)^{2d}\left(\frac{q_j(d_0,\boldsymbol{\theta})}{h_j(d_0,\boldsymbol{\theta})}-\frac{q_j(d,\boldsymbol{\theta})}{h_j(d,\boldsymbol{\theta})}\right)\frac{\exp(h_{wj}(\boldsymbol{\theta}_w))}{m}\sum_{k=1}^{m}\frac{I_z(\lambda_j)}{g_k(d,\boldsymbol{\theta})}\right|. \tag{46}$$

The required results for (41) and (44) follow using the mean value theorem as in (38), whereas the results for (42) and (45) follow as in (37). For (43) and (46) we note that, by inspection of the definition of $q_j(d,\boldsymbol{\theta})$ in (28), c.f. (35), it suffices to prove the result for

$$\exp(h_{wj}(\boldsymbol{\theta}_w))\left(\frac{q_j(d,\boldsymbol{\theta})}{h_j(d,\boldsymbol{\theta})}-\frac{q_j(d_0,\boldsymbol{\theta})}{h_j(d_0,\boldsymbol{\theta})}\right)=\exp(2h_{wj}(\boldsymbol{\theta}_w))\left(\frac{(j/m)^{2d}}{h_j(d,\boldsymbol{\theta})^2}-\frac{(j/m)^{2d_0}}{h_j(d_0,\boldsymbol{\theta})^2}\right)$$
$$=\left(\frac{j}{m}\right)^{2d}\left(\frac{\exp(2h_{wj}(\boldsymbol{\theta}_w))}{h_j(d,\boldsymbol{\theta})^2}-\frac{\exp(2h_{wj}(\boldsymbol{\theta}_w))}{h_j(d_0,\boldsymbol{\theta})^2}\right)+\frac{\exp(2h_{wj}(\boldsymbol{\theta}_w))}{h_j(d_0,\boldsymbol{\theta})^2}\left(\left(\frac{j}{m}\right)^{2d}-\left(\frac{j}{m}\right)^{2d_0}\right).$$

Inserting this into (43) ((46) follows in the same way) we get the bound

$$(43)\quad\leq\quad\sup_{d\in D_n(\eta_n),\boldsymbol{\theta}\in\Theta}\left|\frac{1}{m}\sum_{j=1}^{m}\frac{I_z(\lambda_j)}{g_j(d,\boldsymbol{\theta})}\left(\frac{j}{m}\right)^{4d}\left(\frac{\exp(2h_{wj}(\boldsymbol{\theta}_w))}{h_j(d,\boldsymbol{\theta})^2}-\frac{\exp(2h_{wj}(\boldsymbol{\theta}_w))}{h_j(d_0,\boldsymbol{\theta})^2}\right)\right|$$
$$+\sup_{d\in D_n(\eta_n),\boldsymbol{\theta}\in\Theta}\left|\frac{1}{m}\sum_{j=1}^{m}\frac{I_z(\lambda_j)}{g_j(d,\boldsymbol{\theta})}\left(\frac{j}{m}\right)^{2d}\frac{\exp(2h_{wj}(\boldsymbol{\theta}_w))}{h_j(d_0,\boldsymbol{\theta})^2}\left(\left(\frac{j}{m}\right)^{2d}-\left(\frac{j}{m}\right)^{2d_0}\right)\right|,$$

which we can handle similarly to (37) and (38), respectively.

Appendix D: Auxiliary lemmas

We now state two useful lemmas, which are used in the proofs of the main theorems. The first is stated without proof and gathers some properties of the function $h_j(d,\boldsymbol{\theta})$, which all follow by compactness of Θ and the expansion $\exp(x) = 1 + x + \frac{x^2}{2!} + \ldots$ of the exponential function.

Lemma 2 *Let* $h_{aj}(\boldsymbol{\theta}_a) = \sum_{r=1}^{R_a} \theta_{a,r}\lambda^{2r}$, $a = y, w$, $h_j(d,\boldsymbol{\theta}) = \exp\left(h_{yj}(\boldsymbol{\theta}_y)\right) + \lambda_j^{2d}\theta_\rho \exp\left(h_{wj}(\boldsymbol{\theta}_w)\right)$, $0 < d_1 < d_2 < 1$, *and let* Θ *be compact. Then, as* $n \to \infty$,

(i) $\sup_{\boldsymbol{\theta}\in\Theta} \exp(h_{aj}(\boldsymbol{\theta}_a)) = 1 + O((j/n)^2)$ *for* $a = y, w$,
(ii) $\inf_{\boldsymbol{\theta}\in\Theta} h_j\left(d_0,\boldsymbol{\theta}\right)^c = 1 + O((j/n)^{2d_0})$ *for* $c = 0, 1, 2$,
(iii) $\inf_{d\in[d_1,d_2],\boldsymbol{\theta}\in\Theta} h_j\left(d,\boldsymbol{\theta}\right)^c = 1 + O((j/n)^{2d_2})$ *for* $c = 0, 1, 2$,
(iv) $\sup_{\boldsymbol{\theta}\in\Theta} |h_j(d_0,\boldsymbol{\theta}_0) - h_j(d_0,\boldsymbol{\theta})| = O((j/n)^{2d_0})$,
(v) $\sup_{d\in[d_1,d_2],\boldsymbol{\theta}\in\Theta} |h_j(d_0,\boldsymbol{\theta}) - h_j(d,\boldsymbol{\theta})| = O((j/n)^{2d_1})$.

The next lemma provides approximations of the periodogram of z_t by that of ε_t, following well known results from, e.g., Robinson (1995*a*), Velasco (1999), AS, and Hurvich et al. (2005).

Lemma 3 *Let Assumptions A1-A6 hold. Then, as* $n \to \infty$ *and for all* $k = 1, \ldots, m$,

$$\sum_{j=1}^{k}\left(\frac{I_z\left(\lambda_j\right)}{g_j\left(d_0,\boldsymbol{\theta}_0\right)} - 2\pi I_\varepsilon\left(\lambda_j\right)\right) = O_P\left(\xi_{k,n}(d_0) + k^{\varphi_y+1}n^{-\varphi_y} + k^{d_0+\varphi_w+1}n^{-d_0-\varphi_w} + k^{1+2d_0}n^{-2d_0} + k^{2d_0}n^{-d_0}(\log k)\right)$$

and

$$\sum_{j=1}^{k}\left(\frac{I_z\left(\lambda_j\right)}{g_j\left(d_0,\boldsymbol{\theta}_0\right)} - 2\pi I_\varepsilon\left(\lambda_j\right) - E\left(\frac{I_z\left(\lambda_j\right)}{g_j\left(d_0,\boldsymbol{\theta}_0\right)} - 2\pi I_\varepsilon\left(\lambda_j\right)\right)\right) = O_P\left(\xi_{k,n}(d_0) + k^{\varphi_y+1/2}n^{-\varphi_y} + k^{1/2+2d_0}n^{-2d_0}\right),$$

where

$$\xi_{k,n}(d) = k^{1/3}(\log k)^{2/3} + k^{1/2}n^{-1/4}$$

in the stationary case, and in the nonstationary case

$$\xi_{k,n}(d) = k^{1/(5-4d)}(\log k)^{2/(5-4d)} + k^{2d-1}(\log k) + n^{-1/2}k^{(1+d)/2}(\log n)^{5/4} + n^{-1/4}k^{d}(\log k)^{1/2}.$$

Proof. Define $\tilde{g}_j(d,\boldsymbol{\theta}) = \lambda_j^{-2d}G_0\exp(h_{yj}(\boldsymbol{\theta}_y))$ and write

$$\sum_{j=1}^{k}\left(\frac{I_z\left(\lambda_j\right)}{g_j\left(d_0,\boldsymbol{\theta}_0\right)} - 2\pi I_\varepsilon\left(\lambda_j\right)\right) = \sum_{j=1}^{k}\left(\frac{I_z\left(\lambda_j\right)}{g_j\left(d_0,\boldsymbol{\theta}_0\right)} - \frac{I_y\left(\lambda_j\right)}{\tilde{g}_j(d_0,\boldsymbol{\theta}_0)}\right) + \sum_{j=1}^{k}\left(\frac{I_y\left(\lambda_j\right)}{\tilde{g}_j(d_0,\boldsymbol{\theta}_0)} - 2\pi I_\varepsilon\left(\lambda_j\right)\right). \tag{47}$$

In the stationary case the second term on the right-hand side of (47) is $O_P(k^{1/3}(\log k)^{2/3} + k^{\varphi_y+1}n^{-\varphi_y} + k^{1/2}n^{-1/4})$ by (A.13)(i) of AS, and in the nonstationary case it is $O_P(k^{1/(5-4d_0)}(\log k)^{2/(5-4d_0)} + k^{\varphi_y+1}n^{-\varphi_y} + k^{2d_0-1}(\log k) + n^{-1/2}k^{(1+d_0)/2}(\log n)^{5/4} + n^{-1/4}k^{d_0}(\log k)^{1/2})$ by slight modification of Lemma 1 of Velasco (1999) to account for the better approximation of $f_y(\lambda_j)$ due to the polynomial in $\tilde{g}_j(d_0, \boldsymbol{\theta}_0)$ (the required modification is the same as that used by AS to modify (4.8) of Robinson (1995*a*) to obtain their (A.13)(i)). The first term on the right-hand side of (47) is

$$\frac{I_z(\lambda_j)}{g_j(d_0,\boldsymbol{\theta}_0)} - \frac{I_y(\lambda_j)}{\tilde{g}_j(d_0,\boldsymbol{\theta}_0)} = \frac{\tilde{g}_j(d_0,\boldsymbol{\theta}_0) - g_j(d_0,\boldsymbol{\theta}_0)}{g_j(d_0,\boldsymbol{\theta}_0)}\left(\frac{I_y(\lambda_j)}{\tilde{g}_j(d_0,\boldsymbol{\theta}_0)} - 1\right) \tag{48}$$

$$+\frac{2\sqrt{G_0\theta_{0\rho}\exp(h_{wj}(\boldsymbol{\theta}_{0w}))\tilde{g}_j(d_0,\boldsymbol{\theta}_0)}}{g_j(d_0,\boldsymbol{\theta}_0)}\frac{\operatorname{Re}(I_{yw}(\lambda_j))}{\sqrt{G_0\theta_{0\rho}\exp(h_{wj}(\boldsymbol{\theta}_{0w}))\tilde{g}_j(d_0,\boldsymbol{\theta}_0)}} \tag{49}$$

$$+\frac{I_w(\lambda_j) + \tilde{g}_j(d_0,\boldsymbol{\theta}_0) - g_j(d_0,\boldsymbol{\theta}_0)}{g_j(d_0,\boldsymbol{\theta}_0)}, \tag{50}$$

where $I_{ab}(\lambda) = \frac{1}{2\pi n}\sum_{t=1}^{n}\sum_{s=1}^{n} a_t b_s e^{i(s-t)\lambda}$ denotes the cross-periodogram between the two series a_t and b_t. Using summation by parts on (48) we find that

$$\sum_{j=1}^{k}\frac{\tilde{g}_j(d_0,\boldsymbol{\theta}_0) - g_j(d_0,\boldsymbol{\theta}_0)}{g_j(d_0,\boldsymbol{\theta}_0)}\left(\frac{I_y(\lambda_j)}{\tilde{g}_j(d_0,\boldsymbol{\theta}_0)} - 1\right)$$

$$= \sum_{j=1}^{k-1}\left(\frac{\tilde{g}_j(d_0,\boldsymbol{\theta}_0) - g_j(d_0,\boldsymbol{\theta}_0)}{g_j(d_0,\boldsymbol{\theta}_0)} - \frac{\tilde{g}_{j+1}(d_0,\boldsymbol{\theta}_0) - g_{j+1}(d_0,\boldsymbol{\theta}_0)}{g_{j+1}(d_0,\boldsymbol{\theta}_0)}\right)\sum_{l=1}^{j}\left(\frac{I_y(\lambda_l)}{\tilde{g}_l(d_0,\boldsymbol{\theta}_0)} - 1\right)$$

$$+\frac{\tilde{g}_k(d_0,\boldsymbol{\theta}_0) - g_k(d_0,\boldsymbol{\theta}_0)}{g_k(d_0,\boldsymbol{\theta}_0)}\sum_{j=1}^{k}\left(\frac{I_y(\lambda_j)}{\tilde{g}_j(d_0,\boldsymbol{\theta}_0)} - 1\right),$$

which is $O_P((k/n)^{2d_0}(k^{1/3}(\log k)^{2/3} + k^{\varphi_y+1}n^{-\varphi_y} + k^{1/2}n^{-1/4} + k^{1/2}))$ in the stationary case whereas it is $O_P((k/n)^{2d_0}(k^{1/(5-4d_0)}(\log k)^{2/(5-4d_0)} + k^{\varphi_y+1}n^{-\varphi_y} + k^{2d_0-1}(\log k) + n^{-1/2}k^{(1+d_0)/2}(\log n)^{5/4} + n^{-1/4}k^{d_0}(\log k)^{1/2} + k^{1/2}))$ in the nonstationary case, by the same methods as applied previously and using also (4.9) of Robinson (1995*a*) and that $|\tilde{g}_j(d_0,\boldsymbol{\theta}_0)/g_j(d_0,\boldsymbol{\theta}_0) - 1| \leq C(j/n)^{2d_0}$. Next, (50) is easily seen to be $O_P((j/n)^{2d_0})$ because $E|I_w(\lambda_j)| = O_P(1)$ uniformly in $j = 1,\ldots,m$. Since $\{y_t\}$ and $\{w_t\}$ are independent (49) is $O_P((j/n)^{d_0}(j^{-1}(\log j) + (j/n)^{\min(\varphi_y,\varphi_w)}))$ in the stationary case by Theorem 2 of Robinson (1995*b*), yielding a contribution to (47) of $O_P((k/n)^{d_0}((\log k) + k^{1+\min(\varphi_y,\varphi_w)}n^{-\min(\varphi_y,\varphi_w)}))$. In the nonstationary case we use Theorem 1 of Velasco (1999) which shows that $\operatorname{Re}(I_{yw}(\lambda_j))|\tilde{g}_j(d_0,\boldsymbol{\theta}_0)|^{-1/2}|G_0\theta_{0\rho}\exp(h_{wj}(\boldsymbol{\theta}_{0w}))|^{-1/2} = O_P(j^{2d_0-2}(\log j) + (j/n)^{\min(\varphi_y,\varphi_w)})$, yielding a contribution to (47) of $O_P((k/n)^{d_0}(k^{2d_0-1}(\log k) + k^{1+\min(\varphi_y,\varphi_w)}n^{-\min(\varphi_y,\varphi_w)}))$ (Velasco's result has to be modified to accommodate multivariate time series, but the modification is simple by comparing e.g. his equation (A.1) with equation (4.3) of Robinson (1995*b*), see also the second to last equation on p. 108 of Velasco (1999)). The difference in the remainder terms relative to Robinson (1995*b*) and Velasco (1999) is due to the different remainder term in the approximation of $f_y(\lambda_j)$ by $\tilde{g}_j(d_0,\boldsymbol{\theta}_0)$ due to the polynomial in $\tilde{g}_j(d_0,\boldsymbol{\theta}_0)$.

To prove the second result we write

$$\sum_{j=1}^{k}\left(\frac{I_z(\lambda_j)}{g_j(d_0,\boldsymbol{\theta}_0)}-2\pi I_\varepsilon(\lambda_j)-E\left(\frac{I_z(\lambda_j)}{g_j(d_0,\boldsymbol{\theta}_0)}-2\pi I_\varepsilon(\lambda_j)\right)\right)$$

$$= \sum_{j=1}^{k}\left(\frac{I_z(\lambda_j)}{g_j(d_0,\boldsymbol{\theta}_0)}-\frac{I_y(\lambda_j)}{\tilde{g}_j(d_0,\boldsymbol{\theta}_0)}-E\left(\frac{I_z(\lambda_j)}{g_j(d_0,\boldsymbol{\theta}_0)}-\frac{I_y(\lambda_j)}{\tilde{g}_j(d_0,\boldsymbol{\theta}_0)}\right)\right) \quad (51)$$

$$+\sum_{j=1}^{k}\left(\frac{I_y(\lambda_j)}{\tilde{g}_j(d_0,\boldsymbol{\theta}_0)}-2\pi I_\varepsilon(\lambda_j)-E\left(\frac{I_y(\lambda_j)}{\tilde{g}_j(d_0,\boldsymbol{\theta}_0)}-2\pi I_\varepsilon(\lambda_j)\right)\right). \quad (52)$$

By (A.21) of AS, (52) is $O_P(k^{1/3}(\log k)^{2/3}+k^{\varphi_y+1/2}n^{-\varphi_y}+k^{1/2}n^{-1/4})$ in the stationary case, and by (slight modification of) Lemma 1 of Velasco (1999), (52) is $O_P(k^{1/(5-4d_0)}(\log k)^{2/(5-4d_0)}+k^{\varphi_y+1/2}n^{-\varphi_y}+k^{2d_0-1}(\log k)+n^{-1/2}k^{(1+d_0)/2}(\log n)^{5/4}+n^{-1/4}k^{d_0}(\log k)^{1/2})$ in the nonstationary case. For eq. (51) we write

$$\frac{I_z(\lambda_j)}{g_j(d_0,\boldsymbol{\theta}_0)}-\frac{I_y(\lambda_j)}{\tilde{g}_j(d_0,\boldsymbol{\theta}_0)}-E\left(\frac{I_z(\lambda_j)}{g_j(d_0,\boldsymbol{\theta}_0)}-\frac{I_y(\lambda_j)}{\tilde{g}_j(d_0,\boldsymbol{\theta}_0)}\right)$$

$$= \frac{\tilde{g}_j(d_0,\boldsymbol{\theta}_0)-g_j(d_0,\boldsymbol{\theta}_0)}{g_j(d_0,\boldsymbol{\theta}_0)}\left[\left(\frac{I_y(\lambda_j)}{\tilde{g}_j(d_0,\boldsymbol{\theta}_0)}-2\pi I_\varepsilon(\lambda_j)\right)-E\left(\frac{I_y(\lambda_j)}{\tilde{g}_j(d_0,\boldsymbol{\theta}_0)}-2\pi I_\varepsilon(\lambda_j)\right)\right] \quad (53)$$

$$+\frac{2\sqrt{\tilde{g}_j(d_0,\boldsymbol{\theta}_0)}}{g_j(d_0,\boldsymbol{\theta}_0)}\frac{\mathrm{Re}\left(I_{yw}(\lambda_j)-EI_{yw}(\lambda_j)\right)}{\sqrt{\tilde{g}_j(d_0,\boldsymbol{\theta}_0)}} \quad (54)$$

$$+\frac{G_0\theta_{0\rho}\exp(h_{wj}(\boldsymbol{\theta}_{0w}))}{g_j(d_0,\boldsymbol{\theta}_0)}\left[\left(\frac{I_w(\lambda_j)}{G_0\theta_{0\rho}\exp(h_{wj}(\boldsymbol{\theta}_{0w}))}-2\pi I_\eta(\lambda_j)\right)-E\left(\frac{I_w(\lambda_j)}{G_0\theta_{0\rho}\exp(h_{wj}(\boldsymbol{\theta}_{0w}))}-2\pi I_\eta(\lambda_j)\right)\right] \quad (55)$$

$$+\frac{\tilde{g}_j(d_0,\boldsymbol{\theta}_0)-g_j(d_0,\boldsymbol{\theta}_0)}{g_j(d_0,\boldsymbol{\theta}_0)}\left(2\pi I_\varepsilon(\lambda_j)-1\right)+\frac{G_0\theta_{0\rho}\exp(h_{wj}(\boldsymbol{\theta}_{0w}))}{g_j(d_0,\boldsymbol{\theta}_0)}\left(2\pi I_\eta(\lambda_j)-1\right), \quad (56)$$

using also that $G_0\theta_{0\rho}\exp(h_{wj}(\boldsymbol{\theta}_{0w}))=g_j(d_0,\boldsymbol{\theta}_0)-\tilde{g}_j(d_0,\boldsymbol{\theta}_0)$.

Using summation by parts the contribution from the last term of (56) is $G_0\theta_{0\rho}$ times

$$\sum_{j=1}^{k}\frac{\exp(h_{wj}(\boldsymbol{\theta}_{0w}))}{g_j(d_0,\boldsymbol{\theta}_0)}\left(2\pi I_\eta(\lambda_j)-1\right)$$

$$= \sum_{j=1}^{k-1}\left(\frac{\exp(h_{wj}(\boldsymbol{\theta}_{0w}))}{g_j(d_0,\boldsymbol{\theta}_0)}-\frac{\exp(h_{wj+1}(\boldsymbol{\theta}_{0w}))}{g_{j+1}(d_0,\boldsymbol{\theta}_0)}\right)\sum_{l=1}^{j}\left(2\pi I_\eta(\lambda_l)-1\right)+\frac{\exp(h_{wk}(\boldsymbol{\theta}_{0w}))}{g_k(d_0,\boldsymbol{\theta}_0)}\sum_{j=1}^{k}\left(2\pi I_\eta(\lambda_j)-1\right)$$

$$= \sum_{j=1}^{k-1}\left|\frac{\exp(h_{wj}(\boldsymbol{\theta}_{0w}))g_{j+1}(d_0,\boldsymbol{\theta}_0)-\exp(h_{wj+1}(\boldsymbol{\theta}_{0w}))g_j(d_0,\boldsymbol{\theta}_0)}{g_j(d_0,\boldsymbol{\theta}_0)\,g_{j+1}(d_0,\boldsymbol{\theta}_0)}\right|O_P(j^{1/2})+\frac{\exp(h_{wk}(\boldsymbol{\theta}_{0w}))}{g_k(d_0,\boldsymbol{\theta}_0)}O_P(k^{1/2})$$

$$= O_P\left(\sum_{j=1}^{k-1}j^{2d_0-1/2}n^{-2d_0}\right)+O_P(k^{1/2+2d_0}n^{-2d_0})=O_P(k^{1/2+2d_0}n^{-2d_0}),$$

using (4.9) of Robinson (1995a) for the second equality. The first term of (56) is handled in exactly the same way yielding the same contribution. For the term (54) we can split it up in the same way as (55) and the last term of (56), and the contribution is the same.

Using summation by parts on (53) its contribution to (51) is

$$\sum_{j=1}^{k-1}\left(\frac{\tilde{g}_j(d_0,\boldsymbol{\theta}_0)-g_j\left(d_0,\boldsymbol{\theta}_0\right)}{g_j\left(d_0,\boldsymbol{\theta}_0\right)}-\frac{\tilde{g}_{j+1}(d_0,\boldsymbol{\theta}_0)-g_{j+1}\left(d_0,\boldsymbol{\theta}_0\right)}{g_{j+1}\left(d_0,\boldsymbol{\theta}_0\right)}\right)$$
$$\times\sum_{l=1}^{j}\left[\left(\frac{I_y\left(\lambda_l\right)}{\tilde{g}_l(d_0,\boldsymbol{\theta}_0)}-2\pi I_\varepsilon\left(\lambda_l\right)\right)-E\left(\frac{I_y\left(\lambda_l\right)}{\tilde{g}_l(d_0,\boldsymbol{\theta}_0)}-2\pi I_\varepsilon\left(\lambda_l\right)\right)\right]$$
$$+\frac{\tilde{g}_k(d_0,\boldsymbol{\theta}_0)-g_k\left(d_0,\boldsymbol{\theta}_0\right)}{g_k\left(d_0,\boldsymbol{\theta}_0\right)}\sum_{j=1}^{k}\left[\left(\frac{I_y\left(\lambda_j\right)}{\tilde{g}_j(d_0,\boldsymbol{\theta}_0)}-2\pi I_\varepsilon\left(\lambda_j\right)\right)-E\left(\frac{I_y\left(\lambda_j\right)}{\tilde{g}_j(d_0,\boldsymbol{\theta}_0)}-2\pi I_\varepsilon\left(\lambda_j\right)\right)\right],$$

which is $O_P\left((k/n)^{2d_0}(k^{1/3}(\log k)^{2/3}+k^{\varphi_y+1/2}n^{-\varphi_y}+k^{1/2}n^{-1/4})\right)$ in the stationary case using (A.21) of AS. In the nonstationary case we use Lemma 1 of Velasco (1999) and get that the contribution of (53) to (51) is

$$O_P((k/n)^{2d_0}(k^{1/(5-4d_0)}(\log k)^{2/(5-4d_0)}+k^{\varphi_y+1/2}n^{-\varphi_y}+k^{2d_0-1}(\log k)$$
$$+n^{-1/2}k^{(1+d_0)/2}(\log n)^{5/4}+n^{-1/4}k^{d_0}(\log k)^{1/2})).$$

Finally the term (55) is handled in exactly the same way as the stationary case of (53) yielding the contribution $O_P\left((k/n)^{2d_0}(k^{1/3}(\log k)^{2/3}+k^{\varphi_w+1/2}n^{-\varphi_w}+k^{1/2}n^{-1/4})\right)$ to (51). ■

References

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Ebens, H., 2001. The distribution of realized stock return volatility. Journal of Financial Economics 61, 43–76.

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Labys, P., 2001. The distribution of realized exchange rate volatility. Journal of the American Statistical Association 96, 42–55.

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Labys, P., 2003. Modelling and forecasting realized volatility. Econometrica 71, 579–625.

Andrews, D. W. K. & Guggenberger, P., 2003. A bias-reduced log-periodogram regression estimator for the long memory parameter. Econometrica 71, 675–712.

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

Arteche, J., 2004. Gaussian semiparametric estimation in long memory in stochastic volatility and signal plus noise models. Journal of Econometrics 119, 131–154.

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

Baillie, R. T., 1996. Long memory processes and fractional integration in econometrics. Journal of Econometrics 73, 5–59.

Baillie, R. T., Bollerslev, T. & Mikkelsen, H. O., 1996. Fractionally integrated generalized autoregressive conditional heteroscedasticity. Journal of Econometrics 74, 3–30.

Beran, J., 1994. Statistics for Long-Memory Processes, Chapman-Hall, New York.

Bollerslev, T. & Jubinski, D., 1999. Equity trading volume and volatility: latent information arrivals and common long-run dependencies. Journal of Business and Economic Statistics 17, 9–21.

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

Comte, F. & Renault, E., 1998. Long memory in continuous-time stochastic volatility models. Mathematical Finance 8, 291–323.

Davies, R. B. & Harte, D. S., 1987. Tests for Hurst effect. Biometrika 74, 95–102.

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

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

Doornik, J. A., 2006. Ox: An Object-Oriented Matrix Language. Timberlake Consultants, London.

Fox, R. & Taqqu, M. S., 1986. Large-sample properties of parameter estimates for strongly dependent stationary Gaussian series. Journal of Time Series Analysis 4, 221–238.

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

Fuller, W. A., 1996. Introduction to Statistical Time Series. Wiley, New York.

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

Harvey, A. C., 1989. Forecasting, structural time series models and the Kalman filter. Cambridge University Press, Cambridge.

Harvey, A. C., 1998. Long memory in stochastic volatility, in J. Knight & S. Satchell, eds., Forecasting Volatility in Financial Markets. Butterworth-Heinemann, London, pp. 307–320.

Harvey, A. C., Ruiz, E. & Shephard, N., 1994. Multivariate stochastic variance models. Review of Economics Studies 61, 247–264.

Henry, M. & Zaffaroni, P., 2003. The long range paradigm for macroeconomics and finance, in P. Doukhan, G. Oppenheim & M. S. Taqqu, eds., Theory and Applications of Long-Range Dependence. Birkhäuser, Boston, pp. 417–438.

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

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

Hurvich, C. M. & Ray, B. K., 2003. The local Whittle estimator of long-memory stochastic volatility. Journal of Financial Econometrics 1, 445–470.

Künsch, H. R., 1987. Statistical aspects of self-similar processes, in Y. Prokhorov & V. V. Sazanov, eds., Proceedings of the First World Congress of the Bernoulli Society. VNU Science Press, Utrecht, pp. 67–74.

Liesenfeld, R., 2001. A generalized bivariate mixture model for stock price volatility and trading volume. Journal of Econometrics 104, 141–178.

Phillips, P. C. B. & Shimotsu, K., 2004. Local whittle estimation in nonstationary and unit root cases. Annals of Statistics 32, 656–692.

Ray, B. K. & Tsay, R., 2000. Long-range dependence in daily stock volatility. Journal of Business and Economic Statistics 18, 254–262.

Robinson, P. M., 1995a. Gaussian semiparametric estimation of long range dependence. Annals of Statistics 23, 1630–1661.

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

Robinson, P. M., 2003. Long-memory time series, in P. M. Robinson, ed., Time Series With Long Memory. Oxford University Press, Oxford, pp. 4–32.

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

Velasco, C., 1999. Gaussian semiparametric estimation of non-stationary time series. Journal of Time Series Analysis 20, 87–127.

Wright, J. H., 2002. Log periodogram estimation of long memory volatility dependencies with conditionally heavy tailed returns. Econometric Reviews 21, 397–417.

Table 1: Simulation results for Model I

		LWN		LPWN(1,0)		LPWN(0,1)		LPWN(1,1)	
nsr	n	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
Panel A: $m = \lfloor (3/4)n^{0.8} \rfloor$									
5	2048	0.0025	0.1215	-0.0239	0.1260	0.0149	0.1499	0.0169	0.1551
	4096	0.0007	0.0832	-0.0180	0.0901	0.0056	0.1036	0.0129	0.1069
	8192	0.0024	0.0588	-0.0106	0.0667	0.0024	0.0738	0.0091	0.0755
10	2048	0.0037	0.1663	-0.0294	0.1724	0.0208	0.1906	0.0087	0.1950
	4096	0.0009	0.1067	-0.0235	0.1176	0.0056	0.1265	0.0086	0.1283
	8192	0.0022	0.0723	-0.0137	0.0829	0.0024	0.0880	0.0088	0.0886
15	2048	0.0065	0.1999	-0.0317	0.2064	0.0245	0.2201	-0.0085	0.2238
	4096	0.0009	0.1292	-0.0254	0.1409	0.0088	0.1493	0.0042	0.1512
	8192	0.0015	0.0846	-0.0154	0.0962	0.0035	0.0962	0.0062	0.0967
Panel B: $m = \lfloor n^{0.8} \rfloor$									
5	2048	0.0033	0.1097	-0.0243	0.1176	0.0072	0.1343	0.0123	0.1398
	4096	0.0017	0.0755	-0.0174	0.0850	0.0026	0.0945	0.0100	0.0955
	8192	0.0035	0.0536	-0.0084	0.0619	0.0034	0.0685	0.0111	0.0695
10	2048	0.0080	0.1532	-0.0224	0.1634	0.0165	0.1751	0.0063	0.1810
	4096	0.0012	0.0979	-0.0205	0.1099	0.0024	0.1119	0.0054	0.1140
	8192	0.0035	0.0666	-0.0120	0.0767	0.0035	0.0776	0.0079	0.0784
15	2048	0.0097	0.1893	-0.0218	0.1984	0.0223	0.2070	-0.0018	0.2142
	4096	0.0031	0.1205	-0.0191	0.1337	0.0057	0.1347	0.0024	0.1372
	8192	0.0024	0.0785	-0.0131	0.0896	0.0038	0.0850	0.0041	0.0865
Panel C: $m = \lfloor (5/4)n^{0.8} \rfloor$									
5	2048	0.0041	0.1037	-0.0214	0.1130	0.0054	0.1260	0.0120	0.1295
	4096	0.0036	0.0708	-0.0146	0.0802	0.0018	0.0867	0.0101	0.0877
	8192	0.0043	0.0494	-0.0072	0.0568	0.0029	0.0610	0.0101	0.0624
10	2048	0.0053	0.1429	-0.0212	0.1558	0.0108	0.1634	0.0056	0.1694
	4096	0.0029	0.0918	-0.0163	0.1045	0.0033	0.1066	0.0060	0.1085
	8192	0.0032	0.0625	-0.0102	0.0713	0.0040	0.0725	0.0074	0.0726
15	2048	0.0056	0.1836	-0.0189	0.1928	0.0168	0.1999	0.0008	0.2074
	4096	0.0029	0.1125	-0.0144	0.1245	0.0059	0.1239	0.0044	0.1280
	8192	0.0040	0.0742	-0.0079	0.0835	0.0059	0.0791	0.0053	0.0809

Note: The polynomial approximation used under the heading "LPWN(R_y, R_w)" is (R_y, R_w).

Table 2: Simulation results for Model II with $(\alpha_y, \beta_y) = (0.8, 0)$ and $(\alpha_w, \beta_w) = (0, 0)$

		LWN		LPWN(1,0)		LPWN(0,1)		LPWN(1,1)	
nsr	n	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
Panel A: $m = \lfloor (3/4)n^{0.8} \rfloor$									
5	2048	0.0460	0.1158	-0.0046	0.1115	0.0145	0.1234	0.0106	0.1272
	4096	0.0315	0.0786	-0.0140	0.0800	-0.0022	0.0838	-0.0006	0.0856
	8192	0.0217	0.0555	-0.0195	0.0629	-0.0113	0.0620	-0.0093	0.0637
10	2048	0.0264	0.1486	-0.0212	0.1568	0.0159	0.1695	0.0014	0.1729
	4096	0.0151	0.0955	-0.0259	0.1107	-0.0046	0.1167	-0.0057	0.1149
	8192	0.0081	0.0649	-0.0258	0.0827	-0.0158	0.0834	-0.0110	0.0817
15	2048	0.0206	0.1806	-0.0278	0.1894	0.0190	0.1977	-0.0126	0.2049
	4096	0.0074	0.1156	-0.0314	0.1333	0.0001	0.1369	-0.0062	0.1375
	8192	0.0020	0.0769	-0.0274	0.0958	-0.0082	0.0949	-0.0069	0.0936
Panel B: $m = \lfloor n^{0.8} \rfloor$									
5	2048	0.0667	0.1177	0.0105	0.1109	0.0298	0.1213	0.0197	0.1232
	4096	0.0516	0.0839	0.0001	0.0786	0.0127	0.0822	0.0109	0.0823
	8192	0.0410	0.0621	-0.0057	0.0586	0.0018	0.0562	0.0027	0.0585
10	2048	0.0457	0.1433	-0.0006	0.1530	0.0281	0.1615	0.0085	0.1631
	4096	0.0298	0.0922	-0.0109	0.1049	0.0131	0.1098	0.0057	0.1081
	8192	0.0224	0.0638	-0.0140	0.0759	0.0032	0.0788	0.0015	0.0756
15	2048	0.0368	0.1760	-0.0051	0.1862	0.0289	0.1895	-0.0007	0.1950
	4096	0.0221	0.1100	-0.0132	0.1271	0.0125	0.1270	0.0012	0.1275
	8192	0.0141	0.0723	-0.0153	0.0887	0.0099	0.0857	0.0037	0.0851
Panel C: $m = \lfloor (5/4)n^{0.8} \rfloor$									
5	2048	0.0803	0.1230	0.0283	0.1138	0.0436	0.1217	0.0296	0.1226
	4096	0.0662	0.0910	0.0162	0.0773	0.0293	0.0847	0.0235	0.0817
	8192	0.0552	0.0703	0.0077	0.0544	0.0156	0.0560	0.0149	0.0559
10	2048	0.0546	0.1393	0.0155	0.1505	0.0396	0.1561	0.0195	0.1577
	4096	0.0416	0.0923	0.0056	0.1025	0.0271	0.1082	0.0150	0.1046
	8192	0.0328	0.0654	-0.0014	0.0710	0.0201	0.0785	0.0127	0.0730
15	2048	0.0422	0.1734	0.0087	0.1824	0.0378	0.1869	0.0096	0.1908
	4096	0.0310	0.1058	0.0039	0.1202	0.0261	0.1192	0.0114	0.1206
	8192	0.0241	0.0715	-0.0004	0.0840	0.0234	0.0814	0.0130	0.0817

Note: The polynomial approximation used under the heading "LPWN(R_y, R_w)" is (R_y, R_w).

Table 3: Simulation results for Model III with $(\alpha_y, \beta_y) = (0, 0)$ and $(\alpha_w, \beta_w) = (0, 0.8)$

		LWN		LPWN(1,0)		LPWN(0,1)		LPWN(1,1)	
nsr	n	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
Panel A: $m = \lfloor (3/4)n^{0.8} \rfloor$									
5	2048	-0.0549	0.1202	-0.0565	0.1193	0.0158	0.1543	0.0203	0.1602
	4096	-0.0397	0.0887	-0.0373	0.0868	0.0056	0.1046	0.0160	0.1106
	8192	-0.0251	0.0632	-0.0216	0.0635	0.0022	0.0749	0.0110	0.0779
10	2048	-0.0833	0.1626	-0.0874	0.1627	0.0100	0.1927	0.0083	0.1988
	4096	-0.0602	0.1169	-0.0588	0.1140	0.0005	0.1295	0.0110	0.1352
	8192	-0.0390	0.0811	-0.0353	0.0796	0.0001	0.0889	0.0102	0.0926
15	2048	-0.1009	0.1903	-0.1095	0.1943	0.0061	0.2213	-0.0082	0.2272
	4096	-0.0762	0.1411	-0.0755	0.1377	-0.0051	0.1521	0.0017	0.1577
	8192	-0.0500	0.0971	-0.0465	0.0941	-0.0081	0.0987	0.0030	0.1008
Panel B: $m = \lfloor n^{0.8} \rfloor$									
5	2048	-0.1034	0.1325	-0.0910	0.1237	0.0108	0.1400	0.0199	0.1484
	4096	-0.0768	0.1032	-0.0600	0.0885	0.0034	0.0969	0.0156	0.1012
	8192	-0.0509	0.0729	-0.0352	0.0609	0.0037	0.0696	0.0147	0.0728
10	2048	-0.1496	0.1796	-0.1419	0.1752	0.0158	0.1872	0.0131	0.1928
	4096	-0.1153	0.1434	-0.0978	0.1257	0.0024	0.1215	0.0139	0.1260
	8192	-0.0767	0.1003	-0.0602	0.0842	0.0011	0.0842	0.0132	0.0874
15	2048	-0.1817	0.2113	-0.1801	0.2131	0.0160	0.2202	-0.0006	0.2266
	4096	-0.1416	0.1714	-0.1254	0.1546	-0.0006	0.1450	0.0079	0.1509
	8192	-0.0975	0.1235	-0.0812	0.1055	-0.0048	0.0948	0.0095	0.0985
Panel C: $m = \lfloor (5/4)n^{0.8} \rfloor$									
5	2048	-0.1465	0.1548	-0.1425	0.1572	0.0156	0.1296	0.0236	0.1367
	4096	-0.1222	0.1341	-0.0974	0.1113	0.0057	0.0900	0.0175	0.0935
	8192	-0.0868	0.0987	-0.0619	0.0742	0.0047	0.0633	0.0161	0.0670
10	2048	-0.1971	0.2022	-0.2114	0.2225	0.0226	0.1752	0.0156	0.1808
	4096	-0.1757	0.1856	-0.1545	0.1667	0.0084	0.1170	0.0155	0.1192
	8192	-0.1297	0.1424	-0.1041	0.1151	0.0046	0.0790	0.0165	0.0816
15	2048	-0.2243	0.2279	-0.2543	0.2631	0.0309	0.2154	0.0122	0.2203
	4096	-0.2091	0.2174	-0.1951	0.2065	0.0101	0.1397	0.0077	0.1437
	8192	-0.1603	0.1730	-0.1347	0.1456	0.0057	0.0927	0.0162	0.0942

Note: The polynomial approximation used under the heading "LPWN(R_y, R_w)" is (R_y, R_w).

Table 4: Simulation results for Model IV with $(\alpha_y, \beta_y) = (0, 0)$ and $(\alpha_w, \beta_w) = (0.8, 0)$

		LWN		LPWN(1,0)		LPWN(0,1)		LPWN(1,1)	
nsr	n	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
Panel A: $m = \lfloor (3/4)n^{0.8} \rfloor$									
5	2048	0.0233	0.0375	-0.0622	0.0897	-0.0017	0.0542	-0.0364	0.0707
	4096	0.0123	0.0251	-0.0595	0.0750	0.0027	0.0351	-0.0392	0.0570
	8192	0.0016	0.0161	-0.0569	0.0647	0.0031	0.0181	-0.0457	0.0574
10	2048	0.0384	0.0487	-0.1088	0.1276	-0.0157	0.0615	-0.0597	0.0781
	4096	0.0209	0.0304	-0.1074	0.1174	-0.0005	0.0428	-0.0710	0.0800
	8192	0.0032	0.0170	-0.1048	0.1101	0.0032	0.0179	-0.0829	0.0885
15	2048	0.0490	0.0577	-0.1364	0.1516	-0.0244	0.0694	-0.0720	0.0859
	4096	0.0265	0.0349	-0.1386	0.1466	-0.0024	0.0512	-0.0907	0.0974
	8192	0.0041	0.0172	-0.1376	0.1417	0.0036	0.0186	-0.1069	0.1106
Panel B: $m = \lfloor n^{0.8} \rfloor$									
5	2048	0.0361	0.0441	-0.0486	0.0748	-0.0058	0.0454	-0.0533	0.0859
	4096	0.0291	0.0347	-0.0500	0.0644	-0.0119	0.0339	-0.0480	0.0639
	8192	0.0206	0.0251	-0.0515	0.0595	-0.0141	0.0280	-0.0380	0.0455
10	2048	0.0623	0.0676	-0.0858	0.1039	-0.0161	0.0485	-0.0848	0.1093
	4096	0.0501	0.0538	-0.0916	0.1008	-0.0235	0.0403	-0.0601	0.0718
	8192	0.0356	0.0385	-0.0953	0.1001	-0.0284	0.0383	-0.0568	0.0607
15	2048	0.0806	0.0848	-0.1102	0.1250	-0.0228	0.0520	-0.0866	0.1118
	4096	0.0646	0.0676	-0.1178	0.1252	-0.0315	0.0458	-0.0580	0.0676
	8192	0.0463	0.0486	-0.1239	0.1277	-0.0386	0.0470	-0.0705	0.0736
Panel C: $m = \lfloor (5/4)n^{0.8} \rfloor$									
5	2048	0.0413	0.0472	-0.0322	0.0627	-0.0004	0.0415	-0.0448	0.0818
	4096	0.0369	0.0405	-0.0374	0.0529	-0.0088	0.0296	-0.0488	0.0671
	8192	0.0317	0.0342	-0.0416	0.0499	-0.0123	0.0232	-0.0493	0.0586
10	2048	0.0738	0.0774	-0.0633	0.0841	-0.0062	0.0413	-0.0904	0.1170
	4096	0.0661	0.0684	-0.0737	0.0833	-0.0139	0.0311	-0.0928	0.1057
	8192	0.0559	0.0574	-0.0819	0.0867	-0.0195	0.0282	-0.0719	0.0810
15	2048	0.0972	0.1000	-0.0844	0.1010	-0.0086	0.0415	-0.1180	0.1408
	4096	0.0869	0.0887	-0.0963	0.1038	-0.0163	0.0330	-0.1036	0.1198
	8192	0.0727	0.0740	-0.1073	0.1110	-0.0254	0.0330	-0.0598	0.0717

Note: The polynomial approximation used under the heading "LPWN(R_y, R_w)" is (R_y, R_w).

Table 5: Simulation results for Model V with $(\alpha_y, \beta_y) = (0.8, 0)$ and $(\alpha_w, \beta_w) = (-0.8, 0)$

		LWN		LPWN(1,0)		LPWN(0,1)		LPWN(1,1)	
nsr	n	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
Panel A: $m = \lfloor (3/4)n^{0.8} \rfloor$									
5	2048	0.1119	0.1574	0.0111	0.1297	0.0355	0.1450	0.0016	0.1437
	4096	0.0750	0.1049	-0.0088	0.0857	0.0043	0.0889	0.0001	0.0873
	8192	0.0509	0.0725	-0.0181	0.0646	-0.0088	0.0612	-0.0068	0.0631
10	2048	0.1217	0.1977	0.0226	0.1848	0.0688	0.1935	0.0085	0.1991
	4096	0.0771	0.1236	-0.0037	0.1267	0.0343	0.1378	0.0102	0.1317
	8192	0.0494	0.0818	-0.0163	0.0901	0.0016	0.0949	-0.0010	0.0868
15	2048	0.1334	0.2381	0.0314	0.2162	0.0654	0.2108	-0.0084	0.2241
	4096	0.0848	0.1470	0.0062	0.1513	0.0537	0.1520	0.0216	0.1569
	8192	0.0526	0.0938	-0.0070	0.1068	0.0380	0.1105	0.0244	0.1087
Panel B: $m = \lfloor n^{0.8} \rfloor$									
5	2048	0.1940	0.2187	0.0574	0.1711	0.0999	0.1952	-0.0103	0.1679
	4096	0.1364	0.1522	0.0111	0.0996	0.0456	0.1248	-0.0017	0.1018
	8192	0.0990	0.1098	-0.0038	0.0620	0.0107	0.0658	0.0058	0.0598
10	2048	0.2196	0.2670	0.0781	0.1965	0.1007	0.1947	-0.0061	0.1971
	4096	0.1474	0.1723	0.0540	0.1515	0.1036	0.1684	0.0169	0.1478
	8192	0.1018	0.1185	0.0134	0.1024	0.0898	0.1370	0.0293	0.1103
15	2048	0.2049	0.3152	0.0407	0.2103	0.0683	0.2007	-0.0126	0.2139
	4096	0.1637	0.2003	0.0601	0.1547	0.0834	0.1508	0.0117	0.1564
	8192	0.1095	0.1310	0.0386	0.1202	0.1003	0.1314	0.0494	0.1257
Panel C: $m = \lfloor (5/4)n^{0.8} \rfloor$									
5	2048	0.2858	0.3023	0.1089	0.2088	0.1421	0.2190	-0.0048	0.1535
	4096	0.2068	0.2163	0.0765	0.1676	0.1198	0.1914	-0.0051	0.1062
	8192	0.1524	0.1587	0.0153	0.0740	0.0340	0.0886	0.0012	0.0602
10	2048	0.2544	0.3636	-0.0191	0.1714	0.0519	0.1641	-0.0009	0.1658
	4096	0.2283	0.2446	0.0830	0.1563	0.1113	0.1644	0.0056	0.1314
	8192	0.1618	0.1715	0.0940	0.1521	0.1524	0.1786	0.0247	0.1157
15	2048	0.0766	0.4017	-0.1026	0.2031	0.0195	0.1870	-0.0091	0.1954
	4096	0.2324	0.2818	0.0021	0.1396	0.0390	0.1277	-0.0008	0.1324
	8192	0.1766	0.1894	0.0709	0.1221	0.0863	0.1266	0.0148	0.1061

Note: The polynomial approximation used under the heading "LPWN(R_y, R_w)" is (R_y, R_w).

Table 6: Local Whittle estimation of long memory in volatility of DJIA stocks

	$m = \lfloor (3/4)n^{0.8} \rfloor$			$m = \lfloor n^{0.8} \rfloor$			$m = \lfloor (5/4)n^{0.8} \rfloor$		
Ticker Symbol	LW	LPW	LWN	LW	LPW	LWN	LW	LPW	LWN
AA	0.1554	0.2313	0.6079	0.1379	0.1978	0.6063	0.1349	0.1735	0.5629
	(0.0195)	(0.0292)	(0.0355)	(0.0169)	(0.0253)	(0.0308)	(0.0151)	(0.0226)	(0.0285)
AIG	0.2372	0.3310	0.6216	0.2042	0.2990	0.6471	0.1830	0.2697	0.6603
	(0.0195)	(0.0292)	(0.0352)	(0.0169)	(0.0253)	(0.0299)	(0.0151)	(0.0226)	(0.0265)
AXP	0.2462	0.3411	0.6231	0.2115	0.3089	0.6514	0.1987	0.2747	0.6335
	(0.0195)	(0.0292)	(0.0351)	(0.0169)	(0.0253)	(0.0298)	(0.0151)	(0.0226)	(0.0270)
BA	0.1771	0.2115	0.5124	0.1509	0.2066	0.5336	0.1396	0.1900	0.5142
	(0.0195)	(0.0292)	(0.0385)	(0.0169)	(0.0253)	(0.0327)	(0.0151)	(0.0226)	(0.0298)
C	0.2458	0.3242	0.6138	0.2141	0.2992	0.6310	0.1889	0.2776	0.6557
	(0.0195)	(0.0292)	(0.0354)	(0.0169)	(0.0253)	(0.0302)	(0.0151)	(0.0226)	(0.0266)
CAT	0.1596	0.2229	0.5263	0.1280	0.2022	0.5788	0.1122	0.1781	0.5946
	(0.0195)	(0.0292)	(0.0380)	(0.0169)	(0.0253)	(0.0315)	(0.0151)	(0.0226)	(0.0278)
DD	0.0801	0.1211	0.4093	0.0810	0.0956	0.3367	0.0721	0.0967	0.3482
	(0.0195)	(0.0292)	(0.0433)	(0.0169)	(0.0253)	(0.0420)	(0.0151)	(0.0226)	(0.0368)
DIS	0.1853	0.2477	0.7505	0.1744	0.2134	0.6824	0.1576	0.2050	0.6705
	(0.0195)	(0.0292)	(0.0325)	(0.0169)	(0.0253)	(0.0292)	(0.0151)	(0.0226)	(0.0264)
GE	0.2131	0.2807	0.7272	0.1808	0.2580	0.7545	0.1661	0.2324	0.7463
	(0.0195)	(0.0292)	(0.0329)	(0.0169)	(0.0253)	(0.0281)	(0.0151)	(0.0226)	(0.0252)
GM	0.1803	0.1949	0.3965	0.1602	0.2027	0.4091	0.1358	0.1982	0.4614
	(0.0195)	(0.0292)	(0.0441)	(0.0169)	(0.0253)	(0.0375)	(0.0151)	(0.0226)	(0.0315)
HD	0.1958	0.2724	0.7417	0.1723	0.2432	0.7401	0.1490	0.2249	0.7647
	(0.0195)	(0.0292)	(0.0326)	(0.0169)	(0.0253)	(0.0283)	(0.0151)	(0.0226)	(0.0250)
HON	0.1951	0.2418	0.4460	0.1787	0.2253	0.4242	0.1681	0.2117	0.4073
	(0.0195)	(0.0292)	(0.0414)	(0.0169)	(0.0253)	(0.0368)	(0.0151)	(0.0226)	(0.0336)
HPQ	0.1951	0.2573	0.8454	0.1845	0.2290	0.7584	0.1557	0.2297	0.8045
	(0.0195)	(0.0292)	(0.0310)	(0.0169)	(0.0253)	(0.0280)	(0.0151)	(0.0226)	(0.0245)
IBM	0.2151	0.2915	0.6535	0.1931	0.2638	0.6359	0.1702	0.2464	0.6593
	(0.0195)	(0.0292)	(0.0344)	(0.0169)	(0.0253)	(0.0301)	(0.0151)	(0.0226)	(0.0265)
INTC	0.2106	0.2620	0.6798	0.1807	0.2533	0.6894	0.1541	0.2378	0.7289
	(0.0195)	(0.0292)	(0.0338)	(0.0169)	(0.0253)	(0.0291)	(0.0151)	(0.0226)	(0.0254)
JNJ	0.2182	0.2961	0.6415	0.1940	0.2641	0.6394	0.1600	0.2566	0.7019
	(0.0195)	(0.0292)	(0.0347)	(0.0169)	(0.0253)	(0.0301)	(0.0151)	(0.0226)	(0.0258)
JPM	0.2440	0.3166	0.6062	0.2173	0.2866	0.6058	0.1953	0.2684	0.6164
	(0.0195)	(0.0292)	(0.0356)	(0.0169)	(0.0253)	(0.0308)	(0.0151)	(0.0226)	(0.0273)
KO	0.2032	0.2889	0.8077	0.1833	0.2506	0.7923	0.1600	0.2361	0.8157
	(0.0195)	(0.0292)	(0.0316)	(0.0169)	(0.0253)	(0.0275)	(0.0151)	(0.0226)	(0.0243)
MCD	0.1317	0.2005	0.7193	0.1171	0.1701	0.7117	0.1151	0.1423	0.6718
	(0.0195)	(0.0292)	(0.0330)	(0.0169)	(0.0253)	(0.0287)	(0.0151)	(0.0226)	(0.0263)
MMM	0.1592	0.2235	0.9206	0.1431	0.1944	0.8713	0.1305	0.1770	0.8399
	(0.0195)	(0.0292)	(0.0301)	(0.0169)	(0.0253)	(0.0266)	(0.0151)	(0.0226)	(0.0241)
MO	0.2112	0.2738	0.5148	0.1879	0.2505	0.5163	0.1628	0.2402	0.5498
	(0.0195)	(0.0292)	(0.0384)	(0.0169)	(0.0253)	(0.0332)	(0.0151)	(0.0226)	(0.0288)
MRK	0.1731	0.2031	0.4632	0.1540	0.1931	0.4600	0.1220	0.2003	0.5379
	(0.0195)	(0.0292)	(0.0405)	(0.0169)	(0.0253)	(0.0352)	(0.0151)	(0.0226)	(0.0291)
MSFT	0.2305	0.3173	0.6135	0.2023	0.2883	0.6223	0.1778	0.2632	0.6516
	(0.0195)	(0.0292)	(0.0354)	(0.0169)	(0.0253)	(0.0304)	(0.0151)	(0.0226)	(0.0267)
PFE	0.1892	0.2733	0.6201	0.1644	0.2407	0.6325	0.1450	0.2203	0.6500
	(0.0195)	(0.0292)	(0.0352)	(0.0169)	(0.0253)	(0.0302)	(0.0151)	(0.0226)	(0.0267)
PG	0.2173	0.2391	0.5829	0.1945	0.2435	0.5525	0.1678	0.2397	0.5919
	(0.0195)	(0.0292)	(0.0362)	(0.0169)	(0.0253)	(0.0321)	(0.0151)	(0.0226)	(0.0278)
SBC	0.2088	0.2841	0.5856	0.1866	0.2582	0.5784	0.1683	0.2393	0.5839
	(0.0195)	(0.0292)	(0.0361)	(0.0169)	(0.0253)	(0.0315)	(0.0151)	(0.0226)	(0.0280)
UTX	0.1934	0.2646	0.6712	0.1650	0.2417	0.6893	0.1551	0.2121	0.6709
	(0.0195)	(0.0292)	(0.0340)	(0.0169)	(0.0253)	(0.0291)	(0.0151)	(0.0226)	(0.0263)
VZ	0.2174	0.2788	0.5923	0.1866	0.2642	0.6139	0.1681	0.2438	0.6188
	(0.0195)	(0.0292)	(0.0359)	(0.0169)	(0.0253)	(0.0306)	(0.0151)	(0.0226)	(0.0273)
WMT	0.2005	0.2793	0.7924	0.1668	0.2573	0.8247	0.1496	0.2301	0.8312
	(0.0195)	(0.0292)	(0.0318)	(0.0169)	(0.0253)	(0.0271)	(0.0151)	(0.0226)	(0.0242)
XOM	0.1817	0.2373	0.4535	0.1507	0.2254	0.5013	0.1408	0.1982	0.4863
	(0.0195)	(0.0292)	(0.0410)	(0.0169)	(0.0253)	(0.0337)	(0.0151)	(0.0226)	(0.0306)

Note: Asymptotic standard errors in parentheses.

Table 7: LPWN estimation of long memory in volatility of DJIA stocks

	$m = \lfloor (3/4)n^{0.8} \rfloor$			$m = \lfloor n^{0.8} \rfloor$			$m = \lfloor (5/4)n^{0.8} \rfloor$		
Ticker Symbol	$(1,0)$	$(0,1)$	$(1,1)$	$(1,0)$	$(0,1)$	$(1,1)$	$(1,0)$	$(0,1)$	$(1,1)$
AA	0.5704	0.5702	0.6120	0.6023	0.6029	0.4701	0.5629	0.6200	0.6191
	(0.0549)	(0.0757)	(0.1122)	(0.0464)	(0.0649)	(0.1030)	(0.0428)	(0.0578)	(0.0868)
AIG	0.5941	0.5974	0.6178	0.6056	0.6020	0.6092	0.6283	0.6214	0.6145
	(0.0539)	(0.0751)	(0.1120)	(0.0462)	(0.0649)	(0.0972)	(0.0407)	(0.0578)	(0.0869)
AXP	0.6015	0.6100	0.6028	0.6147	0.6077	0.5962	0.6017	0.6352	0.6099
	(0.0536)	(0.0748)	(0.1125)	(0.0459)	(0.0648)	(0.0976)	(0.0415)	(0.0576)	(0.0870)
BA	0.5081	0.6615	0.6654	0.5334	0.5796	0.6030	0.5157	0.5664	0.5641
	(0.0580)	(0.0739)	(0.1109)	(0.0491)	(0.0654)	(0.0974)	(0.0446)	(0.0588)	(0.0882)
C	0.6160	0.6593	0.6882	0.6284	0.6306	0.6317	0.6398	0.6178	0.6175
	(0.0530)	(0.0740)	(0.1104)	(0.0455)	(0.0645)	(0.0967)	(0.0404)	(0.0579)	(0.0868)
CAT	0.5010	0.5261	0.5255	0.5243	0.5167	0.4979	0.5520	0.5421	0.5402
	(0.0585)	(0.0770)	(0.1156)	(0.0495)	(0.0670)	(0.1014)	(0.0432)	(0.0593)	(0.0890)
DD	0.3581	0.3638	0.3510	0.3483	0.4387	0.4478	0.3666	0.3773	0.4330
	(0.0701)	(0.0863)	(0.1313)	(0.0617)	(0.0700)	(0.1044)	(0.0536)	(0.0659)	(0.0944)
DIS	0.7497	0.8328	0.8286	0.6789	0.8474	0.8383	0.6670	0.7881	0.7878
	(0.0488)	(0.0727)	(0.1091)	(0.0440)	(0.0630)	(0.0945)	(0.0396)	(0.0565)	(0.0847)
GE	0.7269	0.7742	0.8034	0.7550	0.7544	0.7556	0.7466	0.7609	0.7620
	(0.0494)	(0.0729)	(0.1092)	(0.0421)	(0.0632)	(0.0949)	(0.0378)	(0.0566)	(0.0849)
GM	0.3993	0.5542	0.2600	0.4240	0.4506	0.5365	0.4104	0.4114	0.4186
	(0.0659)	(0.0761)	(0.1505)	(0.0552)	(0.0695)	(0.0996)	(0.0503)	(0.0640)	(0.0954)
HD	0.7044	0.7534	0.5764	0.7393	0.7452	0.7040	0.7425	0.7358	0.7042
	(0.0500)	(0.0730)	(0.1134)	(0.0425)	(0.0633)	(0.0954)	(0.0379)	(0.0567)	(0.0854)
HON	0.4456	0.7722	0.8197	0.4315	0.5835	0.6591	0.4194	0.4981	0.5835
	(0.0621)	(0.0729)	(0.1091)	(0.0547)	(0.0653)	(0.0961)	(0.0497)	(0.0605)	(0.0877)
HPQ	0.8446	0.9218	0.9606	0.7540	0.9184	0.8839	0.8127	0.8219	0.8401
	(0.0466)	(0.0727)	(0.1093)	(0.0421)	(0.0630)	(0.0944)	(0.0366)	(0.0564)	(0.0845)
IBM	0.6526	0.6980	0.7179	0.6355	0.6783	0.6768	0.6541	0.6511	0.6488
	(0.0517)	(0.0735)	(0.1100)	(0.0453)	(0.0638)	(0.0958)	(0.0400)	(0.0574)	(0.0862)
INTC	0.6769	0.7809	0.8081	0.6873	0.7155	0.7400	0.6965	0.6900	0.6864
	(0.0509)	(0.0729)	(0.1092)	(0.0437)	(0.0635)	(0.0950)	(0.0389)	(0.0570)	(0.0856)
JNJ	0.6441	0.6649	0.6981	0.6415	0.6689	0.6678	0.6268	0.6086	0.6098
	(0.0520)	(0.0739)	(0.1103)	(0.0451)	(0.0640)	(0.0960)	(0.0407)	(0.0580)	(0.0870)
JPM	0.6090	0.6591	0.6590	0.6084	0.6465	0.6438	0.6201	0.6273	0.6294
	(0.0533)	(0.0740)	(0.1110)	(0.0461)	(0.0642)	(0.0964)	(0.0409)	(0.0577)	(0.0866)
KO	0.7990	0.8041	0.7929	0.7924	0.8299	0.8048	0.7974	0.7922	0.8160
	(0.0476)	(0.0728)	(0.1093)	(0.0413)	(0.0630)	(0.0946)	(0.0369)	(0.0564)	(0.0846)
MCD	0.4935	0.6956	0.4919	0.4920	0.7148	0.4943	0.4878	0.7481	0.5001
	(0.0589)	(0.0735)	(0.1175)	(0.0511)	(0.0635)	(0.1016)	(0.0459)	(0.0566)	(0.0906)
MMM	0.9199	0.9893	0.9879	0.8702	0.9808	0.9611	0.8384	0.9869	0.9645
	(0.0451)	(0.0729)	(0.1094)	(0.0399)	(0.0632)	(0.0947)	(0.0362)	(0.0566)	(0.0847)
MO	0.5207	0.5703	0.5868	0.5226	0.5476	0.5499	0.5060	0.5055	0.5119
	(0.0573)	(0.0757)	(0.1130)	(0.0496)	(0.0661)	(0.0991)	(0.0451)	(0.0602)	(0.0901)
MRK	0.4637	0.5836	0.5988	0.4644	0.5420	0.5481	0.4575	0.4327	0.4280
	(0.0608)	(0.0754)	(0.1126)	(0.0526)	(0.0663)	(0.0992)	(0.0474)	(0.0629)	(0.0947)
MSFT	0.6095	0.6327	0.6694	0.6104	0.6090	0.6060	0.6276	0.6152	0.6078
	(0.0533)	(0.0744)	(0.1108)	(0.0461)	(0.0648)	(0.0973)	(0.0407)	(0.0579)	(0.0870)
PFE	0.6152	0.6012	0.6320	0.6217	0.6153	0.6137	0.6128	0.6058	0.6056
	(0.0530)	(0.0750)	(0.1116)	(0.0457)	(0.0647)	(0.0971)	(0.0411)	(0.0580)	(0.0871)
PG	0.5740	0.8113	0.7787	0.5447	0.6809	0.7760	0.5856	0.6035	0.6219
	(0.0547)	(0.0728)	(0.1093)	(0.0486)	(0.0638)	(0.0947)	(0.0420)	(0.0581)	(0.0867)
SBC	0.5828	0.6089	0.6023	0.5821	0.5864	0.5876	0.5830	0.5752	0.5941
	(0.0544)	(0.0748)	(0.1125)	(0.0471)	(0.0652)	(0.0978)	(0.0421)	(0.0586)	(0.0874)
UTX	0.6734	0.6872	0.6863	0.6687	0.6683	0.6758	0.6713	0.6955	0.6158
	(0.0510)	(0.0736)	(0.1105)	(0.0443)	(0.0640)	(0.0958)	(0.0395)	(0.0570)	(0.0869)
VZ	0.5935	0.7149	0.7093	0.6126	0.6158	0.6764	0.5984	0.6036	0.6046
	(0.0539)	(0.0733)	(0.1101)	(0.0460)	(0.0647)	(0.0958)	(0.0416)	(0.0581)	(0.0871)
WMT	0.7926	0.8095	0.7760	0.7892	0.7729	0.7972	0.8020	0.7882	0.7890
	(0.0477)	(0.0728)	(0.1094)	(0.0414)	(0.0631)	(0.0946)	(0.0368)	(0.0565)	(0.0847)
XOM	0.4575	0.4979	0.5176	0.4436	0.4410	0.4567	0.4690	0.4843	0.4768
	(0.0612)	(0.0780)	(0.1160)	(0.0539)	(0.0699)	(0.1038)	(0.0468)	(0.0609)	(0.0918)

Note: The heading "(R_y, R_w)" indicates the LPWN(R_y, R_w) estimator. Asymptotic standard errors in parentheses.

Table 8: Parametric Whittle estimation of long memory in volatility of DJIA stocks

Ticker Symbol	$\hat{d}$	$\hat{\alpha}_y$	$\hat{\beta}_y$	$\hat{\sigma}^2_\varepsilon$	$\hat{\alpha}_w$	$\hat{\beta}_w$	$\hat{\sigma}^2_\eta$
AA	0.5777	−0.5431	–	0.1634	0.0362	–	3.3266
	(0.0943)	(0.1864)		(0.1122)	(0.0206)		(0.1360)
AIG	0.6377	−0.7946	–	0.2771	−0.8357	0.9185	2.8112
	(0.0748)	(0.1447)		(0.1629)	(0.0925)	(0.3044)	(0.7084)
AXP	0.5915	–	−0.7416	1.9245	0.2959	–	1.2799
	(0.0695)		(0.0863)	(0.2775)	(0.1167)		(0.3486)
BA	0.5532	−0.9275	0.6585	0.1152	–	0.0584	3.2629
	(0.1019)	(0.2558)	(0.2779)	(0.0844)		(0.0215)	(0.1125)
C	0.6201	–	–	0.0913	–	–	3.1170
	(0.0671)			(0.0482)			(0.0866)
CAT	0.4444	–	–	0.2211	0.6814	−0.7236	3.2299
	(0.1982)			(0.4432)	(0.1389)	(0.1823)	(0.5543)
DD	0.2478	−0.7859	0.8867	0.7682	–	–	4.1923
	(0.0978)	(0.3492)	(0.3360)	(0.8560)			(0.7411)
DIS	0.7555	–	–	0.0125	–	0.0604	3.3340
	(0.1331)			(0.0161)		(0.0160)	(0.0758)
GE	0.7509	0.6409	−0.8718	0.1393	–	–	3.1386
	(0.1177)	(0.2936)	(0.0901)	(0.1774)			(0.1740)
GM	0.5040	−0.9730	0.9371	0.1624	0.6010	−0.5739	3.2918
	(0.1438)	(0.0440)	(0.1043)	(0.2038)	(0.2992)	(0.2931)	(0.2419)
HD	0.6211	0.4356	−0.8670	1.5042	–	–	1.8557
	(0.1200)	(0.0676)	(0.0453)	(0.5819)			(0.5603)
HON	0.4166	−0.3490	–	0.6309	−0.6202	0.6475	2.7333
	(0.0672)	(0.3161)		(0.4581)	(0.3682)	(0.3093)	(0.4812)
HPQ	0.9298	−0.9010	–	0.0085	0.6844	−0.6567	3.3724
	(0.1684)	(0.1339)		(0.0133)	(0.1604)	(0.1633)	(0.0762)
IBM	0.6775	−0.6652	–	0.1063	0.0325	–	3.1645
	(0.0978)	(0.1972)		(0.0743)	(0.0196)		(0.1037)
INTC	0.7168	−0.8816	–	0.0712	−0.9341	0.9686	3.0632
	(0.0865)	(0.1020)		(0.0557)	(0.0319)	(0.0836)	(0.3243)
JNJ	0.5824	0.3924	−0.7923	0.9809	–	–	2.4001
	(0.1038)	(0.0799)	(0.0630)	(0.7210)			(0.6923)
JPM	0.5798	–	–	0.1461	–	–	3.1921
	(0.0624)			(0.0688)			(0.1005)
KO	0.8234	−0.7461	–	0.0249	–	0.0423	3.2834
	(0.1214)	(0.2712)		(0.0267)		(0.0166)	(0.0802)
MCD	0.6211	0.5379	−0.8949	0.8105	–	–	2.6296
	(0.1290)	(0.1005)	(0.0414)	(0.4203)			(0.4025)
MMM	0.7032	–	–	0.0128	–	–	3.5654
	(0.1748)			(0.0236)			(0.0871)
MO	0.5410	0.6217	–	0.0185	–	0.0349	3.2044
	(0.0760)	(0.3652)		(0.0339)		(0.0194)	(0.0876)
MRK	0.4903	–	–	0.1430	–	–	3.2380
	(0.0764)			(0.0873)			(0.1142)
MSFT	0.5987	−0.7656	–	0.2990	−0.8105	0.8940	2.8675
	(0.0725)	(0.1451)		(0.1846)	(0.0887)	(0.2491)	(0.6008)
PFE	0.6093	–	–	0.0777	–	–	3.3014
	(0.0850)			(0.0428)			(0.0872)
PG	0.5724	–	–	0.0901	–	–	3.1889
	(0.0741)			(0.0565)			(0.0942)
SBC	0.5518	–	–	0.1294	–	–	3.2578
	(0.0657)			(0.0681)			(0.1018)
UTX	0.6159	0.5142	−0.8598	0.8469	–	–	2.5056
	(0.1041)	(0.0943)	(0.0475)	(0.4559)			(0.4314)
VZ	0.6778	−0.5928	–	0.1039	–	0.0714	3.2410
	(0.1022)	(0.3156)		(0.0747)		(0.0240)	(0.1074)
WMT	0.8328	–	–	0.0078	–	0.0363	3.4027
	(0.1229)			(0.0087)		(0.0154)	(0.0733)
XOM	0.4962	–	–	0.1532	–	–	3.2217
	(0.0698)			(0.0839)			(0.1109)

Note: Asymptotic standard errors (evaluated as the inverse of the negative Hessian) in parentheses.